Distribution Agreement

In presenting this dissertation as a partial fulfillment of the requirements for an advanced degree from Emory University, I agree that the Library of the University shall make it available for inspection and circulation in accordance with its regulations governing materials of this type. I agree that permission to copy from or to publish this dissertation may be granted by the professor under whose direction it was written or, in his absence, by the Dean of the Graduate School when such copying or publication is solely for scholarly purposes and does not involve potential financial gain. It is understood that any copying from or publication of this dissertation that involves potential financial gain will not be allowed without written permission.

____

Lynn Y. Huynh

____

Date

The population genetics of a chromosomal inversion linked to social behavior in the white-throated sparrow (*Zonotrichia albicollis*)

by
Lynn Y. Huynh
Doctor of Philosophy

Graduate Division of Biological and Biomedical Sciences
Program in Population Biology, Ecology and Evolution

____
James W. Thomas, Ph.D.
Advisor

____
Kelly A. Dyer, Ph.D.
Committee Member

____
Donna L. Maney, Ph.D.
Committee Member

____
Todd A. Schlenke, Ph.D.
Committee Member

____
Yun Tao, Ph.D.
Committee Member

Accepted:

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

____
Date

The population genetics of a chromosomal inversion linked to social behavior in the white-throated sparrow (*Zonotrichia albicollis*)

by
Lynn Y. Huynh
B.S., Northern Arizona University, 2003

Advisor: James W. Thomas, 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
Graduate Division of Biological and Biomedical Science
Population Biology, Ecology and Evolution
2010

Abstract

The population genetics of a chromosomal inversion linked to social behavior
in the white-throated sparrow (*Zonotrichia albicollis*)

by
Lynn Y. Huynh

The field of population genetics broadly addresses how genetic diversity in natural populations is generated, shaped and maintained. In this dissertation, I use a population genetic approach to understand the evolution of a chromosomal inversion that influences coloration, social behavior and mate choice in the white-throated sparrow (*Zonotrichia albicollis*). White-striped sparrows (WS) are heterozygous for the ZAL2^m inversion and are more territorially and sexually aggressive than tan-striped sparrows (TS). TS do not have the inversion, invest more in parental care and are less likely to seek extra-pair copulations. The plumage morphs occur at approximately equal frequency and the ZAL2^m inversion is maintained in the population through an exceptionally strong pattern of disassortative mating where almost all breeding pairs are TS × WS. We used targeted sequencing to survey SNPs on the standard chromosome, ZAL2, and the ZAL2^m. We found that the inversion completely suppresses recombination between the homologous chromosomes, except in the short distal region outside of the polymorphism. This results in exceptional linkage disequilibrium, genetic structure and high divergence, and suggests that the ZAL2^m is a rare example of a long-term balanced polymorphism. To understand these patterns in the context of the rest of the genome, we sequenced loci from other autosomes, as well as the ZW sex chromosomes. We describe a strong negative correlation between genetic diversity and chromosome size, which is highly varied in avian genomes. Genetic variation is greatly reduced on the ZAL2 and ZAL2^m chromosomes relative to other similarly sized chromosomes and sex chromosome variation is also exceedingly low. We hypothesize that the low diversity observed for the ZAL2/ZAL2^m and sex chromosomes results from the increased sensitivity to natural selection and genetic drift associated with regions of low recombination. Finally, we identified and characterized polymorphisms in candidate genes that could underlie the social behaviors associated with the inversion. In summary, our data offer insights into the molecular evolution and maintenance of inversions, empirical evidence supporting the relationship between chromosome size and diversity and a set of candidate polymorphisms that could influence social behavior the white-throated sparrow.

The population genetics of a chromosomal inversion linked to social behavior in the white-throated sparrow (*Zonotrichia albicollis*)

by
Lynn Y. Huynh
B.S., Northern Arizona University, 2003

Advisor: James W. Thomas, 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
Graduate Division of Biological and Biomedical Science
Population Biology, Ecology and Evolution
2010

Acknowledgements

My biggest thank you is to my mentor, Jim Thomas, who gave constant support and whose scientific curiosity, careful methods and endless patience inspires me to be more like him.

My research greatly benefited from the thoughtful guidance of my committee members. Kelly supported me with boundless enthusiasm and practical advice. Donna enriched my research with her expertise in hormones and behavior. Todd found the time and patience to teach me population genetics. And Yun, whose office doors were always open, was willing to talk science at all times.

I thank the past and present members of the Thomas lab for their help, guidance, support and friendship. Thank you to the Ford Foundation for their support through a Pre-doctoral Fellowship. I am also grateful for constant presence of all my friends who helped me celebrate every failure and success, both big and small, in science and in life. Thank you.

Thank you to Jack who cheered me on every day, good and bad. Thank you to my big sister, Cathy, who is my constant friend and confidante. Finally, I extend my deepest gratitude to my parents, Hoi and Yen, who taught me the value of an education and hard work. It is because of their unwavering support and selfless sacrifice that I am able to pursue my dreams.

Table of Contents

1 Introduction

Introduction / 2

2 Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*)

2.1 Introduction / 40

2.2 Materials and methods / 45

2.3 Results / 48

2.4 Discussion / 58

3 Contrasting population genetic patterns within the white-throated sparrow genome (*Zonotrichia albicollis*)

3.1 Introduction / 68

3.2 Materials and methods / 73

3.3 Results / 76

3.4 Discussion / 87

4 Characterization of sequence variants in candidate genes linked to behavior in the white-throated sparrow (*Zonotrichia albicollis*)

4.1 Introduction / 96

4.2 Materials and methods / 104

4.3 Results / 105

4.4 Discussion / 106

5 Discussion

Discussion / 108

6 References

References / 118

7 Appendix

7.1 Supplemental Table 1 / 155

7.2 Supplemental Table 2 / 156

7.3 Supplemental Table 3 / 161

7.4 Supplemental Table 4 / 162

7.5 Supplemental Table 5 / 163

Index of Figures

1.1 The alternate plumage morphs of the white-throated sparrow. / 4

1.2 Hill-Robertson interference. / 12

1.3 Mechanisms for inversion formation. / 20

1.4 The formation of an inversion loop in the heterokaryotype. / 21

2.1 ZAL2 and $ZAL2^m$ haplotype networks. / 52

2.2 Linkage disequilibrium across the $ZAL2/ZAL2^m$ chromosomes. / 55

3.1 Average diversity across the sparrow genome. / 79

3.2 F_{ST} values between TS and WS morphs across the genome. / 82

3.3 Tajima's D values across the sparrow genome. / 83

3.4 Linkage disequilibrium patterns within the sparrow genome. / 86

4.1 Predicted locations of our nine candidate genes shown on the ZAL2 chromosome orientation. / 100

Index of Tables

2.1 Diversity, divergence and summary statistics for the standard ZAL2 and inverted ZAL2^{m} chromosomes in the white-throated sparrow. / 49

2.2 Mutations occurring along the ZAL2 and ZAL2^{m} lineages. / 58

3.1 Sampling information and diversity values for autosomal and sex-linked loci compared to previously reported values for ZAL2 and ZAL2^{m} alternative chromosome arrangements. / 77

4.1 Behavioral differences observed in the white- and tan-striped birds. / 97

4.2 Candidate gene characteristics. / 100

4.3 Sequencing results of SNPs from candidate genes within the ZAL2^{m} inversion. / 106

1 Introduction

"The value and utility of any experiment are determined by the fitness of the material to the purpose for which it is used, and thus in the case before us it cannot be immaterial what plants are subjected to experiment and in what manner such experiment is conducted."

Gregor Mendel

Experiments in Plant Hybridization (1866)

The foundation of the science of genetics is based on the experiments of Gregor Mendel, who studied contrasting traits in pea plants. For years, Mendel tracked stem, seed, pod and flower characters in careful crosses of tens of thousands of plants and with these data he established the basic principles of inheritance (Mendel, 1866). Classic studies in the field of genetics mimicked those experiments of Mendel, focusing on tracking mutant genes through generations of laboratory crosses (Sturtevant, 1965). Today, we have the necessary technological advances, wealth of molecular tools and population genetic theory that make it possible to study genetic variation in virtually any organism, inside and outside of the laboratory, without the need for controlled crosses, mutant genes or previous genetic study. Thus, we can choose to study any species of interest and take advantage of systems that may be unconventional, but well-suited for addressing particular evolutionary questions.

This dissertation focuses on such a system. The white-throated sparrow, *Zonotrichia albicollis*, is a migratory North American songbird with two plumage morphs that can be

distinguished by the color of their crown stripes: white-striped (WS) or a tan-striped (TS) (Lowther 1961, Figure 1.1). Coloration is perfectly correlated with the presence of a large chromosomal inversion on chromosome 2, called the ZAL2^{m}. WS birds have at least one copy of the inversion and are almost always inversion heterozygotes. In two instances, WS inversion homozygotes have been found, but they are exceedingly rare occurring in approximately one-tenth of a percent of the population (Falls and Kopachena 1994; Romanov *et al.*2009; Thorneycroft 1975; D. Maney personal communication). TS birds are obligate ZAL2 homozygotes (Thorneycroft, 1975; Falls & Kopachena, 1994; Romanov *et al.*, 2009). The ZAL2^{m} inversion is also linked to differences in social behavior and mate choice of the morphs (reviewed in Tuttle 2003). TS birds adopt a parental strategy and WS birds adopt a competitive strategy that includes increased aggression and mating behaviors (Kopachena & Falls, 1993a; Kopachena & Falls, 1993c; Collins & Houtman, 1999; Tuttle, 2003). The inversion is maintained at a stable frequency in the population by a strong pattern of disassortative mating, where almost all breeding pairs are discordant (WS × TS) and produce half WS and half TS offspring (Lowther, 1961; Falls & Kopachena, 1994). As a result, the ZAL2^{m} inversion is maintained at a population frequency of ~25% from generation to generation and is usually seen in a heterozygous state (Thorneycroft, 1975).

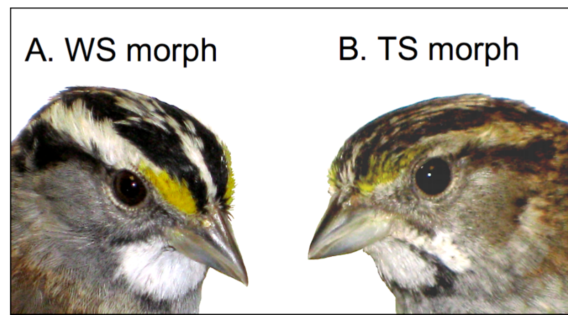

Photo: Allison Reid

Figure 1.1 The alternate plumage morphs of the white-throated sparrow. A. The white-striped morph. B. The tan-striped (TS) morph.

In the first modern genetic characterization of the ZAL2^m polymorphism, Thomas *et al.* identified the presence of two nested inversions that appeared to suppress recombination (Thomas *et al.*, 2008). As the ZAL2^m encompasses ~10% of the white-throated sparrow genome, it is reasonable to suspect that patterns of recombination and related genetic phenomena are dissimilar within the sparrow genome (Thomas *et al.*, 2008). Additionally, avian genomes generally show great variation in chromosome size, which is negatively correlated to recombination rate, providing another reason to believe that the genetic architecture of the white-throated sparrow may be unusual in comparison to the traditional animal model systems, like flies, worms and mice (Hillier *et al.*, 2004).

The white-throated sparrow genome has several properties of interest to population geneticists: the presence of an inversion linked to behavior, the maintenance of that inversion in an almost constant heterozygous state and the extreme variation in chromosome size (Thorneycroft, 1966; Thorneycroft, 1975). Thus, it is a convenient model to directly address long-standing questions in genetics regarding the evolution and

genetic forces that establish and maintain inversions, as well as questions regarding how recombination influences the sequence diversity and genetic architecture of a species. These data can help us resolve the events in a species' evolutionary history, understand how different evolutionary forces shape the genetic landscape and document their relative importance. Overall, studying the population genetic patterns in the white-throated sparrow genome can help elucidate the biological process of evolution at the molecular level. Finally, the linkage of the inversion to several biologically relevant behavioral phenotypes presents a unique opportunity to elucidate the genetic factors underlying social behavior.

What determines genetic diversity in a genome?

Patterns of genetic diversity are shaped by many factors: mutation, selection, genetic drift, population structure and gene flow (Pritchard & Przeworski, 2001). The ultimate source of genetic variation is mutation and in the most simple population genetic model, the neutral model, the fate of a new variant (loss or fixation) is solely determined by random genetic drift (Kimura, 1983). Random genetic drift is expected to play a large role when effective population sizes (N_e) are small, but in populations of large N_e, natural selection becomes more prominent and can act to eliminate or increase the frequency of a variant depending on its selective value (Kimura, 1983). Finally, like drift and selection, gene flow can change allele frequencies within a population by the simple introduction or removal of variants via immigration or emigration. Gene flow can create population structure between groups of individuals, but it can also be used as a tool for understanding structure within a genome (Borge *et al.*, 2005).

Maintenance of genetic variation by balancing selection

Studies of natural populations have identified significant amounts of genetic variation and population geneticists have often invoked balancing selection as a primary force in maintaining genetic diversity in certain genomic regions (Richman & Kohn, 1999; Bamshad *et al.*, 2002; Raymond *et al.*, 2005; Clark *et al.*, 2007; Ferrer-Admetlla *et al.*, 2008). The three primary mechanisms of balancing selection are overdominance (i.e. heterozygote advantage), frequency dependent selection and variation in selection over time and space. Regardless of the mechanism, all forms of balancing selection result in increased allelic diversity at the site under selection (Charlesworth, 2006).

The underlying principle of overdominance is that an allele confers increased fitness only in the heterozygous state. Overdominant alleles can never increase to fixation and it is unlikely to be lost (unless by genetic drift) due to positive selection for heterozygous carriers. The classic case of overdominance is that of sickle-cell anemia, a common hereditary disorder caused by a recessive mutation that occurs at a high frequency in tropical and subtropical human populations (Allison, 1956). The disease is only observed in individuals homozygous for a mutant allele that disrupts hemoglobin function; however, in heterozygous individuals it confers resistance to malaria, resulting in a strong fitness advantage. Thus, the allele is maintained in the population, in spite of its deleterious effects in homozygous individuals. Although this example of overdominance illustrates how variation can be maintained at the population level, examples like these

are actually very rare and the role of overdominance in the maintenance of population level genetic diversity remains unclear, but is likely to be marginal (Fry, 2004).

In frequency dependent selection, the relative fitness of particular genotypes is a function of their frequency in the population relative to other genotypes. Positive frequency dependent selection occurs when the fitness of a genotype increases as it becomes more common and negative frequency dependent selection occurs when the fitness of a genotype is greatest when it is rare (Ayala & Campbell, 1974). Both processes have been shown to maintain stable genetic variation at the population level (Hori, 1993; Sinervo & Lively, 1996; Smithson & Macnair, 2008).

One example of negative frequency dependent selection occurs in *Loxia* cross-bills (Family: Fringillidae) (Benkman, 1996). The characteristic mandible crossing is an adaptation that facilitates access to seeds enclosed in conifer cones. When perch sites are limited and cones cannot be turned or removed, an individual crossbill can only access the seeds on the side of the cone opposite to the side that the lower mandible crosses. As a result, in a population that is monomorphic with respect to bill crossing, a rare variant of the opposite morph is favored by selection because they are uniquely capable of foraging on the seeds that cannot be accessed by the dominating morph. This process of negative frequency dependent selection has been posited as an explanation for the maintenance of both morphs at equal frequency in *Loxia* (Benkman, 1996). Another, very similar example can be found in scale-eating cichlids (Hori, 1993). The common factor in these cases is the complementary exploitation of limited resources. Although

the genetics of these systems have yet to be resolved, self-incompatibility in plants is a form of balancing selection that has been widely investigated at the molecular level (Takayama & Isogai, 2005).

The self-incompatibility loci (SI) in flowering plants and the vertebrate major histocompatibility complex (MHC) are among the best examples of balancing selection (Hughes & Nei, 1988; Kamau & Charlesworth, 2005; Charlesworth, 2006). They also represent the balanced polymorphisms for which we have the greatest understanding at the molecular level. Self-incompatibility is a common genetic mechanism that promotes outbreeding in flowering plants by enabling the pistil to recognize and reject pollen from genetically related individual (Takayama & Isogai, 2005). New alleles will be rare in the population and will have a fertility advantage because pollen bearing them will not be rejected by recipients. In large populations, this frequency dependent selection will favor new alleles until an equilibrium is reached and this process makes SI loci extremely polymorphic (Charlesworth & Guttman, 1997; Charlesworth, 2006).

SI loci are the most highly polymorphic loci known in plants and in vertebrates, the MHC loci are the famous example of extreme genetic diversity (Hughes & Yeager, 1998). The MHC proteins are key factors in vertebrate immune response because they encode cell surface proteins that enable recognition of self and non-self antigens. MHC heterozygosity is expected to confer increased resistance to pathogens as heterozygous carriers are capable of immune recognition of a greater diversity of foreign antigens

(Hughes, 2007). Furthermore, maintenance of diversity is favored by disassortative mating according to MHC genotype (Hughes & Yeager, 1998).

Molecular signatures of balancing selection

At the molecular level, balanced polymorphisms should show increased rate of nonsynonymous substitution as selection and this has been reported for both SI and MHC loci (Hughes & Nei, 1988; Charlesworth & Awadalla, 1998). Additionally, increased genetic diversity is observed at the site under selection as well as linked sites (Charlesworth, 2006). SI regions have been reported to harbor extremely high genetic diversity and in some cases, this diversity extends to neighboring regions due to linkage disequilibrium (Charlesworth *et al.*, 2003; Kamau & Charlesworth, 2005; Kamau *et al.*, 2007; Votintseva & Filatov, 2009). The MHC loci also show extraordinary diversity relative to other regions in the genome (Hughes & Nei, 1988; Raymond *et al.*, 2005). The maintenance of increased genetic diversity is related to another signature of balancing selection, which is a skew in the site frequency spectrum reflecting an excess of common variants (Hughes & Yeager, 1998; Bamshad *et al.*, 2002; Bamshad & Wooding, 2003; Charlesworth, 2006). Finally, because polymorphisms can persist for very long times under balancing selection, alleles under balancing selection will have common ancestors further back than other areas of the genome (Hughes & Yeager, 1998; Charlesworth, 2006). This has been observed both the cases of MHC and for SI loci (Schierup *et al.*, 2001).

With regard to the $ZAL2^m$ inversion polymorphism in the white-throated sparrow, the disassortative mating can be viewed as a type of balancing selection that works to maintain the inversion itself. The effects of balancing selection between the $ZAL2/ZAL2^m$ will maintain genetic variation and potentially result in longer coalescence times to the most recent common ancestor of these chromosomal rearrangements. Under balancing selection, it can be expected that increased heterozygosity and skew of site frequency spectrum will be associated with the $ZAL2/ZAL2^m$ system, in addition to signatures of other forms of selection acting upon the inversion.

Recombination and genetic diversity

Another evolutionary force known to influence the amount of genetic and phenotypic diversity is recombination. Meiotic recombination is a powerful force in generating novelty as it can bring together independently arising beneficial mutations or segregate alleles with contrasting fitness effects, thus facilitating the efficacy of selection. Although recombination is widespread in eukaryotes, the rate at which recombination occurs varies widely across species, between sexes and among different genomic regions (McVean *et al.*, 2004). This variation can have a significant and complex impact on the structure of genetic variation within a genome.

In one of the most widely cited papers in molecular evolution, Begun and Aquadro (1992) demonstrated the positive relationship between local rates of recombination and genetic diversity in *Drosophila melanogaster*. They proposed a selective explanation for this phenomenon and hypothesized that the fixation of beneficial alleles in regions of low

recombination are likely to remove variation from a larger sequence window. In contrast, when selective sweeps occur in regions of high recombination, the affected sequence window will be smaller. Overall, fixation of mutations will reduce the level of standing variation in the surrounding sequence (Maynard Smith & Haigh, 1974). This model of genetic hitchhiking is illustrated in Figure 1.2A. Charlesworth (1993) proposed another selective explanation for this position relationship. In the background selection model, negative selection against deleterious alleles will also remove linked neutral variants. In regions of low recombination, the window of linked sites is larger than in regions of high recombination. This model of background selection is illustrated in Figure 1.2B.

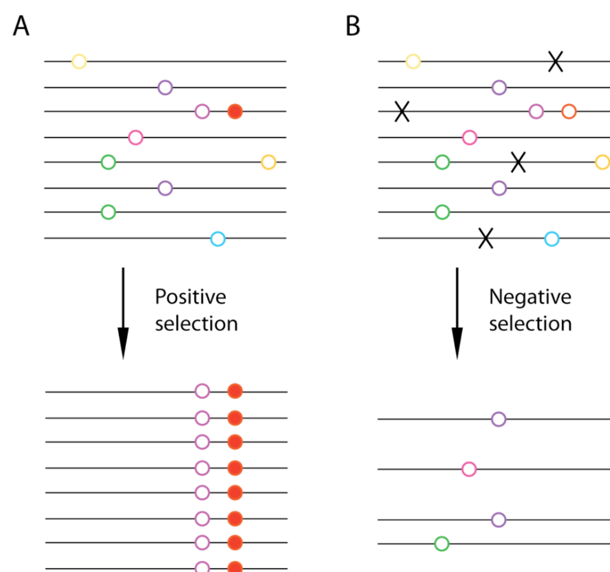

Figure 1.2 Hill-Robertson interference. In the absence of recombination, all sites within a segment are linked and positive and negative selection both result in a decrease in genetic diversity. A. During a selective sweep, selection acts to fix the beneficial (closed red circle) at the population level. Any linked neutral variant (open circles) will also fix. B. Under background selection, deleterious alleles (marked by '×') will be selectively removed from the population along with any linked neutral variant (open circles).

Hill-Robertson interference

The length of the genetic tract influenced by selective sweeps and background selection is dependent on the rate of recombination. As recombination rate increases, the length of the sequence tract in linkage disequilibrium with a target of selection is shorter and natural selection becomes increasingly able to act upon a single target. In contrast, regions with low recombination have larger LD blocks and interference can occur when a selective event at one site determines the fate of the entire LD block. The reduction in the efficacy of selection associated with lowered recombination was used by Hill and Robertson (1966) to argue for the evolutionary advantage of recombination (or sex). Even when selection coefficients are small, Hill-Robertson interference can have a considerable effect across a genome, reducing the level of polymorphism as well as codon usage bias (McVean & Charlesworth, 2000). Thus, Hill-Robertson interference is likely to be an important evolutionary force, especially in the evolution of non-recombining sequences and genomes.

Recombination and mutation as linked processes

An alternative and neutral explanation for the positive relationship between recombination rate and genetic diversity that posits that recombination itself is mutagenic. Esposito and Bruschi (1993) empirically demonstrated that DNA lesions at the LEU1 locus in *Saccharomyces cerevisiae* were both recombinagenic and mutagenic. If mutation and recombination were associated processes, values of both intraspecific differences (polymorphism) and interspecific differences (divergence) would be

positively correlation. In multiple studies, no correlation between recombination rate and divergence has been observed, giving support for a selective explanation for the relationship between recombination and genetic diversity (Begun & Aquadro, 1992; Stephan & Langley, 1998; Roselius *et al.*, 2005). On the other hand, studies in humans have found a positive relationship between recombination rate, diversity and divergence, supporting a relationship between recombination and mutation (Filatov & Gerrard, 2003; Hellmann *et al.*, 2003; Duret & Arndt, 2008).

Although the underlying processes are still a widely debated topic, the positive correlation between recombination and genetic diversity has been observed in a wide-range of organisms, including *Drosophila* spp., *C. elegans*, humans, yeast, tomatoes, corn and mice (Begun & Aquadro, 1992; Kliman & Hey, 1993; Kraft *et al.*, 1998; Stephan & Langley, 1998; Nachman, 2001; Betancourt & Presgraves, 2002; Cutter & Payseur, 2003; Hellmann *et al.*, 2003; Tenaillon *et al.*, 2004; Noor, 2008). It is important to note that these studies have focused on rates of recombination ranging from the kb scale to several Mb; however, recombination rates can vary at a larger scale and it is not yet clear whether the processes that produce the positive correlation between recombination rate and diversity extend to larger scales. The underlying processes generating the relationship between recombination and diversity is a topic of significant importance to population genetics and evolutionary biology as the impact of natural selection on genome evolution is dependent on the rate of recombination.

Recombination deserts, recombination jungles and the avian genome

Avian genomes are likely to be interesting models for understanding the large-scale relationship between recombination rate and diversity. A distinctive feature of the avian genome is the substantial variation in chromosome size, which spans nearly two orders of magnitude (Hillier *et al.*, 2004; Warren *et al.*, 2010). Avian chromosome size negatively correlates with average rate of recombination (Hillier *et al.*, 2004; Backström *et al.*, 2010) and the increased rate of recombination observed for smaller chromosomes is hypothesized to be a consequence of the obligatory cross-over event per chromosome arm during meiosis (Rodionov, 1996). Considering this relationship, it is reasonable to expect that genetic diversity among avian chromosomes is also substantially varied and that diversity would be greatest on the microchromosomes where recombination is the greatest (Ellegren, 2007).

In chicken, Fang *et al.* (2008) reported that the correlation of recombination rate and diversity was observed at the scale of <10 Mb but not at the scale of whole chromosomes. Although the authors did not speculate on the reason behind this, the data are not consistent with predictions made by the hypothesis that recombination is mutagenic. If the increased sequence variation in highly recombining regions is due to the mutagenic properties of recombination, then these effects would be observed at both the local and chromosomal level and microchromosomes should show higher divergence than the macrochromosomes. In a study of chickens and turkeys, Axelsson *et al.* (2005) reported the opposite pattern: a lower rate of divergence on microchromosomes. These data lend support to the role of selection in generating the positive correlation between

recombination rate and genetic diversity; however, it should be noted that the lower rate of divergence could reflect the increased evolutionary constraint on microchromosomes, which are more gene dense than macrochromosomes (Smith *et al.*, 2000). Clearly, there are many factors to be considered and currently, the data addressing these questions are currently limited.

Recent studies of the zebra finch genome suggest that the relationship between chromosome size and recombination rate is complex. Backström *et al.* (2010) reported large regions (10 – 40 Mb) of recombination deserts and jungles in the macrochromosomes. Furthermore, recombination greatly increases towards the telomeres, making all chromosome ends recombination jungles. As the avian microchromosomes are <20 Mb long, they are characterized as recombination jungles. Although macrochromosome telomeres show rates of recombination akin to those observed on microchromosomes, their average rate of recombination is dominated by the presence of recombination deserts as large as 100 Mb. Backström *et al.* (2010) did not report genetic diversity or divergence in these regions, so the relationship between recombination rate and genetic diversity, at the local scale or otherwise, remains to be established in avian genomes.

Overall, the differences in recombination frequency observed between and within avian chromosomes can dramatically influence conclusions drawn from population genetic studies. For example, in a study of nucleotide variation and LD in zebra finch, Balakrishnan and Edwards (2009) observed exceptionally high nucleotide diversity in

wild zebra finch populations but the chromosomal locations of the loci included in the study were not reported. Considering their 50-fold difference in diversity between loci, any bias in sampling from one chromosome size class could influence estimates of diversity, N_e and the understanding of the genetic architecture of the species. In future avian population genetic studies, it may become necessary to interpret data in light of the chromosomal location of sequence data.

Hill-Robertson interference between the avian mitochondria and W chromosomes

The influence of recombination on sequence diversity can most clearly be seen in the case of sex chromosomes. In ZW (female heterogametic) systems, recombination is suppressed over the majority of W chromosome and Hill-Robertson is expected to play a large role in the evolution of the W. Genetic variation on the W chromosome in chicken and seven other bird species is more than 10-fold lower than expected, even after accounting for reductions caused by sex-specific mutation rates and differences in effective population size (Montell *et al.*, 2001; Berlin & Ellegren, 2004). Berlin *et al.* (2007) hypothesize that the extremely low diversity of the W chromosome is a result of Hill-Robertson interference from the mitochondria. As both avian mitochondria and the W chromosome are non-recombining and genetically linked (both are passed from mother to daughter), selection in the mitochondrial genome is expected to influence the W chromosome and vice versa. Although the chicken W chromosome has been found to have exceedingly low genetic diversity (Berlin & Ellegren, 2004), Hill-Robertson interference is not the only explanation. A population genetic bottleneck could have occurred during domestication and would be expected to have more dramatic effects on

sex chromosomes than other regions of the genome (Lindgren *et al.*, 2004). Surprisingly, a SNP survey of the chicken genome did not support the genome-wide loss of diversity that would be expected with population bottleneck (Wong *et al.*, 2004). The linkage of the W chromosome and the mitochondrial genome is interesting, but future studies of both W and mitochondrial diversity, perhaps in other species with female heterogamy such as butterflies, will be necessary to establish the influence and importance of this interaction in reducing genetic diversity.

Recombination, genetic diversity and sex chromosomes

Relative to autosomes, both the Z and the W sex chromosomes show reduced recombination (the W is non-recombining and Z chromosome recombination is primarily limited to the male sex, except in the pseudoautosomal regions). Not surprisingly, both the Z and W have reduced genetic variation. The reduction in diversity expected considering the differences in N_e of autosomes versus sex chromosomes, which predicts that variation of the Z and W will be three-fourths and one-fourth that of autosomes, respectively (Ellegren, 2009a). Studies of Z chromosome diversity report that variation on the Z is one-third that of autosomes (Sundström *et al.*, 2004; Borge *et al.*, 2005). This reduction in diversity could reflect the sensitivity of the Z chromosome to selective sweeps that reduce polymorphism levels because of sexual selection. As the Z is passed directly from father to son, sexual selection for male-specific alleles is increased (unlike the X which is strictly passed to daughters from fathers, Sundström *et al.*2004; Borge *et al.*2005). Evidence for the rapid evolution of Z-linked genes supports the role of

selection in reducing genetic diversity on the Z chromosome (Mank *et al.*, 2007; Ellegren, 2009b).

Genetic degeneration of non-recombining regions

A key consequence to the reduction or loss of recombination is the genetic degeneration associated with the limited power of selection. Studies of non-recombining genomes (Goddard *et al.*, 2005; Dolgin & Charlesworth, 2008; Lockton & Gaut, 2010), non-recombining genomic segments (Silver & Artzt, 1981; Dyer *et al.*, 2007) and sex and neo-sex chromosomes (Rice, 1994; Fridolfsson & Ellegren, 2000; Bachtrog & Charlesworth, 2002; Peichel *et al.*, 2004; Graves, 2006; Kaiser & Charlesworth, 2010) have overwhelmingly demonstrated the progression from accumulation of repetitive elements and deleterious mutations to dramatic physical loss of genetic material. High male-biased mutation rates contribute to the rapid degeneration of Y chromosomes (Aitken & Graves, 2002), and limited recombination makes the Y and the W chromosomes susceptible to degeneration via Hill-Robertson interference (reviewed in Charlesworth and Charlesworth 2000). Another consequence of suppressed recombination is the considerable reduction of W chromosome N_e, which makes the W chromosome especially sensitive to random genetic drift. When the least mutationally burdened W chromosome is lost by drift, it can never be recovered. The action of Muller's Ratchet is irreversible in a non-recombining population and with every click of the ratchet, the W chromosome degrades (Charlesworth & Charlesworth, 2000; Bachtrog, 2006; Graves, 2006).

Factors that modify recombination

Recombination is clearly a powerful evolutionary force shaping patterns of genetic diversity across a genome and it has long been known that cross-overs are not randomly distributed across the genome (i.e. McVean *et al.*2004). A large body of work has addressed the genic and structural causes of recombination rate differences. Recently, three research groups independently reported the role of the Prdm9 gene and a 13-mer DNA motif in recombination hotspots (Baudat *et al.*, 2010; Myers *et al.*, 2010; Parvanov *et al.*, 2010). Although identifying the genic causes of recombination are a major breakthrough in our understanding the non-random distribution of recombination, we have long known about the capability recombination modification of structural changes, especially chromosomal inversions.

Chromosomal inversions

Chromosomal inversions were among the earliest studied genetic markers and were first discovered because of their unique capability to reduce cross-overs within the inverted region as well as in flanking sequence in inversion heterokaryotypes (Sturtevant, 1917; Sturtevant, 1921). A chromosomal inversion is a balanced rearrangement that modifies gene order such that an inverted segment is rotated 180° with respect to its previous standard arrangement. Inversions can occur via two primary processes illustrated in Figure 1.3, breaking and rejoining (Figure 1.2A) or crossing-over between inverted repeats (Figure 1.2B). A third mechanism for inversion formation was recently discovered, in which mobile genetic elements (LINE-1 and Alu) provide fragile sites for chromosome breakage or act as the inverted repeats causing inversions (Lee *et al.*, 2008).

Once an inversion is formed, the relative location of the genes within the inversion is modified with reference to the previous arrangement. No genes are gained or lost unless it is physically disrupted by an inversion breakpoint.

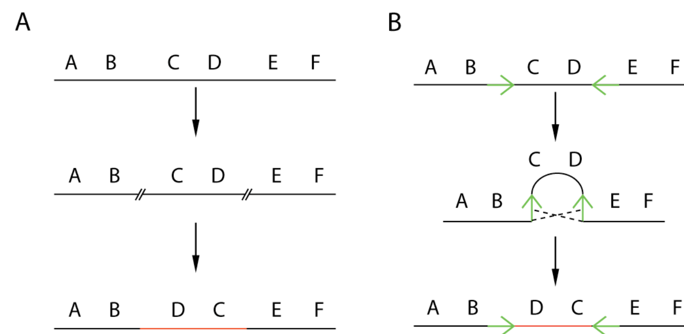

Figure 1.3 Mechanisms for inversion formation. 1.3A An inversion can form when a chromosome experiences two double-stranded breaks and is rejoined in opposite orientation. Alternatively, as shown in Figure 1.3B, inverted repeats, including mobile DNA sequences, can mediate non-allelic homologous crossing over (dashed lines), which can result in an inversion. Note that centromeres are not pictured but inversions can include centromeres (pericentric) or exclude them (paracentric).

Inversions as recombination modifiers

Suppressed recombination between the standard and inverted haplotypes can result from asynapsis (synaptic failure) of the homologous chromosomes during meiosis. Even if the two chromosome arrangements align in an inversion loop, single cross-over events within the inversion interval will produce unbalanced gametes with deleterious duplications and deletions (Stone & Thomas, 1934; Sturtevant & Beadle, 1936). Figure 1.4 illustrates an inversion loop for a single inversion, but when multiple inversions are present, increasingly complex structures can form to accommodate chromosomal alignment

during meiosis (Dobzhansky, 1937). Because recombination is suppressed across the inverted region, gene flow between the standard arrangement and the inverted arrangement is limited. Genetic exchange in heterozygotes between the arrangements can nonetheless occur via double recombination and gene conversion (Navarro *et al.*, 1997). If the inversion reaches appreciable frequencies, recombination can be restored in inversion homozygotes where homologous alignment of homologs is not disrupted.

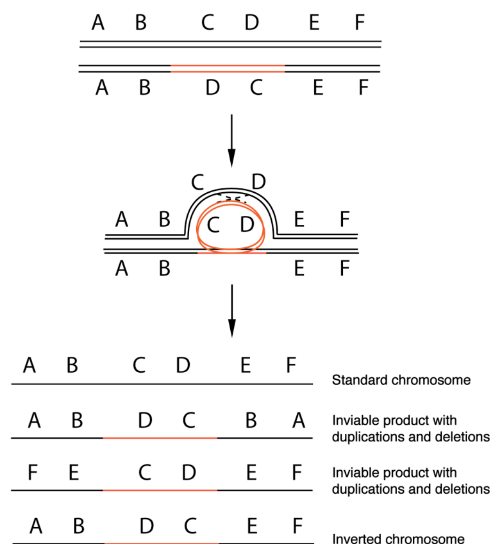

Figure 1.4 The formation of an inversion loop in the heterokaryotype. Single crossing over events (indicated by the dashed lines) within the inversion loop result in recombinants with duplications and deletions. In the case of pericentric inversions, this can result in the gain or loss of a centromere. Increasingly complex structures form in the presence of multiple adjacent and/or overlapping inversions (Dobzhansky, 1937).

Gene flow in the presence of inversions

While studying recombination patterns within Drosophila inversions, Muller reported that the positions of multiple cross-over events associated with the same pair of chromatids

were non-randomly distributed and he termed this phenomenon "cross-over interference" (Muller, 1916). Muller observed that a second cross-over event was more likely to occur in regions farther away from the first cross-over. Thus, genetic exchange due to double recombination events is likely to occur in larger inversions and produce longer tracts of exchanged genetic material (Novitski & Braver, 1954; Navarro *et al.*, 1997). In contrast, gene conversion tracts are generally short, ranging from 2bp to 1kb. Gene conversion will play a larger role in genetic exchange between the alternate arrangements for smaller inversions than crossing-over, which is more likely to facilitate gene flow between alternate arrangements in large inversions (Chovnick, 1973; Navarro *et al.*, 1997; Marais, 2003; Schaeffer & Anderson, 2005).

The age of the inversion as well as the amount of gene flux between alleles from both arrangements will contribute to the LD observed within an inversion. In some *Drosophila* inversions, LD is extreme, spanning >130 cM (Dyer *et al.*, 2007). In contrast, polymorphic inversions in a mosquito, *Anopheles funestus*, show no LD, suggesting that they do not harbor genes involved in coadapted or locally adapted genes (Cohuet *et al.*, 2004). These are two extreme examples of complete LD or lack of LD associated with inversions but for most inversions described in the literature, recombination patterns are intermediate producing variable levels of LD within the region (Schaeffer & Anderson, 2005; Kennington *et al.*, 2007; Machado *et al.*, 2007; White *et al.*, 2007; Nóbrega *et al.*, 2008). In a survey of *D. pseudoobscura* inversions, Schaeffer and Anderson (2005) documented a general pattern of gene flux towards the center of inversions and high levels of differentiation at the breakpoints, consistent with

theoretical predictions (Navarro *et al.*, 1997). Importantly, recombination patterns within an inversion, especially those showing regions of LD interspersed by regions of low association, are key to association mapping of the regions controlling linked phenotypes (Hammer *et al.*, 1991; Schaeffer *et al.*, 2003; Stump *et al.*, 2005; Kennington *et al.*, 2006; White *et al.*, 2007; Wallace & Erhart, 2008). Identifying the genes involved in inversion phenotypes will elucidate the genetic architecture of these traits and can yield insights into the evolution of inversions at the population level.

