Distribution Agreement

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

Signature:

_____ _____

Emily F. Wissel Date

The Gut and Vaginal Microbial Communities during Pregnancy in Patients With and Without Urogenital Infections

By

Emily F. Wissel
Doctor of Philosophy

Nursing

_____

Vicki Hertzberg, Ph.D., FASA
Advisor

_____

Timothy Read, Ph.D.
Co-Chair

_____

Anne Lang Dunlop, MD/MPH
Committee Member

_____

Michelle Wright, Ph.D., RN, FAAN
Committee Member

Accepted:

_____

Kimberly Jacob Arriola, PhD
Dean of the James T. Laney School of Graduate Studies

_____

Date

The Gut and Vaginal Microbial Communities during Pregnancy in Patients With and Without Urogenital Infections

By

Emily F. Wissel
B.S., The University of Texas at Austin, 2017

Advisors:
Vicki Hertzberg, Ph.D., FASA
Timothy Read, Ph.D.

An abstract of
A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy in Nursing
2023

Abstract

The Gut and Vaginal Microbial Communities during Pregnancy in Patients With and Without Urogenital Infections
By Emily F. Wissel

Background: The gut and vaginal microbiome both change over the course of a pregnancy and have been associated with many pregnancy complications. Pregnant individuals are more susceptible to urinary tract infections (UTIs), bacterial vaginosis (BV), and chlamydia infection. It is currently unclear if there are differences in the microbiome or the collection of antimicrobial resistance (AMR) genes for pregnant individuals who develop urogenital infections versus those who don't.
Purpose: The dissertation aims to examine (1) how bacterial species change during pregnancy for those with and without urogenital infections, and (2) how AMR genes change during pregnancy after antibiotic treatment for urogenital infections.
Methods: A subset of the data from the Emory University African American Vaginal, Oral, and Gut Microbiome in Pregnancy Cohort Study were sent for metagenomic sequencing (238 patients, rectal and vaginal swabs at 8-14 weeks & 24 - 30 weeks pregnancy). Taxonomic assignment was done with the metaphlan2 software tool, and AMR genes were detected with the AMR Finder Plus tool. 16S rRNA data from the same samples had taxonomic assignment with the PECAN and DADA2 tools. Associations between the microbiome and urogenital infections were analyzed with a linear decomposition model. Differences in the frequency of AMR genes was analyzed with a chi-squared test for independence. hAMRoaster, a new bioinformatics tool, was created to compare the performance of different AMR gene processing pipelines.
Results: Collectively, this dissertation finds that the gut and vaginal microbial communities are not significantly impacted by urogenital infections or their treatment. Specific microbes and AMR genes tend to be increased in those who developed urogenital infections compared to those who did not, however, these differences do not persist for the entire pregnancy. These findings should reassure most patients that being diagnosed with a urogenital infection and receiving antibiotic therapy for that infection will not have a significant, detrimental impact on their microbiome overall during pregnancy.

The Gut and Vaginal Microbial Communities during Pregnancy in Patients With and Without Urogenital Infections

By

Emily F. Wissel
B.S., The University of Texas at Austin, 2017

Advisors:
Vicki Hertzberg, Ph.D., FASA
Timothy Read, Ph.D.

A dissertation submitted to the Faculty of the
James T. Laney School of Graduate Studies of Emory University
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
in Nursing
2023

Acknowledgements

I am deeply grateful to those who have supported me through my academic journey and the completion of this dissertation. First and foremost, I would like to express my sincere gratitude to my advisors Drs. Timothy Read and Vicki Hertzberg, for their invaluable guidance, support, and encouragement throughout this dissertation. I would also like to thank my committee members Drs. Anne Dunlop and Michelle Wright, for their time, encouragement, and helpful feedback throughout the entire process. The entire committee helped me to create a scientifically rigorous study that is presented in this dissertation. I am grateful to all members of the EMERGENT lab group, past and present, who have provided feedback and troubleshooting support throughout my time at Emory.

I am grateful to the School of Nursing at Emory for providing the resources and opportunities necessary for me to pursue my research interests. I would also like to thank all the faculty members who have supported my work and engaged in conversations about my research with me – Ken Hepburn, Mark Risjord, Sue Ishaq, Katherine Maki, Francisco Parada, Laura Tipton, and countless others. I would like to thank the undergraduate researchers who collaborated with me on this and other projects–Noriko Bracken, Myoungjin Woo, Kemuel Russell, Olive Tambou Nzoutchoum, and Taneecia Natarajan Thirulokachander.

I would like to extend my appreciation to my family and friends for their support and encouragement over the last four years. Finally, I would like to acknowledge all the people who have directly or indirectly contributed to this work. Their contributions, no matter how big or small, have been essential to the completion of this dissertation.

Thank you all for your support and encouragement.

Funding Acknowledgement

This dissertation is supported by the National Science Foundation Graduate Research Fellowship under grant 1937971. This work is also supported by XSEDE XRAS allocation MCB200145 and the Emory Reproductive Health Research in the Southeast (RISE) mini-grant.

Table of Contents

Chapter 1

Introduction

Microbes have been a part of humans' entire evolutionary history. People now rely on our microbes for normative, everyday functioning—immune system development, nutrient digestion, and even mental health. Of increasing interest is the role of the microbiome in pregnancy and birth outcomes. The body goes through many changes in order to support a developing fetus, and it is currently unclear to what extent the human microbiome is changing as a consequence of these broader, body wide changes, or helping to drive these changes. This dissertation explores this knowledge gap to try to understand how the gut and vaginal microbiome change during pregnancy, and how antibiotic administration in early pregnancy may shape this change. Further, I evaluate what type of clinically important information can come from different sequencing technologies with body site swabs and provide recommendations for healthcare providers interested in integrating the microbiome of their patients into patient care.

What is a microbiome?

The microbiome refers to the complete set of genes from microorganisms, bacteria, archaea, fungi, viruses, and other eukaryotic species, which live on and within a host (Shanahan, Ghosh, and O'Toole 2021). Importantly, microbes in the microbial community occupy ecological niches within their host and are sometimes referred to as "colonizers" or "persisters" because they occupy that niche over time (Tipton, Darcy, and Hynson 2019). This is in juxtaposition to a "tourist" in the microbiome–a microbe that passes through the body but does not occupy a niche within the host or persist once it travels through. An example of a persister would be *Prevotella* in the gut microbiome. It almost always colonizes the human gastrointestinal tract (Tett et al. 2021) and can persist in the face of challenge events like

psychosocial stress or antibiotic treatment (Raymond et al. 2016, Hantsoo and Zemel 2021). A tourist is more likely to be a passive environmental microbe, such as soil microbes that might momentarily occupy space on a hand that plants flowers in the garden (Vandegrift et al. 2019), or a probiotic species, as they often do not colonize in the gut (Sanders, Merenstein, Merrifield, and Hutkins 2018).

Some host-microbe associations are quite simple and specific, meaning that there may only be a few microbes that colonize and persist within the particular host, and the taxa that do persist are the same microbes across different hosts. For example, *Pantoa* genus of bacteria almost always mono-colonize *P. stali* stink bugs, regardless of the geographic location of the stink bug (Hosokawa et al. 2016). Such simplicity and specificity are not the case for human and most mammalian microbiomes (Benson 2016). The lack of specificity in the human microbiome has proven to be a challenge in health sciences research, as it means there is no "core" human microbiome or clear microbial markers for health and/or disease status (Caporaso et al. 2011). Interestingly, the human microbiome has a large degree of functional redundancy, or conserved genomic content, across the microbes in the microbiome. While the taxonomic composition of the microbiome varies widely, this functional redundancy helps the microbiome to remain functional similar over time and in the face of disturbance in the human host (Tian et al. 2020).

Lifestyle factors appear to play a much larger role in the composition of the human microbiome than host-specific microbes, making the human microbiome much more malleable than other hosts, such as the stink bug discussed above. Diet, medication (including antimicrobials), physical activity, and geographic location of residence can be some of the largest predictors of human microbiomes, but still explain only a small proportion of the observed taxonomic variance (Dixon et al. 2023), indicating that there is still much to be

discovered about how the human microbiome persists. These factors also predict a varying degree of variance across body sites. The gut microbiome, for example, is much more clearly linked to diet than the vaginal microbiome (Song et al. 2020, Graham et al. 2021, Dong and Gupta 2019). Research on the human microbiome tends to be broken up by body site, so these distinctions can be important. This dissertation focuses primarily on the gut and vaginal microbiome, though oral, skin, and many other organ-microbiome relationships are studied for their role in human health.

Gut Microbiome Sampling

The "gut microbiome" refers specifically to the microbes living within the small and large intestine of the gastrointestinal tract. It is primarily studied via DNA extracted from stool samples or from rectal swabs, though sometimes tissue biopsies or smart, consumable devices are available. Though stool samples are only a proxy for the gut microbiome, they are currently preferred due to the ease with which they can be collected, they are noninvasive (thereby posing less risk to the patient or participant), and there is lower concern for contamination compared to other methods. Rectal swabs are arguably easier or more convenient to collect as researchers do not need to wait for a bowel movement, however, they tend to have low biomass, be even more of a proxy than stool for community composition (Short et al. 2021) and have a higher risk of contamination from skin microbes than stool. More precise measures of the gut microbiome may include biopsies, though they are highly invasive, susceptible to contamination, often with low biomass, expensive, and not suitable when studying a healthy population. Recently, an ingestible sampling device has been developed for collecting fluid from the small intestines, but it is both expensive and poses its own set of challenges and potential patient or participant risks (Tang et al. 2020). These sampling methods and more have been discussed in depth elsewhere (Tang et al.

2018, Song et al. 2020, Short et al. 2021). For the sake of this dissertation, the differences between stool and rectal swabs will be considered.

Collecting stool samples from patients recruited in a clinic is not always feasible. It requires patients to go home, wait for a bowel movement, collect the stool properly, and send it back to the lab or be picked up by lab personnel or for a patient to time their clinical visit with occurrence of a bowel movement. Further, whole stool samples provide logistic challenges for labs. Collecting, processing, and storing whole stool for participants in a study with one hundred participants can quickly become challenging. A second approach is to collect a partial specimen from a stool sample by poking several spots of a stool sample with a sampling tool, putting the sampling tool in a test tube with a DNA stabilizer, and mailing it to the lab or having a researcher retrieve the sample from the patient's home. For some studies, directly swabbing the area of interest while the patient is in the clinic is the most realistic approach, particularly when patient loss to follow up is highly likely. While rectal swabs will not yield the same depth of information (as they inherently only sample rectal bacteria), plenty of studies have effectively used rectal swabs to find meaningful results (Smid et al. 2018, Schlebush et al. 2022, Goltsman et al. 2018).

Vaginal Microbiome Sampling

The vaginal microbiome refers specifically to the microbes that live within the vagina. The vaginal microbiome is different from the microbiome at other body sites in that there is typically much lower diversity in healthy people’s vaginas (Ravel et al. 2010). It should be noted that a significant portion of people across race and ethnicities have healthy, diverse vaginal microbiomes at baseline though (Ravel et al. 2010). Further, cross-kingdom interactions are much more understood within the vagina compared to the gut (Bradford and Ravel 2017; Gupta, Kakkar, and Bhushan 2019). Fungi provide bacteria with alpha amylase, an enzyme that bacteria

need to digest glycogen in the vaginal wall (Miller et al. 2016; Fuochi, Volti, and Furneri 2017; Nunn and Forney 2016). It is thought that there are micro-environments within the vagina (Bartlett et al. 1978, Subramaniam et al. 2016), so researchers must be cautious when discussing how and where samples are collected from. However, samples collected from distinct locations of the vagina do not always differ significantly (Huang et al. 2015), so it is unlikely that this significantly contributes to study variance. Sampling instrument also plays a role in DNA yield from samples, sometimes more so than sample site within the vagina (Virtanen et al. 2017). While there is some discussion between physician collected versus self-collected vaginal swabs, either method is suitable for sampling the vaginal microbiome (Virtanen et al. 2017, Huang et al. 2015).

Bioinformatic Approaches to Microbiome Data

In general, the human microbiome is studied via sequencing technologies, which have advanced the field dramatically as many human-associated microbes are difficult to culture. Short read Illumina sequencing technologies are currently the most common data type for human microbiome studies. Short read sequencing can refer to either 16S rRNA sequencing or shotgun metagenomic sequencing.

16S rRNA Gene Sequencing

16S rRNA gene sequencing, 16S for short, is a short amplicon method which specifically targets the 16S region of the bacterial genome, which is highly conserved across bacterial species. As such, only bacteria can be detected with 16S, typically at the genus level. There are nine hypervariable regions along the 16S segment of the genome, and researchers typically select one or two sections for sequencing, such as the V4 or V3V4 region. Historically, researchers define an operational taxonomic unit, or OTU, by somewhere between 97 and 99% sequence

similarity in 16S data. Cluster methods with tools such as mothur (Schloss 2009) or marker-gene matching in QIIME2 (Bolyen et al. 2019) are used for OTU assignment. As technology and bioinformatic methods have advanced, researchers examine amplicon sequence variants (ASVs), or exact sequences that occur within the 16S data, and count how often the ASV occurs with a correction for sequence errors. Denoising methods from tools such as DADA2 can detect ASVs (Callahan et al. 2016). ASVs are currently regarded as the current "state of the art" method for examining 16S data as it has intrinsic biological meaning, unlike the historical OTU clusters, and better controls for error (Callahan et al. 2017).

Shotgun Metagenome Sequencing

Shotgun metagenome data, often referred to as metagenomics, is an untargeted sequencing approach. All DNA in a sample is sequenced, including host, fungal, protozoal, viral, and bacterial. As such, filtering out host reads from human metagenomics becomes an important quality control step that is avoided with 16S. Host reads can account for anywhere between 1% and 99% of the sequenced reads, depending on the human body site sampled (Pereira-Marques et al. 2019). However, metagenomics allows researchers to examine which microbes are present down to the strain level, as sequencing all the DNA allows one to see all genes and single nucleotide polymorphisms (SNPs) between microbes, allowing for more precise taxonomic and functional analysis.

After human reads are removed, metagenomic data can be processed by two different methods. Reads can be mapped to a reference database, such as with bioBakery (Beghini et al. 2021) or kraken2 (Wood, Lu, and Langmead 2019), or sequences can be assembled into metagenome assembled genomes, or MAGs, such as with metaSPAdes (Nurk et al. 2017) or megaHIT (Li et al. 2015). Assembly-based methods are useful for discovering taxa present in a

sample that may not be described in current databases. This is particularly useful for studies examining environmental microbes, which are often undescribed. However, assembly-based methods can face challenges with uneven sequencing coverage, as is common with metagenomic data, and may be more readily applied to long-read sequencing methods such as PacBio (Bharti and Grimm 2021).

Read mapping methods typically result in a shorter compute processing time as they have relatively smaller databases. These databases are typically sufficient for human microbiome studies, though they are biased towards bacteria and often exclude many fungi or other eukaryotic microbes (Jin et al. 2022, Weissman et al. 2021). Of note is that most of the variation between read mapping bioinformatic tools is due to differences in databases (Balvočiūtė and Huson 2017). However, processing metagenomic data still takes significantly more computational resources than 16S sequencing, as the data size per sample goes from megabytes with 16S to gigabytes with metagenomics. It also requires more bioinformatics training, as the metagenomic bioinformatic processes require knowledge of command line interfaces, whereas 16S processing tools like QIIME2 have graphical user interfaces.

In addition to examining the taxonomy of the microbiome (or examining “who's there?”), microbiome research is also expanding into functional analyses (or “what are they doing?”). There are arguments that favor a functional view of the microbiome, as it may provide a more accurate analytical lens by inherently viewing the microbiome as part of a holobiont and shifting away from trying to single out individual, significant taxa. A holobiont is the idea that the way an individual is conceptualized should be human plus microbes, that individuals are constituted by these multispecies interactions (Bordenstein and Theis 2015; Gilbert, Sapp, and Tauber 2012; Nagpal and Cryan 2021). A functional description of the microbiome may be processed with

tools like HUMAnN (Behini et al. 2021) or MG-RAST (Keegan, Glass, and Meyer 2016) and describes the metabolic potential of the microbes present in a sample based on their genes.

The Microbiome in Health

Both the gut and vaginal microbiome are key to health and development across the lifespan. Of particular importance is the microbiome during pregnancy and birth, as it will impact the health of both the pregnant person and the developing fetus. This section will very briefly overview what is known of the microbiome in human health broadly and specifically in the context of pregnancy.

The Gut Microbiome

The human gut microbiome is responsible for maintaining many aspects of health, including food digestion (Warren et al. 2018), immune system priming (Hitch et al. 2022, Renz et al. 2018), and neurometabolite production (Morris et al. 2017, Maqsood and Stone 2016). There is evidence that the gut microbiome is especially important for mental health and well-being (Smith and Wissel 2019). Some literature describes the gut microbiome as a functional organ as the human body has evolved to delegate certain tasks to, such as those mentioned above.

In pregnancy, these functions of the gut microbiome become especially important. The body goes through natural changes as it prepares to carry the developing fetus to term. There are predictable and necessary immunologic, metabolic, and hormonal changes that allow a pregnant individual to provide nutrients and space for the fetus without mounting an immune response against it (Edwards et al. 2017). The microbiome shifts during pregnancy alongside these changes (Prince et al. 2015, Neuman and Koren 2017), and potentially influences pregnancy complications. Gestational diabetes, hypertension, and gestational weight gain are associated

with distinct changes in the gut microbiome during pregnancy compared to pregnant individuals who carry to term without these complications (Vinturache et al. 2016, DiGuilio et al. 2015). However, there is not a consistently detected trend in how the microbiome changes during pregnancy and health status. It is likely that placing the microbiome in the context of an individual's lifestyle is essential to understanding how and why it changes (e.g., dietary options/choices, daily stress, prenatal vitamins).

The gut microbiome during pregnancy shifts to support the increased metabolic needs during this time. For example, there is a decrease in overall diversity of the gut microbiome and a shift towards carbohydrate digestion (Gosalbes et al. 2019), consistent with increased blood glucose levels in late pregnancy. Research from Pharye's leaf monkeys suggests these shifts in the microbiome are due to reproductive hormones (Mallott et al. 2020). In a small study, pregnant people with a higher proportion of the bacterial phylum Bacteroidetes (compared to Bacillota, formerly known as Firmicutes) have higher weight gain (Aatsinki et al. 2018, Smid et al. 2018), which is on par with trends in nonpregnant people (Muscogiuri et al. 2019). However, there is much more nuance to weight gain and obesity, especially during pregnancy. The proportion of Bacteroidetes to Firmicutes, which was widely reported in early human microbiome studies such as the Human Microbiome Project (Huse et al. 2012), is not sufficient to understand health status. Similarly, while the presence of certain gut microbes during pregnancy are implicated in the development of allergy and asthma in an infant (Gao et al. 2021), this research is still developing and is not conclusive at this time. A major goal of this dissertation is to better understand how the microbiome is shifting during pregnancy, both for those with urogenital infection (UTI, BV, chlamydia) and those without these infections.

The Vaginal Microbiome

The vaginal microbiome is particularly important for vaginal health. Unlike other body sites, the vaginal microbiome is typically characterized by its simplicity and has extremely low diversity. However, a significant proportion of healthy vaginal microbiomes are relatively diverse, and it is not currently clear what drives the difference between single-species dominant healthy communities and diverse healthy communities. Importantly, a diverse vaginal microbiome is not inherently an indicator of disease or health status, contrary to what others suggest (Gupta, Singh, and Goyal 2020).

The inter-kingdom interactions are also better understood in the vaginal microbiome compared to the gut microbiome. Fungi and bacteria rely on each other in this environment and can control the population of one another. The primary food source for bacteria in the vaginal microbiome is glycogen, however, most bacteria need an enzyme produced by fungi called alpha-amylase to be able to digest the glycogen (Miller et al. 2016). If the balance between these two kingdoms swings towards fungi, then yeast infections may develop. If it favors bacteria, then conditions such as bacterial vaginosis may develop. Psychological stress, which can impact glycogen and inflammation availability in the vaginal microbiome, can impact this balance (Witkin and Linhares 2016). What is most often observed is that groups who experience the most stress, either acute psychosocial stress or broader, chronic experiences of stress from structural racism and other forms of oppression, are more likely to have a vaginal microbiome that is diverse at baseline compared to other groups who are less likely to experience chronic oppression and stress (De Wolfe et al. 2021, Benezra 2020, Amato et al. 2021). This likely reflects a functional shift in the vaginal microbiome away from glycogen reliance (towards perhaps mucin degradation), as glycogen becomes depleted when stress is high, though

longitudinal studies will need to confirm this theory. While pregnancy is often considered a physiologic stress test, estrogen levels steadily rise during pregnancy. Higher estrogen levels are often associated with more glycogen in the vagina, so it is unclear if stress during pregnancy will have the same effect as stress when one is not pregnant.

During pregnancy, the vaginal microbiome consistently becomes less diverse as a pregnancy progresses regardless of the presence of pregnancy complications (Freitas et al. 2017). Susceptibility to UTIs and BV is increased during pregnancy (Elkady, Sinha, and Hassan 2019). Changes in the vaginal microbiome that occur during pregnancy could play a role in this increased susceptibility. Specifically, decreased diversity could reflect decreased pathogens through competitive exclusion.

The vaginal microbiome during pregnancy is also thought to be the vaginally born newborn's first exposure to microbes (Singh and Mittal 2020, Kennedy et al. 2023). The microbes seed the newborn with the microbiome of the birthing parent, either vaginal or skin microbes depending on mode of delivery. Interactions between bacteria are important in early life microbiome development (Rao et al. 2020), and infants follow a particular trajectory for gut microbiome development (Moore and Townsend 2019; Enav, Bäckhed, and Ley 2022). Infant gut microbiomes reflect adult microbiomes as they grow up while still being socially considered children (Azad et al. 2013). The impact of birth mode (c-section or vaginal birth) on lifelong health outcomes, like allergy or obesity, is still an open question (Bokulich et al. 2017).

Antibiotics in Pregnancy

Antibiotics also influence the microbial communities during pregnancy. In the United States, approximately one in four individuals are prescribed antibiotics during their pregnancy (such as to treat STIs, UTIs, etc.) (Bookstaver et al. 2015) and individuals undergoing C-section

or who have *Streptococcus agalactiae,* or group b strep infection, receive a continual dose of antibiotics during delivery to reduce risk of infection (Seedat et al. 2017). Administering antibiotics is often an essential life-saving therapy, particularly when treating an infection. However, antibiotic administration increases the pool of antibiotic resistant pathogens (Ma, Forney, and Ravel 2012). In nonpregnant populations, antimicrobial therapies are known to increase the repertoire of AMR genes which increases the risk for drug resistant infections and increase the likelihood of subsequent obesity (Cox and Blaser 2015), though the percent variance explained from these studies is low. Research in rodents suggests that antibiotics either prenatally or right after birth will have a negative effect on long term health outcomes (Cox and Blaser 2015). Antibiotics given to agricultural animals play a role in human AMR genes, and may impact the microbiome as well (Smith 2020). Some work in humans has found that antibiotics in pregnancy can impact the progression of an infant's microbiome (Lemas et al. 2016), however, these trends do not typically persist across the lifespan. It is important to study how antimicrobial therapy impacts AMR genes during pregnancy because antimicrobials may negatively impact the changing microbiome in a manner which limits the microbiome's ability to support a healthy pregnancy or play a role in a newborn's microbiome progression or risk of a maternal or neonatal AMR infection.

Outline of the Dissertation

As described above, there are definite changes in the microbiome during pregnancy that seem to be important for maintaining typical, healthy pregnancies. There are typically differences in the microbiome of those who have healthy pregnancies and those who have pregnancy complications, however, the differences between these groups are not always

consistent. The first aim of this dissertation aims to dig deeper at this point in a large cohort of pregnant people.

Methodology

The Emory University African American Pregnancy Cohort was used for a secondary data analysis for this dissertation (Corwin et al. 2017). The Emory University African American Pregnancy Cohort followed a cohort of African American people throughout their pregnancy (T1 = 8-14 weeks gestation; T2 = 24-30 weeks gestation). Pregnant people provided rectal and vaginal swabs and consent to a medical chart review. From a subset of 238 enrolled individuals, DNA extracted from vaginal and rectal swabs underwent both 16S and shotgun metagenomic sequencing, providing hundreds of metagenome samples for analysis and the unique chance to examine how many host medical factors, like antibiotics, shift alongside changes in the microbiome. **Figure One** depicts these methods.

Aim 1: Examine how bacterial species change in the microbiome during pregnancy in African American patients with and without urogenital infections.

H 1.1: The gut and vaginal microbiome will become less diverse as a pregnancy progresses.

H 1.2: There will be taxonomic differences in the vaginal microbiome of those who are diagnosed with Chlamydia, BV, or a UTI during pregnancy

H1.3: Replicating the LDM analysis in 16S rRNA data will return similar results.

Approach To identify which microbes are present in the metagenomic samples, the bioBakery pipeline will be used to remove human reads, complete quality filtering (kneaddata), assign taxonomy (metaphlan2), and describe the functional capabilities of the microbiome (humann2). This pipeline is widely used by researchers in the field, allowing for comparison of these results to other comparable studies. bioBakery is actively maintained, allowing us to ensure

an accurate up-to-date pipeline will be used for this project. Additionally, the bioBakery pipeline allows for creation of taxonomic and functional profiles for each sample, identifying features detected and genes/metabolic pathways that are present. Data on urogenital infection is collected from medical chart abstraction and is defined as any chlamydia, UTI, or BV diagnosis during their pregnancy.

Data Analysis To test both the global effect of urogenital infection on the microbiome and the association between individual taxa and urogenital infections, the R package LDM will be used to implement a linear decomposition model (Zhu et al. 2021). LDM is beneficial because it can test the global and individual associations in a unified approach while adjusting for the false discovery rate, controlling for continuous and discrete confounding variables, and it can handle samples which are not independent of each other (as is the case with multiple timepoints and body sites per patient included in the study). LDM can also account for multiple samples coming from a single individual with the parameter "cluster.id" corresponding to the subject ID. The LDM analysis will be replicated in the 16S rRNA gene samples that exist for the same data (without exploring the 16S rRNA gene data outside this scope)

Potential Outcomes and Interpretations This analysis is expected to reveal a decrease in diversity as pregnancy progresses for both the gut and vaginal microbiome. Importantly, in the event of a change in diversity, this analysis will reveal the taxa that are changing over the course of pregnancy and in response to urogenital infection during pregnancy. The change in diversity may be impacted by the initial community, especially within the vaginal microbiome where *Lactobacillus* is typically either in high abundances or nearly absent. An analysis of the functional profiles can elucidate if differences in taxonomy are due to functional differences in the microbiome, which is expected as *L. crispatus* dominated vaginal communities produce more

lactic acid than *L. iners* dominated or diverse communities, which produce more acetic acid. The gut microbiome is expected to become less diverse as a pregnancy progresses. Further, in people who experience a urogenital infection during pregnancy, the vaginal microbiome may be more diverse than those without a urogenital infection during pregnancy due to inflammatory response to infections, disturbance to the baseline microbiome when infection is present, and antibiotics used to treat the infection.

Possible Pitfalls and Alternative Approaches A potential limitation of this work is that the individuals who experience one of the three urogenital infections of interest (chlamydia, UTI, or BV) may be too heterogeneous to make inferences about changes in the gut and vaginal microbiome during pregnancy beyond broad generalizations. Heterogeneity in birth outcomes and pregnancy complications may make it difficult to draw conclusions about the shifting microbiome in pregnancy. If the group is too heterogeneous, more variables can be introduced as confounding variables to control for more of the variance in the data. Alternatively, broader grouping variables can be assigned (e.g., any antimicrobial exposure versus no exposure to antimicrobials).

Aim 2: Examine how exposure to antimicrobial therapies change the microbiome over time.

H 2.1: Those who receive antibiotics for urogenital infection (Chlamydia, UTI, and BV) will have a more diverse vaginal microbial community due to antibiotic exposure and less diverse gut microbial community.

H 2.2: Those exposed to antibiotic therapy for these three urogenital infections will have more AMR genes in their gut and vaginal microbiome compared to those not exposed to antimicrobial therapy and those exposed to antimicrobial therapies for other reasons.

Approach Antimicrobials are commonly administered to treat bacterial vaginosis (BV), urinary tract infection (UTI), and Chlamydia infection in this sample (n = 57, 46, 30 respectively). Understanding the influence of antimicrobials on AMR genes is important because antimicrobials are a common therapy during pregnancy and birth, and AMR genes can increase the risk for drug-resistant infection. Particular attention will be paid to antibiotics nitrofurantoin, metronidazole, and Zithromax administered to treat the three most common infections in this cohort - UTIs, BV, and Chlamydia infection.

