Distribution Agreement

In presenting this thesis or dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I hereby grant Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis or dissertation in whole or in part in all forms of media, now or hereafter known, including display on the world wide web. I understand that I may select some access restrictions as part of the online submission of this thesis or dissertation. I retain all ownership rights to the copyright of the thesis or dissertation. I also retain the right to use in future works (such as articles or books) all or part of this thesis or dissertation.

Signature:

_____ _____

Kendra Ann Sirak Date

A Genomic Analysis of Two Early Christian Cemetery Communities from Sudanese Nubia

By

Kendra Ann Sirak
Doctor of Philosophy
Anthropology

_____

Jessica C. Thompson, Ph.D.
Advisor

_____

Dennis P. Van Gerven, Ph.D.
Advisor

_____

John Lindo, Ph.D.
Committee Member

_____

Craig Hadley, Ph.D.
Committee Member

_____

Ron Pinhasi, Ph.D.
Committee Member

Accepted:

_____

Lisa A. Tedesco, Ph.D.
Dean of the James T. Laney School of Graduate Studies

_____

Date

A Genomic Analysis of Two Early Christian Cemetery Communities from Sudanese Nubia

By

Kendra Ann Sirak
B.A., Northwestern University, 2011
M.A., Emory University, 2014

Advisors: Jessica C. Thompson, Ph.D. and Dennis P. Van Gerven, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Anthropology
2018

Abstract

A Genomic Analysis of Two Early Christian Cemetery Communities from Sudanese Nubia

By Kendra Ann Sirak

The Early Christian Period (550–800 C.E.) site of Kulubnarti, located in Sudanese Nubia, is hypothesized to have been home to two distinct subpopulations, with the landowning members of the "R community" inhabiting the west bank of the Nile River, and the more impoverished members of the semi-itinerant "S community" occupying an adjacent Nilotic island. While bioarchaeological analyses of stress-induced lesions and life expectancy consistently found that members of the S community were exposed to more stress, experienced more ill-health, and died younger than members of the R community, biodistance studies based on morphology as a proxy for genetic variation suggested that the two communities were a single, biologically-indistinguishable population. However, without direct genetic analysis, the relationship between the two communities and their biogeographic origins and affinities to present-day populations have remained poorly resolved. This dissertation uses ancient DNA to address these questions from a high-resolution genomic perspective for the first time.

By analyzing genome-wide Single Nucleotide Polymorphisms (SNPs) in 28 individuals from Kulubnarti, this dissertation 1) Explores the genetic relationship between the S and R communities by looking for evidence of community-based genetic population substructure through patterns of clustering and estimation of genetic distance; and 2) Characterizes the genetic composition of individuals from both communities and reveals any community-based differences by exploring biogeographic genetic affinities and components of ancestry and determining the mitochondrial and Y chromosome haplogroup profiles of each community.

While analyses of autosomal SNPs identify no community-based genetic population substructure at Kulubnarti and suggest that from a genomic perspective, the two communities were a single population with the most affinity to present-day populations from Northeastern Africa, analyses of uniparental markers reveal that genetic signatures of Eurasian-associated patrilineal ancestry were overrepresented in the S community and that genetic signatures of Eurasian-associated matrilineal ancestry were overrepresented in the R community. By suggesting that the genetic signatures of sex-specific gene flow were reflected at Kulubnarti in a community-based way, this dissertation reveals a previously unknown pattern of genetic variation and provides a model for interpreting patterns of genetic variation within the context of broader sociocultural dynamics.

A Genomic Analysis of Two Early Christian Cemetery Communities from Sudanese Nubia

By

Kendra Ann Sirak
B.A., Northwestern University, 2011
M.A., Emory University, 2014

Advisors: Jessica C. Thompson, Ph.D. and Dennis P. Van Gerven, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Anthropology
2018

Acknowledgements

First and foremost, I thank George Armelagos, under whose mentorship I began this journey. When I emailed George to tell him that I decided to come to Emory as a graduate student in fall 2012, he sent me back an email that said "Fantastic news. You will have the honor? of being my last and best student." Initially, I thought he had made a typo. I now realize he intentionally put the question mark after the word honor. There is no more appropriate time to answer his question: Yes, George- it was an honor.

George exemplified what it meant to be "good people." He respected my thoughts and opinions, encouraged and supported my endeavors, and never failed to save a seat for me at his dinner table. As a mentor, George was encouraging and empowering- he made me believe that I had the potential do great work. During my first years of graduate school, he took my unrestrained curiosity and focused it on this project. To be able to continue his work on the ancient Nubians has been nothing short of a privilege.

While George inspired my interest in the Nubians, this project would never have come to fruition without Dennis Van Gerven, who has spent the last four years serving as my mentor, a task which he accepted without question. Dennis immediately stepped into George's shoes, both professionally and personally. I am grateful beyond words for the constant guidance he has given me, delivered in the form of both glowing praise as well as constructive criticism. In addition to making me feel proud, confident, and capable, Dennis pushed me to become a critically-thinking scholar. Thank you, Dennis, for having unwavering faith in my capabilities, but also for never failing to point out when you thought I could do better. So much of my academic development and personal growth over these last years is due to you. I am so lucky to have been your student.

I am also extremely grateful to have worked with Jessica Thompson, who stepped into a role as my advisor at Emory and who has impacted my development as a scholar so greatly. Jess has been much more than an advisor and mentor to me- she has been a role model, as both an extremely competent, dedicated, and passionate scholar as well as a driven and successful woman in science. Additional thanks go to John Lindo and Craig Hadley, members of my dissertation committee, for their thoughtful feedback and guidance throughout this dissertation process. I also thank the National Science Foundation (NSF BCS-1613577) for providing the funding that made this work possible.

This dissertation would not have been possible without Ron Pinhasi, who took me into his lab in 2013, allowing me to draw upon both his expertise and his resources to conduct much of the lab work associated with this project. So much of my continuously-growing competency in ancient DNA analysis is thanks to Ron, who has enabled me to work closely with his own group over the last five years and has expanded my horizons and enhanced my professional development. I also thank Jens Carlsson, for allowing me to use (and sometimes monopolize) Archie, and Catherine Moss, for assisting with all things sequencing.

I am indebted to David Reich and the members of his lab, who have provided an incredible amount of assistance and support that enabled the analyses included in this dissertation. In addition to David, who generously dedicated his time and resources to helping with this project, thanks go to Shop Mallick, Iñigo Olalde, Nadin Rohland, Pontus Skoglund, Nasreen Broomandkhoshbacht, and Kristin Stewardson, who completed lab work and offered invaluable feedback.

I've had the luck of being a member of an incredibly talented and wonderful cohort at Emory. Adeem, Amanda, Andrea, Bisan, Hanne, JoAnn, Paul, Sydney, thank you for sharing this experience with me. I'm very grateful to have made this epic journey through graduate school alongside all of you. I'm particularly grateful to Paul for the friendship that we developed and the guidance and love that he has shown to me as I figured out my twenties, as well as Hanne, who has opened my mind to new perspectives and reminded me that sometimes, you just need to dance.

To Éilis, Eppie, and Victoria- in addition to having perfected the art of the prosecco-fueled girls' night, you provided so much love, friendship, and more than occasional bioinformatics advice. I am so thankful that you are in my life. To Bea, Mario, and Olivia- I am so grateful to have had the opportunity to work with you at UCD. I couldn't have asked for a better group of colleagues, who have become some of my closest friends.

To say that this dissertation would not have been completed without having met Daniel Fernandes is an understatement. Daniel, you so graciously assisted me in so many aspects of this work. You patiently taught me bioinformatics (specifically demonstrating the PCA process no less than 900 times), you selflessly shared with me the scripts that took you months of work, and you became my go-to person for every question about every aspect of every analysis. You are not only my colleague and travel buddy, but one of my closest friends.

Marnie, Megan, Olivia, and Sarah- you saw this process from day one through to the end, and you have been there for me at every step of the way (and have documented it through blackmail-style pictures). Knowing that I have you in my corner has helped me

more than I can explain. Thanks for being my best friends, who offer constant love and support and who are always happy to share a Bud Light.

Emily, Gina, Kelsie, Lauren, and Sam- you enhanced my graduate school experience in ways that I never imagined and that are largely unrepeatable in front of polite company. You've been there through the dark days and the glorious days and everything in between, and I am so thankful for our friendship and for the laughter and love we've shared. Thanks for being my best friends and lifelines.

Gratitude beyond measure goes to my parents who have encouraged my every endeavor. I cannot begin to express how thankful I am for your unwavering love and support. I know you have felt every failure and every triumph just as viscerally as I have, and I could not have asked for better guides on this journey of life. Thank you, thank you, thank you for reading every paper I've ever written, for being my go-to phone call at any hour of any day, for always supporting and never pressuring, and for constantly reminding me that life is so much more important than this document will ever be. I could not have done this without you.

Chris, the love, support, and patience you have shown me throughout this process continues to amaze me on a daily basis. You have been my guiding light toward this finish line, and I am so grateful to live both successes and failures by your side. You inspire me, encourage me, and enrich every part of my life every single day. How lucky I am to be on this journey with you.

Table of Contents

List of Figures and Tables

CHAPTER 6: SAMPLE PREPARATION AND SCREENING

CHAPTER 7: GENETIC RELATIONSHIPS AT KULUBNARTI

CHAPTER 8: GENETIC COMPOSITION OF THE KULUBNARTI NUBIANS

SUPPLEMENT 1: SAMPLE PROCESSING DATA FROM NINETY-NINE SCREENED KULUBNARTI SAMPLES

SUPPLEMENT 2: NEXT-GENERATION SQUENCING DATA FROM NINETY-NINE SCREENED KULUBNARTI SAMPLES

SUPPLEMENT 3: MAPDAMAGE PLOTS FROM NINETY-NINE SCREENED KULUBNARTI SAMPLES

SUPPLEMENT 4: SEXING PLOTS FROM NINETY-NINE SCREENED KULUBNARTI SAMPLES

SUPPLEMENT 5: PCA

SUPPLEMENT 6: PAIRWISE F_{ST} ESTIMATES

SUPPLEMENT 7: ADMIXTURE ANALYSIS

CHAPTER 1
INTRODUCTION

"Poor as it was and is, however, the Nile Valley between Aswan and Khartoum offered, for millennia on end, the only dependable way across the great desert barrier of the Sahara... To the extent that the cultures, the products, and, in time, the bloodstreams of black and white men met and mingled, they met and mingled here." (Adams 1977: 19)

1.1 Dissertation Overview

This dissertation presents a genomic analysis of the inhabitants of the Early Christian Period (550–800 C.E.) site of Kulubnarti in Sudanese Nubia. Kulubnarti is hypothesized to have been home to two socially-distinct subpopulations in Early Christian times: landowning members of the "R community" occupied the west bank of the Nile River, while the more impoverished and semi-itinerant "S community" inhabited an adjacent Nilotic island. Until now, investigations of the biological relationship between and identities of the two Kulubnarti communities have relied on the analysis of morphology as a proxy for genetic variation, leaving questions of whether the S and R communities were genetically-distinct subpopulations with different biogeographic genetic affinities or ancestries unresolved. By drawing directly upon the ultimate source of the human genome, this dissertation reveals the genetic relationship between the Kulubnarti communities and characterizes the genetic composition of the Kulubnarti Nubians for the first time.

Ancient DNA (aDNA) was extracted from skeletal remains and analyzed using targeted capture and Next-Generation Sequencing (NGS) of genome-wide Single Nucleotide Polymorphisms (SNPs) to pursue two aims: 1) Explore the genetic

relationship between the S and R communities by looking for evidence of community-based genetic population substructure through patterns of clustering and estimation of genetic distance; and 2) Characterize the genetic composition of individuals from both communities and reveal any community-based differences by exploring biogeographic genetic affinities and components of ancestry and determining the mitochondrial DNA (mtDNA) and Y chromosome haplogroup profiles of each community. Through the analysis of high-resolution genetic data, this dissertation is able to retest long-standing hypotheses about the residents of Early Christian Kulubnarti.

1.2 Background

1.2.1 Nubia

Nubia is a region along the Nile River corridor that stretches from the First Cataract at Aswan in Egypt through the Sixth Cataract of the Nile near Khartoum, the present-day capital of the Republic of the Sudan.[1] As shown in Figure 1.1, Nubia is conventionally divided into two parts: Lower Nubia extends from the First to the Second Cataract, while Upper Nubia extends south from the Second Cataract.

[1] While the northern limit of Nubia is defined as the First Cataract itself, its southern boundary has shifted based on the spread of Egyptian influence over time (Adams 1977). While the southern boundary is currently considered by many to be around the Fourth Cataract (Lobban 2003), it extended as far as Khartoum in the Christian era (Adams 1977) and is defined using that boundary in this dissertation. Cataracts are areas of swift Nile rapids.

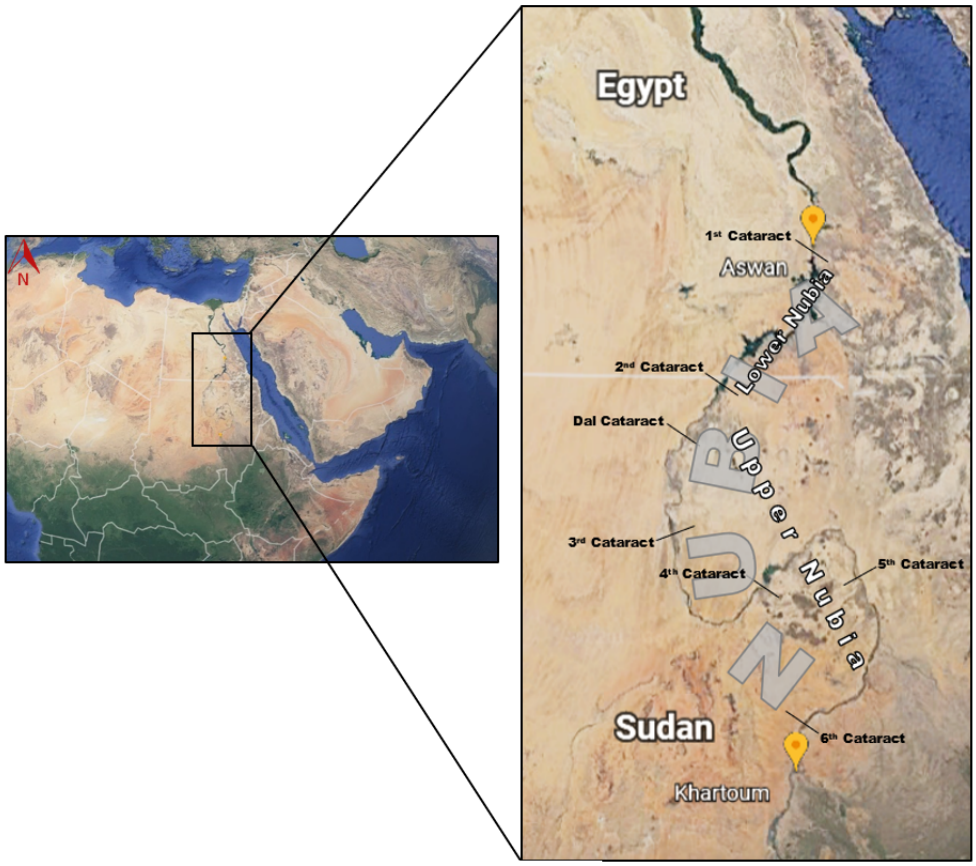


Figure 1.1: Map showing the location and extent of Nubia (Lower and Upper), with the present-day cities of Aswan and Khartoum labeled on the map on the right. The black slashes represent the six numbered cataracts and the Dal Cataract. Maps adapted from Google Earth (version 9.2.47.8).

Nubia has been intermittently occupied since the Paleolithic (Arkell 1961; Adams 1977) and has long been viewed as a key zone of interaction between indigenous African populations, African kingdoms, and international empires (Adams 1977; Edwards 2004). Positioned on the only dependable thoroughfare through the Sahara Desert for many millennia, the Nubian corridor was a primary zone of contact and exchange of ideas, goods, and genes between sub-Saharan Africa, the Horn of Africa, Egypt, the Near East, and the Mediterranean (Adams 1967; Edwards 2007). As such, it has been referred to as a "land between" (Van Gerven 1981: 395) and has been described as having the "…unique position as the meeting ground of two worlds" (Adams, 1977: 19), a place "…which has

never been permanently colonized by foreign peoples but has always been influenced by them" (Adams, 1977: 20). This idea of a "land between" is reflected in the historic characterization of Nubian peoples, both biologically, as "an old, stable blend of African Negro and Mediterranean Caucasian elements, in which the two strains are about equally represented" (Adams, 1977: 46), and culturally, as a group of riverain farmers who, over time, absorbed the Islamic traditions of their northern neighbors while continuing to speak a unique dialect belonging to the Sudanic family of languages.

1.2.2 Kulubnarti

This dissertation focuses specifically on the inhabitants of the small Nubian hamlet of Kulubnarti, located approximately 130 kilometers south of the Sudanese border with Egypt between the Second and Dal Cataracts (Figure 1.2).

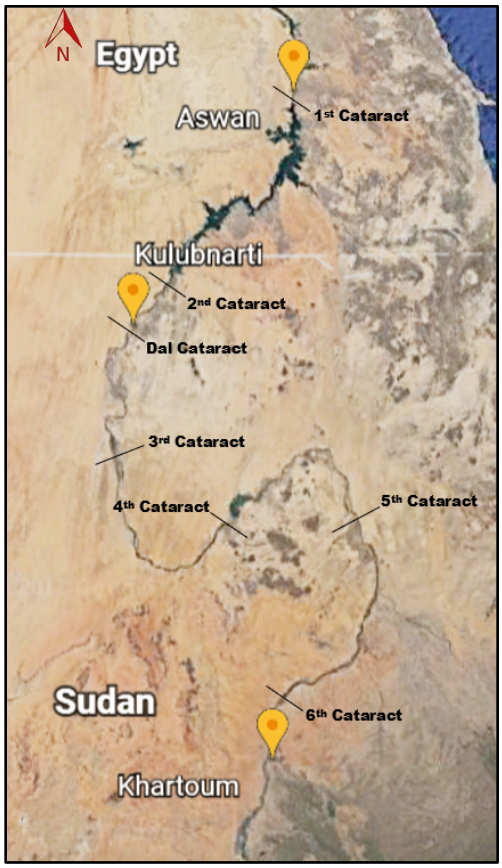


Figure 1.2: Location of Kulubnarti within Nubia. Map adapted from Google Earth (version 9.2.47.8).

Kulubnarti is situated in the heart of a particularly desolate physiographic subdivision of Upper Nubia known as the *Batn el Hajar* ("belly of rock").[2] Described as the "…most barren and forbidding of all Nubian environments" (Adams 1977: 26), the *Batn el Hajar* is dominated by bare granite ridges and gullies, small and scarce patches of arable land, and a dangerously swift and narrow Nile, making subsistence and survival a daily challenge (Adams et al. 1999).

In 1969, an initial survey of Kulubnarti determined that occupation of the hamlet began around the Early Christian Period (approximately 500 C.E.), and a preliminary excavation identified at least ten habitation sites, two churches, and two neighboring cemeteries (Adams and Adams 1998). In 1979, the Colorado-Kentucky Expedition disinterred 406 Early Christian Period individuals from Site 21-S-46 (or the "S cemetery"), located on a Nilotic island, and Site 21-R-2 (or the "R cemetery"), located on the adjacent west bank (Figure 1.3).

[2] Occasionally written "*Batn el Hagar.*"

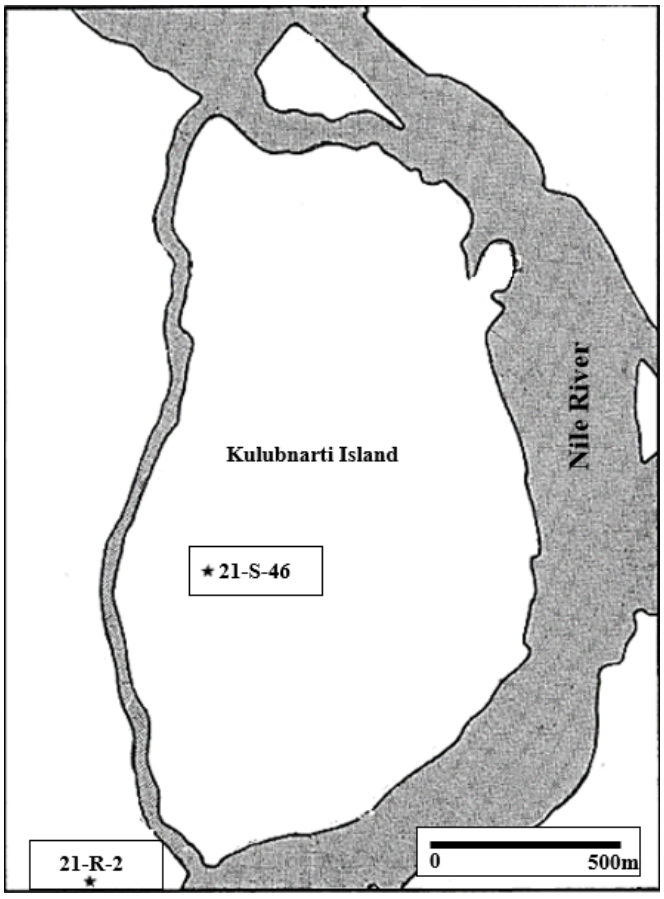


Figure 1.3: Map of cemetery locations at Kulubnarti (adapted from Adams et al. 1999).

Over the past four decades, the Kulubnarti skeletal collection has become the subject of a research program that examined stress-induced lesions (Van Gerven et al. 1981, 1995), subadult growth patterns (Hummert 1983; Hummert and Van Gerven 1983), isotopic signatures of dietary trends (Turner et al. 2007; Sandberg 2012; Basha et al. 2018), and mean life expectancy (Van Gerven et al. 1995). It was these bioarchaeological analyses, which consistently identified inter-cemetery variation in markers of stress, patterns of growth, and mortality rates, that initially generated the hypothesis of two

socially-distinct communities with disparate experiences of health and mortality buried in separate cemeteries at Kulubnarti.[3]

Specifically, bioarchaeological analyses determined that the individuals of the S community, those individuals interred in the S cemetery, were exposed to more stress, experienced more ill-health, and died younger than those of the R community, those individuals buried in the R cemetery (Adams et al. 1999). Archaeological analysis of burial textiles likewise determined that the R community had a higher percentage of more prosperous people than the S community (Adams et al. 1999).

While social stratification is a common feature of complex landowning societies (Davis and Moore 1945; Hayden 2001), the social division detected at Kulubnarti is unique in two ways; first, because of its extent and the severity of the resulting biological consequences, and second, because such stratification has not been detected in any other ancient Nubian population (Adams et al. 1999); however, similar social stratification in present-day Nubian society suggests that the S community may have been a subset of the population known as the "underclass" (Adams and Adams 2006: 11).

Present-day Nubians are mostly farmers with small plots of land living in villages along the Nile River. Among them live groups of impoverished, landless, semi-nomadic persons who act as sharecroppers or seasonal laborers for landowning families, otherwise living off small flocks of sheep and goats; these people are known locally as the Nubian underclass (Adams and Adams 2006). The well-evidenced disparity between the S and R communities at Kulubnarti suggests that such an underclass may have existed in Early

[3] Consistent with previous research on the Kulubnarti Nubians and on the basis of strong data, this dissertation uses the term "community" instead of "cemetery population" to represent those individuals buried in each Kulubnarti cemetery.

Christian times as well, and that the S community was a group of itinerant and disadvantaged individuals who provided labor for the members of the relatively more prosperous R community. However, it has also been noted that the presence of a semi-nomadic, landless underclass in Early Christian Nubian society is a "wholly unexpected possibility" for which there is "neither textual evidence nor archaeological evidence from other sites to support such an interpretation" (Adams et al. 1999: 88). The ambiguity surrounding the relationship between and biogeographic origins and genetic affinities of the S and R communities inspires this dissertation's investigation of the Kulubnarti Nubians from a genomic perspective.

1.3 Research Questions and Rationale

Building upon previous bioarchaeological research, the most salient question about the Kulubnarti Nubians regards the unknown genetic relationship between the S and R communities: **based on direct genetic evidence, were these socially-distinct communities also genetically-distinct subpopulations?** The ability to directly analyze genomic variation in the Kulubnarti Nubians also inspires deeper questions about their biogeographic affinities and potential genetic similarities and differences: **with which present-day populations do the Kulubnarti Nubians share the greatest genetic affinity? Which ancestral components are identified in the Kulubnarti Nubians and how does this fit within the present-day landscape of human genetic diversity? Did members of the S and R communities have similar matrilineal and patrilineal origins?**

Since the beginning of the Nubian research program, hypotheses about the population history of the ancient Nubians have been based on the analysis of morphology, interpreted within a framework of evolving theoretical perspectives. Initially, morphological characteristics were studied from a racial determinist approach that ascribed meaning to "races" and "racial mixtures" (e.g., Elliot Smith 1909); subsequently, they were studied from a racial typological paradigm that classified people into discrete "racial" groups and emphasized "racial" migrations (e.g., Burnor and Harris 1967; Strouhal 1971); finally, a biocultural approach emphasized the role of functional morphology and clinal (i.e., continuous) variation in phenotypic expression (e.g., Greene 1972; Carlson 1976; Van Gerven 1982). Most morphological analyses suggest long-term population continuity throughout Nubia (Adams 1967; Greene 1972; Carlson and Van Gerven 1977; Galland et al. 2016), but also recognize the influence of gene flow along the Nile, which resulted in the admixture seen in present-day populations of Northeastern Africa (Passarino et al. 1998; Hassan et al. 2008; Yousif and Eltayeb 2009; Pagani et al. 2012).

Studies of biological distance ("biodistance") have used morphology as a proxy to specifically investigate the relationship between the Kulubnarti communities as well as between the Kulubnarti Nubians and populations to the north at Wadi Halfa.[4] Craniometric studies suggested a close biological relationship between the S and R communities and found that all Kulubnarti Nubians were also closely related to their contemporaries to the north who inhabited the site of Wadi Halfa near the Second Cataract of the Nile River (Van Gerven 1982; Vollner 2016). Analysis of discrete dental

[4] These biodistance studies are described in greater detail in Chapter 3.

traits corroborated this finding, supporting a close relationship between the Kulubnarti communities as well as biological stability and continuity that extended from Kulubnarti to Wadi Halfa (Greene 1982).

While the most parsimonious assumption is that the Kulubnarti Nubians were direct descendants of autochthonous populations that had inhabited Nubia at least intermittently for thousands of years (Arkell 1961; Adams 1977), interactions between Nubian populations and foreign peoples traveling along the Nile during the Early Christian Period are historically and archaeologically well-evidenced, for example, by the introduction of Christianity into Nubia from Constantinople in 542 C.E. (Kirwan 1959; Adams 1965; Vantini 1975; Säve-Söderbergh 1987) and the Arab invasions that began in the 7th century C.E. and increased in intensity over the next 700 years (Adams 1977; Shinnie 1996; Hollfelder et al. 2017). While these interactions inevitably introduced genetic diversity into ancient Nubia on a broad level, the isolation of the specific site of Kulubnarti within the *Batn el Hajar,* demonstrated in part by the paucity of goods from Egypt and the Near East (Adams et al. 1999), confounds parsimonious expectations surrounding the presence of foreign DNA in the gene pool of the Kulubnarti Nubians.

Though morphological evidence has suggested that the population of Kulubnarti was composed of two closely-related, biologically-indistinguishable communities that were also phenotypically similar to neighboring Nubian populations, all conclusions must be tempered by the recognition that morphology is an indirect proxy for genetic variation. Therefore, morphological expression is not a direct reflection of genetic composition, but is also subject to influence by epigenetic, environmental, and stochastic factors that can confound interpretations of population relationships and affinities (Nystrom 2006;

Stojanowski and Shilliaci 2006; von Cramon-Taubadel and Weaver 2009). Direct analysis of ancient DNA has the potential to provide a more finely-resolved understanding of the genetic relationship between the S and R communities and to offer new insight into the origins, affinities, and interactions of Kulubnarti's two communities with foreign populations as well as with each other.

A more thorough understanding of population relationships at Kulubnarti revealed through the analysis of ancient DNA will help connect the past with the present in Nubia. For example, the structure of present-day Nubian society is loosely caste-based along ethnic lines, with a small elite group of Egyptian, Syrian, and Lebanese descendants at the top, and descendants of sub-Saharan African slaves at the bottom (Adams et al. 1977). This inspires another important question that will be explored from a genomic perspective: **does the ethnically-based social structure that characterizes Nubian society today reflect a similar social division at Early Christian Kulubnarti?**

1.4 Kulubnarti from a Genomic Perspective

DNA-based methods represent a new line of inquiry for directly investigating the Kulubnarti Nubians. By providing a time-specific snapshot of a previously-unknown genetic landscape, the analysis of ancient DNA can effectively resolve the relationship between the S and R communities and elucidate their affinities with present-day peoples inhabiting geographic areas with evidence of historic migration along the Nile Corridor, including the Mediterranean, the Near East, Egypt, and the Horn of Africa.

Ancient DNA has become an increasingly important tool for studying human prehistory by revealing the genetic makeup and relationships of past populations living at

archaeologically-known times and places (Pickrell and Reich 2014). Recent technological developments have facilitated the economically-efficient production of massive amounts of genetic data, particularly using highly-degraded ancient DNA from skeletal samples. These developments, including NGS (Margulies et al. 2005; Mardis 2008; Kircher and Kelso 2010; Knapp and Hofreiter 2010; Rizzi et al. 2012) and targeted enrichment and capture (Mamanova et al. 2010; Maricic et al. 2010; Carpenter et al. 2013; Ávila-Arcos et al. 2015, Mathieson et al. 2015), have opened the door for hypothesis-driven analyses using large sample sizes.

In particular, the analysis of SNPs, variations in single nucleotide bases that occur across the genome, has become widely utilized in anthropological research, including this dissertation (Rasmussen et al. 2010; Haak et al. 2015; Mathieson et al. 2015). This is because the examination of a large number of genome-wide SNPs enables the quantitative analysis of population relationships, while the geographically-patterned distribution of certain SNPs (Hinds et al. 2005; Novembre et al. 2008) allows their use as powerful tools for inferring population history and estimating individual and population-level biogeographic genetic affinities and ancestries (Dupré 2008; Shriver and Kittles 2008).

Addressing the research questions presented in the previous subsection of this chapter demands the generation of genetic data from individuals from each of the Kulubnarti communities. Before these data can be generated, the best-preserved samples must be identified in order to mitigate some of the practical challenges stemming from the degraded nature of ancient DNA. These challenges are particularly prevalent in the analysis of the Kulubnarti skeletal material, because while all DNA retrieved from

ancient skeletal material is degraded due to the termination of the molecular repair mechanisms that function during an organism's life (Lindahl 1993), a hot and open depositional environment, such as that found at Kulubnarti, speeds up the destruction of DNA and decreases the preservation of DNA within the skeletal material (Smith et al. 2001, 2003; Collins et al. 2002; Bollongino et al. 2008). To overcome the added challenge of generating adequate amounts of aDNA data using skeletal material exposed to extreme Nubian temperatures, this dissertation draws upon optimized methodologies and technological innovations, including NGS and targeted enrichment and capture.

1.5 Dissertation Outline

The next three chapters of this dissertation provide the necessary background for interpreting the genetic data presented in this dissertation. Chapter 2 traces the study of the Nubian past in the context of shifting theoretical perspectives in archaeology and explains how the biocultural approach has provided the framework for the exploration of Kulubnarti from a genomic perspective. Chapter 3 situates the site of Kulubnarti in space and time. It reviews the body of bioarchaeological research conducted within the biocultural approach that has elucidated disparate patterns of health, disease, and mortality at Kulubnarti and discusses what is known about the nature of the stratification between the S and R communities. Chapter 4 provides the requisite background for understanding the genetic data generated in this dissertation. It provides an overview of DNA basics, introduces the field of ancient DNA, discusses the integration of ancient DNA and anthropology, and places this dissertation within current research trends in molecular anthropology.

Chapters 5 and 6 discuss the research strategy and methods required for this dissertation to move from "bones to bytes" (i.e., from skeletal samples to molecular data) and present the preliminary screening results that identified the best-preserved samples from Kulubnarti for this dissertation's analysis. Chapter 5 introduces the skeletal sample analyzed in this dissertation and discusses the use of the osseous labyrinth as the optimal biological substrate for aDNA analysis. It revisits the two research aims that guide this dissertation, presents the hypothesis and expectations associated with each aim, and discusses the limitations encountered by this study. Chapter 6 presents the methods used to generate the data necessary to conduct the population genetic analyses that test the hypotheses proposed in Chapter 5. It describes the screening process that identified the best-preserved samples and presents the results of this screening step. It then describes the methods used to perform the targeted capture of genome-wide SNPs for hypothesis testing and culminates by presenting the results of this capture.

The next two chapters provide and interpret the results of population genetic analysis of the selected Kulubnarti Nubian samples. Chapter 7 discusses the analysis of the genetic relationship between the S and R communities at Kulubnarti in association with Aim 1. Chapter 8 discusses the investigation of the genetic composition of the Kulubnarti Nubians in terms of biogeographic genetic affinity and ancestry components, and mtDNA and Y chromosome haplogroup in association with Aim 2. Both chapters present the analytical methods required to achieve their respective aims, discuss the application of these methods to the Kulubnarti data, present the results of these analyses, and explain how the ensuing results are or are not consistent with the expected findings presented in Chapter 5.

Chapter 9 concludes this dissertation by exploring the meaning of the newly generated genetic data from a biocultural perspective. It integrates these new data into the bioculturally-framed anthropological understanding of the Kulubnarti Nubians and reveals how the understanding of Kulubnarti population history has expanded and evolved through the analysis of ancient DNA. It revisits the two aims that guided this dissertation and answers the questions posed in this chapter from a genomic perspective. It culminates with a discussion of pathways for future research and the potential for the integration of anthropology and genetics to enhance and transform our understanding of past populations.

CHAPTER 2

THEORETICAL PERSPECTIVES IN NUBIAN ARCHAEOLOGY

"Over the past half century, the biocultural approach has emerged as an integrative intellectual force in biological anthropology…it provides a unique avenue for synthetic research that extends across the subdisciplines." (Zuckerman and Armelagos 2011: 15)

2.1 Chapter Overview

This chapter reviews the history of archaeology in Nubia and places the expeditions that occurred during the 20th century, one of which recovered the skeletal collection analyzed in this dissertation, within the context of broader theoretical perspectives in anthropology. Beginning with the First Archaeological Survey of Nubia in 1907 and extending through the United Nations Educational, Scientific, and Cultural Organization (UNESCO) International Campaign to Save the Monuments of Nubia, which concluded in 1980, the study of the Nubian past has been shaped by evolving perspectives (used interchangeably with "approaches") that dictated the goals of research and influenced the interpretation of results. The evolution of these perspectives over the last century has set the stage for a genomic investigation of the Kulubnarti Nubians.

This chapter begins by reviewing the First (1907–1911) and Second (1929–1934) Archaeological Surveys of Nubia. The perspective adopted by both Surveys interpreted Nubian history as a series of disconnected episodes dominated by different "racial types"; however, while racial determinism was emphasized in the First Survey, it was absent from the Second. This chapter then moves to a discussion of the UNESCO-funded expeditions that occurred after 1960. Importantly, the UNESCO expeditions adopted a new, progressive theoretical perspective, the biocultural approach, that interpreted the

Nubian past as a "continuous story, encompassing the development of a single people from ancient times to the present, and having an autonomous dynamic of its own" (Adams 1998: 3).

The biocultural approach, which emphasizes the study of human remains within a framework that integrates biological, cultural, environmental, and ecological information, remains the dominant theoretical perspective in anthropology today. This chapter describes the application of the biocultural approach to the study of ancient Nubia on a broad level before focusing on how the specific site of Kulubnarti was investigated from this perspective.

As demonstrated by the early expeditions that studied the ancient Nubians, there was a tendency to conflate population origins, relationships, and affinities with cultural achievement as well as with the concept of "race." Therefore, this chapter concludes with a discussion about how this dissertation's analysis and interpretation of genetic data are conducted from a biocultural approach that uncouples biology and culture but also acknowledges their interplay in shaping the population history of Kulubnarti.

2.2 Introduction

From an archaeological point of view, Nubia is among the best-known regions in Africa. This is the result of 20th century salvage archaeological expeditions that attempted to save the culture history of Nubia ahead of the construction and enlargement of dams along the Nile River that extended cultivatable areas for a growing Egyptian population. While these dam-expansion events served as the impetus for the archaeological expeditions, they also resulted in the flooding, and ultimately the destruction, of Lower

Nubia (Emery 1965; Adams 1967). As such, all archaeology discussed in this chapter is salvage archaeology (Armelagos et al. 1968).

Before the archaeological and bioarchaeological evidence collected during these salvage expeditions can be used to construct a foundation for new genomic investigations of the ancient Nubians, it must be recognized that these expeditions were influenced by theoretical perspectives that dictated the goals of research and impacted the interpretation of results. Throughout the 20th century, the theoretical perspectives that shaped archaeology evolved alongside changing political circumstances, methodological advances, growing professionalization in archaeology, and changing philosophical visions of non-Western cultures (Adams 1981). Various experts on Nubian history present differing opinions on the theoretical perspectives that have specifically shaped Nubian archaeology. For example, Adams (1981) discusses 1) the extractive colonial approach, characterized by a total lack of standards in excavation and a lack of respect for the peoples and countries that were exploited; 2) the enlightened colonial approach, characterized by a greater commitment to responsible archaeology and a desire to protect the surviving antiquities of Nubia; 3) the post-colonial approach, characterized by methodological advances that allowed the reconstruction of culture and culture history and encouraged an enhanced sense of accountability toward the Sudan; and 4) the independent national approach, in which Sudanese archaeologists began to rewrite their own history to suit the needs of emergent nationalism and cultural integration.

Alternatively, Trigger (1994) compares a colonial approach, which reflected the interest of European archaeologists and facilitated interpretations of archaeological data that rationalized the European colonialization of Africa, to a post-colonial approach that

was more objective and considerate of the inhabitants of sub-Saharan Africa. Overall, both Adams and Trigger recognize a trend that moved the interpretive framework from a "Eurocentric" colonial anthropology to a more progressive framework that appreciated the unique and dynamic nature of Nubian culture and emphasized a commitment to proper anthropological stewardship.

This dissertation views the last century of Nubian archaeology as conducted from three primary theoretical perspectives, each associated with a specific dam-building or enlarging event. First, the racial determinist perspective was associated with the First Archaeological Survey of Nubia (1907–1911), an expedition that occurred in response to the construction of the first Aswan Dam ("Aswan Dam" in Figure 2.1) between 1898–1902 and its subsequent heightening in 1908–1910, which created an artificial reservoir that flooded 225 kilometers of Nubia extending south from the First Cataract at Aswan. Second, the racial typological perspective was associated with the Second Archaeological Survey of Nubia (1929–1934), which occurred in response to a further enlargement of the first Aswan Dam during these same years, extending the reservoir over 360 kilometers to Wadi Halfa near the Second Cataract (Emery 1965; Edwards 2004). Third, the biocultural approach was associated with the UNESCO International Campaign to Save the Monuments of Nubia that began in the 1960s in response to the construction of a new dam, the High Aswan Dam ("Sadd al-ʿĀlī" in Figure 2.1), which began four miles upstream from the original dam (Emery 1965; Adams 1977; Abu-Zeid and El-Shibini 1997) and resulted in the permanent flooding of over 500 kilometers of Nubia and the creation of Lake Nasser (Shinnie 1996).[5] These three perspectives and their influence on

[5] Lake Nasser is the Egyptian name for this body of water; its preferred Sudanese name is Lake Nubia.

an evolving understanding of the Nubian past are discussed in the remainder of this chapter.

Figure 2.1: Map showing the location of the city of Aswan, the original Aswan Dam, the High Aswan Dam (Sadd al-ʿĀlī), and the permanently flooded area of Nubia (Lake Nasser) that extends to Wadi Halfa. Map adapted from Google Earth (version 9.2.47.9).

2.3 The First Archaeological Survey of Nubia and the Racial Determinist Perspective

Under the direction of George A. Reisner and later Cecil M. Firth, the First Archaeological Survey of Nubia (1907–1911) explored 150 kilometers of the Nile Valley, beginning at the First Cataract and extending south (Adams 1977). Because "Egyptocentrism" was pervasive, the First Survey was conducted as an extension of

Egyptological research on Egypt's Nubian frontier (Adams 1981; Edwards 2007). As such, Nubia was explored primarily as a "theater of ancient Egyptian exploitation and colonization rather than as a region that was of interest in its own right" (Trigger 1994: 330).

The First Survey had two distinct aims: 1) identify and compare racial "types"; and 2) connect morphological differences to "racial" migrations, assuming that cultural changes evidenced the arrival of new peoples in line with the ideas of racial (or biological) determinism (Adams 1968; Adams 1977). Emphasis was not only placed on studying the "successive races and racial mixtures" who inhabited ancient Nubia (originally written by Reisner, quoted in Emery 1965: 39), but to ascribe value to these "races" and "racial mixtures."

Focusing nearly exclusively on the recovery of mortuary remains (Adams 1977), the First Survey sought out evidence of migration by "Caucasoids" from the north and "Negroids" from the south (e.g. Elliot Smith 1910; Firth 1927), depending on the skull as the primary source of analysis. A heavy emphasis was placed on a small number of characteristics, such as the cephalic index, with the goal of differentiating and, most importantly, ranking the living "varieties" of man (Adams 1977). Sir Grafton Elliot Smith, an anatomist, undertook the analysis and description of Nubian skeletal remains from the First Survey. Elliot Smith maintained that all major cultural changes and developments were attributed to the arrival of superior "Caucasoid" peoples, while "dark ages" were attributed to an influx of indigenous ("Negroid") elements (discussed in Van Gerven et al. 1973; Adams 1981; Adams 1998; Ross 2013). He reported that "…the smallest infusion of Negro-blood immediately manifests itself in a dulling of initiative

and a 'drag' on the further development of the arts of civilization" (Elliot Smith 1909: 25). Nubia came to be regarded as a backward variant of Egyptian civilization (Adams 1998) where the "Negroid" Nubians were not attributed the same capacity for creativity as the "Caucasoid" Egyptians (Adams 1998).

Theories of biological determinism and migrationism led to the representation of Nubian history as a series of disconnected episodes alternating between dark ages and cultural fluorescence that resulted from the influx of "Negroids" and "Caucasoids" into the area, respectively (Derry 1909; Elliot Smith 1910; Firth 1927; Batrawi 1935, 1946; Morant 1935). For example, during the First Archaeological Survey, Reisner branded Nubian cultural stages with no Egyptian counterparts as "A," "B," "C," and "X" Periods, considering these periods to be "dark ages" in Nubian history where Egyptian presence (and therefore, culture) was absent from the region (Adams 1977). Reisner specifically described the B-Group as a "hybrid of Negro and Egyptian stocks" whose decline was "due to a fusion between the lower strata…with Negro slaves" (Firth 1912: 12).

The focus on racial type and the determinist perspective that characterized the First Archaeological Survey of Nubia reflected, in part, the sociocultural attitude of the early decades of the 20th century.[6] As such, the initial understanding of Nubian population history conflated cultural achievement and biological type. While a dominant focus on "race" and "type" extended well into the 20th century, the entanglement of cultural achievement and biological traits was reexamined during the Second Archaeological Survey of Nubia.

[6] Adams (1977) raises the point that the racist point of view shared by nearly all early students of Nubian history condemns the age more than the men themselves.

2.4 The Second Archaeological Survey of Nubia and the Racial Typological Perspective

The enlargement of the first Aswan dam due to rising water levels expanded the reservoir that began around the First Cataract at Aswan well into present-day Sudan. This enlargement provided the impetus for the Second Archaeological Survey of Nubia (1929–1934), led by Walter B. Emery and Laurence P. Kirwan. The Second Survey expanded survey coverage from Egypt's southern border into the Sudanese frontier. Consistent with the First Survey, the Second Survey focused heavily on the recovery of mortuary remains, almost to the exclusion of other archeological evidence (Adams 1977).

The initial picture of Nubian cultural and political history was primarily formed by the First Survey. The Second Survey aimed to contribute additional detail to the existing understanding of ancient Nubia (Adams 1967). While a dedication to the recovery of skeletal remains was consistent between the First and Second Surveys, a personnel change took place prior to the analysis of the skeletal remains collected during the Second Survey. Specifically, medical doctor Ahmed Batrawi was tasked with analyzing the newly recovered skeletal remains, taking over the role of Sir Grafton Elliot Smith.

Though he initially endorsed the prevailing racial determinist approach, Batrawi eventually questioned the empirical foundation of Elliot Smith's determinism and was ultimately responsible for the change in perspective associated with the Second Archaeological Survey of Nubia (Van Gerven et al. 1973). Specifically, Batrawi questioned the ongoing "failure to distinguish clearly between the achievements of populations and their inherent biological characteristics" (Batrawi 1946: 131). He

described the danger of assessing biological relationships from cultural evidence instead of physical characteristics and conducted his analysis of the Nubian skeletal remains by evaluating coefficients of variation and coefficients of racial likeness to estimate racial affinities.

However, though progressive in his denouncement of racial determinism, Batrawi did not reject the concept of "race," which was maintained through the Second Survey and into the mid-20th century. The central features of the second theoretical perspective were a continued orientation toward "racial" typological definition and description, a dependence upon gene flow as an explanatory model (i.e., migrationism), and a commitment to the objective reality of "race" (Van Gerven et al. 1973). As such, the Second Survey is described as having adopted a "racial typological" approach, a small step forward from the "racial deterministic" approach that characterized the previous survey.

Ultimately, while the Second Survey saw the untangling of biological type and cultural achievement, the analysis of human remains still consisted of defining "racial" differences while ignoring the question of whether "race" had any validity as a type-concept (Adams et al. 1978). For example, the analysis of "racial stocks" comprising a northern "type" in Middle Egypt and southern "type" in Upper Egypt, distinguished by variation in cranial index, nasal index, and prognathism, is described, as is the hybridization of these "stocks" (Batrawi 1946: 154).

After the completion of the Second Survey in 1934, there was almost no further fieldwork for nearly a generation. This meant that the picture of Nubian history that was accepted well into the mid-20th century was created entirely before 1935 (Adams 1967).

It was not until the call for the "New Physical Anthropology" (Washburn 1951) that a new generation of researchers came to scrutinize the typological approach, the existence of "racial" traits, and the idea of "racial" superiority in the second half of the 20th century.[7]

2.5 The UNESCO Campaign and the Biocultural Approach

The critical reaction to the racial typological perspective in anthropology associated with the "New Physical Anthropology" (Washburn 1951) emphasized the need for a move from the description of biological "type" to the hypothesis-driven testing of evolution and variation in populations across space and time. This movement inspired the establishment of the biocultural perspective, which provided the theoretical framework for the UNESCO International Campaign to Save the Monuments of Nubia (Carlson and Van Gerven 1979), conducted in response to the impending destruction of Nubia caused by the construction of a new High Aswan Dam, which started in 1947.

Unlike the reservoir created by the first Aswan Dam, which was emptied during part of each year, the High Aswan Dam would create a permanent lake, meaning anything covered by its waters would be lost (Adams 1977). More than forty expeditions were consequently planned and conducted between the Egyptian border and the head of the proposed reservoir. These expeditions discovered over 1,000 archaeological sites and excavated nearly one-third of them.[8]

---

[7] Unfortunately, the outdated idea of racial typology has maintained relevance to some researchers even after the undermining of its empirical foundation (e.g., Emery 1965; Billy 1977)

[8] See Adams (1977) for a listing of the archaeological expeditions to Nubia during this campaign.

The UNESCO International Campaign to Save the Monuments of Nubia was carried out by a group of scientists who sought to understand their data in a way that demonstrated a marked retreat from the examination of biological "types" in favor of more plausible theories of cultural continuity and *in situ* evolution (Van Gerven et al. 1973). Their biocultural approach viewed "the variation in human populations as dependent upon the action of environmental agencies both mediated by and resulting from patterns of cultural adaptation" (Van Gerven et al. 1973: 560).[9]

New expeditions, including the University of Colorado Nubian Expedition (1962–1964) led by Gordon Hewes, the University of Kentucky Expedition to Nubia (1969) directed by W.Y. Adams, and the Joint Colorado-Kentucky Nubian Expedition (1979) led by Dennis Van Gerven, created a contextualized and comprehensive picture of the Nubian past by systematically excavating entire village assemblages instead of focusing on mortuary remains (Van Gerven 1982). These expeditions emphasized the exploration of ecology and settlement patterns in addition to cemeteries, which satisfied the biocultural tenet of examining all possible aspects of an archaeological culture in order to understand the significance of any one part of it (Trigger 1994; Chapman 2013). The data produced by these expeditions can be characterized by two interrelated biocultural themes: 1) a new approach to migrationism; and 2) a view of Nubian history as a continuum of biocultural evolution.

Within the biocultural perspective, Nubian culture change was viewed as a result of the resourcefulness and creativity of the local Nubian population (Trigger 1994). Instead of speculating large-scale migration events, this approach regarded Nubian

[9] Also see Armelagos (1969) for one of the prominent early studies of the ancient Nubians that took place within the biocultural approach.

history as a continuum of biocultural evolution of a single group that spanned from the Neolithic to the Islamic period (Adams 1966; Carlson 1976; Carlson and Van Gerven 1979; Rudney and Greene 1982). While migration was recognized and appreciated, it was not viewed as a main driver of culture change. For example, Adams (1977: 5–6) writes: "Migrations there have undoubtedly been, but they have been for the most part migrations within Nubia: rearrangements of peoples all drawing on a common reservoir of indigenous tradition and experience."

Conclusions drawn from the racial determinist and racial typological perspectives of the First and Second Surveys were re-examined from the biocultural perspective. For example, during the First Archaeological Survey, Reisner proposed that invasion by "alien Negroids" from the south caused a change in the "racial constitution" and replaced the thriving Meroitic culture (350 B.C.E.–350 C.E.) with a more "primitive" X-Group culture (350–550 C.E.) (Firth 1927; Kirwan 1937; Emery 1965; Strouhal 1971). Re-examination of this idea from a biocultural perspective suggested that the individuals of the X-Group were more likely to be the direct descendants of the Meroites (Adams 1967; Adams 1970; Calcagno 1986). Using early childhood stress as a multifactorial biocultural system engendered by synergistic interaction between nutrition and infectious disease as a lens into the Meroitic–X-Group transition, Armelagos (1968), and later Rudney and Greene (1982), determined that the X-Group population was well-adapted to the disease ecology experienced by the Meroites, and therefore most likely represented a continuation of the previous Meroitic population. The only change between these periods affecting cultural development was a decentralization of political power that increased village autonomy (Adams 1966; Adams 1977).

Within the biocultural approach, skeletal morphology was still analyzed, but the analytical aim shifted from pure description to placement of morphological differences within an evolutionary framework (e.g. Carlson 1976; Carlson and Van Gerven 1977; Small 1981; Van Gerven 1982) where morphological changes were attributed to changing biocultural patterns (Greene et al. 1967; Greene 1972; Carlson 1976; Carlson and Van Gerven 1979; Small 1981; Greene 1982; Rudney and Greene 1982; Martin et al. 1984; Brace et al. 1993; Prowse and Lowell 1995; Starling and Stock 2007).

An example of the shift in analytical aim can be seen in the interpretation of diachronic change in the cranial vault and face observed between Mesolithic Nubians and more recent Nubian peoples. While some researchers (e.g., Irish 2005) continued to dichotomize discrete dental traits to conclude that genetic discontinuity, in the form of population replacement or swamping of an indigenous gene pool, occurred in Nubia, others pointed to the pattern of dental reduction and reorientation of facial and cranial morphology as a result of the reduction in masticatory stress associated with the dietary change that accompanied the transition to Neolithic subsistence patterns (Carlson 1976; Carlson and Van Gerven 1977; Small 1981; Martin et al. 1984). Studies in which morphological change was equated with functional-behavioral alteration helped bioarchaeologists move fully into a new era that acknowledged an evolutionary model of biocultural adaptation (Galland et al. 2016).

2.6 Kulubnarti from a New Perspective

As part of the UNESCO Campaign, excavations at the site of Kulubnarti, which occurred in 1969 (excavation of the village of Kulubnarti and survey of the cemeteries)

and 1979 (excavation of the cemeteries), adopted a biocultural approach.[10] For example, when the village of Kulubnarti was excavated, a significant decline in the construction of churches throughout the Christian Period (~550–1400 C.E.) in favor of increasingly defensive settlements was noted (Adams 1994). This change in focus was interpreted as the response to growing pressures from Islamic invaders from the north in the *Batn el Hajar* and a Muslim king at Dongola in the later centuries of the Christian Period that led the local Nubian population to increase their fortifications against foreign invaders (Van Gerven et al. 1981, 1995).

Within the two Kulubnarti cemeteries, grave superstructures, grave orientations, grave shaft types, head and body coverings, shrouding, head and body positioning, and grave goods were investigated in addition to bioarchaeological analysis of the human remains, leading to the recognition that while the appearance of the graves was similar between the two cemeteries, the figures for pathology and mortality were significantly different (Van Gerven et al. 1981; Adams et al. 1999). The biocultural perspective inspired the interpretation of these differences in light of Nubian culture history (Van Gerven et al. 1981). Initially interpreting the S and R cemeteries as diachronic, Van Gerven et al. (1981) looked to a decline in centralized political authority and a reduction in cultural achievement that corresponded to the observed reduction in mortality between the S and R communities.[11]

[10] The aims and achievements of these two Kulubnarti expeditions are discussed in greater detail in Chapter 3.

[11] Until 1999 (Adams et al. 1999), the S cemetery originally was thought to be associated with the Early Christian Period, and the R cemetery was thought to be associated with the Early through Late Christian Periods. The recognition that the cemeteries were synchronic is further discussed in Chapter 3.

Specifically, it was assumed that political changes during the Classic and Late Christian Periods (850–1400 C.E.), which included increasing Muslim military pressure extending southward from Egypt, a growing spirit of "military feudalism" throughout Nubia, and the ascension of a Muslim prince to the Nubian throne, led to a "state of disarray" in Nubia (Adams 1977: 509). These political changes were interpreted as having led to increased autonomy for the Kulubnarti Nubians, which resulted in a decreased amount of stress as evidenced by childhood survival and health (Van Gerven et al. 1981, 1995). Though radiocarbon dating later revealed the cemeteries to be synchronic, effectively invalidating this biocultural interpretation of diachronic change, this example serves as a model for the interpretation of bioarchaeological data within a broader ecological, political, and cultural context.

Though the establishment of the biocultural approach was a progressive step away from outdated perspectives that relied on the biologically-unfounded analysis of "race" or "type," it had a notable consistency with the approaches that came before it: the racial determinist, racial typological, and biocultural approaches all relied on the analysis of morphology and phenotype. This dissertation represents the next shift in perspective by focusing on the direct analysis of the human genome.

2.6.1 Moving Toward a Genomic Perspective

Though its measurements were interpreted from a range of different perspectives, the skull served as the primary source of analysis for anthropologists throughout the 20th and into the 21st century. Because the source of data remained the same, the range of questions that could be asked remained limited.

While maintaining the biocultural framework that emphasizes the interpretation of human biological data within an evolutionary point of view, this dissertation analyzes a novel source of data – the human genome – with a new set of high-resolution methodological techniques. The application of new techniques to a novel source of data allows new questions to be asked. Ultimately, this dissertation demonstrates how research aims, proposed hypotheses, analytical methods, and an interpretive framework can remain cohesive with the biocultural approach while simultaneously taking a step toward exploring longstanding questions of Nubian population history from a genomic perspective.

2.6.2 A Genomic Analysis of Kulubnarti

Consistent with the biocultural approach, this dissertation recognizes that a complex interaction between biology and culture shaped the genomes of the Kulubnarti Nubians. For example, patterns of genetic diversity observed at Kulubnarti are likely to reflect an interaction between biological (e.g., admixture between foreign peoples and an autochthonous Nubian population), social (e.g., separation into socially-distinct communities), cultural (e.g., selection of a partner from within or outside of one's own community), and environmental (e.g., an extreme paucity of available arable land that forced people into close proximity) influences.

To effectively answer both new and long-standing questions regarding Kulubnarti population history and provide high-resolution insight into the genetic identity of the Kulubnarti Nubians, this dissertation interprets the generated data in a way that avoids unfounded causal connections between genetic composition and social status. This involves downplaying migrationism and ignoring the arbitrary classification of people

into discrete populations or groups, and accentuating instead the complexity of population interactions, the influence of the environment (e.g., migration corridors), and the way that social structure can influence biology (e.g., health and well-being) without being caused by biology itself (e.g., genetic variation).

Ultimately, the biocultural perspective provides the framework for quantifying, characterizing, and exploring genetic variation to more clearly elucidate the parallels between culture, in the form of social stratification, and biology, in the form of genetic variation, at Kulubnarti. This same theoretical perspective also inspires this dissertation's commitment to collectively examining, but simultaneously untangling, genetic variation and social status at Kulubnarti. Keeping this biocultural perspective in mind, focus now turns to the site of Kulubnarti.

CHAPTER 3

KULUBNARTI

"It was clear to me [Van Gerven] that Kulubnarti's location in the Batn el Hajar*, combined with mummified human remains, made it more than just another Christian site. It provided an opportunity to explore life and death in ancient Nubia in a setting never studied before. What was life like in the harshest of Nubian environments?" (Armelagos and Van Gerven 2017: 27)*

3.1 Chapter Overview

This chapter situates the site of Kulubnarti in space and time. It begins with a broad discussion of Nubia in the Early Christian Period, narrows in focus to the *Batn el Hajar* region*,* and finally concentrates specifically on Kulubnarti. It reviews the extensive body of bioarchaeological research that explored patterns of health, disease, and mortality at Kulubnarti from a biocultural approach, and discusses how this research has led to the interpretation of two socially-distinct communities at Kulubnarti.

This chapter specifically discusses the bioarchaeology of the impoverished subset of S community Nubians known as the underclass, described as a "floating population" who were likely to have been looked down upon as inferiors by their landowning R community neighbors and not allowed to live among them (Adams et al. 1999: 49). It reviews the biodistance studies that estimated the close biological relationship between these socially-distinct groups using morphology as a proxy for genetics, highlights the lines of evidence that point to the consistent health differences between the two communities, and provides the rationale for a direct genetic exploration of the population relationships and genetic composition of the Kulubnarti Nubians.

3.2 Nubia

As introduced in Chapter 1, the region known as Nubia (Figure 3.1) straddles the present-day boundaries of Egypt and the Republic of the Sudan, stretching from the First Cataract near Aswan in Egypt to the confluence of the White and Blue Nile Rivers near the modern Sudanese capital of Khartoum (Welsby 2002).[12]

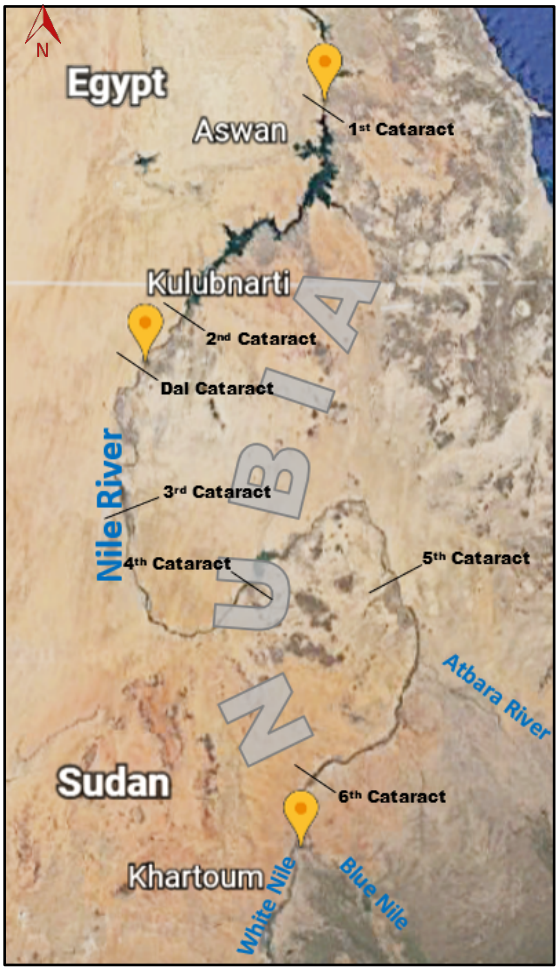


Figure 3.1: Physical map of Nubia, showing the location of Nubia (Lower and Upper) within the present borders of Egypt and the Republic of the Sudan, the location of the six numbered cataracts and the Dal Cataract, the tributaries of the Nile, and the focal site of Kulubnarti. Map adapted from Google Earth (version 9.2.47.8).

[12] As a result of the construction and enlargement of the Aswan dams, approximately 500 kilometers of Nubia from Aswan to the Dal Cataract are now submerged under Lake Nubia/Lake Nasser. In ancient times, Nubia was home to a distinct people who inhabited the arable land along the edge of the Nile. All descriptions of Nubia in this dissertation will focus on the region before the construction of the first Aswan dam in 1898 and all analyses will focus on the culturally-distinct Nubian people who inhabited this area.

The Nile River dominates the Nubian landscape. It is formed by two main tributaries, one of which originates in Lake Victoria (the White Nile), and the other which originates in Lake Tana in the Ethiopian highlands (the Blue Nile). These tributaries unite in the area around Khartoum. Another tributary, the Atbara River, originates in northwest Ethiopia and joins the Nile between the Fifth and Sixth Cataracts. The waters of the Nile are so important to life that habitable Nubian land only extends as far from the Nile as there is arable soil (less than 1 kilometer at some points). As such, since the end of the last glaciation (10,000–15,000 years ago), the inhabitants of Nubia have lived distributed in a one-dimensional pattern and interacted with foreign peoples and with each other along the banks of the Nile (Krings et al. 1999).

3.2.1 Physiographic Subdivisions of Nubia

Traditionally, Nubia is broken up into two unequal halves: Lower Nubia lies between the First Cataract at Aswan and the Second Cataract just south of the modern border between Egypt and the Republic of the Sudan, and Upper Nubia extends south from the Second Cataract into the present-day Republic of the Sudan. However, based on the varying geology and topographic relief of the region, Adams (1977: 21) finds it more valuable to speak "not of two Nubias but of five or six." This is because while Lower Nubia is a relatively homogenous zone of soft sandstone, Upper Nubia consists of well-differentiated physiographic subdivisions of different geological character divided by rocky riverine obstructions that cause areas of white water rapids, known as cataracts. The physiographic subdivisions of Nubia are shown in Figure 3.2.

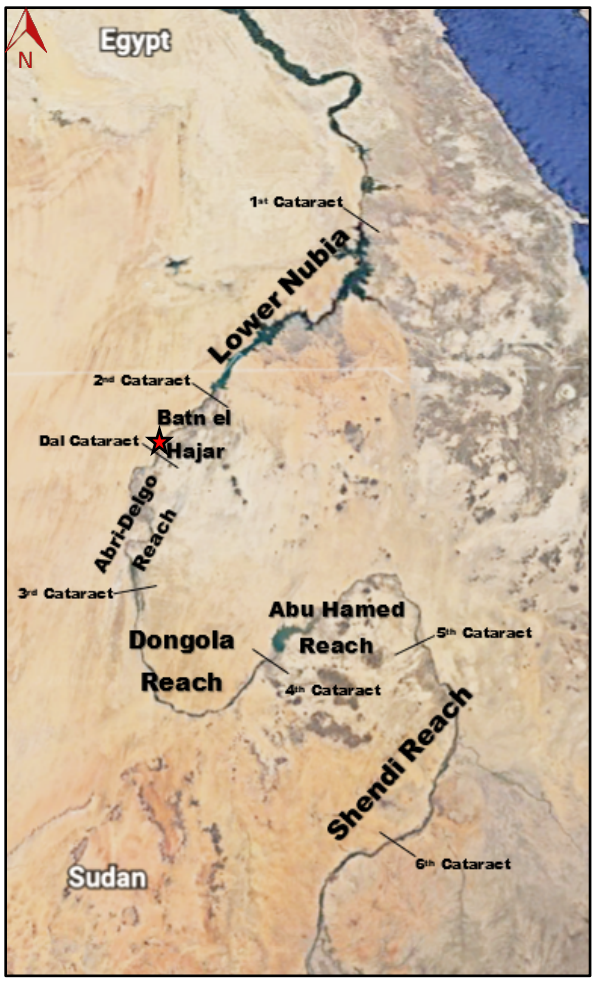


Figure 3.2: Physiographic subdivisions of Nubia showing the six subdivisions of Nubia between the six numbered cataracts and the Dal Cataract. The red star represents the location of Kulubnarti. Map based on Adams (1977), adapted from Google Earth (version 9.2.47.9).

In Nubia's cataract zones, the waters of the Nile are forced through narrow geographic faults and fissures, filling narrow *wadis* (typically dry channels) and creating rocky riverine islands surrounded by dangerous rapids (Säve-Söderbergh 1987). In addition to serving as geographic demarcations, these cataracts have played an important historic role in shaping social, political, and economic practices in Nubia, mainly by impeding prosperous agricultural production and serving as natural boundaries for the movement of people (Adams 1977).

Throughout antiquity, the First Cataract frequently served as a political boundary between Egypt and Nubia by acting as a barrier to river traffic. The Second Cataract gave way to the forbidding *Batn el Hajar* region, filled with smaller cataracts that served as a large natural boundary between Lower Nubia to the north and Upper Nubia to the south. While the Fourth, Fifth, and Sixth cataracts presented a challenge to boat travel, between them lay low gradient basins where Nubian sandstone had been eroded and alluvial sediments abounded; an abundance of arable land made regions such as the Dongola Reach particularly desirable areas for foreign interaction and settlement (Woodward 2007).

3.2.2 Corridor to Africa

For many millennia, the Nile represented the only means of contact across the Sahara Desert, earning Nubia a reputation as the "corridor to Africa" (Adams 1977). It was "a transition zone, or rather *the* transition zone" that linked sub-Saharan Eastern Africa with Egypt, the Near East, the Mediterranean, and greater Eurasia (Adams 1977: 20). Until the opening of the trans-Saharan caravan trade in the last millennium B.C.E., Nubia served as the only link between vastly different worlds, with the Nile facilitating the bidirectional movement of goods, ideas, and peoples along its axis (Adams 1977). Importantly, because of the cataracts that characterized the Nile as it traveled through Nubia, travelers, merchants, missionaries, and invaders often journeyed through Nubia by land alongside the Nile, increasing their opportunity for interaction with indigenous Nubian peoples (Shinnie 1996).[13] Throughout Nubia, local individuals lived, traded, and

[13] Though there is also some evidence for travel by boat along the Nile, Adams (1977) describes the rapids of the *Batn el Hajar* in particular as requiring the transfer of riverine cargoes to donkeys for overland portage while the vessels were laboriously dragged around the rapids.

interacted with foreign peoples at degrees of intensity that varied across space and time, swapping not only their goods and cultures, but genes as well. As such, the cultural, ethnic, and biological character of the Nubians was distinct, largely shaped by their role as "middle men" (Adams 1977: 20) who had a hand in cultural, economic, and biological movement and exchange along the Nile River Corridor (Adams 1977).

3.2.3 Medieval Nubia (550–1500 C.E.)

The beginning of the Nubian medieval period (~500 C.E.) corresponds closely with the dispatch of missions that spread Christianity into Nubia.[14] The term "medieval" encompasses the Early Christian (550–850 C.E.), Classic Christian (850–1100 C.E.), Late Christian (1100–1400 C.E.), and Terminal Christian (1400–1500 C.E.) Periods. These overlapping systems of nomenclature are based on the traditional division of the Nubian medieval period into chronological phases associated with changes in pottery typology, but also with changes in the religious, political, and economic environment.[15] This dissertation uses the term "Early Christian" to refer to the period of interest at Kulubnarti as the individuals that are analyzed in this work inhabited the site throughout this early part of the medieval period.

Two events during the medieval period had particularly profound significance on the course of Nubian culture history: the first was the introduction of Christianity in the mid-6th century, and the second was the fall of Egypt to Muslim armies and the

[14] While the medieval period is described as beginning ~500 C.E., it should be recognized that there is evidence of both cultural (Trigger 1965; Shinnie 1996; Welsby 2002; Edwards 2004, 2007) and biological (Van Gerven et al. 1977; Greene 1982; Calcagno 1986) continuity between the inhabitants of medieval Nubia and the previous inhabitants of this region.

[15] A complete chronology of these phases can be found in Adams et al. (1999) or Welsby et al. (2002). While the term "medieval" is preferred by some to stress continuity with the pagan Nubian kingdoms that pre-dated the introduction of Christianity, the "Christian" terminology is used more frequently (Welsby 2002).

subsequent penetration of the Islamic faith into Nubia in the mid-7th century (Adams 1977). To fully understand the impact of these events, it is necessary to place them within the broader sociopolitical and environmental context of medieval Nubia.

Prior to the medieval period, there was a decline in centralized political authority that remains largely obscure and poorly understood (Shinnie 1996); however, this decentralization is observed in the archaeological record to have incited the development of more regional cultures during the medieval period (Edwards 2007). An abundance of evidence supports the emergence of three distinct Nubian polities (or political states) in the early 6th century C.E. (Figure 3.3). From north to south, these were Nobadia in Lower Nubia and the *Batn el Hajar*, Makuria in the Dongola Reach, and Alwa (alternatively, Alodia) in central Sudan (Edwards 2007).

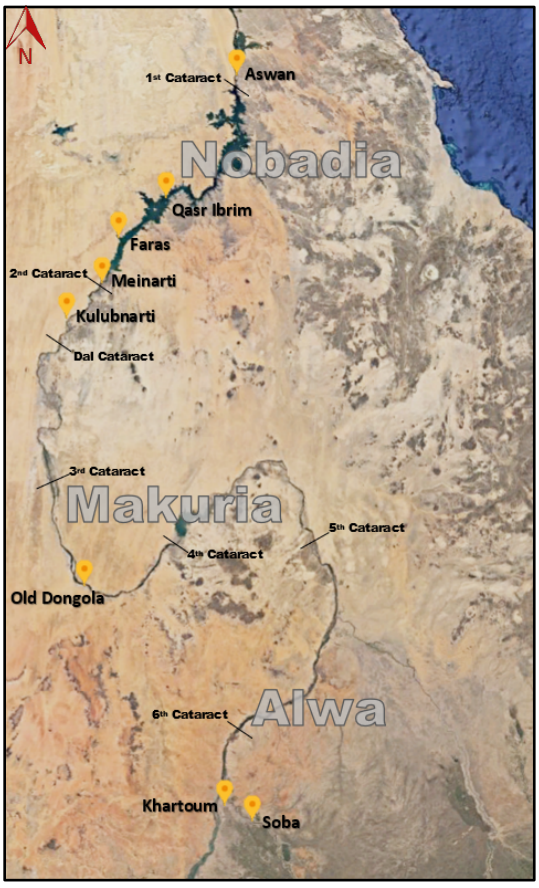


Figure 3.3: Division of Nubia into the political states of Nobadia, Makuria, and Alwa amongst the six numbered cataracts of the Nile and the Dal Cataract. Settlements of importance in medieval Nubia mentioned in this chapter are included on this map; Aswan, Khartoum, and Kulubnarti are provided for reference. Map adapted from Google Earth (version 9.2.47.9).

The development of these three distinct polities represented the new settlement landscape that characterized the medieval period (Edwards 2007). In each polity, particularly powerful regional centers emerged. For example, most likely Faras, but possibly Qasr Ibrim, functioned as the likely capital of Nobadia; the kingdom of Makuria had its capital at Old Dongola; and Soba served as the capital of Alwa (Jakobielski 1987; Shinnie 1996; Edwards 2007). Unfortunately, the dynamics of settlement in the medieval period remain little explored, and very little indigenous documentation is available to shed light on the social and political interactions occurring between the polities (Edwards 2007), though there are some indications of fluid alliances between the three polities across the medieval period and clear evidence pointing toward the merging of Nobadia and Makuria shortly after the introduction of Christianity (Adams 1977; Shinnie 1996).

Variation in the Nubian physical environment is thought to have had a strong influence on each polity's economic development and relations (Jakobielski 1987). Specifically, it is thought that Nobadia (the northernmost polity, and also the location of Kulubnarti), with a narrow Nile Valley and little possibility for agricultural production, developed extensive trade relations with Egypt during the Christian Period as well as a money economy based exclusively on Egyptian currency (Jakobielski 1987). In contrast, based on their geography, the other two political states had greater potential for the development of agriculture and animal husbandry; however, they remained the main commercial agents that supplied goods, gold, and slaves to Egypt (Jakobielski 1987).

The rise of these polities may have been associated with the most significant event of the Nubian medieval period: the conversion from paganism to the new religion of Christianity. Written records document a large-scale conversion to Christianity that took place in Nubia during the mid-6th century, with the first Christian missionaries appearing in 542 C.E. originating from Constantinople in the Near East (Kirwan 1959; Adams 1965; Vantini 1975; Säve-Söderbergh 1987). At the time of the Christian movement into Nubia, there were two major denominations of Christianity that differed on their belief in the physical nature of Jesus Christ, Melkite (the "official" sect) and Monophysite (the "heretical" sect). Both denominations dispatched missionaries simultaneously, and all of Nubia is reported to have been Christianized within forty years (Adams 1965; Edwards 2004).

The motives for Christianization were political as well as religious: rival factions in Egypt sought to strengthen themselves by enlisting the support of the Nubians (Adams 1977). While the conversion to Christianity can be seen archaeologically through changes in burial practices, the presence of Christian motifs in art, and the construction of churches, the depth of infiltration of Christianity into the lives of the general Nubian population remains debated (Adams 1977; Kirwan 1984; Shinnie 1996; Welsby 2002; Edwards 2004). Importantly, the arrival of Christianity encouraged new and productive links with the north, especially Egypt, and was the defining cultural development of the medieval era in Nubia (Edwards 2004).

However, not long after the Christianization of Nubia, Egypt was subjugated by Arab invaders carrying Islam from the Arabian Peninsula beginning in 639 C.E. and culminating with the surrender of Alexandria in 641 C.E. (Adams 1977; Shinnie 1996).

Arab armies then penetrated into Christian Nubia as far south as Dongola (between the Third and Fourth Cataracts) but were met with Nubian resistance and eventually defeated (Hasan 1967). A second invasion by the Arabs into Nubia was mounted in 651–652 C.E., during which the Muslim army reached as far up the Nile as Old Dongola. This attack ultimately ended inconclusively and resulted in a negotiated truce known as the *Baqt* that would remain in place for several centuries (Shinnie 1996).

The *Baqt* treaty is most commonly interpreted as a reciprocal agreement that was equal-part imposed tribute and commercial treaty that involved the delivery of slaves by the Nubians in return for food supplies, wine, and cloth from the Muslims (Spaulding 1995; Shinnie 1996). The treaty effectively institutionalized trade relations between the Nubians and the Islamic world and likely contributed to the prosperity of Nubia (Adams 1977). Importantly, as per the stipulations of the *Baqt*, Arab Muslims were permitted to enter Nubia as traders, but not as settlers (Hasan 1967), though it is noted that the *Baqt* "led in the long term to a prolonged presence of Arab merchants in Nobadia" (Jakobielski 1987: 232). This treaty may have had important implications for contact and gene flow between resident Nubians and Muslim foreigners.

The *Baqt* treaty determined the course of Muslim-Nubian relations for nearly 600 years (Adams 1977; Jakobielski 1987); importantly, it marked the Muslim acceptance of Nubian independence (Welsby 2002) and resulted in at least five centuries of peaceful relations between Nubians and their more powerful and affluent northern neighbors (Shinnie 1996). It was precisely this peace that allowed the development of a distinctive Nubian culture and provided Nubia with the opportunity to follow an increasingly independent line of political and cultural development (Adams 1965; Shinnie 1996).

During the 9th and 10th centuries, relations were particularly amicable between Egypt and Nubia, and these centuries were characterized by cultural florescence, economic prosperity, and a trend toward Nubian cultural independence. All of this was aided by a stable Nile flood that catalyzed agricultural development and resource surplus, which led to larger and more concentrated settlements in certain areas of Nubia, such as at sites such as Debeira and Meinarti in Lower Nubia (Adams 1965, 1967; Jakobielski 1987; Shinnie 1996). Due to a combination of peace and environmental stability during this period, Nubia gained more influence than ever before, expanding contact as far as Persia and Syria (Shinnie 1965).

Important to this dissertation, the introduction of Christianity and the increasing intensity of the relationship between Muslims and Nubians provides evidence for sustained contact with foreign empires during the medieval period (Edwards 2007). Crowfoot (1927) explains that at the time of their Christianization, the Nubians possessed a complex culture which combined Egyptian and Mediterranean influences. This description suggests that the religious, economic, and cultural development of Nubia in the medieval period, along with military conflicts, occupations, and trade along the axis of the Nile and possibly the Red Sea, may have provided the necessary opportunities for admixture and gene flow in ancient Nubia that can be further explored through genomic analysis.

As a caveat, it is necessary to recognize that gene flow throughout Nubia would not have occurred homogenously, due to environmental and political forces that exhibited varying degrees of influence across the subdivisions of Nubia and may have acted as barriers to gene flow in some regions. For example, isolated and small hamlets such as

Kulubnarti situated in desolate regions such as the *Batn el Hajar* may have experienced less contact with foreign peoples than populations in more accessible and desirable regions of Nubia, such as the fertile Dongola Reach between the Third and Fourth Cataracts, though the nature of contact has not yet been studied at this fine-grained scale.

Going forward, this dissertation will focus on the *Batn el Hajar* and the specific site of Kulubnarti, and will discuss the environmental, political, and biocultural forces that shaped population history, influenced contact with foreign peoples, and may have influenced patterns of genetic relationships and variation.

3.3 The *Batn el Hajar*

This dissertation examines the human remains excavated from the site of Kulubnarti, located within the Upper Nubian physiographic subdivision known as the *Batn el Hajar* ("belly of rock"), a region that extends for approximately 160 kilometers from the Second Cataract to the Dal Cataract (Van Gerven et al. 1995; see Figure 3.2).

The *Batn el Hajar* is the inhospitable region that separated Egyptian-influenced Lower Nubia to the north and the rest of Upper Nubia to the south in antiquity (Welsby 2002). While the *Batn el Hajar* technically sits within the realms of Upper Nubia, Adams et al. (1999) note that the cultural features found in this region show more affinity to those in Lower Nubia than they do to those in the southern parts of Upper Nubia. For this reason, the *Batn el Hajar* is considered to be a transition zone between Upper and Lower Nubia that does not fit completely into either region.

This region is described to be the most "barren and forbidding of all Nubian environments," with a "lunar" feeling characterized by steep riverbanks, countless *jebels*

(large outcroppings of rock), and sharp *wadis* (Adams 1977: 26-27; Thurmond et al. 2004). The Nile becomes unnavigable in the *Batn el Hajar*, coursing through granite rapids and hundreds of riverine islands. While the *Batn el Hajar* supported scattered populations based in small villages and hamlets clustered around the region's few floodplains, this terrain simultaneously functioned as a natural deterrent against the infiltration by foreign peoples in the past.

In addition to serving as a natural barrier to migration and invasion, the *Batn el Hajar* was also a political barrier during the medieval period. After the establishment of the *Baqt* treaty, this region was designated as a closed zone through which travel was strictly controlled by the Nubian king who sat further south at Dongola. Based on a combination of environmental and political factors, the *Batn el Hajar* garnered the reputation of the "granite curtain" that separated Upper and Lower Nubia and dissuaded the movement of peoples from Egypt and the Arabic world southward along the Nile corridor (Adams 1977: 26). It is suggested that this "granite curtain" protected the inhabitants of Upper Nubia from the gradual expansion of Islamic influence until the 12th century C.E. (Adams 1977).

It is therefore likely that Nubian populations inhabiting the *Batn el Hajar* experienced less contact with foreign peoples than their northern or southern neighbors who inhabited more geographically-inviting and politically-open regions. As such, it is possible that an overrepresentation of genetic signatures of local ancestry and origins may reflect the environmental and cultural barriers that impeded admixture between populations of the *Batn el Hajar* and non-local peoples in comparison to populations inhabiting more inviting neighboring regions of Nubia.

3.3.1 Climate and Subsistence

In the *Batn el Hajar*, survival was marginal at best, and the hostile environment was embodied in the skeletal record of populations living in this region. The average daily summer temperature in the *Batn el Hajar* is around 30–35°C, falling to an average of 15–20°C during the winter (Adams 1977). Daily temperature extremes can vary by 16–17°C throughout the year. This variation means that in addition to high temperatures during the summer, inhabitants of this region must also deal with extreme temperature fluctuations, which places an increased burden on human physiology (White et al. 2004). Adding to the physiological burden, there has been no appreciable rainfall in the *Batn el Hajar*, at least throughout the recent millennia of the Holocene era (Jackson 1957; Shinnie 1996). Thus, active participants of daily life at a *Batn el Hajar* settlement such as Kulubnarti would be required to make multiple trips per day up and down the boulder-strewn slopes at the edge of the Nile to obtain water. Study of trauma in the Kulubnarti Nubians indicates that they experienced a relatively high level of injury in a pattern that is more consistent with accidental falls and less consistent with incidences of interpersonal aggression (Burrell et al. 1986; Kilgore et al. 1997).

Small hamlets dotted the riverbank wherever alluvial soil was present in the *Batn el Hajar*, but there was no continuous floodplain. As such, population density was consistently low (Van Gerven et al. 1995). Archaeological evidence suggests that the *Batn el Hajar* never supported large populations, and any hamlets or villages it did support were impoverished compared to those to in more fertile regions to the north and south. These hamlets would have been composed of few households engaged in small-

scale farming of sorghum, millet, barley, beans, lentils, peas, dates, and wheat that were completely dependent on the Nile flood for survival (Dafalla 1969; Adams 1977).

The extent to which Nubian populations depended on agriculture for subsistence varied between regions: for example, there is evidence of continued agropastoralism in certain regions long after the adoption of agriculture in others (Marshall and Hildebrand 2002). Particularly in the arid region of the *Batn el Hajar*, the flood level of the Nile carried both seasonal and long-term consequences for agricultural yields, which consequently influenced individual health and well-being as well as population growth and decline (White et al. 2004). Throughout the medieval period, the Nile flood fluctuated substantially. Evidence from Jiang et al. (2002) suggests that the period between 622–758 C.E. saw lower flood levels than the following period of 759–858 C.E. This lower-than-normal flood level inevitably affected the health and well-being of the individuals inhabiting this region, some of whom may be studied in this dissertation.

Agriculture in the *Batn el Hajar* was limited due to the paucity of arable land (Edwards 2004). Elsewhere in Nubia, exploitation of the Nile's waters improved with the ox-driven waterwheel (*saqia* in Arabic, *eskalay* in Nubian), introduced approximately 2,000 years ago, and the man-powered-lever-lifts for irrigation (*shaduf*). However, the steep and rocky riverbanks in the *Batn el Hajar* made these forms of irrigation problematic and rare. Instead, agriculture in the *Batn el Hajar* was primarily reliant on alluvium gradually exposed as the Nile receded from its annual flood in a process known as *seluka* cultivation (Zarroug 1991). Consequently, individual landholdings were very small and highly-valued.

It is possible that the scarcity of arable land created a power dynamic in some populations inhabiting the *Batn el Hajar* that has direct relevance to this dissertation. Specifically, a paucity of arable land could have resulted in the existence of a landless and semi-itinerant underclass community raising small flocks of sheep and goats while intermittently working for more prosperous landowning Nubians, a social relationship observed in present-day Nubia and hypothesized to have existed in Early Christian times as well. Specifically, the presence of "semi-nomadic persons who act as sharecroppers or seasonal laborers for the landowning families" (Adams and Adams 2006: 11) would explain the observed social division at Early Christian Kulubnarti. As such, using genetic data to directly characterize the biogeographic genetic affinities and ancestries of the impoverished S community and determine their genetic relationship to the R community Nubians will help to clarify the details of a social relationship in Early Christian Nubia which may still be in place in Nubia today.

3.4 Kulubnarti

Within the *Batn el Hajar* is the site of Kulubnarti, translated as "island of Kulb" in the Mahasi dialect of Nubian, located on the banks of the Nile River in Sudanese Nubia approximately 130 kilometers south of Sudan's border with Egypt. Before the creation of Lake Nubia, Kulubnarti was a headland projecting into the Nile River from its western bank, characterized by towering rocky ridges alternating with undulating plains devoid of vegetation (Adams et al. 1999).

In this dissertation, the site of Kulubnarti is considered to include the island of Kulubnarti as well as its adjacent west bank, which is presently separated from the island

by a narrow channel; however, before the filling of Lake Nubia, this large island (approximately 1 x 2 kilometers) was not a true island except at the peak of the Nile flood and, as such, was not geographically isolated from the mainland in Early Christian times (Adams et al. 1999) (Figure 3.4).

Figure 3.4: Kulubnarti, now a permanently true island due to the flooding of Lake Nubia. Map adapted from Google Earth (version 9.2.47.8).

3.4.1 Subsistence and Survival at Kulubnarti

The topography of Kulubnarti was consistent with that found throughout the *Batn el Hajar*. Due to its landscape, arable land was sparse, and therefore highly-valued; the implications of the paucity of land are described above.

Small-scale agricultural production was the basis for subsistence, and the Kulubnarti Nubians utilized the same fodder and legume crops and agricultural techniques that had been practiced for centuries. Unlike other villages (such as Wadi Halfa to the north) also located in arid regions of Nubia, there is no evidence for use of

the *saqia* at Kulubnarti before the Late and Terminal Christian Periods (1100-1500 C.E.) (Adams and Adams 1998), likely due to the rugged and steep terrain in the *Batn el Hajar* region that made irrigation difficult at this site (Adams 1977). Instead, the channel that separates the island from the mainland was farmed using *seluka* cultivation to grow legumes and other fodder crops (Adams et al. 1999).[16]

Isotopic data suggest that dietary consumption at Kulubnarti was primarily based on "winter" C_3 plants harvested in April (including barley, legumes, and wheat), with some consumption of "summer" C_4 plants harvested in June (including sorghum and millet) (Turner et al. 2007; Sandberg et al. 2014; Basha et al. 2018); this is consistent with dietary patterns throughout Nubia today (Daffala 1969; Adams 1977). Animals (including goats, cattle, sheep, and pigs) were kept in small numbers, but animal meat was uncommon in the diet (Van Gerven et al. 1995; Turner et al. 2007; Basha et al. 2018); instead the Nubians obtained their protein primarily from plant sources (Adams 1977; Sandford and Kissling 1994). The range of isotopic values and lack of archaeological evidence suggests that consumption of riverine products such as fish was rare, indicating that the Nubians subsisted upon a terrestrially-based diet (Adams 1977; Turner et al. 2007).

There is evidence of mat, basket, and sandal-making at Kulubnarti during the Christian Period; however, the notable lack of specialized craft and imported goods suggests that subsistence agriculture was likely the main activity (Van Gerven et al. 1995; Adams and Adams 1998). Based on the artifact assemblage recovered from Kulubnarti, as well as the paucity of luxury goods, Adams and Adams (1998) concluded that the

[16] *Seluka* cultivation is described previously in this chapter.

Kulubnarti Nubians were relatively isolated from the extensive trade with Egypt practiced at other sites in Lower Nubia and were probably also relatively impoverished compared to their neighbors to the north and south.

3.4.2 Archaeology at Kulubnarti

Archaeological data produced by the first systematic excavation of Kulubnarti in 1969 suggested that the site, occupied since approximately 500 C.E., was one of the last Christian strongholds in Nubia, with its occupants likely practicing the Christian faith with no external political control well into the 16th century (Van Gerven et al. 1981). The very late presence of Islamic influence was proposed to be a consequence of the relative isolation of the site in the *Batn el Hajar* (Van Gerven et al. 1981).

Artifacts diagnostic of Early Christian (550–850 C.E.), Classic Christian (850–1100 C.E.), and Late Christian (1100–1400 C.E.) Periods were recovered from various habitation sites and church ruins across Kulubnarti (Adams and Adams 1998). Throughout the Christian Period, the archaeologists detected fluctuations in foreign trade, an increase in population density, and an increase in the construction of defensive architecture. As mentioned in Chapter 2, architecture at Kulubnarti in the Early Christian Period was characterized by the construction of churches; however, the Late Christian Period saw a shift in focus from church building to the construction of defensive structures. This was largely due to political unrest and subjugation of Christians in Lower Nubia by increasing presence of the Muslim Arabs that led to the influx and settlement of Christian Nubian refugees from the north in the *Batn el Hajar* (Adams 1977; Edwards 2004).

While the initial archaeological expedition in 1969 focused primarily on excavating habitation sites, two cemeteries were also located less than 1 kilometer apart. In 1979, a new Joint Colorado-Kentucky Expedition returned to Nubia to undertake the systematic excavation of these cemeteries. While the data from the excavation of these two cemeteries will be reviewed briefly in the remainder of this chapter, more extensive details of the archaeology and bioarchaeology of Kulubnarti can be found in three monographs detailing the architectural remains, artefactual remains, and human remains recovered from the 1969 and 1979 excavations (Adams 1994; Adams and Adams 1998; Adams et al. 1999).

3.4.3 The Cemeteries

During the initial excavation at Kulubnarti in 1969, a total of 100 graves from two cemeteries were mapped and partially or fully opened. However, since the expedition included no physical anthropologist, the skeletal remains were not studied.[17] The 1979 expedition therefore returned to Nubia with a specific focus to excavate as many graves as possible and remove their human remains for detailed laboratory analysis. Over the course of two months, 406 individuals were disinterred from the two Kulubnarti cemeteries, shown in Figure 3.5.

[17] Unfortunately, all original field records of the 1969 cemetery excavations have been lost. Adams et al. (1999) discuss the challenges stemming from this information loss, though there is no effect on the present dissertation.

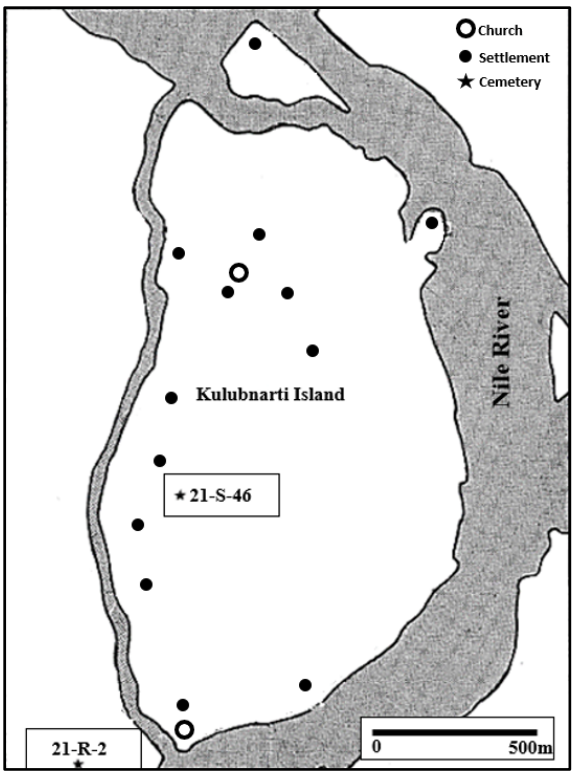


Figure 3.5: Map of site locations at Kulubnarti (redrawn from Adams et al. 1999).

The S cemetery (site 21-S-46 in Figure 3.5) was situated within a dry ancient *wadi* near the west side of the island. The earliest graves in the S cemetery were of pre-Christian type, with a clear transition into Early Christian-style graves, identified by grave orientation, body positioning, and burial shrouds. The total number of graves in the S cemetery remains unknown, though estimations place this number at approximately 300 (Adams et al. 1999). A total of 218 graves were excavated, and 215 bodies were uncovered from the S cemetery in 1979. Most of the graves were slot graves (straight-sided pits with rounded or square ends) that had a covering at the surface, most often simple pavements of flat but unshaped granite slabs arranged in a rectangle over the top of the grave (Adams et al. 1999). Each of the excavated graves had an east-west orientation with the head of the body placed at the west end, as is typical for Christian-

style burials (Adams et al. 1999). Most individuals, regardless of sex or age, were wrapped in a shroud; however, as is common with Christian Period burials, recovery of any personal goods included in the graves was rare (Adams et al. 1999).

The R cemetery (site 21-R-2 in Figure 3.5) was located on the west bank of the mainland just opposite the southern end of the island.[18] This cemetery was located next to a Classic Christian Period domed church as well as an Early Christian Period walled settlement (Adams et al. 1999). In addition to Christian Period graves dug into a barren alluvial surface that merged with the Nile floodplain, Islamic-type graves were found at one end of the R cemetery. It was estimated that the R cemetery contained between 500 and 600 graves (Adams et al. 1999). A total of 188 graves were opened during the 1979 expedition, all but six of which were concentrated in one contiguous area at the far western end of the cemetery. From these graves, a total of 191 bodies were recovered (Adams et al. 1999). Adams et al. (1999) noted that the graves from the R cemetery exhibited no typological distinction from those at the S cemetery.[19] They found that the orientation of graves at the R cemetery was more erratic than at the S cemetery but attributed this difference to the lack of any topographic feature on the western horizon that could serve as an orientation point. Like the S cemetery, the R cemetery most frequently exhibited slot graves, high frequency of burial shrouds, and limited grave goods (Adams et al. 1999).[20]

---

[18] In a strictly technical sense, this site should not be referred to as a Kulubnarti cemetery as it is not located on the island of Kulubnarti. However, it was utilized by the residents of Kulubnarti and as such is considered a Kulubnarti cemetery in all works concerning the site as well as in this dissertation (Adams et al. 1999).

[19] Interestingly, Adams et al. (1999) noted that the majority of graves at the R cemetery had no superstructure (in contrast to the S cemetery where >80% exhibited a superstructure). This was attributed to surface deflation as a result of downwash from the *wadi* that opened at the cemetery.

[20] A detailed comparative summary of the S and R cemeteries can be found in Adams et al. (1999).

The lack of distinctive grave goods presented challenges for the precise dating of the Kulubnarti cemeteries (Adams et al. 1999). The original interpretation of the available archaeological data was that the two cemeteries were used in successive periods with partial overlap. Specifically, analysis of pottery within the graves as well as architectural associations originally suggested that the S cemetery represented a population from the Early Christian Period (550–750 C.E.), while the presence of vaulted brick tombs and both Christian and Muslim burial styles suggested that the R cemetery was in use from the Early Christian through the Terminal Christian Period (550–1400 C.E.) (Van Gerven et al. 1981, 1995). As described in Chapter 2, the significant health differences observed between individuals of the two cemeteries were initially attributed to the changing influences of political and economic forces in greater Nubia that affected individual health and well-being (Van Gerven et al. 1981, 1990; Hummert and Van Gerven 1983; Mittler and Van Gerven 1994).

However, upon analysis of textile remains from the graves and a small sample of radiocarbon dates, it was suggested that both cemeteries may have been used during the Early Christian Period (Adams et al. 1999; per Dr. Dennis Van Gerven, pers. comm.). Analysis of the textiles found in the graves of the S and R cemeteries suggested that the graves excavated from both cemeteries exhibited the characteristics of Nubian textiles from the Early Christian Period, including a high percentage of woolen fabrics, a low percentage of cotton fabrics, an even lower percentage of flax, absence of silk, and rare occurrence of dyed color (Adams et al. 1999). Calibrated ^{14}C dates were subsequently obtained from 9 individuals from the S cemetery and 12 individuals from the R cemetery. The mean date for the S cemetery was determined to be 720 C.E. (SD=97.73,

range=546–848 cal. C.E.), while the mean date for the R cemetery was determined to be 756 C.E. (SD=65.62, range=612–837 cal. C.E.).[21] T-test results indicate that these means are not significantly different (t=1.0075; DF=19; p=0.3263). The non-significant means suggested that the two cemeteries should be considered as contemporaneous. Modified dates of 550–800 C.E. are now widely accepted for both cemeteries (Turner et al. 2007).

The recognition that the S and R cemeteries were contemporaneous further emphasized the need for a biocultural framework in which to interpret the differences in health and mortality observed between the S and R communities, which could no longer be attributed to shifting ecopolitical or cultural forces. The following section both describes these differences in morbidity and mortality and discusses possible explanations for such differences.

3.5 Bioarchaeology of Kulubnarti

As mentioned above, two primary conclusions stemmed from several decades of bioarchaeological research on the Kulubnarti Nubians: 1) the S and R communities were biologically indistinguishable but did not experience comparable levels of well-being or survival (Adams et al. 1999); and 2) the S community co-existed as an underclass relative to their R community neighbors. These conclusions were supported by multiple lines of evidence, including biodistance studies, paleodemographic reconstructions, and paleopathological investigations of dental and skeletal growth, developmental patterns, and degenerative skeletal changes, described in the following subsections.

3.5.1 Relatedness Between the S and R Communities

[21] These calibrated ^{14}C dates are unpublished data, provided to the author by Dr. Dennis Van Gerven.

As introduced in Chapter 1, there is a long history of examining biodistance using dental and osseous variation between populations in physical anthropology (Buikstra et al. 1990). In biodistance studies, population variation in phenotype is appreciated through metric analyses (craniometrics or dental metrics) or nonmetric trait analysis (discrete dental traits or cranial nonmetric traits). Some work exploring transmissibility and heritability of cranial metric traits (e.g., Devor et al. 1987) has asserted that craniometric data reveals approximately the same amount of variation as seen in molecular genetic studies (Relethford 1994), while other work finds that some craniofacial dimensions are better explained by environmental factors than genetic factors (Roseman and Weaver 2004). The use of nonmetric traits has been even more contested because few studies include associated pedigree information to allow for the calculation of heritability of nonmetric traits (Carson 2006).

When interpreting a biodistance study, it should be recognized that human skeletal variation is produced by the interaction of genetic, epigenetic, and environmental influences: neither metric nor nonmetric variation can be assumed to be the result of genetics alone (Relethford 2004; Konigsberg and Ousley 2009). As such, biodistance studies based upon morphology are considered to be an indirect (and potentially imprecise) proxy for the ultimate source of the human genome. They have been used to explore the relatedness of ancient populations across space and time in Nubia (Greene 1982; Van Gerven 1982; Calcagno 1986; Prowse and Lovell 1995; Irish 2005; Godde

2009, 2013; Stynder et al. 2009; Schrader et al. 2014) and have been applied specifically to the Kulubnarti Nubians.[22]

Comparing craniometric data collected from the Kulubnarti R and S communities to a time-series from Wadi Halfa, Van Gerven (1982) determined that the principal discrimination was between Wadi Halfa and Kulubnarti, while the least significant difference was between the S and R communities, indicating their close biological relatedness. They did, however, note that the R community demonstrated more morphological convergence with the Wadi Halfa series than the S community, suggesting that a northern origin for the economically-advantaged mainland community may be worthy of further exploration (Adams et al. 1999). Application of multivariate statistics to discrete dental data (Greene 1982) detected no significant variation between the Kulubnarti communities and also suggested that the mainland R community may be more closely related to Nubian populations from the north (Adams et al. 1999).

Though biodistance studies do not have the resolution necessary to explore biogeographic genetic affinities or ancestries, assign matrilineal or patrilineal origins, or reveal evidence of admixture, they suggest that the Kulubnarti communities were biologically indistinguishable and that this implies a close genetic relationship. While this genetic relationship will be further examined and quantified through direct genomic analysis, previous biodistance studies provide a justification for the expectation of little to no genetic distance between the Kulubnarti communities.

[22] The physical anthropology movement in the 1950s (Washburn 1951) did not invalidate biodistance studies, but did change the way they were interpreted. All studies mentioned here have been conducted within the biocultural approach that unties the concept of human biological variation and cultural character.

3.5.2 Social Disparity at Kulubnarti

In contrast to biodistance studies that support close biological relatedness between the S and R communities, studies of survivorship and health in a biocultural context reveal a very different lived experience for members of the two communities. The idea of neighboring communities that did not experience comparable levels of health and well-being is well-evidenced through analyses of stress-induced lesions (including cribra orbitalia and linear enamel hypoplasias), and perhaps most poignantly, through studies of mortality. All studies to date have led to the conclusion that the S community experienced more stress than their mainland counterparts.

Important differences between the S and R communities have been revealed through the analysis of several different indicators of generalized stress. The first is cribra orbitalia, a type of porotic hyperostosis that manifests on the roof of the eye orbit. Cribra orbitalia is a valuable marker of nutritional stress, and analysis of its prevalence has been applied widely to archaeological remains throughout the world. While its etiology is debated (Wapler et al. 2004; Walker et al. 2009), it is thought that iron-deficiency anemia resulting from endemic causes is most often the cause of the lesion (Carlson et al. 1974); it is almost certainly the cause of cribra orbitalia in the Nubians who have been shown to have very low iron levels (Carlson et al. 1974; Sandford et al. 1983; Sandford and Kissling 1994). Van Gerven and colleagues (1981) find that 94% of S cemetery children display the lesion in comparison with 82% of R cemetery children, indicating a higher degree of childhood stress for the S community. Mittler and Van Gerven (1994) break down this finding, showing high levels of childhood stress for both Kulubnarti communities, with higher lesion frequency, greater lesion severity, and an increased

likelihood of mortality associated with the condition found in members of the S community.

A similar pattern is seen when a second indicator of generalized stress, linear enamel hypoplasias (LEHs), are analyzed. LEHs are growth disruptions in tooth enamel caused by systemic stress events during the formation of the dentition (Goodman and Rose 1990). While Van Gerven and colleagues (1990) found a nearly universal presence of LEH lesions in both cemeteries at Kulubnarti, they also find that LEH lesions appear more frequently and are maintained at a higher frequency for longer in the S community than in the R community, leading to a prolonged period of intensified childhood mortality for members of the S community.[23]

The increased frequency of stress-induced lesions found in the S community corresponds to increased childhood mortality. Discussing research by Van Gerven et al. (1981), Adams et al. (1999: 43) write: "The sharpest, and also the most puzzling difference between Cemetery 21-S-46 and Cemetery 21-R-2 can be seen in the mortality figures from the two sites." Mean life expectancy computed from composite life tables (Moore et al. 1975) reveals an informative pattern: while differences in mortality after childhood are minimal between the S and R communities, mortality between birth and age eight is significantly higher in the S community than the R community (Adams et al. 1999). Thus, probabilities of dying are not only higher for the island children, but chances of dying remain higher for longer (Van Gerven et al. 1995). This results in an

[23] Specifically, it is found that the S community has 8% fewer children with only one hypoplasia than the R community, and 5% more children with 7–8 hypoplasias than the S community (Van Gerven et al. 1990, 1995). This "contrast at the extremes" is interpreted as being suggestive of higher stress among the S community children (Van Gerven et al. 1995: 473).

average life expectancy of 10.6 years for the S community compared to 18.8 years for the R community (Van Gerven et al. 1995).

Taken together, data from analyses of cribra orbitalia, LEH, and mortality indicates that the period of peak physiological stress occurred during childhood and was experienced to a greater degree by the S community. Interestingly, the differences in morbidity and mortality between the S and R communities cannot be attributed to community-based variation in diet (Turner et al. 2007). Analysis of stable carbon, nitrogen, and oxygen isotopes from bone tissue indicates no significant relationships between isotopic indicators related to cemetery of burial, suggesting no isotopically-measurable differences in proportional dietary composition (Turner et al. 2007).[24] Instead, it appears that the source and contribution of dietary protein in both communities was significantly related to age, suggesting age-related differences in dietary intake, with subadults ingesting less protein or more isotopically-depleted protein (Turner et al. 2007). Turner et al. (2007) find that differences in markers of stress between the two Kulubnarti communities are therefore due to factors other than diet, and suggest that future research might consider the role of parasitic infection among other ecological factors.

Based on a combination of archaeological and bioarchaeological evidence discussed in this chapter, Adams et al. (1999: 48) write: "The combination of cultural and biological evidence from the two Kulubnarti cemeteries suggests a wholly unexpected possibility: that this region in Early Christian times was home to two biologically and culturally indistinguishable, but socially distinct, communities, one of which was

[24] Results from a different study suggest that there may be a difference between the two cemeteries in the amount of C_4 dietary carbon sources consumed, but temper these results by noting that they might only appear in metabolically-active soft tissues, but not in the long-term averaged isotope values preserved by bone (Basha et al. 2018).

considerably better off than the other. The grave types, the grave goods, and the skeletal types are virtually indistinguishable as between the two Kulubnarti cemeteries, while the figures for pathology and mortality are significantly different."[25] In order to rationalize the existence of "neighboring Christian communities with powerful genetic ties [that] did not experience comparable levels of survival and biological wellbeing" (Adams et al. 1999: 88), several explanations were raised. Among them was the possible existence of a Kulubnarti "underclass" (Adams and Adams 2006: 11).

3.5.3 The Underclass

Several hypotheses were proposed to explain the relationship between and the identity (or identities) of the S and R communities at Kulubnarti.[26] First, it was proposed that the S community was a slave population that lived and worked at Kulubnarti. However, the significant morphological similarities between members of the S and R communities as revealed by biodistance studies did not support this hypothesis. Slave populations in Nubia were most often drawn from tribal peoples from southern and western Sudan who had physical characteristics distinct from the Nubians, making it unlikely that the individuals interred in the S cemetery represented a population of slaves (Adams et al. 1999). Second, it was proposed that the S cemetery was a group of ethnic immigrants. However, the similarities of burial practices and grave goods between the S and R cemeteries as well as estimations of close biological relatedness did not support

---

[25] The original phrase used by Adams et al. (1999: 88) is "two *racially* and culturally identical, but socially distinct, communities…" However, as Adams and his colleagues subscribed to a biocultural approach (explained in detail in Chapter 2) that did not lend credence to the existence of 'racial type' with the connotations that are associated with it today, a more accurate interpretation of this phrase would be "biologically indistinguishable." I thank Dr. Dennis Van Gerven, a colleague of Adams' and co-author on Adams et al. (1999), for discussions about this quote and his recommendation regarding the change in wording.

[26] These proposed explanations are discussed in greater detail in Adams et al. (1999).

this hypothesis. Third, it was proposed that the S community was a population of refugees from north of the *Batn el Hajar* fleeing persecution. However, there is no record of sustained military disturbance in northern Nubia after the Arab invasion of 652 C.E. until after the end of the Early Christian Period, making this explanation unlikely.

Finally, it was proposed that there existed a "… single culturally homogenous but socially stratified Nubian population" (Adams et al. 1999: 49). This explanation did not look to theories of migration to explain the existence of the S community, but instead approached this enigmatic subpopulation from a biocultural approach that engaged the social realm and explored its correlation with disparities in health and mortality. While there is neither textual evidence nor archaeological evidence from other sites to support this interpretation, it is this fourth proposed explanation that is examined in greater detail in this dissertation.

As introduced in Chapter 1, the social organization of present-day Nubia provides support for the hypothesis of the S community as the Kulubnarti underclass. Adams et al. (1999) describe a personal communication in which it was revealed that in the Dar el-Mahas region spanning roughly from the Second Cataract to the area of Dongola (i.e., the region where Kulubnarti was located), there was, in the recent past, a population of landless Nubians who provided casual labor to landowning Nubians. Adams et al. (1999: 49) describe this population: "These people, although ethnic Nubians, were looked down on as inferiors by their more prosperous fellows, and were not allowed to live among them. They made their abodes, usually temporary, on barren, uninhabited land, and especially on rocky islands. The evidence from Kulubnarti at least raises the possibility that a similar social division prevailed a millennium earlier." Ultimately, through a

combination of ethnographic, historical, archaeological, and bioarchaeological evidence, Adams and Adams (2006: 13) conclude that "the island cemetery [S cemetery], with its high infant mortality and pathology rates was the burial place for the impoverished and malnourished underclass folk." This dissertation now reexamines the hypotheses surrounding the identity of the S community from a genomic perspective.

3.6 Kulubnarti: A Summary

As the "middle men" who inhabited the "corridor to Africa," the lives of the ancient Nubians were shaped by a complex interaction between a demanding physical environment, centered around the life-sustaining Nile flood, and a dynamic social environment, characterized by a mixture of indigenous and foreign influences that varied in intensity across space and time. In particular, the formal arrival of Christianity in the sixth century from Constantinople facilitated new cultural links between the Nubians and foreign peoples, while subsequent contact with Muslims from the Arabian Peninsula resulted in the further development of links between the Nubians and the Arab world starting during the Early Christian Period and extending for hundreds of years after (Edwards 2013). While many questions remain, including the scale of medieval Nubia's external trading contacts, the extent of the medieval Nubian slave trade with the Islamic world, and the amount of contact between the riverine Nubian kingdoms and the Red Sea region, the presence of Christianity and Islam provides important evidence of links between Nubian peoples and empires to the north (Edwards 2013).

In the Early Christian Period, Christian influence was seen throughout Nubia, including at the site of Kulubnarti located in the desolate *Batn el Hajar* region that served

as the “granite curtain” separating Lower and Upper Nubia. This demonstrated that foreign influence did reach even isolated Nubian sites, such as Kulubnarti, though evidence suggests that contact with outsiders may have been less extensive than in neighboring areas of Nubia due to the environmental challenges and political obstacles that characterized the *Batn el Hajar* region.

Research on the human remains recovered from Kulubnarti has provided important insight into the identity and lives of the Nubian peoples living in the harshest of all environments. While evidence of stress is embodied on the majority of the human remains from Kulubnarti, one “anomaly” (Adams et al. 1999: 88) is particularly relevant to this dissertation: archaeological and bioarchaeological evidence suggests a single biological population with shared cultural characteristics, but a population that was also socially stratified, in which the people of the S community inhabiting the island of Kulubnarti were significantly more stressed and experienced a higher level of mortality than their R community neighbors who inhabited the mainland.

Approaching Kulubnarti from a biocultural perspective, Adams et al. (1999) suggest that the disparity in health and well-being may stem from membership in one of two communities that composed the population of Kulubnarti: one of sedentary, landowning, free-hold farmers, and the other of itinerant pastoralists and seasonal laborers positioned in Early Christian society as an underclass. Study of the skeletal remains from Kulubnarti has revealed that the social stratification observed in Nubia today may have also been present 1,000 years ago.

Building upon the prevailing idea of a socially-distinct and economically-deprived underclass buried in a separate cemetery near that of the more prosperous,

biologically-indistinguishable, landowning Kulubnarti Nubians, this dissertation takes the next steps in exploring potential population substructure and characterizing the genetic composition of the inhabitants of Kulubnarti through direct analysis of genetic variation. Switching focus, the analysis of ancient DNA and the challenges associated with the genomic analysis of the Kulubnarti Nubians are presented in the next chapter.

CHAPTER 4
ANCIENT DNA

"Findings from aDNA [ancient DNA] research are currently transforming our understanding of human history at an ever-increasing pace." (Haber et al. 2016: 7)

4.1 Chapter Overview

The analysis of DNA has become an important tool in anthropology. This chapter begins with an introduction to DNA, specifically providing an overview of genetics and human genetic variation. An emphasis is placed on Single Nucleotide Polymorphisms (SNPs), variations in a single base in a DNA sequence, analyzed in this dissertation to investigate the genetic relationship between and explore the genetic composition of the Kulubnarti Nubians.

The focus of this chapter is then narrowed to ancient DNA (aDNA). Though little more than three decades old, the field of ancient DNA has grown rapidly to reveal a human history far richer and more complex than inferred from models based on modern DNA (Haber et al. 2016). It has quickly become the preferred method for reconstructing the population origins, demographic history, migration patterns, and biogeographic affinities of past populations (Morozova et al. 2016). This chapter discusses the development of the field of ancient DNA, reviews the current state of aDNA research, and describes the integration of ancient DNA with anthropology, which provides a contextual framework for deciphering the processes contributing to genetic variation. As aDNA research moves from the analysis of a single archaeological sample to more systematic hypothesis-driven analyses of large sample sizes, effective integration of genetic results with anthropological context becomes increasingly important.

This chapter then discusses the challenges and limitations associated with the analysis of ancient DNA, including contamination from modern and environmental DNA and accumulating *post-mortem* damage, as well as a specific challenge faced by this dissertation: the thermally-driven degradation of DNA due to the extremely hot climate in Nubia. All else being equal, thermal degradation is the most daunting challenge to overcome in aDNA analysis (Collins et al. 2002), but it can be mitigated with the use of optimized methodologies to prepare DNA for Next Generation Sequencing (NGS), as is done in this dissertation.

As the direct analysis of ancient genetic variation is an important tool for exploring questions of anthropological relevance, this chapter concludes with a discussion of how this dissertation fits within current research trends in molecular anthropology. It aligns well with three trends in particular: first, it re-examines anthropological research questions that can benefit from a genomic perspective; second, it thoroughly integrates new genetic data with existing anthropological data; and third, it analyzes samples from a sub-optimal depositional environment along one of the most important migration corridors in Africa, filling a gap in the existing record of aDNA studies.

4.2 DNA Overview

DNA, or deoxyribonucleic acid, is the hereditary material present in humans and almost all other organisms. DNA carries genetic information as a code made up of repeating chemical bases. These bases are adenine, guanine, cytosine, and thymine, abbreviated as A, G, C, and T, respectively. Two bases, A and G, are purines (bases

structurally composed of two closed rings), while the other bases (C and T) are pyrimidines (bases structurally composed of one closed ring). Bases pair with each other, specifically A with T and C with G, to form units called base pairs (bp). In addition to a base pair, each base is also joined to a five-carbon sugar (deoxyribose) and a phosphate molecule. Together, a base, sugar, and phosphate are called a nucleotide. Nucleotides are the subunits that form the DNA molecule.

Nucleotides are arranged in two strands to form a spiral called a double helix (Watson and Crick 1953). This structure is akin to a ladder, with the base pairs forming the rungs of the ladder and the sugar and phosphate molecules forming the vertical sidepieces of the ladder; DNA is referred to as a double-stranded molecule. Because of the specificity of pair bonding (A always with T and C always with G), the two strands in the double helix are complementary: for example, an adenine on one strand is always paired with a thymine on the other. Therefore, knowing the sequence of one strand reveals the sequence of the other strand.

A DNA strand also has directionality, and either end of the strand can be differentiated from the other end. The two ends of the strand are referred to as the 5'-end and the 3'-end. The 3'-end has a free hydroxyl (-OH) group on the 3'-carbon, and the 5'-end has a free hydroxyl group on the 5'-carbon. In the double helix structure, the 3'-end of one strand is parallel to the 5'-end of the other strand, and so the organization of the DNA double helix is consequently referred to as anti-parallel (illustrated in Figure 4.1). Fundamental processes, including DNA replication, occur in one direction only (from the 5'-end to the 3'-end), and differentiating the ends of the DNA strands becomes important during manipulation of the DNA molecules in preparation for NGS.

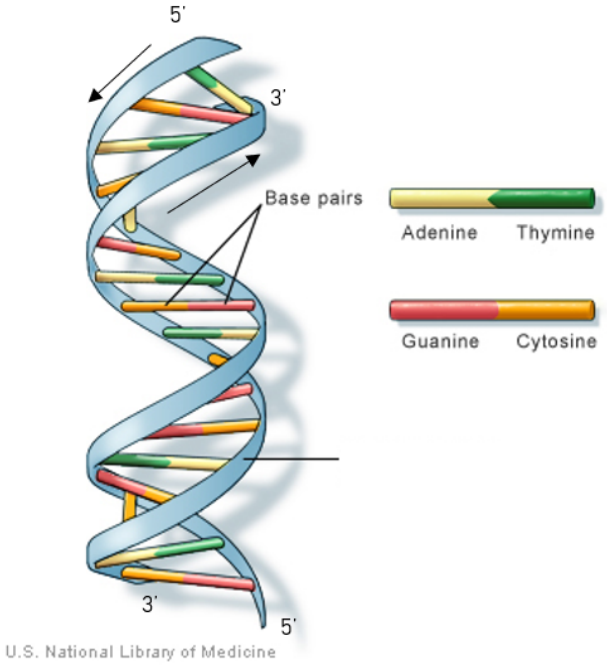


Figure 4.1: DNA is a double-stranded molecule organized as a double helix, with each nucleotide bound by hydrogen bonds to a nucleotide from the complementary strand. Each strand has directionality, with a 5'-end and a 3'-end, and the double helix is anti-parallel. All fundamental processes involving DNA occur from the 5'-end to the 3'-end as indicated by the black arrows. Image adapted from the U.S. National Library of Medicine (NLM).

DNA has two central biological roles. First, it is involved in protein synthesis. Segments of DNA known as genes provide the necessary instructions for the synthesis of proteins, macromolecules that are involved in all chemical processes in living organisms. To synthesize a protein, the nucleotides that compose a gene must be "read" like the letters in a word. Each "word" (called a codon) is three "letters" (bases) long and specifies the production of a particular amino acid. Amino acids are the building blocks of proteins. In order for the nucleotide sequence to be "read," the double helix structure of the DNA must be denatured (meaning that the double helix is opened to form a single-

stranded molecule). It is then possible to "translate" the nucleotide sequence into a defined sequence of amino acids used for building proteins.[27]

The second role of DNA is to provide a means for genetic information to be passed to daughter cells when cell division occurs, enabling genetic transmission from one generation to the next (Jobling et al. 2014). DNA replication is the process of copying each DNA strand into a new complementary strand to be passed onto daughter cells. The complementary arrangement of base pairs along each strand of the double-stranded DNA molecule allows for semiconservative DNA replication, meaning that a new complementary strand is always copied from an existing strand (Meselson and Stahl 1958). This type of replication is both rapid and accurate (Alberts 2003). During DNA replication, two new identical DNA molecules are constructed from the original, and the cell that contains the replicated DNA can now divide into two daughter cells, each containing an identical copy of the entire genome.[28]

4.3 The Human Genome

While DNA is found in all cellular organisms, the specific focus of this dissertation is on human DNA and human genetic variation. DNA is located in two compartments in human cells: within a cell's nucleus and within a cell's mitochondria (organelles located in the cytoplasm of the cell that produce energy for the cell). An individual's genome consists of the entirety of the genetic material in his or her

[27] A detailed description of the processes required to produce a protein from a DNA sequence (transcription and translation) are not described in this dissertation, but can be found in any introductory genetics textbook.
[28] More details about the process of DNA replication can be found in any introductory genetics textbook.

chromosome set (nuclear genome) and in the mitochondria (mitochondrial genome, or mitogenome).

The majority of DNA is found within the nucleus of the cell; this DNA is referred to as nuclear DNA. Nuclear DNA is organized into structures called chromosomes. Each chromosome is made up of a single DNA molecule tightly coiled around proteins known as histones that provide structural support (Kavenoff et al. 1974). Genes, the DNA sequences that contain instructions for protein synthesis, are located on each of the chromosomes at a defined position called a gene locus. An important part of DNA research has been sequencing and mapping the ~19,000 protein-coding genes present at various loci throughout the human genome (Ezkurdia et al. 2014).

Humans have 46 chromosomes arranged in 23 pairs. Both chromosomes in a chromosomal pair have the same structure and gene patterns and are therefore referred to as homologous pairs. This genetic layout makes humans diploid organisms: one copy of each chromosome is inherited from each parent (i.e., there is a maternal and paternal copy of each chromosome in each cell of the body that contains a nucleus). Of these 23 chromosome pairs, 22 are autosomes (any chromosomes that are not sex chromosomes), numbered from largest (chromosome 1) to smallest (chromosome 22) in size. The remaining chromosome pair is the sex chromosomes. Humans have two X chromosomes if they are female, and an X chromosome and Y chromosome if they are male. Altogether, the size of the human nuclear genome is ~3,200,000,000bp, often written as 3.2Gb ("gigabases") (The International Human Genome Mapping Consortium 2001).[29]

[29] This is the size of the haploid genome, meaning a single copy.

In addition to being found in the nucleus, DNA is also found in the mitochondria. Each mitochondrial DNA (mtDNA) molecule is only 16,569bp in size (often written as 16.5kb, or "kilobases"), of which ~7% is a non-coding "control region" that comprises the main focus for studies of mtDNA variation in modern human populations (Anderson et al. 1981; Ingman et al. 2000). mtDNA is a powerful tool for understanding human evolution due to several important characteristics that distinguish it from nuclear DNA. First, though it is only 0.0005% of the size of the nuclear genome, mtDNA is present in high copy number (there are thousands of mitochondria in each cell, each containing 2–10 copies of mtDNA) (Tang et al. 2000), and is not bound in chromosomes, but instead is present as a small circular molecule. Second, mtDNA exhibits a high mutation rate (defined as the number of new mutations per generation), determined to be approximately 10x higher than nuclear DNA (Brown et al. 1979). Finally, mtDNA is passed from generation to generation through a strictly maternal mode of inheritance (Giles et al. 1980).

This maternal mode of inheritance is of particular interest in anthropological research because every living individual can be assigned to an mtDNA haplogroup, defined as a group of people who share a particular set of mutations in the mitogenome that are often inherited together. Because offspring inherit an exact copy of their mother's mtDNA (barring any *de novo* mutations), sequence variation in mtDNA is generated solely by the sequential accumulations of new mutations that can be traced throughout the world along radiating maternal lineages of haplogroups, where shared mtDNA haplogroups suggest shared matrilineal ancestry (Torroni et al. 2006). Worldwide human

mtDNA variation was first characterized in 1987 (Cann et al. 1987) and continues to be important to the reconstruction of population origins and migrations of females today.[30]

mtDNA is known as a uniparental marker, meaning that it is inherited from a single parent (the mother) and is passed down through the maternal lineage. Humans also have another uniparental marker, the Y chromosome, which is passed strictly from fathers to sons (Jobling and Tyler-Smith 1995). Unlike mtDNA (which is located outside the cell nucleus), the Y chromosome is found within the cell's nucleus along with the other chromosomes and is classified as a sex chromosome along with chromosome X. The presence or absence of the Y chromosome determines sex, and since females do not have a Y chromosome, it is only possible to pass chromosome Y from father to son. Like mtDNA, worldwide patterns of variants on the Y chromosome (in particular, the Non-Recombining portion of the Y, or NRY; alternatively, the Male-Specific Region of the Y, or MSY) are used to explore patrilineal origins and migrations through Y chromosome haplogroups (Underhill et al. 2000). Assignment to a Y chromosome haplogroup (or just "Y haplogroup"), defined as a group of males who share a particular set of mutations in the Y chromosome from a common ancestor that are often inherited together, becomes particularly useful to anthropologists as a complement to mtDNA for examination of sex-specific mobility and kinship practices.

4.3.1 Sequencing the Human Genome

DNA sequencing first became possible in 1975 when a method was created for determining the order of nucleotides within a DNA molecule (Sanger and Coulson

[30] It is important to note that while males inherit mtDNA, and therefore, like females, have an mtDNA haplotype, they do not pass their mtDNA onto their offspring.

1975).[31] The first complete human mitochondrial genome (the Cambridge Reference Sequence, or CRS) was sequenced in 1981 (Anderson et al. 1981) with initial sequencing errors corrected in a new version (the revised Cambridge Reference Sequence, or rCRS) in 1999 (Andrews et al. 1999).[32] The first draft sequences of the human nuclear genome were released in 2001 by the Human Genome Project as well as by a competitor group led by Craig Venter (The International Human Genome Mapping Consortium 2001; Venter et al. 2001). Improving upon this initial genome build, updates are continuously being made to the human genome reference sequence. The Genome Reference Consortium (GRC) released the GRCh37 assembly of the human genome in 2009 and the GRCh38 assembly in 2013.[33] In comparison to the first build of the reference genome, which had ~150,000 gaps in its sequence, the GRCh37/hg19 build has only ~250 gaps (E pluribus unum 2010), demonstrating vast improvement under a decade. While more recent versions are available, this dissertation utilizes the GRCh37/hg19 build of the human genome as it is the build referenced by most current projects, such as the main phase of the 1000 Genomes Project (The 1000 Genomes Project Consortium 2012).[34]

4.3.2 Human Genetic Diversity

---

[31] This method is known as Sanger sequencing.

[32] Another version of the mitochondrial genome, the Reconstructed Sapiens Reference Sequence, or "RSRS," can be found in Behar et al. (2012). This dissertation utilizes both the rCRS and RSRS as mitochondrial reference genomes, explained in greater detail in Chapter 8.

[33] Note that the University of California, Santa Cruz (UCSC) maintains another genome browser that uses different nomenclature, although the genome assemblies are identical: in UCSC's system, hg19 is equivalent to GRCh37, and hg38 is equivalent to GRCh38.

[34] It is important to recognize that a reference sequence does not represent the genome of one specific person; instead, it provides a mosaic of DNA sequences from different donors. For example, the GRCh37/hg19 reference genome is a mosaic haploid genome derived from approximately thirteen people (E pluribus unum 2010).

When studying human genetic diversity, it is essential to emphasize that all humans are members of a single species of recent evolutionary origin (Harpending et al. 1998). Genome sequences between any two humans are estimated to be approximately 99.6–99.9% identical (Jorde and Wooding 2004; Tishkoff and Kidd 2004). The other 0.1–0.4% comprises sites or regions of the genome where genetic variation can be detected and studied (Collins et al. 1998; The International SNP Map Working Group 2001; Tishkoff and Kidd 2004).

The origin of genetic variation is mutation. Mutations are random heritable changes in a gene or chromosome resulting from an alteration to the DNA sequence (additions, deletions, or substitutions of bases). They are caused by inaccurate copying of DNA during cellular division or by exposure to specific environmental influences (e.g., chemicals or radiation) and result in a change in DNA sequence. Because mutations occur randomly (i.e., they do not occur to supply what an organism needs to survive and often have no physiological effect on the organism) and regularly, they accumulate in a "clock-like" fashion and can be used to track genetic changes in a population or species through time (The International SNP Map Working Group 2001).

While mutations range in size, the simplest and smallest-scale difference between two DNA sequences affects one base.[35] These single base changes, called SNPs, are the focus of this dissertation's analysis.

4.4 Single Nucleotide Polymorphisms (SNPs)

[35] While there are many larger-scale mutations that can occur, including mutations on a multi-base or whole chromosome scale, this dissertation will focus specifically on the analysis of single-base changes.

SNPs are single-base sites in the genome at which an individual may have either of two or more bases, where each potential base is referred to as an allele. At any SNP site, when a mutation causes a purine to be substituted for another purine (e.g., A for G) or a pyrimidine for another pyrimidine (e.g., C for T), the change is called a transition. When a mutation causes a purine to be substituted for a pyrimidine (e.g., A for C), or vice versa, the change is called a tranärsion. All substitutions are considered in relation to the human reference sequence, and examples are illustrated in Figure 4.2.

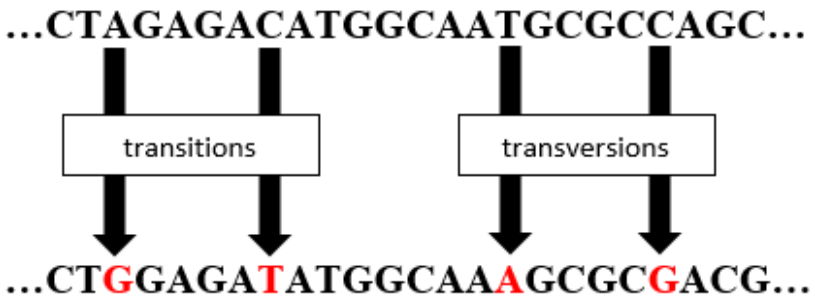


Figure 4.2: Example of SNPs, including both transitions and transversions.

Hypothetically, all individuals in a population may have an A allele at a particular locus for many generations before a mutation arises and results in an individual having a C allele at that locus.[36] It is possible that this C allele will then spread throughout the population through subsequent generations. This locus will be considered a SNP once the minor allele (i.e., the allele found less frequently, C in this hypothetical case) is found at an appreciable frequency (most often, ≥5% of the population; sometimes reduced to ≥1% depending on the sensitivity of the study). This locus can then be evaluated with the

[36] Sometimes SNPs have advantageous or deleterious effects that result in them being selected for or against (respectively); however, this is outside the scope of this dissertation. Neutral alleles, which confer no selective effects, are the focus of this dissertation.

recognition that any individual from a subsequent generation of that population may have either an A or C allele.

SNPs are the most abundant variants in the genome and are present throughout the entire genome at high density (Willing et al. 2012). Importantly, the low mutation rate of base substitutions across the nuclear genome means that the sharing of an allele at any SNP locus is usually the result of Identity By Descent (IBD); specifically, the presence of the same base at a SNP in two independent genome copies implies that that base was inherited from a common ancestor (Jobling et al. 2014).[37] Due to the sheer size of the human genome, it is unlikely that the same variant will arise independently in a different population. As such, SNPs are particularly useful for population genetic analyses.

The number of SNPs identified in the human genome is continuously expanding: in November 2000, 1.42 million SNPs were reported (The International SNP Map Working Group 2001); in 2012, the 1000 Genomes Project Consortium reported 38 million validated SNPs, a number corresponding to one polymorphic site every ~100bp. As of January 2017, build 149 of the dbSNP (Single Nucleotide Polymorphism Database, hosted by the National Center for Biotechnology Information (NCBI); see Sherry et al. 2001) has amassed over 100 million verified SNPs in the human genome, though it is plausible that dbSNP is plagued by false positives, substantially inflating this number (Mitchell et al. 2004; Musumeci et al. 2010).

While an understanding of variant frequency in the human genome is still evolving, these data suggest that while genetic variation represents a very small fraction of the total genome, there exists enough variation to explore population-level genetic

[37] IBD contrasts with Identity by State (IBS), where identical alleles are not identical by descent and do not share a common ancestor.

differences as well as individual genetic uniqueness (Kidd et al. 2004; Tishkoff and Kidd 2004).

4.4.1 Quantifying Genetic Distance Using SNPs

In addition to suggesting IBD inheritance, SNPs are commonly used in population genetics analyses because the frequency of each allele at a SNP site within a population shifts over time and across space due to evolutionary forces, including mutation, genetic drift (which increases genetic diversity), and gene flow (which reduces genetic diversity).[38]

Genetic drift is the result of random fluctuations in allele frequency between successive generations in a group. If panmixia (a situation of random breeding where all individuals are potential partners) is occurring within a group (or between two groups), genetic drift will be shared and there will be no quantifiable genetic distance between the groups; however, if a group is partially or fully isolated from another group, genetic drift will result in the eventual genetic differentiation of the groups through an increase in differences in allele frequencies.

It is critical to recognize that humans do not mate at random, but are influenced by geographic, environmental, and sociocultural factors that structure patterns of gene flow. This can result in statistically significant and quantifiable genetic distance between two populations or subpopulations through the accumulation of group-specific genetic drift. Because mating most often takes place within a localized area, it is recognized that geographically proximate groups often share more genetic drift (i.e., more similar allele

[38] In addition to the evolutionary forces of mutation, genetic drift, and gene flow discussed in this dissertation, natural selection also influences genetic diversity. However, this dissertation does not examine the phenotypic consequences of adaptive variants that would have been subject to natural selection. Therefore, natural selection is not discussed here.

frequencies) than geographically distant populations. However, in addition to geographic proximity, the sharing of genetic drift within may also be influenced by other factors, such as caste-based social stratification or selective mate choice based on phenotype. Regardless of the cause of non-random breeding, genetic drift acts to increase genetic differentiation between groups when a deviation from panmixia is occurring.

Though mating most often takes place within a localized area, a variant that arises in one population can be introduced into another population through gene flow. The evolutionary process of gene flow acts to reduce genetic differentiation between groups by spreading new alleles outside of a localized area or homogenous group. Alleles are carried by people who migrate; however, in a genetic sense, migration does not occur unless people move to a new geographic location *and* transmit the alleles they carry to the next generation. Migration, therefore, can change allele frequencies in a population when a migrant individual contributes their DNA to the next generation in a new location (Jobling et al. 2014). Gene flow must be considered as a process that adds another layer of complexity to any analysis of human genetic variation: a completely isolated population with no admixture over an extended period is an unlikely model for any human population (Feldman et al. 2003; Broushaki et al. 2016).

The genetic distance between two populations or subpopulations is therefore quantifiable, shaped by amount of similarity in allele frequency between the two groups. As such, knowledge of allele frequencies at many SNP sites can be used to quantify the genetic distance between two subpopulations or populations. The quantification of genetic distance using allele frequencies and the fixation index F_{ST} will be discussed in relation to the Kulubnarti Nubians in Chapter 7.

4.4.2 Assessing Biogeographic Affinities and Ancestries Using SNPs

While it has consistently been shown that most genetic diversity (>80%) exists within populations and not between human groups (Lewontin 1972; Barbujani et al. 1997; Jorde et al. 2000; Rosenberg et al. 2002; Jakobsson et al. 2008), the proportion of diversity that exists between human groups is still significantly greater than zero (Barbujani et al. 2013). As mentioned above, individuals living in a localized area are more likely to mate with each other and are therefore likely to share more genetic variants with each other than with people from geographically distant populations. This results in these individuals appearing to group (or "cluster") as a population in comparison to individuals with whom they share less genetic variants when enough SNP sites are examined (thousands to hundreds of thousands). Clustering can therefore provide a qualitative representation of population substructure, genetic homogeneity or heterogeneity, and genetic affinities between individuals and populations (Tishkoff and Kidd 2004).

It is recognized that while patterns of human genetic variation are largely clinal (i.e., continuous), many SNPs sites together can provide evidence of enough genetic diversity to examine humans in terms of geographically-based populations (Rosenberg et al. 2002; Li et al. 2008). This is not to say that human genetic diversity is arranged in distinct or discrete clusters: instead, human genetic variation is best described as "quasicontinuous in clinal patterns related to geography" (Tishkoff and Kidd 2004: S21). These clinal patterns related to geography can be visualized using Principal Component Analysis (PCA), first applied by Menozzi et al. (1978) and widely used in genetic studies

of human variation today (Patterson et al. 2006).[39] For example, Novembre et al. (2008) characterized genetic variation in a sample of Europeans and found a close correspondence between genetic and geographic distances, illustrated in Figure 4.3. They concluded that geographic origin can be accurately inferred using genetic data.