Classic models of inversion evolution

Several models have been put forth to explain how inversions influence fitness and how they are maintained at the population level. The primary models invoke co-adapted gene complexes, local adaptation or position effects. Both the co-adapted gene complex and local adaptation models rely on the capability of an inversion to modify patterns of recombination; in contrast, the position effects occur independently of the recombination effects of the inversion. Another factor to consider is genetic drift by which inversions can fix in a population or subpopulation, independent of their fitness values and, perhaps, in spite of deleterious effects (Lande, 1984; Lande, 1985). Finally, inversions can invade populations through heterozygote advantage and e maintained at an equilibrium frequency by balancing selection, as explained below.

Coadapted gene complexes

While studying *Drosophila pseudoobscura*, Dobzhansky noted the relative adaptive advantage of inversion heterozygotes to inversion homozygotes in natural and artificial

populations. Dobzhansky proposed that the increased fitness could be attributed to positively interacting alleles captured by the inversion (Dobzhansky, 1937). As inversions are known to reduce crossing-over, linkage between these interacting alleles can be maintained and selection favors the suppression of recombination to maintain the high fitness genotype. The fate of this type of inversion is fixation in the population unless it is initially lost by random genetic drift. Dobzhansky went further to propose that the principle role of inversions in evolution was to preserve favorable genetic interactions or "supergene complexes" through the suppression of recombination in the heterozygote.

Although the coadapted gene complex theory has gained much popularity, true coadapted gene complexes require epistatic interactions that favor the reduction of recombination and there is little evidence so far for co-adapted gene complexes as predicted by Dobzhansky (Dobzhansky, 1937). In most cases of proposed coadapted gene complexes, epistasis has not been directly quantified (van Delden & Kamping, 1989; Schaeffer *et al.*, 2003; Kennington *et al.*, 2006). The one clear exception that meets this stringent condition is meiotic drive. Meiotic drive occurs when the transmission of a gene (the distorter) is enhanced in the gamete pool at the expense of its wild-type homolog. Tight linkage is required between the distorter gene and an insensitive responder gene, and in many cases, additional enhancer genes are also linked. The epistatic interactions are evident when the elements are separated by recombination because the extent of drive is greatly reduced or the distorter becomes "suicidal" because it segregates from the insensitive responder (Hartl, 1974).

Because of the requirement for linkage disequilibrium, it is not surprising that chromosomal inversions have been implicated in several cases of meiotic drive (Hartl, 1975; Lande & Wilkinson, 1999; Lyon, 2003; Mroczek *et al.*, 2006; Dyer *et al.*, 2007). The mouse *t* haplotype is the classic example of meiotic drive and has fascinated geneticists for nearly a century. The *t* haplotypes refer to variant forms of mouse chromosome 17 which have been found in wild house mouse populations worldwide at a frequency of up to 25% (Huang *et al.*, 2001; Lyon, 2003). More than 25 haplotypes have been described, most of which are homozygous lethal (Silver, 1985). Males heterozygous for a *t* haplotype will transmit it to an exceptionally high proportion of offspring, up to 99%. In heterozygous females, no transmission ratio distortion (TRD) is observed. This haplotype influences mate choice and researchers have described female avoidance of heterozygous males (Williams & Lenington, 1993). The intensity of this avoidance is dependent on the female genotype, with heterozygous females demonstrating stronger avoidance than wild-type. Similarly, males show increased aggression towards heterozygous females (Williams & Lenington, 1993). Non-random mating in regards to *t* haplotype is a good example of mate choice based on genetic incompatibility, as matings between females with at least one *t* haplotype and heterozygous males will produce recessive lethal offspring (Brown, 1997; Tregenza & Wedell, 2000). Both non-random mating and homozygous lethality in the male work to maintain the *t* haplotypes at intermediate frequencies, thus, balancing selection is among the primary forces acting upon the *t* haplotypes (Hammer & Silver, 1993).

The original *t* haplotype likely started with several closely linked genes showing TRD. Subsequently, selection favored the linkage of TRD enhancing genes enhancing and the suppression of recombination between those alleles and the wild-type haplotype. Today, we know that four non-overlapping inversions suppress recombination in the heterokaryotype across 30-40Mb, or >1%, of the mouse genome (Lyon, 2003). Although regions of lower recombination are expected to show lower diversity, the mouse *t* haplotypes show reduced exceedingly reduced genetic variation considering the age of the inversion and the similar rates of evolution between the *t* and wild-type haplotypes. These data suggest that selection among different *t* haplotypes has resulted in reduced diversity. It is reasonable to believe that positive selection has strongly shaped the evolutionary history of the *t* haplotypes as a result of the evolution of TRD suppressors (Hammer & Silver, 1993). Additionally, modifiers that enhance TRD are likely to quickly sweep to fixation and molecular dating estimates support the step-wise evolution of the *t* haplotypes (Silver, 1993).

Suppression of recombination between the *t* and wild-type haplotypes is increasingly important in order to maintain the capacity for TRD. The presence of four adjacent inversions is likely to greatly inhibit gene flow between the *t* and wild-type haplotypes (Hammer *et al.*, 1989). Although the inversions suppress recombination >100-fold in the *t* haplotypes, surveys in natural populations report that gene conversion occurs between the *t* haplotype and the wild-type chromosomes, as well as between *t* haplotypes that have complementary lethal mutations (Lyon, 1984; Silver & Remis, 1987; Silver, 1993; Wallace & Erhart, 2008). Genetic analysis of partial haplotypes has enabled localization

of candidate genes and demonstrated that at least 5 alleles are required for full TRD (Lyon, 1984; Silver & Remis, 1987). Even infrequent recombination can be a useful tool for mapping genes within inversions.

Local adaptation

There is an alternative selective hypothesis that does not require epistatic interactions. Under the local adaptation model, an inversion can establish itself in a population if it simply captures two or more alleles that increase fitness independently. The conditions for establishment under the local adaptation model are less stringent than coadaptation because the local adaptation model can apply to many combinations of alleles, unlike coadaptation, which requires specific interactions between alleles. An inversion capturing multiple locally adapted alleles will outcompete other haplotypes that have fewer locally adapted alleles. Even if other haplotypes contain all of the advantageous alleles, in the absence of recombination suppression, their association will eventually be lost. As in the coadaptation model, the selective advantage in the local adaptation theory rests in the suppression of recombination caused by the inversion.

Kirkpatrick and Barton's local adaptation argument is wide-reaching and a valid consideration for many examples of inversions showing adaptive phenotypes. Recently, Manoukis *et al.* (2008) suggested that local adaptation played a role in the incipient speciation between differentially adapted "ecotypes" of the mosquito *Anopheles gambiae* (Manoukis *et al.*, 2008). The 2Rj inversion in *A. gambiae* confers adaptation to different larval breeding environments and is found as a fixed difference between *A. gambiae*

subpopulations in peripheral habitats. The empirical data from the 2Rj inversion is consistent with the parameters and predictions of the local adaptation model: locally favored alleles, suppression of recombination causing LD within the inversion, establishment of an inversion in a peripheral population and fixation of an underdominant inversion. Thus, it is a likely example of an inversion that arose through selection for local adaptation (Manoukis *et al.*, 2008).

Position effect

Barring the disruption of a gene with a chromosomal breakpoint, the genes within inversions remain unaffected, but their position relative to genes and regulatory elements outside the inversion changes. Although many inversions have no effect on fitness, phenotypic effects have been found in instances where inversion breakpoints do not interrupt coding sequence. The position effect hypothesis posits that the function of a gene can be modified when it is moved to a new location (Muller, 1930; Dobzhansky, 1936). This can occur when a gene and closely linked modifiers are separated by the inversion. Unlike the coadaptation and local adaptation models, in the positive effect model selection acts upon the inversion itself and not the recombination suppressing properties of the inversion. Thus, no LD within the inversion is required to maintain the phenotype under selection.

A recent example of the position effect is the large inversion on equine chromosome 3 (ECA3). This 43 Mb inversion captures approximately a third of the chromosome and is linked to tobiano coat color (white-spotting) in horses (Brooks *et al.*, 2007). Localization

of the inversion breakpoints indicates that it does not physically disrupt any known genes; however, one of the inversion break points is 70 kb downstream of the KIT gene and it is possible that the inversion modifies the physical proximity between the gene and a regulatory elements (Haase *et al.*, 2008). Previously, this pattern was reported for separate inversions in mice and associated with coat color changes (Nagle *et al.*, 1995; Hough *et al.*, 1998). The ECA3 inversion is of interest to our studies because, similar to the ZAL2^{m} inversion, it influences coloration; however, the avian KIT homolog localizes to chromosome 4 in chicken and zebra finch and is not likely to be a factor in the white-throated sparrow system.

Genetic drift and underdominance

Under the coadaptation, local adaptation and position effect models for inversion evolution, inversions establish in a population as a function of their selective advantage. Genetic drift is another mechanism by which inversions can increase in frequency in a population but under genetic drift, an inversion can invade a population without any fitness advantage and, perhaps, in spite of negative fitness effects (Lande, 1984; Lande, 1985). In small populations or subpopulations, inversions can drift to fixation randomly and genetic drift the key evolutionary force in the stasipatric speciation model that has been proposed to explain why closely related species often differ by underdominant inversions (White, 1978; Rieseberg *et al.*, 1999; Noor *et al.*, 2001; Ortíz-Barrientos *et al.*, 2002). This model is controversial because when an inversion appears in a population, it is likely to be found only in a heterozygous state until it reaches an appreciable frequency. As heterozygote carriers of underdominant inversions will be selectively

removed by definition, the inversion could never reach high enough frequencies to create inversion homozygotes unless populations sizes are extremely small. Furthermore, there are few reports of underdominant inversions as fixed differences between species, obviating the need to explain their presence (Sites & Moritz, 1987; Nachman & Myers, 1989; Coyne *et al.*, 1991; Coyne *et al.*, 1993). Finally, it should be noted that the stasipatric model for speciation would only apply to inversions that are strongly underdominant, as weak ones are likely to also be weak barriers to reproductive isolation (Rieseberg, 2001). Although the role of underdominance in the process of speciation is under debate, an alternative hypothesis to explain the role of inversions in speciation suggests that reduced recombination facilitates speciation by protecting genomic regions from introgression (Rieseberg, 2001)

Overdominance

In contrast to underdominant inversions, overdominant inversions that confer higher fitness in the heterokaryotype are likely to reach a stable equilibrium in the population if they are not initially lost by chance (Kirkpatrick & Barton, 2006). Dobzhansky descdribed a large number of Drosophila inversions with features consistent with heterozygote advantage: their function appeared to be adaptive based on geographical or seasonal variation in frequencies, they reached stable equilibrium frequencies in population cage experiments and they were found as polymorphic at the population level (Dobzhansky, 1937; Levene & Dobzhansky, 1958; Dobzhansky & Pavlovsky, 1960; Dobzhansky, 1970). With these data, he suggested that overdominance was widespread and also a prominent mechanism by which variability is maintained in a population

(Dobzhansky, 1970); however, since Dobzhansky's *Drosophila* studies, little evidence has been put forth supporting the role of overdominance as a major evolutionary force (Fry, 2004; Charlesworth, 2006; Hughes, 2007).

The evolutionary significance of inversions

Models elaborating the evolution of inversions in populations have greatly contributed to our understanding of why inversions are so common in many species across all types of organisms. Although the relative importance of the different models has yet to be established, there is no argument that inversions play a significant role in evolution. They have been implicated in many important biological phenomena and processes, such as speciation, adaptive phenotypes, genetic correlations, disease and disease susceptibility, sex chromosomes, selfish genes and mate choice. Studying inversions with biologically relevant phenotypes will be of great importance in our understanding of their role in evolution.

Inversions and mate choice

The evolutionary significance of mate choice and sexual selection was first realized by Charles Darwin (Darwin, 1871) and has since been elaborated as a powerful force in evolution influencing sexual dimorphism, material resource offerings, species recognition, sex ratios, parental care, evolution of sociality, aggression, speciation, rapid evolution of traits (flowers, pheromones, plumage, song, ornaments) and so on (Emlen & Oring, 1977; West-Eberhard, 1983; Andersson & Iwasa, 1996; Kopp *et al.*, 2000). There are two clear examples described in the literature concerning the association of mate

choice with an inversion, the α/β inversion system in the seaweed fly and the ZAL2/ZAL2^{m} system in the white-throated sparrow (Lowther, 1961; Thorneycroft, 1966).

In the seaweed fly, an inversion has been linked to patterns of mate choice such that mating is disassortative with respect to the α/β inversion on chromosome 1 (Day & Butlin, 1987). The inversion is maintained at an equilibrium frequency in the population because it is overdominant and heterozygotes (α/β) show increased survival over both types of homozygotes (αα or ββ) (Butlin *et al.*, 1982). Homozygous αα flies are larger than ββ and that mating success is dependent on size, both of the male and the female (Butlin *et al.*, 1982). Day and Butlin (1987) further elaborated the indirect influence of the inversion on mate choice and found that like × like mating frequencies are significantly lower than expected. Thus, in the seaweed fly, a combination of heterozygote advantage and influence on mate choice maintains the α/β inversion at an intermediate frequency at the population level (Day & Butlin, 1987).

The textbook example of disassortative mating is the white-throated sparrow (Lowther, 1961; Houtman & Falls, 1994; Bekoff, 2004). These birds are polymorphic for plumage coloration that is linked to the ZAL2^{m} complex inversion consisting of a pair of nested inversions on chromosome 2 (Thomas *et al.*, 2008). The ZAL2^{m} effectively suppresses recombination in the heterokaryotype, a property that may be enhanced by the presence of adjacent and/or overlapping inversions (Dobzhansky, 1937; Lahn & Page, 1999; Lyon, 2003; Munté *et al.*, 2005). Inversion heterokaryotypes (ZAL2/ZAL2^{m}) mate at a

frequency of >96% with standard chromosome homozgyotes (ZAL2/ZAL2), producing an equal frequency of each morph in the next generation (Lowther, 1961). As a result of the extreme disassortative mating patterns, $ZAL2^m$ homokaryotypes are predicted to be extremely rare in the population, <1% (Lowther, 1961; Lowther & Falls, 1968; Thorneycroft, 1975; Falls & Kopachena, 1994) and the mating system may be a mechanism by which recessive deleterious mutations are rarely exposed.

Both the seaweed fly and white-throated sparrow inversions involve a significant portion of the total genome (~10%), spanning >200 and >1,000 genes, respectively (Thorneycroft, 1975; Crocker & Day, 1987; Thomas *et al.*, 2008). Taken together, these data support a theory of mate choice based on genetic incompatibility. In both cases, the mating system works to maintain heterozygosity at the population level in the loci captured within the inversion, supporting the notion that that mate choice may be driven by genetic compatibility (Brown, 1997; Tregenza & Wedell, 2000). Although associations between plumage morph and mate choice is common in birds (Greene *et al.*, 2000; Mundy *et al.*, 2004; Roulin, 2004; Pryke & Griffith, 2009), the white-throated sparrow is the only reported case in which these phenotypes are associated with a chromosomal polymorphism, although it is possible that an inversion is involved male coloration and behavior in the ruff, *Philomachus pugnax* (Lank *et al.*, 1995).

Inversions and sex chromosome evolution

The earliest steps in sex chromosome evolution involve the suppression of recombination around a sex-determining locus (Graves, 2006). Although it remains unclear whether the

initial cause of recombination suppression is genic or structural, in many cases, multiple inversions are associated with sex and neo-sex chromosomes that reduce recombination between the homologs (Lahn & Page, 1999; Iwase *et al.*, 2003; Lawson Handley *et al.*, 2004; Bergero *et al.*, 2008; Ross & Peichel, 2008; Benatti *et al.*, 2010). As previously discussed for the *t* haplotype, the presence of adjacent or overlapping inversions can enhance recombination suppression. Long term balancing selection is believed to be a rare evolutionary process, but sex chromosomes are a unique example of balancing selection imposed by negative assortative mating by sex and the molecular impact of balancing selection is found in the increased divergence between the homologs (Lahn & Page, 1999; Lawson Handley *et al.*, 2004; Charlesworth, 2006; Mank & Ellegren, 2007).

The most readily identifiable characteristic of most sex chromosomes is divergence in gene content and morphology, which both result from the long-term suppression of recombination between the sex chromosomes (reviewed in Graves 2006, Charlesworth and Charlesworth 2000, Charlesworth *et al.*2005). This genetic degeneration is a consequence of the suppression of recombination and the action of Hill-Robertson interference and Muller's Ratchet on the non-recombining chromosome (the Y or the W) and, to a lesser extent, its partner (the X or the Z). Many well-known examples of heteromorphic sex chromosomes have undergone such dramatic loss of genetic material that it is difficult to determine which factors are associated with genetic degeneration (repetitive elements or mutations in coding sequence) and the selective forces initiating and driving the degeneration. To address these gaps in our knowledge, some groups have focused on studying very recently evolved sex chromosomes to understand these

processes (Bachtrog & Charlesworth, 2002; McAllister, 2003; Liu *et al.*, 2004; Peichel *et al.*, 2004; Zhou *et al.*, 2008b; Kaiser & Charlesworth, 2010). We have argued that the ZAL2/ZAL2^{m} system mimics some key aspects of sex chromosomes, including suppression of recombination by multiple inversions, increased divergence and disassortative mating patterns. Because the ZAL2/ZAL2^{m} system bears similiarities to sex chromosomes, we believe that studies of the white-throated sparrow inversion system could help us understand the early steps in the evolution of sex chromosomes (Thomas *et al.*2008, Chapter 2 of this dissertation).

Mapping traits within an inversion

As inversions have been associated with a wide range of adaptive phenotypes, it is of great interest to identify the genes involved. Only a few studies have localized genes with phenotypic effects within inversions. The ability to map traits is dependent on the level of recombination between the two chromosome arrangements and one approach consists of narrowing down regions of interest by studying recombinant haplotypes, or mosaic haplotypes. In the mouse *t* complex, recombination mapping has enabled the localization of regions involved in TRD, as well as elucidation of the number of genes involved and their effect on the extent of TRD (Silver & Remis, 1987; Hammer *et al.*, 1991; Hermann *et al.*, 1999; Lyon, 2003). Only a few other studies of inversions have successfully used recombination mapping to localize candidate regions associated with linked phenotypes.

The 2La inversion is one such example. This inversion has a particularly interesting evolutionary history because it was introgressed into *Anopheles gambiae*, the primary malaria vector in sub-Saharan Africa, from a sibling species and confers aridity tolerance in its adult carriers and thermal tolerance in larvae (Besansky *et al.*, 2003; Rocca *et al.*, 2009). Inversion frequency is known to change seasonally and geographically (Besansky *et al.*, 2003; White *et al.*, 2007). Though the 2La inversion suppresses recombination 4-fold in the heterokaryotype, genetic exchange was observed at 9 out of 10 loci within the >20 Mb inversion (Stump *et al.*, 2007). Recombination mapping resulted in the identification of 2 ~1.5 Mb regions associated with the inversion breakpoints showing high LD (White *et al.*, 2007). Although >200 genes localize to these regions, a candidate gene approach to identifying those involved in aridity tolerance is hampered by poor functional annotation and lack of information regarding the types of genes involved in this phenotype (White *et al.*, 2007). In spite of these limitations, this study demonstrates the utility of recombination mapping for narrowing down regions of interest.

The $ZAL2^{m}$ polymorphism in the white-throated sparrow may afford a unique opportunity to discover the genetic architecture and candidate genes underlying the linked phenotypes; however, the capacity for mapping traits using recombination mapping could be limited in the case that the $ZAL2^{m}$ completely suppresses recombination. This would result in complete LD among all sites within the inversion and the inability to distinguish the effects of a single allele from all others. In this case, a candidate gene approach (as taken in Chapter 3) may be the only way to identify mutations responsible for the phenotypic variation associated with the $ZAL2^{m}$ inversion.

Alternative reproductive strategies, behavior and genetics in the ruff

The stable maintenance of competing reproductive strategies is unusual in animals (Tuttle, 2003). There is another bird species that shows remarkable similarities to the white-throated sparrow as a model for the evolution of alternative reproductive strategies, plumage polymorphism, behavior and mate choice. The ruff, *Philomachus pugnax*, is a species of sandpiper with two kinds of males that differ in plumage as well as mating strategies (van Rhijn, 1973). "Resident" males exhibit dark plumage coloration and occur at a frequency of ~84% and defend mating territories. "Satellite" males are lighter in color and account for the remaining ~16% of males. Satellites do not defend territories, but are recruited to the territories of resident males to attract females to the lek (van Rhijn, 1973; Lank *et al.*, 1995). Females do not show any parallel variation in behavior as seen in the males (Lank *et al.*, 1995; Lank *et al.*, 1999). Male plumage morph and behavior are linked and genetically heritable, though not sex-linked. Sex-linkage is ruled out because males are homogametic (ZZ) and pass their Z chromosomes directly to their sons; therefore, a sex-linked model could not explain how a male could produce sons of the opposite morph (Lank *et al.* 1995). Finally, because the plumage and behavior co-segregate, it is possible that they are genetically linked, perhaps by an inversion, as in the white-throated sparrow, but no studies have reported ruff karyotypes and the genes underlying the traits have yet to be identified.

Objectives of this dissertation

The goal of the research presented in this dissertation is to characterize the evolutionary forces shaping the population genetic patterns in the genome of the white-throated sparrow. In particular, the white-throated sparrow genome is especially suited to answer questions regarding the evolution and maintenance of chromosomal inversions, the capacity for inversions to suppress recombination and the evolutionary forces that shape genetic diversity within a genome. In Chapter 2, we detail the molecular evolution of and describe the distinct evolutionary patterns associated with the ZAL2^m inversion. Chapter 3 provides a genome-wide view of population genetic patterns in this species and reveals novel insights into patterns of variation in avian genomes. We also discuss the findings concerning the population genetic patterns on the ZAL2^m within the context of the entire genome. Additionally, the ZAL2^m inversion in the white-throated sparrow provides an opportunity to characterize polymorphisms in candidate genes that could contribute to the behavioral phenotype. In Chapter 4, we identify a subset of sequence variants in candidate genes that could contribute to the behavioral differences linked to the inversion. Finally, in Chapter 5, we discuss the potential origins of the inversion and the broader implications of this research in the field of inversion evolution as well as genome-wide population genetics and behavioral genetics.

2 Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*)[1]

[1]This chapter has been submitted for publication: Huynh, L.Y., D.L. Maney and J.W. Thomas. 2010. Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (*Zonotrichia albicollis*). Submitted.

2.1 Introduction

Chromosomal inversions are known to occur in a wide variety of organisms where they are associated with instances of adaptive evolution, speciation, selfish genes, sex chromosomes, human disease and disease susceptibility (Hartl, 1975; Lahn & Page, 1999; Noor *et al.*, 2001; Hoffmann *et al.*, 2004; Stefansson *et al.*, 2005; Dyer *et al.*, 2007). Although inversions have been studied for nearly a century, only recently have researchers begun to elucidate their effects on patterns of molecular evolution. Their influence on sequence evolution is derived from their unique ability to suppress recombination within the inverted interval in individuals heterozygous for the inversion (Sturtevant, 1921; Sturtevant & Beadle, 1936). Dobzhansky proposed that natural selection would favor inversions if they captured a set of positively interacting alleles, which he referred to as a “supergene” complex(Dobzhansky, 1937; Dobzhansky & Sturtevant, 1938; Dobzhansky, 1950). Alternatively, Kirkpatrick and Barton (2006) demonstrated that an inversion can enhance the fitness of its carrier, in the absence of positive epistasis, if it simply captures two or more locally adapted alleles. Because inversions suppress recombination, linkage disequilibrium (LD) between the beneficial alleles within the inversion would be preserved even when the alleles are not in close proximity to each other. Thus, the influence of inversions on the rate and pattern of recombination is fundamental to their adaptive significance and evolution.

The suppression of recombination between the alternative chromosomal arrangements along the inverted segment will eventually lead to the formation of two distinct haplotype

groups: the standard and the inverted. In population genetic studies of Drosophila, where inversions have been most intensely studied, genetic differentiation is non-uniform across an inversion and gene flow usually does occur between the standard and inverted chromosomes inside the inversion (Novitski & Braver, 1954; Hasson & Eanes, 1996; Schaeffer *et al.*, 2003; Schaeffer & Anderson, 2005). In these examples, genetic differentiation is typically highest near the inversion breakpoints. Double recombination events or gene conversion mediated by the formation of an inversion loop between the inverted and standard chromosomes will, over time, lead to significant gene flow within the inversion (Navarro *et al.*, 1997; Kovacevic & Schaeffer, 2000; Andolfatto *et al.*, 2001; Schaeffer *et al.*, 2003; Kennington *et al.*, 2006). The patterns of gene flux associated with an inversion can be particularly useful in identifying the targets of selection, which are expected to be in LD with each other as well as the inversion (see White *et al.*2007). Thus, the extent and pattern of gene flow associated with simple inversion polymorphisms will be influenced by the strength of selection to maintain LD, the size of the inversion, the rate of recombination and the age of the inversion.

Chromosomal polymorphisms involving more than one inversion have been found in natural populations and can be associated with patterns of gene flow that are distinct from the pattern described above for simple inversions. For example, the mouse t-complex is comprised of four non-overlapping inversions on mouse chromosome 17 and has been studied for decades with respect to its effect on recombination and association with meiotic drive (Lyon, 2003). Although gene conversion and some rare recombinants have been reported between wild-type and *t* chromosomes (Erhart *et al.*, 2002; Wallace &

Erhart, 2008), strong suppression of recombination extends over the length of the *t* complex. As a result, genetic differentiation between the *t* and wild-type chromosomes is uniformly high across the entire ~30-40 Mb *t* complex (Lyon, 2003). Similarly, the X^D chromosome in *Drosophila recens* is composed of a complex set of inversions and is associated with meiotic drive (Dyer *et al.*, 2007). The X^D completely suppresses recombination between the X^D and its non-distorting homologue, X^{ST}, resulting in dramatic chromosome-wide LD spanning ~130 cM (Dyer *et al.*, 2007). Unlike the classic model of gene flow for simple inversions in Drosophila, chromosomal polymorphisms involving more than one inversion can suppress recombination over the entire length of the rearrangement for prolonged periods of time and lead to genetically distinct haplotypes associated with exceptionally large blocks of LD.

Chromosomal polymorphisms comprised of complex inversions can drastically reduce the frequency of recombination within the rearranged regions in individuals heterozygous for the polymorphism, but recombination is expected to be restored when the inversion is found in the homozgote. This recombination can be prevented if the inversion carries a recessive mutation that causes sterility or early lethality. As a non-recombining segment of the genome, the inverted chromosome will then become subject to a series of population genetic forces that will result in the accumulation of deleterious mutations and genetic degeneration (Rice, 1994; Charlesworth & Charlesworth, 2000). A dramatic example of the long-term consequence of suppressed recombination is the mammalian Y, which has undergone massive genetic degeneration and lost almost all of its original genes except in the small pseudoautosomal region(s) where recombination still occurs

with the X chromosome (Graves, 2006). Newly arisen neo-Y chromosomes also typically show distinct signatures of genetic degeneration (Filatov *et al.*, 2000; Peichel *et al.*, 2004; Kondo *et al.*, 2006; Marais, 2007; Zhou *et al.*, 2008b), as do other rare examples of non-recombining regions of the genome (Slawson *et al.*, 2006). These examples illustrate that under extreme circumstances, complex inversion polymorphisms can eventually lead to regions of the genome in which recombination is rare or never occurs.

Recently we described the first modern genetic and genomic characterization of a chromosomal polymorphism in the white-throated sparrow (*Zonotrichia albicollis*) that is extraordinary in respect to its phenotypic effects and genetic properties (Thomas *et al.*, 2008). In particular, the two alternative arrangements of the 2nd chromosome, which will be referred to here as ZAL2 and ZAL2^{m}, are linked to a plumage polymorphism such that individuals homozygous for the ZAL2 are invariably associated with the tan-stripe (TS) morph, whereas individuals of the white-stripe (WS) morph are either heterozygous for the polymorphism (ZAL2/ ZAL2^{m}) or very rarely ZAL2^{m} homozygotes (Thorneycroft, 1966; Thorneycroft, 1975). In addition to the genetic association with plumage, the chromosomal polymorphism is linked to variation in social behavior such that WS individuals are, on average, more aggressive and less parental than their same-sex TS counterparts (Tuttle, 2003). WS and TS individuals occur at similar frequencies in both sexes and display an exceptionally strong negative assortative mating pattern in which >96% of all breeding pairs are comprised of individuals of opposite morphs (Falls & Kopachena, 1994). As a consequence of this breeding pattern, and perhaps due to

reduced viability (Thorneycroft, 1975), ZAL2^{m} / ZAL2^{m} birds are rare in the population with only a single ZAL2^{m} homozygote having been detected in studies that combined karyotyped more than 600 birds (Thorneycroft, 1975; Romanov *et al.*, 2009). Another consequence of this mating pattern is that the ZAL2^{m} is in a near constant state of heterozygosity maintained at the population level by balancing selection. At the molecular level the ZAL2^{m} differs from the ZAL2 by at least 2 nested inversions that, together, are predicted to span ~100 Mb and encompass ~1000 genes (Thomas *et al.*, 2008). Limited sampling of the genetic diversity and differentiation associated with this system suggested that suppression of recombination between the ZAL2 and ZAL2^{m} within the inverted interval might extend over the length of the inversion and that the two arrangements may have stopped recombining with each other ~2 million years ago (Thomas *et al.*, 2008). Those results led us to propose that because of the predicted lack of recombination between the ZAL2 and ZAL2^{m} inside the inversion and the paucity of ZAL2^{m} homozygotes, the ZAL2^{m} could be a non-recombining segment of the genome and a model for the early stages of sex chromosome evolution (Thomas *et al.*, 2008). Here, we report the results of our study designed to characterize, in detail, the patterns of genetic differentiation and recombination associated with this chromosomal polymorphism.

2.2 Materials and methods

Source of DNA: White-throated sparrows were collected on the campus of Emory University in Atlanta, GA during November and December of 2005, 2006 and 2007. A small blood sample was taken from a wing vein for DNA extraction and the morph of each bird was determined according to previously described criteria (Watt, 1986; Piper & Wiley, 1989) and confirmed by PCR (Michopoulos *et al.*, 2007). DNA from a single dark-eyed junco (*Junco hyemalis*) was collected from a locally captured bird. All procedures involving animals were approved by the Emory University Institutional Animal Care and Use Committee.

DNA sequencing: PCR primers were designed based on publicly available zebra finch genomic sequence (The Genome Center at Washington University, http://genome.wustl.edu) in regions conserved with chicken as described by Thomas *et al.* (2008). Primer sequences, orthologous position in zebra finch and specific annealing temperatures are listed in Supplemental Table 1. Each 25µL PCR contained final concentrations of 1X PCR buffer, 1.5 mM $MgCl_2$, 20 pmol of each primer, 0.2 mM of each dNTP, 1.5 units of Taq or Platinum Taq DNA polymerase (Invitrogen) and approximately 12.5-25 ng of genomic DNA. PCR cycling started with an initial denaturation at 94º for 5 min, followed by 35 cycles of 94º for 30 sec, 55º or 58º for 30 sec, 72º for 1 min and ended with a final extension at 72º for 7 min. PCR products were treated with shrimp alkaline phosphatase and exonuclease I (USB) before direct sequencing using the PCR primers or internal primers.

SNP discovery and sequence annotation: Polymorphisms were automatically called using SNPdetector (Zhang *et al.*, 2005) and all variants were manually confirmed prior to further analyses. Insertion and deletion polymorphisms were scored but not included in the subsequent analyses. The sequence at each locus was annotated for gene features based on annotation of the orthologous chicken genomic segments (Hillier *et al.*, 2004), as well as the presence of evolutionarily conserved regions as predicted by the PhastCons track on the UCSC genome browser (Siepel *et al.*, 2005). Individual loci were ordered and spaced relative to one another based on a previously established white-throated sparrow-chicken comparative map (Thomas *et al.*, 2008) and the assembled zebra finch genome (taeGut1).

DNA sequence analysis: Haplotypes were reconstructed from raw genotype data using PHASE v.2.1.1 under the default parameters (Stephens *et al.*, 2001). Each locus was phased individually and SNPs with < 0.55 phasing confidence (primarily singletons) were assigned randomly to alternative haplotypes which were used to generate population genetic statistics for each locus in Supplemental Table 2. We also phased the concatenated data set to generate a complete haplotype, which was used to build the haplotype network and estimate LD and recombination. Although incorrectly phased haplotypes do not affect measures of diversity, they can subtly influence estimates of LD and recombination. In TS birds, all haplotypes were identified as ZAL2 chromosomes. In WS birds, within the inversion we used fixed differences to distinguish between the ZAL2 and ZAL2^{m} chromosomes.

Population genetic statistics were calculated in DnaSP v.4.50.3 unless otherwise noted (Rozas *et al.*, 2003). We defined neutral sites as non-coding and synonymous sites located outside of segments orthologous to regions detected as being evolutionarily conserved in the chicken genome by PhastCons (Siepel *et al.*, 2005). Summary statistics for the intervals inside and outside the inversion were calculated using the phased haplotype generated from the concatenated sequences except π and θ, which were calculated as a weighted average of the individual loci. The neighbor-joining haplotype network was generated in Splitstree4 (Huson & Bryant, 2006). Pairwise comparisons of non-random association were performed in Haploview and the r^2 color scheme was used to illustrate LD between pairs of sites (Barrett *et al.*, 2005). Estimates and confidence levels for the population recombination rate parameter $\rho = 4N_er$ were assessed by Monte Carlo coalescent simulations in the *interval* algorithm in LDhat (Auton & McVean, 2007) conditioned on θ_W from DnaSP sampling every 2,000 of 10^6 iterations burning the first 100,000 iterations. Finally, we tested for gene conversion using the algorithm developed by (Betrán *et al.*, 1997) implemented in DnaSP (Rozas *et al.*, 2003).

2.3 Results

Data set: Previously we reported an initial survey of the genetic differentiation between the ZAL2 and ZAL2^{m} based on sequencing a small number (n = 10) of loci (Thomas *et al.*, 2008). To carry out a more detailed study of the genetic differentiation, nucleotide diversity, and recombination associated with this chromosomal polymorphism, we expanded our study to include a total of 62 loci that were spaced on average every 1.7 Mb along the ZAL2/ZAL2^{m} chromosome in 4 TS and 8 WS birds, corresponding to a sample size of 16 ZAL2 and 8 ZAL2^{m} chromosomes. Of the sequenced loci, 58 mapped within the inversion and totaled ~35 kb, although 4 loci mapped outside the inversion and totaled ~1.7 kb (Table 2.1). Based on annotation of the loci and our criteria for defining neutral sequence (see Methods for details) the data set included ~14 kb and ~0.8 kb of neutral sites inside and outside the inversion, respectively (Table 2.1). Overall we identified 297 SNPs, of which 279 mapped within the inversion and 18 mapped outside the inversion (Table 2.1). After excluding three tri-allelic SNPs, the final data set available for analyses was comprised of 277 SNPs within and 17 SNPs outside the inversion.

Table 2.1 Diversity, divergence and summary statistics for the standard ZAL2 and inverted ZAL2 ^{m} chromosomes in the white-throated sparrow.

Region	*Loci*	*Length (bp)*	*Segregating sites, S*	*π (All)*	*π (ZAL2)*	*π (ZAL2^{m})*	F_{ST}^{a}	d_{xy}^{b}	*Tajima's D*	*Tajima's D (ZAL2)*	*Tajima's D (ZAL2^{m})*
Inside inversion	58	34,823 (13789)	277 (167)	0.00293 (0.00449)	0.00044 (0.00075)	0.0003 (0.00042)	0.94	0.00583 (0.00871)	0.65	-0.74	-0.84
Outside inversion	4	1,654 (777)	17 (14)	0.00191 (0.00296)	0.0016 (0.00311)	0.00238 (0.00455)	0.21	0.0019 (0.00296)	-0.82	-0.9	-0.22

Numbers in parentheses are calculations for neutral sites, noncoding, synonymous sites outside of most conserved elements.

[a]F_{ST} is calculated by comparing all ZAL2 chromosomes to all ZAL2 ^{m} chromosomes

[b]Divergence (d_{xy}) between the ZAL2 and ZAL2^{m} chromosomes is calculated as the average number of nucleotide substitutions per site.

[c]Tajima's D values were calculated using 4 TS and 4 WS birds, thus representing the natural population frequency of the ZAL2 and ZAL2 ^{m} chromosomes.

Genetic differentiation between the ZAL2 and ZAL2^m: In our earlier study we found high levels of genetic differentiation between the ZAL2 and ZAL2^m within the inversion and evidence of gene flow outside the inversion (Thomas *et al.*, 2008). Consistent with those findings we found that the majority (185/277) of SNPs within the inversion were fixed differences between the chromosomal arrangements (i.e., one allele was invariably linked to the ZAL2 and the alternative allele to the ZAL2^m), and only 1 SNP was a shared polymorphism (i.e., both alleles were present in both chromosomal arrangements). In contrast, outside the inversion we identified 17 SNPs, none of which were classified as fixed differences between the chromosomal arrangements and two were shared based on the complete haplotype phase estimate and not the individual locus phasing.

To formally quantify the level of genetic differentiation between the chromosomal arrangements inside and outside the inversion, we calculated the population genetic statistics F_{ST} and d_{xy} treating the ZAL2 and ZAL2^m as two separate populations (Table 2.1). As expected based on the prevalence of fixed differences between the ZAL2 and ZAL2^m, F_{ST} inside the inversion was near the theoretical maximum of 1 (F_{ST} = 0.94), consistent with a high degree of genetic differentiation and extremely limited gene flow between the two chromosomal arrangements within the inversion. Indeed, the average pairwise divergence in neutral sequence between a ZAL2 and ZAL2^m chromosome inside the inversion was elevated relative to outside the inversion (d_{xy} = 0.0087 versus 0.0029). Outside the inversion, the F_{ST} value between the arrangements was 0.21. Because we had no method to reliably assign ZAL2 and ZAL2^m identity outside the inversion, we tested for significant population structure between TS and WS groups. Inside the inversion,

population differentiation between WS and TS was significant (F_{ST} = 0.41, p = 0.009), whereas outside the inversion, there was no significant structure (F_{ST} = 0.06, p = 0.09) (Excoffier *et al.*, 2005). Thus, these results demonstrate that genetic differentiation between the ZAL2 and ZAL2^m is uniformly high across the entire ~104 Mb inverted interval and low within the region outside the inversion, consistent with gene flow between the chromosomal arrangements being restricted to the small segment of the chromosome outside the inversion.