Data Analysis The repertoire of AMR genes will be assessed for all participants (those exposed to antimicrobial therapy and those not). First, a benchmarking analysis will compare computational tools for identifying AMR genes in metagenomic data - ShortBRED, fARGene, RGI, ResFinder, abricate, AMRFinder Plus, deepARG, sraX, and starAMR. Once these tools are systematically assessed, one will be selected that is the best fit for this data and resources. This tool will be used to characterize the AMR genes in the whole dataset. Then, the catalog of AMR genes will be compared between those who receive antibiotics for the urogenital infections of interest (chlamydia, UTI, and BV), those who receive antimicrobials for other infections, and those who report no antimicrobial exposure during their pregnancy. To test if there is a significant difference between the number and type of AMR genes present in those who do and do not experience each of the three urogenital infections and its treatment during pregnancy, a Chi squared test of independence will be used (Chi squared test for types of AMR genes). To test if there is a significant difference in the microbiome between groups, alpha diversity will be calculated with R package vegan (Dixon 2003) and an ANOVA will be used to test for significant differences between groups.

Potential Outcomes and Interpretations Nitrofurantoin, Metronidazole, and zithromax are administered to treat urogenital infections (UTI, BV, and chlamydia) in this cohort. Nitrofurantoin is thought to only affect the urinary tract and therefore be "safer" to use in pregnant patients. It is expected that those exposed to only this antimicrobial therapy will not have significant changes in the AMR genes in their gut microbiome. However, orally administered metronidazole and zithromax are suspected to impact AMR genes in both the gut and vaginal microbiome, despite the pathogen only residing within the vaginal microbiome. It is expected that the gut microbiome of those who receive any antimicrobials will be less diverse and the vaginal microbiome will be more diverse compared to those who do not receive any antimicrobial therapies, as indicated by Shannon's alpha diversity metric.

Possible Pitfalls and Alternative Approaches One potential pitfall is that there may not be enough individuals exposed to antibiotics to make statistically significant inferences about any antimicrobial therapy and its effect on the gut microbiome. Due to this limitation, special attention is paid to three clinically important urogenital infections and their treatment, which are common in this cohort and will hopefully ensure sufficient data for statistically significant results. It will be important to document what differences do exist - in AMR genes, community composition, and medical factors - for those exposed to any antimicrobial and for those in the targeted antibiotic analysis. It is also possible that the baseline microbiome of those requiring antimicrobials to treat UTIs, STIs, and BV is different from the baseline microbiome than those who do not develop infections as the pathogen may be present in the baseline community or make the baseline community more susceptible to pathogen invasion. To test for this, alpha diversity will be compared across groups at the first sampling time point.

Significance

This study contains a large cohort of pregnant individuals who were followed longitudinally (n = 238) and are typically excluded in microbiome research. The resulting dataset contains over 1,000 metagenomic sequences for analysis, a full medical chart abstraction, and in-depth data from psychometric questionnaires. Analysis of these data will expand our understanding of the healthy microbiome and how it changes during pregnancy.

Further, as metagenomic sequencing is used for this data, it can be determined what strains of microbes are present and what they are doing (through metabolic pathway analysis). As these samples were also sequenced by 16S, information from 16S versus metagenome sequencing can be compared to determine which may be better suited for certain types of clinical use.

As antibiotics are routinely administered during pregnancy and birth, it is critical to understand their role in longitudinal microbiome changes. Antibiotics are a lifesaving therapy, but it is currently unexplored how AMR genes shift during pregnancy in people, and whether this has broader health implications.

Theoretical Model

The conceptual framework for this dissertation is presented in Figure Two. Briefly, microbes interact with each other within the body site they occupy and across body sites (i.e., gut and vaginal microbiome). Aim 1 of this study will examine the microbial communities. Antimicrobials such as antibiotics impact these microbes across body sites, a phenomenon Aim 2 will examine.

Summary

The microbiome shifts during pregnancy alongside expected hormonal, immune, and metabolic changes, but the direction of causal relationships between microbial community structure and broader, body-wide changes in the host is unclear. Understanding of changes in microbial community structure during pregnancy in the absence of complications is necessary to understand causal relationships between microbial communities and complications and to design necessary interventions.

To better understand how the microbiome changes during pregnancy, data from the Emory University African American Pregnancy Cohort were analyzed to examine **(1)** how microbial strains interact both within and across body sites over time to support healthy pregnancies, and **(2)** how antimicrobial therapy impacts these interactions during pregnancy. Results from this dissertation are expected to help further understanding how the microbiome supports healthy pregnancies and provide a pathway for future analysis of the microbiome in pregnancy complications.

Figure One

A visual description of the data collection. Pregnant individuals in the Emory University African American Pregnancy Cohort were enrolled at 8-14 weeks gestation and provided rectal and vaginal swabs, along with medical chart abstraction that provided information from their pregnancy and delivery. There were 238 pregnancy individuals who provided the same swabs and surveys at a second time point in pregnancy (24 - 30 weeks gestation), then consented to a medical chart review to see birth outcomes.

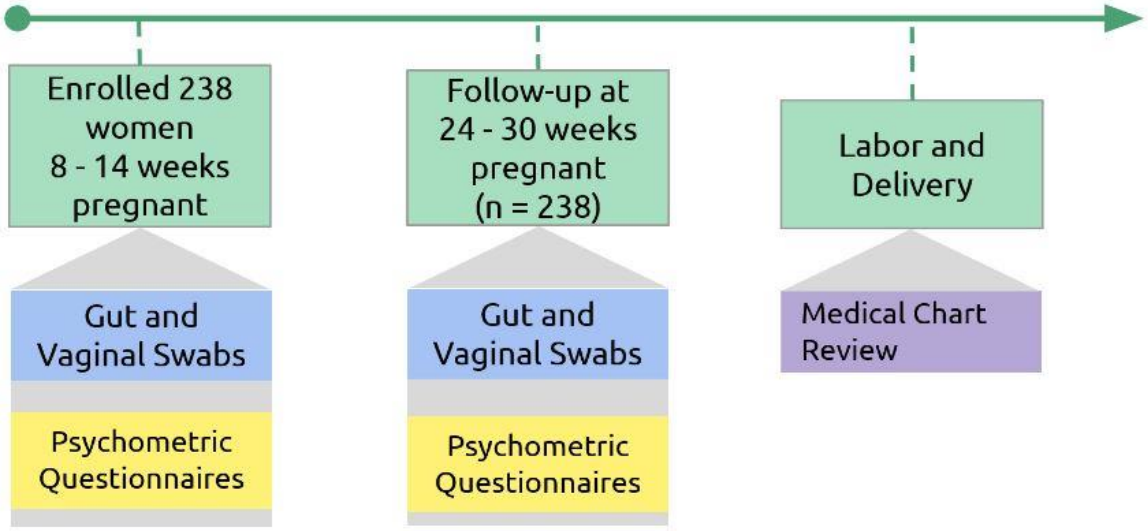
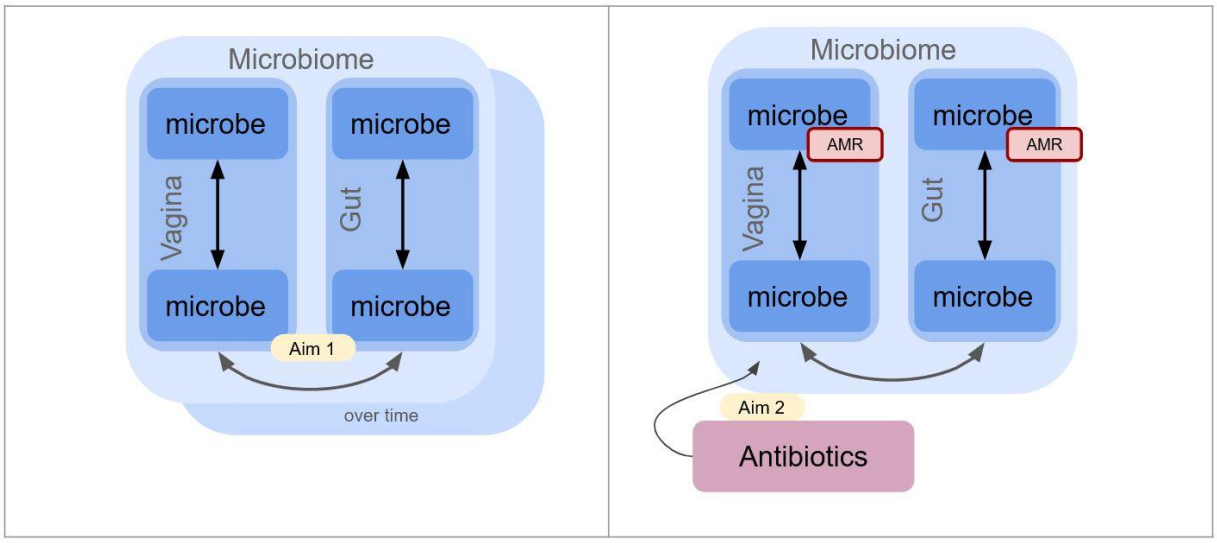


Figure Two

Conceptual Framework, or Theoretical Model, for the phenomenon this dissertation aims to explore. Aim One will explore how microbes interact both within and between body sites. Aim two will explore how antimicrobials for the three most common infections in this cohort, Bacterial Vaginosis (BV), Urinary Tract Infection (UTI), and Chlamydia infection, impact these relationships.



References

Aatsinki, Anna-Katariina, Henna-Maria Uusitupa, Eveliina Munukka, Henri Pesonen, Anniina Rintala, Sami Pietilä, Leo Lahti, Erkki Eerola, Linnea Karlsson, and Hasse Karlsson. 2018. "Gut Microbiota Composition in Mid-Pregnancy Is Associated with Gestational Weight Gain but Not Prepregnancy Body Mass Index." *Journal of Women's Health* 27, no. 10 (October): 1293–1301. https://doi.org/10.1089/jwh.2017.6488.

Amato, Katherine R., Marie-Claire Arrieta, Meghan B. Azad, Michael T. Bailey, Josiane L. Broussard, Carlijn E. Bruggeling, Erika C. Claud, et al. 2021. "The Human Gut Microbiome and Health Inequities." *Proceedings of the National Academy of Sciences* 118, no. 25 (June): e2017947118. https://doi.org/10.1073/pnas.2017947118.

Azad, Meghan B., Theodore Konya, Heather Maughan, David S. Guttman, Catherine J. Field, Radha S. Chari, Malcolm R. Sears, Allan B. Becker, James A. Scott, and Anita L. Kozyrskyj. 2013. "Gut Microbiota of Healthy Canadian Infants: Profiles by Mode of Delivery and Infant Diet at 4 Months." *CMAJ : Canadian Medical Association Journal* 185, no. 5 (March): 385–94. https://doi.org/10.1503/cmaj.121189.

Balvočiūtė, Monika and Daniel H. Huson. 2017. "SILVA, RDP, Greengenes, NCBI and OTT — How Do These Taxonomies Compare?" *BMC Genomics* 18, no. 2 (March): 114. https://doi.org/10.1186/s12864-017-3501-4.

Bartlett, John G., Nancy E. Moon, Paul R. Goldstein, Barbara Goren, Andrew B. Onderdonk, and B.Frank Polk. 1978. "Cervical and vaginal bacterial flora: Ecologic niches in the female lower genital tract." *American Journal of Obstetrics and Gynecology* 130, no. 6 (March): 658-661. https://doi.org/10.1016/0002-9378(78)90323-X.

Beghini, Francesco, Lauren J McIver, Aitor Blanco-Míguez, Leonard Dubois, Francesco Asnicar, Sagun Maharjan, Ana Mailyan, et al. 2021. "Integrating Taxonomic, Functional, and Strain-Level Profiling of Diverse Microbial Communities with BioBakery 3." *ELife* 10 (May): e65088. https://doi.org/10.7554/eLife.65088.

Benezra, Amber. 2020. "Race in the Microbiome." *Science, Technology, & Human Values* 45, no. 5 (September): 877–902. https://doi.org/10.1177/0162243920911998.

Benson, Andrew K. 2016. "The Gut Microbiome—an Emerging Complex Trait." *Nature Genetics* 48, no. 11 (November): 1301–2. https://doi.org/10.1038/ng.3707.

Bharti, Richa and Dominik G Grimm. 2021. "Current Challenges and Best-Practice Protocols for Microbiome Analysis." *Briefings in Bioinformatics* 22, no. 1 (January): 178–93. https://doi.org/10.1093/bib/bbz155.

Bokulich, Nicholas A., Jennifer Chung, Thomas Battaglia, Nora Henderson, Melanie Jay, Huilin Li, Arnon Lieber, et al. 2017. "Antibiotics, Birth Mode, and Diet Shape Microbiome

Maturation during Early Life." *Science Translational Medicine* 8, no. 343 (June): 343ra82. https://doi.org/10.1126/scitranslmed.aad7121.

Bolyen, Evan, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, et al. 2019. "Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2." *Nature Biotechnology* 37, no. 8 (August): 852–57. https://doi.org/10.1038/s41587-019-0209-9.

Bookstaver, P. Brandon, Christopher M. Bland, Brooke Griffin, Kayla R. Stover, Lea S. Eiland, and Milena McLaughlin. 2015. "A Review of Antibiotic Use in Pregnancy." *Pharmacotherapy: The Journal of Human Pharmacology and Drug Therapy* 35, no. 11 (November): 1052–62. https://doi.org/10.1002/phar.1649.

Bordenstein, Seth R., and Kevin R. Theis. 2015. "Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes." *PLOS Biology* 13, no. 8 (August): e1002226. https://doi.org/10.1371/journal.pbio.1002226.

Bradford, L. Latéy and Jacques Ravel. 2017. "The Vaginal Mycobiome: A Contemporary Perspective on Fungi in Women's Health and Diseases." *Virulence* 8, no. 3 (April): 342–51. https://doi.org/10.1080/21505594.2016.1237332.

Callahan, Benjamin J., Paul J. McMurdie, and Susan P. Holmes. 2017. "Exact Sequence Variants Should Replace Operational Taxonomic Units in Marker-Gene Data Analysis." *The ISME Journal* 11, no. 12 (December): 2639–43. https://doi.org/10.1038/ismej.2017.119.

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. "DADA2: High-resolution sample inference from Illumina amplicon data." *Nature Methods* 13 (May): 581-583. https://doi.org/10.1038/nmeth.3869.

Caporaso, J. Gregory, Christian L. Lauber, Elizabeth K. Costello, Donna Berg-Lyons, Antonio Gonzalez, Jesse Stombaugh, Dan Knights, et al. 2011. "Moving Pictures of the Human Microbiome." *Genome Biology* 12, no. 5 (May): R50. https://doi.org/10.1186/gb-2011-12-5-r50.

Corwin, Elizabeth J., Carol J. Hogue, Bradley Pearce, Cherie C. Hill, Timothy D. Read, Jennifer Mulle, and Anne L. Dunlop. 2017. "Protocol for the Emory University African American Vaginal, Oral, and Gut Microbiome in Pregnancy Cohort Study." *BMC Pregnancy and Childbirth* 17, no. 1 (June): 161. https://doi.org/10.1186/s12884-017-1357-x.

Cox, Laura M. and Martin J. Blaser. 2015. "Antibiotics in Early Life and Obesity." *Nature Reviews Endocrinology* 11, no. 3 (March): 182–90. https://doi.org/10.1038/nrendo.2014.210.

De Wolfe, Travis J., Mohammed Rafi Arefin, Amber Benezra, and María Rebolleda Gómez. 2021. "Chasing Ghosts: Race, Racism, and the Future of Microbiome Research." Edited

by Kathryn C. Milligan-Myhre. *MSystems* 6, no. 5 (October): e00604-21. https://doi.org/10.1128/mSystems.00604-21.

DiGiulio, Daniel B., Benjamin J. Callahan, Paul J. McMurdie, Elizabeth K. Costello, Deirdre J. Lyell, Anna Robaczewska, Christine L. Sun, et al. 2015. "Temporal and Spatial Variation of the Human Microbiota during Pregnancy." *Proceedings of the National Academy of Sciences* 112, no. 35 (September): 11060–65. https://doi.org/10.1073/pnas.1502875112.

Dixon, Meredith, Anne L. Dunlop, Elizabeth J. Corwin, and Michael R. Kramer. 2023. "Joint Effects of Individual Socioeconomic Status and Residential Neighborhood Context on Vaginal Microbiome Composition." *Frontiers in Public Health* 11. https://doi.org/10.3389/fpubh.2023.1029741.

Dixon, Philip. 2003. "VEGAN, a Package of R Functions for Community Ecology." *Journal of Vegetation Science* 14, no. 6 (December): 927–30. https://doi.org/10.1111/j.1654-1103.2003.tb02228.x.

Dong, Tien S. and Arpana Gupta. 2019. "Influence of Early Life, Diet, and the Environment on the Microbiome." *Clinical Gastroenterology and Hepatology: The Official Clinical Practice Journal of the American Gastroenterological Association* 17, no. 2 (January): 231–42. https://doi.org/10.1016/j.cgh.2018.08.067.

Edwards, Sara M., Solveig A. Cunningham, Anne L. Dunlop, and Elizabeth J. Corwin. 2017. "The Maternal Gut Microbiome during Pregnancy." *MCN. The American Journal of Maternal Child Nursing* 42, no. 6 (November): 310–17. https://doi.org/10.1097/NMC.0000000000000372.

Elkady, Adel, Prabha Sinha, and Soad Ali Zaki Hassan. 2019. *Infections in Pregnancy: An Evidence-Based Approach*. Cambridge: Cambridge University Press.

Enav, Hagay, Fredrik Bäckhed, and Ruth E. Ley. 2022. "The Developing Infant Gut Microbiome: A Strain-Level View." *Cell Host & Microbe* 30, no. 5 (May): 627–38. https://doi.org/10.1016/j.chom.2022.04.009.

Freitas, Aline C., Bonnie Chaban, Alan Bocking, Maria Rocco, Siwen Yang, Janet E. Hill, Deborah M. Money, et al. 2017. "The Vaginal Microbiome of Pregnant Women Is Less Rich and Diverse, with Lower Prevalence of Mollicutes, Compared to Non-Pregnant Women." *Scientific Reports* 7, no. 1 (August): 9212. https://doi.org/10.1038/s41598-017-07790-9.

Fuochi, Virginia, Giovanni Li Volti, and Pio Maria Furneri. 2017. "Commentary: Lactobacilli Dominance and Vaginal PH: Why Is the Human Vaginal Microbiome Unique?" *Frontiers in Microbiology* 8 (September). https://www.frontiersin.org/articles/10.3389/fmicb.2017.01815.

Gao, Yuan, Ralph Nanan, Laurence Macia, Jian Tan, Luba Sominsky, Thomas P. Quinn, Martin O'Hely, et al. 2021. "The Maternal Gut Microbiome during Pregnancy and Offspring Allergy and Asthma." *Journal of Allergy and Clinical Immunology* 148, no. 3 (September): 669–78. https://doi.org/10.1016/j.jaci.2021.07.011.

Gilbert, Scott F., Jan Sapp, and Alfred I. Tauber. 2012. "A Symbiotic View of Life: We Have Never Been Individuals." *The Quarterly Review of Biology* 87, no. 4 (December): 325–41. https://doi.org/10.1086/668166.

Gosalbes, María José, Joan Compte, Silvia Moriano-Gutierrez, Yvonne Vallès, Nuria Jiménez-Hernández, Xavier Pons, Alejandro Artacho, and M. Pilar Francino. 2019. "Metabolic Adaptation in the Human Gut Microbiota during Pregnancy and the First Year of Life." *EBioMedicine* 39 (January): 497–509. https://doi.org/10.1016/j.ebiom.2018.10.071.

Goltsman, Daniela S. Aliaga, Christine L. Sun, Diana M. Proctor, Daniel B. DiGiulio, Anna Robaczewska, Brian C. Thomas, Gary M. Shaw, David K. Stevenson, Susan P. Holmes, Jillian F. Banfield et al. 2018. "Metagenomic analysis with strain-level resolution reveals fine-scale variation in the human pregnancy microbiome." *Genome Research* 28 (September): 1467-1480. doi:10.1101/gr.236000.118.

Graham, Madeline E., William G. Herbert, Stephanie D. Song, Harshini N. Raman, Jade E. Zhu, Paulina E. Gonzalez, Marina R. S. Walther-António, and Marc J. Tetel. 2021. "Gut and Vaginal Microbiomes on Steroids: Implications for Women's Health." *Trends in Endocrinology & Metabolism* 32, no. 8 (August): 554–65. https://doi.org/10.1016/j.tem.2021.04.014.

Gupta, Parakriti, Mini P. Singh, and Kapil Goyal. 2020. "Diversity of Vaginal Microbiome in Pregnancy: Deciphering the Obscurity." *Frontiers in Public Health* 8 (July). https://doi.org/10.3389/fpubh.2020.00326.

Gupta, Shagun, Vipan Kakkar, and Indu Bhushan. 2019. "Crosstalk between Vaginal Microbiome and Female Health: A Review." *Microbial Pathogenesis* 136 (November): 103696. https://doi.org/10.1016/j.micpath.2019.103696.

Hantsoo, Liisa and Babette S. Zemel. 2021. "Stress Gets into the Belly: Early Life Stress and the Gut Microbiome." *Behavioural Brain Research* 414 (September): 113474. https://doi.org/10.1016/j.bbr.2021.113474.

Hitch, Thomas C. A., Lindsay J. Hall, Sarah Kate Walsh, Gabriel E. Leventhal, Emma Slack, Tomas de Wouters, Jens Walter, and Thomas Clavel. 2022. "Microbiome-based interventions to modulate gut ecology and the immune system." *Mucosal Immunology* 15 (September): 1095-1113. https://doi-org.proxy.library.emory.edu/10.1038/s41385-022-00564-1.

Hosokawa, Takahiro, Yoshiko Ishii, Naruo Nikoh, Manabu Fujie, Nori Satoh, and Takema Fukatsu. 2016. "Obligate Bacterial Mutualists Evolving from Environmental Bacteria in

Natural Insect Populations." *Nature Microbiology* 1, no. 1 (January): 1–7. https://doi.org/10.1038/nmicrobiol.2015.11.

Huang, Yi-E, Yan Wang, Yan He, Yong Ji, Li-Ping Wang, Hua-Fang Sheng, Min Zhang, et al. 2015. "Homogeneity of the Vaginal Microbiome at the Cervix, Posterior Fornix, and Vaginal Canal in Pregnant Chinese Women." *Microbial Ecology* 69, no. 2 (February): 407–14. https://doi.org/10.1007/s00248-014-0487-1.

Huse, Susan M., Yuzhen Ye, Yanjiao Zhou, and Anthony A. Fodor. 2012. "A Core Human Microbiome as Viewed through 16S RRNA Sequence Clusters." *PLOS ONE* 7, no. 6 (June): e34242. https://doi.org/10.1371/journal.pone.0034242.

Jin, Hanbo, Guoru Hu, Chuqing Sun, Yiqian Duan, Zhenmo Zhang, Zhi Liu, Xing-Ming Zhao, and Wei-Hua Chen. 2022. "MBodyMap: A Curated Database for Microbes across Human Body and Their Associations with Health and Diseases." *Nucleic Acids Research* 50, no. D1 (January): D808–16. https://doi.org/10.1093/nar/gkab973.

Keegan, Kevin P., Elizabeth M. Glass, and Folker Meyer. 2016. "MG-RAST, a Metagenomics Service for Analysis of Microbial Community Structure and Function." *Methods in Molecular Biology* 1399: 207–33. https://doi.org/10.1007/978-1-4939-3369-3_13.

Kennedy, Katherine M., Marcus C. de Goffau, Maria Elisa Perez-Muñoz, Marie-Claire Arrieta, Fredrik Bäckhed, Peer Bork, Thorsten Braun, et al. 2023. "Questioning the Fetal Microbiome Illustrates Pitfalls of Low-Biomass Microbial Studies." *Nature* 613, no. 7945 (January): 639–49. https://doi.org/10.1038/s41586-022-05546-8.

Lemas, Dominick J., Shanique Yee, Nicole Cacho, Darci Miller, Michelle Cardel, Matthew Gurka, David Janicke, and Elizabeth Shenkman. 2016. "Exploring the Contribution of Maternal Antibiotics and Breastfeeding to Development of the Infant Microbiome and Pediatric Obesity." *Seminars in Fetal and Neonatal Medicine* 21, no. 6 (December): 406–9. https://doi.org/10.1016/j.siny.2016.04.013.

Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, and Tak-Wah Lam. 2015. "MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct *de Bruijn* Graph." *Bioinformatics* 31, no. 10 (May 15): 1674–76. https://doi.org/10.1093/bioinformatics/btv033.

Ma, Bing, Larry J. Forney, and Jacques Ravel. 2012. "Vaginal Microbiome: Rethinking Health and Disease." *Annual Review of Microbiology* 66, no. 1 (October): 371–89. https://doi.org/10.1146/annurev-micro-092611-150157.

Mallott, Elizabeth K., Carola Borries, Andreas Koenig, Katherine R. Amato, and Amy Lu. 2020. "Reproductive Hormones Mediate Changes in the Gut Microbiome during Pregnancy and Lactation in Phayre's Leaf Monkeys." *Scientific Reports* 10, no. 1 (June): 9961. https://doi.org/10.1038/s41598-020-66865-2.

Maqsood, Raeesah and Trevor W. Stone. 2016. “The Gut-Brain Axis, BDNF, NMDA and CNS Disorders.” *Neurochemical Research* 41, no. 11 (November): 2819–35. https://doi.org/10.1007/s11064-016-2039-1.

Miller, Elizabeth, DeAnna Beasley, Robert Dunn, and Elizabeth Archie. 2016. “Lactobacilli Dominance and Vaginal PH: Why Is the Human Vaginal Microbiome Unique?” *Frontiers in Microbiology* 7 (December). https://www.frontiersin.org/articles/10.3389/fmicb.2016.01936.

Moore, Rebecca E. and Steven D. Townsend. 2019. “Temporal Development of the Infant Gut Microbiome.” *Open Biology* 9, no. 9 (September): 190128. https://doi.org/10.1098/rsob.190128.

Morris, Gerwyn, Michael Berk, Andre Carvalho, Javier R. Caso, Yolanda Sanz, Ken Walder, and Michael Maes. 2017. “The Role of the Microbial Metabolites Including Tryptophan Catabolites and Short Chain Fatty Acids in the Pathophysiology of Immune-Inflammatory and Neuroimmune Disease.” *Molecular Neurobiology* 54, no. 6 (August): 4432–51. https://doi.org/10.1007/s12035-016-0004-2.

Nagpal, Jatin and John F. Cryan. 2021. “Host Genetics, the Microbiome & Behaviour—a ‘Holobiont’ Perspective.” *Cell Research* 31, no. 8 (August): 832–33. https://doi.org/10.1038/s41422-021-00512-x.

Neuman, Hadar and Omry Koren. 2017. “The Pregnancy Microbiome.” In *Nestlé Nutrition Institute Workshop Series* 88: 1-9. https://doi.org/10.1159/000455207.

Nunn, Kenetta L. and Larry J. Forney. 2016. “Unraveling the Dynamics of the Human Vaginal Microbiome.” *The Yale Journal of Biology and Medicine* 89, no. 3 (September): 331–37. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5045142/.

Nurk, Sergey, Dmitry Meleshko, Anton Korobeynikov, and Pavel A. Pevzner. 2017. “MetaSPAdes: A New Versatile Metagenomic Assembler.” *Genome Research* 27, no. 5 (May): 824–34. https://doi.org/10.1101/gr.213959.116.

Muscogiuri, Giovanna, Elena Cantone, Sara Cassarano, Dario Tuccinardi, Luigi Barrea, Silvia Savastano, and Annamaria Colao. 2019. “Gut Microbiota: A New Path to Treat Obesity.” *International Journal of Obesity Supplements* 9, no. 1 (April): 10–19. https://doi.org/10.1038/s41367-019-0011-7.

Pereira-Marques, Joana, Anne Hout, Rui M. Ferreira, Michiel Weber, Ines Pinto-Ribeiro, Leen-Jan van Doorn, Cornelis Willem Knetsch, and Ceu Figueiredo. 2019. “Impact of Host DNA and Sequencing Depth on the Taxonomic Resolution of Whole Metagenome Sequencing for Microbiome Analysis.” *Frontiers in Microbiology* 10 (June). https://doi.org/10.3389/fmicb.2019.01277.

Prince, Amanda L., Derrick M. Chu, Maxim D. Seferovic, Kathleen M. Antony, Jun Ma, and Kjersti M. Aagaard. 2015. "The Perinatal Microbiome and Pregnancy: Moving Beyond the Vaginal Microbiome." *Cold Spring Harbor Perspectives in Medicine* 5, no. 6 (June): a023051. https://doi.org/10.1101/cshperspect.a023051.