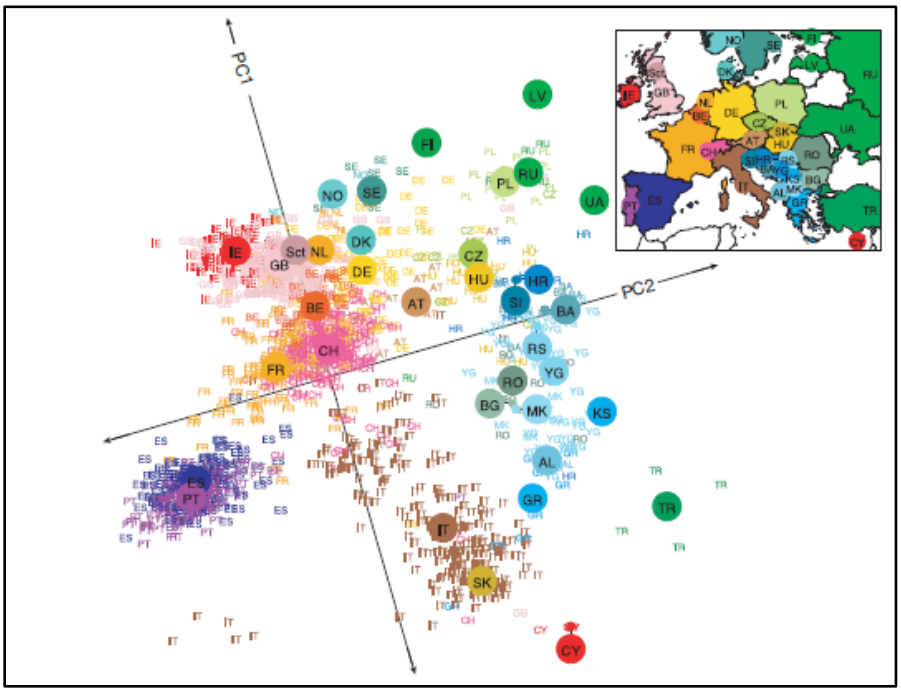


Figure 4.3: A PCA plot of genetic data from 1,387 Europeans in which "a geographical map of Europe arises naturally as an efficient two-dimensional summary of genetic variation in Europeans" (from Novembre et al. 2008: 98).

While there is correspondence between genetic and geographic distances, it is recognized that individuals are rarely classified into a single, main population cluster; instead, each individual genome is more accurately regarded as a mosaic of fragments of different origins (Bamshad et al. 2001; Barbujani et al. 2013). Thus, drawing upon thousands to hundreds of thousands of SNPs for an analysis of "ancestry" using cluster-

[39] The analysis of the Kulubnarti Nubians using PCA is discussed in greater detail in Chapters 7 and 8.

based analyses is a subtler and more complex description which takes into account the continual mixing and migration of human populations throughout history (Jorde and Wooding 2004). Both PCA and cluster-based analyses are pursued in this dissertation, described in greater detail in relation to the Kulubnarti Nubians in Chapter 8.

Ultimately, the principles of analyzing SNPs in modern and ancient DNA are the same; however, the analysis of ancient DNA is characterized by unique limitations and challenges that must be addressed before these analyses can take place.

4.5 Introduction to Ancient DNA

Ancient DNA is a field of study that analyzes DNA sequence data recovered from organisms deceased for hundreds to hundreds of thousands of years (Shapiro and Hofreiter 2013). Akin to the 17th century invention of the light microscope that provided access into the world of microbes and cells, ancient DNA acts as a window into the genetic past, enabling the direct analysis of genetic variation in long-dead organisms (Kaestle and Horsburgh 2002; Pickrell and Reich 2014).

Anthropologists recognize that the present-day inhabitants of many regions of the world have not descended linearly from the populations that lived in the same locations in the distant past (Pickrell and Reich 2014). Instead, human history is characterized by migration, admixture, and major demographic upheavals. Thus, the transformative potential of aDNA analysis lies in its ability to directly reveal the genetic make-up of individuals and populations living at archaeologically-known times and places (Pickrell and Reich 2014).

This field of study has made remarkable advances in a relatively short period of time. The first aDNA study occurred in 1984 when 229bp of mtDNA were extracted and sequenced from a quagga (*Equus quagga quagga*, a zebra-like species) museum specimen (Higuchi et al. 1984). While this study revealed that the quagga was most closely related to the plains zebra and not to the domestic horse, the most important discovery was that DNA survived for long periods of time in deceased organisms. Thirty years later in 2014, approximately 19.4Gb of quagga DNA was sequenced, representing a depth-of-coverage of ~8x and demonstrating the massive advances that have been made in aDNA research in just three decades (Jónsson et al. 2014).[40]

With recent developments, including protocols specifically tailored to the preparation of ancient DNA molecules for sequencing and the revolutionary development of NGS itself, it is argued that aDNA has "come of age" (Leonardi et al. 2016). Using NGS, which generates data from billions of aDNA fragments simultaneously and reduces the time, cost, and amount of sample material needed to produce genetic data, aDNA research has moved from the sequencing of mitochondrial genomes to a primary focus on paleogenomics (the study of genome-wide data from ancient individuals) (Orlando et al. 2015; Leonardi et al. 2016).

In addition to moving from mitochondrial DNA to genomic-level analyses, aDNA research is now embracing an important move from identifying a single well-preserved archaeological sample that yields usable DNA and obtaining a partial or complete genome sequence from it (e.g., as done in Rasmussen et al. 2010; Raghavan et al. 2014; Prüfer et al. 2014) to a more systematic approach involving hypothesis-driven sampling

[40] A "depth-of-coverage ~8x" indicates that each base pair throughout the quagga genome was sequenced an average of 8 times.

across time and space that requires the analysis of larger sample sizes (e.g., as done Allentoft et al. 2015; Haak et al. 2015; Mathieson et al. 2015, and as predicted in Pickrell and Reich 2014). Both "horizontal" (through space) and "vertical" (through time) transects that trace the spread of alleles and changing allele frequencies and connect biological adaptations to environmental or cultural transitions are becoming increasingly popular in aDNA work. These current trends in the field of ancient DNA are introducing an unprecedented level of resolution into the study of the human past (Krause and Pääbo 2016) and are significantly strengthened when integrated with anthropological data produced within a biocultural framework.

4.5.1 Ancient DNA and Anthropology

Since the mid-1980s, the study of DNA has had a significant impact on anthropological theory and practice (Matisoo Smith and Horsburgh 2012), even resulting in the creation of the new specialist subfield of molecular anthropology. Molecular anthropology emphasizes the anthropological dimensions of genetic research and focuses on the study of human history through DNA (Zuckerkandl 1963; Matisoo-Smith and Horsburgh 2012). Within the molecular anthropological approach, DNA is analyzed as one essential part of the "blueprint" of an organism that is ultimately formed through a complex interaction of genetic, epigenetic, cultural, and biological environmental factors, and human genetic variation is considered to be reflective of human behavior throughout time and across space (Matisoo-Smith and Horsburgh 2012).[41]

[41] Epigenetics is the study of heritable changes in gene function that are not changes in the DNA sequence itself. This includes changes in DNA methylation status and histone modification, which are outside of the scope of this dissertation, but nonetheless are an important part of gene expression in an organism.

From the perspective of molecular anthropology, patterns of genetic variation should not only be quantified, but also examined within a context of biological and cultural variables that influence observed patterns of diversity (Matisoo-Smith and Horsburgh 2012). Ancient DNA sequence data, though powerful, do not have any context in and of themselves. As such, while aDNA analyses contribute robust, objective genetic data regarding an individual or population's genetic history, these data should be considered alongside paleontological and bioarchaeological evidence to produce a contextualized and comprehensive evaluation of relevant hypotheses (Destro-Bisol et al. 2010). The addition of ancient DNA data to existing archaeological and bioarchaeological evidence has contributed to a clearer understanding of ancient human migration patterns (Allentoft et al. 2015; Haak et al. 2015; Raghavan et al. 2015; Skoglund et al. 2016; Martiniano et al. 2016; Cassidy et al. 2016), population demography (Skoglund et al. 2014a; Kılınç et al. 2016), population structure and admixture (Skoglund et al. 2012; Lazaridis et al. 2014; Gallego-Llorente et al 2015; Hofmanová et al. 2016), physiological and phenotypic characteristics (Fortes et al. 2013; Olalde et al. 2014; Wilde et al. 2014), and genetic signatures of selection and adaptation (Malmström et al. 2010; Rasmussen et al. 2010; Mathieson et al. 2015).

The incorporation of aDNA as a new line of evidence has also enabled anthropologists to ask new and increasingly in-depth questions regarding specific populations of interest. For example, ancient DNA has been a particularly valuable tool for the exploration of kinship and family structure (Bouwman et al. 2008; Deguilloux et al. 2014; Juras et al. 2017) as well as social organization in past populations (Bolnick and Smith 2007; Haak et al. 2008; Baca et al. 2012; Szécsényi-Nagy et al. 2015).

The integration of aDNA data with anthropological data and the interpretation of these multiple lines of evidence within a biocultural context may help anthropology to overcome some of its inherent methodological limitations (Destrol-Bisol et al. 2010). For example, it is difficult to assign sex for subadults until after the morphological changes associated with puberty occur (Cunningham et al. 2016); however, molecular sex can be determined regardless of age at death, which may aid in the analysis of sex-based patterns of morbidity and mortality in the archaeological record.

It should also be recognized, however, that the analysis of ancient DNA introduces a new suite of limitations as well as obvious benefits. At the most basic level, aDNA analysis is a destructive analysis that requires material from ancient skeletal samples that may be rare or archaeologically-valuable. Its destructive nature may preclude access to important skeletal samples in some cases.

More specifically, there are analytical limitations resulting from the degraded nature of ancient DNA. Many of the early aDNA studies of the 1980–1990s (Pääbo 1985; Pääbo et al. 1988; Golenberg et al. 1991) have now been shown to have reported modern contaminating DNA instead of authentic ancient sequences and have been disproved or effectively disregarded (Willerslev and Cooper 2005). At the turn of the millennium, rigorous controls and methodological improvements were established to ensure that aDNA data were reliable, and the outcome has been largely positive (Cooper and Poinar 2000; Hofreiter et al. 2001). As aDNA analysis is quickly becoming an integral part of anthropological research, it is necessary to address and understand the challenges that it faces.

4.6 Challenges in Ancient DNA

Ancient DNA analysis provides an invaluable window into the genetic past, but also encounters unique challenges, largely because "good biomolecular preservation is the exception rather than the rule" (Llamas et al. 2017: 2). The first challenge faced by ancient DNA analysis is that ancient skeletal samples frequently have low amounts of preserved endogenous DNA (often <1–5%). Instead, the majority of DNA originates from exogenous sources, including microbial and modern DNA contamination (Noonan et al. 2005; Der Sarkissian et al. 2014). A second challenge is that the endogenous aDNA molecules that are present have often been degraded by oxidative and hydrolytic processes. While these challenges are inherent to all aDNA research, this dissertation's analysis (and other studies that utilize skeletal material disinterred from hot environments), must contend with a third challenge: the increased rate of thermally-driven DNA degradation. Especially when conducting aDNA analyses using samples from hot areas, it is particularly important to adopt mitigating strategies and optimized protocols to recover as much DNA as possible in the face of these challenges.

4.6.1 Contamination

Contamination refers to any DNA that is derived from a source other than the target specimen (Matisoo-Smith and Horsburgh 2012). Sources of contamination are diverse and ubiquitous, affecting ancient remains at any time between an organism's death and DNA sequencing (Gilbert et al. 2005; Deguilloux et al. 2011; Pilli et al. 2013; Llamas et al. 2017). Contaminant DNA is commonly present in a much higher quantity than endogenous DNA and will be preferentially amplified by the Polymerase Chain Reaction (PCR), resulting in the availability of large amounts of contaminant DNA for

sequencing (Pääbo et al. 1989; Handt et al. 1994; Gilbert et al. 2005). The two types of contamination that can impact aDNA analyses are microbial contamination and contamination from modern human DNA.

Microbial contamination occurs due to the introduction of bacterial, fungal, and environmental DNA to the ancient sample over time. In many aDNA studies, the significant majority of DNA sequenced comes from environmental sources that colonized the sample after death (Noonan et al. 2005; Green et al. 2010; Der Sarkissian et al. 2014). It has been proposed that exogenous environmental DNA is found mostly near the surface of bones, whereas the crystal aggregates in the interior part of bones constitute endogenous DNA preservation niches (Salamon et al. 2005). Because of this, a superficial surface layer of bone is routinely removed from a bone sample before any bone material is taken for aDNA analysis, and areas of the bone with evidence of microbial activity are avoided altogether (Llamas et al. 2017).

Microbial DNA is easily separated from endogenous DNA when the reads generated during sequencing are aligned to the human genome. However, the relative amount of microbial DNA in comparison to endogenous DNA affects the amount of sequencing required to obtain robust endogenous DNA data and can have significant consequences on the budget required to sequence a sample to a meaningful level of coverage (Knapp and Hofreiter 2010).

To reduce the amount of exogenous microbial DNA prior to sequencing, many studies, including this dissertation, implement a targeted capture step, where pre-designed probes made of nucleotides that are complementary to the sequence of the DNA strands of interest are used to “capture” specific genomic regions so that only DNA molecules of

interest are sequenced (Burbano et al. 2010; Ávila-Arcos et al. 2011; Carpenter et al. 2013; Enk et al. 2014; Fehren-Schmitz et al. 2015).[42] The targets of selective capture can be mitochondrial genomes (Briggs et al. 2009; Maricic et al. 2010; Brotherton et al. 2013; Llamas et al. 2016; Posth et al. 2016); genome-wide SNPs (Lazaridis et al. 2014; Haak et al. 2015; Mathieson et al. 2015; Fu et al. 2016; Skoglund et al. 2016), or complete genomes (Carpenter et al. 2013; Enk et al. 2014; Ávila-Arcos et al. 2015). While microbial contamination of ancient specimens is unavoidable, inclusion of a capture step prior to sequencing improves the cost-effective characterization of genomic regions of interest from a target individual.

Unlike microbial DNA, modern DNA contamination will align to the human genome along with endogenous ancient human DNA sequences and can thereby alter conclusions drawn from the sequencing data. The samples at the greatest risk of modern contamination are those from ancient humans because of the challenge of differentiating ancient and modern sequences from individuals of the same species (Gilbert et al. 2005). Contaminating modern human DNA can come from researchers in the form of dead skin cells, hair, saliva, dandruff, and blood; even breathing can leave trace amounts of DNA behind (Llamas et al. 2017). In addition, laboratory consumables and reagents may not be completely sterile; contamination can occur during production in a manufacturing facility (Champlot et al. 2010; Deguilloux et al. 2011).

Techniques to reduce the amount of contaminant modern human DNA are most effective if they begin during the excavation of the material that will be used for aDNA analysis (Yang and Watt 2005). It is recommended that the archaeologists wear gloves

[42] The science behind and use of targeted capture technologies will be explained in greater detail in Chapter 6.

and a protective suit during excavation, do not wash any remains that will potentially be analyzed for ancient DNA, and place the remains in a sterile plastic bag and refrigerate them prior to shipment to an aDNA facility (Matisoo-Smith and Horsburg 2012; Pickrell and Reich 2014). Unfortunately, much of the material now being analyzed for ancient DNA was collected during past excavations that did not anticipate aDNA studies and did not take such precautions. In many cases, including in this dissertation, the skeletal material used for analysis was stored for decades and handled by hundreds of individuals. Regardless of a sample's history, stringent anti-contamination measures should be implemented at the onset of the aDNA analysis.

Anti-contamination measures include adherence to established guidelines that improve the likelihood of producing authentic aDNA results (e.g., Cooper and Poinar 2000; Hofreiter et al. 2001; Poinar 2003; Pääbo et al. 2004; Willerslev and Cooper 2005; Llamas et al. 2017). At the most basic level, anti-contamination protocols call for the set-up of an aDNA laboratory dedicated strictly to the analysis of low copy number samples that is physically removed from any post-PCR facilities (Cooper and Poinar 2000; Poinar 2003; Knapp et al. 2012a). Chemical cleaning procedures using sodium hypochlorite (bleach) and DNA degrading detergents (such as DNA-ExitusPlus) should be applied to all laboratory surfaces and equipment (Knapp et al. 2012a). Ultra-violet irradiation should be used to decontaminate laboratory equipment and consumables (Champlot et al. 2010). All personnel working within this laboratory space should adhere to a one-way rule for moving between the ancient and modern laboratories, and should wear disposable coveralls, a facemask, shoe coverings, and two layers of disposable gloves (the outer

layer being changed frequently) (Knapp et al. 2012a).[43] Negative controls should be carried throughout all steps of aDNA analysis to identify any potential instances of contamination (Llamas et al. 2017).

If possible, protocols that focus on the reduction or mitigation of contamination should also be employed during aDNA analysis. These include protocols that target the damaged aDNA fragments (Dabney et al. 2013; Gamba et al. 2016), reduce the fraction of contaminating DNA by using buffers or bleach (Malmström et al. 2007; Der Sarkissian et al. 2014; Damgaard et al. 2015; Korlevic et al. 2015; Boessenkool et al. 2016); or enzymatically treat DNA damage (Rohland et al. 2015). The specific anti-contamination measures taken prior to and during this dissertation's analysis will be described in detail in Chapter 6.

4.6.2 Ancient DNA Damage

When an organism dies, chemical damage accumulates in its DNA because the enzymatic repair processes that functioned during life are no longer occurring (Lindahl 1993). Thus, ancient DNA is invariably damaged and degraded, initially by endogenous nucleases that cleave the nucleotide chain and microorganisms that invade the remains, and subsequently by hydrolysis and oxidation reactions that fragment the DNA "backbone" and chemically modify the bases (Pääbo 1989; Lindahl 1993; Höss et al. 1996; O'Rourke et al. 2000; Brotherton et al. 2007). In addition to a fragment length usually <100bp (Sawyer et al. 2012), the most significant and pervasive form of damage in ancient DNA is the deamination of cytosine that results in a cytosine to thymine

[43] The one-way rule for movement between the ancient and modern labs applies to laboratory equipment and reagents as well.

(C→T) substitution occurring at the 5'-end of the DNA molecule (Pääbo 1989; Hofreiter et al. 2001; Briggs et al. 2007; Brotherton et al. 2007; Sawyer et al. 2012).

Cytosine deamination occurs when an amino group (NH_3) is removed from a cytosine base. Upon the removal of the amino group, cytosine (C) is converted to uracil (U) (Lindahl 1993).[44] When the DNA containing this "lesion" (a commonly used term for nucleotide misincorporation) is used as a template for PCR, the U that now exists in the template strand is read as a T, causing DNA polymerase to incorrectly incorporate an A where a G should have been on the complementary 3'-strand (Hofreiter et al. 2001). Thus, deamination ultimately results in C→T and G→A miscoding transitions on the 5'- and 3'-strands (respectively) of newly amplified DNA.

Particularly important to aDNA analysis is that the overrepresentation of C→T substitutions can be used to differentiate ancient DNA from present-day DNA sequences (Skoglund et al. 2014b). Substitutions resulting from miscoding cytosine residues are vastly overrepresented in aDNA sequences and are drastically clustered at the ends of the DNA molecule relative to the interior of the molecule, likely reflecting the presence of single-stranded DNA overhangs resulting from strand fragmentation that increase the effect of water on the chemical structure of DNA (Krause et al. 2010; Stoneking and Krause 2011; Overballe-Petersen et al. 2012). It has been shown that >20% of all cytosines in the human reference sequence are read as thymines in ancient sequences, with up to a 50-fold elevation of C→T substitutions at the 5'-most position and up to 60-fold elevation of G→A substitutions at the 3'-most position (Briggs et al. 2007).

[44] Uracil (U) is a chemical analog of thymine (T).

While sequence inaccuracy due to base misincorporation is a major issue in ancient DNA research (Brotherton et al. 2007), the sequencing of massive amounts of ancient DNA using NGS has allowed researchers to overcome this issue in two ways. First, nucleotide misincorporations can be circumvented by sequencing a DNA library to high coverage so that miscoding lesions will only occur in a minor fraction of the sequencing reads as opposed to a true SNP (Overballe-Petersen et al. 2012). Second, the diagenetic patterns characteristic of aDNA can be exploited to assist with the identification of authentic ancient DNA sequences (e.g., Krause et al. 2010; Sawyer et al. 2012). Primarily, C→T substitution frequencies are quantified using specialist software and often used to distinguish endogenous ancient fragments from modern contaminating DNA. For example, Meyer et al. (2014) stratified deamination signal (C→T substitution frequency) by fragment length in order to differentiate authentic aDNA sequences from modern contamination (shown in Figure 4.4). Likewise, this dissertation will quantify substitution frequencies in order to critically evaluate the authenticity of the DNA extracted from each Kulubnarti sample.

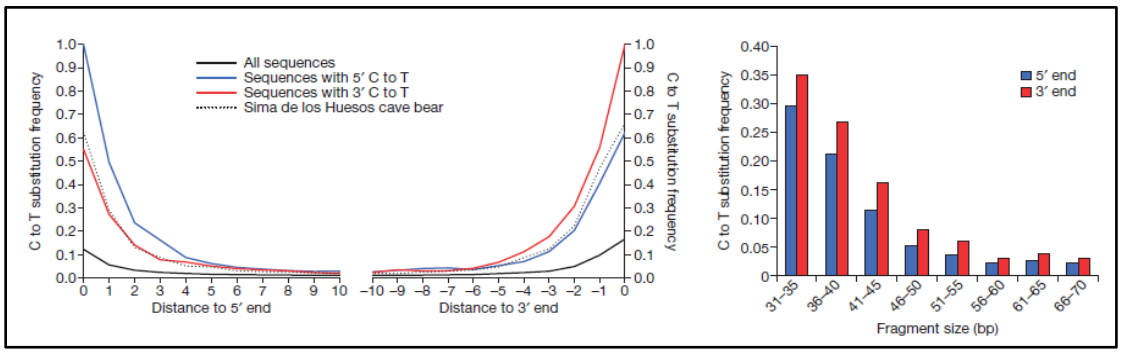


Figure 4.4: Image from Meyer et al. (2014) that shows the patterns of cytosine deamination in the libraries constructed from the Sima de los Huesos specimen (including the substitution frequency when all sequences are included (black), as well as when only sequences carrying a C→T substitution at the 5' end (blue) or 3'end (red) are considered) and the stratification of deamination signal by fragment length. These figures

demonstrate that the first and last base pair of fragments <45bp show substantially higher substitution rates than those >45bp, justifying the removal of longer DNA sequences representing modern contamination. Due to its distinctive manifestation in ancient DNA sequences, the examination of damage patterns has largely replaced the need for independent replications of sequencing results (Gansauge and Meyer 2014).

4.6.3 Environment and Temperature

A unique challenge faced by this dissertation and other aDNA studies that utilize skeletal material from sub-optimal hot environments is the increased rate of molecular degradation due to the thermally-dependent process of depurination (Lindahl and Nyberg 1972). The environment in which an ancient sample was deposited is known to be of paramount importance to the survival of biomolecules including DNA (Höss et al. 1996; Smith et al. 2003). Factors including temperature, pH, and humidity determine the rate of the chemical modifications that affect DNA *post-mortem* (Lindahl 1993, Smith et al. 2003, Willerslev et al. 2004), and it is well-established that temperature plays the most substantial role in defining the envelope of DNA survival (Smith et al. 2001; Willerslev et al. 2004). Invariably, some environments are better for DNA preservation than others: preservation is worst in environments characterized by high heat and humidity and best in permafrost environments (Hofreiter et al. 2001; Smith et al. 2003; Willerslev et al. 2004), defined as regions where soil or rock remains below 0°C throughout the year (Mitchell et al. 2005).

The preservation of DNA in an ancient sample is particularly influenced by high temperatures. High temperatures increase the rate of DNA depurination, a key mechanism of aDNA degradation that exerts a particularly powerful influence on DNA strand fragmentation and base modification (Lindahl and Nyberg 1972; Pääbo et al. 1989; Smith et al. 2001, 2003). It has been demonstrated that a decrease in temperature of 20°C

can result in a 10- to 25-fold reduction in the rate of depurination (Höss et al. 1996). As such, ancient samples recovered from hot environments will have a greater amount of highly fragmented and chemically modified DNA than samples recovered from temperate or cold environments (illustrated in Figure 4.5). For example, Bollongino et al. (2008) recovered amplifiable DNA from 67% of ancient samples from central Europe, but only from 7% of ancient samples from the Near East.

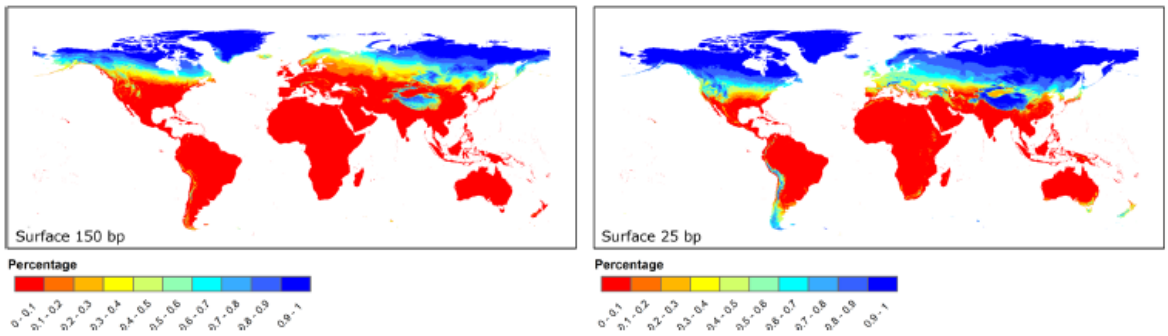


Figure 4.5: Estimated DNA survival after 10,000 years for nuclear DNA fragments 150bp (left) and 25bp (right) in length at open (non-cave) sites (Hofreiter et al. 2015). This figure illustrates the challenges faced by aDNA research utilizing samples from low latitudes and on continents such as Africa and Australia.

While progress has been made in understanding and accounting for the degradation of aDNA (e.g., Briggs et al. 2007; Ginolhac et al. 2011; Jónsson et al. 2013), there are presently no reliable criteria for determining whether samples excavated from a certain environment will yield endogenous DNA except molecular genetic analysis itself (Bollongino et al. 2008). While improvements in DNA extraction techniques (e.g., Dabney et al. (2013) which targets the recovery of DNA fragments <40bp in length) and library preparation techniques (e.g., Gansauge and Meyer (2013) which improves the sequencing of short, degraded, single-stranded DNA molecules) have increased the recovery of genetic material from samples from sub-optimal environments (Gallego-

Llorente et al. 2015, 2016; Skoglund et al. 2016, 2017), screening of many samples is still often necessary to identify the ancient samples that have sufficient quantities of preserved DNA. This is especially critical in samples from hot environments where it still remains to be seen if molecular analyses can successfully be carried out on a consistent basis. Recognizing that skeletal materials from ancient Nubia have been exposed to one of the world's hottest environments, this dissertation considers that most Nubian samples will be too thermally degraded to yield endogenous DNA and implements two screening steps to identify the best-preserved samples for hypothesis testing.[45]

4.7 Direction of this Dissertation

Though it encounters unique limitations that must be addressed and challenges that must be mitigated, aDNA research contributes new and high-resolution data that can be used to test existing hypotheses and resolve long-standing anthropological debates regarding the population history of ancient peoples. Combining aDNA and anthropological data and interpreting these as joint lines of evidence within a biocultural framework has enhanced the understanding of past population history in ways that neither genetic analysis nor anthropological research could achieve alone. This has resulted in the rapid growth of the subfield of molecular anthropology.

This dissertation sits within the trends in molecular anthropology research in several ways. First, it re-examines questions of relevance in anthropology and archaeology and employs novel technologies and methods to produce an additional line of high-resolution evidence that can be applied to resolving these questions. While

[45] These screening steps are explained in detail in Chapters 5 and 6.

existing anthropological evidence points to the co-existence of two communities at Kulubnarti who shared biological and cultural characteristics, but were socially-distinct, the genetic relationship of the two communities to each other, their genetic affinities to present-day populations, and their matrilineal and patrilineal origins have not yet been directly explored. The re-analysis of these ambiguities using aDNA data will offer new insight into the population history of Kulubnarti and, more broadly, will demonstrate the power of pursuing a genomic perspective for answering questions of anthropological importance that have so far only been explored using morphology as a proxy for genetics.

Second, this research considers newly generated genetic data jointly with the extensive body of bioculturally-framed archaeological and bioarchaeological research that explored the architectural remains, artifactual remains, and human remains at Kulubnarti. It approaches these newly generated data as reflective of human behavior as well as cultural and biological influences that shaped the actions and interactions of the Nubians.

Third, it continues the movement toward analysis of samples from sub-optimal African environments and fills a gap in aDNA research, which has not yet widely explored the genetic aspects of past populations inhabiting this region. While optimized protocols and technologies have enabled the recent generation of meaningful genetic data using highly degraded ancient skeletal material (e.g., Gallego-Llorente et al. 2015, 2016; Lazaridis et al. 2016; Skoglund et al. 2016, 2017), there is still a paucity of aDNA research conducted on samples from sub-Saharan Africa in general, and from the Nile Corridor in particular, even though skeletal remains from this region are among the most extensively studied samples in the world. This dissertation takes advantage of a large

skeletal collection disinterred from one of the world's hottest regions and uses optimized protocols and powerful technologies to generate molecular data. Recognizing the potential and challenges of ancient DNA research, the research strategy that guides this dissertation's analysis of the Kulubnarti Nubians from a genomic perspective is presented in the following chapter.

CHAPTER 5
RESEARCH STRATEGY

"Bone is more than a structural tissue; it is a potent record of past lives."
- Turner-Walker et al. (2002)

5.1 Chapter Overview

This chapter discusses the generation of genetic data from the skeletal material recovered from Kulubnarti, presenting the research strategy used to move from "bones to bytes." It begins by introducing the Kulubnarti skeletal collection, consisting of 406 individuals disinterred from Kulubnarti's two cemeteries during the 1979 expedition. It then explains the research plan of this dissertation, which was structured around optimized approaches and methodological advances that improve the outcome of ancient DNA (aDNA) analysis of thermally-degraded skeletal material.

Three aspects of this research strategy in particular were included to mitigate the effects of the poor biomolecular preservation of the Kulubnarti Nubians. First was the use of the osseous labyrinth as the exclusive biological substrate for aDNA analysis. Procuring DNA from the osseous labyrinth of the petrous bone has been shown to improve the results of aDNA analyses using thermally-degraded skeletal material that is otherwise not amenable to molecular studies. Second was the implementation of two screening steps to identify samples with the potential to generate enough data for hypothesis testing. These screening steps made it possible to avert the deeper analysis of samples that had low amounts of endogenous DNA. Third was the use of targeted enrichment and capture of genome-wide Single Nucleotide Polymorphisms (SNPs). This

capture was used to generate the data required to achieve the two research aims of this dissertation.

This chapter then revisits the two research aims, first presented in Chapter 1, and discusses the hypothesis and expected findings for each aim based on the existing research that forms the foundation of this dissertation's analysis. Finally, concluding this chapter is a discussion of the limitations associated with this work and how these limitations influenced interpretations of genetic data.

5.2 The Kulubnarti Skeletal Collection

The Kulubnarti collection contains 406 Early Christian Period individuals disinterred from the S and R cemeteries during the Colorado-Kentucky Expedition to Nubia in 1979.[46] It includes 215 individuals from the S cemetery excavated from 214 graves and 191 individuals from the R cemetery excavated from 188 graves; the skeletal remains from these individuals are currently curated at Arizona State University.[47]

Following their excavation, an assessment of age and sex was conducted by Dr. Dennis Van Gerven and his colleagues at the University of Colorado at Boulder. Developmental age at death was determined based on a seriation technique that examined inter-individual variation in multiple well-established ageing criteria, including stages of dental eruption (Ubelaker 1989; Scheuer and Black 2000), epiphyseal fusion (Buikstra and Ubelaker 1994; Scheuer and Black 2000), and age-related changes in the pubic bones (Meindl et al. 1985; Katz and Suchey 1986; Brooks and Suchey 1990). Population-

[46] More information about the cemeteries and excavation can be found in Chapter 3. Full and individualized details from the excavation of these graves can be found in Adams et al. (1999).
[47] These samples were moved from the University of Colorado at Boulder to Arizona State University in 2017 upon the retirement of Dennis Van Gerven.

specific patterns of dental attrition and skeletal degenerative changes were also considered (Adams et al. 1999). Age estimations were made for 399 individuals by arranging all individuals in a graded developmental series. Sex was determined for adults based on dimorphic skeletal features, including features of the pelvis (Phenice 1969; Klales et al. 2012), cranium (Acsádi and Nemeskéri 1970; Buikstra and Ubelaker 1994), and long bones (Spradley and Jantz 2011).[48] Residual soft tissue occasionally enabled the determination of sex in subadults.

Most individuals appeared well-preserved macroscopically because the heat and dryness of the Nubian environment encouraged the preservation of soft tissue. Examination of the collection by the author revealed that over half of the individuals had some preserved soft tissue in the form of skin, tendons, or muscles, and over one-third of the individuals still had hair. In addition to heat, the observed soft tissue preservation is due in part to the location of both Kulubnarti cemeteries away from the Nile flood. Specifically, though the Nile flooded annually, the location of the S cemetery was at least 10 meters above the level of the Nile floodplain in the Early Christian Period and had not been exposed to flooding in recent millennia. The R cemetery sample was selected from the area of highest ground in the cemetery, making it unlikely that any burials were affected by flooding (Adams et al. 1999). However, the same heat that contributed to the preservation of soft tissue increased the rate and extent of degradation of DNA in the remains. As a result, it was essential to develop a research strategy that would account for the effects of poor molecular preservation.

[48] This dissertation uses the age and sex information published in Adams et al. (1999).

5.3 From Bones to Bytes: The Research Strategy

The research strategy of this dissertation consisted of three sequential phases, outlined in Figure 5.1. Approaching this research in sequential phases represented the best strategy because the poor molecular preservation of the Kulubnarti Nubian samples required the initial identification of a subset of samples that had sufficient endogenous DNA preservation before targeted analyses aimed at data generation could take place. As such, Phase 1 of this analysis consisted of screening petrous samples from the Kulubnarti collection to identify the 30 samples that were best-preserved on a biomolecular level. Phase 2 consisted of ensuring that these 30 samples had adequate levels of authentic and uncontaminated ancient DNA, and Phase 3 consisted of generating robust genome-wide SNP data and subsequently using these data to investigate the research aims of this dissertation.

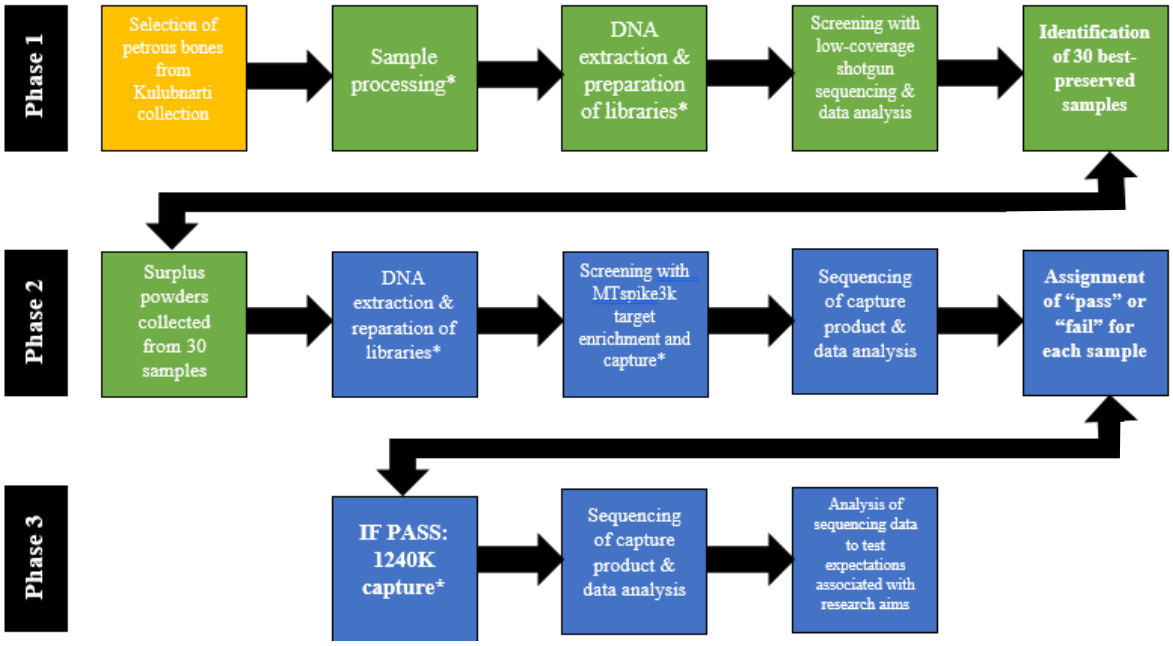


Figure 5.1: This dissertation's research strategy can be divided into three sequential phases. In this schematic, orange boxes represent steps that occurred at the University of

Colorado at Boulder, green boxes represent steps that occurred at University College Dublin, and blue boxes represent steps that occurred at Harvard Medical School; an asterisk next to the text indicates that this step occurred within an ancient DNA-dedicated clean room facility, while bold font represents a step where a decision about quality was made or a reduced number of samples was selected.

While a step-by-step description of the methods used in each phase will be provided in Chapter 6, this chapter discusses how this dissertation's strategy was structured around optimized approaches that improve the outcome of aDNA analysis of highly-degraded ancient skeletal material. Three particularly important aspects of this strategy were 1) the exclusive use of the osseous labyrinth as the biological substrate for aDNA analysis; 2) the screening of samples to identify those best-preserved on a molecular level; and 3) the use of targeted enrichment and capture of genome-wide SNPs to generate the data required to test the expectations associated with each research aim.

5.3.1 Use of the Osseous Labyrinth

Bone is a dynamic living tissue composed of collagen molecules that intertwine to form flexible fibers and hydroxyapatite crystals that impregnate the collagen matrix to stiffen it (White et al. 2012). Biomolecules, including DNA, exist within this weave of protein and mineral (Collins et al. 2002). Because DNA adheres to bone's mineralized matrix, it is better preserved in bone tissue (osseous tissue) than in soft tissue, making archaeological bone material the primary source of ancient DNA (O'Rourke et al. 2000; Salamon et al. 2005; Shapiro and Hofreiter 2014; Schuenemann et al. 2017).[49]

[49] While bone is most frequently used for aDNA analysis, additional work has focused on the use of other biological substrates, including hair (Gilbert et al. 2008; Rasmussen et al. 2010), tooth cementum (Damgaard et al. 2015; Hansen et al. 2017), and dental calculus (Weyrich et al. 2015, 2017).

Consistent with most recent aDNA analyses, bone tissue was the source of all ancient DNA analyzed in this dissertation. In particular, the petrous part of the temporal bone (*pars petrosa ossis temporalis*; shown in Figure 5.2), located at the base of the skull between the sphenoid and occipital bones, has been shown to yield up to 16-fold more endogenous DNA than molar roots or crowns and up to 183-fold more than postcranial bone (Gamba et al. 2014). It has also been shown to yield consistently higher endogenous DNA than tooth cementum, recently demonstrated to be another optimal substrate for aDNA analysis (Hansen et al. 2017). While little research has investigated why the petrous bone preserves DNA better than any postcranial element, it is plausible that its high density in comparison to other skeletal elements (Frisch et al. 1998, Lam et al. 1999) contributes to a reduced rate of bacterially- and chemically-mediated *post-mortem* DNA decay and consequently contributes to superior DNA preservation (Gamba et al. 2014).

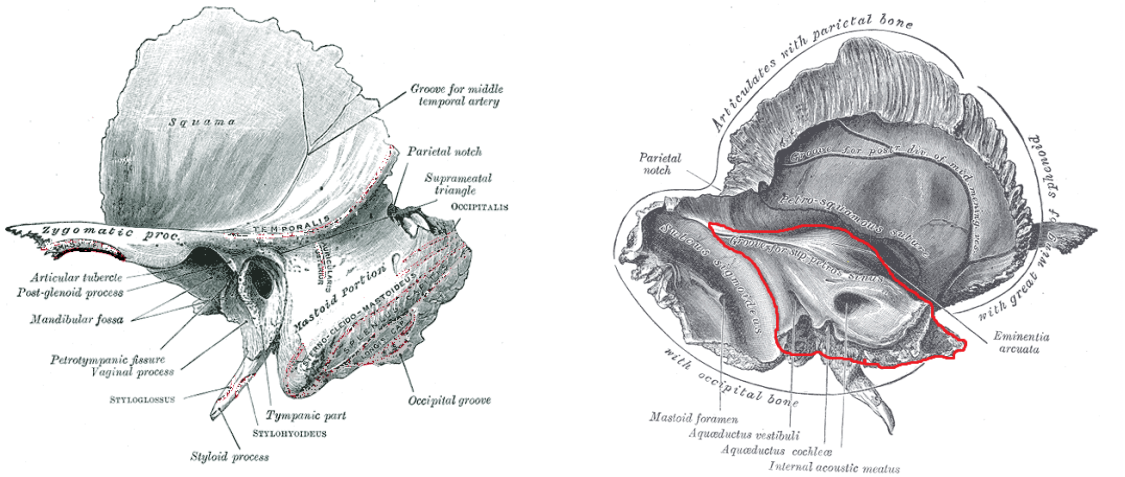


Figure 5.2: Ectocranial view of the left temporal bone (left) and endocranial view of left temporal bone with the petrous part circled in red (right). Images from Gray's Anatomy of the Human Body (Gray 1918), Plates 137 (left) and 138 (right); accessed through Wikimedia Commons.

Building upon the research that revealed superior DNA preservation in the petrous bone in comparison to postcranial bones and teeth, a subsequent intra-petrous comparison determined the osseous labyrinth, or the bony part of the internal ear that houses the organs of hearing and balance, to be the most DNA-dense region of the petrous bone (Pinhasi et al. 2015).[50] As illustrated in Figure 5.3, the osseous labyrinth is contained completely within the petrous bone and is composed of the cochlea, vestibule, and three semi-circular canals.

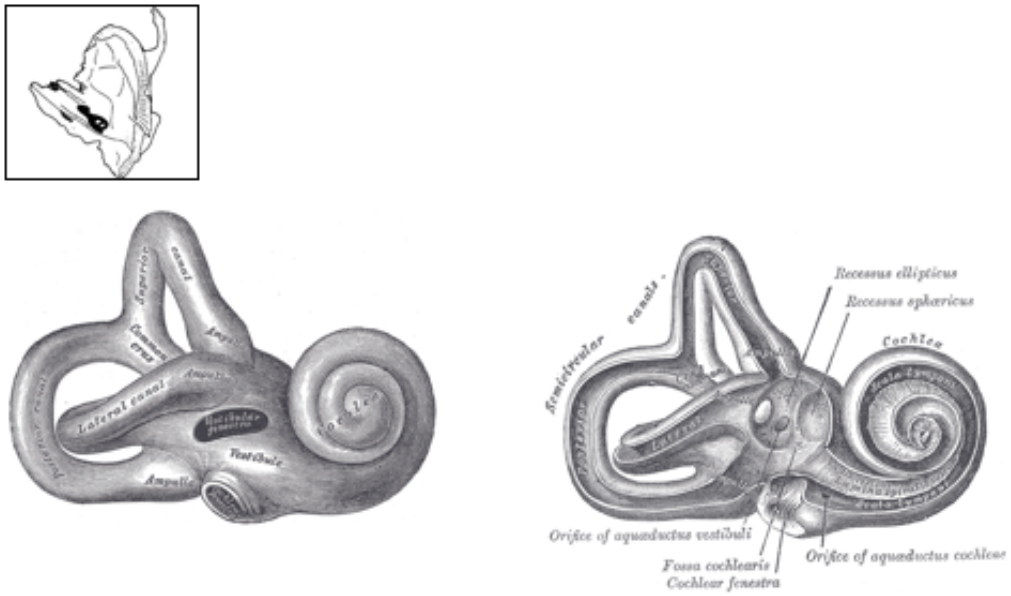


Figure 5.3: External (left) and internal view (right) of the osseous labyrinth. Images from Gray's Anatomy of the Human Body (Gray 1918), Plates 920 (left) and 921 (right); accessed through Wikimedia Commons. Image in box shows osseous labyrinth within petrous bone, adapted from Doden and Halves (1984).

Specifically, when compared to trabecular (spongy) bone from the apex of the petrous and cortical (compact) bone surrounding the osseous labyrinth (shown in Figure

[50] The osseous labyrinth is also referred to as the otic capsule, bony labyrinth, or osseous inner ear.

5.4), the osseous labyrinth yields up to 177-fold and 65-fold more endogenous DNA, respectively (Pinhasi et al. 2015).

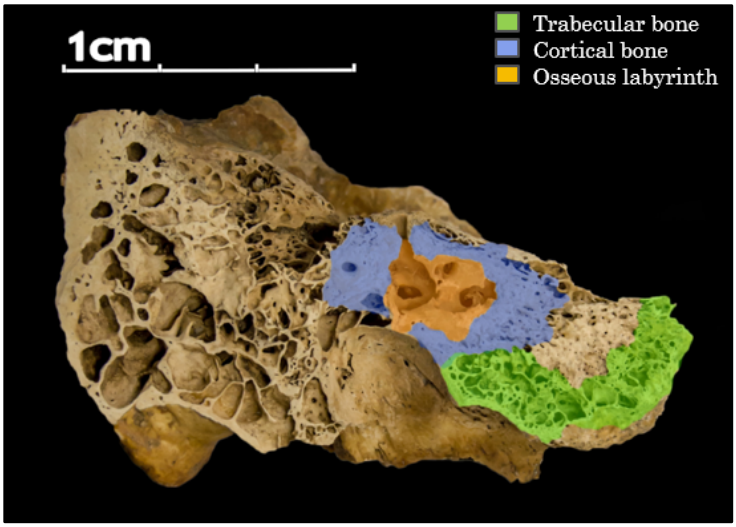


Figure 5.4: Illustration of the three regions of the petrous compared in Pinhasi et al. (2015). The green region identifies the spongy bone at the apex of the petrous; the blue region identifies the cortical bone encircling the osseous labyrinth; the orange region identifies the osseous labyrinth itself. Figure adapted from Pinhasi et al. (2015). Image is licensed under Creative Commons Attribution License.

The reason for the superior preservation of endogenous DNA in the osseous labyrinth in comparison to other regions of the petrous also remains unclear. However, the osseous labyrinth experiences a near-absence of growth, modeling, or remodeling following the completion of ossification by the sixth month *in utero* (Sørensen et al. 1992; Frisch et al. 1998). It is possible that the inhibition in remodeling within the osseous labyrinth and the consequential maintenance of immature woven bone that has a relatively higher proportion of osteocytes (cells responsible for maintenance of bone tissue) contributes to the superior preservation of DNA (Hernandez et al. 2004; White et al. 2012). While the relationship between osteocyte density and DNA preservation has not yet been explored in detail, it is probable that the osteocyte population in woven

bone, estimated to be 4–8 times that of mature lamellar bone (Hernandez et al. 2004), may enhance the preservation of endogenous DNA in the osseous labyrinth and make it an ideal substrate for aDNA analysis.

As the skeletal collection from Kulubnarti was disinterred from one of the hottest environments in the world, it is expected that the biological material will be highly degraded on a molecular level due to an increased rate of DNA depurination (Lindahl and Nyberg 1972; Pääbo et al. 1989; Lindahl 1993; Smith et al. 2001, 2003).[51] Therefore, based on its ability to preserve higher quantities of endogenous DNA than any other skeletal element, the osseous labyrinth of the petrous bone was the exclusive source of skeletal material for this dissertation's analysis.

5.3.2 Screening

The second aspect of the research strategy aimed at improving the outcome of aDNA analysis of highly-degraded ancient skeletal material was the implementation of screening steps that indicated the level of biomolecular preservation of each sample. Recognizing that not all skeletal samples from a geographic region or even within a single archaeological site will have an equivalent level of biomolecular preservation, screening is common in aDNA studies to separate a subset of samples that have nearly no preserved authentic ancient DNA from a subset of samples that have enough preserved authentic ancient DNA to allow population genetic analyses (e.g., Fernández et al. 2014; Haak et al. 2015; Mathieson et al. 2015). Screening becomes particularly important in studies that utilize skeletal material expected to have poor biomolecular preservation (e.g., Lazaridis et al. 2016; Schuenemann et al. 2017; Skoglund et al. 2017).

[51] A detailed description of thermally-driven DNA degradation is presented in Chapter 4.

Pilot research conducted by the author in 2013 and 2014 indicated that only a small subset (~15–25%) of the Kulubnarti skeletal material analysis would have enough preserved endogenous aDNA to generate the data required to test the expectations associated with this dissertation's aims. Based on this pilot research, approximately 120–200 samples would need to be collected in order to identify 30 samples that would be amenable to robust data generation. The best-preserved samples were then identified using two screening steps.

In this dissertation and in most other aDNA studies, low-coverage shotgun sequencing (the sequencing of random short fragments of DNA that are re-assembled upon alignment by using regions of overlap) was used as an initial screening step to avert the deeper analysis (Next-Generation Sequencing (NGS) and targeted enrichment and capture) of samples that had low amounts of preserved endogenous DNA. This dissertation is unique in its inclusion of an additional screening step that used a small capture of ~3,000 genome-wide SNPs to examine ancient DNA authenticity.[52] The threshold for designating a sample as well-preserved enough for targeted analysis will be discussed in Chapter 6.

The desired sample size of 15 individuals from each cemetery (N=30 total) was based on the same pilot research, which suggested that the genome-wide analysis of a larger number of samples was not economically practical due to their generally poor state of biomolecular preservation. The selected sample size was also supported by additional research demonstrating that population-level investigations (such as genetic distance) can be determined just as accurately with a small sample size (N=4–6) as with a higher

[52] This screening capture is discussed in detail in Chapter 6.

number of samples when a large number of genetic markers (>1,000) or whole-genome data are used (Patterson et al. 2012; Willing et al. 2012). As the methods of this dissertation's analysis were designed to produce genome-wide data by targeting SNPs throughout the human genome that are most powerful for population genetic analysis (Mathieson et al. 2015), a sample size of 15 individuals per community provided a balance between research goals and economic feasibility without sacrificing analytical power (per Dr. David Reich, pers. comm.).

5.3.3 Targeted Enrichment and SNP Capture

The third aspect of the research strategy aimed at improving the outcome of aDNA analysis of highly-degraded ancient skeletal material was the use of targeted enrichment and capture of genome-wide SNPs. Due to a low proportion of endogenous DNA relative to exogenous DNA, the implementation of a targeted enrichment and capture step was essential for the generation of adequate amounts of sequencing data for population genetic analysis.[53] In this dissertation, the specific focus was on the enrichment and capture of select genome-wide SNPs known to be most powerful for population genetics analysis (Mathieson et al. 2015).

In contrast to screening by shotgun sequencing, which sequences all DNA present in a metagenomic DNA library, targeted enrichment and capture methods narrow the sequencing focus by hybridizing synthesized "bait" probes specifically designed to target user-selected genomic regions of interest to the molecules in the library that contain those regions. The hybridization of DNA molecules and bait probes occurs either in solution (Enk et al. 2014; Haak et al. 2015; Fu et al. 2016) or on arrays (Burbano et al. 2010;

[53] This target enrichment and capture step will be discussed in detail in Chapter 6.

Fortes and Paijmans 2015) and facilitates the removal of non-target DNA molecules prior to sequencing.

In-solution capture, used in this dissertation, is suggested to be the most effective method for capturing the short DNA molecules present in aDNA libraries (Gnirke et al. 2009). In this method, a reaction is assembled that combines DNA molecules from an amplified library with biotinylated bait probes representing the target sequences (Horn et al. 2012). During a period of incubation, the bait probes hybridize with the target DNA molecules that contain these sequences. Magnetic streptavidin-coated beads then recover the probes, and thereby "capture" the target DNA sequences; non-target DNA is washed away in repeated wash steps.

Because low levels of endogenous DNA made whole genome sequencing of the Kulubnarti Nubians uneconomical, this dissertation obtained all genome-wide data by enriching ancient samples for genome-wide SNPs that had already been genotyped on large panels of present-day individuals representing populations across the world (Mathieson et al. 2015). Ultimately, the analytical strategy described here was optimized to generate the requisite data to test the expectations associated with the two research aims of this dissertation, presented in the next two sections.

5.4 Aim 1: Analyzing Genetic Relationships at Kulubnarti

Aim 1 of this dissertation is to explore the genetic relationship between the S and R communities by looking for evidence of community-based genetic population substructure through patterns of clustering and estimation of genetic distance.

The null hypothesis (H_0) is that there was no community-based genetic population substructure at Kulubnarti. H_0 would fail to be rejected if individuals from both communities were plotted together as a single, distinct cluster when analyzed with Principal Component Analysis (PCA) and projected onto a background dataset of select present-day individuals and if no significant genetic distance was estimated between the communities with the fixation index statistic F_{ST}. The alternative hypothesis (H_a) is that there was community-based genetic population substructure and significant genetic distance between the S and R communities. H_a would be supported if individuals from each community formed discrete and separate community-based clusters when projected onto a background dataset of present-day individuals in PCA space, and if significant genetic distance was estimated between the communities with the fixation index statistic F_{ST}. The analytical methods of PCA and F_{ST} and their specific application to this research will be described in detail in Chapter 7.

Expectation 1 is that H_0 will not be rejected, and that PCA plots and F_{ST} estimates will suggest that there was no genetic population substructure at Kulubnarti.

Expectation 1 is based on archaeological and bioarchaeological evidence that describes Kulubnarti as home to two neighboring communities with socioeconomic differences that influenced their well-being (Van Gerven et al. 1995; Adams et al. 1999; Adams and Adams 2006), but also suggests that these two communities were a single biological population, defined as a "group of interbreeding individuals that exist together in time and space" (Hedrick 2005: 62-63). Specifically, archaeological excavation of the S and R cemeteries that served the small population of Kulubnarti revealed similarities in

grave types, grave goods, and skeletal types found in both cemeteries (Adams et al. 1999), while bioarchaeological analyses that used morphology as a proxy for genetic variation (e.g., Greene 1982; Van Gerven 1982) found no evidence of biological distance between the Kulubnarti communities. If Expectation 1 is supported and no community-based genetic population substructure is found, it would suggest that membership in one community or the other was a social phenomenon with no biological basis and that the S and R communities can be interpreted as socially-distinct groups comprising a single genetic population.

The finding of a single genetic population at Kulubnarti would require that there was not enough community-specific genetic drift to result in population substructure. One possible explanation for a low level of community-specific genetic drift is that inter-community gene flow maintained non-significant genetic distance between the two communities.[54] That some amount of gene flow occurred between the two communities is likely, because while the R and S communities were relatively isolated from outside populations due to their location in the desolate environment of the *Batn el Hajar* (Van Gerven et al. 1995), they were not naturally isolated from one another in a way that would restrict or impede gene flow.[55] As such, the most parsimonious explanation would be that there was no genetically-discernible community-based population substructure at Kulubnarti.

---

[54] Gene flow and other evolutionary processes influencing genetic distance are described in Chapter 4. While inter-community gene flow is one possible explanation for a low genetic distance estimate between the S and R communities, a finding of low genetic distance does not necessarily imply that gene flow was occurring.

[55] It should be kept in mind that Kulubnarti Island was only a true island at the peak of the Nile flood until the filling of Lake Nubia with the construction of the High Aswan Dam (Adams 1977; Adams and Adams 1998; Adams et al. 1999).

However, placing parsimony aside, an expectation of unrestricted bidirectional gene flow at Kulubnarti can be challenged. While there is no evidence that natural barriers to gene flow existed at Kulubnarti, social factors also have the potential to create population substructure (Bamshad et al. 1998; Van Leeuwen and Maas 2005; Mielke et al. 2006; Marchani et al. 2008; Laland 2010; Relethford 2012). In many socially-stratified populations around the world (e.g., Chagnon 1970; Relethford et al. 1983; Majumder 2010), the greatest amount of genetic exchange is observed among individuals within a subpopulation, while the least amount of genetic exchange is observed between subpopulations. As the population of Kulubnarti was consistently small in size due to the subsistence challenges created by the hostile environment of the *Batn el Hajar* (Van Gerven et al. 1995), even a small amount of reproductive isolation could have resulted in the accumulation of community-specific genetic drift that would be detected during analyses of population substructure and genetic distance.

As such, socially-imposed reproductive isolation of either community may have resulted in the genetic differentiation of the two communities. Though it is not expected, it is possible that significant genetic distance between the communities will be detected in this dissertation's analysis, suggesting the presence of population substructure at Kulubnarti.

5.5 Aim 2: Characterizing Genetic Composition at Kulubnarti

Aim 2 of this dissertation is to characterize the genetic composition of individuals from both communities and reveal any community-based differences by exploring biogeographic genetic affinities and components of ancestry and

determining the mitochondrial (mtDNA) and Y chromosome haplogroup profiles of each community.

The null hypothesis (H_0) is that there were no community-based differences between the S and R communities in biogeographic genetic affinities (or just "biogeographic affinities"), ancestry components, or mtDNA or Y chromosome haplogroup profiles. H_0 would fail to be rejected if individuals from both communities showed the most affinity to the same present-day populations based on PCA as evidenced by their projection near to those present-day populations in PCA space, if individuals from both communities showed the same ancestry components in similar proportions as evidenced by ADMIXTURE analysis, and if both communities showed similar signatures of matrilineal and patrilineal ancestries as evidenced by mtDNA and Y haplogroup profiles.

The alternative hypothesis (H_a) is that the S and R communities exhibit community-based differences in genetic affinities, ancestry components, and mtDNA and Y haplogroup profiles. H_a would be supported if individuals from both communities showed affinity to different present-day populations as evidenced by their projection near to those populations in PCA space, if individuals from both communities were determined to have different ancestry components or the same components in different proportions as evidenced by ADMIXTURE analysis, or each community showed different signatures of matrilineal and patrilineal ancestries as evidenced by mtDNA and Y haplogroup profiles. The analytical methods of PCA, ADMIXTURE analysis, and mtDNA and Y haplogroup determination as well as their specific application to this research will be described in detail in Chapter 8.

Expectation 2 is that H_0 will not be rejected, and that no community-level differences in biogeographic genetic affinities, ancestry components, or mtDNA or Y haplogroup profile will be detected between the S and R communities.

As with Expectation 1, Expectation 2 is based on archaeological evidence which suggests that it is most likely that the relatively more impoverished individuals of the S community were part of an autochthonous Nubian population along with the landowning members of the R community based on the similarity of burial practices and grave goods observed in the S and R cemeteries (Adams et al. 1999), as well as on morphological evidence which shows no evidence for community-based phenotypic differentiation between the S and R communities (Greene 1982; Van Gerven 1982). Based on these data, it is expected that the S and R communities will be assessed as having similar genetic affinities and components of ancestry as well as similar mtDNA and Y haplogroup profiles.

It is possible, however, that some aspects of Expectation 2 will be supported by the data while other aspects will not, as autosomal DNA, mtDNA, and the Y chromosome are investigated as part of this research aim. As such, Expectation 2 is separated further into Expectation 2a (biogeographic genetic affinity and ancestry components), Expectation 2b (mtDNA haplogroup profiles), and Expectation 2c (Y chromosome haplogroup profiles).

Expectation 2a is that the both communities from Kulubnarti will show more biogeographic genetic affinity to present-day genotyped populations from Northeastern Africa (specifically, those Eastern African populations inhabiting the Horn of Africa) than any other world region, and will be estimated as having

components of ancestry associated with present-day populations from Northeastern Africa.[56] It is also expected that the Kulubnarti Nubians will be genetically distinguishable from all present-day genotyped populations, indicating that they comprise a population that is no longer present in unadmixed form.

Expectation 2a reflects both the expected uniqueness of the Kulubnarti Nubian genetic make-up as well as the challenges associated with assessing their biogeographic genetic affinities and components of ancestry in relation to present-day populations.[57] Consistent with the biocultural perspective's avoidance of migrationist explanations, it is expected that the Kulubnarti Nubians will have Northeastern African roots based on evidence of nearly-continuous human occupation of Nubia since the Paleolithic (Arkell 1961; Adams 1977) and archaeological and anthropological data suggesting that the same population has occupied this region at least intermittently from the Neolithic through the present day (e.g. Mukherjee et al. 1955; Adams 1966; Van Gerven et al. 1973; Carlson 1976; Carlson and Van Gerven 1979; Rudney and Greene 1982; Adams 1998, but see Irish 2005 for an alternate perspective). Therefore, a component of ancestry associated with Eastern Africa in general, and Northeastern Africa in particular, is expected to be represented in the Kulubnarti Nubians.

In addition to a Northeastern African component, it is also expected that there will be some Northern African (possibly Egyptian) and other Eurasian (most likely Near Eastern, defined in this dissertation as comprising the Arabian Peninsula, Iraq, Iran, Israel, Jordan, Lebanon, Palestine, Cyprus, Turkey, and Syria as in Pagani et al. 2015; see

[56] Egypt is not considered part of Northeastern Africa specifically as it is used in this dissertation, but instead is considered to be part of Northern Africa.
[57] These challenges are explained in detail in Chapter 8.

Supplement 5) components represented in the Kulubnarti Nubians based on their location on the Nile Corridor. However, these foreign components plausibly constitute a minor part of the overall genetic make-up of the Kulubnarti population, because while Nubia is located along the historically-important migration corridor of the Nile River, interactions between local Nubians and migrating peoples varied across space and time in Nubia (Edwards 2004).

Specifically, evidence for extensive inter-population interactions prior to the Early Christian Period is scarce, especially in the isolated *Batn el Hajar*, while further evidence suggests that such interactions may also have been limited during the Early Christian Period (Adams 1977). However, as described in Chapter 3, there is some evidence of foreign contact based on the introduction of Christianity from Constantinople in the mid-6th century C.E. and the evidence of sustained contact with the Arab world, which took place through Islamic invasions and the establishment of the *Baqt* treaty in the mid-7th century C.E (Hasan 1967). Thus, part of Expectation 2a is that the gene pool of both Kulubnarti communities will be mainly comprised of Eastern African components with some lesser influence from Northern Africa and Eurasia.

Based on the initial part of Expectation 2a, it is expected that the Kulubnarti Nubians will share more biogeographic genetic affinity with genotyped populations who presently inhabit Northeastern Africa than with genotyped present-day populations from any other region.[58] However, it is also expected that the Kulubnarti Nubians will not appear genetically identical to any present-day Northeastern African populations. This is

[58] It is important to recognize that biogeographic affinity can only be analyzed using present-day populations that have been genotyped at a genome-wide level. At present, genotyped populations from Northeastern Africa are sparse, limiting the resolution of this analysis. This caveat is further discussed in Chapter 8.

primarily due to the influence of demographic processes, including gene flow and genetic drift, that have occurred between the Early Christian Period and the present-day and shaped the present human genetic landscape.

In the ~1400 years since the Early Christian Period in Nubia, there have been a large number of events that had the potential to influence the present-day human genetic landscape. For example, Lower Nubia became a free-trade zone between Christian Nubia and Muslim Egypt where Egyptian merchants were allowed to trade and settle after the 9th century C.E.; the *Batn el Hajar* became a major refuge for northern Late Christian Period populations fleeing disturbances caused by warring principalities; Mameluke armies, comprised of Turkish and Christian soldiers from Southwestern Asia, established rule in Egypt and eventually gained control of Lower Nubia and the *Batn el Hajar*; extensive Arab migrations catalyzed conversion from Christianity to the Islamic faith throughout Nubia; the Ottoman Turks extended their dominion as far upriver as the Third Cataract after their subjugation of Egypt; the sub-Saharan caravan trade opened and expanded, creating new and consistent links with Northwestern Africa; the Egyptian-controlled sub-Saharan African slave trade, which reached its climax in the 19th century, used the Nile as a route of contact and transportation between Eurasia and the regions surrounding the headwaters of the White Nile and the lake region of Eastern sub-Saharan Africa; and Red Sea trading networks grew, leading to the development of major trade routes across the Eastern Desert and linking the Red Sea coast with Egypt and the Nubian Middle Nile (Hasan 1967; Adams 1977; Shinnie 1996; Welsby 2002; Edwards 2004, 2013).

These events inevitably had substantial effects on the present-day genetic composition of Northeastern Africans as evidenced by the extensive admixture that characterizes present-day populations, such as those in Ethiopia, Somalia, and the Sudan (Passarino et al. 1998; Non et al. 2011; Hollfelder et al. 2017). The caveats associated with examining the genetic affinities and components of ancestry of the Early Christian Period Kulubnarti Nubians using the present-day genetic landscape are addressed in Chapter 8.

Expectation 2b is that the majority of matrilineal lineages of members of both Kulubnarti communities will be primarily of Northeastern African origin and distribution. It is also expected that there will be no community-based differences in mtDNA haplogroup profile.

mtDNA is a powerful tool for exploring human population history and revealing female-mediated migration throughout the world (Brown et al. 1979; Giles et al. 1980; Cann et al. 1987; Ballard and Whitlock 2004); the assignment of mtDNA haplogroups and their implications for the individuals comprising both Kulubnarti communities is presented in detail in Chapter 8.

Importantly, while the African mtDNA landscape was largely shaped by processes occurring long before the Early Christian Period, it has also been changed since the Early Christian Period due to intensifying migration and admixture, as described above in relation to biogeographic genetic affinities determined by autosomal DNA.

As discussed previously, there is some evidence of foreign peoples moving from Eurasia, and particularly from the Near East, into or through Nubia prior to and during the Early Christian Period who could have introduced non-local mtDNA haplogroups that

would be detected in this analysis. However, evidence for contact at the specific site of Kulubnarti is sparse, likely due to its location in the desolate *Batn el Hajar*. Additionally, foreign peoples who did move through the Nubian region (for example, those associated with the spread of Christianity) were likely to be missionaries, soldiers, or merchants, and consequently, more likely to be male than female. As mtDNA is inherited through the female lineage only, foreign males would not have introduced their non-local haplogroup into the autochthonous population, even if genetic exchange with local populations did occur.

While the geographic distribution of mtDNA haplogroups is discussed in greater detail in Chapter 8 and Supplement 8, it is expected that the majority of Kulubnarti Nubians will be assigned mtDNA haplogroups of the African-specific L lineage, and more specifically to branches that originated in Eastern Africa and are found at high frequencies throughout Northeastern Africa today (e.g., branches from the L0 lineage). It is also possible that some members from Kulubnarti will be assigned to a specific haplogroup, M1, presently found at high frequencies in Northeastern Africa as the result of a back-migration from Eurasia into Northeastern Africa that predated the Early Christian Period and is (Passarino et al. 1998; Stevanovitch et al. 2004; González et al. 2007; Schuenemann et al. 2017). This expectation is largely based on the relative isolation of the site of Kulubnarti in the *Batn el Hajar*, which may have protected it from sustained infiltration by or settlement of foreign peoples who would have introduced non-local mtDNA haplogroups.

Though foreign peoples moving along the Nile were most often recorded as being male, it has been demonstrated by genetic studies that there may have been some female-

mediated migration between the Near East and Egypt and sub-Saharan Eastern Africa along the Nile corridor as well (Brace et al. 1993; Fox 1997; Krings et al. 1999; Macaulay et al. 1999; Richards et al. 2000, 2003). Therefore, it is possible that some non-local contribution to the mtDNA haplogroup profile of each community will be detected; for example, it is possible that haplogroups from J or N lineages of Eurasian origin may be detected at Kulubnarti. However, it is not expected that these Eurasian haplogroups would make a substantial contribution to the mtDNA haplogroup profile of either community. It is further possible that even if there was foreign contact between migrating females and the resident Kulubnarti Nubians, these non-local females would have avoided long-term settlement in the *Batn el Hajar* in favor of more desirable regions, such as the fertile Dongola Reach. Additionally, if there was sustained contact or settlement, there is no existing evidence that suggests it would have occurred in a community-based pattern.

Ultimately, it is expected that the haplogroups to which the Kulubnarti Nubians of both the S and R communities are assigned will not suggest a substantial genetic signature of female-mediated gene flow into Kulubnarti from Eurasia. Broadly, it is expected that the Kulubnarti Nubians will show less haplogroup diversity than present-day populations of Northeastern Africa, such as Sudanese, Ethiopians, and Somalis, who exhibit haplogroups of both African and Eurasian origin resulting from increasing amounts of migration and admixture that most likely took place after the Early Christian Period (Passarino et al. 1998; Krings et al. 1999; Fadhlaoui-Zid et al. 2011; Mikkelsen et al. 2012; Hollfelder et al. 2017).

Expectation 2c is that the majority of patrilineal lineages of male members of both Kulubnarti communities will be primarily of Northeastern African origin and distribution. It is also expected that there will be no community-based differences in Y chromosome haplogroup profile.

Complementary to mtDNA, which elucidates the genetic past of the maternal lineage, worldwide patterns of male genetic variation on the Y chromosome are used to explore patrilineal origins and male-mediated migrations, making the Y chromosome another powerful tool for phylogeographic analyses (Underhill et al. 2000; Jobling and Tyler-Smith 2003; Hassan et al. 2008). The assignment of Y chromosome haplogroups and their implications for the males comprising both Kulubnarti communities is presented in detail in Chapter 8.

Similar to the mtDNA landscape, the African Y chromosome landscape was largely shaped by processes occurring long before the Early Christian Period but has also been changed since the Early Christian Period due to intensifying male-mediated migration and admixture.

While the geographic distribution of Y haplogroups is discussed in greater detail in Chapter 8 as well as in Supplement 9, it is expected that the majority of Kulubnarti Nubian males will be assigned to Y chromosome haplogroups within the African-specific A and E lineages, which have origins in sub-Saharan Africa and have been found at high frequency in present-day populations from Northeastern Africa; both haplogroups are essentially restricted to the African continent (Underhill et al. 2000; Semino et al. 2002; Jobling and Tyler-Smith 2003; Arredi et al. 2004; Hassan et al. 2008; Karafet et al. 2008; Cruciani et al. 2011). The assignment of these haplogroups would suggest “local” African

origins for the paternal lineage of the Kulubnarti Nubians. This expectation is largely based on the isolated location of Kulubnarti within the *Batn el Hajar*, a natural environment which has shown to deter infiltration by foreign peoples.

However, is it also possible that there will be more variation in Y chromosome haplogroups than mtDNA haplogroups detected at Kulubnarti. This is because there is well-documented evidence of at least some extra-regional contact throughout Nubia during the Early Christian Period, including with Christian missionaries and Arab invaders, the overwhelming majority of whom were likely males. As the Christian burial style is seen at Kulubnarti, it is evident that the Christian religion was spread to this isolated hamlet in the *Batn el Hajar*. Though it is unclear if it was foreign peoples who introduced Christianity, it is possible that contact with foreign males carrying this religion may have resulted in the introduction of extra-regional Y haplogroups to the Kulubnarti population. For example, haplogroups from the J lineage, found at high frequencies in both the Near East and Northern Africa (Hassan et al. 2008; Abu-Amero et al. 2009) may be detected at low frequencies at Kulubnarti, though it is not expected that these extra-regional haplogroups would be present at Kulubnarti in any community-based pattern or that they would make a substantial contribution to the Y haplogroup profile of either community.

Ultimately, it is expected that the Y haplogroups to which the Kulubnarti Nubians of both the S and R communities are assigned will not suggest a substantial non-local patrilineal genetic signature at Kulubnarti. As with mtDNA haplogroups, it is likely that the Kulubnarti Nubians will show less diversity than present-day populations of Northeastern Africa, such as Sudanese and Ethiopian populations, who have experienced

additional and intensifying migration and admixture with foreign peoples since the Early Christian Period (Semino et al. 2002; Hassan et al. 2008; Hollfelder et al. 2017).

5.6 Limitations

All scientific investigations have a degree of limitation, and it is the task of the author to define these limitations and acknowledge how they may influence the results and conclusions drawn from their data. In addition to encountering limitations associated with aDNA analysis (e.g., thermally-driven degradation, discussed in Chapter 4), this dissertation recognizes two theoretical limitations present in this study: the representativeness of the samples analyzed, and an inability to determine contemporaneity.[59] All conclusions made from the data generated in this dissertation will acknowledge these limitations.

5.6.1 Representativeness

A major concern of both bioarchaeological and aDNA studies is that it is often impossible to know how well a skeletal sample represented the total living population at a site (Howell 1982; Jackes 2011). Before engaging in the interpretation of genetic data at a community- or population-level, the representativeness of the Kulubnarti sample must be considered.

The aim of the 1979 Joint Colorado-Kentucky Nubian expedition was to recover well-preserved burials that included preserved soft tissue for anthropological analyses; therefore, graves were not selected with the intention of being representative of the

[59] Limitations regarding the interpretation of patterns of genetic variation in ancient populations using present-day populations as a reference are presented in detail in Chapter 8.

cemetery population as a whole (Adams et al. 1999). Adams et al. (1999) note that the R cemetery in particular cannot be considered to be a representative sample of the cemetery as a whole because all but six graves excavated in 1979 were concentrated in one contiguous area at the far western end of the cemetery. Armed with this knowledge, it should be recognized that any conclusions regarding the genetic composition of the S and R communities or the singular Kulubnarti population generated from the subset of the Kulubnarti collection analyzed in this dissertation (which itself was a subset of the complete population) may not accurately reflect the genetic composition of either Kulubnarti community or of the population as a whole.

Additional biases may be introduced by unknown factors, such as an exclusion of some individuals from communal burial locations for one reason or another (e.g., Konigsberg 1985; Hadley 2010). While there is no evidence to support selective exclusion from communal burial grounds at Kulubnarti, it remains a possibility that cannot be ruled out.

Characterizing the genetic composition of a community using genetic data from a small number of individuals can be tenuous, though this dissertation utilizes genome-wide data to effectively analyze a much larger effective sample comprising thousands of ancestors for each individual. While this dissertation will consider each individual in the selected sample to contribute to both an individual and a community-level understanding of genetic composition, it is recognized that each individual's genetic profile is, at the most basic level, only representative of that individual.

While aDNA analysis of additional samples would increase the resolution of community-level genetic profiles, the biomolecular degradation of Nubian skeletal

material limits the number of samples suitable for the economically-practical testing of hypotheses requiring genome-wide data. Questions of representativeness will always be present in the analysis of prehistoric skeletal remains and should not deter the aDNA study of the Kulubnarti Nubians, but instead should be "…borne graciously and realized analytically" (Johnston 1962: 249) and should be recognized in all interpretations made using the data.

5.6.2 Contemporaneity

Because both the R and S cemetery samples span approximately 250 years based on the radiocarbon dates obtained from 21 samples from both cemeteries (presented in Chapter 3), it must be recognized that the individuals comprising both the Kulubnarti collection and the subset of the collection utilized in this dissertation lived and died at various times within the period in which the cemetery was in use.[60] Thus, the sample does not represent a sampling of the living population of Kulubnarti at any one point in time; in fact, it is possible that few individuals included in this study interacted in any capacity during life.

Unless the death of each individual can be assigned to a precise date, all data are limited to a period perspective (Souza et al. 2003). This period perspective will be adopted in this dissertation, and all interpretations will refer to the genetics of the Kulubnarti Nubians during the Early Christian Period, spanning 550–800 C.E. These limitations will be recognized as the methods and results of this dissertation's genomic analysis of the Kulubnarti Nubians are presented in the following three chapters.

[60] Accelerator Mass Spectrometry (AMS) radiocarbon dating of the samples included in this analysis is in progress at the time of the submission of this dissertation.

CHAPTER 6
SAMPLE PREPARATION AND SCREENING

"…the field of aDNA research should not be built upon the study of ideally preserved remains." (Dudar et al. 2003: 243)

6.1 Chapter Overview

This chapter discusses the methods used to prepare the Kulubnarti Nubian skeletal samples for ancient DNA (aDNA) analysis and presents the results of the two screening steps used to identify the samples most likely to generate sufficient amounts of genome-wide data using target enrichment and capture of Single Nucleotide Polymorphisms (SNPs). This chapter is divided into four sections describing the progression from the initial selection of petrous bones from 146 individuals from the S and R communities, to the identification of the 28 samples with the best biomolecular preservation used to test the hypotheses associated with the research aims that guide this dissertation.

The first section of this chapter details the methods of sample selection and preparation. It describes the production of bone powder from the cochlea (the most DNA-dense part of the osseous labyrinth) and explains the steps of DNA extraction and construction of DNA sequencing libraries, all of which took place in aDNA-dedicated facilities at University College Dublin (UCD). The methodologies and protocols described in this section were selected based on their success in obtaining authentic endogenous DNA from highly-degraded ancient skeletal material.

The second section describes the initial screening process, also conducted at UCD, which assessed the biomolecular preservation of 99 samples through low-coverage shotgun Next-Generation Sequencing (NGS), a common screening strategy in aDNA

research (Fu et al. 2013a, 2015; Lazaridis et al. 2014; Haak et al. 2015). This section explains how the raw sequencing data were analyzed, presents the results of this initial screening step, and discusses the criteria used to identify the 30 samples that were best-preserved on a molecular level.

The third section describes the process of identifying the samples selected for the generation of genome-wide SNP data using targeted enrichment and capture. All steps in this section took place in the DNA laboratory of Dr. David Reich at Harvard Medical School in Cambridge, Massachusetts, USA (referred to henceforth as the "Reich Lab"). At the Reich Lab, fresh sequencing libraries were prepared from new aliquots (portions) of cochlear bone powder from 29 of the best-preserved Kulubnarti samples.[61] These new libraries were then screened again using the Reich Lab's in-house "MTspike3k" capture, consisting of enrichment and capture of sequences overlapping the mitochondrial genome and ~3000 SNPs from the nuclear genome. A description of the capture and the screening results are presented. Based on these results, 28 samples were identified as having enough uncontaminated endogenous DNA to produce the requisite genome-wide SNP data.

The fourth section describes the generation of genome-wide SNP data by target enrichment and in-solution capture of ~1.24 million SNPs (referred to as the "1240K SNP capture") performed on the 28 best-preserved Kulubnarti Nubian samples. This capture is designed to target variant sites in the human genome that are most powerful for population genetic analysis and ancestry assignment (Mathieson et al. 2015). As all subsequent chapters focus solely on the genome-wide data generated by the 1240K SNP

[61] A single individual from the R cemetery (R195) was not screened.

capture, this chapter includes a discussion of the benefits of using capture methodology, and specifically 1240K SNP capture, for this dissertation's analysis.

6.2 Sample Selection and Preparation

6.2.1 Selection of the Petrous Bone

While the remains recovered during archaeological excavation do not always offer a choice of bone element for aDNA studies, the excavation at Kulubnarti provided a large range of potential samples for this analysis. Because there is no evidence that DNA preservation is affected by age at death or sex, samples were selected from the Kulubnarti collection without consideration of individual demographic information. Instead, the aim of the sampling process was to select the best-preserved samples based on macroscopic appearance, giving preference to individuals with easily accessible petrous bones that were preferably already disarticulated from the rest of the skull.[62] This sampling strategy provided the requisite skeletal material for this dissertation's analysis while simultaneously maintaining as many complete skulls as possible for any future anthropological analyses that require intact specimens.

As discussed in Chapter 5, pilot work suggested that the Kulubnarti Nubian samples would not be amenable to molecular analysis due to high levels of thermally-driven molecular degradation, consistent with other studies of skeletal material from hot environments (e.g. Lazaridis et al. 2016; Schuenemann et al. 2017).[63] This necessitated the screening of a large number of specimens for quality of molecular preservation with

[62] Hansen et al. (2017) confirm some link between visual and molecular preservation. The rationale behind the use of the petrous bone is provided in Chapter 5.
[63] Thermally-driven molecular degradation is discussed in Chapter 4.

the aim of identifying the best-preserved samples to generate genome-wide data for hypothesis testing.

Out of 406 individuals from the Kulubnarti collection, 66 did not have a petrous bone available to be sampled, either due to absence of the skull or, more often, due to the enclosure of the skull in soft tissue or articulation of the skull with the cervical vertebrae. For the remaining 340 individuals, the petrous bone was present and accessible with varying degrees of ease. After an initial survey of the Kulubnarti collection to identify samples that had at least one easily accessible petrous bone, right or left petrous bones from 146 individuals were selected. This included 97 individuals from the S cemetery (representing 45% of the excavated collection) and 49 individuals from the R cemetery (representing 26% of the excavated collection); the breakdown of samples by age category and by sex is provided in Tables 6.1 and 6.2. All sampling took place in the collections at the Department of Anthropology, University of Colorado at Boulder with Dr. Dennis Van Gerven in May 2015.

Table 6.1: Age Category of Individuals Sampled

	S Cemetery	R Cemetery	Totals
Subadults (<18 y.o.)	89	47	136
Adults (≥18 y.o.)	8	2	10
Total	97	49	146

As shown in Table 6.1, most individuals sampled for this analysis were subadults. This was primarily due to their high frequency in the Kulubnarti skeletal collection, particularly in the S cemetery, where the average age at death was 10.6 years of age compared to 18.8 years of age in the R cemetery (Van Gerven et al. 1995). Additionally,

the skulls of young subadults were more fragile and less often fused than those of adults, leading to more frequent disarticulation or breakage *post-mortem*. Because of this, their petrous bones were easily accessible for sampling.

Table 6.2: Sex of Individuals Sampled

	S Cemetery	R Cemetery	Totals
Males	4	5	9
Females	8	5	13
Indeterminate	85	39	124
Total	97	49	146

As shown in Table 6.2, sex was indeterminable for most individuals based on morphological features alone. This was due to difficulty in determining the sex of subadults using dimorphic skeletal traits alone, which typically do not differentiate enough to allow for sex estimation until after puberty (Cunningham et al. 2016); the exception is in cases of soft tissue preservation of sexual organs.

One petrous bone for each individual included in this analysis was selected, sided, photographed, and placed into an individually-labeled plastic bag. Only petrous specimens with no apparent pathology or evidence of microbial degradation were chosen. The side of the petrous was selected at random unless one side of the skull or the petrous bone showed evidence of microbial activity, had a substantial amount of residual soft tissue, or retained sediment from the burial environment. In these cases, the opposite side was selected, as estimates of preservation based on gross appearance have been shown to correlate with success during molecular analyses (Hansen et al. 2017).

To minimize the chance of modern DNA contamination, nitrile gloves were worn during all steps of the sample selection process. After their selection, samples were sent

to UCD and were immediately placed in a freezer room maintained at -20ºC to prevent further molecular degradation (Pruvost et al. 2007; Bollongino et al. 2008; Buś and Allen 2014).

6.2.2 Sample Processing

Ninety-nine of the 146 samples selected from the Kulubnarti collection were processed (here, "processing" refers to the reduction of petrous bone to bone powder) by the author at UCD between July 2015 and July 2016; demographic information for each sample and sample-specific processing details are presented in Supplement 1 as Table S1.1. An additional table that contains detailed bioarchaeological information specific to the 28 samples that were eventually selected for genome-wide analysis is also presented in Supplement 1 as Table S1.2.

All processing of the skeletal material included in this dissertation was conducted in a clean room facility dedicated solely to the processing of ancient skeletal samples for genetic analysis (referred to henceforth as the "grinding lab"). The grinding lab is a specialized and isolated environment that is physically and logistically separated from any post-PCR facilities containing amplified DNA (Cooper and Poinar 2000; O'Rourke 2000; Poinar 2003; Pääbo et al. 2004; Knapp et al. 2012a). This separation of the grinding lab from all post-PCR facilities is critical: because of the low copy number and damaged nature of ancient DNA, any undamaged modern DNA or previously amplified PCR products that are present will be preferentially amplified by PCR instead of the targeted ancient DNA (Willerslev and Cooper 2005; Fulton 2012).

The grinding lab adhered to stringent anti-contamination protocols. Researchers only entered the grinding lab after dressing in disposable microporous coveralls,

disposable shoe covers, a surgical facemask, a hair net, and two pairs of nitrile gloves (the outer layer being changed regularly) (Knapp et al. 2012a; Llamas et al. 2017). A one-way rule for movement from ancient to modern lab facilities was strictly enforced for both personnel and laboratory supplies (Fulton 2012; Knapp et al. 2015; Llamas et al. 2017). Within the lab space, chemical cleaning procedures with ≥10% sodium hypochlorite (bleach) or DNA-degrading detergents (such as DNA-ExitusPlus) were implemented for all laboratory surfaces and equipment (Champlot et al. 2010; Fulton 2012; Knapp et al. 2012a). In addition to chemical cleaning, eligible laboratory consumables, disposable containers (including plastic bags and tubes), small laboratory equipment, and all working surface were exposed to ultra-violet (UV) light for at least 30 minutes before use to contribute to additional decontamination (Champlot et al. 2010).

To begin the sample processing step of this analysis, individual plastic bags containing the Kulubnarti petrous samples were moved from the -20ºC storage freezer to the ante-chamber of the grinding lab, which functioned as an entry area where researchers dressed in the clean room clothing described above as well as an area for preliminary sterilization of any samples, equipment, or consumables that enter the grinding lab space (Knapp et al. 2012a). The exterior surface of each bag was wiped down with bleach and taken into the grinding lab, where it was wiped down again with DNA-ExitusPlus.

Before any sample was removed from its bag, all working surfaces in the grinding lab were wiped down with bleach or DNA-ExitusPlus and all UV lights were turned off. To prevent cross-contamination, each sample was processed completely from start to finish before another sample was removed from its plastic bag; multiple samples were never handled simultaneously.

Once all working surfaces were thoroughly sterilized, the petrous sample was removed from its plastic bag and photographed on a clean surface with a label that included a unique sample identifier. The macroscopic preservation of the bone was recorded based on the author's visual assessment of the bone and is included in Supplement 1. Macroscopic preservation assessments ranged from samples assessed as "very good" (petrous bones that were rich ivory, yellow, or tan in color, were unbroken, and appeared dense) to samples assessed as "bad" (petrous bones that were light ivory or white in color, were light in weight, broken, or had a porous appearance or brittle texture); examples of a "very good" and a "bad" petrous are presented in Figure 6.1.

Figure 6.1: Example of a petrous assessed as "very good" (top picture; S cemetery individual S2) and as "bad" (bottom picture; S cemetery individual S75).

After it was photographed, the sample was placed back into its original bag and put into a cross-linker for 10 minutes. This cross-linker functions to chemically cross-link DNA's double strands by exposure to 254nm UV irradiation, and is commonly used in aDNA research to reduce the potential of amplifying contaminating DNA that may be present on the surface of the sample (O'Rourke et al. 2000; Champlot et al. 2010; Smith et al. 2015). After 10 minutes, the bag was turned over, and the other side was exposed to UV irradiation in the cross-linker.

The petrous sample was then removed from its bag and placed into the previously-sterilized chamber of a dental sandblaster that was located inside its own UV cabinet, shown in Figure 6.2. The placement of the sandblaster inside a UV cabinet allowed for its sterilization through UV-irradiation when it is not in use.

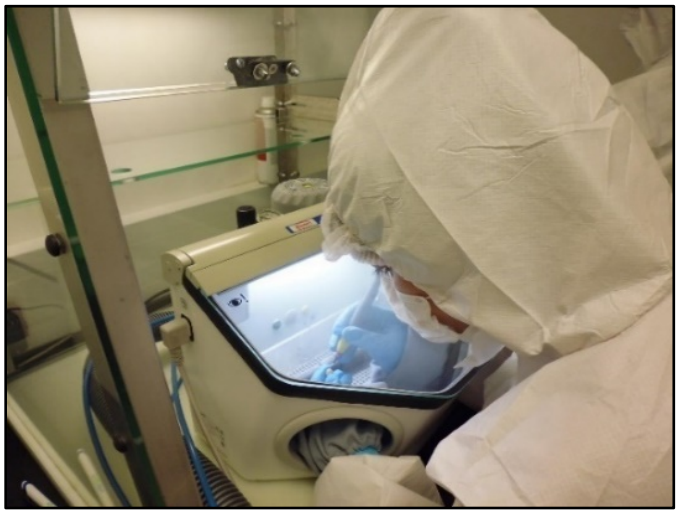


Figure 6.2: Processing a sample inside the sandblaster.

Abrasive aluminum oxide powder (50 microns in size) was added to the tank of the sandblaster and the superficial surface of the petrous bone that had been exposed to the depositional environment and to direct human contact was removed. Once the bone surface was clean, the sandblaster was used to remove additional bone with the purpose of exposing the osseous labyrinth. While not immediately visible upon inspection of a complete petrous bone, the osseous labyrinth was located using the inner auditory meatus and the arcuate eminence (which indicates the location of the superior semicircular canal) as anatomical landmarks (Figure 6.3).

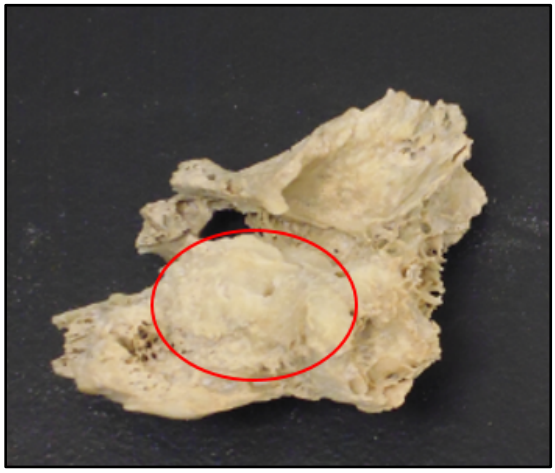


Figure 6.3: The anterior surface of a partially-processed petrous bone with osseous labyrinth exposed by sandblasting; the osseous labyrinth is circled in red.

Once the osseous labyrinth was exposed, it was separated from the rest of the petrous bone. The isolated osseous labyrinth is shown in Figure 6.4.

Figure 6.4: Osseous labyrinth separated from the rest of the petrous bone.

The cochlea was then separated from the rest of the osseous labyrinth through the vestibule area using high-pressure sandblasting to "cut" through this space. The cochlea was the specific substrate targeted for DNA extraction, as unpublished research suggests that it is the most DNA-dense region of the osseous labyrinth. An assessment of the color and state of preservation of each cochlea was made by the author and is included in Table S1.1 in Supplement 1.

The isolated cochlea was then examined and any sediment that had infiltrated the spiral structure of the cochlea was removed using the sandblaster. The clean cochlea was placed in a disposable plastic weigh boat that had been sterilized through chemical cleaning with DNA-ExitusPlus and exposure to UV-irradiation for 10 minutes in the cross-linker. The cochlea was weighed, and the weigh boat containing the cochlea was placed in the cross-linker for 10 minutes. After 10 minutes, the cochlea was flipped over, and the opposite side was exposed to UV irradiation for 10 minutes. This further reduced the chance of contamination.

The cochlea was then placed into a 25mL stainless-steel grinding jar that had been sterilized with chemical cleaning and UV irradiation, along with a medium-sized stainless-steel grinding ball (12mm diameter). The jar was sealed, and a ball mill was used to pulverize the cochlea into fine homogenous bone powder by shaking the jar at a frequency of 30Hz (1800 min^{-1}) for 30 seconds, waiting 10 seconds to allow any heat to dissipate, and shaking again at the same frequency for another 30 seconds. The jar was opened, and the powder was examined for homogeneity. The homogenous bone powder was then poured from the grinding jar into the sterile disposable plastic weigh boat that had originally held the complete cochlea.

Figure 6.5: Homogenous bone powder in weigh boat.

A ~50mg aliquot of powder (range 42–64mg with one outlier at 28mg and a median of 54mg) was poured from the weigh boat into a PCR clean and UV irradiated 2.0mL tube. The tube was labeled with the sample identifier and was placed into a freezer maintained at -20ºC until DNA extraction took place. Additional powder was placed into a separate tube, labeled as "surplus," and placed in a -20ºC storage freezer.

6.2.3 DNA Extraction

DNA was extracted from the ~50mg bone powder aliquots of 99 Kulubnarti samples following the protocol in Dabney et al. (2013; referred to henceforth as the "Dabney protocol"), briefly reviewed here and described in step-by-step detail in Appendix A. The Dabney protocol was selected for this dissertation's analysis over other extraction methods (e.g., Hänni et al. 1995; Yang et al. 1998; Kalmár et al. 2000) because it is optimized to retain ultrashort DNA fragments <40 base pairs (bp) in length. Size distributions of sequences from aDNA studies consistently show a mode of >40bp (e.g., Green et al. 2010; Orlando et al. 2013), suggesting that many molecules <40bp in size are lost prior to sequencing when traditional extraction methods are used (e.g., Yang et al. 1998). This is problematic because the number of DNA fragments in ancient samples increase as fragment length decreases (Allentoft et al. 2012), and most genetic information therefore resides in ultrashort molecules. The Dabney protocol is used in this dissertation because it is very likely that the thermally-driven degradation of the Kulubnarti skeletal material resulted in a high proportion of ultrashort fragments.

This protocol modifies the silica-based extraction technique from Rohland and Hofreiter (2007a, 2007b) in three ways to improve the recovery of ultrashort fragments: 1) the use of a binding buffer comprised of guanidine hydrochloride, sodium acetate, and isopropanol; 2) an increase in the volume of binding buffer relative to that of extraction buffer; and 3) the use of commercially-available silica spin columns and a custom-adapted extension reservoir (referred to as the "custom-made binding apparatus") to load a larger volume of extraction buffer.

All DNA extractions took place in an aDNA-dedicated clean room facility at UCD (referred to henceforth as the "wet lab") that was physically and logistically separated from both the grinding lab as well as any post-PCR facilities. In addition to the physical separation of aDNA facilities from post-PCR facilities, the separation of the wet lab from the grinding lab is an additional precaution against cross-contamination from loose bone powder that may be present in the air and on surfaces in the grinding lab. Consistent with the grinding lab, the wet lab adhered to stringent anti-contamination protocols, including protective dress for all trained researchers entering the lab, a strict one-way rule for movement of people and reagents, chemical cleaning procedures with bleach and DNA-ExitusPlus, the use of UV-irradiation for additional decontamination of eligible consumables and reagents, and two pairs of gloves worn throughout all steps of sample preparation with the outer layer changed regularly. Extractions were completed inside a laminar flow cabinet that was solely dedicated to the extraction of ancient DNA. Between 8–12 extractions were completed simultaneously, with one negative control included for every 10 samples.

Based on unpublished experimentation indicating that a reduction from the 85–120mg of bone powder used in the Dabney protocol was possible, an extraction buffer solution that dissolved part of the hydroxyapatite matrix specific to bone and teeth and denatured contaminating proteins in the solution was added to each tube containing bone powder.[64] The bone powder was then suspended in the buffer by vortexing, and the solution was incubated at 37ºC with rotation for 18 hours in a thermal mixer to release both endogenous and exogenous DNA from the powdered bone into the extraction buffer

[64] The recipe for extraction buffer can be found in Appendix A.

solution; as previously mentioned, endogenous DNA from ancient specimens is present alongside high levels of microbial DNA, making an aDNA extract a metagenomic mixture (Orlando et al. 2015). After 18 hours, the tubes were centrifuged to separate a "pellet" of undissolved bone powder from the supernatant that contained the DNA in solution.

The DNA molecules in solution were then bound to silica particles to retain them during subsequent wash steps used to remove PCR-inhibiting salts from the extract. This required the construction of a custom-made binding apparatus for each sample that combined a large-volume extension reservoir (used to hold large volumes of liquid) with a commercially-available spin column with a silica membrane (used to bind DNA molecules with silica particles to retain them during wash steps).[65] This apparatus was constructed by sterilizing an extension reservoir and forcibly fitting it on to a spin column; each apparatus was placed into a 50mL tube.

For each sample, the supernatant containing the extracted DNA in solution was collected from the centrifuged tubes and combined with binding buffer in the extension reservoir of the binding apparatus.[66] The 50mL tubes (containing the binding apparatus that held the DNA extract and binding buffer) were then centrifuged. Centrifugation forced the DNA extract and binding buffer solution through the silica-based spin column, where the DNA molecules were bound to the silica particles in the spin column membrane and the residual binding buffer solution pooled at the bottom of the 50mL tube.

---

[65] While the MinElute spin column (Qiagen) is the preferred silica-based product for this protocol, the MinElute collection tube itself is too small to accommodate the 13.0mL of binding buffer required by the Dabney protocol. Therefore, the custom-made binding apparatus is required.

[66] The recipe for binding buffer can be found in Appendix A.

Following centrifugation, the spin columns that now contained the DNA molecules bound to silica particles were detached from the extension reservoirs and placed into 2.0mL collection tubes. Two wash steps were performed using an ethanol wash buffer that desalted the silica particles bound to the DNA.[67] The spin columns were then placed into clean 1.5mL tubes and the DNA was detached from the silica particles and eluted into a buffer that solubilizes DNA and maintains the stability of DNA in solution.[68] DNA extracts were stored in a freezer maintained at -20°C until library preparation occurred.

6.2.4 Construction of Next-Generation Sequencing Libraries

The development of NGS technologies has transformed the sample preparation process by requiring that raw DNA fragments in a DNA extract be converted into DNA "libraries" prior to sequencing (Briggs and Heyn 2012; Gansauge and Meyer 2013). This process involves the ligation of adapters (artificially synthesized DNA segments of known sequence) to both ends of the raw DNA fragments, ultimately converting these raw molecules into library molecules that are "visible" to NGS sequencing technologies (Knapp and Hofreiter 2010; Metzker 2010; Gansauge and Meyer 2013). Building a DNA library using universal adapters, as completed for this dissertation, turns amplifiable DNA into a renewable resource that can be used for screening and in experiments beyond shotgun sequencing, such as target enrichment (Briggs and Heyn 2012; Rohland et al. 2015).

---

[67] The clean-up kit containing this wash buffer is described in Appendix A.
[68] The recipe for the elution buffer can be found in Appendix A.

To construct DNA libraries for NGS screening, this dissertation used the library preparation protocol by Meyer and Kircher (2010) as modified by Gamba et al. (2014). In this modified version, blunt-end repair was performed using the NEBNext End Repair Module (New England Biolabs, NEB) and Bst polymerase was inactivated using heat (20 minutes at 80ºC during the adapter fill-in incubation step) instead of an extra purification step.[69] The steps of preparing libraries for multiplexed sequencing on the Illumina platform are briefly reviewed here and are described in step-by-step detail in Appendix B.

Construction of DNA sequencing libraries took place in the wet lab at UCD in adherence with the same stringent anti-contamination protocols already described. All steps of library preparation took place inside a laminar flow cabinet that was specifically dedicated to library preparation (i.e., it was a separate cabinet from the one used for DNA extraction, though both cabinets were located within the wet lab space). Between eight and twelve DNA libraries were prepared simultaneously using DNA extracts that had been stored in a freezer maintained at -20ºC. One negative control for every ten samples was included; in addition, libraries were also built for the negative extraction controls created during the DNA extraction step.

While the first step of library preparation in the Meyer and Kircher (2010) protocol is to shear long DNA molecules to produce molecules of optimal length for the Illumina sequencing approach (200–500bp) (Bronner et al. 2009), this step is not required during ancient DNA analysis. In fact, the shearing of ancient DNA fragments may be detrimental because ancient DNA is already present in the form of short molecules

[69] Studies have shown that each purification step can reduce the amount of template DNA present in the sequencing library between 16-40% (Maricic and Pääbo 2009; Knapp and Hofreiter 2010). The modified protocol used by this dissertation removes one of the purification steps in the original protocol by inactivating Bst with heat, thereby retaining as much DNA as possible.

(Knapp and Hofreiter 2010; Meyer and Kircher 2010; Briggs and Heyn 2012). As such, avoiding the DNA shearing step is recommended practice in aDNA research.

Skipping the DNA shearing step, there were four primary steps (I–IV) involved in the construction of libraries for the Kulubnarti samples in this dissertation, illustrated in Figure 6.6.

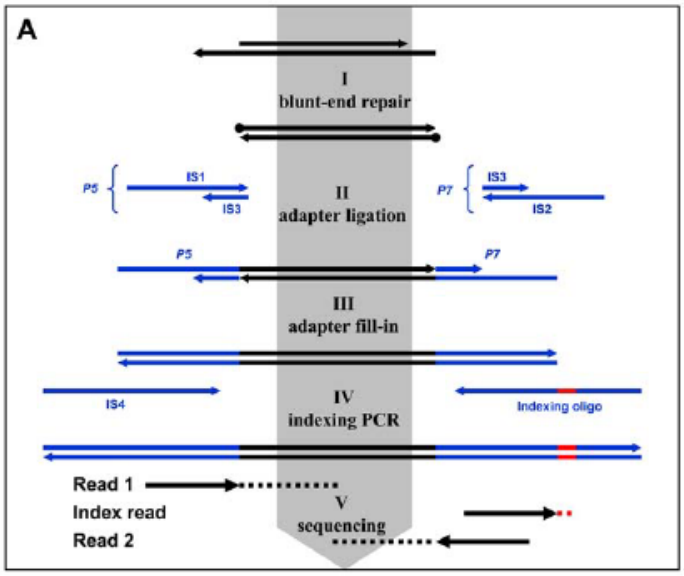


Figure 6.6: Schematic overview from the library preparation protocol of Meyer and Kircher (2010: 4). This schematic does not depict the shearing of sample DNA into small fragments, but details the four primary steps of library preparation, and includes an additional diagram of DNA sequencing (part V).

First, because aDNA extracts do not contain homogenous blunt-ended DNA fragments, but instead contain double-stranded molecules with a mixture of blunt-ends, recessed 3'-ends and recessed 5'-ends, blunt-end repair was required (Bronner et al. 2009). Blunt-end repair works by "filling in" or "chewing back" 3'- and 5'-overhangs respectively and ensuring that 5'-ends are phosphorylated (i.e., contain an added phosphate group), making them suitable for subsequent adapter ligation.

To end repair the sequencing libraries, a blunt-end repair mastermix was assembled that contains a polymerase to repair overhanging 5'- and 3'-ends through 5'→3' polymerase activity and 3'→5' exonuclease activity and a kinase (an enzyme that catalyzes the addition of phosphate groups) to add 5'-phosphates for subsequent adapter ligation.[70] The mastermix was added to the thawed DNA extract, and the reactions were incubated to allow blunt-end repair to occur.

After blunt-end repair, the Meyer and Kircher (2010) protocol calls for the purification of the reaction to remove residual primers, enzymes, salts, and other impurities from DNA samples that may inhibit downstream PCR amplification. The original protocol conducts this reaction clean-up using solid phase reversible immobilization (SPRI), which binds DNA to carboxyl-coated magnetic beads; however, SPRI purification may not retain molecules shorter than 100–150bp. This may be problematic for the Kulubnarti DNA extracts, which are expected to contain a majority of DNA fragments <100bp in length. Therefore, as recommended as an alternate option in the same protocol, a silica-based purification kit specifically designed to retain DNA fragments as small as 70bp in length was used.[71] A binding buffer provided in the kit was first used to bind blunt-end repaired DNA molecules to silica particles, and a wash step with the same ethanol-based wash buffer as used during DNA extraction was performed to remove any potential PCR-inhibiting impurities. After washing, the blunt-end repaired DNA was eluted in a buffer designed specifically for the elution of nucleic acids.

[70] The recipe for blunt-end repair mastermix can be found in Appendix B.
[71] This is the same silica-based purification kit that provided the MinElute spin columns used for the DNA extraction step.

The second step of library preparation was adapter ligation, where two different adapters (referred to as "P5" and "P7" adapters) were ligated onto either end of the blunt-end repaired DNA molecules, flanking the DNA sequence of interest (referred to as the DNA "insert"). The addition of these adapters to the DNA insert allows the DNA fragment to bind to the sequencing flow cell during NGS (using two primer binding sites at the ends of the adapters that are complementary to the two different oligo sequences on the flow cell), allows for PCR enrichment of adapter-ligated DNA fragments, and allows the indexing of samples for multiplex sequencing.[72]

Prior to the adapter ligation step of library preparation, an adapter mix was assembled that contained the requisite adapters.[73] Using the prepared adapter mix, an adapter ligation mastermix was assembled during this step of library preparation that included a ligase to attach the adapters to the ends of the DNA molecules and buffers that increased ligation efficiency and optimized the ligation process.[74] The mastermix was added to the blunt-end repaired eluate and the reactions were incubated in order for adapter ligation to occur. After adapter ligation was completed, the reactions were again purified through another purification step using the same silica-based purification kit.

The third step of library preparation involved the extension of the 3'-ends of the ligated adapters. Because the adapters ligated in the previous step were non-phosphorylated (i.e., lacking 5'-phosphates), they contained a 5'-overhang on one end to ensure ligation in the correct orientation, which resulted in a nick on one strand of the ligated DNA (Yegnasubramanian 2013). A polymerase that contains 5'→3' polymerase

[72] The rationale behind and process of indexing will be explained later in this section.
[73] The recipe for adapter mix can be found in Appendix B. The sequences of the required adapters can be found in Appendix C.
[74] The recipe for adapter ligation mastermix can be found in Appendix B.

activity (Bst polymerase) was therefore used to incorporate dNTPs (deoxyribonucleotide triphosphates, namely the monomeric substrates dATP, dCTP, dGTP, or dTTP) to "fill-in" the complementary sequence to this 5'-overhang.

An adapter fill-in mastermix was assembled that contained the requisite dNTPs in an equimolar solution, Bst polymerase to incorporate the dNTPs to the adapter-ligated strand, and buffers to provide superior reaction conditions.[75] The mastermix was combined with the adapter-ligated eluates, and the reactions were incubated to allow adapter fill-in to occur. Immediately after this incubation step, the reactions were incubated again at a high temperature (80ºC) to inactivate the Bst polymerase; this second incubation represents a modification to the original Meyer and Kircher (2010) protocol which calls for another purification step at this point.

The fourth and final step of library preparation was indexing PCR. This step prepared the DNA template molecules for PCR amplification by adding a PCR-specific primer, IS4, and "indexing" the DNA molecules through the introduction of an adapter containing a unique 7bp fragment of known sequence to each library; this is referred to as the "index." Indexing the DNA molecules in each library allows multiple libraries to be pooled together for multiplex sequencing and subsequently sorted bioinformatically based on their index sequence after NGS (Knapp and Hofreiter 2010; Knapp et al. 2012b).

During the indexing PCR step, an amplification mastermix was assembled using a "hot-start" polymerase and primer IS4, a P5 primer that does not include an index.[76] The

[75] The recipe for adapter fill-in mastermix can be found in Appendix B.
[76] The recipe for amplification mastermix can be found in Appendix B. The sequence for the required primer can be found in Appendix C.

benefit of using a "hot-start" polymerase is that the enzymatic activity does not occur at room temperature, but regains activity after exposure to high temperature during the initial step of PCR amplification.[77] The mastermix was then combined with the filled-in eluate, and the indexing adapter containing the unique 7bp index sequence was added. A unique indexing adapter was used for each sample, beginning with Illumina-specific index B001 and increasing sequentially. At this point, the indexed DNA libraries were immediately removed from the wet lab for amplification by the Polymerase Chain Reaction (PCR).

6.2.5 PCR Amplification

Indexed DNA libraries were subsequently amplified at UCD by PCR. The use of PCR is a fundamental part of preparing aDNA libraries for NGS (Shapiro and Hofreiter 2014; Linderholm 2016). It relies on a thermal cycling method involving three temperatures through which DNA libraries are cycled a number of times defined by the user (Mullis et al. 1986; Mullis and Faloona 1987; Saiki et al. 1988). In the first step of the cycle, double-stranded template DNA is denatured (i.e., the strands are physically separated) at a high temperature. In the second step of the cycle, the temperature is lowered, and the primers present in the reaction anneal to each of the single-stranded DNA templates and the DNA polymerase binds to the primer-template hybrid. In the third step of the cycle, the DNA polymerase synthesizes a new strand of DNA in the 5'→3' direction that is complementary to the DNA template strand by adding free dNTPs

[77] Accuprime *Pfx* Supermix was specifically chosen for this step as an alternative to other commonly used polymerases (e.g. AmpliTaq Gold and Platinum *Taq* High-Fidelity) which preferentially amplify short and GC-rich templates (Orlando et al. 2015).

from the reaction mixture. As PCR progresses, the DNA that is generated in previous cycles is used as a template for replication, resulting in exponential amplification of template DNA. This makes PCR an ideal tool for amplifying the ancient DNA molecules that are often present in small quantity in aDNA libraries (Pääbo et al. 1989).

Based on the temperature cycling profile recommended for the polymerase used in the fourth step of library preparation, PCR amplification was initiated with an extended denaturation step of 5 minutes at 95ºC, allowing the "hot-start" polymerase to regain enzymatic activity (Sharkey et al. 1994). Immediately following this denaturation step, the following temperature cycling profile was used for amplification: 15 seconds at 95ºC, 30 seconds at 60ºC, and 30 seconds at 68ºC. Following the completion of the cycling, a final elongation step was performed at 68ºC for 5 minutes, ensuring that any single-stranded DNA molecules were fully elongated. After the final elongation, temperature was held at 4ºC until the reactions were removed from the thermal cycler. For the Kulubnarti libraries, it was standard to perform 12 cycles of PCR.[78]

After PCR amplification was complete, the PCR products were placed into a storage freezer maintained at -20ºC until post-PCR clean-up took place. Because PCR amplification produces large quantities of amplified and barcoded DNA, this storage freezer was located in a modern DNA lab facility away from any aDNA-dedicated space.

6.2.6 Post-PCR Processing

Before sequencing could occur, it was necessary to purify the PCR products. This purification took place in a modern DNA lab using the same silica-based purification

[78] The subsequent quantification of the library on the Bioanalyzer indicated if the library had been overamplified during PCR. If this was the case, the library was re-indexed and re-amplified using 11 cycles of PCR.

method previously described between steps one and two and two and three of library preparation.

After the PCR products had been purified, it was necessary to determine the concentration of DNA in each amplified library so that multiple libraries could be pooled in equimolar ratios prior to multiplex sequencing. This library quantification was done using an Agilent Bioanalyzer 2100 (referred to henceforth as the "bioanalyzer") according to manufacturer guidelines. The bioanalyzer uses a chip format based on traditional gel electrophoresis principles to perform "on-chip gel electrophoresis." Chips were prepared by loading the amplified library into wells on the chip that were infiltrated by the electrodes of the bioanalyzer. DNA is a charged molecule, and therefore the application of a voltage gradient separated the molecules by size as they were driven electrophoretically across the bioanalyzer chip's sieving polymer matrix (akin to a traditional gel), with smaller fragments migrating faster and farther than larger ones. A standard ladder containing components of known size was used to plot a standard curve of migration time versus fragment size for each sample. The concentration (in ng/μL) and molarity (in nmol/L) of the amplified library was then calculated by comparing the area under the upper marker peak (produced by the ladder) to the sample peak area.

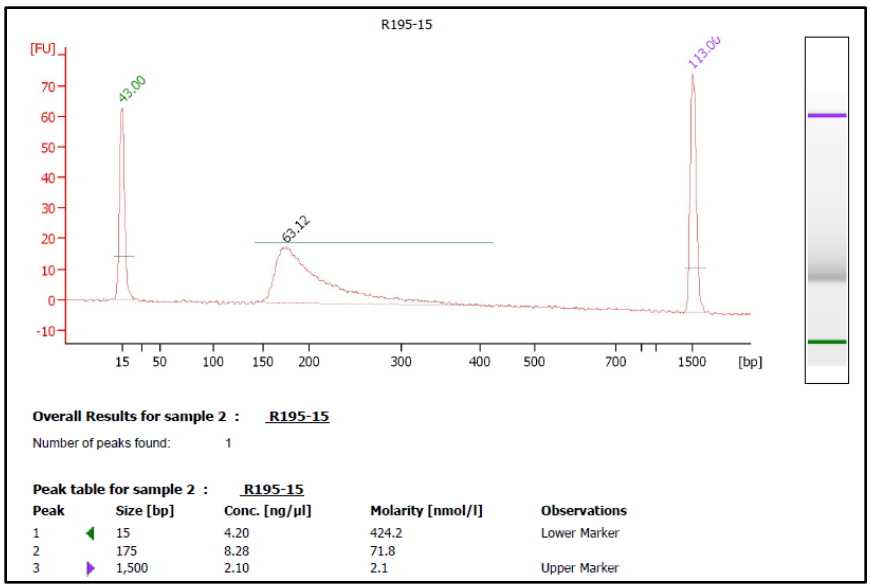


Figure 6.7: Example of output from Bioanalyzer 2100 for R cemetery sample R195 with concentration (8.28ng/µL) and molarity (71.8nmol/L) determined using the Bioanalyzer software. The peak indicating the presence of DNA is visible, and the shape of the peak indicates that the correct number of PCR cycles were used for amplification.

Based on the concentration determined by the bioanalyzer output, the amount of amplified library from each sample to be added to a sequencing library "pool" was determined. The amount of each library to add was calculated by dividing the amount of DNA desired in the pool (in ng) by the concentration (in ng/µL) provided by the bioanalyzer: if the concentration of the library was determined to be >2ng/µL, 20ng of DNA was to be added to the pool; if the concentration was determined to be <2ng/µL, 10x the concentration of the library was to be added to the pool.[79] These calculations resulted in an inverse relationship between the concentration of the library and the amount of library added to the pool. Pooling occurred immediately before sequencing took place.

[79] In cases where the concentration was <2ng/µL, a maximum amount of 10.0µL of library was added to the pool.

6.3 Screening #1: Next-Generation Shotgun Sequencing

6.3.1 Preparation for Next-Generation Shotgun Sequencing

Library pools were created by pipetting the determined amount of each amplified library into a single 1.5mL tube in the modern DNA lab facility at UCD. When all desired samples were pooled, the concentration and molarity of the pool was determined to allow its dilution to an optimal concentration for sequencing.[80] The pool was quantified on the bioanalyzer in triplicate and the average concentration and molarity of the pool were calculated. A Qubit 2.0 Fluorometer, which uses target-specific fluorescence (in this analysis, the target is double-stranded DNA), a method that emits a signal which varies in intensity based on the concentration of the DNA in the pool and minimizes the effects of contaminants, was used for additional precision in quantifying the pool. This analysis was also done in triplicate, and the average concentration of the pool based on Qubit results was calculated. The overall average concentration was then determined using the averages from both quantitation instruments.

This overall average concentration value was ultimately used to determine the initial volume of undiluted pool that was mixed with laboratory-grade water to reach the optimal starting concentration for NGS using the equation $C_1*V_1=C_2*V_2$ (where C represents concentration and V represents volume, and $_1$ represents the initial values and $_2$ represents the desired final values). C_2 and V_2 are standard values in this equation: C_2 is the optimal starting concentration; when the Illumina MiSeq and NextSeq instruments are used, C_2 is 4.0 nM and V_2 is 50.0µL. C_1 is the average concentration of the pool (calculated using both bioanalyzer and Qubit values as described above), and V_1 is the

[80] The Illumina sequencers used in this analysis require a precise concentration to operate properly; this concentration is dependent on the instrument used.

value that is determined with this equation. Once V_1 is determined, 50.0µL – V_1 provides the amount of water that will be required to dilute the library to the desired starting concentration of 4nM.

The protocol used for preparing library pools for sequencing are reviewed briefly here and provided in step-by-step detail in guides published by the manufacturer (Illumina, Inc. 2016b, 2016c). The library pool was diluted to a starting concentration of 4nM directly before sequencing took place. A solution of NaOH was used to denature the DNA molecules in the pool library, and a hybridization buffer was used to further dilute the library pools from a starting concentration of 4nM to a manufacturer-recommended concentration of 11pM when the MiSeq was used, or 1.8pM when the NextSeq instrument was used. Once at its final concentration, a 1% PhiX control was denatured and diluted to a final concentration of 12.5pM and "spiked-in" to the library pool. PhiX is a small viral genome, and a low-concentration "spike-in" (1%) is commonly used as a technical control for clustering assessment and sequencing accuracy; the control libraries generated from this virus can tell the user whether failure of the sequencing run was due to sample preparation failure or a failure at some step in the sequencing process.[81] The final solution comprising library pool and PhiX was added to the instrument's reagent cartridge, the cartridge was loaded onto the instrument for sequencing.

6.3.2 Sequencing

DNA sequencing using NGS is the preferred method for aDNA analysis because it can detect genomic alterations (including Single Nucleotide Polymorphisms, or SNPs) and offers high-throughput (the sequencing of massive amounts of DNA at once), as well

[81] The actual genome sequence of PhiX is not analyzed.

as the option of multiplex sequencing (the sequencing of many samples at once). This dissertation utilizes Illumina NGS technology, which utilizes clonal amplification and sequencing-by-synthesis (SBS) to sequentially read single DNA bases as they are incorporated into a nucleic acid chain during a sequencing run. Illumina's NGS technology will be reviewed briefly here, while more detailed explanations of this technology can be found in the most up-to-date guide to NGS published by Illumina, Inc (Illumina, Inc. 2016a).

The diluted library pool was initially loaded onto a flow cell, a glass slide with lanes containing two different types of "oligos" (synthetic oligonucleotide primers of known sequence) on its surface, or "lawn" (Quail et al. 2012). The first type of oligo was complementary to the adapter region ligated to DNA template strands, enabling hybridization of the adapter-ligated target DNA to the flow cell surface. These oligos captured the adapter-ligated target DNA molecules, enabling each DNA fragment to be amplified into distinct clonal clusters (a process referred to as "cluster generation") during sequential cycles of DNA synthesis through bridge amplification.

In this process, the DNA strand folded over and the adapter region hybridized to the second type of oligo on the lawn. A double-stranded "bridge" was then formed by the generation of the complementary strand by the action of polymerases. The "bridge" was denatured, resulting in two single-stranded copies of the DNA molecule that were attached to the flow cell. The process was repeated and occurred simultaneously for millions of clusters, resulting in clonal amplification of all the DNA fragments.

After the clusters were generated, the reverse strands of DNA attached to the flow cell were washed away, leaving only the forward strands for SBS.[82] SBS occurred when fluorescently-labeled nucleotides were incorporated one-by-one to the growing nucleotide chain during each sequencing cycle; this chain would eventually comprise a strand that was complementary to the template strand. After the addition of each nucleotide, the clusters were "excited" by a light source, and a fluorescent signal was emitted. The emission wavelength and signal intensity were read by the sequencing instrument, consequently creating a "read" by determining the base call (i.e., which nucleotide was incorporated during that particular cycle). All molecules in a given cluster were read simultaneously, resulting in the sequencing of hundreds of millions of clusters in a massively parallel process. The number of cycles determined the length of the read, and the index sequences were subsequently identified in a separate "index read." Decoupling the template read and the index read kept the sequencing error rate low (Kircher et al. 2012).

This dissertation utilized both the Illumina MiSeq and NextSeq instruments in the core genomics facility at UCD for sample screening by low-coverage shotgun sequencing.[83] The instruments were operated by the author following the guidelines of the manufacturer. Screening of 99 Kulubnarti samples took place on eight sequencing runs between September 2015 and October 2016. The MiSeq instrument generated approximately 12–25 million reads per single-end sequencing run with a MiSeq Reagent

[82] This analysis uses single-end sequencing, in which the forward DNA strands are read from one end to the other, generating a sequence of base pairs. Single-end sequencing runs are faster and cheaper than paired-end sequencing runs and are appropriate for this screening step.

[83] All denaturing and diluting steps in the sequencing preparation step of this dissertation follow the manufacturer's guide for the MiSeq and NextSeq systems. The NextSeq was purchased midway through the sample screening process; it was utilized preferentially over the MiSeq as soon as it was available due to its substantially higher data output.

Kit v2 and single-end sequencing for 65 + 7 cycles.[84] Approximately 15 million reads passing quality filter (>Q30) were generated per sequencing run, dispersed between all samples included on the run.[85] The NextSeq instrument generated approximately 400 million reads per run passing quality filter using a NextSeq v2 High Output kit and single-end sequencing for 75 + 7 cycles, dispersed between all samples included on the run.

6.3.3 Processing of Sequencing Data

After each sequencing run, raw sequencing data were uploaded to Illumina's cloud-based storage platform "BaseSpace." All raw data were run through a custom NGS data-processing pipeline that provided the necessary statistics to determine quality of molecular preservation, including number of reads aligned to the human genome and endogenous DNA yield.[86] The authenticity of the data (i.e., the likelihood of the DNA reads being authentically ancient) was assessed and molecular sex and mitochondrial haplogroup were assigned when possible.

FastQC (an Illumina BaseSpace app; https://basespace.illumina.com/home/index) was used to determine the total number of reads sequenced for each sample, including reads of both endogenous and exogenous origin. Following this initial FastQC analysis, manipulation of the raw data was required prior to alignment to the human genome.

---

[84] As Illumina SBS reads 1bp each cycle, 65 cycles were used to create reads that were 65bp in length and indexes were read in a separate 7 cycle index read; this is written as 65 + 7 cycles.

[85] The "Q score" is a measure of base calling accuracy; a score of Q30 is equivalent to the probability of an incorrect base call 1 in 1000 times, meaning that the base call accuracy is 99.9% (Illumina Inc., 2011). Q30 is considered the benchmark for quality in NGS because when sequencing quality reaches Q30, virtually all reads will have no errors or ambiguities (Illumina Inc., 2011).

[86] Additional metrics, such as analysis of library complexity using Quantitative PCR (qPCR), will be included in future work as further indication of molecular preservation; this technology was not established in the ancient DNA lab at UCD during the time of screening the Kulubnarti Nubian libraries.

Because aDNA molecules are frequently ~50bp in length, and the sequencing runs were set to either 65 cycles (MiSeq) or 75 cycles (NextSeq), the sequencing instrument often sequenced part of the adapter ligated to the 3' end of the DNA template molecule, resulting in the partial inclusion of the adapter sequence in the output reads.[87] This adapter needed to be removed ("trimmed") from all reads before alignment to the human reference genome. The cutadapt v1.5 software (Martin 2011) was used to trim the 3' adapter sequence (-a AGATCGGAAGAGCACACGTCTGAACTCCAGTCAC) from all sequencing reads. Minimum overlap between the read and the adapter was set to 1 using the option '-O 1', reducing the number of bases trimmed due to their similarity to the adapter sequence, and minimum read length was set to 17bp using the option '-m 17'.

Trimmed reads were aligned to the UCSC hg19 (GRCh37) build of the human genome with the mitochondrial sequence replaced by the revised Cambridge reference sequence (Andrews et al. 1999; NCBI accession number NC_012920.1) using the Burrows-Wheeler Aligner (BWA) v0.7.5a (Li and Durbin 2009).[88] The command *samse* and the parameters '-n 0.01 -o 2 -1 1000' were used to allow for more sequences mismatches and nucleotide insertions or deletions ("indels"), as well as to turn off seeding.[89] These are "relaxed parameters" common in aDNA analysis, as discussed in Meyer et al. (2012), which ensure that the high frequency of C to T substitutions at the

---

[87] Keeping in mind that the number of cycles determines the length of the read (1 cycle = 1bp).

[88] The replacement of the mitochondrial reference sequence was based on the recommendation of UCSC: "Since the release of the UCSC hg19 assembly, the *Homo sapiens* mitochondrion sequence (represented as "chrM" in the Genome Browser) has been replaced in GenBank with the record NC_012920 [the Revised Cambridge Reference Sequence]. We have not replaced the original sequence, NC_001807, in the hg19 Genome Browser. We plan to use the Revised Cambridge Reference Sequence (rCRS) in the next human assembly release."

[89] Seeding is used to decrease the time required for sequence alignment by looking for exact matches of part of the read with part of the target (here, the reference genome). It requires that the designated seed length sequence is an exact match to the reference, and therefore is not recommended for damaged and degraded ancient DNA.

5'-end of ancient DNA molecules do not lead to the exclusion of the read during alignment (Schubert et al. 2012).[90]

FastQC analysis was then performed again on each sample to determine the number of reads aligned to the human genome. Percent endogenous DNA yield was subsequently determined by dividing the number of aligned reads by the total number of reads and multiplying this value by 100. This preliminary data manipulation provided the requisite metrics for assessing the quality of molecular preservation for the Kulubnarti Nubian samples.

SAMtools v.0.1.19 (Li et al. 2009), a set of utilities to manipulate alignments, was used to remove PCR duplicates and any reads with mapping quality <Q30. The software mapDamage2.0 (Jónsson et al. 2013) was then used to assess whether the aligned reads exhibited a *post-mortem* degradation pattern typical of ancient samples. As discussed in Chapter 4, because authentic ancient DNA exhibits characteristic nucleotide misincorporations in its sequencing reads, the damage patterns that typify ancient DNA can be used to argue for sequence validity (Krause et al. 2010; Ginolhac et al. 2011). As *post-mortem* degradation in ancient samples results in an increase of C→T substitutions at the 5'-end of the reads and G→A substitutions at the 3'-end of the reads, the frequency of nucleotide misincorporation at the terminal nucleotide at the ends of the reads was determined, and the pattern and frequency of nucleotide misincorporation throughout the reads was plotted and assessed.[91]

---

[90] The default parameter is to use the first 32 nucleotides as a seed region, allowing two mismatches at most in this region in order to reduce the alignment runtime (Schubert et al. 2012). It is recommended to disable the seed when analyzing ancient sequences.

[91] This hydrolytic damage occurs primarily at the ends of the DNA molecule because single-stranded overhangs of a few bases (Briggs et al. 2007; Brotherton et al. 2007) result in faster cytosine deamination than in double-stranded DNA. *Post-mortem* DNA degradation is discussed in detail in Chapter 4.

When possible, sex was assigned using a method that plots the number of reads sequenced on each chromosome against the length of the chromosome.[92] Chromosomes of longer length contribute more DNA molecules to the DNA library and consequently provide more reads during shotgun sequencing. In the human genome, chromosome 1 is longest, with each sequential autosomal chromosome decreasing in length through chromosome 22.[93] The X chromosome is between chromosomes 7 and 8 in size, while the Y chromosome is approximately the size of chromosome 19 (i.e., chromosome X is substantially larger than chromosome Y). Because humans, as diploid organisms, have two copies of each autosome, there is a linear relationship between the size of the chromosome and the number of reads it receives during shotgun sequencing. Sex can thereby be determined based on the number of reads received by the sex chromosomes. If there are two copies of the X chromosome (indicating that the individual is female), the total number of reads will fall as expected within the linear relationship of chromosome length versus number of reads when chromosomes are present in diploid; however, if there is one copy of the X chromosome and one copy of the Y chromosome (indicating that the individual is male), the total number of reads will fall below the linear relationship of chromosome length versus number of reads when chromosomes are present in diploid. Along with estimation of sex, the likelihood of the correct sex assignment was predicted and a p-value associated with the likelihood was calculated.[94]

---

[92] Sexing method developed by Dr. Eppie Jones, Trinity College Dublin. Method is not yet published.
[93] Information regarding the size of each chromosome can be found on the National Center for Biotechnology Information's Genome Reference Consortium website or in many basic human genetics textbooks.
[94] A script for calculating these values is implemented directly into the sex assignment script, and is available upon request from Dr. Jones.

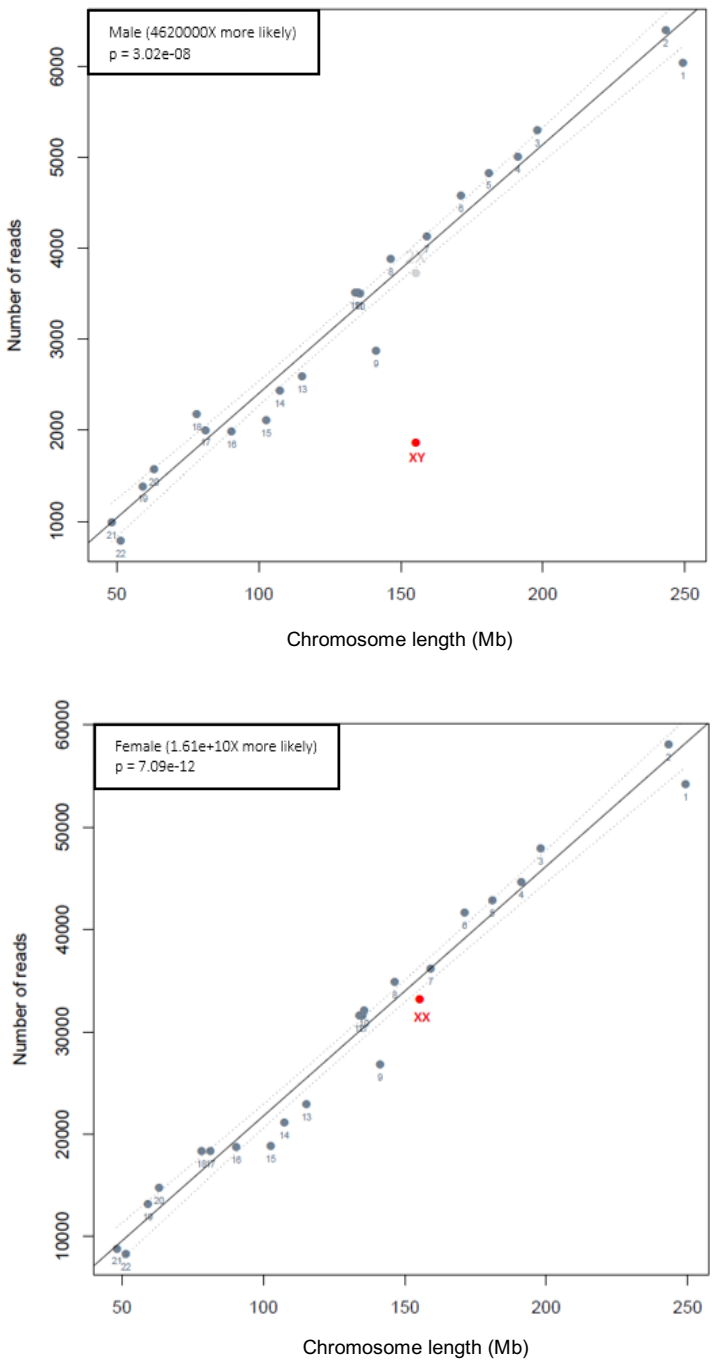


Figure 6.8: Example of sexing plots produced for male individual (XY, top) and female individual (XX, bottom).

Also when possible, mitochondrial haplogroup was assigned using the program Phy-Mer (Navarro-Gomez et al. 2015). Phy-Mer decomposes a mitochondrial sequence

into all possible k-mers (all possible subsequences of length k from a sequencing read; Phy-Mer specifically uses 12mers) and then compares all 12mers against each of the 12mer sets of all haplogroups. A score is derived for each haplogroup, and the top score is assigned as the haplogroup classification.[95]

6.3.4 Screening Results

All processed screening data from 99 Kulubnarti samples are presented in Supplement 2. Upon the completion of data processing, the 99 samples were divided into two groups by cemetery (N=59 for the S cemetery; N=40 for the R cemetery). The samples in these two groups were then ordered from highest to lowest endogenous DNA yield.[96] Endogenous DNA yield from S cemetery samples ranged from 0.04–24.36%, with a median of 0.74%, while endogenous DNA yield from R cemetery samples ranged from 0.10–25.84%, with a median of 2.67%.

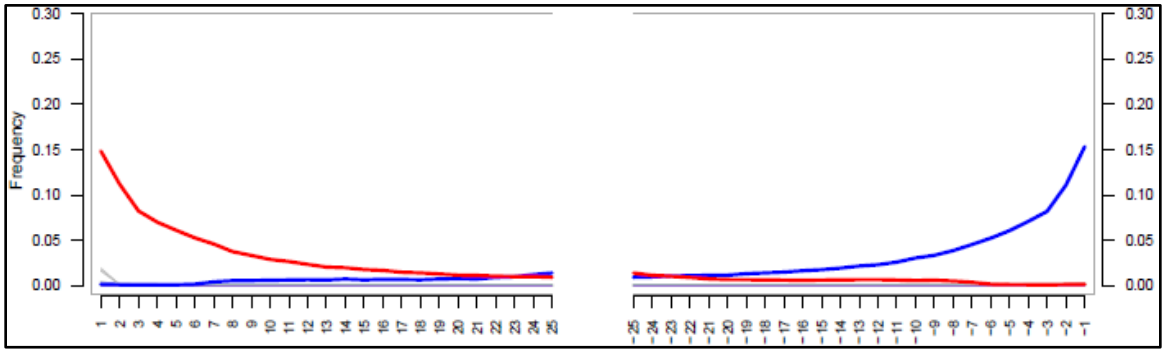
For each group, the fifteen samples with the highest endogenous yield (ranging from 3.85–25.84%, with a median of 9.49%) and at least 100,000 total reads were selected for further assessment of their molecular damage patterns.

To be considered a candidate for target enrichment and capture, a sample was also required to exhibit a rate of nucleotide misincorporation at the terminal nucleotide at the 5'-end of approximately 5% or greater with a damage pattern across the DNA molecule consistent with expectations for aDNA (i.e., a clear increase of C→T substitutions in the last three base pairs at the 5'-end and corresponding G→A substitutions at the 3'-end and

[95] Phy-Mer is described in greater detail in Chapter 8.
[96] When a relatively high (>3%) endogenous yield was determined, the number of aligned reads was also examined to ensure that no anomalies occurred during the sequencing process leading to the misinterpretation of endogenous yield due to a low number (<100,000) of total reads.

little to no damage in the base pairs comprising the interior of the read, illustrated in Figure 6.9).

a)



b)

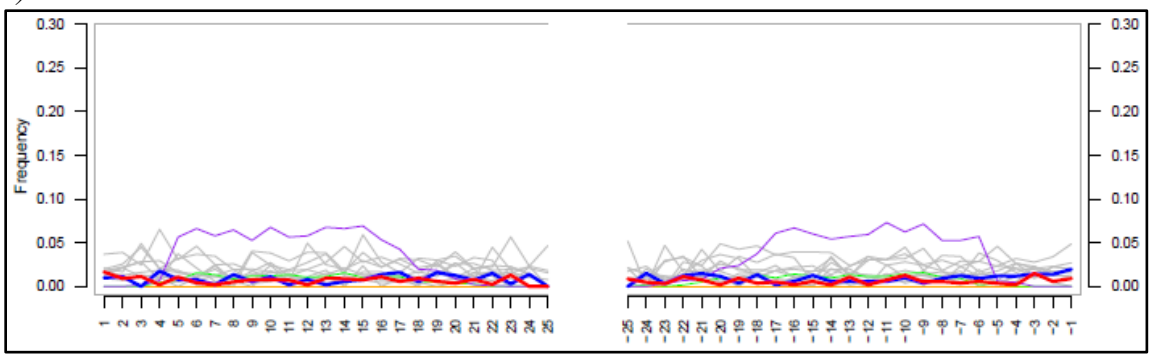


Figure 6.9: MapDamage plots: a) a plot consistent with typical patterns of aDNA damage; b) a plot inconsistent with typical patterns of aDNA damage.