Patterns of haplotype and nucleotide diversity within and between the ZAL2 and ZAL2^m: The striking genetic differentiation between the chromosomal rearrangements inside compared to outside the inversion is consistent with the presence of two highly divergent haplotype groups each associated with the inverted segments of the ZAL2 and ZAL2^m chromosomes. To visualize the relationships among the haplotypes, we constructed separate haplotype networks based on the phased concatenated haplotype inside and outside the inversion (Figure 2.1). As expected, the ZAL2 and ZAL2^m haplotypes clustered into two distinct groups that were clearly differentiated from one another whereas outside the inversion no clustering based on chromosomal arrangement was observed (Figure 2.1). Moreover, the short branches of the ZAL2 and ZAL2^m groups illustrate the relatively low diversity observed inside the inversion within both chromosomal rearrangements compared to outside the inversion (Table 2.1 and Figure 2.1A).

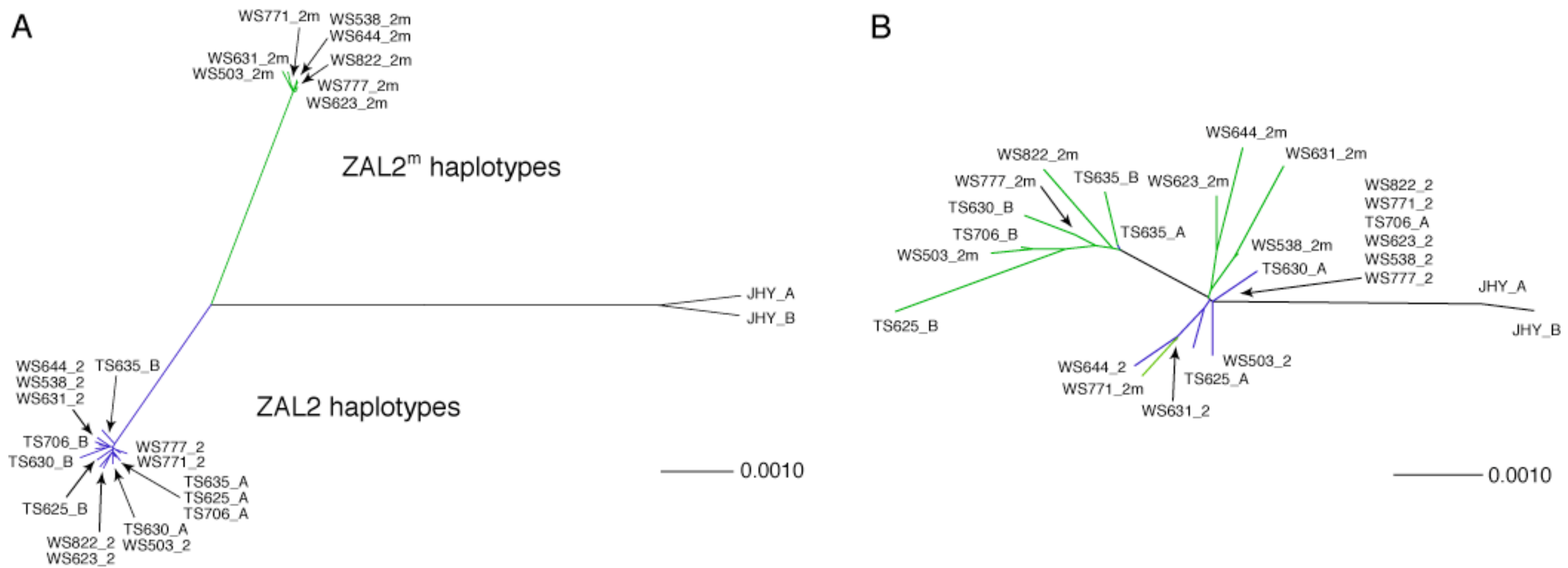

Figure 2.1 ZAL2 and ZAL2^{m} haplotype networks. Haplotype networks were constructed using the concatenated phased haplotypes from all loci within (A) and outside (B) the inversion. For the white-striped (WS) birds the ZAL2 (blue lines) and ZAL2^{m} haplotypes (green lines) are labeled as 2 or 2m. For the tan-striped (TS) birds and Junco (JHY) the haplotypes for each individual were arbitrarily labeled A (blue) and B (green).

The haplotype networks clearly indicate distinct differences in the genetic diversities of the ZAL2^{m}, the ZAL2 and the region outside the inversion. To assess this difference we considered that natural populations of white-throated sparrows are comprised of ~50% TS (ZAL2/ZAL2) individuals and ~50% WS (ZAL2/ZAL2^{m}) individuals (Lowther, 1961). At the population level the inverted ZAL2^{m} region, the homologous standard ZAL2 region and the region outside the inversion exist at a 1:3:4 ratio. Under the neutral model, in which we assume the absence of selection and comparable rates of recombination, nucleotide diversity is directly proportional to the effective population size ($\theta = 4N_e\mu$). In this context, we predicted that the ratio of neutral nucleotide diversity of the ZAL2^{m}, the ZAL2 and the region outside the inversion would also be 1:3:4. The observed ratio of neutral nucleotide diversity between the ZAL2^{m} and ZAL2 was ~1:2, similar to our expected ratio. The ratio of diversity of the ZAL2^{m}, the ZAL2 and the region outside the inversion was 1:2:10 (Table 2.1), lower than predicted by the model. This result could be caused by lower than expected diversity for the ZAL2^{m} and ZAL2, higher than expected diversity outside the inversion, or both.

Linkage disequilibrium: The suppression of recombination between the chromosomal arrangements across the length of the inverted interval is expected to result in a strong signature of linkage disequilibrium (LD) extending over the ~100 Mb inversion. To quantify the extent of LD associated with this chromosomal polymorphism, we measured LD between pairs of informative sites along the entire length of the chromosome (Figure 2.2). A total of 240 SNPs, 232 inside the inversion and 8 outside the inversion, were informative and as expected under a scenario of long-term suppressed recombination

between the arrangements, a large fraction of the pairwise comparisons inside the inversion showed significant nonrandom association (17,217 out of 26,796, $p < 0.01$). Outside the inversion we did not detect significant patterns of LD (Figure 2.2), only 1 pairwise comparison was significant ($p < 0.01$) and the physical distance separating the two SNPs is only 52 bp. To summarize the level of LD inside and outside the inversion we calculated ZnS between informative sites (Kelly, 1997), which was high inside the inversion (ZnS = 0.69) and much lower outside (ZnS = 0.10) the inversion. Thus, these results establish that the suppression of recombination between the ZAL2 and $ZAL2^m$ has resulted in widespread LD that based on our most proximal and distal markers spans a minimum of ~104 Mb inside the inversion

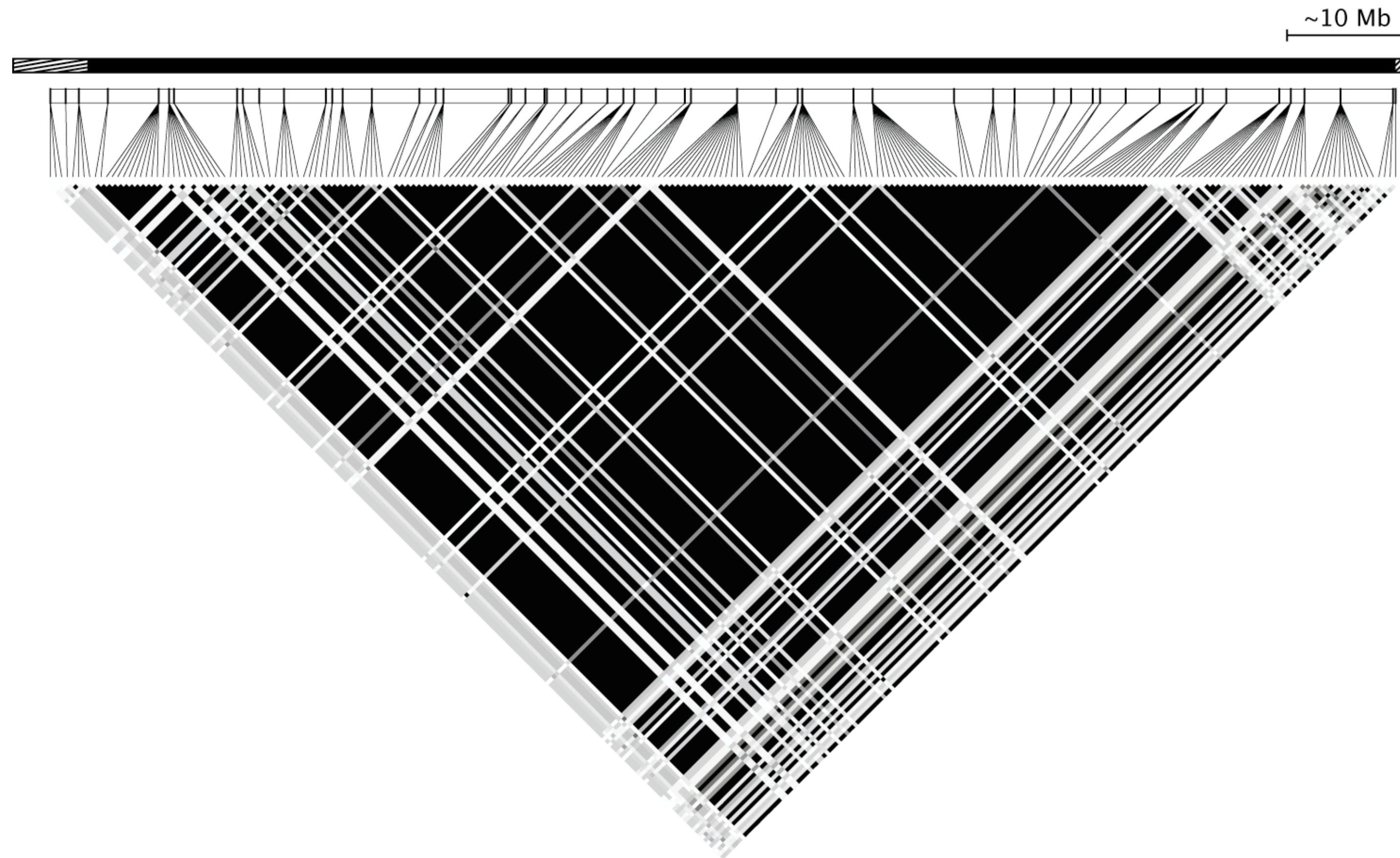

Figure 2.2 Linkage disequilibrium across the ZAL2/ZAL2^{m} chromosomes. The concatenated phased haplotypes were used to generate an LD plot across the ZAL2/ZAL2^{m} chromosomes. Each of the 239 informative SNPs are plotted to scale with respect to their predicted position along the ZAL2 chromosome. Black squares indicate perfect LD between pairs of SNPs ($r^2 = 1$), gray squares indicate pairs of SNPs with r^2 between 0 and 1 and white squares show no LD. The striped segments of the line above the LD plot indicate regions outside the inversion and the solid (black) segment of the line indicates the region inside the inversion.

Rates of recombination within the ZAL2 and ZAL2^m haplotype groups: Our data strongly support a scenario in which recombination in WS (ZAL2/ZAL2^m) birds is restricted to the interval outside the inversion. Because ZAL2^m/ZAL2^m individuals are extremely rare, $< 1\%$ of the population, we had previously predicted that the inverted segment of the ZAL2^m chromosome might be a non-recombining autosome (Thomas *et al.*, 2008). To test this hypothesis we applied the four-gamete test, 4GT (Hudson & Kaplan, 1985), to look for evidence of recombination within ZAL2^m haplotypes, as well as to the ZAL2 haplotypes and the haplotypes outside the inversion on both chromosomes. As expected, we detected recombination within the ZAL2 haplotype ($\geq$ 12 events), and within haplotypes outside of the inversion ($\geq$ 1 event). In contrast to our prediction, the 4GT indicated that recombination had also occurred within the ZAL2^m haplotypes ($\geq$ 2 events).

Given the evidence for recombination within both ZAL2 and ZAL2^m haplotypes, we next attempted to quantify the difference in population rate of recombination, ρ, between the ZAL2 and ZAL2^m. In particular, because of the large difference in the population frequency of ZAL2^m/ZAL2^m birds ($< 1\%$) to ZAL2/ZAL2 birds ($\sim 50\%$) (Lowther, 1961; Thorneycroft, 1975; Falls & Kopachena, 1994; Tuttle, 2003), we expected a lower rate of recombination inside the inversion within the ZAL2^m haplotypes versus within the ZAL2 haplotypes. Estimates of the population recombination rate amongst all ZAL2^m chromosomes was similar to that of the ZAL2 chromosomes ($\rho \pm SD = 8.48 \times 10^{-9} \pm 8.59 \times 10^{-9}$ and $4.40 \times 10^{-8} \pm 5.05 \times 10^{-8}$ per base, respectively), though it should be noted that diversity levels within the inversion on the ZAL2 and the ZAL2^m were low and our small

sample size did not support accurate measurements of ρ (i.e., note standard errors). Finally, we also predicted that outside the inversion, the rate of recombination would be greater than within the inversion because the presumed obligatory recombination event required for the proper chromosome segregation in the ~50% of the population that are ZAL2/ ZAL2^{m} heterozygotes (WS) (Lowther, 1961; Thorneycroft, 1975), must occur in this small interval. Consistent with our prediction, ρ for this segment was estimated to be higher than observed inside the inversion ($1.55 \times 10^{-7} \pm 2.18 \times 10^{-7}$ per base), again, note the large confidence intervals.

Relative rate of evolution on the ZAL2 versus ZAL2^{m} chromosomes: The pattern of extensive recombination suppression between the ZAL2 and ZAL2^{m} within the inverted region suggests that Hill-Robertson interference among the linked loci could work to reduce the efficacy of selection and subsequently lead to the genetic degeneration of the ZAL2^{m} chromosome (Hill & Robertson, 1966; Rice, 1994; Gordo & Charlesworth, 2001). We were interested in whether the ZAL2^{m} exhibited any signs of degeneration. Specifically, we tested for differences in the rates of evolution in the ZAL2 versus ZAL2^{m} lineages applying Tajima's relative rate method using the χ^2 test (Tajima, 1993) to fixed differences that mapped inside the inversion and could be unambiguously polarized using sequence from a closely related outgroup (*J. hyemalis*, Table 2.2). Although a greater number of mutations were inferred to have occurred in the ZAL2^{m} versus ZAL2 lineage in all sequences classes examined, including nonsynonomous and other presumably functional sites, the observed differences were not statistically significant (Table 2.2). Thus, though the ZAL2^{m} lineage may be associated with a

slightly higher overall mutation rate than the ZAL2 lineage, we did not observe any overt signs of genetic degeneration expected on a chromosome with reduced levels of recombination.

Table 2.2 Mutations occurring along the ZAL2 and ZAL2^m lineages.

	All	*Nonsynon*	*Synon*	*Conserved*	*Neutral*
ZAL2	78	1	6	31	47
ZAL2^m	98	5	8	36	62
P-value*	0.13	0.102	0.593	0.54	0.15

This test includes the subset of fixed differences where ancestral state could be unambiguously inferred using junco sequence.

Conserved category includes all SNPs within most conserved elements and neutral category includes all SNPs outside of most conserved elements.

*P-values were calculated using the c^2-test to determine if the rate of substitution on the ZAL2 and ZAL2^m branches differed.

2.4 Discussion

Gene flow, genetic differentiation and population structure between the ZAL2 and ZAL2^m: Unlike most simple inversion polymorphisms where, over time, gene flow occurs between the standard and inverted arrangements except near the inversion breakpoints, complex inversion polymorphisms can be effective suppressors of recombination (Lyon, 2003; Munté *et al.*, 2005; Dyer *et al.*, 2007). For example, the O_{3+4} inversion polymorphism in *Drosophila subobscura* differs from the standard O_{ST} arrangement by two overlapping inversions and suppresses gene flow between the two chromosome types producing uniformly high genetic differentiation across the ~4 Mb region (Munté *et al.*, 2005). Although evidence for gene conversion between the alternative arrangements is prevalent in this system, the limited gene flow in the absence of double-recombination events is clearly demonstrated by the high proportion of fixed

differences (35%) and the level of divergence (d_{xy} = 1.5%) between the alternate arrangements (Munté *et al.*, 2005). Similarly, in the case of an even larger complex inversion polymorphism, the X_D chromosome in *D. recens*, over half of the polymorphisms identified on the X_D chromosome were fixed differences between the distorting and wild-type chromosomes (Dyer *et al.*, 2007). We also observed a lack of gene flow between the ZAL2 and ZAL2$^{\mathbf{m}}$ as indicated by a high frequency of fixed differences (67%) and divergence between the chromosomes approaching 1%. Moreover, the frequency of shared polymorphisms within the inversion was very low (< 1%), and unlike the O_{3+4} and O_{ST} system, we failed to detect any gene conversion tracts between the alternative arrangements or between ZAL2^{m} chromosomes within the inversion interval. In these examples from Drosophila and the ZAL2^{m} polymorphism, there is no evidence for genetic exchange towards the center of the inversion polymorphisms as predicted by models developed by (Navarro *et al.*, 1997) for single inversions, suggesting that multiple overlapping inversions are likely to completely suppress double crossovers and can prevent gene flow across the entire inversion interval. Moreover, the very high F_{ST} value (0.94) between the ZAL2 and ZAL2^{m} inside the inversion is also quite uncommon. Indeed, similar values of F_{ST} in excess of > 0.8 have been associated with special circumstances, such as homologous loci in non-recombining segments of sex chromosomes (Ironside & Filatov, 2005) or between cryptic species (Terry *et al.*, 2000). Thus, the degree of genetic differentiation observed between the ZAL2 and the ZAL2$^{\mathbf{m}}$ chromosomes is very high even for complex inversions, and is comparable to values observed in the most extreme cases of suppressed gene flow.

Dating and origin of the ZAL2^{m}: Using the level of neutral divergence between the ZAL2 and ZAL2^{m} chromosomes, we estimated the age of the current ZAL2^{m} haplotypes (i.e., time since recombination stopped) assuming a genomic mutation rate equivalent to that described for zebra finch, 2.95×10^{-9} substitutions/site/year (Balakrishnan & Edwards, 2009). Our estimate of ~2.95± 0.3 MY is similar to our previous estimate based on phylogenetic comparisons of noncoding and synonymous sequence between white-throated sparrows and the *J. hyemalis* outgroup (Thomas *et al.*, 2008). The time of the ZAL2^{m} origin predates the estimated time of divergence of the white-throated sparrow from other birds in the genus and is at apparent odds with the hypothesis that the inversion occurred in the white-throated sparrow lineage (Thorneycroft, 1975). Although the polymorphism may be ancient and was simply lost in the other *Zonotrichia* sparrow lineages, an alternative explanation for its restricted presence in the white-throated sparrow is through introgression as the result of hybridization. Hybridization in birds is well-established (Grant & Grant, 1992; Mallet, 2005) and fertile offspring are common in cases where the genetic divergence between the species is on the order observed between the ZAL2 and ZAL2^{m}, i.e. < 2% (Price & Bouvier, 2002). Although there is precedent for hybridization between the white-throated sparrow and the Junco (Dickerman, 1961), our haplotype networks do not support a recent introgression of either chromosome arrangement from that species. Relatively recent introgression resulting from the hybridization with another species might explain the high level of genetic divergence between the ZAL2 and ZAL2^{m}, as well as the lack of recombination within the inversion. Studies of hybrid genomes have shown that inversions can produce patterns of heterogeneous recombination and high genetic divergence

(Rieseberg *et al.*, 1999; Feder *et al.*, 2003; Panithanarak *et al.*, 2004; Stump *et al.*, 2005; Noor *et al.*, 2007). Additionally, because collinear regions are expected to homogenize more quickly than non-collinear regions (Rieseberg *et al.*, 1999), time since hybridization is another factor to take into consideration. Future studies focused on genome-wide patterns of nucleotide diversity and haplotype structures should help clarify whether or not the ZAL2/ZAL2^{m} polymorphism could have resulted from a recent hybridization event.

Extreme LD associated with the ZAL2/ZAL2^{m} system: Theoretically, the adaptive significance of inversions lies in their capability to suppress recombination, thereby maintaining LD between multiple alleles favored by natural selection (Dobzhansky, 1937; Kirkpatrick & Barton, 2006), and the pattern of LD is dependent on the rate of recombination as well as the strength of selection to maintain linkage between the alleles. In the case of the ZAL2/ZAL2^{m} system we detected two dominant haplotypes representing each arrangement that spanned the ~100 Mb inside the inversion, linking alleles from ~1,000 genes into a super-gene complex. The extended LD we observed in this system is more extreme than that observed in the mouse *t* complex, which spans ~30-40 Mb (Lyon, 2003), and comparable to that observed in the distorting (X_D) in *D. recens* (Dyer *et al.*, 2007). In that case, LD was found to essentially extend across the entire X chromosome, ~130 cM (Dyer *et al.*, 2007). As far as we are aware, the pattern of LD associated with the ZAL2/ZAL2^{m} is the most extreme example of long-range LD yet to be reported in vertebrates.

Comparison of the ZAL2/ZAL2^{m} chromosome pair to sex chromosomes: Previously, we proposed that the ZAL2/ZAL2^{m} chromosomes were mimicking the early stages of sex chromosome evolution and that the ZAL2^{m} shared a number of features with the Y (W) chromosome (Thomas *et al.*, 2008). In particular, these features included: suppression of recombination between the ZAL2 and ZAL2^{m} over most of their length due to inversions, a negative assortative mating system in which > 96% of breeding pairs consist of heterogametic (ZAL2/ZAL2^{m}) × homogametic (ZAL2/ZAL2) individuals, and the maintenance of the ZAL2^{m} in a near constant state of heterozygosity (Thomas *et al.*, 2008). Although our results from this study do not require us to reconsider those shared features, the detection of recombination within ZAL2^{m} haplotypes refutes the additional prediction that the region inside the inversion on the ZAL2^{m} may represent a non-recombining autosome. Thus, though ZAL2^{m}/ZAL2^{m} individuals in the population are rare, the detection of historical recombination events associated with this haplotype suggest that at least some of those individuals are fertile. Indeed, a male ZAL2^{m} homozygote has been described that was able to reproduce (Falls & Kopachena, 1994).

Given the presence of recombination, however rare it may be on the ZAL2^{m}, it was therefore not surprising that we did not detect a significant difference in the accumulation of potentially deleterious mutations on the ZAL2^{m} compared to the ZAL2. The complete cessation of recombination, as with neo-Y chromosomes like those in *Drosophila miranda* (Bachtrog & Charlesworth, 2002), stickleback and medaka fish (Peichel *et al.*, 2004; Kondo *et al.*, 2006), the black muntjac (Zhou *et al.*, 2008b) and *Silene latifolia* (Filatov *et al.*, 2000; Marais *et al.*, 2007), leads to reduced efficacy of selection and

accumulation of mutations (Charlesworth & Charlesworth, 2000). Even very low levels of recombination can effectively prevent the process of genetic degeneration observed in non-recombining chromosomes (Haddrill *et al.*, 2007). In this context, and in light of the results from the loci sampled in this study, we can conclude that the ZAL2^{m} is unlikely to be mimicking the initial phases of genetic deterioration observed on neo-Y chromosomes.

It should be noted that our inference of recombination within the ZAL2^{m} chromosomes relies on the accurate reconstruction of haplotypes from genotype information. Because the number of predicted recombination events within the ZAL2^{m} chromosomes is small and the frequency of the minor alleles of these sites is low, incorrect phase estimation could lead to false positives. Additionally, the 4GT relies on the assumption of infinite sites, which excludes the possibility of independent recurrent mutations. Considering that the minor allele frequency of the sites involved is generally low, independent mutations within the ZAL2 and ZAL2^{m} lineages could produce patterns of variation that resemble those resulting from recombination. Given that we did not detect the genetic degeneration that is characteristic of non-recombining sequences, we believe that these are true signatures of recombination generated by the rare ZAL2^{m} homozygotes.

Although we no longer consider the ZAL2^{m} a non-recombining chromosome, we did note potentially low levels of diversity associated with the ZAL2/ZAL2^{m} system that are typical of sex chromosomes. For example, the mammalian and plant Y and the avian W chromosomes, as well as other regions of low recombination, can show 20-30 fold reductions in diversity even after correcting for differences in N_e (Filatov *et al.*, 2000;

Jensen *et al.*, 2002; Berlin & Ellegren, 2004; Hellborg & Ellegren, 2004; Betancourt *et al.*, 2009). Similarly, the W chromosome has been associated with lower than expected levels of nucleotide diversity in other birds (Montell *et al.*, 2001; Berlin & Ellegren, 2004). In the case of the ZAL2/ZAL2^{m} polymorphism we observed that diversity within the inversion on the ZAL2 and ZAL2^{m} was 5 and 10 times lower, respectively, than outside the inversion. It has been argued that intense sexual selection could be responsible for reductions in diversity on sex chromosomes by further reducing N_e of the Y chromosome (Caballero, 1995; Nagylaki, 1995; Charlesworth, 1996); however, such a mechanism cannot explain the reduced diversity observed on the ZAL2^{m}. Because sexual selection is not a factor in the ZAL2/ZAL2^{m} system, natural selection is the only factor that can explain the reduced variability of sex chromosomes and regions of low recombination alike (Hellborg & Ellegren, 2004). Further efforts to quantify patterns of recombination and diversity in the white-throated sparrow genome will provide the necessary context to understand how rates of recombination have influenced the regions inside and outside the inversion.

Implications for identifying specific genes underlying the phenotypes associated with the ZAL2^{m} polymorphism: One of the compelling reasons to study the ZAL2^{m} is the opportunity to identify the genetic basis of the phenotypic variation associated with the inversion. Previous studies have localized candidate regions within inversions by examining patterns of recombination between the wild-type and inverted arrangements and identifying region in LD as targets of selection. For example, in a study of the 2La inversion in *Anopheles gambiae*, (White *et al.*, 2007) examined patterns of divergence

and LD to localize candidate regions involved in aridity tolerance. In the case of the ZAL2/ZAL2^{m} polymorphism, LD across the entire inversion will preclude the further localization of candidate regions or genes by standard recombination-based mapping. In addition, if we consider that the divergence between the ZAL2 and ZAL2^{m} within the inversion is on the order of 1%, conservatively estimate that 50% of the differences are fixed or near-fixed differences between the arrangements, and that the inversion is ~100 Mb, then we can predict there are at least 500,000 single-nucleotide differences that distinguish the ZAL2 from the ZAL2^{m}. Thus, it is possible and likely that the TS and WS birds differ from each other in hundreds of traits. Nonetheless, understanding that we expect to observe some differences between the ZAL2 and ZAL2^{m} alleles in cis-regulatory or protein coding portions of essentially all ~1,000 genes within the inversion, it would be of significant interest to characterize genes that have been previously established to modulate social behavior in other systems that map within the inversion. A candidate gene approach may be the only way to identify mutations responsible for the phenotypic variation associated with the ZAL2^{m} inversion.

Conclusions: Our population genetic analysis reveals that the ZAL2^{m} arrangement suppresses recombination in the heterokaryotype, resulting in reduced gene flow, high levels of genetic differentiation, extensive population structure and LD between the alternate arrangements. Although we no longer consider the ZAL2^{m} a non-recombining chromosome, we believe that it will be valuable as a model for understanding how selection acts to reduce diversity in genomic regions with low recombination rates.

Future studies of recombination and diversity at unlinked autosomal and sex-linked loci will likely shed light on the evolution of the ZAL2/ZAL2^{m} system.

3 Contrasting population genetic patterns within the white-throated sparrow genome (*Zonotrichia albicollis*)[1]

[1]This chapter has been submitted for publication: Huynh, L.Y., D.L. Maney and J.W. Thomas. 2010. Contrasting population genetic patterns within the white-throated sparrow genome (*Zonotrichia albicollis*). Submitted.

3.1 Introduction

The genomic landscape is influenced by the combined interactions of mutation, recombination, natural selection, genetic drift, and demographics. Within a genome, signatures of these forces and their relative importance can be inferred by examining local levels and patterns of genetic variation. In general, purifying selection is the dominating selective force in molecular evolution and results in a reduction of genetic diversity as new variants are selectively removed from the population (Kimura, 1983). Reductions in diversity can also arise from positive directional selection, in which a particular variant is favored by natural selection and rises to fixation resulting in the loss of diversity, referred to as a selective sweep (Maynard Smith & Haigh, 1974). By contrast, other forms of positive selection, called balancing selection, can increase the amount of genetic diversity through heterozygote advantage or frequency-dependent selection (Charlesworth, 2006). Finally, genetic drift can randomly influence the evolutionary fate of new mutations regardless of their selective benefit, such that in larger populations, variants are more likely to be lost due to drift (Charlesworth, 2009).

Although selection and drift may act upon a single variant, the rate of recombination will determine whether closely linked sites will also be affected. In genomic locations where recombination is absent or infrequent, there will be linkage disequilibrium (LD), or non-random association of genetic variants. If purifying selection removes a deleterious mutation, any variation in LD with that mutation will also be removed (Charlesworth *et al.*, 1995). If positive selection fixes a beneficial mutation, any linked variation will also

be fixed (Maynard Smith & Haigh, 1974). Together, background selection and genetic hitchhiking reduce overall levels of variation and inhibit the efficacy of selection because selection cannot act on mutations independently as a result of LD. This phenomenon is known as Hill-Robertson Interference or HRI (Hill & Robertson, 1966). When recombination is frequent, the impact of HRI is minimized and standing variation is less susceptible to loss due to selection at another site and local levels of nucleotide diversity are positively correlated with local recombination rates (Begun & Aquadro, 1992).

The rate of recombination has been shown to vary substantially between different genomic regions in many species (Nachman, 2002). In the chicken, the rate of recombination is known to systematically vary between chromosomes of differing sizes. Most avian genomes are organized into several large macrochromosomes, several intermediate sized chromosomes and many tiny microchromosomes (Hillier *et al.*, 2004). In chicken, the rate of recombination on microchromosomes is eight times greater than on macrochromosomes (Hillier *et al.*, 2004). Recent studies in zebra finch have revealed that the recombination rate can be 5–10 times greater on microchromosomes than macrochromosomes (Backström *et al.*, 2010). The higher rate of recombination could be attributed to the obligatory cross-over during meiosis which would cause the frequency of crossing-over (i.e. the recombination rate) to increase as chromosome size decreases (Rodionov, 1996). In zebra finch, Backström *et al.* (2010) found a pronounced telomere effect such that recombination rates are highly elevated within <20 Mb of the telomeres. As microchromosome sizes are generally <20 Mb in length, their recombination landscape is characterized as a recombination jungle, like the telomeres (Backström *et*

al., 2010). Considering that smaller chromosomes have higher recombination rates and that higher recombination rates are associated with increased nucleotide diversity, it is reasonable to expect that smaller chromosomes will harbor increased genetic variation (Ellegren, 2005), though this prediction has not been investigated in depth in avian genomes (Wong *et al.*, 2004; Fang *et al.*, 2008). Furthermore, because recombination rate increases so dramatically toward the chromosome ends (Backström *et al.*, 2010), intrachromosomal variation in nucleotide polymorphism may also be considerable.

Polymorphism levels are also expected to differ between autosomes and sex chromosomes because standing variation is directly proportional to the effective population size (N_e), which varies between sex chromosomes and autosomes (Ellegren, 2009a). The expected ratio of N_e under a neutral model between autosomes and the Z and W chromosomes is 4:3:1, because, on average in an idealized population, one W and three Z chromosomes are observed for every four autosomes (Ellegren, 2009a). Levels of polymorphism are predicted to show the same relationship, and reduced diversity is expected for both sex chromosomes, but is predicted to be more dramatic on the W. Studies of avian sex chromosomes (Montell *et al.*, 2001; Berlin & Ellegren, 2004) have shown that the diversity on the Z and W is lower than expected based on relative N_e when compared to autosomes, which is a pattern that has been widely observed in XY systems as well (Ellegren, 2009a). Although sex chromosomes differ from autosomes in their N_e, they also show a different pattern of recombination, such that recombination is reduced on the Z and absent on the W in the non-recombining regions. Because of these differences in recombination, it is likely that HRI plays a role in further reduction of

genetic diversity on both sex chromosomes with a more dramatic effect on the W chromosome (Charlesworth *et al.*, 2005).

Because females are the heterogametic sex in birds, the W chromosome is genetically linked to the mitochondrial genome, which is also non-recombining. As a result, any selective events on the W will affect the mitochondrial genome and vice versa(Berlin *et al.*, 2007), which could explain why W chromosome diversity is at least 100-fold lower than on the autosomes (Montell *et al.*, 2001; Berlin & Ellegren, 2004). The Z chromosome may also experience a further reduction in genetic diversity due to the fact that two-thirds of the time it is passed through the male germ line. When there is high variation in male mating success, the N_e of the Z chromosome will be reduced, as well as the amount of standing variation (Sundström *et al.*, 2004; Borge *et al.*, 2005). Thus, both the Z and W chromosomes are expected to experience reductions in diversity due to factors that do not play a role in XY chromosome diversity.

Previously, we described unusual patterns of polymorphism and recombination in remarkable chromosomal polymorphism in the white-throated sparrow (*Zonotrichia albicollis*) that shares many characteristics of sex chromosomes, though it is not sex-linked (Thomas *et al.*, 2008; Huynh *et al.*, 2010). The ZAL2 and $ZAL2^m$ are heteromorphic chromosomes, and together they comprise the second-largest chromosome pair in the white-throated sparrow genome (Thorneycroft, 1966). The two chromosomes differ from each other by a pair of nested inversions that suppress recombination across the majority of the chromosome, >100 Mb (Thorneycroft, 1975; Thomas *et al.*, 2008).

The inversion is of particular interest because it is linked to differences in plumage coloration, social behavior and mate choice (Tuttle, 2003). Approximately half the population is homozygous for the standard arrangement (ZAL2/ZAL2), and they have tan-striped (TS) crowns. The other half of the population is heterozygous for the inverted arrangement ($ZAL2/ZAL2^{m}$), and they have white-striped (WS) crowns (Falls & Kopachena, 1994). In general, the TS birds display more parental behaviors than their sex-matched WS counterparts whereas the WS birds display more aggressive sexual and territorial behavior than their sex-matched TS counterparts (Falls & Kopachena, 1994; Tuttle, 2003). The $ZAL2^{m}$ is maintained in the population through a strong pattern of disassortative mating, where the majority of matings are between a TS bird (ZAL2/ZAL2) and a WS bird ($ZAL2/ZAL2^{m}$) (Lowther, 1961). Because the $ZAL2^{m}$ is inherited in a Mendelian fashion, these pairings produce TS and WS offspring in equal proportions (Falls & Kopachena, 1994). This system is analogous to the ZW, where matings consist of one ZZ (male) and ZW (female) bird and produce ZZ and ZW offspring in approximately equal proportions (Falls & Kopachena, 1994).

Like the Z chromosome, which only recombines in males, the ZAL2 recombines in approximately half of the population, the TS birds. In the other half, the WS birds, recombination is restricted to a small (~5-Mb) collinear segment outside the inversion that is analogous to the pseudoautosomal region of sex chromosomes (Thorneycroft, 1975; Thomas *et al.*, 2008). Because no recombination occurs between the alternative chromosome arrangements within the inverted region, LD extends for >100 Mb (Huynh *et al.*, 2010). The unusually high LD would make the $ZAL2^{m}$ and the ZAL2 sensitive to

HRI and, in our previous study, we found that genetic diversity within the inversion region on both arrangements to be reduced relative to the region outside the inversion. Because our previous studies have only included a very limited sampling of the rest of the genome, we were not able to interpret the population genetic signatures associated with the chromosomal polymorphism in the context of normal patterns of diversity and LD in this species. Here we report the general population genetic patterns across the white-throated sparrow genome and evaluate those patterns in the context of chromosome size, autosomes versus sex chromosomes and in comparison to the ZAL2/ZAL2^{m} polymorphism.

3.2 Materials and methods

White-throated sparrow samples: White-throated sparrows and a female dark-eyed junco (*Junco hyemalis*) were collected in mist nets on the campus of Emory University in Atlanta, GA during November and December of 2005-2008. Blood was taken and plumage was determined visually and by PCR according to the methods described in Thomas *et al.* (2008). The Emory University Institutional Animal Care and Use Committee approved all procedures involving animals.

DNA sequencing: PCR primers for autosomal and Z chromosomes were designed in Primer3 (Rozen & Skaletsky, 2000) using white-throated sparrow BAC-end sequences that mapped to unique locations in the zebra finch genome (taeGut1) by MEGABLAST searches (-t 16, -N 2, -W 11, -e 1e-30) (Zhang *et al.*, 2000). For the W-linked loci, we

designed primers from a completely sequenced white-throated sparrow BAC from the W chromosome (GenBank Accession No.: AC236562) containing the *CHD1W* locus. To avoid potential amplification of gametologous Z chromosome loci, the W chromosome primers were designed to include at least three mismatches compared to the corresponding Z chromosome sequence and were used on a female (ZW)-only panel of individuals. A complete list of PCR primers, their orthologous positions in zebra finch and orthologous chromosome assignments in chicken are listed in Supplemental Table 3. Each 25μL PCR contained final concentrations of 1X PCR buffer, 1.5 mM $MgCl_2$, 20 pM of each primer, 0.2 mM of each dNTP, 1.5 U of Taq or Platinum Taq DNA polymerase (Invitrogen) and ~25 ng of genomic DNA. PCR cycling parameters were as follows: 94° for 5 min, 35 cycles of 94° for 30 sec, 55° for 30 sec and 72° for 1 min, followed by 72° for 7 min. Amplicons were subsequently purified and directly sequenced using the PCR primers.