Rao, Chitong, Katharine Z. Coyte, Wayne Bainter, Raif S. Geha, Camilia R. Martin, and Seth Rakoff-Nahoum. 2020. "Multi-kingdom quantitation reveals distinct ecological drivers of predictable early-life microbiome assembly." *bioRxiv* (March). https://doi.org/10.1101/2020.03.02.970061.

Ravel, Jacques, Pawel Gajer, Zaid Abdo, G. Maria Schneider, Sara S. K. Koenig, Stacey L. McCulle, Shara Karlebach, et al. 2010. "Vaginal Microbiome of Reproductive-Age Women." *Proceedings of the National Academy of Sciences* 108, no. 1 (March): 4680–87. https://doi.org/10.1073/pnas.1002611107.

Raymond, Frédéric, Amin A. Ouameur, Maxime Déraspe, Naeem Iqbal, Hélène Gingras, Bédis Dridi, Philippe Leprohon, et al. 2016. "The Initial State of the Human Gut Microbiome Determines Its Reshaping by Antibiotics." *The ISME Journal* 10, no. 3 (2016): 707–20. https://doi.org/10.1038/ismej.2015.148.

Renz, Harald, Becky D. Adkins, Sina Bartfeld, Richard S. Blumberg, Donna L. Farber, Johan Garssen, Peter Ghazal, et al. 2018. "The Neonatal Window of Opportunity—Early Priming for Life." *The Journal of Allergy and Clinical Immunology* 141, no. 4 (April): 1212–14. https://doi.org/10.1016/j.jaci.2017.11.019.

Sanders, M. E., D. Merenstein, C. A. Merrifield, and R. Hutkins. 2018. "Probiotics for Human Use." *Nutrition Bulletin* 43, no. 3 (September): 212–25. https://doi.org/10.1111/nbu.12334.

Schlebush, Sanmarié, Rikki M. A. Graham, Amy V. Jennison, Melissa M. Lassig-Smith, Patrick N. A. Harris, Jeffrey Lipman, Páraic Ó Cuív, and David L. Paterson. 2022. "Standard rectal swabs as a surrogate sample for gut microbiome monitoring in intensive care." *BMC Microbiology* 22 (April): 99. https://doi-org.proxy.library.emory.edu/10.1186/s12866-022-02487-0.

Schloss, Patrick D., Sarah L. Westcott, Thomas Ryabin, Justine R. Hall, Martin Hartmann, Emily B. Hollister, Ryan A. Lesniewski, et al. 2009. "Introducing Mothur: Open-Source, Platform-Independent, Community-Supported Software for Describing and Comparing Microbial Communities." *Applied and Environmental Microbiology* 75, no. 23 (December): 7537–41. https://doi.org/10.1128/AEM.01541-09.

Seedat, Farah, Chris Stinton, Jacoby Patterson, Julia Geppert, Bee Tan, Esther R. Robinson, Noel Denis McCarthy, et al. 2017. "Adverse Events in Women and Children Who Have Received Intrapartum Antibiotic Prophylaxis Treatment: A Systematic Review." *BMC Pregnancy and Childbirth* 17, no. 1 (July): 247. https://doi.org/10.1186/s12884-017-1432-3.

Shanahan, Fergus, Tarini S. Ghosh, and Paul W. O'Toole. 2021. "The Healthy Microbiome—What Is the Definition of a Healthy Gut Microbiome?" *Gastroenterology* 160, no. 2 (January): 483–94. https://doi.org/10.1053/j.gastro.2020.09.057.

Short, Meghan I., Robert Hudson, Benjamin D. Besasie, Kelly R. Reveles, Dimpy P. Shah, Susannah Nicholson, Teresa L. Johnson-Pais, Korri Weldon, Zhao Lai, Robin J. Leach, et al. 2021. "Comparison of rectal swab, glove tip, and participant-collected stool techniques for gut microbiome sampling." *BMC Microbiology* 21 (January): 26. https://doi-org.proxy.library.emory.edu/10.1186/s12866-020-02080-3.

Singh, Anchala and Mahima Mittal. 2020. "Neonatal Microbiome – a Brief Review." *The Journal of Maternal-Fetal & Neonatal Medicine* 33, no. 22 (November): 3841–48. https://doi.org/10.1080/14767058.2019.1583738.

Smid, Marcela, Nitasha Ricks, Alexis Panzer, Amber Mccoy, M. Azcarate-Peril, Temitope Keku, and Kim Boggess. 2018. "Maternal Gut Microbiome Biodiversity in Pregnancy." *American Journal of Perinatology* 35, no. 1 (January): 024–030. https://doi.org/10.1055/s-0037-1604412.

Smith, Hunter Jackson. 2020. "An Ethical Investigation into the Microbiome: The Intersection of Agriculture, Genetics, and the Obesity Epidemic." *Gut Microbes* 12, no. 1 (November): 1760712. https://doi.org/10.1080/19490976.2020.1760712.

Smith, Leigh K. and Emily F. Wissel. 2019. "Microbes and the Mind: How Bacteria Shape Affect, Neurological Processes, Cognition, Social Relationships, Development, and Pathology." *Perspectives on Psychological Science* 14, no. 3 (May): 397–418. https://doi.org/10.1177/1745691618809379.

Song, Eun-Ji, Eun-Sook Lee, and Young-Do Nam. 2018. "Progress of analytical tools and techniques for human gut microbiome research." *Journal of Microbiology* 56, no. 10 (June): 693-705. DOI 10.1007/s12275-018-8238-5.

Song, Stephanie D., Kalpana D. Acharya, Jade E. Zhu, Christen M. Deveney, Marina R. S. Walther-Antonio, Marc J. Tetel, and Nicholas Chia. 2020. "Daily Vaginal Microbiota Fluctuations Associated with Natural Hormonal Cycle, Contraceptives, Diet, and Exercise." Edited by Krishna Rao. *MSphere* 5, no. 4 (August): e00593-20. https://doi.org/10.1128/mSphere.00593-20.

Subramaniam, Akila, Ranjit Kumar, Suzanne P. Cliver, Degui Zhi, Jeff M. Szychowski, Adi Abramovici, Joseph R. Biggio, Elliot J. Lefkowitz, Casey Morrow, and Rodney K. Edwards. 2016. "Vaginal Microbiota in Pregnancy: Evaluation Based on Vaginal Flora, Birth Outcome, and Race." *American Journal of Perinatology* 33, no. 4 (October): 401–8. https://doi.org/10.1055/s-0035-1565919.

Tang, Qiang, Ge Jin, Gang Wang, Tianyu Liu, Xiang Liu, Bangmao Wang, and Hailong Cao. 2020. “Current Sampling Methods for Gut Microbiota: A Call for More Precise Devices.” *Frontiers in Cellular and Infection Microbiology* 10 (April). https://doi.org/10.3389/fcimb.2020.00151.

Tett, Adrian, Edoardo Pasolli, Giulia Masetti, Danilo Ercolini, and Nicola Segata. 2021. “Prevotella Diversity, Niches and Interactions with the Human Host.” *Nature Reviews Microbiology* 19, no. 9 (September): 585–99. https://doi.org/10.1038/s41579-021-00559-y.

Tian, Liang, Xu-Wen Wang, Ang-Kun Wu, Yuhang Fan, Jonathan Friedman, Amber Dahlin, Matthew K. Waldor, George M. Weinstock, Scott T. Weiss, and Yang-Yu Liu. 2020. “Deciphering Functional Redundancy in the Human Microbiome.” *Nature Communications* 11, no. 1 (December): 6217. https://doi.org/10.1038/s41467-020-19940-1.

Tipton, Laura, John L. Darcy, and Nicole A. Hynson. 2019. “A Developing Symbiosis: Enabling Cross-Talk Between Ecologists and Microbiome Scientists.” *Frontiers in Microbiology* 10 (February): 292. https://doi.org/10.3389/fmicb.2019.00292.

Vandegrift, Roo, Ashkaan K. Fahimipour, Mario Muscarella, Ashley C. Bateman, Kevin Van Den Wymelenberg, and Brendan J. M. Bohannan. 2019. “Moving Microbes: The Dynamics of Transient Microbial Residence on Human Skin.” *bioRxiv* (March). https://doi.org/10.1101/586008.

Vinturache, Angela E., Cynthia Gyamfi-Bannerman, Joseph Hwang, Indira U. Mysorekar, and Bo Jacobsson. 2016. “Maternal Microbiome – A Pathway to Preterm Birth.” *Seminars in Fetal and Neonatal Medicine* 21, no. 2 (April): 94–99. https://doi.org/10.1016/j.siny.2016.02.004.

Virtanen, Seppo, Ilkka Kalliala, Pekka Nieminen, and Anne Salonen. 2017. “Comparative Analysis of Vaginal Microbiota Sampling Using 16S RRNA Gene Analysis.” *PLoS ONE* 12, no. 7 (July): e0181477. https://doi.org/10.1371/journal.pone.0181477.

Warren, Frederick J., Naoki M. Fukuma, Deirdre Mikkelsen, Bernadine M. Flanagan, Barbara A. Williams, Allan T. Lisle, Páraic Ó Cuív, Mark Morrison, and Michael J. Gidley. 2018. “Food Starch Structure Impacts Gut Microbiome Composition.” *MSphere* 3, no. 3 (June): e00086-18. https://doi.org/10.1128/mSphere.00086-18.

Weissman, J. L., Sonia Dogra, Keyan Javadi, Samantha Bolten, Rachel Flint, Cyrus Davati, Jess Beattie, et al. 2021. “Exploring the Functional Composition of the Human Microbiome Using a Hand-Curated Microbial Trait Database.” *BMC Bioinformatics* 22, no. 1 (June): 306. https://doi.org/10.1186/s12859-021-04216-2.

Witkin, S. S. and Linhares, I. M. 2016. “Why do lactobacilli dominate the human vaginal microbiota?” *BJOG: An International Journal of Obstetrics and Gynecology* 124, no. 4 (March): 606-11. https://doi.org/10.1111/1471-0528.14390.

Wood, Derrick E., Jennifer Lu, and Ben Langmead. 2019. “Improved Metagenomic Analysis with Kraken 2.” *Genome Biology* 20, no. 1 (November): 257. https://doi.org/10.1186/s13059-019-1891-0.

Zhu, Zhengyi, Glen A. Satten, Caroline Mitchell, and Yi-Juan Hu. 2021. “Constraining PERMANOVA and LDM to Within-Set Comparisons by Projection Improves the Efficiency of Analyses of Matched Sets of Microbiome Data.” *Microbiome* 9, no. 1 (June): 133. https://doi.org/10.1186/s40168-021-01034-9.

Chapter 2

The Impact of Urogenital Infections on the Gut and Vaginal Microbiome in Pregnancy

Abstract

Background: During pregnancy, the gut and vaginal microbiome change as the pregnancy progresses. Differences in specific dynamics in the microbiome between those who develop urogenital infection during pregnancy and those who do not develop urogenital infections are unclear.

Methods: A subset of the data from the Emory University African American Vaginal, Oral, and Gut Microbiome in Pregnancy Cohort Study were sent for metagenomic sequencing (238 patients, rectal and vaginal swabs at 8-14 weeks & 24 - 30 weeks pregnancy). Taxonomic assignment was done with metaphlan2. To test for association between UTIs, BV, and Chlamydia infection and the microbiome, a linear decomposition model was used with R package LDM. Data was analyzed separately for rectal and vaginal samples.

Results: There was no significant association between the overall gut and vaginal microbiome and chlamydia, urinary tract infection (UTI), or bacterial vaginosis (BV) in pregnant patients. However, specific taxa in the gut and vaginal microbiome were found to be significantly associated with each of the infections. Alpha papilloma virus was found to be significantly associated with bacterial vaginosis, and *Lactobacillus jensenii, Lactobacillus phage Lv 1*, and *Megasphaera genomosp type 1* were all significantly associated with UTI. In addition, *Mobiluncus mulieris* in the vagina and *Mycoplasma hominis* in the gut were found to be significantly associated with chlamydia infection. The results of the LDM analysis were replicated using 16S rRNA gene sequencing data, though results of specific taxa associated with each urogenital infection differed between metagenome and 16S data.

Keywords: antimicrobial resistance, chlamydia, urinary tract infection, bacterial vaginosis, pregnancy, shotgun metagenomics

Introduction

Gut and vaginal microbial communities can be influenced by many factors, including diet, lifestyle, and hormonal changes. During pregnancy, the gut and vaginal microbiome are of heightened importance because they can influence the health of a pregnant person and developing fetus. Studies have shown that the gut microbiome of pregnant people generally becomes less diverse than that of non-pregnant people (Koren et al. 2012, Neuman and Koren 2017). This is thought to be due to the hormonal, immunological, and metabolic changes that occur to help the body support a healthy pregnancy (Fuhler 2020). The change in the vaginal microbiome over the course of a pregnancy is not very well understood. However, the vaginal microbiome shifts towards *Lactobacillus* dominance for most individuals during pregnancy as estrogen levels stabilize (Fuhler 2020). Sometimes, a diverse microbiome is associated with pregnancy complications such as preterm birth (Mysorekar and Cao 2014), however, this research is not always replicated.

Antibiotics are prescribed during 1 in 4 pregnancies in the United States. During pregnancy, susceptibility to urogenital infections such as chlamydia, UTIs, and BV increases compared to nonpregnant individuals (Elkady, Sinha, and Hassan 2019). Immunological changes occur so that there is not an immune response against the developing fetus, and this change, in addition to changes in hormones and physiology during pregnancy, could play a role in this increased susceptibility. It is not well understood how the gut and vaginal microbiome recover after urogenital infections during pregnancy and whether these urogenital infections or their treatment have any broader impact on the microbiome during pregnancy. However, evidence suggests that certain infections, like high-risk human papilloma viruses (HPV) and bacterial vaginosis (BV), are associated with a more complex, diverse vaginal microbiome (Chen et al.

2019, Severgnini et al. 2022). A blinded study found that some vaginal microbes in pregnant people were associated with BV, and that this was independent of race and birth timing (full vs. preterm; Subramaniam 2016).

Despite the progress that has been made in our understanding of the gut and vaginal microbiome during pregnancy, there is still much that remains unknown. The aim of this study is to examine how bacterial species change in the gut and vaginal microbiome during pregnancy with and without urogenital infections, and whether potential mechanisms can be hypothesized from these associations.

Methods

Rectal and vaginal swabs were self-collected at two time points during pregnancy as previously described (Corwin et al. 2017). Participants consented to a medical chart review so that medications, complications, and birth outcomes could be known. 980 swabs from 238 patients were included in this study.

Swabs were sent to Omega Bioservices Inc. for metagenome sequencing with NextSeq. The resulting FASTQ files were processed using the bioBakery pipeline. FASTQC was used for quality filtering the FASTQ files (Babraham Bioinformatics, n.d.). Trimmaomatic (Bolger, Lohse, and Usadel 2014) was used to remove adaptor sequences. Briefly, human reads were removed with kneaddata v0.7.4 (Huttenhower Lab, n.d.) and taxonomy was assigned with metaphlan v2.6.0 (Truong et al. 2015). After quality control, rectal samples had a median of 16.12 million reads, and vaginal samples had a median of 4.46 million reads. Bray-Curtis dissimilarity (Faith, Minchin, and Belbin 1987) was used to compare the relative abundance of taxa detected in our positive controls to known relative abundance from Zymo (Zymo Research, n.d.) and ensured these numbers were within accepted thresholds (<30% dissimilarity). Batch

effect by sequencing plate was observed and controlled for within the LDM package in R as a confounder (Yi-Juan and Satten 2020). Decontam [c] in R was used to detect contaminants from the negative control samples. Of 27 taxa detected in the negative control samples, only 4 were plausible gut or vaginal microbes, and only 3 were indicated to be contaminants from decontam. Read counts in the contaminated negative control samples indicated that contamination was minimal and not distributed across all sequencing plates. As such, the sequencing plate was controlled for in all subsequent analyses. Taxa which were detected at less than 1% relative abundance were excluded from further analysis. This 1% threshold was selected based on results from positive control samples, in which many false positives were detected under 1% by metaphlan2. Of note, metaphlan2 could correctly identify some taxa below 1% relative abundance, however there are proportionally many more false positives than true positives between 1% and 0.01%.

Since relatively few taxa were identified at the species level within each sample (**Figure One)**, a small subset of samples were processed with metaphlan2 (Truong et al. 2015), metaphlan3 (Beghini et al. 2021, and kraken2 (Wood, Lu, and Langmead 2019) to compare the number of taxa identified per sample. Three samples with high DNA concentration and 3 samples with low DNA concentration and few taxa identified by metaphlan2 were selected per body site for this subsample. High DNA concentration samples were E0594-2- VagM1, G0040-1-VagM1, E0502-1-VagM1, E0632-2-RecM1, E0588-2-RecM1, and G0139-1- RecM1. Low DNA concentration samples with few identified species were G0296-1-Vag, E0683-1-Vag, E0627-2-Vag, G0050-1-Rec, G0148-2-Rec, and E0671-2-Rec. Four samples from an external study (Goltsman et al. 2018) with a similar self-collection protocol for pregnant participants were also included (rectal sample SRR6747978 & SRR6748112; vaginal swabs SRR6747938 &

SRR6747978). FASTQ files from the external study were subsampled so that the total number of reads was the same as the median number of reads from the FASTQ files per body site from this study.

The 16S rRNA gene sequence data were analyzed as previously described (Wright et al. 2022, Dunlop et al. 2021). Briefly, the V3-V4 region was sequenced using Illumina HiSeq at the University of Maryland. QIIME (Bolyen et al. 2019) and DADA2 (Callahan et al. 2016) were used for quality control and data processing. Closed-reference operational taxonomic units (OTUs) were assigned using PECAN (Wright 2018). The purpose of the 16S data in this study was to see if LDM analysis in metagenomic data could be replicated.

A data matrix of the relative abundance, with samples as rows and taxa as columns, was used for further analysis with the R package LDM (Zhu et al. 2021). LDM is a preferable analysis approach as it can handle compositional data, can test for a global effect of the microbiome and effect of any individual taxa while controlling for the false discovery rate, and handle multiple covariates at once (numerical and categorical). Only taxa at the same hierarchy level were included as input per LDM analysis (e.g., only species included as columns for the species analysis, only genera for the genus-level analysis). Confounders were selected based on prior literature and significant variables from the parent study with the full 16S dataset. Confounders were body site, parity, BMI, age, tobacco use, alcohol consumption, marijuana use, and income (categorical). Covariates of interest were timepoint (matched LDM), chlamydia infection, UTI, and BV diagnosed during pregnancy (cluster LDM). All code is available on GitHub (Wissel 2023). A separate LDM analysis was done per body site and was replicated for species and genera from the metagenomic data, and at the species level from the 16S data for the same samples.

Results

Pregnant patients had a mean age of 24.9 years old (sd = 4.73; range 18 - 40) in this study. On average, patients had given birth once before (parity median = 1; sd = 1.05; range 0 - 4). Few patients report tobacco (17.6%), alcohol (8.8%), or marijuana (37%) consumption during their pregnancy. For the urogenital infections of interest, 57 patients (31.5%) received a BV diagnosis during their pregnancy, 46 (24%) received a UTI diagnosis, and 30 (14.4%) received a Chlamydia diagnosis (**Figure Two)**.

After quality control and filtering out human reads, there were a median of 16.12 million reads per rectal sample and a median of 4.46 million reads per vaginal sample. A median of 11 species were identified in the rectal swabs, with a median of 2 species in the vaginal swabs (**Figure One)**. Since there are relatively few species identified per sample, a small subset of samples was processed with metaphlan2, metaphlan3, and kraken2 to compare the number of species taxa identified per sample. Overall, any particular species was not found in more than three samples; most species were only found in two samples. Differences in the total number of taxa identified could be mostly explained by difference in the database used by each bioinformatic tool (e.g., kraken2 identified more species in vaginal samples, but the kraken2 database also reflects the latest taxonomy changes which splits Lactobacillus into many genera; Table 1). While this subsample was too small to test for significant differences between bioinformatic methods, the comparable results show that low taxonomic yield is likely due to the samples themselves and not a bioinformatic artifact. Overall, the few species per sample is attributed to the relative abundance threshold (1%), and subsequent analyses can include a lower threshold (0.01%) when the number of false positives is more permissible. Lowering the 0.01% threshold could improve the number of species detected, but also provide different kinds of

insight into the dynamics of microbes in the gut and vagina, as low abundance microbes are often important in infection contexts (Costello et al. 2012; Hajishengallis et al. 2011; Lloyd-Price, Abu-Ali, and Huttenhower 2016).

LDM analysis of the shotgun metagenome data revealed no significant associations between the gut or vaginal microbiome and time point at the species or genus level ($p > 0.05$, Table 1). No individual taxa were significantly associated with time point, indicating that the microbiome as a whole and no individual microbe was significantly associated with early or late pregnancy at either body site. This finding was replicated in the 16S rRNA gene sequence ASV LDM analysis. The cluster LDM analysis revealed no significant associations between the microbiome and a diagnosis of chlamydia, UTI, or BV during pregnancy at the species and genus level with either body site ($p > 0.05$ Table 1). Individual species and genera from both the gut and vaginal microbiome were associated with these urogenital infections (Table 2). Significant species which were present in fewer than 5 samples are not reported. Notably, *Mycoplasma hominis* from the gut, an opportunistic pathogen in humans (Taylor-Robinson 1996; Haggerty and Taylor 2011), and *Mobiluncus mulieris* in the vagina, a microbe associated with urogenital infections (Mastromarino, Vitali, and Mosca 2013), were significantly associated with chlamydia infection during pregnancy. *Lactobacillus jensenii*, *Lactobacillus phage Lv 1,* and *Megasphaera genomosp type 1* were all significantly associated with UTI during pregnancy, while *Alpha Papillomavirus 14* was significantly associated with BV.

Analysis with genera revealed that genus *Haemophilus* in the gut and *Begomovirus* and *Badnavirus* in the vagina were significantly associated with UTI. These are two likely novel viruses that are placed in this taxa group due to a lack of other information. *Megasphaera,*

Alphapapillomavirus, Begomovirus, and *Campylobacter* in the vagina were associated with BV. No genera were significantly associated with chlamydia at either body site.

The LDM analysis was replicated with the ASVs from the 16S data. As with metagenome data, there were no global associations between the microbiome and any of the three urogenital infections for any body site. However, the vaginal microbiome from 16SrRNA gene sequence data was significantly associated with time point, indicating there is a difference in the vaginal microbiome as a whole as a pregnancy progresses. There was no association between the gut microbiome and time point. A UTI diagnosis was associated with g__Klebsiella.s__Kosakonia_ in the gut microbiome (note that this taxa has undergone recent taxonomy changes; Kosakonia is now its own genus, but the species identified by DADA2 is given in this manuscript). Many individual taxa from the 16S rRNA gene sequence data were significantly associated with the three urogenital infections, however, none of these taxa were significant in the metagenome data. This is likely due to differences in what can be detected by different sequencing approaches and differences in the database used by the bioinformatic tools used for taxonomic assignment.

Discussion

Overall, there was no significant association between the microbiome as a whole and chlamydia, UTI, or BV, a good finding for pregnant patients, as it indicates that these common infections are not significantly changing the entire microbiome over the course of pregnancy. These results should be replicated in a broader population with more sampling timepoints before being assumed to be broadly true, as some research has found associations between the diversity of the microbiome and these infections at various timepoints in pregnancy (Severgnini et al. 2022). Some individual taxa were associated with diagnosis of specific urogenital infections

(**Table 3**), discussed below. While few taxa per sample were detected after the 1% quality control, we have confidence in the few taxa that were significant as they are replicated from prior research studies and have biologic plausibility in the context of infection. Results from the matched LDM analyses examining timepoint and the microbiome are not shown because those results were not significant.

There was also no significant change in the microbiome detected as a pregnancy progresses, while much of the literature does detect a decrease of diversity in the vaginal microbiome as a pregnancy progresses (Nuriel-Ohayon, Neuman, and Koren 2016; Aagaard et al. 2012) (though the microbiome in late pregnancy looks more like the microbiome of nonpregnant individuals (MacIntyre et al. 2015, Romero et al. 2014). It is unclear if the lack of detection of a difference over time in this study is an artifact such as from sampling method, due differences in sampling time point during pregnancy between this and other studies, or some other unknown cause. Further, the LDM analysis of 16S rRNA gene sequences revealed an association between the vaginal microbiome and time point, indicating that there is likely a significant change in the microbiome as a whole over time during pregnancy, in line with prior literature, but not detected in the shotgun metagenome data due to low resolution data.

Bacterial Vaginosis

While there was no significant association between any of the three infections and the microbiome, specific taxa in both the gut and vaginal microbiome were significantly associated with each of the three infections. Alphapapillomma viruses are a type of human papillomavirus that can cause genital warts and cervical cancer. In this study, the presence of Alpha Papilloma virus was significantly associated with BV at the genus and species level in the vagina. This is in line with prior literature which detected Alpha Papilloma in pooled samples from pregnant

people with BV (Zhang et al. 2021). At the genus level, Megasphaera in the vagina was significantly associated with BV, which is not surprising as Megasphaera is a genus of BV-associated bacteria. A common rectal genus and common cause of diarrhea, Campylobacter, when found in the vaginal environment was significantly associated with BV diagnosis as well. One likely misclassified virus genus in the vagina was associated with BV, Begomovirus. This suggests that the role of viruses in this polymicrobial infection is underappreciated and likely plays a role in the development of BV symptoms. Of note, these taxa were not significantly associated with BV in the 16S analysis. However, 16S data lacks the ability to detect viruses. Further, the bacteria taxa were detected in the genus-level analysis of the metagenome data, while the 16S data was only analyzed at the OTU level to try and replicate findings in the species level analysis.

Urinary Tract Infection

UTIs are the most common bacterial infections in pregnancy (Foxman 2002) and can increase risk for complications like preeclampsia and low birth weight if left untreated (Kalinderi et al. 2018). In this study, *Lactobacillus jensenii*, *Lactobacillus phage Lv 1*, and *Megasphaera genomosp type 1* were all significantly associated with UTI during pregnancy. *Megasphaera* is typically associated with BV in pregnant and nonpregnant populations (Zozaya-Hinchliffe, Martin, and Ferris 2008; Glascock et al. 2021) (and was associated with BV at the genus level in this study), but it and other BV associated bacteria have been found to be associated with risk for UTI in pregnancy before (Yoo et al. 2022). While it is typically *E. coli* that causes UTI, Megasphaera species may make the urinary tract more susceptible to infection. Two likely misclassified viral genera in the vagina, Begomovirus and Badnavirus, were significantly associated with UTI. These are likely misclassified because they are plant pathogens, and

databases do not describe viruses as well as they describe bacteria. Haemophilus in the gut microbiome was significantly associated with UTI, though this taxa is only present in 5 samples in this cohort. Pathogenic species like *H. influenzae* and *H. parainfluenzae* belong to this genus, and this genus typically colonizes the nasopharynx and requires metabolites from broken down blood cells to grow (Musher 1996). As it is an opportunistic pathogen which can be a part of the commensal microbiome, it is likely that the immune response to UTI causes changes to the gut microbiome that allows *Haemophilus* to proliferate in the gut without causing other illness.

Chlamydia Infection

Chlamydia species are obligate intracellular species that grow inside other mammalian cells. Chlamydia infection during pregnancy confers high risk for multiple measures of preterm birth, low birth weight babies, and babies small for their gestational age (He et al. 2020). As such, pregnant patients are routinely screened for chlamydia at pregnancy onset and in the third trimester so they can be treated quickly to avoid the development of these complications (Majeroni and Ukkadam 2007). In this study, *Mobiluncus mulieris* in the vagina was significantly associated with chlamydia infection during pregnancy and was previously found to be associated with chlamydia infections (Bommana et al. 2022) and BV (Bautista et al. 2016). *Mycoplasma hominis* in the gut microbiome was significantly associated with chlamydia infection. *M. hominis* is often associated with urogenital infections, especially in immunocompromised patients (McMahon et al. 1990). While *M. hominis* was detected in rectal samples and not vaginal samples in this study, these body sites are very close to one another and it is feasible that *M. hominis* originating from the rectum can exist in the extragenital area, especially in the context of another infection. Further, this microbe is unlikely to be a

contaminant as it was not detected in any of the negative control samples or in the contaminant analysis with decontam in R.