The value of 5% is based on research showing that the frequency of nucleotide misincorporation at the terminal base of the 5'-end exhibits a strong and positive correlation with age (Sawyer et al. 2012). A recent analysis of ancient Egyptian mummy genomes ranging from the late New Kingdom to the Roman Period (cal. 1388 B.C.E.–426 C.E.) showed damage at both ends of sequencing reads ranging from 7–26% (Schuenemann et al. 2017). Because the samples analyzed in this dissertation are more recent, samples designated as having adequate molecular preservation are expected to exhibit nucleotide misincorporation at the last base at a frequency of approximately 5% at

the 5'-end. In the 30 samples with the highest endogenous DNA, this frequency ranged from 4–14%, with a median of approximately 9%. MapDamage plots for all screened samples are included in Supplement 3.

While assignment of molecular sex or mitochondrial haplogroup was not required for a sample to be considered a candidate for targeted capture, this information was determined when possible and is included in Supplement 2, with sexing plots included in Supplement 4.

Ultimately, based on 1) relatively high endogenous DNA yield and 2) a *post-mortem* damage pattern suggestive of authentic ancient DNA, 30 Kulubnarti samples (15 from each cemetery) were identified as best-preserved on a molecular level. The screening data for these 30 samples are presented in Table 6.3. All sample IDs starting with "R" designate the individual to be a member of the R cemetery; sample IDs starting with "S" designate the individual to be a member of the S cemetery.

Table 6.3: Screening metrics for the 30 best-preserved samples from Kulubnarti.

Sample ID	Macroscopic Preservation Assessment	Sequenc-ing Platform	Total Number of Reads	Reads Aligned to Human Genome	% Endo-genous DNA	Molecular Sex*	mtDNA Haplo-group Estimat-ion and Score	mapDamage (5' | 3')**
R101	V. GOOD	MiSeq	628647	33463	5.32	Male (p = 3.55e−08)	?	0.12 | 0.10 (+)
R124	GOOD	MiSeq	355955	26490	7.44	Male (p = 2.71e−07)	K1a12 (0.59)	0.09 | 0.09 (-/+)
R152	OKAY	NextSeq	10425922	2694260	25.84	Female (p = 1.02e−11)	L0a1a (0.83)	0.08 | 0.08 (+/-)
R169	GOOD	NextSeq	8783510	471162	5.36	Male (p = 1.02e−08)	U5b2b (0.86)	0.08 | 0.08 (+/-)
R181	V. GOOD	NextSeq	9390239	505956	5.39	Male (p = 1.99e−08)	H2a (0.86)	0.08 | 0.09 (+/-)

R182	GOOD	MiSeq	872056	128204	14.7	Female (p = 2.6e−11)	U5b2b3 (0.53)	0.09 | 0.08 (+/-)
R186	GOOD	MiSeq	446974	35252	7.89	Female (p = 2.89e−10)	L2a1m1 (0.52)	0.08 | 0.09 (+)
R195	OKAY	NextSeq	10472395	967756	9.24	Female (p = 1.3e−11)	U3b2 (0.84)	0.11 | 0.11 (+)
R196	OKAY	MiSeq	209000	13467	6.44	Male (p = 2.81e−08)	?	0.09 | 0.10 (+)
R201	V. GOOD	MiSeq	857390	33044	3.85	Male (p = 2.64e−08)	J2a2c (0.57)	0.09 | 0.09 (++)
R202	OKAY	NextSeq	1160922	82469	7.10	Female (p = 2.09e−11)	H65 (0.62)	0.13 | 0.13 (++)
R5	OKAY	MiSeq	302561	39045	12.9	Female (p = 2.23e−10)	N1b1a2 (0.60)	0.11 | 0.11 (+)
R59	V. GOOD	NextSeq	9765699	450370	4.61	Male (p = 3.38e−08)	H2a (0.76)	0.13 | 0.13 (+)
R79	GOOD	MiSeq	566069	55262	9.76	Female (p = 1.26e−10)	U1a1 (0.54)	0.12 | 0.12 (++)
R93	GOOD	MiSeq	439737	28205	6.41	Male (p = 1.6e−07)	H65 (0.73)	0.11 | 0.09 (+/-)
S133	OKAY	MiSeq	275266	13612	4.95	Female (p = 5.93e−10)	L2a1d1 (0.55)	0.04 | 0.04 (+/-)
S149	OKAY	NextSeq	12738113	690030	5.42	Female (p = 7.09e−12)	H2a (0.83)	0.07 | 0.07 (+/-)
S159	OKAY	MiSeq	312079	61918	19.84	Male (p = 9.51e−07)	L2a1d1 (0.72)	0.09 | 0.08 (+/-)
S17	OKAY	MiSeq	351496	19435	5.53	Female (p = 1.56e−08)	T1b2 (0.57)	0.10 | 0.09 (+)
S198	GOOD	MiSeq	679307	80760	11.89	Male (p = 5.5e−08)	L1b1a8 (0.54)	0.06 | 0.06 (+/-)
S208	BAD	MiSeq	1375987	310255	22.55	Female (p = 1.89e−11)	J2a2 (0.82)	0.14 | 0.14 (+)
S27	GOOD	MiSeq	316464	59261	18.73	Male (p = 8.21e−09)	U5b2b (0.58)	0.07 | 0.06 (-)
S33	GOOD	NextSeq	9597747	614629	6.40	Male (p = 7.12e−09)	L0a1a (0.80)	0.09 | 0.09 (+/-)

S50	GOOD	NextSeq	9695589	1431080	14.76	Male (p = 1.15e−08)	L0a1a (0.85)	0.05 | 0.05 (+/-)
S68a	GOOD	NextSeq	11762033	1836921	15.62	Female (p = 4.38e−11)	L2a1d1 (0.85)	0.09 | 0.09 (+)
S73	GOOD	MiSeq	429216	103680	24.16	Male (p = 5.45e−08)	L2a1d1 (0.57)	0.08 | 0.08 (++)
S79	GOOD	MiSeq	399354	38864	9.73	Male (p = 3.23e−07)	L5a1b (0.58)	0.07 | 0.07 (+)
S81	OKAY	NextSeq	9221125	1297503	14.07	Male (p = 4.26e−09)	L2a1d1 (0.85)	.08 | .09 (+/-)
S87	GOOD	MiSeq	457836	111512	24.36	Female (p = 5.06e−10)	L2a5 (0.52)	0.10 | 0.09 (+)
S89	V. GOOD	MiSeq	700103	135517	19.36	Female (p = 4.01e−12)	H2a1e1 (0.58)	0.07 | 0.07 (+/-)

** sexing plots are provided in Supplement 4*
*** mapDamage plots are provided in Supplement 3; the author assessed all MapDamage results as ++ if the damage pattern was strongly consistent with ancient DNA, + if the damage pattern was consistent with ancient DNA, - if the damage pattern either consistent with ancient DNA but exhibited a low frequency of damage (<5%) at the ends of the molecule or if the frequency at damage at the end of the molecule was >5% but the damage pattern throughout the molecule was inconsistent with ancient DNA, and x if there was no evident damage.*

For each of these 30 samples, an aliquot of ~100mg of cochlear bone powder was separated from the surplus supply created when the samples were first processed. This new powder aliquot was transferred into a sterile 2.0mL tube in the grinding lab at UCD, placed in a plastic bag, and sent to the Reich Lab for another round of processing and screening leading to target enrichment and capture. An initial set of 19 powders were brought to the Reich Lab by the author in July 2016, and the final 11 powders were sent by mail in October 2016. The lab and analytical work presented in the remainder of this chapter took place in the Reich Lab.

6.4 Sample Preparation and Screening #2: MTspike3k Capture

6.4.1 DNA Extraction

Using the new bone powder aliquot, DNA extraction took place in the Reich Lab's dedicated aDNA facility that adhered to the same stringent anti-contamination standards followed at UCD and discussed in the previous sections. DNA was extracted from ~75mg (±10mg) of bone powder for 29 of the 30 Kulubnarti samples following the Dabney protocol, reviewed previously in this chapter and described in step-by-step detail in Appendix A, but replacing the custom-made binding apparatus comprising MinElute spin columns and Zymo-spin V extension reservoirs with a pre-assembled High Pure Extender Assembly from Roche's High Pure Viral Nucleic Acid Large Volume Kit (Korlević et al. 2015).[97]

6.4.2 Construction of Libraries

Library preparation followed the protocol from Rohland et al. (2015) with slight modification to enzymatic reactions and by replacing MinElute spin column clean-ups between the various preparation steps with silica-coated magnetic bead clean-ups (manuscript in preparation) and SPRI purification after the library amplification (Rohland and Reich 2012). An Agilent Bravo Workstation was used for the entire library preparation and a Perkin Elmer EP3 or Agilent Bravo Workstation was used for the final SPRI purification to allow the processing of up to 96 libraries in parallel.

The Rohland et al. (2015) protocol is optimized for the construction of libraries from damaged DNA molecules that are likely to contain nucleotide misincorporations due to the removal of an amino group (NH_3) that converts cytosine (C) to uracil (U), a demethylated form of thymine (T). These misincorporations result in "errors" during the blunt-end repair step of library preparation, both on the 5'-strand where the nucleotide

[97] Individual R195 was not screened.

misincorporation takes place, as well as on the 3'-strand where an adenine (A) is incorporated onto the complementary strand instead of guanine (G). To contend with these misincorporations, this protocol treats the DNA molecules with the enzymes uracil-DNA-glycosylase (UDG), which cleaves deaminated cytosines (i.e., uracils) and leaves abasic (baseless) sites (Lindahl et al. 1977), and Endonuclease VIII (Endo VIII), which splices the DNA at the resulting abasic sites. Together these enzymes are used to reduce DNA sequencing errors due to nucleotide misincorporations resulting from DNA damage; however, this treatment also "silences" the damage signal provided by nucleotide misincorporations that are predominantly present at the terminal ends of the DNA molecule that are used to assess aDNA authenticity (Ginolhac et al. 2011; Jónsson et al. 2013). Thus, the Rohland et al. (2015) protocol also introduces a "partial" UDG treatment that inefficiently removes terminal uracils, but repairs miscoding damage throughout the interior of the DNA strand.[98] By restricting aDNA damage to the terminal nucleotides while nearly eliminating it from the interior of the DNA molecule, the damage signal characteristic of ancient DNA is maintained, and a single library can be used to test both for aDNA authenticity as well as to carry out population genetics analysis. This partial UDG treatment was applied to the Kulubnarti samples.

Partial UDG treatment occurred in the blunt-end repair step of library preparation. Uracil Specific Excision Reagent (USER) enzyme, comprising both UDG and Endo VIII, was added to a mastermix containing buffers to ensure optimum reaction conditions and dNTPs to provide the nucleotides required by blunt-end repair. The mastermix was added to the DNA extracts and the reactions were incubated to allow the excision of deaminated

[98] As for all UDG treatments, damage in CpG sites (regions of DNA where a cytosine base occurs next to a guanine base on the same DNA strand) is never removed.

cytosines by UDG and cutting of abasic sites by Endo VIII. After the initial incubation period, the activity of USER enzyme was inhibited by the addition of Uracil Glycosylase Inhibitor (UGI) and the reactions were incubated again. After this second incubation, polymerase to repair overhanging 5'- and 3'-ends through 5'→3' polymerase activity and 3'→5' exonuclease activity and kinase to add 5'-phosphates for subsequent adapter ligation were added to the reactions, and the reactions were incubated for a third time to allow blunt-end repair to occur. The omission of the DNA polymerase and kinase during the UDG treatment retained some of the deaminated cytosines at the terminal ends of the DNA molecule due to unphosphorylated uracils at the 3'-end that were not efficiently excised by UDG (Varshney and van de Sande 1991); this resulted in the partial treatment.

To allow highly-parallelized library processing, the reactions were cleaned using silica-coated magnetic beads and a magnet instead of centrifugation. With this method, silica beads were added to the blunt-end repaired reactions and binding buffer was added to allow the DNA molecules in solution to hybridize with these beads. After a period of incubation, magnetic silica particles were separated on a magnet and were washed twice with ethanol wash buffer. After an air-drying step, the beads were eluted and were again separated on a magnet. The supernatant containing the blunt-end repaired DNA in solution was collected.

Adapters were then ligated to either end of the blunt-end repaired eluates. The Rohland et al. (2015) protocol implements two unique optimizations not present in other Illumina-based library preparation protocols. First, internal 7mer (7bp) barcodes are attached directly to the DNA insert to eliminate undetected cross-contamination from the ligation step onward and permit highly multiplexed sample pooling in a process known as

"dual barcoding."[99] Second, the adapters flanking these internal barcodes are incomplete adapters of short length ("short adapters"), a specific optimization for target enrichment and capture based on research showing that the presence of short adapters increases the specificity of hybridization-based target enrichment (i.e., the desired genomic region is enriched and captured more often) (Rohland and Reich 2012, Rohland et al. 2015).[100]

Prior to library preparation, 100 barcodes of 7bp length were designed as in Meyer and Kircher (2010) and partial double-stranded adapters (P5 and P7) were prepared by hybridizing the long oligonucleotide (with truncated Illumina-specific universal adapter sites) to the reverse complementary short oligonucleotide and including 7mer barcodes at both ends. For each library, a barcoded, partially double-stranded P5-adapter and a barcoded, partially double-stranded P7-adapter were combined with the purified blunt-end repaired eluates and mixed to bring the DNA molecules and adapters into close proximity; a unique barcode combination was chosen for each sample and no barcode was repeated in the same batch of samples (Rohland et al. 2015). A ligation mastermix containing a ligase to attach the adapters onto either end of the DNA molecules and buffers to optimize the ligation process was then assembled and combined with the blunt-end repair eluate and adapters. The reactions were incubated to allow adapter ligation to occur. Reaction clean-up with silica-coated beads was again performed.

---

[99] This differs from the Meyer and Kircher (2010) protocol used to prepare libraries for sample screening because instead of ligating universal adapters and relying on indexing to differentiate samples, barcoded Illumina adapters are ligated directly to DNA molecules.

[100] Higher specificity results in less sequencing necessary to generate adequate coverage of sequence data for downstream analysis, thereby making sequencing more economical (Mamanova et al. 2010). After target enrichment and hybridization capture takes place, these short adapters are completed in a dual-indexing PCR reaction before sequencing as in Kircher et al. (2012).

An adapter fill-in mastermix was then assembled that contained the requisite dNTPs in an equimolar solution, Bst polymerase to incorporate the dNTPs to the adapter-ligated strand, and buffers to provide superior reaction conditions. The mastermix was combined with the adapter-ligated eluates, and the reactions were incubated to allow adapter fill-in to occur. Immediately after this incubation step, the reactions were incubated again at a high temperature (80ºC) to inactivate the Bst polymerase.

The entire fill-in reaction was then amplified by PCR. A PCR mastermix was prepared comprising two universal primers that did not extend the short adapter sites, but kept them truncated, resulting in the presence of short adapters for post-PCR target enrichment. Based on the temperature cycling profile recommended for the polymerase used in the PCR mastermix, PCR amplification was initiated with a denaturation stage for 2 minutes at 95ºC , allowing the "hot-start" polymerase to regain enzymatic activity. The following temperature cycling profile was then used for amplification: 30 cycles of 30 seconds at 95ºC, 30 seconds at 55ºC, and 1 minute at 72ºC.[101] Following the completion of the cycling, a final elongation step was performed at 72ºC for 10 minutes.

PCR products were then purified using SPRI purification as in Rohland and Reich (2012). The use of paramagnetic carboxyl-coated beads (known as SPRI technology, DeAngelis et al. 1995) allowed up to 96 samples to be purified in parallel and permitted size selection, allowing the removal of PCR primer or adapter dimers (primer or adapter molecules that have hybridized to each other because of strings of complementary bases) and long contaminant molecules before capture or sequencing (Quail et al. 2008; Lennon

[101] This temperature cycling profile differs slightly from that used in the PCR of the libraries used in the screening step. This is because different enzymes were used, and each enzyme has its own optimal temperature cycling profile.

et al. 2010). After purification, the PCR products were stored in a freezer maintained at -20ºC until further screening took place.

6.4.3 MTspike3k Capture

The amplified and purified libraries were then screened for biomolecular preservation using the Reich Lab's in-house screening method consisting of target enrichment and in-solution hybridization capture of selected sequences overlapping the complete mitochondrial genome (Meyer et al. 2014) and ~3000 SNPs from the nuclear genome (including ~2500 ancestry informative markers and ~500 SNPs on the Y chromosome) (Fu et al. 2015; Haak et al. 2015). This capture is referred to as the "MTspike3k" capture.

The MTspike3k screening step was used to assess contamination, aDNA authenticity, and library complexity (the expected number of distinct molecules that can be observed in a set of sequencing reads) to identify promising samples for 1240K capture. Because the reagents and details surrounding the specifics of the MTspike3k capture are exclusive to the Reich Lab at present, specific details of this screening step are not provided in this dissertation; however, a general overview of the capture process is presented, and relevant protocols are cited.

Each purified library was enriched and captured using the MTspike3k capture following the methodology described in Fu et al. (2013a) with modifications described in Meyer et al. (2014) including the lowering of hybridization and wash temperatures to facilitate the enrichment of short library molecules (Dabney et al. 2013).

An aliquot of each library (including ~500ng DNA) was enriched for the desired sequences in two consecutive rounds of targeted enrichment using a custom-designed set

of biotinylated capture probes (details of probe design found in Fu et al. 2013a). For each reaction, a sample library pool was created by combining an aliquot of sample library with blocking agents that reduce non-specific hybridization and improve on-target capture percentage. The reaction was incubated, allowing hybridization of library and blocking agents to occur. Subsequently, hybridization reactions were assembled by combining the sample library pool with the bait probe pool that contained single-stranded biotinylated probes representing the target DNA sequences. These hybridization reactions were incubated to allow hybridization of target strands and bait probes to occur.

Multiple wash steps followed this incubation to remove all non-target DNA using streptavidin-coated beads that captured the biotinylated probes hybridized to the target DNA fragments. After the last wash step, the target molecules were released into solution, purified using Qiagen's QIAquick Nucleotide Removal Kit (following the manufacturer's instructions) and eluted in elution buffer.

An aliquot of capture eluate was quantified with quantitative PCR (qPCR) to verify the successful retrieval of library molecules using a SYBR-Green Assay from a KAPA library quantification kit following the guidelines of the manufacturer.[102] Amplifying PCR was then performed on the capture eluate with the same two universal primers that kept short adapter sites truncated. Based on the temperature cycling profile recommended for the polymerase used, PCR amplification was initiated with a denaturation stage for 2 minutes at 95ºC, allowing the "hot-start" polymerase to regain enzymatic activity. The following temperature cycling profile was then used for

[102] While the P5 and P7 primers in the KAPA kit are designed to bind to the full-length universal adapters ligated to the ends of Illumina NGS libraries, the library preparation protocol followed instead uses truncated adapters to optimize for target enrichment. Therefore, the primers supplied with the KAPA kit were replaced with the primers IS7_short_amp.P5 and IS8_short_amp.P7 from Meyer and Kircher (2010).

amplification: 30 cycles of 30 seconds at 95ºC, 30 seconds at 55ºC, and 30 seconds at 72ºC. Following the completion of the cycling, a final elongation step was performed at 72ºC for 5 minutes. PCR products were then cleaned using SPRI purification following the protocol in Rohland and Reich (2012).

A second round of hybridization was then performed using an aliquot of capture eluate from the first round of hybridization as a template. This second round of hybridization was performed under the same reaction conditions as the first round, except for a reduced incubation time for sample library/probe hybridization (Fu et al. 2013a). Capture eluates were again quantified by qPCR, amplified by PCR, and purified with SPRI purification.

6.4.4 Indexing PCR and Sequencing of Capture Products

After two rounds of target enrichment and capture, the eluates were diluted to a concentration of ~5ng/μL. Indexing PCR was then performed on the diluted capture product using a "hot-start" polymerase and two indexing primers (Kircher et al. 2012). These primers added dual indices to each library and extended the short adapter sites to their final length. Because the original amplification was completed with universal primers that maintained the short adapters in order to optimize specificity during target enrichment, enriched libraries were amplified a second time with indexing primers that encoded an index sequence in the oligonucleotide sequence of the primer and extended the short adapters to full-length adapter sites (Kircher et al. 2012).

Based on the temperature cycling profile recommended for the polymerase used, PCR amplification was initiated with a denaturation stage for 2 minutes at 95ºC, and the following temperature cycling profile was used for amplification: 6 cycles of 30 seconds

at 95ºC, 30 seconds at 55ºC, and 30 seconds at 72ºC. Following the completion of the cycling, a final elongation step was performed at 72ºC for 5 minutes. PCR products were then cleaned using SPRI purification following the protocol in Rohland and Reich (2012). DNA concentration was determined, and samples were pooled in equimolar ratios for sequencing.

The enriched capture products were then sequenced with an Illumina NextSeq500 instrument using identical sequencing technology as described previously in this chapter. However, instead of single-end sequencing, in which only the forward DNA strand was read, the current screening step utilized paired-end sequencing for 2 x 76 + 2 x 7 cycles (paired-end sequencing of the DNA molecule + paired-end sequencing of the index sequence; here, the "2" denotes paired-end sequencing), which consists of sequencing both the forward and reverse strands of each DNA molecule, generating approximately twice the amount of data as single-end sequencing (Kircher and Kelso 2010).

6.4.5 Processing of Sequencing Data

Raw sequencing data were run through the Reich Lab's custom NGS data-processing pipeline that provided the necessary information to assess contamination, aDNA authenticity, and library complexity to identify promising samples for 1240K SNP capture.[103]

The sequencing data were demultiplexed (i.e., reads were sorted by their index sequence combinations), requiring that the P5 and P7 indices matched and allowing a 1bp mismatch. Accurate demultiplexing was further ensured by requiring that the 7mer

[103] Details in this subsection were provided by members of the Reich Lab.

internal barcodes added during the adapter ligation step of library preparation matched, again allowing a 1bp mismatch. Forward and reverse reads were merged using *SeqPrep* (github.com/jstjohn/SeqPrep), ensuring that the highest quality base was retained in any overlapping region of merged reads. Forward and reverse reads were required to have a minimum length of 30bp with at least a 15bp overlap (allowing a 1bp mismatch). The merging of forward and reverse reads generated single-ended reads with reduced error rates in their overlapping sequence (Briggs et al. 2010; Kircher and Kelso 2010). Adapter sequences were also trimmed using *SeqPrep*.

Merged reads were mapped to the human reference genome with BWA v0.6.1 (Li and Durbin 2009) using the *samse* command and parameters '-n 0.01 -o 2 -1 16500' to allow for more sequence mismatches and nucleotide insertions or deletions ("indels"), as well as to turn off seeding (Meyer et al. 2012). For mitochondrial analysis, the Reconstructed Sapiens Reference Sequence (Behar et al. 2012), or RSRS, mitochondrial genome was used (human mtDNA sequence NC_001807); for nuclear analysis, the hg19 build of the human genome was used.[104] Duplicate sequences were removed by identifying clusters of reads with the same stop and start position and the same mapped orientation.

Evaluation of ancient DNA authenticity and estimation of contamination was then performed. The software mapDamage2.0 (Jónsson et al. 2013) was used to determine the frequency of nucleotide misincorporation at the ends of the DNA molecules and to assess whether the aligned reads exhibited a *post-mortem* degradation pattern typical of ancient

[104] The rationale behind the use of the RSRS can be found in Behar et al. (2012). Briefly, the RSRS uses a hypothetical ancestral reference sequence as a root in contrast to the European mitogenome that serves as a root for the rCRS.

samples. The software contamMix 1.0-10, a likelihood-based method of contamination estimation, was used to assess levels of mitochondrial contamination (Fu et al. 2013b). This method compares the mapping affinities of an individual's mtDNA sequences to its own consensus mitochondrial genome relative to the mapping affinity of its mtDNA sequences to a dataset of potential contaminants represented by 311 mitochondrial genomes from worldwide populations encompassing all potential contaminating sequences.

In addition, mtDNA haplogroup was assigned for each individual using HaploGrep2 (Kloss-Brandstätter et al. 2011; Weissensteiner et al. 2016), a program that determines haplogroup by traversing the most current version of the known phylogenetic tree, PhyloTree Build 17, comprising 5,437 haplogroups (van Oven 2015). Sex was assessed using the method in Skoglund et al. (2013) that examines the ratio of reads aligned to chromosome Y to those aligned to both sex chromosomes (X and Y).[105]

The processed sequencing data were then examined for three key criteria determined by Dr. David Reich to indicate the sample's likelihood to be amenable to 1240K SNP capture (per Dr. David Reich, pers. comm.). First, it was preferred that a sample exhibit a >0.98 (98%) match rate between a sample's mtDNA sequences and its consensus mitochondrial genome as analyzed using contamMix, though values >0.80 (80%) were accepted. A high value here indicated that the sample library was mainly composed of endogenous aDNA sequences that were likely to provide an accurate

[105] In contrast to the previously described method of assigning sex based on the number of sequenced reads plotted against the length of the chromosome (Dr. Eppie Jones, unpublished research), which uses only X-chromosomal genetic material, the presently described approach utilizes both sex chromosomes based on the expectation that males have half the amount of X chromosomal genetic material as females in addition to Y chromosomal genetic material.

assessment of an individual's genetic composition. Second, the frequency of nucleotide misincorporation at the terminal molecules was required to be at least 3% for any libraries partially treated with UDG (Rohland et al. 2015), indicating that the DNA present is likely ancient.[106] Third, library complexity was used as a proxy for projected coverage after 1240K capture. Here, complexity estimations were made using multiple regression with the number of reads mapping to the target prior to and after duplication as the independent variable and the genome-wide coverage at 10% marginal uniqueness as the dependent variable. The coefficients of this model were trained and subsequently used in the prediction of coverages of new samples based on screening data. Here, it was preferred that a sample exhibit a complexity value of >0.1, though values >0.01 were accepted.

6.4.6 Screening Results

Of the 30 samples sent to the Reich Lab, 29 were screened between May and July 2017. Screening data for these samples are presented in Table 6.4. Samples were tagged as "warn" if the metric being measured was below the preferred value and "fail" if the metric being measured was below the accepted value; if a sample received a "fail" in any metric, the other metrics were examined to determine if that sample would be amenable to 1240K SNP capture and a final assessment was made. All assessments were made by the Reich Lab. Any sample assessed as "pass" or "warn" was promoted to 1240K SNP capture. Consistent with all previous tables, sample IDs starting with "R" designate the

[106] While Sawyer et al. (2012) suggest using a value of ≥10% as a threshold to call a library plausibly authentic, Rohland et al. (2015) note a threefold reduction of substitutions seen in partially UDG-treated libraries and propose using a damage threshold of 3% or higher when UDG is included in the library preparation protocol.

individual to be a member of the R cemetery; sample IDs starting with "S" designate the individual to be a member of the S cemetery.

Table 6.4: MTspike3k screening metrics for the 29 Kulubnarti samples. Yellow cells indicate sub-optimal but passable screening result ("warn"); red cells indicate sample failure at a particular screening metric ("fail").

Sample ID	Total Number of Reads (nuclear)	% Endo-genous DNA	Total Number of Reads (mtDNA)	% Sequences Aligning to mtDNA Reference	% Damage in Last Base	mtDNA Match to Consensus Sequence [95% CI]	Comp-lexity	**Assess-ment**
R101	926127	5.82	2059578	1.27	6.4	0.998 [0.996,0.999]	0.06	**Pass**
R124	1201980	12.49	1748838	8.5	4.1	1.0 [0.998,1.0]	0.10	**Pass**
R152	494693	46.07	1444096	27.11	4.4	1.0 [0.999,1.000]	0.10	**Pass**
R169	497224	16.22	953584	24.53	3.1	1.0 [0.998,1.000]	0.07	**Pass**
R181	659805	11.41	1732678	22.69	3.3	1.0 [0.998,1.000]	0.11	**Pass**
R182	1002663	20.21	1371290	2.32	7.3	0.993 [0.989,0.996]	0.31	**Pass**
R186	1010750	13.24	1727473	5.68	4.7	1.0 [0.999,1.000]	0.32	**Pass**
R196	18005	14.5	1177229	15.75	5.6	1.0 [0.999,1.000]	0.26	**Pass**
R201	415307	7.52	1045687	12.08	3.3	1.0 [0.999,1.000]	0.09	**Pass**
R202	810843	10.79	1152313	15.39	4	1.0 [0.996,1.000]	0.14	**Pass**
R5	512069	18.29	1345999	17.09	7.1	0.989 [0.981,0.995]	0.10	**Pass**
R59	490297	6.07	1891885	2.44	7.6	1.0 [0.998,1.000]	0.07	**Pass**
R79	378129	19.43	1346497	7.92	5.9	1.0 [0.998,1.000]	0.31	**Pass**
R93	591350	12.59	1729798	26.26	3.8	1.0 [0.998,1.000]	0.07	**Pass**
S133	664173	6.23	1745259	15.28	2.5	1.0 [0.999,1.000]	0.34	**Warn**
S149	1141047	12.36	2130402	26.05	2.5	1.0 [0.998,1.000]	0.10	**Warn**
S159	411178	43.03	1091787	32.03	2.8	0.993 [0.989,0.995]	0.12	**Warn**
S17	491345	8.6	1897264	7.33	3	1.0 [0.999,1.000]	0.07	**Pass**
S198	475975	17.46	1539048	0.05	1.3	0.982 [0.646,0.997]	0.004	**Warn**

S208	72479	3.55	567363	4.99	0.3	1.0 [0.999,1.000]	0.01	**Fail**
S27	390271	28.56	1397507	18.38	2.9	1.0 [0.999,1.000]	0.42	**Warn**
S33	486850	12.15	1382246	9.59	3.5	1.0 [0.999,1.000]	0.31	**Pass**
S50	572029	22.79	1527001	18.13	2.6	0.997 [0.995,0.998]	0.42	**Warn**
S68a	435688	22.21	1727949	18.31	3	1.0 [0.999,1.000]	0.32	**Pass**
S73	619258	39.02	1251903	14.9	3.7	1.0 [0.999,1.000]	0.40	**Pass**
S79	461605	20.95	1340162	18.24	2.8	1.0 [0.998,1.000]	0.34	**Warn**
S81	633023	23.46	1528708	15.47	3.3	1.0 [0.999,1.000]	0.30	**Pass**
S87	445104	37.74	728750	12.03	5.3	0.998 [0.997,0.999]	0.39	**Pass**
S89	424643	33.82	873641	13.95	3.1	1.0 [0.998,1.000]	0.34	**Pass**

Of the 29 samples screened, 28 were assigned a “warn” or “pass” and were promoted to 1240K SNP capture. One sample, S cemetery individual S208, exhibited a very low damage rate at the terminal nucleotide, suggesting that the library contained a high amount of contaminating modern DNA molecules. Additionally, the complexity of the library was very low, suggesting that there were few unique library molecules and that additional capture and sequencing would produce redundant data (Daley and Smith 2013). Interestingly, S208 was the only sample designated as “bad” when macroscopic preservation was assessed. Because of these poor screening results, 1240K SNP capture was not performed on this sample.

6.5 1240K SNP Capture

To generate the data necessary to test the proposed hypotheses regarding community relationships and genetic composition at Kulubnarti, target enrichment and capture of 1,237,207 (i.e., ~1.24 million, or 1240K) genome-wide SNPs was performed

between January and June 2017 on the 28 Kulubnarti samples assigned a "warn" or "pass" and assessed as likely to be amenable to genome-wide capture.

This 1240K SNP capture was first reported in Mathieson et al. (2015) and has been utilized widely in recent aDNA studies (e.g., Lazaridis et al. 2016; Skoglund et al. 2016, 2017; Mathieson et al. 2018; Olalde et al. 2017; Lipson et al. 2018). The targeted SNP set merges 394,577 SNPs first reported in Haak et al. (2015) with 842,630 SNPs first reported in Fu et al. (2015). It comprises nearly all SNPs on the Affymetrix Human Origins and Illumina 610-Quad arrays, ~50,000 SNPs on chromosome X, ~33,000 SNPs on chromosome Y, and ~47,000 SNPs with evidence of functional importance (Mathieson et al. 2015).

6.5.1 Benefits of 1240K SNP Capture

As low levels of endogenous yield and high levels of microbial contaminant DNA render shotgun sequencing for anything other than low-coverage sample screening uneconomical for most ancient samples (Pickrell and Reich 2014), a targeted approach was required to generate adequate amounts of sequencing data for population genetic analysis in this dissertation.

The benefits of the 1240K SNP capture used in this dissertation's analysis are twofold. First, this capture targets only the SNP sites from across the genome that will be utilized in population genetics analysis, increasing analytic efficiency and reducing the amount of sequencing required to generate adequate coverage of sequencing data (Mathieson et al. 2015). Second, this capture provides access to genome-wide data from ancient samples with small fractions of endogenous human ancient DNA that would not be otherwise amenable to whole-genome shotgun sequencing. The analysis of genome-

wide SNP data represents the theoretical analysis of a large effective sample size (i.e., hundreds to thousands of the individual's ancestors), making these SNP data a powerful tool for population genetics analysis. The use of this capture enables this dissertation's population genetic analysis of the Kulubnarti Nubians.

6.5.2 1240K SNP Capture

Target enrichment and capture were performed following the protocol in Fu et al. (2013a) with modifications described in Meyer et al. (2014); the general capture methodology is discussed in greater detail previously in this chapter in reference to the MTspike3k capture. Once the 1240K SNP capture products were amplified by PCR and purified by SPRI technology, they were pooled in equimolar ratios and sequenced on an Illumina NextSeq 500 with 2 x 76 + 2 x 7 cycles (as for the MTspike3k capture, paired-end sequencing was used). The enriched products were sequenced up to the point that it was estimated that the generation of 100 new sequences was expected to add data of approximately one or fewer than one new SNP (Mathieson et al. 2015).

6.5.3 Data Processing and Alignment

Raw sequencing data were processed at the Reich Lab using their automated data analysis pipeline. Data processing and alignment used the same procedure as described for the analysis of MTspike3k capture data, but without requiring the mitochondrial alignment to the RSRS; instead, only the hg19 build of the human genome was used as a reference for alignment.[107] Sex was assigned using the method in Skoglund et al. (2013), described previously in this chapter. If the individual was determined to be male, Y

[107] The mitochondrial genome was not included in the 1240K capture. All mtDNA haplogroup data comes from the MTspike3k capture.

chromosome haplogroup was assigned using the nomenclature in version 11.110 of the International Society of Genetic Genealogy (version 11.110, April 2016; http://www.isogg.org). For Y chromosome haplogroup assignment, SAMtools (Li et al. 2009) was used to restrict analysis to sites with mapping quality Q>30 and two bases at the end of each sequenced fragment were excluded due to a high probability of *post-mortem* damage at these sites.

Aligned capture data were subsequently provided to the author in bam format, a binary format for storing sequence data. All additional population genetics analyses of the Kulubnarti Nubians using the 1240K capture data were performed by the author. The results of these analyses were used to test the hypotheses associated with the two research aims that guide this dissertation, presented in the following two chapters.

CHAPTER 7
GENETIC RELATIONSHIPS AT KULUBNARTI

7.1 Chapter Overview

While the archaeological and bioarchaeological analyses discussed previously in this dissertation have consistently suggested that two biologically-indistinguishable but socially-distinct communities lived side-by-side at Kulubnarti, the relationship between the communities has not yet been explored from a genomic perspective. Through the analysis of genome-wide Single Nucleotide Polymorphisms (SNPs), this dissertation investigated the genetic relationship between the Kulubnarti S and R communities at a higher resolution than has previously been possible. Specifically, **Aim 1 of this dissertation was to explore the genetic relationship between the S and R communities by looking for evidence of community-based genetic population substructure through patterns of clustering and estimation of genetic distance.**

This chapter introduces the analytical methods used to achieve Aim 1, discusses the application of these methods to the Kulubnarti data, and presents and interprets the ensuing results. A discussion of how these results shape a broader understanding of Early Christian Kulubnarti from a genomic perspective is provided in Chapter 9, following the presentation of all results.

7.2 The 1240K Human Origins Dataset

The initial dataset prepared by the author to investigate the genetic relationships and genetic composition of the Kulubnarti Nubians contained SNP data from 3,341 present-day individuals from 299 worldwide populations genotyped on the Affymetrix

Human Origins array (Patterson et al. 2012) at 597,573 SNPs (including 592,924 autosomal SNPs) for which ascertainment is well known (Keinan et al. 2007); it is referred to in this dissertation as the 1240K Human Origins ("1240K_HO") dataset.[108] The 1240K_HO dataset was chosen because of this dissertation's primary focus on modern populations genotyped on the Human Origins array as a comparative reference sample.[109]

For the analyses described in this chapter and in Chapter 8, genome-wide SNP data from the 28 best-preserved Kulubnarti Nubians were overlapped with the 1240K_HO dataset, resulting in 40,532–390,045 SNPs covered at least once with a median coverage of 162,577 SNPs. Table 7.1 presents the number of SNP hits for each Kulubnarti sample using the 1240K_HO dataset. It has been recommended that each sample have a minimum coverage of 10,000–20,000 SNPs for accurate and high-resolution population genetics analysis (Allentoft et al. 2015); all Kulubnarti samples analyzed in this dissertation met this threshold. Consistent with previous tables, sample IDs starting with "R" designate the individual to be a member of the R cemetery and sample IDs starting with "S" designate the individual to be a member of the S cemetery.

Table 7.1: Number of SNP hits for each Kulubnarti sample when overlapped with the 1240K_HO dataset.

Sample ID	SNP hits
R101	40,532

[108] The samples included in this dataset were first published in various manuscripts, including Patterson et al. (2012); Pickrell et al. 2012; Qin and Stoneking (2015); Lazaridis et al. (2014, 2016); Broushaki et al. (2016); and Skoglund et al. (2015, 2016). All present-day individuals that were genotyped provided individual informed consent consistent with studies of population history, following protocols approved by the ethical review committees of the institutions of the researchers who collected the samples.

[109] Use of the Reich Lab's complete 1240K SNP dataset would only provide additional information if most of the desired reference dataset comprised ancient individuals who had been sequenced initially on the 1240K capture array (Mathieson et al. 2015). The ~600,000 SNPs on the Human Origins array are also present in the Reich Lab's dataset.

R124	70,671
R152	158,897
R169	157,713
R181	166,257
R182	139,395
R186	144,814
R196	195,497
R201	94,911
R202	148,070
R5	120,673
R59	47,532
R79	143,403
R93	199,760
S133	272,831
S149	189,499
S159	330,410
S17	60,031
S198	189,720
S27	327,352
S33	150,912
S50	310,847
S68a	182,871
S73	278,256
S79	308,968
S81	152,938
S87	390,045
S89	303,592

SNP data from the Altai Neandertal (Prüfer et al. 2014) and Denisovan (Meyer et al. 2012) were also overlapped with the 1240K_HO dataset to provide a potential outgroup.

The genotypes of the Kulubnarti Nubians were called by randomly sampling a single non-duplicate sequence read at each position (Green et al. 2010) to create “pseudo-

haploid" genotypes. The use of pseudo-haploid genotypes is standard when analyzing low-coverage ancient DNA (aDNA) data with insufficient coverage for the assignment of a diploid genotype (Green et al. 2010; Lazaridis et al. 2014; Raghavan et al. 2014; Allentoft et al. 2015; Haak et al. 2015). The SNP data from present-day individuals included in the reference dataset are also made pseudo-haploid to match the nature of the genotypes of the ancient individuals through adjustments of associated parameters for each analysis.

The initial 1240K_HO dataset prepared by the author was manipulated according to the requirements of each analysis; for example, a reduced dataset was created for Principal Components Analysis (PCA) that included only autosomal SNPs from present-day individuals from geographic regions surrounding Kulubnarti. The dataset created for each analysis is described alongside the analytical methods.

7.3 Assessing Population Substructure at Kulubnarti

7.3.1 Analytical Method: Principal Component Analysis (PCA)

PCA is a non-parametric multivariate approach to studying population structure that was first applied to genetic data by Menozzi et al. (1978) and is now a standard tool in genetic analysis (Patterson et al. 2006). The PCA method works by reducing the dimensions of a large number of measurements (in the case of this dissertation, hundreds of thousands of genotyped SNPs) to a few Principal Components (PCs) that explain the main patterns of variation in those measurements (Reich et al. 2008). Unlike cluster-based analyses (such as ADMIXTURE, discussed in Chapter 8), PCA does not attempt to classify individuals into discrete clusters, but instead compares every individual against

every other individual for all genotyped SNPs through the generation of a covariance matrix where each number in the matrix represents the degree of similarity between two individuals as measured by genotyped SNPs (Patterson et al. 2006). It then projects each individual separately into PCA space based on the eigenvalues of the first two PCs of the data calculated from the covariance matrix of the genome-wide SNP set (Jobling et al. 2014).

PCA is most commonly used to provide an approximation of population biogeographic genetic affinities (Leonardi et al. 2016), but can also be used to qualitatively explore population substructure. When applied to the analysis of ancient samples, PCA is particularly useful for displaying the distribution of ancient human genetic variation in comparison to that of modern populations as well as for identifying individuals who are outliers to that genetic variation, as discussed in this chapter.

To assess potential population substructure at Kulubnarti based on the genetic similarity of the individuals from the two communities in relation to a background dataset of geographically-relevant present-day individuals, the SNP data from the present-day individuals included in the dataset were first used to compute PCs. The first PC (PC1, on the x-axis) represented the combination of measurements that accounted for the largest amount of variation in the data, and the second PC (PC2, on the y-axis) represented the combination of measurements that accounted for the second largest amount of variation in the data (Reich et al. 2008).

While it has been shown that placing ancient individuals within PCA plots produced using modern data is possible, even when only low-coverage genetic data are available (Skoglund et al. 2012), ancient individuals should not be included when

computing PCs due to large amounts of missing data relative to present-day individuals.[110] Instead, the PCA space is built using the SNP data from present-day individuals, and the SNP data from ancient individuals are subsequently merged with the dataset and projected into the PCA space.

Therefore, PCs were computed using SNP data from present-day individuals only, and each present-day individual was projected in PCA space as a separate data point according to their eigenvalues for the first two PCs of the data calculated from the covariance matrix of the SNP dataset.[111] As discussed in Chapter 4, individuals that shared more genetic similarity appeared to group as a population in the PCA space in comparison to individuals with whom they shared less genetic similarity, who appeared separated in the PCA space. When studying human genetic variation, it has been observed that PCs often correlate with geography (e.g., Li et al. 2008; Novembre et al. 2008), reflecting major migrations as well as isolation by distance, whereby genes are only exchanged between neighboring populations (i.e., the geographic proximity of the populations is the primary determinant of how closely the populations are related) (Reich et al. 2008).

Following the creation of the PCA space using SNP data from present-day individuals, each ancient individual was merged with the dataset and plotted as a separate data point in PCA space according to their eigenvalues for the first two PCs of the data calculated from the covariance matrix of the SNP dataset. As with present-day individuals, each ancient individual was represented by a separate data point. In this

[110] The use of present-day individuals only to compute PCs is common practice in ancient DNA research.
[111] This covariance matrix is analogous to a distance matrix between populations and is visualized in two dimensions using PCA.

dissertation, the projected location of the ancient individuals in PCA space provided a visual representation of their genetic similarity, and thereby allowed an assessment of community-based population substructure as represented by patterns of clustering or non-clustering in overlapping PCA space. Additionally, PCA allowed a qualitative assessment of the affinities of the Kulubnarti Nubians to various present-day individuals included in the dataset, further discussed in Chapter 8 in relation to Aim 2.

PCA plots were created to explore the following three possibilities: 1) the Kulubnarti Nubians from both communities clustered together in PCA space, suggesting no community-based population substructure at Kulubnarti; 2) the S community and R community formed two distinct and separate clusters in PCA space, suggesting community-based population substructure at Kulubnarti; or 3) the Kulubnarti Nubians did not cluster in PCA space, suggesting genetic heterogeneity at Kulubnarti. In addition to an assessment of clustering, PCA outliers were identified. Potential explanations for their outlier status will be further explored when the components of their ancestry are assessed in greater detail during ADMIXTURE analysis.

7.3.2 PCA of the Kulubnarti Nubians

To make inferences about population substructure at Kulubnarti by assessing the clustering of individuals from the S and R communities in relation to a background dataset of present-day individuals, projection PCAs were performed using the *smartpca* program v13050 from the EIGENSOFT package (Patterson et al. 2006). Options used included 'lsqproject: YES' to project samples with a large amount of missing data (ancient samples) onto samples with a small amount of missing data (present-day samples) (Lazaridis et al. 2014), 'shrinkmode: YES' to account for the "stretching" of the

PC axes when some samples are used to compute axes and others are not, and 'numoutlieriter: 0' to turn off the removal of outliers.[112] As previously discussed, genotype data from low-coverage ancient samples were made pseudo-haploid by randomly sampling a single non-duplicate sequence read at each position; therefore, to match the pseudo-haploid nature of aDNA genomes, the option 'inbreed: YES' was used to randomly assign a single allele from each present-day individual (Skoglund and Jakobsson 2011).

A PCA dataset of 579,601 autosomal SNPs from 1,023 present-day individuals genotyped on the Human Origins array comprising 86 populations from the Caucasus, the Near East, Northern Africa, Eastern Africa, Western Africa, Central Africa, and Southern Africa was used to compute PC1 and PC2.[113] One SNP from each pair in linkage disequilibrium (LD) with $r^2 > 0.2$ was removed, and only autosomal SNPs were included in the dataset. PLINK v1.9b3.41 (Chang et al. 2015; www.cog-genomics.org/plink/1.9/) was used to merge the 28 samples from the Kulubnarti S and R communities with the PCA dataset.[114] The ancient Kulubnarti Nubians were then projected onto the eigenvectors computed using the present-day populations and the PCA was plotted with R v3.3.1. This plot is presented as Figure 7.1.[115]

---

[112] The shrinkmode option is a new parameter available in Eigensoft version 6.0beta, and while not yet widely implemented, has been used with success and is recommended for use by the Reich lab (see the README text file available with the Eigensoft package).
[113] The list of populations used for the PCAs presented in this chapter and their geographic grouping is provided in Supplement 5 as Table S5.1.
[114] Two archaic individuals (Neandertal and Denisovan) were not included.
[115] A PCA plot with present-day individuals labeled with the abbreviation of the population to which they belong is provided as Figure S5.1 in Supplement 5.

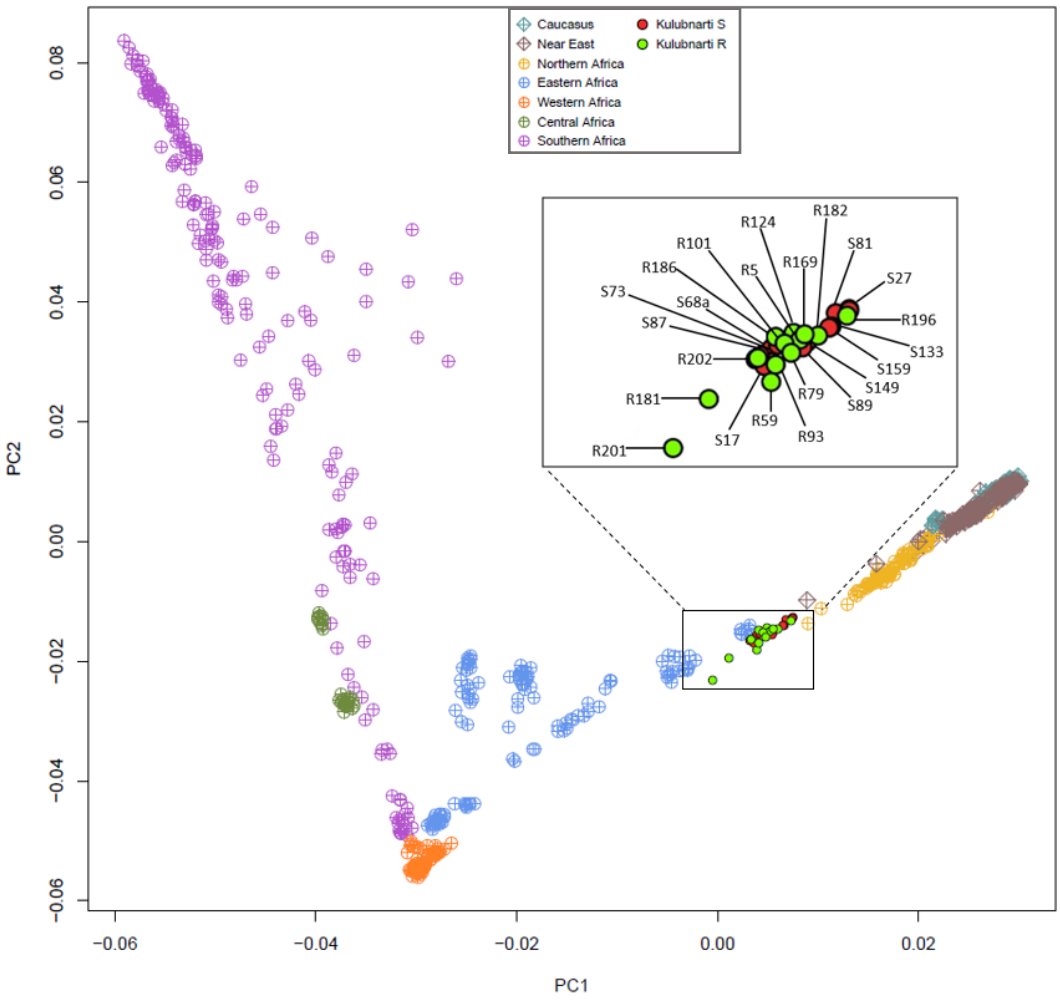
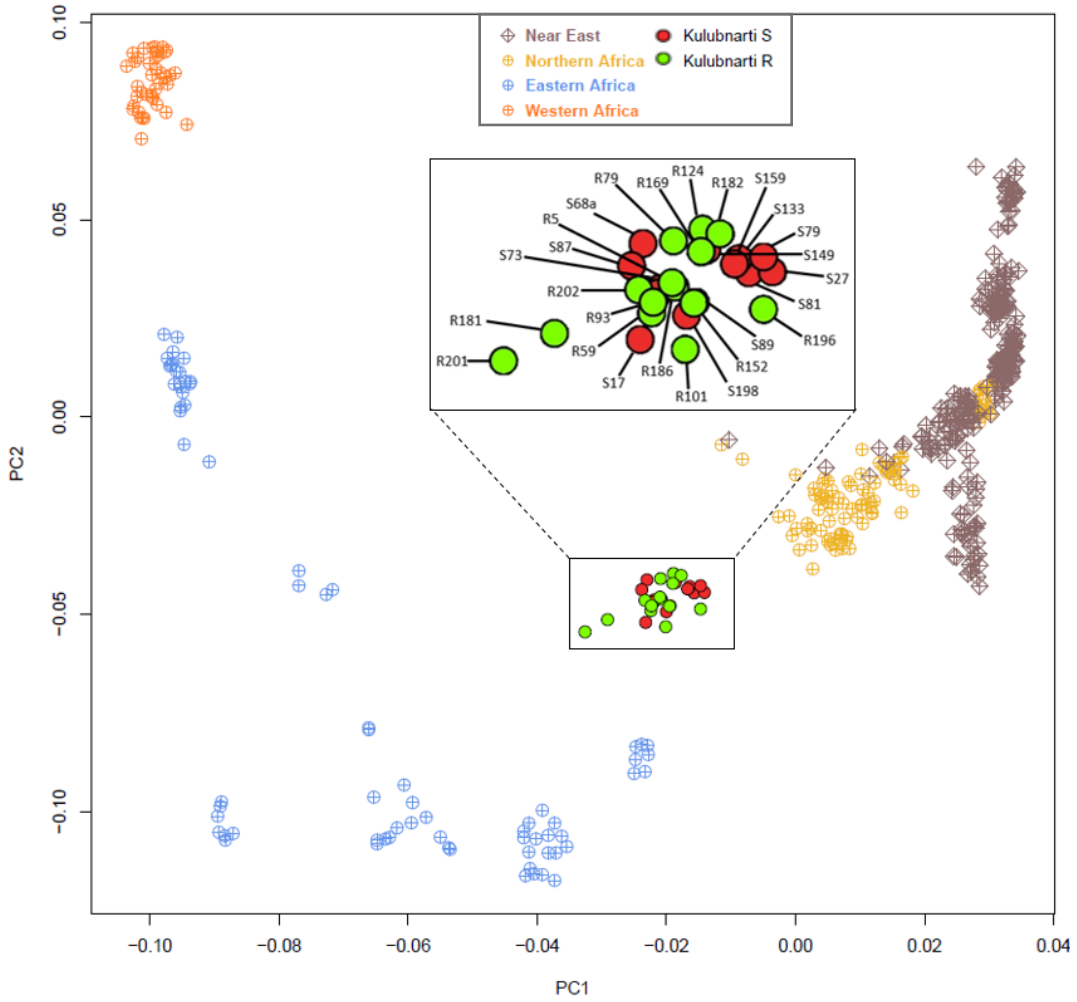


Figure 7.1: PCA plot showing the individuals from the S and R communities at Kulubnarti projected onto Caucasian, Near Eastern, and African populations from the 1240K_HO dataset. A zoomed depiction of the projected location of the Kulubnarti Nubians is located within the black box with all visible individuals labeled. Note that the clustering of the Kulubnarti Nubians completely obscures several samples: R152 and R198 are obscured by the cluster including R5, R124, and R169; S33 and S50 are obscured by the cluster containing R202, S87, S73, and S68a; and S79 is obscured by S27.

For a higher resolution plot depicting the clustering of the Kulubnarti Nubians in PCA space, a second projection PCA was performed using options identical to those described above, but further reducing the geographic distribution of the present-day populations included as the background dataset. A PCA dataset of 549,194 autosomal

SNPs from 466 present-day individuals genotyped on the Affymetrix Human Origins Array comprising 38 populations from the Near East, Northern Africa, Eastern Africa, and Western Africa selected by the author based on their proximity to the Kulubnarti Nubians in the initial PCA was compiled. Their pseudo-haploid SNP data were used to compute PC1 and PC2, and the 28 Kulubnarti Nubian samples were projected onto the modern variation. PCA was again plotted with R v3.3.1. This plot is presented as Figure 7.2.[116]



[116] A PCA plot with present-day individuals labeled with the abbreviation of the population to which they belong is provided as Figure S5.2 in Supplement 5.

Figure 7.2: PCA plot showing the individuals from the S and R communities at Kulubnarti projected onto select Near Eastern and African populations from the 1240K_HO dataset. A zoomed depiction of the projected location of the Kulubnarti Nubians is located within the black box with all visible individuals labeled. Note that S33 and S50 are obscured by the cluster of samples containing R5, R202, R93, R59, S17, and R186.

7.3.3 Interpretation of PCAs

As discussed in Chapter 5, it was expected that individuals from both communities would be projected into overlapping PCA space and would appear as a distinct and single cluster within that space, consistent with expectations of no community-based genetic population substructure at Kulubnarti.

The projections of the Kulubnarti Nubians on the two PCA plots presented as Figures 7.1 and 7.2 in this chapter supported the expectation of no community-based population substructure at Kulubnarti.[117] Individuals from both communities clustered together in the first two dimensions of both PCAs with no evidence of community-based clustering. Instead, individuals from both communities clustered together to the extent that some individuals are obscured by others. As such, a visual assessment of both PCA plots failed to reject the null hypothesis that there was no community-based population substructure at Kulubnarti. This finding suggests that there was no accumulation of community-specific genetic drift that would lead to the distinct clustering of each of the two communities when individuals from each community were plotted in PCA space.

Two individuals from the R cemetery (R201 and R181) appeared as PCA outliers, separated from the rest of the Kulubnarti Nubians in PCA space.[118] These individuals

[117] A lack of community-based genetic population substructure is also supported by ADMIXTURE clustering analysis, which will be presented in Chapter 8.
[118] These individuals do not appear as outliers in terms of archaeological or bioarchaeological evidence.

were shifted along PC1 away from the Kulubnarti cluster and toward Eastern African populations; however, it is unlikely that this small amount of separation merits their consideration as individuals from another population. Instead, these outlier individuals were projected close enough to the rest of the Kulubnarti Nubians to be considered as more divergent members of the Kulubnarti population. While the PCAs presented in this chapter identify R201 and R181 as outliers, their PCA outlier status will be explored in greater detail during analyses of their biogeographic genetic affinity and ancestry components in Chapter 8.

While the two PCAs presented in this chapter suggest that community-based population substructure did not exist at Kulubnarti, a visual assessment of clustering is a low-resolution and potentially imprecise analysis. As such, assumptions based on PCA must be explored further with additional statistical analyses. The assessment of population substructure between the S and R communities was therefore further tested by estimating the genetic distance between the communities using the fixation index F_{ST}, described in the next section of this chapter, as well as by ADMIXTURE clustering analysis, considered to be a more robust procedure for examining genetic similarity in terms of ancestry components compared between individuals based on a restricted number of parameters (Leonardi et al. 2016) and presented in Chapter 8. The broader implications of finding no evidence of community-based genetic population substructure at Kulubnarti when PCA was performed will be discussed in Chapter 9.

7.4 Quantifying Genetic Distance at Kulubnarti

7.4.1 Analytical Method: F_{ST}

F_{ST} (or "pairwise F_{ST}") is an estimate of genetic structure in natural populations (Wright 1949; Malécot 1968; Weir and Cockerham 1984), also described as a measure of the proportion of genetic differentiation (interchangeable here with "genetic distance") between two populations or subpopulations due to differences in allele frequencies (Jobling et al. 2014). F_{ST} values range from 0 to 1, where values close to 0 indicate that there is very little genetic differentiation between two groups (i.e., allele frequencies between each group are similar) and values close to 1 indicate that the groups are highly differentiated (i.e., allele frequencies between each group are different).

The estimation of pairwise F_{ST} is an important component in studies of population structure in present-day human populations (e.g., The International HapMap Consortium 2005, 2007, 2010; 1000 Genomes Project Consortium 2012). In human populations, estimates of F_{ST} <0.05 are considered to indicate a low level of genetic differentiation, estimates between 0.05–0.15 are considered to indicate a moderate level of genetic differentiation, estimates between 0.15–0.25 are considered to indicate a great level of genetic differentiation, and estimates >0.25 are considered to indicate a very great amount of genetic differentiation (Wright 1978).

It is well-established that genetic distance between human populations is generally low (Holsinger and Weir 2009; 1000 Genomes Project Consortium 2010, 2012; Elhaik 2012). Multiple studies showed that the majority of human genetic variation is actually found between individuals within groups, with only ~5–15% of the total variation found between continental groups, and even less (~3–8%) found between groups inhabiting different continents (Lewontin 1972; Cavalli-Sforza et al. 1994; Barbujani et al. 1997; Jorde et al. 2000; Rosenberg et al. 2002; Shriver et al. 2004; 1000

Genomes Project Consortium 2010, 2012; The International HapMap Consortium 2010; Elhaik 2012). Phase 3 of the International HapMap project estimated F_{ST} values indicating low genetic differentiation (based on Wright 1978) between populations of European ancestry (CEU + TSI, F_{ST}=0.004) as well as between populations of East Asian ancestry (JPT + CHB + CHD, F_{ST}=0.001–0.008). Relatively higher, but still low F_{ST} values were estimated between populations of African ancestry (ASW + MKK + LWK + YRI, $F_{ST}\leq$0.027) (The International HapMap Consortium 2010). Research by the 1000 Genomes Project Consortium demonstrated consistent results, with F_{ST} values from related populations within broader geographic regions typically estimated to be <0.01, and intra-continental populations described as being "weakly differentiated" (1000 Genomes Project Consortium 2012: 58).[119]

F_{ST} can also be used to determine genetic distance between subpopulations comprising a single structured or stratified population. These subpopulations can differ from each other in various ways, including in cultural practices, residence patterns, or opportunity for contact with foreign people. For example, a study of two neighboring subpopulations on Santa Catarina Island, Brazil were detected to have an F_{ST} value of 0.0028, indicating very low genetic distance despite potential partial reproductive isolation due to social practices (de Souza et al. 2003).

In addition to quantifying genetic distance between various present-day populations or subpopulations, estimates of F_{ST} have also been used to quantify genetic distance between ancient groups as well as between ancient and present-day populations

[119] International HapMap Project abbreviations include CEU: Utah residents with Northern and Western European ancestry; TSI: Toscani in Italy; JPT: Japanese in Tokyo, Japan; CHB: Han Chinese in Beijing, China; CHD: Chinese in Metropolitan Denver, Colorado; MKK: Maasai in Kinyawa, Kenya; LWK: Luhya in Webuye, Kenya; and YRI: Yoruba in Ibadan, Nigeria.

(e.g., Malmström et al. 2009; Fehren-Schmitz et al. 2010; Haak et al. 2015; Mathieson et al. 2015; Skoglund et al. 2016; Schuenemann et al. 2017). For example, a recent study of ancient Egyptians and present-day Egyptians and Ethiopians found pairwise F_{ST} values between 0.01–0.10 (low to moderate genetic differentiation) when ancient Egyptians were compared with present-day populations from neighboring geographic regions, and F_{ST} values 0.03–0.06 (low to moderate genetic differentiation) when present-day Ethiopians and Egyptian groups were compared to each other (Schuenemann et al. 2017).

Pairwise F_{ST} was estimated in this dissertation to assess genetic distance between the ancient Kulubnarti S and R communities, as well as between both communities and various present-day populations from around the world. In this analysis, the use of genome-wide SNP data enabled the determination of genetic differentiation with high precision, even in the case of small sample sizes (N=2–6) because F_{ST} can be measured accurately without large sample sizes when many SNPs are used (k>1000) (Willing et al. 2012).

7.4.2 Genetic Distance at Kulubnarti

Pairwise F_{ST} was estimated using the python program *popstats* (Skoglund et al. 2015; https://github.com/pontussk/popstats). The Hudson estimator (Hudson et al. 1992) for F_{ST} and a corresponding weighted block-jackknife estimator for the standard error of F_{ST} were implemented in the program. The Hudson estimator was selected over other estimators (e.g. Weir and Cockerham 1984) because it produced F_{ST} estimates that were independent of sample size and did not systematically overestimate F_{ST} (Bhatia et al.

2013); this is especially important when small sample sizes are analyzed.[120] Z-scores for the F_{ST} statistic were obtained using the computed standard error to indicate if the estimation of F_{ST} deviated significantly from zero; $|Z| \geq 3.0$ was interpreted as a statistically significant deviation.

Quantification of genetic differentiation is maximally robust when ascertainment of SNPs for the analysis is performed in the presence of an outgroup (Patterson et al. 2012; Wang and Nielsen 2012) so that there are no biases in allele frequencies between the analyzed populations and the polymorphisms that appeared by mutation in the ancestral population of all analyzed populations (Skoglund et al. 2017). The first modern human divergences are estimated to have occurred between 260–350 thousand years ago (kya) between the Khoe-San and other groups (Schlebusch et al. 2017), which allows their use as an outgroup in most analyses because they diverged from other lineages before those lineages separated from each other. However, because the dataset used in this dissertation also included SNP data from the archaic Denisovan (Meyer et al. 2012) and Altai Neandertal (Prüfer et al. 2014) genomes (combined in this analysis and denoted as "Archaics"), who diverged from all modern human lineages before any human lineages split from each other (Green et al. 2010; Meyer et al. 2012; Prüfer et al. 2014), the Archaics were selected as the outgroup to ensure that all ascertained SNPs were present as polymorphisms before the divergence of any African populations.[121]

---

[120] F_{ST} estimates can vary substantially based on choice of estimator (Bhatia et al. 2013). The Hudson estimator is preferentially selected for this analysis over the more commonly-used Weir-Cockerham estimator (Weir and Cockerham 1984) because the Weir-Cockerham estimator has been shown to overestimate F_{ST} when analyzing small sample sizes (Willing et al. 2012; Bhatia et al. 2013).

[121] The use of an Archaic outgroup was recommended by Dr. Pontus Skoglund. Experimentation by the author demonstrated that use of the Yoruba as an outgroup resulted in the same F_{ST} estimate as when the Archaics were used as an outgroup.

Because consanguinity may increase F_{ST} estimates (Reich et al. 2009), one individual from each of three definitively related dyads (R59/R152; R93/R101; S17/S50) was selected at random and removed prior to estimation of F_{ST}.[122] All present-day individuals included in the reference dataset were made pseudo-haploid to match the nature of the genotypes of the ancient individuals using the '--haploidize' option implemented in the *popstats* program. The use of this parameter is common in ancient DNA research (e.g., Mathieson et al. 2015; Skoglund et al. 2016). Finally, only autosomal SNPs were included in this analysis.

Pairwise F_{ST} was estimated between the S and R communities as well as between either community and 260 present-day populations from around the world. All pairwise F_{ST} estimates were based on ≥10,000 SNPs and required present-day populations to have ≥5 individuals. The full set of estimates for both communities are provided in Supplement 6 as Tables S6.1 and S6.2, while the 20 lowest F_{ST} estimates for each community are presented in Tables 7.2 (S community) and 7.3 (R community).

Table 7.2: Twenty-one lowest pairwise F_{ST} estimates (with standard error and Z-score) for the Kulubnarti S community based on a minimum of 10,000 SNP sites and ≥5 individuals in present-day populations.[123] Cell background color represents geographic region of test population as classified in this dissertation: grey represents Near East, yellow represents Northern Africa, light blue represents Eastern Africa. The Kulubnarti R community is represented by a white cell.

	FST (Kulubnarti S, Test)		
Test Population	Estimate	SE	Z-score
Kulubnarti R	0.0024^{+}	0.0009^{+}	2.52*$^{+}$
Jew_Ethiopian	0.016	0.0007	22.14
Somali	0.020	0.0006	33.96

[122] The method used to assess relatedness at Kulubnarti is briefly described in Supplement 10, and the corresponding relatedness coefficients and plots showing relatedness between these three dyads are provided.

[123] Twenty-one estimates are included due to identical F_{ST} estimations for Iranian, Kikuyu, and Saudi populations.

Moroccan	0.022	0.0007	30.76
Egyptian	0.023	0.0007	34.29
Tunisian	0.023	0.0008	29.50
Yemeni	0.023	0.0009	28.07
Libyan	0.024	0.0009	27.74
BedouinA	0.030	0.0007	42.49
Saharawi	0.030	0.0009	30.89
Jordanian	0.033	0.0009	38.05
Mozabite	0.033	0.0007	48.07
Masai	0.033	0.0007	47.58
Lebanese	0.035	0.0009	39.97
Palestinian	0.035	0.0007	49.02
Syrian	0.035	0.0009	39.08
Algerian	0.037	0.0009	42.07
Lebanese_Muslim	0.037	0.0008	45.07
Iranian	0.038	0.0007	58.23
Kikuyu	0.038	0.0010	35.52
Saudi	0.038	0.0009	42.55

* indicates that Z-score is not significant
$^{+}$ based on 1,000 permutations

Table 7.3: Twenty lowest pairwise F_{ST} estimates (with standard error and Z-score) for the Kulubnarti R community based on a minimum of 10,000 SNP sites and ≥5 individuals in present-day populations. Cell background color represents geographic region of test population as classified in this dissertation: grey represents Near East, yellow represents Northern Africa, light blue represents Eastern Africa. The Kulubnarti S community is represented by a white cell.

	FST (Kulubnarti R, Test)		
Test Population	Estimate	SE	Z-score
Kulubnarti S	0.0024^{+}	0.0009^{+}	2.52*$^{+}$
Jew_Ethiopian	0.013	0.0009	15.33
Somali	0.017	0.0008	21.45
Moroccan	0.020	0.0009	22.50
Tunisian	0.020	0.0009	23.19
Egyptian	0.023	0.0009	26.57
Libyan	0.024	0.0011	22.07
Yemeni	0.024	0.0010	21.86
Masai	0.028	0.0009	32.13
BedouinA	0.029	0.0009	33.87
Saharawi	0.029	0.0011	25.61
Mozabite	0.032	0.0009	35.91
Jordanian	0.033	0.0010	32.29

Kikuyu	0.035	0.0012	26.90
Lebanese	0.035	0.0011	35.19
Palestinian	0.035	0.0009	40.49
Syrian	0.035	0.0010	35.02
Algerian	0.036	0.0011	34.01
Lebanese_Muslim	0.038	0.0010	38.46
Sandawe	0.039	0.0009	44.46

* indicates that Z-score is not significant
+ based on 1,000 permutations

7.4.3. Interpretation of F_{ST} Results

As discussed in Chapter 5, it was expected that there would be no statistically significant genetic distance between the S and R communities at Kulubnarti, consistent with an expectation of no community-based genetic population substructure at Kulubnarti. This expectation was supported by a non-significant pairwise F_{ST} estimate of 0.0024 (s.e.=0.0009, Z=2.52) between the S community and R community using 403,172 SNPs and based on the average of 1,000 permutations (Tables 7.2 and 7.3). This estimate indicated that 0.24% of the total allele frequency variance existed between the S and R communities at Kulubnarti, while 99.76% of allele frequency variance existed within the communities themselves. The small standard error suggested that this F_{ST} estimate was made with high accuracy, and the non-significant Z-score (Z=2.52) associated with the pairwise F_{ST} value between the S and R communities indicated that there was no statistically significant deviation from zero, and thus failed to reject the null hypothesis that there was no significant genetic distance between the S and R communities. These results suggested that there was no genetically-discernible accumulation of genetic drift that resulted in significant population substructure at Kulubnarti, and therefore that these two socially-distinct communities comprised a single genetic population.

The F_{ST} estimate of 0.0024 between the S and R communities at Kulubnarti was contextualized through a comparison to recent studies that estimated F_{ST} in present-day Nubian population. For example, a similar pairwise F_{ST} of 0.004513 was estimated between the Mahas and the Halfawieen, two present-day Nubian groups living along the Nile in Upper Nubia (Hollfelder et al. 2017). The F_{ST} estimate between the Kulubnarti communities was also contextualized through a comparison to the F_{ST} estimates for the 20 (or 21 in the case of the S community) present-day populations that had the next lowest F_{ST} estimates with the Kulubnarti communities (shown in Tables 7.2 and 7.3). The F_{ST} estimate between the S and R communities was shown to be >4-fold lower than the next lowest F_{ST} estimate between either Kulubnarti community and any present-day group, further emphasizing the genetic closeness of these two communities. A broader examination of the data revealed that the only non-significant pairwise F_{ST} estimate was between the Kulubnarti communities.

Changing focus to F_{ST} estimates between either Kulubnarti community and 260 present-day populations from around the world, genotyped populations from throughout the Near East, Northern Africa, and Eastern Africa showed the next-lowest levels of genetic differentiation from both Kulubnarti communities, though it is necessary to keep in mind that F_{ST} estimates between the Kulubnarti Nubians and all present-day populations were statistically significant. Tables 7.2 and 7.3 display the F_{ST} estimates for the 20 (or 21 in the case of the S community) populations that had the lowest F_{ST} estimates with the Kulubnarti communities; here, all estimates were $\leq$0.039, indicating low (but significant) genetic differentiation with each Kulubnarti community as per Wright (1978).

Apart from each other, both the S and R communities were estimated as having the least amount of genetic distance with present-day genotyped populations from Northeastern Africa, including Ethiopian Jews (F_{ST}=0.016 for the S community, and 0.013 for the R community) and present-day Somalis (F_{ST}=0.020 for the S community and 0.016 for the R community).[124] This result is likely to reflect the geographic proximity of Ethiopia and Somalia with Nubia in relation to other populations separated by greater distance or by geographic barriers, such as the Sahara Desert. It must be cautioned that this result may reflect the extensive amounts of genetic exchange that have occurred in the centuries since the Early Christian Period due to increasing amounts of interaction between populations throughout Northern Africa, Eastern Africa, and the Near East as more extensive trade routes developed; these results do not imply that there is any direct genetic relationship between the Kulubnarti Nubians and any of these present-day populations.[125]

Returning to the research question that inspired Aim 1, the non-significant genetic distance between the S and R communities suggests that there is no community-based genetic population substructure at Kulubnarti and that both communities are part of a single genetic population. The broader implications of these F_{ST} results on the understanding of Kulubnarti from a genomic perspective will be discussed in Chapter 9.

7.5 Chapter Summary

[124] The Ethiopian Jews included in this dataset are Jews currently living in Israel who migrated from Ethiopia. The genetic composition of Ethiopian Jews and their relationship to non-Jewish Ethiopians is further discussed in Chapter 9. It should be noted that the Oromo Ethiopians were not included in this estimation of F_{ST} because they did not meet the requirement of ≥5 individuals in present-day populations.
[125] Caveats surrounding the use of a reference dataset of present-day populations are discussed further in association with interpretation of PCA and ADMIXTURE analysis in Chapter 8.

Several research questions that inspired this genomic analysis of the Kulubnarti Nubians were presented in Chapter 1. The question that served as the cornerstone for this dissertation was explored in this chapter: were the socially-disparate S and R communities that comprised the population of Kulubnarti also genetically-distinct subpopulations? Aim 1 of this dissertation specifically focused on answering this question, first by looking for evidence of community-based patterns of clustering in PCA space, and second by quantifying the genetic distance between the communities using the fixation index F_{ST}.

As discussed in Chapter 5, the null hypothesis associated with Aim 1 is that there was no genetic population substructure at Kulubnarti. It was therefore expected that the individuals from both communities would cluster together in overlapping space when projected into PCA space and that the genetic distance between the communities quantified using pairwise F_{ST} would not be statistically significant. The results of the analyses presented in this chapter consistently failed to reject the null hypothesis.

The Kulubnarti Nubians from both communities were projected into overlapping PCA space and appeared as a cluster within that space, suggesting that they comprised a single genetic population. Conclusions made from the projection PCAs were supported by quantification of genetic differentiation using F_{ST}, which indicated no statistically significant genetic differentiation between the S and R communities, which was also consistent with their interpretation as a single genetic population. These results were consistent with previous craniometric analyses and analyses of discrete dental traits that suggested a close biological relationship between the Kulubnarti communities. However,

for the first time, this relationship was assessed at a high resolution by investigating patterns of human genetic variation.

The analyses presented in this chapter provided an answer to the initial question posed in Chapter 1. The analysis of autosomal SNP data suggested that the socially-distinct S and R communities that comprised the population of Kulubnarti were not genetically-distinct subpopulations. More broadly, these results suggested that the community structure observed in the archaeological and bioarchaeological records at Kulubnarti did not have a genetically-discernible basis. The implications of these results will be discussed in the final chapter of this dissertation.

CHAPTER 8
GENETIC COMPOSITION OF THE KULUBNARTI NUBIANS

8.1 Chapter Overview

While archaeological and bioarchaeological evidence suggested that Kulubnarti was home to two communities that were likely the descendants of peoples who had inhabited Nubia for millennia (Arkell 1961; Adams 1977), the biogeographic genetic affinities (or simply "biogeographic affinities") and origins of the Kulubnarti Nubians have not been investigated from a genomic perspective. Ancient DNA (aDNA) analysis represents a direct line of inquiry from which to explore the genetic composition of the Kulubnarti Nubians in reference to the present-day human genetic landscape. Therefore, **Aim 2 of this dissertation was to characterize the genetic composition of individuals from both Kulubnarti communities and reveal any community-based differences by exploring biogeographic genetic affinities and components of ancestry and determining the mitochondrial DNA (mtDNA) and Y chromosome haplogroup profiles of each community**.

This chapter introduces the methods used to achieve Aim 2, discusses the application of these methods to the Kulubnarti data, and presents and interprets the ensuing results. A discussion of how these results influence the broader understanding of Early Christian Kulubnarti from a genomic perspective is provided in Chapter 9.

8.2 Assessing Biogeographic Genetic Affinities and Components of Ancestry at Kulubnarti

8.2.1 Analytical Methods: Principal Component Analysis and ADMIXTURE Clustering Analysis

The analytical method of Principal Components Analysis (PCA) was first presented in Chapter 7 as a method for assessing patterns of clustering of members of the Kulubnarti S and R communities in relation to a background dataset of present-day individuals.[126] While initially used in this dissertation to explore the potential existence of community-based genetic population substructure at Kulubnarti, a more common use of PCA is to provide an exploratory approximation of affinities to ancient or present-day populations prior to focused cluster-based analyses of ancestry components based on a restricted number of parameters (Leonardi et al. 2016). The methods of PCA (Patterson et al. 2006) and ADMIXTURE clustering analysis (Alexander et al. 2009) are the most frequently used methods in population genetics analysis to study global ancestry (Patterson et al. 2012) and are used in this dissertation to explore the biogeographic genetic affinities and ancestry components of the Kulubnarti Nubians within the context of genotyped present-day populations from around the world.

As introduced in Chapter 4, individuals who share more genetic similarity (i.e., who have a greater number of identical alleles at variant sites across the genome) will cluster together in PCA space. In human populations, the observed pattern of clustering often correlates closely with geography, reflecting migrations and admixture as well as isolation by distance (i.e., an increased exchange of genes between geographically-proximate populations and a decreased exchange of genes between geographically-separate populations) (Li et al. 2008; Novembre et al. 2008; Reich et al. 2008; Patterson et al. 2012). This phenomenon of isolation by distance has been shown to be

[126] Details of the PCA analytical method can be found in Chapter 7.

characteristic of human populations (Tishkoff and Kidd 2004).[127] Therefore, it is possible to suggest genetic affinity between individuals and/or populations by identifying the individuals or groups that are projected in close proximity to each other in PCA space. Importantly, suggestions of affinity do not imply direct genetic connections; this caveat is discussed later in this chapter.[128]

In addition to suggesting affinity between individuals or populations, it is also possible to identify genetic clines using PCA plots. For example, Skoglund et al. (2017) identify an ancient African genetic cline that runs along a north-south axis with southern and eastern hunter-gatherers showing increasing affinity to ancient and present-day southern Africans in a way that is correlated closely to the geographic origin of the ancient individuals.

While it is possible to place low-coverage ancient individuals within PCA plots produced using modern data (Skoglund et al. 2012) to assess their genetic affinities to present-day worldwide populations, it is also necessary to recognize that PCAs (as well as clustering analyses) are sensitive to the genetic drift that occurred after the time the ancient individuals lived, and therefore may not always provide an accurate view of shared ancestry between ancient and present-day individuals (Patterson et al. 2012; Skoglund et al. 2014b).

PCAs often reveal that ancient samples do not precisely resemble any present-day populations, but instead are shifted along a particular axis. This shift likely corresponds to evolutionary processes (including changes in gene flow or accumulating genetic drift)

---

[127] While populations do cluster by geography based on genetic variation, it is important to reiterate that the distribution of these clusters is quasi-continuous in clinal patterns related to geography (Tishkoff and Kidd 2004).

[128] This discussion can be found in subsection 8.2.3.

that influenced the genetic landscape in the time since the ancient individual lived (e.g., Jones et al. 2015; Mathieson et al. 2015). This is why assessments of biogeographic affinity using PCA are best communicated by descriptions of "affinity" to present-day populations near to whom ancient individuals cluster and descriptions of the shifts observed in relation to present-day populations from the same geographic area. Despite this limitation, assessments of genetic affinities using PCA are now standard in nearly all recent aDNA research (e.g., Allentoft et al. 2015; Haak et al. 2015; Gallego-Llorente et al. 2015; Mathieson et al. 2015; Skoglund et al. 2017).

Following an exploratory analysis of population affinities using PCA, cluster-based analyses such as ADMIXTURE (Alexander et al. 2009) are frequently used as a more robust procedure for examining genetic similarity in terms of ancestry components based on PCA approximations of genetic affinities (Leonardi et al. 2016). In contrast to PCA, which does not attempt to classify individuals into discrete populations, model-based clustering approaches, such as the approach implemented in the ADMIXTURE software, assign individuals to corresponding ancestral "clusters" based on allele frequencies by assuming K clusters representing K putative ancestral populations with different allele frequencies.[129] Individuals are probabilistically assigned to one or more ancestral clusters based on their alleles; for each individual, estimates of the genetic fraction descending from each of the K ancestral populations is provided.

Clustering approaches are most commonly used in aDNA research to place ancient individuals within a framework of present-day human diversity (Leonardi et al. 2016). Specifically, they are used to provide additional support for approximate genetic

[129] Here, K is a user-defined value ≥2.

affinities approximated by PCA by identifying the various ancestral components of individuals or populations, which may be present in different quantities. In this dissertation, the ancestral components of the Kulubnarti Nubians were estimated using present-day populations from across Eurasia and throughout Africa as a reference panel.

The analysis of ancestral components can be used to decipher subtle differences in genetic ancestry between contemporaneous individuals (e.g., Martiniano et al. 2016; Olalde et al. 2017) or to identify changes in the ancestral components of individuals across space and time (e.g., Mathieson et al. 2018; Skoglund et al. 2017). Especially when a diachronic time series of samples is used, it is often possible to directly link the introduction of ancestral components to specific historical events, such as large-scale migrations (Haak et al. 2015; Mathieson et al. 2015); however, it must be cautioned that clustering methods can be ill-suited to complex admixture histories based on their assumption of a defined number of ancestral clusters (Patterson et al. 2012). Despite this limitation, the use of model-based clustering analysis represents a higher-resolution exploration of PCA results.

8.2.2 Biogeographic Genetic Affinities and Ancestry Components of the Kulubnarti Nubians

To explore the genetic affinities of the Kulubnarti Nubians, projection PCAs were performed as described in Chapter 7 (all details regarding dataset assembly and PCA options used are provided in section 7.3.2). To briefly reiterate, Principal Components (PCs) 1 and 2 in the PCA plot presented as Figure 8.1 were computed using 579,601 autosomal SNPs from 1,023 present-day individuals comprising 86 populations from the Caucasus, Near East, Northern Africa, Eastern Africa, Western Africa, Central Africa,

and Southern Africa selected from the 1240K_HO dataset.[130] The 28 Kulubnarti Nubians samples were merged with this reduced dataset and projected onto the eigenvectors computed using the selected present-day individuals.[131]

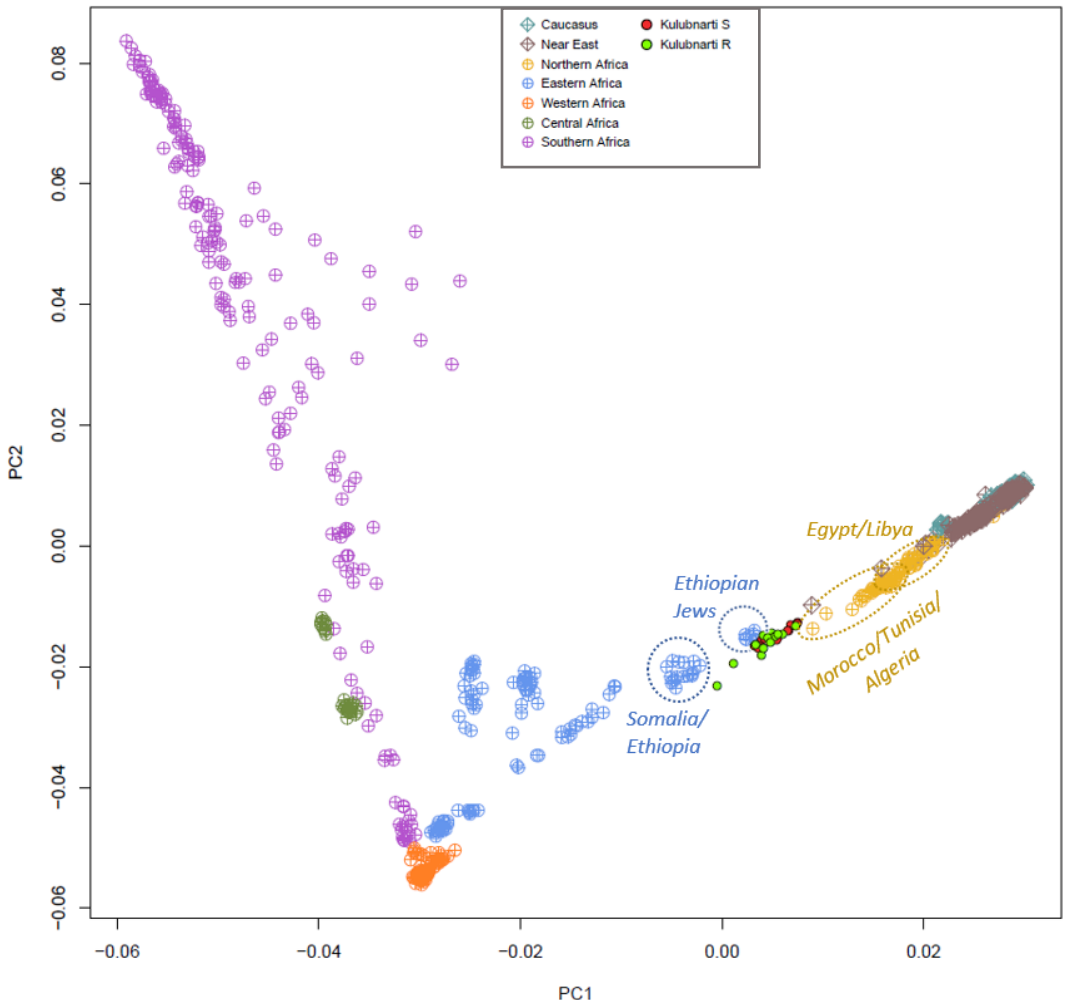


Figure 8.1: PCA plot showing the individuals from the S and R communities at Kulubnarti projected onto Caucasian, Near Eastern, and African populations from the 1240K_HO dataset. The present-day populations to whom the Kulubnarti Nubians are projected closest to are identified by dashed circles and labels, colored to correspond to the geographic region to which they are assigned.

[130] The list of populations used for the PCAs presented in this chapter and their geographic grouping is provided in Supplement 5 as Table S5.1.
[131] A PCA plot with present-day individuals labeled with the abbreviation of the population to which they belong is provided as Figure S5.1 in Supplement 5.

As in Chapter 7, this initial PCA was supplemented by a higher-resolution plot that used options identical to those used to produce the PCA plot above but that also eincluded a reduced number of present-day populations from a concentrated geographic area to compute PC1 and PC2. To reiterate, Figure 8.2 was created using 549,194 autosomal SNPs from 466 present-day individuals comprising 38 populations from the Near East, Northern Africa, Eastern Africa, and Western Africa selected based on their proximity to the Kulubnarti Nubians in the initial PCA and used to compute PC1 and PC2. Again, the 28 Kulubnarti Nubian samples were merged with this reduced dataset and projected onto the eigenvectors computed using the selected present-day individuals.[132]

[132] A PCA plot with present-day individuals labeled with the abbreviation of the population to which they belong is provided as Figure S5.2 in Supplement 5.

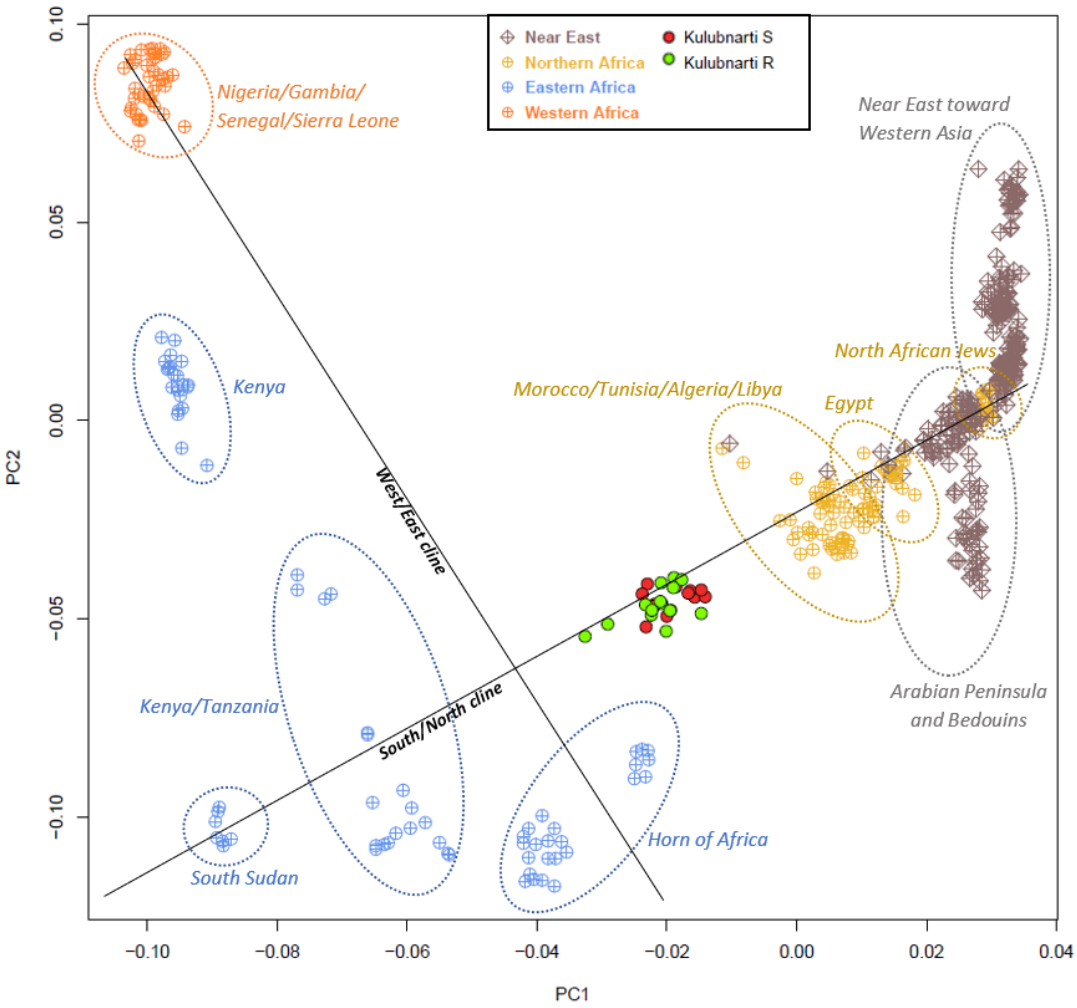


Figure 8.2: PCA plot showing the individuals from the S and R communities at Kulubnarti projected onto select Near Eastern and African populations from the 1240K_HO dataset. The present-day populations included on this plot are identified by dashed circles and labels, colored to correspond to the geographic region to which they are assigned. In addition, two approximate genetic clines are labeled.

Following PCA, model-based clustering analysis was performed using ADMIXTURE (Alexander et al. 2009). A reduced version of the 1240K_HO dataset that included 1,897 present-day individuals from 146 populations spanning Asia, Europe, the Caucasus, the Near East, Northern Africa, Eastern Africa, Western Africa, Central Africa, and Southern Africa was used as a reference panel to place the 28 Kulubnarti Nubian samples within a landscape of present-day human genetic diversity.

Autosomal SNPs were used for this analysis, and one SNP from each pair in linkage disequilibrium (LD) with $r^2 > 0.4$ was pruned using PLINK v1.9b3.41 (Chang et al. 2015; www.cog-genomics.org/plink/1.9/) with the options '--indep-pairwise 200 25 0.4' (indicating that LD was calculated in windows of 200 SNPs with a step size of 25 SNPs and a pairwise r^2 threshold of 0.4).[133] This resulted in the inclusion of 251,809 SNPs in ADMIXTURE analysis.

The number of ancestral clusters expected to contribute to the genomes of the Kulubnarti Nubians was unknown; therefore, multiple runs of the clustering analysis with increasing values of K were required. Here, each run gave a different probability reflecting the fit of the K clusters to the data, and the K value that was determined to be the best fit was the best estimate of the true number of clusters. The best-fit K was determined by the cross-validation (CV) procedure provided by the ADMIXTURE program, where the best value of K exhibited the lowest CV error when compared to other K values (Alexander et al. 2009).

ADMIXTURE was run with the cross validation ('--cv') flag for values of K ranging from K=2 to K=14, with 10 replicates for each value of K with different random seeds. CV errors for all tested values of K can be found in Supplement 7 as Table S7.1 and Figure S7.1. The minimal CV error was found at K=11, suggesting that the data were best represented by 11 ancestral components; however, the error started plateauing from K=8, suggesting little improvement from this point onward. Q matrices (ancestral cluster proportions) for values of K between K=2 and K=12 were plotted using R v3.3.1 and can

[133] LD-pruning is recommended by the authors of ADMIXTURE, and a relatively high r^2 value of 0.4 retains a larger number of SNPs than more aggressive pruning, which would reduce the SNPs available for analysis for the individuals with most missing data (Haak et al. 2015).

be found in Supplement 7 as Figure S7.2 (present-day and ancient individuals) and Figure S7.3 (ancient individuals only). In this chapter, all plots correspond to K=11, determined to be the number of ancestral components that best fit the data. Figure 8.3 presents the ADMIXTURE plot for 1,897 present-day individuals and the 28 Kulubnarti Nubians, Figure 8.4 focuses specifically on the 28 Kulubnarti Nubian samples included in the ADMIXTURE dataset, and Figure 8.5 focuses on the similarity in ancestral components between present-day Ethiopian Jews, Somali, Oromo Ethiopians and the two Kulubnarti Nubian communities.

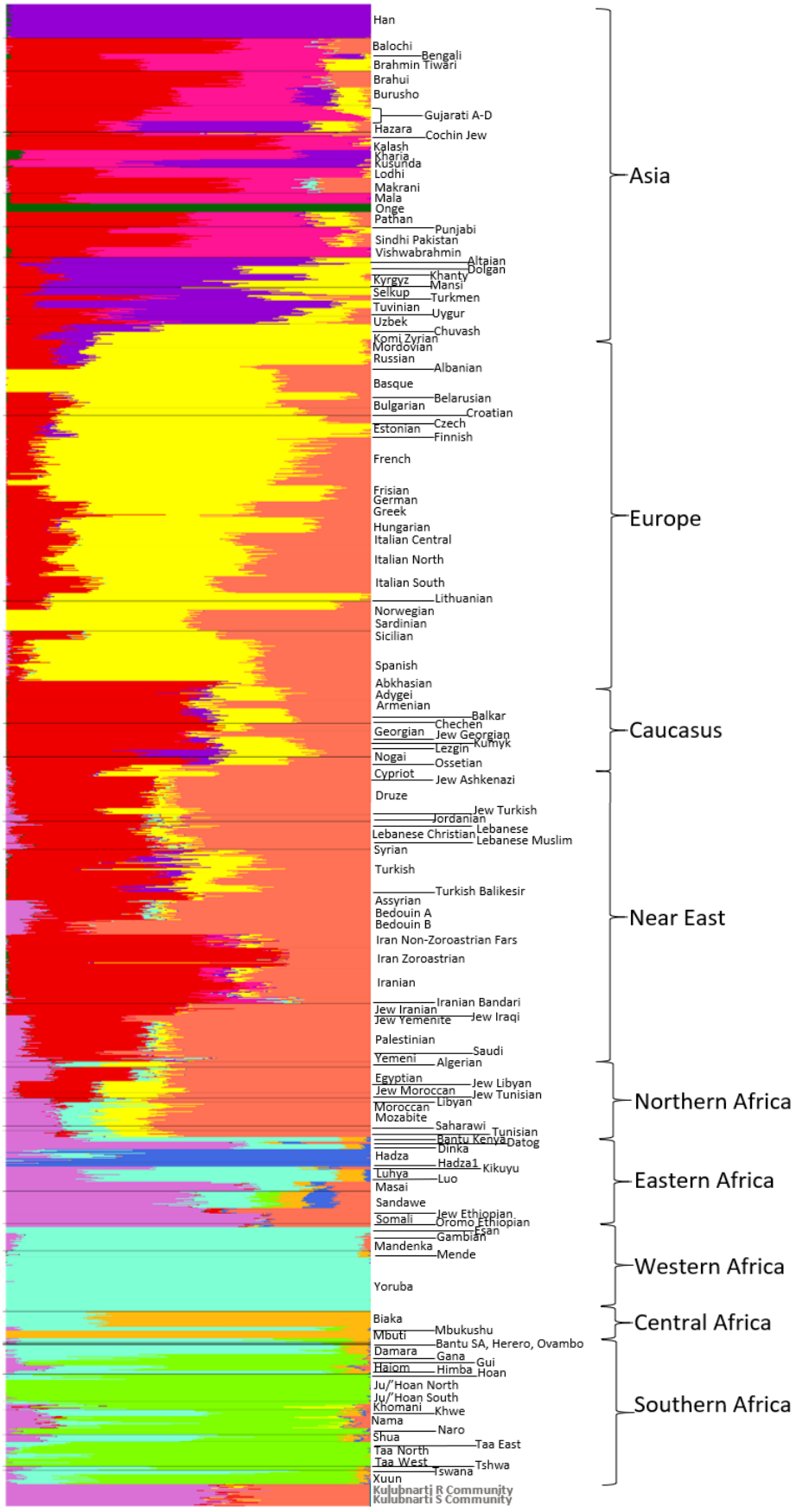

Han
Balochi
Bengali
Brahmin Tiwari
Brahui
Burusho
Gujarati A-D
Hazara
Cochin Jew
Kalash
Kharia
Kusunda
Lodhi
Makrani
Mala
Onge
Pathan
Punjabi
Sindhi Pakistan
Vishwabrahmin
Altaian
Dolgan
Kyrgyz
Khanty
Mansi
Selkup
Turkmen
Tuvinian
Uygur
Uzbek
Chuvash
Komi Zyrian
Mordovian
Russian
Albanian
Basque
Belarusian
Bulgarian
Croatian
Czech
Estonian
Finnish
French
Frisian
German
Greek
Hungarian
Italian Central
Italian North
Italian South
Lithuanian
Norwegian
Sardinian
Sicilian
Spanish
Abkhasian
Adygei
Armenian
Balkar
Chechen
Georgian
Jew Georgian
Kumyk
Lezgin
Nogai
Ossetian
Cypriot
Jew Ashkenazi
Druze
Jew Turkish
Jordanian
Lebanese
Lebanese Christian
Lebanese Muslim
Syrian
Turkish
Turkish Balikesir
Assyrian
Bedouin A
Bedouin B
Iran Non-Zoroastrian Fars
Iran Zoroastrian
Iranian
Iranian Bandari
Jew Iranian
Jew Iraqi
Jew Yemenite
Palestinian
Saudi
Yemeni
Algerian
Egyptian
Jew Libyan
Jew Moroccan
Jew Tunisian
Libyan
Moroccan
Mozabite
Saharawi
Tunisian
Bantu Kenya
Datog
Dinka
Hadza
Hadza1
Kikuyu
Luhya
Luo
Masai
Sandawe
Jew Ethiopian
Oromo
Somali
Ethiopian
Esan
Gambian
Mandenka
Mende
Yoruba
Biaka
Mbuti
Mbukushu
Bantu SA, Herero, Ovambo
Damara
Gana
Gui
Haiom
Himba
Hoan
Ju/'Hoan North
Ju/'Hoan South
Khomani
Khwe
Nama
Naro
Shua
Taa East
Taa North
Taa West
Tshwa
Tswana
Xuun
Kulubnarti R Community
Kulubnarti S Community
Asia
Europe
Caucasus
Near East
Northern Africa
Eastern Africa
Western Africa
Central Africa
Southern Africa

Figure 8.3: ADMIXTURE results showing ancestry proportions of 1,897 present-day individuals and 28 Kulubnarti Nubians assuming 11 ancestral components (K=11). Present-day individuals are hierarchically clustered by population and region, both labeled in black. The Kulubnarti R and S communities are positioned at the bottom of the plot and are labeled in bold grey text.

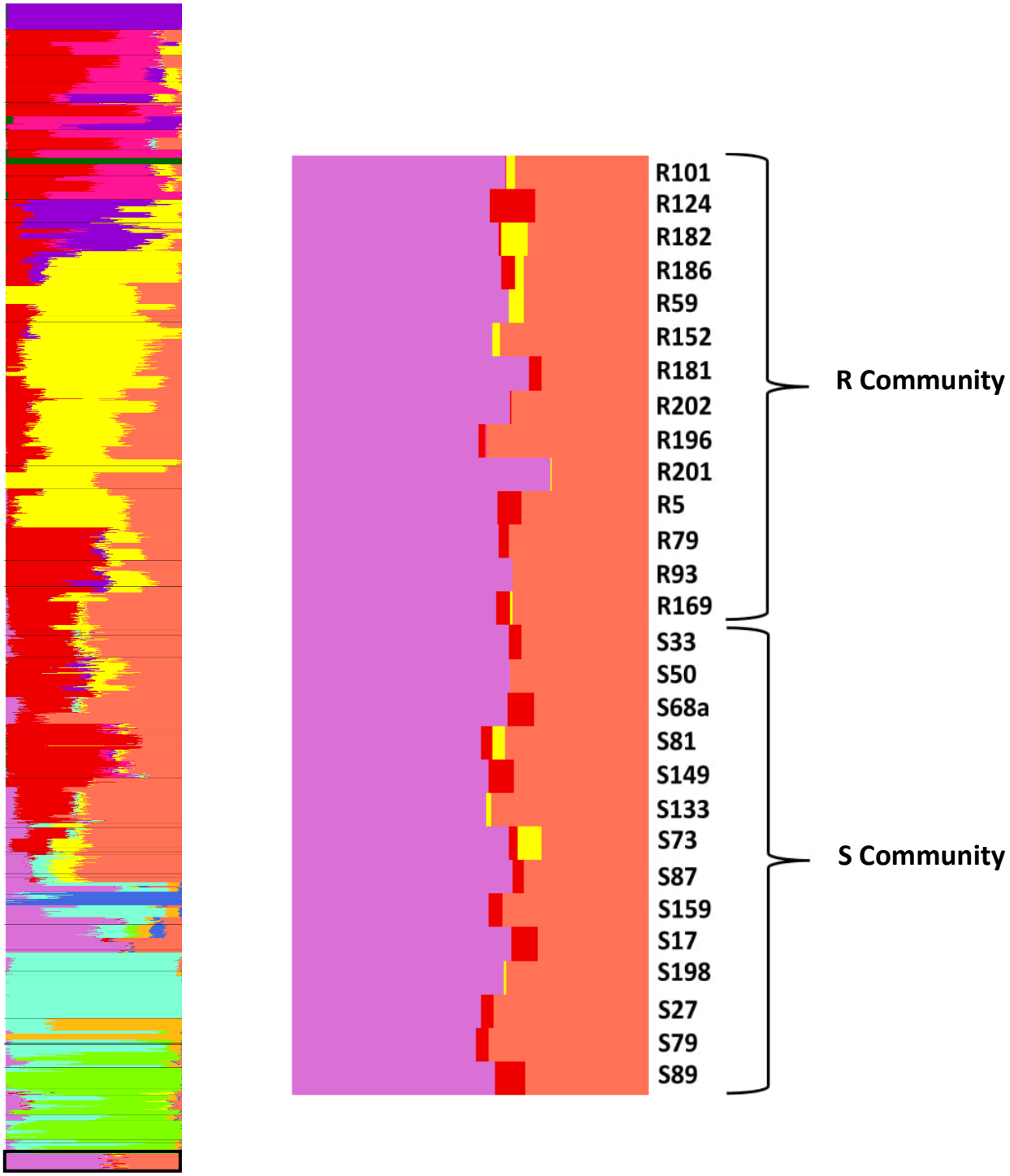


Figure 8.4: ADMIXTURE results showing ancestry proportions at K=11 for 28 Kulubnarti Nubians. This figure contains the ADMIXTURE plot representing ancestry proportions of 1,897 present-day and 28 ancient Kulubnarti Nubians (left), with the Kulubnarti Nubians outlined in black, and the focused ADMIXTURE plot of Kulubnarti R and S communities (right). Each individual from Kulubnarti is labeled.

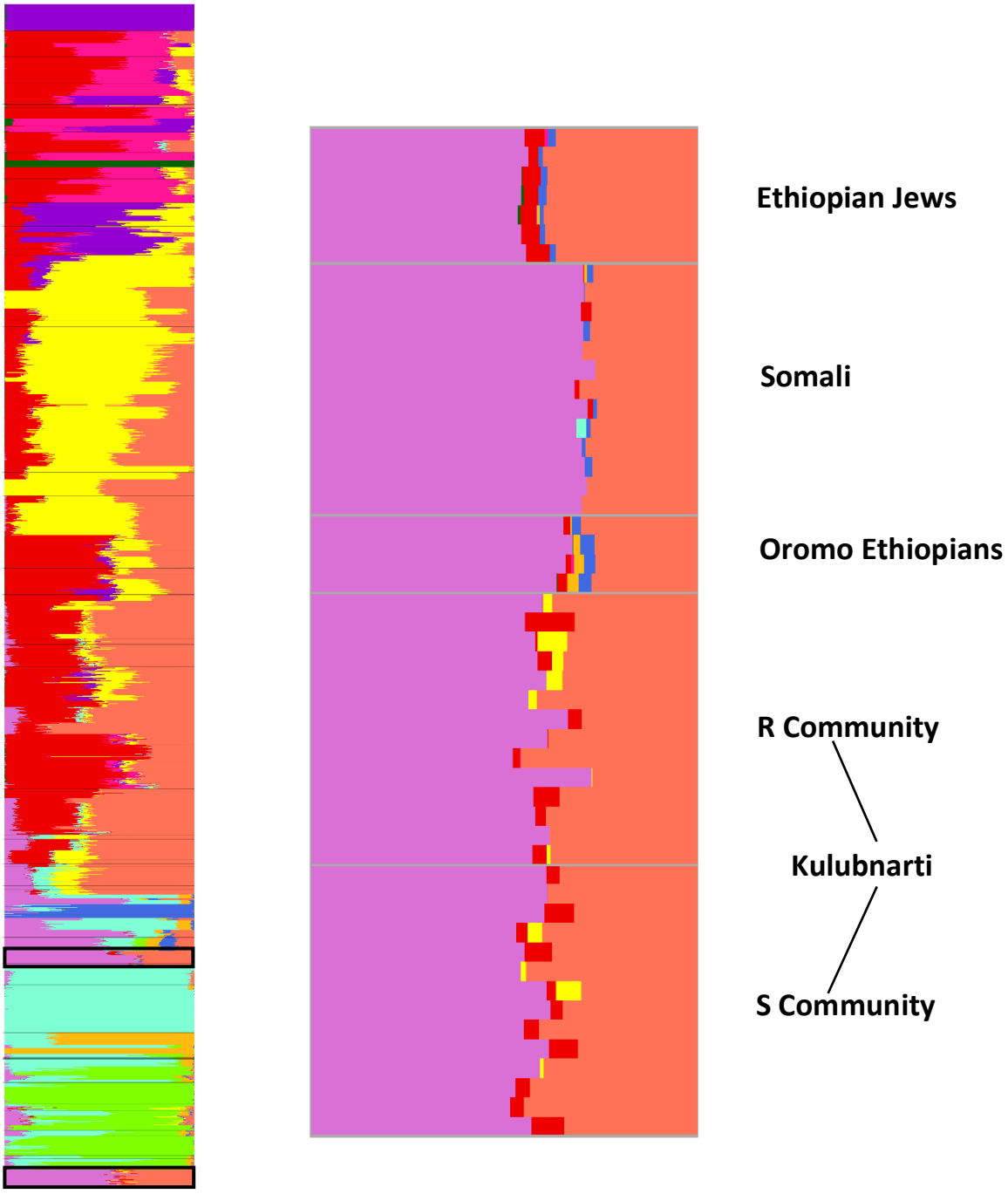


Figure 8.5: ADMIXTURE results showing genetic affinity in terms of ancestral components at K=11 between Kulubnarti Nubian and present-day Ethiopian Jews, Somali, and Oromo Ethiopians. This figure contains the ADMIXTURE plot representing ancestry proportions of 1,897 present-day and 28 ancient Kulubnarti Nubians (left), with the Ethiopian Jews, Somali, and Oromo Ethiopians and the Kulubnarti Nubians outlined in black, and the focused ADMIXTURE plot of Ethiopian Jew, Somali, and Oromo, Kulubnarti R community, and Kulubnarti S community (right, from top to bottom respectively).

8.2.3 Caveats to PCAs and ADMIXTURE Analysis

Prior to interpreting PCA and ADMIXTURE results, it is necessary to discuss several caveats, previously mentioned in Chapters 5 and 7. First, all assessments of genetic affinity between the Kulubnarti Nubians and present-day populations and estimations of the ancestral components of the Nubians are limited by available data from genotyped populations worldwide. Compared to other regions of the world (for example, Eurasia), there remains a paucity of genome-level data from present-day African populations. This limits the resolution of this dissertation's analyses, and consequently affects the specificity of the interpretations of data presented in this chapter.

Though limitations are present, analyses investigating the genetic affinities and ancestry components of the Kulubnarti Nubians have the potential to reveal new information about the genetic composition of an ancient Nubian population at a higher resolution than previously possible. These analyses represent the first step in the exploration of Kulubnarti from a genomic perspective and will be enhanced as additional and increasingly dense genome-wide data from present-day Africans as well as from ancient African samples becomes available.[134]

Second, the genomes of present-day populations that serve as references for estimating affinities and ancestry components for the Kulubnarti Nubians have been subject to up to 1200 years of additional demographic and evolutionary processes since the Early Christian Period which have influenced their genomic composition; examples of demographic events which may have influenced the genetic composition of present-day Nubians are discussed in greater detail in Chapter 5. As such, all results must be

[134] New datasets that include a larger number of present-day African populations will be utilized for future genomic analyses of the Kulubnarti Nubians (e.g., Schuenemann et al. 2017; Skoglund et al. 2017).

interpreted with caution: it is as likely that any detected genetic affinities reflect the extensive interactions and admixture that has occurred since the Early Christian Period as reflect the role of the Kulubnarti Nubians as the "middlemen" of the Nile who interacted and admixed with geographically-diverse people. Because of this limitation, all assessments of genetic affinity and estimations of ancestry components will be interpreted with the understanding that these results do not imply a direct genetic contribution or linear connection.

8.2.4 Interpretation of PCAs and ADMIXTURE Analysis

First discussed in Chapter 5 in association with Aim 2, Expectation 2 was that no community-based differences in genetic composition (including genetic affinities or ancestry components) will be detected at Kulubnarti. Both PCA and ADMIXTURE results were consistent with this expectation and identified no community-based differences in genetic composition between the S and R communities at Kulubnarti in terms of biogeographic genetic affinities or components of ancestry.

As described in relation to the PCAs discussed in Chapter 7, individuals from both Kulubnarti communities were projected together in a single distinct cluster in PCA space, meaning that they were assessed in this chapter as showing the most genetic affinity to the same present-day populations.[135] Consistent with this result, no community-based differences were detected between the S and R communities in terms of ancestry components or proportions of those components at any value of K when ADMIXTURE analysis was performed. Instead, individuals from both communities were

[135] The present-day populations to which the Kulubnarti individuals showed the greatest affinity will be discussed later in this section.

represented by the same two primary ancestry components in similar proportions, with very minor contributions from two additional components.

As shown in the context of the landscape of present-day human genetic diversity in Figure 8.3 and as shown as the specific focus of Figure 8.4, the most dominant ancestral signal detected in the Kulubnarti Nubians from both communities (lavender in color in both figures; referred to as the "lavender component") was maximized in some present-day sub-Saharan Eastern African populations (Datog, Dinka, Masai, Sandawe, Ethiopian Jews, Somali, and Oromo Ethiopians), present at a lesser extent in present-day Northern African populations (except in Northern African Jews), and additionally detected in some present-day populations from the Near East, especially Bedouins and some populations inhabiting the Arabian Peninsula (including Yemenite Jews, Yemenites, Palestinians, and Saudis). The second dominant ancestral signal detected in the Kulubnarti Nubians from both communities (coral in color in both figures; referred to as the "coral component") was maximized in many present-day Northern African populations (including Northern African Jews) as well as many Near Eastern populations, with the exception of Turkish, Iranian, and Iraqi populations who had a substantial signal found more prevalently in present-day populations from the Caucasus (the "red component" in both figures). The coral component was found at a less substantial amount in populations throughout Europe, the Caucasus, the Near East, Northern Africa, and some Eastern African populations (Datog, Kikuyu, Masai, Sandawe, Ethiopian Jews, Somali, and Oromo Ethiopians), and was notably absent from other Eastern African populations (Bantu from Kenya, Dinka, Hadza, Luhya, and Luo), Western African populations, Central African populations, and Southern African populations. In addition,

a very minor signal from a present-day Caucasian component (the "red component") and/or European component (the "yellow component") was detected in the majority of individuals from Kulubnarti and did not appear in any community-based pattern.[136]

By failing to reject the null hypothesis of no community-based differences between the S and R communities in terms of genetic affinities or ancestry components, these results supported the expectation of genetic similarity between the two socially-distinct communities.[137]

Exploring the genetic composition of the Kulubnarti Nubians at a deeper level, Expectation 2a was that both Kulubnarti communities would show more genetic affinity to present-day genotyped populations from Northeastern Africa than to present-day populations from any other world region and will be estimated as having components of ancestry associated with present-day populations from Northeastern Africa. In addition, because the genome-shaping demographic and evolutionary processes that occurred in the ~1200 years separating the Early Christian Period Nubians from present-day populations cannot be accounted for using PCA (Skoglund et al. 2012), it was also expected that the Kulubnarti Nubians would appear genetically distinct from all present-day populations.

The PCA plot presented as Figure 8.1 partially supported Expectation 2a, suggesting that the Kulubnarti Nubians from both communities showed more genetic affinity to genotyped present-day populations from Northeastern Africa (supporting Expectation 2a) as well as genotyped present-day populations across Northern Africa (not

[136] There were also no sex-based patterns detected.
[137] These results are also consistent with the findings of genetic similarity suggested by the overlap of the S and R communities in PCA space and non-significant pairwise F_{ST} value (0.0024 ± 0.0009, discussed in Chapter 7).

supporting Expectation 2a) than to present-day populations from any other world region. In particular, the Kulubnarti Nubians were plotted in closest proximity to present-day Northeastern African populations from Somalia (Somali) and Ethiopia (including Oromo Ethiopians and Ethiopian Jews) as well as to various Northern African populations, including those from Morocco, Tunisia, Algeria, as well as Egypt and Libya, though the latter two were located slightly father away in PCA space, suggesting less genetic affinity.

Consistent with Figure 8.1, the PCA plot presented as Figure 8.2 also partially supported Expectation 2a. Figure 8.2 presented a higher-resolution PCA plot that was structured around two genetic clines that broadly correlated to geography in Africa and the Near East. One cline ran along an approximate west-east axis with Western African populations from Nigeria, Gambia, Senegal, and Sierra Leone at one end, and Eastern African populations from Kenya, Tanzania, Somalia, and Ethiopia at the other, while the other cline ran along an approximate north-south axis with Near Eastern populations, particularly those proximate to the Arabian Peninsula, at one end, and Eastern African populations from Kenya, Tanzania, and South Sudan at the other. The Kulubnarti Nubians were projected along this north-south axis in a space that corresponded roughly to their geographic position between Northern Africa and Eastern sub-Saharan Africa. Their projected location represented an intermediate position in relation to present-day populations from Northern Africa and Northeastern Africa, suggesting they shared more affinity with present-day Somali, Oromo Ethiopians, and Ethiopian Jews from the Horn of Africa, and present-day Moroccans, Tunisians, Algerians, and Libyans, who were plotted in closer proximity to the Kulubnarti Nubians than present-day Egyptians or any

Northern African Jews. It also suggested more genetic affinity to Near Eastern populations from the Arabian Peninsula and Bedouin groups than to populations from Western Africa and to some other Eastern African groups, including the Luo, Luhya, and Bantu from Kenya and the Dinka from South Sudan.

It is very likely that the biogeographic affinities detected using PCA are a reflection of evolutionary processes (e.g., gene flow, genetic drift) that occurred since the Early Christian Period. It is particularly possible that the perceived genetic similarity between the Kulubnarti Nubians and Northern African populations (i.e., those from Morocco, Tunisia, Algeria, and Libya) are a remnant of the genetic exchanges associated with the development and expansion of the trans-Saharan caravan trade over the last millennium and that this suggestion of affinity would not be observed if Early Christian Period individuals from across Northern Africa were included in this analysis.

The impact of demographic and evolutionary processes was also reflected by the projected location of the Kulubnarti Nubians in Figure 8.2 in a space that did not overlap with any present-day populations. This suggested that the Kulubnarti Nubians were a population that no longer exists in unadmixed form today. The recognition of the genetic uniqueness of the Kulubnarti Nubians through their projected location in otherwise unoccupied PCA space was consistent with Expectation 2a.

Initial exploratory approximations of biogeographic affinity made using PCA between the Kulubnarti Nubians and present-day genotyped populations were investigated in greater detail by assessing the ancestral components of the Kulubnarti Nubians along with a reference panel of worldwide present-day populations to place the Nubian samples within a landscape of human genetic diversity. Broadly, while the PCA

plots (Figures 8.1 and 8.2) presented in this chapter ambiguously suggested that the Kulubnarti Nubians shared genetic affinity to present-day populations from both Northern and Northeastern Africa, ADMIXTURE analysis refined this approximation, detecting two primary ancestry signals in the Kulubnarti Nubians (lavender and coral components, maximized in various present-day populations described earlier in this section) in proportions most closely associated with present-day populations from Northeastern Africa.

Thereby, unlike the results from PCA, the results of ADMIXTURE analysis were clearly consistent with Expectation 2a, finding that the Kulubnarti Nubians were estimated as having components of ancestry in proportions associated more closely with present-day populations from Northeastern Africa than with present-day populations from any other world region.

Figure 8.3 illustrated that the ancestry signals detected in the Kulubnarti Nubians were associated with a range of present-day populations, including populations spanning Europe, the Caucasus, the Near East, and Northern and Eastern Africa; however, placing the Kulubnarti Nubians within the present-day genetic landscape also revealed that present-day genotyped populations from Northeastern Africa, including Ethiopian Jews, Oromo Ethiopians, and Somalis, exhibited the same components of ancestry in proportions most closely related to those associated with the Kulubnarti Nubians. This result was consistent with the location of Ethiopian Jews, Somali, and Oromo Ethiopians in close proximity to the Kulubnarti Nubians in PCA space and was also supported by low levels of genetic distance between the Nubians and these present-day Northeastern

African populations using F_{ST}.[138] The geographic locations of these three present-day Northeastern African populations are shown in relation to Kulubnarti in Figure 8.6.

Figure 8.6: Geographic location of Ethiopian Oromo, Ethiopian Jews, and Somali in relation to Kulubnarti.

Based on the results presented in Figure 8.3, Figure 8.5 presented a focused comparison of the ancestry components at K=11 for the Ethiopian Jews, Somali, Oromo Ethiopians, and Kulubnarti S and R communities. For all individuals in these populations, the lavender component maximized in some populations from present-day Eastern Africa was the dominant ancestral signal, closely followed by the coral component maximized in some populations from present-day Northern Africa and the Near East.

[138] PCA results can be found in this chapter, while F^{ST} results can be found in Chapter 7.

Though they shared these primary ancestry components in similar proportions, the Kulubnarti Nubians were also distinguishable from these three present-day Northeastern African groups, further emphasizing their genetic uniqueness as initially suggested by their location in PCA space. Specifically, a very minor signal from a present-day Caucasian component (red component) and/or European component (yellow component) was detected in most individuals from Kulubnarti. While the red component was also detected as a very minor ancestral component in Ethiopian Jews and Oromo Ethiopians, the yellow component was unique to the Kulubnarti Nubians. Additionally, a minor signal from an Eastern African component maximized in the Hadza was detected in the Ethiopian Jews, Somali, and Oromo Ethiopians, but was not present in any individual from Kulubnarti. This finding of the genetic uniqueness of the Kulubnarti Nubians in comparison to present-day populations from around the world was consistent with Expectation 2a.

Though the PCA plot in Figure 8.2 suggested that the Kulubnarti Nubians showed affinity to present-day Moroccans, Tunisians, Algerians and Libyans, a more highly-resolved ADMIXTURE analysis suggested that these Northern African populations had both a greater number of ancestry components as well as different proportions of these components when compared to the Kulubnarti Nubians and illustrated in Figure 8.3. Specifically, in comparison to the Kulubnarti Nubians, a component maximized in some Western, Central, and Eastern African populations (the “teal component” in Figure 8.3) was detected in these Northwestern African groups, the yellow component (detected primarily in present-day Europeans) was represented at a much more substantial level, and the lavender component (maximized in some Eastern African populations) was

represented at a much less substantial level. It is therefore likely that the similar proportion of the Northern African/Near Eastern coral component contributed to the suggestion of biogeographic genetic affinity between the Kulubnarti Nubians and present-day Moroccans, Tunisians, Algerians, and Libyans when PCA was performed.

The projected location of the Kulubnarti Nubians on the PCA was likewise nearer to specific Near Eastern populations (Bedouins and populations from the Arabian Peninsula), suggesting greater biogeographic affinity with these particular groups than with other Near Eastern populations. While Near Eastern populations were generally represented on the ADMIXTURE output by a more substantial signal associated with Northern African and Near Eastern ancestry (coral component in Figure 8.3) than the signal maximized in Caucasian populations (red component in Figure 8.3), as well as a marked reduction in the component associated with European populations (yellow component in Figure 8.3), the specific populations that shared greater biogeographic affinity with the Kulubnarti Nubians also contained a signal from the lavender component maximized in some Eastern African populations in addition to the dominant coral component. In particular, this lavender component was detected in the Bedouin groups as well as in populations from the Arabian Peninsula (Yemeni, Yemenite Jews, Palestinians, and Saudis). It is likely this small Eastern African ancestral component was reflected in the projection of these particular Near Eastern groups in closer proximity to the Kulubnarti Nubians on PCA plots than other Near Eastern groups.

In addition, it was detected using PCA that while the Kulubnarti Nubians plotted closest to Northeastern African populations, including Ethiopian Jews, Somali, and Oromo Ethiopians, they plotted further from other Eastern African populations, including

the Bantu from Kenya, Luo, Luhya, and Dinka. In addition, no population from the latter grouping of Eastern African populations was determined to be particularly closely related to the Kulubnarti Nubians as analyzed through F_{ST} (i.e., not included in the top 20 or 21 closest related present-day populations). The results of ADMIXTURE analysis were consistent with these findings and suggested that some Eastern African populations contained very different ancestry signals in comparison to the Kulubnarti Nubians. Specifically, these populations were not associated with having the Northern African/Near Eastern signal represented by the coral component, and instead were determined to have a dominant component found throughout many Western, Central, and Eastern African populations (teal component in Figure 8.3) or a unique component maximized in the Hadza (royal blue component in Figure 8.3).

Finally, in the initial PCAs presented in Chapter 7, two PCA-outliers were detected: R201 and R181. While these two individuals were not diverse enough to be considered part of a population separate from the rest of the Kulubnarti population as a whole, an analysis of their biogeographic genetic affinities and ancestry components may shed additional light on their outlier status. These two outlier individuals appear on the PCAs to be shifted toward Eastern African populations including Ethiopian Jews, Somali, and Oromo Ethiopians, and away from Northern African/Near Eastern populations. From the ADMIXTURE plot presented as Figure 8.4, it can be seen that both of these individuals are detected as having a more prominent signal from the lavender component maximized in some Eastern African populations and a less prominent signal from the coral component maximized in some Northern African/Near Eastern populations than any other R or S community individuals, consistent with the observations made using the

PCA plots that suggested two R community individuals had more affinity to Eastern African populations than any other Kulubnarti Nubian individual. The results from ADMIXTURE, while providing additional detail on their separation from the rest of the Kulubnarti Nubians in PCA space, also confirms that they were genetically similar enough to the other Nubians from Kulubnarti to be considered members of that population.

Ultimately, the null hypothesis of no difference in biogeographic affinities or ancestry components between the S and R communities was not rejected: the results of both PCA and ADMIXTURE analysis supported Expectation 2 and were consistent with the results presented in Chapter 7.

At a more detailed level, while PCA was unable to provide evidence in support of Expectation 2a, this expectation was clearly supported by cluster-based ADMIXTURE analysis which suggested that the Kulubnarti Nubians had components of ancestry in proportions associated more closely with present-day populations from Northeastern Africa than with present-day populations from any other world region. Following the investigation of biogeographic genetic affinities and ancestry components using autosomal DNA, additional expectations regarding the mtDNA and Y chromosome haplogroup profiles of the Kulubnarti communities were explored through the analysis of uniparental markers.

8.3 Determining Haplogroup Profiles at Kulubnarti

8.3.1 Analytical Method: Calling mtDNA Haplogroups

While mitochondrial DNA (mtDNA) is, in effect, a single genetic locus that comprises only a small fraction of the human genome (Wilkins 2006), it is a powerful tool for studying female demographic history due to its presence in high copy number in the cell (Satoh and Kuroiwa 1991), its high mutation rate (Brown et al. 1979), its lack of recombination (Olivo et al. 1983), and its maternal mode of transmission from mother to offspring of both sexes (Giles et al. 1980).[139] Because offspring inherit an exact copy of their mother's mtDNA barring *de novo* mutations, sequential accumulations of mutations can be traced throughout the world along radiating maternal lineages of haplogroups, phylogenetic clusters represented by combinations of alleles inherited together that pertain to a single line of descent (Torroni et al. 1993, 2006). These mutations are rarely influenced by natural selection, and instead are only influenced by the evolutionary process of genetic drift.[140] This results in the geographic structure of mtDNA haplogroup distribution (Chen et al. 1995), such that analyzing the mutations found throughout the mitogenome can provide a record of an individual's matrilineal biogeographic ancestry (Kivisild et al. 2005; Torroni et al. 2006; Wilkins 2006).

All contemporary human mtDNA branched off from a common ancestor (commonly referred to as "Mitochondrial Eve") who existed ~150,000–200,000 years ago in Africa, where the highest mtDNA diversity is detected (Cann et al. 1987; Ingman et al. 2000; Soares et al. 2009; Fu et al. 2013b). An initial split occurred between haplogroups L0 and L1'2'3'4'5'6 (L1-6) (Van Oven and Kayser 2009), the two "daughter branches" of human mtDNA phylogeny (Behar et al. 2008); each of these haplogroups contains numerous sub-haplogroups that exhibit complex geographic patterning (Gonder

[139] Additional detail about mtDNA can be found in Chapter 4.
[140] Evolutionary forces, including genetic drift, are discussed in detail in Chapter 4.

et al. 2007). While the L0 haplogroup is confined primarily to Eastern and Southeastern Africa, the L1–6 branch is geographically widespread across the African continent (Behar et al. 2008). Further, the L3 haplogroup split off from the L1–6 branch approximately 60,000–70,000 years ago (Soares et al. 2012) and gave rise to two clades, M and N, that were the progenitors of all non-African haplogroups, and the various haplogroups that descended from them now represent the mtDNA pool of all non-Africans (Torroni et al. 2006).[141] Specifically, descendants of haplogroups M, N, and R (which diverged from haplogroup N) can be found all over Eurasia, America, Australasia, and Oceania, with sub-branches that are roughly specific to major geographic regions (Torroni et al. 2006).

Broadly, individuals assigned to macrohaplogroup L, defined traditionally by the presence of a T allele at nucleotide pair (np) 3954 and G allele at np 10398, are assessed as having matrilineal ancestry from sub-Saharan Africa, making this an "African-specific" macrohaplogroup (Denaro et al. 1981; Chen et al. 1995; Fox 1997; Wallace 1999; Lell and Wallace 2000), while individuals assigned to all other haplogroups are evaluated as having non-African matrilineal ancestry. Within Africa, the various lineages of haplogroup L are distributed in ways that reflect specific regional and population histories (e.g., Watson et al. 1997; Chen et al. 2000; Salas et al. 2002, Silva et al. 2015), with the richest haplogroup diversity found in Eastern Africa (Salas et al. 2002; Mishmar et al. 2003; Torroni et al. 2006; Gonder et al. 2007). Unfortunately, many deep African mtDNA lineages, particularly L0 and L1, have complex histories that are poorly understood (Gonder et al. 2007); however, the understanding of the phylogeny and present-day geographic distribution of mtDNA haplogroups has benefitted from analyses

[141] Excluding the descendants of migrations from Africa that occurred within the past several thousand years (Torroni et al. 2006).

of ancient individuals, which provide a new window into the reconstruction of when and where branching events occurred.

Complete and high-coverage sequencing of human mitogenomes has identified a unique set of mutations diagnostic of each haplogroup (van Oven 2015). An individual can therefore be assigned to a haplogroup based on the detection of the specific set of mutations using techniques such as whole-genome mitogenome sequencing or capture of haplogroup-determining SNPs from throughout the mitogenome. Determination of mtDNA haplogroup is an important component in studies of the genetic composition of an individual or of a population.

Specifically, analysis of mtDNA haplogroups can provide information about matrilineal ancestry because the geographic distribution of mtDNA haplogroups is structured by the behavior of females, including female-mediated migration (e.g., Kivisild et al. 2004; Kujanová et al. 2009; Gallego-Llorente et al. 2015; Hervella et al. 2016; Molto et al. 2017) and culturally-based patterns of residence. In addition, the diversity and geographic origins of mtDNA haplogroups within a population can be investigated on the population level to provide a clearer picture of population-wide matrilineal demography and can be compared with patrilineal demography as determined by Y-chromosome haplogroup in order to explore possible sex-specific demographic processes (Passarino et al. 1998; Marchani et al. 2008; Szécsényi-Nagy et al. 2015).

In this dissertation, mtDNA haplogroups are called for each individual from Kulubnarti. These haplogroups are then used to assess matrilineal ancestry at an individual level, explore possible community-based patterns of diversity in matrilineal

ancestry at a population level, and assess biogeographic differences between matrilineal and patrilineal origins when investigated alongside Y-chromosome data.

8.3.2 mtDNA Haplogroups of the Kulubnarti Nubians

To explore the matrilineal biogeographic ancestry of the Kulubnarti Nubians, the mitogenome sequencing data generated for each sample by the MTspike3k screening capture (described in Chapter 6) were analyzed with the software Phy-Mer (Navarro-Gomez et al. 2015). Phy-Mer is a haplogroup-defining algorithm that uses a k-mer approach, decomposing a mitochondrial sequence that has been aligned to the revised Cambridge Reference Sequence (rCRS, GenBank NC_012920; Anderson et al. 1981; Andrews et al. 1999) into a set of all possible k-mers which are then compared against each of the k-mer sets of all haplogroups in a library of representative sequences corresponding to each haplogroup. The rCRS-oriented PhyloTree Build 16 (Van Oven and Kayser 2009; www.phylotree.org) provided this k-mer library of representative sequences.

Phy-Mer decomposed mitogenome sequencing data from the 28 Kulubnarti Nubian samples into a set of all possible 12-mers and compared them against each of the 12-mer libraries of all haplogroups. The option '--def-snp=file.csv' was used to display the haplogroup-defining SNPs in the output, and the option '--min-DoC=10' was used to restrict the consideration to 12-mers that occurred a minimum of 10 times in the set of all 12-mers. For each individual, a "best match" haplogroup was called along with a haplogroup prediction score that indicated the quality of the input data in terms of the amount of sequence or polymorphisms an input mtDNA sequence was missing, and how

reliable the Phy-Mer-generated haplogroup call was; a higher score (maximum score of 1) indicated a more accurate haplogroup call (Navarro-Gomez et al. 2015).

In addition to Phy-Mer analysis, mtDNA haplogroups for each Kulubnarti sample were concurrently called by the Reich Lab's automated data-analysis pipeline using the software HaploGrep2 (Kloss-Brandstätter et al. 2011; Weissensteiner et al. 2016) and the Reconstructed Sapiens Reference Sequence (RSRS; Behar et al. 2012) as the mitochondrial reference genome.[142] HaploGrep2 is another haplogroup classification method that considers pre-calculated phylogenetic weights that reflect the mutational stability of the haplogroup-defining variants in the mitogenome and is based on Phylotree 17, the newest build of the phylogenetic tree of human mtDNA variation (www.phylotree.org; van Oven and Kayser 2009; van Oven 2015). While Phy-Mer and HaploGrep2 results were consistently in agreement regarding the major haplogroup of each individual, haplogroup calls differed slightly for some individuals, mainly in the specificity of sub-lineage level; for example, L0a1a (Phy-Mer) versus L0a1a1 (HaploGrep2). Supplement 8 presents the haplogroup calls from both Phy-Mer (including haplogroup prediction score and haplogroup-defining SNPs) and HaploGrep2 (including an estimated match to the consensus sequence with 95% confidence interval). Table 8.1 presents the mtDNA haplogroup for each individual with the reported level of sub-lineage specificity determined by the agreement of both Phy-Mer and HaploGrep2 analyses along with the approximate geographic association of that haplogroup.

Table 8.1: mtDNA haplogroup calls for 28 Kulubnarti Nubians determined by consistency between Phy-Mer and HaploGrep2 output and the approximate geographic association of each haplogroup.

[142] The difference between the rCRS and RSRS is explained in Chapter 6.

Sample ID	mtDNA Haplogroup	Approximate Geographic Association
R101	H2a	Europe/Caucasus/Near East
R124	K1a19	Italy, Iran, Armenia, Turkey, Ashkenazi Jews
R182	U5b2b	Europe
R186	L2a1	Across African continent, peaks in Ghana, Mozambique, Sudan
R59	H2a	Europe/Caucasus
R152	L0a1a	Southern/Eastern Africa
R181	H2a	Europe/Caucasus/Near East
R202	H2a	Europe/Caucasus/Near East
R196	L2a1d	Eastern Africa
R201	J2a2	Near East/Northern Africa
R5	N1b1	Near East/Caucasus/Arabian Peninsula/Northern Africa
R79	U1a	Eastern Europe/Near East
R93	H2a	Europe/Caucasus/Near East
R169	U5b2b	Europe
S33	L0a1a	Southern/Eastern Africa
S50	L0a1a	Southern/Eastern Africa
S68a	L2a1d	Eastern Africa
S81	L2a1d	Eastern Africa
S149	H2a	Europe/Caucasus/Near East
S133	L2a1d	Eastern Africa
S73	L2a1d	Eastern Africa
S87	L2a1d	Eastern Africa
S159	L2a1d	Eastern Africa
S17	T1a	Europe/Near East/North Africa/Central and North Asia
S198	L1b1a2	Western/Central Africa; also found in Ethiopians/Nubians
S27	U5b2b	Europe
S79	L5a1b	Eastern/Southeastern Africa
S89	H2a	Europe/Caucasus/Near East

8.3.3 Interpretation of mtDNA Haplogroups

As discussed in Chapter 5, the null hypothesis is that there was no community-based difference in genetic composition between the S and R communities at Kulubnarti,

including in terms of mtDNA haplogroups. Expectation 2b of this dissertation was that the null hypothesis would not be rejected, and that the majority of matrilineal lineages called for members of both Kulubnarti communities would be primarily of Northeastern African origin and distribution (e.g., broadly, macrohaplogroup L; more specifically, sub-lineages such as L0a).

Eleven different haplogroups were identified among the 28 analyzed individuals from Kulubnarti. Inconsistent with Expectation 2b, African-specific mtDNA lineages and non-African-specific mtDNA lineages were represented in the Kulubnarti Nubians at similar frequencies. African-specific L haplogroups were called in 13 individuals (46% of the Kulubnarti population), while non-African-specific haplogroups were called in 15 individuals (54% of the Kulubnarti population). All non-African-specific haplogroups were geographically associated with Eurasia.