SNP discovery and sequence annotation: Nucleotide polymorphisms were automatically called using SNPdetector (Zhang *et al.*, 2005) and manually confirmed prior to further analyses. Annotation of the gene features for each locus was based on the annotation of orthologous zebra finch genomic segments (taeGut1). Insertion-deletion polymorphisms and sites with more than two segregating alleles were excluded from our analyses. We assigned each locus to orthologous zebra finch and chicken chromosomes assuming conserved synteny (chromosome location). To examine population genetic patterns between differently sized autosomes, we grouped our sequence data by chromosome type according to the classification convention of the International Chicken

Genome Sequencing Consortium that divides chicken autosomes into three size classes: macrochromosomes (GGA 1-5), intermediate chromosomes (GGA 6-10) and microchromosomes (GGA 11-38) (Hillier *et al.*, 2004).

Population genetic analysis: Where we had sequence from paired BAC ends, genotypes were concatenated and phased as a single locus using Phase v2.1.1 (Stephens *et al.*, 2001). Haplotypes were then split into individual loci for population genetic analysis, except for calculation of LD and construction of haplotype networks. Analysis of polymorphism and tests of neutrality based on allele frequency spectrum were performed in DnaSP v5.1 (Librado & Rozas, 2009).

Using Splitstree v4.10 (Huson & Bryant, 2006), we generated haplotype networks from individual loci and when possible, the concatenated paired BAC-ends using the neighbor-joining algorithm and *J. hyemalis* sequence as an outgroup. To quantify population structure, F_{ST} values were calculated in Arlequin v3.11 (Excoffier *et al.*, 2005) and statistical significance was assessed with exact tests for genetic differentiation. In order to make a direct comparison between F_{ST} within the ZAL2/ZAL2^m system and the autosomal and sex-linked loci reported here, we calculated F_{ST} between the TS and WS groups. Estimates and confidence levels for the population recombination rate parameter $\rho = 4N_er$ were assessed using paired BAC end sequences by Monte Carlo coalescent simulations in the *interval* algorithm in LDhat (Auton & McVean, 2007) conditioned on θ_W from DnaSP sampling every 2,000 of 10^6 iterations burning the first 100,000 iterations. LD and r^2 values for all pairwise comparisons between informative sites,

including paired BAC-end sequences when possible, were calculated with Haploview v4.2 (Barrett *et al.*, 2005). We identified four-gamete pairs within all loci using DnaSP v5.1 (Librado & Rozas, 2009) and estimated the average haplotype block size using the four-gamete rule in Haploivew v4.2 (Barrett *et al.*, 2005).

3.3 Results

Data set: Summary statistics by chromosome category are shown in Table 3.1 (with previously reported ZAL2/ZAL2^{m} data included for reference) and detailed statistics for each locus can be found in Supplemental Table 4. In total, we sequenced ~10 kb from 27 autosomal loci (nine pairs of BAC-end sequences and nine additional loci), as well as ~6 kb from the 12 sex-linked loci, three of which were from the W chromosome (1.7 kb). The majority of the sequence was intergenic or intronic with a small fraction of protein coding positions (~2%). For autosomal and Z-linked loci, we sampled 9 - 12 birds with a minimum of four TS and four WS birds. For W-linked loci we sampled 23 - 24 females, with data from at least 14 TS and nine WS birds. In parallel, sequence data from a female *J. hyemalis* data was collected for all loci.

Table 3.1 Sampling information and diversity values for autosomal and sex-linked loci compared to previously reported values for ZAL2 and ZAL2^{m} alternative chromosome arrangements.

Chromosome Class	*Orthologous chromosome in chicken*	*Number of loci sampled*	*Number of chromosomes sampled*	*Total length sampled (bp)*	*S*	*Silent π^{a}*	*± SD*
Sex chromosomes	Z	9	18-22	3832 (3603)[b]	12	0.0005	± 0.00009
	W	3	46-48	1729	1	0.00005	± 0.00005
ZAL2[c]	3	58	16	34827 (13804)[b]	61	0.00072	± 0.00007
ZAL2^{m} [c]	3	58	8	34827 (13804)[b]	32	0.00039	± 0.00006
Chr2 outside of inversions[c]	3	4	24	1654 (777)[b]	17	0.00296	± 0.00041
Macrochromosomes	1, 2, 4	12	16-22	4773	44	0.00198	± 0.00033
Intermediate chromosomes	6, 7, 8, 9, 10	8	18-24	3101	87	0.00485	± 0.00054
Microchromosomes	17, 24, 27, 28	7	16-22	2083 (1980)[b]	136	0.01591	± 0.00115

[a]Average diversity is calculated from concatenated haplotypes from all loci in each category.

[b]Numbers in parentheses indicates the number of synonymous and noncoding positions.

[c]Data previously reported in a survey of standard ZAL2, inverted ZAL2^{m} and a region outside of the inversion polymorphism.

Patterns of nucleotide diversity: To understand patterns of diversity as they relate to chromosome size, we grouped our autosomal data into three categories: macrochromosomes, intermediate chromosomes, and microchromosomes. Based on the 267 identified polymorphic autosomal sites (excluding three tri-allelic SNPs), we observed a trend of increasing diversity with decreasing chromosome size (Figure 3.1). Overall, the amount of nucleotide polymorphism across the sparrow genome is highly variable. The nucleotide diversity across the classes of autosomes varied up to 8-fold, with the lowest diversity observed on the macrochromosomes ($\pi \pm SD = 1.98 \times 10^{-3} \pm 3.3 \times 10^{-4}$) and the highest diversity on the microchromosomes ($\pi \pm SD = 1.59 \times 10^{-2} \pm 1.0 \times 10^{-3}$). For all autosome types, the 95% confidence intervals (calculated as the mean $\pm$ $1.96 \times SD$) on diversity values did not overlap (Figure 3.1B).

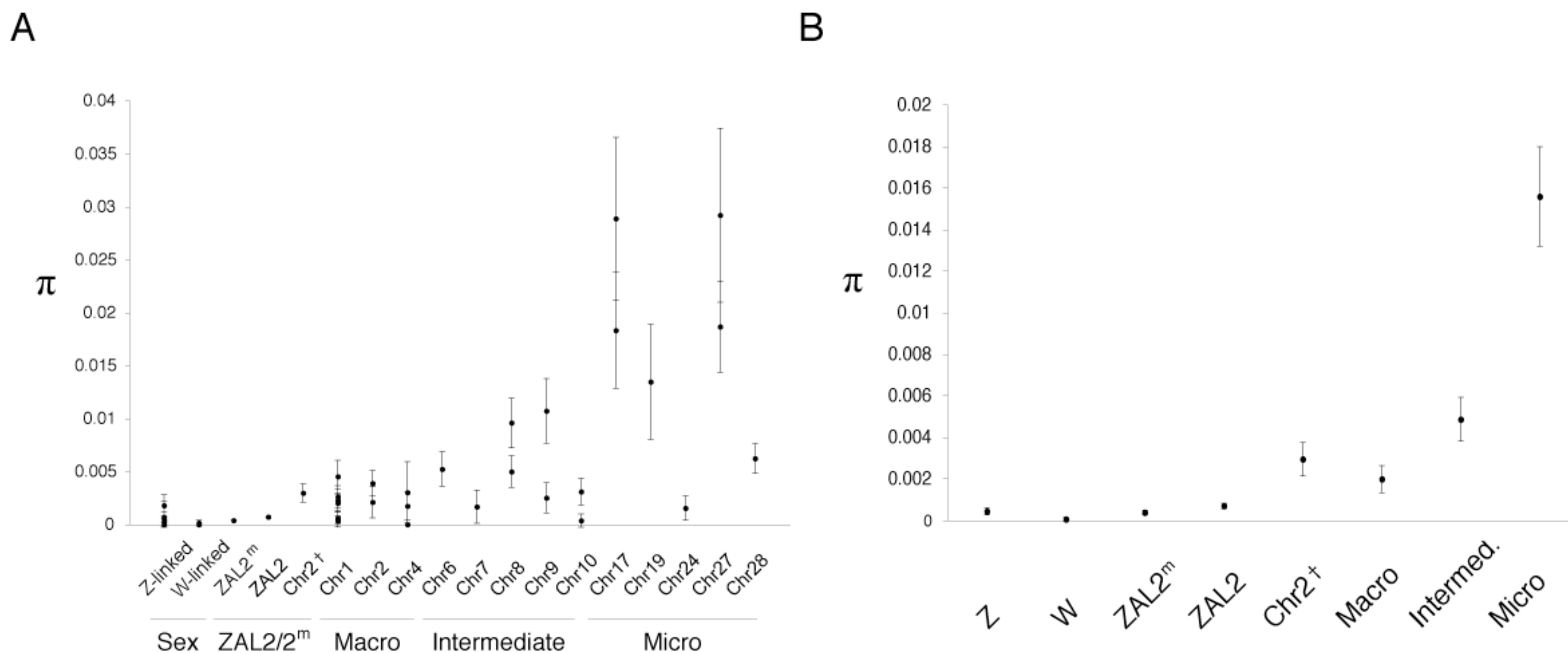

Figure 3.1 Average diversity across the sparrow genome. ZAL2 and $ZAL2^m$ refer to average π within the inversion interval on each chromosome arrangement and $Chr2^{\dagger}$ indicates the region outside the inversion on white-throated sparrow chromosome 2. They are shown for reference. A. Average genetic diversity (π) per locus. B. Average genetic diversity (π) for each chromosome type is calculated from concatenated sequences from all loci within that chromosome type. Error bars represent 95% CI (± 1.96 × SD). Note that the Chr2 label among the macrochromosomes refers to loci that map to chicken Chr2, which is not orthologous to white-throated sparrow chromosome 2.

To confirm the general relationship we observed in nucleotide diversity and chromosome size, we calculated the pairwise divergence between available pairs of complete BAC clone sequences representing alternative haplotypes from three macrochromosomes and a microchromosome. Consistent with our results from the sequencing of multiple small loci in multiple individuals, nucleotide diversity was lower on the macrochromosomes than the sampled microchromosome. Specifically, after excluding low-quality sites and annotated protein coding exons and UTRs, the π values for the macrochromosomes were 0.00120 ± 0.0006 (145,536 sites, orthologous to chicken chromosome 1, GenBank Ac. AC235523 and AC236253), 0.00172 ± 0.00086 (142,469 sites, orthologous to chicken chromosome 2, GenBank Ac. AC237008 and AC236908), 0.00081 ± 0.0004 (143,722 sites, orthologous to chicken chromosome 5, GenBank Ac. AC236607 and AC237119), whereas the π on the sampled microchromosome was 0.01632 ± 0.00816 (106,615 sites, orthologous to chicken chromosome 17, GenBank Ac. AC235993 and AC235934).

For the sex chromosomes we identified 12 Z-linked SNPs and one polymorphic site on the W chromosome. As expected, the Z and W sex chromosomes showed reduced diversity compared to the sampled autosomal loci ($\pi \pm SD = 5.0 \times 10^{-4} \pm 9.0 \times 10^{-5}$ and $5.0 \times 10^{-5} \pm 5.0 \times 10^{-5}$, respectively; see Table 3.1 and Figure 3.1) and were lower than all three classes of autosomes. Our estimate of π for the W chromosome was 10-fold lower than that of the Z chromosome and between 40- and 300-fold lower than diversity on the autosomes, depending on which chromosome type was used for comparison. Note that the high variance in the estimate was due to the observation of just a single SNP.

Previously, we reported that diversity levels observed within the inversion interval on the ZAL2 and ZAL2^{m} ($\pi \pm SD = 7.2 \times 10^{-4} \pm 7.0 \times 10^{-5}$ and $3.9 \times 10^{-4} \pm 6.0 \times 10^{-5}$, respectively) were lower than the region outside of the inversion ($\pi \pm SD = 3.0 \times 10^{-3} \pm 4.1 \times 10^{-4}$) (Huynh *et al.*, 2010). Having sampled other regions of the white-throated sparrow genome and considering our finding that nucleotide diversity is correlated with chromosome size, we believe that it is most appropriate to compare diversity within the ZAL2/ZAL2^{m} system to that of other macrochromosomes. We found that, although diversity within the inversion interval on both the ZAL2 and ZAL2^{m} is lower than observed on other macrochromosomes, it is similar to diversity on the Z chromosome, as indicated by overlapping confidence intervals. Outside the inversion, diversity is higher and similar to levels on other macrochromosomes.

Population structure: We previously reported extreme genetic differentiation between the ZAL2 and ZAL2^{m} chromosomes as a result of suppressed recombination between the alternate chromosome types (Thomas *et al.*, 2008; Huynh *et al.*, 2010), which is also apparent in comparisons of TS and WS individuals. To determine how extreme this pattern of genetic differentiation was, we calculated F_{ST} between the groups of TS and WS birds at all the sampled loci (Figure 3.2 and Supplemental Table 4). Most F_{ST} values for the autosomal and sex-linked loci clustered around zero, and we found no significant signal of population structure across the genome except within the ZAL2/ZAL2^{m} inversion interval (Figure 3.2). Additionally, no haplotype networks in this data set showed patterns of population structure (data not shown), indicating the structure

observed between TS and WS birds within the ZAL2/ZAL2^{m} system is exceptional with respect to the rest of the genome-wide sampled loci.

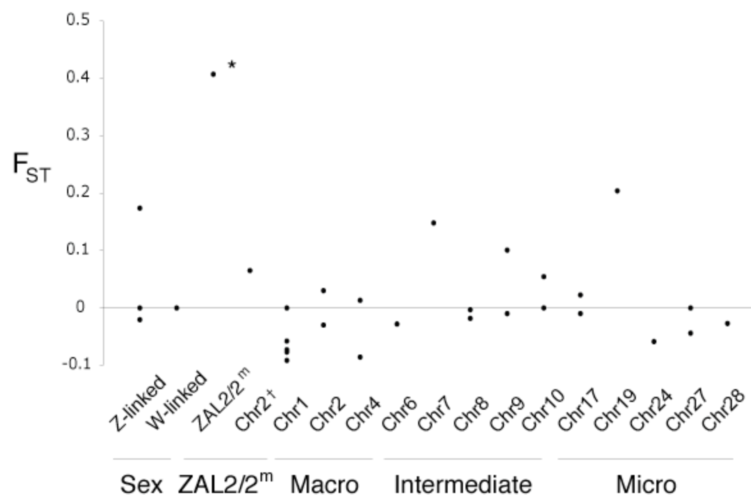

Figure 3.2 F_{ST} values between TS and WS morphs across the genome. The ZAL2/ZAL2^{m} data point represents F_{ST} within the inversion interval between the arrangements and Chr2† is the only value to show significant population structure ($p < 0.01$), indicated by the asterisk. The ZAL2 data point represents the F_{ST} outside the inversion on white-throated sparrow chromosome 2. Note that the Chr2 label among the macrochromosomes refers to loci that map to chicken Chr2, which is not orthologous to white-throated sparrow chromosome 2.

Allele frequency spectra: Tests based on allele frequency distributions can yield insights into demographic, population genetic and evolutionary processes acting upon a particular genomic region. Previously, we reported a skew towards intermediate frequency alleles within the ZAL2/ZAL2^{m} system, as a result of population structure between the two chromosome arrangements. This excess of intermediate frequency alleles led to a high, positive Tajima's D (Thomas *et al.*, 2008; Huynh *et al.*, 2010). To

establish whether this pattern was unique to the ZAL2/ZAL2^m system within the sparrow genome, we calculated Tajima's D for all sampled loci (Figure 3.3, Supplemental Table 4). As with the F_{ST} values, the Tajima's D associated with the ZAL2/ZAL2^m chromosomal polymorphism was a clear outlier compared to the other regions of the white-throated sparrow genome, where all but two autosomal loci showed negative values. Thus, with the exception of the inverted segment of the ZAL2/ZAL2^m system, the allele frequency spectrum across the rest of the white-throated sparrow genome revealed a general excess of rare alleles, consistent with neutral expectations.

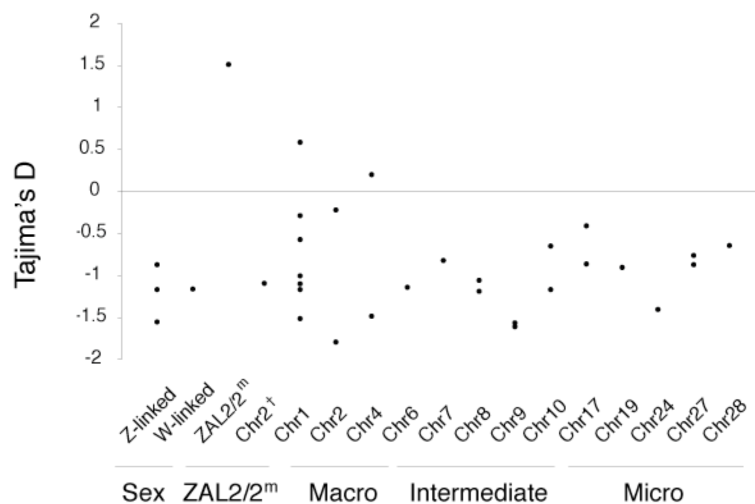

Figure 3.3 Tajima's D values across the sparrow genome. Tajima's D values are generally negative, indicating an excess of rare polymorphisms and consistent with neutral expectations. The ZAL2/ZAL2^m data point represents data from within the inversion on both chromosome arrangements, Chr2† represents Tajima's D for outside the inversion on white-throated sparrow chromosome 2. Note that the Chr2 label among the macrochromosomes refers to loci that map to chicken Chr2, which is not orthologous to white-throated sparrow chromosome 2.

Linkage disequilibrium and recombination: Recombination is expected to reduce LD and, in general, LD decreases with increasing distance between sites. Recombination rate estimates for paired BAC-ends grouped by chromosome size classes suggests that recombination rate is negatively correlated with chromosome size but estimates were associated with high error rates, suggesting that our small sample size did not support accurate measurements. For macrochromosomes $\rho \pm SD = 5.04 \times 10^{-6} \pm 4.95 \times 10^{-6} – 6.80 \times 10^{-6} \pm 6.52 \times 10^{-6}$ per base, for intermediate chromosomes $\rho \pm SD = 5.28 \times 10^{-6} \pm 5.29 \times 10^{-6} – 6.11 \times 10^{-6} \pm 6.05 \times 10^{-6}$ per base and the one microchromosome estimate $\rho \pm SD = 9.62 \times 10^{-6} \pm 9.86 \times 10^{-6}$ per base.

To compare LD associated with the ZAL2/ZAL2^{m} chromosomal polymorphism to that observed in the rest of the genome, we pooled r^2 values from informative pairwise comparisons within autosomal loci in this study and visualized the distribution of the r^2 values by generating a histogram of the r^2 values and as a function of distance (Figure 3.4A and B). The r^2 values from our autosomal data were generally low, as measured by ZnS = 0.14, a summary statistic which is based on the average correlation between pairs of sites (Kelly, 1997). Among the autosomal sites, LD decayed rapidly such that only limited LD was observed even within 500 bp (Figure 3.4). Furthermore, the proportion of statistically significant pairwise comparisons ($p < 0.01$) was lower on the macrochromosomes (8/42), intermediate chromosomes (5/229) and microchromosomes (54/794) relative to that within the ZAL2/ZAL2^{m} system (17,217/26,796). Consistent with the overall low levels of LD in the sparrow genome outside of the ZAL2/ZAL2^{m} system, the four-gamete test revealed evidence for recombination across all our

autosomal BAC-paired end loci (data not shown). Note that due to the limited number of informative pairwise comparisons on the Z (n = 3), no similar analyses could be performed on that chromosome. In contrast, we previously observed perfect LD between the majority of pairwise comparisons within the ZAL2/ZAL2^{m} inversion, independent of distance between sites and extending to >100 Mb (Figure 3.4C and D). Thus, in direct comparison to other regions of the white-throated sparrow genome the LD associated with the chromosomal polymorphism was extreme

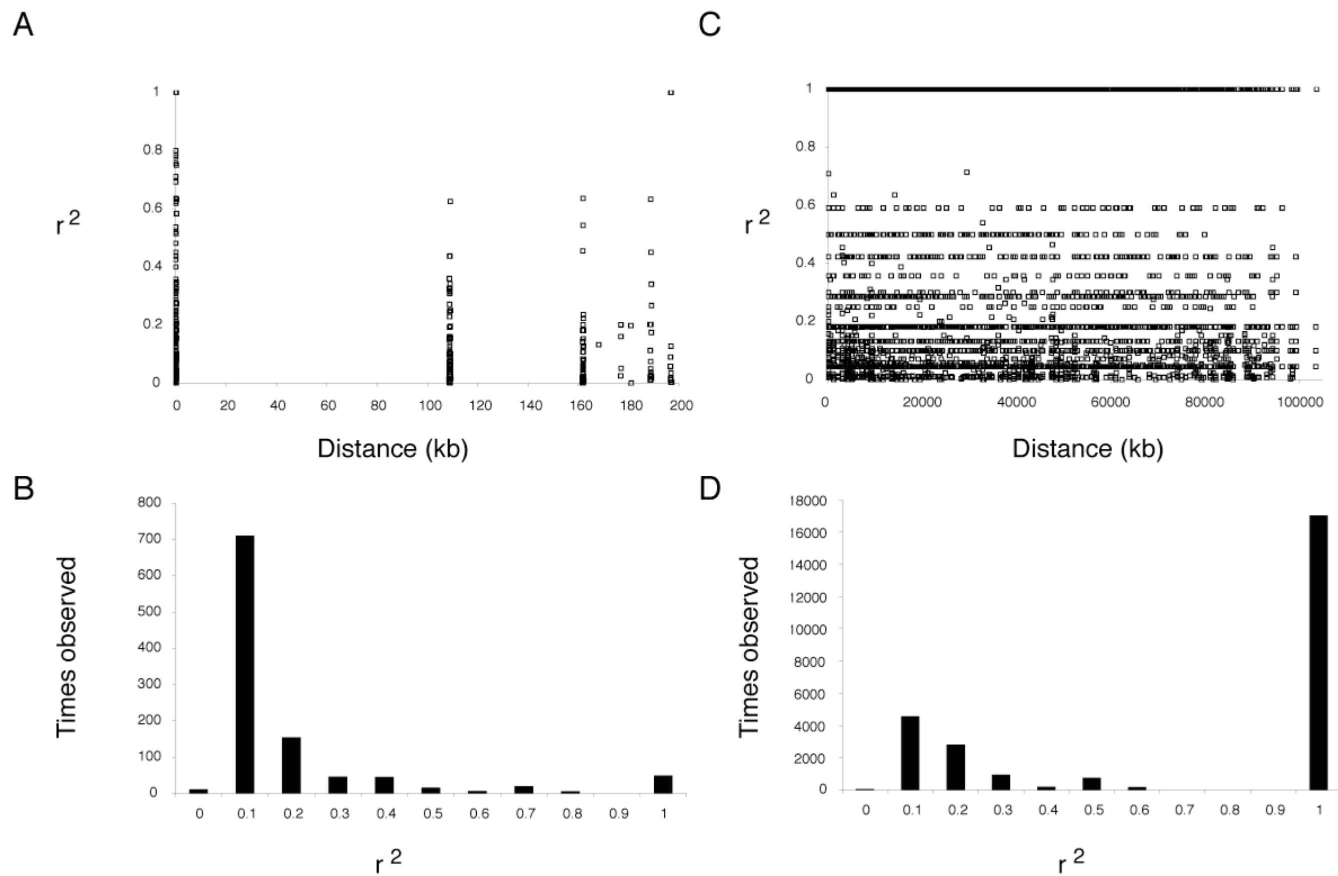

Figure 3.4 Linkage disequilibrium patterns within the sparrow genome. A. Patterns of LD between alleles among autosomal loci (excluding ZAL2/ZAL2^m data) show that LD is generally very low and slightly decreases with distance. B. A histogram indicates that the majority of pairwise comparisons among autosomal loci have $0 < r^2 \leq 0.1$. C. Patterns of LD between alleles among ZAL2/ZAL2^m loci show that high LD spanning >100 Mb. D. A histogram indicates that the vast majority of pairwise comparisons show high r^2, $0.9 < r^2 \leq 1$.

3.4 Discussion

Nucleotide variation, linkage disequilibrium and genetic structure are all fundamental parameters that describe the population genetics of a species. Initial studies of the white-throated sparrow population genetics focused on characterizing the unusual ZAL2 and $ZAL2^m$ chromosome system (Thomas *et al.*, 2008; Huynh *et al.*, 2010). In this study, we sequenced loci from other autosomes, as well as sex chromosomes to understand population genetic patterns elsewhere in the genome and to compare and contrast these patterns with those previously reported for the ZAL2 and $ZAL2^m$ chromosomes.

Contrasting patterns of nucleotide diversity within the sparrow genome: On average across all sparrow autosomes, one segregating site is observed every ~180 bp ($\pi = 5.62 \times 10^{-3} \pm 4.7 \times 10^{-4}$). This estimate is of the same magnitude of diversities calculated in other natural populations of passerine birds, such as *Ficedula* flycatchers, $\pi = 2.7 – 3.6 \times 10^{-3}$ (Primmer *et al.*, 2002; Borge *et al.*, 2005), *Carpodacus* finches, $\pi = 5.7 – 8.5 \times 10^{-3}$ (Wang *et al.*, 2003), the great reed warbler, $\pi = 1.2 \times 10^{-3}$ (Backström *et al.*, 2008) and the blue tit, $\pi = 1.8 \times 10^{-3}$ (Backström *et al.*, 2008). One factor that could account for the relatively high genetic diversity among avian species is large effective population size (N_e). We calculated N_e of the white-throated sparrow using the standard population genetic relationship $\theta = 4N_e\mu$, where μ is the mutation rate and π is used as an estimator of θ. Applying previously established avian mutation rates (2.95×10^{-9} substitutions/base/year (Balakrishnan & Edwards, 2009) and 1.5×10^{-9} substitutions/base/year (Ellegren, 2007), we estimate N_e to be on the order of 10^5–10^6,

similar to estimates for other birds, supporting the notion that natural bird populations tend to have larger effective population sizes than mammals (Berlin & Ellegren, 2004; Ellegren, 2007).

We stratified our data by chromosome size to study the differences in diversity among different chromosome types. Previous studies of avian genetic diversity have not stratified data by chromosome type and those data may represent a biased sampling of chromosome classes that could influence data interpretation. That we found significant variation in levels of polymorphism between differently sized chromosomes suggests that it may be necessary to account for chromosome location in future population genetic studies of avian genomes.

Nucleotide diversity across the sparrow autosomes varied substantially, spanning a range of nearly two orders of magnitude between individual loci ($\pi = 3.0 \times 10^{-4} - 2.9 \times 10^{-2}$). The microchromosome group showed the greatest diversity with eight times the genetic variation on the macrochromosomes. Polymorphism levels at two loci (chr24 and chr28) were noticeably lower than the other microchromosomal loci. Because the sparrow chromosome location is based on the assumption of conserved synteny (chromosome location) between the sparrow and zebra finch, it is possible that these loci were misassigned to the microchromosomes. Although synteny between avian genomes is generally conserved (Schmid *et al.*, 2000), chromosome reshuffling via fussion, fission and translocation can convolute these assignments (de Oliveira *et al.*, 2005; Hansson *et al.*, 2010). Despite the low genetic diversity observed in two of our seven

microchromosome loci, there is still strong evidence for increased diversity on the smaller chromosomes.

The negative correlation between nucleotide diversity and chromosome size class observed in our study was previously predicted for avian genomes based on two well-established observations. The first is that the rate of recombination on avian microchromosomes can be 5–10 times greater than on macrochromosomes (Hillier *et al.*, 2004). The second observation is that regions of high recombination tend to harbor greater genetic variation (Begun & Aquadro, 1992). Nevertheless, conclusions from previous studies that looked for this correlation in nucleotide diversity in the chicken genome reported conflicting results (Hillier *et al.*, 2004; Fang *et al.*, 2008; Megens *et al.*, 2009), and to our knowledge, no other systematic empirical study of genome-wide patterns of nucleotide diversity in birds has been reported. Our data clearly demonstrate a general negative correlation between chromosome size and diversity in the white-throated sparrow. Future studies will therefore be necessary to establish if this trends holds in other bird lineages. Because the avian karyotype is characterized by a high variation in chromosome size, bird genomes are likely to be informative models for this area of research and future studies in avian population genetics should take note of the chromosomal locations of their loci and interpret patterns in the context of known correlations between chromosome size and population genetic parameters.

As expected, the white-throated sparrow sex chromosomes showed the least diversity out of all the loci sampled. The neutral theory predicts a ratio of 1:3:4 for the genetic

diversities of the W, the Z and autosomes based on their relative N_es. Thus, Z chromosome to autosome diversity ($\pi_Z : \pi_A$) is expected to equal 0.75 (Sundström *et al.*, 2004). Because sex chromosomes are transmitted differently from autosomes, the standard models of how mutation, selection, drift, demography and mating system influence population genetic patterns may not apply (Ellegren, 2009a). The combined effects of increased HRI and lowered N_e due to variation in male mating success can reduce Z chromosome diversity relative to autosomal diversity (Sundström *et al.*, 2004; Borge *et al.*, 2005). Indeed, reports of avian $\pi_Z : \pi_A$ are lower than the expected 0.75: in *Ficedula* flycatchers $\pi_Z : \pi_A \approx 0.4$ (Borge *et al.*, 2005) and in chicken $\pi_Z : \pi_A \approx 0.25$ (Sundström *et al.*, 2004). Because the authors did not indicate the chromosome size class of the autosomal loci used for comparison, the degree to which the $\pi_Z : \pi_A$ is skewed is uncertain. Our finding that chromosome size and genetic diversity are negatively correlated suggests that it is appropriate to consider Z chromosome diversity in reference to similarly sized chromosomes (macrochromosomes). In the white-throated sparrow the $\pi_Z : \pi_A \approx 0.08$ when comparing Z chromosome diversity to average diversity across all autosomes. If Z chromosome diversity is compared to the diversity of the average macrochromosome, then $\pi_Z : \pi_A \approx 0.25$, illustrating the importance of considering chromosome type when establishing these kinds of relationships among data within a genome. Nevertheless, our data are consistent with other avian estimates, as well as with the hypothesis that sexual selection works to reduce diversity on the Z chromosome.

Under the neutral model, the W chromosome-to-autosome diversity is predicted to be 1:4 or 0.25; however, studies of the avian W chromosomes report far lower diversity values.

For example, in a survey of ~3.4 kb from >150 W chromosomes in seven avian species, Montell *et al.* (2001) did not observe any polymorphisms. Berlin and Ellegren (2004) identified a single segregating site in a survey of ~8 kb in 47 chickens from divergent breeds and estimated $\pi_W = 7 \times 10^{-5}$, ~1/100 of their π_A estimate (Sundström *et al.*, 2004). Similarly, we observed exceedingly low levels of variation on the W, 100-fold lower than average autosome diversity and 30-fold lower than the macrochromosomes. Although it is likely that the reduced variation on the non-recombining sex chromosome (Y or W) is universal for both male and female heterogametic systems, as Y chromosomes also show lower than expected diversity levels (Hellborg & Ellegren, 2004; Charlesworth *et al.*, 2005), the chicken W diversity could be exceptionally low due to a founder effect associated with domestication (Berlin & Ellegren, 2004). Furthermore, the genetic linkage between the W chromosome and the non-recombining mitochondrial genome further promotes the loss of genetic diversity on the W through HRI (Berlin *et al.*, 2007). The genetic diversity we observed on the white-throated sparrow W chromosome is on the order of the chicken W diversity and is consistent with a recent population bottleneck.

Within the white-throated sparrow genome, dramatic HRI is probably not limited to the sex chromosomes. We previously reported patterns of reduced recombination and extensive LD within the ZAL2/ZAL2^m system, which bears striking similarities to sex chromosomes (Thomas *et al.*, 2008; Huynh *et al.*, 2010). In addition to the recombination suppression within the inversion interval due to the inversions, avian macrochromosomes are disproportionately associated with recombination deserts, except close to the chromosome ends (Backström *et al.*, 2010). This low rate of recombination

for macrochromosomes is likely to result in intrinsically low diversity values, which are further reduced by HRI on the ZAL2 and ZAL2^{m}.

Outside the inversion on white-throated sparrow chromosome 2, diversity is similar to the macrochromosome average. This finding is surprisingly low considering the amount of recombination predicted to occur in this small region. There are two reasons to believe that recombination rates are greatly elevated outside of the inversions. First, obligate crossing-over in half of the white-throated sparrow population (WS birds, ZAL2/ZAL2^{m}) is restricted to this ~5 Mb interval. Second, Backström *et al.* (2010) reported a strong telomere effect in the zebra finch genome, where chromosome ends show highly elevated rates of crossing-over making the recombination landscape at all chromosome ends equivalent to that of microchromosomes. Although increased recombination in this interval may preserve standing variation, this region is subjected to HRI from selection on both the ZAL2 and ZAL2^{m} when in ZAL2/ZAL2^{m} heterozygotes, which will reduce variation. The balance of these forces results in a level of genetic diversity lower than expected, but similar to other macrochromosomes.

Although the ZAL2/ZAL2^{m} system bears many similarities to the XY (ZW) sex chromosomes, including differences in patterns of recombination, disassortative mating, reduced genetic diversity, the ZAL2/ZAL2^{m} are not sex-linked meaning that sexual selection (Caballero, 1995), sex-biased demographic history (Hammer *et al.*, 2008; Keinan *et al.*, 2009), and mitochondrial linkage (Berlin *et al.*, 2007) cannot easily explain patterns of evolution within the ZAL2/ZAL2^{m} system. For these reasons, the

ZAL2/ZAL2^{m} system remains an informative point of comparison for the study of population genetic patterns of sex chromosomes and can help distinguish between sex-specific and fundamental molecular evolutionary processes that shape their evolution.

LD within the sparrow genome: The extent of LD is dependent on two key parameters: N_e and recombination rate (Hillier *et al.*, 2004; Stapley *et al.*, 2008; Stapley *et al.*, 2010). Although the extent of LD within avian genomes from natural bird populations is largely unknown, the few studies currently available indicate that LD is generally low on avian autosomes as a consequence of overall high levels of recombination (Edwards & Dillon, 2004; Balakrishnan & Edwards, 2009; Megens *et al.*, 2009). Our autosomal data are consistent with these reports and we find no signals of LD even between sites within a few hundred bases of each other. Furthermore, almost nothing is known about the relationship between chromosome size and LD. Megens *et al.* (2009) reported a correlation between LD and chromosome size, such that macrochromosomes exhibit consistently higher LD and larger haplotype blocks when compared to microchromosomes. Data from our study support previous findings that contrasting patterns of LD on the macro- and microchromosomes are consistent with their respective differences in recombination rate (Hillier *et al.*, 2004). Additionally, our white-throated sparrow data are consistent with other reports demonstrating generally low levels of LD throughout the avian genome (Backström *et al.*, 2006).

There is one striking exception to the general pattern of low LD within the sparrow genome. Our previous study of the ZAL2 and ZAL2^{m} chromosomes revealed extensive

LD spanning >100 Mb, thus representing one of the largest segregating haplotypes to be reported (Huynh *et al.*, 2010). Given our findings here, as well as the general patterns of low LD reported in other bird autosomes (Edwards & Dillon, 2004; Balakrishnan & Edwards, 2009; Megens *et al.*, 2009), we conclude that the LD observed between the ZAL2 and ZAL2^{m} is unusual within the sparrow genome as well as amongst avian genomes in general.

Population structure: Our previous study of the ZAL2/ZAL2^{m} system indicated extensive population structure within the inversion polymorphism. The majority of the segregating sites (~70%) were fixed differences and $F_{ST} = 0.94$ ($p < 0.01$) within the inversion between the alternate arrangements (Huynh *et al.*, 2010). The high proportion of fixed differences also produced large and positive Tajima's D values, as it resulted in the presence of many intermediate frequency variants. In this study of autosomal and sex-linked loci, we found no significant population structure between WS and TS birds outside of the ZAL2/ZAL2^{m} inversion.

Conclusions: Our population-based sequencing survey of autosome and sex chromosome loci in the white-throated sparrow provides empirical evidence for the predicted negative correlation between chromosome size and nucleotide diversity predicted in avian genomes. In addition, we found that the patterns of nucleotide diversity, population structure, and LD previously associated with the ZAL2/ZAL2^{m} chromosomal polymorphism in this species are atypical compared to other macrochromosomes.

4 Characterization of sequence variants in candidate genes linked to behavior in the white-throated sparrow (*Zonotrichia albicollis*)

4.1 Introduction

The genetic basis for social behavior has long been of interest to biologists. In some cases, social behavior has shown high heritability, but our understanding of the role of specific genes in social behavior are limited to a few species (Robinson *et al.*, 2008). The white-throated sparrow (*Zonotrichia albicollis*) has recently emerged as a promising system in which to study genes and behavior. This species exhibits two plumage morphs that are associated with different suites of social behavior, summarized in Table 4.1. These behaviors have been described as alternative reproductive strategies. Individuals with white-striped crowns (WS) engage in a competitive reproductive strategy and those with tan-striped (TS) crowns engage in a parental reproductive strategy (reviewed in Tuttle 2003 and summarized in Table 4.1). WS males sing more and they are more likely to attract multiple females to their territory; they also are more likely to invade other males' territories to engage in extra-pair copulations (Tuttle, 2003). WS females also exhibit "aggressive behavior" as they sing and defend territory (Kopachena & Falls, 1993a; Tuttle, 2003). In contrast, TS males invest less in aggressive and intrusion behaviors and spend more time to feeding nestlings than WS males (Knapton & Falls, 1983; Kopachena & Falls, 1993c). TS females do not sing nor defend territory (Tuttle, 2003). Considering that behavioral differences observed between the WS and TS morphs, the white-throated sparrow is a particularly attractive model for studying aggression and parental care.

Table 4.1 Behavioral differences observed in the white- (WS) and tan- (TS) striped birds (modified from Tuttle 2003).

	White-striped		Tan-striped
MALES			
Aggressive behavior	WS	>	TS
Intrusion behavior	WS	>	TS
Pursuit of extra-pair copulations	WS	>	TS
Extra-pair fertilization	WS	>	TS
Parental care	WS	<	TS
Mate-guarding	WS	<	TS
FEMALES			
Aggression	WS	>	TS
Sexual solicitation rate	WS	>	TS
Parental care	WS	<	TS

Although many species show polymorphisms in behavior, the white-throated sparrow provides a unique to link genes and behavior because the differences in social behavior are associated with a large inversion on chromosome 2. WS birds, which display more aggressive behaviors, have at least one copy of the inversion and are almost always inversion heterozygotes (ZAL2/ZAL2^{m}); however, two rare WS birds in a collective sample of >1300 white-throated sparrows were found to be homozygous for the inversion (Falls and Kopachena 1994, D. Maney personal communication, Romanov *et al.*2009, Thorneycroft 1975). TS birds, which allocate more time on parental care, are invariably ZAL2 homozygotes. Because the ZAL2^{m} polymorphism occurs on an autosome, the differences in coloration and social behavior in the white-throated sparrow are not sex-specific, making it a unique model in which the genetics and neuroendocrine basis of behavior can be studied independent of sex (Maney, 2008).