16S versus shotgun metagenome LDM analysis

The purpose of the 16S rRNA gene sequencing data in this study was to see if results from the shotgun metagenome data could be replicated with the same analysis with the 16S data. Comparison between the 16S rRNA gene sequencing data and metagenome data revealed each type of sequencing data is better suited for detecting different microbes. For example, viruses can be detected in metagenome data but not 16S data, while the 16S data has no sequences lost to human DNA reads, and a larger portion of the total reads can be classified.

While overall, there were no significant differences between the microbiome over time and body site for any of the analyses, analysis of the 16S rRNA gene sequence data did find a significant association between the vaginal microbiome at the ASV level and time point. Many ASVs from the vaginal microbiome and a single ASV from the gut microbiome were significant in the LDM analysis, however, there was no overlap between the individual taxa detected with the 16S analysis and the metagenome analysis. All significantly associated taxa from the LDM analyses have biological plausibility and relevance, indicating that all these taxa are likely playing a role in the dynamics of the microbiome during infection in pregnancy. Overall, using different sequencing technologies will provide different advantages, answer different kinds of questions, and require different resources to process. Clinicians will need to consider patients' needs and restraints (i.e., cost and time) when considering which approach to use as microbiome data becomes integrated into healthcare practice.

Conclusion

Overall, the specific taxa that were associated with urogenital infections were taxa that prior literature also found to be associated with these infections. We did not find any associations between the any of these urogenital infections and the microbiome, indicating that these infections are not causing widespread, global changes in the microbiome during pregnancy and suggesting that pregnant individuals are not at severe risk of microbiome disruption as a result of these infections and their treatment. 16S rRNA gene sequencing and metagenome sequencing can provide different information on the microbiome, and clinicians will have to consider the cost and resource requirement for each when deciding to integrate microbiome profiling into patient care.

Figure 1: Number of Species Identified per Sample by MetaPhlAn2

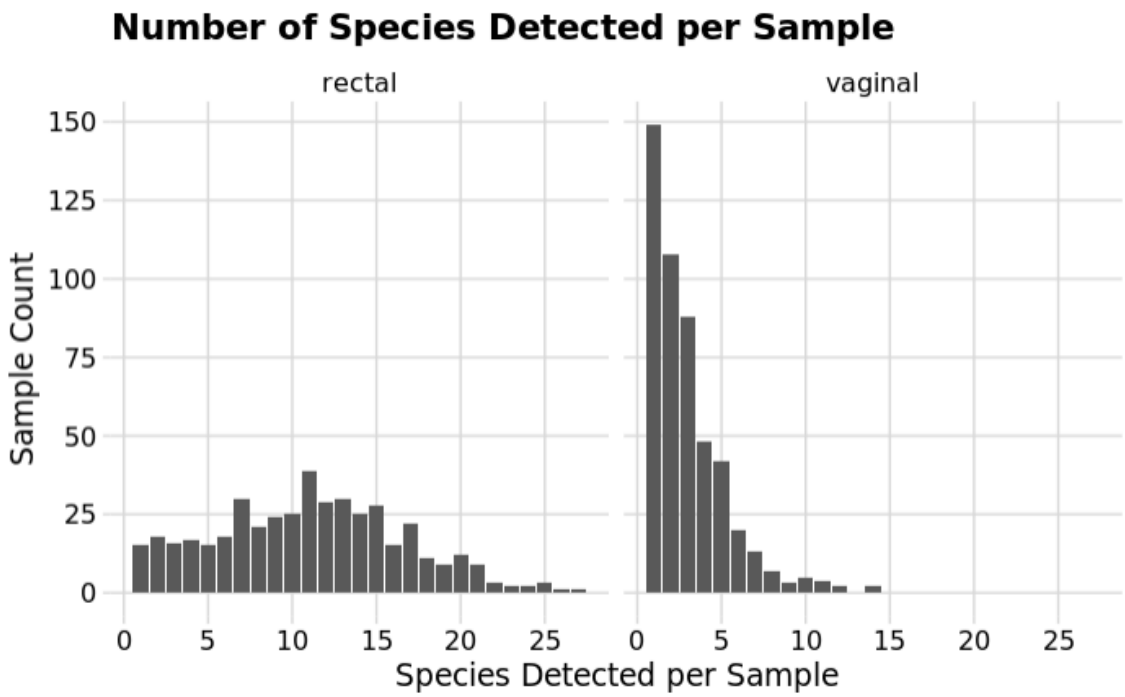
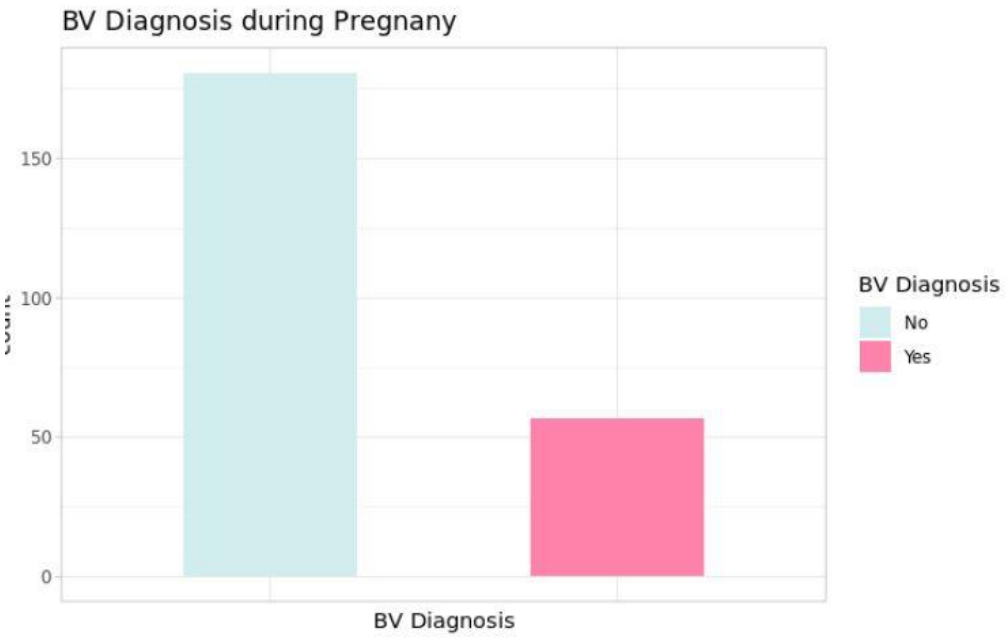
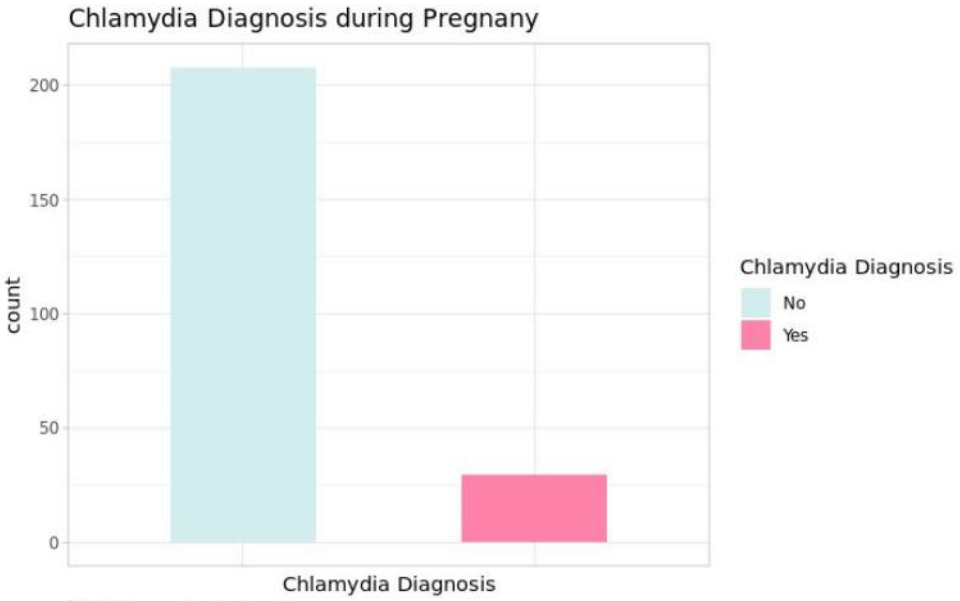


Figure 2: Prevalence of BV, UTI, and Chlamydia Diagnosis in this sample

(I'm planning to have three plots to show the prevalence of BV, UTI, and Chlamydia diagnosis in the study. My undergrad is working on it currently but here is an example of one of the figures. Final plot will have the same Y axis. Also playing with stacked bar charts.)




Chlamydia Diagnosis during Pregnany
200
150
100
50
0
count
Chlamydia Diagnosis
Chlamydia Diagnosis
No
Yes

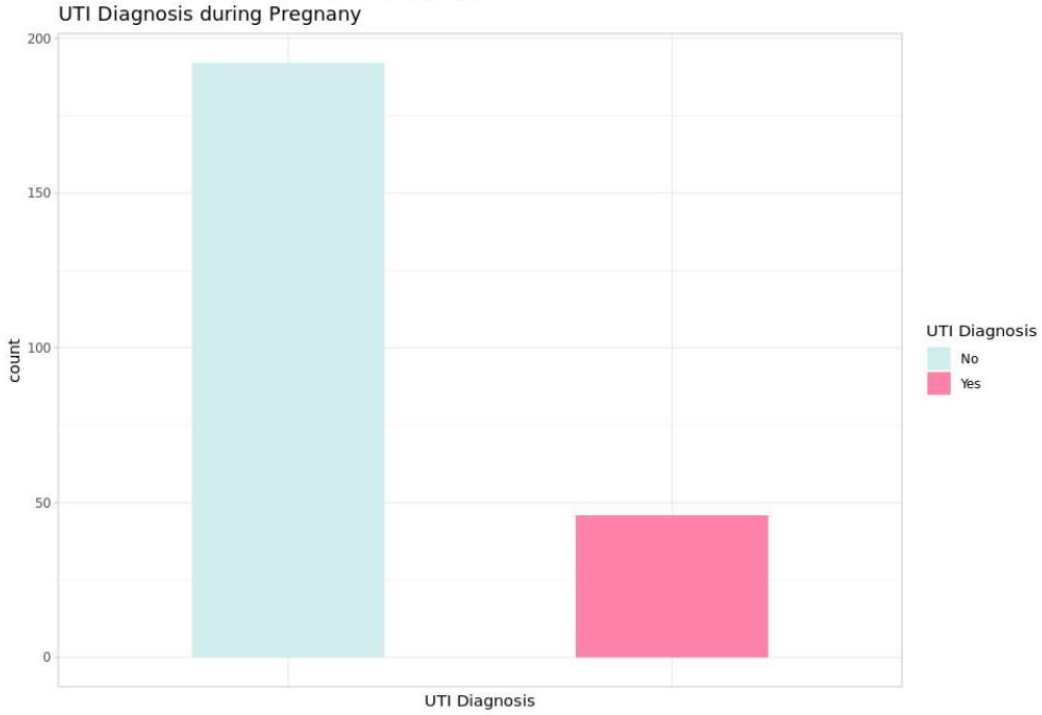

UTI Diagnosis during Pregnany
200
150
100
50
0
count
UTI Diagnosis
UTI Diagnosis
No
Yes

Table 1: Comparison of Total Number of Species Identified by different bioinformatic tools

This table shows how many species were identified in ten metagenome samples (5 per body site) using the default settings of the taxonomic assignment tool. Overall, no taxa were found in more than three samples; most species were only found in two samples. Differences in the taxa identified could be mostly explained by difference in the database used (e.g., kraken2 identified more species in vaginal samples, but the kraken2 database also reflects the latest taxonomy changes which splits *Lactobacillus* into many genera).

	Species Identified		
body site	**metaphlan2**	**metaphlan3**	**kraken2**
rectal	47	93	66
vaginal	24	19	45

Table 2 Significant p values

The results from the LDM analysis are presented below. The table shows the p-values for testing an association between the microbiome as a whole and the covariate of interest. Significant p-values are in bold. In the LDM analysis, confounding variables controlled for were sequencing plate, maternal age, income, parity, tobacco use, alcohol use, and marijuana use. Chlamydia infection, UTI infection, and BV were tested in a cluster LDM while timepoint was tested in a matched LDM while controlling for the other covariates of interest.

body site	**data type**	**preg_chlam**	**preg_uti**	**preg_BV**	**timepoint**
rectal	species, metagenomic	0.6880	0.9910	0.2960	0.8560
vaginal	species, metagenomic	0.7760	0.4250	0.9430	0.9560
rectal	genus, metagenomic	0.8820	0.9760	0.4980	0.2120
vaginal	genus, metagenomic	0.1660	0.7090	0.2410	0.4960
rectal	16S, OTU	0.4230	0.7070	0.5780	0.5020
vaginal	16S, OTU	0.1720	0.3410	0.5980	**0.0036**

Table 3 Significant taxa from aim 1 analysis

This table displays significant results from LDM analysis in R. Each body site and data type were run in its own LDM model (e.g., rectal species from metagenome, rectal 16S ASVs, and rectal genus from metagenome). Note that only significant results came from the cluster LDM analyses with covariates Chlamydia, UTI, and BV. Results from the matched LDM analyses examining timepoint and the microbiome are not shown because those results were not significant. Taxa from the metagenomic data was assigned with metaphlan2. OTUs from 16S data are assigned with DADA2 for rectal samples and PECAN for vaginal samples.

body site	Data Type	covariate	species	Raw p value	Adj p value	Number samples it is in
rectal	Metagenome species	Chlamydia	Mycoplasma_hominis	0.000128	0.00362	25
rectal	Metagenome genus	UTI	Haemophilus	0.00027	0.00901	5
rectal	16S, OTU	UTI	.g__Klebsiella.s__Kosakonia_sp.	0.000256	0.0104	17
vaginal	Metagenome species	UTI	Lactobacillus_jensenii	0.007	**0.0461**	51
vaginal	Metagenome species	UTI	Megasphaera_genomosp_type_1	0.00197	**0.0245**	42

vaginal	Metagenome species	Chlamydia	Mobiluncus_mulieris	0.000405	**0.0454**	11
vaginal	Metagenome species	UTI	Lactobacillus_phage_Lv_1	0.000282	**0.00631**	9
vaginal	Metagenome species	BV	Alphapapillomavirus_14	0.005	**0.0149**	5
Vaginal	Metagenome genus	BV	Megasphaera	0.00075	0.0046	95
Vaginal	Metagenome genus	BV	Alphapapillomavirus	0.00647	0.0291	19
Vaginal	Metagenome genus	UTI	Begomovirus	0.000975	0.0346	11
Vaginal	Metagenome genus	BV	Begomovirus	0.0015	0.00779	11
Vaginal	Metagenome genus	UTI	Badnavirus	0.00171	0.0404	5
Vaginal	Metagenome genus	BV	Campylobacter	0.00175	0.00844	5
vaginal	16S, OTU	UTI	Streptococcus_infantis	0.000333	0.0067	394
vaginal	16S, OTU	BV	Phascolarctobacterium_fa	0.000322	0.00312	250

			ecium			
vaginal	16S, OTU	Chlamydia	Phascolarctobacterium_faecium	0.000333	0.00332	250
vaginal	16S, OTU	UTI	Phascolarctobacterium_faecium	0.000333	0.0067	250
vaginal	16S, OTU	Chlam	Prevotellaceae_bacterium	0.000333	0.00332	155
vaginal	16S, OTU	UTI	Prevotellaceae_bacterium	0.000333	0.0067	155
vaginal	16S, OTU	UTI	Fusobacterium_equinum	0.000333	0.0067	142
vaginal	16S, OTU	Chlam	Bifidobacterium_adolescentis	0.00467	0.0416	139
vaginal	16S, OTU	UTI	Bifidobacterium_adolescentis	0.000666	0.0119	139
vaginal	16S, OTU	Chlamydia	Prevotella_copri	0.000333	0.00332	138
vaginal	16S, OTU	UTI	Prevotella_copri	0.000333	0.0067	138
vaginal	16S, OTU	BV	Enterococcus_faecalis	0.000322	0.00312	94

vaginal	16S, OTU	Chlamydia	Aggregatibacter_segnis_Aggregatibacter_actinomycetemcomitans	0.000333	0.00332	79
vaginal	16S, OTU	BV	Parabacteroides_merdae	0.000322	0.00312	78
vaginal	16S, OTU	BV	Prevotella_stercorea	0.000322	0.00312	74
vaginal	16S, OTU	Chlamydia	Prevotella_stercorea	0.000333	0.00332	74
vaginal	16S, OTU	UTI	Prevotella_stercorea	0.000333	0.0067	74
vaginal	16S, OTU	BV	Bacillus_subtilis_Bacillus_vallismortis_Bacillus_mojavensis_Baci	0.000322	0.00312	56
vaginal	16S, OTU	BV	Lactobacillus_fermentum	0.000322	0.00312	51
vaginal	16S, OTU	BV	Listeria_monocytogenes	0.000322	0.00312	42
vaginal	16S, OTU	BV	Staphylococcus_aureus	0.000322	0.00312	42
vaginal	16S, OTU	BV	Corynebacterium_amycolatum	0.000645	0.00476	41
vaginal	16S, OTU	Chlamydi	Corynebacterium_amycol	0.000333	0.00332	41

		a	atum			
vaginal	16S, OTU	BV	Erwinia_billingiae	0.000322	0.00312	37
vaginal	16S, OTU	Chlamydia	Erwinia_billingiae	0.000333	0.00332	37
vaginal	16S, OTU	UTI	Erwinia_billingiae	0.000333	0.0067	37
vaginal	16S, OTU	BV	Pseudomonas_aeruginosa	0.000322	0.00312	35
vaginal	16S, OTU	BV	Neisseria_cinerea_3	0.000322	0.00312	34
vaginal	16S, OTU	BV	Alistipes_finegoldii	0.000322	0.00312	27
vaginal	16S, OTU	Chlamydia	Alistipes_finegoldii	0.000333	0.00332	27
vaginal	16S, OTU	BV	Lactobacillus_ultunensis_1	0.000322	0.00312	13
vaginal	16S, OTU	Chlamydia	Lactobacillus_ultunensis_1	0.004	0.0376	13
vaginal	16S, OTU	BV	Eubacterium_coprostanoligenes	0.000322	0.00312	12

vaginal	16S, OTU	Chlamydia	Eubacterium_coprostanoligenes	0.000333	0.00332	12
vaginal	16S, OTU	BV	Bifidobacterium_ruminantium	0.000322	0.00312	11
vaginal	16S, OTU	Chlamydia	Bifidobacterium_ruminantium	0.000333	0.00332	11
vaginal	16S, OTU	BV	Corynebacterium_imitans	0.000645	0.00476	7
vaginal	16S, OTU	BV	Corynebacterium_appendicis	0.000645	0.00476	7
vaginal	16S, OTU	Chlamydia	Arthrobacter_cumminsii	0.000333	0.00332	7
vaginal	16S, OTU	Chlamydia	Corynebacterium_imitans	0.000333	0.00332	7
vaginal	16S, OTU	Chlamydia	Corynebacterium_appendicis	0.000333	0.00332	7
vaginal	16S, OTU	BV	Brevibacterium_ravenspurgense	0.000645	0.00476	5
vaginal	16S, OTU	Chlamydia	Oligella_urethralis	0.000333	0.00332	5
vaginal	16S, OTU	Chlamydi	Brevibacterium_ravenspur	0.000333	0.00332	5

		a	gense			

References

Aagaard, Kjersti, Kevin Riehle, Jun Ma, Nicola Segata, Toni-Ann Mistretta, Cristian Coarfa, Sabeen Raza, et al. 2012. "A Metagenomic Approach to Characterization of the Vaginal Microbiome Signature in Pregnancy." *PloS One* 7, no. 6: e36466. https://doi.org/10.1371/journal.pone.0036466.

Babraham Bioinformatics. n.d. "FastQC." Accessed April 5, 2023. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bautista, Christian T., Eyako Wurapa, Warren B. Sateren, Sara Morris, Bruce Hollingsworth, and Jose L. Sanchez. 2016. "Bacterial Vaginosis: A Synthesis of the Literature on Etiology, Prevalence, Risk Factors, and Relationship with Chlamydia and Gonorrhea Infections." *Military Medical Research* 3, no. 1 (February): 4. https://doi.org/10.1186/s40779-016-0074-5.

Beghini, Francesco, Lauren J. McIver, Aitor Blanco-Míguez, Leonard Dubois, Francesco Asnicar, Sagun Maharjan, Ana Mailyan, et al. 2021. "Integrating Taxonomic, Functional, and Strain-Level Profiling of Diverse Microbial Communities with BioBakery 3." *ELife* 10 (May): e65088. https://doi.org/10.7554/eLife.65088.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. 2014. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30, no. 15 (August): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Bolyen, Evan, Jai Ram Rideout, Matthew R. Dillon, Nicholas A. Bokulich, Christian C. Abnet, Gabriel A. Al-Ghalith, Harriet Alexander, et al. 2019. "Reproducible, Interactive, Scalable and Extensible Microbiome Data Science Using QIIME 2." *Nature Biotechnology* 37, no. 8 (August): 852–57. https://doi.org/10.1038/s41587-019-0209-9.

Bommana, Sankhya, Gracie Richards, Mike Kama, Reshma Kodimerla, Kenan Jijakli, Timothy D. Read, and Deborah Dean. 2022. "Metagenomic Shotgun Sequencing of Endocervical, Vaginal, and Rectal Samples among Fijian Women with and without Chlamydia Trachomatis Reveals Disparate Microbial Populations and Function across Anatomic Sites: A Pilot Study." Edited by Salika M. Shakir. *Microbiology Spectrum* 10, no. 3 (June): e00105-22. https://doi.org/10.1128/spectrum.00105-22.

Callahan, Benjamin J., Paul J. McMurdie, Michael J. Rosen, Andrew W. Han, Amy Jo A. Johnson, and Susan P. Holmes. 2016. "DADA2: High-Resolution Sample Inference from Illumina Amplicon Data." *Nature Methods* 13, no. 7 (July): 581–83. https://doi.org/10.1038/nmeth.3869.

Chen, Yulian, Zubei Hong, Wenjing Wang, Liying Gu, Hua Gao, Lihua Qiu, and Wen Di. 2019. "Association between the Vaginal Microbiome and High-Risk Human Papillomavirus Infection in Pregnant Chinese Women." *BMC Infectious Diseases* 19, no. 1 (August): 677. https://doi.org/10.1186/s12879-019-4279-6.

Corwin, Elizabeth J., Carol J. Hogue, Bradley Pearce, Cherie C. Hill, Timothy D. Read, Jennifer Mulle, and Anne L. Dunlop. 2017. "Protocol for the Emory University African American Vaginal, Oral, and Gut Microbiome in Pregnancy Cohort Study." *BMC Pregnancy and Childbirth* 17 (June): 161. https://doi.org/10.1186/s12884-017-1357-x.

Costello, Elizabeth K., Keaton Stagaman, Les Dethlefsen, Brendan J. M. Bohannan, and David A. Relman. 2012. "The Application of Ecological Theory towards an Understanding of the Human Microbiome." *Science* 336, no. 6086 (June): 1255–62. https://doi.org/10.1126/science.1224203.

Dunlop, Anne L., Glen A. Satten, Yi-Juan Hu, Anna K. Knight, Cherie C. Hill, Michelle L. Wright, Alicia K. Smith, Timothy D. Read, Bradley D. Pearce, and Elizabeth J. Corwin. 2021. "Vaginal Microbiome Composition in Early Pregnancy and Risk of Spontaneous Preterm and Early Term Birth Among African American Women." *Frontiers in Cellular and Infection Microbiology* 11: 641005. https://doi.org/10.3389/fcimb.2021.641005.

Elkady, Adel, Prabha Sinha, and Soad Ali Zaki Hassan. 2019. *Infections in Pregnancy: An Evidence-Based Approach*. Cambridge: Cambridge University Press.

Faith, Daniel P., Peter R. Minchin, and Lee Belbin. 1987. "Compositional Dissimilarity as a Robust Measure of Ecological Distance." *Vegetatio* 69, no. 1/3: 57–68. https://www.jstor.org/stable/20038103.

Foxman, Betsy. 2002. "Epidemiology of Urinary Tract Infections: Incidence, Morbidity, and Economic Costs." *The American Journal of Medicine* 113, Suppl 1A (July): 5S-13S. https://doi.org/10.1016/s0002-9343(02)01054-9.

Fuhler, G. M. 2020. "The Immune System and Microbiome in Pregnancy." *Best Practice & Research Clinical Gastroenterology*, 44–45 (February): 101671. https://doi.org/10.1016/j.bpg.2020.101671.

Glascock, Abigail L., Nicole R. Jimenez, Sam Boundy, Vishal N. Koparde, J. Paul Brooks, David J. Edwards, Jerome F. Strauss III, et al. 2021. "Unique Roles of Vaginal Megasphaera Phylotypes in Reproductive Health." *Microbial Genomics* 7, no. 12 (December): 000526. https://doi.org/10.1099/mgen.0.000526.

Goltsman, Daniela S. Aliaga, Christine L. Sun, Diana M. Proctor, Daniel B. DiGiulio, Anna Robaczewska, Brian C. Thomas, Gary M. Shaw, et al. 2018. "Metagenomic Analysis with Strain-Level Resolution Reveals Fine-Scale Variation in the Human Pregnancy Microbiome." *Genome Research* 28, no. 10 (October): 1467–80. https://doi.org/10.1101/gr.236000.118.

Haggerty, Catherine L. and Brandie D. Taylor. 2011. "Mycoplasma Genitalium: An Emerging Cause of Pelvic Inflammatory Disease." *Infectious Diseases in Obstetrics and Gynecology* 2011: 959816. https://doi.org/10.1155/2011/959816.

Hajishengallis, George, Shuang Liang, Mark A. Payne, Ahmed Hashim, Ravi Jotwani, Mehmet A. Eskan, Megan L. McIntosh, et al. 2011. "Low-Abundance Biofilm Species Orchestrates Inflammatory Periodontal Disease through the Commensal Microbiota and Complement." *Cell Host & Microbe* 10, no. 5 (November): 497–506. https://doi.org/10.1016/j.chom.2011.10.006.

He, Weihua, Yue Jin, Haibin Zhu, Yan Zheng, and Jianhua Qian. 2020. "Effect of Chlamydia Trachomatis on Adverse Pregnancy Outcomes: A Meta-Analysis." *Archives of Gynecology and Obstetrics* 302, no. 3 (September): 553–67. https://doi.org/10.1007/s00404-020-05664-6.

Hu, Yi-Juan and Glen A Satten. 2020. "Testing Hypotheses about the Microbiome Using the Linear Decomposition Model (LDM)." *Bioinformatics* 36, no. 14 (April): 4106–15. https://doi.org/10.1093/bioinformatics/btaa260.

Huttenhower Lab. n.d. "Kneaddata." Accessed April 5, 2023. https://huttenhower.sph.harvard.edu/kneaddata/.

Kalinderi, Kallirhoe, Dimitrios Delkos, Michail Kalinderis, Apostolos Athanasiadis, and Ioannis Kalogiannidis. 2018. "Urinary Tract Infection during Pregnancy: Current Concepts on a Common Multifaceted Problem." *Journal of Obstetrics and Gynaecology* 38, no. 4 (May): 448–53. https://doi.org/10.1080/01443615.2017.1370579.

Koren, Omry, Julia K. Goodrich, Tyler C. Cullender, Aymé Spor, Kirsi Laitinen, Helene Kling Bäckhed, Antonio Gonzalez, et al. 2012. "Host Remodeling of the Gut Microbiome and Metabolic Changes during Pregnancy." *Cell* 150, no. 3 (August): 470–80. https://doi.org/10.1016/j.cell.2012.07.008.

Lloyd-Price, Jason, Galeb Abu-Ali, and Curtis Huttenhower. 2016. "The Healthy Human Microbiome." *Genome Medicine* 8, no. 1 (April): 51. https://doi.org/10.1186/s13073-016-0307-y.

MacIntyre, David A., Manju Chandiramani, Yun S. Lee, Lindsay Kindinger, Ann Smith, Nicos Angelopoulos, Benjamin Lehne, et al. 2015. "The Vaginal Microbiome during Pregnancy and the Postpartum Period in a European Population." *Scientific Reports* 5 (March): 8988. https://doi.org/10.1038/srep08988.

Majeroni, Barbara A. and Sreelatha Ukkadam. 2007. "Screening and Treatment for Sexually Transmitted Infections in Pregnancy." *American Family Physician* 76, no. 2 (July): 265–70. https://www.aafp.org/pubs/afp/issues/2007/0715/p265.html.

Mastromarino, Paola, Beatrice Vitali, and Luciana Mosca. 2013. "Bacterial Vaginosis: A Review on Clinical Trials with Probiotics." *The New Microbiologica* 36, no. 3 (July): 229–38.