Across the Kulubnarti population, African-specific mtDNA lineages from macrohaplogroup L included L0a1a (11%), L1b1a2 (4%), L2a1/L2a1d (29%), and L5a1b (4%); a more detailed description of the approximate geographic distribution of each of these haplogroups is provided in Supplement 8.

Briefly, these lineages are largely restricted to sub-Saharan Africa (Rosa et al. 2004; Salas et al. 2004) or regions with historical gene flow from sub-Saharan Africa (Kivisild 2015). L0a1a, L2a1/L2a1d, and L5a1 (the parent branch of L5a1b) originated in or are historically and presently found at a high frequency throughout Eastern and Northeastern Africa; therefore, the detection of such sub-lineages at Kulubnarti was expected. The detection of an L1b sub-lineage was more difficult to reconcile due to its primary concentration in Western, Central, and Northern Africa and near absence in

Eastern and Southern Africa. Upon further investigation, the specific L1b1a2a sub-lineage assigned to this individual by HaploGrep2 was postulated to have originated in Eastern Africa based on its representation by multiple divergent sequences from Ethiopia (Behar et al. 2008; Cerezo et al. 2012; Messina et al. 2017), Egypt (Behar et al. 2008), and Somalia (Mikkelsen et al. 2012). As such, it is possible that the Kulubnarti individual assigned to this haplogroup was a member of a local L2b1a2a matrilineal lineage, though assignment to this sub-lineage was not confirmed by Phy-Mer. Ultimately, the presence of these L lineages, originating or historically and presently found at high frequencies in Northeastern Africa, suggested a matrilineal genetic signature of "local" ancestry for some of the Kulubnarti Nubian population.

While a genetic signature of local matrilineal ancestry was suggested for some members of the Kulubnarti population, other members of the population were assigned to haplogroups of non-African-specific origin presently distributed widely throughout Eurasia, a result inconsistent with Expectation 2b, but nonetheless unsurprising based on previous evidence of historic admixture between Nubian populations and Near Eastern groups. Non-African-specific Eurasian-associated mtDNA lineages detected at Kulubnarti included H2a (25%), J2a2 (4%), K1a19 (4%), N1b1 (4%), T1a (4%), U1a (4%), and U5b2b (11%); a more detailed description of the approximate geographic distribution of each of these haplogroups is provided in Supplement 8.

The detection of non-African-specific haplogroups with origins traced to and with a geographic distribution radiating out from a Near Eastern epicenter at Kulubnarti suggested that female-mediated gene flow from Eurasia into Northeastern Africa occurring prior to or during the Early Christian Period influenced even isolated regions of

Upper Nubia, such as the *Batn el Hajar*. Haplogroups J2a2, K1a, N1b1, T1a, and U1a were suggestive of a genetic signature of Near Eastern matrilineal ancestry at Kulubnarti. Haplogroup H2a, found at the highest frequency of any Eurasian haplogroup at Kulubnarti, is widely distributed throughout Eurasia and is found at an appreciable frequency on the Arabian Peninsula (Ennafaa et al. 2009). It would therefore be most parsimonious to suggest that this haplogroup may have been introduced through Arab migration. The detection of haplogroup U5b2b at Kulubnarti was more difficult to reconcile because the U5b lineage is found at its most appreciable frequency in northern Europe and the Iberian Peninsula, though it has a geographically-broad distribution and has also been detected at notable frequencies in a wide range of populations; for example, it has been detected in Egyptian Berbers (Rando et al. 1998; Richards et al. 1998, 2000; Coudray et al. 2009). As the diversification and distribution of U5b and specifically U5b2b remains poorly understood, its presence at Kulubnarti is perplexing; however, the assignment of three Kulubnarti individuals to haplogroup U5b2b is suggestive of a genetic signature of European matrilineal ancestry at Kulubnarti.

Ultimately, it was revealed that the while the Kulubnarti Nubians harbored some of the most ancestral African-specific mtDNA lineages (e.g., L0a, L5), they also harbored non-African-specific Eurasian-associated lineages as well (e.g., J, K). The presence of both local and non-local matrilineal lineages suggested the influence of female-mediated gene flow from Eurasia into isolated regions of Nubia. Interestingly, mtDNA haplogroups associated with an early Eurasian backflow into Eastern Africa, including U6 and M1 (Macaulay et al. 1999; Maca-Meyer et al. 2003; González et al. 2007), were

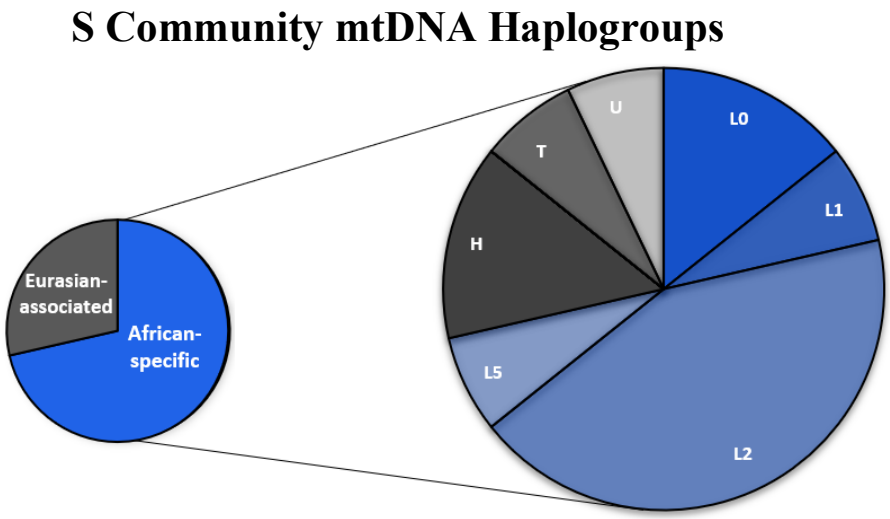
absent from Kulubnarti, suggesting that the genetic signature of this specific back-migration is not observed in the Kulubnarti Nubians.

Just as the Kulubnarti Nubians were assessed as having more affinity to present-day genotyped populations from Northeastern African than any other world region when autosomal DNA was analyzed, the presence of both African-specific and non-African-specific Eurasian-associated mtDNA lineages was similar to the pattern observed in many present-day populations from Northeastern Africa (Passarino et al. 1998; Non et al. 2011). This observation suggests that female-mediated migration has been occurring in this region for millennia and affects even more isolated populations than previously thought. Further implications of these mtDNA haplogroup assignments will be discussed in greater detail in Chapter 9.

While Expectation 2b, which predicted the majority presence of African-specific mtDNA haplogroups that originated and are presently found at high frequencies in Northeastern Africa, was not supported by the Kulubnarti mtDNA data, the null hypothesis of no community-based differences in genetic composition between the S and R communities in terms of mtDNA haplogroup profiles required community-level analysis. If this hypothesis is not rejected, both communities would be revealed to have an approximately even proportion of African-specific and non-African-specific Eurasian-associated mtDNA haplogroups. To test this hypothesis, individuals were separated by community and mtDNA haplogroup profiles were created for each community and examined at the community level.

Although the analysis of autosomal DNA discussed earlier in this chapter suggested similar genetic composition between the S and R communities in terms of

biogeographic affinities and ancestry components and failed to reject the null hypothesis, the mtDNA haplogroup profiles for the S and R communities were inconsistent with these results. Instead, the analysis of Kulubnarti Nubian mtDNA haplogroups suggested an overrepresentation of genetic signatures from either African-specific or non-African-specific Eurasian-associated mtDNA haplogroups in each of the communities, rejecting the null hypothesis of no genetic differences between the S and R communities in terms of mtDNA haplogroup profiles. These results are illustrated in Figure 8.7.



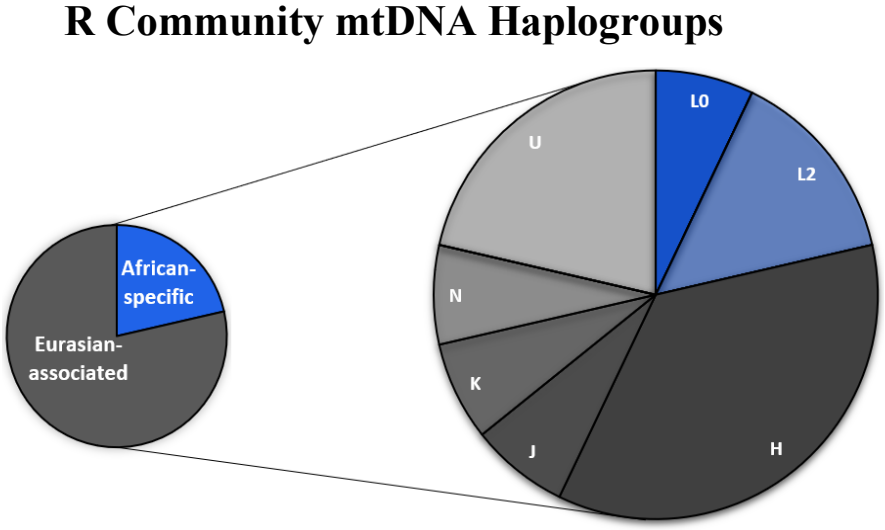


Figure 8.7: mtDNA haplogroup calls for the S and R communities, and their distributions in terms of African-specific or Eurasian-associated haplogroups. A blue background indicates an African-specific haplogroup, while a gray background indicates a Eurasian-associated haplogroup.

When examined at the community level, African-specific mtDNA haplogroups were overrepresented in the mtDNA profile of the S community, comprising 71% of mtDNA haplogroups, and non-African-specific Eurasian-associated haplogroups were underrepresented in the mtDNA profile of the S community, comprising only 29%. The opposite pattern was seen in the mtDNA profile of the R community, where African-specific mtDNA haplogroups were underrepresented, comprising only 21% of mtDNA haplogroups, and non-African-specific Eurasian-associated haplogroups were overrepresented, comprising 79% of mtDNA haplogroups.

Inconsistent with Expectation 2b, these data suggest that genetic signatures of female-mediated gene flow from Eurasia were not represented evenly within the two communities comprising the population of Kulubnarti. Instead, a Eurasian-associated matrilineal genetic signature was overrepresented in the R community and underrepresented in the S community, with the latter comprised of individuals assigned to mtDNA haplogroups of a more "local" African-specific origin.

In addition to being inconsistent with expectations, these mtDNA chromosome data reject the null hypothesis of genetic similarity between the S and R communities in terms of mtDNA chromosome haplogroup profiles. This discrepancy in the mtDNA profiles will be explored further in light of Y-chromosome haplogroup data at the end of this chapter, and the implications of the different genetic compositions of the Kulubnarti

communities in terms of mtDNA haplogroups and the consequential rejection of the null hypothesis will be discussed in greater detail in Chapter 9.

8.3.4 Analytical Method: Calling Y Chromosome Haplogroups

Complementary to analyses of mtDNA detailing female demographic histories, investigations of the Y chromosome are essential to the reconstruction of male demographic histories (Jobling and Tyler-Smith 2003). The Y chromosome is unique because it is male-specific (Heyer et al. 2012), haploid (present in a single copy), and passed directly from father to male offspring only (Underhill et al. 2000, 2001; Jobling and Tyler-Smith 2003; Jobling et al. 2014). Similar to mtDNA, which does not undergo recombination, the majority (~95%) of the Y chromosome is non-recombining; this region is known as the Non-Recombining portion of the Y (NRY) or the Male-Specific region of the Y (MSY) (Jobling and Tyler-Smith 1995; Bachtrog and Charlesworth 2001; Tilford et al. 2001). Because of its lower effective population size (i.e., it constitutes only one-quarter of the sex chromosomes in an entire population) (Jobling and Tyler-Smith 2003; Lippold et al. 2014), the Y chromosome is especially prone to genetic drift (Underhill et al. 2001), rapidly accruing mutations that can be studied along radiating lineages in terms of Y chromosome haplogroups (or just "Y haplogroups") which represent sets of mutations inherited together, like mtDNA (Underhill and Kivisild 2007). High genetic drift has led to geographic structuring of Y haplogroups, characterized by large differences between populations; this makes the Y chromosome the most geographically-informative single locus in the human genome (Jobling et al. 2014).

Combinations of SNPs on the Y chromosome are most commonly used to identify haplogroups within Y chromosome phylogeny (Lippold et al. 2014). A tree of known

polymorphisms in the NRY has provided important insights into the demographic history of and paternal genetic relationships among human populations (Heyer et al. 2012; Reguig et al. 2014). The phylogenetic tree of the NRY has been progressively refined through the discovery and mapping of new SNPs resulting in the increased resolution of a growing number of distinct haplogroups (Y Chromosome Consortium 2002; Jobling and Tyler-Smith 2003, 2017; Karafet et al. 2008). The first branching of the tree separated the African-specific A clade (Hammer et al. 2001; Underhill et al. 2001) from clade BT (Underhill et al. 2000), while the second split divided clade B (almost entirely restricted to sub-Saharan Africa) (Underhill and Kivisild 2007; Karafet et al. 2008) from CT, which comprises the majority of African-specific as well as all non-African-specific haplogroups (Underhill et al. 2001; Underhill and Kivisild 2007; Karafet et al. 2008); subsequent branching events have led to further geographic patterning of haplogroups D–T throughout the world. As with mtDNA, analysis of branching events and the current geographic distribution of Y haplogroups supports an African origin for anatomically modern humans (Hammer et al. 1997, 1998; Underhill et al. 2000), with Southern and Eastern Africans sharing the deepest lineages of phylogeny (Semino et al. 2002).

As with mtDNA, the utility of the NRY comes from its uniparental mode of inheritance, which can provide insight into past social structure and the potential behavior of males (Jobling and Tyler-Smith 2017). In addition to genetic drift, the geographic distribution of Y haplogroups is structured by the behavior of males (Jobling and Tyler-Smith 2003), including male-mediated demographic expansions (e.g., Batini et al. 2015) as well as culturally-based residence patterns (e.g., Oota et al. 2001). As such, the study of Y haplogroups provides important insights into the paternal histories of human

populations, primarily by elucidating patrilineal biogeographic ancestry and allowing the reconstruction of the demographic processes that led to the observed Y haplogroup profile of a population.

The data obtained from analysis of the NRY is particularly informative when interpreted alongside analyses of mtDNA variation (Lippold et al. 2014). It is widely observed that human behavior and culture results in different patterns of mobility for males and females (Passarino et al. 1998; Oota et al. 2001; Destro-Bisol et al. 2004; Nasidze et al. 2005; Heyer et al. 2012), shaping the geographic distribution of their associated uniparental markers. As such, the comparison of mtDNA and NRY variation is critical to the development of a complete picture of human populations that includes both maternal and paternal demographic histories (Wilkins 2006; Heyer et al. 2012; Lippold et al. 2014).

In this dissertation, Y haplogroups are called for each individual from Kulubnarti. These haplogroups are then used to assess patrilineal ancestry at an individual level and to explore possible community-based patterns of diversity in Y haplogroup profiles at Kulubnarti. The Y haplogroup profiles of the Kulubnarti Nubians are examined independently as well as interpreted alongside mtDNA haplogroup data.

8.3.5 Y Chromosome Haplogroups of the Kulubnarti Nubians

To explore the patrilineal biogeographic ancestries of the Kulubnarti Nubians, 32,681 SNPs from the Y chromosome included on the 1240K SNP capture (Mathieson et al. 2015) were extracted and alleles were called at each covered SNP position. SAMtools v0.1.19 (Li et al. 2009) was used to restrict analysis to sites with mapping quality $Q \geq 30$ and two bases at the end of each sequenced fragment that were excluded due to the high

probability of *post-mortem* damage at these sites.[143] The majority allele was taken at each covered SNP.[144] The most derived SNP as well as downstream SNPs for which the sample was ancestral were used to call Y haplogroups based on the Y-DNA Chromosome Tree of the International Society of Genetic Genealogy (version 11.110, April 2016; http://www.isogg.org).[145] Different levels of resolution are the result of different numbers of SNPs covered. The Y haplogroups and their approximate geographic associations are presented in Table 8.2 for each of the 16 males from Kulubnarti.

Table 8.2: Y chromosome haplogroup calls for 16 male Kulubnarti Nubians and the approximate geographic associations of each haplogroup; if haplogroup is listed as "N/A," the individual was female.

Sample ID	Y-chromosome Haplogroup	Approximate Geographic Association
R101	E1b1b1b2	Northwestern Africa, Northern Africa
R124	E1b1b1a1b2	Northeastern Africa, Eastern Africa, Western Eurasia, Europe
R182	N/A	N/A
R186	N/A	N/A
R59	E1b1b1a1a1b	Eastern Africa (Horn of Africa, Sudan)
R152	N/A	N/A
R181	E1b1b1a	Northern Africa; Northeastern Africa
R202	N/A	N/A
R196	E1b1b1a1a1	Northern Africa, especially Southern Egypt; Sudan
R201	T1a1a	Near East, Eastern Africa
R5	N/A	N/A
R79	N/A	N/A
R93	E1b1b1b2	Northwestern Africa, Northern Africa
R169	T1a1a	Near East, Eastern Africa
S33	R2a3a	Central/Southern Asia
S50	J2a1	Near East, Caucasus, Europe

[143] A discussion of *post-mortem* damage is included in Chapter 4.
[144] Only one allele should occur at each position of the Y chromosome in males. The presence of a minority allele may represent aDNA damage, sequencing error, or contamination (Lazaridis et al. 2016).
[145] Y chromosome haplogroup calls were made by the Reich Lab.

S68a	N/A	N/A
S81	J1a	Near East, Caucasus, Europe
S149	N/A	N/A
S133	N/A	N/A
S73	J2a1	Near East, Caucasus, Europe
S87	N/A	N/A
S159	E2	Western and Eastern sub-Saharan Africa
S17	N/A	N/A
S198	G2a2b2a	Near East, Caucasus, Europe
S27	E1b1b1a1a1	North Africa, especially Southern Egypt; Sudan
S79	G2a2b2a1	Near East, Caucasus, Europe
S89	N/A	N/A

8.3.6 Interpretation of Y Chromosome Haplogroups

As discussed in Chapter 5, the null hypothesis is that there was no community-based difference in genetic composition between the S and R communities at Kulubnarti, including in terms of Y haplogroups. Expectation 2c of this dissertation was that the null hypothesis would not be rejected, and that the majority of patrilineal lineages called for male members of both Kulubnarti communities will be primarily of Northeastern African origin and distribution (e.g., broadly, macrohaplogroup E; more specifically, sub-lineages such as E1b1b1).[146]

Seven different haplogroups were identified among the 16 analyzed males from Kulubnarti. Inconsistent with Expectation 2c, African-specific Y chromosome lineages (lineages of the macrohaplogroups A, B, or E) and non-African-specific Y chromosome lineages were represented in the Kulubnarti Nubians at similar frequencies. Specifically, African-specific haplogroups from macrohaplogroup E were called in 8 individuals (50%

[146] All haplogroups are discussed according to ISOGG 2016 nomenclature.

of the Kulubnarti population), and non-African-specific haplogroups were called in 8 individuals (50% of the Kulubnarti population). All non-African haplogroups were geographically associated with Eurasia.

While no lineages from African-specific macrohaplogroups A or B were detected at Kulubnarti, lineages from African-specific macrohaplogroup E included E1b1b1a-V68 (6.25%) and its sub-haplogroups E1b1b1a1a1-V12 (12.5%), E1b1b1a1a1b-V32 (6.25%), and E1b1b1a1b2-L677 (6.25%), as well as E1b1b1b2-Z830 (12.5%) and E2-M75 (6.25%); a more detailed description of the approximate geographic distribution of each of these haplogroups is provided in Supplement 9.

Briefly, Y chromosome haplogroup E most likely originated in the northern hemisphere of Africa based on the concentration and variety of E subclades in this region (Cruciani et al. 2002; Semino et al. 2004) and is primarily found at high frequencies across Africa and lower frequencies in the Near East, Southern Europe, and into Western Asia (Underhill et al. 2001; Jobling and Tyler-Smith 2003; Semino et al. 2004; Wood et al. 2005; Cruciani et al. 2007; Karafet et al. 2008). It is found throughout the present-day Sudan, including in the Nubians (Hassan et al. 2008). One individual at Kulubnarti was assigned to the basal haplogroup E2, distributed throughout Africa and frequently found in Eastern Africans (Underhill et al. 2000; Luis et al. 2004), and all other individuals assigned E haplogroups were assigned to a branch of E1b1, a sub-haplogroup of the other basal haplogroup E1 and the most represented human Y haplogroup throughout Africa (Semino et al. 2004; Trombetta et al. 2011). More specifically, these individuals were assigned to the E1b1 sub-lineage E1b1b, found most frequently in Nilo-Saharan-speaking Eastern African groups from the Sudan as well as Ethiopia (Hassan et al. 2008), and most

specifically, to the E1b1b sub-lineage E1b1b1 and its branches, a haplogroup of Northeastern African origin (Cruciani et al. 2007) with a broad geographic distribution and a primary distribution in Northeastern Africa (Trombetta et al. 2015).

Assignment to lineages of macrohaplogroup E suggested a patrilineal genetic signature of "local" ancestry for half of the analyzed Kulubnarti Nubians. However, while a local patrilineal ancestry was suggested for some members of the population, other members were assigned to Y haplogroups of non-African-specific origin presently distributed throughout Eurasia. While this result was inconsistent with Expectation 2c, it is suggestive of a greater-than-expected genetic contribution to the Kulubnarti gene pool from males of Eurasian biogeographic ancestry who were historically evidenced to have traveled along the Nile Corridor. Non-African-specific Eurasian-associated Y chromosome lineages detected at Kulubnarti included G2a2b2a (6.25%) and its sub-lineage G2a2b2a1 (6.25%), J2a1 (12.5%), J1a (6.25%), R2a3a (6.25%), and T1a1a (12.5%); a more detailed description of the approximate geographic distribution of each of these haplogroups is provided in Supplement 9.

The detection of these non-African-specific Y haplogroups at Kulubnarti are suggestive of male-mediated gene flow from Eurasia into Northeastern Africa that occurred prior to or during the Early Christian Period and influenced settlements in isolated regions such as the *Batn el Hajar*. The detection of some of these non-African-specific Y haplogroup lineages was parsimonious with expectations of migration from the Near East into Nubia, such as the J lineages that have a strong presence throughout the Near East as well as Northeastern Africa today (Hammer et al. 2001; Quintana-Murci et al. 2001; Cinnioğlu et al. 2004; Hassan et al. 2008), and the G lineages, which have

peak frequencies in certain Caucasian and Near Eastern groups (Balanovsky et al. 2011; Rootsi et al. 2012) and have been detected among Arab groups in North Africa (Luis et al. 2004).

The presence of a lineage from haplogroup R2 at Kulubnarti was more difficult to explain due to its Asian origins and primary concentration in Southern and Central Asia with only low-level presence in the Caucasus, Near East, and Europe (Cinnioğlu et al. 2004; Di Cristofaro et al. 2013); so far, only R1 lineages have been detected at appreciable frequencies in areas where populations were most likely to have admixed with the Nubians (Luis et al. 2004; Wood et al. 2005; Hassan et al. 2008; Kujanová et al. 2009). However, the distribution of the R2 lineage outside of Central Asia has not been studied extensively, making it possible that the specific R2a3a lineage found in one individual at Kulubnarti was introduced through migration from a small pocket of individuals carrying this sub-lineage.

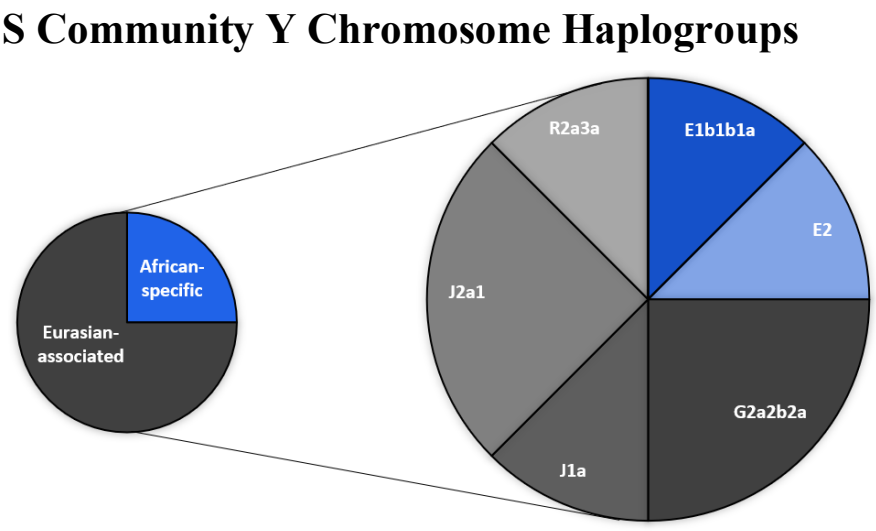
The biogeographic assignment of two unrelated individuals from Kulubnarti to T1a1a was also difficult to interpret, as the T haplogroup, rare and geographically widespread (Chiaroni et al. 2009; Mendez et al. 2011) likely arose in the Near East (Nogueiro et al. 2010) but is presently found at high frequency both there and in Eastern Africa (Somalia in particular) as well (King et al. 2007). Because the T lineage is found throughout the Arabian Peninsula, was most likely introduced to Eastern Africa via migration from the Near East, and has not been detected previously in Nubian populations (Hassan et al. 2008), this dissertation will categorize Y haplogroup T as suggestive of a genetic signature of Eurasian patrilineal ancestry plausibly introduced

through Arab migration. However, it should be kept in mind that this patrilineal lineage may trace its biogeographic ancestry back to the Northeastern Africa.

Ultimately, because previous studies using Y chromosome data have suggested that Northeastern Sudanese populations including Nubians were most extensively affected by Eurasian migrations since the introduction of Islam from the Arabian Peninsula through Egypt and possibly via the Red Sea starting in 651 C.E. (Hassan et al. 2008; Hollfelder et al. 2017), it was expected that the gene pool of the small and isolated population of Kulubnarti would not reflect the genetic signature of male-mediated migration from Eurasia. However, inconsistent with this expectation, the presence of both local and non-local patrilineal lineages suggested that male-mediated gene flow from Eurasia influenced isolated regions of Nubia prior to the mid-7th century. Further implications of these Y haplogroup assignments will be discussed in greater detail in Chapter 9.

While Expectation 2c, which predicted the majority presence of African-specific Y haplogroups that originated in and that are presently found at high frequencies in Northeastern Africa, was not supported by the Y chromosome data, the null hypothesis of no community-based differences in genetic composition between the S and R communities in terms of Y haplogroup profiles remained to be tested. If this hypothesis is not rejected, both communities would be revealed to have an approximately even proportion of African-specific and non-African-specific Eurasian-associated Y haplogroups. To test this hypothesis, individuals were separated by community and Y haplogroup profiles were created for each community and examined at the community level.

Inconsistent with the finding of similar genetic composition between the S and R communities in terms of biogeographic affinities and ancestry components based on autosomal DNA, which failed to reject the null hypothesis, the analysis of Y haplogroup profiles for the S and R communities suggested community-based differences in terms of Y haplogroup profile. Like the analysis of mtDNA haplogroups, the analysis of Kulubnarti Nubian Y haplogroups suggested an overrepresentation of genetic signatures from either African-specific or non-African-specific Eurasian-associated haplogroups in each of the communities, rejecting the null hypothesis of no genetic differences between the S and R communities in terms of Y chromosome haplogroup profiles. These results are illustrated in Figure 8.8.



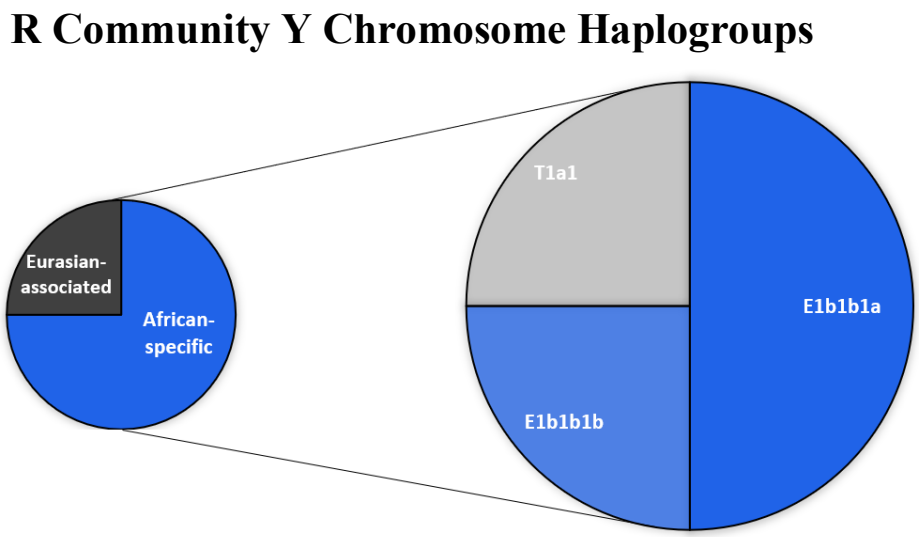


Figure 8.8: Y chromosome haplogroup calls for the S and R communities, and their distributions in terms of African-specific or Eurasian-associated haplogroups. A blue background indicates an African-specific haplogroup, while a gray background indicates a Eurasian-associated haplogroup.

When examined at the community level, African-specific haplogroups were underrepresented in the Y haplogroup profile of the S community, comprising 25% of assigned Y haplogroups, and Eurasian-associated haplogroups were overrepresented in the Y haplogroup profile of the S community, comprising 75% of assigned Y haplogroups. The opposite pattern was seen in the Y haplogroup profile of the R community, where African-specific haplogroups were overrepresented, comprising 75% of assigned Y haplogroups, and Eurasian-associated haplogroups were underrepresented, comprising 25% of assigned Y haplogroups.

Inconsistent with Expectation 2c, these data suggest that genetic signatures of male-mediated gene flow from Eurasia were not represented evenly within the two communities comprising the population of Kulubnarti. A Eurasian-associated patrilineal genetic signature was instead overrepresented in the S community and underrepresented in the R community, with the latter comprised of individuals assigned to Y haplogroups

of more "local" African-specific origin. In addition to being inconsistent with expectations, these Y chromosome data reject the null hypothesis of genetic similarity between the S and R communities in terms of Y chromosome haplogroup profiles. The implications of the different genetic composition of the Kulubnarti communities in terms of Y haplogroups and the consequential rejection of the null hypothesis presented in Chapter 5 will be discussed in greater detail in Chapter 9.

8.3.7 Comparing mtDNA and Y Chromosome Haplogroup Profiles of the Kulubnarti Nubians

Both the mtDNA and Y haplogroup profiles of the Kulubnarti Nubian population rejected the null hypothesis of no community-based differences in genetic composition between the S and R communities. When examined together, analysis of the mtDNA and Y chromosome data revealed that the two communities had contrasting patterns of uniparental genetic signatures. Specifically, while a Eurasian-associated matrilineal genetic signature was detected to a greater extent in the R community and to a lesser extent in the S community, the opposite pattern was seen when patrilineal genetic signatures were examined: a Eurasian-associated patrilineal genetic signature was instead detected to a greater extent in the S community and to a lesser extent in the R community. This contrasting pattern of genetic signatures detected by community-level analysis of mtDNA and Y haplogroup distribution is presented as Table 8.3.

Table 8.3: mtDNA and Y chromosome haplogroup profiles for the Kulubnarti S and R communities. A blue background indicates an African-specific haplogroup, while a gray background indicates a Eurasian-associated haplogroup. A white background in the "Y chromosome haplogroup" column indicates that the individual was female.

Sample ID	mtDNA Haplogroup	Y Haplogroup	
R101	H2a	E1b1b1b2	**R community**
R124	K1a19	E1b1b1a1b2	
R152	L0a1a		
R169	U5b2b	T1a1a	
R181	H2a	E1b1b1a	
R182	U5b2b		
R186	L2a1		
R196	L2a1d	E1b1b1a1a1	
R201	J2a2	T1a1a	
R202	H2a		
R5	N1b1b		
R59	H2a	E1b1b1a1a1b	
R79	U1a		
R93	H2a	E1b1b1b2	
S133	L2a1d		**S community**
S149	H2a		
S159	L2a1d	E2	
S17	T1a		
S198	L1b1a2	G2a2b2a	
S27	U5b2b	E1b1b1a1a1	
S33	L0a1a	R2a3a	
S50	L0a1a	J2a1	
S68a	L2a1d		
S73	L2a1d	J2a1	
S79	L5a1b	G2a2b2a1 (L140)	
S81	L2a1d	J1a	
S87	L2a1d		
S89	H2a		

Overall, the comparison of mtDNA and Y haplogroup profiles of the two Kulubnarti communities suggests that the genetic signal of Eurasian-associated uniparental ancestry is detected within the population of Kulubnarti in both a sex-based and a community-based pattern. As a result, the two socially-distinct communities at Kulubnarti did not have the same genetic compositions in terms of uniparental haplogroup profiles, rejecting the null hypothesis of no community-based genetic

differences presented in Chapter 5. The implications of this contrasting pattern of uniparental ancestries at Kulubnarti will be explored in greater detail in Chapter 9.

8.4 Chapter Summary

After establishing in Chapter 7 that there was no evidence of community-based genetic population substructure at Kulubnarti, a finding which failed to reject the null hypothesis that the S and R communities at Kulubnarti comprised a single genetic population, the focus was shifted to answering more detailed questions about the genetic composition of the Kulubnarti Nubians. These questions, initially presented in Chapter 1, investigated the genetic composition of the Kulubnarti Nubians in terms of biogeographic genetic affinities and components of ancestry and matrilineal and patrilineal origins. Aim 2 of this dissertation focused on answering these questions by exploring the biogeographic affinities and components of ancestry of the S and R communities using autosomal DNA and by creating mtDNA and Y haplogroup profiles for the communities using SNPs from the mitogenome and Y chromosome.

As discussed in Chapter 5, the null hypothesis associated with Aim 2 is that there were no community-based differences between the S and R communities in biogeographic affinities or ancestry components, mtDNA haplogroup profile, or Y haplogroup profile. It was expected that the null hypothesis would not be rejected (i.e., no community-based differences in genetic composition would be identified), and further, that the Kulubnarti Nubians from both communities would share more biogeographic affinity with present-day genotyped populations from Northeastern Africa than from any other world region, but would be genetically distinguishable from all present-day

populations, and that the majority of individuals from both communities would be assigned mtDNA and Y haplogroups with origins and distribution in Northeastern African.

The results presented in this chapter do not consistently fail to reject the null hypothesis, but instead indicate that community-based differences in genetic composition are only apparent when uniparental markers are analyzed. Specifically, the analysis of biogeographic genetic affinities and ancestry components supported the aforementioned expectations and suggested that the Kulubnarti Nubians from both communities shared greater genetic affinity to present-day genotyped populations in Northeastern Africa than to populations in any other world region. These results, based on analysis of autosomal SNPs, suggested no community-based differences, failing to reject the null hypothesis. In contrast, the analysis of mtDNA and Y haplogroup profiles for the S and R communities were found to reject the null hypothesis, instead suggesting that each community had its own unique uniparental genetic composition. The implications of these findings will be further discussed in the final chapter.

CHAPTER 9

A SUMMARY: KULUBNARTI FROM A GENOMIC PERSPECTIVE

"Bones are like old books in strange languages. Learn how to read them and they have wonderful tales to tell." (Armelagos and Van Gerven 2017: 1)

9.1 Chapter Overview

The skeletal remains recovered from the Early Christian Period (550–800 C.E.) site of Kulubnarti in the *Batn el Hajar* region of Sudanese Nubia have been the subject of a decades-long biocultural research program. Biodistance studies using craniometrics and discrete dental traits have suggested a high level of biological similarity between two cemetery communities from Kulubnarti and have detected morphological convergence between the Kulubnarti Nubian population and neighboring Nubian populations (Greene 1982; Van Gerven 1982; Vollner 2016). In contrast, analyses of health and survival have provided clear evidence that one of these cemetery communities (the S community) was substantially more stressed than their counterparts of the landowning R community (e.g., Van Gerven et al. 1981, 1995; Hummert 1983; Hummert and Van Gerven 1983).

Specifically, it appeared that members of the S community experienced greater amounts of stress for longer periods, resulting in skeletally-discernible consequences and an average age at death of only 10.6 years (Van Gerven et al. 1995).[147] Interpreted within the context of ethnographic evidence that describes the presence of a semi-itinerant, landless "underclass" in present-day Nubia, these data supported the hypothesis that such an underclass may have existed in Early Christian times as well, and specifically, that the

[147] Compared to an average of 18.8 years for the R community.

S community comprised the impoverished underclass that provided occasional labor for landowning members of the R community (Adams et al. 1999; Adams and Adams 2006).

Ultimately, the hypothesis of a single population at Kulubnarti, intimately related by blood and culture, but divided into two socially-distinct communities, has been supported widely by anthropological research that used cranial morphology as a low-resolution proxy for genetics. However, because morphological expression is subject to influence by environmental and stochastic factors (Relethford 2004; Konigsberg and Ousley 2009), and analysis of morphological phenotype is one step removed from direct analysis of genetic variation (Nystrom 2006; Stojanowski and Shilliaci 2006; von Cramon-Taubadel and Weaver 2009), the use of morphology as a proxy may be problematic. In contrast, the use of ancient DNA (aDNA) has the potential to provide a high resolution and time-specific snapshot of a previously-unknown landscape of population relationships and genetic variation.

Taking a genomic perspective, this dissertation analyzed genome-wide Single Nucleotide Polymorphisms (SNPs) to examine the genetic relationship between, biogeographic affinities and origins of, and patterns of genetic variation within the two Kulubnarti communities. A single question formed the keystone of this dissertation: **based on direct genetic evidence, were the socially-distinct S and R communities also genetically-distinct subpopulations?** While this was the primary question driving this dissertation, direct genomic analysis of the Kulubnarti Nubians also inspired more detailed questions about the genetic similarities and differences of the two communities: **with which present-day populations do the Kulubnarti Nubians share the greatest genetic affinity? Which ancestral components are identified in the Kulubnarti**

Nubians and how do they fit within the present-day landscape of human genetic diversity? Did the two communities have similar matrilineal and patrilineal origins? Were any differences in genetic composition community-based? Finally, the ability to generate high-resolution aDNA data evoked a question that attempted to connect the Nubian past and present: **does the ethnically-based social structure that characterizes Nubian society today reflect a similar social division at Early Christian Kulubnarti?**

This chapter discusses the meaning of the genetic data presented in Chapters 7 and 8. Assigning meaning to these data involved interpretation from within a biocultural framework. Longstanding hypotheses about the Kulubnarti Nubians could then be retested and broader conclusions regarding the population of Kulubnarti from a genomic perspective could be explored.

Additionally, because the data presented in this dissertation represent only the first step toward a more comprehensive understanding of Kulubnarti from a genomic perspective, this chapter provides an overview of potential pathways for future research. This chapter concludes with final thoughts on the impact of this dissertation and, more broadly, on how the direct analysis of human genetic variation using ancient DNA and the interpretation of patterns of genetic variation within the context of broader sociocultural dynamics provides an invaluable tool for resolving ancient population histories.

9.2 A Single Genetic Population

To assess whether population substructure at Kulubnarti separated the socially-distinct S and R communities at Kulubnarti into genetically-distinct subpopulations, Aim

1 of this dissertation was to explore the genetic relationship between the S and R communities by looking for evidence of community-based genetic population substructure through patterns of clustering and estimation of genetic distance.

Based on existing archaeological and bioarchaeological evidence, Expectation 1 was that no community-based genetic population substructure would be detected at Kulubnarti, consistent with the interpretation that the S and R communities were a single genetic population. This expectation was supported by the data generated in this dissertation.

Finding #1: From a genomic perspective, no community-based genetic population substructure was detected at Kulubnarti, suggesting that the S and R communities were a single genetic population.

Finding #1 was supported by the results of Principal Component Analysis (PCA) applied to genome-wide nuclear SNP data from the Kulubnarti Nubians to look for potential community-based genetic population substructure at Kulubnarti through a visual assessment of patterns of clustering in PCA space. PCA results, illustrated in Figures 7.1 and 7.2 (Chapter 7), showed that the members of both Kulubnarti communities were projected into overlapping PCA space as a single discrete cluster within that space, and further, that no community-based patterns of clustering were observed. PCA therefore provided no evidence of community-based population substructure at Kulubnarti, suggesting that the individuals from the S and R communities shared enough genetic similarity to be interpreted as a single genetic population.

Finding #1 was also supported by an estimation of pairwise F_{ST}, a statistical measure of genetic distance between two populations or subpopulations (Weir and

Cockerham 1984). F_{ST} estimates suggested very low and statistically non-significant genetic distance between the S and R communities. These results were consistent with an interpretation of no community-based genetic population substructure at Kulubnarti, and further, with the interpretation that the two Kulubnarti communities comprised a single genetic population.

Finding #1, based primarily on the assessment of PCA plots and the F_{ST} estimate between the Kulubnarti communities, was also supported by the results of ADMIXTURE clustering analysis, commonly used to explore genetic similarity in terms of components of ancestry. As shown in Figure 8.4 (Chapter 8), the Kulubnarti Nubians from both communities were determined to have the same two dominant signals of ancestry in similar proportions, suggesting no community-based differences in ancestry components.[148]

Ultimately, the results from these analyses were consistent with the following conclusion:

Broader conclusion #1: Genetic data suggest that the community structure detected at Kulubnarti was a social phenomenon without a biological basis.

Because no natural (geographic) boundaries isolated the S and R communities from each other, inter-community gene flow at Kulubnarti would not have been impeded by physical barriers. However, anthropological research has shown that, in addition to natural barriers, social factors can modulate evolutionary forces including gene flow (Laland et al. 2010). Often in human populations, socioculturally-influenced systems of mating and locality influence gene flow in ways that generate and structure patterns of

[148] More details regarding the results of ADMIXTURE analysis are provided in the following section.

genetic distance between subpopulations within a population (Bamshad et al. 1998; Oota et al. 2001). Therefore, the finding of no community-based genetic population substructure at Kulubnarti had a notable implication: it suggested that despite the social divisions present at Kulubnarti, members of each community appeared genetically similar enough to be considered a single population in this analysis.

The data generated by this dissertation suggested that even though the S community may have been positioned as the Nubian "underclass," (Adams and Adams 2006: 11), a "floating population of landless Nubians who provided casual labor at the time of the date harvest or whenever there were extra labor needs" (Adams et al. 1999: 49) and who were "looked down on as inferiors by their more prosperous fellows, and were not allowed to live among them…" (Adams et al. 1999: 49), their social position in relation to the more prosperous R community did not result in the generation of genetically-discernible population substructure at Kulubnarti.

The prevailing interpretation of Kulubnarti as home to a single population in terms of blood and culture that was divided along social lines into two distinct communities was supported by the results of this dissertation's analysis. In addition to supporting this existing interpretation, the analysis of ancient DNA further elucidated the nature and extent of inter-community relations at Kulubnarti. Though many aspects of the relationship between the S and R communities may never be fully resolved, these two communities can be assumed to have maintained an intimate biological connection reflected by a lack of community-based genetic population substructure.

9.3 Biogeographic Genetic Affinity to Present-Day Northeastern Africans

In addition to enabling the analysis of the genetic relationship between the S and R communities, the direct analysis of genome-wide SNP data permitted the investigation of several deeper research questions surrounding the genetic composition of the Kulubnarti Nubians. Consequently, Aim 2 of this dissertation was to characterize the genetic composition of individuals from both communities and reveal any community-based differences by exploring biogeographic affinities and components of ancestry and determining the mtDNA and Y chromosome haplogroup profiles of each community.

Consistent with an expectation of no community-based genetic population substructure at Kulubnarti (Expectation 1), Expectation 2 was that no community-based differences in biogeographic affinities, ancestry components, or mtDNA or Y chromosome haplogroups would be detected between the S and R communities. Because Expectation 2 required the analysis of autosomal SNP data as well as SNPs from the uniparental markers of mitochondrial DNA (mtDNA) and the Y chromosome, it was broken down further into three parts, Expectations 2a-2c, that were then examined independently to provide support for or evidence against Expectation 2.

Expectation 2a was that both communities from Kulubnarti would show more genetic affinity to present-day genotyped populations from Northeastern Africa than any other world region and that they will be estimated to have components of ancestry associated with present-day populations from Northeastern Africa. Expectation 2a was supported by the data generated in this dissertation.

Finding #2a: Kulubnarti Nubians from both communities showed greater genetic affinity to present-day genotyped populations from Northeastern Africa than to populations from any other world region and had components of ancestry associated with the same present-day populations.

Finding #2a was supported by the application of PCA to genome-wide nuclear SNP data from the Kulubnarti Nubians to provide an exploratory approximation of biogeographic genetic affinities to present-day genotyped populations from around the world, though the results were somewhat ambiguous.[149] As illustrated in Figures 8.1 and 8.2 (Chapter 8), the PCA results suggested that the Kulubnarti Nubians showed genetic affinity to present-day genotyped populations from Northeastern Africa (i.e., the Horn of Africa) as well as to present-day genotyped populations from across Northern Africa. Specifically, the Kulubnarti Nubians were shown in Figure 8.2 to be projected into an intermediate position in PCA space in relation to present-day populations from Northern Africa and Northeastern Africa, suggesting that they shared more affinity with present-day Oromo Ethiopians, Ethiopian Jews, and Somalis from the Horn of Africa, and present-day Moroccans, Tunisians, Algerians, and Libyans from across Northern Africa, than with any other genotyped populations.

In contrast, a more robust and nuanced clustering analysis using ADMIXTURE provided unambiguous support for Finding #2a. Here, ADMIXTURE was used to assess the ancestral components of the Kulubnarti Nubians alongside a reference panel of present-day genotyped populations from around the world to place the Nubian samples within the landscape of modern human genetic diversity. As illustrated in Figures 8.3 and 8.4 (Chapter 8), two dominant ancestral components were detected in the Kulubnarti Nubians from both communities. The most dominant component (the "lavender component") was maximized in some present-day sub-Saharan Eastern African

[149] It must be acknowledged that the resolution of this analysis was limited by a relatively sparse reference dataset of present-day populations from Africa and a lack of other ancient African individuals for direct comparison.

populations; the second most dominant component (the "coral component") was maximized in many present-day Northern African populations and some Near Eastern populations. Most Kulubnarti Nubians also exhibited one or both of two very minor components maximized in many present-day Caucasian or European populations (the "red component" and "yellow component," respectively).

ADMIXTURE analysis thereby clarified the ambiguity of the exploratory PCAs. The Kulubnarti Nubians were not found to share the same ancestry components in similar proportions to populations from across Northern Africa, but were instead clearly determined to be most similar in terms of ancestry components to present-day genotyped populations from Northeastern Africa, including Ethiopian Jews, Oromo Ethiopians, and Somalis. These present-day Northeastern African populations showed the same dominant presence of the lavender and coral components in approximately the same proportions of the Kulubnarti Nubians, only differing in the contribution from very minor components. This result was consistent with the following conclusion:

Broader conclusion #2a: Genetic data suggest that the Kulubnarti Nubians were likely to have been an admixed population.

The Nile Corridor has longstanding geographic and historical links to both sub-Saharan Africa and Eurasia. These links have brought peoples of varying ancestries into contact with genetically-discernible consequences. As present-day populations from Northeastern Africa consistently show admixture with Eurasian sources (Hollfelder et al. 2017), the detected affinity between the Kulubnarti Nubians and present-day genotyped Northeastern African populations suggested that the Kulubnarti Nubians may have been an admixed population as well.

Specifically, genetic analyses of present-day Northeastern African populations have consistently identified both African-specific and non-African components in Ethiopians (Lovell et al. 2005; Campbell and Tishkoff 2010; Pagani et al. 2012), as well as Nubians (Hollfelder et al. 2017).[150] As a result of millennia of demographic movement, it is unsurprising that the majority of present-day Northeastern African populations are admixed, containing varying amounts of sub-Saharan African as well as Northern African and non-African ancestral components (Dobon et al. 2015).

In particular, the Oromo Ethiopians have been detected to have been extensively impacted by Islamic expansion events, with the present-day population demonstrating a mixture of non-African as well as African-specific genetic components (Lovell et al. 2005; Pagani et al. 2012). Ethiopian Jews, Jewish individuals presently living in Israel but who were born (or whose parents or grandparents were born) in Ethiopia (Thomas et al. 2002), are also assessed as having a mixture of "typically African and typically Caucasian haplotypes" (Ritte et al. 1993: 435).[151] The identification of African and non-African genetic components makes Ethiopian Jews genetically distinct from other African and Near Eastern Jewish populations (Ritte et al. 1993). Instead, there is remarkable genetic similarity between Ethiopian Jews and Ethiopian non-Jews (Hammer et al. 2000; Behar et al. 2010).

While present-day Nubian individuals were not included in the reference dataset used in this dissertation, recent studies have detected strong signals of admixture from Eurasian populations into present-day Nubian populations, leading to the conclusion that

[150] As the publication of the Nubian dataset occurred very recently, these individuals were not included in this dissertation's reference dataset. However, these samples will be included in future work and will increase the resolution of analyses and improve the interpretation of results.
[151] All Ethiopian Jews included in this dataset were themselves migrants from Ethiopia.

the Nubians presently comprise "a group with substantial genetic material relating to Nilotes [autochthonous populations living along the Nile] that later have received much gene-flow from Eurasians (likely Middle [Near] Eastern) and from East Africans" (Hollfelder et al. 2017: 7).

The results suggesting biogeographic affinity and sharing of ancestry components in similar proportions between the Kulubnarti Nubians and present-day admixed populations from Ethiopia and Somalia therefore supported an interpretation of the Kulubnarti Nubians as an admixed population, with non-local Northern African or Eurasian influences detected in their genomes despite considerable geographic isolation in the *Batn el Hajar*. This interpretation was further supported by a recent analysis, which used genetic data from present-day Nubians to estimate admixture dates for several Nubian populations. This analysis found that admixture dates likely pre-dated the Arab expansion and suggested that migrations and admixture between Eurasian populations and autochthonous Nubians predate Islamic conquest (Hollfelder et al. 2017).

While further research using ancient DNA is needed to more clearly elucidate precisely when, and possibly why, foreign genetic influences were introduced into the Kulubnarti population, the results in this dissertation inspire a reconsideration of the extent of genetic impact from peoples moving along the Nile, which appears to have influenced even the smallest and most isolated of Early Christian Period communities.

9.4 Community-Based Variation in mtDNA and Y Chromosome Haplogroup Profiles

In addition to exploring genetic composition in the Kulubnarti Nubians using data from across the autosomal genome, Aim 2 also included the analysis of uniparental genetic markers, including mtDNA and the Y chromosome. Though they represent only the genetic contribution of a single lineage, the analysis of mtDNA and the Y chromosome can be used to infer sex-biased patterns of migration and systems of matrilocal and patrilocal residence.[152]

Unfortunately, little information is available on the mobility of ancient Nubian males and females in regard to marriage patterns (Edwards 2004; Vollner 2016). Ethnohistoric studies have suggested that Nubia, like much of pre-Islamic Africa, may have been matrilineal and matrilocal (Crowfoot 1927; Awad 1971; Adams 1977; Salih 2004; Ruffini 2012; Fleuhr-Lobban and Lobban 2016), though there is contrasting evidence suggesting that succession passed from father to eldest son in early Medieval Nubia in "the usual Christian fashion" (Adams 1977: 463), suggesting a patrilineal system of inheritance and plausibly, patrilocality. This would be in line with a possible erosion of matrilineal kinship systems during Christian times (Lobban 2003). While the Arabization and Islamization of Nubia following the Christian period resulted in the development of a patrilineal system of succession (Awad 1971), no evident rules of succession or locality existed during Christian times (Edwards 2004).

Without clear expectations for sex-based succession or locality suggesting otherwise, Expectation 2b was that the majority of matrilineal lineages of members of both Kulubnarti communities would be primarily of Northeastern African origin and

[152] In matrilocal societies, females remain in their birthplace, and males move to their mate's residence after marriage, while in patrilocal societies, males remain in their birthplace, and females move to their mate's residence after marriage. These patterns of movement produce specific patterns of genetic variance.

distribution and that there would be no community-based differences in mtDNA haplogroup profile. Likewise, Expectation 2c was that the majority of patrilineal lineages of male members of both Kulubnarti communities would be primarily of Northeastern African origin and distribution and that there would be no community-based differences in Y chromosome haplogroup profile. Neither Expectation 2b nor Expectation 2c were supported by the uniparental genetic data generated in this dissertation.

Finding #2b: African-specific mtDNA lineages and non-African-specific Eurasian-associated mtDNA lineages were represented at similar frequencies at Kulubnarti when examined at a population level. At a community level, the S and R communities were characterized by different matrilineal genetic signatures; specifically, an African-specific matrilineal genetic signature was overrepresented in the S community, while a Eurasian-associated matrilineal genetic signature was overrepresented in the R community.

Finding #2b was supported by the assignment of mtDNA haplogroup for each individual at Kulubnarti. At a population level, African-specific mtDNA lineages and non-African-specific Eurasian-associated mtDNA lineages were found to be represented at similar frequencies. While inconsistent with Expectation 2b, this result was unsurprising due to thousands of years of potential admixture between autochthonous Nubians and non-local females. These mtDNA data suggested that female-biased migration into Nubia from Eurasia occurred prior to or during the Early Christian Period and influenced the mtDNA haplogroup profile at Kulubnarti.

When mtDNA haplogroup data were examined at a community level, an African-specific matrilineal genetic signature was found to be overrepresented in the S community, while a Eurasian-associated matrilineal genetic signature was found to be overrepresented in the R community, illustrated in Figure 8.7 (Chapter 8). These genetic

data best supported a scenario where female-mediated gene flow from Eurasia was reflected to a greater extent in members of the R community than members of the S community. When taken alone, results of mtDNA haplogroup assignment suggested that matrilocality may have been practiced in the S community, but that exogamous marriages, or possibly patrilocality, may have been practiced in the R community.

The detection of community-based differences in matrilineal genetic signatures was an unexpected finding that has not been observed in any other ancient Nubian population. To this regard, the analysis of Y chromosome haplogroups became even more essential for investigating potential differences in patrilineal genetic signatures. This led to the next finding.

Finding #2c: African-specific Y chromosome lineages and non-African-specific Eurasian-associated Y chromosome lineages were represented at similar frequencies at Kulubnarti when examined at a population level. At a community level, the S and R communities were characterized by different patrilineal genetic signatures; specifically, a Eurasian-associated patrilineal genetic signature was overrepresented in the S community, while an African-specific patrilineal genetic signature was overrepresented in the R community.

Finding #2c was supported by the assignment of Y chromosome haplogroup for each male individual at Kulubnarti. At a population level, African-specific and non-African-specific Eurasian-associated Y chromosome lineages were represented at a frequency of 50% each. This result was inconsistent with Expectation 2c and suggested that extensive male-mediated gene flow from Eurasia occurred prior to or during the Early Christian Period and influenced patterns of Y chromosomal genetic variation at Kulubnarti. While the occurrence of male-mediated gene flow in the millennia preceding the Christian Period as well as in the Christian Period itself is well-documented, it is

notable that a village as small and isolated as Kulubnarti would demonstrate such a high representation of non-local Y haplogroups.

As with mtDNA, Y haplogroups were then examined at a community level. When divided by community, the data indicated that a Eurasian-associated patrilineal genetic signature was found to be overrepresented in the S community, while an African-specific patrilineal genetic signature was found to be overrepresented in the R community, illustrated in Figure 8.8 (Chapter 8). These genetic data best supported a scenario where male-mediated gene flow from Eurasia was reflected to a greater extent in members of the S community than members of the R community. When taken alone, results of Y chromosome haplogroup analysis suggested that patrilocality, or possibly patrilineal primogeniture (succession of land ownership by the eldest son, a possibility suggested by Dr. Dennis Van Gerven) may have been practiced in the R community, but not in the landless S community. While this interpretation is speculative, it would be consistent with the hypothesis that the S community consisted of semi-itinerant laborers who did not own land that would be transferred through intergenerational inheritance. As with mtDNA, the detection of contrasting community-based residence patterns using Y chromosome haplogroups at Kulubnarti was also unexpected and has not been observed elsewhere in ancient Nubia.

When examined together, a contrasting pattern of geographically-associated matrilineal and patrilineal genetic signatures was detected between the Kulubnarti communities. Specifically, genetic signatures of Eurasian-associated patrilineal ancestry were overrepresented in the S community, while genetic signatures of Eurasian-associated matrilineal ancestry were overrepresented in the R community; the opposite

pattern is seen for genetic signatures of African-specific ancestries, with signatures of African-specific matrilineal ancestry overrepresented in the S community, and signatures of African-specific patrilineal ancestry overrepresented in the R community. This pattern is illustrated in Table 8.3 (Chapter 8), and led to a wholly unexpected broader conclusion:

Broader Conclusion #2b/c: From a genomic perspective, genetic signatures of sex-specific gene flow were reflected within the Kulubnarti population in a community-based way.

Studies of mtDNA and Y chromosome haplogroups have clearly demonstrated the genetic influence of sex-biased migration into past and present populations. However, because studies usually focus on haplogroup diversity for either mtDNA (e.g., Krings et al. 1999; Kivisild et al. 2004; Mikkelsen et al. 2012) or the Y chromosome (e.g., Semino et al. 2002; Luis et al. 2004; Hassan et al. 2008), the opportunity to integrate the results from both uniparental markers for a single population is precluded. Those that do investigate both uniparental markers (e.g., Passarino et al. 1998; Destro-Bisol et al. 2004; Wood et al. 2005) often use many diverse populations instead of a single socially-stratified population. This dissertation had the unique opportunity to examine the signatures of both uniparental genetic markers in an ancient population that comprised two socially-distinct communities.

Expectation 2 of this dissertation, which predicted that no community-based differences in biogeographic genetic affinities, ancestry components, or mtDNA or Y chromosome haplogroups would be detected between the S and R communities, was supported by the autosomal SNP data presented as Finding #2a. Taken alone, autosomal SNP data suggested that the Kulubnarti Nubians from either community were

genetically-indistinguishable in terms of genetic affinities and components of ancestry. However, Expectation 2 was not supported by the analysis of either uniparental marker, presented as Findings #2b and #2c. In fact, the analysis of these uniparental markers produced the most unanticipated result of this dissertation: that genetic signatures of sex-specific gene flow were reflected in a community-based way at Kulubnarti. Though these data alone are unable to reveal the reasons behind this unexpected pattern of mtDNA and Y chromosome haplogroups in the Kulubnarti communities, they can be taken to suggest a practice of matrilocality for the S community and patrilocality, perhaps associated with paternal primogeniture, for the R community.

Unfortunately, while genetic data support the hypothesis that both matrilocal and patrilocal systems co-existed within the socially-stratified Kulubnarti population, this hypothesis finds little support in historical evidence, primarily because little evidence detailing patterns of locality in Christian Nubia is available (per Dr. David N. Edwards, pers. comm.) and archaeological evidence neither supports nor refutes this genetic finding. Interestingly, recent research investigating possible patterns of residence in three Nubian populations (Kulubnarti being one of them) using craniometric variance (where greater variance was akin to greater mobility), found the greatest amount of variance to be present in both males and females of Kulubnarti compared to the other Nubian populations (Vollner 2016). While the interpretation of Vollner's data was unable to decipher the possibility of matrilocal or patrilocal systems at Kulubnarti using cranial variance, the finding of higher craniometric variance in both males and females from Kulubnarti compared to any other Nubian society is consistent with the detection of the presence of both local and non-local mtDNA and Y haplogroups in the Kulubnarti

population. However, as Vollner did not divide her sample from Kulubnarti along community lines, her analysis was unable to provide additional evidence that could support or contest the genetic findings of this dissertation.

9.5 A Hypothetical Scenario

In an attempt to connect the past with the present in Nubia, the final question asked in this dissertation was: does the ethnically-based social structure that characterizes Nubian society today reflect a similar social division at Early Christian Kulubnarti?

While this question may never be fully answered, the genetic data generated in this dissertation enable the envisioning of a hypothetical situation where females of Eurasian ancestry migrating into Kulubnarti or along the Nile through the *Batn el Hajar* were more likely to introduce their genes into the R community, potentially admixing with males of local ancestry who owned small, valuable plots of arable land based on a system of patrilineal inheritance associated with the R community. If males of Eurasian ancestry migrating into Kulubnarti or along the Nile through the *Batn el Hajar* did not have the same opportunity to inherit or own plots of land, they may have been more likely to become (either temporarily or permanently) members of the semi-itinerant and landless S community.

Furthermore, if marriage and mating with females of Eurasian ancestry was preferred in Kulubnarti (an idea which is based on Adams's (1977) description of the elite status of individuals with Egyptian or Near Eastern ancestry in Nubia today), locally-born females may not have been provided the same opportunity to marry and mate with the relatively more advantaged landholding males of the R community, and

may have therefore more frequently found their partners among the more impoverished members of the S community.

It should be recognized that this hypothetical situation may or may not be an accurate reflection of interpersonal interactions occurring at Early Christian Kulubnarti; however, this scenario may be an ancient correlate of the social interactions that are presently observed in Nubia, where society is currently loosely caste-based along ethnic lines, with a small elite group of Egyptian, Syrian, and Lebanese descendants comprising the top castes, and descendants of sub-Saharan African slaves comprising the lowest castes (Adams 1977).

The discussion of the findings of this dissertation concludes with an intriguing possibility: the Kulubnarti Nubians, evaluated to be a single genetic population, also may have been an admixed population with genomic contributions from Eastern Africa as well as Northern Africa and Western Eurasia. A fascinating analysis would be to determine whether the Eastern African genomic component was introduced into the S community gene pool by females and to the R community gene pool by males, and whether the Northern African or Eurasian genomic component was introduced into the S community gene pool by males and to the R community gene pool by females. Future research will undoubtedly further elucidate the nuances of population relationships at Kulubnarti from a genomic perspective.

9.6 Future Directions

The genetic findings presented in this dissertation speak as powerfully to the need for additional research as they do to the insights gained thus far. This section highlights

only some of the most pressing analyses that may build upon and clarify the findings presented here.

This dissertation's study of Kulubnarti from a genomic perspective is only a first step toward a thorough understanding of the dynamic and complex population history of the Kulubnarti Nubians using ancient DNA. A particularly valuable future analysis would be the dedicated study of the sex-specifically-inherited X chromosome and comparison of genetic differentiation between the X chromosome and autosomes. As has been demonstrated in recent studies using ancient DNA (e.g., Skoglund et al. 2016; Goldberg et al. 2017), the X chromosome can be used as a tool for exploring variation in the proportion of males and females contributing to admixture processes in a past population. As such, the analysis of the X chromosome could be used to more clearly illustrate the impact of sex-biased processes in each community at Kulubnarti.

In a broader sense, it is imperative for future research to continue to expand both the temporal and spatial reach of studies exploring ancient Nubian populations from a genomic perspective while maintaining a dedication to interpreting patterns of genetic variation within the context of social dynamics. Expanding this reach is particularly important, because in addition to providing a more detailed understanding of biological interactions and population history at additional Nubian sites, data from these studies can eventually be combined to generate a more comprehensive picture of genetic diversity across space and time in ancient Nubia. A continued adherence to the biocultural approach will ensure that genetic diversity is contextualized in terms of the unique sociocultural characteristics of each population under study.

A particularly powerful starting point would be to conduct genome-wide analyses of Nubian populations inhabiting Nilotic regions both north and south of Kulubnarti; for example, the generation of a comparative dataset using the human remains excavated from three Christian Period cemeteries on Mis Island in the Fourth Cataract (Soler 2012) would provide a powerful complement to the present work. These results would further develop an understanding of population interactions between sub-Saharan Africa and North Africa and Eurasia, as well as within Lower Nubia, the *Batn el Hajar*, and Upper Nubia. The continued growth of genetic analyses using Egyptian populations (e.g., Schuenemann et al. 2017) will also aid the further elucidation of ancient patterns of genetic diversity between Egypt and Nubia.

Finally, the continued development of more complete and dense reference datasets that incorporate the large amount of regional- and local-scale genetic diversity in Northeastern Africa will greatly benefit future research on the population history of the ancient Nubians and other ancient African populations. Ancient DNA studies are limited in resolution by the reference datasets they use to represent present-day human genetic variation. As Africa is the most genetically diverse region in the world (Cann et al. 1987), and Northeastern Africa in particular has been shaped by a myriad of historic and ongoing demographic migrations along the Nile, across the Sahara, and via the Red Sea, extensive sampling of a greater number of present-day populations in concentrated regions will revolutionize the resolution of the conclusions that can be made using ancient DNA.

At a time when aDNA research in Africa is beginning to offer unprecedented insight into past populations, the time is ripe to continue focusing specifically on

revealing the dynamic population histories of ancient Nubia, and more broadly on interpreting patterns of genetic variation within the context of sociocultural dynamics.

9.7 Concluding Thoughts

This dissertation was inspired by the suggestion that ancient DNA can contribute towards a more accurate and refined understanding of ancient population history (Schuenemann et al. 2017) and aimed to demonstrate how the interpretation of patterns of genetic variation are strengthened when conducted with an appreciation of sociocultural dynamics. Recognizing that ancient DNA can be used as a tool to supplement archaeological and historical data that may not be able to fully evaluate the biological consequences of social differences in the past, this dissertation represents a step forward from the studies of the Kulubnarti Nubians that relied upon morphology as a proxy for genetic variation. Drawing upon the hypotheses proposed by these analyses of morphology, this dissertation explored the genetic relationship of the Kulubnarti communities as well as patterns of genetic variation and their relationship to the social stratification at Kulubnarti from the ultimate source of the human genome for the first time.

While it presented a high-resolution genomic analysis of one small, impoverished population living in a single, isolated location at a brief, but significant, point in time, this dissertation more broadly demonstrated the power of analyzing ancient DNA in the context of known sociocultural dynamics. It revealed a previously unknown pattern of genetic variation in an ancient Nubian population surviving in the most desolate of environments and demonstrated how the consideration of sociocultural factors facilitates

a clearer understanding of how the social and biological realms may have overlapped in a past population in unexpected ways. This dissertation thus serves as a model that demonstrates how the integration of anthropology and genomics can contribute to a more refined understanding of ancient population history and illustrates how new types of analyses have the potential to inform, expand, and transform our understanding of populations in the past.

REFERENCES

Abu-Amero KK, González AM, Larruga JM, Bosley TM, and Cabrera VM. 2007. Eurasian and African mitochondrial DNA influences in the Saudi Arabian population. BMC Evolutionary Biology 7: 32.

Abu-Amero KK, Hellani A, González AM, Larruga JM, Cabrera VM, and Underhill PA. 2009. Saudi Arabian Y-Chromosome diversity and its relationship with nearby regions. BMC Genetics 10(1): 59.

Abu-Amero KK, Larruga JM, Cabrera VM, and González AM. 2008. Mitochondrial DNA structure in the Arabian Peninsula. BMC Evolutionary Biology 8(1): 45.

Abu-Zeid M, and El-Shibini F. 1997. Egypt's high Aswan dam. International Journal of Water Resources Development 13(2): 209-218.

Achilli A, Rengo C, Magri C, Battaglia V, Olivieri A, Scozzari R, Cruciani F, Zevani M, Briem E, Carelli V, Moral P, Dugoujon J-M, Roostalu U, Loogväli E-L, Kivisild T, Bandelt H-J, Richards M, Villems R, Santachiara-Benerecetti AS, Semino O, and Torroni A. 2004. The molecular dissection of mtDNA haplogroup H confirms that the Franco-Cantabrian glacial refuge was a major source for the European gene pool. The American Journal of Human Genetics 75(5): 910-918.

Acsádi G, and Nemeskéri J. 1970. History of human life span and mortality. Budapest, Hungary: Akadémiai Kiadó.

Adams WY. 1965. Post-Pharaonic Nubia in the Light of Archaeology II. The Journal of Egyptian Archaeology 51: 160-178.

Adams WY. 1966. The Nubian Campaign: Retrospect and Prospect. In: Mélanges offerts à K. Michalowski. Warsaw. p. 13-30.

Adams WY. 1967. Continuity and Change in Nubian Cultural History. Sudan Notes and Records 48: 1-32.

Adams WY. 1968. Invasion, diffusion, evolution? Antiquity 42(167): 194-215.

Adams WY. 1970. A re-appraisal of Nubian culture history. Orientalia 39(2): 269-277.

Adams WY. 1977. Nubia: Corridor to Africa. Princeton: Princeton University Press.

Adams WY. 1981. Paradigms in Sudan Archaeology. Africa Today 28(2): 15-24.

Adams WY. 1994. Kulubnarti I: The Architectural Remains. Lexington: Program for Cultural Resource Assessment, University of Kentucky.

Adams WY. 1998. The misappropriation of Nubia. Sudan Studies 21: 1-9.

Adams WY, and Adams NK. 1998. Kulubnarti II: The Artifactual Remains. London: Sudan Archaeological Research Society.

Adams WY, and Adams NK. 2006. The Kulubnarti Underclass. In: Gratien B (editor). Cahiers de Recherches de l'Institut de Papyrologie et d'Égyptologie de Lille. Lille. p. 11-16.

Adams WY, Van Gerven DP, and Guise D. 1999. Kulubnarti III: The Cemeteries. Oxford: Archaeopress.

Adams WY, Van Gerven DP, and Levy RS. 1978. The retreat from migrationism. Annual Review of Anthropology 7(1): 483-532.

Alberts B. 2003. DNA replication and recombination. Nature 421(6921): 431-435.

Alexander DH, Novembre J, and Lange K. 2009. Fast model-based estimation of ancestry in unrelated individuals. Genome research 19(9): 1655-1664.

Allentoft ME, Collins M, Harker D, Haile J, Oskam CL, Hale ML, Campos PF, Samaniego JA, Gilbert MTP, Willerslev E, Zhang G, Scofield RP, Holdaway RN, and Bunce M. 2012. The half-life of DNA in bone: measuring decay kinetics in 158 dated fossils. Proceedings of the Royal Society B 279: 4724-4733.

Allentoft ME, Sikora M, Sjögren K-G, Rasmussen S, Rasmussen M, Stenderup J, Damgaard PB, Schroeder H, Ahlström T, Vinner L, Malaspinas A-S, Margaryan A, Higham T, Chivall D, Lynerrup N, Harvig L, Baron J, Della Casa P, Dabrowski P, Duffy PR, Ebel AV, Epimakhov A, Frei K, Furmanek M, Gralak T, Gromov A, Gronkiewicz S, Grupe G, Hajdu T, Jarysz R, Khartanovich V, Khokhlov A, Kiss V, Kolář J, Krisska A, Lasak I, Longhi C, McGlynn G, Merkevicius A, Merkyte I, Metspalu M, Mkrtchyan R, Moiseyev V, Paja L, Pálfi G, Pokutta D, Pospieszny L, Price TD, Saag L, Sablin M, Shishlina N, Smrčka V, Soenov VI, Szeverényi V, Tóth G, Trifanova SV, Varul L, Vicze M, Yepiskoposyan L, Zhitenev V, Orlando L, Sicheritz-Pontén T, Brunak S, Nielsen R, Kristiansen K, and Willerslev E. 2015. Population genomics of Bronze Age Eurasia. Nature 522(7555): 167-172.

Anderson S, Bankier AT, Barrell BG, de Bruijn MHL, Coulson AR, Drouin J, Eperon IC, Nierlich DP, Roe BA, Sanger F, Schreier PH, Smith AJH, Staden R, and Young IG. 1981. Sequence and organization of the human mitochondrial genome. Nature 290: 457-465.

Andrews RM, Kubacka I, Chinnery PF, Lightowlers RN, Turnbull DM, and Howell N. 1999. Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. Nature Genetics 23(2): 147-147.

Arkell AJ. 1961. A History of the Sudan From the Earliest Times to 1821. London: The Athlone Press.

Armelagos GJ. 1968. Paleopathology of three archeological populations from Sudanese Nubia. PhD Dissertation: University of Colorado.

Armelagos GJ. 1969. Disease in ancient Nubia. Science 163(3864): 255-259.

Armelagos GJ, Ewing GH, and Greene DL. 1968. Physical Anthropology and Man-Made Lakes. Anthropological Quarterly: 122-131.

Armelagos GJ, and Van Gerven DP. 2017. Life and Death on the Nile: A Biothenography of Three Ancient Nubian Communities. Gainesville: University Press of Florida.

Arredi B, Poloni ES, Paracchini S, Zerjal T, Fathallah DM, Makrelouf M, Pascali VL, Novelletto A, and Tyler-Smith C. 2004. A Predominantly Neolithic Origin for Y-Chromosomal DNA Variation in North Africa. The American Journal of Human Genetics 75(2): 338-345.

Ávila-Arcos MC, Cappellini E, Romero-Navarro JA, Wales N, Moreno-Mayar JV, Rasmussen M, Fordyce SL, Montiel R, Vielle-Calzada J-P, Willerslev E, and Gilbert MTP. 2011. Application and comparison of large-scale solution-based DNA capture-enrichment methods on ancient DNA. Scientific Reports 1: 74.

Ávila-Arcos, MC, Sandoval-Velasco M, Schroeder H, Carpenter ML, Malaspinas AS, Wales N, Peñaloza F, Bustamante CD, Gilbert MTP. 2015. Comparative performance of two whole-genome capture methodologies on ancient DNA Illumina libraries. Methods in Ecology and Evolution 6: 725-734.

Awad MH. 1971. The Evolution of Landownership in the Sudan. Middle East Journal 25(2): 212-228.

Baca M, Doan K, Sobczyk M, Stankovic A, and Węgleński P. 2012. Ancient DNA reveals kinship burial patterns of a pre-Columbian Andean community. BMC Genetics 13(1): 30.

Bachtrog D, and Charlesworth B. 2001. Towards a complete sequence of the human Y chromosome. Genome Biology 2(5): reviews1016.1.

Badro DA, Douaihy B, Haber M, Youhanna SC, Salloum A, Ghassibe-Sabbagh M, Johnsrud B, Khazen G, Matisoo-Smith E, Soria-Hernanz DF, Wells RS, Tyler-Smith C, Platt DE, Zalloua PA, and The Genographic Consortium. 2013. Y-chromosome and mtDNA genetics reveal significant contrasts in affinities of modern Middle Eastern populations with European and African populations. PLoS One 8(1): e54616.

Balanovsky O, Dibirova K, Dybo A, Mudrak O, Frolova S, Pocheshkhova E, Haber M, Platt D, Schurr T, Haak W, Kuznetsova M, Radzhabov M, Balaganskaya O, Romanov A, Zakharova T, Soria Hernanz DF, Zalloua P, Koshel S, Ruhlen M, Renfrew C, Wells RS, Tyler-Smith C, Balanovska E, and The Genographic Consortium. 2011. Parallel evolution of genes and languages in the Caucasus region. Molecular Biology and Evolution 28(10): 2905-2920.

Ballard JWO, and Whitlock MC. 2004. The incomplete natural history of mitochondria. Molecular Ecology 13(4): 729-744.

Bamshad MJ, Kivisild T, Watkins WS, Dixon ME, Ricker CE, Rao BB, Naidu JM, Prasad BVR, Reddy PG, Rasanayagam A, Papiha SS, Villems R, Redd AJ, Hammer MF, Nguyen SV, Carroll ML, Batzer MA, and Jorde LB. 2001. Genetic evidence on the origins of Indian caste populations. Genome Research 11(6): 994-1004.

Bamshad MJ, Watkins WS, Dixon ME, Jorde LB, Rao BB, Naidu JM, Prasad BVR, Rasanayagam A, and Hammer MF. 1998. Female gene flow stratifies Hindu castes. Nature 395(6703): 651-652.

Barbujani G, Ghirotto S, and Tassi F. 2013. Nine things to remember about human genome diversity. Tissue Antigens 82(3): 155-164.

Barbujani G, Magagni A, Minch E, and Cavalli-Sforza LL. 1997. An apportionment of human DNA diversity. Proceedings of the National Academy of Sciences 94(9): 4516-4519.

Basha WA, Lamb AL, Zaki ME, Kandeel WA, Fares NH, and Chamberlain AT. 2018. Dietary seasonal variations in the Medieval Nubian population of Kulubnarti as indicated by the stable isotope composition of hair. Journal of Archaeological Science: Reports 18: 161-168.

Batini C, Hallast P, Zadik D, Delser PM, Benazzo A, Ghirotto S, Arroyo-Pardo E, Cavalleri GL, de Knijff P, Dupuy BM, Eriksen HA, King TE, López de Munain A, López-Parra AM, Loutradis A, Milasin J, Novelletto A, Pamjav H, Sajantila A, Tolun A, Winney B, and Jobling MA. 2015. Large-scale recent expansion of European patrilineages shown by population resequencing. Nature Communications 6: 7152.

Batini C, Lopes J, Behar DM, Calafell F, Jorde LB, van der Veen L, Quintana-Murci L, Spedini G, Destro-Bisol G, and Comas D. 2011. Insights into the Demographic History of African Pygmies from Complete Mitochondrial Genomes. Molecular Biology and Evolution 28(2): 1099-1110.

Batrawi A. 1946. The Racial History of Egypt and Nubia: Part II. The Racial Relationships of the Ancient and Modern Populations of Egypt and Nubia. The

Journal of the Royal Anthropological Institute of Great Britain and Ireland 76(2): 131-156.

Batrawi AM. 1935. Report on the Human Remains: Mission Archeologique de Nubie, 1929-34. Government Press.

Behar DM, Hammer MF, Garrigan D, Villems R, Bonne-Tamir B, Richards M, Gurwitz D, Rosengarten D, Kaplan M, Pergola SD, Quintana-Murci L, and Skorecki K. 2004. MtDNA evidence for a genetic bottleneck in the early history of the Ashkenazi Jewish population. European Journal of Human Genetics 12: 355-364.

Behar DM, Metspalu E, Kivisild T, Achilli A, Hadid Y, Tzur S, Pereira L, Amorim A, Quintana-Murci L, Majamaa K, Hernstadt C, Howell N, Balanovsky O, Kutuev I, Pshenichnov A, Gurwitz D, Bonne-Tamir B, Torroni A, Villems R, and Skorecki K. 2006. The Matrilineal Ancestry of Ashkenazi Jewry: Portrait of a Recent Founder Event. American Journal of Human Genetics 78(3): 487-497.

Behar DM, van Oven M, Rosset S, Metspalu M, Loogväli E-L, Silva NM, Kivisild T, Torroni A, and Villems R. 2012. A "Copernican" Reassessment of the Human Mitochondrial DNA Tree from its Root. The American Journal of Human Genetics 90(4): 675-684.

Behar DM, Villems R, Soodyall H, Blue-Smith J, Pereira L, Metspalu E, Scozzari R, Makkan H, Tzur S, Comas D, Bertranpetit J, Quintana-Murci L, Tyler-Smith C, Wells RS, Rosset S, and The Genographic Consortium. 2008. The Dawn of Human Matrilineal Diversity. The American Journal of Human Genetics 82(5): 1130-1140.

Behar DM, Yunusbayev B, Metspalu M, Metspalu E, Rosset S, Parik J, Rootsi S, Chaubey G, Kutuev I, Yudkovsky G, Khusnutdinova EK, Balanovsky O, Semino O, Pereira L, Comas D, Gurwitz D, Bonne-Tamir B, Parfitt T, Hammer MF, Skorecki K, and Villems R. 2010. The genome-wide structure of the Jewish people. Nature 466(7303): 238-242.

Bekada A, Fregel R, Cabrera VM, Larruga JM, Pestano J, Benhamamouch S, and González AM. 2013. Introducing the Algerian Mitochondrial DNA and Y-Chromosome Profiles into the North African Landscape. PLoS One 8(2): e56775.

Bhatia G, Patterson N, Sankararaman S, and Price AL. 2013. Estimating and interpreting FST: The impact of rare variants. Genome Research 23(9): 1514-1521.

Billy G. 1977. Population changes in Egypt and Nubia. Journal of Human Evolution 6(8): 697-704.

Boessenkool S, Hanghøj K, Nistelberger HM, Der Sarkissian C, Gondek AT, Orlando L, Barrett GH, and Star B. 2016. Combining bleach and mild pre-digestion improves ancient DNA recovery from bones. Molecular Ecology Resources 17(4): 742-751.

Bollongino R, Tresset A, and Vigne J-D. 2008. Environment and excavation: Pre-lab impacts on ancient DNA analyses. Comptes Rendus Palevol 7(2): 91-98.

Bolnick DA, and Smith DG. 2007. Migration and social structure among the Hopewell: Evidence from ancient DNA. American Antiquity: 627-644.

Bouwman AS, Brown KA, Prag A, and Brown TA. 2008. Kinship between burials from Grave Circle B at Mycenae revealed by ancient DNA typing. Journal of Archaeological Science 35(9): 2580-2584.

Brace CL, Tracer DP, Yaroch LA, Robb J, Brandt K, and Nelson AR. 1993. Clines and clusters versus "race:" a test in ancient Egypt and the case of a death on the Nile. American Journal of Physical Anthropology 36(S17): 1-31.

Brandstätter A, Salas A, Niederstätter H, Gassner C, Carracedo A, and Parson W. 2006. Dissection of mitochondrial superhaplogroup H using coding region SNPs. Electrophoresis 27(13): 2541-2550.

Briggs AW, Good JM, Green RE, Krause J, Maricic T, Stenzel U, Lalueza-Fox C, Rudan P, Brajković D, Kućan Ž, Gušić I, Schmitz R, Doronichev VB, Golovanova LV, de la Rasilla M, Fortea J, Rosas A, and Pääbo S. 2009. Targeted retrieval and analysis of five Neandertal mtDNA genomes. Science 325(5938): 318-321.

Briggs AW, and Heyn P. 2012. Preparation of next-generation sequencing libraries from damaged DNA. In: Shapiro B, and Hofreiter M (editors). Ancient DNA: Methods and Protocols. New York: Humana Press. p.143-154.

Briggs AW, Stenzel U, Johnson PL, Green RE, Kelso J, Prüfer K, Meyer M, Krause J, Ronan MT, Lachmann M, and Pääbo S. 2007. Patterns of damage in genomic DNA sequences from a Neandertal. Proceedings of the National Academy of Sciences 104(37): 14616-14621.

Briggs AW, Stenzel U, Meyer M, Krause J, Kircher M, and Pääbo S. 2010. Removal of deaminated cytosines and detection of in vivo methylation in ancient DNA. Nucleic Acids Research 38(6): e87.

Bronner IF, Quail MA, Turner DJ, and Swerdlow H. 2009. Improved Protocols for Illumina Sequencing. Current Protocols in Human Genetics 62: 18.2.1-18.2.27.

Brooks S, and Suchey JM. 1990. Skeletal age determination based on the os pubis: A comparison of the Acsádi-Nemeskéri and Suchey-Brooks methods. Human Evolution 5(3): 227-238.

Brotherton P, Endicott P, Sanchez JJ, Beaumont M, Barnett R, Austin J, and Cooper A. 2007. Novel high-resolution characterization of ancient DNA reveals C > U-type base modification events as the sole cause of post mortem miscoding lesions. Nucleic Acids Research 35(17): 5717-5728.

Brotherton P, Haak W, Templeton J, Brandt G, Soubrier J, Adler CJ, Richards SM, Der Sarkissian C, Ganslmeier R, Friederich S, Dresely V, van Oven M, Kenyon R, Van der Hoek MB, Korlach J, Luong K, Ho SYW, Quintana-Murci L, Behar DM, Meller H, Alt KW, Cooper A, and The Genographic Consortium. 2013. Neolithic mitochondrial haplogroup H genomes and the genetic origins of Europeans. Nature Communications 4: 1764.

Broushaki F, Thomas MG, Link V, López S, van Dorp L, Kirsanow K, Hofmanová Z, Diekmann Y, Cassidy LM, Díez-del-Molino D, Kousathanas A, Sell C, Robson HK, Martiniano R, Blöcher J, Scheu A, Kreutzer S, Bollongino R, Bobo D, Davudi H, Munoz O, Currat M, Abdi K, Biglari F, Craig OE, Bradley DG, Shennan S, Veeramah KR, Mashkour M, Wegmann D, Hellenthal G, and Burger J. 2016. Early Neolithic genomes from the eastern Fertile Crescent. Science: aaf7943.

Brown WM, George M, and Wilson AC. 1979. Rapid evolution of animal mitochondrial DNA. Proceedings of the National Academy of Sciences 76(4): 1967-1971.

Buikstra JE, Frankenberg SR, and Konigsberg LW. 1990. Skeletal biological distance studies in American physical anthropology: Recent trends. American Journal of Physical Anthropology 82(1): 1-7.

Buikstra JE, and Ubelaker DH. 1994. Standards for data collection from human skeletal remains: Proceedings of a seminar at the Field Museum of Natural History. Arkansas Archaeology Research Series 44.

Burbano HA, Hodges E, Green RE, Briggs AW, Krause J, Meyer M, Good JM, Maricic T, Johnson PLF, Xuan Z, Rooks M, Bhattacharjee A, Brizuela L, Albert FW, de la Rasilla M, Fortea J, Rosas A, Lachmann M, Hannon GJ, and Pääbo S. 2010. Targeted investigation of the Neandertal genome by array-based sequence capture. Science 328(5979): 723-725.

Burnor DR, and Harris J. 1967. Racial continuity in lower Nubia: 12,000 BC to the present. Paper presented at the Proceedings of the Indiana Academy of Science.

Burrell LL, Maas MC, and Van Gerven DP. 1986. Patterns of long-bone fracture in two Nubian cementeries. Human Evolution 1(6): 495-506.

Buś MM, and Allen M. 2014. Collecting and Preserving biological Samples from Challenging Environments for DNA Analysis. Biopreservation and Biobanking 12(1): 17-22.

Cabrera VM, Abu-Amero KK, Larruga J, and González A. 2009. The Arabian peninsula: Gate for Human Migrations Out of Africa or Cul-de-Sac? A Mitochondrial DNA Phylogeographic Perspective. In: Petraglia M, and Rose J (editors). The Evolution of Human Populations in Arabia. New York: Springer.

Calcagno JM. 1986. Odontometrics and biological continuity in the Meroitic, X-Group, and Christian phases of Nubia. Current Anthropology 27(1): 66-69.

Campbell MC, and Tishkoff SA. 2010. The evolution of human genetic and phenotypic variation in Africa. Current Biology 20(4): R166-R173.

Cann RL, Stoneking M, and Wilson AC. 1987. Mitochondrial DNA and Human Evolution. Nature 325: 31-36.

Caratti S, Gino S, Torre C, and Robino C. 2009. Subtyping of Y-chromosomal haplogroup E-M78 (E1b1b1a) by SNP assay and its forensic application. International Journal of Legal Medicine 123(4): 357-360.

Carlson DS. 1976. Temporal variation in prehistoric Nubian crania. American Journal of Physical Anthropology 45(3): 467-484.

Carlson DS, Armelagos GJ, and Van Gerven DP. 1974. Factors influencing the etiology of cribra orbitalia in prehistoric Nubia. Journal of Human Evolution 3(5): 405-410.

Carlson DS, and Van Gerven DP. 1977. Masticatory function and post-Pleistocene evolution in Nubia. American Journal of Physical Anthropology 46(3): 495-506.

Carlson DS, and Van Gerven DP. 1979. Diffusion, biological determinism, and biocultural adaptation in the Nubian corridor. American Anthropologist 81(3): 561-580.

Carpenter ML, Buenrostro JD, Valdiosera C, Schroeder H, Allentoft ME, Sikora M, Rasmussen M, Gravel S, Guillén S, Nekhrizov G, Leshtakov K, Dimitrova D, Theodossiev N, Pettener D, Luiselli D, Sandoval K, Moreno-Estrada M, Li Y, Wang J, Gilber MTP, Willerslev E, Greenleaf WJ, and Bustamante CD. 2013. Pulling out the 1%: Whole-Genome Capture for the Targeted Enrichment of Ancient DNA Sequencing Libraries. The American Journal of Human Genetics 93(5): 852-864.

Carson EA. 2006. Maximum-likelihood variance components analysis of heritabilities of cranial nonmetric traits. Human Biology 78(4): 383-402.

Cassidy LM, Martiniano R, Murphy EM, Teasdale MD, Mallory J, Hartwell B, and Bradley DG. 2016. Neolithic and Bronze Age migration to Ireland and establishment of the insular Atlantic genome. Proceedings of the National Academy of Sciences 113(2): 368-373.

Cavalli-Sforza LL, Menozzi P, and Piazza A. 1994. The History and Geography of Human Genes. Princeton: Princeton University Press.

Cerezo M, Achilli A, Olivieri A, Perego UA, Gómez-Carballa A, Brisighelli F, Lancioni H, Woodward SR, López-Soto M, Carracedo Á, Capelli C, Torroni A, and Salas A. 2012. Reconstructing ancient mitochondrial DNA links between Africa and Europe. Genome Research 22(5): 821-826.

Chagnon NA, Neel JV, Weitkamp L, Gershowitz H, and Ayres M. 1970. The influence of cultural factors on the demography and pattern of gene flow from the Makiritare to the Yanomama Indians. American Journal of Physical Anthropology 32(3): 339-349.

Champlot S, Berthelot C, Pruvost M, Bennett EA, Grange T, and Geigl E-M. 2010. An efficient multistrategy DNA decontamination procedure of PCR reagents for hypersensitive PCR applications. PLoS One 5(9): e13042.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, and Lee JJ. 2015. Second-generation PLINK: rising to the challenge of larger and richer datasets. Gigascience 4(1): 7.

Chapman R. 2013. Death, Burial, and Social Representation In: Tarlow S, and Stutz LN (editors). The Oxford Handbook of the Archaeology of Death and Burial. Oxford: Oxford University Press.

Chen J, Sokal RR, and Ruhlen M. 1995. Worldwide analysis of genetic and linguistic relationships of human populations. Human Biology 67(4): 577-594.

Chen Y-S, Olckers A, Schurr TG, Kogelnik AM, Huoponen K, and Wallace DC. 2000. mtDNA variation in the South African Kung and Khwe—and their genetic relationships to other African populations. The American Journal of Human Genetics 66(4): 1362-1383.

Cherni L, Loueslati BY, Pereira L, Ennafaa H, Amorim A, and el Gaaied ABA. 2005. Female Gene Pools of Berber and Arab Neighboring Communities in Central Tunisia: Microstructure of mtDNA Variation in North Africa. Human Biology 77(1): 61-70.

Chiaroni J, Underhill PA, and Cavalli-Sforza LL. 2009. Y chromosome diversity, human expansion, drift, and cultural evolution. Proceedings of the National Academy of Sciences 106(48): 20174-20179.

Cinnioğlu C, King R, Kivisild T, Kalfoğlu E, Atasoy S, Cavalleri GL, Lillie AS, Roseman CC, Lin AA, Prince K, Oefner PJ, Shen P, Semino O, Cavalli-Sforza LL, and Underhill PA. 2004. Excavating Y-chromosome haplotype strata in Anatolia. Human genetics 114(2): 127-148.

Collins FS, Brooks LD, and Chakravarti A. 1998. A DNA polymorphism discovery resource for research on human genetic variation. Genome Research 8(12): 1229-1231.

Collins MJ, Nielsen-Marsh CM, Hiller J, Smith CI, Roberts JP, Prigodich RV, Wess TJ, Csapó TJ, Millard AR, and Turner-Walker G. 2002. The Survival of Organic Matter in Bone: A Review. Archaeometry 44(3): 383-394.

Cooper A, and Poinar HN. 2000. Ancient DNA: Do It Right or Not at All. Science 289(5482): 1139-1139.

Costa MD, Pereira JB, Pala M, Fernandes V, Olivieri A, Achilli A, Perego UA, Rychkov S, Naumova O, Hatina J, Woodward SR, Eng KK, Macaulay V, Carr M, Soares P, Pereira L, and Richards MB. 2013. A substantial prehistoric European ancestry amongst Ashkenazi maternal lineages. Nature Communications 4: 2543.

Coudray C, Olivieri A, Achilli A, Pala M, Melhaoui M, Cherkaoui M, El-Chennawi F, Kossmann M, Torroni A, and Dugoujon J-M. 2009. The complex and diversified mitochondrial gene pool of Berber populations. Annals of Human Genetics 73(2): 196-214.

Crowfoot JW. 1927. Christian Nubia. The Journal of Egyptian Archaeology 13(3/4): 141-150.

Cruciani F, La Fratta R, Santolamazza P, Sellitto D, Pascone R, Moral P, Watson E, Guida V, Colomb EB, Zaharova B, Lavinha J, Vona G, Aman R, Calì F, Akar N, Richards M, Torroni A, Novelletto A, and Scozzari R. 2004. Phylogeographic analysis of haplogroup E3b (E-M215) Y chromosomes reveals multiple migratory events within and out of Africa. The American Journal of Human Genetics 74(5): 1014-1022.

Cruciani F, La Fratta R, Trombetta B, Santolamazza P, Sellitto D, Colomb EB, Dugoujon J-M, Crivellaro F, Benincasa T, Pascone R, Moral P, Watson E, Melegh B, Barbujani G, Fuselli S, Vona G, Zagradisnik B, Assum G, Brdicka R, Kozlov AI, Efremov GD, Coppa A, Novelletto A, and Scozzari R. 2007. Tracing Past Human Male Movements in Northern/Eastern Africa and Western Eurasia: New Clues from Y-Chromosomal Haplogroups E-M78 and J-M12. Molecular Biology and Evolution 24(6): 1300-1311.

Cruciani F, Santolamazza P, Shen P, Macaulay V, Moral P, Olckers A, Modiano D, Holmes S, Destro-Bisol G, Coia V, Wallace DC, Oefner PJ, Torroni A, Cavalli-

Sforza LL, Scozzari R, and Underhill PA. 2002. A back migration from Asia to sub-Saharan Africa is supported by high-resolution analysis of human Y-chromosome haplotypes. The American Journal of Human Genetics 70(5): 1197-1214.

Cruciani F, Trombetta B, Massaia A, Destro-Bisol G, Sellitto D, and Scozzari R. 2011. A Revised Root for the Human Y Chromosomal Phylogenetic Tree: The Origin of Patrilineal Diversity in Africa. The American Journal of Human Genetics 88(6): 814-818.

Cunningham C, Scheuer L, and Black S. 2016. Developmental juvenile osteology (2nd edition). London: Academic Press.

da Silva MS. 2014. Phylogeography of mtDNA haplogroup L2. MSc Dissertation: Universidade do Porto.

Dabney J, Knapp M, Glocke I, Gansauge M-T, Weihmann A, Nickel B, Valdiosera C, García N, Pääbo S, Arsuaga J-L, and Meyer M. 2013. Complete mitochondrial genome sequence of a Middle Pleistocene cave bear reconstructed from ultrashort DNA fragments. Proceedings of the National Academy of Sciences 110(39): 15758-15763.

Dafalla H. 1969. LAND ECONOMY OF OLD HALFA. Sudan Notes and Records 50: 63-74.

Daley T, and Smith AD. 2013. Predicting the molecular complexity of sequencing libraries. Nature Methods 10(4): 325-327.

Damgaard PB, Margaryan A, Schroeder H, Orlando L, Willerslev E, and Allentoft ME. 2015. Improving access to endogenous DNA in ancient bones and teeth. Scientific Reports 5: 11184.

Davis K, and Moore WE. 1945. Some principles of stratification. American Sociological Review 10(2): 242-249.

DeAngelis MM, Wang DG, and Hawkins TL. 1995. Solid-phase reversible immobilization for the isolation of PCR products. Nucleic Acids Research 23(22): 4742.

de Souza IR, Muniz YCN, Saldanha GdM, Alves L, da Rosa FC, Maegawa FAB, Susin MF, Lipinksi MdS, and Petzl-Erler ML. 2003. Demographic and genetic structures of two partially isolated communities of Santa Catarina Island, southern Brazil. Human Biology 75(2): 241-253.

Deguilloux MF, Pemonge M, Mendisco F, Thibon D, Cartron I, and Castex D. 2014. Ancient DNA and kinship analysis of human remains deposited in Merovingian

necropolis sarcophagi (Jau Dignac et Loirac, France, 7th–8th century AD). Journal of Archaeological Science 41: 399-405.

Deguilloux MF, Ricaud S, Leahy R, and Pemonge M-H. 2011. Analysis of ancient human DNA and primer contamination: One step backward one step forward. Forensic Science International 210(1): 102-109.

Denaro M, Blanc H, Johnson MJ, Chen KH, Wilmsen E, Cavalli-Sforza LL, and Wallace DC. 1981. Ethnic variation in Hpa I endonuclease cleavage patterns of human mitochondrial DNA. Proceedings of the National Academy of Sciences 78(9): 5768-5772.

Derry D. 1909. Anatomical report (B). Archaeological Survey of Nubia, Cairo, Egyptian Ministry of Finance: 40-42.

Der Sarkissian C, Ermini L, Jónsson H, Alekseev A, Crubezy E, Shapiro B, and Orlando L. 2014. Shotgun microbial profiling of fossil remains. Molecular Ecology 23(7): 1780-1798.

Destro-Bisol G, Donati F, Coia V, Boschi I, Verginelli F, Caglia A, Tofanelli S, Spedini G, and Capelli C. 2004. Variation of Female and Male Lineages in Sub-Saharan Populations: The Importance of Sociocultural Factors. Molecular Biology and Evolution 21(9): 1673-1682.

Destro-Bisol G, Jobling MA, Rocha J, Novembre J, Richards MB, Mulligan C, Batini C, and Manni F. 2010. Molecular anthropology in the genomic era. Journal of Anthropological Sciences 88: 93-112.

Devor EJ. 1987. Transmission of human craniofacial dimensions. Journal of Craniofacial Genetics and Developmental Biology 7(2): 95-106.

Di Cristofaro J, Pennarun E, Mazières S, Myres NM, Lin AA, Temori SA, Metspalu M, Metspalu E, Witzel M, King RJ, Underhill PA, Villems R, and Chiaroni J. 2013. Afghan Hindu Kush: Where Eurasian Sub-Continent Gene Flows Converge. PLoS One 8(10): e76748.

Dobon B, Hassan HY, Laayouni H, Luisi P, Ricaño-Ponce I, Zhernakova A, Wijmenga C, Tahir H, Comas D, Netea MG, and Bertranpetit J. 2015. The genetics of East African populations: a Nilo-Saharan component in the African genetic landscape. Scientific Reports 5: 09996

Doden E, and Halves R. 1984. On the functional morphology of the human petrous bone. American Journal of Anatomy 169(4): 451-462.

Dudar JC, Waye JS, and Saunders SR. 2003. Determination of a Kinship System Using Ancient DNA, Mortuary Practice, and Historic Records in an Upper Canadian Pioneer Cemetery. International Journal of Osteoarchaeology 13(4): 232-246.

Dupré J. 2008. What Genes Are and Why There Are No Genes for Race. In: Koenig BA, Lee SS-J, and Richardson SS (editors). Revisiting Race in a Genomic Age. New Jersey: Rutgers University Press.

E pluribus unum (editorial). 2010. Nature Methods 7(5): 331-331.

Edwards DN. 2004. The Nubian Past: An Archaeology of the Sudan. New York: Routledge.

Edwards DN. 2007. The Archaeology of Sudan and Nubia. Annual Review of Anthropology 36: 211-228.

Edwards DN. 2013. Medieval and Post-Medieval States of the Nile Valley. In: The Oxford Handbook of African Archaeology. Mitchell P, and Lane PJ (editors). Printed from Oxford Handbooks Onlines.

Elhaik E. 2012. Empirical distributions of FST from large-scale human polymorphism data. PLoS One 7(11): e49837.

Elkamel S, Boussetta S, Khodjet-El-Khil H, Benammar Elgaaied A, and Cherni L. 2018. Ancient and recent Middle Eastern maternal genetic contribution to North Africa as viewed by mtDNA diversity in Tunisian Arab populations. American Journal of Human Biology: e23100.

Elliot Smith G. 1909. Archaeological Survefy of Nubia, Bulletin No. 3. Cairo: National Printing Department.

Elliot Smith G. 1910. The Archaeological Survey of Nubia Report for 1907-1908. Cairo: National Printing Department.

Emery WB. 1965. Egypt in Nubia. London: Hutchinson.

Enk JM, Devault AM, Kuch M, Murgha YE, Rouillard J-M, and Poinar HN. 2014. Ancient whole genome enrichment using baits built from modern DNA. Molecular Biology and Evolution 31(5): 1292-1294.

Ennafaa H, Cabrera VM, Abu-Amero KK, González AM, Amor MB, Bouhaha R, Dzimiri N, Elgaaïed AB, and Larruga JM. 2009. Mitochondrial DNA haplogroup H structure in North Africa. BMC Genetics 10(1): 8.

Ezkurdia I, Juan D, Rodriguez JM, Frankish A, Diekhans M, Harrow J, Vazquez J, Valencia A, and Tress ML. 2014. Multiple evidence strands suggest that there

may be as few as 19 000 human protein-coding genes. Human Molecular Genetics 23(22): 5866-5878.

Fadhlaoui-Zid K, Plaza S, Calafell F, Ben Amor M, Comas D, Bennamar A, and Gaaied E. 2004. Mitochondrial DNA heterogeneity in Tunisian Berbers. Annals of Human Genetics 68(3): 222-233.

Fadhlaoui-Zid K, Rodríguez-Botigué L, Naoui N, Benammar-Elgaaied A, Calafell F, and Comas D. 2011. Mitochondrial DNA structure in North Africa reveals a genetic discontinuity in the Nile Valley. American Journal of Physical Anthropology 145(1): 107-117.

Fehren-Schmitz L, Llamas B, Lindauer S, Tomasto-Cagigao E, Kuzminsky S, Rohland N, Santos FR, Kaulicke P, Valverde G, Richards SM, Nordenfelt S, Seidenberg V, Mallick S, Cooper A, Reich D, and Haak W. 2015. A re-appraisal of the early Andean human remains from Lauricocha in Peru. PLoS One 10(6): e0127141.

Fehren-Schmitz L, Reindel M, Cagigao ET, Hummel S, and Herrmann B. 2010. Pre-Columbian population dynamics in coastal southern Peru: A diachronic investigation of mtDNA patterns in the Palpa region by ancient DNA analysis. American Journal of Physical Anthropology 141(2): 208-221.

Feldman MW, Lewontin RC, and King M-C. 2003. Race: a genetic melting-pot. Nature 424(6947): 374-374.

Fernandes D, Sirak K, Novak M, Finarelli JA, Byrne J, Connolly E, Carlsson JEL, Ferretti E, Pinhasi R, and Carlsson J. 2017. The Identification of a 1916 Irish Rebel: New Approach for Estimating Relatedness From Low Coverage Homozygous Genomes. Scientific Reports 7: 41529.

Fernandes V, Alshamali F, Alves M, Costa MD, Pereira JB, Silva NM, Cherni L, Harich N, Černy V, Soares P, Richards MB, Pereira L. 2012. The Arabian Cradle: Mitochondrial Relicts of the First Steps along the Southern Route out of Africa. American Journal of Human Genetics 90(2): 347-355.

Fernandes V, Triska P, Pereira JB, Alshamali F, Rito T, Machado A, Fajkošová Z, Cavadas B, Černy V, Soares P, Richards MB, and Pereira L. 2015. Genetic Stratigraphy of Key Demographic Events in Arabia. PLoS One 10(3): e0118625.

Fernández E, Pérez-Pérez A, Gamba C, Prats E, Cuesta P, Anfruns J, Molist M, Arroyo-Pardo E, and Turbón D. 2014. Ancient DNA Analysis of 8000 BC Near Eastern Farmers Supports an Early Neolithic Pioneer Maritime Colonization of Mainland Europe through Cyprus and the Aegean Islands. PLoS Genetics 10(6): e1004401.

Firth CM. 1912. The Archaeological Survey of Nubia Report for 1908-1909. Cairo: Government Press.

Firth CM. 1927. The Archaeological Survey of Nubia Report for 1910-1911. Cairo: Government Press.

Fleuhr-Lobban C, and Lobban RA. 2016. New Social Movements in Nubian Identity Among Nubians in Egypt, Sudan, and the United States. In: Mullings L (editor). New social movements in the African diaspora: Challenging global apartheid. New York: Palgrave Macmillan.

Fortes GG, and Paijmans JL. 2015. Analysis of whole mitogenomes from ancient samples. Whole Genome Amplification: Methods and Protocols: 179-195.

Fortes GG, Speller CF, Hofreiter M, and King TE. 2013. Phenotypes from ancient DNA: Approaches, insights and prospects. BioEssays 35(8): 690-695.

Fox CL. 1997. mtDNA analysis in ancient Nubians supports the existence of gene flow between sub-Sahara and North Africa in the Nile Valley. Annals of Human Biology 24(3): 217-227.

Frisch T, Sørensen M, Overgaard S, Lind M, and Bretlau P. 1998. Volume-referent bone turnover estimated from the interlabel area fraction after sequential labeling. Bone 22(6): 677-682.

Fu Q, Hajdinjak M, Moldovan OT, Constantin S, Mallick S, Skoglund P, Patterson N, Rohland N, Lazaridis I, Nickel B, Viola B, Prüfer K, Meyer M, Kelso J, Reich D, and Pääbo S. 2015. An early modern human from Romania with a recent Neanderthal ancestor. Nature 524(7564): 216-219.

Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, and Pääbo S. 2013a. DNA analysis of an early modern human from Tianyuan Cave, China. Proceedings of the National Academy of Sciences 110(6): 2223-2227.

Fu Q, Mittnik A, Johnson PLF, Bos K, Lari M, Bollongino R, Sun C, Giemsch L, Schmitz R, Burger J, Ronchitelli AM, Martini F, Cremonesi RG, Svoboda J, Bauer P, Caramelli D, Castellano S, Reich D, Pääbo S, and Krause J. 2013b. A Revised Timescale for Human Evolution Based on Ancient Mitochondrial Genomes. Current Biology 23(7): 553-559.

Fu Q, Posth C, Hajdinjak M, Petr M, Mallick S, Fernandes D, Furtwängler A, Haak W, Meyer M, Mittnik A, Nickel B, Peltzer A, Rohland N, Slon V, Talamo S, Lazaridis I, Lipson M, Mathieson I, Schiffels S, Skoglund P, Derevianko AP, Drozdov N, Slavinsky V, Tsybankov A, Cremonesi RG, Mallegni F, Gély B, Vacca E, Morales MRG, Straus LG, Neugebauer-Maresch C, Teschler-Nicola M, Constantin S, Moldovan OT, Benazzi S, Peresani M, Coppola D, Lari M, Ricci S, Ronchitelli A, Valentin F, Thevenet C, Wehrberger K, Grigorescu D, Rougier H, Crevecoeur I, Flas D, Semal P, Mannino MA, Cupillard C, Bocherens H, Conard NJ, Harvati K, Moiseyev V, Drucker DG, Svoboda J, Richards MP, Caramelli D,

Pinhasi R, Keleso J, Patterson N, Krause J, Pääbo S, and Reich D. 2016. The genetic history of Ice Age Europe. Nature 534(7606): 200-205.

Fulton TL. 2012. Setting up an ancient DNA laboratory. In: Shapiro B, and Hofreiter M (editors). Ancient DNA: Methods and Protocols. New York: Humana Press. p.1-11.

Galland M, Van Gerven DP, Von Cramon-Taubadel N, and Pinhasi R. 2016. 11,000 years of craniofacial and mandibular variation in Lower Nubia. Scientific Reports 6: 31040.

Gallego-Llorente M, Connell S, Jones ER, Merrett DC, Jeon Y, Eriksson A, Siska V, Gamba C, Meiklejohn C, Beyer R, Jeon S, Cho YS, Hofreiter M, Bhak J, Manica A, and R Pinhasi. 2016. The genetics of an early Neolithic pastoralist from the Zagros, Iran. Scientific Reports 6: 31326.

Gallego-Llorente M, Jones ER, Eriksson A, Siska V, Arthur KW, Arthur JW, Curtis MC, Stock JT, Coltorti M, Pieruccini P, Stretton S, Brock F, Higham T, Park Y, Hofreiter M, Bradley DG, Bhak J, Pinhasi R, and A Manica. 2015. Ancient Ethiopian genome reveals extensive Eurasian admixture throughout the African continent. Science 350(6262): 820-822.

Gamba C, Hanghøj K, Gaunitz C, Alfarhan AH, Alquraishi SA, Al-Rasheid KA, Bradley DG, and Orlando L. 2016. Comparing the performance of three ancient DNA extraction methods for high-throughput sequencing. Molecular Ecology Resources 16(2): 459-469.

Gamba C, Jones ER, Teasdale MD, McLaughlin RL, Gonzalez-Fortes G, Mattiangeli V, Domboróczki L, Kővári I, Pap I, Anders A, Whittle A, Dani J, Raczky P, Higham TFG, Hofreiter M, Bradley DG, and Pinhasi R. 2014. Genome flux and stasis in a five millennium transect of European prehistory. Nature Communications 5: 5257.

Gansauge M-T, and Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nature Protocols 8(4): 737-748.

Gansauge M-T, and Meyer M. 2014. Selective enrichment of damaged DNA molecules for ancient genome sequencing. Genome Research 24(9): 1543-1549.

Gebremeskel EI, and Ibrahim ME. 2014. Y-chromosome E haplogroups: their distribution and implication to the origin of Afro-Asiatic languages and pastoralism. European Journal of Human Genetics 22(12): 1387.

Gilbert MT., Bandelt H-J, Hofreiter M, and Barnes I. 2005. Assessing ancient DNA studies. Trends in Ecology and Evolution 20(10): 541-544.

Gilbert MTP, Kivisild T, Grønnow B, Andersen PK, Metspalu E, Reidla M, Tamm E, Axelsson E, Götherström A, Campos PF, Rasmussen M, Metspalu M, Higham TFG, Schwenninger J-L, Nathan R, De Hoog C-J, Koch A, Møller LN, Andreasen C, Meldgaard M, Villems R, Bendixen C, and Willerslev E. 2008. Paleo-Eskimo mtDNA genome reveals matrilineal discontinuity in Greenland. Science 320(5884): 1787-1789.

Giles RE, Blanc H, Cann HM, and Wallace DC. 1980. Maternal inheritance of human mitochondrial DNA. Proceedings of the National Academy of Sciences 77(11): 6715-6719.

Ginolhac A, Rasmussen M, Gilbert MTP, Willerslev E, and Orlando L. 2011. mapDamage: testing for damage patterns in ancient DNA sequences. Bioinformatics 27(15): 2153-2155.

Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, and Nusbam C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nature Biotechnology 27(2): 182-189.

Godde K. 2009. An examination of Nubian and Egyptian biological distances: Support for biological diffusion or in-situ development? Homo-Journal of Comparative Human Biology 60(5): 389-404.

Godde K. 2013. An Examination of the Spatial–Temporal Isolation Model in a Nilotic Population: Variation across Space and Time in Nubians Using Cranial Discrete Traits. International Journal of Osteoarchaeology 23(3): 324-333.

Goldberg A, Günther T, Rosenberg NA, and Jakobsson M. 2017. Ancient X chromosomes reveal contrasting sex bias in Neolithic and Bronze Age Eurasian migrations. Proceedings of the National Academy of Sciences 114(10): 2657-2662.

Golenberg EM, Brown T, Bada J, Westbroek P, Bishop M, and Dover G. 1991. Amplification and analysis of Miocene plant fossil DNA [and discussion]. Philosophical Transactions of the Royal Society B: Biological Sciences 333(1268): 419-427.

Gonder MK, Mortensen HM, Reed FA, de Sousa A, and Tishkoff SA. 2007. Whole-mtDNA Genome Sequence Analysis of Ancient African Lineages. Molecular Biology and Evolution 24(3): 757-768.

González AM, Larruga JM, Abu-Amero KK, Shi Y, Pestano J, and Cabrera VM. 2007. Mitochondrial lineage M1 traces an early human backflow to Africa. BMC Genomics 8: 223-223.

Goodman AH, and Rose JC. 1990. Assessment of systemic physiological perturbations from dental enamel hypoplasias and associated histological structures. American Journal of Physical Anthropology 33(S11): 59-110.

Gray H. 1918. Anatomy of the human body. Philadelphia: Lea and Febiger.

Green RE, Krause J, Briggs AW, Maricic T, Stenzel U, Kircher M, Patterson N, Li H, Zhai W, Fritz MH-Y, Hansen NF, Durand EY, Malaspinas A-S, Jensen JD, Marques-Bonet T, Alkan C, Prüfer K, Meyer M, Burbano HA, Good JM, Schultz R, Aximu-Petri A, Buthof A, Höffner B, Siegemund M, Weihmann A, Nusbaum C, Lander ES, Russ C, Novod N, Affourtit J, Egholm M, Verna C, Rudan P, Brajkovic D, Kucan Ž, Gušis I, Doronichev VB, Golovanova LV, Lalueza-Fox C, de la Rasilla M, Fortea J, Rosas A, Schmitz RW, Johnson PLF, Eichler EE, Falush D, Birney E, Mullikin JC, Slatkin M, Nielsen R, Kelso J, Lachmann M, Reich D, and Pääbo S. 2010. A draft sequence of the Neandertal genome. Science 328(5979): 710-722.

Greene DL. 1972. Dental anthropology of early Egypt and Nubia. Journal of Human Evolution 1(3): 315-324.

Greene DL. 1982. Discrete dental variations and biological distances of Nubian populations. American Journal of Physical Anthropology 58(1): 75-79.

Greene DL, Ewing GH, and Armelagos GJ. 1967. Dentition of a mesolithic population from Wadi Halfa, Sudan. American Journal of Physical Anthropology 27(1): 41-55.

Grugni V, Battaglia V, Hooshiar Kashani B, Parolo S, Al-Zahery N, Achilli A, Olivieri A, Gandini F, Houshmand M, Sanati MH, Torroni A, and Semino O. 2012. Ancient Migratory Events in the Middle East: New Clues from the Y-Chromosome Variation of Modern Iranians. PLoS One 7(7): e41252.

Haak W, Brandt G, de Jong HN, Meyer C, Ganslmeier R, Heyd V, Hawkesworth C, Pike AWG, Meller H, and Alt KW. 2008. Ancient DNA, Strontium isotopes, and osteological analyses shed light on social and kinship organization of the Later Stone Age. Proceedings of the National Academy of Sciences 105(47): 18226-18231.

Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, Brandt G, Nordenfelt S, Harney eE, Stewardson K, Fu Q, Mittnik A, Bánffy E, Economou C, Francken M, Friederich S, Pena RG, Hallgren F, Khartanovich V, Khokhlov A, Kunst M, Kuznetsov P, Meller H, Mochalov O, Moiseyev V, Nicklisch N, Pichler SL, Risch R, Guerra MAR, Roth C, Szécsényi-Nagy A, Wahl J, Meyer M, Krause J, Brown D, Anthony D, Cooper A, Alt KW, and Reich D. 2015. Massive migration from the steppe is a source for Indo-European languages in Europe. Nature 522(7555): 207.

Haber M, Mezzavilla M, Xue Y, and Tyler-Smith C. 2016. Ancient DNA and the rewriting of human history: be sparing with Occam's razor. Genome Biology 17(1): 1-8.

Hadley D. 2010. Burying the socially and physically distinctive in later Anglo-Saxon England. In: Buckberry J, and Cherryson A (editors). Burial in Later Anglo-Saxon England c. 650-1100 AD. Oxbow Books. p.103-116.

Hammer MF, Karafet T, Rasanayagam A, Wood ET, Altheide TK, Jenkins T, Griffiths RC, Templeton AR, and Zegura SL. 1998. Out of Africa and Back Again: Nested Cladistic Analysis of Human Y Chromosome Variation. Molecular Biology and Evolution 15(4): 427-441.

Hammer MF, Karafet TM, Redd AJ, Jarjanazi H, Santachiara-Benerecetti S, Soodyall H, and Zegura SL. 2001. Hierarchical patterns of global human Y-chromosome diversity. Molecular Biology and Evolution 18(7): 1189-1203.

Hammer MF, Redd AJ, Wood ET, Bonner MR, Jarjanazi H, Karafet T, Santachiara-Benerecetti S, Oppenheim A, Jobling MA, Jenkins T, Ostrer H, and Bonné-Tamir B. 2000. Jewish and Middle Eastern non-Jewish populations share a common pool of Y-chromosome biallelic haplotypes. Proceedings of the National Academy of Sciences 97(12): 6769-6774.

Hammer MF, Spurdle A, Karafet T, Bonner M, Wood E, Novelletto A, Malaspina P, Mitchell RJ, Horai S, Jenkins T, Zegura SL. 1997. The geographic distribution of human Y chromosome variation. Genetics 145(3): 787-805.

Handt O, Höss M, Krings M, and Pääbo S. 1994. Ancient DNA: Methodological challenges. Cellular and Molecular Life Sciences 50(6): 524-529.

Hänni C, Brousseau T, Laude V, and Stehelin D. 1995. Isopropanol precipitation removes PCR inhibitors from ancient bone extracts. Nucleic Acids Research 23(5): 881.

Hansen HB, Damgaard PB, Margaryan A, Stenderup J, Lynnerup N, Willerslev E, and Allentoft ME. 2017. Comparing Ancient DNA Preservation in Petrous Bone and Tooth Cementum. PLoS One 12(1): e0170940.

Hardy OJ, and Vekemans X. 2002. SPAGeDi: a versatile computer program to analyse spatial genetic structure at the individual or population levels. Molecular Ecology Notes 2: 618-620.

Harich N, Costa MD, Fernandes V, Kandil M, Pereira JB, Silva NM, and Pereira L. 2010. The trans-Saharan slave trade – clues from interpolation analyses and high-

resolution characterization of mitochondrial DNA lineages. BMC Evolutionary Biology 10(1): 1-18.

Harpending HC, Batzer MA, Gurven M, Jorde LB, Rogers AR, and Sherry ST. 1998. Genetic traces of ancient demography. Proceedings of the National Academy of Sciences 95(4): 1961-1967.

Hasan YF. 1967. Main aspects of the Arab migration to the Sudan. Arabica 14: 14-31.

Hassan HY, Underhill PA, Cavalli-Sforza LL, and Ibrahim ME. 2008. Y-chromosome variation among Sudanese: Restricted gene flow, concordance with language, geography, and history. American Journal of Physical Anthropology 137(3): 316-323.

Hayden B. 2001. Richman, poorman, beggarman, chief: The dynamics of social inequality. In: Feinman GM, and Price TD (editors). Archaeology at the Millennium. New York: Springer. p.231-272.

Hedrick P. 2005. Genetics of Populations (3rd edition). Sudbury: Jones and Bartlett.

Hernandez CJ, Majeska RJ, and Schaffler MB. 2004. Osteocyte density in woven bone. Bone 35(5): 1095-1099.

Hervella M, Svensson E, Alberdi A, Günther T, Izagirre N, Munters A, Alonso S, Ioana M, Ridiche F, Soficaru A, Jakobsson M, Netea MG, and de-la-Rua C. 2016. The mitogenome of a 35,000-year-old Homo sapiens from Europe supports a Palaeolithic back-migration to Africa. Scientific Reports 6: 25501.

Heyer E, Chaix R, Pavard S, and Austerlitz F. 2012. Sex-specific demographic behaviours that shape human genomic variation. Molecular Ecology 21(3): 597-612.

Higuchi R, Bowman B, Freiberger M, Ryder OA, and Wilson AC. 1984. DNA sequences from the quagga, an extinct member of the horse family. Nature 312: 282-284.

Hinds DA, Stuve LL, Nilsen GB, Halperin E, Eskin E, Ballinger DG, Frazer KA, Cox DR. 2005. Whole-genome patterns of common DNA variation in three human populations. Science 307(5712): 1072-1079.

Hirbo JB. 2011. Complex Genetic History of East African Human Populations. PhD Dissertation: University of Maryland, College Park.

Hofmanová Z, Kreutzer S, Hellenthal G, Sell C, Diekmann Y, Díez-del-Molino D, van Dorp L, López S, Kousathanas A, Link V, Kirsanow K, Cassidy LM, Martiniano R, Strobel M, Scheu A, Kotsakis K, Halstead P, Triantaphyllou S, Kyparissi-Apostolika N, Urem-Kotsou D, Ziota C, Adaktylou F, Gopalan S, Bobo DM,

Winkelbach L, Blöcher J, Unterländer M, Leuenberger C, Çilingiroğlu Ç, Horejs B, Gerritsen F, Shennan SJ, Bradley DG, Currat M, Veeramah KR, Wegnmann D, Thomas MG, Papageorgopolou C, and Burger J. 2016. Early farmers from across Europe directly descended from Neolithic Aegeans. Proceedings of the National Academy of Sciences 113(25): 6886-6891.

Hofreiter M, Jaenicke V, Serre D, von Haeseler A, and Pääbo S. 2001. DNA sequences from multiple amplifications reveal artifacts induced by cytosine deamination in ancient DNA. Nucleic Acids Research 2923: 4793-4799.

Hofreiter M, Paijmans JL, Goodchild H, Speller CF, Barlow A, Fortes GG, Thomas JA, Ludwig A, and Collins MJ. 2015. The future of ancient DNA: Technical advances and conceptual shifts. BioEssays 37(3): 284-293.

Hollfelder N, Schlebusch CM, Günther T, Babiker H, Hassan HY, and Jakobsson M. 2017. Northeast African genomic variation shaped by the continuity of indigenous groups and Eurasian migrations. PLoS Genetics 13(8): e1006976.

Holsinger KE, and Weir BS. 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. Nature Reviews Genetics 10(9): 639-650.

Horn S. 2012. Target enrichment via DNA hybridization capture. In: Shapiro B, and Hofreiter M (editors). Ancient DNA: Methods and Protocols. New York: Humana Press. p.177-188.

Höss M, Jaruga P, Zastawny TH, Dizdaroglu M, and Pääbo S. 1996. DNA damage and DNA sequence retrieval from ancient tissues. Nucleic Acids Research 24(7): 1304-1307.

Howell N. 1982. Village Composition Implied by a Paleodemographic Life Table: The Libben Site. American Journal of Physical Anthropology 59(3): 263-269.

Hudson RR, Slatkin M, and Maddison W. 1992. Estimation of levels of gene flow from DNA sequence data. Genetics 1322: 583-589.

Hummert JR. 1983. Cortical Bone Growth and Dietary Stress Among Subadults From Nubia's Batn El Hajar. American Journal of Physical Anthropology 62(2): 167-176.

Hummert JR, and Van Gerven DP. 1983. Skeletal growth in a medieval population from Sudanese Nubia. American Journal of Physical Anthropology 60(4): 471-478.

Illumina, Inc. 2011. Quality Scores for Next-Generation Sequencing: Assessing sequencing accuracy using Phred quality scoring.

(https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf).

Illumina, Inc. 2016a. An Introduction to Next-Generation Sequencing Technology. (https://www.illumina.com/content/dam/illumina-marketing/documents/products/illumina_sequencing_introduction.pdf).

Illumina, Inc. 2016b. MiSeq System: Denature and Dilute Libraries Guide. (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-03.pdf).

Illumina, Inc. 2016c. NextSeq System: Denature and Dilute Libraries Guide. (https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/nextseq/nextseq-denature-dilute-libraries-guide-15048776-02.pdf).

Ingman M, Kaessmann H, Pääbo S, and Gyllensten U. 2000. Mitochondrial genome variation and the origin of modern humans. Nature 408(6813): 708-713.

Irish JD. 2005. Population continuity vs. discontinuity revisited: Dental affinities among late Paleolithic through Christian-era Nubians. American Journal of Physical Anthropology 128(3): 520-535.

Jackes M. 2011. Representativeness and bias in archaeological skeletal samples. In Agarwal SC, and Glencross BA (editors). Social Bioarchaeology. West Sussex, UK: Blackwell Publishing Ltd. p. 107-146.

Jackson J. 1957. Changes in the climate and vegetation of the Sudan. Sudan Notes and Records 38: 47-66.

Jakobielski S. 1987. North and South in Christian Nubian Culture: Archaeology and History. Paper presented at the Sixth International Conference for Nubian Studies, Uppsala, Sweden.

Jakobsson M, Scholz SW, Scheet P, Gibbs JR, VanLiere JM, Fung H-C, Szpiech ZA, Degnan JH, Wang K, Guerreiro R, Bras JM, Schymick JC, Hernandez DG, Traynor BJ, Simon-Sanchz J, Matarin M, Britton A, van de Leemput J, Rafferty I, Bucan M, Cann HM, Hardy JA, Rosenberg NA, and Singleton AB. 2008. Genotype, haplotype and copy-number variation in worldwide human populations. Nature 451(7181): 998.

Jiang J, Mendelssohn R, Schwing F, and Fraedrich K. 2002. Coherency detection of multiscale abrupt changes in historic Nile flood levels. Geophysical Research Letters 29(8): 112-1–112-4.

Jobling M, Hollox E, Hurles M, Kivisild T, and Tyler-Smith C. 2014. Human Evolutionary Genetics. New York: Garland Science.

Jobling MA, and Tyler-Smith C. 1995. Fathers and sons: the Y chromosome and human evolution. Trends in Genetics 11(11): 449-456.

Jobling MA, and Tyler-Smith C. 2003. The Human Y Chromosome: An Evolutionary Marker Comes of Age. Nature Reviews Genetics 4(8): 598.

Jobling MA, and Tyler-Smith C. 2017. Human Y-chromosome variation in the genome-sequencing era. Nature Reviews Genetics 18(8): 485.

Johnston FE. 1962. Growth of the Long Bones of Infants and Young Children at Indian Knoll. American Journal of Physical Anthropology 20(3): 249-254.

Jones ER, Fortes GG, Connell S, Siska V, Eriksson A, Martiniano R, McLaughlin RL, Gallego-Llorente M, Cassidy LM, Gamba C, Meshveliani T, Bar-Yosef O, Müller W, Belfer-Cohen A, Matskevich Z, Jakeli N, Higham TFG, Currat M, Lordkipanidze D, Hofreiter M, Manica A, Pinhasi R, and Bradley DG. 2015. Upper Palaeolithic genomes reveal deep roots of modern Eurasians. Nature Communications 6: 8912.

Jónsson H, Ginolhac A, Schubert M, Johnson PL, and Orlando L. 2013. mapDamage2.0: fast approximate Bayesian estimates of ancient DNA damage parameters. Bioinformatics 29(13): 1682-1684.

Jónsson H, Schubert M, Seguin-Orlando A, Ginolhac A, Petersen L, Fumagalli M, Albrechtsen A, Petersen B, Korneliussen TS, Vilstrup JT, Lear T, Myka JL, Lundquist J, Miller DC, Alfarhan AH, Alquraishi SA, Al-Rasheid KAS, Stagegaard J, Strauss G, Bertelsen MF, Sicheritz-Ponten T, Antczak DF, Bailey E, Nielsen R, Willerslev E, and Orlando L. 2014. Speciation with gene flow in equids despite extensive chromosomal plasticity. Proceedings of the National Academy of Sciences 111(52): 18655-18660.

Jorde LB, Watkins W, Bamshad M, Dixon M, Ricker C, Seielstad M, and Batzer M. 2000. The distribution of human genetic diversity: a comparison of mitochondrial, autosomal, and Y-chromosome data. The American Journal of Human Genetics 66(3): 979-988.

Jorde LB, and Wooding SP. 2004. Genetic variation, classification and 'race'. Nature Genetics 36(11): S28-S33.

Juras A, Chyleński M, Krenz-Niedbała M, Malmström H, Ehler E, Pospieszny Ł, Łukasik S, Bednarczyk J, Piontek J, Jakobsson M, and Dabert M. 2017. Investigating kinship of Neolithic post-LBK human remains from Krusza Zamkowa, Poland using ancient DNA. Forensic Science International: Genetics 26: 30-39.

Kaestle FA, and Horsburgh K. 2002. Ancient DNA in Anthropology: Methods, Applications, and Ethics. American Journal of Physical Anthropology 45: 92-130.

Kalmár T, Bachrati CZ, Marcsik A, and Raskó I. 2000. A simple and efficient method for PCR amplifiable DNA extraction from ancient bones. Nucleic Acids Research 28(12): e67.

Karafet TM, Mendez FL, Meilerman MB, Underhill PA, Zegura SL, and Hammer MF. 2008. New binary polymorphisms reshape and increase resolution of the human Y chromosomal haplogroup tree. Genome Research 18(5): 830-838.

Katz D, and Suchey JM. 1986. Age determination of the male os pubis. American Journal of Physical Anthropology 69(4): 427-435.

Kavenoff R, Klotz LC, and Zimm BH. 1974. On the nature of chromosome-sized DNA molecules. Cold Spring Harbor Symposia on Quantitative Biology 38: 1-8.

Keinan A, Mullikin JC, Patterson N, and Reich D. 2007. Measurement of the human allele frequency spectrum demonstrates greater genetic drift in East Asians than in Europeans. Nature Genetics 39(10): 1251.

Kidd KK, Pakstis A, Speed W, and Kidd J. 2004. Understanding human DNA sequence variation. Journal of Heredity 95(5): 406-420.

Kilgore L, Jurmain R, and Van Gerven DP. 1997. Palaeoepidemiological Patterns of Trauma in a Medieval Nubian Skeletal Population. International Journal of Osteoarchaeology 72: 103-114.

Kılınç GM, Omrak A, Özer F, Günther T, Büyükkarakaya AM, Bıçakçı E, Baird D, Dönertaş HM, Ghalichi A, Yaka R, Koptekin D, Açan SC, Parvizi P, Krzewińska M, Daskalaki EA, Yüncü E, Dağtas ND, Fairbairn A, Pearson J, Mustafaoğlu G, Erdal YG, Çakan YG, Togan İ, Somel M, Storå J, Jakobsson M, and Götherström A. 2016. The demographic development of the first farmers in anatolia. Current Biology 26(19): 2659-2666.

King TE, Bowden GR, Balaresque PL, Adams SM, Shanks ME, and Jobling MA. 2007. Thomas Jefferson's Y chromosome belongs to a rare European lineage. American Journal of Physical Anthropology 132(4): 584-589.

Kircher M, and Kelso J. 2010. High-throughput DNA sequencing– concepts and limitations. BioEssays 32(6): 524-536.

Kircher M, Sawyer S, and Meyer M. 2012. Double indexing overcomes inaccuracies in multiplex sequencing on the Illumina platform. Nucleic Acids Research 40(1): e3.

Kirwan LP. 1937. A SURVEY OF NUBIAN ORIGINS. Sudan Notes and Records 20(1): 47-62.

Kirwan LP. 1959. THE INTERNATIONAL POSITION OF SUDAN IN ROMAN AND MEDIEVAL TIMES. Sudan Notes and Records 40: 23-37.

Kirwan LP. 1984. THE BIRTH OF CHRISTIAN NUBIA: SOME ARCHAEOLOGICAL PROBLEMS. Rivista degli studi orientali 58(1/4): 119-134.

Kivisild T. 2015. Maternal ancestry and population history from whole mitochondrial genomes. Investigative genetics 6(1): 3.

Kivisild T, Reidla M, Metspalu E, Rosa A, Brehm A, Pennarun E, Parik J, Geberhiwot T, Usanga E, and Villems R. 2004. Ethiopian mitochondrial DNA heritage: tracking gene flow across and around the gate of tears. The American Journal of Human Genetics 75(5): 752-770.

Kivisild T, Shen P, Wall DP, Do B, Sung R, Davis K, Passarino G, Underhill PA, Scharfe C, Torroni A, Scozzari R, Modiano D, Coppa A, de Knijff P, Feldman M, Cavalli-Sforza LL, and Oefner PJ. 2005. The role of selection in the evolution of human mitochondrial genomes. Genetics 172(1): 372-387.

Klales AR, Ousley SD, and Vollner JM. 2012. A revised method of sexing the human innominate using Phenice's nonmetric traits and statistical methods. American Journal of Physical Anthropology 149(1): 104-114.

Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R, Specht G, and Kronenberg F. 2011. HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Human Mutation 32(1): 25-32.

Knapp M, Clarke AC, Horsburgh KA, and Matisoo-Smith EA. 2012a. Setting the stage–building and working in an ancient DNA laboratory. Annals of Anatomy-Anatomischer Anzeiger 194(1): 3-6.

Knapp M, and Hofreiter M. 2010. Next generation sequencing of ancient DNA: requirements, strategies and perspectives. Genes 1: 227-243.

Knapp M, Lalueza-Fox C, and Hofreiter M. 2015. Re-inventing ancient human DNA. Investigative Genetics 6(1): 4.

Knapp M, Stiller M, and Meyer M. 2012b. Generating barcoded libraries for multiplex high-throughput sequencing. In: Shapiro B, and Hofreiter M (editors). Ancient DNA: Methods and Protocols. New York: Humana Press. p.155-170.

Knight A, Underhill PA, Mortensen HM, Zhivotovsky LA, Lin AA, Henn BM, Louis D, Ruhlen M, and Mountian JL. 2003. African Y chromosome and mtDNA

divergence provides insight into the history of click languages. Current Biology 13(6): 464-473.

Konigsberg LW. 1985. Demography and Mortuary Practice at Siep Mound One. Midcontinental Journal of Archaeology 10(1): 123-148.

Konigsberg LW, and Ousley SD. 2009. Multivariate quantitative genetics of anthropometric traits from the Boas data. Human Biology 81(5/6): 579-594.

Korlević P, Gerber T, Gansauge M-T, Hajdinjak M, Nagel S, Aximu-Petri A, and Meyer M. 2015. Reducing microbial and human contamination in DNA extractions from ancient bones and teeth. BioTechniques 58: 87-93.

Korneliussen TS, and Moltke I. 2015. NgsRelate: a software tool for estimating pairwise relatedness from next-generation sequencing data. Bioinformatics 31(24): 4009-4011.

Krause J, Briggs AW, Kircher M, Maricic T, Zwyns N, Derevianko A, and Pääbo S. 2010. A complete mtDNA genome of an early modern human from Kostenki, Russia. Current Biology 20(3): 231-236.

Krause J, and Pääbo S. 2016. Genetic time travel. Genetics 203(1): 9-12.

Krings M, Salem AH, Bauer K, Geisert H, Malek AK, Chaix L, Simon C, Welsby D, Di Rienzo A, Utermann G, Sajantila A, Pääbo S, and Stoneking M. 1999. mtDNA analysis of Nile River Valley populations: A genetic corridor or a barrier to migration? The American Journal of Human Genetics 64(4): 1166-1176.

Kujanová M, Pereira L, Fernandes V, Pereira JB, and Černý V. 2009. Near Eastern Neolithic genetic input in a small oasis of the Egyptian Western Desert. American Journal of Physical Anthropology 140(2): 336-346.