Inversions have been associated with a variety of phenotypes and previous studies have localized candidate regions within inversions by examining patterns of recombination between the wild-type and inverted arrangements and identifying region in LD as targets

of selection. For example, in the mouse *t* haplotypes, the characterization of rare recombinant chromosomes has greatly elucidated the genetic architecture of the TRD phenotype, as well as identified critical genes (Silver & Remis, 1987; Hermann *et al.*, 1999; Wallace & Erhart, 2008). In another example, White *et al.* (2007) examined patterns of divergence, recombination and LD to localize two candidate regions, ~1.5 Mb each, involved in aridity tolerance in the 2La inversion in *Anopheles gambiae* mosquito.

In the case of the ZAL2/ZAL2^{m} polymorphism, the sustained LD across the entire inversion described in Chapter 2 precludes the further localization of candidate regions or genes by standard recombination-based mapping. Taking these limitations into consideration, a candidate gene approach may be the only way to identify mutations underlying the phenotypic variation associated with the ZAL2^{m} inversion. We know that the genetic alterations underlying the behavior will show certain characteristics. The mutations occur within the inversion or are associated with the inversion breakpoints. The mutations are dominant, as the differences in behavior are apparent in the heterozygote. Finally, the mutations will be fixed differences observed between the ZAL2 and ZAL2^{m} chromosomes.

If we consider that ZAL2/ZAL2^{m} divergence is on the order of 1%, make a conservative estimate that 50% of the differences are fixed between the arrangements and that the inversion is ~100 Mb, we can predict that there are at least 500,000 single-nucleotide differences that distinguish the ZAL2 from the ZAL2^{m}. Thus, it is possible and likely that the TS and WS birds differ from each other in hundreds of traits. Nonetheless,

understanding that we expect to observe some differences between the ZAL2 and ZAL2^m alleles in cis-regulatory or protein coding portions of essentially all ~1,000 genes within the inversion, it would be of significant interest to characterize candidate genes (those which have been previously established to modulate social behavior in other systems) that map within the inversion.

Taking into consideration that aggressive, parental and social bonding behaviors are influenced by gonadal steroids in humans and birds (Feder, 1984; Wingfield, 1994; Nelson, 2000), we identified a set of candidate genes involved in these pathways within the ZAL2^m inversion of the white-throated sparrow. We have selected candidate genes based on their association with the hypothalamic-pituitary-gonadal (HPG) axis (Table 4.2) and include genes that encode 1. enzymes in the gonadal steroid synthesis pathway; 2. receptors of gonadal steroid hormones; and 3. pituitary hormones with pro-gonadal effects. We also selected genes based on their previous association with aggressive or parental behavior. Using information from the orthologous zebra finch chromosome 3, we identified those genes with behavioral effects that localized within the ZAL2^m inversion. Our set of nine promising candidate genes and their predicted locations on the white-throated sparrow ZAL2 chromosome are shown in Figure 4.1. Their characteristics are listed in Table 4.2.

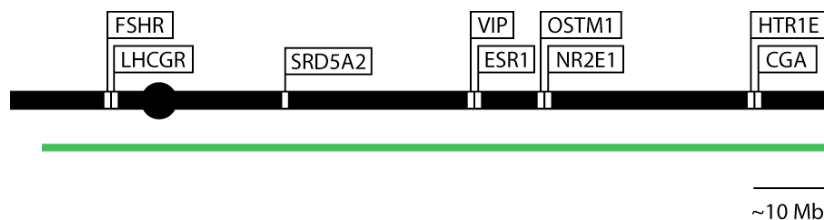

Figure 4.1 Predicted locations of our nine candidate genes shown on the ZAL2 chromosome orientation. The centromere is represented by the closed circle and the green line indicates the segment of chromosome involved in the ZAL2^m inversion polymorphism.

Table 4.2 Candidate gene characteristics.

Gene Symbol	*Gene name*	*HPG*[a]	*Aggression*	*Parenting*
ESR1	Estrogen receptor 1, alpha	√	√	√
CGA	Glycoprotein hormones, alpha polypeptide	√		
FSHR	Follicle stimulating hormone receptor	√		
LHCGR	Luteinizing hormone/choriogonadotropin receptor	√	√	
SRD5A2[b]	Steroid-5-alpha-reductase, alpha polypeptide 2	√		
VIP	Vasoactive intestinal peptide	√	√	√
NR2E1	Nuclear receptor subfamily 2, group E, member 1		√	
HTR1E	5-hydroxytryptamine (serotonin) receptor 1E		√	
OSTM1	Osteopetrosis-associated transmembrane protein 1 (plumage candidate)			

[a]HPG refers to genes associated with the hypothalamic-pituitary-gonadal axis.
[b]The five exons of SRD5A2 were directly sequenced.

Candidate gene descriptions

CGA: The gonadotropic hormones follicle stimulating hormone (FSH), luteinizing hormone (LH) and thyroid stimulating hormone (TSH) are all dimers that consist of two subunits, alpha and beta. The alpha subunit of all proteins is encoded by the same gene, CGA (Fiddes & Goodman, 1979). FSH and LH are known to stimulate the secretion of gonadal steroids, which have been associated with reproductive and aggressive behaviors, and TSH stimulates the release of thyroid hormones, which have been previously

associated with mood disorders (reviewed by Nelson, 2000). Because CGA is associated with molecules known to be active in the HPG axis and the thyroid axis, sequence variants of the CGA subunit could potentially modify social and sexual behavior through both axes.

FSHR and LHCGR: FSH and LH are two principle gonadotropins found in all vertebrates, including birds and humans (Farner & Wingfield, 1980), and the genes encoding their receptors (FSHR and LHCGR) are captured by the $ZAL2^m$ inversion in the white-throated sparrow. Proper expression and function of these hormones as well as their receptors are necessary for reproductive development in both males and females. In the male, LH stimulates the production of testosterone, a hormone widely known to influence aggressive behaviors (Nelson, 2000).

SRD5A2: 5-alpha reductase (SRD5A2) is an enzyme that metabolizes testosterone to dihydrotestosterone (another androgen) and is crucial in maintaining levels of androgens and estrogens. Mutations of SRD5A2 are associated with pseudohermaphroditism and have been shown to influence gender identity in humans (Thigpen *et al.*, 1992; Wilson *et al.*, 1993). In birds, it is believed that expression levels of androgens and estrogens play a role in sexual differentiation and behavior (Schlinger, 2001). In studies of the song sparrow (*Melospiza melodia*), Soma *et al.* found that SRD5A2 activity varies seasonally, demonstrating that that steroid synthesis is a dynamic process that could influence behavior in birds (Soma *et al.*, 2003).

VIP: Vasoactive intestinal polypeptide (VIP), found in gastrointestinal tissues as well as neural tissues, is an important developmental mediator in all vertebrates (Dietl *et al.*, 1990; Sherwood *et al.*, 2000). In birds, VIP influences parental behaviors like incubation, broodiness and care of young (Zhou *et al.*, 2008a). In the white-throated sparrow, Maney *et al.* reported both sex and morph differences in VIP expression in the brain, making this gene a good candidate to explore as a behavioral modifier (Maney *et al.*, 2005).

ESR1: Estrogen receptor 1 (ESR1) is essential for fertility, development and sexual differentiation. Because of this crucial role, the ESR1 gene is highly conserved among all vertebrates, including birds and humans (Kohno *et al.*, 2008). In zebra finch, estrogen modifies neural circuitry and masculinizes neurological features in females (Arnold, 1998). Sex differences in song circuitry in birds could be due to differences in estrogen level as well as differential estrogen receptor expression (Jacobs *et al.*, 1996; Holloway & Clayton, 2001). ESR1 also has an established role in influencing aggression in mice (Ogawa *et al.*, 1996; Ogawa *et al.*, 1997; Ogawa *et al.*, 1998). When ESR1 is knocked-out in the female mouse, the resulting phenotype shows increased aggression and decreased parental behavior (Ogawa *et al.*, 1998). In males ESR1 knock-out produces a less aggressive phenotype when compared to normal wild type male mice (Ogawa *et al.*, 1997). The association of ESR1 with increased aggression, decreased parental behavior and the $ZAL2^{m}$ inversion make this gene our most promising candidate for modifying behavior in the white-throated sparrow.

HTR1E: The HTR1E gene codes for one of several receptors for the neurotransmitter serotonin or 5-hydroxytryptamine (5-HT). In the brain, the serotonergic system modulates various cognitive and behavioral functions, including mood, anxiety, learning, sleep, feeding and respiratory activity (Olivier & van Oorschot, 2005; Sperry *et al.*, 2005). Because HTR1E is highly expressed in the human brain and shows high sequence conservation among human populations, it has also been hypothesized to have an important physiological role (Shimron-Abarbanell *et al.*, 1995). Although little else is known about this receptor, other closely related serotonin receptors have been implicated changes in aggressive and sexual behaviors (Saudou *et al.*, 1994; Popova & Amstislavskaya, 2002). Additionally, in sparrows, serotonin has been reported to influence aggressive behavior (Sperry *et al.*, 2003; Sperry *et al.*, 2005); thus, any of its receptors are good candidates for studying the tendency for aggressive behavior associated with WS white-throated sparrows.

NR2E1: Nuclear receptor 2E1 (NR2E1) is expressed in vertebrate forebrains at all developmental stages (Shi *et al.*, 2004). NR2E1 regulates cortical neurogenesis and the proliferation and differentiation of neural progenitor cells (Shi *et al.*, 2004). Interestingly, mutations in NR2E1 in mice are associated with pathological aggression and violence, as well as ocular abnormalities (Young *et al.*, 2002). Although the mechanism for modifying behavior has not yet been established, human NR2E1 is known to rescue the mouse knock-out and restore normal behavior (Abrahams *et al.*, 2005), demonstrating the conserved function of NR2E1 between mouse and human.

OSTM1: The OSTM1 gene encodes for the osteopetrosis-associated transmembrane protein 1 and defects in this gene have been associated with osteopetrosis, a rare recessive disorder characterized by dense bones (Chalhoub *et al.*, 2003). In studies of mice, a single point mutation in the OSTM1 gene has been shown to disrupt the regulation of pigment synthesis and melanocyte survival as well as genes involved in bone degradation (Chalhoub *et al.*, 2003). This gene is highly expressed in the mouse melanocyte and it is our only plumage coloration candidate in this study.

4.2 Materials and methods

White-throated sparrow samples: White-throated sparrows and a dark eyed junco (*Junco hyemalis*) were collected using mist nets on Emory University's campus in Atlanta, GA in 2005. DNA was extracted from blood or tissue and used to determine plumage morph according to the methods described in Thomas *et al.* (2008). All procedures were approved by Emory University's Institutional Animal Care and Use Committee.

DNA sequencing and analysis: BAC clone sequencing was conducted by the NIH Intramural Sequencing Center (NISC), downloaded from GenBank and annotated using the chicken genome as a reference. To screen for fixed SNPs, we used a panel of 2 TS and 2 WS; we sequenced the Junco for reference. We developed primers directly from white-throated sparrow sequence from BAC clones using Primer3 (Rozen & Skaletsky, 2000). As the BAC library was created from a WS bird (#822), we identified BAC

clones representing both haplotypes to identify SNPs between the ZAL2 and $ZAL2^m$ arrangements. The primers were designed to amplify 400-800 bp flanking the SNP or SNPs of interest. A complete list of PCR primers and their orthologous locations on in zebra finch are listed in Supplemental Table 5.

Each 25μL PCR contained final concentrations of 1X PCR buffer, 1.5 mM $MgCl_2$, 20 pM of each primer, 0.2 mM of each dNTP, 1.5 U of Taq or Platinum Taq DNA polymerase (Invitrogen) and ~25 ng of genomic DNA. PCR cycling parameters for all primers, except ESR1_543 primers, were as follows: 94° for 5 min, 35 cycles of 94° for 30 sec, 55° for 30 sec and 72° for 1 min, followed by 72° for 7 min. PCR parameters for ESR1_543 were as follows: 94° for 5 min, 15 cycles of 94° for 30 sec, 55° for 30 sec and 72° for 1 min, 20 cycles of 94° for 30 sec, 55° for 30 sec and 72° for 1 min and a final extension of 72° for 7 min. Amplicons were subsequently purified and directly sequenced with the PCR primers.

Sequences were aligned using PhredPhrap (Ewing & Green, 1998; Ewing *et al.*, 1998) and consed (Gordon *et al.*, 1998) and manually analyzed to determine whether any of the detected nonsynonymous variants represented likely fixed differences.

4.3 Results

Among the 9 candidate genes within the $ZAL2^m$ inversion, 16 nonsynonymous SNPs were identified through BAC sequence comparisons (Table 4.3). No protein coding

changes were observed in the genes CGA, VIP, NR2E1 or the four exons of SRD5A2 that we were able to sequence. We were unable to re-sequence exon 1 of SRD5A2 as well as 3 other SNPs within the LHCGR and OSTM1 genes and as a result, we have no data on our only candidate for plumage coloration differences. Of the 13 SNPs we were able to sequence in our panel of 2 TS and 2 WS birds, 11 were fixed differences.

Table 4.3 Sequencing results of SNPs from candidate genes within the ZAL2^{m} inversion.

Gene symbol	*# of candidate SNPs identified*	*Amino acid position*[a]	*Amino acid change* $(2 \rightarrow 2^{m})$[b]	*Fixed difference?*
ESR1	1	543	A → T	Yes
CGA	0			
FSHR	5	35	S → R	Yes
		42	Q → R	Yes
		240	L → F	Yes
		484	R / C	No
		621	S → A	Yes
LHCGR[c]	6	21	H / R	no data
		47	A / P	no data
		301	S → N	Yes
		595	R → K	Yes
		639	V → I	Yes
		693	A → T	Yes
SRD5A2[d]	0			
VIP	0			
NR2E1	0			
HTR1E	3	25	V → I	Yes
		244	S / L	No
		273	L → P	Yes
OSTM1[c]	1	33	L / V	no data

[a]Amino acid positions based on chicken orthologs downloaded from GenBank, ESR1 NM_205183, FSHR P79763, LHCGR NP_990267, OSTM1 NP_001026248,
[b]SNPs were polarized to the ZAL2 and ZAL2^{m} only if they were fixed differences.
[c]Unable to sequence region flanking 2 SNPs in LHCGR and the OSTM1 SNP.
[d]Unable to sequence exon 1 of SRD5A2. No SNPs were identified in the remaining 4 exons.

4.4 Discussion

It is not surprising that we did not identify any nonsynonymous mutations in several genes in this study because the coding sequences of these genes are highly conserved.

Additionally, our high frequency of fixed differences is consistent with our chromosome-wide study (Chapter 2), indicating that ~70% of all the sequence differences between the ZAL2 and ZAL2^m arrangements are indeed fixed differences. Future studies of these polymorphisms and other nonsynonymous changes should consider other attributes of the polymorphisms, such as whether they occur in highly conserved residues or active sites of functional domains.

We provide a small subset of candidate mutations for future studies, but in order to understand the genetic basis of the social behavior and plumage polymorphisms, parallel studies of gene expression and protein activity will likely be necessary. For example, protein specific assays that compare the activity of the ZAL2 and ZAL2^m protein variants could be used to determine if the nonsynonymous differences we identified affect the function of the proteins. In cases where no differences in the ZAL2 and ZAL2^m coding sequence were detected, it is possible that differences in cis-regulatory elements may alter the expression patterns between the WS and TS birds. Thus, it will be important to characterize and compare the gene expression patterns and levels in WS and TS birds in conjunction with analysis of the coding sequence from the ZAL2 and ZAL2^m chromosomes.

In summary, we present a set of 11 protein-coding mutations in genes known to influence social behavior that are likely to be fixed between the ZAL2 and ZAL2^m chromosome arrangements. These polymorphisms are appropriate candidates for future studies of the genetic basis of the behavioral in the white-throated sparrow.

5 Discussion

The white-throated sparrow (*Zonotrichia albicollis*) is a species of interest to backyard birders and ornithologists. Recently, it has drawn much attention from biologists because the white-throated sparrow harbors an extraordinary chromosomal polymorphism that influences plumage coloration, mate choice and social behavior. Although the polymorphism was first identified in the 1960s (Thorneycroft, 1966), the first modern genetic characterization of the $ZAL2^m$ polymorphism was conducted in 2008 (Thomas *et al.*, 2008). Our lab characterized the $ZAL2^m$ as a complex polymorphism consisting of a pair of nested inversions spanning >100 Mb and encompassing ~10% of the entire genome. In a study of 10 loci, Thomas *et al.*identified patterns of increased divergence between the homologous ZAL2 and $ZAL2^m$ chromosomes and proposed this system as a model for the study of sex chromosome evolution as well as the evolution of social behavior (Thomas *et al.*, 2008). The work presented in this dissertation delves deeper into the patterns of molecular evolution both within the ZAL2/$ZAL2^m$ system, as well as in the white-throated sparrow genome. We propose that this model can address broader questions of how basic processes in molecular evolution shape patterns of genetic variation and ultimately shed light on the process of evolution at the molecular level. Additionally, we used a candidate gene approach to characterize protein-coding changes that could underlie the social behaviors linked to the inversion.

The coadaptation and local adaptation models for inversion evolution depend on an inversion's capacity to suppress recombination and maintain LD among the adaptive genes. Considering the sheer number of genes captured by the $ZAL2^m$, there is considerable opportunity for coadaptation among genes captured by the inversion or the

presence of multiple beneficial alleles within the haplotype. Selection for the ZAL2^{m} inversion is compatible with either the coadapted or locally adapted gene complex theories for inversion evolution. It should be noted that these data do not rule out the action of position effects associated with selection for the ZAL2^{m}. Breakpoint mapping of the inversions has not revealed physical interruption of any gene and it is possible that some genes may be disassociated from regulatory elements by the inversions. The role of possible effect in the phenotypes linked to the ZAL2^{m} inversion is a topic for future research.

In Chapter 2, our study of the patterns of genetic diversity within the ZAL2/ZAL2^{m} system suggests that two types of forces dominate the evolution the ZAL2/ZAL2^{m} system: balancing selection maintaining by disassortative mating and Hill-Robertson interference, which includes both positive and negative selection on the variants captured by the inversion. Interestingly, these processes have contrasting outcomes, which can be most clearly seen by their effects on the site frequency spectrum. Balancing selection is evidenced by the excess of intermediate frequency variants observed on the ZAL2 and ZAL2^{m} chromosomes. The balancing selection is a result of the disassortative mating in the white-throated sparrow, which also ensures that ZAL2^{m} homozygotes are rare in the population and can facilitate the accumulation of recessive deleterious mutations causing increase variation on the ZAL2^{m}. We predicted that Hill-Robertson effects play a strong role in the ZAL2/ZAL2^{m} system and because we did not observed increased variation on the ZAL2^{m}, we hypothesize that Hill-Robertson interference overcomes the potential accumulation of recessive lethal variants. Within each group, Hill-Robertson interference

produces a dramatic excess of rare variants. The role of Hill-Robertson interference in shaping the genetic diversity within inversion interval on both chromosomes implies a weakened efficiency of selection in both of these regions. Although we did not detect the increased rate of mutation that characterizes regions of very low or no recombination, other characteristics associated with reduced efficacy of selection could also be explored, such as reduced GC content, lowered codon bias or divergence at nonsynonymous sites.

The complete linkage within the ZAL2^{m} polymorphism creates a difficulty in identifying regions under selection. If a sweep were detected within the inversion, it would be impossible to localize the gene or even the region under selection because all sites in the ZAL2^{m} are linked. Interestingly, the same may be true for the region outside the inversion but for a different reason. If a selective sweep were to occur in the short distal segment outside the inversion, signatures of that sweep would be quickly erased because of the increased frequency of recombination in this region (Wang *et al.*, 1999). It is a possibility that one could detect a recent selective event by identifying regions of reduced genetic diversity, however, in some cases of extreme recombination, the length of selective sweeps is exceedingly small (Wang *et al.*, 1999).

Although selective sweeps and background selection are hypothesized to play a large role in ZAL2^{m} evolution, it is notoriously difficult to distinguish between these two processes in regions of low or no recombination (Charlesworth & Charlesworth, 2000; Innan & Stephan, 2003). It has previously been reported that an extreme excess of rare variants is uniquely associated with positive selection (Charlesworth & Charlesworth, 2000;

Bachtrog, 2004) . Furthermore, it may be possible to distinguish between these two models by examining rates of protein evolution (Gerrard & Filatov, 2005). The selective sweep model and the background selection models make opposing predictions on the fixation of advantageous alleles (Charlesworth, 1994). It should be noted that in a region of extremely low recombination, such as the ZAL2^{m}, the current haplotypes only contain information dating back to the most recent selective sweep.

The reduced diversity we observed in the ZAL2/ZAL2^{m} system is one of many parallels between it and sex chromosomes. Prior to the research presented in this dissertation, our lab detailed characteristics of the ZAL2/ZAL2^{m} system that were analogous to sex chromosomes, including suppression of recombination by inversions and disassortative mating (Thomas *et al.*, 2008). Our data in Chapter 2 suggest that the ZAL2^{m} differs from the W chromosome with respect to recombination, as we detected recombination within our sample of ZAL2^{m} chromosomes. This is not unexpected considering that among the birds sampled in several independent studies, two ZAL2^{m} homozygotes (one male and one female) have been observed and they comprise 0.10% (2/1318) of the population (Falls and Kopachena 1994; D. Maney personal communication; Romanov *et al.* 2009; Thorneycroft 1975). The extreme rarity of ZAL2^{m} homozygotes is consistent with estimates of ZAL2^{m} homozygote frequency at ~0.4% based on reports of WS × WS mating pairs (Lowther, 1961; Lowther & Falls, 1968; Tuttle, 1993), and suggests that it is not necessary to invoke a recessive lethal mutation associated with the ZAL2^{m} to explain the rarity of ZAL2^{m} homozygotes, as previously suggested (Thorneycroft, 1975; Tuttle, 1993).

Another feature that differs between the ZAL2/ZAL2^{m} and the sex chromosomes is the lack of genetic degeneration observed on the ZAL2^{m} chromosome. This lack of degeneration is especially surprising considering that our dating estimates, based on the divergence between the ZAL2 and ZAL2^{m}, suggests that recombination between the chromosome types ceased at least ~2.5 mya. Based on data from studies of neo-sex chromosomes, it seems unlikely that this chromosome has experienced suppressed recombination between its homolog for such an extended time without any sign of degeneration. Additionally, the estimated age of the *Zonotrichia* species complex is ~1.5 my and the inversion is not observed in any other *Zonotrichia* species. Based on these data, we hypothesize that the chromosomal polymorphism is a genetic legacy of the past and originated by a hybridization event. This would account for the divergence observed between the ZAL2/ZAL2^{m}, the timing of recombination suppression as well as the lack of evidence for genetic degeneration. Furthermore, hybridization is fairly common among birds, even across genera and white-throated sparrow is known to hybridize with the dark-eyed junco (Kellogg, 1959; Short & Simon, 1965), the white-crowned sparrow (Kellogg, 1959) and the golden-crowned sparrow (Payne, 1979).

The role of hybridization in evolution is a controversial issue among zoologists because species hybrids are presumed to be less fit under the biological species concept. If the ZAL2^{m} introgressed into the white-throated sparrow genome, it would add the short list of examples where introgressive hybridization has potentially facilitated evolutionary adaptation (Besansky *et al.*, 2003; Rieseberg *et al.*, 2003). Because a chromosome

similar to the ZAL2^{m} has not been observed in limited sampling of the *Zonotrichia* and other related species, the originating species under the hypothetical introgression scenario remains unknown. In the future, more intense sampling of related species could shed light on the origin of the inversion.

The origin of the white-throated sparrow remains a mystery and there are several possible explanations to be explored. One option is that the inversion was introduced into a population of individuals with only ZAL2-like chromosomes. Because we found no evidence that the inversions occurred sequentially, these proto-ZAL2 and proto-ZAL2^{m} chromosomes would have differed from each other by both inversions. Subsequently, the ZAL2^{m} could only increase in frequency if it conferred higher fitness to the carrier. For example, if it was a highly preferred mate or if it was a better competitor than those birds in the original population. As the ancestral population would have consisted of only ZAL2/ZAL2 individuals, the mating advantage of the new ZAL2/ZAL2^{m} could easily increase in frequency, thus generating a pattern of disassortative mating. If the behavioral traits subsequently evolved to be associated with the ZAL2^{m} polymorphism, they could reinforce the pattern of disassortative mating which would maintain the inversion at an intermediate frequency in the population.

An alternative hybridization scenario consistent with our data is homoploid hybrid speciation, in which the white-throated sparrow species originated by the hybridization between two ancestral species with the same chromosome number (Coyne & Orr, 2004). This mechanism of speciation is thought to be rare and appears to be more prominent in

plant species than in animals (Rieseberg *et al.*, 2003), though a recent study identified a case in *Rhagoletis* flies (Schwarz *et al.*, 2005). An important difference in this scenario, with reference to the aforementioned hybridization scenario, is that the ZAL2^{m} haplotype would have a single origin and this is consistent with the single dominating haplotype that we observed in our studies. However, under homoploid hybrid speciation there may have been multiple ZAL2^{m} haplotypes in the originating species but any selective sweep on the ZAL2^{m} could result in a single dominating haplotype in the present-day population. The primary difficulties with the homoploid hybrid speciation are that hybrids need to be fit, reproductive isolation is necessary to prevent backcrossing with parental taxa and speciation must occur in sympatry or parapatry, thus, the hybrids may need to be specially adapted to an ecological niche that is distinct from that of the parental taxa. If the white-throated sparrow species originated by homoploid hybrid speciation then there must have been a strong fitness advantage associated with the disassortative mate pairs.

Falls and Kopachena (1994) suggested that selection acts against monomorphic pairs as TS × TS pairs cannot successfully acquire/defend territory and WS × WS pairs do not sufficiently provision to their offspring. The long-term stability of this system depends on the equal fitness of the TS and WS genotypes, but equally fit reproductive strategies are exceedingly rare in nature. Furthermore, there are no examples of alternative mating strategies that occur in equal frequency and generally, there is a dominating strategy (Lank *et al.*, 1995; Sinervo & Lively, 1996; Shuster & Sassaman, 1997). The white-throated sparrow appears to be the only example of an alternative reproductive strategy

where both strategies are equally fit and roughly equally represented in the population. In regards to fitness, equal fitness is evidenced by reports showing that, in total, both types of disassortative mating pairs spend the equal time on territorial defense and offspring provisioning resulting in identical hatchling survival rates (Knapton & Falls, 1983; Knapton *et al.*, 1984; Kopachena & Falls, 1993b). The equal frequency of the morphs is necessitated by the extreme pattern of disassortative mating. Currently, it is unclear what mechanism underlies the disassortative mating and this remains one of the most intriguing aspects of the system.

Concluding remarks

The research presented in this dissertation demonstrates the utility of this species for understanding the evolution and maintenance of inversions, as well as elucidating the relationship between recombination, linkage disequilibrium and genetic diversity. This species also has the potential to shed light on the genetic basis of behavior through investigations of candidate genes and there are still many unresolved questions in the white-throated sparrow to be addressed in future research. What are the relative influences of positive and negative selection within the $ZAL2^m$, the ZAL2, as well as the region outside of the inversion? How important is recombination in the $ZAL2^m$ homozygotes? Does it rescue the $ZAL2^m$ from the genetic degeneration that characterizes regions of extremely low or no recombination? How variable is recombination on a local level across the genome? How does recombination shape diversity on a local level? What are the genes involved in and what is the genetic architecture of the behavioral differences? Finally, the answer to the most compelling

questions in this system remains elusive. What is the origin of this extraordinary inversion? In the future, these questions will likely lead to answers that will illuminate our basic understanding of the biological process of evolution.

Although the white-throated sparrow is a common bird in North America, it is far from ordinary. The data from this dissertation confirms exceptional and widely contrasting patterns in evolution and molecular evolution within the genome. Considering the behavioral and genetic characteristics of the white-throated sparrow, this species is a compelling model for future research in the fields of molecular evolution, evolution of inversions, sex chromosome evolution and the genetics of social behavior.

6 References

Abrahams BS, Kwok MCH, Trinh S, Budaghzadeh S, Hossain SM & Simpson EM (2005) Pathological aggression in "Fierce" mice corrected by human nuclear receptor 2E1. *J Neurosci* **25**, 6263-6270.

Aitken RJ & Graves JAM (2002) The future of sex. *Nature* **415**, 963.

Allison AC (1956) The sickle-cell and Haemoglobin C genes in some African populations. *Ann Human Genet* **21**, 67-89.

Andersson M & Iwasa Y (1996) Sexual selection. *Trends Ecol Evol* **11**, 53-58.

Andolfatto P, DePaulis F & Navarro A (2001) Inversion polymorphisms and nucleotide variability in Drosophila. *Genet. Res. Camb.* **77**, 1-8.

Arnold AP (1998) Sexual differentiation of the zebra finch song system: positive evidence, negative evidence, null hypothesis and paradigm shift. *J Neurobiol* **33**, 572-584.

Auton A & McVean G (2007) Recombination rate estimatation in the presence of hotspots. *Genome Res* **17**, 1219-1227.

Ayala FJ & Campbell CA (1974) Frequency-dependent selection. *Ann Rev Ecol Syst* **5**, 115-138.

Bachtrog D (2004) Evidence that positive selection drives Y-chromosome degeneration in *Drosophila miranda*. *Nature Genetics* **36**, 518-522.

Bachtrog D (2006) A dynamic view of sex chromosome evolution. *Curr Opin Genet & Devel* **15**, 578-585.

Bachtrog D & Charlesworth B (2002) Reduced adaptation of a non-recombining neo-Y chromosome. *Nature* **416**, 323-326.

Backström N, Fagerberg S & Ellegren H (2008) Genomics of natural bird populations: a gene-based set of reference markers evenly spread across the avian genome. *Mol Ecol* **17**, 964-980.

Backström N, Forstmeier W, Schielzeth H, Mellenius H, Nam K, Bolund E, Webster MT, Öst T, Schneider M, Kempenaers B & Ellegren H (2010) The recombination landscape of the zebra finch Taeniopygia guttata genome. *Genome Res* **20**, 485-495.

Backström N, Qvarnström A, Gustafsson L & Ellegren H (2006) Levels of linkage disequilibrium in a wild bird population. *Biol Lett* **2**, 435-438.

Balakrishnan CN & Edwards SV (2009) Nucleotide variation, linkage disequilibrium and founder-facilitated speciation in wild populations of the Zebra Finch (*Taeniopygia guttata*). *Genetics* **181**, 645-660.

Bamshad M, Mummidi S, Gonzalez E, Ahuja SS, Dunn DM, Watkins WS, Wooding SP, Stone AC, Jorde LB, Weiss RB & Ahuja SK (2002) A strong signature of balancing selection in the 5' *cis*-regulatory region of CCR5. *Proc Natl Acad Sci USA* **99**, 10539-10544.

Bamshad M & Wooding SP (2003) Signatures of natural selection in the human genome. *Nature Rev Genet* **4**, 99-111.

Barrett JC, Fry B, Maller J & Daly MJ (2005) Haploview: analysis and visualization of LD and haplotype maps. *Bioinformatics* **21**, 263-265.

Baudat F, Buard J, Grey C, Fledel-Alon A, Ober C, Przeworski M, Coop G & de Massy B (2010) PRDM9 is a major determinant of meiotic recombination hotspots in humans and mice. *Science* **327**, 836-840.

Begun DJ & Aquadro CF (1992) Levels of naturally occurring DNA polymorphism correlate with recombination rates in *D. melanogaster*. *Nature* **356**, 519-520.

Bekoff M (2004) *Encyclopedia of animal behavior*. Westport, CT: Greenwood Press.

Benatti TR, Valicente FH, Aggarwal R, Zhao C, Walling JG, Chen M-S, Cambron SE, Schemerhorn BJ & Stuart JJ (2010) A neo-sex chromosome that drive postzygotic sex determination in the Hessian fly (*Mayetiola destructor*). *Genetics* **184**, 769-777.

Benkman CW (1996) Are the ratios of bill crossing morphs in crossbills the result of frequency-dependent selection? *Evol Ecol* **10**, 119-126.

Bergero R, Charlesworth D, Filatov DA & Moore RC (2008) Defining regions and rearrangements of the *Silene latifolia* Y chromosome. *Genetics* **178**, 2045-2053.

Berlin S & Ellegren H (2004) Chicken W: A genetically uniform chromosome in a highly variable genome. *Proc. Natl. Acad. Sci. USA* **101**, 15967-15969.

Berlin S, Tomaras D & Charlesworth B (2007) Low mitochondrial variability in birds may indicate Hill-Robertson effects on the W chromosome. *Heredity* **99**, 389-396.

Besansky NJ, Krzywinski J, Lehmann T, Simard F, Kern M, Mukabayire O, Fontenille D, Touré Y & Sagnon NF (2003) Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA sequence variation. *Proc Natl Acad Sci USA* **100**, 10818-10823.

Betancourt A & Presgraves DC (2002) Linkage limits the power of natural selection in *Drosophila*. *Proc Natl Acad Sci USA* **99**, 13616-13620.

Betancourt AJ, Welch JJ & Charlesworth B (2009) Reduced effectiveness of selection caused by a lack of recombination. *Curr. Biol.* **19**, 655-660.

Betrán E, Rozas J, Navarro A & Barbadilla A (1997) The estimation of the number and length distribution of gene conversion tracts from population DNA sequence data. *Genetics* **146**, 89-99.

Borge T, Webster MT, Andersson G & Saetre G-P (2005) Contrasting patterns of polymorphism and divergence on the Z chromosome and autosomes in two Ficedula flycatcher species. *Genetics* **171**, 1861-1873.

Brooks SA, Lear TL, Adelson DL & Bailey E (2007) A chromosome inversion near the KIT gene and the Tobiano spotting pattern in horses. *Cytogenet Genome Res* **119**, 225-230.

Brown JL (1997) A theory of mate choice based on heterozygosity. *Behav Ecol* **8**, 60-65.

Butlin RK, Read IL & Day TH (1982) The effects of a chromosomal inversion on adult size and male mating success in the seaweed fly, *Coelopa frigida*. *Heredity* **49**, 51-62.

Caballero A (1995) On the effective size of populations with separate sexes, with particular reference to sex-linked genes. *Genetics* **139**, 1007-1011.

Chalhoub N, Benachenhou N, Rajapurohitam V, Pata M, Ferron M, Frattini A, Villa A & Vacher J (2003) Grey-lethal mutation induces severe malignant autosomal recessive osteopetrosis in mouse and human *Nature Med* **9**, 399-406.

Charlesworth B (1994) The effect of background selection against deleterious mutations on weakly selected, linked variants *Genet Res* **63**, 213-227.

Charlesworth B (1996) Background selection and patterns of genetic diversity in *Drosophila melanogaster*. *Genet. Res.* **68**, 131-149.

Charlesworth B (2009) Effective population size and patterns of molecular evolution and variation. *Nat Genet Rev* **10**, 195-205.

Charlesworth B & Charlesworth D (2000) The degeneration of Y chromosomes. *Phil. Trans. R. Soc. B* **355**, 1563-1572.

Charlesworth D (2006) Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet* **2**, e64.

Charlesworth D & Awadalla P (1998) Flowering plant self-incompatibility: the molecular population genetics of *Brassica* S-loci. *Heredity* **81**, 1-9.

Charlesworth D, Bartolomé C, Schierup MH & Mable BK (2003) Haplotype structure of the stigmatic self-incompatibility gene in natural populations of *Arabidopsis lyrata*. *Mol Biol Evol* **20**, 1741-1753.

Charlesworth D, Charlesworth B & Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. *Heredity* **95**, 118-128.

Charlesworth D, Charlesworth B & Morgan MT (1995) The pattern of neutral molecular variation under the background selection model. *Genetics* **141**, 1619-1632.

Charlesworth D & Guttman DS (1997) Plant genetics: seeing selection in S allele sequences. *Curr Biol* **7**, R34-R37.

Chovnick A (1973) Gene conversion and transfer of genetic information within the inverted region of inversion heterozygotes. *Genetics* **75**, 123-131.

Clark MR, Schweikert G, Toomajian C, Ossowski S, Zeller G, Shinn P, Warthmann N, Hu TT, Fu G, Hinds DA, Chen H, Frazer KA, Huson DH, Schölkopf B, Nordborg M, Rätsch G, Ecker JR & Weigel D (2007) Common sequence polymorphisms shaping genetic diversity in *Arabidopsis thaliana*. *Science* **317**, 338-342.

Cohuet A, Dia I, Simard F, Raymond M & Fontenille D (2004) Population structure of the malaria vector *Anopeles funestus* in Senegal based on microsatellite and cytogenetic data. *Insect Mol Biol* **13**, 251-258.

Collins CE & Houtman AM (1999) Tan and white color morphs of the white-throated sparrows differ in their song vocal responses to territorial intrusion. *Condor* **101**, 842-845.

Coyne JA, Aulard S & Berry A (1991) Lack of underdominance in a naturally occurring pericentric inversion in *Drosophila melanogaster* and its implications for chromosome evolution. *Genetics* **129**, 791-802.

Coyne JA, Meyers W, Crittenden AP & Sniegowski P (1993) The fertility effects of pericentric inversions in *Drosophila melanogaster*. *Genetics* **134**, 487-496.

Coyne JA & Orr HA (2004) *Speciation*. Sunderland, MA, USA: Sinauer.

Crocker G & Day T (1987) An advantage to mate choice in the seaweed fly, *Coelopa frigida*. *Behav Ecol Sociobiol* **20**, 295-301.

Cutter AD & Payseur BA (2003) Selection at linked sites in the partial selfer *Caenorhabditis elegans*. *Mol Biol Evol* **20**, 665-673.

Darwin C (1871) *The descent of man, and selection in relation to sex*. Longon: John Murra.