McMahon, Deborah K., J.Stephen Dummer, A. William Pasculle, and Gail Cassell. “Extragenital Mycoplasma Hominis Infections in Adults.” *The American Journal of Medicine* 89, no. 3 (September 1990): 275–81. https://doi.org/10.1016/0002-9343(90)90338-e.

Musher, Daniel M. 1996. “Haemophilus Species.” In *Medical Microbiology*, edited by Samuel Baron, 4th ed. Galveston: University of Texas Medical Branch at Galveston, 1996. http://www.ncbi.nlm.nih.gov/books/NBK8458/.

Mysorekar, Indira U., and Bin Cao. 2014. “Microbiome in Parturition and Preterm Birth.” *Seminars in Reproductive Medicine* 32, no. 1 (January): 50–55. https://doi.org/10.1055/s-0033-1361830.

Neuman, Hadar, and Omry Koren. “The Pregnancy Microbiome.” In *Nestlé Nutrition Institute Workshop Series*, edited by E. Isolauri, P.M. Sherman, and W.A. Walker, 88:1–10. S. Karger AG, 2017. https://doi.org/10.1159/000455207.

Nuriel-Ohayon, Meital, Hadar Neuman, and Omry Koren. 2016. “Microbial Changes during Pregnancy, Birth, and Infancy.” *Frontiers in Microbiology* 7. https://www.frontiersin.org/articles/10.3389/fmicb.2016.01031.

Romero, Roberto, Sonia S. Hassan, Pawel Gajer, Adi L. Tarca, Douglas W. Fadrosh, Lorraine Nikita, Marisa Galuppi, et al. 2014. “The Composition and Stability of the Vaginal Microbiota of Normal Pregnant Women Is Different from That of Non-Pregnant Women.” *Microbiome* 2, no. 1 (February): 4. https://doi.org/10.1186/2049-2618-2-4.

Severgnini, Marco, Sara Morselli, Tania Camboni, Camilla Ceccarani, Melissa Salvo, Sara Zagonari, Giulia Patuelli, et al. 2022. “Gardnerella Vaginalis Clades in Pregnancy: New Insights into the Interactions with the Vaginal Microbiome.” *PLOS ONE* 17, no. 6 (June): e0269590. https://doi.org/10.1371/journal.pone.0269590.

Subramaniam, Akila, Ranjit Kumar, Suzanne P. Cliver, Degui Zhi, Jeff M. Szychowski, Adi Abramovici, Joseph R. Biggio, Elliot J. Lefkowitz, Casey Morrow, and Rodney K. Edwards. 2015. “Vaginal Microbiota in Pregnancy: Evaluation Based on Vaginal Flora, Birth Outcome, and Race.” *American Journal of Perinatology* 33, no. 4 (October): 401–8. https://doi.org/10.1055/s-0035-1565919.

Taylor-Robinson, D. 1996. “Infections Due to Species of Mycoplasma and Ureaplasma: An Update.” *Clinical Infectious Diseases* 23, no. 4 (October): 671–84. https://doi.org/10.1093/clinids/23.4.671.

Truong, Duy Tin, Eric A. Franzosa, Timothy L. Tickle, Matthias Scholz, George Weingart, Edoardo Pasolli, Adrian Tett, Curtis Huttenhower, and Nicola Segata. 2015. “MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling.” *Nature Methods* 12, no. 10 (October): 902–3. https://doi.org/10.1038/nmeth.3589.

Wissel, Emily. “Dissertation Analysis Code.” GitHub. Accessed April 5, 2023. https://github.com/ewissel/dissertation_analysis_code.

Wood, Derrick E., Jennifer Lu, and Ben Langmead. 2019. “Improved Metagenomic Analysis with Kraken 2.” *Genome Biology* 20, no. 1 (November): 257. https://doi.org/10.1186/s13059-019-1891-0.

Wright, Michelle L. 2018. “PECAN.” GitHub. Accessed April 5, 2023. https://github.com/mlwright97/PECAN.

Wright, Michelle L., Anne L. Dunlop, Alexis B. Dunn, Rebecca M. Mitchell, Emily F. Wissel, and Elizabeth J. Corwin. 2022. “Factors Associated with Vaginal *Lactobacillus* Predominance Among African American Women Early in Pregnancy.” *Journal of Women’s Health* 31, no. 5 (May): 682–89. https://doi.org/10.1089/jwh.2021.0148.

Yoo, Jeong-Ju, Ju Sun Song, Woong Bin Kim, Jina Yun, Hee Bong Shin, Mi-Ae Jang, Chang Beom Ryu, et al. 2022. “Gardnerella Vaginalis in Recurrent Urinary Tract Infection Is Associated with Dysbiosis of the Bladder Microbiome.” *Journal of Clinical Medicine* 11, no. 9 (January): 2295. https://doi.org/10.3390/jcm11092295.

Zhang, He-Teng, Hao Wang, Hai-Sheng Wu, Jian Zeng, and Yan Yang. 2021. “Comparison of Viromes in Vaginal Secretion from Pregnant Women with and without Vaginitis.” *Virology Journal* 18, no. 1 (January): 11. https://doi.org/10.1186/s12985-020-01482-z.

Zhu, Zhengyi, Glen A. Satten, Caroline Mitchell, and Yi-Juan Hu. 2021. “Constraining PERMANOVA and LDM to Within-Set Comparisons by Projection Improves the Efficiency of Analyses of Matched Sets of Microbiome Data.” *Microbiome* 9, no. 1 (June): 133. https://doi.org/10.1186/s40168-021-01034-9.

Zozaya-Hinchliffe, Marcela, David H. Martin, and Michael J. Ferris. 2008. “Prevalence and Abundance of Uncultivated Megasphaera-like Bacteria in the Human Vaginal Environment.” *Applied and Environmental Microbiology* 74, no. 5 (March): 1656–59. https://doi.org/10.1128/AEM.02127-07.

Zymo Research. n.d. “Instruction Manual: ZymoBIOMICS Microbial Community DNA Standard.” Accessed April 5, 2023. https://files.zymoresearch.com/protocols/_d6305_d6306_zymobiomics_microbial_community_dna_standard.pdf.

Chapter 3

hAMRoaster: a tool for comparing performance of AMR gene detection software

Abstract

Background. The use of shotgun metagenomics for AMR detection is appealing because data can be generated from clinical samples with minimal processing. Detecting antimicrobial resistance (AMR) in clinical genomic data is an important epidemiological task, yet a complex bioinformatic process. Many software tools exist to detect AMR genes, but they have mostly been tested in their detection of genotypic resistance in individual bacterial strains. Further, these tools use different databases, or even different versions of the same databases. Understanding the comparative performance of these bioinformatics tools for AMR gene detection in shotgun metagenomic data is important because this data type is increasingly used in public health and clinical settings.

Methods. We developed a software pipeline, hAMRoaster (Harmonized AMR Output compAriSon Tool ER; Wissel 2022), for assessing accuracy of prediction of antibiotic resistance phenotypes. For evaluation purposes, we simulated a highly resistant mock community and several low resistance metagenomic short read (Illumina) samples based on sequenced strains with known phenotypes. We benchmarked nine open-source bioinformatics tools for detecting AMR genes that 1) were conda or Docker installable, 2) had been actively maintained, 3) had an open-source license, and 4) took FASTA or FASTQ files as input. hAMRoaster calculated sensitivity, specificity, precision, and accuracy for each tool, comparing detected AMR genes to susceptibility testing.

Conclusion. Overall, all tools were precise and accurate at all genome coverage levels tested (5x, 50x, 100x sequenced bases / genome length) in the highly resistant mock community with

more variability in the low resistance community (1x coverage). This study demonstrated that different bioinformatic tools and pipelines yield differences in AMR gene identification across drug classes, and that these differences become important if researchers are interested in resistance to specific drug classes.

Significance. Software selection for metagenomic AMR prediction should be driven by the context of the clinical/research questions and tolerance for true and false negative results. The ability to assess which bioinformatics tool best fits a particular dataset prior to beginning a large-scale project allows for more efficient processing and analysis using optimal tools for a particular research question. As prediction software and databases are in a state of constant refinement, the approach used here—creating synthetic communities containing taxa and phenotypes of interest along with using hAMRoaster to assess performance of candidate software—offers a template to aid researchers in selecting the most appropriate strategy at the time of analysis.

Keywords: antimicrobial resistance, bioinformatics, metagenomics

Introduction

Antibiotic resistant infections pose a serious threat not only to public health but to the agricultural, veterinary, and food safety industries. The misuse of antibiotics in healthcare and livestock production has led to widespread antimicrobial resistance in diverse environments and has emerged as a threat to global health (Shao et al. 2021, Poole and Sheffield 2013). The burden of multi-drug resistant pathogens is increasing globally, creating complex clinical scenarios in which there are limited (if any) therapeutic options, resulting in increased mortality and healthcare costs for common medical procedures (Teillant et al. 2015). Genes that confer antimicrobial resistance (AMR) are increasingly present in commensal members of the human microbiome and are recognized as an important reservoir for conferring pathogen resistance through horizontal gene transfer (Nji et al. 2021, Brinkac et al. 2017).

Two key approaches to mitigating AMR infections are antibiotic stewardship and AMR surveillance. While antibiotic stewardship focuses on using antibiotics appropriately, AMR surveillance focuses on describing AMR genes already present in a community. Currently, AMR surveillance typically relies on phenotypic characterization through culture or genotypic characterization through molecular diagnostics based on PCR and hybridization techniques (Anjum, Zankari, and Hasman 2017). However, there is a move toward genome-based methods with the Illumina short-read platform being the dominant platform for data generation at the present time (Porter and Hajibabaei 2018).

Sequencing technology has revolutionized research across many disciplines, with more applications found every year as both the technologies and analysis methods advance. This is particularly evident in the use of metagenomic data for the microbial surveillance of antimicrobial resistance (AMR), as microbial communities can be characterized without the need

to first isolate and culture the specimen prior to analysis (Kraemer, Ramachandran, and Perron 2019; Hendriksen et al. 2019; Kumar, Pornsukarom, and Thakur 2019). As the cost and time of sequencing has dramatically decreased, petabytes of data are quickly generated, with Illumina short reads becoming more prevalent (Porter and Hajibabaei 2018; Robinson, Harkin, and Shukla 2021; GenBank, n.d.). Detecting AMR genes potential through non-culture based, high throughput DNA sequencing and bioinformatic approaches is of growing relevance and importance.

There are many bioinformatic tools available to process large amounts of data while following open-science principles (de Abreu, Perdigão, and Almeida 2021). Open science is a term used to describe data that is Findable, Accessible, Interoperable, and Reusable or (FAIR) and that are open-source (Wilkinson et al. 2016). With so many options available, investigators need to determine the open-source tool best suited for their research question. One way to address issues with replicability and variance across studies is to establish standardized bioinformatics pipelines and best practices, as has been done, for example, by the National Microbiome Data Collaborative (NMDC) (Eloe-Fadrosh et al. 2022). However, for many researchers, a standardized bioinformatics pipeline may not the best suited for their data or research question (de Abreu, Perdigão, and Almeida 2021).

As shotgun metagenomic sequencing is emerging as a powerful tool for detecting AMR (Oniciuc et al. 2018), it is essential to evaluate how well different tools perform. In addition to testing AMR gene prediction tools against widely available metagenome samples, they should be compared in samples with extensive phenotypic resistance (acquired and mutational AMR genes). Here, we describe a software pipeline, hAMRoaster, that provides metrics on tool performance in detecting AMR genes from known resistant phenotypes and can therefore help in

decision-making about which tools will be adequate for detecting resistance to the drug classes being studied.

Methods

For a schematic overview of the methods, see **Figure One.**

Development of a software pipeline, hAMRoaster, to assess results of antibiotic resistance prediction

hAMRoaster was written as a conda installable command line tool in a Python script and requires three inputs: a) the text output of AMR tool on a FASTQ or FASTA test file, such as a text file processed through hAMRonization (Public Health Alliance for Genomic Epidemiology 2022), b) a list of known phenotypes associated with the test file or samples names, and c) (optional) a tab formatted table which matches antibiotic drugs with their drug class. If option c) is not specified a default table is used. The output of the program is a set of performance metrics that include sensitivity and specificity. A conda installable version of the software was deposited in the Bioconda (Grüning et al. 2018) database. The GitHub site for the software is https://github.com/ewissel/hAMRoaster.

hAMRoaster requires, as input, a formatted results table of runs by AMR detection tools. This table is identical to that produced by the hAMRonization (Public Health Alliance for Genomic Epidemiology 2022) software**.** hAMRonization is conda installable and can compile the output of many AMR tools into a unified format. shortBRED (Kaminski et al. 2015) and fARGene (fannyhb 2019) are not included in hAMRonization at the time of analysis, so hAMRoaster can take the path to the raw output for these tools and partially match it to the hAMRonization output.

hAMRoaster requires an input to the “known” phenotypic resistance in the mock community (--AMR_key flag of hAMRoaster), such susceptibility testing tables that are available from NCBI Biosamples. Antibiotics in the table of known phenotypic resistances are matched to their respective drug classes. Results classified as “susceptible” in susceptibility testing are considered “susceptible”, and “intermediate” results are ignored. In cases where susceptibility testing occurred with two or more agents, each agent is considered independently (e.g., resistance to “amoxicillin-tetracycline” is treated as resistance to “amoxicillin” and “tetracycline” independently). Each identified AMR gene is labeled with its corresponding drug class for comparison. In instances where a gene confers resistance to multiple drug classes, the detected gene is split into multiple rows so that each conferred resistance can be independently compared to the susceptibility testing. Gene to drug class linkage is verified using the CARD database (Alcock et al. 2020) when applicable by accession ID. Any genes corresponding to ‘unknown’ or ‘other’ drug classes (including hypothetical resistance genes) are excluded from further analysis. Genes that confer resistance to an antibiotic that is only administered and effective in combination with another drug (e.g., clavulanic acid in amoxicillin-clavulanic acid) are classified as ‘Other’ and excluded from analysis.

A detected AMR gene is labeled as a true positive by hAMRoaster if the drug class matched to an AMR gene corresponds to a drug class that tests “resistant” in the susceptibility testing for the mock community. Similarly, a false positive is coded as a drug class that is called by the software, but tested as susceptible in the mock community (--AMR key parameter). Observed AMR genes are labeled “unknown” if the corresponding drug class is not tested in the mock community and is not included in the AMR key file. Once true/false positives and

true/false negatives are determined per tool, hAMRoaster calculates sensitivity, specificity, precision, accuracy, and percent unknown.

Creation of multiple synthetic mock communities of antibiotic resistance bacteria

Simple synthetic community with high *resistance*

Bacterial members of the base mock community were chosen from NCBI's BioSample Database (Barrett et al. 2012) and met the following criteria: (1) the strain had extensive antibiotic susceptibility testing data using CLSI or EUCAST testing standards as part of the public NCBI BioSample record; (2) the strain was isolated from human tissue; (3) the strain was the cause of a clinical infection; (4) the FASTA was available to download from NCBI BioSample Database (Barrett et al. 2012). Eight bacteria, each representing a different species, with overlapping resistance to 43 antibiotics across 18 drug classes, were selected for the mock community (**Table 1**). The included taxa were *Acinetobacter baumannii* MRSN489669, *Citrobacter freundii* MRSN12115, *Enterobacter cloacae* 174, *Escherichia coli* 222, *Klebsiella pneumoniae* CCUG 70742, *Pseudomonas aeruginosa* CCUG 70744, *Neisseria gonorrhoeae* SW0011, and *Staphylococcus aureus* LAC (Table 1).

Paired-end FASTQs were simulated by NCBI's ART (Huang et al. 2012) using default parameters for HiSeq 2500 at three levels of average sequence coverage (5x, 50x, and 100x sequenced bases / genome length) and are available on FigShare (https://figshare.com/account/home#/projects/125974). Simulated FASTQs were subsequently concatenated to resemble shotgun metagenomics reads, and metaSPAdes (Nurk et al. 2017) was used to create assembled contigs. The FASTQs were simulated with approximately equal numbers of reads of each genome.

Complex synthetic clinical mock community with low resistance

We created a community profile with previously simulated human metagenomes (Fritz et al. 2019) and added a single AMR isolate collected from a human infection at 1x coverage to simulate a human metagenome with restrictive phenotypic resistance. We included samples 0 through 5 from CAMISIM (Fritz et al. 2019), a set of previously simulated human metagenomes, and combined these with simulated fastqs from one of two isolates from human infections, SRR17789825 (Biggs et al. 2022) for even sample numbers and SRR16683675 (Sequence Read Archive, n.d.) for odd sample numbers.

Running antibiotic prediction software on mock communities

All tools for AMR prediction were run on the mock community and restrictive samples at all coverage levels using default settings for either simulated FASTQ or assembled contigs. Default settings were used as it is what most users use and understand to be the developer recommendations. When both options were available, assembled contigs were run.

Statistical Analysis

Data were analyzed in Python v3.7.7 and plotted in R v4.0.4. hAMRoaster calculated all performance metrics reported in Table 3. Unweighted Cohen's kappa was calculated using R package IRR (Gamer et al. 2022) for each pairwise combination of tools to test agreement between tools.

Data Availability

All data and code is available on the hAMRoaster GitHub repository (https://github.com/ewissel/hAMRoaster) and figshare (for large files; https://figshare.com/account/home#/projects/125974)

Results

Selection of nine open-source, conda-installable tools for detection of antibiotic resistance phenotypes

To identify tools for antibiotic resistance prediction, we used a multi-headed search strategy. We searched PubMed using terms “AMR”, “antibiotic resistance genes”, “bioinformatics”, and “antimicrobial resistance”. We also searched GitHub using the same set of terms. Once an initial list of tools was compiled, we performed a second PubMed literature review including the search terms from above plus the names of the tools (“tool 1” OR “tool 2”). We also used Twitter to ask the research community what bioinformatic tools they use to identify AMR (supplementary text 1). These searches identified 16 potential tools to identify AMR genes (**Table 2**). The search for tools concluded on March 1, 2021.

For an identified tool to be considered eligible for comparison, it had to meet the following criteria: (1) be conda or Docker installable; (2) have source code publicly available in a data repository and be actively maintained (defined as tool updates or GitHub responses within the last year); (3) have an open-source license; and (4) take FASTQs or FASTAs as input files. Nine tools met the criteria to be included in this analysis: ABRIcate (Seeman 2020), fARGene (Berglund et al. 2019), ResFinder (Bortolaia et al. 2020), shortBRED (Kaminski et al. 2015), RGI (Alcock et al. 2020), AMRFinderPlus (National Center for Biotechnology Information 2023), starAMR (National Microbiology Laboratory 2022) , sraX (Panunzi 2020), and deepARG (Arango-Argoty 2018). PointFinder also qualified (Zankari et al. 2017) but was a subtool of ResFinder and only identified mutational resistance for some organisms, so it was excluded from analysis. The code used to install and run all tools is available on the hAMRoaster GitHub.

ABRIcate

ABRIcate v.1.0.1 took contig FASTA files as inputs and compared reads against NCBI AMRFinder Plus (National Center for Biotechnology Information 2023) by default, though there are options to compare against CARD (Alcock et al. 2020), ResFinder (Bortolaia et al. 2020), ARG-ANNOT (Gupta et al. 2014), MEGARES (Doster et al. 2020), EcOH (Ingle et al. 2016), PlasmidFinder (Carattoli and Hasman 2020), VFDB (Chen et al. 2016), and Ecoli_VF (National Microbiology Laboratory 2017), which are also pre-downloaded. ABRIcate reported on acquired AMR genes and not mutational resistance.

shortBRED

shortBRED (Kaminski et al. 2015) v0.9.3 used a set of marker genes to search metagenomic data for protein families of interest. The bioBakery (McIver et al. 2018) team published an AMR gene marker database built from 849 AR protein families derived from the ARDB (Liu and Pop 2009) v1.1 and independent curation alongside shortBRED, which is used in this study.

fARGene

fARGene (fannyhb 2019, Berglund et al. 2019) v.0.1 used Hidden Markov Models to detect AMR genes from short metagenomic data or long read data. This was a different approach from most other tools which compare the reads directly. fARGene has three pre-built models for detecting resistance to quinolone, tetracycline, and beta lactamases, which were tested in this study. fARGene can predict unknown ARGs using its gene models.

RGI

RGI (Alcock et al. 2020) v5.1.1 used protein homology and SNP models to predict ‘resistomes’. It used CARD’s protein homolog models as a database. RGI predicts open reading frames (ORFs) using Prodigal (Hyatt 2010), detects homologs with BLAST (McGinnis and Madden 2004), and matches to CARD’s database and model cut off values.

ResFinder

ResFinder (Bortolaia et al. 2020) v4.0 was available both as a web-based application or the command line. We used ResFinder 4 in this study, which was specifically designed for detecting genotypic resistance in phenotypically resistant samples. ResFinder aligned reads directly to its own curated database without need for assembly.

deepARG

deepARG (Arango-Argoty et al. 2018) v.2.0 used a supervised deep learning based approach for antibiotic resistance gene annotation of metagenomic sequences. It combines three databases—CARD, ARDB, and UNIPROT—and categorizes them into resistance categories.

sraX

sraX (Panunzi 2020) v.1.5 was built as a one-step tool; in a single command, sraX downloaded a database and aligned contigs to this database with DIAMOND (Buchfink, Reuter, and Drost 2021). By default, sraX used CARD, though other options can be specified. As we use default settings for all tools, only CARD was used in this study for sraX. It should be noted that the one step aspect is convenient but can become lengthy if there are multiple runs and databases need to be downloaded multiple times.

starAMR

starAMR (National Microbiology Laboratory 2022, Zankari et al. 2012) v.0.7.2 used BLAST+ (Camacho et al. 2009) to compare contigs against a combined database with data from ResFinder, PointFinder, and PlasmidFinder.

AMR Finder Plus

AMR Finder Plus (National Center for Biotechnology Information 2023) v.3.9.3 used BLASTX (McGinnis and Madden 2004) translated searches and hierarchical tree of gene families to detect AMR genes. The database was derived from the Pathogen Detection Reference Gene Catalog (National Center for Biotechnology Information 2023) and was compiled as part of the National Database of Antibiotic Resistant Organisms (NDARO).

Performance of software on synthetic metagenomes with high- and -low-prevalence of AMR phenotypes

Each software tool was run against a synthetic mock community of 8 bacteria at three coverage levels that expressed 43 antibiotic resistance phenotypes. Overall, the number of AMR genes detected across all tools ranged from 13 to over 700 at 100x coverage (**Table 3**). For some tools, genes detected did not correspond to a tested phenotype in the mock community, so the prediction fell into the "unknown" category. Among the tools tested, AMR Finder Plus had the highest degree of unclassifiable/unknown results (observed AMR gene not tested in the mock community; **Figure 3).** An overview of these results are available in **Figure 2A**.

After filtering out the AMR genes detected in the simulated human metagenomes (for which AMR phenotypes were unknown), detected AMR genes were examined per sample. None of the tools detected true or false positives for one of the AMR isolates in the low resistance samples (**Figure 2b**). Fewer genes were detected overall compared to the highly resistant

sample, as expected for samples with a limited resistance phenotype (**Table 3**), though many of these corresponded to unknown AMR phenotypes and not those included in susceptibility testing.

Sensitivity and Specificity

Sensitivity tests what portion of AMR genes are correctly identified by a tool when phenotypic resistance to the drug class that gene confers resistance to is present in the mock community. Specificity tests what portion of known negatives (i.e., susceptible drugs from phenotypic testing) do not have AMR genes detected for that drug class. Sensitivity for phenotype detection ranged from >0.99 (RGI) to 0.23 (sraX) at the lowest coverage levels for the highly resistant, antibiotic resistance gene (ARG)-rich dataset sample (**Fig. 2a**). In general, genome coverage did not greatly affect sensitivity, except for that of sraX, which increased to 0.53 at the highest level. fARGene and deepARG had a high sensitivity value (>0.90) at all coverage levels. RGI, deepARG, and fARGene are all tools that compare reads to a model of AMR instead of aligning reads directly to a database, indicating that this method may be appropriate when high sensitivity values are preferred. As a note, in this ARG-rich dataset, there were only 2 possible true negatives because only two drug classes were always susceptible to antibiotics in those two drug classes when tested (nitrofuran and polypeptide).

In samples with lower numbers of resistance genes, sensitivity and specificity were variable within- and across-tools for samples, with sensitivity much lower than the high resistance community **(**(0 - <0.45; **Fig. 1b)** Specificity was much higher overall, though variable across samples depending on whether any true positives were detected by the tools (**Table 3**). Precision was highly variable across tools with no consistent trend across tools (range 0 - 1), while accuracy was less variable, with most tools having an accuracy between .50 and 0.75

Concordance between tools

An analysis of the agreement between tools of detected resistance to drug classes revealed that overall, agreement was highly variable across tools (0.02 - 0.72 at 5x coverage, **Fig. 5A**) between tools at all coverage levels for the ARG-rich dataset (**Figure 5A, 5B, 5C**). Low agreement was found between most tools in the low AMR samples except for AMR Finder Plus, abricate, and ResFinder4, which had kappa values > 0.80 (**Figure 5D**).

Discussion

Development of a framework for assessing AMR prediction software performance using synthetic data

There is a considerable research effort to develop new software for predicting AMR using DNA sequence alone. In this dynamic environment, there is a need for researchers and epidemiologists to understand the relative performance of open-source software tools. While some tools currently exist for compiling the results of several AMR tools together (hAMRonizer and chARMedDb [Underwood 2021]), this study was motivated by the lack of an open-source pipeline for comparing the results once compiled.

The central challenge in developing this software was to compare detected AMR genes to resistance phenotypes. Detected AMR genes needed to be classified by their corresponding drug class(es) so they could be matched to the known phenotypically resistant drug classes. One hurdle in this translation is that tools use different databases, and some databases classify genes differently. For example, shortBRED classifies gene families, while CARD classifies specific genes. While this analysis checked the drug classification via the DNA/Protein Accession value in CARD, only around 300 of the >1,000 genes detected could directly map to genes in CARD by accession value. The hAMRonization tool overcomes this challenge by providing a drug class

column and filling in the values from ChEBI ontology (Hastings et al. 2016) when possible. The hAMRoaster strategy is to assign a CARD drug class value to every detected AMR gene first by accession number, then by gene name. If neither of these methods assign a drug class for an AMR gene, then the drug class provided by hAMRonization is used. Another challenge in converting detected AMR genes to drug classes is that some drugs are only administered in combination, such as clavulanic acid with amoxicillin. For these instances, resistance to the drug only used in combination (e.g., clavulanic acid) is treated as an "other" drug class and excluded from analysis in hAMRoaster. In these cases, we incorporated the experience of practicing clinicians to identify combination antibiotics into the hAMRoaster antibiotic key.

The analysis presented here used synthetic data to compare tool performance. Synthetic data has the benefit of allowing controlled input with known ground truth. Therefore users can focus on the types of organisms and phenotypes they need to detect in their own datasets, perform experiments with real samples, and manipulate a range of factors such as relative abundance and sequencing error. The NCBI BioSample repository (used in this study) is an invaluable resource for creating such datasets as it contains many samples with AMR phenotypes determined by international standards. Researchers could also sequence and phenotype culturable organisms in their own laboratories to provide testing standards to evaluate software. Here, we exclusively examined synthetic short read Illumina data, but this analysis strategy could be adapted to understand the effect of using data generated on long read technologies such as the Pacific Bioscience and Oxford Nanopore platforms.

Overall trends in performance and reasons for variability between tools

We found the sensitivity of almost all tools to be very good in a highly resistant sample (>0.80), except for that of sraX, which had a proportionally high number of false negatives

compared to true positives. However, sensitivity was lower in low-resistance samples (0 - <0.45), indicating that tool selection plays an important role in results for targeted AMR studies. All tools except shortBRED and starAMR detected many genes that were not associated with a lab-determined phenotype in our highly resistant mock community, while this was true for all tools except starAMR in the low-resistance sample. In practice, researchers and epidemiologists may be only interested in a narrow range of AMR phenotypes. Overall, these results indicate when researchers are interested in resistance to a particular drug class as opposed to resistance to a broad range of drug classes, tool selection becomes very important.

We calculated Cohen's kappa to capture the agreement at the drug class level between AMR tools to see if all AMR tools detected resistance to the same drug classes across samples. We found that agreement at the drug class level was surprisingly low across all tools in the high and low resistance data, though some pairs of tools have higher agreement than others (e.g., AMR Finder Plus, abricate, and ResFinder4 in the low resistance samples; **Figure 5**), indicating that some tools may be better suited for detecting different types of resistance. As such, hAMRoaster provided a table with the number of genes detected per drug class for each tool that may help researchers in selecting an AMR gene detection tool that is best suited for their research question.