Lacan M, Keyser C, Ricaut F-X, Brucato N, Duranthon F, Guilaine J, Crubézy E, and Ludes B. 2011. Ancient DNA reveals male diffusion through the Neolithic Mediterranean route. Proceedings of the National Academy of Sciences 108(24): 9788-9791.

Laland KN, Odling-Smee J, and Myles S. 2010. How culture shaped the human genome: bringing genetics and the human sciences together. Nature Reviews Genetics 11(2): 137-148.

Lam YM, Chen X, and Pearson OM. 1999. Intertaxonomic variability in patterns of bone density and the differential representation of bovid, cervid, and equid elements in the archaeological record. American Antiquity 64(2): 343-362.

Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, Fernandes D, Novak M, Gamarra B, Sirak K**,** Connell S, Stewardson K, Harney E, Fu Q, Gonzalez-Fortes G, Jones ER, Alpaslan Roodenberg S, Lengyel G, Bocquentine F, Gasparian B, Monge JM, Gregg M, Eshed V, Mizrahi A-S, Mieklejohn C, Gerritsen F, Bejenaru L, Blüher M, Campbell A, Cavalleri G, Comas D, Froguel P, Gilbert E, Kerr SM, Kovacs P, Krause J, McGettigan D, Merrigan M, Merriwether DA, O'Reilly S, Richards MB, Semino O, Shamoon-Pour M, Stefanescu G, Stumvoll M, Tönjes A, Torroni A, Wilson JF, Yengo L, Hovhannisyan NA, Patterson N, Pinhasi R, and Reich D. 2016. Genomic insights into the origin of farming in the ancient Near East. Nature 536(7617): 419-424.

Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH, Schraiber JG, Castellano S, Lipson M, Berger B, Economou C, Bollongino R, Fu Q, Bos KI, Nordenfelt S, Li H, de Filippo C, Prüfer K, Sawyer S, Posth C, Haak W, Hallgren F, Fornander E, Rohland N, Delsate D, Francken M, Guinet J-M, Wahl J, Ayodo G, Babiker HA, Bailliet G, Balanovska E, Balanovsky O, Barrantes R, Bedoya G, Ben-Ami H, Bene J, Berrada F, Bravi CM, Brisighelli F, Busby GBJ, Cali F, Churnosov M, Cole DEC, Corach D, Damba L, van Driem G, Dryomov S, Dugoujon J-M, Fedorova SA, Romero IG, Gubina M, Hammer M, Henn BM, Jervig T, Hodolugil U, Jha AR, Karachanak-Yankova S, Khusainova R, Khusnutdinova E, Kittles R, Kivisild T, Klitz W, Kučinskas V, Kushniarevich A, Laredj L, Litvinov S, Loukidis T, Mahley RW, Belegh B, Metspalu E, Molina J, Mountain J, Näkkäläjärvi K, Nesheva D, Nyambo N, Osipova L, Parik J, Platonov F, Posukh O, Romano V, Rothhammer F, Rudan I, Ruizbakiev R, Sahakyan H, Sajantila A, Salas A, Starikovskaya EB, Tarekegn A, Toncheva D, Turdikulova S, Uktveryte I, Utevska O, Vasquez R, Villena M, Voevoda M, Winkler CA, Yepiskoposyan L, Zalloua P, Zemunik T, Cooper A, Capelli C, Thomas MG, Ruiz-Linares A, Tishkoff SA, Singh L, Thangaraj K, Villems R, Comas D, Sukernik R, Metspalu M, Meyer M, Eichler EE, Burger J, Slatkin M, Pääbo S, Kelso J, Reich D, and Krause J. 2014. Ancient human genomes suggest three ancestral populations for present-day Europeans. Nature 513(7518): 409-413.

Lell JT, and Wallace DC. 2000. The peopling of Europe from the maternal and paternal perspectives. American Journal of Human Genetics 67(6): 1376.

Lennon NJ, Lintner RE, Anderson S, Alvarez P, Barry A, Brockman W, Daza R, Erluch RL, Giannoukos G, Green L, Hollinger A, Hoover CA, Jaffe DB, Juhn F, McCarthy D, Perrin D, Ponchner K, Powers TL, Rizzolo K, Robbins D, Ryan E, Russ C, Sparrow T, Stalker J, Steelman S, Weiand M, Zimmer A, Henn MR, Nusbaum C, and Nicol R. 2010. A scalable, fully automated process for construction of sequence-ready barcoded libraries for 454. Genome Biology, 11(2): R15.

Leonardi M, Librado P, Der Sarkissian C, Schubert M, Alfarhan AH, Alquraishi SA, Al-Rasheid KAS, Gamba C, Willerslev E, and Orlando L. 2016. Evolutionary

Patterns and Processes: Lessons from Ancient DNA. Systematic Biology 66(1): e1-e29.

Lewontin RC. 1972. The apportionment of human diversity. Evolutionary Biology 6: 381-398.

Li H, and Durbin R. 2009. Fast and accurate short read alignment with Burrows–Wheeler transform. Bioinformatics 25(14): 1754-1760.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and 1000 Genomes Project Data Processing Subgroup. 2009. The sequence alignment/map format and SAMtools. Bioinformatics 25(16): 2078-2079.

Li JZ, Absher DM, Tang H, Southwick AM, Casto AM, Ramachandran S, Cann HM, Barsh GS, Feldman M, Cavalli-Sforza LL, and Meyers RM. 2008. Worldwide human relationships inferred from genome-wide patterns of variation. Science 319(5866): 1100-1104.

Lindahl T. 1993. Instability and decay of the primary structure of DNA. Nature 362(6422): 709-715.

Lindahl T, Ljungquist S, Siegert W, Nyberg B, and Sperens B. 1977. DNA N-glycosidases: properties of uracil-DNA glycosidase from Escherichia coli. Journal of Biological Chemistry 252(10): 3286-3294.

Lindahl T, and Nyberg B. 1972. Rate of depurination of native deoxyribonucleic acid. Biochemistry 11(19): 3610-3618.

Linderholm A. 2016. Ancient DNA: the next generation–chapter and verse. Biological Journal of the Linnean Society 117(1): 150-160.

Lippold S, Xu H, Ko A, Li M, Renaud G, Butthof A, Schröder R, and Stoneking M. 2014. Human paternal and maternal demographic histories: insights from high-resolution Y chromosome and mtDNA sequences. Investigative Genetics 5(1): 13.

Lipson M, Skoglund P, Spriggs M, Valentin F, Bedford S, Shing R, Buckley H, Phillip I, Ward GK, Mallick S, Rohland N, Broomandkhoshbacht N, Cheronet O, Ferry M, Harper TK, Michel M, Oppenheimer J, Sirak K, Stewardson K, Auckland K, Hill AVS, Maitland K, Oppenheimer SJ, Parks T, Robson K, Williams TN, Kennett DJ, Mentzer AJ, Pinhasi R, and Reich D. 2018. Population Turnover in Remote Oceania Shortly after Initial Settlement. Current Biology. doi:https://doi.org/10.1016/j.cub.2018.02.051.

Llamas B, Fehren-Schmitz L, Valverde G, Soubrier J, Mallick S, Rohland N, Nordenfelt S, Valdiosera C, Richards SM, Rohrlach A, Romero MIB, Espinoza IF, Cagigao

ET, Jiménez LW, Makowski K, Reyna ISL, Lory JM, Torrez JAB, Rivera MA, Burger RL, Ceruti MC, Reihard J, Wells RS, Politis G, Santoro GM, Standen VG, Smith C, Reich D, Ho SYW, Cooper A, and Haak W. 2016. Ancient mitochondrial DNA provides high-resolution time scale of the peopling of the Americas. Science Advances 2(4): e1501385.

Llamas B, Valverde G, Fehren-Schmitz L, Weyrich LS, Cooper A, and Haak W. 2017. From the field to the laboratory: Controlling DNA contamination in human ancient DNA research in the high-throughput sequencing era. STAR: Science and Technology of Archaeological Research 3(1): 1-14.

Lobban RA. 2003. Historical Dictionary of Ancient and Medieval Nubia (Vol. 10). Oxford: Scarecrow Press.

Loogväli E-L, Roostalu U, Malyarchuk BA, Derenko MV, Kivisild T, Metspalu E, Tambets K, Reidla M, Tolk H-V, Parik J, Pennarun E, Laos S, Lunkina A, Golubenko M, Barać L, Peričić M, Balanovsky OP, Gusar V, Khusnutdinova EK, Stepanov V, Puzyrev V, Rudan P, Balanovska EV, Grechanina E, Richard C, Moisan J-P, Chaventré A, Anagnou NP, Pappa KI, Michalodimitrakis EN, Claustres M, Gölge M, Mikerezi I, Usanga E, and Villems R. 2004. Disuniting Uniformity: A Pied Cladistic Canvas of mtDNA Haplogroup H in Eurasia. Molecular Biology and Evolution 21(11): 2012-2021.

Lovell A, Moreau C, Yotova V, Xiao F, Bourgeois S, Gehl D, Bertranpetit J, Schurr E, and Labuda D. 2005. Ethiopia: between Sub-Saharan Africa and Western Eurasia. Annals of Human Genetics 69(3): 275-287.

Luis JR, Rowold DJ, Regueiro M, Caeiro B, Cinnioğlu C, Roseman C, Underhill PA, Cavalli-Sforza LL, and Herrera RJ. 2004. The Levant versus the Horn of Africa: Evidence for Bidirectional Corridors of Human Migrations. The American Journal of Human Genetics 74(3): 532-544.

Maca-Meyer N, González AM, Pestano J, Flores C, Larruga JM, and Cabrera VM. 2003. Mitochondrial DNA transit between West Asia and North Africa inferred from U6 phylogeography. BMC Genetics 4(1): 15.

Macaulay V, Richards M, Hickey E, Vega E, Cruciani F, Guida V, Scozzari R, Bonné-Tamir B, Sykes B, and Torroni A. 1999. The emerging tree of West Eurasian mtDNAs: a synthesis of control-region sequences and RFLPs. The American Journal of Human Genetics 64(1): 232-249.

Majumder PP. 2010. The Human Genetic History of South Asia. Current Biology 20(4): R184-R187.

Malécot G. 1968. The Mathematics of Heredity. San Francisco: Freeman.

Malmström H, Gilbert MTP, Thomas MG, Brandström M, Storå J, Molnar P, Andersen PK, Bendixen C, Holmlund G, Götherstrom A, and Willerslev E. 2009. Ancient DNA reveals lack of continuity between neolithic hunter-gatherers and contemporary Scandinavians. Current Biology 19(20): 1758-1762.

Malmström H, Linderholm A, Lidén K, Storå J, Molnar P, Holmlund G, Jakobsson M, and Götherstrom A. 2010. High frequency of lactose intolerance in a prehistoric hunter-gatherer population in northern Europe. BMC Evolutionary Biology 10(1): 89.

Malmström H, Svensson EM, Gilbert MTP, Willerslev E, Götherström A, and Holmlund G. 2007. More on contamination: The use of asymmetric molecular behavior to identify authentic ancient human DNA. Molecular Biology and Evolution 24(4): 998-1004.

Malyarchuk B, Derenko M, Grzybowski T, Perkova M, Rogalla U, Vanecek T, and Tsybovsky I. 2010. The Peopling of Europe from the Mitochondrial Haplogroup U5 Perspective. PLoS One 5(4): e10285.

Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, Howard E, Shendure J, and Turner DJ. 2010. Target-enrichment strategies for next-generation sequencing. Nature methods 7(2): 111-118.

Marchani EE, Watkins WS, Bulayeva K, Harpending HC, and Jorde LB. 2008. Culture creates genetic structure in the Caucasus: Autosomal, mitochondrial, and Y-chromosomal variation in Daghestan. BMC Genetics 9(1): 1.

Mardis ER. 2008. Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics 9: 387-402.

Margulies M, Egholm M, Altman WE, Attiya S, Bader JS, Bemben LA, Berka J, Braverman MS, Chen Y-J, Chen Z, Detwell SB, Du L, Fierro JM, Gomes XV, Godwain BC, He W, Helgesen S, Ho CH, Irzyk GP, Jando SC, Alenquer MLI, Jarvie TP, Jirage KB, Kim J-B, Knight JR, Lanza JR, Leamon JH, Lefkowitz SM, Lei M, Li J, Lohman KL, Lu H, Makhijani VB, McDade KE, McKenna MP, Myers EW, Nickerson E, Nobile JR, Plant R, Puc BP, Ronan MT, Roth GT, Sarkis GJ, Simons JF, Simpson JW, Srinivasan M, Tartaro KR, Tomasz A, Vogt KA, Volkmer GA, Wang SH, Wang Y, Weiner MP, Yu P, Begley RF, and Rothberg JM. 2005. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437(7057): 376-380.

Maricic T, and Pääbo S. 2009. Optimization of 454 sequencing library preparation from small amounts of DNA permits sequence determination of both DNA strands. BioTechniques 46(1): 51.

Maricic T, Whitten M, and Pääbo S. 2010. Multiplexed DNA sequence capture of mitochondrial genomes using PCR products. PLoS One 5(11): e14004.

Marshall F, and Hildebrand E. 2002. Cattle before crops: The beginnings of food production in Africa. Journal of World Prehistory 16(2): 99-143.

Martin DL, Armelagos GJ, Goodman AH, and Van Gerven DP. 1984. The effects of socioeconomic change in prehistoric Africa: Sudanese Nubia as a case study. In: Cohen MN, and Armelagos GJ (editors). Paleopathology at the Origins of Agriculture. Gainesville: University Press of Florida. p.193-214

Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal 17(1): 10-12.

Martiniano R, Caffell A, Holst M, Hunter-Mann K, Montgomery J, Müldner G, McLaughlin RL, Teasdale MD, van Rheenen W, Veldink JH, van den Berg LH, Hardiman O, Carroll M, Roskams S, Oxley J, Morgan C, Thomas MG, Barnes I, McDonnell C, Collins MJ, and Bradley DG. 2016. Genomic signals of migration and continuity in Britain before the Anglo-Saxons. Nature Communications 7: 10326.

Mathieson I, Alpaslan Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N, Mallick S, Olalde I, Broomandkhoshbacht N, Cheronet O, Fernandes D, Ferry M, Gamarra B, González Fortes G, Haak W, Harney E, Krause-Kyora B, Kucukkalipci I, Michel M, Mittnik A, Nägele K, Novak M, Oppenheimer J, Patterson N, Pfrengle S, Sirak K, Stewardson K, Vai S, Alexandrov S, Alt KW, Andreescu R, Antonović D, Ash A, Atanassova N, Bacvarov K, Balázs Gusztáv M, Bocherens H, Bolus M, Boroneanţ A, Boyadzhiev Y, Budnik A, Burmaz J, Chohadzhiev S, Conard NJ, Cottiaux R, Čuka M, Cupillard C, Drucker DG, Elenski N, Francken M, Galabova B, Ganetovski G, Gely B, Hajdu T, Handzhyiska V, Harvati K, Higham T, Iliev S, Janković I, Karavanić I, Kennett DJ, Komšo D, Kozak A, Labuda D, Lari M, Lazar C, Leppek M, Leshtakov K, Lo Vetro D, Los D, Lozanov I, Malina M, Martini F, McSweeney K, Meller H, Menđušić M, Mirea P, Moiseyev V, Petrova V, Price TD, Simalcsik A, Sineo L, Šlaus M, Slavchev V, Stanev P, Starović A, Szeniczey T, Talamo S, Teschler-Nicola M, Thevenet C, Valchev I, Valentin F, Vasilyev S, Veljanovska F, Venelinova S, Veselovskaya E, Viola B, Virag C, Zaninović J, Zäuner S, Stockhammer PW, Catalano G, Krauß R, Caramelli D, Zarina G, Gaydarska B, Lillie M, Nikitin AG, Potekhina I, Papathanasiou A, Borić D, Bonsall C, Krause J, Pinhasi R, and Reich D. 2018. The Genomic History Of Southeastern Europe. Nature 555(7695): 197-203.

Mathieson I, Lazaridis I, Rohland N, Mallick S, Patterson N, Alpaslan Roodenberg S, Harney E, Stewardson K, Fernandes D, Novak M, Sirak K, Gamba C, Jones ER, Llamas B, Dryomov S, Pickrell J, Arsuaga JL, Bermúdez de Castro JM, Carbonell E, Gerritsen F, Khokhlov A, Kuznetsov P, Lozano M, Meller H, Mochalov O, Moiseyev V, Rojo Guerra MA, Roodenberg J, Vergès JM, Krause J, Cooper A,

Alt KW, Brown D, Anthony D, Lalueza-Fox C, Haak W, Pinhasi R, and Reich D. 2015. Genome-wide patterns of selection in 230 ancient Eurasians. Nature 528(7583): 499-503.

Matisoo-Smith E, and Horsburgh KA. 2012. DNA for Archaeologists. Walnut Creek: Left Coast Press, Inc.

The International Human Genome Mapping Consortium. 2001. A physical map of the human genome. Nature 409(6822): 934-941.

Meindl RS, Lovejoy CO, Mensforth RP, and Walker RA. 1985. A revised method of age determination using the os pubis, with a review and tests of accuracy of other current methods of pubic symphyseal aging. American Journal of Physical Anthropology 68(1): 29-45.

Mendez FL, Karafet TM, Krahn T, Ostrer H, Soodyall H, and Hammer MF. 2011. Increased resolution of Y chromosome haplogroup T defines relationships among populations of the Near East, Europe, and Africa. Human Biology 83(1): 39-53.

Menozzi P, Piazza A, and Cavalli-Sforza LL. 1978. Synthetic maps of human gene frequencies in Europeans. Science 201(4358): 786-792.

Meselson M, and Stahl FW. 1958. THE REPLICATION OF DNA IN ESCHERICHIA COLI. Proceedings of the National Academy of Sciences 44(7): 671-682.

Messina F, Scano G, Contini I, Martínez-Labarga C, De Stefano GF, and Rickards O. 2017. Linking between genetic structure and geographical distance: Study of the maternal gene pool in the Ethiopian population. Annals of Human Biology 44(1): 53-69.

Metzker ML. 2010. Sequencing technologies—the next generation. Nature Reviews Genetics 11(1): 31-46.

Meyer M, Fu Q, Aximu-Petri A, Glocke I, Nickel B, Arsuaga J-L, Martínez I, Gracia A, de Castro JMB, Carbonell E, and Pääbo S. 2014. A mitochondrial genome sequence of a hominin from Sima de los Huesos. Nature 505(7483): 403.

Meyer M, and Kircher M. 2010. Illumina sequencing library preparation for highly multiplexed target capture and sequencing. Cold Spring Harbor Protocols 2010(6): pdb.prot5448.

Meyer M, Kircher M, Gansauge M-T, Li H, Racimo F, Mallick S, Schraiber JG, Jay F, Prüfer K, de Filippo C, Sudmant PH, Alkan C, Fu Q, Do R, Rohland N, Randon A, Siebauer M, Green RE, Bryc K, Briggs AW, Stenzel U, Dabney J, Shendure J, Kitzman J, Hammer MF, Shunkov MV, Derevianko AP, Patterson N, Andrés AM, Eichler EE, Slatkin M, Reich D, Kelso J, and Pääbo S. 2012. A high-

coverage genome sequence from an archaic Denisovan individual. Science 338(6104): 222-226.

Mielke JH, Konigsberg LW, and Relethford J. 2006. Human Biological Variation. Oxford: Oxford University Press.

Mikkelsen M, Fendt L, Röck AW, Zimmermann B, Rockenbauer E, Hansen AJ, Parson W, and Morling N. 2012. Forensic and phylogeographic characterisation of mtDNA lineages from Somalia. International Journal of Legal Medicine 126(4): 573-579.

Mishmar D, Ruiz-Pesini E, Golik P, Macaulay V, Clark AG, Hosseini S, Brandon M, Easley K, Chen E, Brown MD, Sukernik RI, Olkers A, and Wallace DC. 2003. Natural selection shaped regional mtDNA variation in humans. Proceedings of the National Academy of Sciences 100(1): 171-176.

Mitchell AA, Zwick ME, Chakravarti A, and Cutler DJ. 2004. Discrepancies in dbSNP confirmation rates and allele frequency distributions from varying genotyping error rates and patterns. Bioinformatics 20(7): 1022-1032.

Mitchell D, Willerslev E, and Hansen A. 2005. Damage and repair of ancient DNA. Mutation Research 571(1): 265-276.

Mittler DM, and Van Gerven DP. 1994. Developmental, diachronic, and demographic analysis of cribra orbitalia in the medieval Christian populations of Kulubnarti. American Journal of Physical Anthropology 93(3): 287-297.

Molto JE, Loreille O, Mallott EK, Malhi RS, Fast S, Daniels-Higginbotham J, Marshall C, and Parr R. 2017. Complete Mitochondrial Genome Sequencing of a Burial from a Romano–Christian Cemetery in the Dakhleh Oasis, Egypt: Preliminary Indications. Genes 8(10): 262.

Moore JA, Swedlund AC, and Armelagos GJ. 1975. The Use of Life Tables in Paleodemography. Memoirs of the Society for American Archaeology (30): 57-70.

Morant G. 1935. A study of predynastic Egyptian skulls from Badari based on measurements taken by Miss BN Stoessiger and Professor DE Derry. Biometrika 27(3/4): 293-309.

Morozova I, Flegontov P, Mikheyev AS, Bruskin S, Asgharian H, Ponomarenko P, Klyuchnikov V, ArunKumar G, Prokhortchouk E, Gankin Y, Rogaev E, Nikolsky Y, Baranova A, Elhaik E, and Tatarinova TV. 2016. Toward high-resolution population genomics using archaeological samples. DNA Research, 23(4): 295-310.

Mukherjee R, Trevor J, and Rao C. 1955. The Ancient Inhabitants of Jebel Moya. Cambridge: Cambridge University Press.

Mullis K, Faloona F, Scharf S, Saiki R, Horn G, and Erlich H. 1986. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harbor Symposia on Quantitative Biology 51: 263-273.

Mullis KB, and Faloona FA. 1987. Specific synthesis of DNA in vitro via a polymerase-catalyzed chain reaction. Methods in Enzymology 155: 335.

Musumeci L, Arthur JW, Cheung FSG, Hoque A, Lippman S, and Reichardt JKV. 2010. Single Nucleotide Differences (SNDs) in the dbSNP Database May Lead to Errors in Genotyping and Haplotyping Studies. Human Mutation 31(1): 67-73.

Myres NM, Rootsi S, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Pshenichnov A, Yunusbayev B, Balanovsky O, Balanovska E, Rudan P, Baldovic M, Herrera RJ, Chiaroni J, Di Cristofaro J, Villems R, Kivisild T, and Underhill PA. 2010. A major Y-chromosome haplogroup R1b Holocene era founder effect in Central and Western Europe. European Journal of Human Genetics 19: 95.

Nasidze I, Quinque D, Ozturk M, Bendukidze N, and Stoneking M. 2005. MtDNA and Y-chromosome Variation in Kurdish Groups. Annals of Human Genetics 69(4): 401-412.

Navarro-Gomez D, Leipzig J, Shen L, Lott M, Stassen AP, Wallace DC, Wiggs JL, Falk MJ, van Oven M, and Gai X. 2015. Phy-Mer: a novel alignment-free and reference-independent mitochondrial haplogroup classifier. Bioinformatics 31(8): 1310-1312.

Neves da Nova Fernandes VC. 2013. High-resolution characterization of genetic markers in the Arabian Peninsula and Near East. PhD Dissertation: University of Leeds.

Nogueiro I, Manco L, Gomes V, Amorim A, and Gusmão L. 2010. Phylogeographic analysis of paternal lineages in NE Portuguese Jewish communities. American Journal of Physical Anthropology 141(3): 373-381.

Non A. 2010. Analyses of Genetic Data Within an Interdisciplinary Framework to Investigate Recent Human Evolutionary History and Complex Disease. PhD Thesis: University of Florida.

Non AL, Al-Meeri A, Raaum RL, Sanchez LF, and Mulligan CJ. 2011. Mitochondrial DNA reveals distinct evolutionary histories for Jewish populations in Yemen and Ethiopia. American Journal of Physical Anthropology 144(1): 1-10.

Noonan JP, Hofreiter M, Smith D, Priest JR, Rohland N, Rabeder G, Krause J, Detter JC, Pääbo S, and Rubin EM. 2005. Genomic sequencing of Pleistocene cave bears. Science 309(5734): 597-599.

Novembre J, Johnson T, Bryc K, Kutalik Z, Boyko AR, Auton A, Indap A, King KS, Bergmann S, Nelson MR, Stephens M, and Bustamante CD. 2008. Genes mirror geography within Europe. Nature 456(7218): 98-101.

Nystrom KC. 2006. Late Chachapoya population structure prior to Inka conquest. American Journal of Physical Anthropology 131(3): 334-342.

Olalde I, Allentoft ME, Sánchez-Quinto F, Santpere G, Chiang CW, DeGiorgio M, Prado-Martinez J, Rodríguez JA, Rasmussen S, Quilez J, Ramírez O, Marigorta UM, Fernández-Callejo M, Prada ME, Encinas JMV, Nielsen R, Netea MG, Novembre J, Sturm RA, Sabeti P, Marquès-Bonet T, Navarro A, Willerslev E, and Lalueza-Fox C. 2014. Derived immune and ancestral pigmentation alleles in a 7,000-year-old Mesolithic European. Nature 507(7491): 225-228.

Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Rohland N, Mallick S, Booth T, Szécsény-Nagy A, Mittnik A, Altena E, Lipson M, Lazaridis I, Patterson N, Broomandkhoshbacht N, Diekmann Y, Faltyskova Z, Fernandes D, Ferry M, Harney E, de Knijff P, Michel M, Oppenheimer J, Stewardson K, Barclay A, Alt KW, Fernández AA, Bánffy E, Bernabò-Brea M, Billoin D, Blasco C, Bonsall C, Bonsall L, Allen T, Büster L, Carver S, Navarro LC, Craig OE, Cook GT, Cunliffe B, Denaire A, Dinwiddy KE, Dodwell N, Ernée M, Evans C, Kuchařik M, Farre JF, Fokkens H, Fowler C, Gazenbeek M, Pena RG, Haber-Uriate M, Haduch E, Hey G, Jowett N, Knowles T, Massy K, Pfrengle S, Lefranc P, Lemercier O, Lefebvre A, Maurandi JL, Majó T, McKinley JI, McSweeney K, Gusztáv MB, Modi A, Kulcsár G, Kiss V, Czene A, Patay R, Endrödi A, Köhler K, Hajdu T, Cardoso JL, Liesau C, Pearson MP, Włodarczak P, Price TD, Prieto P, Rey P-J, Ríos P, Risch R, Guerra MAR, Schmitt A, Serralongue J, Silva AM, Smrčka V, Vergnaud L, Zilhão J, Caramelli D, Higham T, Heyd V, Sheridan A, Sjögren K-G, Thomas MG, Stockhammer PW, Pinhasi R, Krause J, Haak W, Barnes I, Lalueza-Fox C, and Reich D. 2017. The Beaker Phenomenon And The Genomic Transformation Of Northwest Europe. Nature 555: 190-196.

Olivo PD, Van de Walle MJ, Laipis PJ, and Hauswirth WW. 1983. Nucleotide sequence evidence for rapid genotypic shifts in the bovine mitochondrial DNA D-loop. Nature 306(5941): 400-402.

Oota H, Settheetham-Ishida W, Tiwawech D, Ishida T, and Stoneking M. 2001. Human mtDNA and Y-chromosome variation is correlated with matrilocal versus patrilocal residence. Nature genetics 29(1): 20.

Orlando L, Gilbert MTP, and Willerslev E. 2015. Reconstructing ancient genomes and epigenomes. Nature Reviews Genetics 16(7): 395-408.

Orlando L, Ginolhac A, Zhang G, Froese D, Albrechtsen A, Stiller M, Schubert M, Cappellini E, Petersen B, Moltke I, Johnson PLF, Fumagalli M, Vilstrup JT, Raghavan M, Korneliussen T, Malaspinas A-S, Vogt J, Szklarczyk D, Kelstrup CD, Vinther J, Dolocan A, Stenderup J, Velazquez AMV, Cahill J, Rasmussen M, Wang X, Min J, Zazula GD, Seguin-Orlando A, Mortensen C, Magnussen K, Thompson JF, Weinstock J, Gregersen K, Røed KH, Eisenmann V, Rubin CJ, Miller DC, Antczak DF, Bertelsen MF, Brunak S, Al-Rasheid KAS, Ryder O, Andersson L, Mundy J, Krogh A, Gilbert MTP, Kjær K, Sicheritz-Ponten T, Jensen LJ, Olsen JV, Hofreiter M, Nielsen R, Shapiro B, Wang J, and Willerslev E. 2013. Recalibrating Equus evolution using the genome sequence of an early Middle Pleistocene horse. Nature 499(7456): 74-78.

O'Rourke DH, Hayes MG, and Carlyle SW. 2000. Ancient DNA studies in physical anthropology. Annual Review of Anthropology (29): 217-242.

Overballe-Petersen S, Orlando L, and Willerslev E. 2012. Next-generation sequencing offers new insights into DNA degradation. Trends in Biotechnology 30(7): 364-368.

Pääbo S. 1985. Molecular cloning of Ancient Egyptian mummy DNA. Nature 314(6012): 644-645.

Pääbo S. 1989. Ancient DNA: Extraction, characterization, molecular cloning, and enzymatic amplification. Proceedings of the National Academy of Sciences 86(6): 1939-1943.

Pääbo S, Gifford JA, and Wilson AC. 1988. Mitochondrial DNA sequences from a 7000-year old brain. Nucleic acids research 16(20): 9775-9787.

Pääbo S, Higuchi RG, and Wilson AC. 1989. Ancient DNA and the Polymerase Chain Reaction: The Emerging Field of Molecular Archaeology. The Journal of Biological Chemistry 264(17): 9709-9712.

Pääbo S, Poinar H, Serre D, Jaenicke-Després V, Hebler J, Rohland N, Kuch M, Krause J, Vigilant L, and Hofreiter M. 2004. Genetic analyses from ancient DNA. Annual Review of Genetics 38: 645-679.

Pagani L, Kivisild T, Tarekegn A, Ekong R, Plaster C, Gallego Romero I, Ayub Q, Mehdi Q, Thomas MG, Luiselli D, Bekele E, Bradman N, Balding DJ, and Tyler-Smith C. 2012. Ethiopian Genetic Diversity Reveals Linguistic Stratification and Complex Influences on the Ethiopian Gene Pool. The American Journal of Human Genetics 91(1): 83-96.

Pagani L, Schiffels S, Gurdasani D, Danecek P, Scally A, Chen Y, Xue Y, Haber M, Ekong R, Oljira T, Mekonnen E, Luiselli D, Bradman N, Bekele E, Zalloua P, Durbin R, Kivisild T, and Tyler-Smith C. 2015. Tracing the route of modern

humans out of Africa by using 225 human genome sequences from Ethiopians and Egyptians. The American Journal of Human Genetics 96(6): 986-991.

Pala M, Olivieri A, Achilli A, Accetturo M, Metspalu E, Reidla M, Tamm E, Karmin M, Reisberg T, Kashani BH, Perego UA, Carossa V, Gandini F, Pereira JB, Soares P, Angerhofer N, Rychkov S, Al-Zahery N, Carelli V, Sanati MH, Moushmand M, Hatina J, Macaulay V, Pereira L, Woodward SR, Davies W, Gamble C, Baird D, Semino O, Villems R, Torroni A, and Richards MB. 2012. Mitochondrial DNA Signals of Late Glacial Recolonization of Europe from Near Eastern Refugia. The American Journal of Human Genetics 90(5): 915-924.

Passarino G, Semino O, Quintana-Murci L, Excoffier L, Hammer M, and Santachiara-Benerecetti AS. 1998. Different genetic components in the Ethiopian population, identified by mtDNA and Y-chromosome polymorphisms. The American Journal of Human Genetics 62(2): 420-434.

Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, Genschoreck T, Webster T, and Reich D. 2012. Ancient admixture in human history. Genetics 192(3): 1065-1093.

Patterson N, Price AL, and Reich D. 2006. Population structure and eigenanalysis. PLoS Genetics 2(12): e190.

Pereira L, Macaulay V, Torroni A, Scozzari R, Prata M-J, and Amorim A. 2001. Prehistoric and historic traces in the mtDNA of Mozambique: insights into the Bantu expansions and the slave trade. Annals of Human Genetics 65(5): 439-458.

Pereira L, Richards M, Goios A, Alonso A, Albarrán C, Garcia O, Behar DM, Gölge M, Hatina J, Al-Gazali L, Bradley DG, Macaulay V, and Amorim A. 2005. High-resolution mtDNA evidence for the late-glacial resettlement of Europe from an Iberian refugium. Genome Research 15(1): 19-24.

Phenice TW. 1969. A newly developed visual method of sexing the os pubis. American Journal of Physical Anthropology 30(2): 297-301.

Pickrell JK, Patterson N, Barbieri C, Berthold F, Gerlach L, Güldemann T, Kure B, Mpoloka SW, Nakagawa H, Naumann C, Lipson M, Loh P-R, Lachance J, Mountain J, Bustamante CD, Berger B, Tishkoff SA, Henn BM, Stoneking M, Reich D, and Pakendorf B. 2012. The genetic prehistory of southern Africa. Nature Communications 3: 1143.

Pickrell J, and Reich D. 2014. Towards a new history and geography of human genes informed by ancient DNA. Trends in Genetics 30(9): 377-389.

Pilli E, Modi A, Serpico C, Achilli A, Lancioni H, Lippi B, Bertoldi F, Gelichi S, Lari M, and Caramelli D. 2013. Monitoring DNA Contamination in Handled vs. Directly Excavated Ancient Human Skeletal Remains. PLoS One 8(1): e52524.

Pinhasi R, Fernandes D, Sirak K, Novak M, Connell S, Alpaslan-Roodenberg S, Gerritsen F, Moiseyev V, Gromov A, Raczky P, Anders A, Pietrusewsky M, Rollefson G, Jovanovic M, Trinhhoang H, Bar-Oz G, Oxenham M, Matsumura H, and Hofreiter M. 2015. Optimal Ancient DNA Yields from the Inner Ear Part of the Human Petrous Bone. PLoS One 10(6): e0129102.

Poinar HN. 2003. The top 10 list: Criteria of authenticity for DNA from ancient and forensic samples. International Congress Series 1239: 575-579.

Posth C, Renaud G, Mittnik A, Drucker DG, Rougier H, Cupillard C, Valentin F, Thevenet C, Furtwängler A, Wißing C, Francken M, Malina M, Bolus M, Lari M, Gigli E, Capecchi G, Crevecoeur I, Beauval C, Flas D, Germonpré M, van der Plicht J, Cottiaux R, Gély B, Ronchitelli A, Wehrberger K, Grigorescu D, Svoboda J, Semal P, Caramelli D, Bocherens H, Harvati K, Conrad NJ, Haak W, Powell A, and Krause J. 2016. Pleistocene mitochondrial genomes suggest a single major dispersal of non-Africans and a Late Glacial population turnover in Europe. Current Biology 26(6): 827-833.

Prowse TL, and Lovell NC. 1995. Biological continuity between the A-and C-groups in lower Nubia: Evidence from cranial non-metric traits. International Journal of Osteoarchaeology 5(2): 103-114.

Prüfer K, Racimo F, Patterson N, Jay F, Sankararaman S, Sawyer S, Heinze A, Renaud G, Sudmant PH, de Filippo C, Li H, Mallick S, Dannemann M, Fu Q, Kircher M, Kuhlwilm M, Lachmann M, Meyer M, Ongyerth M, Siebauer M, Theunert C, Tandon A, Moorjani P, Pickrell J, Mullikin JC, Vohr SH, Green RE, Hellmann I, Johnson PLF, Blanche H, Cann H, Kitzman JO, Shendure J, Eichler EE, Lein ES, Bakken TE, Golovanova LV, Doronichev VB, Shunkov MV, Derevianko AP, Viola B, Slatkin M, Reich D, Kelso J, and Pääbo S. 2014. The complete genome sequence of a Neanderthal from the Altai Mountains. Nature 505(7481): 43-49.

Pruvost M, Schwarz R, Correia VB, Champlot S, Braguier S, Morel N, Fernandez-Jalvo Y, Grange T, and Geigl E-M. 2007. Freshly Excavated Fossil Bones Are Best for Amplification of Ancient DNA. Proceedings of the National Academy of Sciences 104(3): 739-744.

Qin P, and Stoneking M. 2015. Denisovan Ancestry in East Eurasian and Native American Populations. Molecular Biology and Evolution 32(10): 2665-2674.

Quail MA, Kozarewa I, Smith F, Scally A, Stephens PJ, Durbin R, Swerdlow H, and Turner DJ. 2008. A large genome center's improvements to the Illumina sequencing system. Nature Methods 5(12): 1005-1010.

Quail MA, Smith M, Coupland P, Otto TD, Harris SR, Connor TR, Bertoni A, Swerdlow HP, and Gu Y. 2012. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. BMC Genomics 13(1): 341.

Queller DC, and Goodnight KF. 1989. Estimating relatedness using genetic markers. Evolution 43(2): 258-275.

Quintana-Murci L, Chaix R, Wells RS, Behar DM, Sayar H, Scozzari R, Rengo C, Al-Zahery N, Semino O, Santachiara-Benerecetti AS, Coppa A, Ayub Q, Mohyuddin A, Tyler-Smith C, Medhi SQ, Torroni A, and McElreavey K. 2004. Where West Meets East: The Complex mtDNA Landscape of the Southwest and Central Asian Corridor. The American Journal of Human Genetics 74(5): 827-845.

Quintana-Murci L, Krausz C, Zerjal T, Sayar SH, Hammer MF, Mehdi SQ, Ayub Q, Qamar R, Mohyuddin A, Radhakrishna U, Jobling MA, Tyler-Smith C, and McElreavey K. 2001. Y-chromosome lineages trace diffusion of people and languages in southwestern Asia. The American Journal of Human Genetics 68(2): 537-542.

Raghavan M, Skoglund P, Graf KE, Metspalu M, Albrechtsen A, Moltke I, Rasmussen S, Stafford Jr TW, Orlando L, Metspalu E, Karmin M, Tambets K, Rootsi S, Mägi R, Campos PF, Balanovska E, Balanovsky O, Khusnutdinova E, Litvinov S, Osipova LP, Fedorova SA, Voevoda MI, DeGiorgio M, Sicheritz-Ponten T, Brunak S, Demeshchenko S, Kivisild T, Villems R, Nielsen R, Jakobsson M, and Willerslev E. 2014. Upper Palaeolithic Siberian genome reveals dual ancestry of Native Americans. Nature 505: 87-91.

Raghavan M, Steinrücken M, Harris K, Schiffels S, Rasmussen S, DeGiorgio M, Albrechtsen A, Valdiosera C, Ávila-Arcos MC, Malaspinas A-S, Eriksson A, Moltke I, Metspalu M, Homburger JR, Wall J, Cornejo OE, Moreno-Mayar JV, Korneliussen TS, Pierre T, Rasmussen M, Campos PF, Damgaard PdB, Allentoft ME, Lindo J, Metspalu E, Rodríguez-Varela R, Mansilla J, Henrickson C, Seguin-Orlando A, Malmström H, Stafford Jr T, Shringarpure SS, Moreno-Estrada A, Karmin M, Tambets K, Bergström A, Xue Y, Warmuth V, Friend AD, Singarayer J, Valdes P, Balloux F, Leboreiro I, Vera JL, Rangel-Villalobos H, Pettener D, Luiselli D, Davis LG, Heyer E, Zollikofer CPE, Ponce de León MS, Smith CI, Grimes V, Pike K-A, Deal M, Fuller BT, Arriaza B, Standen V, Luz MF, Ricaut F, Guidon N, Osipova L, Voevoda MI, Posukh OL, Balanovsky O, Lavryashina M, Bogynov Y, Khusnutdinova E, Gubina M, Balanovska E, Fedorova S, Litvinov S, Malyarchuk B, Derenko M, Mosher MJ, Archer D, Cybulski J, Petzelt B, Mitchell J, Worl R, Norman PJ, Parham P, Kemp BM, Kivisild T, Tyler-Smith C, Sandhu MS, Crawford M, Villems R, Smith DG, Waters MR, Goebel T, Johnson JR, Malhi RS, Jakobsson M, Meltzer DJ, Manica A, Durbin R, Bustamante CD, Song YS, Nielsen R, and Willerslev E. 2015. Genomic evidence

for the Pleistocene and recent population history of Native Americans. Science, 349(6250): aab3884.

Rando J, Pinto F, Gonzalez A, Hernandez M, Larruga J, Cabrera V, and Bandelt HJ. 1998. Mitochondrial DNA analysis of Northwest African populations reveals genetic exchanges with European, Near-Eastern, and sub-Saharan populations. Annals of Human Genetics 62(6): 531-550.

Rasmussen M, Li Y, Lindgreen S, Pedersen JS, Albrechtsen A, Moltke I, Metspalu M, Metspalu E, Kivisild T, Gupta R, Bertalan M, Nielsen K, Gilbert MTP, Wang Y, Raghaban M, Campos PF, Kamp HM, Wilson AS, Gledhill A, Tridico S, Bunce M, Lorensen ED, Binladen J, Guo X, Zhao J, Zhang X, Zhang H, Li Z, Chen M, Orlando L, Kristiansen K, Bak M, Tommerup N, Bendixen C, Pierre TL, Grønnow B, Meldgaard M, Andreasen C, Fedorova SA, Osipova LP, Higham TFG, Ramsey CB, Hansen TvO, Nielsen FC, Crawford MH, Brunak S, Sicheritz-Pontén, Villems R, Nielsen R, Krogh A, Wang J, and Willerslev E. 2010. Ancient human genome sequence of an extinct Palaeo-Eskimo. Nature 463(7282): 757-762.

Regueiro M, Cadenas A, Gayden T, Underhill PA, and Herrera R. 2006. Iran: Tricontinental Nexus for Y-Chromosome Driven Migration. Human Heredity 61(3): 132-143.

Reguig A, Harich N, Barakat A, and Rouba H. 2014. Phylogeography of E1b1b1b-M81 Haplogroup and Analysis of Its Subclades in Morocco. Human Biology 86(2): 105-112.

Reich D, Price AL, and Patterson N. 2008. Principal component analysis of genetic data. Nature Genetics 40(5): 491-492.

Reich D, Thangaraj K, Patterson N, Price AL, and Singh L. 2009. Reconstructing Indian population history. Nature 461(7263): 489-494.

Relethford JH. 1994. Craniometric variation among modern human populations. American Journal of Physical Anthropology 95(1): 53-62.

Relethford JH. 2004. Boas and beyond: migration and craniometric variation. American Journal of Human Biology 16(4): 379-386.

Relethford JH. 2012. Human Population Genetics. Hoboken: Wiley-Blackwell.

Relethford JH, Stern MP, Gaskill SP, and Hazuda HP. 1983. Social class, admixture, and skin color variation in Mexican-Americans and Anglo-Americans living in San Antonio, Texas. American Journal of Physical Anthropology 61(1): 97-102.

Richards MB, Macaulay VA, Bandelt H-J, and Sykes BC. 1998. Phylogeography of mitochondrial DNA in western Europe. Annals of Human Genetics 62(3): 241-260.

Richards M, Macaulay V, Hickey E, Vega E, Sykes B, Guida V, Rengo C, Sellitto D, Cruciani F, Kivisild T, Villems R, Thomas M, Rychkov S, Rychkov O, Rychkov Y, Gölge M, Dimitrov D, Hill E, Bradley D, Romano V, Calì F, Vona G, Demaine A, Papiha S, Triantaphyllidis C, Stefanescu G, Hatina J, Belledi M, Di Rienzo A, Novelletto A, Oppenheim A, Nørby S, Al-Zaheri N, Santachiara-Benerecetti S, Scozzari R, Torroni A, and Bandelt H-J. 2000. Tracing European founder lineages in the Near Eastern mtDNA pool. The American Journal of Human Genetics 67(5): 1251-1276.

Richards M, Rengo C, Cruciani F, Gratrix F, Wilson JF, Scozzari R, Macaulay V, and Torroni A. 2003. Extensive female-mediated gene flow from sub-Saharan Africa into near eastern Arab populations. The American Journal of Human Genetics 72(4): 1058-1064.

Ritte U, Neufeld E, Broit M, Shavid D, and Motro U. 1993. The Differences Among Jewish Communities – Maternal and Paternal Contributions. Journal of Molecular Evolution 37: 435-440.

Rizzi E, Lari M, Gigli E, De Bellis G, and Caramelli D. 2012. Ancient DNA studies: new perspectives on old samples. Genetics Selection Evolution 44: 21.

Rohland N, Harney E, Mallick S, Nordenfelt S, and Reich D. 2015. Partial uracil–DNA–glycosylase treatment for screening of ancient DNA. Philosophical Transactions of the Royal Society of London B: Biological Sciences 370(1660): 20130624.

Rohland N, and Hofreiter M. 2007a. Ancient DNA extraction from bones and teeth. Nature Protocols 2(7): 1756-1762.

Rohland N, and Hofreiter M. 2007b. Comparison and optimization of ancient DNA extraction. BioTechniques 42(3): 343.

Rohland N, and Reich D. 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Research 22(5): 939-946.

Roostalu U, Kutuev I, Loogväli E, Metspalu E, Tambets K, Reidla M, Khusnutdinova EK, Usanga E, Kivisild T, and Villems R. 2007. Origin and Expansion of haplogroup H, the Dominant Human Mitochondrial DNA Lineage in West Eurasia: The Near Eastern and Caucasian Perspective. Molecular Biology and Evolution 24(2): 436-448.

Rootsi S, Myres NM, Lin AA, Järve M, King RJ, Kutuev I, Cabrera VM, Khusnutdinova EK, Varendi K, Sahakyan H, Behar DM, Khusainova R, Balanovsky O,

Balanovska E, Rudan P, Yepiskoposyan L, Bahmanimehr A, Farjadian S, Kushniarevich A, Herrera RJ, Grugni V, Battaglia V, Nici C, Crobu F, Karachanak S, Kashani BH, Houshmand M, Sanati MH, Toncheva D, Lisa A, Semino O, Chiaroni J, Di Cristofaro J, Villems R, Kivisild T, and Underhill PA. 2012. Distinguishing the co-ancestries of haplogroup G Y-chromosomes in the populations of Europe and the Caucasus. European Journal of Human Genetics 20(12): 1275-1282.

Rosa A, Brehm A, Kivisild T, Metspalu E, and Villems R. 2004. MtDNA Profile of West Africa Guineans: Towards a Better Understanding of the Senegambia Region. Annals of Human Genetics 68(4): 340-352.

Roseman CC, and Weaver TD. 2004. Multivariate apportionment of global human craniometric diversity. American Journal of Physical Anthropology 125(3): 257-263.

Rosenberg NA, Pritchard JK, Weber JL, Cann HM, Kidd KK, Zhivotovsky LA, and Feldman MW. 2002. Genetic structure of human populations. Science 298(5602): 2381-2385.

Ross L. 2013. Nubia and Egypt 10,000 BC to 400 AD: From Pre-history to the Meroitic Period. Lewiston, NY: Edwin Mellen Press.

Rudney JD, and Greene DL. 1982. Interpopulation differences in the severity of early childhood stress in ancient Lower Nubia: Implications for hypotheses of X-group origins. Journal of Human Evolution 11(7): 559-565.

Ruffini G. 2012. Medieval Nubia: A Social and Economic History. Oxford: Oxford University Press.

Saiki RK, Gelfand DH, Stoffel S, Scharf SJ, Higuchi R, Horn GT, Mullis KB, and Erlich HA. 1988. Primer-directed enzymatic amplification of DNA with a thermostable DNA polymerase. Science 239(4839): 487-491.

Salamon M, Tuross N, Arensburg B, and Weiner S. 2005. Relatively well preserved DNA is present in the crystal aggregates of fossil bones. Proceedings of the National Academy of Sciences 102(39): 13783-13788.

Salas A, Richards M, De la Fe T, Lareu M-V, Sobrino B, Sánchez-Diz P, MaCaulay V, and Carracedo Á. 2002. The making of the African mtDNA landscape. The American Journal of Human Genetics 71(5): 1082-1111.

Salas A, Richards M, Lareu M-V, Scozzari R, Coppa A, Torroni A, Macaulay V, and Carracedo Á. 2004. The African Diaspora: Mitochondrial DNA and the Atlantic Slave Trade. American Journal of Human Genetics 74(3): 454-465.

Salih AOM. 2004. Archaeology and settlement in the Third Cataract region during the Medieval and Post-Medieval periods. AZANIA: Journal of the British Institute in Eastern Africa 39(1): 34-49.

Sandberg PA. 2012. Investigating Childhood Diet and Early Life History Events in the Archaeological Record Using Biogeochemical Techniques. PhD Dissertation: University of Colorado at Boulder.

Sandberg PA, Sponheimer M, Lee-Thorp J, and Van Gerven DP. 2014. Intra-tooth stable isotope analysis of dentine: A step toward addressing selective mortality in the reconstruction of life history in the archaeological record. American Journal of Physical Anthropology 155(2): 281-293.

Sandford MK, and Kissling GE. 1994. Multivariate analyses of elemental hair concentrations from a medieval Nubian population. American Journal of Physical Anthropology 95(1): 41-52.

Sandford MK, Van Gerven DP, and Meglen RR. 1983. Elemental hair analysis: new evidence on the etiology of cribra orbitalia in Sudanese Nubia. Human Biology 55(4): 831-844.

Sanger F, and Coulson AR. 1975. A rapid method for determining sequences in DNA by primed synthesis with DNA polymerase. Journal of Molecular Biology 94(3): 441-448.

Satoh M, and Kuroiwa T. 1991. Organization of multiple nucleoids and DNA molecules in mitochondria of a human cell. Experimental Cell Research 196(1): 137-140.

Saunier JL, Irwin JA, Strouss KM, Ragab H, Sturk KA, and Parsons TJ. 2009. Mitochondrial control region sequences from an Egyptian population sample. Forensic Science International: Genetics 3(3): e97-e103.

Säve-Söderbergh T. 1987. Temples and tombs of ancient Nubia: The international rescue campaign at Abu Simbel, Philae and other sites. London: Thames and Hudson.

Sawyer S, Krause J, Guschanski K, Savolainen V, and Pääbo S. 2012. Temporal patterns of nucleotide misincorporations and DNA fragmentation in ancient DNA. PLoS One 7(3): e34131.

Scheuer L, and Black S. 2000. Development and ageing of the juvenile skeleton. In: Cox, M, and Mays S. Human osteology in archaeology and forensic science. Cambridge: Cambridge University Press. p. 9-22.

Schlebusch CM, Malmström H, Günther T, Sjödin P, Coutinho A, Edlund H, Munters AR, Vicente M, Steyn M, Soodyall H, Lombard M, and Jakobsson M. 2017.

Southern African ancient genomes estimate modern human divergence to 350,000 to 260,000 years ago. Science 358(6363):aao6266.

Schrader S, Buzon M, and Irish J. 2014. Illuminating the Nubian 'Dark Age': A bioarchaeological analysis of dental non-metric traits during the Napatan Period. HOMO - Journal of Comparative Human Biology 65(4): 267-280.

Schubert M, Ginolhac A, Lindgreen S, Thompson JF, Al-Rasheid KA, Willerslev E, Krogh A, and Orlando L. 2012. Improving ancient DNA read mapping against modern reference genomes. BMC Genomics 13(1): 1.

Schuenemann VJ, Peltzer A, Welte B, van Pelt WP, Molak M, Wang C-C, Furtwängler A, Urban C, Reiter E, Nieselt K, Teßmann B, Francken M, Harvati K, Haak W, Schiffels S, and Krause J. 2017. Ancient Egyptian mummy genomes suggest an increase of Sub-Saharan African ancestry in post-Roman periods. Nature Communications 8: 15694.

Semino O, Magri C, Benuzzi G, Lin AA, Al-Zahery N, Battaglia V, Maccioni L, Triantaphyllidis C, Shen P, Oefner PJ, Zhivotovsky LA, King R, Torroni A, Cavalli-Sforza LL, Underhill PA, and Santachiara-Benerecetti AS. 2004. Origin, Diffusion, and Differentiation of Y-Chromosome Haplogroups E and J: Inferences on the Neolithization of Europe and Later Migratory Events in the Mediterranean Area. The American Journal of Human Genetics 74(5): 1023-1034.

Semino O, Santachiara-Benerecetti AS, Falaschi F, Cavalli-Sforza LL, and Underhill PA. 2002. Ethiopians and Khoisan share the deepest clades of the human Y-chromosome phylogeny. The American Journal of Human Genetics 70(1): 265-268.

Sengupta S, Zhivotovsky LA, King R, Mehdi S, Edmonds CA, Chow C-ET, Lin AA, Mitra M, Sil SK, Ramesh A, Usha Rani MV, Thakur CM, Cavalli-Sforza LL, Majumder PP, and Underhill PA. 2006. Polarity and Temporality of High-Resolution Y-Chromosome Distributions in India Identify Both Indigenous and Exogenous Expansions and Reveal Minor Genetic Influence of Central Asian Pastoralists. The American Journal of Human Genetics 78(2): 202-221.

Shapiro B, and Hofreiter M. 2013. Ancient DNA. In: Losos JB (editor). The Princeton Guide to Evolution. Princeton: Princeton University Press. p. 475-481.

Shapiro B, and Hofreiter M. 2014. A paleogenomic perspective on evolution and gene function: new insights from ancient DNA. Science 343(6169): 1236573.

Sharkey DJ, Scalice ER, Christy KG, Atwood SM, and Daiss JL. 1994. Antibodies as Thermolabile Switches: High Temperature Triggering for the Polymerase Chain Reaction. Nature Biotechnology 12(5): 506-509.

Shen P, Lavi T, Kivisild T, Chou V, Sengun D, Gefel D, Shpirer I, Woolf E, Hillel J, Feldman MW, and Oefner PJ. 2004. Reconstruction of patrilineages and matrilineages of Samaritans and other Israeli populations from Y-Chromosome and mitochondrial DNA sequence Variation. Human Mutation 24(3): 248-260.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, and Sirotkin K. 2001. dbSNP: the NCBI database of genetic variation. Nucleic Acids Research, 29(1): 308-311.

Shinnie PL. 1996. Ancient Nubia. New York: Columbia University Press.

Shriver MD, Kennedy GC, Parra EJ, Lawson HA, Sonpar V, Huang J, Akey JM, and Jones KW. 2004. The genomic distribution of population substructure in four populations using 8,525 autosomal SNPs. Human Genomics 1(4): 274.

Shriver M, and Kittles R. 2008. Genetic Ancestry and the Search for Personalized Genetic Histories. In: Koenig BA, Lee SS-J, and Richardson SS (editors). Revisiting Race in a Genomic Age. Revisiting Race in a Genomic Age: New Brunswick, NJ: Rutgers University Press. p. 201-214.

Silva M, Alshamali F, Silva P, Carrilho C, Mandlate F, Trovoada MJ, Černý V, Pereira L, and Soares P. 2015. 60,000 years of interactions between Central and Eastern Africa documented by major African mitochondrial haplogroup L2. Scientific Reports 5: 12526.

Skoglund P, and Jakobsson M. 2011. Archaic human ancestry in East Asia. Proceedings of the National Academy of Sciences 108(45): 18301-18306.

Skoglund P, Mallick S, Bortolini MC, Chennagiri N, Hünemeier T, Petzl-Erler ML, Salzano FM, Patterson N, and Reich D. 2015. Genetic evidence for two founding populations of the Americas. Nature 525(7567): 104-108.

Skoglund P, Malmström H, Omrak A, Raghavan M, Valdiosera C, Günther T, Hall P, Tambets K, Parik J, Sjögren K-G, Apel J, Willerslev E, Storå J, Götherstrom A, and Jakobsson M. 2014a. Genomic diversity and admixture differs for Stone-Age Scandinavian foragers and farmers. Science 344(6185): 747-750.

Skoglund P, Malmström H, Raghavan M, Storå J, Hall P, Willerslev E, Gilbert MTP, Götherstrom A, and Jakobsson M. 2012. Origins and genetic legacy of Neolithic farmers and hunter-gatherers in Europe. Science 336(6080): 466-469.

Skoglund P, Northoff BH, Shunkov MV, Derevianko AP, Pääbo S, Krause J, and Jakobsson M. 2014b. Separating endogenous ancient DNA from modern day contamination in a Siberian Neandertal. Proceedings of the National Academy of Sciences 111(6): 2229-2234.

Skoglund P, Posth C, Sirak K, Spriggs M, Valentin F, Bedford S, Clark GR, Reepmeyer C, Petchey F, Fernandes D, Fu Q, Harney E, Lipson M, Mallick S, Novak M, Rohland N, Stewardson K, Abdullah S, Cox MP, Friedlaender FR, Friedlaender JS, Kivisild T, Koki G, Kusuma P, Merriwether DA, Ricaut F-X, Wee JTS, Patterson N, Krause J, Pinhasi R, and Reich D. 2016. Genomic insights into the peopling of the Southwest Pacific. Nature 538(7626): 510.

Skoglund P, Storå J, Götherström A, and Jakobsson M. 2013. Accurate sex identification of ancient human remains using DNA shotgun sequencing. Journal of Archaeological Science 40(12): 4477-4482.

Skoglund P, Thompson JC, Prendergast ME, Mittnik M, Sirak K, Hajdinjak M, Salie T, Rohland N, Mallick S, Peltzer A, Heinze A, Olalde I, Ferry M, Harney E, Michel M, Stewardson K, Cerezo-Román J, Chiuma C, Crowther A, Gomani-Chindebvu E, Gidna AO, Grillo KM, Helenius T, Hellenthal G, Helm R, Horton M, López S, Mabulla AZP, Parkington J, Shipton C, Thomas MG, Tibesasa R, Welling M, Hayes VM, Kennett DJ, Ramesar R, Meyer M, Pääbo S, Patterson N, Morris AG, Boivin N, Pinhasi R, Krause J, Reich D. 2017. Reconstructing Prehistoric African Population Structure. Cell 171(1): 59-71.

Small MF. 1981. The Nubian Mesolithic: A Consideration of the Wadi Halfa Remains. Journal of Human Evolution 10(2) 159-162.

Smith CI, Chamberlain AT, Riley MS, Cooper A, Stringer CB, and Collins MJ. 2001. Neanderthal DNA: Not just old but old and cold? Nature 410(6830): 771-772.

Smith CI, Chamberlain AT, Riley MS, Stringer C, and Collins MJ. 2003. The thermal history of human fossils and the likelihood of successful DNA amplification. Journal of Human Evolution 45(3): 203-217.

Smith RW, Monroe C, and Bolnick DA. 2015. Detection of Cytosine Methylation in Ancient DNA from Five Native American Populations Using Bisulfite Sequencing. PLoS One 10(5): e0125344.

Soares P, Alshamali F, Pereira JB, Fernandes V, Silva NM, Afonso C, Costa MD, Musilová E, Macaulay V, Richards MB, Černý V, and Pereira L. 2012. The Expansion of mtDNA Haplogroup L3 within and out of Africa. Molecular Biology and Evolution 29(3): 915-927.

Soares P, Ermini L, Thomson N, Mormina M, Rito T, Röhl A, Salas A, Oppenheimer S, Macaulay V, and Richards MB. 2009. Correcting for Purifying Selection: An Improved Human Mitochondrial Molecular Clock. The American Journal of Human Genetics 84(6): 740-759.

Soler A. 2012. Life and death in a medieval Nubian farming community: the experience at Mis Island. Azania: Archaeological Research in Africa 47(4): 539-540.

Sørensen MS, Bretlau P, and Jørgensen MB. 1992. Quantum Type Bone Remodeling in the Human Otic Capsule: Morphometric Findings. Acta Oto-Laryngologica 112(sup496): 4-10.

Souza SMFMd, Carvalho DMd, and Lessa A. 2003. Paleoepidemiology: is there a case to answer? Memórias do Instituto Oswaldo Cruz 98: 21-27.

Spaulding J. 1995. Medieval Christian Nubia and the Islamic World: A Reconsideration of the Baqt Treaty. The International Journal of African Historical Studies 28(3): 577-594.

Spradley MK, and Jantz RL. 2011. Sex estimation in forensic anthropology: skull versus postcranial elements. Journal of Forensic Sciences 56(2): 289-296.

Starling AP, and Stock JT. 2007. Dental indicators of health and stress in early Egyptian and Nubian agriculturalists: a difficult transition and gradual recovery. American Journal of Physical Anthropology 134(4): 520-528.

Stevanovitch A, Gilles A, Bouzaid E, Kefi R, Paris F, Gayraud R, Spadoni JL, El-Chenawi F, and Béraud-Colomb E. 2004. Mitochondrial DNA sequence diversity in a sedentary population from Egypt. Annals of Human Genetics 68(1): 23-39.

Stojanowski CM, and Schillaci MA. 2006. Phenotypic approaches for understanding patterns of intracemetery biological variation. American Journal of Physical Anthropology 131(S43): 49-88.

Stoneking M, and Krause J. 2011. Learning about human population history from ancient and modern genomes. Nature Reviews Genetics 12(9): 603-614.

Strouhal E. 1971. Evidence of the Early Penetration of Negroes into Prehistoric Egypt. The Journal of African History 12(1): 1-9.

Stynder DD, Braga J, and Crubezy E. 2009. Craniometric evidence for biological continuity between Meroitic and post-Meroitic populations buried at the necropolis of Missiminia, Middle Nubia. The South African Archaeological Bulletin: 122-129.

Szécsényi-Nagy A, Brandt G, Haak W, Keerl V, Jakucs J, Möller-Rieker S, Köhler K, Mende BG, Oross K, Marton T, Osztás A, Kiss V, Fecher M, Pálfi G, Molnár E, Sebők K, Czene A, Paluch T, Šlaus M, Novak M, Pećina-Šlaus N, Ősz B, Voicsek V, Somogyi K, Tóth G, Kromer B, Bánffy E, and Alt KW. 2015. Tracing the genetic origin of Europe's first farmers reveals insights into their social organization. Proceedings of the Royal Society of London B: Biological Sciences 282(1805): 2015.0339.

Tang Y, Schon EA, Wilichowski E, Vazquez-Memije ME, Davidson E, and King MP. 2000. Rearrangements of Human Mitochondrial DNA (mtDNA): New Insights into the Regulation of mtDNA Copy Number and Gene Expression. Molecular Biology of the Cell 11(4): 1471-1485.

The 1,000 Genomes Project Consortium. 2010. A map of human genome variation from population-scale sequencing. Nature 467(7319): 1061-1073.

The 1,000 Genomes Project Consortium. 2012. An integrated map of genetic variation from 1,092 human genomes. Nature 491(7422): 56-65.

The International HapMap Consortium. 2005. A haplotype map of the human genome. Nature 437(7063): 1299-1320.

The International HapMap Consortium. 2007. A second generation human haplotype map of over 3.1 million SNPs. Nature 449(7164): 851-861.

The International HapMap Consortium. 2010. Integrating common and rare genetic variation in diverse human populations. Nature 467(7311): 52-58.

The International SNP Map Working Group. 2001. A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. Nature 409(6822): 928-933.

Thomas, MG, Weale ME, Jones AL, Richards M, Smith A, Redhead N, Torroni A, Scozzari R, Gratrix F, Tarekegn A, Wilson JF, Capelli C, Bradman N, and Goldstein DB. 2002. Founding mothers of Jewish communities: geographically separated Jewish groups were independently founded by very few female ancestors. The American Journal of Human Genetics 70(6): 1411-1420.

Thurmond AK, Stern RJ, Abdelsalam MG, Nielsen KC, Abdeen MM, and Hinz E. 2004. The Nubian Swell. Journal of African Earth Sciences 39(3): 401-407.

Tilford CA, Kuroda-Kawaguchi T, Skaletsky H, Rozen S, Brown LG, Rosenberg M, McPherson JD, Wylie K, Sekhon M, Kucaba TA, Waterston RH, and Page DC. 2001. A physical map of the human Y chromosome. Nature 409(6822): 943-945.

Tishkoff SA, and Kidd KK. 2004. Implications of biogeography of human populations for 'race' and medicine. Nature Genetics 36: S21-S27.

Torroni A, Achilli A, Macaulay V, Richards M, and Bandelt H-J. 2006. Harvesting the fruit of the human mtDNA tree. Trends in Genetics 22(6): 339-345.

Torroni A, Schurr TG, Cabell MF, Brown MD, Neel JV, Larsen M, Smith DG, Vullo CM, and Wallace DC. 1993. Asian affinities and continental radiation of the four

founding Native American mtDNAs. American Journal of Human Genetics 53(3): 563.

Trigger BG. 1965. History and settlement in Lower Nubia. New Haven: Department of Anthropology, Yale University.

Trigger BG. 1994. Paradigms in Sudan archaeology. International Journal of African Historical Studies 27(2): 323-345.

Trombetta B, Cruciani F, Sellitto D, and Scozzari R. 2011. A New Topology of the Human Y Chromosome Haplogroup E1b1 (E-P2) Revealed through the Use of Newly Characterized Binary Polymorphisms. PLoS One 6(1): e16073.

Trombetta B, D'Atanasio E, Massaia A, Ippoliti M, Coppa A, Candilio F, Coia V, Russo G, Dugoujon J-M, Akar PMN, Sellitto D, Valesini G, Novelletto A, Scozzari R, and Cruciani F. 2015. Phylogeographic Refinement and Large Scale Genotyping of Human Y Chromosome Haplogroup E Provide New Insights into the Dispersal of Early Pastoralists in the African Continent. Genome Biology and Evolution 7(7): 1940-1950.

Turner BL, Edwards JL, Quinn EA, Kingston JD, and Van Gerven DP. 2007. Age-related variation in isotopic indicators of diet at medieval Kulubnarti, Sudanese Nubia. International Journal of Osteoarchaeology 17(1): 1-25.

Turner-Walker G, Nielsen-Marsh C, Syversen U, Kars H, and Collins M. 2002. Sub-micron spongiform porosity is the major ultra-structural alteration occurring in archaeological bone. International Journal of Osteoarchaeology 12(6): 407-414.

Ubelaker DH. 1989. Human Skeletal Remains: Excavation, Analysis, Interpretation (2nd edition). Washington D.C.: Taraxacum.

Underhill PA, and Kivisild T. 2007. Use of Y chromosome and mitochondrial DNA population structure in tracing human migrations. Annual Review of Genetics 41: 539-564.

Underhill PA, Passarino G, Lin AA, Shen P, Lahr MM, Foley RA, Oefner PJ, and Cavalli-Sforza LL. 2001. The phylogeography of Y chromosome binary haplotypes and the origins of modern human populations. Annals of Human Genetics 65(1): 43-62.

Underhill PA, Shen P, Lin AA, Jin L, Passarino G, Yang WH, Kauffman E, Bonné-Tamir B, Bertranpetit J, Francalacci P, Ibrahim M, Jenkins T, Kidd JR, Mehi SQ, Seielstad MT, Wells RS, Piazza A, Davis RW, Feldman MW, Cavalli-Sforza LL, and Oefner PJ. 2000. Y chromosome sequence variation and the history of human populations. Nature Genetics 26(3): 358-361.

Van Gerven DP. 1982. The contribution of time and local geography to craniofacial variation in Nubia's Batn el Hajar. American Journal of Physical Anthropology 59(3): 307-316.

Van Gerven DP, Armelagos GJ, and Rohr A. 1977. Continuity and change in cranial morphology of three Nubian archaeological populations. Man 12(2): 270-277.

Van Gerven DP, Beck R, and Hummert JR. 1990. Patterns of enamel hypoplasia in two medieval populations from Nubia's *Batn el Hajar*. American Journal of Physical Anthropology 82(4): 413-420.

Van Gerven DP, Carlson DS, and Armelagos GJ. 1973. Racial History and Bio-Cultural Adaptation of Nubian Archaeological Populations. The Journal of African History 14(4): 555-564.

Van Gerven DP, Sandford MK, and Hummert JR. 1981. Mortality and culture change in Nubia's Batn el Hajar. Journal of Human Evolution 10(5): 395-408.

Van Gerven DP, Sheridan SG, and Adams WY. 1995. The health and nutrition of a Medieval Nubian population. American Anthropologist 97(3): 468-480.

van Leeuwen MHD, and Maas I. 2005. Endogamy and social class in history: An overview. International Review of Social History 50: 1-23.

van Oven M. 2015. PhyloTree Build 17: Growing the human mitochondrial DNA tree. Forensic Science International: Genetics Supplement Series 5: e392-e394.

van Oven M, and Kayser M. 2009. Updated comprehensive phylogenetic tree of global human mitochondrial DNA variation. Human Mutation 30(2): E386-E394.

Vantini G. 1975. Oriental Sources Concerning Nubia. Warsaw: Heidelberg.

Varshney U, and van de Sande JH. 1991. Specificities and kinetics of uracil excision from uracil-containing DNA oligomers by Escherichia coli uracil DNA glycosylase. Biochemistry 30(16): 4055-4061.

Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M, Subramanian G, Thomas PD, Zhang J, Miklos GLG, Nelson C, Broder S, Clark AG, Nadeau J, McKusick VA, Zinder N, Levine AJ, Roberts RJ, Simon M, Slayman C, Hunkapiller M, Bolanos R, Delcher A, Dew I, Fasulo D, Flanigan M, Florea L, Halpern A, Hannenhalli S, Kravitz S, Levy S, Mobarry C, Reinert K, Remington K, Abu-Threideh J, Beasley E, Biddick K, Bonazzi V, Brandon R, Cargill M, Chandramouliswaran I, Charlab R, Chaturvedi K, Deng Z, Di Francesco V, Dunn P, Eilbeck K, Evangelista C, Gabrielian AE, Gan W, Ge W, Gong F, Gu Z, Guan

P, Heiman TJ, Higgins ME, Ji R-R, Ke Z, Ketchum KA, Lai Z, Lei Y, Li Z, Li J, Liang Y, Lin Z, Lu F, Merkulov GV, Milshina N, Moore HM, Naik AK, Narayan VA, Neelam B, Nusskern D, Rusch DB, Salzberg S, Shao W, Shue B, Sun J, Wang ZY, Wang A, Wang X, Wang J, Wei M-H, Wide R, Xiao C, Yan C, Yao A, Ye J, Zhan M, Zhang W, Zhang H, Zhao Q, Zheng L, Zhong F, Zhong W, Zhu SC, Zhao S, Gilbert D, Baumhueter S, Spier G, Carter C, Cravchik A, Woodage T, Ali F, An H, Awe A, Baldwin D, Baden H, Barnstead M, Barrow I, Beeson K, Busam D, Carver A, Cener A, Cheng ML, Curry L, Danaher S, Davenport L, Desilets R, Dietz S, Dodson K, Doup L, Ferriera S, Garg N, Gluecksmann A, Hart B, Haynes J, Haynes C, Heiner C, Hladun S, Hostin D, Houck J, Howland T, Ibegwam C, Johnson J, Kalush F, Kline L, Koduru S, Love A, Mann F, May D, McCawley S, McIntosh T, McMullen I, Moy M, Moy L, Murphy B, Nelson K, Pfannkoch C, Pratts E, Puri V, Qureshi H, Reardon M, Rodriguez R, Rogers Y-H, Romblad D, Ruhfel B, Scott R, Sitter C, Smallwood M, Stewart E, Strong R, Suh E, Thomas R, Tint NN, Tse S, Vech C, Wang G, Wetter J, Williams S, Williams M, Windsor S, Winn-Deen E, Wolfe K, Zaveri J, Zaveri K, Abril JF, Guigó R, Campbell MJ, Sjolander KV, Karlak B, Kejariwal A, Mi H, Lazareva B, Hatton T, Narechania A, Diemer K, Muruganujan A, Guo N, Sato S, Bafna V, Istrail S, Lippert R, Schwarts R, Walenz B, Yooseph S, Allen D, Basu A, Baxendale J, Blick L, Caminha M, Carnes-Stine J, Caulk P, Chiang Y-H, Coyne M, Dahlke C, Mays AD, Dombroski M, Donnelly M, Ely D, Esparham S, Fosler C, Gire H, Glanowski S, Glasser K, Glodek A, Gorokhov M, Graham K, Gropman B, Harris M, Heil J, Henderson S, Hoover J, Jennings D, Jordan C, Jordan D, Kasha J, Kagan L, Kraft C, Levitsky A, Lewis M, Lui X, Lopez J, Ma D, Majoros W, McDaniel D, Murphy S, Newman M, Nguyen T, Nguyen N, Nodell M, Pan S, Peck J, Peterson M, Rowe W, Sanders R, Scott J, Simpson M, Smith T, Sprague A, Stockwell T, Turner R, Venter E, Wang M, Wen M, Wu D, Wu M, Xia A, Zandieh A, and Zhu X. 2001. The sequence of the human genome. Science 291(5507): 1304-1351.

Vollner JM. 2016. Examining the population history of three medieval Nubian sites through craniometric analyses. PhD Dissertation: Michigan State University.

Von Cramon-Taubadel N, and Weaver TD. 2009. Insights from a quantitative genetic approach to human morphological evolution. Evolutionary Anthropology 18: 237-240.

Walker PL, Bathurst RR, Richman R, Gjerdrum T, and Andrushko VA. 2009. The causes of porotic hyperostosis and cribra orbitalia: A reappraisal of the iron-deficiency-anemia hypothesis. American Journal of Physical Anthropology 139(2): 109-125.

Wallace DC. 1999. Aging and Degenerative Diseases: A Mitochondrial Paradigm. In: Papa S, Guerrieri F, and Tager JM (editors). Frontiers of Cellular Biogenetics: Molecular Biology, Biochemistry, and Physiopathology. New York: Springer US.

Wang Y, and Nielsen R. 2012. Estimating population divergence time and phylogeny from single-nucleotide polymorphisms data with outgroup ascertainment bias. Molecular Ecology 21(4): 974-986.

Wapler U, Crubézy E, and Schultz M. 2004. Is cribra orbitalia synonymous with anemia? Analysis and interpretation of cranial pathology in Sudan. American Journal of Physical Anthropology 123(4): 333-339.

Washburn SL. 1951. Section of anthropology: the new physical anthropology. Transactions of the New York Academy of Sciences 13(Issue 7 Series II): 298-304.

Watson JD, and Crick FH. 1953. Molecular structure of nucleic acids. Nature 171(4356): 737-738.

Watson E, Forster P, Richards M, and Bandelt H-J. 1997. Mitochondrial footprints of human expansions in Africa. The American Journal of Human Genetics 61(3): 691-704.

Weir BS, and Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. Evolution 38: 1358-1370.

Weissensteiner H, Pacher D, Kloss-Brandstätter A, Forer L, Specht G, Bandelt H-J, Kronenberg F, Salas A, and Schönherr S. 2016. HaploGrep 2: mitochondrial haplogroup classification in the era of high-throughput sequencing. Nucleic Acids Research 44: W58-W63.

Wells RS, Yuldasheva N, Ruzibakiev R, Underhill PA, Evseeva I, Blue-Smith J, Jin L, Su B, Pitchappan R, Shanmugalakshmi S, Balakrishnan K, Read M, Pearson NM, Zerjal T, Webster MT, Zholoshvili I, Jamarjashvili E, Gambarov S, Nikbin B, Dostiev A, Aknazarov O, Zalloua P, Tsoy I, Kitaev M. Mirrakhimov M, Chariev A, and Bodmer WF. 2001. The Eurasian Heartland: A continental perspective on Y-chromosome diversity. Proceedings of the National Academy of Sciences 98(18): 10244-10249.

Welsby D. 2002. The Medieval Kingdoms of Nubia: Pagans, Christians, and Muslims along the Middle Nile. London: The British Museum Press.

Weyrich LS, Dobney K, and Cooper A. 2015. Ancient DNA analysis of dental calculus. Journal of Human Evolution 79: 119-124.

Weyrich LS, Duchene S, Soubrier J, Arriola L, Llamas B, Breen J, Morris AG, Alt KW, Caramelli D, Dresely V, Farrell M, Farrer AG, Francken M, Gully N, Haak W, Hardy K, Harvati K, Held P, Holmes EC, Kaidonis J, Lalueza-Fox C, de la rasilla M, Rosas A, Semal P, Soltysiak A, Townsend G, Usai D, Wahl J, Huson DH,

Dobney K and Cooper A. 2017. Neanderthal behaviour, diet, and disease inferred from ancient DNA in dental calculus. Nature 544(7650): 357-361.

White TD, Black MT, and Folkens PA. 2012. Human Osteology (3rd Edition). Burlington: Elsevier Academic Press.

White C, Longstaffe FJ, and Law KR. 2004. Exploring the effects of environment, physiology and diet on oxygen isotope ratios in ancient Nubian bones and teeth. Journal of Archaeological Science 31(2): 233-250.

Wilde S, Timpson A, Kirsanow K, Kaiser E, Kayser M, Unterländer M, Hollfelder N, Potekhina ID, Schier W, Thomas MG, and Burger J. 2014. Direct evidence for positive selection of skin, hair, and eye pigmentation in Europeans during the last 5,000 y. Proceedings of the National Academy of Sciences 111(13): 4832-4837.

Wilkins JF. 2006. Unraveling male and female histories from human genetic data. Current Opinion in Genetics and Development 16(6): 611-617.

Willerslev E, and Cooper A. 2005. Ancient DNA. Proceedings of the Royal Society B: Biological Sciences 272(1558): 3-16.

Willerslev E, Hansen AJ, and Poinar HN. 2004. Isolation of nucleic acids and cultures from fossil ice and permafrost. Trends in Ecology and Evolution 19(3): 141-147.

Willing E-M, Dreyer C, and Van Oosterhout C. 2012. Estimates of genetic differentiation measured by FST do not necessarily require large sample sizes when using many SNP markers. PLoS One 7(8): e42649.

Wood ET, Stover DA, Ehret C, Destro-Bisol G, Spedini G, McLeod H, Louie L, Bamshad M, Strassmann BI, Soodyall H, and Hammer MF. 2005. Contrasting patterns of Y chromosome and mtDNA variation in Africa: evidence for sex-biased demographic processes. European Journal of Human Genetics 13(7): 867.

Woodward JC, Macklin MG, Krom MD, and Williams MA. 2007. The Nile: Evolution, Quaternary River Environments and Material Fluxes. Large Rivers: Geomorphology and Management 13: 712.

Wright S. 1949. The genetical structure of populations. Annals of Eugenics 15(1): 323-354.

Wright S. 1978. Evolution and the Genetics of Populations (Vol. 4: Variability within and among natural populations). Chicago: University of Chicago Press.

Y Chromosome Consortium. 2002. A nomenclature system for the tree of human Y-chromosomal binary haplogroups. Genome Research 12(2): 339-348.

Yousif H, and Eltayeb M. 2009. Genetic Patterns of Y-chromosome and Mitochondrial DNA Variation, with Implications to the Peopling of the Sudan. PhD Dissertation: University of Khartoum.

Yang DY, Eng B, Waye JS, Dudar JC, and Saunders SR. 1998. Technical Note: Improved DNA Extraction from Ancient Bones Using Silica-Based Spin Columns. American Journal of Physical Anthropology 105(4): 539-543.

Yang DY, and Watt K. 2005. Contamination controls when preparing archaeological remains for ancient DNA analysis. Journal of Archaeological Science 32(3): 331-336.

Yegnasubramanian S. 2013. Preparation of Fragment Libraries for Next-Generation Sequencing on the Applied Biosystems SOLiD Platform. Methods in Enzymology 529: 185-200.

Yunusbayev B, Metspalu M, Järve M, Kutuev I, Rootsi S, Metspalu E, Behar DM, Varendi K, Sahakyan H, Khusainova R, Yepiskoposyan L, Khusnutdinova EK, Underhill PA, Kivisild T, and Villems R. 2011. The Caucasus as an asymmetric semipermeable barrier to ancient human migrations. Molecular Biology and Evolution 29(1): 359-365.

Zarroug Me-DA. 1991. The Kingdom of Alwa. Calgary: University of Calgary Press.

Zuckerkandl E. 1963. Perspectives in Molecular Anthropology. In: Washburn SL (editor). Classification and Evolution. Chicago: Aldine. p. 243-272.

Zuckerman MK, and Armelagos GJ. 2011. The Origins of Biocultural Dimensions in Bioarchaeology. In: Agarwal SC, and Glencross BA (editors). Social Bioarchaeology. p.15-43.

SUPPLEMENT 1: SAMPLE PROCESSING DATA FROM NINETY-NINE SCREENED KULUBNARTI SAMPLES

Table S1.1 presents the demographic and sample processing data for each of the 99 screened Kulubnarti Nubian samples. Demographic data were obtained from Adams et al. (1999). Sample processing data include the side and specific part of the petrous bone processed for ancient DNA analysis, the author's visual assessment of macroscopic preservation of the bone part selected for analysis, a brief description of the bone part, and the amount of bone powder used in DNA extraction. An asterisk next to the sample ID designates the sample as one of the 28 best-preserved individuals that were used to generate the genome-wide data analyzed in this dissertation.

Table S1.2 presents detailed archaeological and bioarchaeological information for the 28 samples identified as being best preserved during the initial screening step (indicated by an asterisk in Table S1.1). All demographic data (age, morphological sex) and archaeological information (grave information, head/body coverings, burial position) are from Adams et al. (1999); description of remains and evidence of cribra orbitalia are based on notes taken by the author during sample collection in 2015.

For both tables, all sample IDs starting with "R" designate the individual to be a member of the R cemetery; sample IDs starting with "S" designate the individual to be a member of the S cemetery.

Table S1.1: Demographic and sample processing data for each of the 99 screened Kulubnarti samples.

Sample ID	Age of Individual (yrs)	Morph-ological Sex	Bone part	Macroscopic Preservation Assessment	Description of bone part used	Amount of Bone Powder Used
R101*	4	?	Right petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	64 mg
R103	18 mo.	?	Right petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	52 mg
R123	10	Male	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion; only woven bone included	51 mg
R124*	5	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	55 mg
R125	2	Male	Right petrous bone: cochlea	GOOD	Dense tan cochlear portion	56 mg
R133	4	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	58 mg
R143	45	Female	Left petrous bone: cochlea + osseous labyrinth	GOOD	Dense yellow cochlear portion; additional osseous labyrinth included	59 mg
R144	34	Female	Left petrous bone: cochlea	V. GOOD	Extremely dense yellow cochlear portion	52 mg
R15	22	Male	Left petrous bone: cochlea	V. GOOD	Dense tan cochlear portion	53 mg
R150	3	?	Right petrous bone: cochlea + osseous labyrinth	GOOD	Dense yellow cochlear portion; additional osseous labyrinth included	55 mg
R152*	3	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion isolated	49 mg
R153	38	Male	Right petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	51 mg
R157	4	?	Left petrous bone: cochlea	OKAY	Dense yellow cochlear portion only woven bone included	50 mg
R159	2	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	59 mg

R164	9 mo.	?	Right petrous bone: cochlea	OKAY	Dense tan cochlear portion	57 mg
R169*	18 mo.	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	53 mg
R173	12 mo.	?	Right petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	58 mg
R181*	12 mo.	?	Right petrous bone: cochlea	V. GOOD	Dense yellow cochlea portion	55 mg
R182*	3	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	55 mg
R183	18 mo.	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	53 mg
R184	2	?	Left petrous bone: cochlea	OKAY	Dense yellow cochlear portion	50mg
R186*	2	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	63 mg
R195*	18 mo.	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	53 mg
R196*	18 mo.	Male	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	58 mg
R201*	7	?	Right petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	59 mg
R202*	9 mo.	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	55 mg
R21	6	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion; only woven bone included	50 mg
R5*	6	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	58 mg
R57	9 mo.	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	?
R59*	9	?	Right petrous bone: cochlea	V. GOOD	Dense yellow/tan cochlear portion	54 mg
R76	5	?	Right petrous bone: cochlea	GOOD	Dense tan cochlear portion	56 mg
R79*	4	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	51 mg
R8	7	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	54 mg
R84	2	Female	Right petrous bone: cochlea	OKAY	Dense tan cochlear portion	60 mg

R86	6	Male	Left petrous bone: cochlea	OKAY	Dense tan cochlear portion	60 mg
R91	2	Female	Left petrous bone: cochlea	GOOD, POWDER OKAY	Dense tan cochlear portion	57 mg
R93*	10	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	51 mg
R94	12	?	Right petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	5 3mg
R95	14	Female	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	54 mg
R96	3	?	Left petrous bone: cochlea	OKAY	Dense yellow cochlear portion	54 mg
S106	5	?	Left petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	51 mg
S109	31	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	55 mg
S110	4	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	28 mg
S114	5	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	46 mg
S115	3	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	54 mg
S120	9 mo.	?	Left petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	58 mg
S126	16	?	Left petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	58 mg
S131a	18 mo.	?	Right petrous bone: cochlea	GOOD	Dense tan cochlear portion isolated	57 mg
S132	3	?	Left petrous bone: cochlea	GOOD	Dense tan cochlear portion	55 mg
S133*	6 mo.	?	Left petrous bone: cochlea	OKAY	Dense yellow cochlear portion	61 mg
S136	4	?	Right petrous bone: cochlea	BAD	Dense yellow cochlear portion; only woven bone included	55 mg
S143	neo- nate	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	61 mg

S144	7 mo. in utero	?	Right petrous bone: cochlea	BAD (powder GOOD)	Dense yellow cochlear portion; additional osseous labyrinth included	52 mg
S149*	neo-nate	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	58 mg
S159*	9 mo.	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	44 mg
S17*	5	?	Right petrous bone: cochlea	OKAY	Dense tan cochlear portion	50 mg
S171	36	Female	Left petrous bone: cochlea	GOOD	Bone was crumbly and not well-preserved, but cochlea was dense and rich tan in color	55 mg
S173	42	Male	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	49 mg
S174	51+	Female	Left petrous bone: cochlea	OKAY	Dense yellow cochlear portion	50mg
S175	18 mo.	?	Right petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	56 mg
S181	5	?	Left petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	53 mg
S182	4	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	55 mg
S183	neo-nate	?	Left petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; cochlear portion used, but was not dense or yellow	47 mg
S190	8	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion; only woven bone included	53 mg

S193	6 mo.	?	Left petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; cochlear portion isolated, but was not dense or yellow	49 mg
S198*	2	Male	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	53 mg
S199a	8	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	50 mg
S199b	6	?	Right petrous bone: cochlea	GOOD	Dense yellow/tan cochlear portion	56 mg
S2	10	?	Left petrous bone: cochlea	V. GOOD	Dense tan cochlear portion	51 mg
S201	3	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	51 mg
S208*	18 mo.	Female	Right petrous bone: cochlea	BAD	Dense dark brown cochlear portion	58 mg
S217	4	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion; only woven bone included	51 mg
S218	6 mo.	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	56 mg
S230	6	?	Right petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	53 mg
S235	3	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	57 mg
S238	12	?	Right petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	53 mg
S24	4	?	Left petrous bone: cochlea	OKAY	Dense yellow cochlear portion	59 mg
S242	6	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	52 mg
S25	12 mo.	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	52 mg
S27*[153]	9 mo.	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	50 mg

[153] While writing up the results of this analysis, it was discovered in Adams et al. (1999) that individual S27 may have been buried in a grave dating to the previous X-Group period, though a definitive assessment was

S33*	12 mo.	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	57 mg
S37	12 mo.	?	Left petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	56 mg
S38	4	?	Left petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	57 mg
S42a	neo-nate	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	49 mg
S42b	5	?	Right petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	53 mg
S50*	5	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	56 mg
S53	11	Female	Right petrous bone: cochlea	OKAY	Dense tan cochlear portion	55 mg
S57	1 mo.	?	Right petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	42 mg
S60	12 mo.	?	Right petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	57 mg
S68a*	4	Male	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	53 mg
S69	9 mo.	?	Left petrous bone: cochlea	OKAY	Dense tan cochlear portion	58 mg
S70	1	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	52 mg
S73*	3	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	51 mg
S75	5	?	Left petrous bone: cochlea	BAD	Bone was crumbly and not well-preserved; densest cochlear portion used	53 mg
S78	4	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	60 mg
S79*	3	?	Right petrous bone: cochlea	GOOD	Dense yellow cochlear portion	54 mg

not made. Based on the east-west orientation of the grave and the placement of the head at the west end (consistent with a Christian style of burial), a lack of any descriptive features or body positioning classifying the burial of S27 as part of the X-Group period, and no direct dates of this grave, this individual will be considered in this analysis.

S81*	3	?	Right petrous bone: cochlea	OKAY	Dense yellow cochlear portion	59 mg
S87*	6 mo.	?	Left petrous bone: cochlea	GOOD	Dense yellow cochlear portion	47 mg
S89*	9 mo.	?	Right petrous bone: cochlea	V. GOOD	Dense yellow cochlear portion	51 mg

Table S1.2: Available archaeological and bioarchaeological information for 30 samples identified as best preserved during the initial screening step.

Sample ID	Age of Ind. (yrs)	Morph-ological Sex	Grave information	Head/body coverings	Burial position	Description of Remains	Evidence of Cribra Orbitalia?
R101*	4	?	Slot grave; no superstructure	3 stones over head	Buried on L side; head facing L; arms at pubis; legs sl. flexed	Preserved soft tissue; brown curly hair	Yes
R124*	5	?	Side niche grave; no superstructure	Stones over head and body	Buried on L side; head facing L; arms at pubis; legs sl. flexed	Preserved soft tissue; brown straight hair; meninges present	Yes
R152*	3	?	Slot grave; no superstructure	3 stones over head	Buried on L side; head facing L; arms at sides; legs straight	Little soft tissue; straight brown hair	Yes
R169*	18 mo.	?	Slot grave; stone pavement superstructure	None	Buried on L side; head facing L; arms at pubis; legs sl. flexed	Preserved soft tissue	Yes
R181*	12 mo.	?	Side niche grave; stone pavement superstructure	None	Buried on L side; head facing L; arms at pubis; legs sl. flexed	Preserved soft tissue; brown hair	No
R182*	3	?	Side niche grave; stone pavement superstructure	Bricks over head and body	Buried on L side; head facing L; arms at pubis; legs sl. flexed	Preserved soft tissue; straight brown hair	Yes
R186*	2	?	Slot grave; stone outline superstructure	3 stones over head	Buried on L side; head facing L; R arm at chin/L arm at pubis; legs flexed	Preserved soft tissue; brown hair	Yes

R196*	18 mo.	Male	Slot grave; no superstructure	None	Buried on L side; head facing L; arms at pubis; legs sl. flexed	Preserved soft tissue; brown hair	Yes
R201*	7	?	Slot grave; brick pavement superstructure	3 stones over head	Buried on L side; head facing L; arms at pubis; legs sl. flexed	No soft tissue	Yes
R202*	9 mo.	?	Superstructure and shaft disturbed	1 brick over head	Buried on L side; head facing L; ? arms; legs straight	No soft tissue	No
R5*	6	?	Slot grave; no superstructure	None	Buried on R side; head facing R; arms at sides; legs straight	Preserved soft tissue; straight dark hair	Yes
R59*	9	?	Slot grave; no superstructure	3 stones over head	Buried ¼ L; head facing up; arms at pubis; legs straight	Preserved soft tissue; straight dark brown hair	No
R79*	4	?	Slot grave; no superstructure	Stones over head and body	Buried on L side; head facing L; arms at pubis; legs sl. flexed	No soft tissue	Yes
R93*	10	?	Slot grave; no superstructure	None	Buried on back; head facing R; arms at sides; legs sl. flexed	No soft tissue	Yes
S133*	6 mo.	?	Slot grave; stone pavement superstructure	3 stones over head	Buried on back; head facing R; arms at sides; legs flexed	Preserved soft tissue; straight brown hair	Yes
S149*	neo-nate	?	Slot grave; no record of superstructure	None	Buried on back?; head facing L; arms ?; legs ?	Preserved soft tissue; dark brown hair	?
S159*	9 mo.	?	Slot grave; no record of superstructure	None	Buried on back; head facing L; arms at sides; legs straight	Preserved soft tissue; brown hair; burial shroud attached	Yes

S17*	5	?	Slot grave; stone pavement superstructure	None	Buried on L side; head facing L; arms at pubis; legs straight	Preserved soft tissue; brown straight hair	Yes
S198*	2	Male		Stones over head and body	Buried on back; head facing L; arms at sides; legs straight	Preserved soft tissue; brown curly hair	No
S27*	9 mo.	?	Side niche grave; stone pavement superstructure; X-group grave?	None	Buried on L side; head facing down; arms at sides; legs flexed	Preserved soft tissue; burial shroud attached	No
S33*	12 mo.	?	Slot grave; stone pavement superstructure	Stones over head and body	Buried on L side; head facing L; arms at sides; legs flexed	Preserved soft tissue; brown hair; burial shroud attached	Yes
S50*	5	?	Slot grave; stone pavement superstructure	3 stones over head	Buried on L side; head facing L; arms at sides; legs flexed	Preserved soft tissue; light brown hair	?
S68a*	4	Male	Slot grave; stone pavement superstructure	Stones over head	Buried on L side; head facing L; arms at sides; legs sl. flexed	Preserved soft tissue; brown hair	Yes
S73*	3	?	Slot grave; stone outline superstructure	Stones over head and body	Buried on back; head facing up; ? arms; legs straight	Preserved soft tissue; straight brown hair	Yes
S79*	3	?	Side niche grave; stone pavement superstructure	None	Buried on back; head facing up; arms at pubis; legs mod. Flexed	Preserved soft tissue; brown hair	Yes

S81*	3	?	Side niche grave; stone pavement superstructure	None	Buried on back; head facing L; arms ?; legs straight	Preserved soft tissue; reddish-brown hair; burial shroud attached	No
S87*	6 mo.	?	Slot grave; stone pavement superstructure	Stones over head and body	Buried on L side; head facing left; arms at pubis; legs ?	No soft tissue	Yes
S89*	9 mo.	?	Slot grave; stone pavement superstructure	Stones over head and body	Buried on back; head facing L; arms ?; legs flexed	Preserved soft tissue; meninges present; braided brown hair	?

SUPPLEMENT 2: NEXT-GENERATION SQUENCING DATA FROM NINETY-NINE SCREENED KULUBNARTI SAMPLES

Table S2.1 presents the screening data obtained by Next-Generation Sequencing (NGS) for each of the 99 screened Kulubnarti samples. For each sample, these data include the NGS platform (MiSeq or NextSeq) used for screening, the total number of reads, the number of reads aligned to the human genome (build hg19, with the mitochondrial sequence replaced by the revised Cambridge reference sequence) (Andrews et al. 1999), and the percent endogenous DNA. Additionally, when possible, molecular sex is provided along with an associated p-value; sex is listed as "unable to be determined" when not enough reads aligned to the X chromosome for sex assignment. Mitochondrial haplogroup is assigned when possible, along with a score that indicates the likelihood that the haplogroup estimation is correct (provided by the Phy-Mer software) (Navarro-Gomez et al. 2015); "no data" indicates that there were not enough mtDNA reads to make a haplogroup estimation; and "no consensus" indicates that too few reads prevented the accurate determination of macrohaplogroup. Finally, the quantified frequency of damage at the last base pair of the 5'- and 3'-end of the DNA molecules is provided, along with the author's visual assessment of the damage pattern across the DNA molecule.

For this visual assessment, "++" indicates that the damage pattern is strongly consistent with that expected for ancient DNA; "+" indicates that the damage pattern is consistent with that expected for ancient DNA; "-" indicates either a damage pattern consistent with that expected for ancient DNA but a low frequency of damage (<5%) at the ends of the molecule or an expected frequency of damage (>5%) at the ends of the

molecule but a damage pattern throughout the molecule inconsistent with ancient DNA; and "x" indicates no evident damage, suggesting that the majority of DNA is modern, contaminating DNA. "No data" indicates that there were not enough reads for mapDamage analysis to take place.

All sample IDs starting with "R" designate the individual to be a member of the R cemetery; sample IDs starting with "S" designate the individual to be a member of the S cemetery. An asterisk next to the sample ID designates the sample as one of the 30 best-preserved individuals that were sent to the Reich Lab for additional screening and potential 1240K SNP capture.

Table S2.1: Screening data for each of the 99 Kulubnarti samples.

Sample ID	Sequencing Platform	Total Reads	Reads Aligned to Human Genome	% Endogenous DNA	Molecular Sex	Mitochondrial DNA Haplogroup Estimation and Score	mapDamage (5' | 3')
R101*	MiSeq	628647	33463	5.32	Male (p = 3.55e−08)	no consensus	0.12 | 0.10 (+)
R103	NextSeq	12883422	360825	2.80	Female (p = 3.41e−11)	L0a1a1 (0.82)	0.15 | 0.15 (++)
R123	MiSeq	435455	8416	1.93	Male (p = 7.39e−09)	L5a1b (0.56)	0.10 | 0.08 (+)
R124*	MiSeq	355955	26490	7.44	Male (p = 2.71e−07)	K1a12 (0.59)	0.09 | 0.09 (+)
R125	NextSeq	9549043	73992	0.77	Male (p = 3.02e−08)	L0a1a1 (0.63)	0.12 | 0.12 (++)
R133	NextSeq	9290130	111768	1.20	Female (p = 2.16e−11)	T (0.82)	0.10 | 0.10 (+)
R143	MiSeq	281721	557	0.2	unable to be determined	no data	0.03 | 0.02 (x)

R144	NextSeq	6376060	9019	0.14	Female (p = 1.13e−06)	no consensus	0.12 | 0.12 (+)
R15	NextSeq	8862581	199262	2.25	Male (p = 7.4e−09)	H2a (0.72)	0.11 | 0.12 (++)
R150	MiSeq	268429	1742	0.65	Male (p = 1.64e−05)	L2a1d1 (0.52)	0.07 | 0.06 (-)
R152*	NextSeq	10425922	2694260	25.84	Female (p = 1.02e−11)	L0a1a (0.83)	0.08 | 0.08 (+)
R153	NextSeq	14290027	14293	0.10	unable to be determined	no consensus	0.01 | 0.01 (x)
R157	MiSeq	393280	579	0.15	Female (p = 2.48e−05)	no data	0.05 | 0.07 (x)
R159	MiSeq	505071	14790	2.93	Female (p = 1.22e−10)	I2b (0.53)	0.09 | 0.09 (+)
R164	NextSeq	9472926	254709	2.69	Female (p = 1.65e−10)	H16 (0.61)	0.13 | 0.14 (++)
R169*	NextSeq	8783510	471162	5.36	Male (p = 1.02e−08)	U5b2b (0.86)	0.08 | 0.08 (+)
R173	MiSeq	634762	12302	1.94	Male (p = 0.00047)	no consensus	0.12 | 0.10 (+)
R181*	NextSeq	9390239	505956	5.39	Male (p = 1.99e−08)	H2a (0.86)	0.08 | 0.09 (+)
R182*	MiSeq	872056	128204	14.7	Female (p = 2.6e−11)	U5b2b3 (0.53)	0.09 | 0.08 (+)
R183	NextSeq	8142804	46347	0.57	Female (p = 6.32e−10)	U5b2b3 (0.55)	0.10 | 0.11 (+)
R184	NextSeq	9354986	225781	2.41	Female (p = 2.29e−11)	U5b2b (0.59)	0.08 | 0.08 (+)
R186*	MiSeq	446974	35252	7.89	Female (p = 2.89e−10)	L2a1m1 (0.52)	0.08 | 0.09 (+)
R195*	NextSeq	10472395	967756	9.24	Female (p = 1.3e−11)	U3b2 (0.84)	0.11 | 0.11 (++)
R196*	MiSeq	209000	13467	6.44	Male (p = 2.81e−08)	no consensus	0.09 | 0.10 (+)

R201*	MiSeq	857390	33044	3.85	Male (p = 2.64e−08)	J2a2c (0.57)	0.09 | 0.09 (+)
R202*	NextSeq	1160922	82469	7.10	Female (p = 2.09e−11)	H65 (0.62)	0.13 | 0.13 (++)
R21	MiSeq	392898	7378	1.88	Male (p = 3.52e−09)	U5b2b (0.61)	0.07 | 0.06 (+)
R5*	MiSeq	302561	39045	12.9	Female (p = 2.23e−10)	N1b1a2 (0.60)	0.11 | 0.11 (++)
R57	MiSeq	618618	7801	1.26	Male (p = 2.88e−08)	no consensus	0.08 | 0.07 (+)
R59*	NextSeq	9765699	450370	4.61	Male (p = 3.38e−08)	H2a (0.76)	0.13 | 0.13 (++)
R76	NextSeq	8665936	128331	1.48	Male (p = 2.34e−07)	H2a (0.85)	0.09 | 0.09 (+)
R79*	MiSeq	566069	55262	9.76	Female (p = 1.26e−10)	U1a1 (0.54)	0.12 | 0.12 (++)
R8	NextSeq	9963423	15786	0.16	Female (p = 0.00029)	R0a2l (0.52)	0.10 | 0.10 (+)
R84	MiSeq	410271	10925	2.66	Female (p = 4.08e−10)	L0a2 (0.56)	0.12 | 0.12 (+)
R86	MiSeq	680562	3508	0.52	Male (p = 1.63e−05)	J2a2a (0.57)	0.11 | 0.12 (+)
R91	MiSeq	1176266	40693	3.46	Female (p = 2.69e−10)	R0a1a (0.79)	0.09 | 0.09 (+)
R93*	MiSeq	439737	28205	6.41	Male (p = 1.6e−07)	H65 (0.73)	0.11 | 0.09 (+)
R94	NextSeq	10109623	228890	2.26	Female (p = 8.73e−12)	L5a1b (0.82)	0.15 | 0.16 (++)
R95	MiSeq	552430	4971	0.9	Female (p = 6.85e−08)	no consensus	0.13 | 0.12 (+)
R96	MiSeq	208614	6389	3.06	Male (p = 3.49e−07)	no consensus	0.10 | 0.11 (+)
S106	NextSeq	2388448	2129	0.09	unable to be determined	no consensus	0.07 | 0.09 (-)

S109	MiSeq	403143	1090	0.27	Female (p = 0.000195)	no data	0.09 | 0.05 (x)
S110	NextSeq	13202803	16428	0.12	Male (p = 0.00251)	U2eh1 (0.54)	0.02 | 0.02 (x)
S114	MiSeq	346032	5003	1.45	Female (p = 1.41e−09)	X2e1a1 (0.50)	0.11 | 0.12 (-)
S115	MiSeq	1065705	29148	2.73	Male (p = 6.48e−09)	U5b2b (0.57)	0.07 | 0.07 (+)
S120	NextSeq	13353963	9272	0.07	Male (p = 0.0069)	M68a2a (0.50)	0.01 | 0.02 (x)
S126	NextSeq	9744830	10632	0.11	Female (p = 0.023)	H13a1a1d1 (0.53)	0.03 | 0.03 (x)
S131a	NextSeq	11346702	4011	0.04	Male (p = 0.000247)	U1a1b (0.52)	0.02 | 0.02 (x)
S132	NextSeq	8482568	135643	1.60	Female (p = 3.19e−11)	L2a1d1 (0.56)	0.08 | 0.08 (+)
S133*	MiSeq	275266	13612	4.95	Female (p = 5.93e−10)	L2a1d1 (0.55)	0.04 | 0.04 (-)
S136	MiSeq	618732	4152	0.67	Female (p = 1.17e−08)	no consensus	0.06 | 0.08 (-)
S143	MiSeq	261256	1323	0.51	Female (p = 2.4e−06)	no consensus	0.07 | 0.03 (x)
S144	MiSeq	327332	1731	0.53	Male? (p = 0.042)	L2a1d2 (0.52)	0.06 | 0.08 (-)
S149*	NextSeq	12738113	690030	5.42	Female (p = 7.09e−12)	H2a (0.83)	0.07 | 0.07 (+)
S159*	MiSeq	312079	61918	19.84	Male (p = 9.51e−07)	L2a1d1 (0.72)	0.09 | 0.08 (+)
S17*	MiSeq	351496	19435	5.53	Female (p = 1.56e−08)	T1b2 (0.57)	0.10 | 0.09 (+)
S171	NextSeq	11700	119	1.02	unable to be determined	no consensus	no data
S173	MiSeq	614621	2323	0.38	unable to be determined	no data	0.04 | 0.03 (x)

S174	NextSeq	9853295	14011	0.14	Male (p = 0.0367)	no data	0.01 | 0.01 (x)
S175	NextSeq	19606359	9933	0.05	Male? (p = 0.0506)	no consensus	0.01 | 0.01 (x)
S181	NextSeq	14233826	7689	0.05	Male (p = 0.0472)	no data	0.01 | 0.01 (x)
S182	MiSeq	520242	2346	0.45	Male (p = 0.000322)	no consensus	0.06 | 0.07 (-)
S183	NextSeq	28767039	24499	0.09	Male (p = 0.0301)	L5 (0.51)	0.01 | 0.01 (x)
S190	NextSeq	8640852	3958	0.05	Male? (p = 0.0528)	L0a2a2a1 (0.51)	0.03 | 0.04 (x)
S193	NextSeq	13954509	19609	0.14	Male (p = 1.6e−05)	H35 (0.59)	0.05 | 0.04 (-)
S198*	MiSeq	679307	80760	11.89	Male (p = 5.5e−08)	L1b1a8 (0.54)	0.06 | 0.06 (+)
S199a	MiSeq	321649	1146	0.36	Male (p = 0.00224)	L2a1m1a (0.50)	0.05 | 0.06 (x)
S199b	NextSeq	5553	196	3.53	Female (p = 0.0122)	no data	no data
S2	NextSeq	11437044	22053	0.19	Female (p = 1.34e−07)	N1b1a4 (0.58)	0.08 | 0.07 (+)
S201	NextSeq	43928	651	1.48	Female (p = 0.000328)	L0a2c (0.50)	0.10 | 0.08 (-)
S208*	MiSeq	1375987	310255	22.55	Female (p = 1.89e−11)	J2a2 (0.82)	0.06 | 0.06 (+)
S217	MiSeq	436082	4624	1.06	Male (p = 1.85e−08)	no consensus	0.09 | 0.09 (+)
S218	NextSeq	10336579	260028	2.52	Female (p = 1.27e−11)	N1b2 (0.58)	0.07 | 0.07 (+)
S230	NextSeq	9026041	3586	0.04	unable to be determined	no consensus	0.01 | 0.01 (x)

S235	NextSeq	8564855	162991	1.90	Female (p = 2.85e−11)	L2a1d1 (0.74)	0.06 | 0.06 (+)
S238	MiSeq	294574	428	0.15	Male (p = 0.0177)	no data	0.03 | 0.04 (x)
S24	MiSeq	234277	1079	0.46	Female (p = 0.00113)	no consensus	0.14 | 0.06 (-)
S242	MiSeq	661118	1238	0.19	Male (p = 0.0118)	no data	0.01 | 0.03 (x)
S25	MiSeq	577935	2484	0.43	Male (p = 2.15e−05)	no consensus	0.11 | 0.12 (-)
S27*	MiSeq	316464	59261	18.73	Male (p = 8.21e−09)	U5b2b (0.58)	0.07 | 0.06 (+)
S33*	NextSeq	9597747	614629	6.40	Male (p = 7.12e−09)	L0a1a (0.80)	0.09 | 0.09 (+)
S37	NextSeq	4168392	171096	4.10	Female (p = 1.58e−11)	U5b2b (0.79)	0.11 | 0.12 (++)
S38	NextSeq	17681811	10173	0.06	Male? (p = 0.0573)	no consensus	0.01 | 0.01 (x)
S42a	MiSeq	1502941	2059	0.14	unable to be determined	no consensus	0.07 | 0.06 (-)
S42b	MiSeq	865066	8782	1.02	Female (p = 5.76e−09)	no consensus	0.10 | 0.08 (+)
S50*	NextSeq	9695589	1431080	14.76	Male (p = 1.15e−08)	L0a1a (0.85)	0.05 | 0.05 (+)
S53	MiSeq	361869	3800	1.05	Female (p = 2.05e−06)	L1c2a3a (0.51)	0.09 | 0.06 (+)
S57	NextSeq	2648	43	1.62	Female? (p = 0.0506)	no consensus	no data
S60	MiSeq	413804	1499	0.36	unable to be determined	no data	0.02 | 0.02 (x)
S68a*	NextSeq	11762033	1836921	15.62	Female (p = 4.38e−11)	L2a1d1 (0.85)	0.09 | 0.09 (+)
S69	NextSeq	6441	16	0.25	Female (p = 0.00044)	no data	no data

S70	NextSeq	1885	14	0.74	unable to be determined	no data	no data
S73*	MiSeq	429216	103680	24.16	Male (p = 5.45e−08)	L2a1d1 (0.57)	0.08 | 0.08 (+)
S75	MiSeq	318763	813	0.26	Male? (p = 0.0457)	no data	0.01 | 0.01 (x)
S78	MiSeq	224174	8793	3.92	Male (p = 5.96e−08)	L2a1m1a (0.51)	0.09 | 0.09 (+)
S79*	MiSeq	399354	38864	9.73	Male (p = 3.23e−07)	L5a1b (0.58)	0.07 | 0.07 (+)
S81*	NextSeq	9221125	1297503	14.07	Male (p = 4.26e−09)	L2a1d1 (0.85)	.08 | .09 (+)
S87*	MiSeq	457836	111512	24.36	Female (p = 5.06e−10)	L2a5 (0.52)	0.10 | 0.09 (+)
S89*	MiSeq	700103	135517	19.36	Female (p = 4.01e−12)	H2a1e1 (0.58)	0.07 | 0.07 (+)

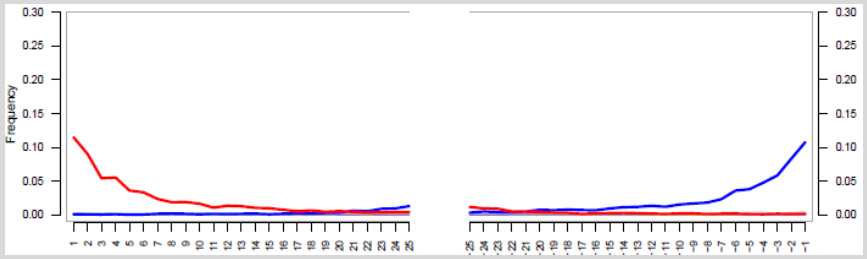
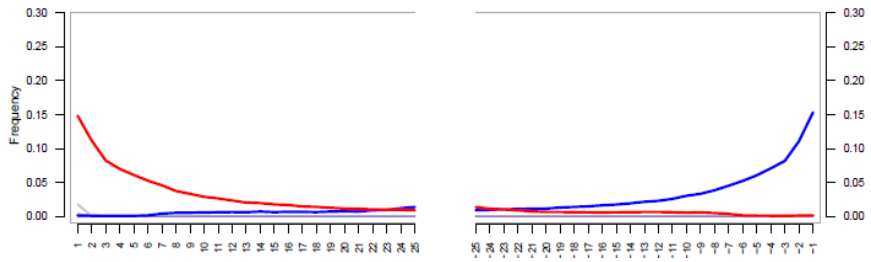
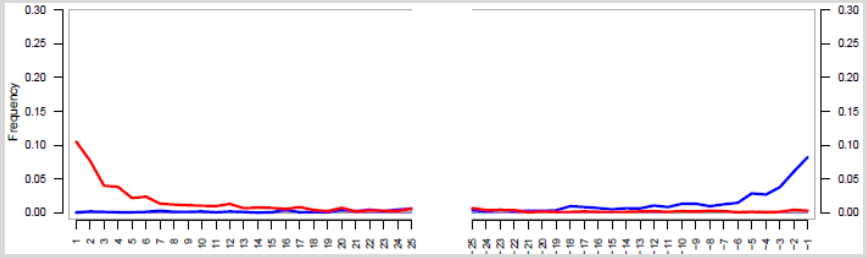
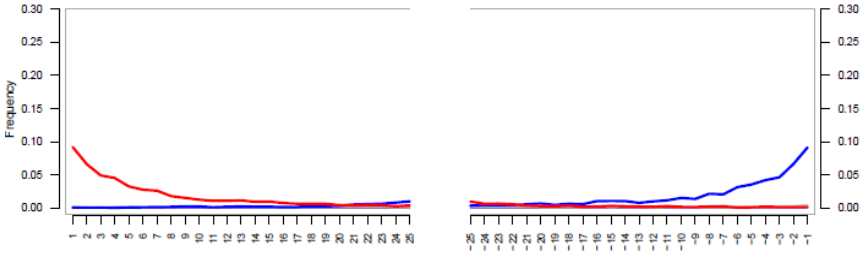
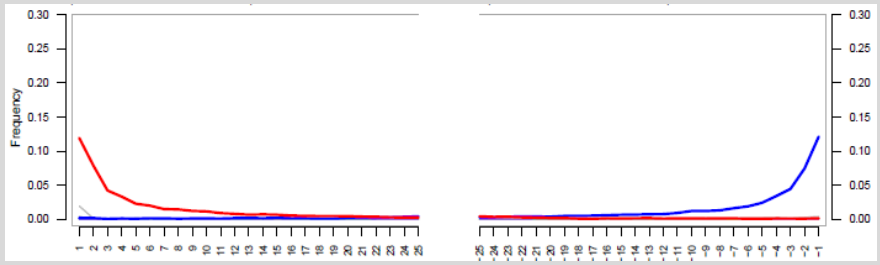
SUPPLEMENT 3: MAPDAMAGE PLOTS FROM NINETY-NINE SCREENED KULUBNARTI SAMPLES

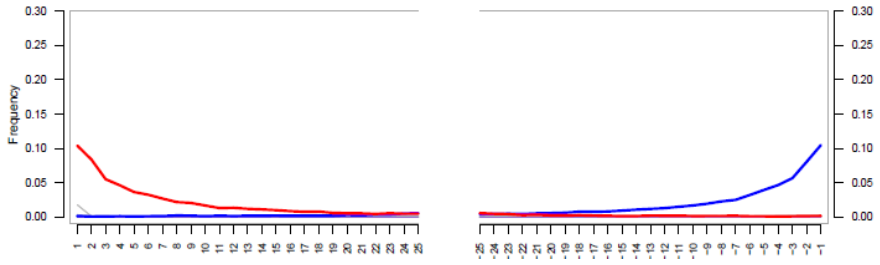
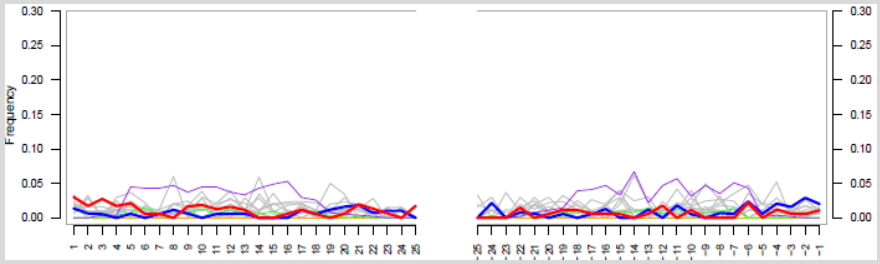
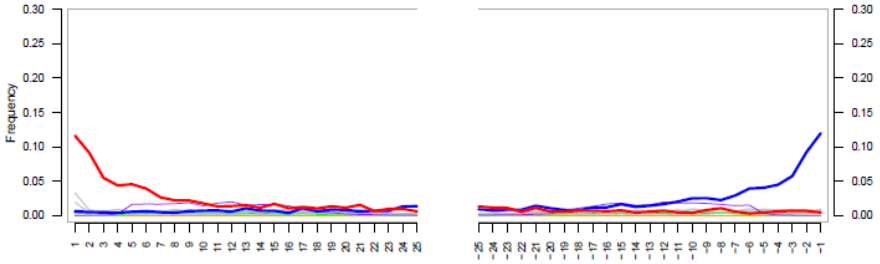
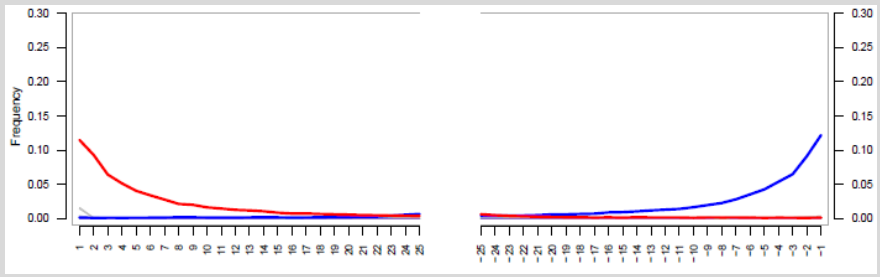
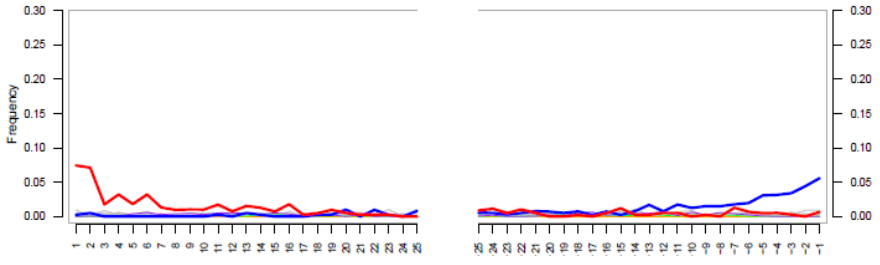
Table S3.1 presents the mapDamage plot (Ginolhac et al. 2011; Jónsson et al. 2013) for each of the 99 screened Kulubnarti Nubian samples. Along with the plot, it provides the quantified frequency of damage at the last base pair of the 5'- and 3'-ends of the DNA molecules as well as the author's visual assessment of the damage pattern across the DNA molecule.

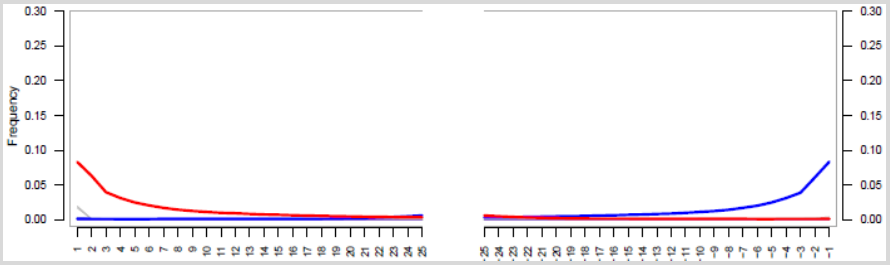
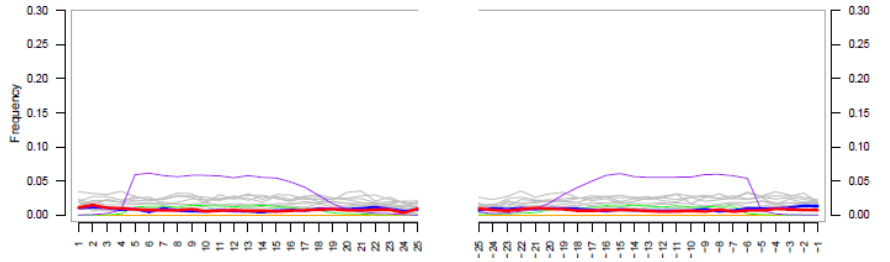
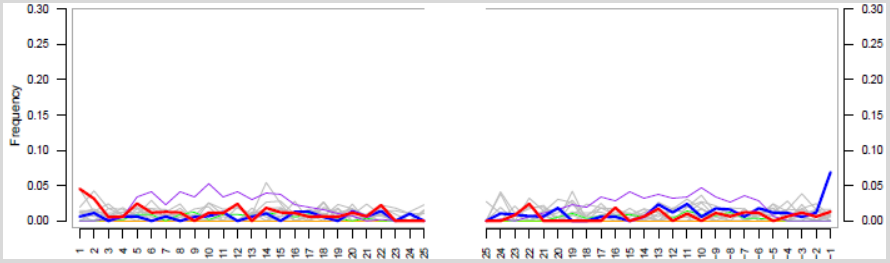
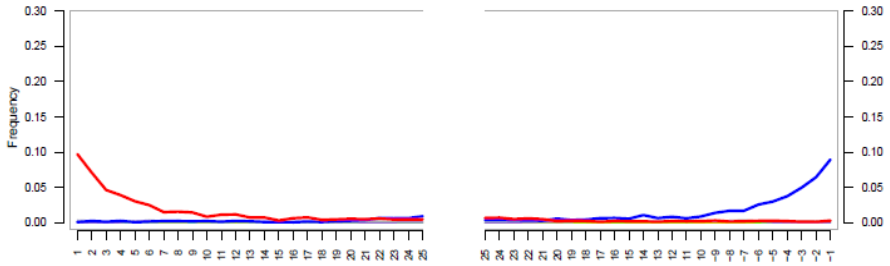
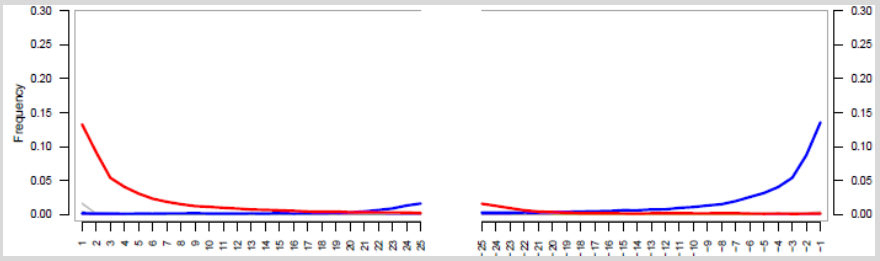
For this visual assessment, "++" indicates that the damage pattern is strongly consistent with that expected for ancient DNA; "+" indicates that the damage pattern is consistent with that expected for ancient DNA; "-" indicates either a damage pattern consistent with that expected for ancient DNA but a low frequency of damage (<5%) at the ends of the molecule or an expected frequency of damage (>5%) at the ends of the molecule but a damage pattern throughout the molecule inconsistent with ancient DNA; and "x" indicates no evident damage, suggesting that the majority of DNA is modern, contaminating DNA. "No data" indicates that there were not enough reads for mapDamage analysis to take place.

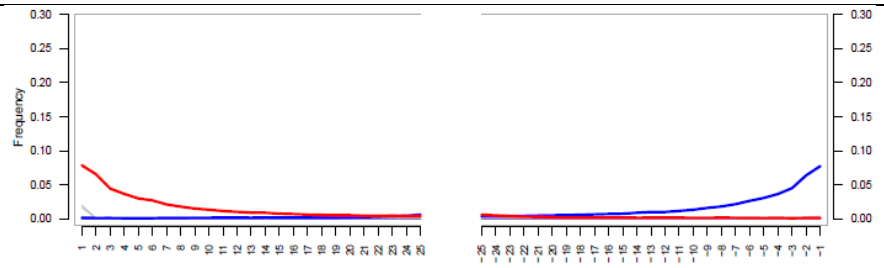
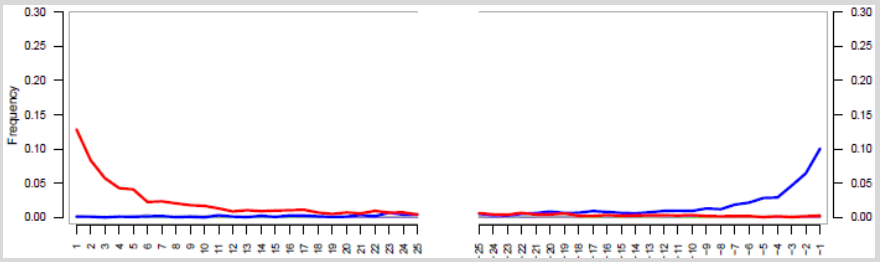
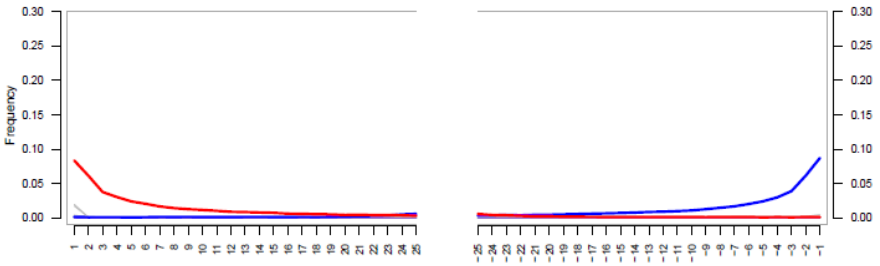
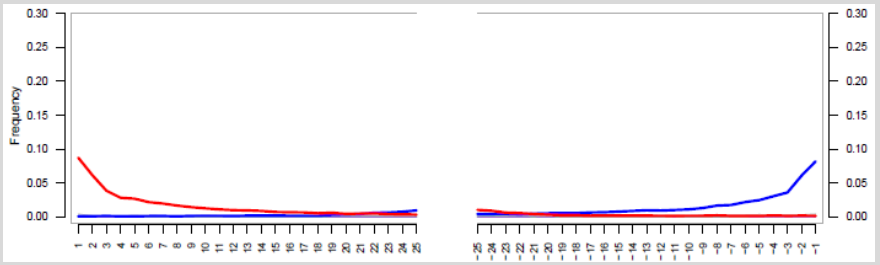
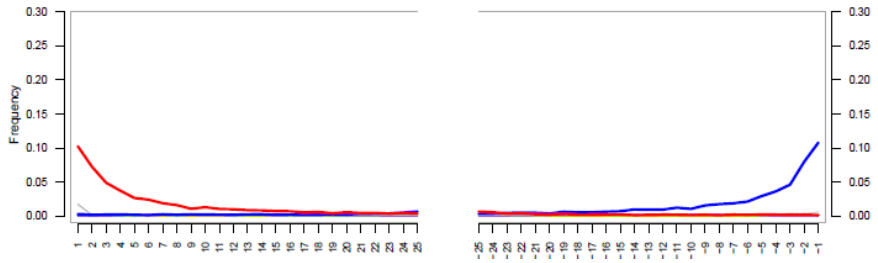
All sample IDs starting with "R" designate the individual to be a member of the R cemetery; sample IDs starting with "S" designate the individual to be a member of the S cemetery. An asterisk next to the sample ID designates the sample as one of the 30 best-preserved individuals that were sent to the Reich Lab for additional screening and potential 1240K SNP capture.

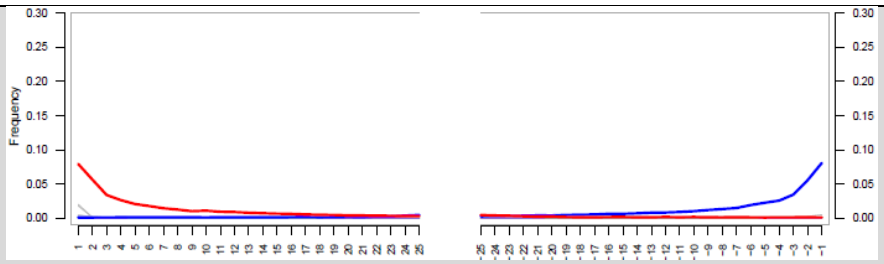
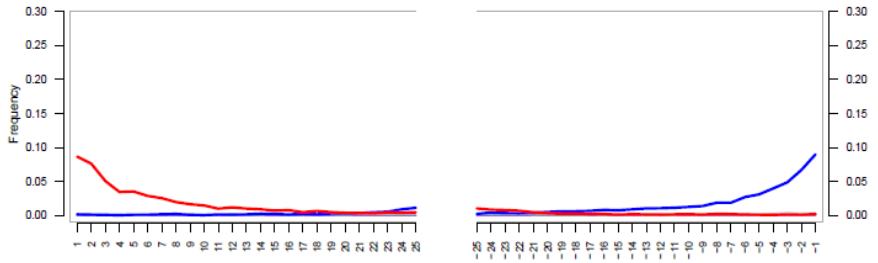
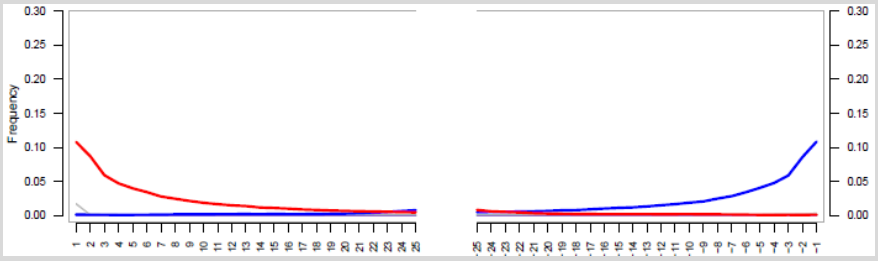
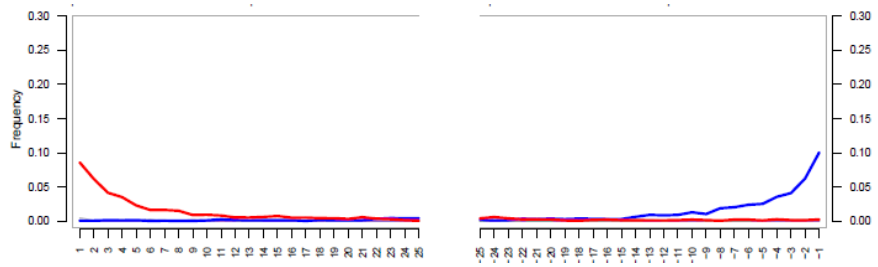
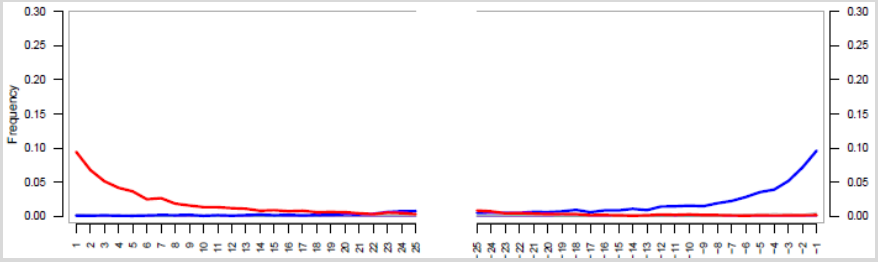
Table S3.1: mapDamage data for each of the 99 screened Kulubnarti samples.