Day TH & Butlin RK (1987) Non-random mating in natural populations of seaweed fly, *Coelopa frigida*. *Heredity* **58**, 213-220.

de Oliveira EH, Habermann FA, Lacerda O, Sbalqueiro IJ, Wienberg J & Müller S (2005) Chromosome reshuffling in birds of prey: the karyotype of the world's

largest eagle (Harpy eagle, Harpia harpyja) compared to that of the chicken (Gallus gallus). *Chromosoma* **114**, 338-343.

Dickerman RW (1961) Hybrids among the fringillid genera *Junco-Zonotrichia* and *Melospiza*. *Auk* **78**, 627-632.

Dietl MM, Hof PR, Martin JL, Magistretti PJ & Palacios JM (1990) Autoradiographic analysis of the distribution of vasoactive intestinal peptide binding sites in the vertebrate central nervous system: a phylogenetic study. *Brain Res* **520**, 14-26.

Dobzhansky T (1936) Position effects on genes. *Biol Rev* **11**, 364-384.

Dobzhansky T (1937) *Genetics and the origin of species*. New York: Columbia University Press.

Dobzhansky T (1950) Genetics of natural populations. XIX. Origin of heterosis through natural selection in populations of *Drosophila pseudoobscura*. *Genetics* **35**, 288-302.

Dobzhansky T (1970) *Genetics of the evolutionary process*. New York, New York: Columbia University Press.

Dobzhansky T & Pavlovsky O (1960) How stable is balanced polymorphism? *Proc Natl Acad Sci USA* **46**, 41-47.

Dobzhansky T & Sturtevant AH (1938) Inversions in the chromosomes of *Drosophila pseudoobscura*. *Genetics* **23**.

Dolgin ES & Charlesworth B (2008) The effects of recombination rate on the distribution and abundance of transposable elements. *Genetics* **178**, 2169-2177.

Duret L & Arndt PF (2008) The impact of recombination on nucleotide substitutions in the human genome. *PLoS Genetics* **4**, e1000071.

Dyer K, Charlesworth B & Jaenike J (2007) Chromosome-wide linkage disequilibrium as a consequence of meiotic drive. *Proc. Natl. Acad. Sci. USA* **104**, 1587-1592.

Edwards SV & Dillon M (2004) Hitchhiking and recombination in birds: evidence from Mhc-linked and unlinked loci in red-winged blackbirds (*Agelaius phoeniceus*). *Genet Res* **84**, 175-192.

Ellegren H (2005) The avian genome uncovered. *TREE* **20**, 180-186.

Ellegren H (2007) Molecular evolutionary genomics of birds. *Cytogenet Genome Res* **117**, 120-130.

Ellegren H (2009a) The different levels of genetic diversity in sex chromosome and autosomes. *Trends Genet* **25**, 278-284.

Ellegren H (2009b) Genomic evidence for a large-Z effect. *Proc Biol Sci* **276**, 361-366.

Emlen ST & Oring LW (1977) Ecology, sexual selection and the evolution of mating systems. *Science* **197**, 215-223.

Erhart MA, Lekgothoane S, Grenier J & Nadeau JH (2002) Pattern of segmental recombination in the distal inversion of the mouse t-haplotypes. *Mamm. Genome* **13**, 438-444.

Ewing B & Green P (1998) Basecalling of automated sequencer traces using phred. II. Error probabilities. *Genome Res* **8**, 186-194.

Ewing B, Hillier LW, Wendl M & Green P (1998) Basecalling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* **1986**, 175-185.

Excoffier L, Laval G & Schneider S (2005) Arlequin ver. 3.0: An integrated software package for population genetic data analysis *Evolutionary Bioinformatics Online* **1**, 47-50.

Falls JB & Kopachena JG (1994) White-throated sparrow (*Zonotrichia albicollis*). In *The Birds of North America* [A Poole and F Gill, editors]. Philadelphia: The Academy of Natural Sciences.

Fang L, Ye J, Li N, Zhang Y, Li S, Wong GK-S & Wang J (2008) Positive correlation between recombination rate and nucleotide diversity is shown under domestication selection in the chicken genome. *Chinese Science Bull* **53**, 746-750.

Farner DS & Wingfield JC (1980) Reproductive endocrinology of birds. *Ann Rev Physiol* **42**, 457-472.

Feder HH (1984) Hormones and sexual behavior. *Ann Rev Psych* **35**, 165-200.

Feder JL, Roethele JB, Filchak KE, Niedbalski J & Romero-Severson J (2003) Evidence for inversions related to sympatric host race formation in the apple maggot fly, *Rhagoletis pomonella* (Diptera: Tephritidae). *Genetics* **163**.

Ferrer-Admetlla A, Bosch E, Sikora M, Marquès-Bonet T, Ramírez-Soriano A, Muntasell A, Navarro A, Lazarus R, Calafell F, Bertranpetit J & Casals F (2008) Balancing selection is the main force shaping the evolution of innate immunity genes. *J Immunol* **181**, 1315-1322.

Fiddes JC & Goodman HM (1979) Isolation, cloning and sequence analysis of the cDNA for the α-subunit of human chorionic gonadotropin. *Nature* **281**, 351-356.

Filatov DA & Gerrard DT (2003) High mutation rates in human and ape pseudoautosomal genes. *Gene* **317**, 67-77.

Filatov DA, Moneger F, Negrutiu I & Charlesworth D (2000) Low variability in a Y-linked plant gene and its implications for Y-chromosome evolution. *Nature* **404**, 388-390.

Fridolfsson A-K & Ellegren H (2000) Molecular evolution of the avian *CHD1* genes on the Z and W sex chromosomes. *Genetics* **155**, 1903-1912.

Fry JD (2004) How common are overdominant mutations? *Genetics* **167**, 1031-1032.

Gerrard DT & Filatov DA (2005) Positive and negative selection on mammalian Y chromosomes. *Mol Biol Evol* **22**, 1423-1432.

Goddard MR, Godfray HCJ & Burt A (2005) Sex increases the efficacy of natural selection in experimental yeast populations. *Nature* **434**, 636-640.

Gordo I & Charlesworth B (2001) Genetic linkage and molecular evolution. *Curr. Biol.* **11**, R684-R686.

Gordon D, Abajian C & Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Res* **8**, 195-202.

Grant PR & Grant BR (1992) Hybridization of bird species. *Science* **256**.

Graves JAM (2006) Sex chromosome specialization and degeneration in mammals. *Cell* **124** 901-914.

Greene E, Lyon BE, Muehter VR, Ratcliffe L, Oliver SJ & Boag PT (2000) Disruptive sexual selection for plumage coloration in a passerine bird. *Science* **407**, 1000-1003.

Haase B, Jude R, Brooks SA & Leeb T (2008) An equine chromosome 3 inversion is associated with the tobiano spotting pattern in German horse breeds. *Anim Genet* **39**, 306-309.

Haddrill PR, Halligan DL, Tomaras D & Charlesworth B (2007) Reduced efficacy of selection in regions of the *Drosophila* genome that lack crossing over. *Genome Biol.* **8**, R18.

Hammer MF, Bliss S & Silver LM (1991) Genetic exchange across a paracentric inversion of the mouse t complex. *Genetics* **128**, 799-812.

Hammer MF, Mendez FL, Cox MP, Woerner AE & Wall JD (2008) Sex-biased evolutionary forces shape genomic patterns of human diversity. *PLoS Genet* **2008**, e1000202.

Hammer MF, Schimenti J & Silver LM (1989) Evolution of mouse chromosome 17 and the origin of inversions associated with *t* haplotypes. *Proc Natl Acad Sci USA* **86**, 3261-3265.

Hammer MF & Silver LM (1993) Phylogenetic analysis of the alpha-globin pseudogene-4 (*Hba9s4*) locus in the house mouse species complex reveals a stepwise evolution of *t* haplotypes. *Mol Biol Evol* **10**, 971-1001.

Hansson B, Ljungqvist M, Dawson DA, Mueller JC, Olano-Marin J, Ellegren H & Nilsson J-Å (2010) Avian genome evolution insights from a linkage map of the blue tit (*Cyanistes caeruleus*). *Heredity* **104**, 67-78.

Hartl DL (1974) Genetic dissection of segregation distortion. I. Suicide combinations of *SD* genes. *Genetics* **76**, 477-486.

Hartl DL (1975) Genetic dissection of segregation distortion II. Mechanism of suppression of distortion by certain inversions. *Genetics* **80**, 539-547.

Hasson E & Eanes WF (1996) Contrasting histories of three gene regions associated with *In(3L)Payne* of *Drosophila melanogaster*. *Genetics* **144**, 1565-1575.

Hellborg L & Ellegren H (2004) Low levels of nucleotide diversity in mammalian Y chromosomes. *Mol. Biol. Evol.* **21**, 158-163.

Hellmann I, Ebersberger I, Ptak SE, Pääbo S & Przeworski M (2003) A neutral explanation for the correlation of diversity with recombination rates in humans. *Am J Hum Genet* **72**, 1527-1535.

Hermann BG, Koschorz B, Wertz K, McLaughlin KJ & Kispert A (1999) A protein kinase encoded by the *t complex responder* gene causes non-mendelian inheritance. *Nature* **402**, 141-146.

Hill WG & Robertson A (1966) The effect of linkage on limits to artificial selection. *Genet. Res.* **8**, 269-294.

Hillier LW, Miller W, Birney E, Warren W, Hardison RC, Ponting CP, Bork P, Burt DW, Groeen MAM, Delany ME, Dodgson JB, Chinwalla AT, Cliften PF, Clifton SW, Delehaunty KD, Fronick C, Fulton RS, Graves TA, Kremitzki C, Layman D, Magrini V, McPherson JD, Miner TL, Minx P, Nash WE, Nhan MN, Nelson JO, Oddy LG, Pohl CS, Randall-Maher J, Smith SM, Wallis JW, Yan S-P, Romanov MN, Rondelli CM, Paton B, Smith J, Morrice D, Daniels L, Tempest HG, Roberston L, Masabanda JS, Griffin DK, Kierzek AM, McLaren SR, Overton IM, Arakawa H, Beattie KJ, Bezzubov Y, Boardman PE, Bonfield JK, Croning MDR, Davies RM, Francis MD, Humphray SJ, Scott CE, Taylor RG, Tickle C, Brown WRA, Rogers J, Buerstedde J-M, Wilson SA, Stubbs L, Ovcharenko I, Gordon L, Lucas S, Miller MM, Inoko H, Shiina T, Kaufman J, Salomonsen J, Skjoedt K, Wong GK-S, Wang J, Liu B, Wang J, yu J, Yang H, Nefedov M, Koriabine M, deJong PJ, Goodstadt L, Webber C, Dickens NJ, Letunic I, Suyama M, Torrents D, Mering Cv, Zdobnov EM, Makova K, Nekrutenko A, Elnitski L, Eswara P, King DC, Yan S, Tyekucheva S, Radakrishnan A, Harris RS, Chiaromonte R,

Taylor J, He J, Rijnkels M, Griffiths-Jones R, Ureta-Vidal A, Hoffman MM, Severin J, Searle SMJ, Law AS, Speed D, Waddington D, Cheng Z, Tuzun E, Eichler E, Bao Z, Flicek P, Shteynberg DD, Brent MR, Bye JM, Huckle EJ, Chatterji S, Dewey C, Pachter L, Kouranov A, Mourelatos S, Hatzigeorgiou AG, Paterson AH, Ivarie R, Brandstrom M, Axelsson E, Backström N, Berlin S, Webster MT, Pourquie O, Reymond A, Ucla C, Antonarakis SE, Long M, Emerson JJ, Bétran E, Dupanloup I, Kaessmann H, Hinrichs AS, Bejerano G, Furey TS, HArte RA, Raney B, Siepel A, Kent WJ, Haussler D, Eyras E, Castelo R, Abril JF, Castellano S, Camara F, Parra G, Guigo R, Bourque G, Tesler G, Pevzner PA, Smit A, Fulton LA, Mardis ER & Willson RK (2004) Sequence and comparative analysis of the chicken genome provide unique perspectives on vertebrate evolution. *Nature* **432**, 695-716.

Hoffmann AA, Sgró CM & Weeks AR (2004) Chromosomal inversion polymoprhisms and adaptation. *TREE* **19**, 482-488.

Holloway CC & Clayton DF (2001) Estrogen synthesis in the male brain triggers development of the avian song control pathway *in vitro*. *Nature Neurosci* **4**, 170-175.

Hori M (1993) Frequency-dependent natural selection in the handedness of scale-eating cichlid fish. *Science* **260**, 216-219.

Hough RB, Lengeling A, Bedian V, Lo C & Bucan M (1998) *Rump white* inversion in the mouse disrupts dipeptidyl aminopeptidase-like protein 6 and causes dysregulation of *Kit* expression *Proc Natl Acad Sci USA* **95**, 13800-13805.

Houtman AM & Falls JB (1994) Negative assortative mating in the white-throated sparrow, *Zonotrichia albicollis*: the role of mate-choice and intersexual competition. *Anim Behav* **48**, 377-383.

Huang SW, Ardlie KG & Yu HT (2001) Frequeny and distribution of t-haplotypes in Southeast Asian house mouse (*Mus musculus castaneus*) in Taiwan. *Mol Ecol* **10**, 2349-2354.

Hudson RR & Kaplan NL (1985) Statistical properties of the number of recombination events in the history of a sample of DNA sequences. *Genetics* **111**, 147-164.

Hughes AL (2007) Looking for Darwin in all the wrong places: the misguided quest for positive selection at the nucleotide sequence level. *Heredity* **99**, 364-373.

Hughes AL & Nei M (1988) Pattern of nucleotide substitution at major histocompatibility complex class I loci reveals overdominant selection. *Nature* **335**, 167-170.

Hughes AL & Yeager M (1998) Natural selection at major histocompatibility complex loci of vertebrates. *Ann Rev Genet* **32**, 415-435.

Huson DH & Bryant D (2006) Application of phylogenetic networks in evolutionary studies. *Mol. Biol. Evol.* **23**, 254-267.

Huynh LY, Maney DL & Thomas JW (2010) Chromosome-wide linkage disequilibrium caused by an inversion polymorphism in the white-throated sparrow (Zonotrichia albicollis). In *Submitted*.

Innan H & Stephan W (2003) Distinguishing the hitchhiking and background selection models. *Genetics* **165**, 2307-2312.

Ironside JE & Filatov DA (2005) Extreme population structure and high interspecific divergence of the Silene Y chromosome. *Genetics* **171**, 705-713.

Iwase M, Satta Y, Hirai Y, Hirai H, Imai H & Takahata N (2003) The amelogenin loci span an ancient pseudoautosomal boundary in diverse mammalian species. *Proc Natl Acad Sci USA* **100**, 5258-5263.

Jacobs EC, Arnold AP & Campagnoni AT (1996) Zebra finch estrogen receptor cDNA: cloning and mRNA expression. *J Steroid Biochem Mol Biol* **59**, 135-145.

Jensen MA, Charlesworth B & Kreitman M (2002) Patterns of genetic variation at a chromosome 4 locus of *Drosophila melanogaster*. *Genetics* **160**, 493-507.

Kaiser VB & Charlesworth B (2010) Muller's Ratchet and the degeneration of the *Drosophila miranda* neo-Y chromosome. *Genetics* **[Epub ahead of print]**.

Kamau E, Charlesworth B & Charlesworth D (2007) Linkage disequilibrium and recombination rate estimates in the self-incompatibility region of *Arabidopsis lyrata*. *Genetics* **176**, 2357-2369.

Kamau E & Charlesworth D (2005) Balancing selection and low recombination affect diversity near the self-incompatibility loci of the plant *Arabidopsis lyrata*. *Curr Biol* **15**, 1773-1778.

Keinan A, Mullikin JC, Patterson N & Reich D (2009) Accelerated genetic drift on chromosome X during the human dispersal out of Africa. *Nat Genet* **41**, 66-77.

Kellogg PP (1959) A hybrid white-crowned x white-throated sparrow. *Wilson Bull* **71**, 282-283.

Kelly JK (1997) A test of neutrality based on interlocus associations. *Genetics* **146**, 1197-1206.

Kennington JW, Partridge L & Hoffmann AA (2006) Patterns of diversity and linkage disequilibrium within the cosmopolitan inversion *In(3R)Payne* in *Drosophila melanogaster* are indicative of coadaptation. *Genetics* **172**, 1655-1663.

Kennington WJ, Hoffmann AA & Partridge L (2007) Mapping regions within cosmopolitan inversion *In)3R)Payne* associated with natural variation in body size in *Drosophila melanogaster*. *Genetics* **177**, 549-556.

Kimura M (1983) *The neutral theory of molecular evolution*. Cambridge, UK: Cambridge University Press.

Kirkpatrick M & Barton N (2006) Chromosome inversions, local adaptation and speciation. *Genetics* **173**, 419-434.

Kliman RM & Hey J (1993) Reduced natural selection associated with low recombination in *Drosophila melanogaster*. *Mol Biol Evol* **10**, 1239-1258.

Knapton RW, Carter V & Falls JB (1984) A comparison of breeding ecology and reproductive success between morphs of the white-throated sparrow. *Wilson Bull* **96**, 60-71.

Knapton RW & Falls JB (1983) Differences in parental contribution among pair types in the polymorphic white-throated sparrow. *Can J Zool* **61**, 1288-1292.

Kohno S, Katsu Y, Iguchi T & Guillette LJ (2008) Novel approaches for the study of vertebrate steroid hormone receptors. *Integ Compar Biol* **48**, 527-534.

Kondo M, Hornung U, Nanda I, Imai S, Sasaki T, Shimizu A, Asakawa S, Hori H, Schmid M, Shimizu N & Schartl M (2006) Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Res* **16**, 815-826.

Kopachena JG & Falls JB (1993a) Aggressive performance as a behavioral correlate of plumage polymorphism in the white-throated sparrow. *Behavior* **124**, 249-266.

Kopachena JG & Falls JB (1993b) Postfledgling parental care in the white-throated sparrow (*Zonotrichia albicollis*). *Can J Zool* **71**, 227-232.

Kopachena JG & Falls JB (1993c) Re-evaluation of morph-specific variations in parental behavior of the white-throated sparrow. *Wilson Bull* **105**, 48-59.

Kopp A, Duncand I & Carroll SB (2000) Genetic control and evolution of sexually dimorphic characters in *Drosophila*. *Nature* **408**, 553-559.

Kovacevic M & Schaeffer SW (2000) Molecular population genetics of *X*-linked genes in *Drosophila pseudoobscura*. *Genetics* **156**, 155-172.

Kraft T, Säll T, Magnusson-Rading I, Nilsson N-O & Halldén C (1998) Positive correlation between recombination rates and levels of genetic variation in natural populations of sea beet (*Beta vulgaris* subsp. *maritima*). *Genetics* **150**, 1239-1244.

Lahn BT & Page DC (1999) Four evolutionary strata on the human X chromosome. *Science* **286**, 964-967.

Lande R (1984) The expected fixation rate of chromosomal inversions. *Evolution* **38**, 743-752.

Lande R (1985) The fixation of chromosomal rearrangements in a subdivided population with local extinction and colonization. *Heredity* **54**, 323-332.

Lande R & Wilkinson GS (1999) Models of sex-ratio meiotic drive and sexual selection in stalk-eyed flies. *Genetical Res* **74**, 245-253.

Lank DB, Coupe M & Wynne-Edwards KK (1999) Testosterone-induced male traits in female ruffs (*Philomachus pugnax*): autosomal inheritance and gender differentiation. *Proc Biol Sci* **266**, 2323-2330.

Lank DB, Smith CM, Hanotte O, Burke T & Cooke F (1995) Genetic polymorphism for alternative mating behavior in a lekking male ruff *Philomachus pugnax*. *Nature* **378**, 59-62.

Lawson Handley L-J, Ceplitis H & Ellegren H (2004) Evolutionary strata on the chicken Z chromosome: implications for sex chromosome evolution. *Genetics* **167**, 367-376.

Lee J, Han K, Meyer TJ, Kim H-S & Batzer MA (2008) Chromosomal inversions between human and chimpanzee lineages caused by retrotransposons. *PLoS ONE* **2**, e4047.

Levene H & Dobzhansky T (1958) New evidence of heterosis in naturally occurring inversion heterozygotes in *Drosophila pseudoobscura*. *Heredity* **12**, 37-49.

Librado P & Rozas J (2009) DnaSP v5: A software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**, 1451-1452.

Lindgren G, Backström N, Swinburne J, Hellborg L, Einarsson A, Sandberg K, Cothran G, Villà C, Binns M & Ellegren H (2004) Limited number of patrilines in horse domestication. *Nature Genet* **36**, 335-336.

Liu Z, Moore PH, Ma H, Ackerman CM, Ragiba M, Yu Q, Pearl HM, Kim MS, Charlton JW, Stiles JI, Zee FT, Paterson AH & Ming R (2004) A primitive Y chromosome in papaya marks incipient sex chromosome evolution. *Nature* **427**, 348-352.

Lockton S & Gaut BS (2010) The evolution of transposable elements in natural populations of self-fertilizing *Arabidopsis thaliana* and its outcrossing relative *Arabidopsis lyrata*. *BMC Evol Biol* **10**, 10.

Lowther JK (1961) Polymorphism in the white-throated sparrow, *Zonotrichia albicollis* Gmelin. *Can. J. Zool.* **39**, 281-292.

Lowther JK & Falls JB (1968) *Zonotrichia albicollis* (Gmelin) White-throated sparrow. *US Natl Mus Bull* **237**, 1364-1392.

Lyon MF (1984) Transmission ratio distortion in mouse t-haplotypes is due to multiple distorter genes acting on a responder locus. *Cell* **37**, 621-628.

Lyon MF (2003) Transmission ratio distortion in mice. *Annu. Rev. Genet.* **37**, 393-408.

Machado CA, Haselkorn TS & Noor MAF (2007) Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *D. persimilis*. *Genetics* **175**, 1289-1306.

Mallet J (2005) Hybridization as an invasion of the genome. *TREE* **20**, 229-237.

Maney DL (2008) Endocrine and genomic architecture of life history trade-offs in an avian model of social behavior. *Gen Comp Endocrin* **157**, 275-282.

Maney DL, Erwin KL & Goode CT (2005) Neuroendocrine correlates of behavioral polymorphism in the white-throated sparrow. *Horm Behav* **48**, 196-206.

Mank JE, Axelsson E & Ellegren H (2007) Fast-X on the Z: Rapid evolution of sex-linked genes in birds. *Genome Res* **17**, 618-624.

Mank JE & Ellegren H (2007) Parallel divergence and degradation of the avian W sex chromosome. *Trends Ecol Evol* **22**, 389-391.

Manoukis NC, Powell JR, Touré Y, Sacko A, Edillo FE, Coulibaly MB, Traoré SF, Taylor CE & Besansky NJ (2008) A test of the chromosomal theory of ecotypic speciation in *Anopheles gambiae*. *Proc Natl Acad Sci USA* **105**, 2940-2945.

Marais G (2003) Biased gene conversion: implications for genome and sex evolution. *Trends Genet* **6**, 330-338.

Marais GAB, Nicolas M, Bergero R, Chambrier P, Kejnovsky E, Monéger F, Hobza R, Widmer A & Charlesworth D (2007) Evidence for degeneration of the Y chromosome in the dioecious plant *Silene latifolia*. *Curr Biol* **18**, 545-549.

Maynard Smith J & Haigh J (1974) The hitch-hiking effect of a favourable gene. *Genet Res* **23**, 23-35.

McAllister BF (2003) Sequence differentiation associated with an inversion on the neo-X chromosome in *Drosophila americana*. *Genetics* **165**, 1317-1328.

McVean GAT & Charlesworth B (2000) The effects of Hill-Robertson interference between weakly selected mutations on patterns of molecular evolution and variation. *Genetics* **155**, 929-944.

McVean GAT, Myers SR, Hunt S, Deloukas P, Bentley D & Donnelly P (2004) The fine-scale structure of recombination rate variation in the human genome. *Science* **304**, 581-584.

Megens H-J, Croojimans RPMA, Bastiaansen JWM, Kerstens HHD, Coster A, Jalving R, Vereijken A, Sila P, Muir WM, Cheng HH, Hanotte O & Groenen MAM (2009) Comparison of linkage disequilibrium and haplotype diversity on macro- and microchromsomes in chicken. *BMC Genetics* **10**, 86.

Mendel G (1866) Versuche über Plflanzenhybriden (Translation by W. Bateson). Verhandlungen des naturforschenden Vereines in Brünn, Bd. IV für das Jahr 1865. *Abhandlungen*, 3-47.

Michopoulos V, Maney DL, Morehouse CB & Thomas JW (2007) A genotyping assay to determine plumage morph in the white-throated sparrow, *Zonotrichia albicollis*. *Auk* **124**, 1330-1335.

Montell H, Fridolfsson A-K & Ellegren H (2001) Contrasting levels of nucleotide diversity on the avian Z and W sex chromosomes. *Mol. Biol. Evol.* **18**, 2010-2016.

Mroczek RJ, Melo JR, Luce AC, Hiatt EN & Dawe RK (2006) The maize Ab10 meiotic drive system maps to supernumerary sequences in a large complex haplotype. *Genetics* **174**, 145-154.

Muller HJ (1916) The mechanism of crossing-over. II. IV. The manner of occurrence of crossing-over. *Am Nat* **50**, 284-305.

Muller HJ (1930) Types of visible variations induced by X-rays in *Drosophila*. *J Genet* **3**, 299-334.

Mundy NI, Badcock NS, Hart T, Scribner K, Janssen K & Nadeau NJ (2004) Conserved genetic basis of a quantitative plumage trait involved in mate choice. *Science* **303**, 1870-1873.

Munté A, Rozas J, Aguadé M & Segarra C (2005) Chromosomal inversion polymorphism leads to extensive genetic structure: A multilocus survey in *Drosophila pseudoobscura*. *Genetics* **39**, 1573-1581.

Myers S, Bowden R, Tumian A, Bontrop RE, Freeman C, MacFie TS, McVean G & Donnelly P (2010) Drive against hotspot motifs in primates implicates the *PRDM9* gene in meiotic recombination. *Science* **327**, 876-879.

Nachman MW (2001) Single nucleotide polymorphisms and recombination rate in humans. *Trends Genet* **17**, 481-485.

Nachman MW (2002) Variation in recombination rate across the genome: evidence and implications. *Curr Opin Genet & Devel* **12**, 657-663.

Nachman MW & Myers P (1989) Exceptional chromosomal mutations in a rodent population are not strongly underdominant. *Proc Natl Acad Sci USA* **86**, 6666-6670.

Nagle DL, Kozak CA, Mano H, Chapman VM & Bucan M (1995) Physical mapping of the *Tec* and *Gabrb1* loci reveals that the Wsh mutation on mouse chromosome 5 is associated with an inversion. *Hum Mol Genet* **4**, 2073-2079.

Nagylaki T (1995) The inbreeding effective population number in dioecious populations. *Genetics* **148**, 1245-1255.

Navarro A, Betrán E, Barbadilla A & Ruiz A (1997) Recombination and gene flux caused by gene conversion and crossing over in inversion heterozygotes. *Genetics* **146**, 695-709.

Nelson RJ (2000) *An introduction to behavioral endocrinology*. Sunderland, MA: Sinauer.

Nóbrega C, Khadem M, Aguadé M & Segarra C (2008) Genetic exchange versus genetic differentiation in a medium-sized inversion of *Drosophila*: the A2/Ast arrangements of *Drosophila subobscura*. *Mol Biol Evol* **25**, 1534-1543.

Noor MAF (2008) Mutagenesis from meiotic recombination is not a primary driver of sequence divergence between *Saccharomyces* species. *Mol Biol Evol* **25**, 2439-2444.

Noor MAF, Garfield DA, Schaeffer SW & Machado CA (2007) Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. *Genetics* **177**, 1417-1428.

Noor MAF, Grams KL, Bertucci LA & Reiland J (2001) Chromosomal inversions and the reproductive isolation of species. *Proc. Natl. Acad. Sci. USA* **98**, 12084-12088.

Novitski E & Braver G (1954) An analysis of crossing over within a heterozygous inversion in *Drosophila melanogaster*. *Genetics* **39**, 197-209.

Ogawa S, Eng V, Taylor DB, Lubahn DB, Korach KS & Pfaff SW (1998) Roles of estrogen receptor-alpha gene expression in reproduction-related behaviors in female mice. *Endocrin* **139**, 5070-5081.

Ogawa S, Lubahn DB, Korach KS & Pfaff SW (1996) Aggressive behaviors of transgenic estrogen-receptor knockout male mice. *Ann NY Acad Sci* **794**, 384-385.

Ogawa S, Taylor JA, Lubahn DB, Korach KS & Pfaff DW (1997) Behavioral effects of estrogen receptor gene disruption in male mice. *Proc Natl Acad Sci USA* **94**, 1476-1481.

Olivier B & van Oorschot R (2005) 5-HT1B receptors and aggression: a review. *Europ J Pharm* **526**, 207-217.

Ortíz-Barrientos D, Reiland J, Hey J & Noor MAF (2002) Recombination and the divergence of hybridizing species. *Genetica* **116**, 167-178.

Panithanarak T, Hauffe HC, Dallas JF, Glover A, Ward RG & Searle JB (2004) Linkage-dependent gene flow in a house mouse chromosomal hybrid zone. *Evolution* **58**, 37-47.

Parvanov ED, Petkov MP & Paigen K (2010) Prdm9 controls activation of mammalian recombination hotspots. *Science* **327**, 835.

Payne RB (1979) Two apparent hybrid Zonotrichia sparrows. *The Auk* **96**, 595-599.

Peichel CL, Ross JA, Matson CK, Dickson M, Grimwood J, Schmutz J, Myers RM, Mori S, Schluter D & Kingsley DM (2004) The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr. Biol.* **14**, 1416-1424.

Piper WH & Wiley RH (1989) Distinguishing morphs of the white-throated sparrow in basic plumage. *J Field Ornithol* **60**, 73-83.

Popova N & Amstislavskaya TG (2002) Involvement of the 5-HT(1A) and 5-HT(1B) serotonergic receptor subtypes in sexual arousal in male mice. *Psychoneuroendocrin* **27**, 609-618.

Price TD & Bouvier MM (2002) The evolution of F1 postzygotic incompatibilities in birds. *Evolution* **56**, 2083-2089.

Primmer CR, Borge T, Lindell J & Saetre G-P (2002) Single-nucleotide polymorphism characterization in species with limited available sequence information: high nucleotide diversity revealed in the avian genome. *Mol Ecol* **11**, 603-612.

Pritchard J & Przeworski M (2001) Linkage disequilibrium in humans: models and data. *Am J Hum Genet* **69**, 1-14.

Pryke SR & Griffith SC (2009) Socially mediated trade-offs between aggression and parental effort in competing color morphs. *Am Nat* **174**, 455-464.

Raymond CK, Kas A, Paddock M, Qiu R, Zhou Y, Subramanian S, Chang J, Palmieri A, Haugen E, Kaul R & Olson MV (2005) Ancient haplotypes of the HLA class II region. *Genome Res* **15**, 1250-1257.

Rice WR (1994) Degeneration of a nonrecombining chromsome. *Science* **263**, 230-232.

Richman AD & Kohn JR (1999) Self-incompatibility alleles from *Physalis*: Implications for historical inference from balanced genetic polymorphisms. *Proc Natl Acad Sci USA* **96**, 168-172.

Rieseberg LH (2001) Chromosomal rearrangements and speciation. *Trends Ecol Evol* **16**, 351-358.

Rieseberg LH, Raymond O, Rosenthal DM, Lai Z, Livingstone K, Nakazato T, Durphy JL, Schwarzbach AE, Donovan LA & Lexer C (2003) Major ecological transitions in wild sunflowers facilitated by hybridization *Science* **301**, 1211-1216.

Rieseberg LH, Whitton J & Gardner K (1999) Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. *Genetics* **152**, 713-727.

Robinson GE, Fernald RD & Clayton DF (2008) Genes and social behavior. *Science* **322**, 896-900.

Rocca KAC, Gray EM, Costantini C & Besansky NJ (2009) 2La chromosomal inversion enhances thermal tolerance of *Anopheles gambiae* larvae. *Malaria Journal* **8**, 147.

Rodionov AV (1996) Micro vs. macro: Structural-funtional organization of avian micro- and macrochromosomes. *Genetika* **32**, 597-608.

Romanov MN, Tuttle EM, Houck ML, Modi WS, Chemnick LG, Korody ML, Mork EMS, Otten CA, Renner T, Jones KC, Dandekar S, Papp JC, Da Y, Program

NCS, Green ED, Magrini V, Hickenbotham MT, Glasscock J, McGrath S, Mardis E & Ryder OA (2009) The value of avian genomics to the conservation of wildlife. *BMC Genomics* **10**, S10.

Roselius K, Stephan W & Städler T (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* **171**, 753-763.

Ross JA & Peichel CL (2008) Molecular cytogenetic evidence of rearrangements on the Y chromosome of the threespine stickleback fish. *Genetics* **179**, 2173-2182.

Roulin A (2004) The evolution, maintenance and adaptive function of genetic colour polymorphism in birds. *Biol Rev* **79**, 815-848.

Rozas J, Sánchez-DelBarrio JC, Messeguer X & Rozas R (2003) DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**, 2496-2497.

Rozen S & Skaletsky H (2000) Primer3 on the WWW for general users and for biologist programmers. *Methods Mol Biol* **132**, 365-386.

Saudou F, Amara DA, Dierich A, LeMeur M, Ramboz S, Segu L, Buhot MC & Hen R (1994) Enhanced aggressive behavior in mice lacking 5-HT1B receptor. *Science* **23**, 1875-1878.

Schaeffer SW & Anderson WW (2005) Mechanisms of genetic exchange within the chromosomal inversions of *Drosophila pseudoobscura*. *Genetics* **171**, 1729-1739.

Schaeffer SW, Goetting-Minesky MP, Kovacevic M, Peoples JR, Graybill JL, Miller JM, Kim K, Nelson JG & Anderson WW (2003) Evolutionary genomics of inversions in *Drosophila pseudoobscura*: Evidence for epistasis. *Proc. Natl. Acad. Sci. USA* **100**, 8319-8324.

Schierup MH, Mikkelsen AM & Hein J (2001) Recombination, balancing selection and phylogenies in MHC and self-incompatibility genetics. *Genetics* **159**, 1833-1844.

Schlinger BA (2001) Neurosteroids and brain sexual differentiation. *Trends Neurosci* **24**, 429-431.

Schmid M, Nanda I, Guttenbach M, Steinlein C, Hoehn M, Schartl M, Haaf T, Weigend S, Fries R & Buerstedde J-M (2000) First report on chicken genes and chromosomes 2000. *Cytogenet Genome Res* **90**, 169-218.

Schwarz D, Matta BM, Shakir-Botteri NL & McPheron BA (2005) Host shift to an invasive plant trigges rapid animals speciation. *Nature* **436**, 546-549.

Sherwood NM, Krueckl SL & McRory JE (2000) The origin and function of pituitary adenylate cyclase-activating polypeptide PACAP/glucagons superfamily. *Endocrin Rev* **21**, 619-670.

Shi Y, Chichung LD, Taupin P, Nakashima K, Ray J, Yu RT, Gage FH & Evans RM (2004) Expression and function of orphan nuclear receptor TLX in adult neural stem cells. *Nature* **427**, 78-83.

Shimron-Abarbanell D, Nöthen MM, Erdmann J & Propping P (1995) Lack of genetically determined structural variants of the human serotonin-1E (5-HT1E) receptor protein points to its evolutionary conservation. *Mol Brain Res* **29**, 387-390.

Short LL & Simon SW (1965) Additional hybrids of the slate-colored junco and the white-throated sparrow. *Condor* **67**.

Shuster SM & Sassaman C (1997) Genetic interaction between male strategy and sex ratio in a marine isopod. *Nature* **388**, 373-377.

Siepel A, Bejerano G, Pedersen JS, Hinrichs AS, Hou M, Rosenbloom K, Clawson H, Cpieth J, Hillier LW, Richards S, Weinstock GM, Wilson RK, Gibbs RA, Kent WJ & Haussler D (2005) Evolutionarily conserved elements in vertebrate, insect, worm and yeast genomes. *Genome Res* **15**, 1034-1050.

Silver LM (1985) Mouse *t* haplotypes. *Ann Rev Genet* **19**, 179-208.

Silver LM (1993) The peculiar journey of a selfish chromosome: mouse *t* haplotypes and meiotic drive. *Trends Genet* **9**, 250-254.

Silver LM & Artzt K (1981) Recombination suppression of mouse *t*-haplotypes due to chromatin mismatching. *Nature* **290**, 68-70.

Silver LM & Remis D (1987) Five of the nine genetically defined regions of mouse t haplotypes are involved in transmission ratio distortion. *Genet Res* **49**, 51-56.

Sinervo B & Lively CM (1996) The rock–paper–scissors game and the evolution of alternative male strategies. *Nature* **380**, 240-243.

Sites JW & Moritz C (1987) Chromosomal evolution and speciation revisited. *Syst Zool* **36**, 153-174.

Slawson EE, Shaffer CD, Malone CD, Leung W, Kellmann E, Shevchek RB, Craig CA, Bloom SM, Bogenphol J, Dee J, Morimoto ETA, Myoung J, Nett AS, Ozsolak F, Tittiger ME, Zeug A, Pardue M-L, Buhler J, Mardis ER & Elgin SCR (2006) Comparison of dot chromosome sequences from *D. melanogaster* and *D. virilis* reveals an enrichment of DNA transposon sequences in heterochromatic regions. *Genome Biol.* **7**, R15.

Smith J, Bruley CK, Paton IR, Dunn I, Jones CT, Windsor D, Morrice DR, Law AS, Masabanda J, Sazanov A, Waddington D, Fries R & Burt DW (2000) Differences

in gene density on chicken macrochromosomes and microchromosomes. *Anim Genet* **31**, 96-103.

Smithson A & Macnair MR (2008) Density-dependent and frequency-dependent selection by bumblebees *Bombus terrestris* (L.) (Hymenoptera: Apidae). *Biol J Linn Soc* **60**, 401-417.

Soma KK, Schlinger BA, Wingfield JC & Saldanha CJ (2003) Brain aromatase, 5-alpha reductase and 5-beta reductase change seasonally in wild male song sparrows (*Melospiza melodia morphna*). *J Neuroendocrin* **15**, 150-160.