This research underscores the need for the further development of software tools for the detection of AMR genes in the human microbiome. It is increasingly recognized that the confined location and genetic diversity of this microbial population provides ideal conditions for genetic exchange among residential microbes and between residential and transient microbes, including pathogenic microbes. Notably, rates of horizontal gene transfer among bacteria in the human microbiome (especially the gastrointestinal tract) are estimated to be many times higher

than among bacteria in other diverse ecosystems, such as soil (Human Microbiome Project Consortium 2012). Refined tools appropriate for use in shotgun metagenomic data will be important for tracking the spread of AMR genes from diverse environmental sources to the human microbiome and across sites in the human body and understanding whether AMR genes are derived from vertical inheritance or via horizontal gene transfer.

In conclusion, this study compared bioinformatics tools for detecting AMR genes in a simulated short read metagenomic sample at three coverage levels at one time point. While tools use slightly different methods and databases, these tools overall had high sensitivity for detection of AMR genes, indicating all of these tools perform well for a broad-resistance approach. Moreover, agreement between tools was sometimes low, indicating the importance of careful tool selection when investigating AMR to specific drug classes. We advocate that researchers should test these software tools using pipelines such as hAMRoaster with a synthetic community that highlights the resistance profiles and sample of interest.

Acknowledgements

We thank Jon Moller for helping to create the hAMRoaster name.

Figure 1: Schematic l Methods

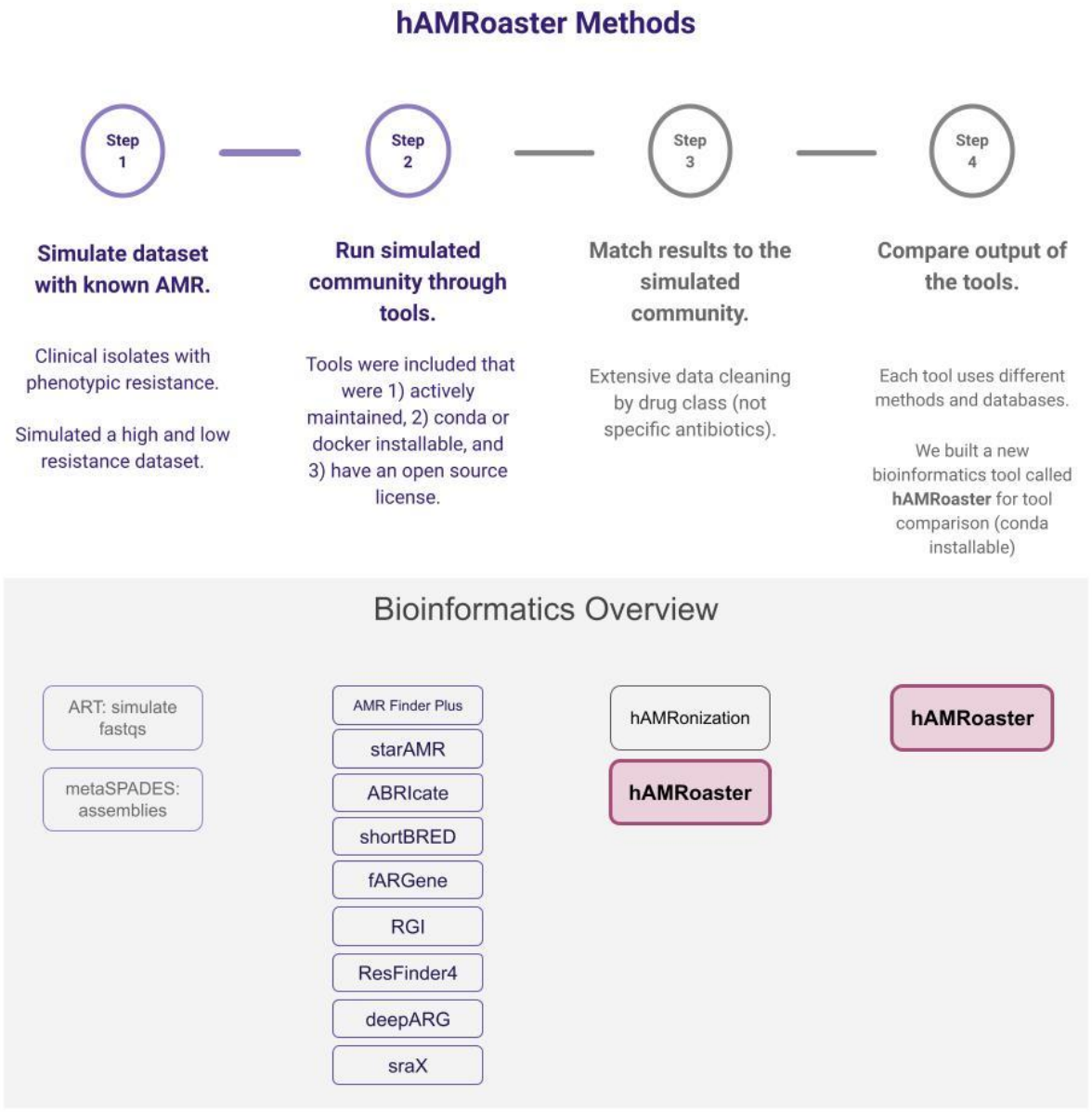
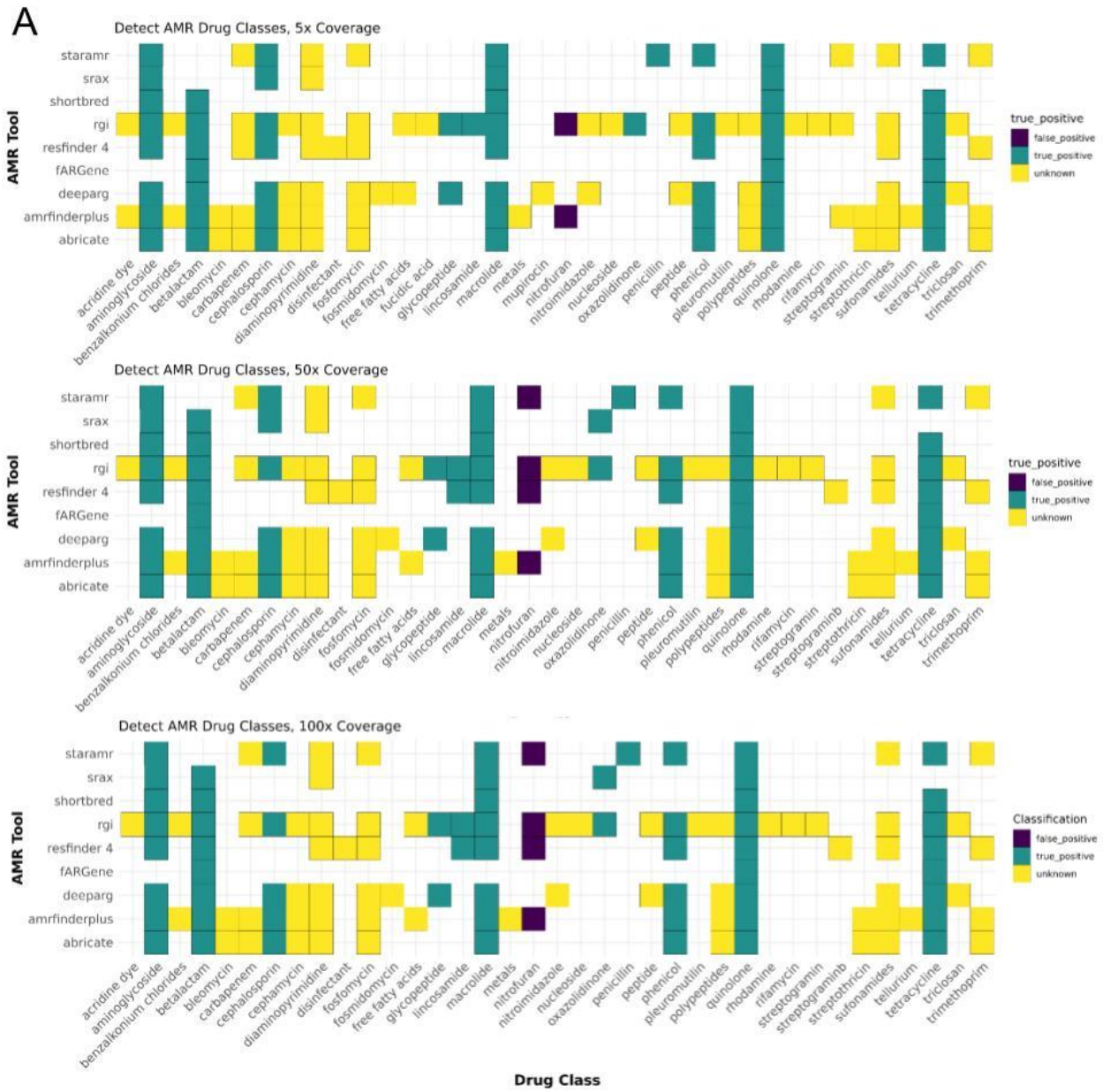


Figure 2: Antimicrobial Resistance (AMR) Genes Detected By Software Tools by Drug Class

AMR Genes detected by each tool across coverage levels, grouped into drug class to which the genes confer resistance with the color coding indicating whether the detection was true positive (green), false positive (purple) or unknown (yellow). Clear spaces in the plot indicate that AMR genes were not detected for the drug class on the x-axis by the tool on the y-axis. Plot A contains the high AMR Data, while plot B contains the low AMR data.


A
Detect AMR Drug Classes, 5x Coverage
Detect AMR Drug Classes, 50x Coverage
Detect AMR Drug Classes, 100x Coverage
AMR Tool
staramr
srax
shortbred
rgi
resfinder 4
fARGene
deeparg
amrfinderplus
abricate
true_positive
false_positive
true_positive
unknown
Classification
acridine dye
aminoglycoside
benzalkonium chlorides
betalactam
bleomycin
carbapenem
cephalosporin
cephamycin
diaminopyrimidine
disinfectant
fosfomycin
fosmidomycin
free fatty acids
fucidic acid
glycopeptide
lincosamide
macrolide
metals
mupirocin
nitrofuran
nitroimidazole
nucleoside
oxazolidinone
penicillin
peptide
phenicol
pleuromutilin
polypeptides
quinolone
rhodamine
rifamycin
streptogramin
streptograminb
streptothricin
sufonamides
tellurium
tetracycline
triclosan
trimethoprim
Drug Class

B

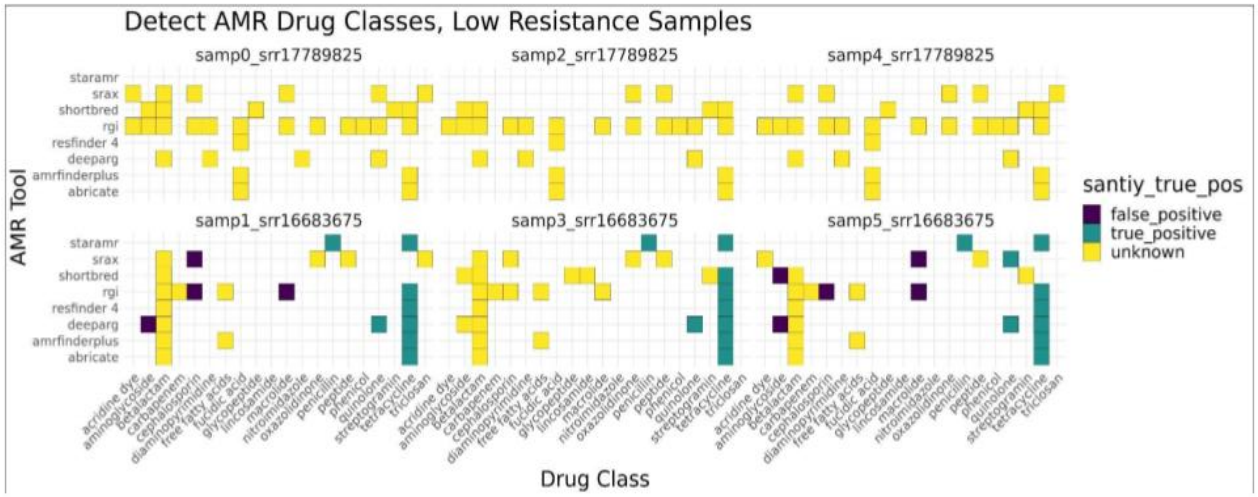
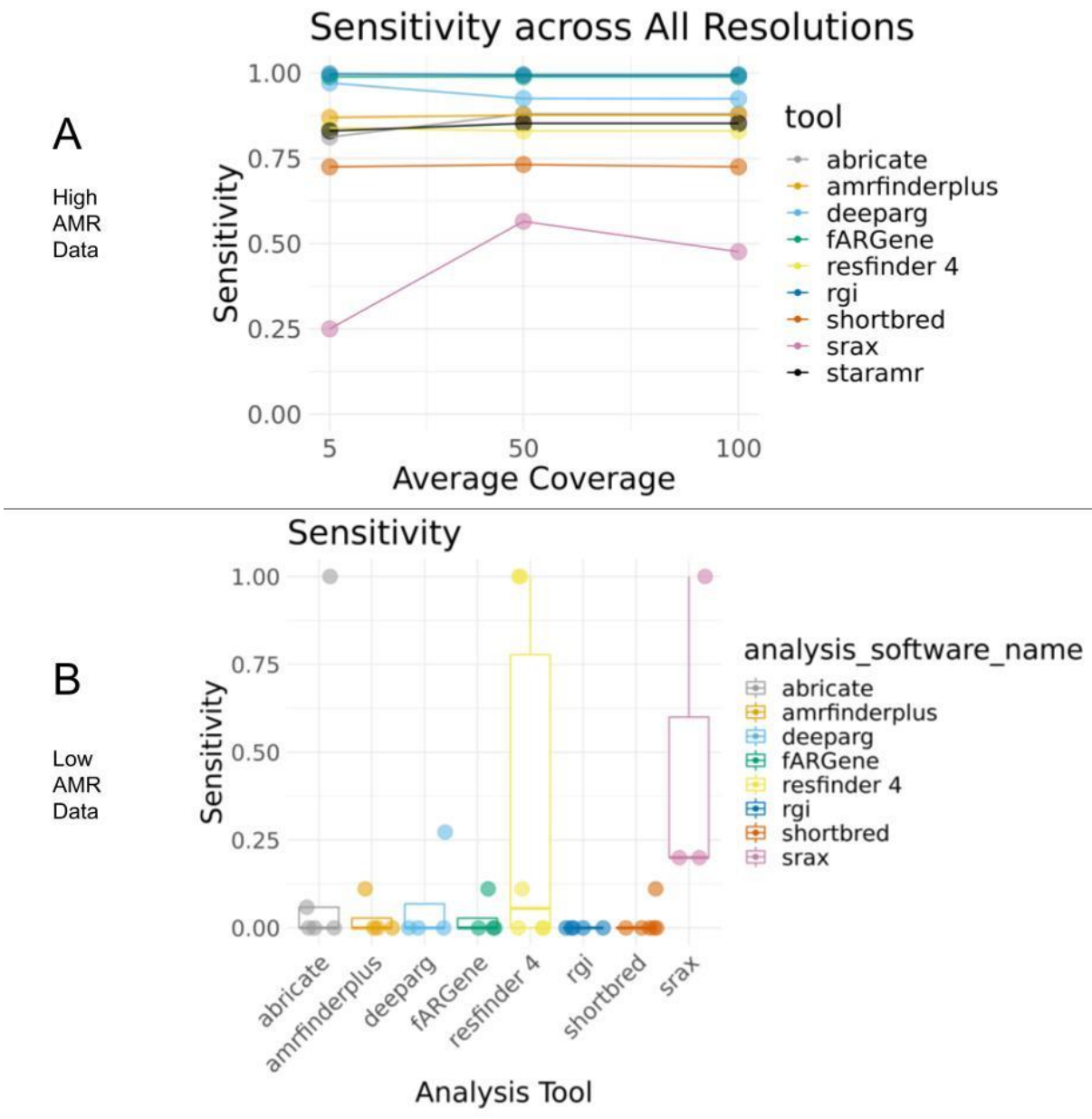

Detect AMR Drug Classes, Low Resistance Samples
samp0_srr17789825
samp2_srr17789825
samp4_srr17789825
samp1_srr16683675
samp3_srr16683675
samp5_srr16683675
staramr
srax
shortbred
rgi
resfinder 4
deeparg
amrfinderplus
abricate
AMR Tool
acridine dye
aminoglycoside
betalactam
carbapenem
cephalosporin
diaminopyrimidine
free fatty acids
fusidic acid
glycopeptide
lincosamide
macrolide
nitroimidazole
oxazolidinone
penicillin
peptide
phenicol
quinolone
streptogramin
tetracycline
triclosan
Drug Class
santiy_true_pos
false_positive
true_positive
unknown

Figure 3: Sensitivity of Software Tools for Detection of Antimicrobial Resistance (AMR) Genes Across Coverage Levels

Sensitivity was calculated as (true positives) / (true positives + false negatives). Most tools were highly sensitive (greater than 0.80). All genes corresponding to “Other” or “Unknown” drug classes were not included in these calculations. Similarly, AMR genes corresponding to phenotypic resistance that was not tested in the mock community was considered “Unknown”



and not included in the sensitivity analysis.

Figure 4: Percent Detection of Unknown Antimicrobial (AMR) Resistance Genes Across Coverage

The percent detection of AMR genes that could not be classified because the drug class the gene confers resistance to was not tested for the high AMR (A) and low AMR (b) data. A black dashed line is placed at 20%, indicating where at least 20% of the detected AMR genes could not be classified.

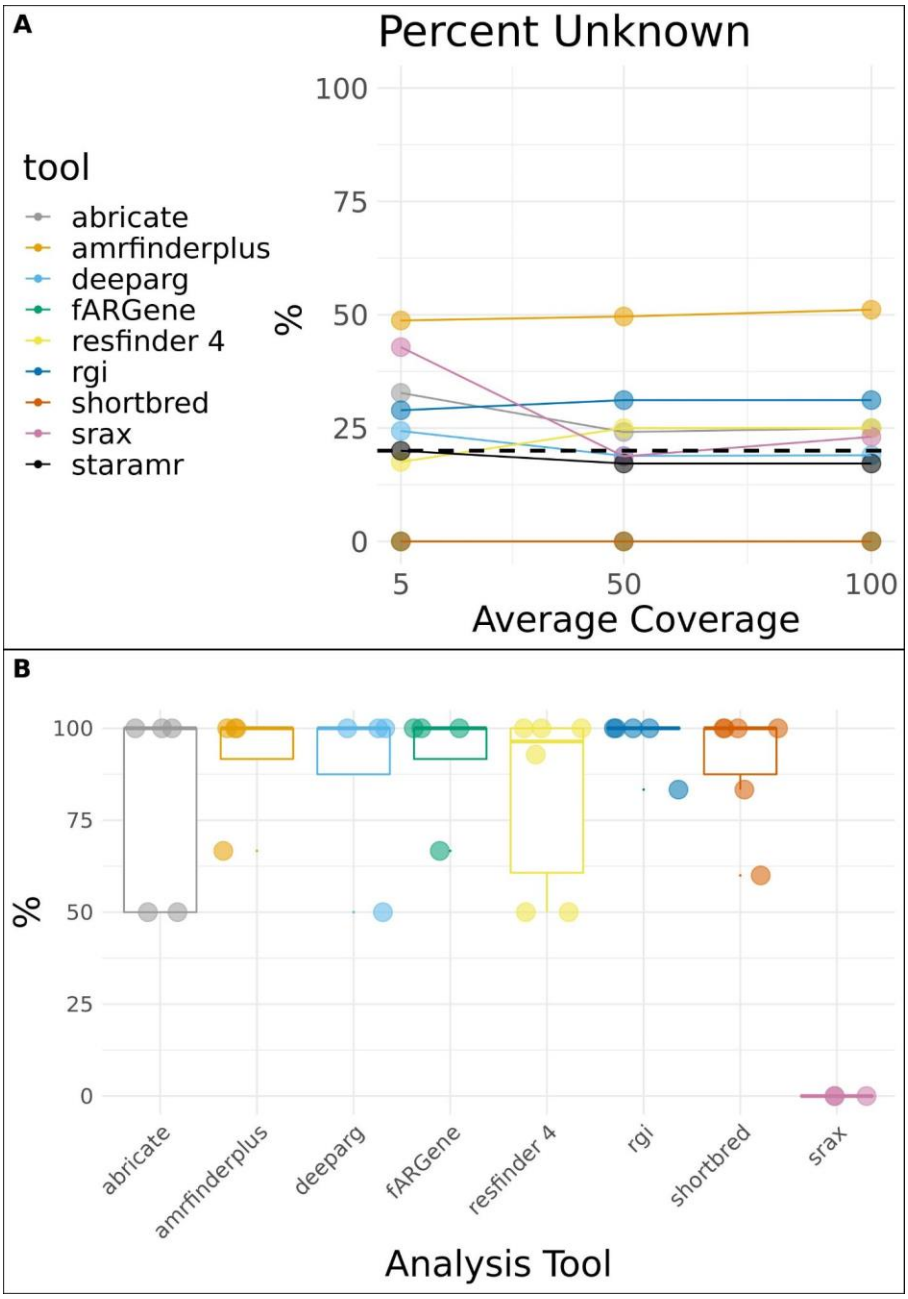


Figure 5: Agreement (Cohen's Kappa) values between tools across coverage levels calculated in R using the kappa2 function

Agreement between tools in detecting resistance to drug classes is shaded across all plots while kappa values are bolded when the p-value is less than 0.05. A, B, and C display the agreement between tools for the 5x, 50x, and 100x coverage high AMR datasets, respectively. D displays the agreement between tools for the low AMR samples.

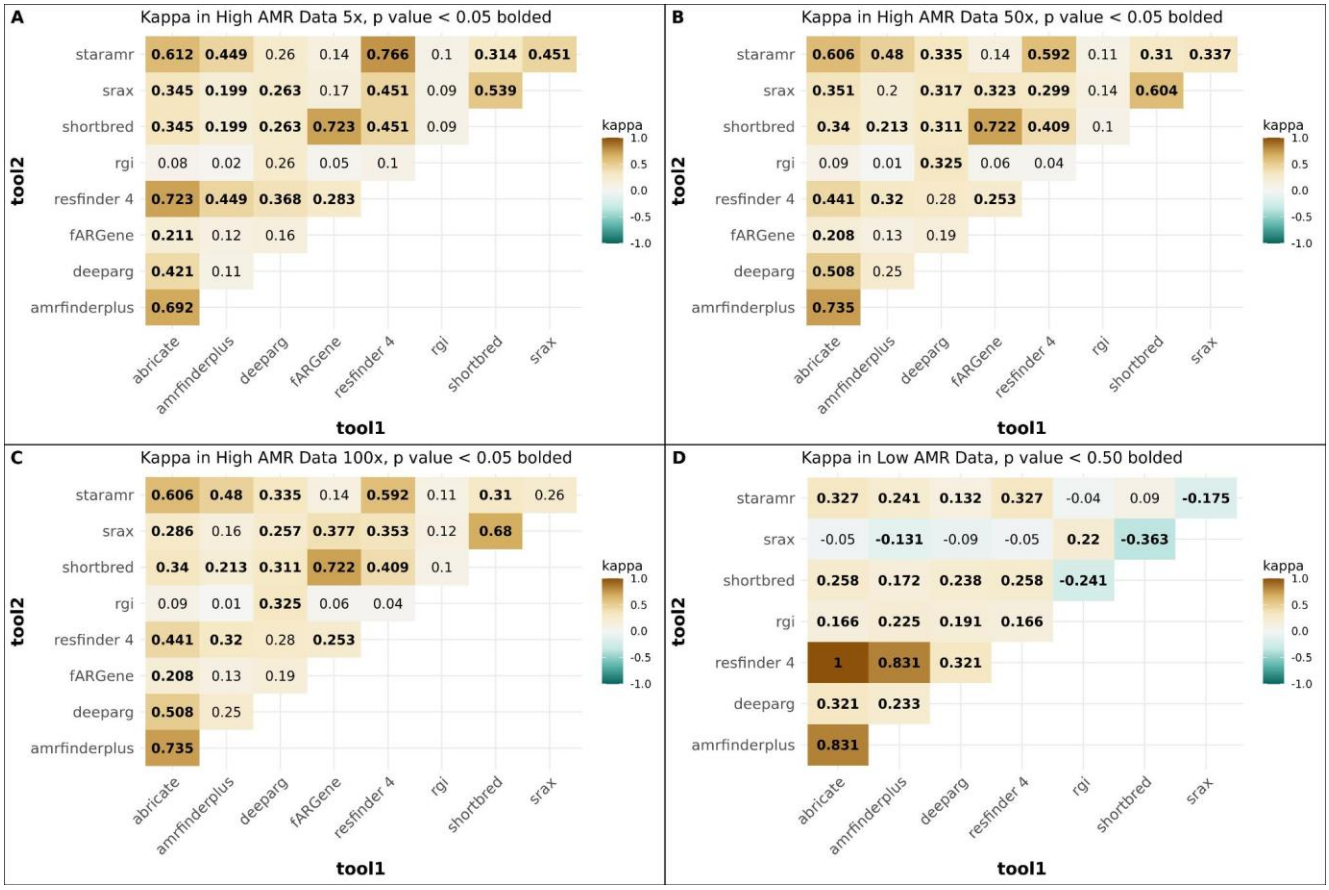


Table 1A: Clinical isolates included in the high resistance simulated community.

Strain	Testing Standard (CLSI or EUCAST)	BioSample ID	Link
Neisseria gonorrhoeae SW0011	CLSI	SAMN15960549	https://www.ncbi.nlm.nih.gov/biosample/SAMN15960549
Klebsiella pneumoniae CCUG 70742	EUCAST	SAMN07602587	https://www.ncbi.nlm.nih.gov/biosample/SAMN07602587
Pseudomonas aeruginosa CCUG 70744	EUCAST	SAMN07602569	https://www.ncbi.nlm.nih.gov/biosample/SAMN07602569 /
Acinetobacter baumannii MRSN489669	CLSI	SAMN12087686	https://www.ncbi.nlm.nih.gov/biosample/SAMN12087686
Enterobacter cloacae 174	CLSI	SAMN04456586	https://www.ncbi.nlm.nih.gov/biosample/SAMN04456586
Citrobacter freundii MRSN12115	CLSI	SAMN13412315	https://www.ncbi.nlm.nih.gov/biosample/SAMN13412315

Staphylococcus aureus LAC	CLSI	SAMN08391108	https://www.ncbi.nlm.nih.gov/biosample/SAMN08391108
Escherichia coli 222	CLSI	SAMN05194390	https://www.ncbi.nlm.nih.gov/biosample/SAMN05194390

Table 1B: Clinical isolates included in the low resistance simulated community.

Strain	Testing Standard (CLSI or EUCAST)	BioSample ID	Link
Staphylococcus aureus	EUCAST	SAMN25295985	https://www.ncbi.nlm.nih.gov/biosample/25295985
Neisseria gonorrhoeae	CLSI	SAMN22824038	https://www.ncbi.nlm.nih.gov/biosample/22824038

Table 2: Tools identified from search methods with the selection criteria and whether they subsequently worked or not.