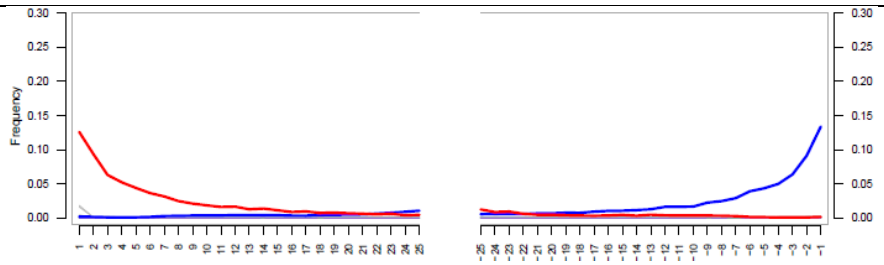
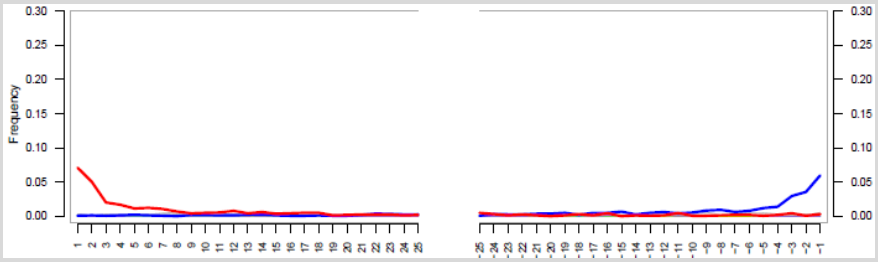
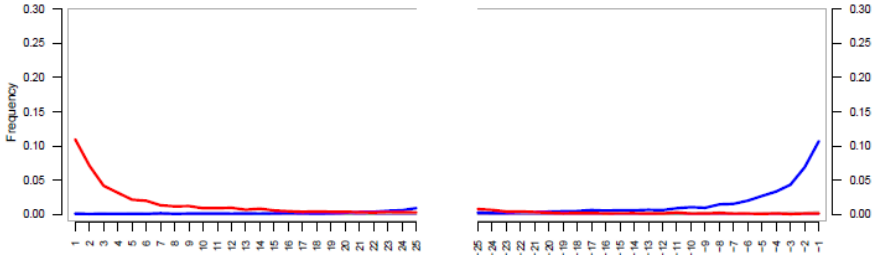
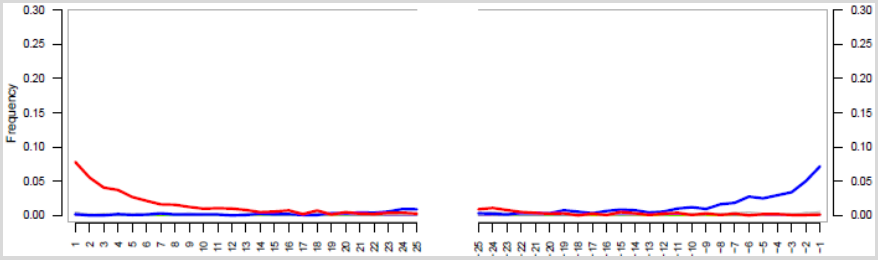
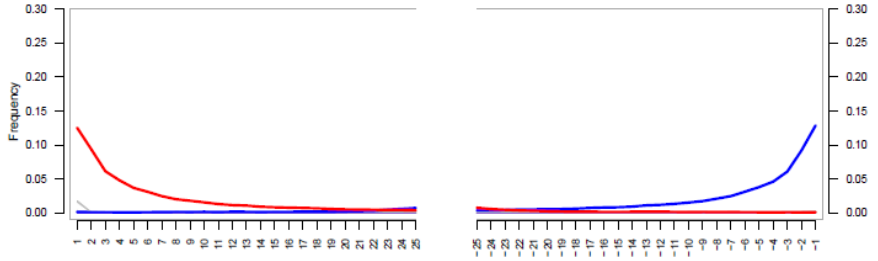
Sample ID	mapDamage plot	map-Damage (5' | 3')
R101*		0.12 | 0.10 (+)
R103		0.15 | 0.15 (++)
R123		0.10 | 0.08 (+)
R124*		0.09 | 0.09 (+)
R125		0.12 | 0.12 (++)

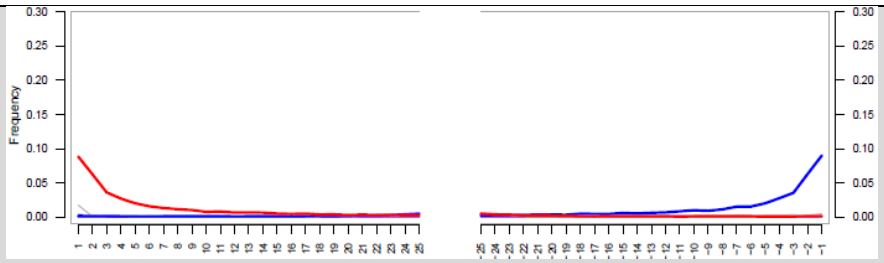
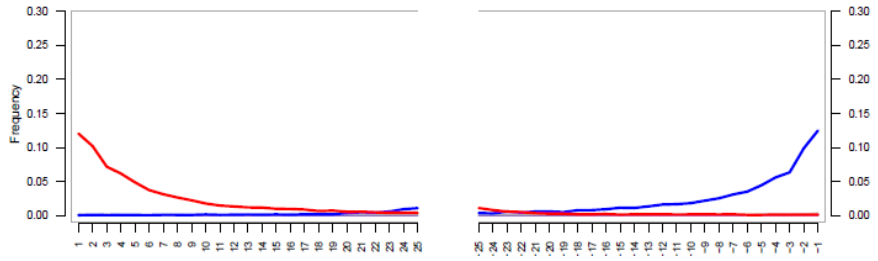
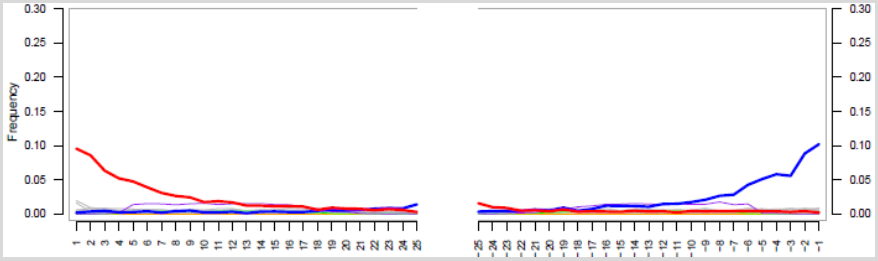
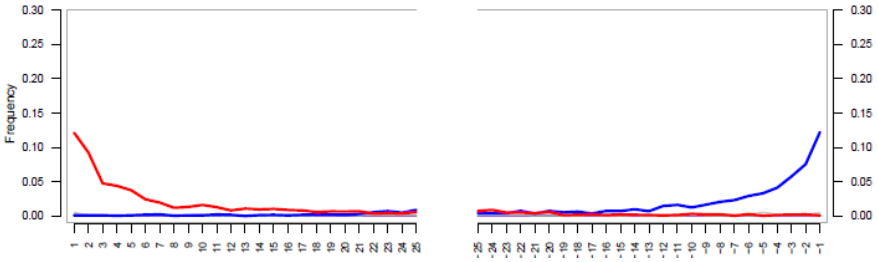
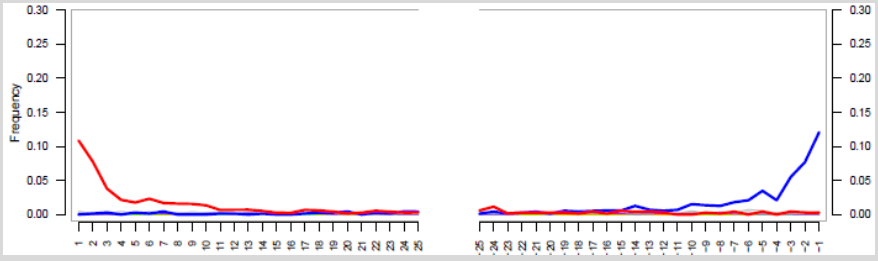
R133		0.10 | 0.10 (+)
R143		0.03 | 0.02 (x)
R144		0.12 | 0.12 (+)
R15		0.11 | 0.12 (++)
R150		0.07 | 0.06 (-)

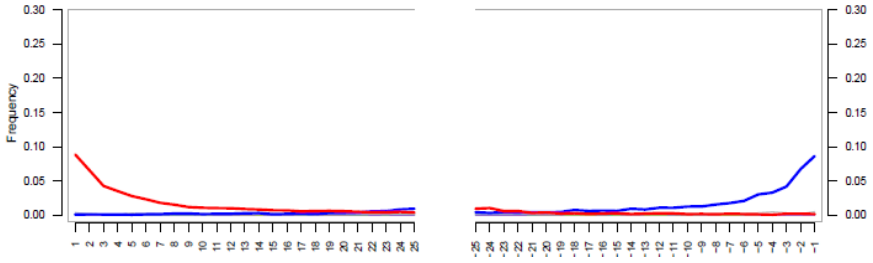
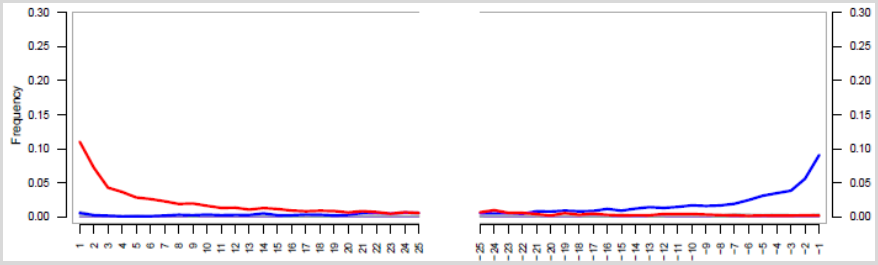
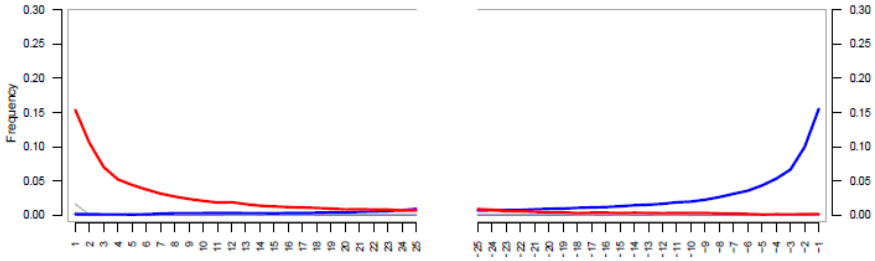
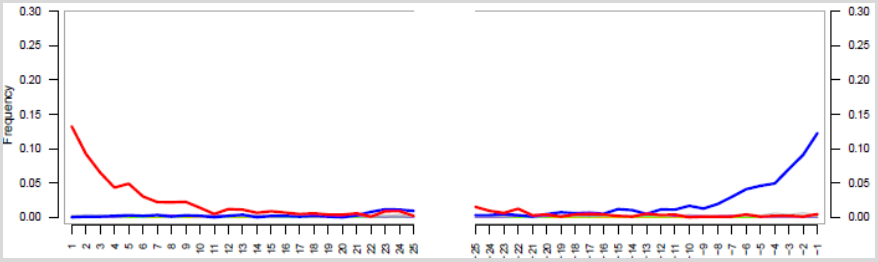
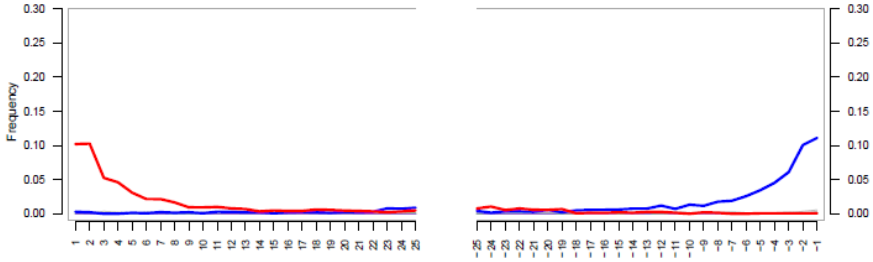
R152*		0.08 | 0.08 (+)
R153		0.01 | 0.01 (x)
R157		0.05 | 0.07 (x)
R159		0.09 | 0.09 (+)
R164		0.13 | 0.14 (++)
R169*		0.08 | 0.08 (+)

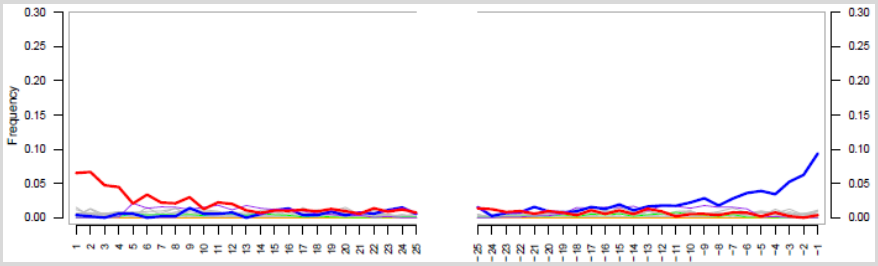
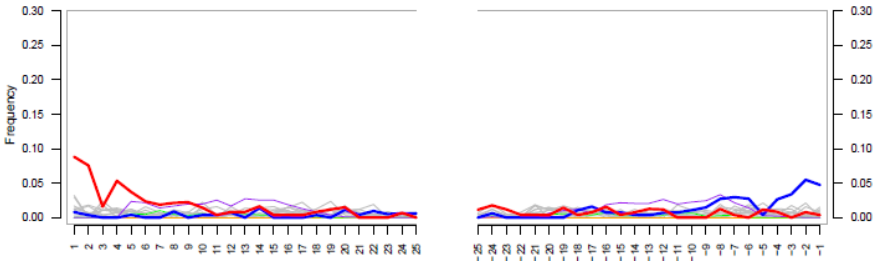
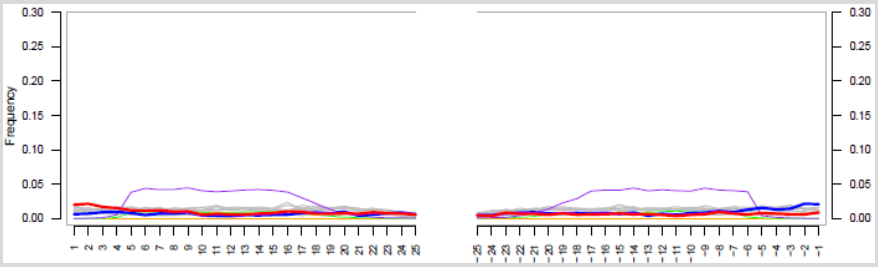
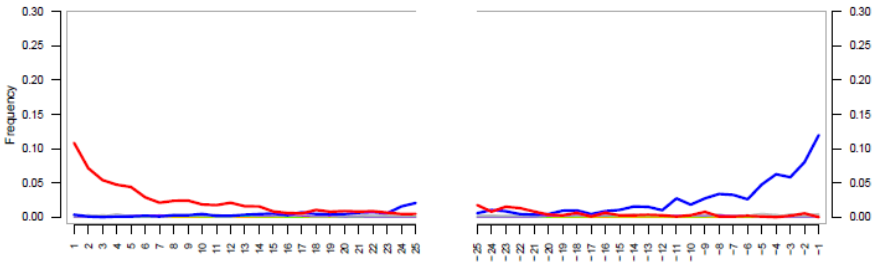
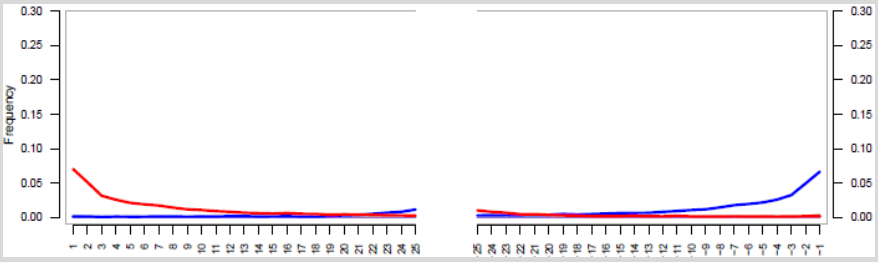
		
R173		0.12 | 0.10 (+)
R181*		0.08 | 0.09 (+)
R182*		0.09 | 0.08 (+)
R183		0.10 | 0.11 (+)
R184		0.08 | 0.08 (+)

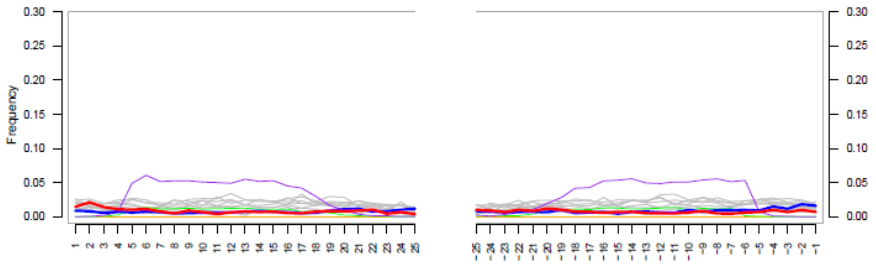
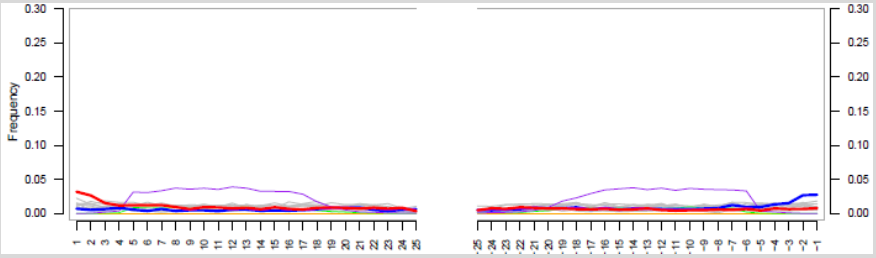
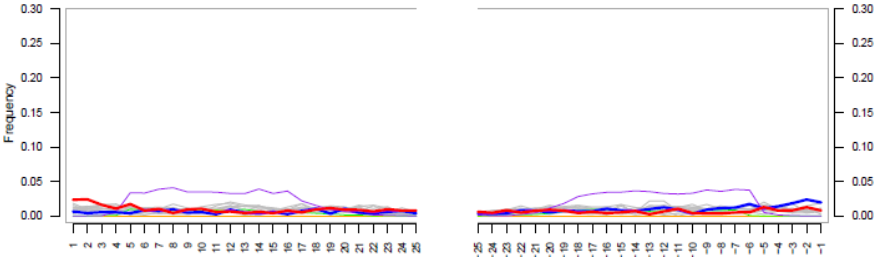
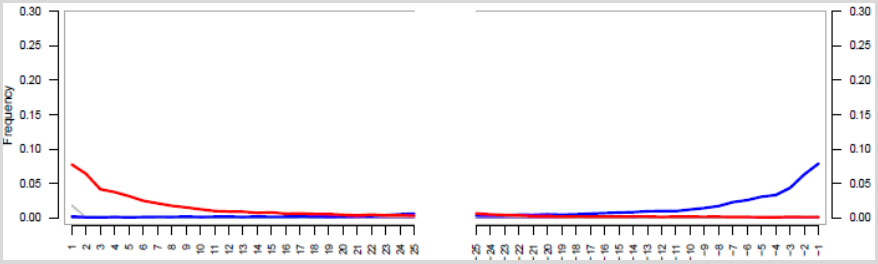
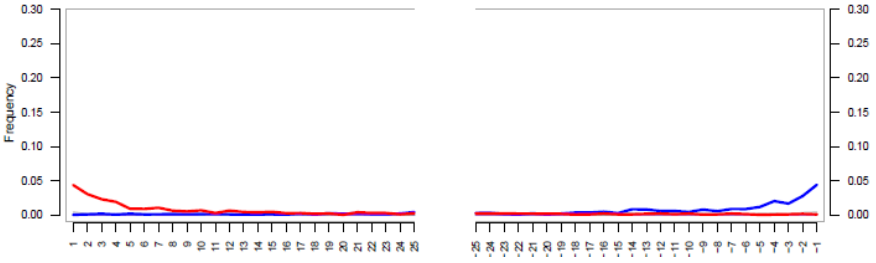
		
R186*		0.08 | 0.09 (+)
R195*		0.11 | 0.11 (++)
R196*		0.09 | 0.10 (+)
R201*		0.09 | 0.09 (+)
R202*		0.13 | 0.13 (++)

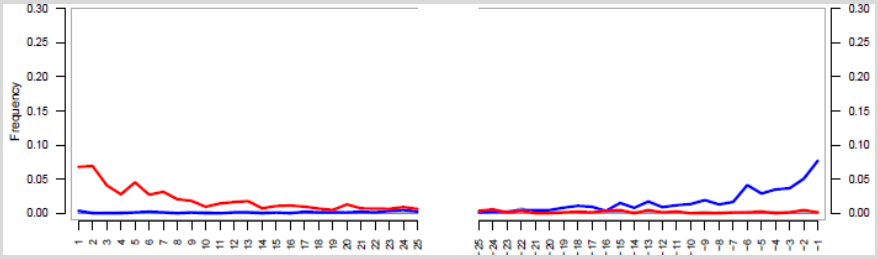
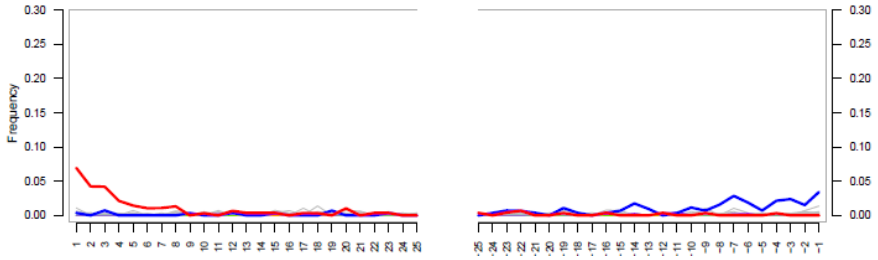
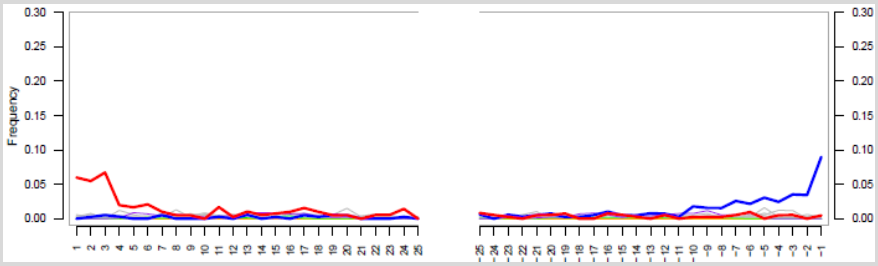
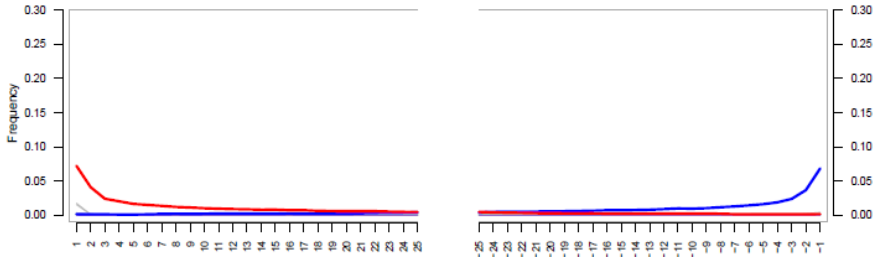
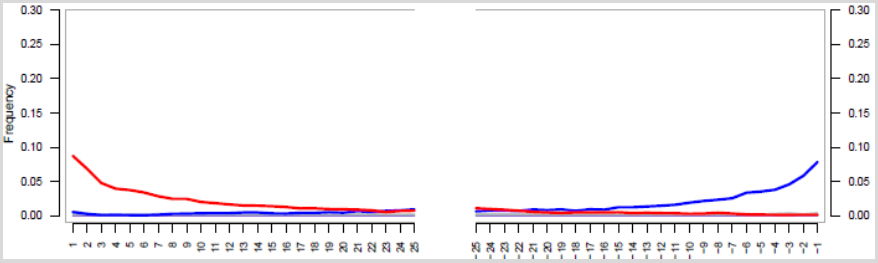
		
R21		0.07 | 0.06 (+)
R5*		0.11 | 0.11 (++)
R57		0.08 | 0.07 (+)
R59*		0.13 | 0.13 (++)
R76		0.09 | 0.09 (+)

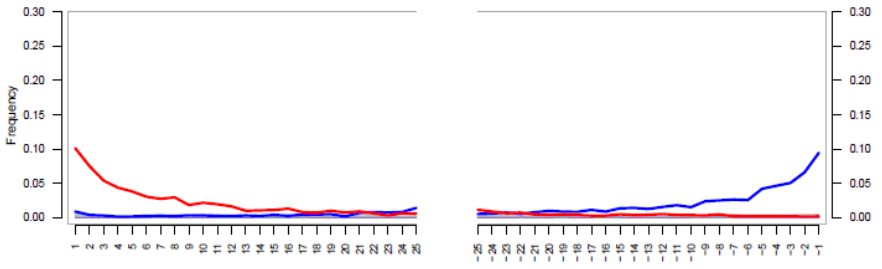
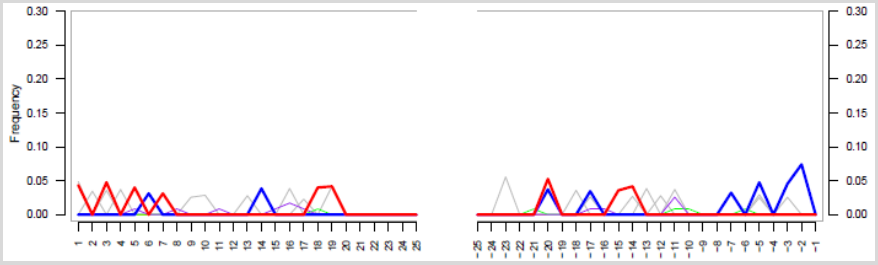
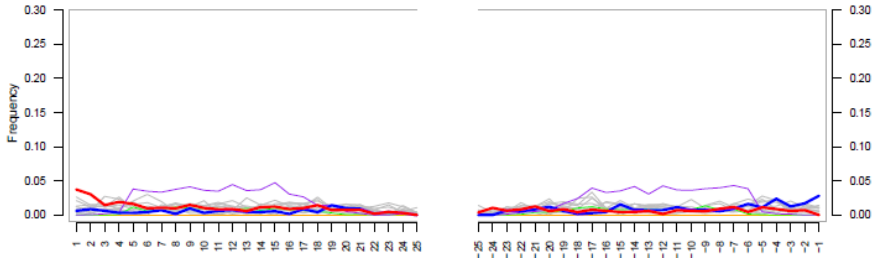
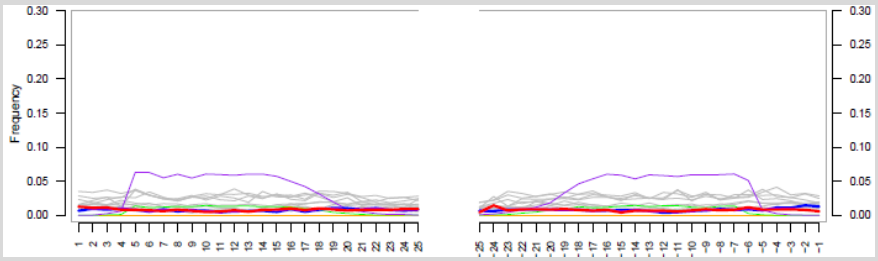
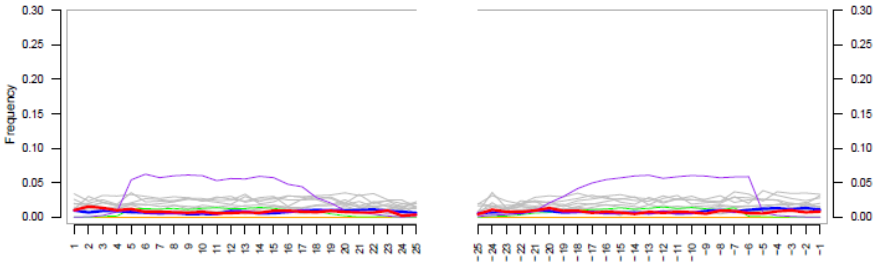
		
R79*		0.12 | 0.12 (++)
R8		0.10 | 0.10 (+)
R84		0.12 | 0.12 (+)
R86		0.11 | 0.12 (+)

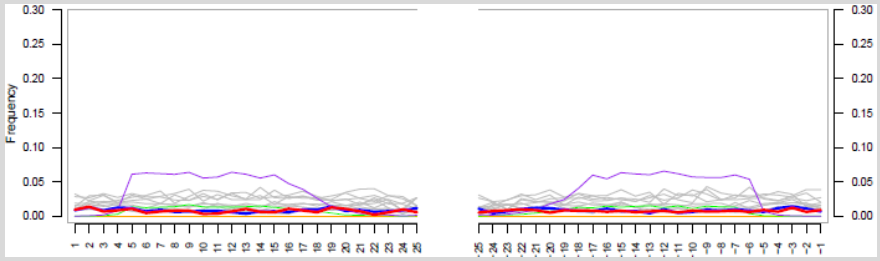
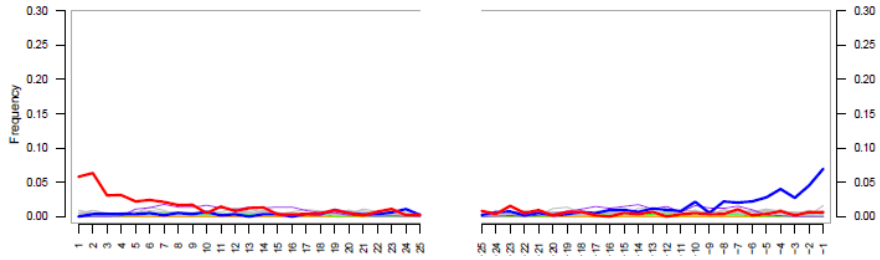
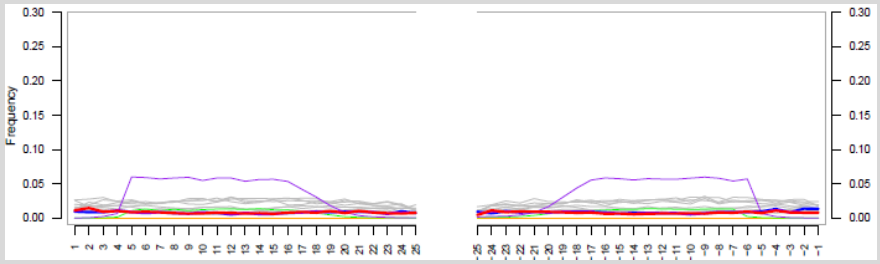
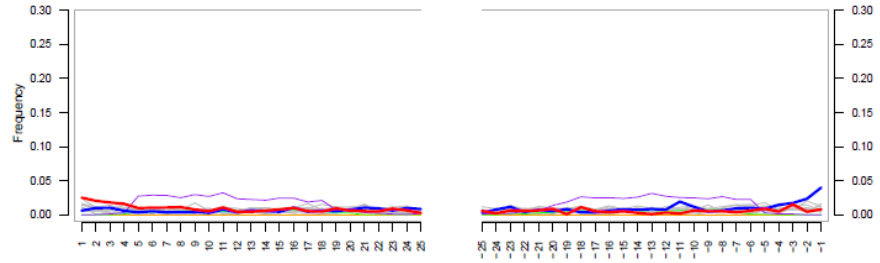
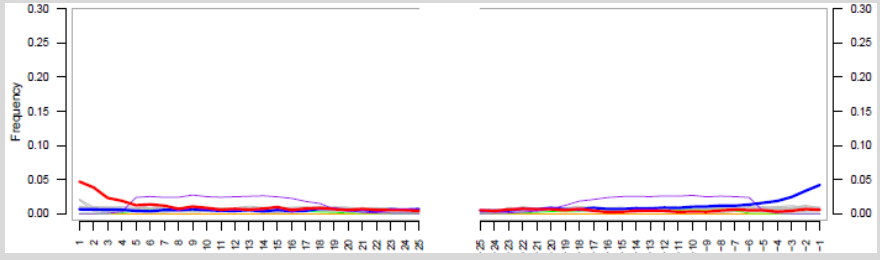
R91		0.09 | 0.09 (+)
R93*		0.11 | 0.09 (+)
R94		0.15 | 0.16 (++)
R95		0.13 | 0.12 (+)
R96		0.10 | 0.11 (+)

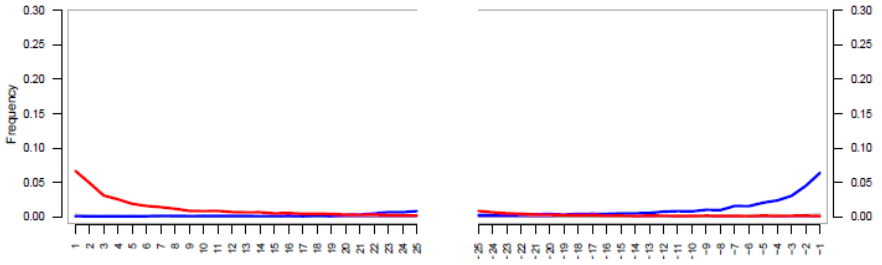
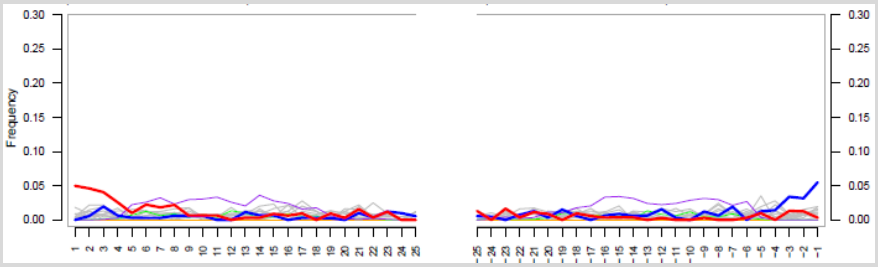
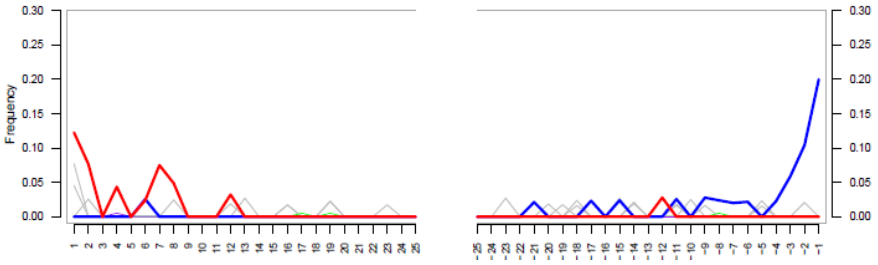
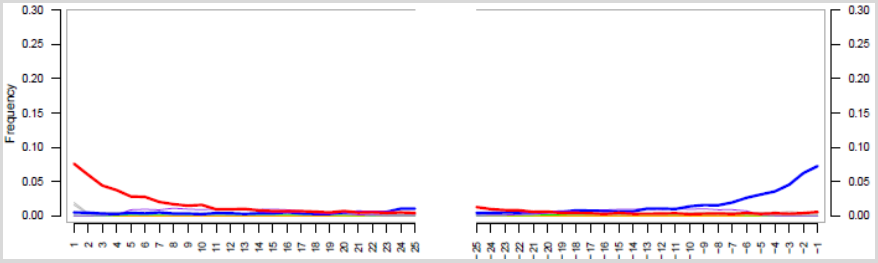
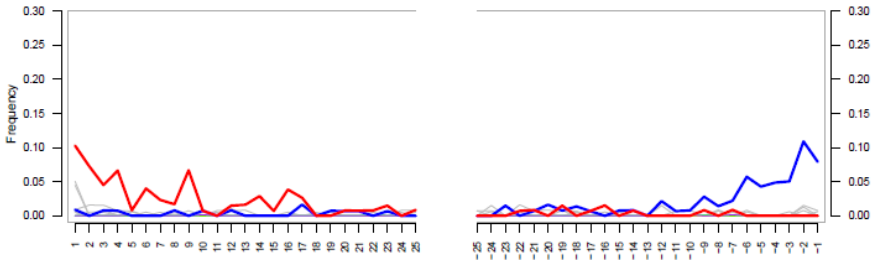
S106		0.07 | 0.09 (-)
S109		0.09 | 0.05 (x)
S110		0.02 | 0.02 (x)
S114		0.11 | 0.12 (-)
S115		0.07 | 0.07 (+)

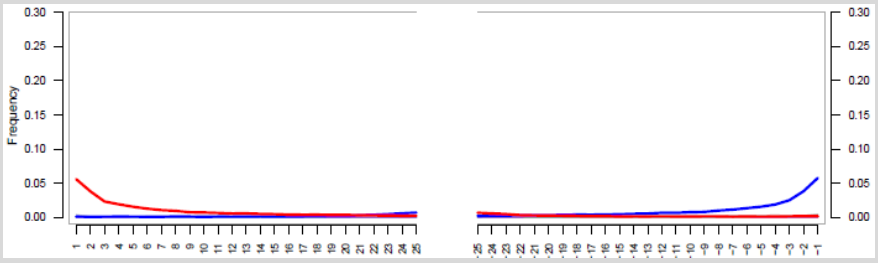
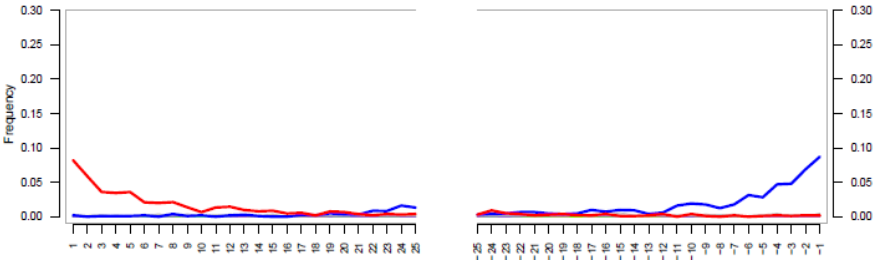
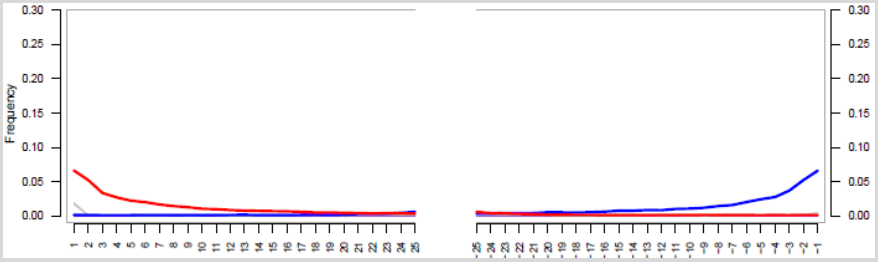
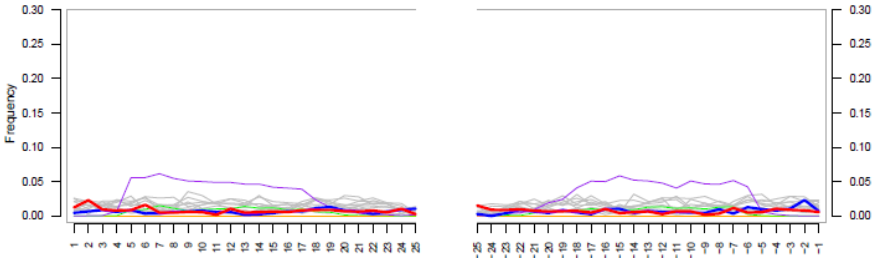
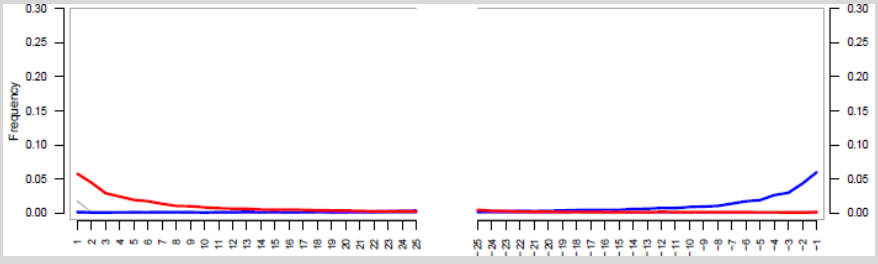
S120		0.01 | 0.02 (x)
S126		0.03 | 0.03 (x)
S131a		0.02 | 0.02 (x)
S132		0.08 | 0.08 (+)
S133*		0.04 | 0.04 (-)

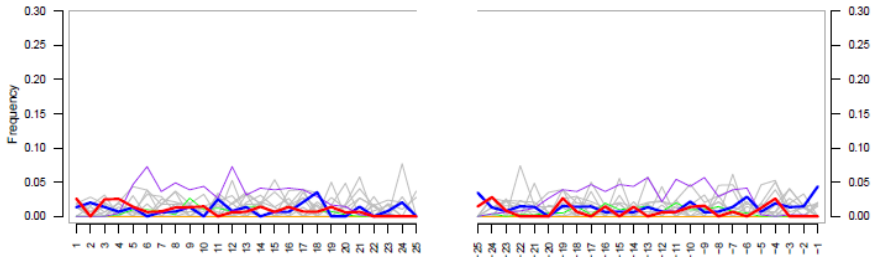
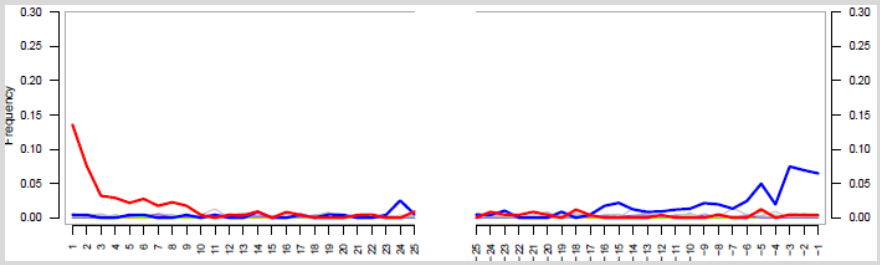
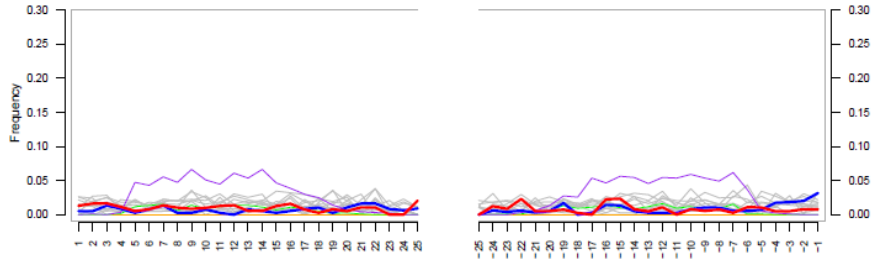
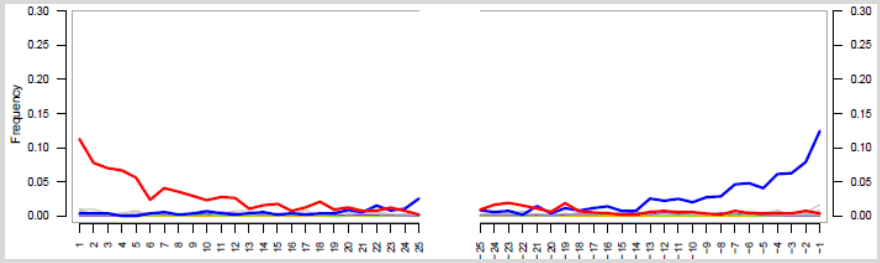
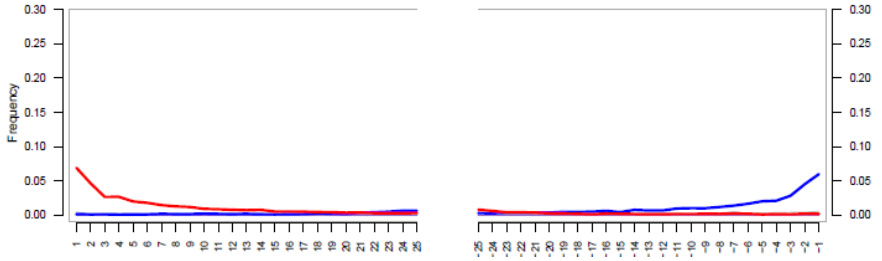
S136		0.06 | 0.08 (-)
S143		0.07 | 0.03 (x)
S144		0.06 | 0.08 (-)
S149*		0.07 | 0.07 (+)
S159*		0.09 | 0.08 (+)

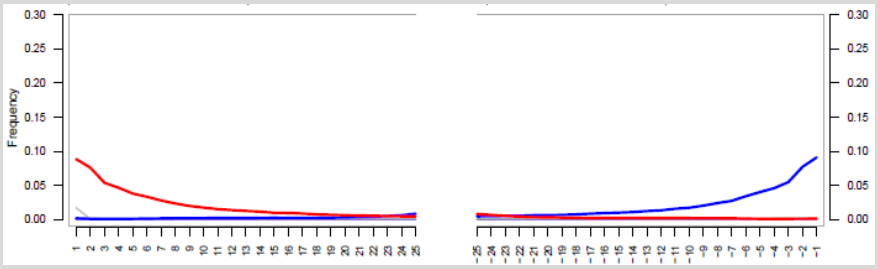
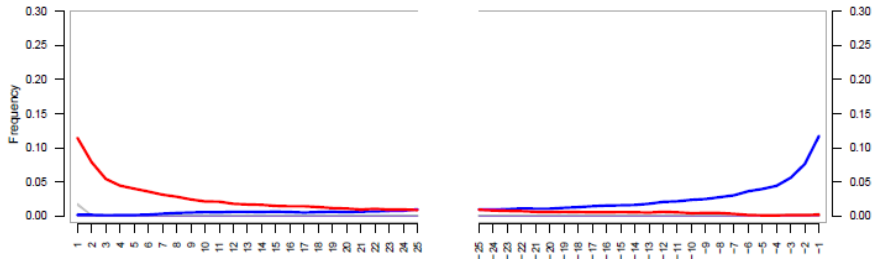
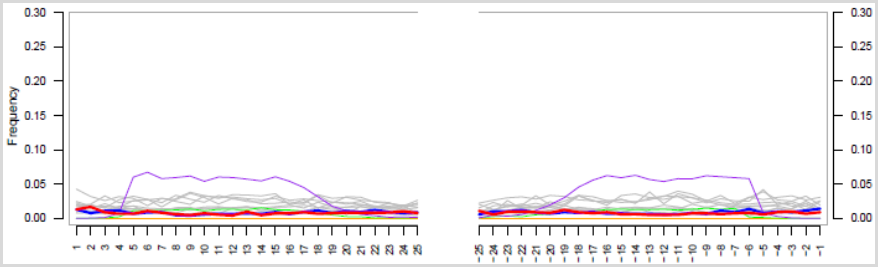
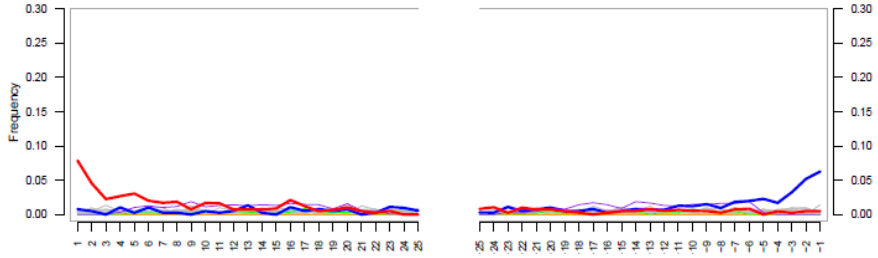
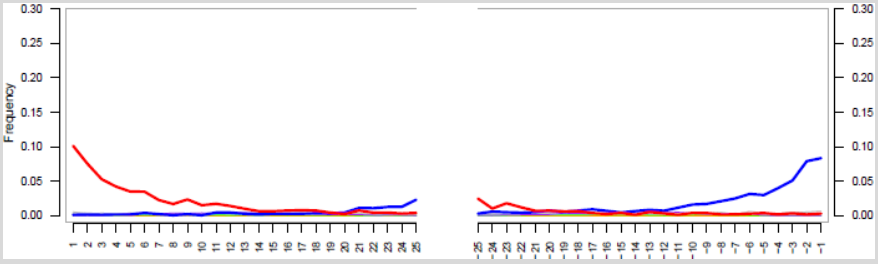
S17*		0.10 | 0.09 (+)
S171		no data
S173		0.04 | 0.03 (x)
S174		0.01 | 0.01 (x)
S175		0.01 | 0.01 (x)

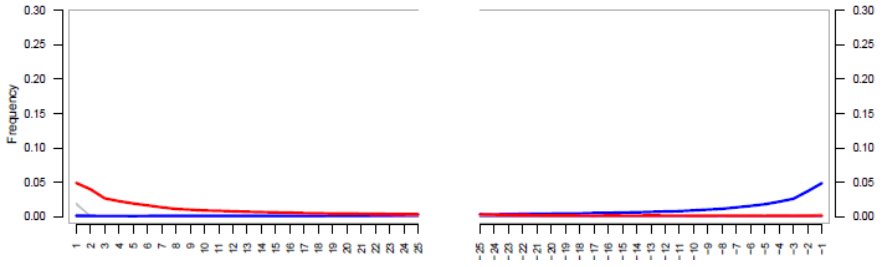
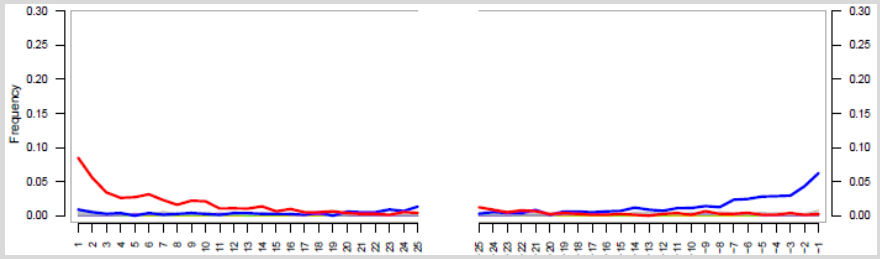
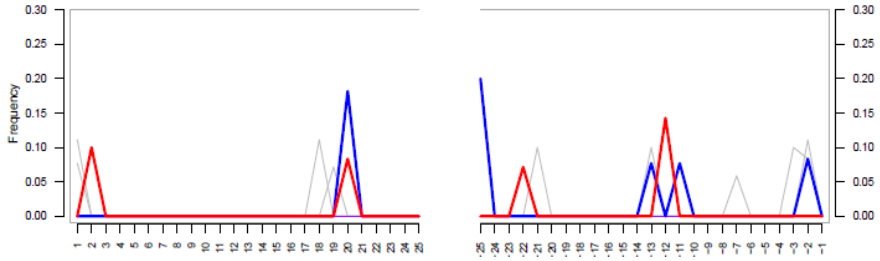
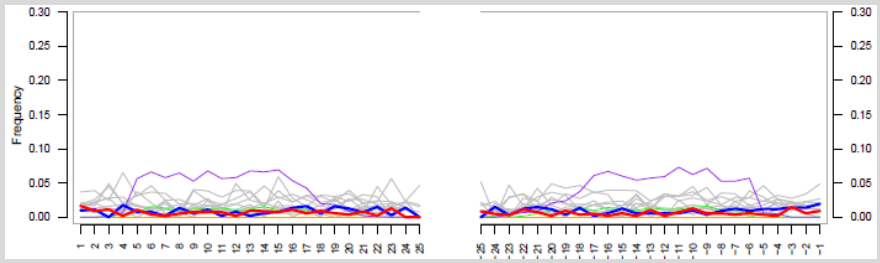
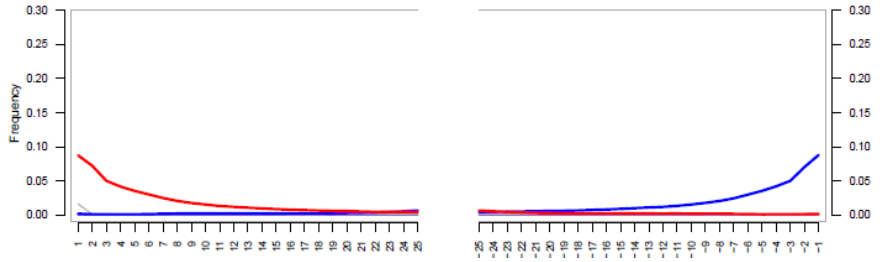
S181		0.01 | 0.01 (x)
S182		0.06 | 0.07 (-)
S183		0.01 | 0.01 (x)
S190		0.03 | 0.04 (x)
S193		0.05 | 0.04 (-)

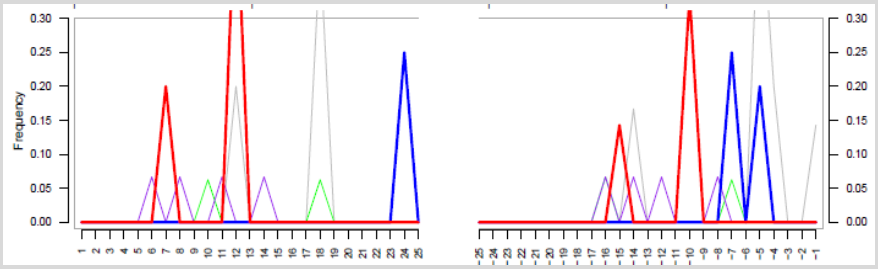
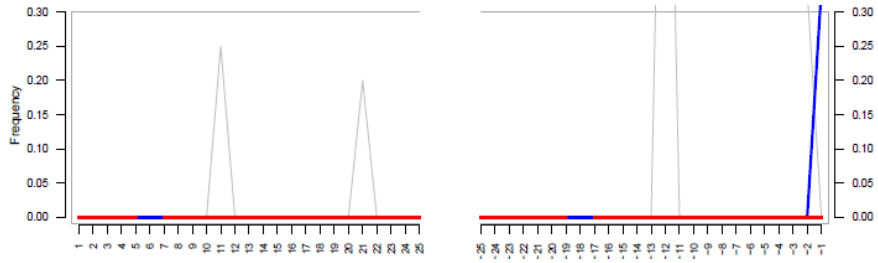
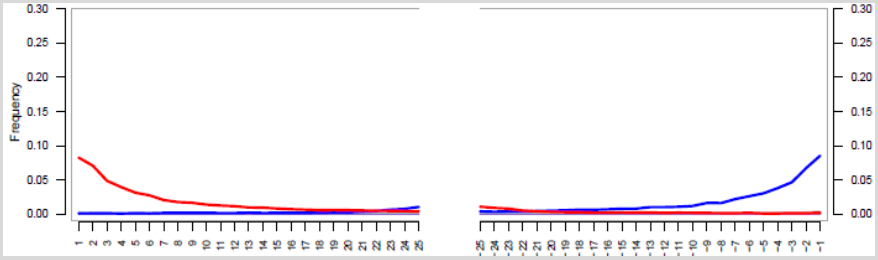
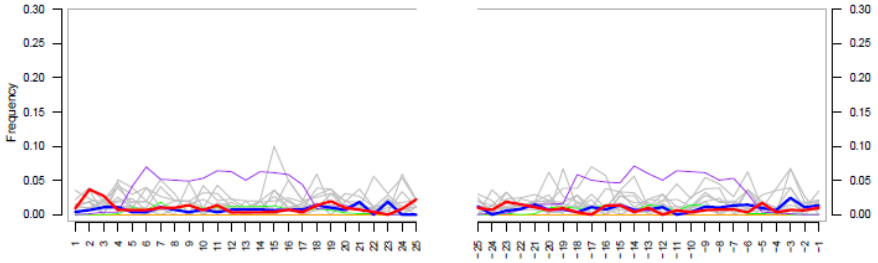
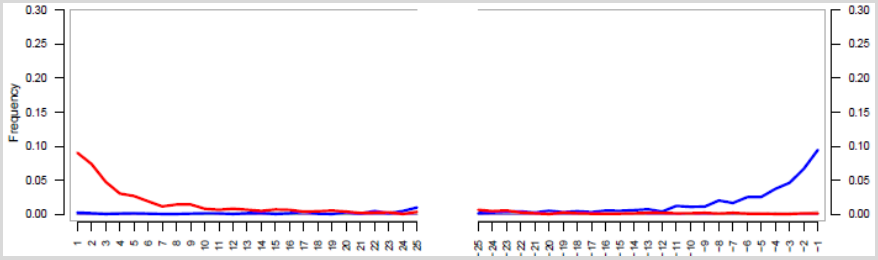
S198*		0.06 | 0.06 (+)
S199a		0.05 | 0.06 (x)
S199b		no data
S2		0.08 | 0.07 (+)
S201		0.10 | 0.08 (-)

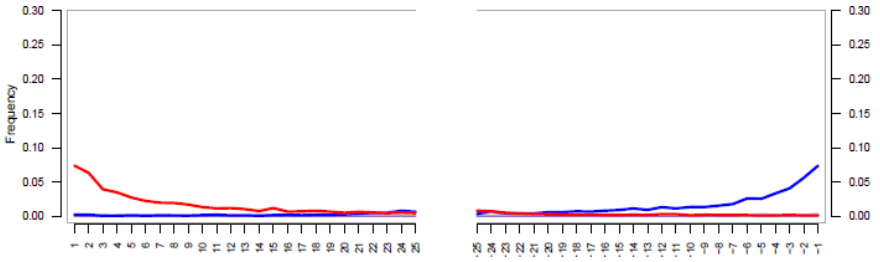
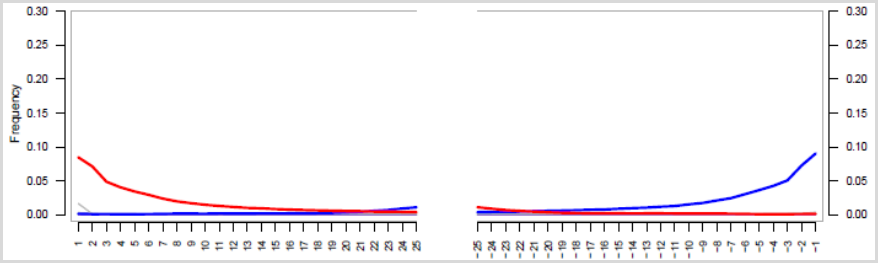
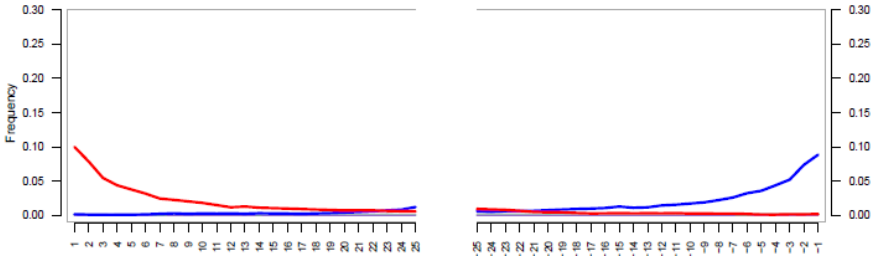
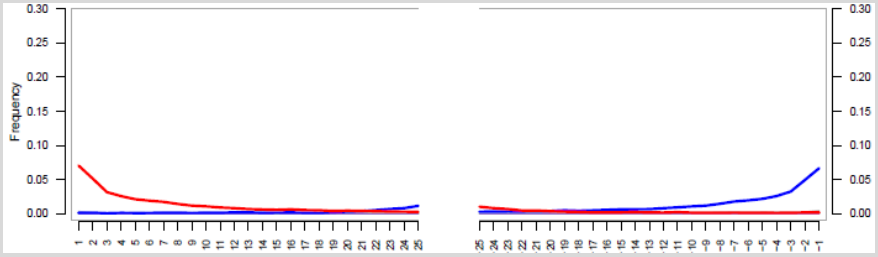
S208*		0.06 | 0.06 (+)
S217		0.09 | 0.09 (+)
S218		0.07 | 0.07 (+)
S230		0.01 | 0.01 (x)
S235		0.06 | 0.06 (+)

S238		0.03 | 0.04 (x)
S24		0.14 | 0.06 (-)
S242		0.01 | 0.03 (x)
S25		0.11 | 0.12 (-)
S27*		0.07 | 0.06 (+)

S33*		0.09 | 0.09 (+)
S37		0.11 | 0.12 (++)
S38		0.01 | 0.01 (x)
S42a		0.07 | 0.06 (-)
S42b		0.10 | 0.08 (+)

S50*		0.05 | 0.05 (+)
S53		0.09 | 0.06 (+)
S57		no data
S60		0.02 | 0.02 (x)
S68a*		0.09 | 0.09 (+)

S69		no data
S70		no data
S73*		0.08 | 0.08 (+)
S75		0.01 | 0.01 (x)
S78		0.09 | 0.09 (+)

S79*		0.07 | 0.07 (+)
S81*		.08 | .09 (+)
S87*		0.10 | 0.09 (+)
S89*		0.07 | 0.07 (+)

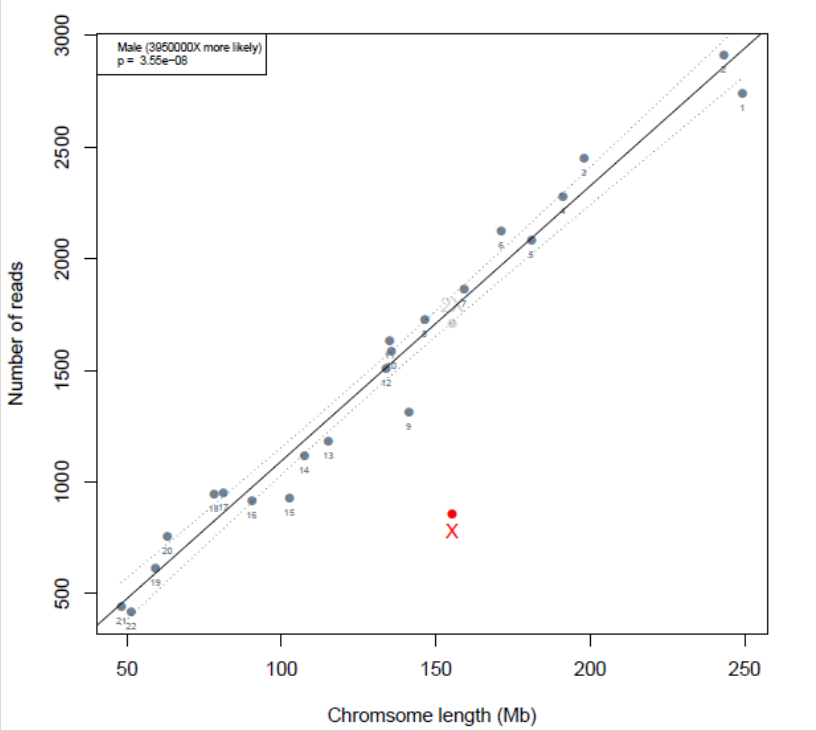
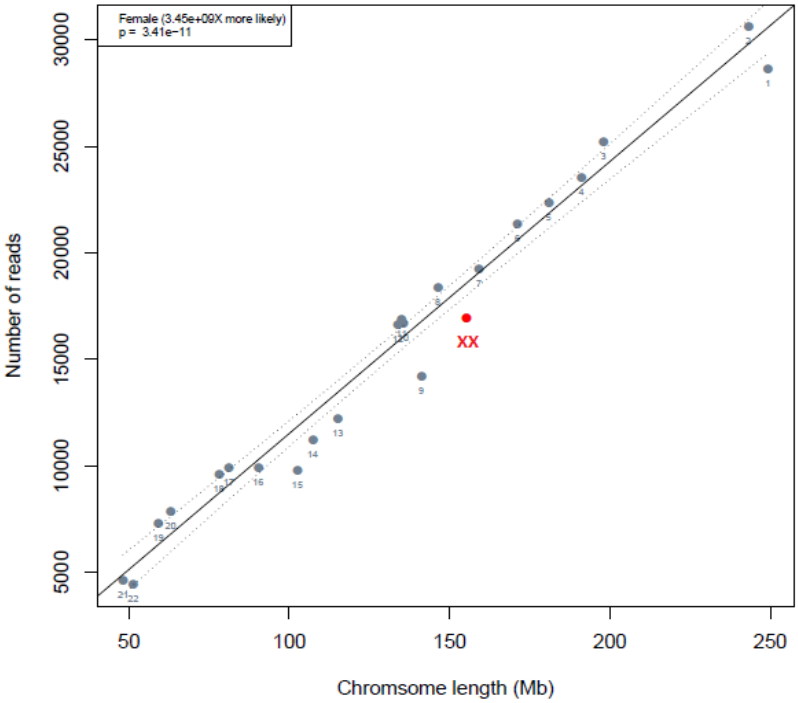
SUPPLEMENT 4: SEXING PLOTS FROM NINETY-NINE SCREENED KULUBNARTI SAMPLES

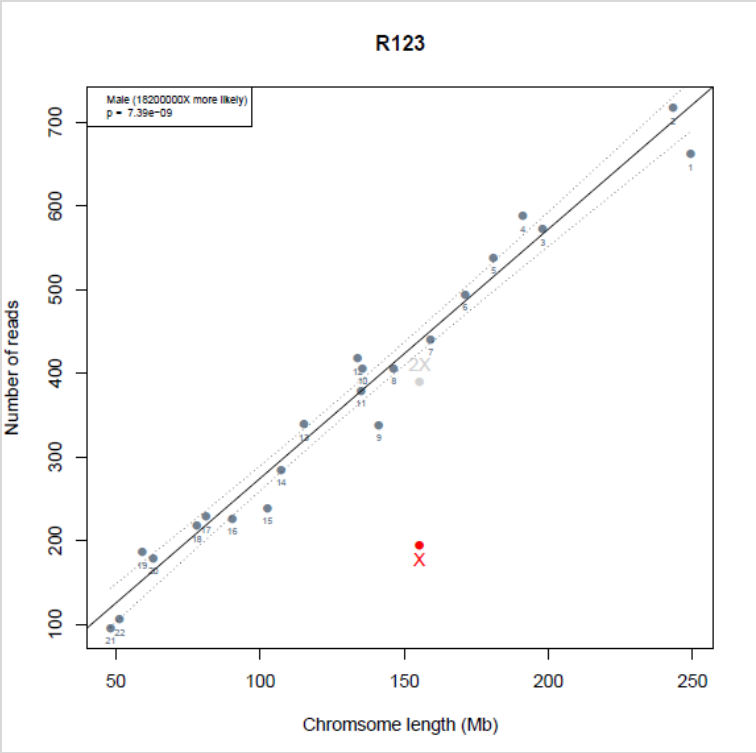
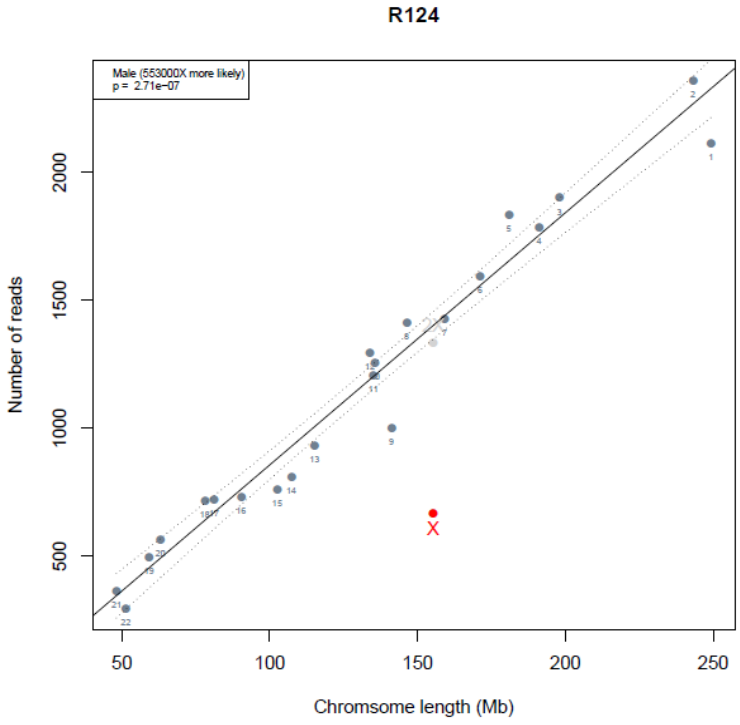
Table S4.1 presents the sexing plots for each of the 99 screened Kulubnarti samples. By plotting the number of reads for each chromosome by chromosome length, sex can be determined by assessing if an individual has one X chromosome (and is therefore male), or two X chromosomes (and is therefore female). An individual's sex is represented by a red dot with either one X (male) or two X's (female) on each plot.

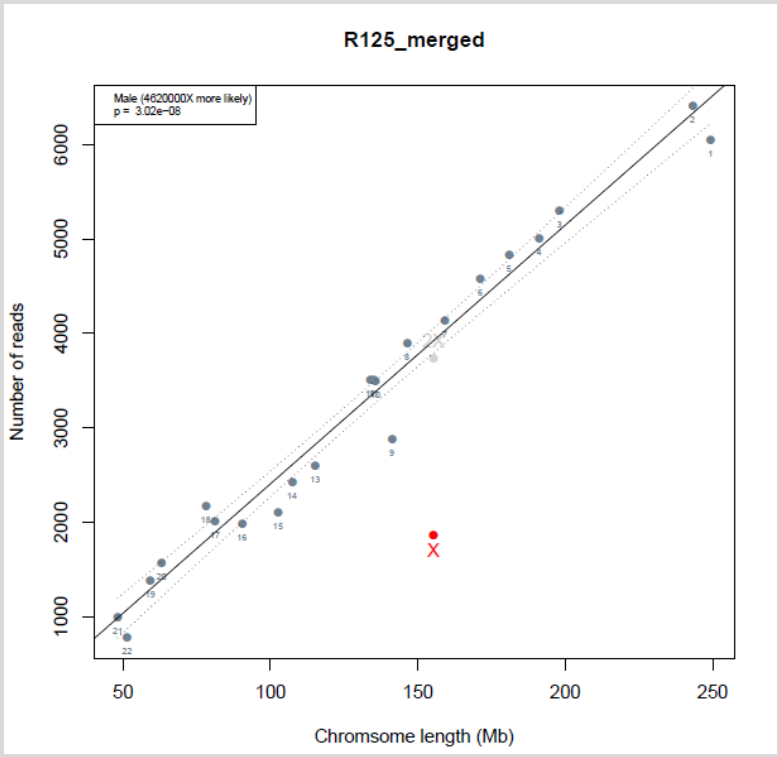
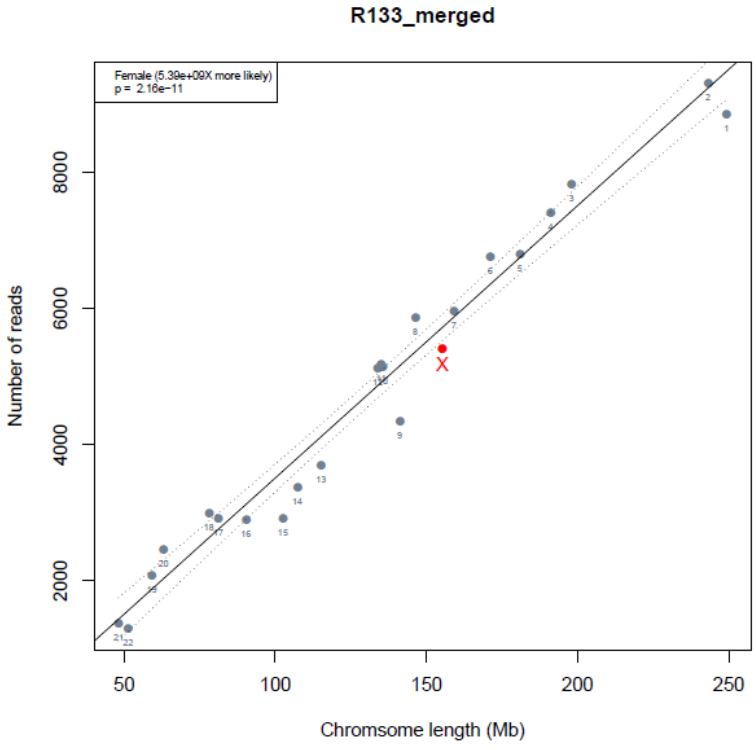
Along with the sexing plot, a p-score associated with the assignment of an individual to a particular sex is provided. Sex is listed as "unable to be determined" when not enough reads aligned to the X chromosome for sex assignment.

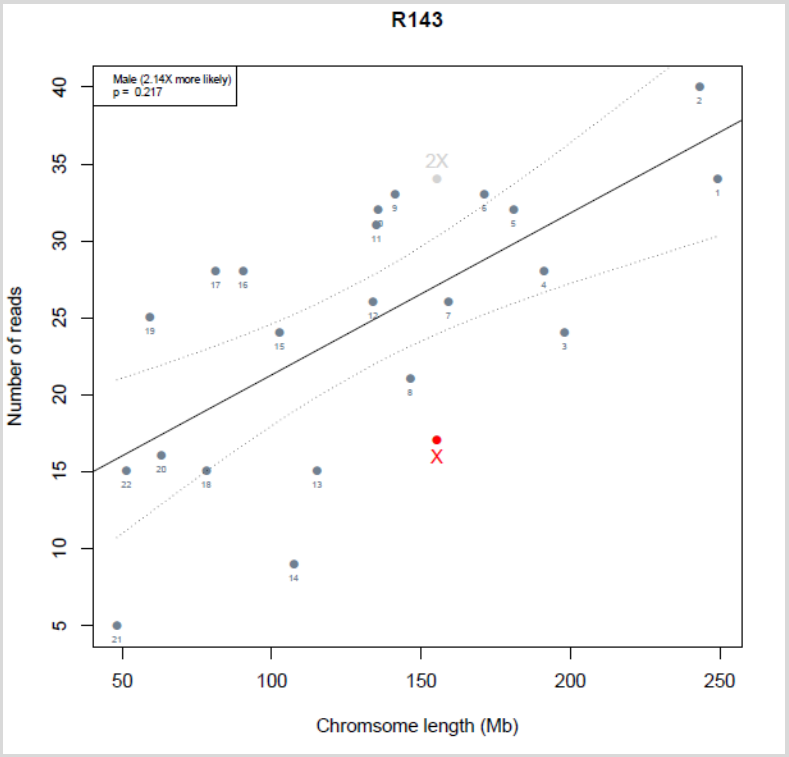
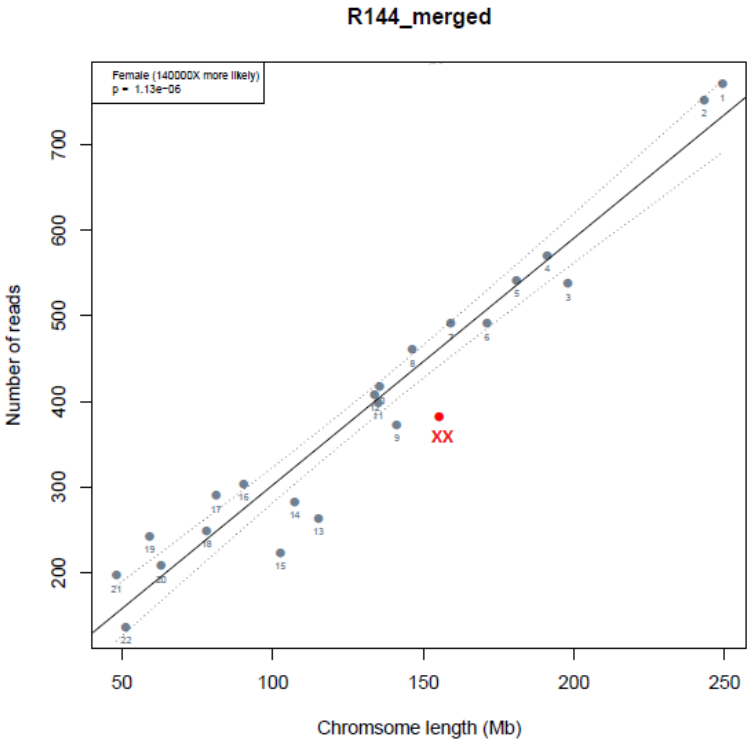
All sample IDs starting with "R" designate the individual to be a member of the R cemetery; sample IDs starting with "S" designate the individual to be a member of the S cemetery. An asterisk next to the sample ID designates the sample as one of the 30 best-preserved individuals that were sent to the Reich Lab for additional screening and potential 1240K SNP capture.

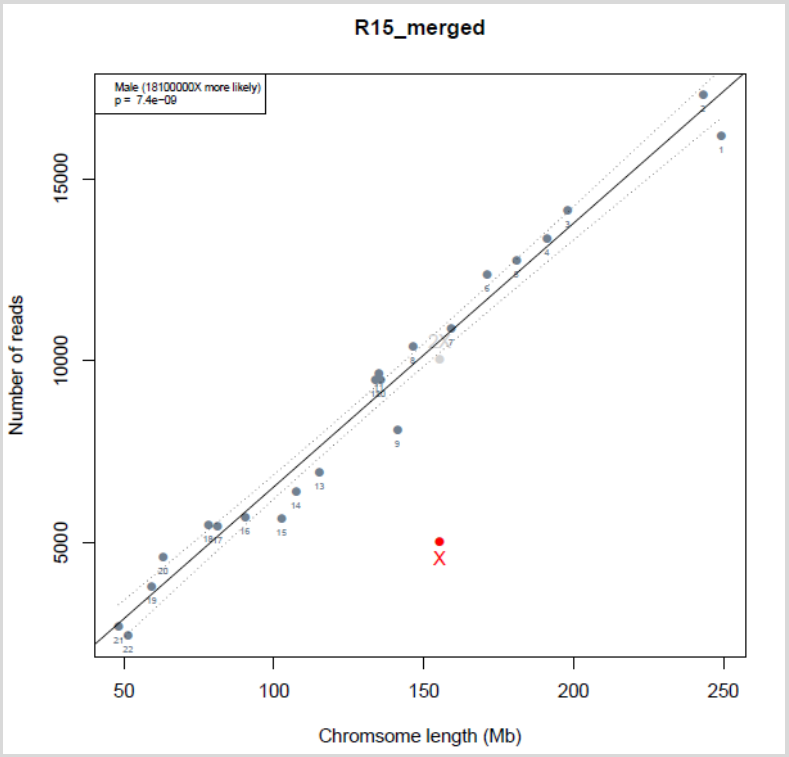
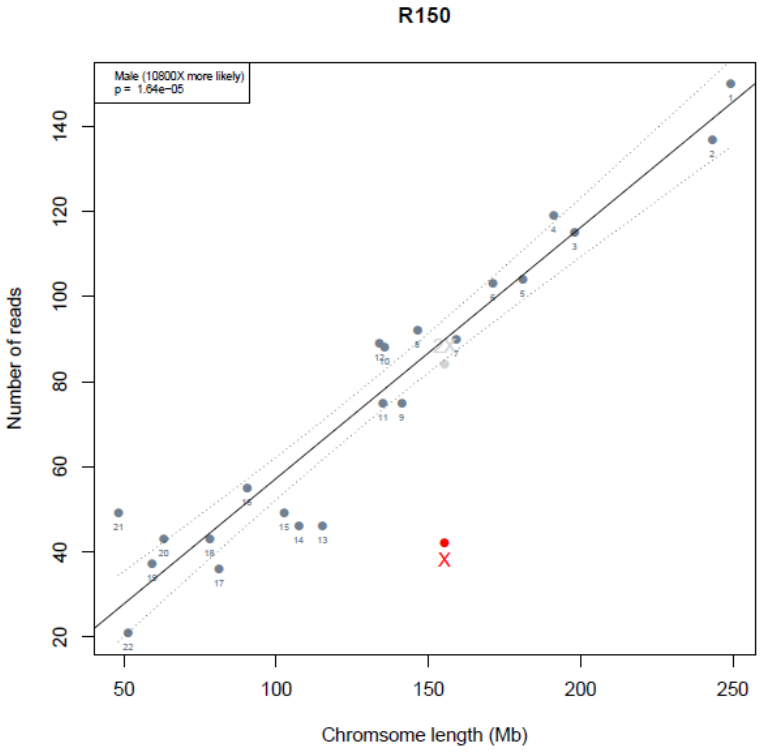
Table S4.1: Sexing plots for each of the 99 screened Kulubnarti samples.

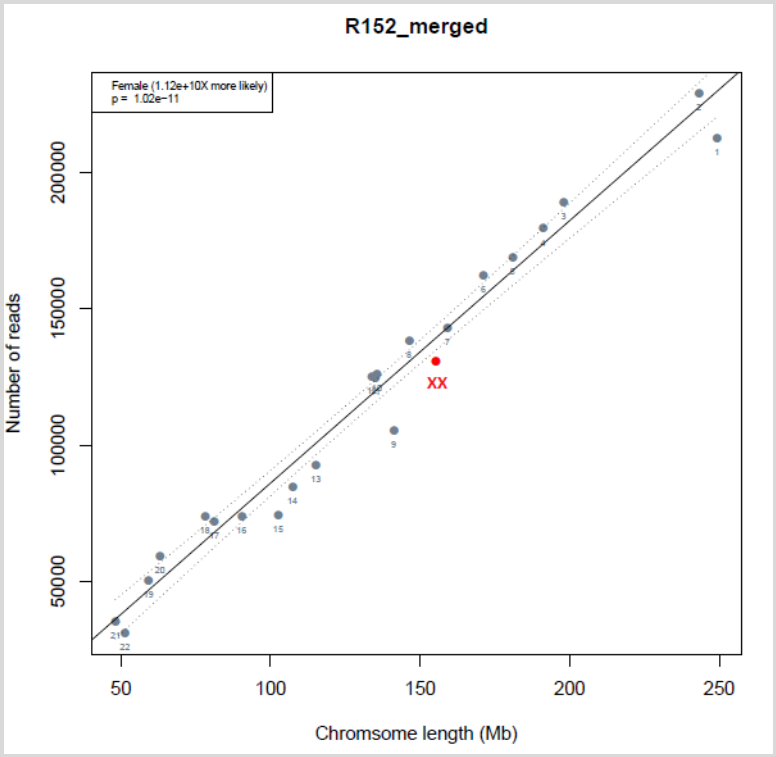
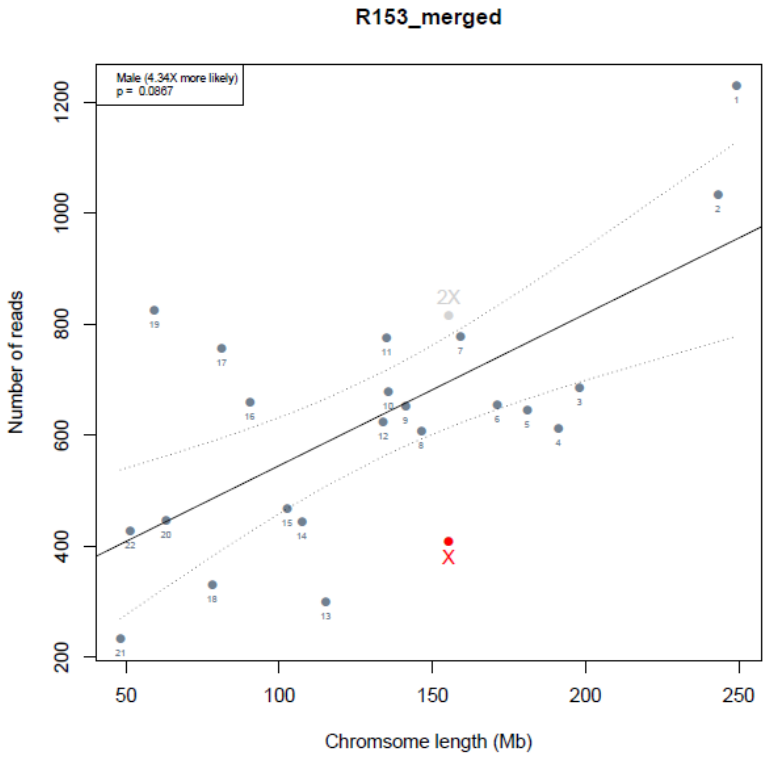
Sample ID	Sexing plot	Molecular Sex
R101*		Male (p = 3.55e−08)
R103		Female (p = 3.41e−11)

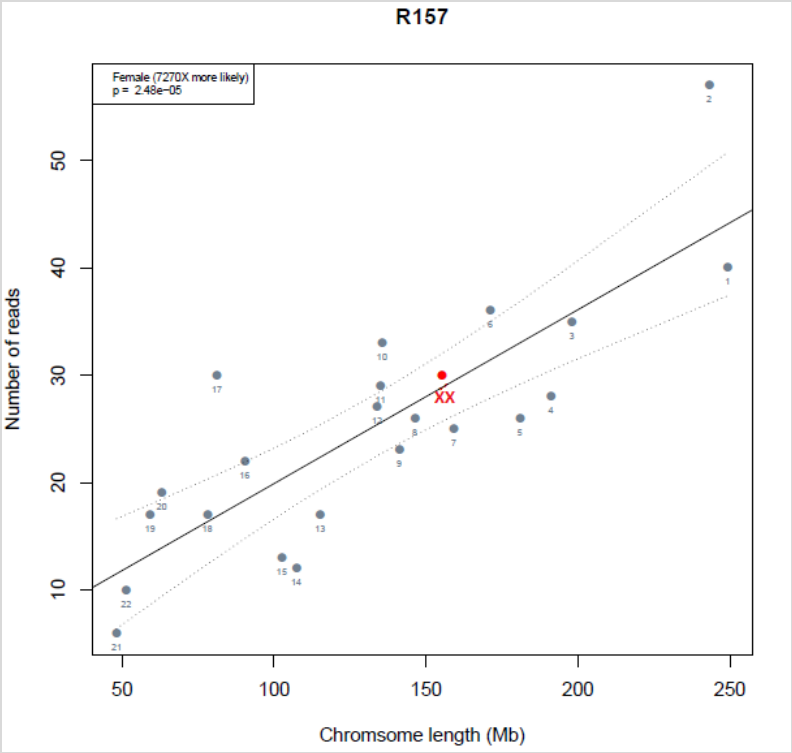
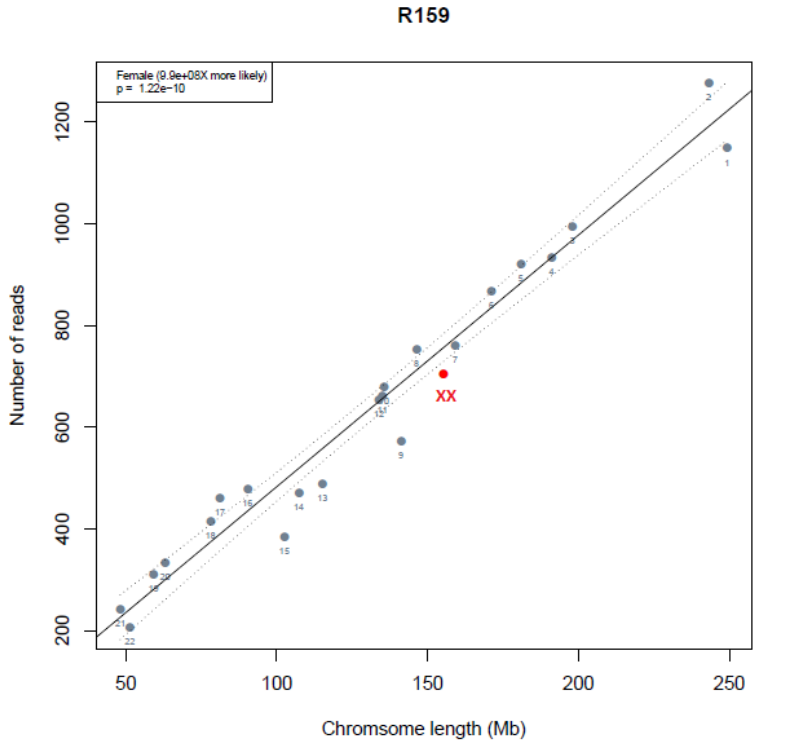
R123		Male (p = 7.39e−09)
R124*		Male (p = 2.71e−07)

R125		Male (p = 3.02e−08)
R133		Female (p = 2.16e−11)

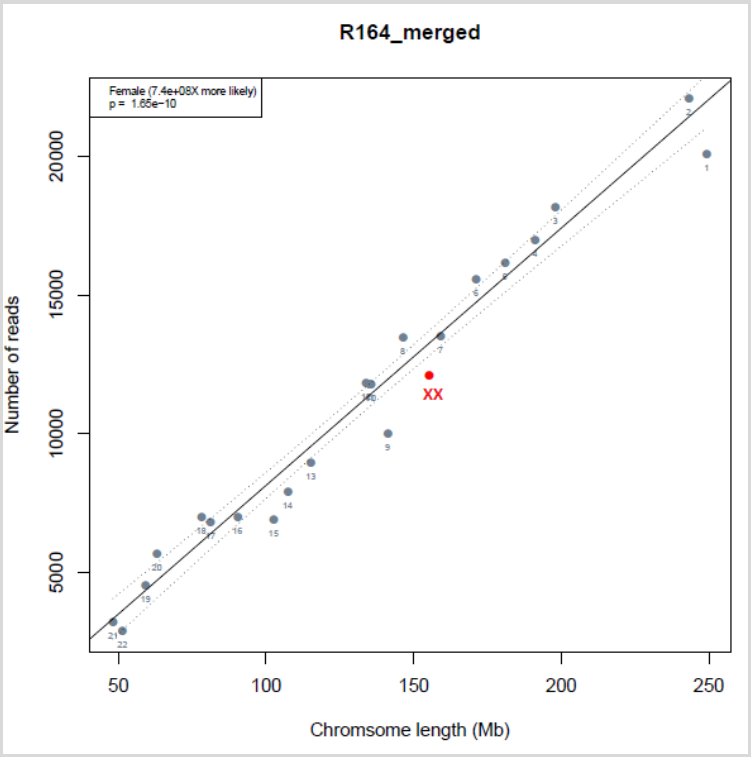
R143		unable to be determined
R144		Female (p = 1.13e−06)

R15		Male (p = 7.4e−09)
R150		Male (p = 1.64e−05)

R152*		Female (p = 1.02e−11)
R153		unable to be determined

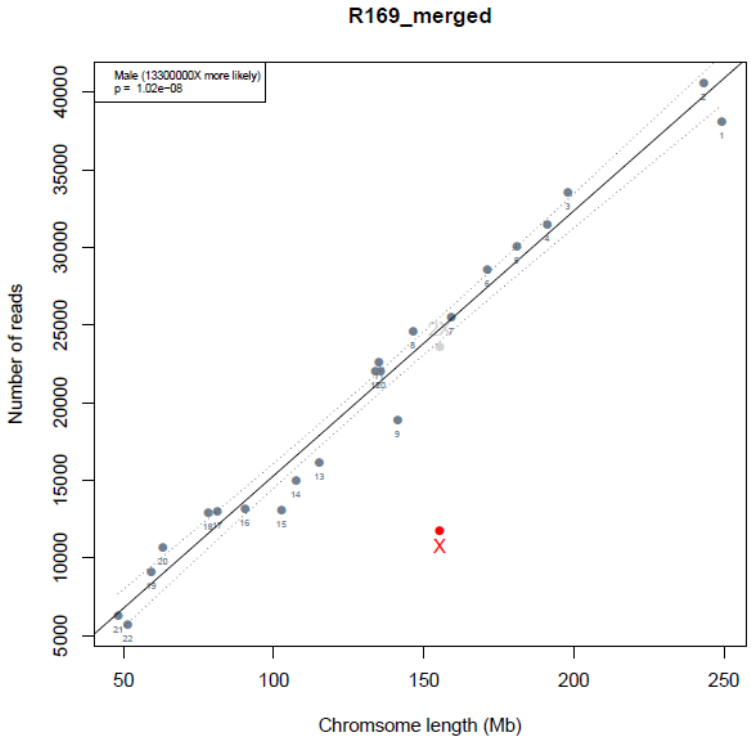
R157		Female (p = 2.48e−05)
R159		Female (p = 1.22e−10)

R164

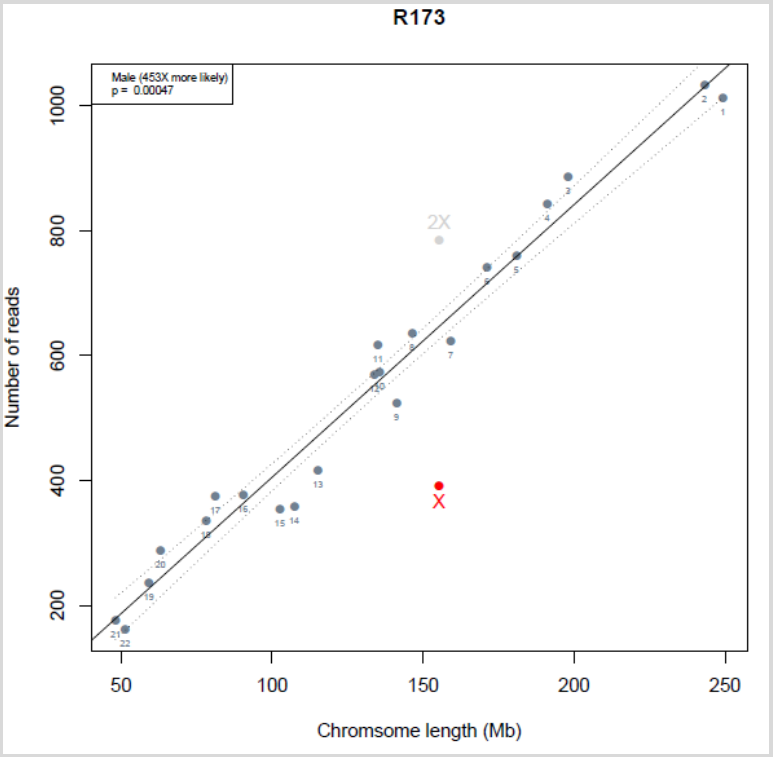
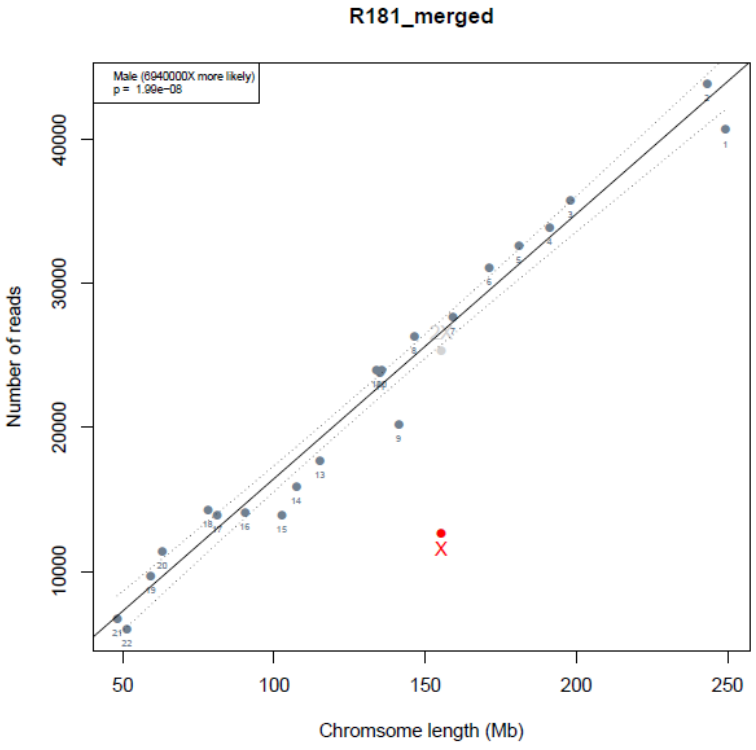


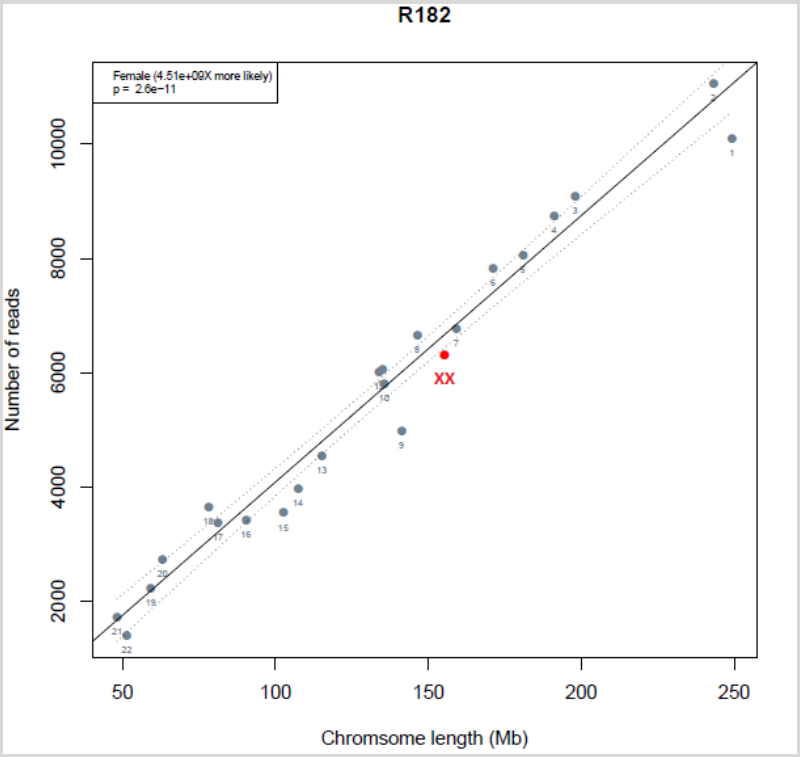
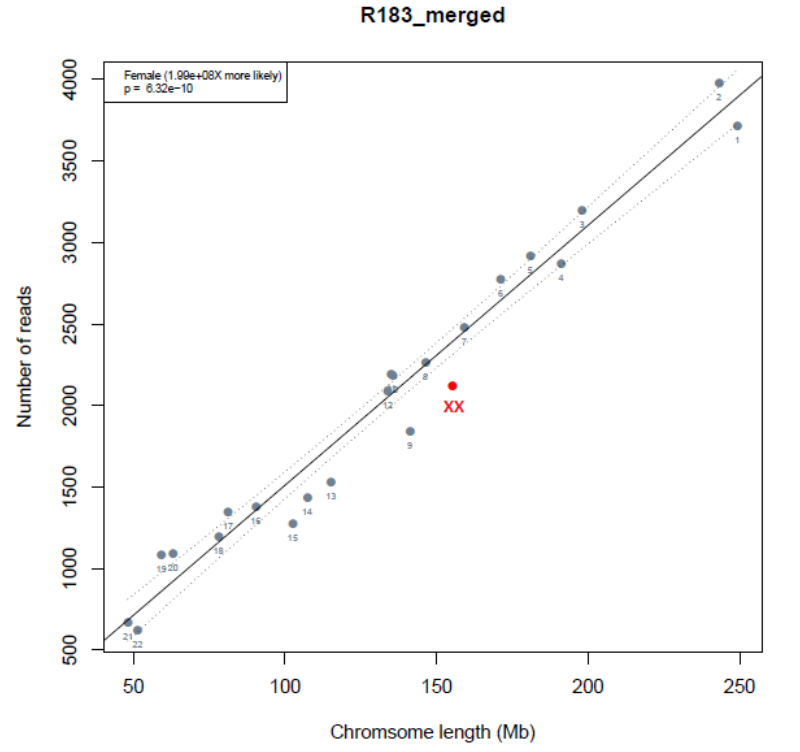
Female
(p = 1.65e−10)

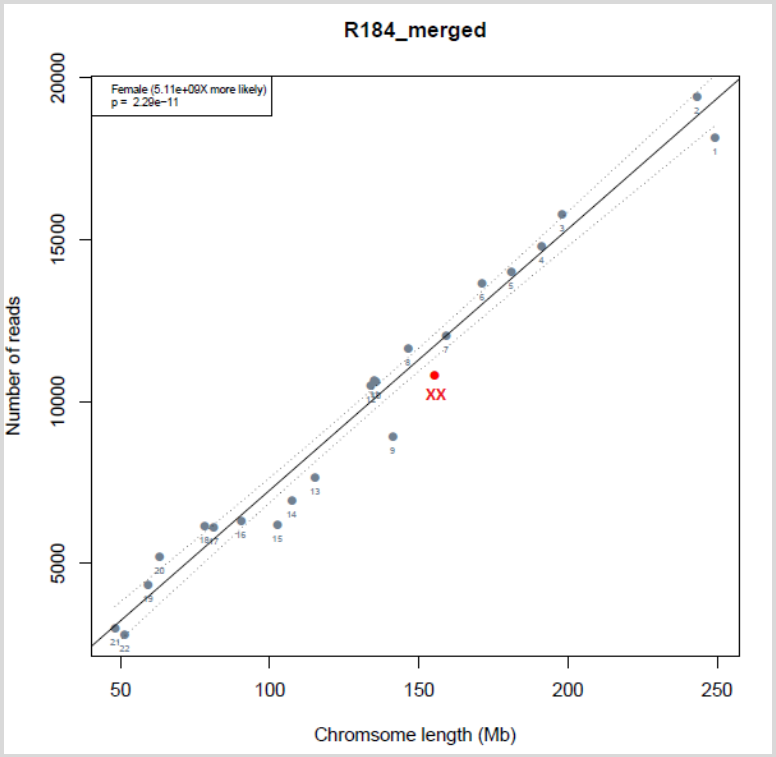
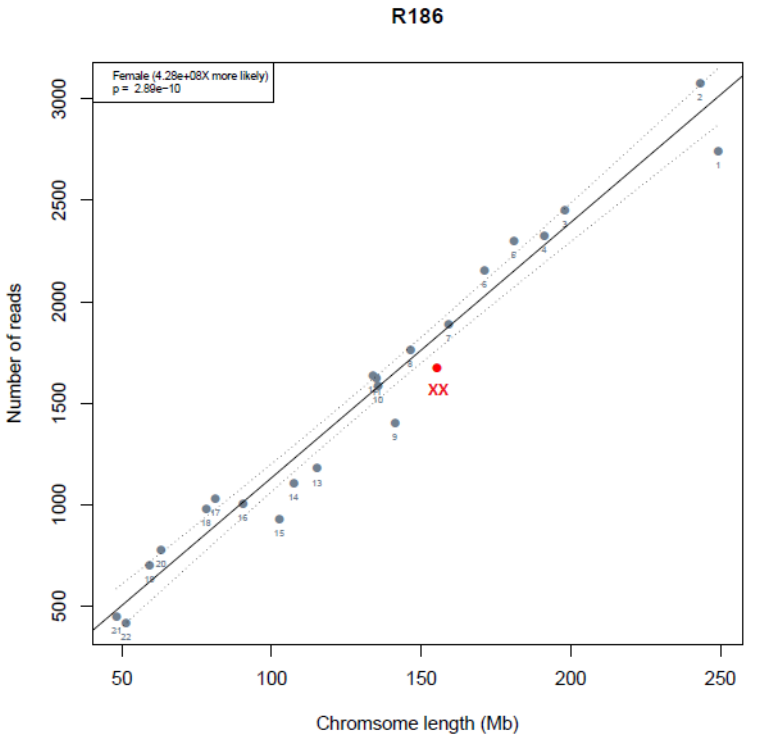
R169*



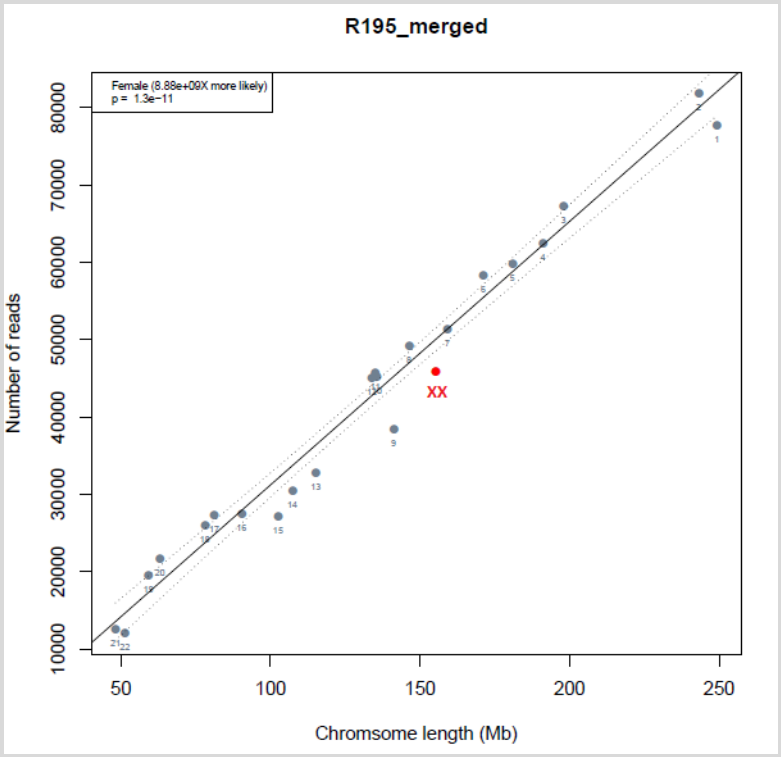
Male
(p = 1.02e−08)

R173		Male (p = 0.00047)
R181*		Male (p = 1.99e−08)

R182*		Female (p = 2.6e−11)
R183		Female (p = 6.32e−10)

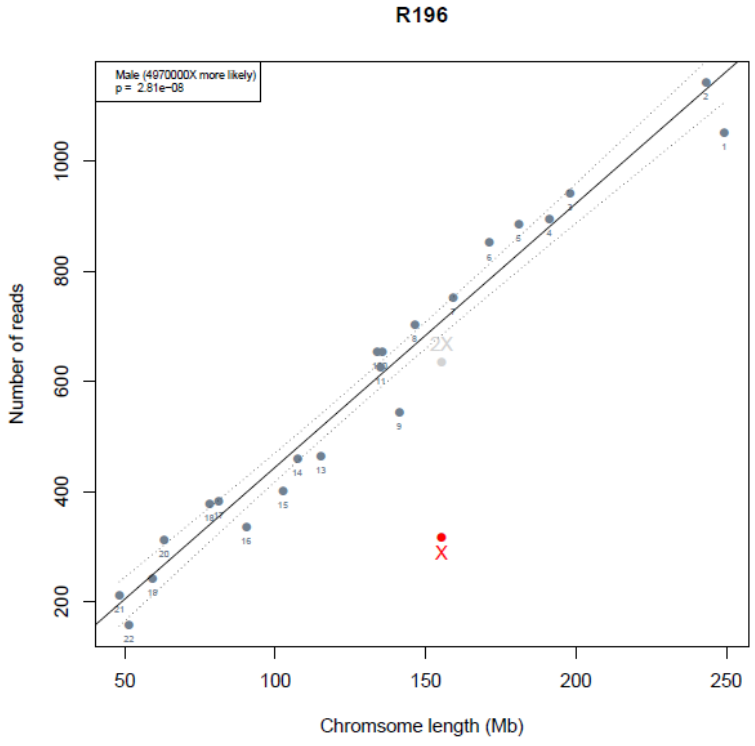
R184		Female (p = 2.29e−11)
R186*		Female (p = 2.89e−10)

R195*

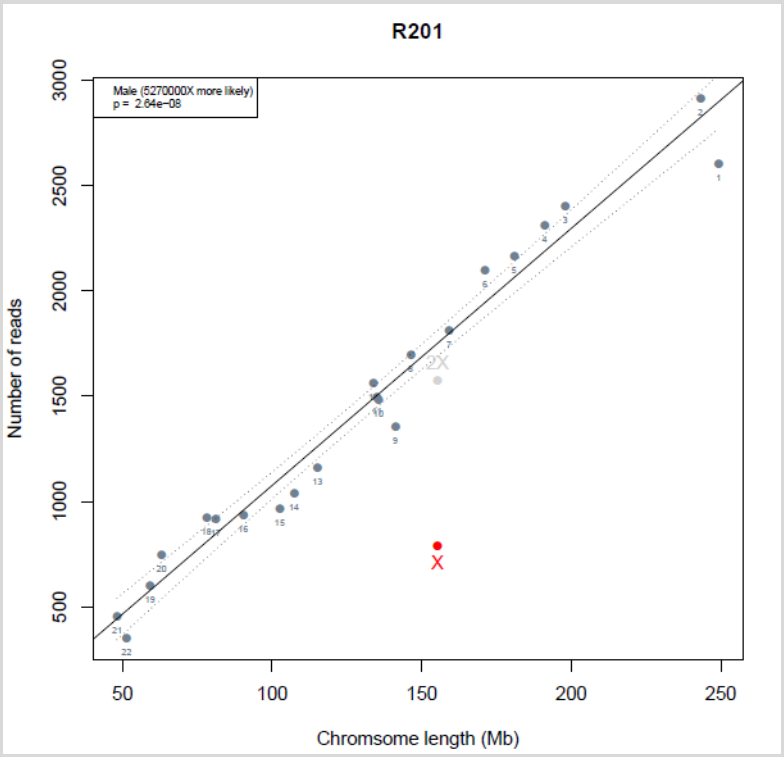
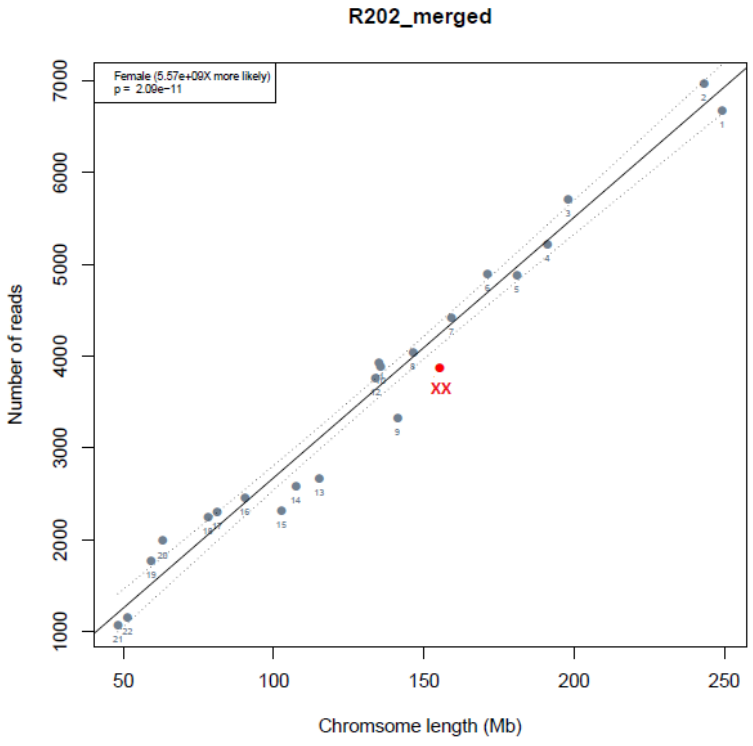


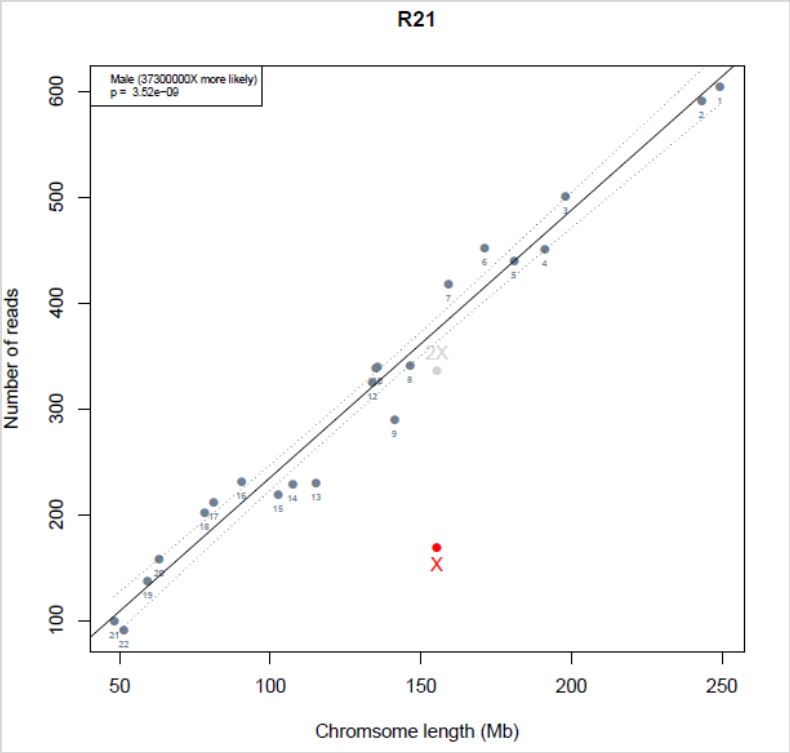
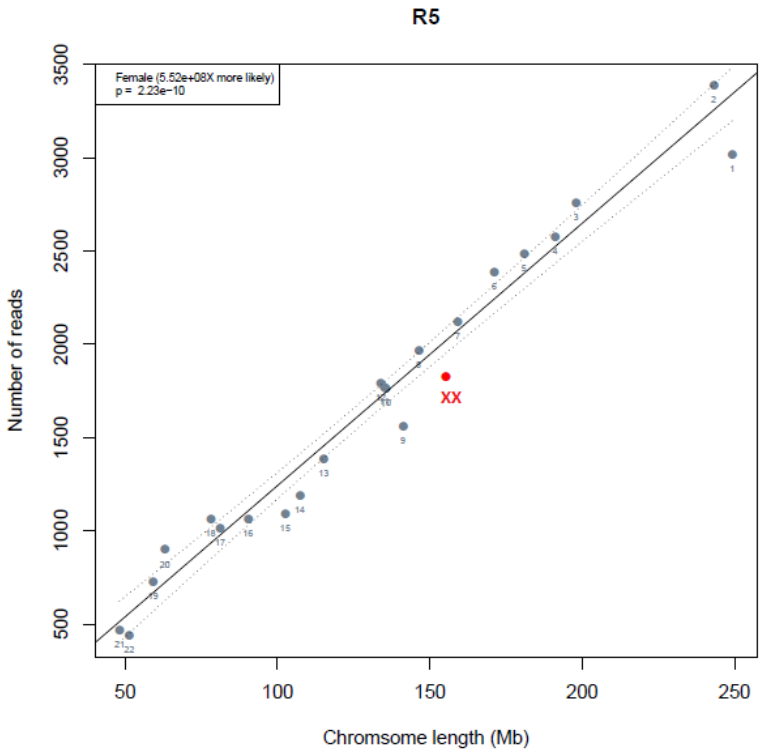
Female
(p = 1.3e−11)

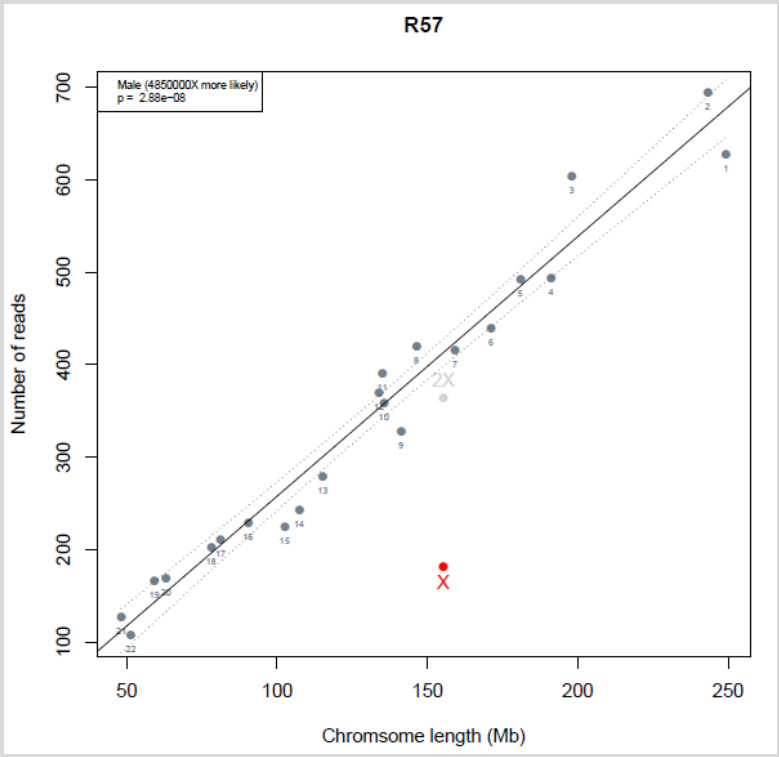
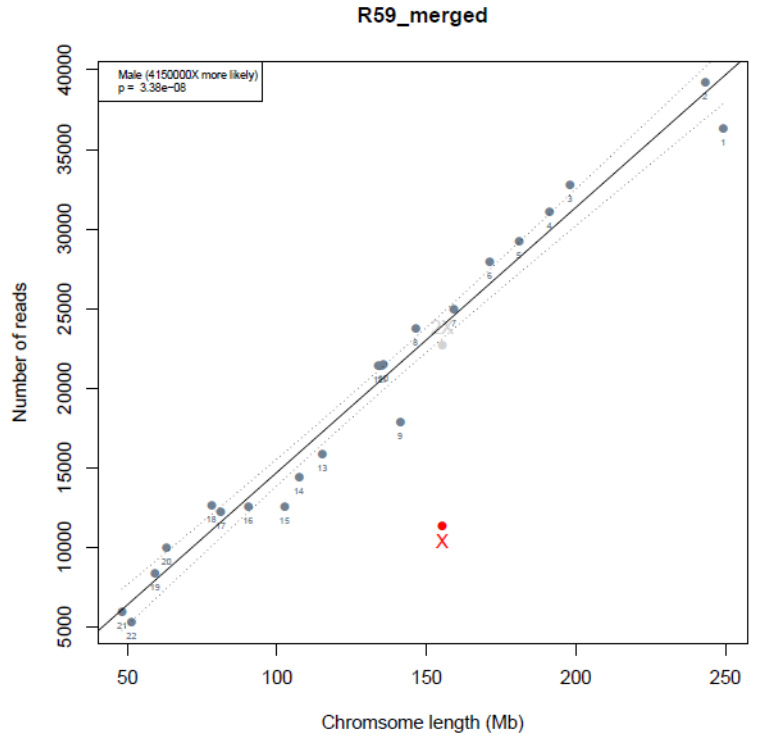
R196*

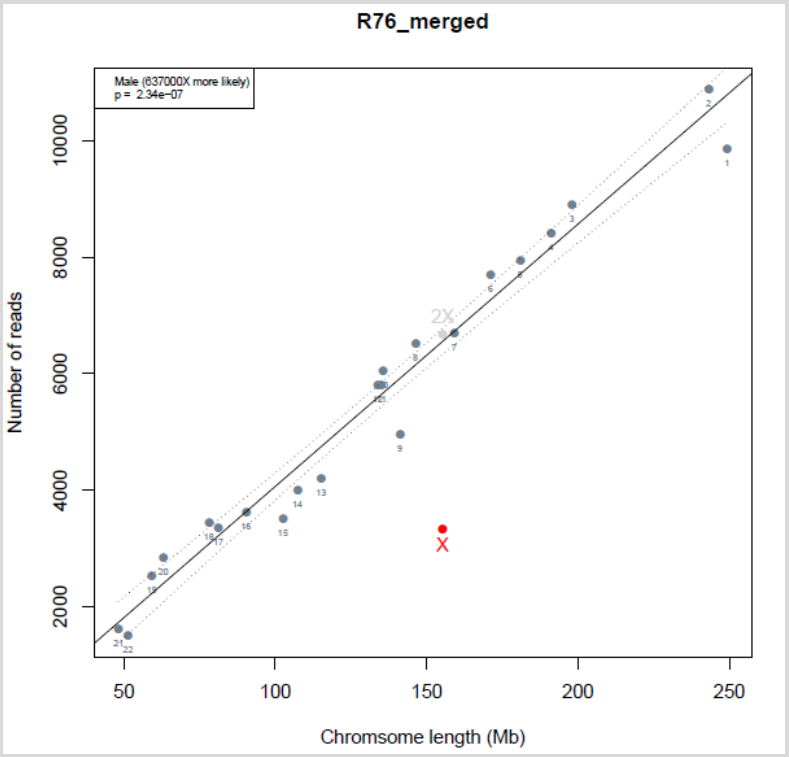
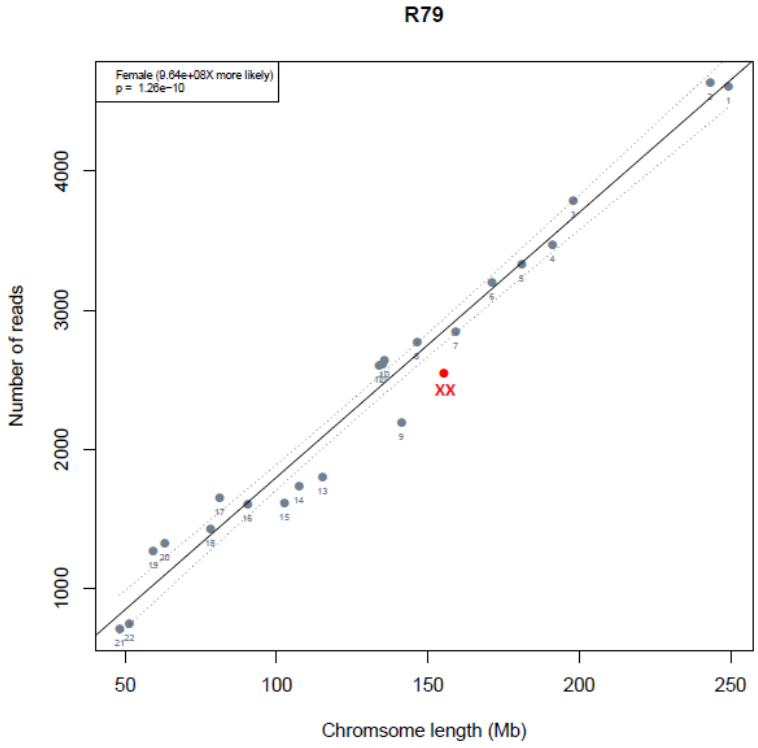


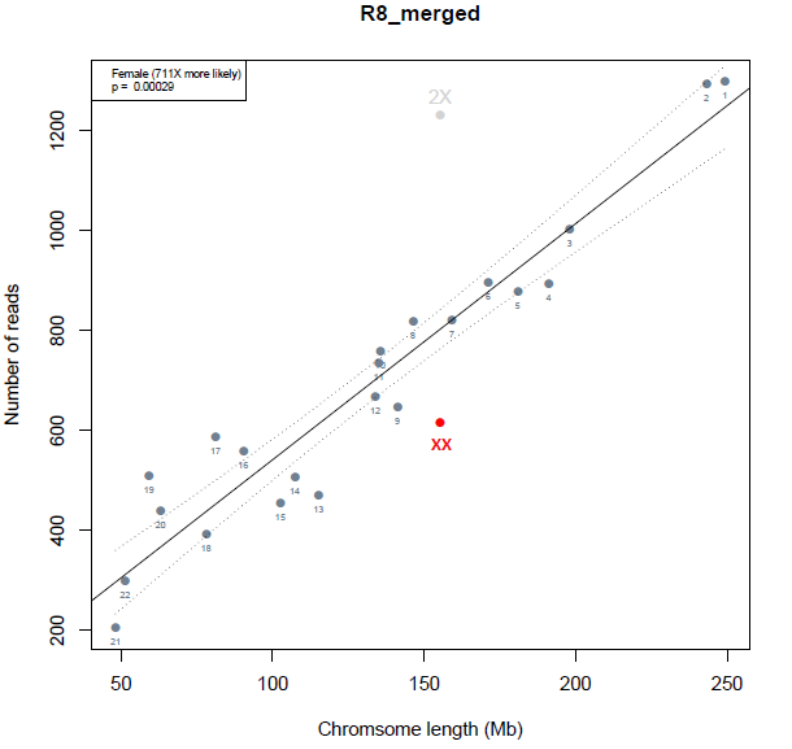
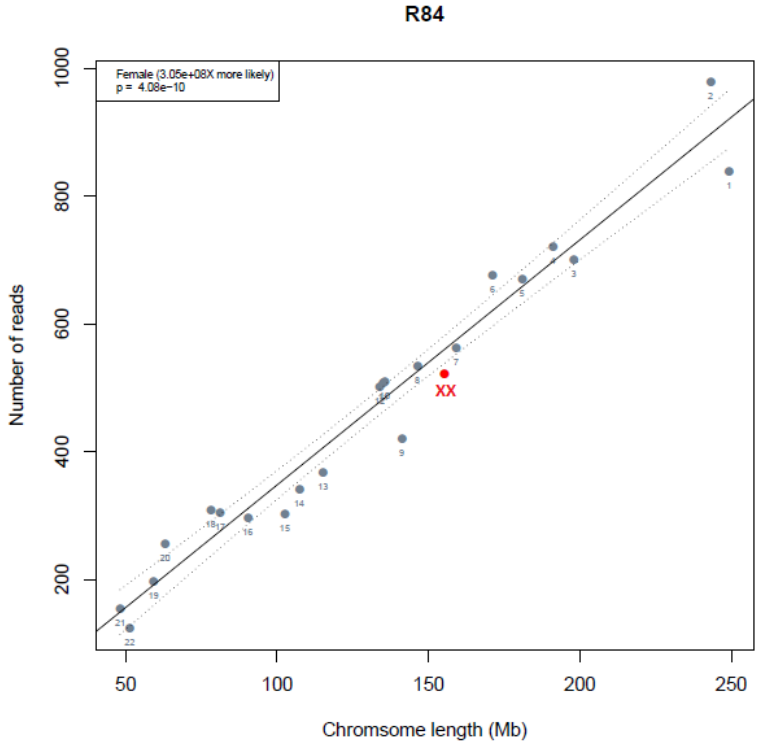
Male
(p = 2.81e−08)

R201*		Male (p = 2.64e−08)
R202*		Female (p = 2.09e−11)

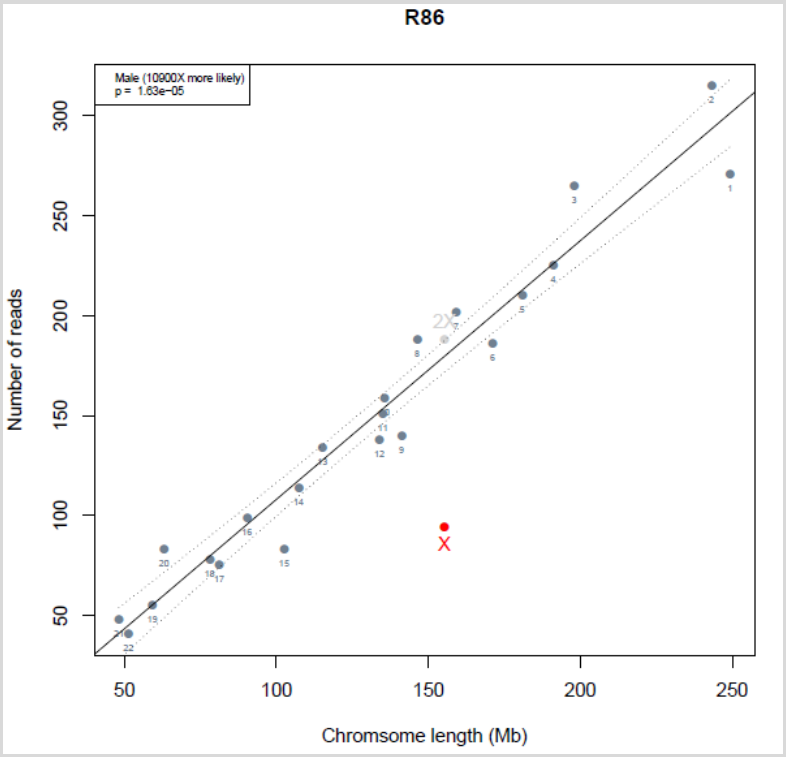
R21		Male (p = 3.52e−09)
R5*		Female (p = 2.23e−10)

R57		Male (p = 2.88e−08)
R59*		Male (p = 3.38e−08)

R76		Male (p = 2.34e−07)
R79*		Female (p = 1.26e−10)

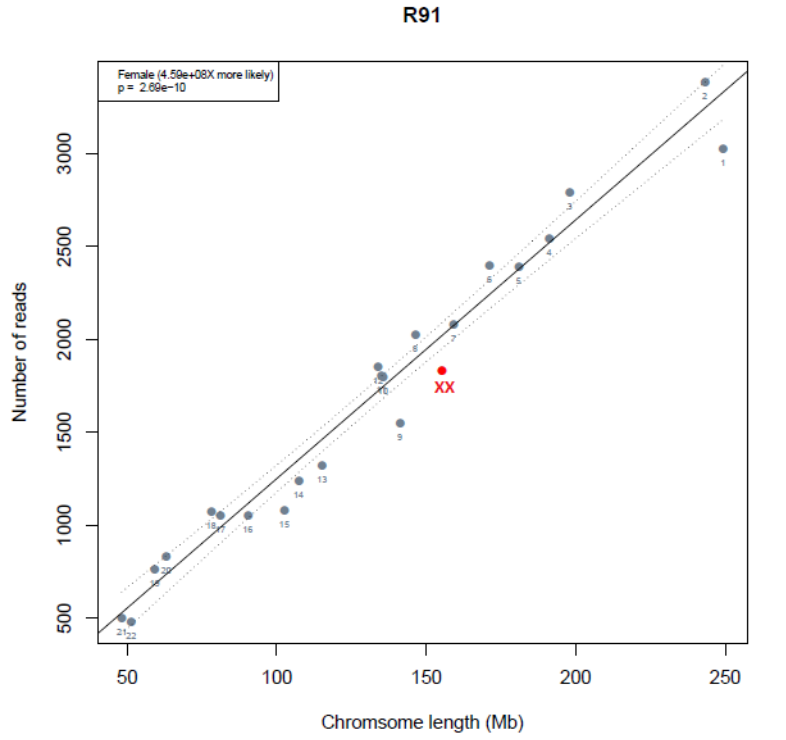
R8		Female (p = 0.00029)
R84		Female (p = 4.08e−10)

R86

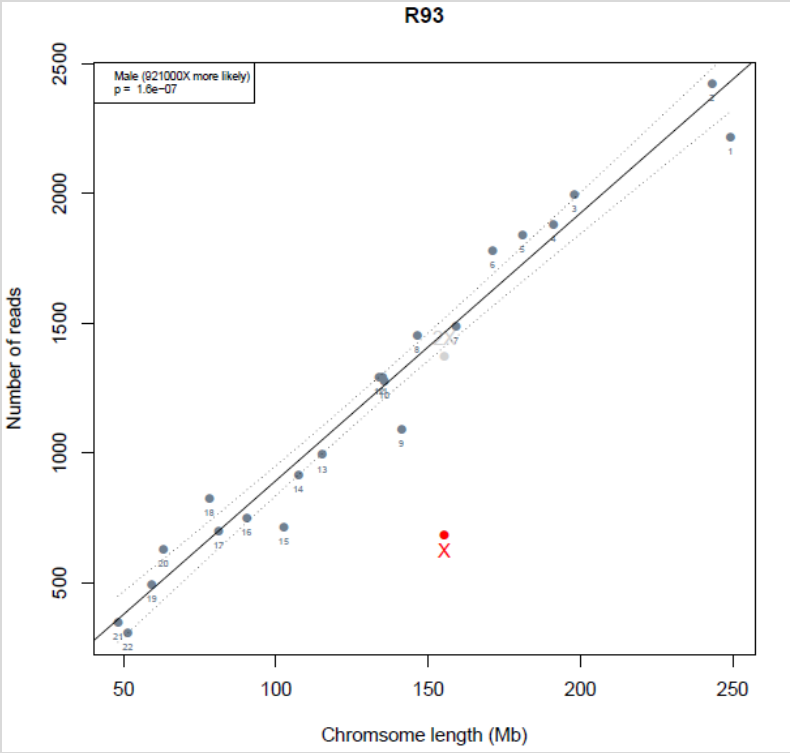
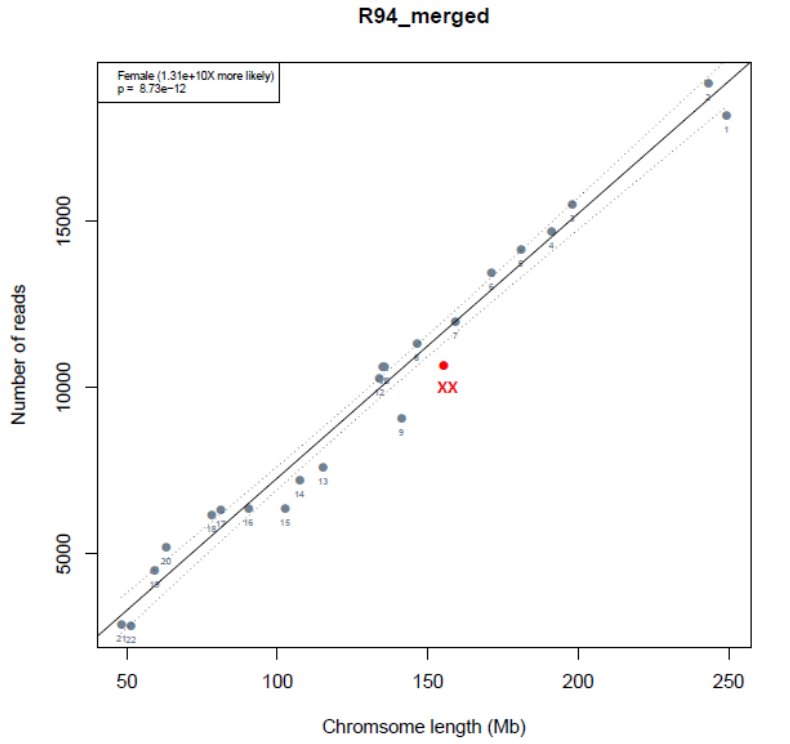


Male
(p = 1.63e−05)

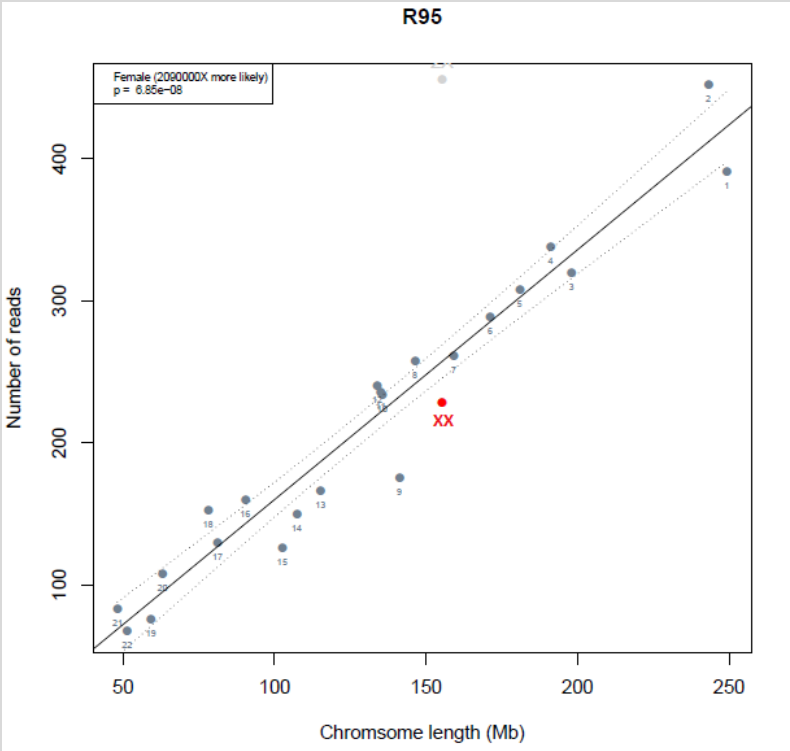
R91



Female
(p = 2.69e−10)

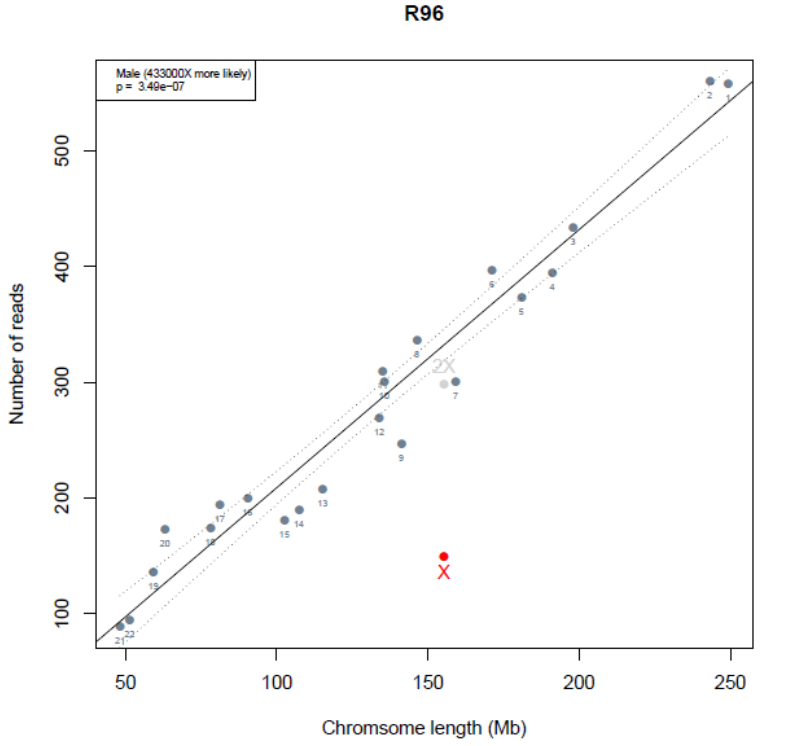
R93*		Male (p = 1.6e−07)
R94		Female (p = 8.73e−12)