Sperry TS, Moore IT, Meddle SL, Benowitz-Fredericks ZM & Wingfield JC (2005) Increased sensitivity of the serotonergic system during the breeding season in free-living American tree sparrows. *Behav Brain Res* **157**, 119-126.

Sperry TS, Thompson CK & Wingfield JC (2003) Effects of acute treatment with 8-OH-DPAT and fluoxetine on aggressive behaviour in Male song sparrows (*Melospiza melodia morphna*). *J Neuroendocrin* **15**, 150-160.

Stapley J, Birkhead TR, Burke T & Slate J (2008) A linkage map of the Zebra Finch *Taeniopygia guttata* provides new insights into avian genome evolution. *Genetics* **179**, 651-667.

Stapley J, Birkhead TR, Burke T & Slate J (2010) Pronounced inter- and intrachromosomal variation in linkage disequilibrium across the zebra finch genome. *Genome Res* **20**, 496-502.

Stefansson H, Helgason A, Thorlelfsson G, Steinthorsdottir V, Masson G, Barnard J, Baker A, Jonasdottir A, Ingason A, Gudnadottir VG & al. e (2005) A common inversion under selection in Europeans. *Nat. Genet.* **37**, 129-137.

Stephan W & Langley CH (1998) DNA polymorphism in lycopersicon and crossing-over per physical length. *Genetics* **150**, 1585-1593.

Stephens M, Smith NJ & Donnelly P (2001) A new statistical method for haplotype reconstruction from population data. *Am. J. Hum. Genet.* **68**, 978-989.

Stone W & Thomas I (1934) Crossover and disjunctional properties of X chromosome inversions in *Drosophila melanogaster*. *Genetica* **17**, 170-184.

Stump AD, Fitzpatrick MC, Lobo NF, Traoré SF, Sagon N, Costantini C, Collins FH & Besansky NJ (2005) Centromere-proximal differentiation and speciation in *Anopheles gambiae*. *Proc. Natl. Acad. Sci. USA* **102**, 15930-15935.

Stump AD, Pombi M, Goeddel L, Ribeiro JMC, Wilder JA, della Torre A & Besansky NJ (2007) Genetic exchange in 2La inversion heterokaryotypes of *Anopheles gambiae*. *Insect Mol Biol* **16**, 703-709.

Sturtevant AH (1917) Genetic factors affecting the strength of linkage in Drosophila. *Proc Natl Acad Sci USA* **3**, 555-558.

Sturtevant AH (1921) A case of rearrangement of genes in Drosophila. *Proc. Natl. Acad. Sci. USA* **7**, 235-237.

Sturtevant AH (1965) *A history of genetics*. Cold Spring Harbor, NY: Cold Spring Harbor Laboratory Press.

Sturtevant AH & Beadle GW (1936) The relations of inversions in the X chromosome of *Drosophila melanogaster* to crossing over and disjunction. *Genetics* **21**, 554-604.

Sundström H, Webster MT & Ellegren H (2004) Reduced variation on the chicken Z chromosome. *Genetics* **167**, 377-385.

Tajima F (1993) Simple methods for testing the molecular evolutionary clock hypothesis. *Genetics* **135**, 599-607.

Takayama S & Isogai A (2005) Self-incompatibility in plants. *Ann Rev Plant Biol* **56**, 467-489.

Tenaillon MI, U'Ren J, Tenaillon O & Gaut BS (2004) Selection versus demography: a multilocus investigation of the domstication process in maize. *Mol Biol Evol* **21**, 1214-1225.

Terry A, Bucciarelli G & Bernardi G (2000) Restricted gene flow and incipient speciation in disjunct Pacific Ocean and Sea of Cortez populations of a reef fish species, *Girella nigricans*. *Evolution* **54**, 652-659.

Thigpen AE, Davis DL, Milatovich A, Mendonca BB, Imperato-McGinley J, Griffin JE, Francke U, Wilson JD & Russell DW (1992) Molecular genetics of steroid 5α-reductase 2 deficiency. *J Clin Invest* **90**, 799-809.

Thomas JW, Cáceres M, Lowman JL, Morehouse CB, Short ME, Baldwin EL, Maney DL & Martin CL (2008) The chromosomal polymorphism linked to variation in social behavior in the white-throated sparrow (*Zonotrichia albicollis*) is a complex rearrangement and suppressor of recombination. *Genetics* **179**, 1455-1468.

Thorneycroft HB (1966) Chromosomal polymorphism in the white-throated sparrow, *Zonotrichia albicollis* (Gmelin). *Science* **154**, 1571-1572.

Thorneycroft HB (1975) A cytogenetic study of the white-throated sparrow, Zonotrichia alibcollis Gmelin. *Evolution* **29**, 611-621.

Tregenza T & Wedell N (2000) Genetic compatibility, mate choice and patterns of parentage: invited review. *Mol Ecol* **9**, 1013-1027.

Tuttle E (1993) Mate choice and stable polymorphism in the white-throated sparrow, State University of New York at Albany.

Tuttle EM (2003) Alternative reproductive strategies in the white-throated sparrow: behavioral and genetic evidence. *Behav Ecol* **14**, 425-432.

van Delden W & Kamping A (1989) The association between the polymorphisms at the *Adh* and *αGpdh* loci in the *In(2L)t* inversion in *Drosophila melanogaster* in relation to temperature. *Evolution* **43**.

van Rhijn JG (1973) Behavioural dimorphism in male ruffs, *Philomachus pugnax*. *Behaviour* **47**, 153-229.

Votintseva AA & Filatov DA (2009) Evolutionary strata in a small mating-type-specific region of the smut fungus *Microbotryum violaceum*. *Genetics* **182**, 1391-1396.

Wallace LT & Erhart MA (2008) Recombination within mouse t haplotypes has replaced significant segments of t-specific DNA. *Mamm. Genome* **19**, 263-271.

Wang RL, Stec A, Hey J, Lukens L & Doebley J (1999) The limits of selection during maize domestication. *Nature* **398**, 236-239.

Wang Z, Baker AJ, Hill GE & Edwards SV (2003) Reconciling actual and inferred population histories in the house finch (*Carpodacus mexicanus*) by AFLP analysis. *Evolution* **57**, 2852-2864.

Warren WC, Clayton DF, Ellegren H, Arnold AP, Hillier LW, Kunstner A, Searle S, White S, Vilella AJ, Fairley S, Heger A, Kong L, Ponting CP, Jarvis ED, Mello CV, Minx P, Lovell P, Velho TA, Ferris M, Balakrishnan CN, Sinha S, Blatti C,

London SE, Li Y, Lin YC, George J, Sweedler J, Southey B, Gunaratne P, Watson M, Nam K, Backstrom N, Smeds L, Nabholz B, Itoh Y, Whitney O, Pfenning AR, Howard J, Volker M, Skinner BM, Griffin DK, Ye L, McLaren WM, Flicek P, Quesada V, Velasco G, Lopez-Otin C, Puente XS, Olender T, Lancet D, Smit AF, Hubley R, Konkel MK, Walker JA, Batzer MA, Gu W, Pollock DD, Chen L, Cheng Z, Eichler EE, Stapley J, Slate J, Ekblom R, Birkhead T, Burke T, Burt D, Scharff C, Adam I, Richard H, Sultan M, Soldatov A, Lehrach H, Edwards SV, Yang SP, Li X, Graves T, Fulton L, Nelson J, Chinwalla A, Hou S, Mardis ER & Wilson RK (2010) The genome of a songbird. *Nature* **464**, 757-762.

Watt DJ (1986) Plumage brightness index for white-throated sparrows. *J Field Ornithol* **57**, 105-115.

West-Eberhard MJ (1983) Sexual selection, social competition and speciation. *Quart Rev Biol* **58**, 155-183.

White BJ, Hahn MW, Pombi M, Cassone BJ, Lobo NF, Simard F & Besansky NJ (2007) Localization of candidate regions maintaining a common polymorphic inversion (2La) in *Anopheles gambiae*. *PloS Genet.* **3**, e217 doi:210.1371/journal.pgen.0030217.

White MJD (1978) Chain processes in chromosomal speciation. *Syst Zool* **27**, 285-298.

Williams JR & Lenington S (1993) Factors modulating preferences of female house mice for males differing in t-complex genotype: Role of t-complex genotype, genetic background, and estrous condition of females. *Behav Genet* **23**, 51-58.

Wilson JD, Griffin JE & Russell DW (1993) Steroid 5-alpha reductase 2 deficiency. *Endocrin Rev* **14**, 577-593.

Wingfield JC (1994) Hormone-behavior interactions and mating systems in male and female birds. In *The differences between the sexes* [RV Short and E Balaban, editors]. Cambridge: Cambridge University Press.

Wong GKS, Liu B, Wang J, Zhang Y, Yang X, Zhang ZJ, Meng QS, Zhou J, Li DW, Zhang JJ, Ni PX, Li SG, Ran LH, Li H, Zhang JG, Li RQ, Li ST, Zheng HK, Lin W, Li GY, Wang XL, Zhao WM, Li J, Ye C, Dai MT, Ruan J, Zhou Y, Li YZ, He XM, Zhang YZ, Wang J, Huang XG, Tong W, Chen J, Ye J, Chen C, Wei N, Li GQ, Dong L, Lan FD, Sun YQ, Zhang ZP, Yang Z, Yu YP, Huang YQ, He DD, Xi Y, Wei D, Qi QH, Li WJ, Shi JP, Wang MH, Xie F, Wang JJ, Zhang XW, Wang P, Zhao YQ, Li N, Yang N, Dong W, Hu SN, Zeng CQ, Zheng WM, Hao BL, Hillier LW, Yang SP, Warren WC, Wilson RK, Brandstrom M, Ellegren H, Crooijmans RPMA, van der Poel JJ, Bovenhuis H, Groenen MAM, Ovcharenko I, Gordon L, Stubbs L, Lucas S, Glavina T, Aerts A, Kaiser P, Rothwell L, Young JR, Rogers S, Walker BA, van Hateren A, Kaufman J, Bumstead N, Lamont SJ, Zhou HJ, Hocking PM, Morrice D, de Koning DJ, Law A, Bartley N, Burt DW, Hunt H, Cheng HH, Gunnarsson U, Wahlberg P, Andersson L, Kindlund E, Tammi MT, Andersson B, Webber C, Ponting CP, Overton IM, Boardman PE, Tang HZ, Hubbard SJ, Wilson SA, Yu J, Wang J & Yang HM (2004) A genetic variation map for chicken with 2.8 million single-nucleotide polymorphisms. *Nature* **432**, 717-722.

Young KA, Berry ML, Maharrey CL, Saionz JR, Hawes NL, Chang B, Zheng QY, Smith RS, Bronson RT, Nelson RJ & Simpson EM (2002) Fierce: a new mouse deletion of Nr2e1: violent behavior and ocular abnormalities are background-dependent. *Behav Brain Res* **132**, 145-158.

Zhang J, Wheeler DA, Yakub I, Wei S, Sood R, Rowe W, Liu PP, Gibbs RA & Buetow KH (2005) SNPdetector: A software tool for sensitive and accurate SNP detection. *PloS Comput Biol* **1**, e53.

Zhang Z, Schwartz S, Wagner L & Miller W (2000) A greedy algorithm for aligning DNA sequences. *J. Comput. Biol.* **7**, 203-214.

Zhou M, Lei M, Rao Y, Nie Q, Zeng H, Xia M, Liang F, Zhang D & Zhang X (2008a) Polymorphisms of vasoactive intestinal peptide receptor-1 gene and their genetic effects on broodiness in chickens. *Poultry Sci* **87**, 893-903.

Zhou Q, Wang J, Huang L, Nie W, Wang J, Liu Y, Zhao X, Yang F & Wang W (2008b) Neo-sex chromosomes in the black muntjac recapitulate incipient evolution of mammalian sex chromosomes. *Genome Biol.* **9**, R98.

7 Appendix

Supplementary Table 1. PCR primers used in Chapter 2.

Locus	*Orthologous position on Zebra finch chr3*[a]	*Forward primer*	*Reverse primer*	T_m
26	28,582,485-28,582,967	ACATGAGGCACAGAGAGTAACG	CAATGAATACCTCTTCCAAGCC	58
36 (C20orf74)[b]	29,415,389-29,416,561	CCTGGGATAGAAGGTAGGAAGAG	AAATCTTCATAGGCCTGGCTTC	58
87	21,384,038-21,384,842	AGGCTGTTGATTACCCACTGTC	TGAGGTGTTGACACTATGGAGC	55
92	1,278,462-1,278,872	AGGAAGCCATCAGAAGATGAAG	GCTGATCTGGATAATGGAAAGG	55
99	2,121,187-2,121,728	AGTGACAGAAATACTGTGTTGTGG	TGTATGTGCTGTTACTTGCCG	55
104	2,750,169-2,750,716	TCCTAGGATTTCAGCCATGC	CATGCCAACTCATACCAAACTG	55
144	24,482,179-24,483,053	CAGCAGTTGCTGGTAAATGAAG	AAGCTCAATAATCTCCACGCTC	55
164	23,831,637-23,832,540	AGCCCAGATCAAGCCATATTAG	GCATAAAGCATGAATGCCAAG	55
182	18,757,043-18,757,561	ACAGGAACAGCACAGAAGACAG	CAAGTGAAGCCTACTTGGTTGG	55
186	18,290,962-18,291,385	CTGTGAGAGTGTTCCTCCCATC	GTTAACTAGTTCCACCACAACTCC	55
204	10,256,808-10,257,452	AGGTTCATGAATCCATTTCCAG	GTGCTCTAGATGCTGCCTATTG	55
212	11,083,581-11,084,078	GAGCTCTCTAGTGTGCCATGC	CACTGCAATTATTGACGAGGAC	55
217	11,589,424-11,589,841	GCAGTGCAGGTTCAAGAGTTC	ACTCCATAATGCCTTTGTTCTG	55
249	14,965,374-14,965,784	AAACAAATGGACTCCCAAAGG	CCAGCAATTGAAAGCTGTAAGG	55
264	16,553,731-16,554,127	AGTATTCGAAGGGCGTTACAAG	AGCATTCAAGAATCTGAACCG	55
268	16,974,376-16,974,905	ACCATTCTCTGGGTTCTGAGTC	GCTTGTCTGTGTGGAAGTGAAG	55
283	7,901,130-7,901,708	ACAACTTGGAATGGATTTGAGG	ACAAATGTTGCATGTGGCTTAC	55
333	35,725,276-35,725,905	ACCGAAGATATGTTTCAGCCTC	CATATCTTGCAGCTTTGACCTG	55
335	35,539,025-35,539,305	AAGCATACTGGGTATCTGTGTCTC	TAAGCGTTGCTCTTGCTACTTG	55
349	34,001,976-34,002,398	TTTGAATTACGCTCAGTCAAGC	CTGCAAAGAGTGCAAATTTAGG	55
359	32,853,166-32,853,636	AAACCAATCTCAGGGCTACAAC	TCCTTGGTGATAATGCTGACTC	55
361	32,634,444-32,634,899	TCACTGCTGAAGCTGCAATATC	TTCCTTGAACTGAAAGTTTGGC	55
374	46,848,557-46,849,035	GGAAATCGCTCACATCAAAGAG	GACAACAGACAAAGACACAGGC	60
400	44,505,212-44,505,563	CTGCATCACCAAGCAGACTTC	ACACAATGTAGTGGGTGTGCTC	55
418	42,766,484-42,767,291	GGGTGTGACAGAATGACTGATG	AGAAACCCTCTTCCAGCCTTC	55
427	41,897,166-41,897,721	TGAGTGCATCATAGCAACTGAG	AATCCTTGAACACAGACCCTTG	55
439	40,599,532-40,600,019	ACACATATTCACACTTGACCCG	ATTTGACAAAGCCCTTTACTGG	55
459	38,502,808-38,503,430	AAGCCCAAGAAAGACAACACTC	CGACACTCGAACATTTGAAGAG	55
471 (MAP3K4)[b]	37,235,355-37,235,709	TGCTAAGCTTCCAGTCTCTGTG	GATTTGCAAACTCCTTCTCCTG	58
499	47,329,827-47,330,396	GTGATTAAGCACGATCATGTCC	TGAGGTTGCTTCAGTGTTCAAG	55
510 (ESR1)[b]	56,300,786-56,303,685	AGCACCTTGAAATCTCTGGAAG	ACTGCAAGGAATGAGATGAAGC	58
513 (VIP)[b]	55,933,393-55,934,140	ACTGACAACTACAGCCGCTTTC	CAAGATGTTCTGCTTCATCTCG	55
530	54,193,083-54,193,283	AAATTGTTTCAGGTGGGACTTG	GTCTGGCAATGAGGTTCTCAG	55
547	chr3_random	CCAGCTTCTGTTTCTGCTCTTC	GCTCACCTAGAGGAACAGCTTC	55
560 (REPS1)[b]	51,058,127-51,059,760	CAGACTTCAGCCAGTTTGAGG	CATCTGCTTTGGCAACTCTAAG	58
571	49,945,846-49,946,394	TTACACGGCTTATGGTGGATG	AGATCTGCCTTTCTCTGAGGC	55
603	60,420,861-60,421,570	CTGGATATTCTACCCGTCCAAC	GTGCCAAGTTTGTGAGTCTGTC	55
621 (TRMT11)[b]	61,971,176-61,972,312	TCCTTCAGAAGAGATGGCAAG	GTTCTTCAAAGATGCATAGAGC	58
689	68,537,309-68,537,760	GGCTTCATGAAATTCTTACGTG	AGCTTGCTAGGGAGTACAGTGG	55
703	69,993,462-69,993,874	AGAGATTGTCTCACTGCCTTCC	TACTGGAGTAGCAACGCTTCAG	55
715	71,181,934-71,182,378	TCACAATATCCATGACTGACAGG	TTGCTACCATTGTCATTGAACC	55
748	74,467,617-74,468,010	CAATGAGGAACAGAGTATAGCCAC	CCTGACCTATTTCTTTCAGTGC	55
750	74,686,689-74,687,223	CCATCAAAGCAAATGGCTAAC	ATTAACAGTGGAAGCAGTGGAG	55
793 (CGA)[b]	78,911,832-78,912,869	CCAGATGGAGAGTTTCTCATGC	GTCTTCTTGGACCTCATTGGAG	55
822	81,821,859-81,822,532	GCTGCAAAGGAAGGTACTACAC	GCACTGTTATGAGAAGGGAAATG	55
833	82,924,696-82,925,373	TAATGTGGGCAATCTGCTATTC	ATCTGTTGCATGTTAGCTCCAC	55
842	83,860,159-83,860,713	AAACCCATGTGTGTTGAGAGC	TCCACAAACATTCCTTTCCTTC	55
885	88,117,728-88,118,266	ACAGCCTGTGGGATCTGAATG	CCAGGCATTTATTCCTTTCAC	55
905 (FAM83B)[b]	90,016,994-90,017,384	TTCCTTGGCTTCAGTTTCTAGC	GCAAGATTTACCTTTGGTCGTG	58
910	90,547,701-90,548,286	ACGAGCACCTAATAAAGGAAGG	CCTCAAACAAAGTCAATGGGTC	55
938	93,483,873-93,484,595	CATGTGGTAGGTAAGAATCATGC	GGACTTCAGAGTTGTGTTGCTG	55
965	96,229,720-96,230,388	GACCTTGACAGCTGATTGACTG	GACCTCCACAGCTGAACAGAG	55
974	97,110,021-97,110,556	ATCCGATTCTTAACAAAGCTGC	AGTTTCATACCTTCCATCTGGG	55
985	98,299,172-98,299,682	ATAGGACCCATTCAATTTGCAC	CCCTGGAGCTTCTTTACTGG	55
991	98,873,626-98,873,986	ATGACATGCACTTTAGTCTGCC	TCCTTGATTTGTGTTTGTCCTG	55
1007	100,621,194-100,621,859	GATACTGGCTTCCATATCTGCC	GAAACATCCTCTTGAGAGTGGG	55
1021	102,010,865-102,011,402	AGGACAGTCATTGCCACTTTG	AAAGTGCTATAAGACGCAGCAC	55
1051	105,186,307-105,187,070	TCAAACTTGTTGAACCATCCAG	TGGCAGATTCAATCAGTACAGG	55
1068	106,901,144-106,901,925	AAGTACAAGCCCAAATTCAAGC	GTAACAGCAGCACTTGTGGAAG	55
1097	111,407,149-111,407,592	GCCAAATTTCACCTTCATCTTC	TTTCACCTTGATGAGAAGCCAC	55
1108	110,167,617-110,167,697	ATTCTGAAAGCGAGATGAAACC	GACTAATCTTCAGCAAGCCAGG	55
1119 (SUPT3H)[b]	109,107,200-109,107,939	CAGATATGGCATGACTGAATGG	CTGACAGCAAAGAAAGCATCC	58

[a]Coordinates are based on the taeGut1 assembly of the zebra finch genome.

[b]Loci previously described in Thomas et al., 2008, Genetics **179:** 1455-1468. Locus names from Thomas *et al.*, 2008 are shown in parentheses.

Supplementary Table 2. Population genetic statistics of individual loci in Chapter 2.

Locus Name	1097	1108	1119 (SUPT3H)	715	Total (outside inversion)	Total (inversion)	144	26	36 (C20orf74)[b]	164	87	182
Locus location												
Estimated Z. albicollis position in Mb ZAL2 (ZAL2^m)	1.21	2.45	3.51	5.21	NA	NA	5.84 (33.49)	9.94 (37.59)	10.77 (38.43)	11.18 (38.83)	13.62 (41.28)	16.25 (43.90)
Orthologous chicken position (Mb)	109.72	110.83	111.94	71.55	NA	NA	14.43	2.62	3.69	16.46	8.70	18.26
Orhologous zebra finch position (Mb)	111.41	110.17	109.11	71.18	NA	NA	24.48	28.58	29.42	23.83	21.38	18.76
DNA polymorphism, divergence and gene flow												
Number of sites	444	79	685	446	1654	4492	875	628	754	897	809	529
Noncoding & synonymous sites	444	79	606.5	446	1575.5	4225.625	875	461.459	754	867.333	809	458.833
Segregating sites, S	3	1	12	1	17[a]	41	4	12	11	4	2	8
Fixed polymorphisms	0	0	0	0	0[a]	21	2	10	5	2	0	2
Shared shared polymorphisms	1	1	3	0	2[a]	0	0	0	0	0	0	0
Polymorphic in ZAL2, monomorphic in ZAL2^m	0	0	5	1	13[a]	13	2	2	5	2	1	1
Polymorphic in ZAL2^m, monomorphic in ZAL2	2	0	4	0	2[a]	7	0	0	1	0	1	5
F_{ST} (2 vs 2^m)	0.18207	0.19524	-0.02241	0	0.21232	0.93641	0.94118	0.98765	0.88498	0.87719	0	0.70517***
F_{ST} (WS vs TS)	-0.03619	-0.10476	0.12149	0	0.06738	0.40688	0.39298	0.4444	0.34792	0.31895	0	0.30933
d_{xy} (2 vs 2^m)	0.00179	0.00633	0.0031	0.00014	0.0019	0.00583	0.00243	0.01612	0.00754	0.00265	0.00023	0.00555
Number of haplotypes	5	2	11	2	17	24	4	4	6	4	3	7
Haplotype diversity, Hd	0.57971	0.3913	0.75362	0.083	0.92029	1	0.56884	0.569	0.717	0.699	0.163	0.598
H_d (ZAL2)	0.54167	0.23333	0.625	0.125	0.94167	1	0.24167	0.24167	0.525	0.54167	0.125	0.125
H_d (ZAL2^m)	0.60714	0.57143	0.096429	0	0.89286	1	0	0	0.25	0	0.25	0.78571
π	0.0016	0.00495	0.00289	0.00019	0.00191	0.00293	0.00125	0.00765	0.00414	0.00151	0.00021	0.00299
π (ZAL2)	0.00141	0.00295	0.00243	0.00028	0.00160	0.00044	0.00029	0.0004	0.0014	0.00065	0.00015	0.00024
π (ZAL2^m)	0.00153	0.00723	0.00391	0	0.00238	0.00030	0	0	0.00033	0	0.00031	0.00304
θ	0.00181	0.00339	0.00469	0.0006	0.00275	0.00213	0.00122	0.00512	0.00391	0.00119	0.00066	0.00405
θ (ZAL2)	0.00136	0.00381	0.00352	0.00068	0.00219	0.00054	0.00069	0.00096	0.002	0.00067	0.00037	0.00057
θ (ZAL2^m)	0.00174	0.00488	0.00394	0	0.00233	0.00035	0	0	0.00051	0	0.00048	0.00365
Tests of neutrality												
TajimaD (16 ZAL2 + 8 ZAL2^m)	-0.29186	0.77632	-1.31485	-1.1593	-1.12844	1.50768	0.05885	1.6985	0.2041	0.7249	-1.5147	-0.8426
Tajima's D (ZAL2)	0.09547	-0.44843	-1.11292	-1.16221	-0.89877	-0.99663	-1.49796	-1.49796	-0.98088	-0.08238	-1.16221	-1.16221
Tajima's D (ZAL2^m)	-0.44794	1.44416	-0.03729	NA	-0.22175	-0.83892	NA	NA	-1.05482	NA	-1.05482	-0.7554
FuLiD* (16 ZAL2 + 8 ZAL2^m)	0.97946	0.62273	-2.02945	-1.6058	-1.39749	0.78754	-0.85599	0.591	0.5023	0.1142	-2.1591	-1.0527
FuLiF* (16 ZAL2 + 8 ZAL2^m)	0.71917	0.76213	-2.11623	-1.7042	-1.53687	1.19564	-0.68949	1.075	0.4816	0.3326	-2.2809	-1.1518
FuFs (16 ZAL2 + 8 ZAL2^m)	-1.784	1.06	-5.213	-1.028	-11.898	-2.437	0.4	6.085	1.426	0.944	-2.078	-1.582
FuLiD* (ZAL2)	-0.50381	0.68829	-1.80787	-1.45287	-1.30735	-0.51567	-1.9147	-1.9147	0.45883	-0.50381	-1.45287	-1.45287
FuLiF* (ZAL2)	-0.39619	0.45036	-1.85871	-1.5682	-1.37507	-0.67039	-2.06018	-2.06018	0.08156	-0.44901	-1.5682	-1.5682
FuFs (ZAL2)	0.037	0.083	-2.291	-0.7	-6.087	-5.999	-1.615	-1.615	-0.133	-0.094	-0.7	-0.7
FuLiD* (ZAL2^m)	-0.14931	0.88779	0.03986	NA	-0.17637	-0.79331	NA	NA	-1.12639	NA	-1.12639	-0.49407
FuLiF* (ZAL2^m)	-0.23785	1.10037	0.02424	NA	-0.20541	-0.89586	NA	NA	-1.20353	NA	-1.20353	-0.61102
FuFs (ZAL2^m)	-0.478	0.966	-3.588	NA	-3.515	-2.094	NA	NA	-0.182	NA	-0.182	-1.643
ZnS (ZAL2 + ZAL2^m)	0.023	NA	0.0694	NA	0.0613	0.4999	0.1815	0.6884	0.2458	0.2197	0.0019	0.153
ZnS (ZAL2)	0.1111	NA	0.1282	NA	0.0952	0.0865	0.0044	0.0044	0.212	0.0303	NA	NA
ZnS (ZAL2^m)	0.0476	NA	0.2167	NA	0.0769	0.1747	NA	NA	NA	NA	NA	0.2109
Data for synonymous and non-coding sites outside of most conserved elements (MCE)[c]												
Total sites outside MCE	139	0	588	55	782	1709	29	386	593	687	0	14
Synonymous and noncoding sites outside of MCE	139	0	583	55	777	1698	29	375	593	687	0	14
S, outside of MCE	2	0	8	0	14	25	0	11	9	4	0	1
d_{xy}	0.0036	0	0.00361	0	0.00294	0.00871	0	0.02607	0.00622	0.00346	0	0.00893
Silent divergence (K_s, 2 vs 2^m)	0.0036	0	0.00364	0	0.00296	0.00888	0	0.02683	0.00622	0.00346	0	0.00893
Silent π	0.00229	0	0.0034	0	0.00296	0.00449	0	0.01259	0.0037	0.00197	0	0.00595
Silent θ	0.00385	0	0.00551	0	0.00482	0.00320	0	0.00786	0.00406	0.00156	0	0.01913
Silent π (ZAL2)	0	0	0.00286	0	0.00216	0.00075	0	0.00033	0.00178	0.00085	0	0
Silent θ (ZAL2)	0	0	0.00414	0	0.00311	0.00083	0	0.0008	0.00254	0.00088	0	0
Silent π (ZAL2^m)	0.00617	0	0.00459	0	0.00455	0.00042	0	0	0.00042	0	0	0.01786
Silent θ (ZAL2^m)	0.00555	0	0.00463	0	0.00447	0.00051	0	0	0.00065	0	0	0.02755

*0.05 > p > 0.01, **0.01 > p > 0.001, *** p < 0.001

[a] Phasing assignments outside the inversion varied when phasing individual loci versus the total haplotype (all loci inside and outside the inversion together). Within the text of this paper, we report data on the total haplotype unless otherwise noted.

[b] Loci previously described in Thomas et al., 2008, Genetics 179: 1455-1468. Locus names from Thomas et al. 2008 shown in parentheses.

[c] Highly conserved elements identified by the "most conserved" track in the chicken assembly (galGal3) on the UCSC Genome browser (see Siepel A et al., 2005, Genome Res 15:1034-1050)

Supplementary Table 2 (continued). Population genetic statistics of individual loci in Chapter 2.

Locus Name	186	268	264	249	217	212	204	283	547	104	99	92	361
Locus location													
Estimated Z. albicollis position in Mb ZAL2 ($ZAL2^m$)	16.72 (44.37)	18.03 (45.69)	18.45 (46.11)	20.04 (47.69)	23.42 (51.07)	23.92 (51.58)	24.75 (52.40)	27.11 (54.76)	30.95 (58.61)	32.26 (59.91)	32.89 (60.54)	33.73 (61.38)	38.13 (65.78)
Orthologous chicken position (Mb)	18.68	26.87	26.47	24.92	21.72	21.23	20.48	28.34	5.47	10.49	9.97	92.60	36.15
Orhologous zebra finch position (Mb)	18.29	16.97	16.55	14.97	11.59	11.08	10.26	7.90	4.05	2.75	2.12	1.28	32.63
DNA polymorphism, divergence and gene flow													
Number of sites	650	530	397	418	419	497	627	567	740	548	532	411	461
Noncoding & synonymous sites	650	530	328.667	418	419	497	571.333	567	510.111	548	479.833	411	461
Segregating sites, S	3	1	0	5	3	2	5	5	5	2	6	0	3
Fixed polymorphisms	2	1	0	3	2	2	5	4	1	2	6	0	1
Shared shared polymorphisms	0	0	0	0	0	0	0	0	0	0	0	0	0
Polymorphic in ZAL2, monomorphic in $ZAL2^m$	1	0	0	2	0	0	0	0	1	0	0	0	0
Polymorphic in $ZAL2^m$, monomorphic in ZAL2	0	0	0	0	1	0	0	1	3	0	0	0	2
F_{ST} (2 vs 2^m)	0.90090***	1***	NA	0.90929***	0.88722***	1***	1***	0.93878***	0.67222***	1***	1***	NA	0.8***
F_{ST} (WS vs TS)	0.3473	0.46667*	NA	0.36841	0.41404*	0.46667*	0.46667*	0.43810*	0.28698	0.46667*	0.46667*	NA	0.37333
d_{xy} (2 vs 2^m)	0.00356	0.00189	0	0.00912	0.00567	0.00402	0.00797	0.00772	0.00203	0.00365	0.01128	0	0.00271
Number of haplotypes	4	2	1	4	3	2	2	3	5	2	2	1	4
Haplotype diversity, Hd	0.66304	0.464	0	0.685	0.518	0.464	0.464	0.518	0.612	0.464	0.464	0	0.511
H_d (ZAL2)	0.45833	0	0	0.50833	0	0	0	0	0.23333	0	0	0	0
H_d ($ZAL2^m$)	0	0	0	0	0.53571	0	0	0.53571	0.46429	0	0	0	0.46429
π	0.00196	0.00088	0	0.00495	0.00276	0.00187	0.0037	0.00367	0.00118	0.00169	0.00523	0	0.00137
π (ZAL2)	0.00082	0	0	0.00165	0	0	0	0	0.00032	0	0	0	0
π ($ZAL2^m$)	0	0	0	0	0.00128	0	0	0.00094	0.00101	0	0	0	0.00108
θ	0.00124	0.00051	0	0.0032	0.00192	0.00108	0.00214	0.00236	0.00181	0.00098	0.00302	0	0.00174
θ (ZAL2)	0.00046	0	0	0.00144	0	0	0	0	0.00041	0	0	0	0
θ ($ZAL2^m$)	0	0	0	0	0.00092	0	0	0.00068	0.00156	0	0	0	0.00167
Tests of neutrality													
TajimaD (16 ZAL2 + 8 $ZAL2^m$)	1.46643	1.2318	NA	1.5797	1.1034	1.6093	2.1208*	1.6111	-1.008	1.6093	2.2134*	NA	-0.5414
Tajima's D (ZAL2)	1.03439	NA	NA	0.37769	NA	NA	NA	NA	-0.44832	NA	NA	NA	NA
Tajima's D ($ZAL2^m$)	NA	NA	NA	NA	1.1665	NA	NA	1.1665	-1.44751	NA	NA	NA	-1.31009
FuLiD* (16 ZAL2 + 8 $ZAL2^m$)	0.97946	0.6227	NA	1.1663	0.9795	0.8373	1.1663	1.1663	-1.338	0.8373	1.2325	NA	-1.3573
FuLiF* (16 ZAL2 + 8 $ZAL2^m$)	1.2879	0.9033	NA	1.4884	1.1705	1.2089	1.6682*	1.4988	-1.4401	1.2089	1.7589*	NA	-1.3026
FuFs (16 ZAL2 + 8 $ZAL2^m$)	2.04	1.362	NA	2.198	1.776	2.923	6.423	3.699	-1.238	2.923	7.417	NA	-0.876
FuLiD* (ZAL2)	0.68829	NA	NA	0.90708	NA	NA	NA	NA	0.68829	NA	NA	NA	NA
FuLiF* (ZAL2)	0.88463	NA	NA	0.8769	NA	NA	NA	NA	0.45036	NA	NA	NA	NA
FuFs (ZAL2)	1.096	NA	NA	0.233	NA	NA	NA	NA	0.083	NA	NA	NA	NA
FuLiD* ($ZAL2^m$)	NA	NA	NA	NA	0.88779	NA	NA	0.88779	-1.56533	NA	NA	NA	-1.4098
FuLiF* ($ZAL2^m$)	NA	NA	NA	NA	1.0316	NA	NA	1.0316	-1.68583	NA	NA	NA	-1.51361
FuFs ($ZAL2^m$)	NA	NA	NA	NA	0.866	NA	NA	0.866	-0.305	NA	NA	NA	-0.999
ZnS (ZAL2 + $ZAL2^m$)	0.4211	NA	NA	0.4483	0.5238	1	1	0.7143	0.1322	1	1	NA	0.0586
ZnS (ZAL2)	NA	NA	NA	0.3143	NA	NA	NA	NA	NA	NA	NA	NA	NA
ZnS ($ZAL2^m$)	NA	NA	NA	NA	NA	NA	NA	NA	0.3469	NA	NA	NA	0.0204
Data for synonymous and non-coding sites outside of most conserved elements (MCE)[c]													
Total sites outside MCE	307	0	0	301	15	14	537	235	273	0	362	0	49
Synonymous and noncoding sites outside of MCE	307	0	0	301	15	14	537	235	208.278	0	362	0	49
S, outside of MCE	1	0	0	4	0	0	4	3	2	0	4	0	0
d_{xy}	0.00326	0	0	0.01038	0	0	0.00745	0.01011	0.00412	0	0.01105	0	0
Silent divergence (K_s, 2 vs 2^m)	0.00326	0	0	0.01038	0	0	0.00745	0.01011	0.00092	0	0.01105	0	0
Silent π	0.00151	0	0	0.00515	0	0	0.00345	0.00492	0.00062	0	0.00512	0	0
Silent θ	0.00087	0	0	0.00356	0	0	0.00199	0.00342	0.00198	0	0.00296	0	0
Silent π (ZAL2)	0	0	0	0.00078	0	0	0	0	0	0	0	0	0
Silent θ (ZAL2)	0	0	0	0.001	0	0	0	0	0	0	0	0	0
Silent π ($ZAL2^m$)	0	0	0	0	0	0	0	0.00228	0.00185	0	0	0	0
Silent θ ($ZAL2^m$)	0	0	0	0	0	0	0	0.00164	0.00285	0	0	0	0

*$0.05 > p > 0.01$, **$0.01 > p > 0.001$, *** $p < 0.001$

[a] Phasing assignments outside the inversion varied when phasing individual loci versus the total haplotype (all loci inside and outside the inversion together). Within the text of this paper, we report data on the total haplotype unless otherwise noted.

[b] Loci previously described in Thomas et al., 2008, Genetics 179: 1455-1468. Locus names from Thomas et al. 2008 shown in parentheses.

[c] Highly conserved elements identified by the "most conserved" track in the chicken assembly (galGal3) on the UCSC Genome browser (see Siepel A et al. 2005. Genome Res 15:1034-1050)

Supplementary Table 2 (continued). Population genetic statistics of individual loci in Chapter 2.