Tool	Conda / Docker Installable?	Actively Maintained?	Input format?	Included in Analysis?	Implementation Method	Database
ABRIcate	Yes - conda	Yes	FASTA	Yes	Align reads to specified database	NCBI (default), AMRFinder Plus, CARD, ResFinder, ARG-ANNOT, MEGARES, EcOH, PlasmidFinder, VFDB, and Ecoli_VF
shortBRED	Yes - docker & conda	Yes	FASTA	Yes	Align reads to database	AMR gene marker database from 849 AR protein families from the ARDB19 and independent curation

fARGene	Yes - conda	Yes	FASTQ	Yes	Compare to AMR model	Hidden markov models for quinolone, tetracycline, and beta lactamases
RGI	Yes -docker (conda outdated)	Yes	FASTQ	Yes	Compare to AMR model	Prodigal predicts ORF and compared to CARD and WildCARD
ResFinder 4	Yes - docker (conda broken)	Yes	FASTA	Yes	Align reads to database	ResFinder 4 database
DeepARG	Yes - docker	Unclear	FASTA	Yes	Compare to AMR model	Supervised deep learning compares reads to antibiotic resistance categories created from CARD, ARDB, and UNIPROT
sraX	Yes - docker &	Yes	FASTA	Yes	Align reads to database	CARD by default

	conda					
starAMR	Yes - conda	Yes	FASTA	Yes	Align reads to database	ResFinder, PointFinder, and PlasmidFinder
AMR Finder Plus	Yes - conda	Yes	FASTA	Yes	Align reads to database	Pathogen Detection Reference Gene Database
ResPipe	No	Yes	FASTQ or BAM	No		
PointFinder	Yes - docker	Yes	FASTA	No		
PCM: Pairwise Comparative Modelling	No	Yes	FASTA - protein	No		
SRST2	No	No	FASTQ	No		
Arg_Ranker	Yes - conda	Yes	Requires special metadata input	No		
MetaCherchant	Yes - conda	No	FASTA -	No		

			genomic			
ARIBA	Yes - docker	No	Paired end FASTQ	No		
ARG-ANNOT	No	No	Unclear	No		
kmerresistance	No	No	-	No		
c-sstar	No	No	Unknown	No - could not track down github		

Table 3A: Summary Statistics for the high resistance data from hAMRoaster: These are the counts and metrics as calculated by the hAMRoaster pipeline. Formulas for all metrics are as follows:

Specificity = TN / (TN + FP)

Sensitivity = TP / (TP + FN)

Precision = TP / (TP + FP)

Accuracy = (TP + TN) / (TP + FP + TN + FN)

Proportion Unknown = unknown / (TP + FP + unknowns)

High Resistance Data, 100x Coverage										
tool	**False positive**	**True positive**	**unknown**	**False negative**	**True negative**	**sensitivity**	**specificity**	**precision**	**accuracy**	**Proportion Unknown**
abricate	0	66	22	9	2	0.8800	1.0000	1.0000	0.8831	0.2500
amrfinderplus	2	62	71	9	1	0.8732	0.3333	0.9688	0.8514	0.5259
deeparg	0	98	23	8	2	0.9245	1.0000	1.0000	0.9259	0.1901
fARGene	0	713	0	13	2	0.9821	1.0000	1.0000	0.9821	0.0000
resfinder4	1	43	15	9	1	0.8269	0.5000	0.9773	0.8148	0.2542
rgi	4	559	255	6	1	0.9894	0.2000	0.9929	0.9825	0.3117
shortbred	0	29	0	11	2	0.7250	1.0000	1.0000	0.7381	0.0000
srax	0	10	3	11	2	0.4762	1.0000	1.0000	0.5217	0.2308
staramr	1	52	11	9	1	0.8525	0.5000	0.9811	0.8413	0.1719
High Resistance Data, 50x Coverage										

tool	False positive	True positive	unknown	False negative	True negative	sensitivity	specificity	precision	accuracy	Proportion Unknown
abricate	0	66	21	9	2	0.8800	1.0000	1.0000	0.8831	0.2414
amrfinderplus	2	62	67	9	1	0.8732	0.3333	0.9688	0.8514	0.5115
deeparg	0	99	23	8	2	0.9252	1.0000	1.0000	0.9266	0.1885
fARGene	0	702	0	13	2	0.9818	1.0000	1.0000	0.9819	0.0000
resfinder4	1	43	15	9	1	0.8269	0.5000	0.9773	0.8148	0.2542
rgi	4	557	254	6	1	0.9893	0.2000	0.9929	0.9824	0.3117
shortbred	0	30	0	11	2	0.7317	1.0000	1.0000	0.7442	0.0000
srax	0	13	3	10	2	0.5652	1.0000	1.0000	0.6000	0.1875
staramr	1	52	11	9	1	0.8525	0.5000	0.9811	0.8413	0.1719

High Resistance Data, 5x Coverage										
tool	**False positive**	**True positive**	**unknown**	**False negative**	**True negative**	**sensitivity**	**specificity**	**precision**	**accuracy**	**Proportion Unknown**
abricate	0	9	39	19	2	0.8125	1.0000	1.0000	0.8200	0.3276
amrfinderplus	1	9	60	58	1	0.8696	0.5000	0.9836	0.8592	0.4874
deeparg	0	8	267	86	2	0.9709	1.0000	1.0000	0.9711	0.2436
fARGene	0	13	470	0	2	0.9731	1.0000	1.0000	0.9732	0.0000
resfinder4	0	9	43	10	2	0.8269	1.0000	1.0000	0.8333	0.1887
rgi	12	6	1015	418	1	0.9941	0.0769	0.9883	0.9826	0.2893
shortbred	0	11	29	0	2	0.7250	1.0000	1.0000	0.7381	0.0000
srax	0	12	4	3	2	0.2500	1.0000	1.0000	0.3333	0.4286
staramr	0	9	44	11	2	0.8302	1.0000	1.0000	0.8364	0.2000

Table 3B: Summary Statistics for the low resistance data from hAMRoaster: These are the counts and metrics as calculated by the hAMRoaster pipeline.

input_file_name	AMR Isolate	toool	True Positive	False Positive	Unknown	True Negative	False Negative	Sensitivity	Specificity	Precision	Accuracy	Proportion Unknown
samp0_srr1789825	SRR17789825	amrfinderplus	0	0	4	5	4	1.0000	0.0000	0.0000	0.5556	1.0000
samp0_srr1789825	SRR17789826	deeparg	0	0	7	5	4	1.0000	0.0000	0.0000	0.5556	1.0000
samp0_srr1789825	SRR17789827	resfinder4	0	0	1	5	4	1.0000	0.0000	0.0000	0.5556	1.0000
samp0_srr1789825	SRR17789828	rgi	0	0	22	5	4	1.0000	0.0000	0.0000	0.5556	1.0000
samp0_srr17	SRR17789	shortbred	0	0	8	5	4	1.000	0.000	0.00	0.55	1.0000

789825	9829							0	0	00	56	
samp0_srr17789825	SRR17789830	srax	0	0	6	5	4	1.0000	0.0000	0.0000	0.5556	1.0000
samp1_srr16683675	SRR16683675	amrfinderplus	1	0	2	4	4	1.0000	0.2000	1.0000	0.5556	0.6667
samp1_srr16683675	SRR16683676	deeparg	3	1	2	4	4	0.8000	0.4286	0.7500	0.5833	0.3333
samp1_srr16683675	SRR16683677	resfinder4	1	0	2	4	4	1.0000	0.2000	1.0000	0.5556	0.6667
samp1_srr16683675	SRR16683678	rgi	1	6	7	4	4	0.4000	0.2000	0.1429	0.3333	0.5000
samp1_srr16683675	SRR16683679	shortbred	0	0	4	4	4	1.0000	0.0000	0.0000	0.5000	1.0000
samp1_srr16683675	SRR16683680	srax	0	1	5	4	4	0.8000	0.0000	0.0000	0.4444	0.8333

samp1_srr16683675	SRR16683681	staramr	2	0	0	4	4	1.0000	0.3333	1.0000	0.6000	0.0000
samp2_srr17789825	SRR17789830	amrfinderplus	0	0	4	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp2_srr17789825	SRR17789831	deeparg	0	0	6	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp2_srr17789825	SRR17789832	resfinder4	0	0	1	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp2_srr17789825	SRR17789833	rgi	0	0	22	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp2_srr17789825	SRR17789834	shortbred	0	0	4	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp2_srr17789825	SRR17789835	srax	0	0	3	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp3_srr16	SRR1668	amrfinder	1	0	2	4	4	1.000	0.200	1.00	0.55	0.6667

683675	3675	plus						0	0	00	56	
samp3_srr16683675	SRR16683676	deeparg	3	0	3	4	4	1.0000	0.4286	1.0000	0.6364	0.5000
samp3_srr16683675	SRR16683677	resfinder4	1	0	2	4	4	1.0000	0.2000	1.0000	0.5556	0.6667
samp3_srr16683675	SRR16683678	rgi	1	0	13	4	4	1.0000	0.2000	1.0000	0.5556	0.9286
samp3_srr16683675	SRR16683679	shortbred	2	0	9	4	4	1.0000	0.3333	1.0000	0.6000	0.8182
samp3_srr16683675	SRR16683680	srax	0	0	5	4	4	1.0000	0.0000	0.0000	0.5000	1.0000
samp3_srr16683675	SRR16683681	staramr	2	0	0	4	4	1.0000	0.3333	1.0000	0.6000	0.0000
samp4_srr17789825	SRR17789830	amrfinderplus	0	0	4	12	4	1.0000	0.0000	0.0000	0.7500	1.0000

samp4_srr17789825	SRR17789831	deeparg	0	0	6	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp4_srr17789825	SRR17789832	resfinder4	0	0	1	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp4_srr17789825	SRR17789833	rgi	0	0	22	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp4_srr17789825	SRR17789834	shortbred	0	0	7	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp4_srr17789825	SRR17789835	srax	0	0	6	12	4	1.0000	0.0000	0.0000	0.7500	1.0000
samp5_srr16683675	SRR16683675	amrfinderplus	1	0	2	4	4	1.0000	0.2000	1.0000	0.5556	0.6667
samp5_srr16683675	SRR16683676	deeparg	3	1	2	4	4	0.8000	0.4286	0.7500	0.5833	0.3333
samp5_srr16	SRR1668	resfinder	1	0	2	4	4	1.000	0.200	1.00	0.55	0.6667

683675	3677	4						0	0	00	56	
samp5_srr16683675	SRR16683678	rgi	1	6	7	4	4	0.4000	0.2000	0.1429	0.3333	0.5000
samp5_srr16683675	SRR16683679	shortbred	0	1	5	4	4	0.8000	0.0000	0.00	0.4444	0.8333
samp5_srr16683675	SRR16683680	srax	1	1	3	4	4	0.8000	0.2000	0.5000	0.5000	0.6000
samp5_srr16683675	SRR16683681	staramr	2	0	0	4	4	1.0000	0.3333	1.00	0.6000	0.0000

Supplementary text 1: URL link to tweet

https://twitter.com/emily_wissel/status/1336013892116488195

Supplementary table 1: access link to tidy table of data

https://docs.google.com/spreadsheets/d/1bfACqEh0nkS65vCUj5DfMg4PvW0fHxbtrv0PgKt1gT4/edit#gid=53644837

References

Abreu, Vinicius A. C. de, José Perdigão, and Sintia Almeida. 2021. “Metagenomic Approaches to Analyze Antimicrobial Resistance: An Overview.” *Frontiers in Genetics* 11. https://www.frontiersin.org/articles/10.3389/fgene.2020.575592.

Alcock, Brian P., Amogelang R. Raphenya, Tammy T. Y. Lau, Kara K. Tsang, Mégane Bouchard, Arman Edalatmand, William Huynh, et al. 2020. “CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database.” *Nucleic Acid Research* 48, no. D1 (January): D517–D525. https://doi.org/10.1093/nar/gkz935.

Anjum, Muna F., Ea Zankari, and Henrik Hasman. 2017. “Molecular Methods for Detection of Antimicrobial Resistance.” *Microbiology Spectrum* 5, no. 6 (December). https://doi.org/10.1128/microbiolspec.ARBA-0011-2017.

Arango-Argoty, Gustavo, Emily Garner, Amy Pruden, Lenwood S. Heath, Peter Vikesland, and Liqing Zhang. 2018. “DeepARG: A Deep Learning Approach for Predicting Antibiotic Resistance Genes from Metagenomic Data.” *Microbiome* 6, no. 1 (December): 23. https://doi.org/10.1186/s40168-018-0401-z.

Barrett, T., K. Clark, R. Gevorgyan, V. Gorelenkov, E. Gribov, I. Karsch-Mizrachi, M. Kimelman, et al. 2012. “BioProject and BioSample Databases at NCBI: Facilitating Capture and Organization of Metadata.” *Nucleic Acids Research* 40, no. D1 (January): D57–63. https://doi.org/10.1093/nar/gkr1163.

Berglund, Fanny, Tobias Österlund, Fredrik Boulund, Nachiket P. Marathe, D. G. Joakim Larsson, and Erik Kristiansson. 2019. “Identification and Reconstruction of Novel Antibiotic Resistance Genes from Metagenomes.” *Microbiome* 7, no. 1 (December): 52. https://doi.org/10.1186/s40168-019-0670-1.

Biggs, Sharon L., Amy V. Jennison, Haakon Bergh, Rikki Graham, Graeme Nimmo, and David Whiley. 2022. “Limited Evidence of Patient-to-Patient Transmission of Staphylococcus Aureus Strains between Children with Cystic Fibrosis, Queensland, Australia.” Edited by Herminia De Lencastre. *PLOS ONE* 17, no. 10 (October): e0275256. https://doi.org/10.1371/journal.pone.0275256.

Bortolaia, Valeria, Rolf S Kaas, Etienne Ruppe, Marilyn C Roberts, Stefan Schwarz, Vincent Cattoir, Alain Philippon, et al. 2020. “ResFinder 4.0 for Predictions of Phenotypes from Genotypes.” *Journal of Antimicrobial Chemotherapy* 75, no. 12 (December): 3491–3500. https://doi.org/10.1093/jac/dkaa345.

Brinkac, Lauren, Alexander Voorhies, Andres Gomez, and Karen E. Nelson. 2017. “The Threat of Antimicrobial Resistance on the Human Microbiome.” *Microbial Ecology* 74, no. 4 (November): 1001–8. https://doi.org/10.1007/s00248-017-0985-z.

Buchfink, Benjamin, Klaus Reuter, and Hajk-Georg Drost. 2021. “Sensitive Protein Alignments at Tree-of-Life Scale Using DIAMOND.” *Nature Methods* 18, no. 4 (April): 366–68. https://doi.org/10.1038/s41592-021-01101-x.

Camacho, Christiam, George Coulouris, Vahram Avagyan, Ning Ma, Jason Papadopoulos, Kevin Bealer, and Thomas L Madden. 2009. “BLAST+: Architecture and Applications.” *BMC Bioinformatics* 10, no. 1 (December): 421. https://doi.org/10.1186/1471-2105-10-421.

Carattoli, Alessandra and Henrik Hasman. 2019. “PlasmidFinder and In Silico pMLST: Identification and Typing of Plasmid Replicons in Whole-Genome Sequencing (WGS).” *Horizontal Gene Transfer, Methods in Microbiology* 2075 (October): 285-94. DOI: 10.1007/978-1-4939-9877-7_20.

Chen, Lihong, Dandan Zheng, Bo Liu, Jian Yang, and Qi Jin. 2016. “VFDB 2016: Hierarchical and Refined Dataset for Big Data Analysis—10 Years On.” *Nucleic Acids Research* 44, no. D1 (January): D694–97. https://doi.org/10.1093/nar/gkv1239.

Doster, Enrique, Steven M Lakin, Christopher J Dean, Cory Wolfe, Jared G Young, Christina Boucher, Keith E Belk, Noelle R Noyes, and Paul S Morley. 2020. “MEGARes 2.0: A Database for Classification of Antimicrobial Drug, Biocide and Metal Resistance Determinants in Metagenomic Sequence Data.” *Nucleic Acids Research* 48, no. D1 (January): D561–69. https://doi.org/10.1093/nar/gkz1010.

Eloe-Fadrosh, Emiley A., Faiza Ahmed, Anubhav, Michal Babinski, Jeffrey Baumes, Mark Borkum, Lisa Bramer, et al. 2022. “The National Microbiome Data Collaborative Data Portal: an integrated multi-omics microbiome data resource.” *Nucleic Acids Research* 50, no. D1 (January): D828–D836. https://doi.org/10.1093/nar/gkab990.

Fannyhb. 2019. “fARGene.” GitHub. Accessed April 5, 2023. https://github.com/fannyhb/fargene.

Fritz, Adrian, Peter Hofmann, Stephan Majda, Eik Dahms, Johannes Dröge, Jessika Fiedler, Till R. Lesker, et al. 2019. “CAMISIM: Simulating Metagenomes and Microbial Communities.” *Microbiome* 7, no. 1 (December): 17. https://doi.org/10.1186/s40168-019-0633-6.

Gamer, Matthias, Jim Lemon, Ian Fellows, and Puspendra Singh. 2022. “Package ‘irr’: Various Coefficients of Interrater Reliability and Agreement.” Accessed April 4, 2023. http://bioconductor.statistik.tu-dortmund.de/cran/web/packages/irr/irr.pdf.

Grüning, Björn, Ryan Dale, Andreas Sjödin, Brad A. Chapman, Jillian Rowe, Christopher H. Tomkins-Tinch, Renan Valieris, and Johannes Köster. 2018. “Bioconda: Sustainable and Comprehensive Software Distribution for the Life Sciences.” *Nature Methods* 15, no. 7 (July): 475–76. https://doi.org/10.1038/s41592-018-0046-7.

Gupta, Sushim Kumar, Babu Roshan Padmanabhan, Seydina M. Diene, Rafael Lopez-Rojas, Marie Kempf, Luce Landraud, and Jean-Marc Rolain. 2014. "ARG-ANNOT, a New Bioinformatic Tool to Discover Antibiotic Resistance Genes in Bacterial Genomes." *Antimicrobial Agents and Chemotherapy* 58, no. 1 (January): 212–20. https://doi.org/10.1128/AAC.01310-13.

Hastings, Janna, Gareth Owen, Adriano Dekker, Marcus Ennis, Namrata Kale, Venkatesh Muthukrishnan, Steve Turner, Neil Swainston, Pedro Mendes, and Christoph Steinbeck. 2016. "ChEBI in 2016: Improved services and an expanding collection of metabolites." *Nucleic Acids Research* 44, no. D1 (January): D1214–D1219. https://doi.org/10.1093/nar/gkv1031.

Hendriksen, Rene S., Valeria Bortolaia, Heather Tate, Gregory H. Tyson, Frank M. Aarestrup, and Patrick F. McDermott. 2019. "Using Genomics to Track Global Antimicrobial Resistance." *Frontiers in Public Health* 7 (September): 242. https://doi.org/10.3389/fpubh.2019.00242.

Huang, Weichun, Leping Li, Jason R. Myers, and Gabor T. Marth. 2012. "ART: A Next-Generation Sequencing Read Simulator." *Bioinformatics* 28, no. 4 (February): 593–94. https://doi.org/10.1093/bioinformatics/btr708.

Human Microbiome Project Consortium. 2012. "Structure, function and diversity of the healthy human microbiome." *Nature* 486 (June): 207-14. https://doi.org/10.1038/nature11234.

Hyatt, Doug, Gwo-Liang Chen, Philip F LoCascio, Miriam L Land, Frank W Larimer, and Loren J Hauser. 2010. "Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification." *BMC Bioinformatics* 11, no. 1 (December): 119. https://doi.org/10.1186/1471-2105-11-119.

Ingle, Danielle J., Mary Valcanis, Alex Kuzevski, Marija Tauschek, Michael Inouye, Tim Stinear, Myron M. Levine, Roy M. Robins-Browne, and Kathryn E. Holt. 2016. "In Silico Serotyping of E. Coli from Short Read Data Identifies Limited Novel O-Loci but Extensive Diversity of O:H Serotype Combinations within and between Pathogenic Lineages." *Microbial Genomics* 2, no. 7 (July). https://doi.org/10.1099/mgen.0.000064.

Kaminski, James, Molly K. Gibson, Eric A. Franzosa, Nicola Segata, Gautam Dantas, and Curtis Huttenhower. 2015. "High-Specificity Targeted Functional Profiling in Microbial Communities with ShortBRED." Edited by William Stafford Noble. *PLOS Computational Biology* 11, no. 12 (December): e1004557. https://doi.org/10.1371/journal.pcbi.1004557.

Kraemer, Susanne A., Arthi Ramachandran, and Gabriel G. Perron. 2019. "Antibiotic Pollution in the Environment: From Microbial Ecology to Public Policy." *Microorganisms* 7, no. 6 (June): 180. https://doi.org/10.3390/microorganisms7060180.

Kumar, Deepak, Suchawan Pornsukarom, and Siddhartha Thakur. 2019. “Antibiotic Usage in Poultry Production and Antimicrobial-Resistant Salmonella in Poultry.” In *Food Safety in Poultry Meat Production*, edited by Kumar Venkitanarayanan, Siddhartha Thakur, and Steven C. Ricke, 47–66. Cham: Springer International Publishing, https://doi.org/10.1007/978-3-030-05011-5_3.

Liu, Bo and Mihai Pop. 2009. “ARDB—Antibiotic Resistance Genes Database.” *Nucleic Acids Research* 37, no. 1 (January): D443–D447. https://doi.org/10.1093/nar/gkn656.

McGinnis, Scott and Thomas L. Madden. 2004. “BLAST: at the core of a powerful and diverse set of sequence analysis tools.” *Nucleic Acids Research* 32, no. 1 (July): W20-25. https://doi.org/10.1093/nar/gkh435.

McIver, Lauren J, Galeb Abu-Ali, Eric A Franzosa, Randall Schwager, Xochitl C Morgan, Levi Waldron, Nicola Segata, and Curtis Huttenhower. 2018. “BioBakery: A Meta’omic Analysis Environment.” Edited by John Hancock. *Bioinformatics* 34, no. 7 (April): 1235–37. https://doi.org/10.1093/bioinformatics/btx754.

National Center for Biotechnology Information. 2021. “WGS of Neisseria gonorrhoeae: DDD027 (SRR16683675).” Sequence Read Archive. Accessed April 4, 2023. https://trace.ncbi.nlm.nih.gov/Traces/?view=run_browser&acc=SRR16683675&display=metadata.

National Center for Biotechnology Information. 2023. “Bacterial Antimicrobial Resistance Reference Gene Database.” Reference Gene Catalog. Accessed April 5, 2023. https://www.ncbi.nlm.nih.gov/pathogens/refgene/#.

National Center for Biotechnology Information. 2023. “GenBank and WGS Statistics.” Genbank. Accessed April 5, 2023. https://www.ncbi.nlm.nih.gov/genbank/statistics/.

National Center for Biotechnology Information. 2023. “NCBI Antimicrobial Resistance Gene Finder (AMRFinderPlus).” GitHub. Accessed April 5, 2023. https://github.com/ncbi/amr.

National Microbiology Laboratory. 2017. “Escherichia coli virulence factors.” GitHub. Accessed April 4, 2023. https://github.com/phac-nml/ecoli_vf.

National Microbiology Laboratory. 2022. “staramr.” GitHub. Accessed April 5, 2023. https://github.com/phac-nml/staramr.

Nji, Emmanuel, Joseph Kazibwe, Thomas Hambridge, Carolyn Alia Joko, Amma Aboagyewa Larbi, Lois Afua Okyerewaa Damptey, Nana Adoma Nkansa-Gyamfi, Cecilia Stålsby Lundborg, and La Thi Quynh Lien. 2021. “High Prevalence of Antibiotic Resistance in Commensal Escherichia Coli from Healthy Human Sources in Community Settings.” *Scientific Reports* 11, no. 1 (February): 3372. https://doi.org/10.1038/s41598-021-82693-4.

Nurk, Sergey, Dmitry Meleshko, Anton Korobeynikov, and Pavel A. Pevzner. 2017. "MetaSPAdes: A New Versatile Metagenomic Assembler." *Genome Research* 27, no. 5 (May): 824–34. https://doi.org/10.1101/gr.213959.116.

Oniciuc, Elena, Eleni Likotrafiti, Adrián Alvarez-Molina, Miguel Prieto, Jesús Santos, and Avelino Alvarez-Ordóñez. 2018. "The Present and Future of Whole Genome Sequencing (WGS) and Whole Metagenome Sequencing (WMS) for Surveillance of Antimicrobial Resistant Microorganisms and Antimicrobial Resistance Genes across the Food Chain." *Genes* 9, no. 5 (May): 268. https://doi.org/10.3390/genes9050268.

Panunzi, Leonardo G. 2020. "SraX: A Novel Comprehensive Resistome Analysis Tool." *Frontiers in Microbiology* 11 (February): 52. https://doi.org/10.3389/fmicb.2020.00052.

Porter, Teresita M. and Mehrdad Hajibabaei. 2018. "Over 2.5 Million COI Sequences in GenBank and Growing." Edited by Wolfgang Arthofer. *PLOS ONE* 13, no. 9 (September): e0200177. https://doi.org/10.1371/journal.pone.0200177.

Public Health Alliance for Genomic Epidemiology. 2022. "hAMRonization." GitHub. Accessed April 4, 2023. https://github.com/pha4ge/hAMRonization.

Robinson, Tony, Jim Harkin, and Priyank Shukla. 2021. "Hardware Acceleration of Genomics Data Analysis: Challenges and Opportunities." *Bioinformatics* 37, no. 13 (July): 1785–95. https://doi.org/10.1093/bioinformatics/btab017.

Seemann, Tortsen. 2020. "ABRicate." GitHub. Accessed April 5, 2023. https://github.com/tseemann/abricate.

Shao, Yitian, Yiping Wang, Yiwen Yuan, and Yujing Xie. 2021. "A Systematic Review on Antibiotics Misuse in Livestock and Aquaculture and Regulation Implications in China." *Science of The Total Environment* 798 (December): 149205. https://doi.org/10.1016/j.scitotenv.2021.149205.

Teillant, Aude, Sumanth Gandra, Devra Barter, Daniel J Morgan, and Ramanan Laxminarayan. 2015. "Potential Burden of Antibiotic Resistance on Surgery and Cancer Chemotherapy Antibiotic Prophylaxis in the USA: A Literature Review and Modelling Study." *The Lancet Infectious Diseases* 15, no. 12 (December): 1429–37. https://doi.org/10.1016/S1473-3099(15)00270-4.

Toni Poole and Cynthia Sheffield. 2013. "Use and Misuse of Antimicrobial Drugs in Poultry and Livestock: Mechanisms of Antimicrobial Resistance." *Pakistan Veterinary Journal* 33, no. 3: 266–71. https://handle.nal.usda.gov/10113/58191.

Underwood, Anthony. 2021. "chAMReDb." GitLab. Accessed April 5, 2023. https://gitlab.com/antunderwood/chamredb.

Wilkinson, Mark D., Michel Dumontier, IJsbrand Jan Aalbersberg, Gabrielle Appleton, Myles Axton, Arie Baak, Niklas Blomberg, et al. 2016. “The FAIR Guiding Principles for Scientific Data Management and Stewardship.” *Scientific Data* 3, no. 1 (March): 160018. https://doi.org/10.1038/sdata.2016.18.

Wissel, Emily. 2022. “hAMRoaster.” GitHub. Accessed April 5, 2023. https://github.com/ewissel/hAMRoaster.

Zankari, Ea, H. Hasman, S. Cosentino, M. Vestergaard, S. Rasmussen, O. Lund, F. M. Aarestrup, and M. V. Larsen. 2012. “Identification of Acquired Antimicrobial Resistance Genes.” *Journal of Antimicrobial Chemotherapy* 67, no. 11 (November): 2640–44. https://doi.org/10.1093/jac/dks261.

Zankari, Ea, Rosa Allesøe, Katrine G. Joensen, Lina M, Cavaco, Ole Lund, and Frank M. Aarestrup. 2017. “PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens.” *Journal of Antimicrobial Chemotherapy* 72, no. 10 (October): 2764–68. https://doi.org/10.1093/jac/dkx217.

Chapter 4

Frequency of Antimicrobial Resistance Genes in a Pregnancy Cohort for People with and without Urogenital Infections during Pregnancy

Abstract

Background: Understanding how the collection of AMR genes during pregnancy changes over time is important because antimicrobials are commonly administered during pregnancy for infections such as UTIs, BV, and chlamydia.

Methods: A subset of the data from the Emory University African American Vaginal, Oral, and Gut Microbiome in Pregnancy Cohort Study were sent for metagenomic sequencing (238 patients, rectal and vaginal swabs at 8-14 weeks & 24 - 30 weeks pregnancy). The taxonomic assignment was done with metaphlan2, and AMR genes were detected with AMR Finder Plus. An ANOVA was used to detected significant differences in alpha diversity while a chi squared test was used to detect differences in the frequencies of AMR genes in those with urogenital infections (BV, UTI, chlamydia), those exposed to antimicrobials for other reasons, and those not exposed to any antimicrobials during their pregnancy.

Results: Individuals who had urogenital infections tend to have more antimicrobial resistance (AMR) genes compared to those without exposure to antimicrobials, although there was no significant difference in alpha diversity of the gut and vaginal microbiome across groups. The gut microbiome in early pregnancy had significantly more copper and copper/silver resistance genes in those with urogenital infections, though there was no difference in the AMR from the gut microbiome across groups in later pregnancy. The vaginal microbiome had significantly more virulence genes in those with urogenital infections in later pregnancy but not in early

pregnancy. Overall, this study helps understand the dynamics in AMR genes across pregnancies for those with and without urogenital infections and antimicrobial exposure.