R95

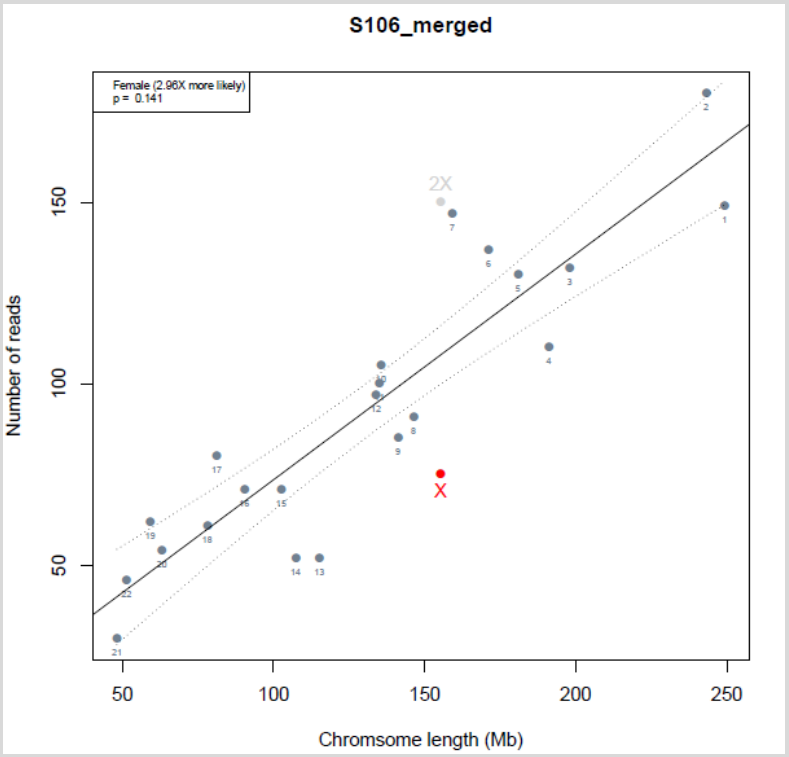
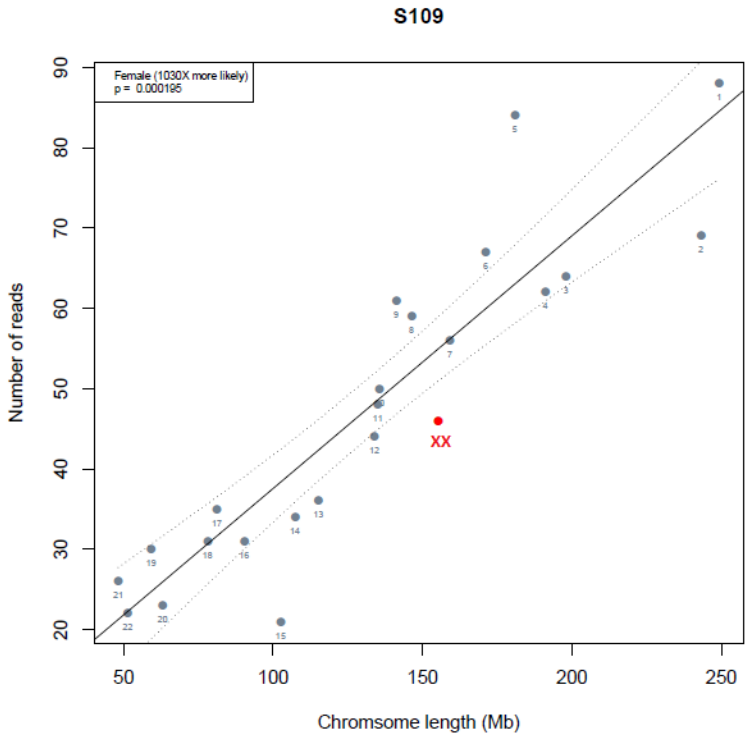


Female
(p = 6.85e−08)

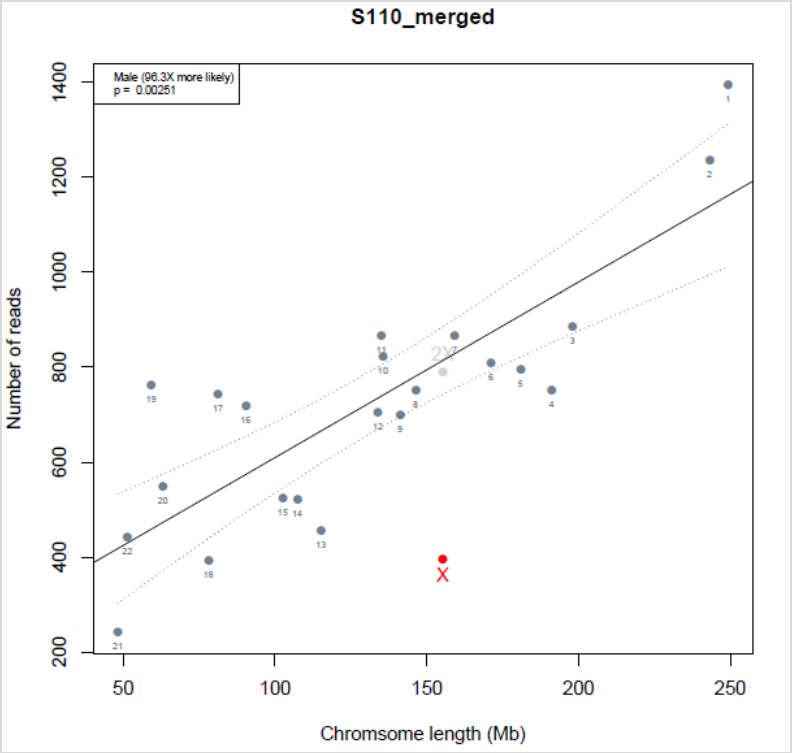
R96



Male
(p = 3.49e−07)

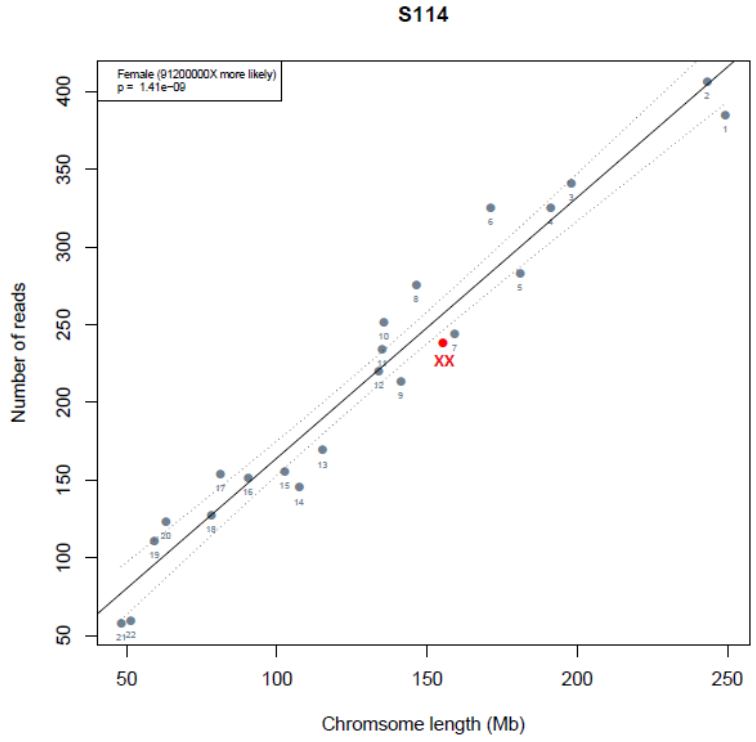
S106		unable to be determined
S109		Female (p = 0.000195)

S110

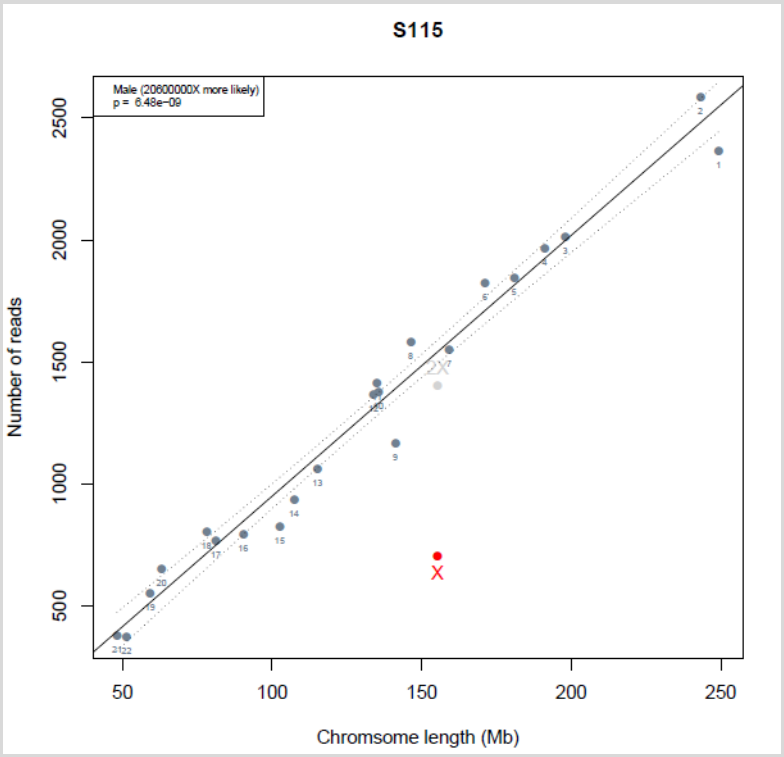
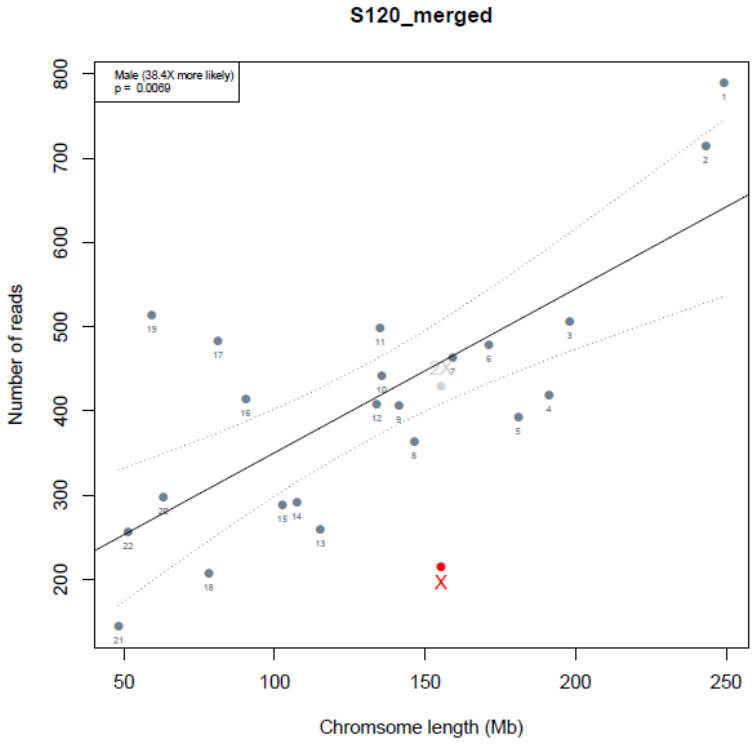


Male
(p = 0.00251)

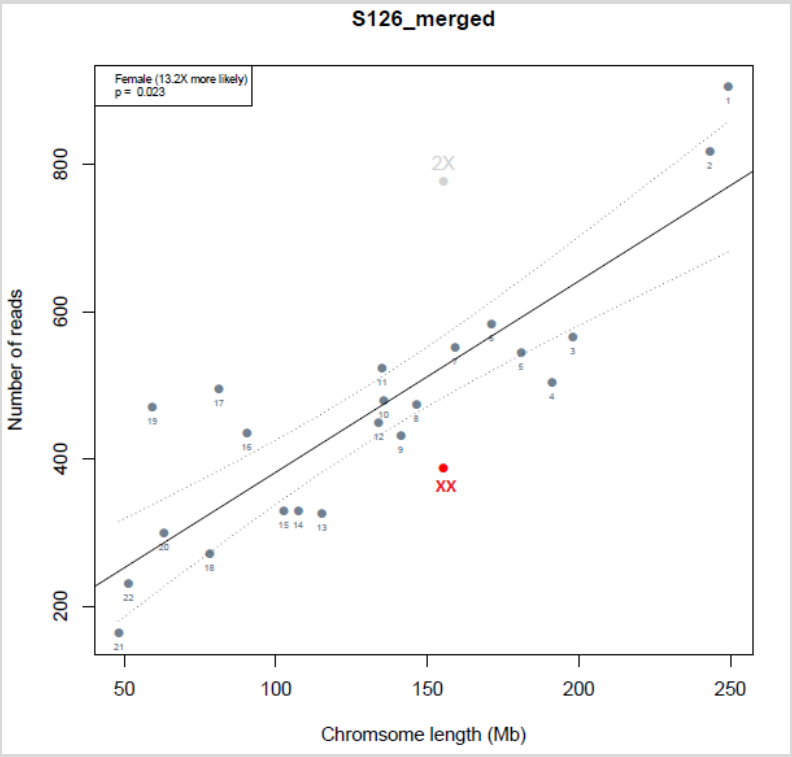
S114



Female
(p = 1.41e−09)

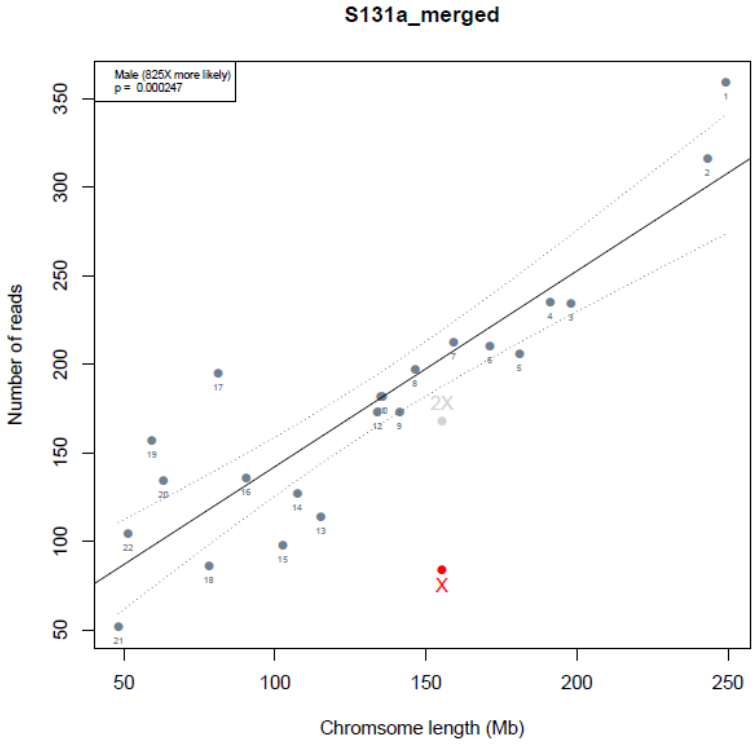
S115		Male (p = 6.48e−09)
S120		Male (p = 0.0069)

S126

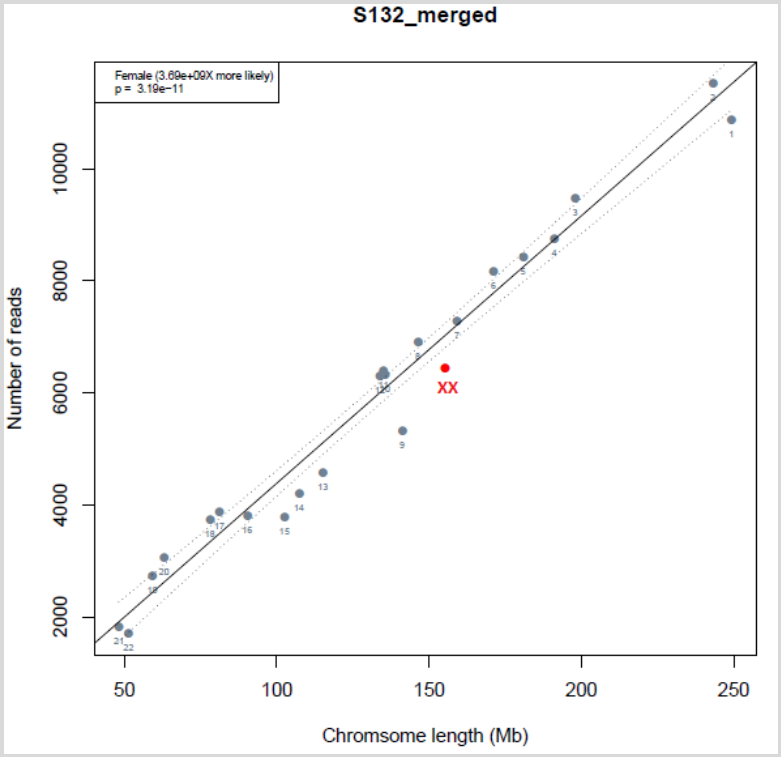
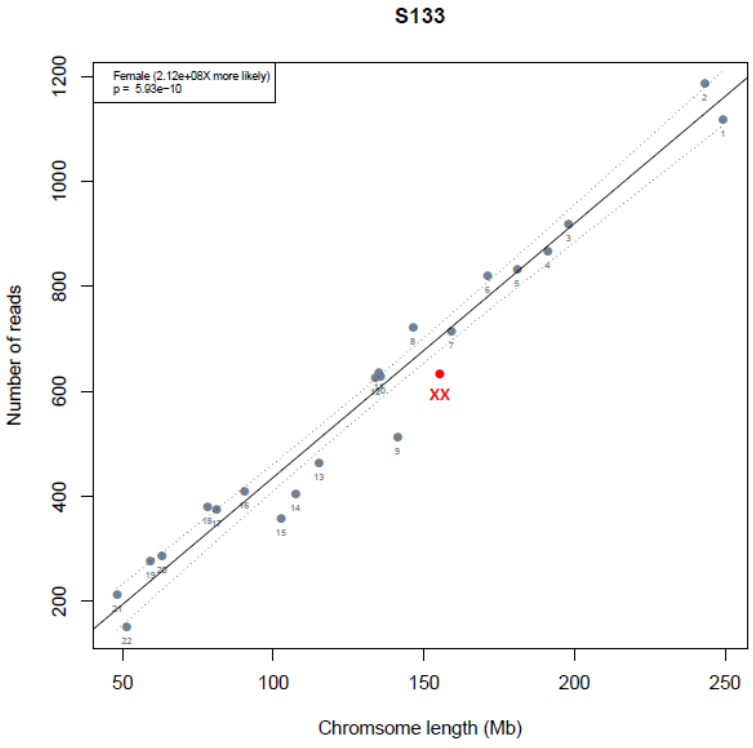


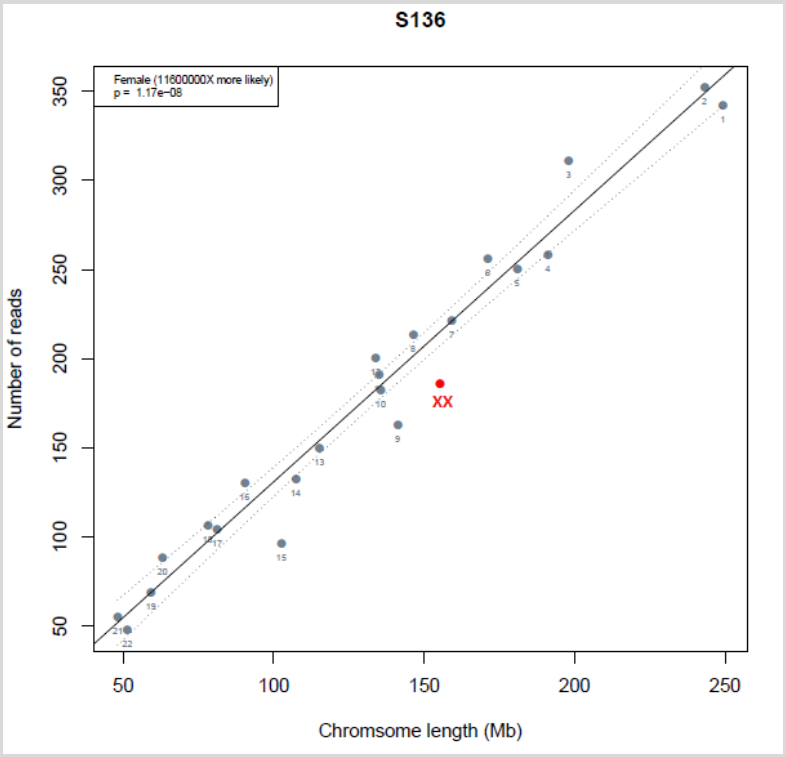
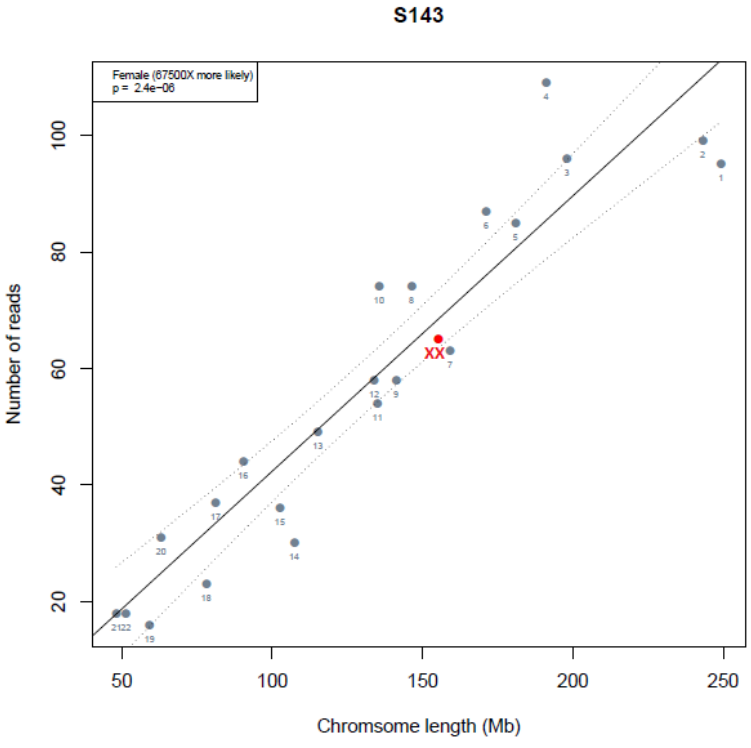
Female
(p = 0.023)

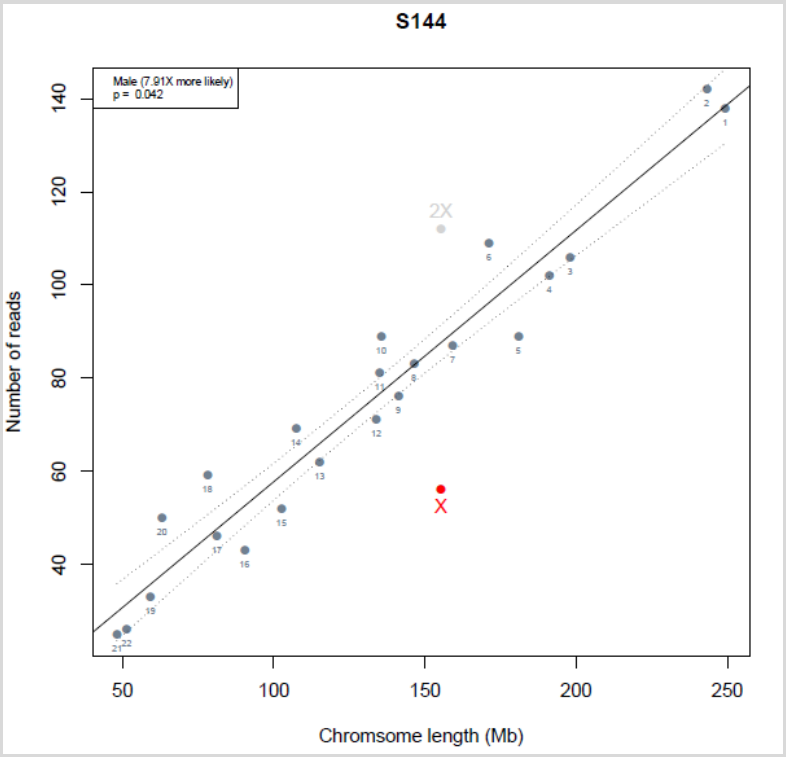
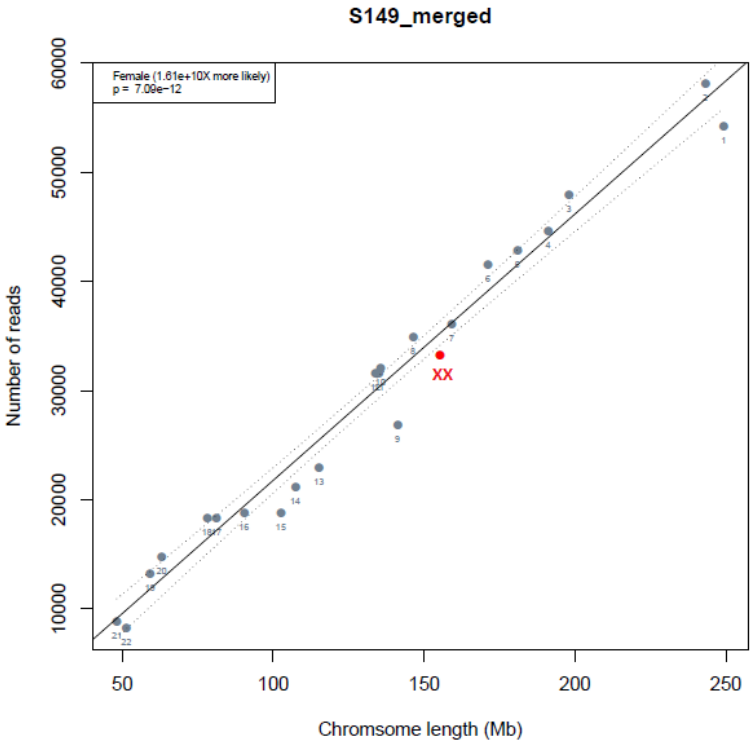
S131a



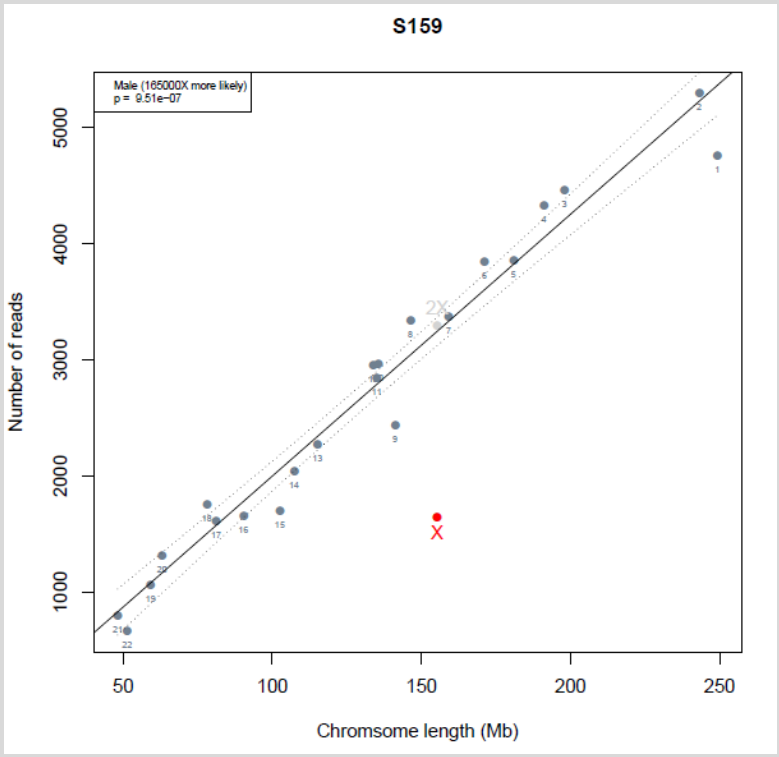
Male
(p = 0.000247)

S132		Female (p = 3.19e−11)
S133*		Female (p = 5.93e−10)

S136		Female (p = 1.17e−08)
S143		Female (p = 2.4e−06)

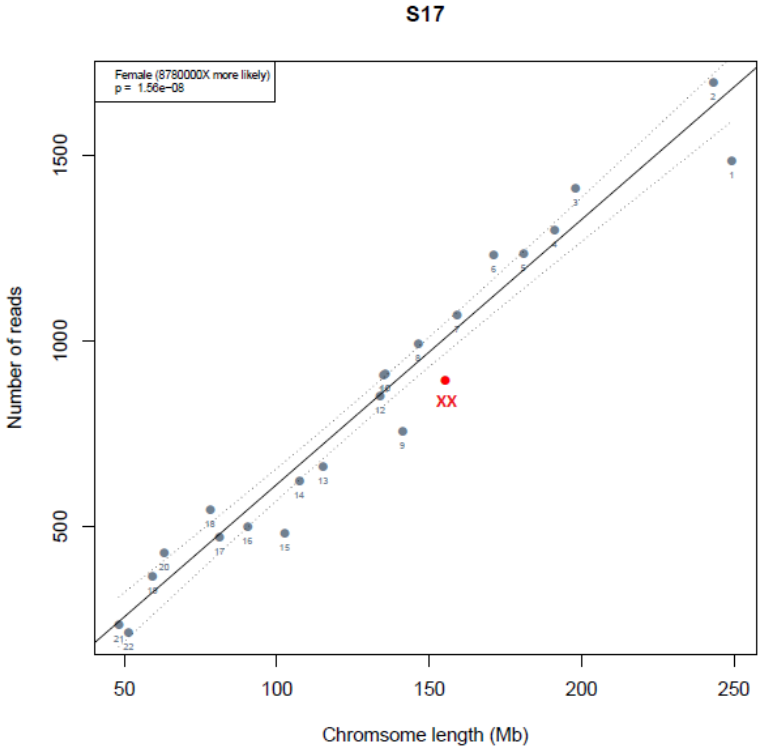
S144		Male? (p = 0.042)
S149*		Female (p = 7.09e−12)

S159*

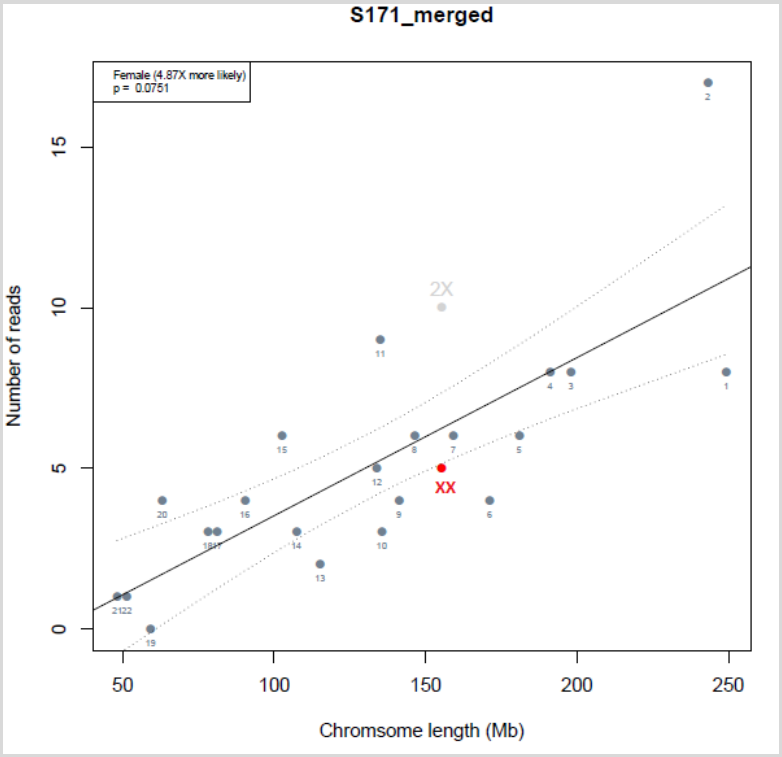
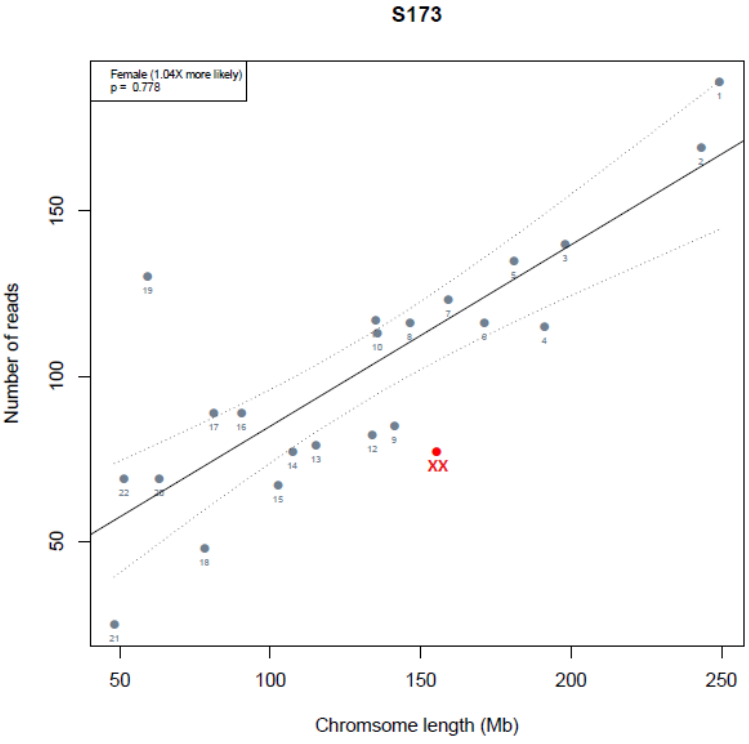


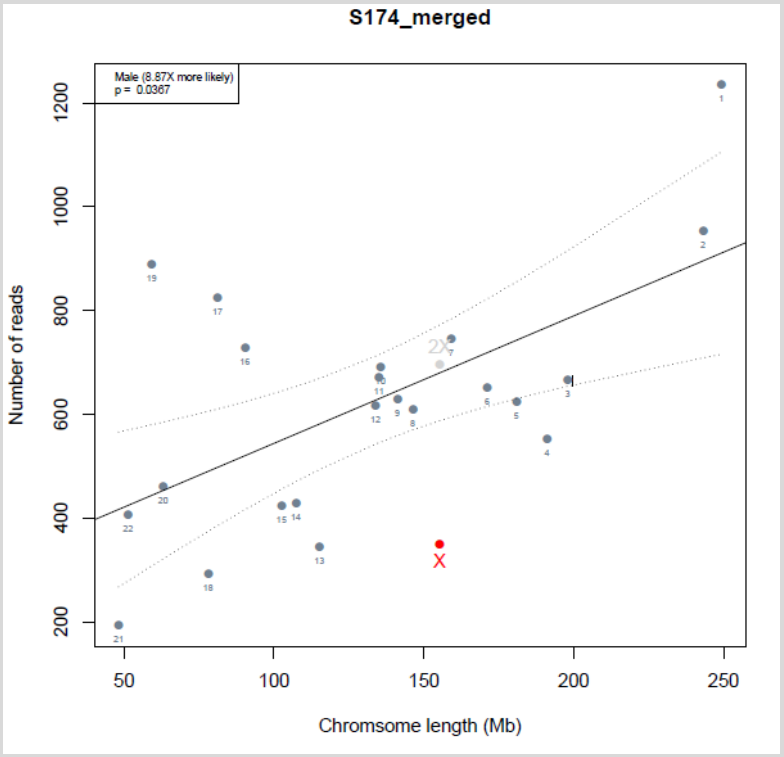
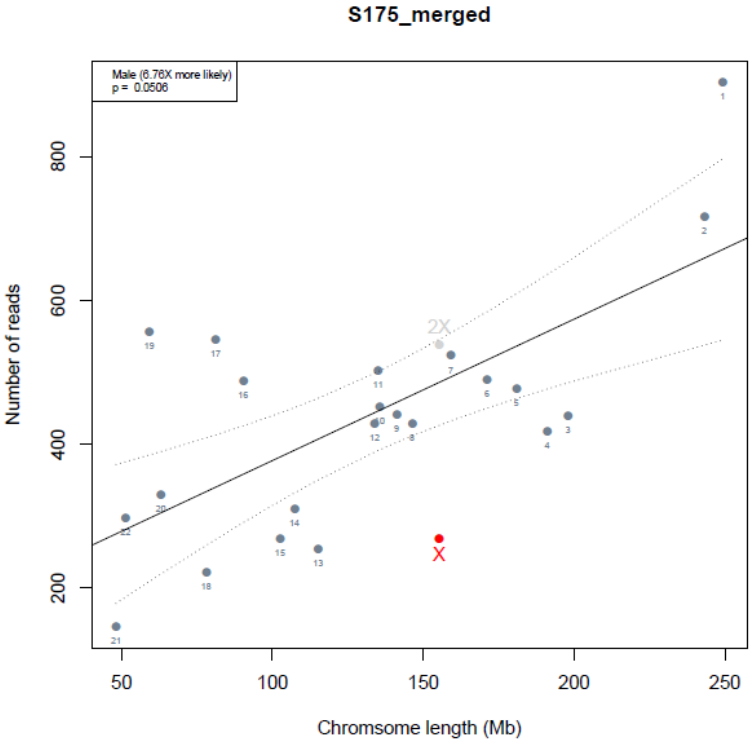
Male
(p = 9.51e−07)

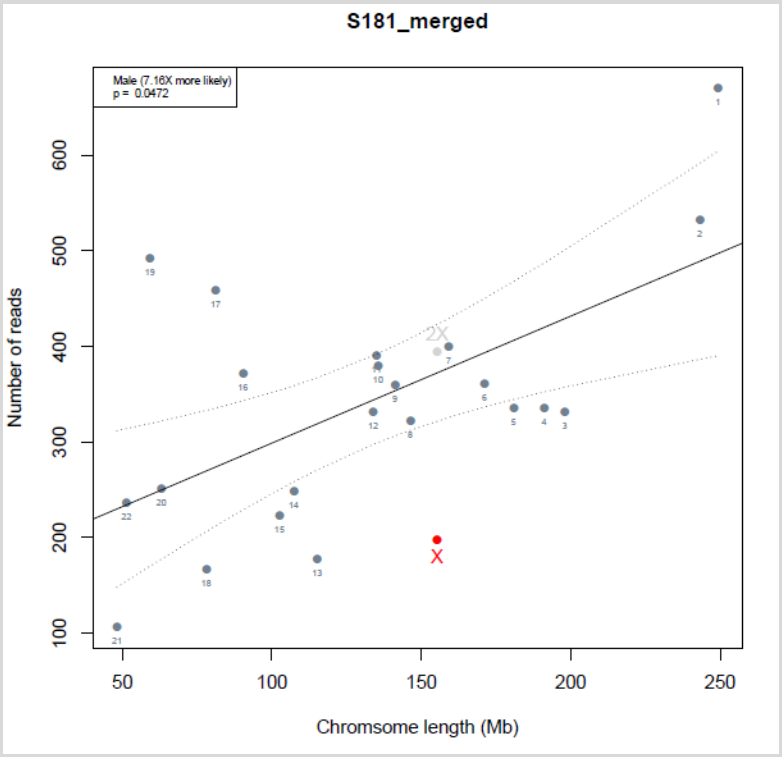
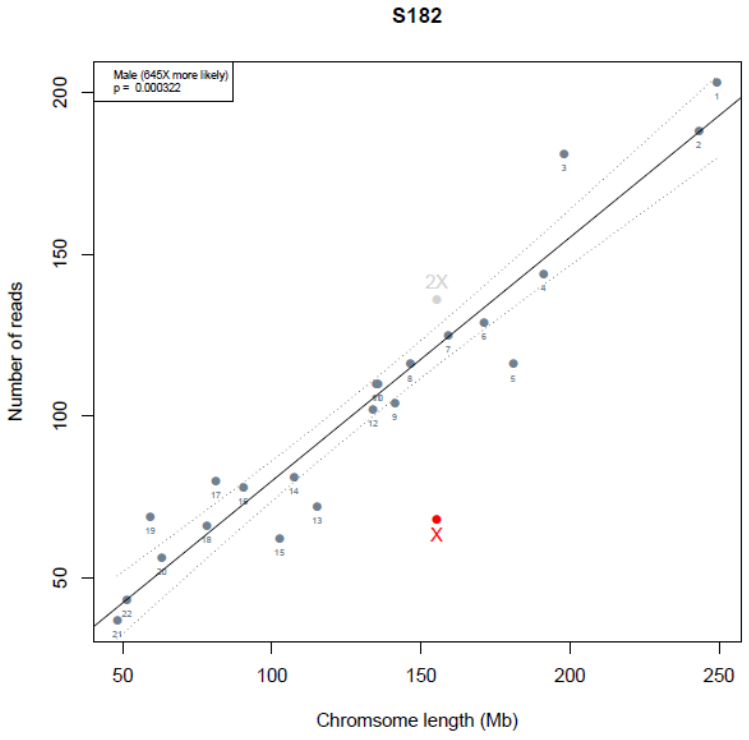
S17*

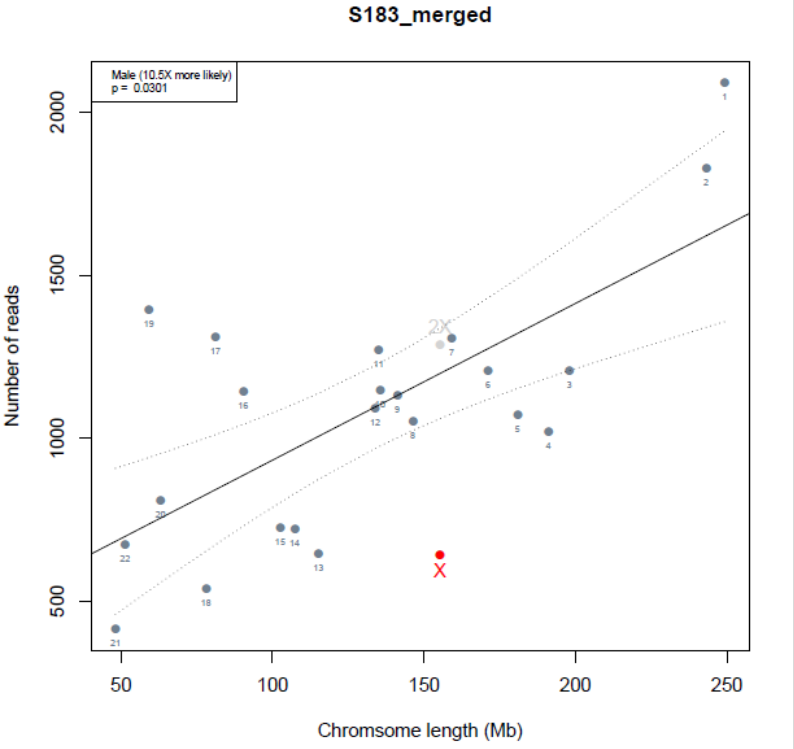
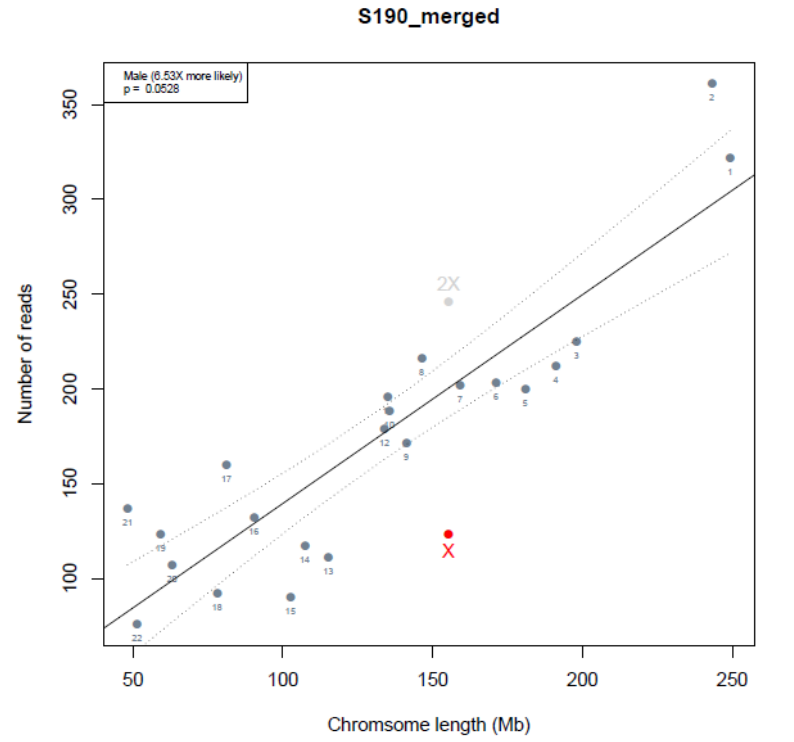


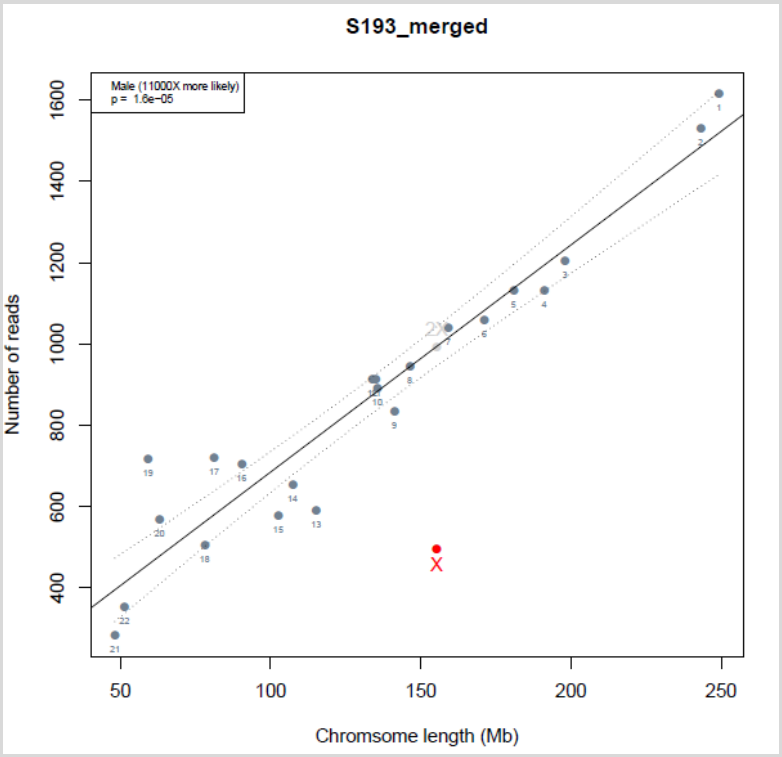
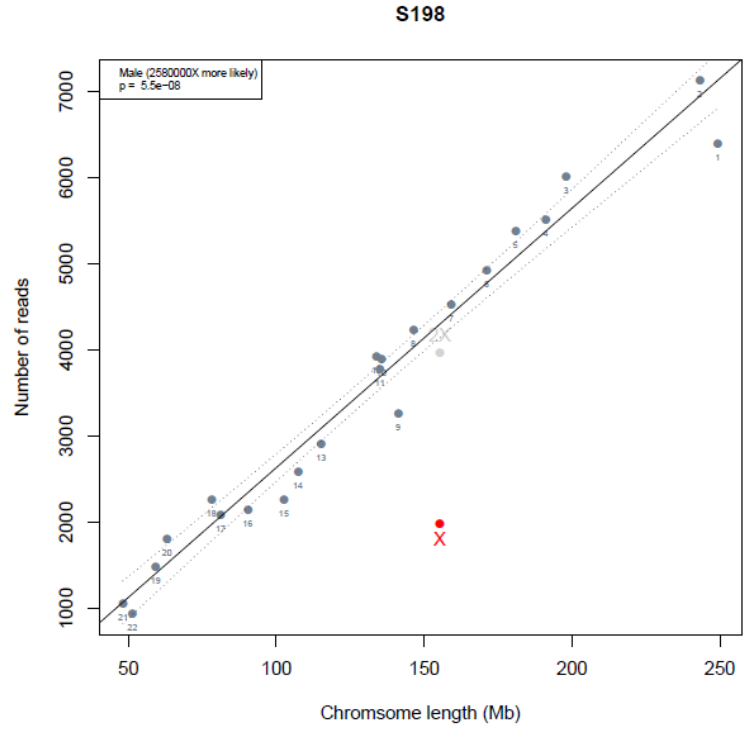
Female
(p = 1.56e−08)

S171		unable to be determined
S173		unable to be determined

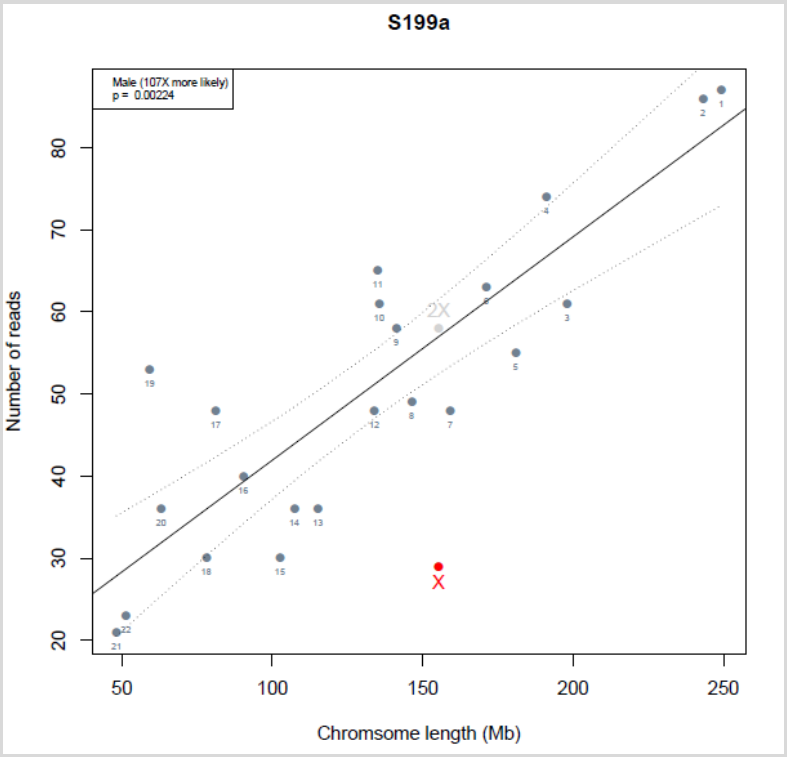
S174		Male (p = 0.0367)
S175		Male? (p = 0.0506)

S181		Male (p = 0.0472)
S182		Male (p = 0.000322)

S183		Male (p = 0.0301)
S190		Male? (p = 0.0528)

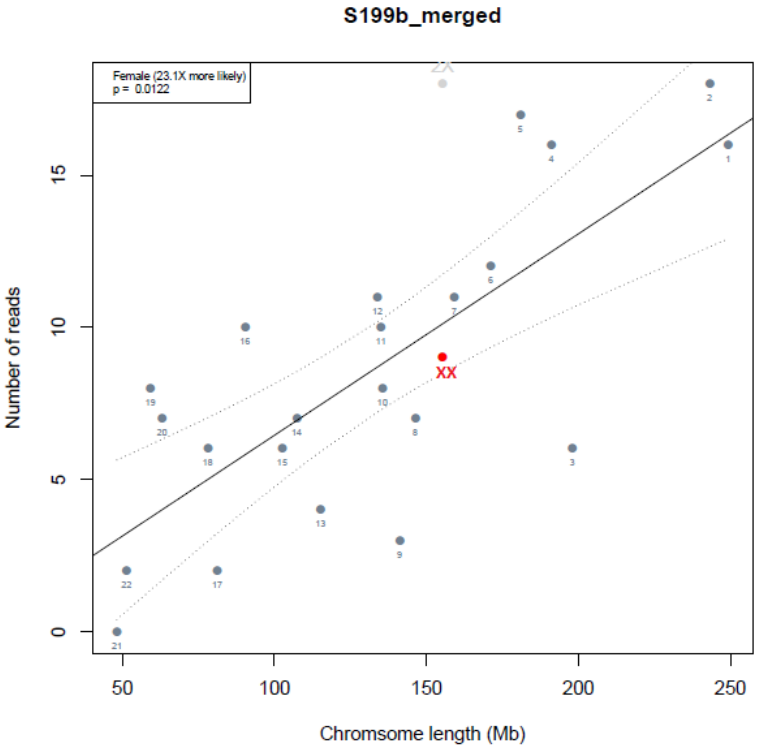
S193		Male (p = 1.6e−05)
S198*		Male (p = 5.5e−08)

S199a

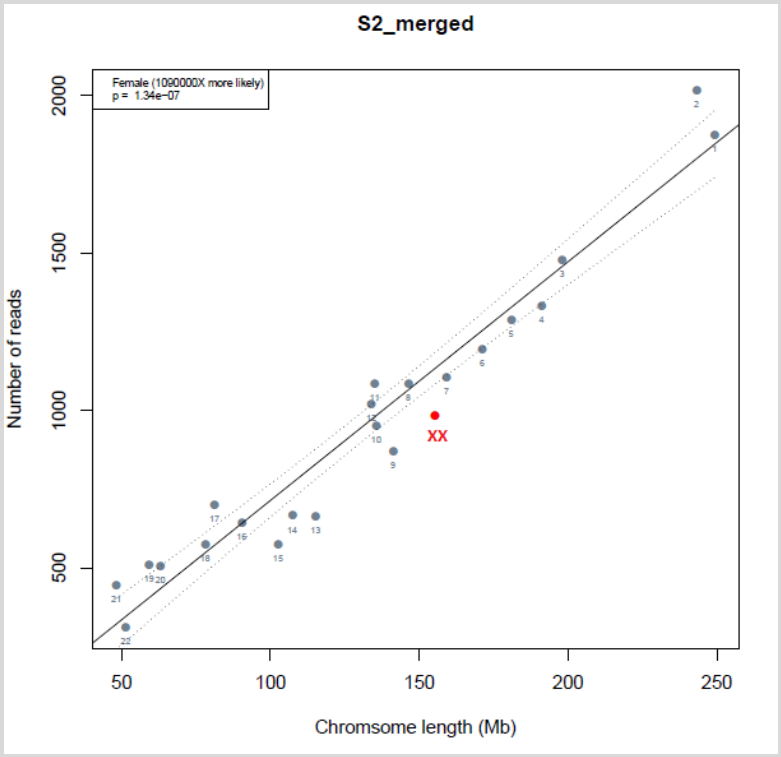
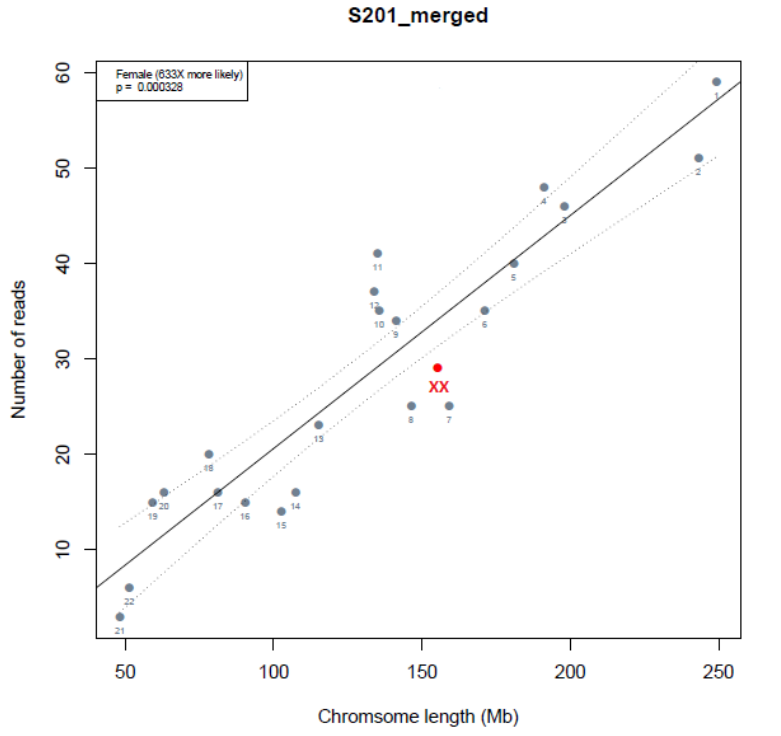


Male
(p = 0.00224)

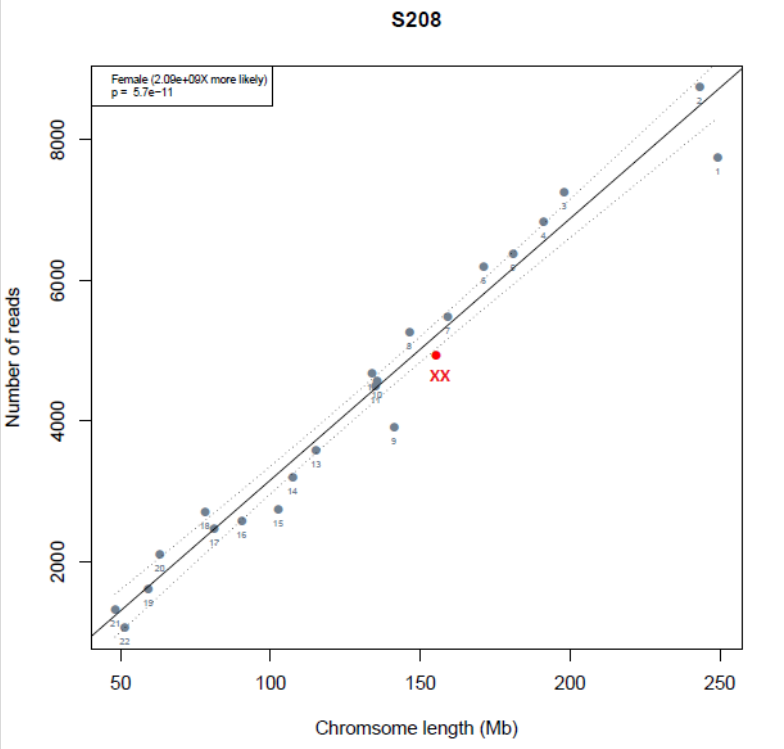
S199b



Female
(p = 0.0122)

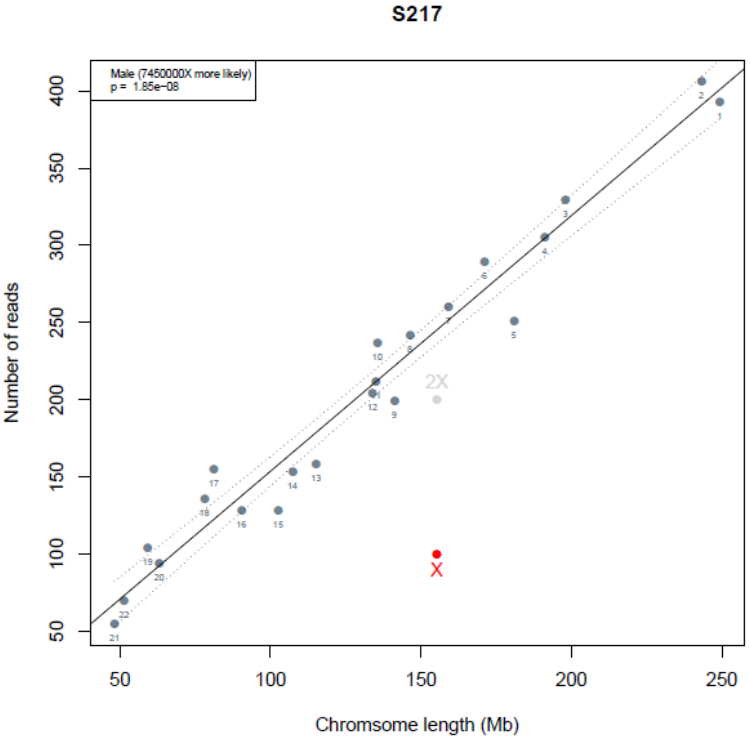
S2		Female (p = 1.34e−07)
S201		Female (p = 0.000328)

S208*

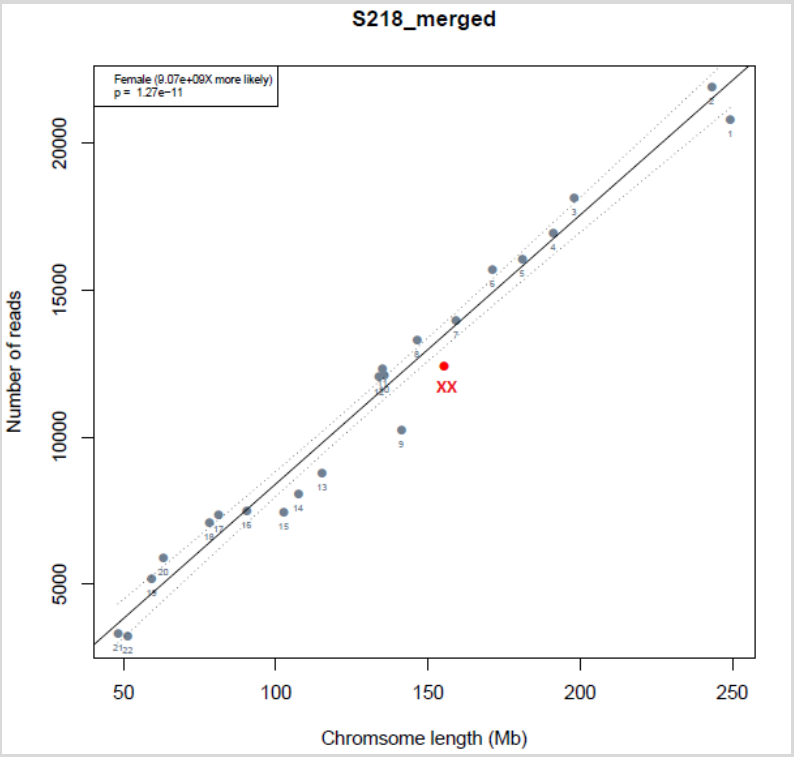
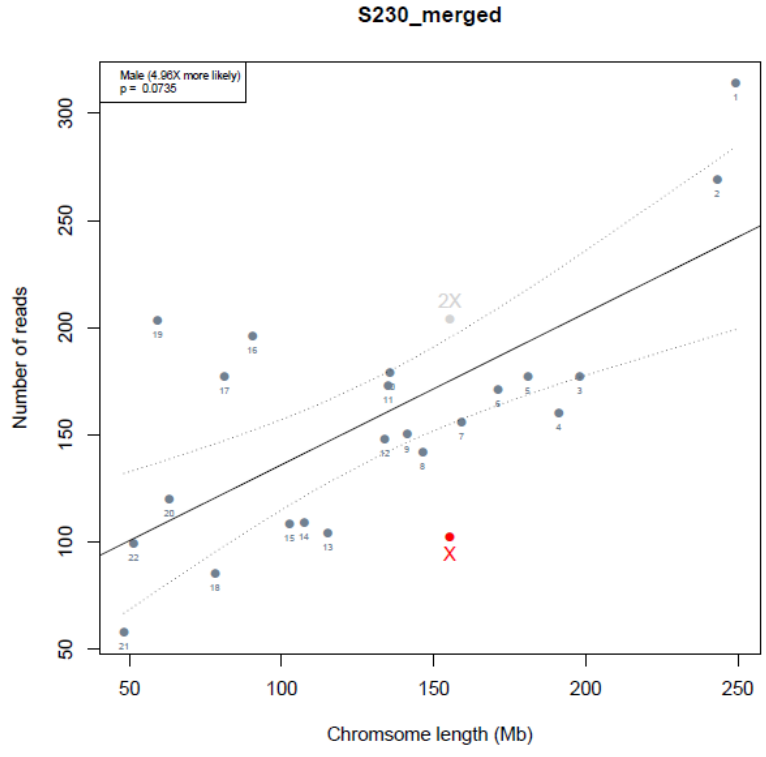


Female
(p = 1.89e−11)

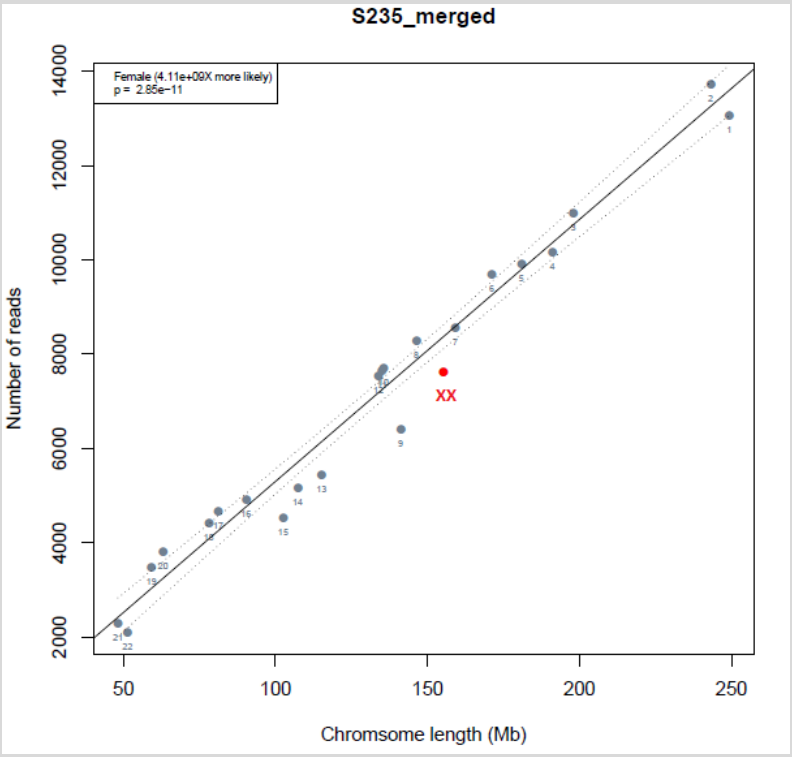
S217



Male
(p = 1.85e−08)

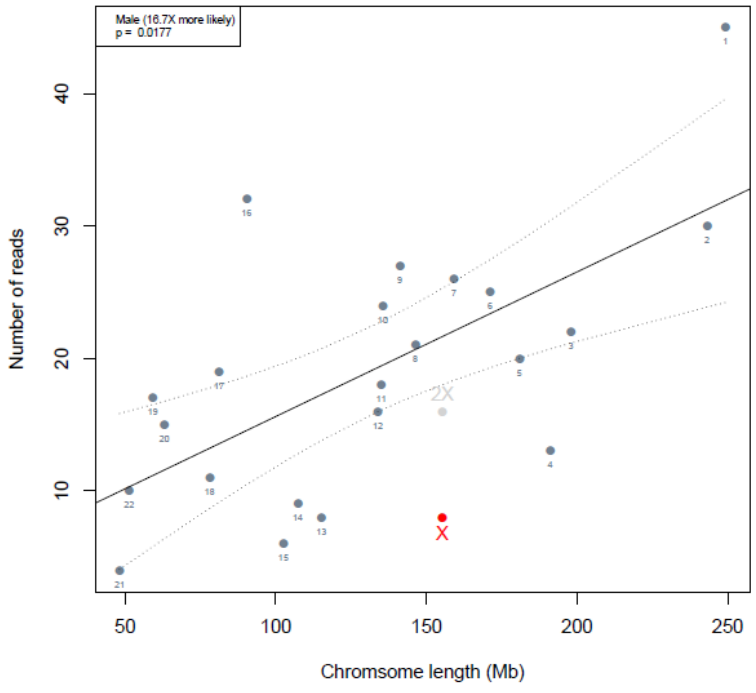
S218		Female (p = 1.27e−11)
S230		unable to be determined

S235



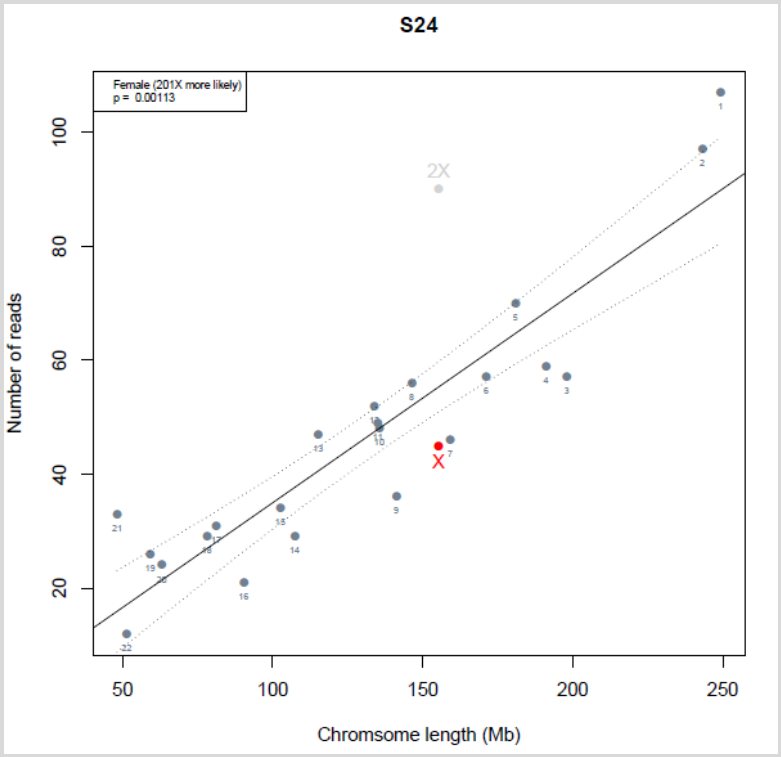
Female
(p = 2.85e−11)

S238



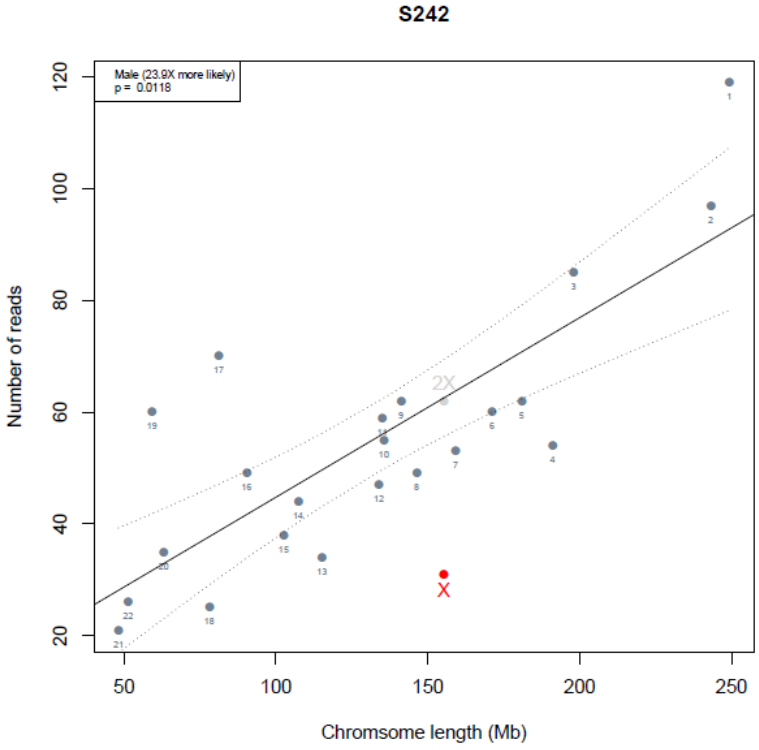
Male
(p = 0.0177)

S24

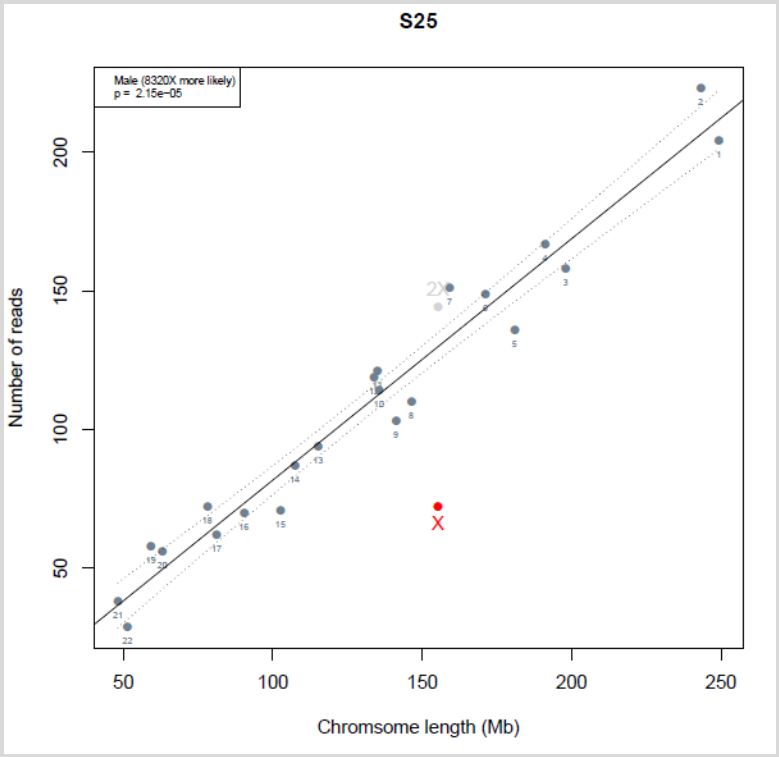
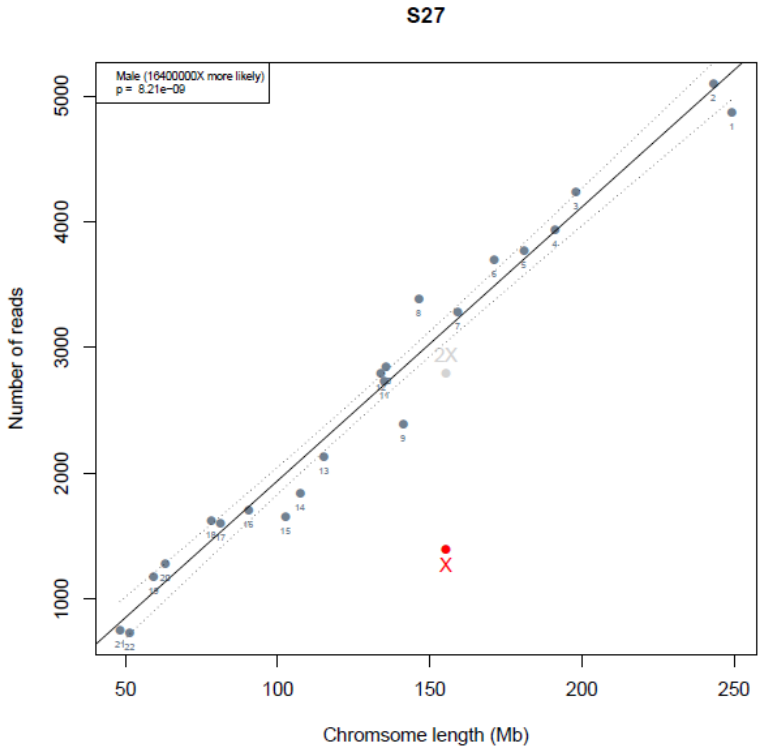


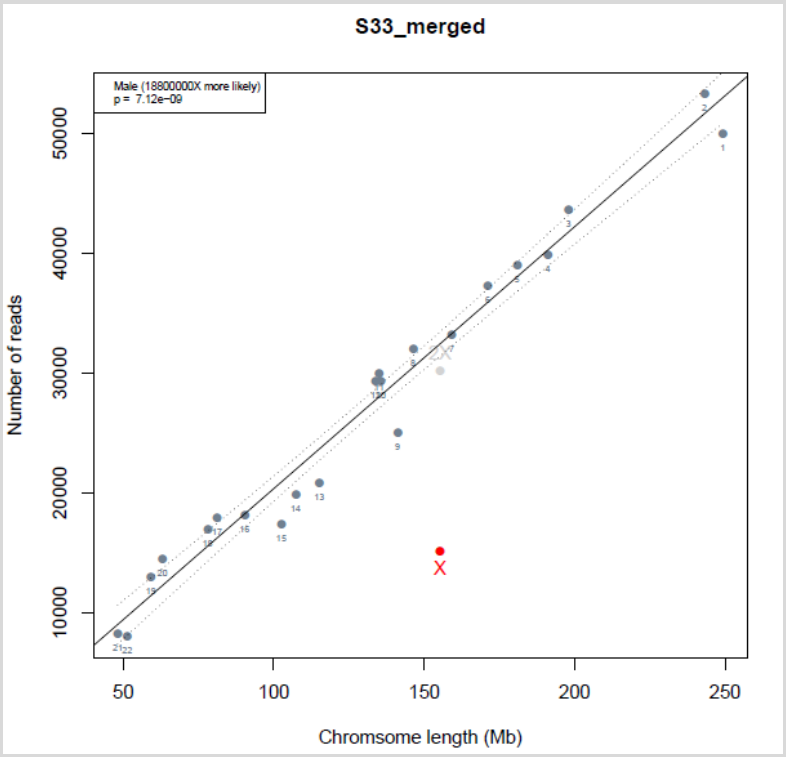
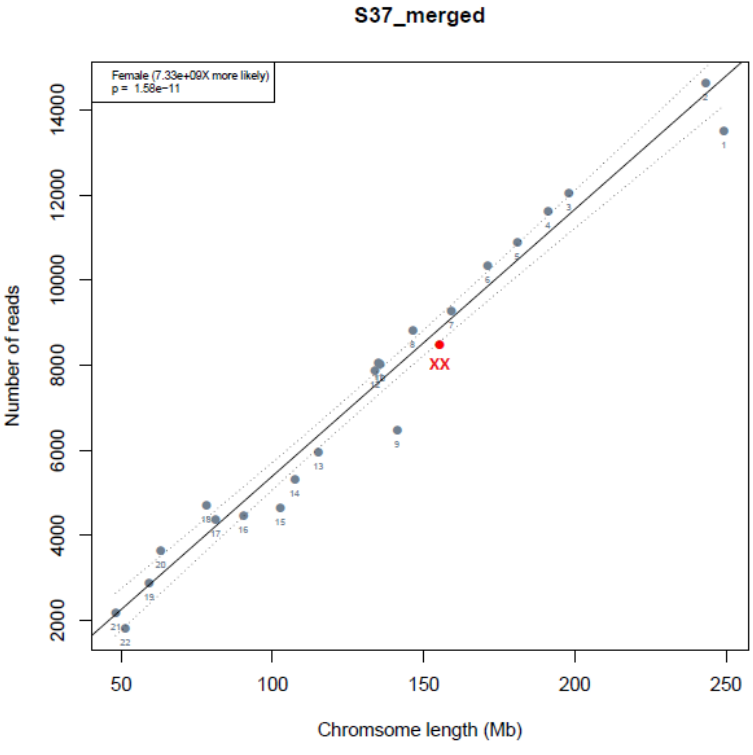
Female
(p = 0.00113)

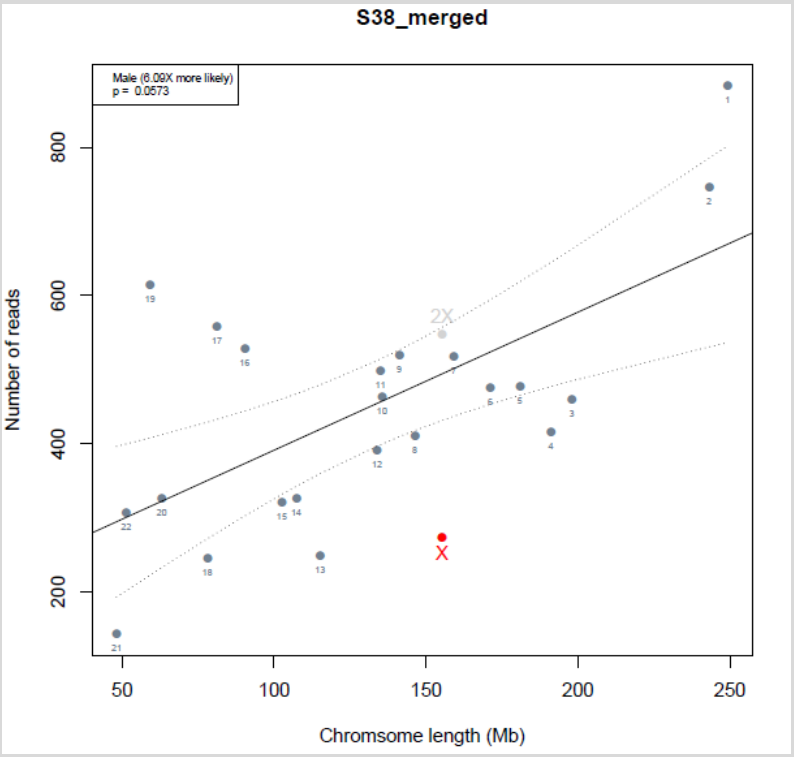
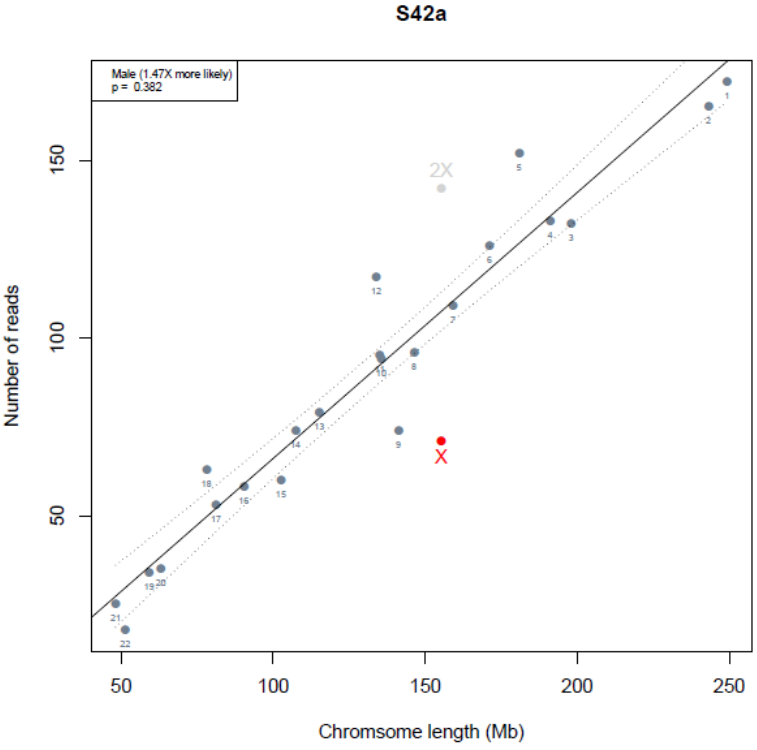
S242



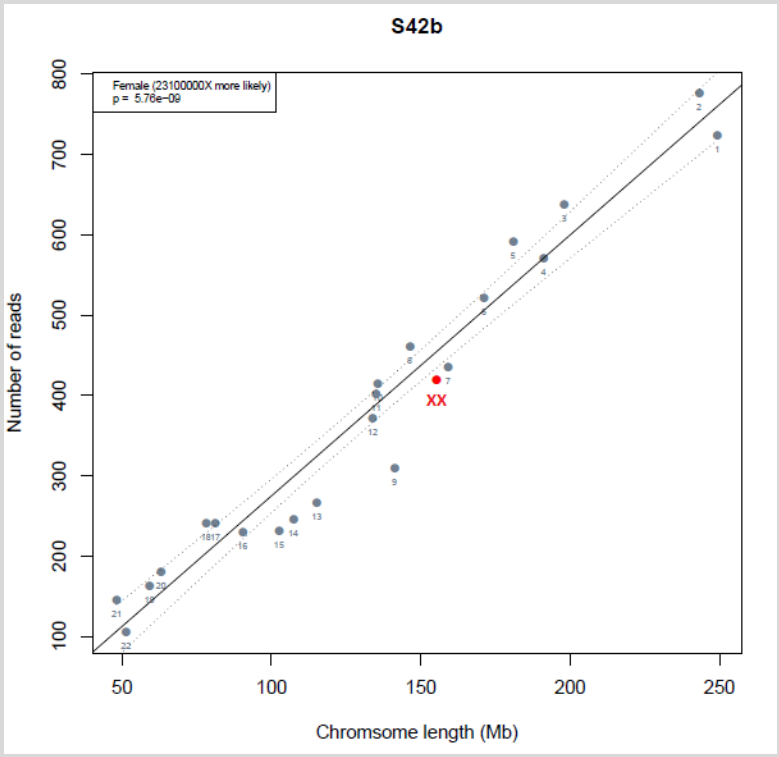
Male
(p = 0.0118)

S25		Male (p = 2.15e−05)
S27*		Male (p = 8.21e−09)

S33*		Male (p = 7.12e−09)
S37		Female (p = 1.58e−11)

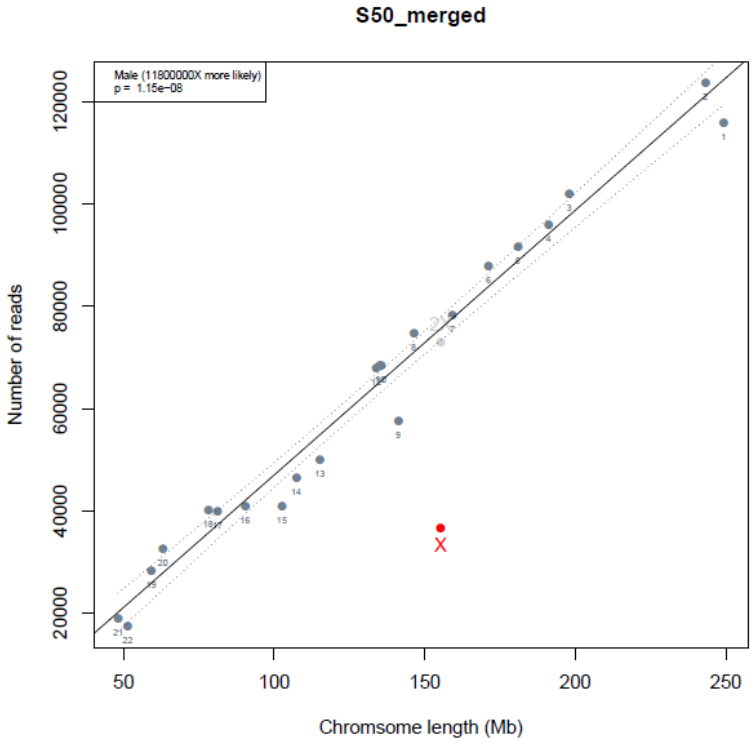
S38		Male? (p = 0.0573)
S42a		unable to be determined

S42b

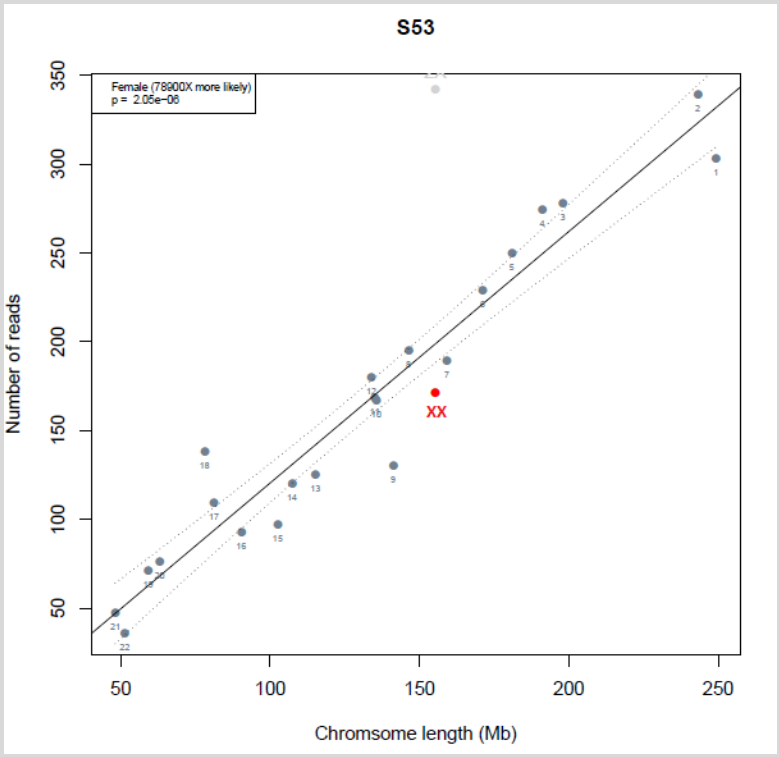
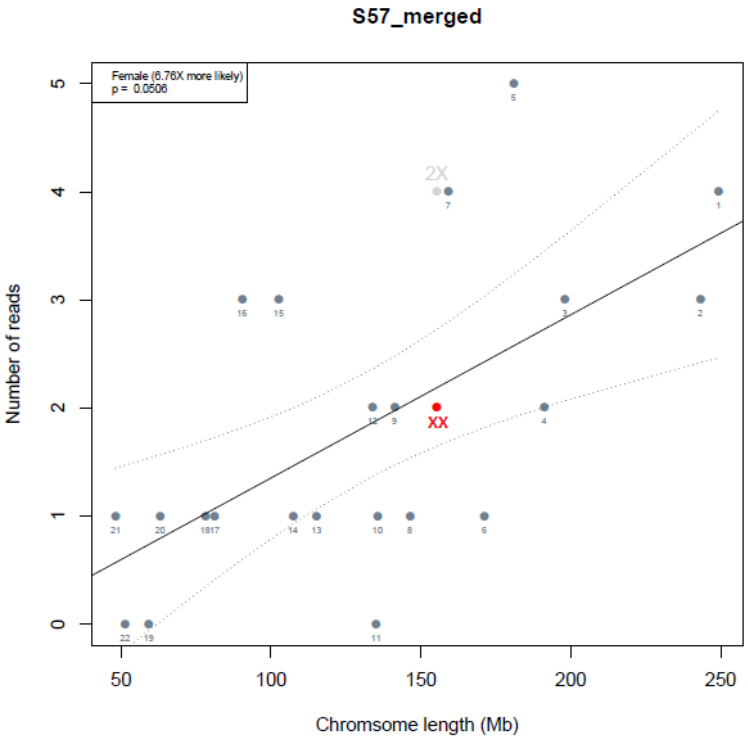


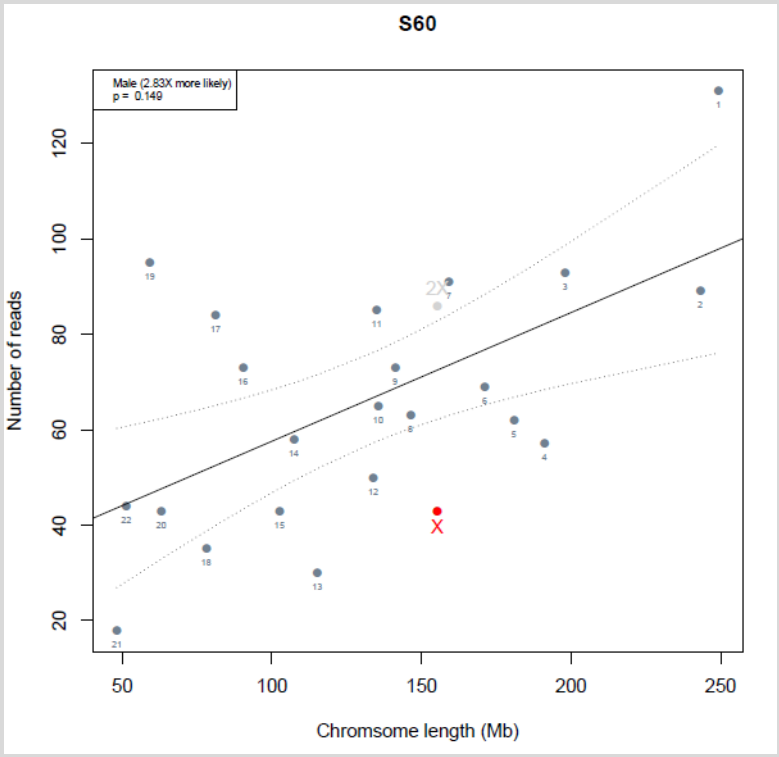
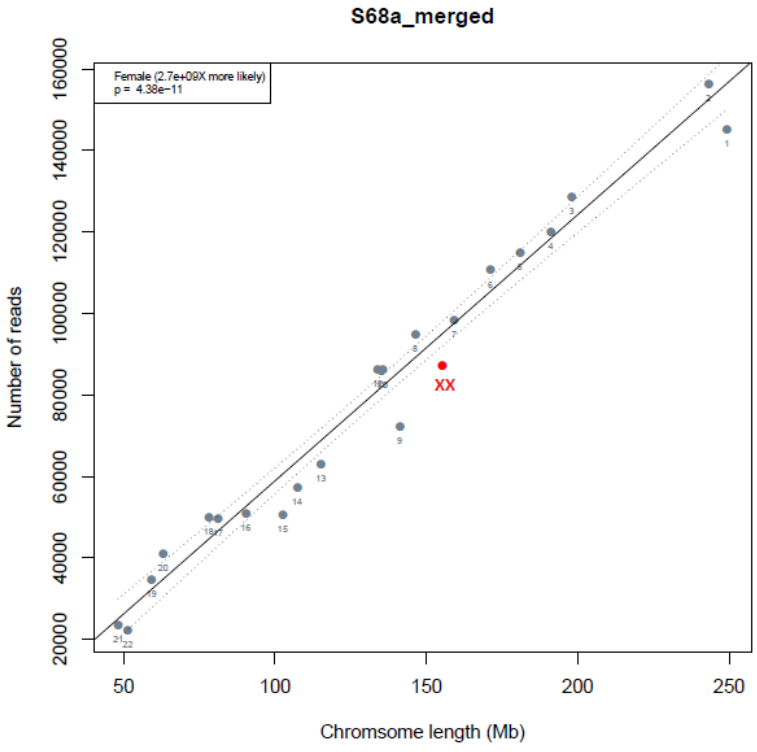
Female
(p = 5.76e−09)

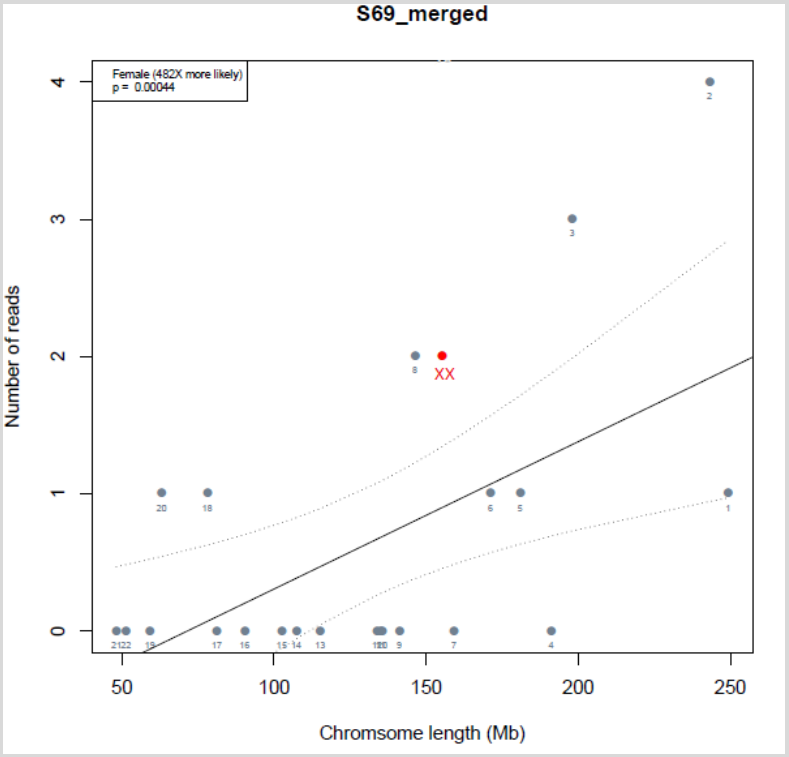
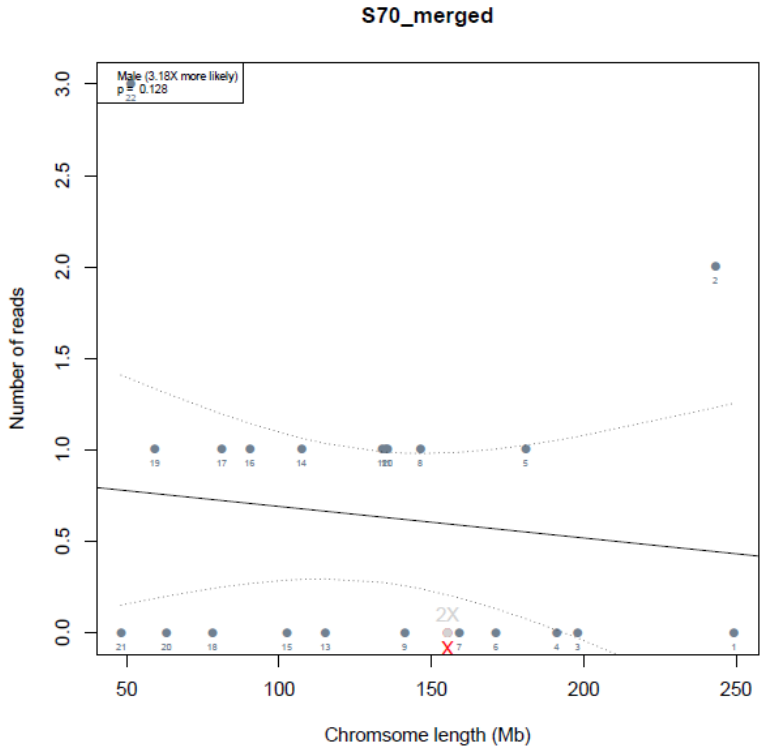
S50*



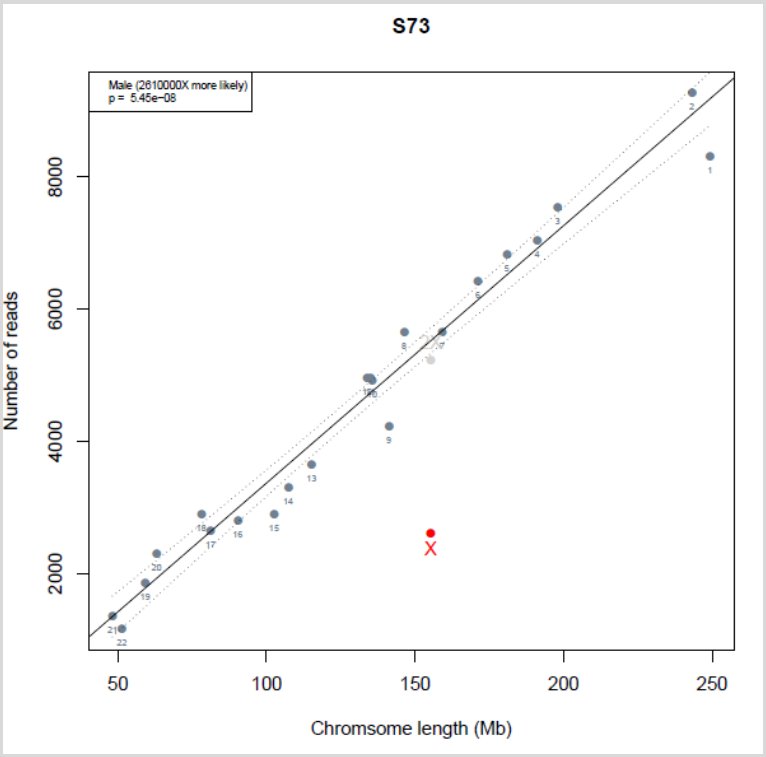
Male
(p = 1.15e−08)

S53		Female (p = 2.05e−06)
S57		Female? (p = 0.0506)

S60		unable to be determined
S68a*		Female (p = 4.38e−11)

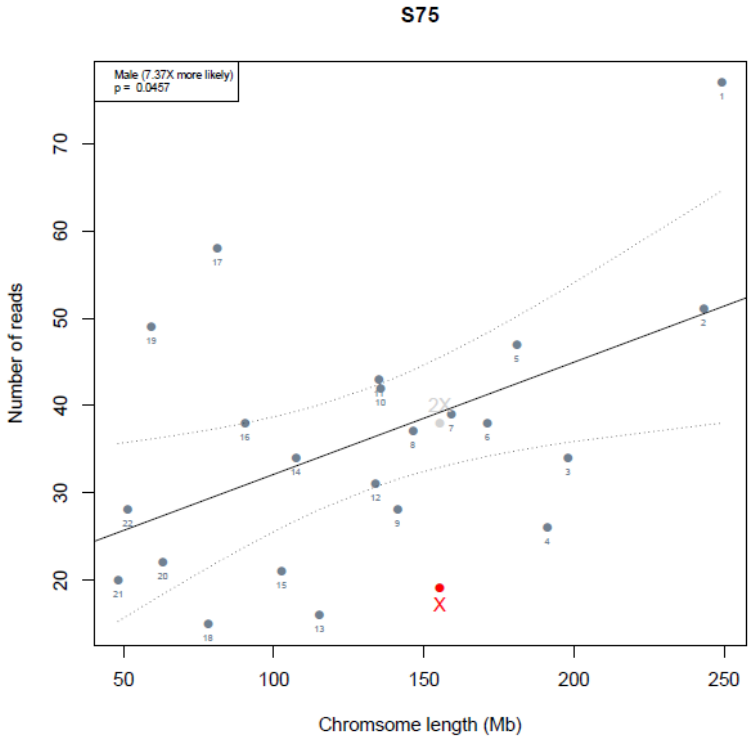
S69		Female (p = 0.00044)
S70		unable to be determined

S73*

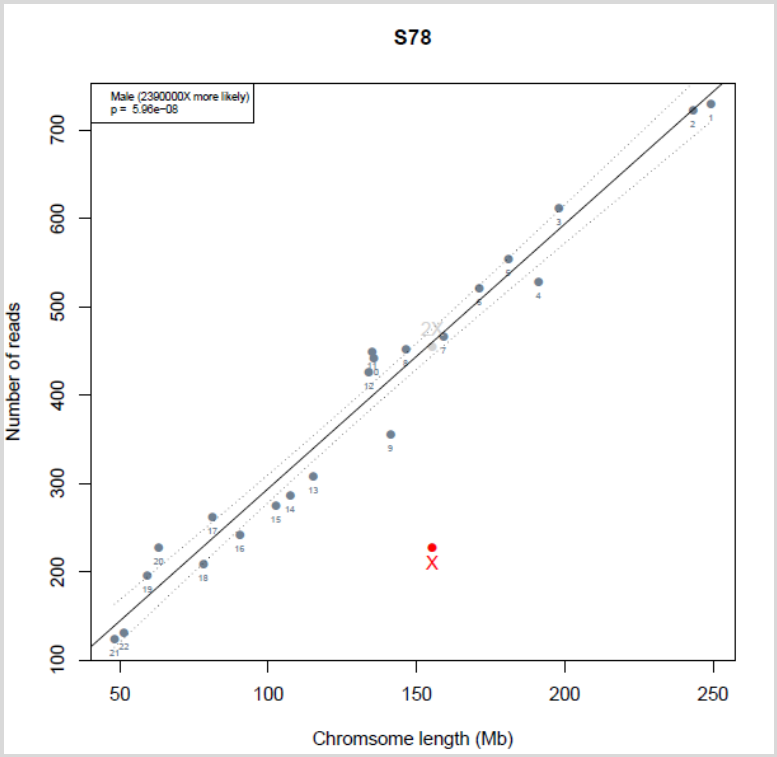
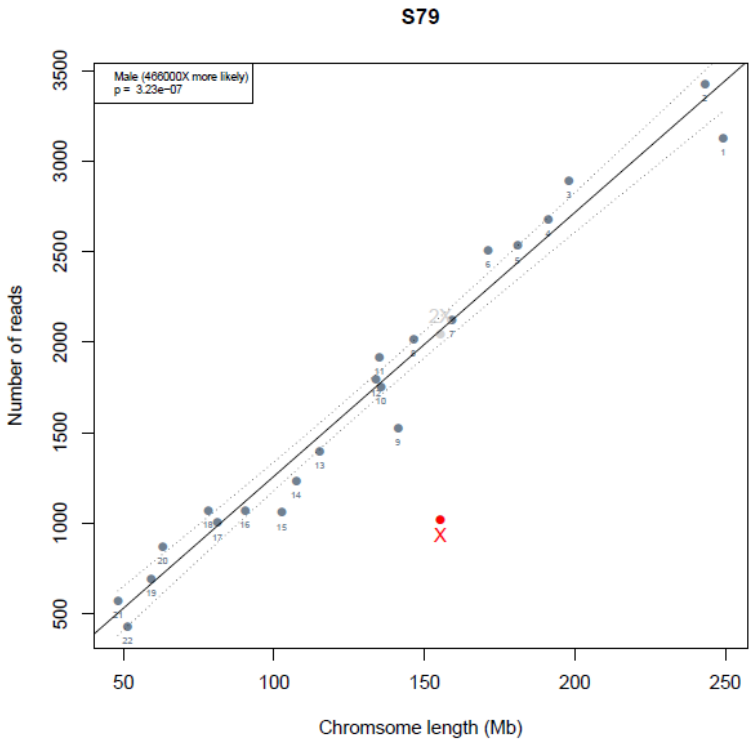


Male
(p = 5.45e−08)

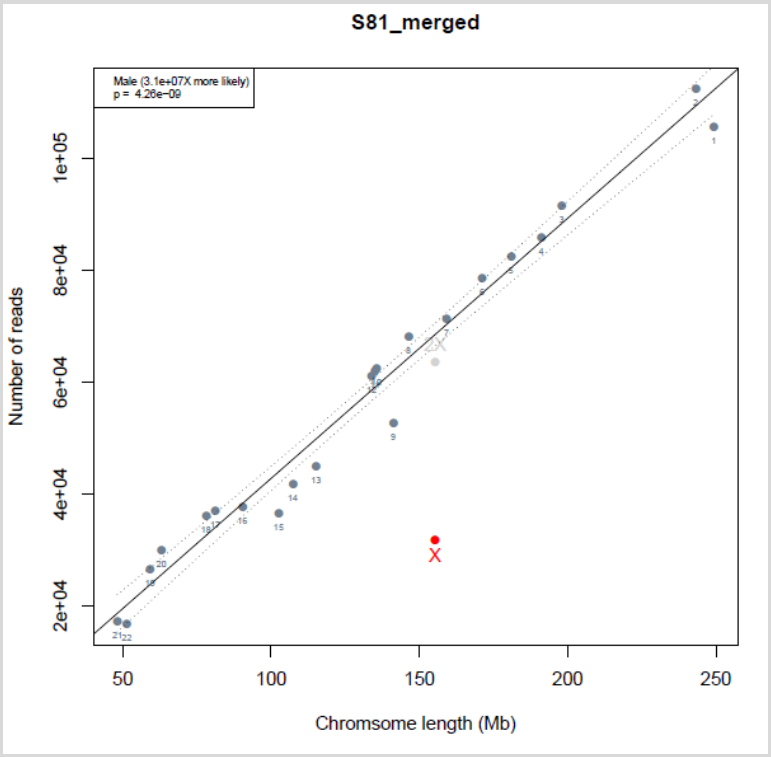
S75



Male?
(p = 0.0457)

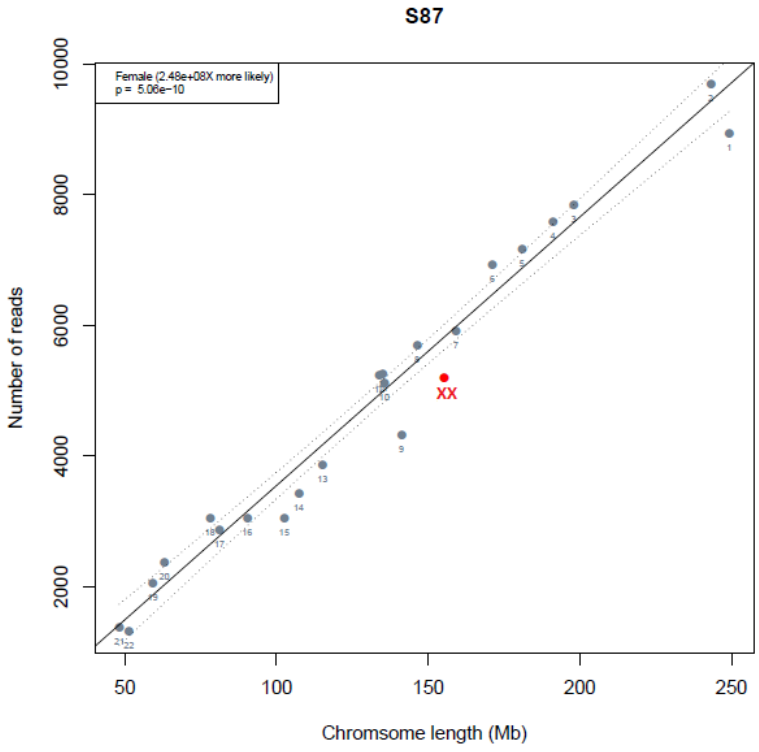
S78		Male (p = 5.96e−08)
S79*		Male (p = 3.23e−07)

S81*



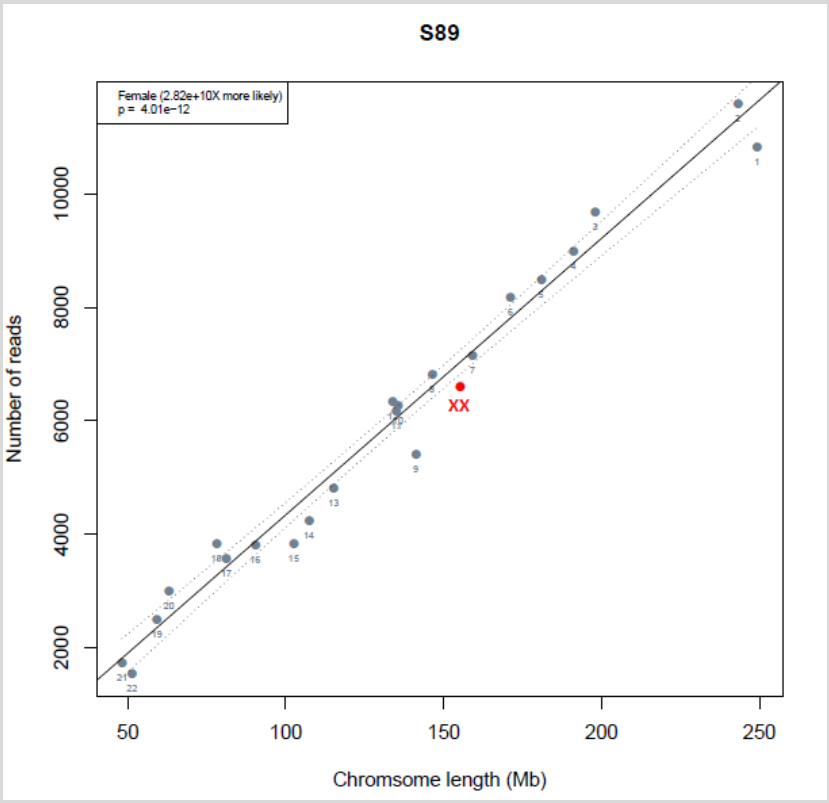
Male
(p = 4.26e−09)

S87*



Female
(p = 5.06e−10)

S89*



Female
(p = 4.01e−12)

SUPPLEMENT 5: PCA

The following table presents the information for the populations included in the PCA plots in Figures 7.1 and 7.2 in Chapter 7 and Figures 8.1 and 8.2 in Chapter 8. This information includes the name of the population, the population abbreviation used for Figures S5.1 and S5.2, the number of individuals included in that population, whether the population was included in the zoomed PCA (Figures 7.2, 8.2 and S5.2), and the geographic grouping for PCA analysis.

Figures S5.1 and S5.2 are identical to the plots presented in Chapter 7 as Figures 7.1 and 7.2 and Chapter 8 as Figures 8.1 and 8.2; however, instead of using symbols to represent each individual in PCA space, Figures S5.1 and S5.2 represent each individual using the population abbreviation listed in Table S5.1.

Table S5.1: Abbreviation, geographic grouping, and number of individuals included in the PCA in Figure 7.1; the column titled "Included in Zoomed PCA" indicates whether the samples are present in the PCA in Figure 7.2.

Population Name	Population Abbreviation	Number of Individuals (full PCA)	Included in Zoomed PCA	Geographic Grouping for PCA
Abkhasian	ABK	9	no	Caucasus
Adygei	ADY	16	no	Caucasus
Algerian	ALG	7	yes	Northern Africa
Armenian	ARM	10	no	Caucasus
Assyrian	ASS	11	yes	Near East
Balkar	BLK	10	no	Caucasus
BantuKenya	BAN	6	yes	Eastern Africa
BantuSA	BSA	5	no	Southern Africa
Bantu_Herero	BHE	2	no	Southern Africa
Bantu_Ovambo	BOV	1	no	Southern Africa
BedouinA	BEDA	25	yes	Near East
BedouinB	BEDB	19	yes	Near East
Biaka	BIA	20	no	Central Africa

Chechen	CHE	9	no	Caucasus
Cypriot	CYP	8	no	Near East
Damara	DAM	12	no	Southern Africa
Datog	DAT	3	yes	Eastern Africa
Dinka	DIN	7	yes	Eastern Africa
Druze	DRU	39	yes	Near East
Egyptian	EGY	18	yes	Northern Africa
Esan	ESA	8	yes	Western Africa
Gambian	GAM	6	yes	Western Africa
Gana	GAN	8	no	Southern Africa
Georgian	GEO	10	no	Caucasus
Gui	GUI	7	no	Southern Africa
Hadza	HAD	22	no	Eastern Africa
Haiom	HAI	7	no	Southern Africa
Himba	HIM	4	no	Southern Africa
Hoan	HOA	7	no	Southern Africa
Iran_Non-Zoroastrian_Fars	INZ	17	yes	Near East
Iran_Zoroastrian	IZO	26	yes	Near East
Iranian	IRA	38	yes	Near East
Jew_Ethiopian	ETHJew	7	yes	Eastern Africa
Jew_Georgian	GEOJew	7	no	Caucasus
Jew_Iranian	IRAJew	9	yes	Near East
Jew_Iraqi	IRQJew	6	yes	Near East
Jew_Libyan	LIBJew	9	yes	Northern Africa
Jew_Moroccan	MORJew	6	yes	Northern Africa
Jew_Tunisian	TUNJew	7	yes	Northern Africa
Jew_Turkish	TURJew	8	yes	Near East
Jew_Yemenite	YEMJew	8	yes	Near East
Jordanian	JOR	9	no	Near East
Ju_hoan_North	JHN	22	no	Southern Africa
Ju_hoan_South	JHS	6	no	Southern Africa
Kalmyk	KAL	10	no	Caucasus
Khomani	KHO	11	no	Southern Africa
Khwe	KHW	8	no	Southern Africa
Kikuyu	KIK	4	yes	Eastern Africa
Kumyk	KUM	8	no	Caucasus
Lebanese	LEB	8	no	Near East
Lebanese_Christian	LEC	9	no	Near East
Lebanese_Muslim	LEM	11	no	Near East
Lezgin	LEZ	9	no	Caucasus

Libyan	LIB	5	yes	Northern Africa
Luhya	LUH	8	yes	Eastern Africa
Luo	LUO	8	yes	Eastern Africa
Mandenka	MAN	17	yes	Western Africa
Masai	MAS	12	yes	Eastern Africa
Mbuti	MBT	10	no	Central Africa
Mende	SLE	8	yes	Western Africa
Moroccan	MOR	10	yes	Northern Africa
Mozabite	MOZ	21	yes	Northern Africa
Nama	NAM	16	no	Southern Africa
Naro	NAR	8	no	Southern Africa
Nogai	NOG	9	no	Caucasus
Oromo	ORM	4	yes	Eastern Africa
Ossetian	OSS	10	no	Caucasus
Palestinian	PAL	38	yes	Near East
Saharawi	SAH	6	yes	Northern Africa
Sandawe	TAN	22	no	Eastern Africa
Saudi	SAU	8	yes	Near East
Shua	SHU	9	no	Southern Africa
Somali	SOM	13	yes	Eastern Africa
Syrian	SYR	8	no	Near East
Taa_East	TAE	7	no	Southern Africa
Taa_North	TAA	9	no	Southern Africa
Taa_West	TAW	16	no	Southern Africa
Tshwa	TSH	5	no	Southern Africa
Tswana	TSW	5	no	Southern Africa
Tunisian	TUN	8	yes	Northern Africa
Turkish	TUR	50	no	Near East
Turkish_Balikesir	BAL	5	no	Near East
Xuun	XUU	13	no	Southern Africa
Yemeni	YEM	6	yes	Near East
Yoruba	YOR	70	no	Western Africa

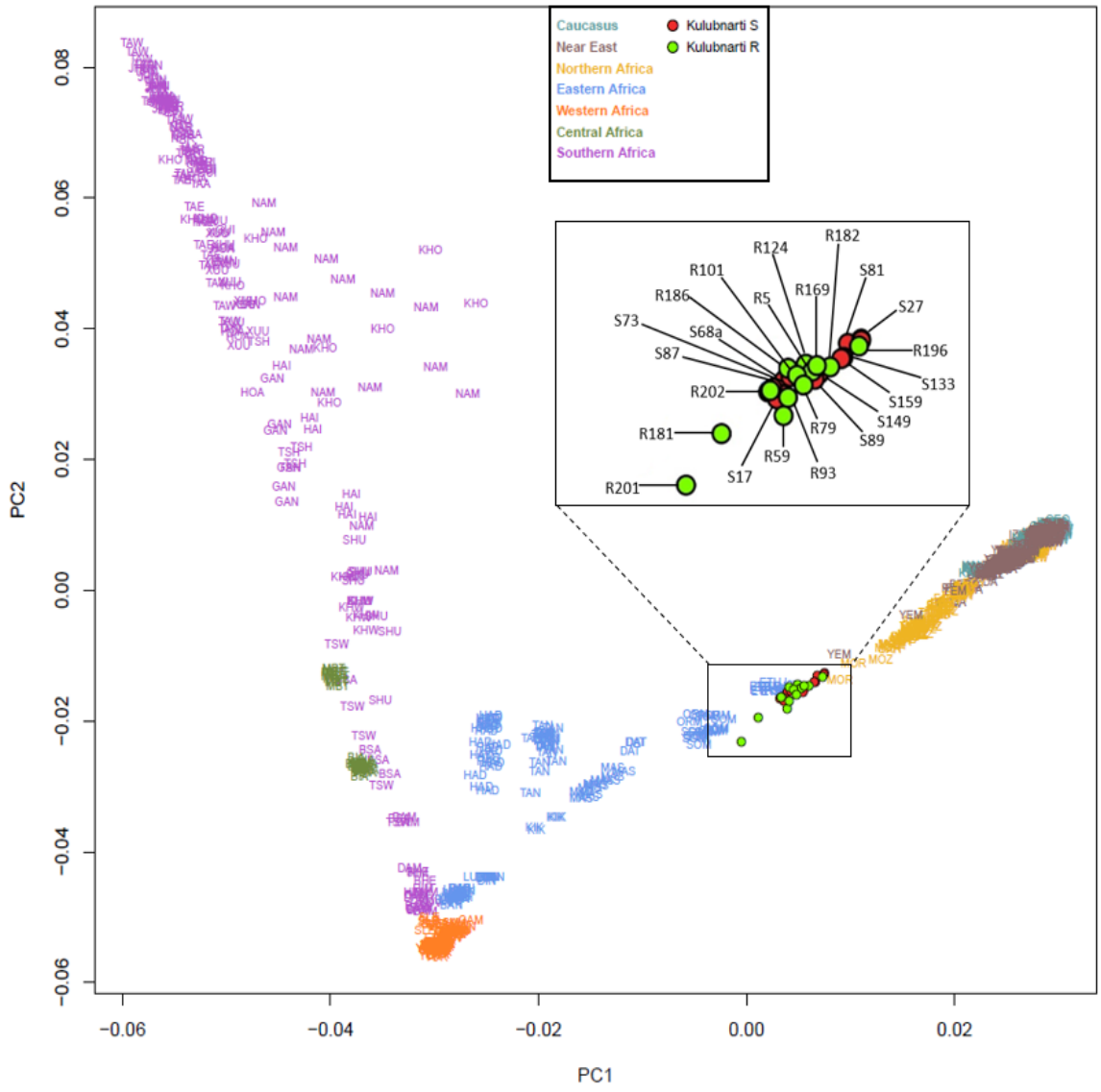


Figure S5.1: PCA projecting the individuals from the S and R communities at Kulubnarti onto Caucasian, Near Eastern, and African populations from the Human Origins dataset. Note that the clustering of the Kulubnarti Nubians obscures several samples: R152 and R198 are obscured by the cluster including R5, R124, and R169; S33 and S50 are obscured by the cluster containing R202, S87, S73, and S68a; and S79 is obscured by S27.

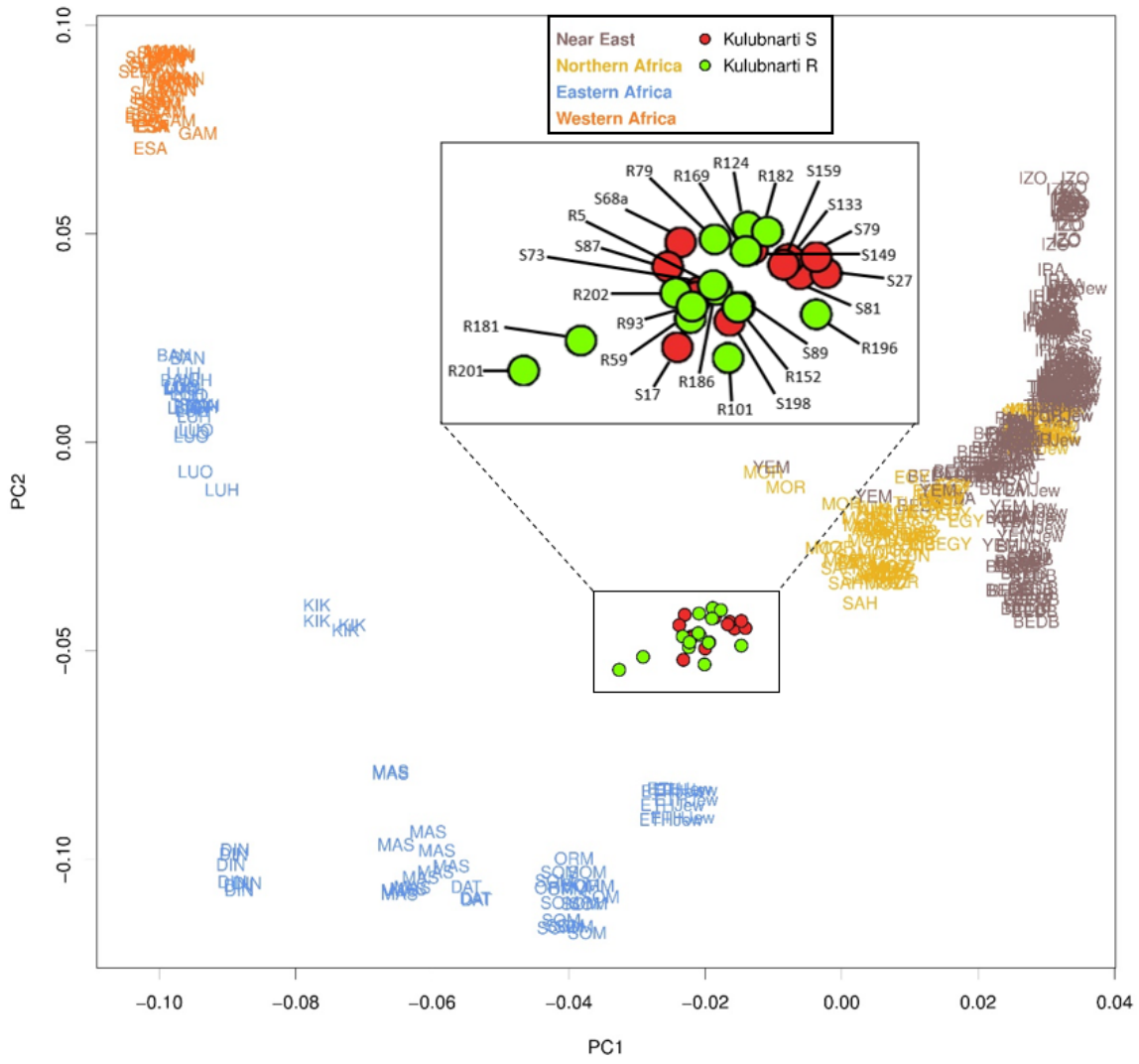


Figure S5.2: PCA projecting the individuals from the S and R communities at Kulubnarti onto selected Near Eastern and African populations from the Human Origins dataset. Note that S33 and S50 are obscured by the cluster of samples containing R5, R202, R93, R59, S17, and R186.

SUPPLEMENT 6: PAIRWISE F_{ST} ESTIMATES

The following two tables present the pairwise F_{ST} estimates for the S community and R community and 260 present-day populations from around the world (Tables S6.1 and S6.2 respectively). The associated standard error and Z-score are reported along with each F_{ST} estimate. All pairwise F_{ST} estimates are based on a minimum of 10,000 SNPs and require that present-day populations have a minimum of 5 individuals. $|Z| \geq 3.0$ is interpreted to indicate a statistically significant deviation from zero.

The pairwise F_{ST} estimate and associated standard error and Z-score reported for Kulubnarti S and Kulubnarti R communities is the average of 1,000 permutations, indicated in the tables by the "+" symbol. A non-significant Z-score is indicated by the "*" symbol.

Table S6.1: Pairwise F_{ST} estimates between the Kulubnarti S community and the Kulubnarti R community, as well as 260 present-day populations. The table is ordered from lowest to highest F_{ST} estimates.

	FST (Kulubnarti S, Test)		
Test Population	**Estimate**	**SE**	**Z-score**
Kulubnarti R	0.0024^{+}	0.0009^{+}	2.52*$^{+}$
Jew_Ethiopian	0.016	0.0007	22.14
Somali	0.020	0.0006	33.96
Moroccan	0.022	0.0007	30.76
Egyptian	0.023	0.0007	34.29
Tunisian	0.023	0.0008	29.50
Yemeni	0.023	0.0009	28.07
Libyan	0.024	0.0009	27.74
BedouinA	0.030	0.0007	42.49
Saharawi	0.030	0.0009	30.89
Jordanian	0.033	0.0009	38.05
Mozabite	0.033	0.0007	48.07
Masai	0.033	0.0007	47.58
Lebanese	0.035	0.0009	39.97
Palestinian	0.035	0.0007	49.02

Syrian	0.035	0.0009	39.08
Algerian	0.037	0.0009	42.07
Lebanese_Muslim	0.037	0.0008	45.07
Iranian	0.038	0.0007	58.23
Kikuyu	0.038	0.0010	35.52
Saudi	0.038	0.0009	42.55
Iranian_Bandari	0.039	0.0009	45.76
Jew_Yemenite	0.039	0.0009	42.96
Jew_Turkish	0.041	0.0009	47.79
Lebanese_Christian	0.041	0.0009	48.21
Jew_Moroccan	0.042	0.0009	45.35
Cypriot	0.043	0.0009	46.03
Iran_Non-Zoroastrian_Fars	0.043	0.0008	54.81
Jew_Libyan	0.043	0.0008	51.74
Jew_Tunisian	0.043	0.0009	45.46
Maltese	0.043	0.0009	47.82
Sandawe	0.043	0.0007	63.38
Sicilian	0.043	0.0008	52.21
Turkish	0.043	0.0008	57.13
Druze	0.044	0.0008	57.77
Jew_Iraqi	0.044	0.0009	47.59
Armenian	0.045	0.0009	52.08
Jew_Ashkenazi	0.045	0.0009	48.48
Assyrian	0.046	0.0009	52.19
Italian_South	0.046	0.0008	56.70
Makrani	0.046	0.0008	58.99
Greek	0.047	0.0008	56.57
Italian_Central	0.047	0.0008	55.73
Kumyk	0.047	0.0009	50.94
Albanian	0.048	0.0010	49.80
Balkar	0.048	0.0009	56.81
Balochi	0.048	0.0008	62.96
Nogai	0.048	0.0009	53.46
Bulgarian	0.049	0.0009	56.23
Jew_Georgian	0.049	0.0010	49.70
Spanish	0.049	0.0008	61.54
Abkhasian	0.050	0.0009	58.62
Adygei	0.050	0.0008	60.28
BedouinB	0.050	0.0008	58.61
Brahui	0.050	0.0008	65.15
Georgian	0.050	0.0009	55.94
Italian_North	0.050	0.0008	66.33
Lezgin	0.050	0.0009	55.49

Ossetian	0.050	0.0009	58.02
Pathan	0.050	0.0008	64.01
Turkmen	0.050	0.0009	52.89
Sindhi_Pakistan	0.051	0.0008	63.47
Tajik	0.051	0.0009	59.32
Jew_Cochin	0.052	0.0011	48.61
Romanian	0.052	0.0009	56.30
Brahmin_Tiwari	0.053	0.0008	66.56
Chechen	0.053	0.0009	56.23
Croatian	0.053	0.0009	59.37
French	0.053	0.0008	65.74
Hungarian	0.053	0.0008	61.80
Uzbek	0.053	0.0009	61.64
Burusho	0.055	0.0008	70.01
Czech	0.055	0.0009	60.44
English	0.055	0.0009	59.15
German	0.055	0.0009	59.39
Sardinian	0.055	0.0008	66.93
Scottish	0.055	0.0008	66.81
Ukrainian	0.055	0.0009	55.57
Polish	0.056	0.0008	68.14
Belarusian	0.057	0.0009	63.53
Frisian	0.057	0.0010	59.61
Iran_Zoroastrian	0.057	0.0008	69.56
Irish	0.057	0.0009	65.52
Mordovian	0.057	0.0009	60.94
Norway_Southeast	0.057	0.0010	57.48
Norway_West	0.057	0.0011	51.17
Norwegian	0.057	0.0010	60.30
Russian	0.057	0.0008	68.14
Icelandic	0.058	0.0009	63.52
Orcadian	0.058	0.0009	65.50
Shetlandic	0.058	0.0010	58.79
Sorb	0.058	0.0009	61.68
Turkish_Balikesir	0.058	0.0012	48.13
Basque	0.059	0.0009	67.34
Chuvash	0.059	0.0009	63.10
Dinka	0.060	0.0010	60.62
Estonian	0.060	0.0010	62.81
Finnish	0.060	0.0010	59.96
Uygur	0.060	0.0010	63.97
Bengali	0.061	0.0010	61.85
Punjabi	0.061	0.0009	66.04

Lithuanian	0.062	0.0010	64.24
Luhya	0.062	0.0009	67.96
Luo	0.062	0.0009	65.55
Hazara	0.063	0.0009	73.68
Lodhi	0.063	0.0009	73.93
Vishwabrahmin	0.065	0.0009	74.28
BantuKenya	0.066	0.0010	64.54
Mala	0.067	0.0009	74.11
Gambian	0.070	0.0010	67.34
Yoruba	0.072	0.0008	89.37
Mandenka	0.072	0.0009	81.95
Aleut	0.073	0.0011	65.44
Mende	0.073	0.0009	79.91
Esan	0.074	0.0009	78.61
Wambo	0.074	0.0011	68.10
Mbukushu	0.074	0.0010	71.21
Kyrgyz	0.075	0.0010	72.74
BantuSA	0.076	0.0011	68.01
Kalash	0.077	0.0010	80.15
Komi_Zyrian	0.077	0.0010	73.83
Tswana	0.078	0.0011	70.94
Shua	0.079	0.0010	78.59
Khwe	0.081	0.0010	76.92
Even	0.082	0.0010	83.29
Mansi	0.082	0.0011	74.49
Damara	0.083	0.0009	88.12
Altaian	0.085	0.0011	75.88
Himba	0.085	0.0012	68.36
Tubalar	0.085	0.0009	91.01
Haiom	0.088	0.0011	80.31
Kharia	0.088	0.0010	83.30
Nama	0.090	0.0011	83.58
Kalmyk	0.091	0.0011	84.64
Khanty	0.091	0.0010	86.41
Selkup	0.092	0.0011	84.97
Kgalagadi	0.093	0.0012	74.91
Hadza	0.094	0.0012	81.64
Tuvinian	0.098	0.0011	86.44
Ojibwa	0.100	0.0011	94.17
Biaka	0.102	0.0010	102.22
Mon	0.102	0.0012	88.18
Cambodian	0.103	0.0012	82.91
Cree	0.103	0.0012	87.97

Thai	0.104	0.0011	92.02
Tu	0.104	0.0011	94.41
Yukagir	0.104	0.0010	106.85
Kusunda	0.105	0.0012	90.18
Mongola	0.107	0.0012	86.27
Gana	0.108	0.0012	93.30
Khomani	0.109	0.0012	90.21
Xibo	0.110	0.0012	91.50
Yakut	0.111	0.0011	102.99
Tshwa	0.111	0.0016	69.21
Daur	0.112	0.0012	93.47
RapaNui	0.112	0.0018	62.20
Khmer	0.113	0.0012	91.43
Lao	0.113	0.0012	92.01
Kinh	0.115	0.0012	93.52
Yi	0.115	0.0011	100.65
Hezhen	0.116	0.0012	98.96
Hakka_Taiwan	0.117	0.0011	103.00
Han	0.117	0.0011	103.36
Naxi	0.117	0.0012	97.21
Nyah_Kur	0.117	0.0012	96.58
Han_Taiwan	0.118	0.0012	96.33
Tai_Lue	0.118	0.0012	97.32
Tujia	0.118	0.0012	98.56
Japanese	0.119	0.0011	109.17
Korean	0.119	0.0011	105.15
Micronesian	0.119	0.0012	102.53
Xuun	0.119	0.0012	103.77
Dai	0.120	0.0012	97.89
Borneo	0.121	0.0013	95.39
Kuy_Suay	0.121	0.0012	99.85
Miao	0.121	0.0012	97.55
Lawa	0.122	0.0012	99.09
Samoan	0.123	0.0012	92.98
Makira	0.124	0.0014	90.80
Ulchi	0.126	0.0011	113.24
Lahu	0.127	0.0013	101.43
Malaita	0.127	0.0014	90.71
Nggela	0.127	0.0015	86.95
Savo	0.127	0.0014	90.52
Semende	0.127	0.0013	97.85
Karen_Sgaw	0.128	0.0013	97.74
Santa_Isabel	0.129	0.0014	92.97

Chipewyan	0.131	0.0012	112.11
Buka	0.132	0.0015	90.87
Mbuti	0.133	0.0012	109.04
Ranongga	0.135	0.0016	84.52
Vella_Lavella	0.135	0.0015	90.73
Ontong_Java	0.136	0.0014	99.03
Saposa	0.136	0.0014	95.86
Ami	0.137	0.0013	104.09
Kolombangara	0.137	0.0016	86.72
Algonquin	0.138	0.0016	84.60
Mamanwa	0.138	0.0014	97.07
Tongan	0.138	0.0014	97.75
Chukchi	0.139	0.0012	111.95
Hoan	0.139	0.0015	91.35
Manus	0.139	0.0023	60.82
Choiseul	0.140	0.0015	93.97
Eskimo_ChaplinSireniki	0.140	0.0014	101.32
Mussau	0.140	0.0015	94.36
Notsi	0.140	0.0015	95.99
Hmong	0.141	0.0013	107.59
Taa_East	0.141	0.0014	99.31
Naro	0.142	0.0014	104.72
Mangseng	0.143	0.0016	90.26
Nailik	0.143	0.0015	96.02
Htin_Mal	0.144	0.0014	103.49
Mayan	0.144	0.0014	104.27
Madak	0.145	0.0016	89.96
Tigak	0.146	0.0014	101.84
Koryak	0.147	0.0014	107.91
Teop	0.147	0.0015	97.07
Gui	0.148	0.0016	93.44
Kuot_Lamalaua	0.148	0.0016	92.42
Lavongai	0.148	0.0014	104.14
Taa_North	0.148	0.0014	106.24
Taa_West	0.148	0.0013	114.85
Tikopia	0.148	0.0016	94.94
Tolai	0.148	0.0014	107.05
Bolivian	0.149	0.0016	91.43
Nganasan	0.149	0.0013	117.63
Itelmen	0.150	0.0014	109.17
Melamela	0.150	0.0015	102.59
Quechua	0.150	0.0016	94.34
Mengen	0.151	0.0015	101.66

Kove	0.152	0.0014	107.14
Kuot_Kabil	0.152	0.0016	97.35
Nakanai_Bileki	0.152	0.0016	96.66
Ju_hoan_South	0.153	0.0015	104.64
Eskimo_Naukan	0.154	0.0014	106.82
Zapotec	0.154	0.0014	108.35
Ju_hoan_North	0.155	0.0013	122.70
Sulka	0.155	0.0014	108.82
Santa_Cruz	0.156	0.0016	97.89
Atayal	0.158	0.0014	111.51
Mixtec	0.158	0.0015	104.46
Kaqchikel	0.161	0.0016	98.38
Nasioi	0.163	0.0016	100.95
Aymara	0.167	0.0017	96.77
Mamusi_Paleabu	0.168	0.0018	92.86
Guarani	0.169	0.0017	99.46
Ata	0.173	0.0017	102.06
Onge	0.174	0.0015	112.84
Rennell_and_Bellona	0.174	0.0017	99.83
Nakanai_Loso	0.177	0.0018	101.14
Mixe	0.178	0.0015	116.47
Mamusi	0.181	0.0016	114.21
Pima	0.183	0.0016	117.66
New_Guinea	0.186	0.0017	111.12
Papuan	0.187	0.0016	113.70
Gimi	0.196	0.0017	114.58
Xavante	0.204	0.0017	121.67
Karitiana	0.221	0.0017	129.66
Mlabri	0.229	0.0021	107.16
Surui	0.230	0.0018	125.00
Cabecar	0.233	0.0020	116.90

Table S6.2: Pairwise F_{ST} estimates between the Kulubnarti R community and Kulubnarti S community, as well as 260 present-day populations. The table is ordered from lowest to highest F_{ST} estimates.