Locus Name	359	349	335	333	471 (MAP3K4)[b]	459	439	427	418	400	374	499	571
Locus location													
Estimated Z. albicollis position in Mb ZAL2 (ZAL2^m)	38.35 (66.00)	39.49 (67.15)	41.03 (68.68)	41.22 (68.87)	42.73 (70.38)	43.99 (71.65)	46.09 (73.74)	47.39 (75.04)	48.26 (75.91)	50.00 (77.65)	52.34 (79.99)	52.82 (80.47)	55.44 (83.09)
Orthologous chicken position (Mb)	35.95	34.95	33.53	33.37	47.16	45.97	43.95	42.72	41.88	40.01	37.41	49.98	57.15
Orhologous zebra finch position (Mb)	32.85	34.00	35.54	35.73	37.24	38.50	40.60	41.90	42.77	44.51	46.85	47.33	49.95
DNA polymorphism, divergence and gene flow													
Number of sites	472	417	290	636	355	623	489	543	680	345	479	570	553
Noncoding & synonymous sites	429	355.333	290	636	83.167	623	489	543	518	345	454	570	553
Segregating sites, S	3	2	1	4	1	4	2	7	7	1	1	9	0
Fixed polymorphisms	3	2	1	2	1	2	2	7	4	1	1	5	0
Shared shared polymorphisms	0	0	0	0	0	0	0	0	0	0	0	0	0
Polymorphic in ZAL2, monomorphic in ZAL2^m	0	0	0	1	0	1	0	0	2	0	0	4	0
Polymorphic in ZAL2^m, monomorphic in ZAL2	0	0	0	1	0	1	0	0	1	0	0	0	0
F_{ST} (2 vs 2^m)	1***	1***	1***	0.86065***	1***	0.91429***	1***	1***	0.90741***	1***	1***	0.90640***	NA
F_{ST} (WS vs TS)	0.46667*	0.46667*	0.46667*	0.35422	0.46667*	0.41481	0.46667*	0.46667*	0.36822	0.46667*	0.46667*	0.38884	NA
d_{xy} (2 vs 2^m)	0.00636	0.0048	0.00345	0.00373	0.00282	0.00351	0.00409	0.01289	0.00662	0.0029	0.00209	0.01086	0
Number of haplotypes	2	2	2	4	2	4	2	2	5	2	2	5	1
Haplotype diversity, Hd	0.464	0.464	0.464	0.609	0.464	0.543	0.464	0.464	0.725	0.464	0.464	0.743	0
H_d (ZAL2)	0	0	0	0.23333	0	0.125	0	0	0.54167	0	0	0.64167	0
H_d (ZAL2^m)	0	0	0	0.42857	0	0.25	0	0	0.25	0	0	0	0
π	0.00295	0.00222	0.0016	0.00196	0.00131	0.00176	0.0019	0.00598	0.00348	0.00134	0.00097	0.00592	0
π (ZAL2)	0	0	0	0.00037	0	0.0002	0	0	0.00086	0	0	0.00203	0
π (ZAL2^m)	0	0	0	0.00067	0	0.0004	0	0	0.00037	0	0	0	0
θ	0.0017	0.00128	0.00092	0.00168	0.00075	0.00172	0.0011	0.00345	0.00276	0.00078	0.00056	0.00423	0
θ (ZAL2)	0	0	0	0.00047	0	0.0048	0	0	0.00089	0	0	0.00211	0
θ (ZAL2^m)	0	0	0	0.00061	0	0.00062	0	0	0.00057	0	0	0	0
Tests of neutrality													
TajimaD (16 ZAL2 + 8 ZAL2^m)	1.8408	1.6093	1.2318	0.4474	1.2318	0.0588	1.6093	2.2874*	0.8194	1.2318	1.2318	1.3098	NA
Tajima's D (ZAL2)	NA	NA	NA	-0.44832	NA	-1.16221	NA	NA	-0.08238	NA	NA	-0.12231	NA
Tajima's D (ZAL2^m)	NA	NA	NA	0.3335	NA	-1.05482	NA	NA	-1.05482	NA	NA	NA	NA
FuLiD* (16 ZAL2 + 8 ZAL2^m)	0.9795	0.8373	0.6227	1.0844	0.6227	-0.856	0.8373	1.2873	-0.0289	0.6227	0.6227	0.8272	NA
FuLiF* (16 ZAL2 + 8 ZAL2^m)	1.409	1.2089	0.9033	1.0449	0.9033	-0.6895	1.2089	1.8336**	0.2538	0.9033	0.90326	1.128	NA
FuFs (16 ZAL2 + 8 ZAL2^m)	4.207	2.923	1.362	0.726	1.362	0.4	2.923	8.354	1.499	1.362	1.362	2.78	NA
FuLiD* (ZAL2)	NA	NA	NA	0.68829	NA	-1.45287	NA	NA	-0.50381	NA	NA	0.25371	NA
FuLiF* (ZAL2)	NA	NA	NA	0.45036	NA	-1.5628	NA	NA	-0.44901	NA	NA	0.17539	NA
FuFs (ZAL2)	NA	NA	NA	0.083	NA	-0.7	NA	NA	-0.094	NA	NA	0.071	NA
FuLiD* (ZAL2^m)	NA	NA	NA	0.88779	NA	-1.12639	NA	NA	-1.12639	NA	NA	NA	NA
FuLiF* (ZAL2^m)	NA	NA	NA	0.82528	NA	-1.20353	NA	NA	-1.20353	NA	NA	NA	NA
FuFs (ZAL2^m)	NA	NA	NA	0.536	NA	-0.182	NA	NA	-0.182	NA	NA	NA	NA
ZnS (ZAL2 + ZAL2^m)	1	1	1	0.2438	1	0.2032	1	1	0.3327	1	1	0.3806	NA
ZnS (ZAL2)	NA	NA	NA	NA	NA	NA	NA	NA	0.0303	NA	NA	0.1915	NA
ZnS (ZAL2^m)	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA	NA
Data for synonymous and non-coding sites outside of most conserved elements (MCE)[c]													
Total sites outside MCE	276	56	132	418	29	2	138	349	207	318	94	132	152
Synonymous and noncoding sites outside of MCE	276	56	132	418	6	0	138	349	207	318	94	132	152
S, outside of MCE	3	0	1	4	0	0	0	4	2	1	0	3	0
d_{xy}	0.01087	0	0.00758	0.00568	0	0	0	0.01146	0.00543	0.00314	0	0.02083	0
Silent divergence (K_s, 2 vs 2^m)	0.01087	0	0.00758	0.00568	0	0	0	0.01146	0.00543	0.00314	0	0.02083	0
Silent π	0.00504	0	0.00351	0.00298	0	0	0	0.00532	0.00264	0.00146	0	0.01098	0
Silent θ	0.00291	0	0.00203	0.00256	0	0	0	0.00307	0.00259	0.00084	0	0.00609	0
Silent π (ZAL2)	0	0	0	0.00056	0	0	0	0	0	0	0	0.00303	0
Silent θ (ZAL2)	0	0	0	0.00072	0	0	0	0	0	0	0	0.00228	0
Silent π (ZAL2^m)	0	0	0	0.00103	0	0	0	0	0.00121	0	0	0	0
Silent θ (ZAL2^m)	0	0	0	0.00092	0	0	0	0	0.00186	0	0	0	0

*0.05 > p > 0.01, **0.01 > p > 0.001, *** p < 0.001

[a] Phasing assignments outside the inversion varied when phasing individual loci versus the total haplotype (all loci inside and outside the inversion together). Within the text of this paper, we report data on the total haplotype unless otherwise noted.

[b] Loci previously described in Thomas et al., 2008, Genetics 179: 1455-1468. Locus names from Thomas et al. 2008 shown in parentheses.

[c] Highly conserved elements identified by the "most conserved" track in the chicken assembly (galGal3) on the UCSC Genome browser (see Siepel A et al., 2005, Genome Res 15:1034-1050)

Supplementary Table 2 (continued). Population genetic statistics of individual loci in Chapter 2.

Locus Name	560 (REPS1)[b]	530	513 (VIP)[b]	510 (ESR1)[b]	603	621 (TRMT11)[b]	689	703	1068	1051	1021	1007	991
Locus location													
Estimated Z. albicollis position in Mb ZAL2 (ZAL2^m)	56.55 (84.20)	59.69 (87.34)	61.80 (89.08)	65.91 (89.45)	67.46 (93.56)	74.03 (95.12)	75.49 (101.68)	77.18 (103.14)	78.89 (104.83)	82.07 (106.55)	83.46 (109.72)	85.21 (111.11)	85.78 (32.90)
Orthologous chicken position (Mb)	56.09	53.08	51.39	51.04	60.32	62.16	68.97	70.36	106.90	105.11	102.12	100.76	99.14
Orhologous zebra finch position (Mb)	51.06	54.19	55.93	56.30	60.42	61.97	68.54	69.99	106.90	105.19	102.01	100.62	98.87
DNA polymorphism, divergence and gene flow													
Number of sites	1242	227	733	1136	723	1168	480	413	795	764	540	665	361
Noncoding & synonymous sites	1207.167	227	667.5	918.666	701.333	1157.667	480	413	795	181.944	540	665	361
Segregating sites, S	16	1	5	12	7	16	5	0	6	3	1	2	1
Fixed polymorphisms	16	1	5	10	5	14	3	0	5	3	1	2	1
Shared shared polymorphisms	0	0	0	1	0	0	0	0	0	0	0	0	0
Polymorphic in ZAL2, monomorphic in ZAL2^m	0	0	0	1	1	0	1	0	1	0	0	0	0
Polymorphic in ZAL2^m, monomorphic in ZAL2	0	0	0	0	1	2	1	0	0	0	0	0	0
F_{ST} (2 vs 2^m)	1***	1***	1***	0.96282***	0.96386***	0.97640***	0.94118***	NA	0.98765***	1***	1***	1***	1***
F_{ST} (WS vs TS)	0.46667*	0.46667*	0.46667*	0.42726	0.43411	0.45565	0.40377	NA	0.44444*	0.46667*	0.46667*	0.46667*	0.46667*
d_{xy} (2 vs 2^m)	0.01288	0.00441	0.00682	0.00917	0.00717	0.01231	0.00664	0	0.00637	0.00393	0.00185	0.00301	0.00277
Number of haplotypes	2	2	2	5	4	4	4	1	3	2	2	2	2
Haplotype diversity, Hd	0.46377	0.464	0.464	0.703	0.543	0.511	0.543	0	0.518	0.464	0.464	0.464	0.464
H_d (ZAL2)	0	0	0	0.49167	0.125	0	0.125	0	0.125	0	0	0	0
H_d (ZAL2^m)	0	0	0	0.25	0.25	0.46429	0.25	0	0	0	0	0	0
π	0.00597	0.00204	0.00316	0.00448	0.00344	0.00577	0.00325	0	0.00302	0.00182	0.00086	0.00139	0.00128
π (ZAL2)	0	0	0	0.00046	0.00017	0	0.00026	0	0.00016	0	0	0	0
π (ZAL2^m)	0	0	0	0.00022	0.00035	0.00058	0.00052	0	0	0	0	0	0
θ	0.00345	0.00118	0.00183	0.00283	0.00259	0.00367	0.00279	0	0.00202	0.00105	0.0005	0.00081	0.00074
θ (ZAL2)	0	0	0	0.00053	0.00042	0	0.00063	0	0.00038	0	0	0	0
θ (ZAL2^m)	0	0	0	0.00034	0.00053	0.00066	0.0008	0	0	0	0	0	0
Tests of neutrality													
TajimaD (16 ZAL2 + 8 ZAL2^m)	2.60593**	1.2318	2.1208*	2.0002	1.01874	2.0368*	0.4741	NA	1.4973	1.8408	1.2318	1.6093	1.2318
Tajima's D (ZAL2)	NA	NA	NA	-0.3301	-1.16221	NA	-1.16221	NA	-1.16221	NA	NA	NA	NA
Tajima's D (ZAL2^m)	NA	NA	NA	-1.05482	-1.05482	-0.44794	-1.05482	NA	NA	NA	NA	NA	NA
FuLiD* (16 ZAL2 + 8 ZAL2^m)	1.54623*	0.6227	1.1663	1.4645*	-0.02894	1.2004	-0.5032	NA	0.4974	0.9795	0.6227	0.8373	0.6227
FuLiF* (16 ZAL2 + 8 ZAL2^m)	2.17995*	0.9033	1.6682*	1.8937*	0.32128	1.6972*	-0.2581	NA	0.9127	1.409	0.9033	1.2089	0.9033
FuFs (16 ZAL2 + 8 ZAL2^m)	15.107	1.362	6.423	4.741	2.853	8.428	1.326	NA	4.295	4.207	1.362	2.923	1.362
FuLiD* (ZAL2)	NA	NA	NA	-0.50381	-1.45287	NA	-1.45287	NA	-1.45287	NA	NA	NA	NA
FuLiF* (ZAL2)	NA	NA	NA	-0.52297	-1.5682	NA	-1.5682	NA	-1.5682	NA	NA	NA	NA
FuFs (ZAL2)	NA	NA	NA	-0.29	-0.7	NA	-0.7	NA	-0.7	NA	NA	NA	NA
FuLiD* (ZAL2^m)	NA	NA	NA	-1.12639	-1.12639	-0.14931	-1.12639	NA	NA	NA	NA	NA	NA
FuLiF* (ZAL2^m)	NA	NA	NA	-1.20353	-1.20353	-0.23785	-1.20353	NA	NA	NA	NA	NA	NA
FuFs (ZAL2^m)	NA	NA	NA	-0.182	-0.182	-0.478	-0.182	NA	NA	NA	NA	NA	NA
ZnS (ZAL2 + ZAL2^m)	1	1	1	0.699	0.5022	0.7937	0.3328	NA	0.6739	1	1	1	1
ZnS (ZAL2)	NA	NA	NA	0.0222	NA	NA	NA	NA	NA	NA	NA	NA	NA
ZnS (ZAL2^m)	NA	NA	NA	NA	NA	0.4286	NA	NA	NA	NA	NA	NA	NA
Data for synonymous and non-coding sites outside of most conserved elements (MCE)[c]													
Total sites outside MCE	1198	200	142	515	284	1137	196	0	303	229	78	180	107
Synonymous and noncoding sites outside of MCE	1162.167	200	136	467.667	284	1133.333	196	0	303	50.667	78	180	107
S, outside of MCE	15	1	0	7	4	16	2	0	4	1	0	0	1
d_{xy}	0.01252	0.005	0	0.01053	0.00748	0.01264	0.00574	0	0.01011	0.00437	0	0	0.00935
Silent divergence (K_s, 2 vs 2^m)	0.01291	0.005	0	0.00732	0.00748	0.01268	0.00574	0	0.01011	0	0	0	0.00935
Silent π	0.00599	0.00232	0	0.00394	0.00385	0.00594	0.00279	0	0.00487	0	0	0	0.00433
Silent θ	0.00346	0.00134	0	0.00286	0.00377	0.00378	0.00273	0	0.00354	0	0	0	0.0025
Silent π (ZAL2)	0	0	0	0.00112	0.00088	0	0	0	0.00041	0	0	0	0
Silent θ (ZAL2)	0	0	0	0.00129	0.00212	0	0	0	0.00099	0	0	0	0
Silent π (ZAL2^m)	0	0	0	0.00053	0	0.0006	0.00128	0	0	0	0	0	0
Silent θ (ZAL2^m)	0	0	0	0.00082	0	0.00068	0.00197	0	0	0	0	0	0

*0.05 > p > 0.01, **0.01 > p > 0.001, *** p < 0.001

[a] Phasing assignments outside the inversion varied when phasing individual loci versus the total haplotype (all loci inside and outside the inversion together). Within the text of this paper, we report data on the total haplotype unless otherwise noted.

[b] Loci previously described in Thomas et al., 2008, Genetics 179: 1455-1468. Locus names from Thomas et al. 2008 shown in parentheses.

[c] Highly conserved elements identified by the "most conserved" track in the chicken assembly (galGal3) on the UCSC Genome browser (see Siepel A et al., 2005, Genome Res 15:1034-1050)

Supplementary Table 2 (continued). Population genetic statistics of individual loci in Chapter 2.

Locus Name	985	974	965	938	910	905 (FAM83B)[b]	885	842	833	822	793 (CGA)[b]	750	748
Locus location													
Estimated Z. albicollis position in Mb ZAL2 (ZAL2^m)	86.97 (32.33)	87.85 (31.14)	87.85 (30.26)	90.60 (27.51)	93.53 (24.58)	94.06 (24.04)	95.96 (22.15)	100.22 (17.89)	101.16 (16.95)	102.26 (15.85)	105.17 (12.94)	109.39 (8.71)	109.61 (8.50)
Orthologous chicken position (Mb)	98.59	97.43	96.54	93.83	91.02	90.51	88.53	84.26	83.30	82.22	79.32	75.05	74.83
Orhologous zebra finch position (Mb)	98.30	97.11	96.23	93.48	90.55	90.01	88.12	83.86	82.93	81.82	78.91	74.69	74.47
DNA polymorphism, divergence and gene flow													
Number of sites	511	534	668	667	592	391	539	585	683	672	1045	553	394
Noncoding & synonymous sites	511	534	668	420.986	546.833	83.389	126.375	476.667	683	643.333	967.333	553	394
Segregating sites, S	3	0	2	4	7	5	7	9	12	10	14	5	1
Fixed polymorphisms	1	0	1	3	7	5	1	6	6	2	5	0	1
Shared shared polymorphisms	0	0	0	0	0	0	0	0	0	0	0	0	0
Polymorphic in ZAL2, monomorphic in ZAL2^m	1	0	1	1	0	0	5	2	4	7	8	3	0
Polymorphic in ZAL2^m, monomorphic in ZAL2	1	0	0	0	0	0	1	1	2	1	1	2	0
F_{ST} (2 vs 2^m)	0.82667***	NA	0.94118***	0.97959***	1***	1***	0.60504**	0.93188**	0.88010**	0.75322**	0.79702**	0.11575	1***
F_{ST} (WS vs TS)	0.20276	NA	0.37333*	0.43077*	0.46667*	0.46667*	0.22646	0.41187	0.34818	0.37762	0.22563	0.02925	0.46667*
d_{xy} (2 vs 2^m)	0.00367	0	0.00159	0.00459	0.01182	0.01279	0.00394	0.01122	0.01071	0.0053	0.00801	0.00147	0.00254
Number of haplotypes	4	1	3	3	2	2	8	5	8	8	12	6	2
Haplotype diversity, Hd	0.663	0	0.518	0.518	0.464	0.464	0.848	0.76667	0.81522	0.786	0.88043	0.543	0.464
H_d (ZAL2)	0.4	0	0.125	0.125	0	0	0.78333	0.34167	0.66667	0.68333	0.9	0.575	0
H_d (ZAL2^m)	0.25	0	0	0	0	0	0.42857	0.53571	0.60714	0.25	0.25	0.46429	0
π	0.00209	0	0.00082	0.00221	0.00548	0.00593	0.00292	0.00556	0.00575	0.00347	0.00505	0.0014	0.00118
π (ZAL2)	0.00078	0	0.00019	0.00019	0	0	0.00232	0.00061	0.00157	0.00224	0.00301	0.00137	0
π (ZAL2^m)	0.00049	0	0	0	0	0	0.008	0.00092	0.00099	0.00037	0.00024	0.00123	0
θ	0.00157	0	0.0008	0.00161	0.00317	0.00342	0.00348	0.00412	0.0047	0.00398	0.00359	0.00242	0.00068
θ (ZAL2)	0.00059	0	0.00045	0.00045	0	0	0.0028	0.00103	0.00176	0.00314	0.00231	0.00163	0
θ (ZAL2^m)	0.00075	0	0	0	0	0	0.00072	0.00066	0.00113	0.00057	0.00037	0.00139	0
Tests of neutrality													
TajimaD (16 ZAL2 + 8 ZAL2^m)	0.8312	NA	0.0473	1.0302	2.2874*	2.1208*	-0.5036	1.14716	0.76236	-0.4293	1.42964	-1.2198	1.23177
Tajima's D (ZAL2)	0.64998	NA	-1.16221	-1.16221	NA	NA	-0.5617	-1.03789	-0.33859	-1.00422	1.10518	-0.46739	NA
Tajima's D (ZAL2^m)	-1.05482	NA	NA	NA	NA	NA	0.3335	1.1665	-0.44794	-1.05482	-1.05482	-0.44794	NA
FuLiD* (16 ZAL2 + 8 ZAL2^m)	-0.1889	NA	-0.6609	0.1142	1.2873	1.1663	-0.0289	0.82719	0.15425	-0.1036	0.7378	-0.5032	0.62273
FuLiF* (16 ZAL2 + 8 ZAL2^m)	0.1119	NA	-0.536	0.4328	1.8336*	1.6682*	-0.1944	1.07215	0.39182	-0.2331	1.10383	-0.8211	0.90326
FuFs (16 ZAL2 + 8 ZAL2^m)	0.343	NA	0.089	2.481	8.354	6.423	-2.626	2.635	0.469	-1.23	-1.577	-2.767	1.362
FuLiD* (ZAL2)	0.68829	NA	-1.45287	-1.45287	NA	NA	-0.29672	-0.50381	-0.63393	0.14654	0.82878	-0.03858	NA
FuLiF* (ZAL2)	0.77204	NA	-1.5682	-1.5682	NA	NA	-0.42186	-0.73427	-0.63536	-0.1914	1.03892	-0.17393	NA
FuFs (ZAL2)	0.872	NA	-0.7	-0.7	NA	NA	-2.002	-0.979	-1.234	-1.463	-3.564	-0.869	NA
FuLiD* (ZAL2^m)	-1.12639	NA	NA	NA	NA	NA	0.88779	0.88779	-0.14931	-1.12639	-1.12639	-0.14931	NA
FuLiF* (ZAL2^m)	-1.20353	NA	NA	NA	NA	NA	0.82528	1.0316	-0.23785	-1.20353	-1.20353	-0.23785	NA
FuFs (ZAL2^m)	-0.182	NA	NA	NA	NA	NA	0.536	0.866	-0.478	-0.182	-0.182	-0.478	NA
ZnS (ZAL2 + ZAL2^m)	0.2101	NA	0.0217	0.5109	1	1	0.067	0.4761	0.2998	0.1085	0.2749	0.1028	1
ZnS (ZAL2)	NA	NA	NA	NA	NA	NA	0.0939	0.0095	0.0952	0.1547	0.2252	0.1788	NA
ZnS (ZAL2^m)	NA	NA	NA	NA	NA	NA	NA	NA	0.0476	NA	NA	0.4286	NA
Data for synonymous and non-coding sites outside of most conserved elements (MCE)[c]													
Total sites outside MCE	406	0	573	407	369	112	150	289	314	158	815	144	14
Synonymous and noncoding sites outside of MCE	406	0	573	357	369	26.667	34.167	281	314	158	809.333	144	14
S, outside of MCE	3	0	2	1	5	2	1	8	10	7	11	5	0
d_{xy}	0.00462	0	0.000185	0.00246	0.01355	0.01786	0.00333	0.02249	0.02269	0.01068	0.00817	0.00564	0
Silent divergence (K_s, 2 vs 2^m)	0.00462	0	0.00185	0.0028	0.01355	0	0.01463	0.02313	0.02269	0.01068	0.00822	0.00564	0
Silent π	0.00263	0	0.00095	0.0013	0.00628	0	0.01357	0.01128	0.01198	0.00803	0.00545	0.00538	0
Silent θ	0.00198	0	0.00093	0.00075	0.00363	0	0.00784	0.00762	0.00853	0.01186	0.00364	0.0093	0
Silent π (ZAL2)	0.00099	0	0.00022	0	0	0	0.01561	0.00083	0.00303	0.0067	0.00362	0.00527	0
Silent θ (ZAL2)	0.00074	0	0.00053	0	0	0	0.00882	0.00107	0.00288	0.00954	0.00298	0.00628	0
Silent π (ZAL2^m)	0.00062	0	0	0	0	0	0	0.00191	0.00136	0.00158	0	0.00471	0
Silent θ (ZAL2^m)	0.00095	0	0	0	0	0	0	0.00137	0.00123	0.00244	0	0.00536	0

*0.05 > p > 0.01, **0.01 > p > 0.001, *** p < 0.001

[a] Phasing assignments outside the inversion varied when phasing individual loci versus the total haplotype (all loci inside and outside the inversion together). Within the text of this paper, we report data on the total haplotype unless otherwise noted.

[b] Loci previously described in Thomas et al., 2008, Genetics 179: 1455-1468. Locus names from Thomas et al. 2008 shown in parentheses.

[c] Highly conserved elements identified by the "most conserved" track in the chicken assembly (galGal3) on the UCSC Genome browser (see Siepel A et al. 2005, Genome Res 15:1034-1050)

Supplemental Table 3. Primers used for sequencing BAC clones in Chapter 3.

White-throated sparrow BAC clone name	Orthologous position in zebra finch (taeGut1)	Orthologous chromosome in chicken (galGal3)	Forward primer	Reverse primer
CH264-077A17C7	chr1:18904021-18904367	1	GGACAACCTAATCCAGAAATG	AGCTATCTGTCTTTGGCTGAG
CH264-077A17S6Wu	chr1:19101128-19101560	1	AGAGGTGCTATAGTCCCAGTG	CGACTGTTTAATCAAGGGAAC
CH264-523H22C7	chr1:96963900-96964368	1	TTTAGATTTCCAGGGACTCAG	GATATAAGCACCTTGGTCCAG
CH264-172F07C7	chr1A:71724269-71724655	1	AGACAAAGACCTCCAGATGAG	AGAGTGAATCTCCCACTCAAG
CH264-172F07S6Wu	chr1A:71905680-71905986	1	CAGTCTCTTGATTTCCCTCAG	GTTCATCCCATCAGATAATCC
CH264-271I04C7	chr1A:36072541-36073068	1	TTAGTGCCTCACACTGACATC	AGCCAAAGTAACAAACTCCTG
CH264-271I04S6Wu	chr1A:35924191-35924551	1	AAGACTAAGGCCTCAACTCAG	TCGTGTATTGTTGCTGCTAAG
CH264-128L17C7	chr2:12151777-12152255	2	TTTGACACACTGCTTTAGCTC	CCCTAAATTTGCACTGTAAGC
CH264-128L17S6Wu	chr2:11974192-11974641	2	GAATCCAGTGAAATGTCCTTC	CCCATCTCTATTAGCATGTGAC
CH264-006N19S6Wu	chr4:30833869-30834199	4	AGGTGAAAGGAGCTGTTATTG	TGACTGAGCACAAAGGTGTAG
CH264-215E08C7	chr4:44229579-44229962	4	AGCCTCAAGAATAACAACACC	CAGGACTGCTTTATCACACTG
CH264-215E08S6Wu	chr4:44061228-44061597	4	CGGCATCTTATTACACATCAG	AGTGTCTTCCCTTGTAACAGC
CH264-106A13C7	chr6:24970066-24970488	6	AATCACTTGTAATCCTGAGCAC	TTGTCATGGAGACATAACCAG
CH264-274A08C7	chr7:27817587-27817952	7	AACAGAGAAGTCCAACCAAAC	CACCTCTAGGATGCAAAGTG
CH264-134P10S6Wu	chr8:24896047-24896526	8	CAGCTTAATTTCTGCCCTATC	GTTACAGGCAGGGTTAGATTC
CH264-134P10C7	chr8:25059175-25059656	8	ATCTACTTCAGCGAGGTAAGC	TGATCTGTTACACAGCACTCC
CH264-510P14C7	chr9:123877-124174	9	AACCTATCATCACCCAGACAG	GAAATCCTGTTTCAGTTGACC
CH264-510P14S6Wu	chr9:313958-314345	9	TGGTTCTCTAGGTACTCTCAGC	CAGTTCTTAAACAGATGCACAG
CH264-423C02C7	chr10:14004257-14004530	10	AGCAGAGGGAAGACTCTCAG	CCACCTTTACTCTCCTGTAGC
CH264-423C02S6Wu	chr10:14168328-14168731	10	GCTCTGTATTCAGCAACACAC	AACAGATCCCTTGAACATCTC
CH264-004B01C7	chr17:3831071-3831292	17	GGCATCAAACAGAGATTAAGG	TACAATTCTGAGCCCAGTTTC
CH264-004B01S6Wu	chr17:9361745-9361926	17	TGGTTTCACAGTGACTCCTC	CTGCTCTAAACAAGTCCAGAAG
CH264-171B18C7	chr19:88123-88405	19	AGACTGACAAAGCATTTAGGG	AATCTGTGAATTCCTGCTCTC
CH264-436P16S6Wu	chr24: unknown[a]	24	TAACAAAGGCTTTCTCCTCTG	CCTCTACCTGCCTTTACACC
CH264-246N16S6Wu	chr27:2886212-2886485	27	AGAACTGTGAAAGATGCACTG	AATGGGACAGACTTAGCAGAG
CH264-246N16C7	chr27:2994875-2995251	27	TCAGTGTCAAACCAAGTAAGTG	TGCAGTATTTCCTTCTACCAAG
CH264-378L22C7	chr28:1654232-1654727	28	AGCAGGGCTAAGAACTAATTG	CACACACAAAGAAACAAGTGG
CH264-028F08S6Wu	chrZ:21562548-21563048	Z	TCTTCTCCTGGTTCCATAAAC	TGATATGACTCATTGGCTCAG
CH264-111E01S6Wu	chrZ:34164982-34165463	Z	ATTCCCAAGTAGTCTTGTTCG	GGAGAGGAAACCAAATTCAC
CH264-173G13C7	chrZ:20795813-20796484	Z	AGCCAACATCAGTTTAGAACC	CTTCAGGACTGGATGAATCTC
CH264-184O11S6wu	chrZ:17126280-17126433	Z	AAGAACTTGGAACAGCAATG	CTTGGCAATACTGTTGTTGAG
CH264-184O11C7	chrZ:21565574-21565888	Z	GCTGCTCACAAAGTTAAACAG	GAGGACACTGCTGTCACATAG
CH264-350I20C7	chrZ:14572467-14572858	Z	AACACAGCCACAGTGTATTTC	TTGTAGCCCTTCATTCTCTTC
CH264-350I20S6Wu	chrZ:14725447-14725889	Z	GGATGATGCACTACACAACAC	CAAAGGACTTTACCCAGTCAC
CH264-458C15S6Wu	chrZ:24572726-24573190	Z	AACATTAACTGCACCACTGAG	AAGTTAAGTCCATCCATGAGG
CH264-458C15C7	chrZ:24717126-24717571	Z	ATGACTACAGGAGCAGAGACC	CTTCTGTTGGGATAACAGACC
CH264-448L22 (131200)[b]	chrW:unknown[c]	W	GAGCAGATTGATGCTTCTTGAG	TCTAGCTGCACAGCAAACATTAC
CH264-448L22 (133000)[b]	chrW:unknown[c]	W	CTTCAGAAGGTGAAGCAGAAGG	TATCGGAGAGACAGCTCTGATG
CH264-448L22 (135000)[b]	chrW:unknown[c]	W	GCTCTTTGGGATCAGTTATTGG	GCAGTCGTTTCTCACACTGATG

[a]While this BAC end did not map to a specific location in zebra finch, the other end mapped to chr24:2960785-2961147.

[b]All W loci, named 131200, 133000 and 135000, were designed from the same BAC clone.

[c]The orthologous location of the W loci could not be determined in the zebra finch as the genome sequence is from a male (ZZ).

Supplemental Table 4. Population genetic analysis and summary statistics of autosomal and sex-linked loci (Chapter 3) in the white-throated sparrow (*Zonotrichia albicollis*).

White-throated sparrow BAC clone name	*Orthologous chromosome in chicken*	*Length (bp)*	*Number of individuals sampled*	*S*	*Silent π*	*Silent θ*	*Tajima's D*	F_{ST}	*ZnS*
Macrochromosomes									
CH264-077A17C7	1	362	11	3	0.00202	0.00227	-0.29	-0.09	0.02
CH264-077A17S6Wu	1	433	11	6	0.00452	0.00380	0.59	-0.07	0.34
CH264-523H22C7	1	410	10	2	0.00049	0.00138	-1.51	0.00	n.a.
CH264-172F07C7	1	382	9	5	0.00260	0.00381	-1.00	-0.06	0.03
CH264-172F07S6Wu	1	305	9	3	0.00227	0.00286	-0.57	-0.08	0.05
CH264-271I04C7	1	513	9	2	0.00062	0.00113	-1.10	-0.08	n.a.
CH264-271I04S6Wu	1	367	9	1	0.00030	0.00079	-1.16	0.00	n.a.
CH264-128L17C7	2	489	9	7	0.00389	0.00416	-0.22	0.03	0.16
CH264-128L17S6Wu	2	458	9	7	0.00252	0.00386	-1.79	-0.03	0.20
CH264-006N19S6Wu	4	340	11	0	0.00000	0.00000	n.a.	n.a.	n.a.
CH264-215E08C7	4	339	8	6	0.00302	0.00533	-1.48	0.01	0.47
CH264-215E08S6Wu	4	375	8	2	0.00173	0.00161	0.20	-0.09	n.a.
Total[a]		*4773*		*44*	*0.00198 ± 0.00033*	*0.00259 ± 0.00040*			
Intermediate chromosomes									
CH264-106A13C7	6	443	10	12	0.00522	0.00764	-1.14	-0.03	0.09
CH264-274A08C7	7	367	9	3	0.00167	0.00238	-0.82	0.15	0.46
CH264-134P10C7	8	485	12	13	0.00499	0.00718	-1.06	0.00	0.07
CH264-134P10S6Wu	8	495	12	26	0.00960	0.01407	-1.19	-0.02	0.07
CH264-510P14C7	9	297	11	20	0.01071	0.01847	-1.56	-0.01	0.06
CH264-510P14S6Wu	9	381	11	7	0.00251	0.00504	-1.61	0.10	0.07
CH264-423C02C7	10	274	10	1	0.00037	0.00103	-1.16	0.00	n.a.
CH264-423C02S6Wu	10	359	10	5	0.00309	0.00393	-0.65	0.05	0.09
Total[a]		*3101*		*87*	*0.00485 ± 0.00054*	*0.00713 ± 0.00082*			
Microchromosomes									
CH264-004B01C7	17	221	8	15	0.01833	0.02045	-0.41	-0.01	0.14
CH264-004B01S6Wu	17	191	11	26	0.02888	0.03734	-0.86	0.02	0.12
CH264-171B18C7	19	294	8	17	0.01346	0.01743	-0.90	0.20	0.20
CH264-436P16S6Wu	24	280	9	3	0.00154	0.00312	-1.40	-0.06	0.01
CH264-246N16C7	27	359	9	29	0.01836	0.02342	-0.87	-0.04	0.10
CH264-246N16S6Wu	27	275	9	34	0.02921	0.03595	-0.76	0.00	0.10
CH264-378L22C7	28	463 (360)[b]	9	12	0.00623	0.00754	-0.64	-0.03	0.10
Total[a]		*2083 (1980)*[b]		*136*	*0.01591 ± 0.00115*	*0.001962 ± 0.00164*			
Sex chromosomes									
CH264-028F08S6Wu	Z	498	11	0	0.00000	0.00000	n.a.	n.a.	n.a.
CH264-111E01S6Wu	Z	484	10	6	0.00179	0.00349	-1.55	-0.02	0.07
CH264-173G13C7	Z	487	11	2	0.00069	0.00113	-0.87	0.17	n.a.
CH264-184O11C7	Z	314	9	1	0.00035	0.00093	-1.16	0.00	n.a.
CH264-184O11S6Wu	Z	249	9	1	0.00084	0.00117	-0.53	-0.12	n.a.
CH264-350I20C7	Z	407	9	1	0.00027	0.00071	-1.16	0.00	n.a.
CH264-350I20S6Wu	Z	491	9	0	0.00000	0.00000	n.a.	n.a.	n.a.
CH264-458C15C7	Z	447 (218)[b]	10	0	0.00000	0.00000	n.a.	n.a.	n.a.
CH264-458C15S6Wu	Z	455	10	1	0.00022	0.00062	-1.16	0.00	n.a.
Total[a]		*3832 (3603)*[b]		*12*	*0.00050 ± 0.00009*	*0.00079 ± 0.00025*			
CH264-448L22 (131200)[c]	W	620	23	0	0.00000	0.00000	n.a.	n.a.	n.a.
CH264-448L22 (133000)[c]	W	564	23	1	0.00015	0.00048	-1.16	0.00	n.a.
CH264-448L22 (135000)[c]	W	545	24	0	0.00000	0.00000	n.a.	n.a.	n.a.
Total[a]		*1729*		*1*	*0.000005 ± 0.00005*	*0.00016 ± 0.00016*			

[a]Average diversity (π and θ) is calculated from concatenated haplotypes from all loci in each category. Averages are reported ± SD.

[b]Numbers in parentheses indicates the number of synonymous and noncoding positions.

[c]All W loci, named 131200, 133000 and 135000, were designed from the same BAC clone.

Supplemental Table 5. Primer sequences, orthologous locations and Tm for loci in candidate SNP study.

SNP Target (Gene symbol and SNP position)	*Orthologous position on Zebra finch chr3*[a]	*Forward primer*	*Reverse primer*	T_m
ESR1_543	chr3:56,300,867-56,302,155	AAGGTGCTCCTGTCTGCTTAAC	AAATGCTGGTGTCTGTTGTCAC	58/55[a]
FSHR_36_43	chr3:21,855,905-21,856,635	GTCTCACCAAGCAAGACATGAA	AGAGACCTCACAATGAGCACAA	55
FSHR_241	chr3:21,781,537-21,781,902	ACACCACTTGATTGTTGTTTGC	GCAATGGCTAGGATAGGTGAAG	55
FSHR_485_622_668	chr3:21,774,636-21,775,300	ATGCCATAGACTGGCAGACC	AATCTGGGCTTGCATTTCAC	55
LH_306	chr3:21,659,479-21,660,193	CGTGTCTCTGAAGATCACTTGC	ACAGCCAACTCAACAGTTCTGA	55
LH_570_614_668	chr3:21,656,555-21,657,255	CTGGCACACCATCACCTATG	GAGATGTCTTGTCCCGTGGT	55
SRD5A2_exon2	chr3:33,935,927-33,936,572	TGGAGCATTAAATCAAGTCTGG	TGCTGATTTCCATCACTACAGG	55
SRD5A2_exon3	chr3:33,934,009-33,934,715	TGATGTGCTGCATAACTCTGTG	ATACTCTTGCTGCTTTCCTTGG	55
SRD5A2_exon4	chr3:33,933,475-33,934,086	TTGTGAAAGTGCTGGAAACATC	AATGAGGCTCAGTGAAGAGAGG	55
SRD5A2_exon5	chr3:33,932,171-33,932,692	TTAAGGTTGCTCAGCTGTTCTG	TTGCAGAAGTGCTGTAAATTGG	55
HTR1E_25	chr3:78,947,816-78,948,432	GAATAGCCATTTCACATCAGCA	ACAGCAACAAGGAGATCTGTCA	55
HTR1E_244_273	chr3:78,946,908-78,947,539	CGTCAGTATTCCCAGCGAGT	GTCATGCCAGCATCATCATC	55

[a]ESR1 PCR consisted of 15 cyles at 58°C annealing temperature followed by 20 cycles at 55°C annealing temperature.