Test Population	FST (Kulubnarti R, Test)		
	Estimate	SE	Z-score
Kulubnarti S[&]	0.0024[+]	0.0009[+]	2.52*[+]
Jew_Ethiopian	0.013	0.0009	15.33
Somali	0.017	0.0008	21.45
Moroccan	0.020	0.0009	22.50
Tunisian	0.020	0.0009	23.19
Egyptian	0.023	0.0009	26.57
Libyan	0.024	0.0011	22.07
Yemeni	0.024	0.0010	21.86
Masai	0.028	0.0009	32.13
BedouinA	0.029	0.0009	33.87
Saharawi	0.029	0.0011	25.61
Mozabite	0.032	0.0009	35.91
Jordanian	0.033	0.0010	32.29
Kikuyu	0.035	0.0012	26.90
Lebanese	0.035	0.0011	35.19
Palestinian	0.035	0.0009	40.49
Syrian	0.035	0.0010	35.02
Algerian	0.036	0.0011	34.01
Lebanese_Muslim	0.038	0.0010	38.46
Sandawe	0.039	0.0009	44.46
Iranian	0.040	0.0009	44.33
Iranian_Bandari	0.040	0.0011	39.04
Jew_Yemenite	0.040	0.0011	36.31
Saudi	0.040	0.0011	36.41
Jew_Turkish	0.042	0.0011	39.04
Lebanese_Christian	0.042	0.0010	41.85
Cypriot	0.043	0.0011	40.05
Iran_Non-Zoroastrian_Fars	0.043	0.0010	44.94
Jew_Libyan	0.043	0.0010	43.92
Jew_Moroccan	0.043	0.0012	35.44
Jew_Tunisian	0.043	0.0010	40.59
Maltese	0.044	0.0011	40.52
Sicilian	0.044	0.0010	42.81
Turkish	0.044	0.0009	47.36
Armenian	0.045	0.0010	45.55
Druze	0.045	0.0009	48.29
Assyrian	0.046	0.0010	44.82

Italian_South	0.046	0.0010	47.23
Jew_Iraqi	0.046	0.0011	42.25
Makrani	0.046	0.0009	49.67
Italian_Central	0.047	0.0010	48.37
Jew_Ashkenazi	0.047	0.0011	41.77
Albanian	0.048	0.0012	42.60
Greek	0.048	0.0010	48.83
Nogai	0.048	0.0011	46.52
Balochi	0.049	0.0010	52.56
Bulgarian	0.049	0.0010	46.24
Italian_North	0.049	0.0009	52.40
Kumyk	0.049	0.0011	46.29
Spanish	0.049	0.0009	52.06
Adygei	0.050	0.0010	51.99
Balkar	0.050	0.0011	48.95
BedouinB	0.050	0.0010	48.16
Brahui	0.050	0.0009	52.87
Jew_Georgian	0.050	0.0011	44.21
Ossetian	0.050	0.0010	49.08
Pathan	0.050	0.0010	51.99
Turkmen	0.050	0.0011	48.06
Abkhasian	0.051	0.0010	48.05
Georgian	0.051	0.0011	47.90
Sindhi_Pakistan	0.051	0.0009	53.77
Croatian	0.052	0.0010	50.23
Lezgin	0.052	0.0011	49.70
Romanian	0.052	0.0010	49.45
Uzbek	0.052	0.0010	50.50
Chechen	0.053	0.0010	50.82
French	0.053	0.0009	56.41
Hungarian	0.053	0.0009	53.72
Jew_Cochin	0.053	0.0013	41.17
Tajik	0.053	0.0011	47.45
Dinka	0.054	0.0011	48.47
Brahmin_Tiwari	0.055	0.0010	55.78
Czech	0.055	0.0011	51.14
German	0.055	0.0010	52.33
Sardinian	0.055	0.0010	55.22
Burusho	0.056	0.0010	57.64
English	0.056	0.0011	50.99
Scottish	0.056	0.0010	55.94
Ukrainian	0.056	0.0011	51.25

Luhya	0.056	0.0011	50.29
Belarusian	0.057	0.0010	56.21
Frisian	0.057	0.0011	53.39
Iran_Zoroastrian	0.057	0.0010	57.17
Irish	0.057	0.0010	55.18
Mordovian	0.057	0.0011	52.60
Norwegian	0.057	0.0011	52.61
Polish	0.057	0.0010	55.46
Russian	0.057	0.0010	56.69
Shetlandic	0.057	0.0011	52.83
Luo	0.057	0.0011	52.07
Norway_Southeast	0.058	0.0011	48.44
Norway_West	0.058	0.0012	46.51
Orcadian	0.058	0.0011	56.17
Sorb	0.058	0.0011	50.96
Basque	0.059	0.0010	58.33
Chuvash	0.059	0.0011	54.67
Icelandic	0.059	0.0011	53.44
Turkish_Balikesir	0.059	0.0014	43.24
BantuKenya	0.060	0.0012	50.32
Finnish	0.060	0.0011	56.97
Uygur	0.060	0.0011	55.18
Bengali	0.061	0.0012	54.30
Estonian	0.061	0.0011	57.99
Punjabi	0.061	0.0011	55.32
Lithuanian	0.062	0.0011	56.10
Hazara	0.063	0.0010	61.71
Gambian	0.064	0.0012	53.64
Lodhi	0.064	0.0011	59.79
Vishwabrahmin	0.066	0.0011	61.19
Yoruba	0.067	0.0010	66.09
Mandenka	0.067	0.0011	62.63
Esan	0.068	0.0012	58.05
Mala	0.068	0.0011	61.31
Mende	0.068	0.0012	61.51
Mbukushu	0.069	0.0012	55.20
Wambo	0.070	0.0012	56.88
BantuSA	0.072	0.0013	53.83
Aleut	0.073	0.0013	58.49
Shua	0.073	0.0012	64.59
Tswana	0.074	0.0012	58.28
Kyrgyz	0.075	0.0011	63.74

Kalash	0.077	0.0011	71.26
Khwe	0.077	0.0012	65.44
Komi_Zyrian	0.077	0.0012	65.49
Damara	0.078	0.0011	70.66
Himba	0.079	0.0015	52.16
Even	0.082	0.0013	69.14
Mansi	0.082	0.0012	65.62
Haiom	0.083	0.0012	67.23
Altaian	0.084	0.0014	61.08
Tubalar	0.085	0.0011	76.44
Nama	0.086	0.0012	75.74
Kharia	0.087	0.0013	70.93
Kgalagadi	0.089	0.0014	65.27
Hadza	0.090	0.0013	67.90
Kalmyk	0.091	0.0013	72.70
Khanty	0.091	0.0013	69.74
Selkup	0.091	0.0013	72.03
Biaka	0.097	0.0012	83.76
Tuvinian	0.098	0.0013	76.16
Ojibwa	0.101	0.0012	83.21
Mon	0.102	0.0014	74.63
Cambodian	0.103	0.0014	71.03
Thai	0.103	0.0013	78.84
Tu	0.103	0.0013	81.45
Yukagir	0.103	0.0012	84.79
Cree	0.104	0.0013	80.18
Gana	0.104	0.0013	77.71
Khomani	0.104	0.0013	78.55
Kusunda	0.105	0.0014	76.54
Mongola	0.106	0.0015	73.47
Tshwa	0.106	0.0016	64.81
Xibo	0.109	0.0014	76.46
Daur	0.111	0.0015	75.84
Yakut	0.111	0.0012	89.02
Khmer	0.112	0.0015	76.44
Lao	0.112	0.0014	80.83
RapaNui	0.112	0.0020	56.90
Kinh	0.114	0.0014	81.85
Hezhen	0.115	0.0014	84.20
Xuun	0.115	0.0013	89.47
Yi	0.115	0.0014	83.48
Han	0.116	0.0013	89.73

Naxi	0.116	0.0014	83.63
Hakka_Taiwan	0.117	0.0014	83.95
Nyah_Kur	0.117	0.0014	82.41
Tujia	0.117	0.0014	85.58
Han_Taiwan	0.118	0.0014	83.44
Korean	0.118	0.0013	90.13
Tai_Lue	0.118	0.0014	84.28
Borneo	0.119	0.0015	81.79
Dai	0.119	0.0014	83.18
Japanese	0.119	0.0013	91.97
Micronesian	0.119	0.0014	84.05
Miao	0.120	0.0013	85.52
Kuy_Suay	0.121	0.0014	87.45
Lawa	0.121	0.0014	84.22
Samoan	0.122	0.0015	78.53
Makira	0.125	0.0016	79.02
Ulchi	0.126	0.0013	93.28
Karen_Sgaw	0.127	0.0015	85.61
Lahu	0.127	0.0015	86.31
Nggela	0.127	0.0016	77.22
Semende	0.127	0.0015	87.45
Malaita	0.128	0.0015	83.32
Savo	0.128	0.0016	78.54
Mbuti	0.128	0.0014	93.02
Santa_Isabel	0.129	0.0016	82.09
Chipewyan	0.131	0.0014	97.42
Buka	0.132	0.0016	81.45
Hoan	0.134	0.0017	79.92
Ranongga	0.134	0.0018	74.22
Ontong_Java	0.135	0.0016	85.58
Vella_Lavella	0.135	0.0017	80.82
Kolombangara	0.136	0.0017	78.87
Tongan	0.136	0.0016	85.70
Saposa	0.137	0.0016	84.92
Taa_East	0.137	0.0015	87.98
Algonquin	0.138	0.0017	79.27
Ami	0.138	0.0017	79.70
Mamanwa	0.138	0.0015	88.55
Manus	0.138	0.0025	56.33
Chukchi	0.139	0.0014	97.12
Mussau	0.139	0.0017	83.69
Naro	0.139	0.0015	93.54

Choiseul	0.140	0.0017	82.34
Eskimo_ChaplinSireniki	0.140	0.0016	87.19
Notsi	0.140	0.0016	85.90
Hmong	0.141	0.0015	92.72
Gui	0.143	0.0017	85.49
Htin_Mal	0.143	0.0016	89.95
Mangseng	0.143	0.0018	78.05
Nailik	0.143	0.0017	84.64
Madak	0.144	0.0018	81.91
Taa_North	0.144	0.0015	92.43
Taa_West	0.144	0.0014	101.36
Mayan	0.145	0.0016	90.89
Tigak	0.145	0.0016	90.81
Koryak	0.147	0.0016	94.41
Kuot_Lamalaua	0.147	0.0018	82.35
Tolai	0.147	0.0016	93.92
Teop	0.148	0.0017	87.63
Bolivian	0.149	0.0019	80.65
Lavongai	0.149	0.0016	90.03
Melamela	0.149	0.0016	90.60
Tikopia	0.149	0.0018	84.51
Ju_hoan_North	0.150	0.0014	108.91
Ju_hoan_South	0.150	0.0016	95.99
Nganasan	0.150	0.0015	101.27
Quechua	0.150	0.0018	82.32
Itelmen	0.151	0.0016	94.29
Mengen	0.151	0.0017	88.73
Kove	0.152	0.0016	93.52
Kuot_Kabil	0.152	0.0017	87.85
Nakanai_Bileki	0.152	0.0017	87.51
Eskimo_Naukan	0.154	0.0016	97.74
Sulka	0.155	0.0016	95.31
Zapotec	0.155	0.0017	89.58
Atayal	0.157	0.0016	97.36
Santa_Cruz	0.157	0.0018	87.61
Mixtec	0.159	0.0017	92.27
Kaqchikel	0.161	0.0019	87.26
Nasioi	0.162	0.0018	90.20
Mamusi_Paleabu	0.166	0.0019	87.43
Aymara	0.169	0.0019	87.75
Guarani	0.169	0.0019	88.79
Onge	0.173	0.0017	103.50

Rennell_and_Bellona	0.173	0.0019	89.22
Ata	0.174	0.0012	89.30
Nakanai_Loso	0.175	0.0019	91.22
Mamusi	0.180	0.0018	101.24
Mixe	0.180	0.0018	101.35
Pima	0.183	0.0017	106.17
New_Guinea	0.186	0.0019	100.10
Papuan	0.188	0.0018	102.85
Gimi	0.196	0.0019	101.25
Xavante	0.204	0.0019	108.92
Karitiana	0.222	0.0020	112.76
Mlabri	0.227	0.0023	98.81
Surui	0.232	0.0020	113.52
Cabecar	0.234	0.0022	105.07

SUPPLEMENT 7: ADMIXTURE ANALYSIS

The following table and figures present results of ADMIXTURE analysis for values of K=2–14 and plots for values of K=2–12. Table S7.1 presents the CV errors for 10 replicates of each value of K ranging from K=2 to K=14 as well as the average CV error for each value of K. Figure S7.1 plots the average CV error for each value of K from K=2–14, illustrating that the lowest CV score is at K=11 and that CV scores plateau beginning at K=8.

Figure S7.2 presents ADMIXTURE plots for values of K=2–12 using a reference panel of 1,897 present-day individuals from 146 world-wide populations and 28 ancient individuals from Kulubnarti. Figure S7.3 presents focused ADMIXTURE results for K=2–12 of the 28 ancient Kulubnarti Nubian individuals from the R and S communities.

Figure S7.2 shows that an African component (teal) was distinguished at K=2, followed by an Eastern Asian component (purple) at K=3. At K=4, another African component (light green) was introduced, maximized in Southern African populations and present in most populations from sub-Saharan Africa. At K=5, a component (magenta) maximized in Central South Asians, but also found in Caucasian, Near Eastern, and some Northern African populations, was introduced. At K=6, a geographically broad component (yellow) was introduced, maximized in European and Western Eurasian populations, and found at a substantial frequency in most non-African populations.

Higher values of K worked out more regional structures in various world regions. An East African component (blue) was introduced at K=7, maximized in the Hadza from Tanzania. At K=8, an additional Central/Southern Asian component (dark green) was introduced that was maximized in the Onge from the Andaman Islands and separated

them from other Central South Asian groups. K=9 introduced another sweeping component (red) maximized in Europeans, but also present in most non-sub-Saharan African populations. At K=10, a component (lavender) maximized in East African populations outside of the Hadza and also present in North African populations was introduced. At K=11, a Central African component (orange) was introduced, maximized in the Mbuti. At K=12, a northern Russian component (navy blue) maximized in the Dolgan and Khanty is introduced. K=13 and K=14 were not plotted due to their association with higher CV errors and the determination that higher values of K would only continue to work out increasingly regional variation in focused world regions.

Table S7.1: CV errors for 10 iterations each of K=2–14 and average CV error for each value of K.

Run	K=2	K=3	K=4	K=5	K=6	K=7	K=8	K=9	K=10	K=11	K=12	K=13	K=14
1	0.35909	0.35355	0.34849	0.34783	0.34623	0.34560	0.34542	0.34494	0.34470	0.34457	0.34450	0.34497	0.34483
2	0.35912	0.35355	0.34849	0.34722	0.34656	0.34575	0.34512	0.34492	0.34469	0.34451	0.34448	0.34469	0.34484
3	0.35911	0.35354	0.34849	0.34723	0.34623	0.34577	0.34509	0.34487	0.34474	0.34455	0.34453	0.34498	0.34486
4	0.35910	0.35356	0.34849	0.34722	0.34656	0.34558	0.34510	0.34487	0.34479	0.34465	0.34486	0.34470	0.34489
5	0.35911	0.35355	0.34849	0.34722	0.34623	0.34561	0.34508	0.34490	0.34469	0.34465	0.34483	0.34558	0.34479
6	0.35910	0.35355	0.34850	0.34722	0.34622	0.34576	0.34509	0.34494	0.34480	0.34468	0.34485	0.34465	0.34688
7	0.35911	0.35355	0.34850	0.34723	0.34656	0.34558	0.34514	0.34495	0.34469	0.34456	0.34543	0.34494	0.34484
8	0.35910	0.35357	0.34848	0.34722	0.34659	0.34556	0.34509	0.34486	0.34468	0.34457	0.34490	0.34468	0.34561
9	0.35910	0.35403	0.34848	0.34723	0.34658	0.34560	0.34545	0.34497	0.34471	0.34457	0.34472	0.34470	0.34518
10	0.35911	0.35356	0.34850	0.34723	0.34623	0.34560	0.34509	0.34499	0.34469	0.34454	0.34451	0.34468	0.34482
Avg.	**0.35911**	**0.35360**	**0.34849**	**0.34729**	**0.34640**	**0.34564**	**0.34517**	**0.34492**	**0.34472**	**0.34459**	**0.34476**	**0.34486**	**0.34515**

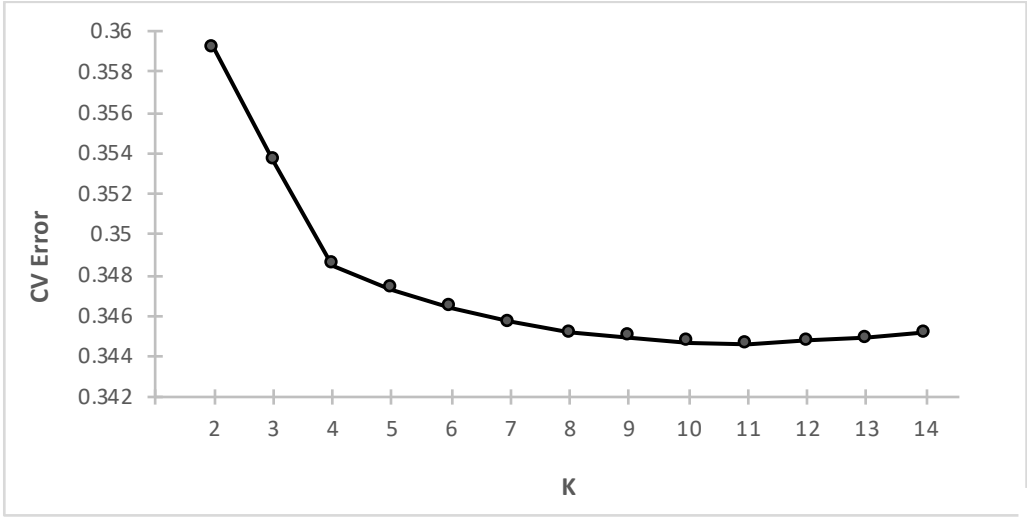


Figure S7.1: Line plot of average CV error for each value of K between K=2 and K=14.

K=9
K=10
K=11
K=

K=4
K=5
K=6
K=7

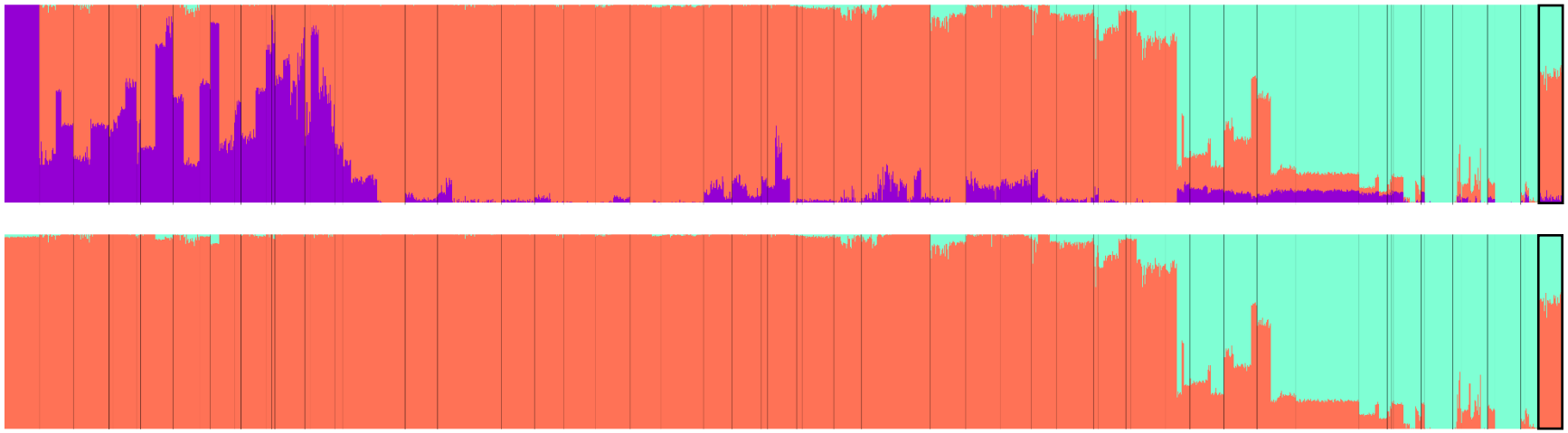


Figure S7.2: ADMIXTURE plots representing ancestry proportions of 1,897 present-day individuals and 28 Kulubnarti Nubians for 2–12 ancestral clusters (K=2–12). Present-day individuals are hierarchically clustered by population and geographic region as follows: Han, Balochi, Bengali, Brahmin Tiwari, Brahui, Burusho, Gujarati A, Gujarati B, Gujarati C, Gujarati D, Hazara, Jew Cochin, Kalash, Kharia, Kusunda, Lodhi, Makrani, Mala, Onge, Pathan, Punjabi, Sindhi Pakistan, Vishwabrahmin, Altaian, Dolgan, Khanty, Kyrgyz, Mansi, Selkup, Turkmen, Tuvinian, Uygur, Uzbek, Chuvash, Komi Zyrian, Mordovian, Russian, Albanian, Basque, Belarusian, Bulgarian, Croatian, Czech, Estonian, Finnish, French, Frisian, German, Greek, Hungarian, Italian Central, Italian North, Italian South, Lithuanian, Norwegian, Sardinian, Sicilian, Spanish, Abkhasian, Adygei, Armenian, Balkar, Chechen, Georgian, Jew Georgian, Kumyk, Lezgin, Nogai, Ossetian, Cypriot, Jew Ashkenazi, Druze, Jew Turkish, Jordanian, Lebanese, Lebanese Christian, Lebanese Muslim, Syrian, Turkish, Turkish Balikesir, Assyrian, Bedouin A, Bedouin B, Iran Non-Zoroastrian Fars, Iran Zoroastrian, Iranian, Iranian Bandari, Jew Iranian, Jew Iraqi, Jew Yemenite, Palestinian, Saudi, Yemeni, Algerian, Egyptian, Jew Libyan, Jew Moroccan, Jew Tunisian, Libyan, Moroccan, Mozabite, Saharawi, Tunisian, Bantu Kenya, Datog, Dinka, Hadza, Hadza1, Kikuyu, Luhya, Luo, Masai, Sandawe, Jew Ethiopian, Somali, Oromo, Esan, Gambian, Mandenka, Mende, Yoruba, Biaka, Mbukushu, Mbuti, Bantu South Africa, Bantu South Africa Herero, Bantu South Africa Ovambo, Damara, Gana, Gui, Haiom, Himba, Hoan, Ju/'Hoan North, Ju/'Hoan South, Khomani, Khwe, Nama, Naro, Shua, Taa East, Taa North, Taa West, Tshwa, Tswana, Xuun. Kulubnarti Nubian samples are outlined in a black box at the right end of each plot; the R community samples precede the S community samples.

K=9 K=10 K=11 K=12

K=5 K=6 K=7

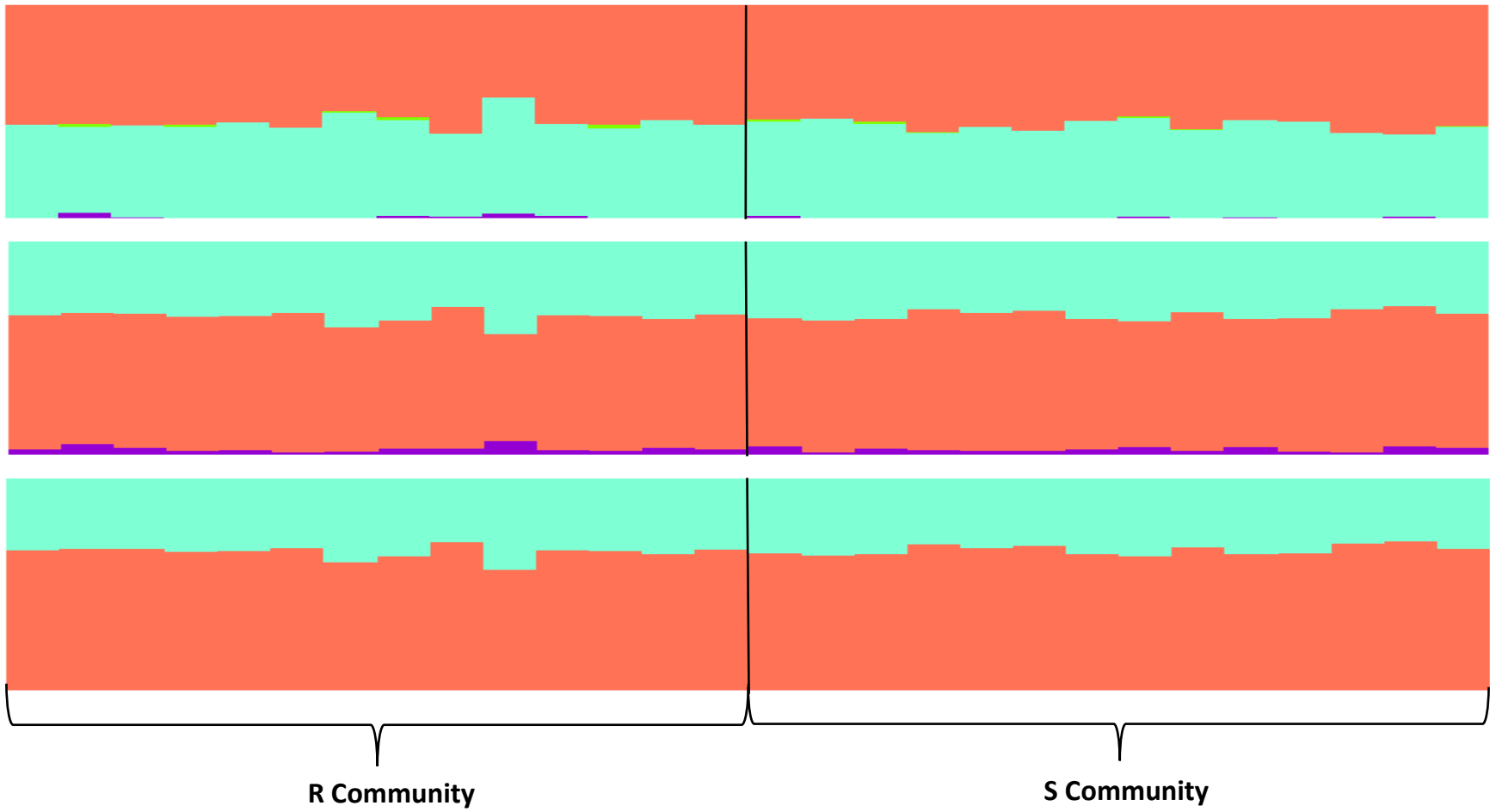


Figure S7.3: ADMIXTURE plots showing ancestry proportions of 28 Kulubnarti Nubian individuals for 2-12 ancestral clusters (K=2–12). The R and S communities are labeled below the K=2 plot and are separated on each ADMIXTURE plot by a solid black line, with R community individuals to the left of the line, and S community individuals to the right of the line.

SUPPLEMENT 8: MITOCHONDRIAL HAPLOGROUPS

Table S8.1 presents the mtDNA haplogroup assessment for each Kulubnarti Nubian sample. The coverage of the mitogenome (mtDNA coverage) after capture is reported, followed by the haplogroup call made using the Phy-Mer software and the associated haplogroup prediction score and list of haplogroup-defining positions as called by Phy-Mer. The HaploGrep2 haplogroup call is reported, along with an estimate of the sample's match to the RSRS consensus sequence (Behar et al. 2012) and the 95% confidence interval. Yellow highlighting represents cases in which the Phy-Mer and HaploGrep2 haplogroup calls were not in agreement. A brief description of the geographic origin and/or present geographic distribution of each haplogroup assigned to an individual or individuals at Kulubnarti is provided below the table.

Table S8.1: mtDNA haplogroup calls for the 28 Kulubnarti Nubian samples determined by Phy-Mer and HaploGrep2.

Sample ID	mtDNA Coverage	Phy-Mer Haplo-group Call	Phy-Mer Haplogroup Prediction Score	Defining Positions	HaploGrep2 Haplogroup Call	Match to Consensus [95% C.I.]
R101	79x	H2a	0.95	'263', '750', '8860', '15326'	H2a	0.998 [0.996-0.999]
R124	456x	K1a19	0.98	'73', '263', '497', '750', '1189', '1438', '1811', '2706', '3480', '4769', '7028', '8860', '9055', '9698', '10398', '10550', '11299', '11467','11719', '12308', '12338','12372', '14167', '14766','14798', '15326', '16093','16224', '16311'	K1a19	1 [0.998-1.0]
R182	89x	U5b2b	0.93	'73', '150', '263', '750', '1438', '1721', '2706', '3197', '4769', '7028', '7768', '8860', '9477', '11467', '11653','11719', '12308', '12372','12634',	U5b2b	0.993 [0.989-0.996]

				'13617', '13630','13637', '14182', '14766','15326', '16270'		
R186	296x	L2a1	0.98	'73', '143', '146', '152', '195', '263', '750', '769', '1018', '1438', '2416', '2706', '2789', '3594', '4104', '4769', '7028', '7175', '7256', '7274', '7521', '7771', '8206', '8701', '8860', '9221', '9540', '10115', '10398', '10873', '11719','11914', '11944', '12693','12705', '13590', '13650','13803', '14566', '14766','15301', '15326', '15784','16189', '16192', '16223','16278', '16294', '16309', '16390'	L2a1	1 [0.999,1.000]
R59	133x	H2a	0.97	'263', '750', '8860', '15326'	H2a	1 [0.998,1.000]
R152	1200x	L0a1a	0.94	'64', '93', '95C', '152', '185', '189', '236', '247', '263', '750', '769','825A', '1018', '1048', '1438', '2245', '2706', '2758', '2885', '3516A', '3594', '3866', '4104', '4312', '4586', '4769', '5096', '5231', '5442', '5460', '5603', '6185', '7028', '7146', '7256', '7521', '8428', '8468', '8566', '8655', '8701', '8860', '9042', '9347', '9540', '9755', '9818', '10398', '10589', '10664','10688', '10810', '10873','10915', '11176', '11641','11719', '11914', '12007','12705', '12720', '13105','13276', '13506', '13650','14308', '14766', '15136','15326', '15431', '16129','16148', '16168', '16172','16187', '16188G', '16189', '16223', '16230','16311', '16320'	L0a1a1	1 [0.999,1.000]
R181	1230x	H2a	0.94	'263', '750', '8860', '15326'	H2a	1 [0.998,1.000]
R202	526x	H2a	0.97	'263', '750', '8860', '15326'	H2a	1 [0.996,1.000]
R196	498x	L2a1d1	0.97	'73', '143', '146', '152', '182', '195', '263', '750', '769', '1018', '1438', '2416', '2706', '2789',	L2a1d	1 [0.999,1.000]

				'3594', '4104', '4769', '5196', '7028', '7175', '7256', '7274', '7521', '7771', '8206', '8701', '8860', '9221', '9530', '9540', '10115', '10398', '10873', '11386','11719', '11914', '11944','12007', '12612', '12693','12705', '13395', '13590','13650', '13803', '13934','14566', '14766', '15301','15326', '15784', '16209','16223', '16278', '16294','16301', '16354', '16390'		
R201	382x	J2a2	0.98	'73', '150', '195', '263', '295', '489', '750', '1438', '2706', '4216', '4769', '6671', '7028', '7476', '8860', '10398', '10499', '11002', '11251', '11377', '11719', '12570', '12612', '13708', '14766', '15257', '15326', '15452A', '15679', '16069', '16126'	J2a2	1 [0.999,1.000]
R5	701x	N1b1a2	0.97	'73', '152', '263', '750', '1438', '1598', '1703', '1719', '2639', '2706', '3921A', '4769', '4904', '4960', '5471', '7028', '8251', '8472', '8836', '8860', '9335', '10238', '11362', '11719','12501', '12705', '12822','14766', '15326', '16145', '16176G', '16223', '16390'	N1b1b	0.989 [0.981,0.995]
R79	305x	U1a1	0.97	'73', '263', '285', '750', '1438', '2218', '2706', '4769', '4991', '6026', '7028', '7581', '8860', '11467', '11719','12308', '12372', '12879','13104', '14070', '14364','14766', '15148', '15326', '15954C', '16189', '16249'	U1a	1 [0.998,1.000]
R93	1490	H2a	0.93	'263', '750', '8860', '15326'	H2a	1 [0.998,1.000]
R169	724	U5b2b	0.96	'73', '150', '263', '750', '1438', '1721', '2706', '3197', '4769', '7028', '7768', '8860', '9477', '11467', '11653','11719',	U5b2b	1 [0.998,1.000]

				'12308', '12372','12634', '13617', '13630','13637', '14182', '14766','15326', '16270'		
S33	420	L0a1a	0.97	'64', '93', '95C', '152', '185', '189', '236', '247', '263', '750', '769', '825A', '1018', '1048', '1438', '2245', '2706', '2758', '2885', '3516A', '3594', '3866', '4104', '4312', '4586', '4769', '5096', '5231', '5442', '5460', '5603', '6185', '7028', '7146', '7256', '7521', '8428', '8468', '8566', '8655', '8701', '8860', '9042', '9347', '9540', '9755', '9818', '10398', '10589','10664', '10688', '10810','10873', '10915', '11176','11641', '11719', '11914','12007', '12705', '12720','13105', '13276', '13506','13650', '14308', '14766','15136', '15326', '15431','16129', '16148', '16168','16172', '16187', '16188G', '16189', '16223','16230', '16311', '16320'	L0a1a1	1 [0.999,1.000]
S50	920	L0a1a	0.96	'64', '93', '95C', '152', '185', '189', '236', '247', '263', '750', '769', '825A', '1018', '1048', '1438', '2245', '2706', '2758', '2885', '3516A', '3594', '3866', '4104', '4312', '4586', '4769', '5096', '5231', '5442', '5460', '5603', '6185', '7028', '7146', '7256', '7521', '8428', '8468', '8566', '8655', '8701', '8860', '9042', '9347', '9540', '9755', '9818', '10398', '10589','10664', '10688', '10810','10873', '10915', '11176','11641', '11719', '11914','12007', '12705', '12720','13105', '13276', '13506','13650', '14308', '14766','15136', '15326', '15431','16129', '16148', '16168','16172', '16187', '16188G',	L0a1a1	0.997 [0.995,0.998]

				'16189', '16223','16230', '16311', '16320'		
S68a	1040	L2a1d	0.96	'73', '143', '146', '152', '182', '195', '263', '750', '769', '1018', '1438', '2416', '2706', '2789', '3594', '4104', '4769', '5196', '7028', '7175', '7256', '7274', '7521', '7771', '8206', '8701', '8860', '9221', '9530', '9540', '10115', '10398', '10873', '11386','11719', '11914', '11944','12007', '12612', '12693','12705', '13395', '13590','13650', '13803', '13934','14566', '14766', '15301','15326', '15784', '16209','16223', '16278', '16294','16301', '16354', '16390'	L2a1d	1 [0.999,1.000]
S81	771	L2a1d1	0.96	73', '143', '146', '152', '182', '195', '263', '750', '769', '1018', '1438', '2416', '2706', '2789', '3594', '4104', '4769', '5196', '7028', '7175', '7256', '7274', '7521', '7771', '8206', '8701', '8860', '9221', '9530', '9540', '10115', '10398', '10873', '11386','11719', '11914', '11944','12007', '12612', '12693','12705', '13395', '13590','13650', '13803', '13934','14566', '14766', '15301','15326', '15784', '16209','16223', '16278', '16294','16301', '16354', '16390'	L2a1d	1 [0.999,1.000]
S149	1800	H2a	0.92	'263', '750', '8860', '15326'	H2a	1 [0.998,1.000]
S133	861	L2a1d1	0.97	'73', '143', '146', '152', '182', '195', '263', '750', '769', '1018', '1438', '2416', '2706', '2789', '3594', '4104', '4769', '5196', '7028', '7175', '7256', '7274', '7521', '7771', '8206', '8701', '8860', '9221', '9530', '9540', '10115', '10398', '10873', '11386','11719', '11914', '11944','12007', '12612', '12693','12705', '13395', '13590','13650',	L2a1d	1 [0.999,1.000]

				'13803', '13934','14566', '14766', '15301','15326', '15784', '16209','16223', '16278', '16294','16301', '16354', '16390'		
S73	577	L2a1d1	0.97	'73', '143', '146', '152', '182', '195', '263', '750', '769', '1018', '1438', '2416', '2706', '2789', '3594', '4104', '4769', '5196', '7028', '7175', '7256', '7274', '7521', '7771', '8206', '8701', '8860', '9221', '9530', '9540', '10115', '10398', '10873', '11386','11719', '11914', '11944','12007', '12612', '12693','12705', '13395', '13590','13650', '13803', '13934','14566', '14766', '15301','15326', '15784', '16209','16223', '16278', '16294','16301', '16354', '16390'	L2a1d	1 [95%CI: 0.999,1.000]
S87	257	L2a1d1	0.97	73', '143', '146', '152', '182', '195', '263', '750', '769', '1018', '1438', '2416', '2706', '2789', '3594', '4104', '4769', '5196', '7028', '7175', '7256', '7274', '7521', '7771', '8206', '8701', '8860', '9221', '9530', '9540', '10115', '10398', '10873', '11386','11719', '11914', '11944','12007', '12612', '12693','12705', '13395', '13590','13650', '13803', '13934','14566', '14766', '15301','15326', '15784', '16209','16223', '16278', '16294','16301', '16354', '16390'	L2a1d	0.998 [95%CI: 0.997,0.999]
S159	1070	L2a1d1	0.94	'73', '143', '146', '152', '182', '195', '263', '750', '769', '1018', '1438', '2416', '2706', '2789', '3594', '4104', '4769', '5196', '7028', '7175', '7256', '7274', '7521', '7771', '8206', '8701', '8860', '9221', '9530', '9540', '10115', '10398', '10873', '11386','11719', '11914', '11944','12007', '12612', '12693','12705',	L2a1d	0.993 [0.989,0.995]

				'13395', '13590','13650', '13803', '13934','14566', '14766', '15301','15326', '15784', '16209','16223', '16278', '16294','16301', '16354', '16390'		
S17	454	T1a7	0.98	'73', '263', '512', '709', '750', '1438', '1888', '2706', '4216', '4769', '4917', '7028', '8697', '8860', '10463', '11251', '11719', '12633A', '13368', '14500','14766', '14905', '15326', '15452A', '15607', '15928', '16126','16163', '16186','16189', '16274', '16294'	T1a	1 [0.999,1.000]
S198	2	L1b1a2	0.98	'73', '152', '182', '185T', '195', '247', '263', '357', '709', '710', '750', '769', '825A', '1018', '1738', '2352', '2706', '2758', '2768', '2885', '3308', '3594', '3666', '3693', '4104', '4769', '5036', '5046', '5393', '5655', '6548', '6827', '6989', '7028', '7055', '7146', '7256', '7389', '7521', '7867', '7954', '8248', '8468', '8655', '8701', '8860', '9540', '10398', '10688', '10810','10873', '11719', '12519','12705', '13105', '13506','13650', '13789', '13880A', '14178', '14203','14560', '14766', '14769','15115', '15326', '16126','16187', '16189', '16223','16264', '16270', '16278','16293', '16311'	L1b1a2a	0.982 [0.646,0.997]
S27	822	U5b2b	0.97	'73', '150', '263', '750', '1438', '1721', '2706', '3197', '4769', '7028', '7768', '8860', '9477', '11467', '11653','11719', '12308', '12372','12634', '13617', '13630','13637', '14182', '14766','15326', '16270'	U5b2b	1 [0.999,1.000]
S79	763	L5a1b	0.96	'73', '152', '182', '195', '247', '263', '455.1TT', '459.1C', '709', '750', '769', '825A', '851',	L5a1b	1 [0.999,1.000]

				'930', '1018', '1438', '1822', '2706', '3423', '3594', '4104', '4496', '4769', '5111', '5147', '5656', '6182', '6297', '7028', '7256', '7424', '7521', '7972', '8155', '8188', '8582', '8655', '8701', '8754', '8860', '9305', '9329', '9540', '10398', '10688','10810', '10873', '11025','11719', '11881', '12236','12432', '12705', '12950','13105', '13506', '13650','13722', '14212', '14239','14581', '14668', '14766','14819', '14905', '14971','15217', '15326', '15884','16129', '16148', '16166','16187', '16189', '16223','16278', '16311', '16355', '16362'		
S89	358	H2a	0.98	'263', '750', '8860', '15326'	H2a	1 [0.998,1.000]

Geographic distribution of African-specific mtDNA haplogroups:

L0a1a

Haplogroup L0a is an ancient mtDNA lineage that originated in Eastern Africa and is presently found at a high frequency throughout Eastern, Central, and Southeastern Africa (Salas et al. 2002). The L0a1 branch is the most widely distributed L0 haplogroup in Africa and accounts for the majority of L0 lineages found in present-day populations (Salas et al. 2002; Hirbo 2011). L0a1 has specifically been detected in present-day Northeastern African populations shown to share affinity with the Kulubnarti Nubians, including Ethiopians and Ethiopian Jews (Kivisild et al. 2004; Hirbo 2011), and has also been detected at an appreciable frequency in present-day Nubian populations (Hirbo 2011).

L2a1d

Haplogroup L2 is one of the most prevalent haplogroups in Africa today, most likely originating in Western Africa but spreading throughout the continent as both the result of the Bantu expansion and trans-Saharan slave trade as well as earlier migrations that may have brought the haplogroup to Eastern Africa (Pereira et al. 2001; Salas et al. 2002; Silva et al. 2015). L2a is the most dominant and widespread mtDNA cluster of L2 in Africa (Salas et al. 2002; Harich et al. 2010) and is presently found at rates up to 30% in Sudan and Ethiopia (da Silva 2014). L2a1 is the most complex branch of L2a (da Silva 2014; Silva et al. 2015), and though it has a Western African origin (Bekada et al. 2013), it has also been detected in East African populations (Kivisild et al. 2004) including Nubians (Hirbo 2011; Messina et al. 2017). The sub-lineage L2a1d1 has been found in individuals from Eastern Africa (Ethiopia, Somalia, and the Sudan) as well as Egypt (Behar et al. 2008; da Silva 2014; Silva et al. 2015; Messina et al. 2017) and may reflect historic contact between Eastern Africans and Egypt (da Silva 2014).

L5a1b

The L5 haplogroup is a small haplogroup predominantly observed in Eastern Africa and presently detected at its highest frequency among Nilo-Saharan-speaking populations (Hirbo 2011). The L5a lineage is found most commonly in Eastern and Southeastern Africa, though it is sparsely distributed (Pereira et al. 2001; Salas et al. 2002; Kivisild et al. 2004; Gonder et al. 2007; Cabrera et al. 2009; Batini et al. 2011; Silva et al. 2015). It has been detected in Egypt (Stevanovitch et al. 2004), Ethiopia (Kivisild et al. 2004), and Tanzania (Knight et al. 2003), and has been used as evidence of the expansion of Nilo-

Saharan populations into Egypt (Hirbo 2011). The sub-lineage L5a1 has been detected at particularly high frequencies among Ethiopian Jews (Kivisild et al. 2004; Non et al. 2011), has been observed among Cushitic speaking populations of northern Kenya and southern Ethiopia as well as Afroasiatic speaking populations from Ethiopia (Kivisild et al. 2004; Hirbo 2011), and has also been found in Saudi Arab populations with exact matches in Egypt, Ethiopia and Kenya (Abu-Amero et al. 2008).

L1b1a2

A less parsimonious haplogroup assignment at Kulubnarti is L1b1a2, which may have originated in Central/Western Africa (Salas et al. 2002). Presently, the L1b lineage is found primarily in Western, Central, and Northern Africa (Watson et al. 1997; Rando et al. 1998; Salas et al. 2002; Rosa et al. 2004; Harich et al. 2010; Hirbo 2011) and is nearly absent in Eastern and Southern Africa (Gonder et al. 2007). Despite this distribution, lineages L1b and L1b1a have been previously detected in Nubian populations (Kivisild et al. 2004; Hirbo 2011), and the specific L1b1a2a sub-lineage has been proposed to have originated in Eastern Africa based on its representation by multiple divergent sequences from Ethiopia (Behar et al. 2008; Cerezo et al. 2012; Messina et al. 2017), Egypt (Behar et al. 2008), and Somalia (Mikkelsen et al. 2012). It has been suggested that this lineage may have spread subsequently from Eastern to Northern Africa down the Nile.

Geographic distribution of non-African-specific mtDNA haplogroups:

H2a

Haplogroup H originated in the Near East and displays an extremely wide geographic distribution as well as a high frequency throughout most of its range (Achilli et al. 2004).

It encompasses >40% of mtDNA variation in Europe and 10–30% in the Near East, Caucasus, and Central Asia (Richards et al. 2000; Achilli et al. 2004; Quintana-Murci et al. 2004; Loogväli et al. 2004; Roostalu et al. 2007; Pereira et al. 2005; Brandstätter et al. 2006), making it the most frequent haplogroup throughout western Eurasia (Roostalu et al. 2007). It has additionally been found throughout Northern Africa, primarily in Northwestern Africa (Ennafaa et al. 2009; Fadhlaoui-Zid et al. 2011; Badro et al. 2013), but also in present-day and ancient Egyptians (Stevanovitch et al. 2004; Schuenemann et al. 2017). Sub-lineage H2a is widely spread between Eastern Europe and Central Asia (Loogväli et al. 2004) and is found at an appreciable frequency on the Arabian Peninsula (Ennafaa et al. 2009) and throughout the Near East (Loogväli et al. 2004). H2a has not been detected until now in Northeastern Africa, where it is found at the highest frequency of any Eurasian mtDNA haplogroup in the Kulubnarti Nubians.

J2a2

Haplogroup J originated and diversified in the Near East and is presently distributed homogenously throughout Europe, as well as the Near East, Caucasus, and North Africa (Pala et al. 2012; Neves da Nova Fernandes 2013). It has been found in Northeastern Africa in both ancient and present-day Egyptians (Stevanovitch et al. 2004; Schuenemann et al. 2017). The rare haplogroup J2 is predominantly Near Eastern in its distribution (Pala et al. 2012), but has also been detected in Northeastern Africans, including Ethiopians (Kivisild et al. 2004), while the sub-lineage J2a2 has been found in Northeastern Africa as well as in the Near East (Pala et al. 2012; Fernandes et al. 2015).

K1a19

Haplogroup K originated in Europe or the Near East and presently appears in West Eurasia, Northern Africa, Eastern Africa, and also in South Asia. It is more common in specific populations inhabiting these areas, including the Druze and Palestinians of the Near East as well as Ashkenazi Jews (Macaulay et al. 1999; Richards et al. 2000; Behar et al. 2004; Non 2010; Costa et al. 2013). The specific sub-lineage K1a19 has been found in Italian, Iranian, Armenian, Turkish, and Ashkenazi Jewish individuals (Behar et al. 2006, 2010, 2012).

N1b1

Haplogroup N1b is the western-Eurasian branch of the first non-African founder node, N (Abu-Amero et al. 2007, Fernandes et al. 2012). N1b is a rare lineage with patchy distributions throughout Eurasia, the Arabian Peninsula, and particularly Southwest Asia (Cabrera 2009; Fernandes et al. 2012), though it has also been detected in Northern (Stevanovitch et al. 2004) and Eastern Africans (Fadhalaoui-Zid et al. 2011; Mikkelsen et al. 2012). N1b has also been reported at high frequency in Ashkenazi Jews (Behar et al. 2004). The main sub-lineage of N1b, N1b1, has been found most frequently in the Near East, particularly on the Arabian Peninsula, and in Northern Africa (Saunier et al. 2009; Harich et al. 2010; Fadhalaoui-Zid et al. 2011; Fernandes et al. 2012; Neves da Nova Fernandes 2013).

T1a

Haplogroup T presently makes up almost 10% of mtDNA haplogroups in Europe and ~8% in the Near East (Pala et al. 2012). The T1 lineage appears to have originated in the Near East (Pala et al. 2012) and is distributed from Northwestern Africa throughout

Europe, the Caucasus, the Near East, and across Asia into Siberia, while the T1a sub-lineage that represents ~90% of T1 has a wide but patchy distribution, with peak frequencies in the South Caucasus, Iran, Tunisia, and Romania (Pala et al. 2012; Elkamel et al. 2018). T1a has been detected throughout Northern and Eastern Africa, including identical T1a sequences detected in Ethiopians (Kivisild et al. 2004) and Nubians (Krings et al. 1999). T1a7 has been found in ancient Egyptians prior to the Early Christian Period (Schuenemann et al. 2017) as well as in present-day Near Easterners and North African Egyptians (Elkamel et al. 2018).

U1a

Haplogroup U is most likely of Western Asian origin, with various subclades now distributed widely throughout Eurasia as well as in Northern and Eastern Africa (Hervella et al. 2016). The U1a sub-lineage originated in the Near East and spread from India through Europe; it is presently found at highest frequency in the Near East and Northern Africa (Macaulay et al. 1999; Richards et al. 2000). U1a has been detected in populations from the Arabian Peninsula (Abu-Amero et al. 2008), Yemenites (Kivisild et al. 2004; Shen et al. 2004), Palestinians (Shen et al. 2004), Berbers, and Tunisian Arabs (Fadhlaoui-Zid et al. 2004). Several individuals assigned to the U1a lineage were detected among ancient Egyptians from varying time periods (Molto et al. 2017; Schuenemann et al. 2017).

U5b2b

Haplogroup U5 has been found only at appreciable frequencies in Europe, where it is proposed to have arisen (Rando et al. 1998; Richards et al. 1998, 2000; Achilli et al.

2004). It is found at lower frequencies in the northern Caucasus, is rarer in the southern Caucasus, and is found at only trace frequencies on the Arabian Peninsula (Malyarchuk et al. 2010). The U5b lineage is primarily distributed throughout Europe, but has also been detected in Egyptian individuals from Alexandria (Saunier et al. 2009), at a high frequency in a Berber population from Egypt (Coudray et al. 2009), and in a Berber individual from Tunisia (Cherni et al. 2005), showing that it has penetrated into Northern Africa. Sub-haplogroup U5b2b requires further phylogeographic studies.

SUPPLEMENT 9: Y CHROMOSOME HAPLOGROUPS

Table S9.1 presents the Y chromosome haplogroup assessment for each male Kulubnarti Nubian. A brief description of the approximate geographic origin and/or present geographic distribution of each Y chromosome haplogroup assigned to a male individual or individuals at Kulubnarti is provided below the table.

Table S9.1: Y chromosome haplogroup calls for16 male Kulubnarti Nubian samples.

Sample ID	Y-chromosome Haplogroup	Approximate Geographic Association
R101	E1b1b1b2	Northwestern Africa, Northern Africa
R124	E1b1b1a1b2	Northeastern Africa, Eastern Africa, Western Eurasia, Europe
R59	E1b1b1a1a1b	Eastern Africa (Horn of Africa, Sudan)
R181	E1b1b1a	Northern Africa; Northeastern Africa
R196	E1b1b1a1a1	Northern Africa, especially Southern Egypt; Sudan
R201	T1a1a	Near East, Eastern Africa
R93	E1b1b1b2	Northwestern Africa, Northern Africa
R169	T1a1a	Near East, Eastern Africa
S33	R2a3a	Central/Southern Asia
S50	J2a1	Near East, Caucasus, Europe
S81	J1a	Near East, Caucasus, Europe
S73	J2a1	Near East, Caucasus, Europe
S159	E2	Western and Eastern sub-Saharan Africa
S198	G2a2b2a	Near East, Caucasus, Europe
S27	E1b1b1a1a1	North Africa, especially Southern Egypt; Sudan
S79	G2a2b2a1	Near East, Caucasus, Europe

Geographic distribution of African-specific Y chromosome haplogroups

Haplogroup E

Y chromosome haplogroup E most likely originated somewhere between the Red Sea and Lake Chad in the northern hemisphere of Africa (Gebremeskel and Ibrahim 2014). It is the most common Y chromosome clade within Africa (Cruciani et al. 2002; Chiaroni et al. 2009). Its various subclades are presently widely distributed in clinal patterns within Africa, the Near East, Europe, and into Western Asia likely reflecting a mosaic of small-scale, regional population movements, replacements and expansions (Semino et al. 2004; Trombetta et al. 2015). In comparison to other African-dominant Y chromosome haplogroups, such as A and B, haplogroup E is found at a much higher frequency in places such as Sudan (Hassan et al. 2008).

E2

E2 is one of the two deepest haplogroup E subclades (Karafet et al. 2008; Trombetta et al. 2015) and is much less frequent than the E1 clade (Jobling and Tyler-Smith 2017) and may be of Central African origin (isogg.org). It is largely restricted to sub-Saharan Africa and is virtually absent in European populations (Semino et al. 2004). This haplogroup has been detected in Khoisan and Niger-Congo speaking peoples (Wood et al. 2005), while its descendants have been detected widely, specifically in Sudanese and Ethiopian populations (Underhill et al. 2000).

E1b1b1a/E1b1b1b2/E1b1b1a1a1/E1b1b1a1b2/E1b1b1a1a1b

E1b1 is the most widespread lineage of Y chromosome haplogroup E. It has been observed at a particularly high frequency in Eastern and Northeastern Africans (Cruciani et al. 2004, 2007; Semino et al. 2004; Hassan et al. 2008; Caratti et al. 2009). One of the sub-lineages of E1b1, E1b1b, probably evolved in Northeast Africa or the Near East and

is found at high frequencies throughout Northern Africa, Eastern Africa, as well as on the Arabian Peninsula and in the Near East (Semino et al. 2004; Hassan et al. 2008; isogg.org). All lineages assigned to haplogroup E at Kulubnarti (with the exception of a single individual assigned to subclade E2), are assigned to various lineages of E1b1b1. Interestingly, the E1b1b1a1b2 lineage detected in a single individual at Kulubnarti was also detected in a Ptolemaic Period individual from Egypt (Schuenemann et al. 2017).

Geographic distribution of non-African-specific Y chromosome haplogroups

G2a2b2a/G2a2b2a1

Y chromosome haplogroup G is found in populations spanning the Caucasus and Near East to southern Europe; though widespread, it is found in most places outside of the Caucasus at low frequencies (Rootsi et al. 2012). The G2 subclade likely originated near Anatolia, Armenia, or Iran, and is largely restricted to southern Europe, the Near East, and the Caucasus; it is found commonly in the Caucasus and decreases in frequency further northward and eastward (Balanovsky et al. 2011; Rootsi et al. 2012). The vast majority of haplogroup G individuals belong to subclade G2a-related lineages, though these lineages are unevenly distributed throughout the Caucasus and Near East, likely the result of demographic complexities (Luis et al. 2004; Yunusbayev et al. 2011; Rootsi et al. 2012). The specific G2a2b2a has been detected at especially high frequencies amongst particular Caucasian groups as well as among Palestinians (Rootsi et al. 2012), Iranians (Grugni et al. 2012), and has also been detected among Arabs in Egypt and Oman (Luis et al. 2004).

J1a and J2a1

Y chromosome haplogroup J most likely arose in the Near East, where it has the highest frequency and diversity. It exhibits a decreasing clinal pattern from the Near East to Europe, North Africa, the Caucasus, Iran, Central Asia, and India (Cinnioğlu et al. 2004; Semino et al. 2004; Regueiro et al. 2006; Sengupta et al. 2006). Two sister subclades, J1 and J2 make up haplogroup J and are distributed differently within the Near East, North Africa, and Europe (Semino et al. 2004). J1 lineages have been found at high frequencies in Nubian, Egyptian, Arab, and Near Eastern populations, and is thought to have spread most recently in association with the diffusion of Arab peoples (Semino et al. 2004; Hassan et al. 2008). J2 is noted to be very frequent in Turkey and the Levant (Hassan et al. 2008), as well as into Central Asia (Di Cristofaro et al. 2013), and may reflect the spread of Anatolian farmers (Semino et al. 2004). J2a is a main subclade of J2, and its various subclades are found primarily throughout the Near East and the Caucasus (Nasidze et al. 2005; Hassan et al. 2008).

R2a3a (now R2a2b1)

As one of the most widespread Y chromosome haplogroups (Chiaroni et al. 2009), Y chromosome haplogroup R spans Europe and the Caucasus, as well as West, Central, and South Asia (Sengupta et al. 2006). R2 is one of the two primary descendants of haplogroup R. It is primarily concentrated in West and Central Asia and appears at low levels in the Caucasus, Near East (including Iran, Turkey, and Palestine), and Europe. Its complex and multimodal distribution indicates demographic complexity that is inconsistent with a single recent history (Sengupta et al. 2006; Myres et al. 2010; Yunusbayev et al. 2011). Haplogroup R2a is a main branch of R2 and has been detected in populations spanning from the Near East and Caucasus into Central Asia, where it

appears to reach its highest frequencies (Wells et al. 2001; Sengupta et al. 2006; Bekada et al. 2013; Di Cristofaro et al. 2013).

T1a1a

Y chromosome haplogroup T originated in the Near East (Nogueiro et al. 2010) and is presently found at variable frequencies across West Asia, Northern and Eastern Africa, and Europe (Underhill et al. 2001; Luis et al. 2004; Chiaroni et al. 2009; Mendez et al. 2011). Like haplogroup R, it has a multimodal distribution (Chiaroni et al. 2009), with maximal frequencies observed in Eastern Africa (particularly in the Horn of Africa) and in the Near East (especially Iraq, Iran, Oman, Palestine, and Lebanon). T1a is the largest subclade of haplogroup T, and is seen in relatively high frequencies in Egypt, Tanzania, Ethiopia, and Morocco (Luis et al. 2004). The internal branches of the T haplogroup require more in-depth exploration to improve the resolution of their phylogeography.

SUPPLEMENT 10: RELATEDNESS ANALYSIS

Prior to estimation of F_{ST}, it is important to remove one individual from any pair ("dyad") of related individuals ("kin"), as consanguinity may increase F_{ST} estimates (Reich et al. 2009). While autosomal STRs are commonly used to study kinship in archaeological populations (Haak et al. 2008; Lacan et al. 2011; Baca et al. 2012), their utility is frequently hampered by the fragmented nature of ancient DNA (Dudar et al. 2003; Haak et al. 2008; Deguilloux et al. 2014). Therefore, this dissertation explores kinship at Kulubnarti through the analysis of Single Nucleotide Polymorphisms (SNPs) shared between two individuals ("intersecting SNPs") using a novel computational technique.

This technique was initially developed to assess relatedness between an ancient individual and potential living descendants using low-coverage SNP data (Fernandes et al. 2017) Similar to other relatedness estimation methods implemented in popular software such as PLINK (Chang et al. 2015) and NGSRelate (Korneliussen and Moltke 2015), this technique (referred to as the "forced homozygote approach") uses intersecting SNPs between two individuals to estimate their relatedness through the determination of a relatedness coefficient (here denoted "Rxy" as in Fernandes et al. 2017).

However, unlike other relatedness estimation methods, the forced homozygote approach is unique in its treatment of all SNP data from ancient samples as homozygous, accounting for the fact that most SNP loci in aDNA studies are represented by only one base call (i.e., 1x read coverage) making it impossible to identify sites that may be heterozygous (i.e., carry two different alleles). As such, this method preemptively "forces" a homozygote structure on each genotyped SNP locus by repeating the single

base call to generate a homozygous diploid locus prior to estimating the relatedness coefficient for a dyad of ancient individuals; if there is >1x coverage at any SNP locus, one base call of Q>30 is randomly selected and repeated. Consequently, only one half of the genome is analyzed with this approach; therefore, the relatedness coefficient calculated for each dyad using forced homozygote SNP data must be doubled prior to interpreting the degree of relatedness between the two individuals.

Where Rxy determined using other relatedness estimation methods ranges from 0 (unrelated individuals) to 1 (identical twins), with first-degree relatives (parents/children and full siblings) estimated at Rxy = ~0.50, and second-degree relatives (grandparents/grandchildren, half-siblings, uncles and aunts/nephews and nieces, and double cousins) estimated at Rxy = ~0.25, the forced homozygote approach produces Rxy estimations that are half of what is expected. For example, an Rxy estimation of ~0.25 using the forced homozygote approach would be equivalent to an Rxy estimation of ~0.50 if heterozygous data were used, and would therefore be consistent with first-degree relatives (Fernandes et al. 2017).

Following the calculation of a relatedness coefficient, the forced homozygote approach assigns posterior probabilities associated with membership in each of three relatedness "classes" (unrelated individuals (Rxy= ~0.0), second-degree relatives, (Rxy= ~0.125), and first-degree relatives (Rxy= ~0.25)). By their nature, posterior probabilities can only be assigned after a background of relevant evidence is taken into account; using the forced homozygote approach, the background of relevant evidence is produced by the *in silico* generation of 200 dyads each of virtual unrelated individuals, second-degree relatives, and first-degree relatives, generated using genotype frequency information from

~100 individuals who are likely to share recent ancestry or high amounts of genetic drift with the population under investigation. Further insight into the nature of specific relationships is then obtained through consideration of demographic information, specifically age and sex, of the individuals comprising the dyad.

To estimate kinship in dyads of Kulubnarti Nubians, genome-wide SNP data from 102 present-day individuals genotyped on the Human Origins array comprising 8 populations from throughout Africa, the Caucasus, and the Near East (selected because they were believed to share some biogeographic genetic affinity with the Kulubnarti Nubians) were extracted from the 1240K_HO dataset and used to create a genotype frequency file using PLINK v1.9b3.41 (Chang et al. 2015; www.cog-genomics.org/plink/1.9/) and the command --freq. [154] Prior to its use in the estimation of relatedness coefficients between dyads of Kulubnarti Nubians, the SNP frequencies in this file were used for the *in silico* generation of simulated datasets of dyads of known relationships to provide the background evidence for the assignment of posterior probabilities associated with membership of the Kulubnarti dyads in each of the three aforementioned relatedness classes.

The SNP frequencies in this genotype frequency file were used to generate virtual unrelated individuals *in silico* using R v3.3.1 and a custom script from Fernandes et al. (2017) (available on GitHub: https://github.com/danimag/tkrelated). These unrelated individuals were "forced" homozygous, and relatedness coefficients were estimated for

[154] As previously mentioned, the populations selected for the genotype frequency file should be populations known or expected to share common ancestry and/or high amounts of genetic drift with the population being analyzed. In most cases, the selected populations will be geographically-proximate populations. Alternatively, these populations can be selected by examining which groups fall closest on a PCA to the population of interest.

200 dyads of unrelated individuals using a symmetrical Rxy estimator algorithm developed by Queller and Goodnight (1989) incorporated in the software SPAGeDi1-5a build04-03-2015 (Hardy and Vekemans 2002) along with the correspondent African, Middle Eastern, and Caucasian genotype frequencies from the 102 individuals in the genotype frequency file. All relatedness coefficients were calculated based on "windows" of 15,000 shared SNPs analyzed simultaneously. To create "windows," shared SNPs between two individuals were selected at random until the desired number of SNPs (15,000) was reached. At this point, another "window" was created until no more windows of 15,000 shared SNPs were left. All leftover SNPs were placed into their own window; this window was only included in the analysis if it contained >1500 SNPs based on experimentation indicating that the inclusion of windows of <1500 SNPs skewed the estimation of the relatedness coefficient. This "window" approach created the normal curve required for posterior probability calculations when the relatedness coefficients for the 200 unrelated dyads were plotted; specifically, the Rxy of the majority of unrelated dyads was estimated to be 0.0 (the expected Rxy value), but there existed some variation (+/- 0.05) around this expected value.[155]

Unrelated individuals were then "mated" *in silico* to create virtual first- and second-degree relatives. The relatedness coefficients between 200 dyads each of first- and second-degree relatives were then calculated using the same "window" approach, Queller and Goodnight (1989) estimator algorithm, and correspondent genotype

[155] Because the Queller and Goodnight (1989) estimator was developed to analyze small amounts of genetic data, it became over-sensitive if >15,000 SNPs were analyzed simultaneously. This over-sensitivity resulted in the estimation of relatedness coefficients for the simulated dyads that were always exactly at the expected value of Rxy with little to no variation around this value, thereby precluding the creation of normal curves and preventing the calculation of posterior probabilities. However, by analyzing "windows" of SNPs, this over-sensitivity was mitigated.

frequencies. Again, the "window" approach created the normal curves required for posterior probability calculations, where Rxy was estimated to be 0.125 for the majority of second-degree relatives and 0.25 for the majority of first-degree relatives with some variation (+/- 0.05) around these expected values.[156]

Once relatedness coefficients were calculated for the 600 total simulated dyads of unrelated individuals, second-degree relatives, and first-degree relatives to create the background of relevant evidence for calculating posterior probabilities associated with the assignment of Kulubnarti Nubian dyads into each of the three relatedness classes, relatedness coefficients were estimated for 378 possible Nubian dyads using the same script from Fernandes et al. (2017). For each dyad, all intersecting SNPs were extracted and forced homozygous. A relatedness coefficient was then estimated using the same symmetrical Rxy estimator algorithm (Queller and Goodnight 1989) incorporated in the software SPAGeDi1-5a build04-03-2015 (Hardy and Vekemans 2002) along with the correspondent African, Near Eastern, and Caucasian genotype frequencies. As with the simulated datasets, "windows" of 15,000 SNPs were analyzed simultaneously. Posterior probabilities associated with membership in each of the three relatedness classes were then assigned for each dyad based on the relatedness coefficients calculated for the 600 total simulated dyads as background evidence.

From a set of 593,125 autosomal SNPs genotyped and with known allele frequencies, between 3,807 and 225,146 intersecting SNPs were analyzed for each dyad from Kulubnarti, with a median of 52,908 intersecting SNPs. Table S10.1 presents the results of relatedness analysis for 378 dyads at Kulubnarti, including the number of

[156] As previously mentioned, the homozygote structure imposed upon the data necessarily impacts the Rxy estimation so that it is approximately half of the expected value if heterozygous data had been used.

intersecting SNPs, the homozygous relatedness coefficient, the doubled relatedness coefficient, posterior probabilities associated with assignment into each of the three relatedness classes, and the relatedness class as assigned by the author. Consistent with previous tables and figures, all sample IDs starting with "R" designate the individual to be a member of the R cemetery; sample IDs starting with "S" designate the individual to be a member of the S cemetery.

Table S10.1: Results from relatedness analysis for 378 dyads from Kulubnarti. Cells of dyads estimated as having a first- or second-degree relationship have a yellow background.

Dyad		Inter-secting SNPs	Homo-zygous Related-ness Coefficient	Doubled Related-ness Coefficient	Posterior Probability			Related-ness Class
Ind 1	Ind 2				Unrelated	First Deg.	Second Deg.	
R5	R59	10385	0.0308	0.0616	1.000	0.000	0.000	unrelated
R5	R79	31521	0.0157	0.0314	1.000	0.000	0.000	unrelated
R5	R93	43887	0.0172	0.0344	1.000	0.000	0.000	unrelated
R5	S17	12776	0.0292	0.0584	1.000	0.000	0.000	unrelated
R5	S27	69963	0.0155	0.0310	1.000	0.000	0.000	unrelated
R5	S33	33807	0.0334	0.0668	1.000	0.000	0.000	unrelated
R5	S50	67525	0.0285	0.0570	1.000	0.000	0.000	unrelated
R5	S68a	40365	0.0294	0.0588	1.000	0.000	0.000	unrelated
R5	S73	59826	0.0081	0.0162	1.000	0.000	0.000	unrelated
R5	S79	66500	0.0289	0.0578	1.000	0.000	0.000	unrelated
R5	S81	33697	0.0227	0.0454	1.000	0.000	0.000	unrelated
R5	S87	82722	0.0274	0.0548	1.000	0.000	0.000	unrelated
R5	S89	65584	0.0187	0.0374	1.000	0.000	0.000	unrelated
R5	S133	58732	0.0262	0.0524	1.000	0.000	0.000	unrelated
R5	S148	42086	0.0195	0.0390	1.000	0.000	0.000	unrelated
R5	S159	70337	0.0132	0.0264	1.000	0.000	0.000	unrelated
R5	S198	40692	0.0174	0.0348	1.000	0.000	0.000	unrelated
R59	R79	12585	0.0185	0.0370	1.000	0.000	0.000	unrelated
R59	R93	16910	0.0408	0.0816	1.000	0.000	0.000	unrelated
R59	S17	5052	0.0443	0.0886	1.000	0.000	0.000	unrelated
R59	S27	27595	0.0154	0.0308	1.000	0.000	0.000	unrelated
R59	S33	13895	0.0373	0.0746	1.000	0.000	0.000	unrelated
R59	S50	27265	0.0322	0.0644	1.000	0.000	0.000	unrelated

R59	S68a	16552	0.0179	0.0358	1.000	0.000	0.000	unrelated
R59	S73	23491	0.0663	0.1326	0.023	0.977	0.000	uncertain
R59	S79	26266	0.0176	0.0352	1.000	0.000	0.000	unrelated
R59	S81	13758	0.0296	0.0592	1.000	0.000	0.000	unrelated
R59	S87	32614	0.0172	0.0344	1.000	0.000	0.000	unrelated
R59	S89	25902	0.0300	0.0600	1.000	0.000	0.000	unrelated
R59	S133	22987	0.0254	0.0508	1.000	0.000	0.000	unrelated
R59	S148	16988	0.0004	0.0008	1.000	0.000	0.000	unrelated
R59	S159	27527	0.0138	0.0276	1.000	0.000	0.000	unrelated
R59	S198	16127	0.0400	0.0800	1.000	0.000	0.000	unrelated
R79	R93	51930	0.0530	0.1060	1.000	0.000	0.000	unrelated
R79	S17	15403	0.0205	0.0410	1.000	0.000	0.000	unrelated
R79	S27	84651	0.0364	0.0728	1.000	0.000	0.000	unrelated
R79	S33	41792	0.0326	0.0652	1.000	0.000	0.000	unrelated
R79	S50	82291	0.0352	0.0704	1.000	0.000	0.000	unrelated
R79	S68a	49571	0.0250	0.0500	1.000	0.000	0.000	unrelated
R79	S73	72157	0.0228	0.0456	1.000	0.000	0.000	unrelated
R79	S79	80741	0.0270	0.0540	1.000	0.000	0.000	unrelated
R79	S81	41285	0.0241	0.0482	1.000	0.000	0.000	unrelated
R79	S87	100530	0.0241	0.0482	1.000	0.000	0.000	unrelated
R79	S89	79464	0.0184	0.0368	1.000	0.000	0.000	unrelated
R79	S133	70947	0.0228	0.0456	1.000	0.000	0.000	unrelated
R79	S148	51383	0.0072	0.0144	1.000	0.000	0.000	unrelated
R79	S159	84776	0.0241	0.0482	1.000	0.000	0.000	unrelated
R79	S198	48943	0.0183	0.0366	1.000	0.000	0.000	unrelated
R93	S17	20941	0.0286	0.0572	1.000	0.000	0.000	unrelated
R93	S27	115615	0.0227	0.0454	1.000	0.000	0.000	unrelated
R93	S33	55248	0.0274	0.0548	1.000	0.000	0.000	unrelated
R93	S50	111188	0.0263	0.0526	1.000	0.000	0.000	unrelated
R93	S68a	66162	0.0249	0.0498	1.000	0.000	0.000	unrelated
R93	S73	98638	0.0354	0.0708	1.000	0.000	0.000	unrelated
R93	S79	110060	0.0451	0.0902	1.000	0.000	0.000	unrelated
R93	S81	54893	0.0212	0.0424	1.000	0.000	0.000	unrelated
R93	S87	137077	0.0261	0.0522	1.000	0.000	0.000	unrelated
R93	S89	108342	0.0224	0.0448	1.000	0.000	0.000	unrelated
R93	S133	96966	0.0344	0.0688	1.000	0.000	0.000	unrelated
R93	S148	69029	0.0211	0.0422	1.000	0.000	0.000	unrelated
R93	S159	116228	0.0215	0.0430	1.000	0.000	0.000	unrelated
R93	S198	67026	0.0240	0.0480	1.000	0.000	0.000	unrelated
R101	R5	8966	0.0133	0.0266	1.000	0.000	0.000	unrelated
R101	R59	3807	-0.0087	-0.0174	1.000	0.000	0.000	unrelated
R101	R79	10863	0.0595	0.1190	0.905	0.095	0.000	unrelated
R101	**R93**	**14729**	**0.2671**	0.5342	**0.000**	**0.000**	**1.000**	**1st degree**

R101	R124	5539	0.0508	0.1016	0.993	0.007	0.000	unrelated
R101	R152	12519	0.0293	0.0586	1.000	0.000	0.000	unrelated
R101	R169	11656	0.0228	0.0456	1.000	0.000	0.000	unrelated
R101	R181	13214	0.0081	0.0162	1.000	0.000	0.000	unrelated
R101	R182	10854	0.0283	0.0566	1.000	0.000	0.000	unrelated
R101	R186	11612	0.0137	0.0274	1.000	0.000	0.000	unrelated
R101	R196	14299	0.0039	0.0078	1.000	0.000	0.000	unrelated
R101	R201	7248	0.0127	0.0254	1.000	0.000	0.000	unrelated
R101	R202	11986	0.0402	0.0804	1.000	0.000	0.000	unrelated
R101	S17	4361	-0.0004	-0.0008	1.000	0.000	0.000	unrelated
R101	S27	23948	0.0249	0.0498	1.000	0.000	0.000	unrelated
R101	S33	12050	0.0408	0.0816	1.000	0.000	0.000	unrelated
R101	S50	23687	0.0294	0.0588	1.000	0.000	0.000	unrelated
R101	S68a	14141	0.0298	0.0596	1.000	0.000	0.000	unrelated
R101	S73	20385	0.0300	0.0600	1.000	0.000	0.000	unrelated
R101	S79	22906	0.0440	0.0880	1.000	0.000	0.000	unrelated
R101	S81	11890	0.0263	0.0526	1.000	0.000	0.000	unrelated
R101	S87	28329	0.0153	0.0306	1.000	0.000	0.000	unrelated
R101	S89	22312	0.0252	0.0504	1.000	0.000	0.000	unrelated
R101	S133	19965	0.0201	0.0402	1.000	0.000	0.000	unrelated
R101	S149	14856	0.0168	0.0336	1.000	0.000	0.000	unrelated
R101	S159	23813	0.0030	0.0060	1.000	0.000	0.000	unrelated
R101	S198	14000	0.0435	0.0870	1.000	0.000	0.000	unrelated
R124	R5	15315	0.0120	0.0240	1.000	0.000	0.000	unrelated
R124	R59	6399	0.0223	0.0446	1.000	0.000	0.000	unrelated
R124	R79	18969	0.0301	0.0602	1.000	0.000	0.000	unrelated
R124	R93	25404	0.0173	0.0346	1.000	0.000	0.000	unrelated
R124	R152	21492	0.0361	0.0722	1.000	0.000	0.000	unrelated
R124	R169	19942	0.0156	0.0312	1.000	0.000	0.000	unrelated
R124	R181	22733	0.0134	0.0268	1.000	0.000	0.000	unrelated
R124	R182	18400	0.0296	0.0592	1.000	0.000	0.000	unrelated
R124	R186	20030	0.0225	0.0450	1.000	0.000	0.000	unrelated
R124	R196	24640	0.0000	0.0000	1.000	0.000	0.000	unrelated
R124	R201	12292	0.0314	0.0628	1.000	0.000	0.000	unrelated
R124	R202	20430	0.0282	0.0564	1.000	0.000	0.000	unrelated
R124	S17	7594	0.0149	0.0298	1.000	0.000	0.000	unrelated
R124	S27	41326	0.0107	0.0214	1.000	0.000	0.000	unrelated
R124	S33	20781	0.0248	0.0496	1.000	0.000	0.000	unrelated
R124	S50	40838	0.0150	0.0300	1.000	0.000	0.000	unrelated
R124	S68a	24761	0.0403	0.0806	1.000	0.000	0.000	unrelated
R124	S73	35390	0.0238	0.0476	1.000	0.000	0.000	unrelated
R124	S79	39557	0.0282	0.0564	1.000	0.000	0.000	unrelated
R124	S81	20776	0.0347	0.0694	1.000	0.000	0.000	unrelated

R124	S87	49037	0.0163	0.0326	1.000	0.000	0.000	unrelated
R124	S89	38898	0.0354	0.0708	1.000	0.000	0.000	unrelated
R124	S133	34625	0.0118	0.0236	1.000	0.000	0.000	unrelated
R124	S149	25663	0.0217	0.0434	1.000	0.000	0.000	unrelated
R124	S159	41378	0.0120	0.0240	1.000	0.000	0.000	unrelated
R124	S198	23696	0.0238	0.0476	1.000	0.000	0.000	unrelated
R152	R5	35572	0.0289	0.0578	1.000	0.000	0.000	unrelated
R152	**R59**	**14543**	**0.1511**	**0.3022**	**0.000**	**1.000**	**0.000**	**2nd degree**
R152	R79	43136	0.0223	0.0446	1.000	0.000	0.000	unrelated
R152	R93	57916	0.0149	0.0298	1.000	0.000	0.000	unrelated
R152	R169	45753	0.0319	0.0638	1.000	0.000	0.000	unrelated
R152	R181	51790	0.0246	0.0492	1.000	0.000	0.000	unrelated
R152	R182	42720	0.0354	0.0708	1.000	0.000	0.000	unrelated
R152	R186	45583	0.0280	0.0560	1.000	0.000	0.000	unrelated
R152	R196	56740	0.0326	0.0652	1.000	0.000	0.000	unrelated
R152	R201	28217	0.0335	0.0670	1.000	0.000	0.000	unrelated
R152	R202	46938	0.0335	0.0670	1.000	0.000	0.000	unrelated
R152	S17	17203	0.0484	0.0968	1.000	0.000	0.000	unrelated
R152	S27	94046	0.0272	0.0544	1.000	0.000	0.000	unrelated
R152	S33	47305	0.0551	0.1102	0.997	0.003	0.000	unrelated
R152	S50	93006	0.0751	0.1502	0.000	1.000	0.000	uncertain
R152	S68a	56283	0.0358	0.0716	1.000	0.000	0.000	unrelated
R152	S73	80526	0.0563	0.1126	0.993	0.007	0.000	unrelated
R152	S79	90359	0.0359	0.0718	1.000	0.000	0.000	unrelated
R152	S81	47149	0.0146	0.0292	1.000	0.000	0.000	unrelated
R152	S87	111512	0.0160	0.0320	1.000	0.000	0.000	unrelated
R152	S89	88393	0.0299	0.0598	1.000	0.000	0.000	unrelated
R152	S133	79018	0.0183	0.0366	1.000	0.000	0.000	unrelated
R152	S149	58368	0.0315	0.0630	1.000	0.000	0.000	unrelated
R152	S159	94381	0.0304	0.0608	1.000	0.000	0.000	unrelated
R152	S198	54219	0.0329	0.0658	1.000	0.000	0.000	unrelated
R169	R5	34107	0.0279	0.0558	1.000	0.000	0.000	unrelated
R169	R59	13401	0.0142	0.0284	1.000	0.000	0.000	unrelated
R169	R79	40897	0.0228	0.0456	1.000	0.000	0.000	unrelated
R169	R93	55923	0.0231	0.0462	1.000	0.000	0.000	unrelated
R169	R181	48324	0.0179	0.0358	1.000	0.000	0.000	unrelated
R169	R182	39747	0.0527	0.1054	1.000	0.000	0.000	unrelated
R169	R186	42293	0.0456	0.0912	1.000	0.000	0.000	unrelated
R169	R196	55010	0.0130	0.0260	1.000	0.000	0.000	unrelated
R169	R201	27371	0.0241	0.0482	1.000	0.000	0.000	unrelated
R169	R202	43291	0.0280	0.0560	1.000	0.000	0.000	unrelated
R169	S17	16584	0.0014	0.0028	1.000	0.000	0.000	unrelated
R169	S27	91457	0.0235	0.0470	1.000	0.000	0.000	unrelated

R169	S33	44360	0.0347	0.0694	1.000	0.000	0.000	unrelated
R169	S50	88073	0.0289	0.0578	1.000	0.000	0.000	unrelated
R169	S68a	52800	0.0277	0.0554	1.000	0.000	0.000	unrelated
R169	S73	77600	0.0325	0.0650	1.000	0.000	0.000	unrelated
R169	S79	86878	0.0182	0.0364	1.000	0.000	0.000	unrelated
R169	S81	44008	0.0114	0.0228	1.000	0.000	0.000	unrelated
R169	S87	108421	0.0139	0.0278	1.000	0.000	0.000	unrelated
R169	S89	85529	0.0268	0.0536	1.000	0.000	0.000	unrelated
R169	S133	76695	0.0434	0.0868	1.000	0.000	0.000	unrelated
R169	S149	54499	0.0212	0.0424	1.000	0.000	0.000	unrelated
R169	S159	91475	0.0269	0.0538	1.000	0.000	0.000	unrelated
R169	S198	52952	0.0131	0.0262	1.000	0.000	0.000	unrelated
R181	R5	37226	0.0322	0.0644	1.000	0.000	0.000	unrelated
R181	R59	15062	0.0263	0.0526	1.000	0.000	0.000	unrelated
R181	R79	45631	0.0239	0.0478	1.000	0.000	0.000	unrelated
R181	R93	60798	0.0265	0.0530	1.000	0.000	0.000	unrelated
R181	R182	44975	0.0216	0.0432	1.000	0.000	0.000	unrelated
R181	R186	48063	0.0371	0.0742	1.000	0.000	0.000	unrelated
R181	R196	59364	0.0163	0.0326	1.000	0.000	0.000	unrelated
R181	R201	29725	0.0310	0.0620	1.000	0.000	0.000	unrelated
R181	R202	49612	0.0284	0.0568	1.000	0.000	0.000	unrelated
R181	S17	18089	0.0050	0.0100	1.000	0.000	0.000	unrelated
R181	S27	99191	0.0100	0.0200	1.000	0.000	0.000	unrelated
R181	S33	50239	0.0162	0.0324	1.000	0.000	0.000	unrelated
R181	S50	98453	0.0217	0.0434	1.000	0.000	0.000	unrelated
R181	S68a	59706	0.0185	0.0370	1.000	0.000	0.000	unrelated
R181	S73	84935	0.0168	0.0336	1.000	0.000	0.000	unrelated
R181	S79	95205	0.0186	0.0372	1.000	0.000	0.000	unrelated
R181	S81	49893	0.0129	0.0258	1.000	0.000	0.000	unrelated
R181	S87	118043	0.0222	0.0444	1.000	0.000	0.000	unrelated
R181	S89	93682	0.0308	0.0616	1.000	0.000	0.000	unrelated
R181	S133	83347	0.0182	0.0364	1.000	0.000	0.000	unrelated
R181	S149	62154	0.0154	0.0308	1.000	0.000	0.000	unrelated
R181	S159	99295	0.0169	0.0338	1.000	0.000	0.000	unrelated
R181	S198	57360	0.0212	0.0424	1.000	0.000	0.000	unrelated
R182	R5	30673	0.0316	0.0632	1.000	0.000	0.000	unrelated
R182	R59	12449	0.0234	0.0468	1.000	0.000	0.000	unrelated
R182	R79	37555	0.0346	0.0692	1.000	0.000	0.000	unrelated
R182	R93	50394	0.0303	0.0606	1.000	0.000	0.000	unrelated
R182	R186	39268	0.0279	0.0558	1.000	0.000	0.000	unrelated
R182	R196	49172	0.0174	0.0348	1.000	0.000	0.000	unrelated
R182	R201	24472	0.0503	0.1006	1.000	0.000	0.000	unrelated
R182	R202	40663	0.0297	0.0594	1.000	0.000	0.000	unrelated

R182	S17	14813	0.0258	0.0516	1.000	0.000	0.000	unrelated
R182	S27	81965	0.0286	0.0572	1.000	0.000	0.000	unrelated
R182	S33	41190	0.0317	0.0634	1.000	0.000	0.000	unrelated
R182	S50	80620	0.0403	0.0806	1.000	0.000	0.000	unrelated
R182	S68a	48841	0.0247	0.0494	1.000	0.000	0.000	unrelated
R182	S73	69742	0.0288	0.0576	1.000	0.000	0.000	unrelated
R182	S79	78263	0.0493	0.0986	1.000	0.000	0.000	unrelated
R182	S81	40721	0.0296	0.0592	1.000	0.000	0.000	unrelated
R182	S87	96987	0.0262	0.0524	1.000	0.000	0.000	unrelated
R182	S89	76818	0.0088	0.0176	1.000	0.000	0.000	unrelated
R182	S133	68539	0.0403	0.0806	1.000	0.000	0.000	unrelated
R182	S149	50575	0.0366	0.0732	1.000	0.000	0.000	unrelated
R182	S159	81879	0.0253	0.0506	1.000	0.000	0.000	unrelated
R182	S198	47425	0.0330	0.0660	1.000	0.000	0.000	unrelated
R186	R5	32517	0.0308	0.0616	1.000	0.000	0.000	unrelated
R186	R59	13322	0.0284	0.0568	1.000	0.000	0.000	unrelated
R186	R79	39931	0.0379	0.0758	1.000	0.000	0.000	unrelated
R186	R93	52864	0.0178	0.0356	1.000	0.000	0.000	unrelated
R186	R196	51533	0.0321	0.0642	1.000	0.000	0.000	unrelated
R186	R201	26039	0.0453	0.0906	1.000	0.000	0.000	unrelated
R186	R202	43599	0.0365	0.0730	1.000	0.000	0.000	unrelated
R186	S17	15817	0.0356	0.0712	1.000	0.000	0.000	unrelated
R186	S27	86327	0.0268	0.0536	1.000	0.000	0.000	unrelated
R186	S33	44433	0.0287	0.0574	1.000	0.000	0.000	unrelated
R186	S50	86290	0.0305	0.0610	1.000	0.000	0.000	unrelated
R186	S68a	52594	0.0344	0.0688	1.000	0.000	0.000	unrelated
R186	S73	73872	0.0268	0.0536	1.000	0.000	0.000	unrelated
R186	S79	83233	0.0242	0.0484	1.000	0.000	0.000	unrelated
R186	S81	43942	0.0287	0.0574	1.000	0.000	0.000	unrelated
R186	S87	102743	0.0206	0.0412	1.000	0.000	0.000	unrelated
R186	S89	81548	0.0283	0.0566	1.000	0.000	0.000	unrelated
R186	S133	72538	0.0250	0.0500	1.000	0.000	0.000	unrelated
R186	S149	54395	0.0309	0.0618	1.000	0.000	0.000	unrelated
R186	S159	86513	0.0307	0.0614	1.000	0.000	0.000	unrelated
R186	S198	49815	0.0333	0.0666	1.000	0.000	0.000	unrelated
R196	R5	42700	0.0271	0.0542	1.000	0.000	0.000	unrelated
R196	R59	16497	0.0259	0.0518	1.000	0.000	0.000	unrelated
R196	R79	50749	0.0241	0.0482	1.000	0.000	0.000	unrelated
R196	R93	70686	0.0200	0.0400	1.000	0.000	0.000	unrelated
R196	R201	33746	0.0283	0.0566	1.000	0.000	0.000	unrelated
R196	R202	53424	0.0305	0.0610	1.000	0.000	0.000	unrelated
R196	S17	20611	0.0033	0.0066	1.000	0.000	0.000	unrelated
R196	S27	113001	0.0256	0.0512	1.000	0.000	0.000	unrelated

R196	S33	53850	0.0368	0.0736	1.000	0.000	0.000	unrelated
R196	S50	108295	0.0286	0.0572	1.000	0.000	0.000	unrelated
R196	S68a	64297	0.0276	0.0552	1.000	0.000	0.000	unrelated
R196	S73	96345	0.0233	0.0466	1.000	0.000	0.000	unrelated
R196	S79	107155	0.0245	0.0490	1.000	0.000	0.000	unrelated
R196	S81	53577	0.0155	0.0310	1.000	0.000	0.000	unrelated
R196	S87	133632	0.0179	0.0358	1.000	0.000	0.000	unrelated
R196	S89	105909	0.0282	0.0564	1.000	0.000	0.000	unrelated
R196	S133	94737	0.0324	0.0648	1.000	0.000	0.000	unrelated
R196	S149	66991	0.0260	0.0520	1.000	0.000	0.000	unrelated
R196	S159	113654	0.0213	0.0426	1.000	0.000	0.000	unrelated
R196	S198	65766	0.0290	0.0580	1.000	0.000	0.000	unrelated
R201	R5	20773	0.0122	0.0244	1.000	0.000	0.000	unrelated
R201	R59	8399	0.0464	0.0928	1.000	0.000	0.000	unrelated
R201	R79	25320	0.0262	0.0524	1.000	0.000	0.000	unrelated
R201	R93	34386	0.0151	0.0302	1.000	0.000	0.000	unrelated
R201	R202	26726	0.0322	0.0644	1.000	0.000	0.000	unrelated
R201	S17	10152	0.0237	0.0474	1.000	0.000	0.000	unrelated
R201	S27	55393	0.0199	0.0398	1.000	0.000	0.000	unrelated
R201	S33	27183	0.0239	0.0478	1.000	0.000	0.000	unrelated
R201	S50	53769	0.0114	0.0228	1.000	0.000	0.000	unrelated
R201	S68a	32252	0.0238	0.0476	1.000	0.000	0.000	unrelated
R201	S73	47430	0.0192	0.0384	1.000	0.000	0.000	unrelated
R201	S79	52808	0.0198	0.0396	1.000	0.000	0.000	unrelated
R201	S81	26817	0.0136	0.0272	1.000	0.000	0.000	unrelated
R201	S87	65667	0.0240	0.0480	1.000	0.000	0.000	unrelated
R201	S89	52342	0.0321	0.0642	1.000	0.000	0.000	unrelated
R201	S133	46830	0.0081	0.0162	1.000	0.000	0.000	unrelated
R201	S149	33523	0.0428	0.0856	1.000	0.000	0.000	unrelated
R201	S159	55718	0.0166	0.0332	1.000	0.000	0.000	unrelated
R201	S198	32485	0.0181	0.0362	1.000	0.000	0.000	unrelated
R202	R5	33595	0.0176	0.0352	1.000	0.000	0.000	unrelated
R202	R59	13675	0.0218	0.0436	1.000	0.000	0.000	unrelated
R202	R79	41006	0.0325	0.0650	1.000	0.000	0.000	unrelated
R202	R93	54857	0.0247	0.0494	1.000	0.000	0.000	unrelated
R202	S17	16365	0.0250	0.0500	1.000	0.000	0.000	unrelated
R202	S27	89176	0.0313	0.0626	1.000	0.000	0.000	unrelated
R202	S33	45467	0.0500	0.1000	1.000	0.000	0.000	unrelated
R202	S50	88869	0.0328	0.0656	1.000	0.000	0.000	unrelated
R202	S68a	53485	0.0430	0.086	1.000	0.000	0.000	unrelated
R202	S73	76431	0.0282	0.0564	1.000	0.000	0.000	unrelated
R202	S79	85382	0.0326	0.0652	1.000	0.000	0.000	unrelated
R202	S81	44890	0.0349	0.0698	1.000	0.000	0.000	unrelated

R202	S87	105764	0.0216	0.0432	1.000	0.000	0.000	unrelated
R202	S89	83928	0.0231	0.0462	1.000	0.000	0.000	unrelated
R202	S133	74657	0.0395	0.0790	1.000	0.000	0.000	unrelated
R202	S149	55936	0.0253	0.0506	1.000	0.000	0.000	unrelated
R202	S159	89113	0.0304	0.0608	1.000	0.000	0.000	unrelated
R202	S198	51446	0.0281	0.0562	1.000	0.000	0.000	unrelated
S17	S27	34242	0.0334	0.0668	1.000	0.000	0.000	unrelated
S17	S33	16441	0.0434	0.0868	1.000	0.000	0.000	unrelated
S17	**S50**	**33080**	**0.1728**	**0.3456**	**0.000**	**1.000**	**0.000**	**2nd degree**
S17	S68a	19842	0.0242	0.0484	1.000	0.000	0.000	unrelated
S17	S73	29317	0.0379	0.0758	1.000	0.000	0.000	unrelated
S17	S79	32565	0.0432	0.0864	1.000	0.000	0.000	unrelated
S17	S81	16645	0.0046	0.0092	1.000	0.000	0.000	unrelated
S17	S87	40730	0.0228	0.0456	1.000	0.000	0.000	unrelated
S17	S89	31973	0.0101	0.0202	1.000	0.000	0.000	unrelated
S17	S198	19879	0.0117	0.0234	1.000	0.000	0.000	unrelated
S27	S33	90757	0.0398	0.0796	1.000	0.000	0.000	unrelated
S27	S50	182098	0.0235	0.0470	1.000	0.000	0.000	unrelated
S27	S68a	108691	0.0405	0.0810	1.000	0.000	0.000	unrelated
S27	S73	160607	0.0332	0.0664	1.000	0.000	0.000	unrelated
S27	S79	179633	0.0328	0.0656	1.000	0.000	0.000	unrelated
S27	S81	90186	0.0190	0.0380	1.000	0.000	0.000	unrelated
S27	S87	224515	0.0231	0.0462	1.000	0.000	0.000	unrelated
S27	S89	176718	0.0235	0.0470	1.000	0.000	0.000	unrelated
S27	S50	90442	0.0366	0.0732	1.000	0.000	0.000	unrelated
S33	S68a	54978	0.0313	0.0626	1.000	0.000	0.000	unrelated
S33	S73	77281	0.0314	0.0628	1.000	0.000	0.000	unrelated
S33	S79	87430	0.0218	0.0436	1.000	0.000	0.000	unrelated
S33	S81	45803	0.0101	0.0202	1.000	0.000	0.000	unrelated
S33	S87	107694	0.0290	0.0580	1.000	0.000	0.000	unrelated
S33	S89	85153	0.0348	0.0696	1.000	0.000	0.000	unrelated
S50	S68a	107442	0.0602	0.1204	0.850	0.150	0.000	unrelated
S50	S73	155409	0.0353	0.0706	1.000	0.000	0.000	unrelated
S50	S79	174434	0.0288	0.0576	1.000	0.000	0.000	unrelated
S50	S81	89409	0.0228	0.0456	1.000	0.000	0.000	unrelated
S50	S87	217024	0.0160	0.0320	1.000	0.000	0.000	unrelated
S50	S89	170852	0.0278	0.0556	1.000	0.000	0.000	unrelated
S68a	S73	92676	0.0349	0.0698	1.000	0.000	0.000	unrelated
S68a	S79	104065	0.0361	0.0722	1.000	0.000	0.000	unrelated
S68a	S81	54224	0.0331	0.0662	1.000	0.000	0.000	unrelated
S68a	S87	129278	0.0456	0.0912	1.000	0.000	0.000	unrelated
S68a	S89	101996	0.0259	0.0518	1.000	0.000	0.000	unrelated
S73	S79	153070	0.0232	0.0464	1.000	0.000	0.000	unrelated

S73	S81	77059	0.0479	0.0958	1.000	0.000	0.000	unrelated
S73	S87	190844	0.0379	0.0758	1.000	0.000	0.000	unrelated
S73	S89	150466	0.0321	0.0642	1.000	0.000	0.000	unrelated
S79	S81	86587	0.0199	0.0398	1.000	0.000	0.000	unrelated
S79	S87	214126	0.0192	0.0384	1.000	0.000	0.000	unrelated
S79	S89	168236	0.0185	0.0370	1.000	0.000	0.000	unrelated
S81	S87	107272	0.0521	0.1042	1.000	0.000	0.000	unrelated
S81	S89	84936	0.0177	0.0354	1.000	0.000	0.000	unrelated
S87	S89	210233	0.0268	0.0536	1.000	0.000	0.000	unrelated
S133	S17	28807	0.0267	0.0534	1.000	0.000	0.000	unrelated
S133	S27	158464	0.0752	0.1504	0.000	1.000	0.000	uncertain
S133	S33	75919	0.0326	0.0652	1.000	0.000	0.000	unrelated
S133	S50	152819	0.0409	0.0818	1.000	0.000	0.000	unrelated
S133	S68a	91151	0.0272	0.0544	1.000	0.000	0.000	unrelated
S133	S73	134877	0.0401	0.0802	1.000	0.000	0.000	unrelated
S133	S79	150854	0.0329	0.0658	1.000	0.000	0.000	unrelated
S133	S81	75810	0.0278	0.0556	1.000	0.000	0.000	unrelated
S133	S87	188204	0.0178	0.0356	1.000	0.000	0.000	unrelated
S133	S89	148113	0.0270	0.0540	1.000	0.000	0.000	unrelated
S133	S149	94574	0.0237	0.0474	1.000	0.000	0.000	unrelated
S133	S159	158761	0.0364	0.0728	1.000	0.000	0.000	unrelated
S133	S198	91369	0.0214	0.0428	1.000	0.000	0.000	unrelated
S149	S17	20414	0.0225	0.0450	1.000	0.000	0.000	unrelated
S149	S27	112554	0.0189	0.0378	1.000	0.000	0.000	unrelated
S149	S33	57118	0.0191	0.0382	1.000	0.000	0.000	unrelated
S149	S50	112167	0.0217	0.0434	1.000	0.000	0.000	unrelated
S149	S68a	67576	0.0342	0.0684	1.000	0.000	0.000	unrelated
S149	S73	96048	0.0310	0.0620	1.000	0.000	0.000	unrelated
S149	S79	108192	0.0253	0.0506	1.000	0.000	0.000	unrelated
S149	S81	56323	0.0204	0.0408	1.000	0.000	0.000	unrelated
S149	S87	134515	0.0209	0.0418	1.000	0.000	0.000	unrelated
S149	S89	106033	0.0180	0.0360	1.000	0.000	0.000	unrelated
S149	S159	112513	0.0266	0.0532	1.000	0.000	0.000	unrelated
S149	S198	64687	0.0313	0.0626	1.000	0.000	0.000	unrelated
S159	S17	34513	0.0214	0.0428	1.000	0.000	0.000	unrelated
S159	S27	189553	0.0160	0.0320	1.000	0.000	0.000	unrelated
S159	S33	90572	0.0270	0.0540	1.000	0.000	0.000	unrelated
S159	S50	182503	0.0398	0.0796	1.000	0.000	0.000	unrelated
S159	S68a	108628	0.0239	0.0478	1.000	0.000	0.000	unrelated
S159	S73	161450	0.0242	0.0484	1.000	0.000	0.000	unrelated
S159	S79	180615	0.0254	0.0508	1.000	0.000	0.000	unrelated
S159	S81	90246	0.0200	0.0400	1.000	0.000	0.000	unrelated
S159	S87	225146	0.0251	0.0502	1.000	0.000	0.000	unrelated

S159	S89	177430	0.0214	0.0428	1.000	0.000	0.000	unrelated
S159	S198	109087	0.0160	0.0320	1.000	0.000	0.000	unrelated
S198	S27	108896	0.0298	0.0596	1.000	0.000	0.000	unrelated
S198	S33	52175	0.0257	0.0514	1.000	0.000	0.000	unrelated
S198	S50	104989	0.0292	0.0584	1.000	0.000	0.000	unrelated
S198	S68a	62373	0.0358	0.0716	1.000	0.000	0.000	unrelated
S198	S73	93127	0.0299	0.0598	1.000	0.000	0.000	unrelated
S198	S79	103461	0.0328	0.0656	1.000	0.000	0.000	unrelated
S198	S81	52440	0.0160	0.0320	1.000	0.000	0.000	unrelated
S198	S87	129085	0.0302	0.0604	1.000	0.000	0.000	unrelated
S198	S89	102048	0.0133	0.0266	1.000	0.000	0.000	unrelated

Out of 378 dyads, one pair of first-degree relatives (R101/R93) and two pairs of second-degree relatives (R152/R59 and S17/S50) were identified.[157] To visually represent the relatedness of these related dyads, their Rxy values were projected onto a background comprising relatedness coefficients from 600 virtual dyads of unrelated individuals, second-degree relatives, and first-degree relatives created *in silico* (Figures S10.1-S10.3).

The Rxy for the dyad of individuals R101 and R93 was estimated to be 0.267 based on 14,729 intersecting SNPs. This would be equivalent to an Rxy of 0.534 if a forced homozygote approach was not used. This Rxy value classified these individuals as first-degree relatives and was supported by posterior probabilities, as illustrated in Figure S10.1.

[157] Three additional dyads (R59/S73, R152/S50, and S133/S27) were identified as "plausibly related" based on a relatedness coefficient outside the expected range for unrelated individuals (-0.05>Rxy>0.05); these dyads are not assessed further in this dissertation, but will be explored in greater detail in the future.

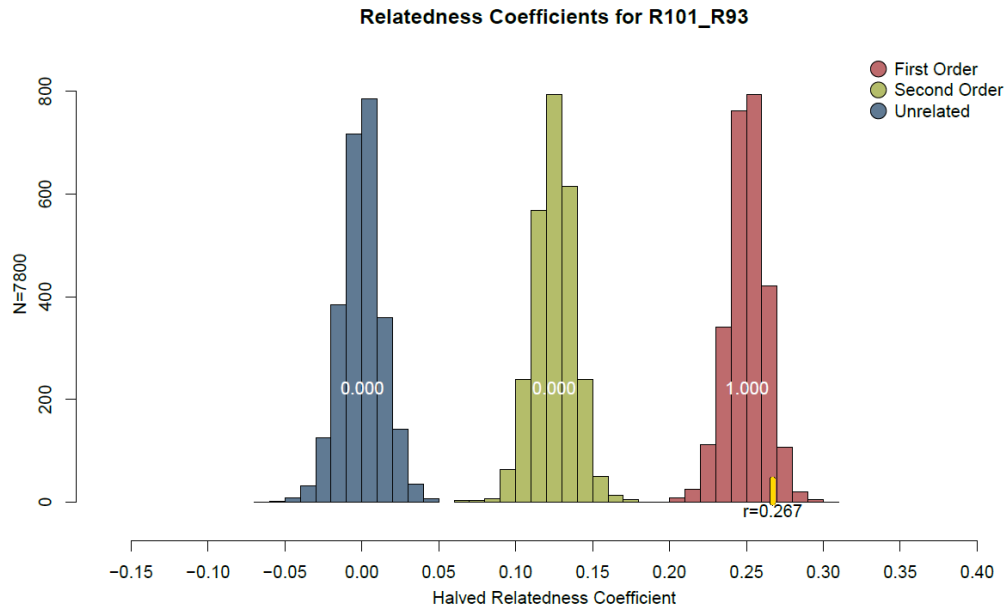


Figure S10.1: Projection of relatedness coefficient for the dyad R101/R93 onto a background of three normal and non-overlapping curves generated from the relatedness coefficients of 600 virtual dyads of known relatedness created in silico.

R101 and R93 were both determined to be male using the molecular sexing method in Skoglund et al. (2013) and were aged as 4 years old and 10 years old, respectively (Adams et al. 1999). This led to the conclusion that R101 and R93 were brothers, an interpretation which was by the assignment of both individuals to mtDNA haplogroup H2a and Y chromosome haplogroup E1b1b1b2.[158]

The layout of the Kulubnarti cemeteries was not recorded accurately during excavation, and we are therefore unable to determine the proximity of the burials of related individuals.[159] However, recorded details of each burial reveal that these brothers were both interred in slot graves, but were buried in different positions: R101 was buried on his left side with his head facing to the left, his arms on his pelvis, his legs slightly

[158] Additional detail regarding both mtDNA and Y chromosome haplogroup of all captured Kulubnarti individuals is provided in Chapter 8.
[159] Dr. Dennis Van Gerven confirmed the lack of an accurate record of grave locations.

flexed, three stones covering his head, and no stones covering his body, while R93 was buried on his back (dorsal) with his head facing right, his arms at his sides, his legs slightly flexed, and no stones covering his head or body (Adams et al. 1999). While the motives behind the different positioning of these two brothers remain unclear, both burial positions are found at relatively high frequencies at Kulubnarti.

In addition to the R cemetery dyad identified as brothers, two pairs of second-degree relatives were also identified, one dyad from the R cemetery and the other from the S cemetery. R152 and R59 were determined to be second-degree relatives with a relatedness coefficient of 0.1511 based on 14,543 intersecting SNPs. This Rxy estimate is equivalent to an Rxy estimate of 0.3022 if the forced homozygote approach was not used. This Rxy value classified these individuals as second-degree relatives and was supported by posterior probabilities, as illustrated in Figure S10.2.

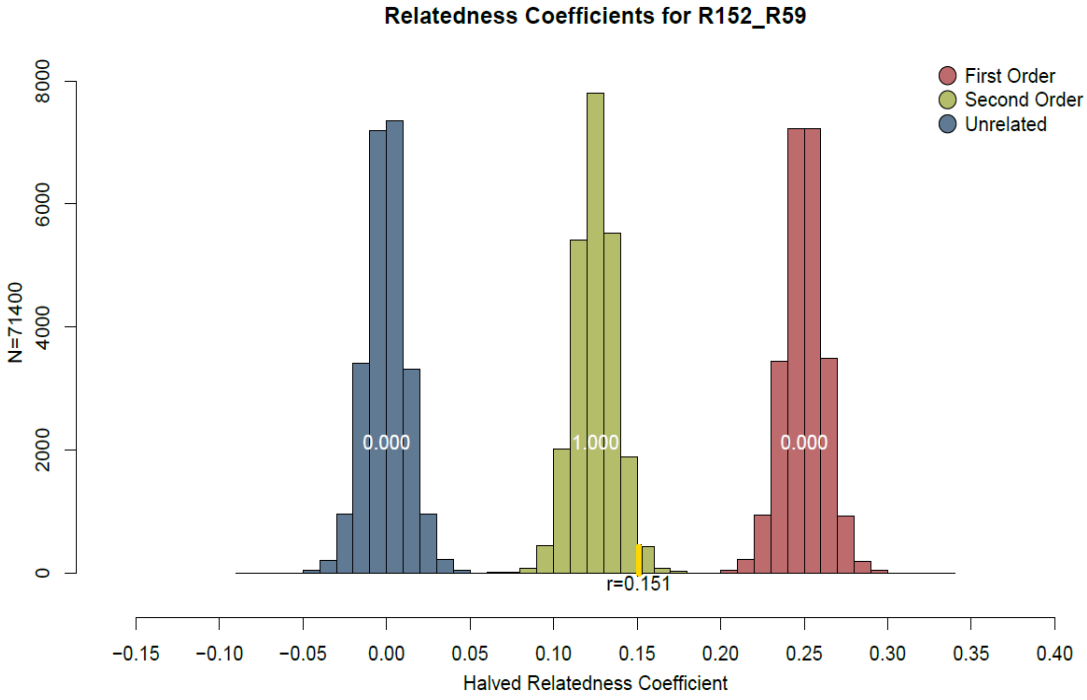


Figure S10.2: Projection of relatedness coefficient for the dyad R152/R59 onto a background of three normal and non-overlapping curves generated from the relatedness coefficients of 600 virtual dyads of known relatedness created in silico.

The molecular sexing technique in Skoglund et al. (2013) determined that R152 was female and R59 was male, aged at 3 years old and 9 years old, respectively (Adams et al. 1999). This leads us to conclude that these individuals were most likely half-siblings or double cousins (the result of two siblings from one family mating with two siblings of another family); we are unable to rule out the possibility that they are uncle/niece or aunt/nephew, though it is unlikely based on their nearness in age and the relatively short expected lifespan of the Kulubnarti Nubians (Adams et al. 1999). R152 and R59 do not share an mtDNA haplogroup, with R152 assigned to haplogroup L0a1a1 and R59 assigned to haplogroup H2a; it is interesting to note that mtDNA haplogroup L is of African origin, while H2a is of non-African origin, suggesting that there may have been individuals of various biogeographic backgrounds in this family. As R152 is a female, a comparison of Y chromosome haplogroup is not possible.

Based on the record by Adams et al. (1999), R152 and R59 were both buried in slot graves. R152 was buried on her left side with her head facing to the left, her arms at her sides, her legs straight, three stones covering her head, and no stones covering her body. R59 was buried in a quarter-left position (partially on his left side, but neither fully on his side nor on his back) with his head facing to the left, his arms at his pelvis, his legs slightly flexed, three stones covering his head, and no stones covering his body. Again, diversity in burial position is observed between these two related individuals, though neither is positioned in a way that would be unexpected at Early Christian Kulubnarti.

A second dyad of second-degree individuals was identified in the S cemetery. S17 and S50 were determined to be second-degree relatives with a relatedness coefficient of 0.1728 based on 33,080 intersecting SNPs. This Rxy estimate is equivalent to an Rxy estimate of 0.3456 if the forced homozygote approach was not used. This Rxy value classified these individuals as second-degree relatives and was supported by posterior probabilities, as illustrated in Figure S10.3.

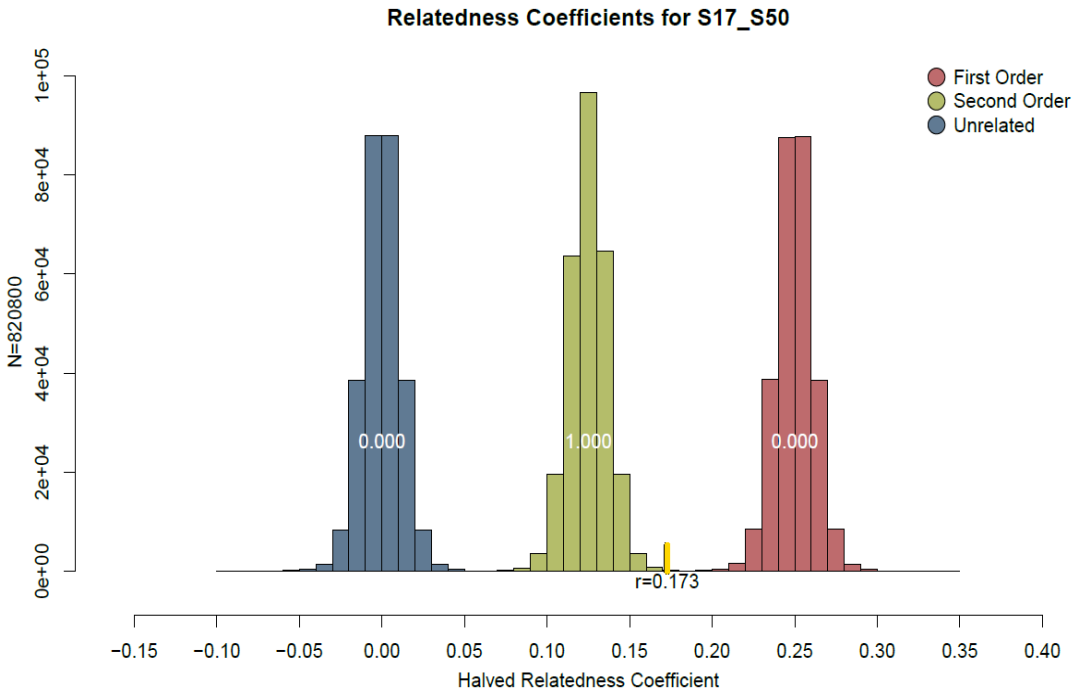


Figure S10.3: Projection of relatedness coefficient for the dyad S17/S50 onto a background of three normal and non-overlapping curves generated from the relatedness coefficients of 600 virtual dyads of known relatedness created in silico.

The molecular sexing technique in Skoglund et al. (2013) determined that S17 was female and S50 was male, and these individuals were both aged at 5 years old. Again, these individuals were most likely half-siblings or double cousins, though the possibility cannot be ruled out that they were uncle/niece or aunt/nephew. S17 and S50

do not share an mtDNA haplogroup, with S17 assigned to T1a and S50 assigned to L0a1a1. As with second-degree relatives R152 and R59, there is evidence of various biogeographic backgrounds in this family. As S17 is a female, a comparison of Y chromosome haplogroup is not possible.

Both individuals were buried in slot graves, and Adams et al. (1999) recorded S17 to have been buried on her left side with her head facing to the left, her arms at her pelvis, her legs straight, and no coverings over either her head or body, while S50 was buried on his left side with his head facing to the left, his arms at his sides, his legs straight, three stones covering his face, and no stones covering his body. Again, slight differences in burial position were evident, though both individuals were buried in commonly observed positions at Kulubnarti.

Due to unknown contemporaneity between 28 individuals analyzed, relatedness analysis cannot contribute to broad conclusions regarding the genetic relationship between the S and R communities at Kulubnarti; however, the identification of relatives provides additional insight into relationships at Kulubnarti and enables the accurate estimation of F_{ST} in this dissertation.

APPENDIX A: DABNEY EXTRACTION PROTOCOL

PART 1: RECIPES

Note: all recipes are written for one sample unless noted; initial concentration of reagent stock is in parentheses alongside reagent name.

Extraction Buffer

	final conc.	1x
EDTA (0.5 M)	0.45 M	900 µL
Proteinase K (15 mg/mL)	0.25 mg/mL	16.7 µL
ddH_2O*	-	83.3 µL

***laboratory grade double-distilled H_2O should be used in all solutions.**

- Combine the ddH_2O and EDTA
- UV irradiate the solution for 30 minutes
- Add the Proteinase K and pipette mix gently
- Store at room temperature

Binding Buffer

Note: although only 13 mL of binding buffer is required for each sample, it is recommended to make 50 mL at a time.

	final conc.	amount for 50mL
Guanidine hydrochloride (MW 95.53)	5 M	23.9 g
Isopropanol	40%	20 mL
Tween-20 (10%)	0.05%	250 µL
Sodium acetate (3.0 M)	90 mM	1.5 mL
Make up to 50mL with ddH_2O	-	-

- In a 50 mL Falcon tube, combine isopropanol and Tween-20; mix well by inverting.
- Add sodium acetate; mix well by inverting
- Add guanidine hydrochloride slowly, and fill the Falcon tube to the 50 mL line with ddH_2O.
- Shake the tube as long as is necessary to dissolve all the guanidine hydrochloride; this can take several minutes

- UV irradiate for 30 minutes
- Store at room temperature

TET Elution Buffer

Note: because TET buffer can be stored indefinitely, it is recommended to make 50 mL at a time.

	final conc.	amount for 50mL
Tris-HCL solution (1 M)	10 mM	500 µL
EDTA solution (0.5 M)	1 mM	100 µL
Tween-20 solution (10%)	0.05%	250 µL
Make up to 50mL with ddH_2O	-	49 mL + 150 µL

- Combine all reagents in 50 mL Falcon tube and mix by inverting.
- UV irradiate for 30 minutes
- Store at room temperature

PART 2: ASSEMBLY OF CUSTOM-MADE BINDING APPARATUS

Consumables Required:

- Zymo-spin V column extension reservoir (1 per sample)
- Qiagen MinElute silica spin column (1 per sample)
- 50mL Falcon tube (1 per sample)

Assembly Procedure:

- Remove Zymo-spin V column from the extension reservoir, and submerge the extension reservoirs in ≥10% bleach solution for 20 minutes
- Remove extension reservoirs from bleach solution and rinse well with laboratory-grade ddH_2O
- Dry extension reservoir with paper towels, and wipe with DNA-ExitusPlus
- Place clean extension reservoirs in cross-linker and UV-irradiate for 30 minutes
- Remove the MinElute kit from the refrigerator and separate MinElute silica spin column from MinElute collection tube; keep the collection tube
- Forcibly fit the Zymo-spin V extension reservoir on to the MinElute silica spin column to make the custom-made binding apparatus; this will require some force!
- Place the custom-made binding apparatus into a sterile 50mL Falcon tube
- Cut off the lid from the MinElute collection tubes and keep the lid

---

PART 3: EXTRACTION PROCEDURE

Note: all steps should take place inside a laminar flow cabinet dedicated to DNA extraction.

- Thoroughly clean all working surfaces with DNA-ExitusPlus or bleach solution
- Prepare the extraction buffer according to the recipe in Part 1; note that the addition of Proteinase K occurs *after* the buffer has been UV-irradiated for 30 minutes
- Remove all tubes containing ~50mg bone powder from refrigerator and place them in tube rack; prepare an extraction control (an empty 1.5mL tube)
- Add 1mL of extraction buffer into each tube; be sure to change pipette tip between each sample to avoid cross-contamination
- Vortex the tubes to suspend the bone powder in the extraction buffer; flick the bottom of the tube with your fingers if the bone powder is not suspending by vortexing alone
- Place the tubes onto a thermal mixer set at 37ºC and rotation at 1200rpm for 18 hours
- *OPTIONAL*: Prepare the binding buffer according to the recipe in Part 1 so that it is ready for the following day
- *OPTIONAL*: Sterilize the Zymo-spin V extension reservoirs by following the instructions in Part 2 up to and including the UV-irradiation step; note that the MinElute spin columns should not be removed from the refrigerator until the extractions have finished their 18 hours in the thermal mixer

----------overnight----------

- After 18 hours, remove the tubes from the thermal mixer
- Centrifuge the tubes for 2 minutes at max speed on mini-centrifuge (13,300rpm)
- Assemble a custom-made binding apparatus for each sample following the instructions in Part 2
- Add 13mL of binding buffer to each extension reservoir
- Remove the tubes from the mini-centrifuge and collect 1mL of DNA extract supernatant from each tube, changing pipette tips between each sample to avoid cross-contamination and avoiding collecting any bone powder from the pellet at the bottom of the tube (this powder will clog the membrane of the spin column and will inhibit future centrifugation steps)
- Add the supernatant to the 13mL of binding buffer in the extension reservoir and pipette-mix; repeat for each sample
- Secure the lids of the 50mL Falcon tubes containing the binding apparatus using paper tape (this step is done because once the binding apparatus is added to the 50mL Falcon tube, the cap of the tube no longer attaches securely to the tube body)
- Centrifuge the 50mL Falcon tubes for 4 minutes at 2,500rpm on a large centrifuge
- After 4 minutes, visually examine each Falcon tube to determine if the solution has passed through the MinElute spin column completely; if the solution has passed through completely, do not centrifuge further

- If the solution has not passed through the MinElute spin column completely, rotate the Falcon tube 90º and centrifuge again for 2 minutes at 3,000rpm on large centrifuge
- Remove Falcon tubes from centrifuge and disassemble the binding apparatus by detaching the MinElute spin column from the Zymo-spin V extension reservoir
- Discard the extension reservoir and 50mL Falcon tube and place each MinElute spin column in its original collection tube and replace the cap
- Dry-spin the MinElute tubes (now containing the spin columns) for 1 minute at 6,000rpm on the mini-centrifuge
- Add 650µL of PE washing buffer to the MinElute tube; note that PE wash buffer is part of the QIAquick PCR Purification Kit by Qiagen)
- Centrifuge for 1 minute at 6,000rpm on mini-centrifuge
- Discard flow-through
- Add 650µL of PE washing buffer to MinElute tubes
- Centrifuge at 1 minute at 6,000rpm on mini-centrifuge
- Discard flow-through
- Dry-spin the MinElute tube for 1 minute at 13,300rpm
- Remove MinElute spin column from collection tube and place in a clean 2.0mL tube
- Add 12.5µL TET Elution buffer to the spin column by pipetting it directly onto the silica membrane
- Incubate for 10 minutes at room temperature
- Centrifuge for 30 seconds at 13,300rpm
- Add 12.5µL TET Elution buffer to the spin column by pipetting it directly onto the silica membrane
- Incubate for 10 minutes at room temperature
- Centrifuge for 30 seconds at 13,300rpm; this will give a total of 25µL DNA extract
- Store extract in freezer maintained at -20ºC until library preparation takes place

APPENDIX B: MEYER AND KIRCHER LIBRARY PREPARATION PROTOCOL

PART 1: PREPARATION OF NECESSARY REAGENTS

Note: all recipes are written for one sample unless noted; initial concentration of reagent stock is in parentheses alongside reagent name.

Oligo Hybridization Buffer

	Amount for 1mL
Sodium Chloride solution (5M)	100 µL
TrisHCl (1M)	10 µL
EDTA (0.5M)	2 µL
ddH_2O	888 µL

- Combine all reagents in 1.5mL tube and mix by pipetting up and down
- Store in a freezer maintained at -20ºC

Adapter Mix

Note: the sequences of all adapters required for the adapter mix can be found in Appendix C.

- Set thermal mixer to 95ºC
- Mix P5 and P7 adapters in separate 0.5mL tubes according to the following table:

P5 Adapter Mix (200µM)

	Amount for 0.5 µL
IS1_adapter_P5.F (500µM)	0.2 µL
IS3_adapter_P5+P7.R (500µM)	0.2 µL
Oligo Hybridization Buffer (10x)	0.05 µL
ddH_2O	0.05 µL

P7 Adapter Mix (200mM)

	Amount for 0.5 µL
IS2_adapter_P7.F (500µM)	0.2 µL
IS3_adapter_P5+P7.R (500µM)	0.2 µL
Oligo Hybridization Buffer (10x)	0.05 µL
ddH₂O	0.05 µL

- Mix each adapter mix by pipetting up and down
- Incubate the reactions in a thermal mixer for 10 seconds at 95ºC followed by a ramp from 95ºC to 12ºC at a rate of 0.1ºC/second; note that this will take >2 hours
- Combine both P5 and P7 reactions in a new 0.5mL tube to obtain a ready to use adapter mix (100µM each adapter)
- Store in a freezer maintained at -20ºC

Primer IS4

Note: the sequence of primer IS4 can be found in Appendix C.

	final conc.	1x
IS4 (100µM)	10 µM	10 µL
ddH₂O	-	90 µL

- Combine all reagents in 0.5mL tube and mix by pipetting up and down
- Store in a freezer maintained at -20ºC

Indexing Primers

	final conc.	1x
Index (100µM)	5 µM	5 µL
ddH₂O	-	95 µL

- Combine all reagents in 0.5mL tube and mix by pipetting up and down
- Store in a freezer maintained at -20ºC

EBT Elution Buffer

Note: EB buffer is provided in the Qiagen MinElute PCR Purification kit

	final conc.	1x
EB Buffer	-	59.97 µL
Tween-20	0.05%	0.03 µL

- Combine all reagents in 1.5mL tube and mix by pipetting up and down
- UV irradiate for 30 minutes
- Store at room temperature

PART 2: LIBRARY PREPARATION PROCEDURE

Note: all recipes are written for one sample

STEP 1: BLUNT-END REPAIR

- Before beginning blunt-end repair, set one thermal mixer to 12ºC and another to 25ºC
- Place tubes with 25µL DNA extracts in tube rack and allow them to thaw; prepare a negative library control by adding 25µL of laboratory-grade ddH_2O to a sterile 1.5mL tube
- Prepare the blunt-end repair mastermix in a 1.5mL tube according to the following table:

Blunt-End Repair Mastermix

	1x
End Repair Buffer	7.0 µL
End Repair Enzyme Mix	3.5 µL
ddH_2O*	34.5 µL

***laboratory grade double-distilled H_2O should be used in all solutions.**

- Mix mastermix by pipetting up and down; do not vortex after the addition of enzymes
- Transfer 45µL of mastermix to each tube that contains 25µL DNA extract and the tube that contains the library control; mix by pipetting up and down
- Incubate reactions for 15 minutes at 25ºC on thermal mixer
- Transfer reactions from thermal mixer set at 25ºC to thermal mixer set at 12ºC
- Incubate reactions for 5 minutes at 12ºC
- Proceed immediately to Step 1A, or place reactions in freezer maintained at -20ºC

STEP 1A: SAMPLE CLEAN-UP

- Take out the necessary reagents from the Qiagen MinElute PCR Purification kit: PB (binding) buffer, PE (wash) buffer, and EBT (elution) buffer
- Take out one MinElute tube with silica spin column for each sample from the 4ºC refrigerator and label one for each sample
- Add 350µL of PB binding buffer to each MinElute tube and then add the 70µL blunt-end repaired sample; to ensure total sample volume is collected, samples may need to be spun down with quick spin in mini-centrifuge prior to being added to MinElute tube
- Centrifuge for 1 minute at 13,300rpm on mini-centrifuge
- Discard PB buffer flow-through collected at bottom of MinElute tube
- Add 750µL PE wash buffer to each MinElute tube
- Centrifuge for 1 minute at 13,300rpm on mini-centrifuge
- Discard PE buffer flow-through collected at bottom of MinElute tube
- Centrifuge again (dry-spin) for 1 minute at 13,300rpm to remove any residual wash buffer
- Remove the silica spin column from the MinElute tube and place in a sterile 1.5mL tube
- Add 20µL of EBT elution buffer directly to the silica membrane of the MinElute tube
- Incubate for 1 minute at room temperature
- Centrifuge for 1 minute at 13,300rpm on mini-centrifuge
- Remove MinElute spin columns from 1.5mL tubes, and discard spin columns, keeping only the eluate at the bottom of the tubes
- Proceed immediately to Step 2 or place samples in freezer maintained at -20ºC

STEP 2: ADAPTER LIGATION

- Before beginning adapter ligation, set thermal mixer to 22ºC
- Prepare adapter ligation mastermix in a 1.5mL tube according to the following table:

Adapter Ligation Mastermix*

	1x
ddH_2O	10.0 µL
T4 DNA ligase buffer (10x)	4.0 µL
PEG-4000 (50%)	4.0 µL
Adapter mix (100uM each)	1.0 µL
T4 DNA ligase (5U/µL)	1.0 µL

***Reagents should be added in the order listed here; vortex after the addition of PEG-4000 and pipette mix ONLY after the addition of enzymes**

- Mix mastermix by pipetting up and down
- Transfer 20µL of mastermix to each tube that contains 20µL eluate; mix by pipetting up and down

- Incubate reactions for 30 minutes at 22ºC on thermal mixer
- Proceed immediately to Step 2A, or place reactions in freezer maintained at -20ºC

STEP 2A: SAMPLE CLEAN-UP

- Take out the necessary reagents from the Qiagen MinElute PCR Purification kit: PB (binding) buffer, PE (wash) buffer, and EBT (elution) buffer
- Take out one MinElute tube with silica spin column for each sample from the 4ºC refrigerator and label one for each sample
- Add 200µL of PB binding buffer to each MinElute tube and then add the 40µL adapter-ligated sample; to ensure total sample volume is collected, samples may need to be spun down with quick spin in mini-centrifuge prior to being added to MinElute tube
- Centrifuge for 1 minute at 13,300rpm on mini-centrifuge
- Discard PB buffer flow-through collected at bottom of MinElute tube
- Add 750µL PE wash buffer to each MinElute tube
- Centrifuge for 1 minute at 13,300rpm on mini-centrifuge
- Discard PE buffer flow-through collected at bottom of MinElute tube
- Centrifuge again (dry-spin) for 1 minute at 13,300rpm to remove any residual wash buffer
- Remove the silica spin column from the MinElute tube and place in a sterile 1.5mL tube
- Add 20µL of EBT elution buffer directly to the silica membrane of the MinElute tube
- Incubate for 1 minute at room temperature
- Centrifuge for 1 minute at 13,300rpm on mini-centrifuge
- Remove MinElute spin columns from 1.5mL tubes, and discard spin columns, keeping only the eluate at the bottom of the tubes
- Proceed immediately to Step 3 or place samples in freezer maintained at -20ºC

STEP 3: ADAPTER FILL-IN

- Before beginning adapter fill-in, set one thermal mixer to 37ºC and another to 80ºC
- Prepare adapter fill-in mastermix in a 1.5mL tube according to the following table:

Adapter Fill-In Mastermix

	1x
ddH_2O	13.5 µL
ThermoPol Reaction Buffer (10x)	4.0 µL
dNTPs* (10mM each)	1.0 µL
Bst polymerase, large fragment (8U/µL)	1.5 µL

***Let dNTPs thaw naturally, do not try to warm with your hands or using a thermal mixer**

- Mix mastermix by pipetting up and down
- Transfer 20µL of mastermix to each tube that contains 20µL eluate; mix by pipetting up and down
- Incubate reactions for 30 minutes at 37ºC on thermal mixer
- Transfer reactions from thermal mixer set at 37ºC to thermal mixer set at 80ºC
- Incubate reactions for 20 minutes at 80ºC
- Proceed immediately to Step 4, or place reactions in freezer maintained at -20ºC

STEP 4: INDEXING PCR

- Label one clean 0.5mL PCR tube for each sample
- Prepare amplification mastermix in a 1.5mL tube according to the following table:

Amplification Mastermix

	1x
AccuPrime Pfx Supermix	20.5 µL
Primer IS4 (10uM)	0.5 µL

- Mix mastermix by pipetting up and down
- Transfer 21µL of mastermix to clean 0.5mL PCR tubes
- Transfer 3µL of each filled-in reaction to its corresponding 0.5mL PCR tube and mix by pipetting up and down
- Transfer 1µL of selected indexing adapter to each 0.5mL PCR tube and mix by pipetting up and down; use a different index for each sample, starting with B001 and increasing sequentially
- Remove samples from ancient DNA lab immediately for PCR amplification

---

PART 3: PCR AMPLIFICATION

- Perform PCR amplification according to the following cycling profile:

95ºC	5 minutes
95ºC	15 seconds
60ºC	30 seconds
68ºC	30 seconds
68ºC	5 minutes
4ºC	hold

12x

- Remove samples from thermal cycler and take to modern DNA lab; note that samples cannot enter any ancient DNA-dedicated space from this point forward

---

PART 4: PURIFICATION OF PCR PRODUCTS

Note: this must take place in modern lab facility

- Take out the necessary reagents from the Qiagen MinElute PCR Purification kit: PB (binding) buffer, PE (wash) buffer, and EBT (elution) buffer
- Take out one MinElute tube with silica spin column for each sample from the 4ºC refrigerator and label one for each sample
- Add 125µL of PB binding buffer to each MinElute tube and then add the 25µL amplified sample; to ensure total sample volume is collected, samples may need to be spun down with quick spin in mini-centrifuge prior to being added to MinElute tube
- Centrifuge for 1 minute at 13,300rpm on mini-centrifuge
- Discard PB buffer flow-through collected at bottom of MinElute tube
- Add 750µL PE wash buffer to each MinElute tube
- Centrifuge for 1 minute at 13,300rpm on mini-centrifuge
- Discard PE buffer flow-through collected at bottom of MinElute tube
- Centrifuge again (dry-spin) for 1 minute at 13,300rpm to remove any residual wash buffer
- Remove the silica spin column from the MinElute tube and place in a sterile 1.5mL tube
- Add 20µL of EBT elution buffer directly to the silica membrane of the MinElute tube
- Incubate for 1 minute at room temperature
- Centrifuge for 1 minute at 13,300rpm on mini-centrifuge
- Remove MinElute spin columns from 1.5mL tubes, and discard spin columns, keeping only the eluate at the bottom of the tubes
- Place samples in freezer maintained at -20ºC until needed for sequencing

APPENDIX C: OLIGO SEQUENCES

Name	Sequence
IS1_adapter.P5	a*c*a*c*TCTTTCCCTACACGACGCTCTTCCg*a*t*c*t
IS2_adapter.P7	g*t*g*a*CTGGAGTTCAGACGTGTGCTCTTCCg*a*t*c*t
IS3_adapter.P5+P7	a*g*a*t*CGGAa*g*a*g*c
IS4	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCT ACACGACGCTCTT