Keywords: antimicrobial resistance genes, pregnancy, UTI, BV, chlamydia, metagenome, microbiome

Introduction

Antimicrobial resistance (AMR) is a global health issue in which previously treatable pathogens become resistant to drugs, which can lead to death in the case of multiple-drug resistance infections and increase the cost on healthcare systems. In the United States, approximately one in four individuals are prescribed antibiotics during their pregnancy (such as to treat STIs, UTIs, etc.) (Bookstaver et al. 2015), and individuals undergoing C-section or who have group B streptococcus often receive a continual dose of antibiotics during delivery to reduce risk of infection (Seedat et al. 2017). Administering antibiotics is often an essential life-saving therapy, particularly when treating an infection. In nonpregnant populations, antimicrobial therapies are known to increase AMR genes which increases the risk for drug resistant infections (DeLong et al. 2021; Ma, Forney, and Ravel 2012). It is important to study how antimicrobial therapy impacts microbial composition and AMR genes during pregnancy because antimicrobials may negatively impact the changing microbiome in a manner which limits the microbiome's ability to support a healthy pregnancy. These AMR genes could also be passed to the newborn during birth, creating a generational risk for drug-resistant infections.

Understanding the distribution of antibiotic resistant genes (AMR) during pregnancy is important for several reasons. First, pregnant people are at increased risk of developing certain infections, or certain infections may pose higher risk and timely treatment becomes more important. For example, urinary tract infections (UTIs) become more common during pregnancy as hormones and the immune system change alongside pregnancy development. Bacterial vaginosis (BV) during pregnancy can increase the risk for multiple adverse outcomes, such as preterm birth and amniotic fluid infection (McGregor and French, 2000). Chlamydia infections are also routinely screened at pregnancy onset and in the third trimester because of the wide

array of adverse outcomes it can lead to, including preterm birth and stillbirth (He et al. 2020, Majeroni and Ukkadam 2007). If the bacteria causing the infection are antibiotic resistant, treatment may be more difficult, less effective, and there is a greater probability that the infection can have a negative effect on the pregnant person or fetus. Secondly, the transmission of antibiotic resistant genes from parent to baby during birth can lead to the development of antibiotic resistant infections in the newborn. While the presence of AMR genes does not automatically mean there will be AMR phenotypes, AMR genes can give rise to antibiotic resistant infections under the right conditions, which poses undue risk for newborns. Notably, some other classes of genes such as genes which increase pathogen virulence, called virulence genes, and stress genes, which decide resistance genes to non-antimicrobials such as metal or heat, can be elevated and important for understanding gene dynamics during urogenital infection.

Understanding AMR gene frequency during pregnancy for those with urogenital infections compared to those with no antibiotic exposure during pregnancy can inform what type of AMR genes tend to be present in "typical" healthy pregnancies, and how urogenital infection and treatment impacts the collection of AMR genes. To this aim, we described the AMR genes present in the subset of data from the Emory University African American Pregnancy Cohort that had metagenomic data available (Corwin et al. 2017). Antimicrobials are commonly administered to treat Bacterial Vaginosis (BV), Urinary Tract Infection (UTI), and chlamydia infection in this cohort (n = 57, 46, 30 respectively). Particular attention will be paid to antibiotics nitrofurantoin, metronidazole, and zithromax administered to treat the three most common infections in this cohort - UTIs, BV, and chlamydia infection. We hypothesized that those who had one of these three urogenital infections would have a higher total number and higher diversity of AMR genes in their gut and vaginal microbiome than those with no reported

antibiotic/antifungal exposure and those with antibiotic/antifungal exposure but none of these infections during their pregnancy.

Methods

The study is part of the Emory University African American Vaginal, Oral, and Gut Microbiome in Pregnancy Cohort Study (Corwin et al. 2017). Briefly, rectal, and vaginal swabs were self-collected at two time points during pregnancy (8-14 weeks & 24 - 30 weeks). A total of 814 samples from 238 pregnant individuals who provided swabs at both time points were sent for metagenomic sequencing. Patients self-reported whether they took any antibiotics, antifungals, or antivirals within one month prior to the first sample time point and anytime between the two sampling timepoints. Patients also consented to a medical chart review. Swabs were sent to Omega Bioservices Inc. for metagenome sequencing with NextSeq. FASTQ files were downloaded using BASESPACE and assembled into contigs with megahit (Li et al. 2015). Quality filtering on the FASTQ files was done with FASTQC (Babraham Bioinformatics, n.d.), and Trimmaomatic (Bolger, Lohse, and Usadel 2014) was used to remove adaptor sequences. Human contaminant reads were removed with kneaddata (Huttenhower Lab, n.d.), and taxonomy was assigned from FASTQs using metaphlan2 (Truong et al. 2015). Alpha diversity was calculated in R using the vegan package (Dixon 2003).

A preliminary study comparing the performance of nine bioinformatics tools for detecting AMR genes was conducted. Specifically, we tested the ability of these tools to detect resistance to metronidazole, azithromycin, and nitrofurantoin. Antibiotics from these drug classes are commonly used to treat bacterial vaginosis (BV), urinary tract infections (UTI), and chlamydia infections during pregnancy, which are common in this cohort (**Table 1**). Specifically, a simulated dataset from clinical isolates with phenotypic resistance to these three drug classes

was used to compare the performance of ShortBRED, fARGene, RGI, ResFinder, abricate, AMRFinder Plus, deepARG, sraX, and starAMR. The clinical isolates used for the simulated data are in **Supplementary Table 1**. Results from all tools were compiled with hamronization (Public Health Alliance for Genomic Epidemiology 2022) and analyzed with hAMRoaster (Wissel et al. 2023) **(Supplementary Table 2**). Overall, only RGI, AMR Finder Plus, and deepARG detected any resistance genes at all. AMR Finder Plus was selected for this analysis as it returned the most realistic results for the drug classes of interest **(Supplementary Table 3),** and it performed well in a general resistance approach (Chapter 3).

Assembled contigs were analyzed for AMR genes with AMR Finder Plus using default settings. Detected genes with less than 80% coverage of the reference sequence and less than 90% identity of the reference sequence were filtered out and excluded from analysis. Drug classes with AMR genes detected in fewer than 10 samples were not included in subsequent analysis to ensure sufficient observations for chi squared analysis.

An ANOVA was used to test for a significant difference in alpha diversity across groups while controlling for sampling time point. To test if there is a significant difference between the frequencies of AMR genes across all drug classes for those who have one of the urogenital infections of interest (chlamydia, UTI, BV; n = 111), those who receive an antibiotic or antifungal for any other reason between timepoints (n = 23), and those who report no antibiotics or antifungals (n = 103), a chi squared test was used. A separate chi squared test was run for each body site and time point. All statistical tests were run, and all figures were created in R version 4.0.4. All code is available on github (https://github.com/ewissel/dissertation_analysis_code).

Results

Mean Shannon alpha diversity was 1.9 (sd = 0.66) for rectal samples and 0.78 (sd = 0.54) for vaginal samples. An ANOVA showed no significant difference in Shannon alpha diversity across antimicrobial exposure groups for both body sites. The distribution of Shannon alpha diversity scores per sample across timepoints is displayed in **Figure One.**

There were a median of 4 AMR genes detected per sample for rectal samples and 3 AMR genes for the vaginal samples, though there was a large degree of variance (range = 1 -166 for rectal samples; 1-30 for vaginal samples). The total number of AMR genes detected per drug class by body site is shown in **Figure Two** (9,417 total AMR genes for rectal samples and 1,831 total AMR genes in vaginal samples). After excluding drug classes which had resistance genes detected in fewer than 10 samples, only the drug classes beta-lactam, macrolide, tetracycline, and virulence remained for analysis in the vaginal samples. Note that virulence is an element type label given by AMR Finder Plus which is substituted for "drug class" when the detected gene codes a virulence factor and not an antimicrobial resistance gene. Virulence factors are those which can help an infection spread more rapidly and cause more damage to the host. As such, it is not surprising that virulence factors are elevated in those with a recent infection. Macrolide, tetracycline, beta-lactam, lincosamide, efflux, arsenic, aminoglycoside, phenicol, mercury, trimethoprim, quaternary ammonium, streptothricin, sulfonamide, copper/silver, copper, glycopeptide, macrolide/lincosamide/streptogramin, virulence, and stress are drug classes included in analysis from the rectal samples. While stress and virulence is an element type label given by AMR Finder Plus when a detected gene does not confer resistance to antibiotics, these element types provide insight into the genetic landscape of the microbiome during these infections and are included in analysis. Two of the antibiotics commonly used to treat urogenital

infections in this cohort, nitrofurantoin and metronidazole, have no observed resistance genes to their drug class after quality filtering.

A chi-squared test was conducted for each time point comparing the infection status (urogenital infection; antibiotics or antifungals for other reasons; no antibiotic/antifungal exposure) to the frequency of AMR genes across drug classes. Note there is not data on when the infection occurred during pregnancy, only that a patient received these diagnoses during pregnancy. For the rectal samples, there was a significant difference in the AMR genes detected across infection status in early pregnancy ($p < 0.01$) but not at the second time point ($p = 0.106$). Specifically, resistance genes for copper and copper/silver drug classes were more frequently detected at the first timepoint in samples from those with one of the three urogenital infections compared to those exposed to antibiotics for other reasons and those not exposed to antibiotics (based on standardized residual values greater than |2|, **Table 3**). Other drug classes with standardized residuals less than |2| are also reported in **Table 3**, though standard practice assumes that a standardized residual less than |2| is not significant (Sharpe 2023).

In the vaginal samples, there was no significant difference in the AMR genes detected across infection status in early pregnancy ($p = 0.547$), but there was a significant difference at the second time point ($p < 0.01$). Specifically, virulence genes were much more frequent in samples from those with urogenital infections than from samples of those exposed to antibiotics/antifungals for other reasons and those not exposed to antibiotics in pregnancy. The other three drug classes had residuals less than |2| across all cells, indicating that the observed frequency of AMR genes in these drug classes was not significantly different across groups (**Table 3**).

Discussion

Overall, we found AMR genes in all samples from both body sites in this cohort. In line with prior literature on nonpregnant populations, even healthy individuals with no recent reported exposure to antimicrobial medicines carry AMR genes in their microbiome (Pereira-Dias et al. 2021; Afridi, Ali, and Chang 2021). While we expected that the diversity of the microbiome would be different between antimicrobial exposure groups, presumably due to effects of broad-spectrum antibiotics and infections on the microbiome, there are no significant differences in alpha diversity across groups at either body sites in this cohort. This could be because the antimicrobial exposure is not necessarily right before the sampling timepoints, so the microbiome has time to recover and stabilize after antimicrobial perturbation. This also implies that there is not a lasting significant difference of the microbiome when infections are present during pregnancy. This is in line with the results from Aim 1 of this dissertation, which found no significant difference in the microbiome for each of the three urogenital conditions at either body site using a linear decomposition model. Overall, this is a reassuring finding as it indicates that both an infection and its subsequent treatment do not significantly impact the gut or vaginal microbiome, so pregnant patients can be reassured that treating these infections with conventional antimicrobial therapy is effective and low risk, particularly when compared to the risk of not receiving timely treatment for an infection.

We found that, while the microbiome was not significantly different, individuals who had urogenital infections tended to have significantly more AMR genes compared to those with no antimicrobial exposure and those who receive antimicrobials for other infections, though only a few drug classes were significantly different at each body site. Interestingly, AMR genes in drug classes with the antibiotics used to treat urogenital infections did not differ significantly across

groups. Most notably, only virulence genes, which increase the ability of a pathogen to infect a host, in the vaginal microbiome were significantly elevated in those with urogenital infections, while conventional AMR genes to other drug classes had no significant difference across groups in the vaginal microbiome. It makes sense that there would be higher virulence genes in the vaginal environment for those who had urogenital infections, as virulence factors tend to be elevated in those with BV and other urogenital infections (Africa, Nel, and Stemmet 2014; Vornhagen, Waldorf, and Rajagopal 2017). Perhaps, as mentioned previously, the microbiome has time to recover after antibiotic exposure and thus the collection of AMR genes also returns to baseline. There is a fitness cost for bacteria associated with holding on to AMR genes (Newbury et al. 2022, Buckner et al. 2018), and though the cost varies across organisms and the specific AMR gene, microbes may not hold on to AMR genes when there is not a selective pressure for those genes. Further, it is presumed that the urogenital infections in this cohort were not drug resistant (i.e., no AMR phenotypes) as the antibiotics administered according to the medical chart data were antibiotics for the associated infection. As such, it is possible that there would not be AMR genes to the drug classes that were administered to treat infections.

The gut microbiome in early pregnancy had no significant differences in frequency of AMR genes across groups, while copper and copper/silver resistance genes were more common in later pregnancy for those with urogenital infections. While the link between copper and silver resistance to urogenital infections is not necessarily obvious, copper resistance genes have been associated with drug-resistance UTIs previously (Saenkham-Huntsinger et al. 2021). Copper is an essential micronutrient for prokaryotes and eukaryotes, and plays a role in mammalian immune systems (Focarelli, Giachino, and Waldron 2022). Specifically, copper will increase in the microenvironment where a pathogen is detected, leading to an increase in copper-resistant

pathogens. Due to the immune response increasing copper in response to urogenital infections, there may be an increased selective pressure for copper resistance genes in the gut microbiome so that commensal microbes do not get impacted by the increased copper levels. Of note is that gut microbiome data is collected from rectal swabs, and the rectum is physiologically close to the urogenital infections.

Future Directions

While this study does not examine the microbiome of newborns, it is known that AMR genes can transfer to newborns from the mother via several routes, including during labor, from breast milk, and via salivary exchange when parents kiss their babies (Patangia et al. 2022). In this study, birthing parents with urogenital infections during pregnancy tend to have more AMR genes in their microbiome (albeit only significantly different at different timepoints), indicating that the newborns of these parents could be at increased risk for harboring AMR genes in their microbiome compared to the newborns of parents not exposed to antimicrobial therapies during pregnancy. However, this is not to imply that antimicrobial treatment is harmful during pregnancy; antimicrobials like antibiotics are administered to treat infections which would otherwise pose an elevated risk to the parents and/or newborn, including preterm birth and death. Further the presence of AMR genes does not translate directly to AMR phenotypes. Rather, understanding the transmission and distribution of AMR genes in newborns' microbiome is the first step in understanding disease landscape and the role of parent-baby AMR transmission in subsequent health outcomes for the infant. Future studies interested in studying AMR gene transmission from parents to newborns should select sampling timepoints that are closer to when transmission would occur (e.g., right before birth, during the breastfeeding period).

Limitations

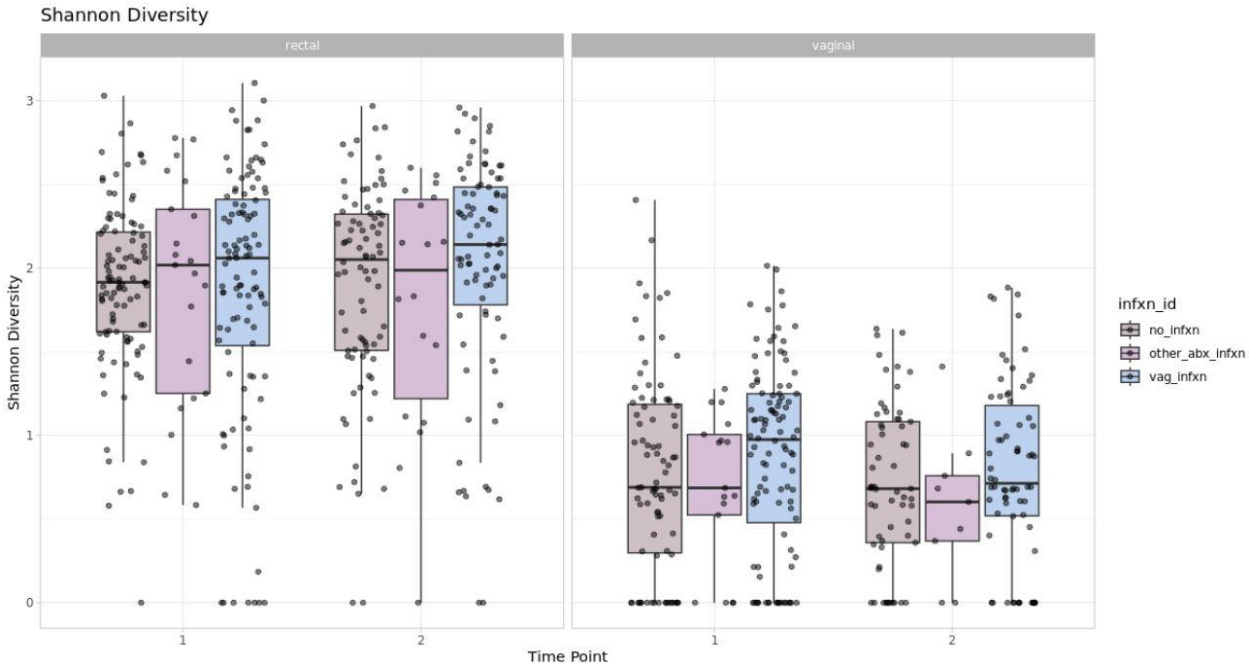
Vaginal samples are typically considered low biomass samples, as it is difficult to get a high concentration of DNA from vaginal swabs. Self-collected swabs from pregnant individuals were used for this study. Many individuals are uncomfortable with self-collecting these swabs as this is a new experience for many individuals. Additionally, as the body changes and grows during pregnancy, it may be increasingly difficult for pregnant individuals to collect these swabs. It is possible that this insufficient swabbing at the sampling site impacted our ability to detect AMR genes in this population. However, it is also plausible that AMR genes wouldn't differ significantly between drug classes in this cohort as timing of antibiotic administration related to sample collection was variable, impacting the perturbations that could be observed. A limit of the current study is that it does not control the number of AMR genes detected by the read count per sample. Such a correction will be key to future research studies. Some AMR tools, such as shortBRED, include the number of gene copies for an AMR gene in a particular sample. Tools like AMRFinderPlus will need to correct for this by seeing which contig the AMR gene is detected on, then going into the FASTA file to see how many times that contig was observed. Then, it is possible to normalize the detected genes by read counts, such as by taking the read count of a gene in a sample and dividing by the total read count of that sample

An additional limit of this study is that there is limited data on the timing of the urogenital infection diagnoses that are being examined. For example, a patient may have contracted a UTI after the initial timepoint but still weeks prior to the second sampling time point. Future studies examining the effect of urogenital infections on the microbiome during pregnancy should aim to recruit patients at the visit where the infection is diagnosed and aim to follow up after treatment of the infection.

Conclusion

Overall, there was no significant difference in the gut and vaginal microbiome for those who contracted urogenital infections during pregnancy or were exposed to antimicrobials for other reasons compared to pregnant patients who were not exposed to any antimicrobials or have any urogenital infections. There were significantly more virulence genes in the vaginal microbiome in later pregnancy in those who developed urogenital infections. There were more resistance genes to copper and silver in the gut microbiome in early pregnancy in those who developed urogenital infections, though this difference disappears in later pregnancy.

Figure One: Alpha Diversity of Vaginal and Rectal Samples for all Groups

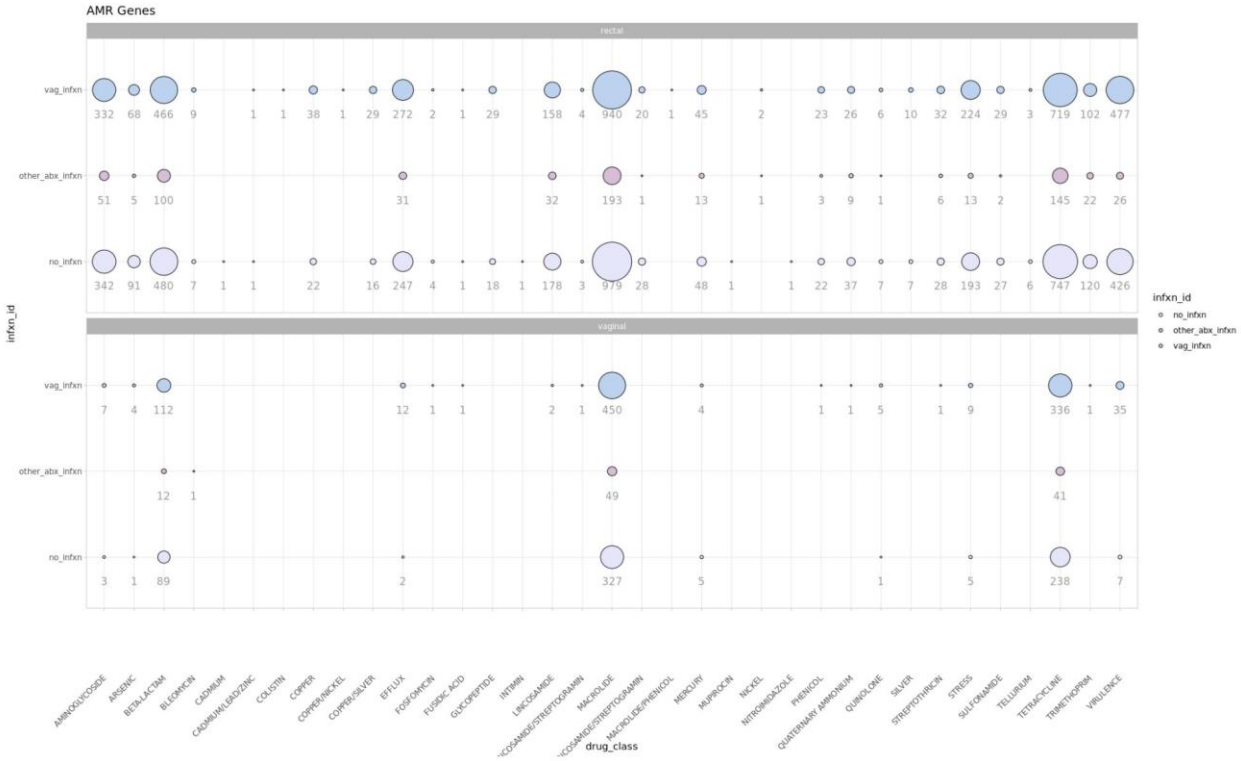


Shannon alpha diversity as calculated using R package vegan. Samples from patients with chlamydia, UTI, or BV diagnosis during pregnancy are shown in light blue, samples from patients with exposure to antimicrobials for other reasons are in light purple, and samples from patients with no reported antimicrobial exposure are in grey. There was no significant difference in alpha diversity across these groups while controlling for sampling time point.

Figure Two: Total Number of AMR Genes per drug class across body sites

This figure shows the total number of AMR genes detected by AMR Finder Plus (or AMR Class if not an AMR gene, e.g., virulence and stress) in (A), while showing the distribution of AMR genes detected per sample in (B). On the Y axis is "vag_infxn" for those who had either a UTI, BV, and Chlamydia infection; "other_abx_infxn" for those who receive antibiotics or antifungals for other conditions during pregnancy, and "no_infxn" for those who don't have any of the three conditions or any antimicrobial exposure in their pregnancy. Both (A) and (B) have the same x axis, drug class.

A:



B:

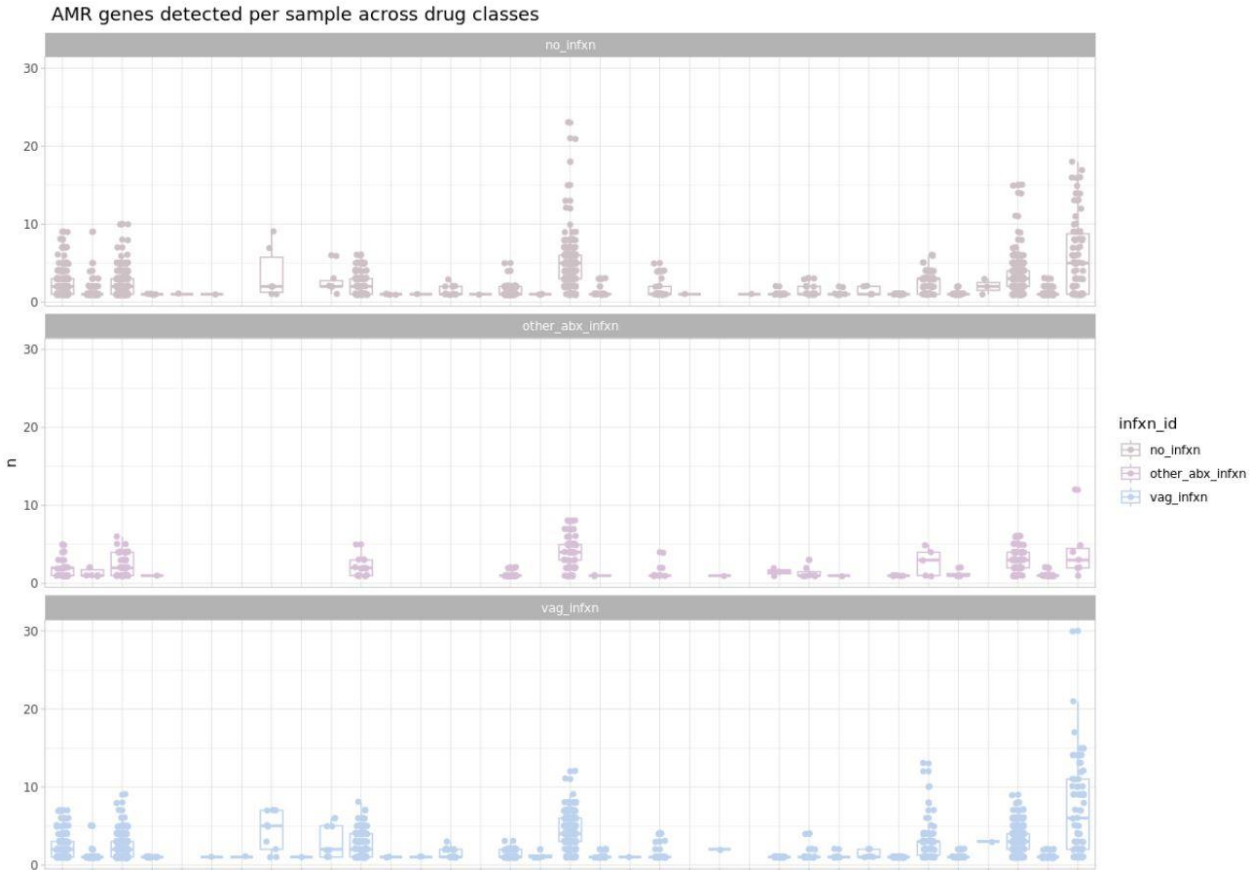

AMR genes detected per sample across drug classes
no_infxn
other_abx_infxn
vag_infxn
n
0
10
20
30
infxn_id
no_infxn
other_abx_infxn
vag_infxn

Table 1: Vaginal Infection Rates in this cohort

This table shows how many patients received a diagnosis of BV, UTI, and/or Chlamydia infection during pregnancy, the percentage of the cohort with that diagnosis in parenthesis.

	No Diagnosis	**Yes, diagnosis**
BV	181	57 (31.5 %)
UTI	192	46 (24 %)
Chlamydia	208	30 (14.4 %)
BV + UTI	-	11
UTI + Chlamydia	-	9
BV + Chlamydia	-	3
BV + UTI + Chlamydia	-	1

Table 2: Standardized residual Values from Chi Squared

https://docs.google.com/spreadsheets/d/1JWC3N-pEOMAWLbFr6pGfYhML5DibIaSHVh4P8-7Lsbg/edit#gid=694029774

Supplementary Table 1: Clinical Isolates used for simulated data

Strain	Abx	classification
B513: gardnerella	**Metrodiazole**	resistant
B482 gardnerella	**Metrodiazole**	resistant
B483 gardnerella	**Metrodiazole**	resistant
Gonorrhea GCGS096	**Azithromycin**	resistant
Gonorrhea GCGS 142 (weird alleles	**Azithromycin**	resistant
Gonorrhea GCGS 226	**Azithromycin**	resistant
INF299 klebsiella	**Nitrofurantoin**	resistant
INF158 klebsiella	**Nitrofurantoin**	resistant
INF157 klebsiella	**Nitrofurantoin**	resistant

Supplementary Table 2: Preliminary Data Results from hAMRonization to Compare AMR Tool Performance

https://docs.google.com/spreadsheets/d/1wTYBZue3VZjTnVXZg7fNJZRpXoxsKljQWIsB5UeDKoE/edit#gid=0

Supplementary Table 3: hAMRoaster Output for preliminary data analysis

tool	true_pos	false-neg	unknown	true-neg	false_pos	sensitivity	precision	specificity	accuracy	percent_unclassified
abricate	0.000	2.000	2.000	0.000	0.000	0.000	NA	0.000	0.000	1.000
amrfinderplus	2.000	1.000	2.000	0.000	0.000	0.667	1.000	0.000	0.667	0.500
deeparg	0.000	2.000	4.000	0.000	0.000	0.000	NA	0.000	0.000	1.000
rgi	6.000	1.000	8.000	0.000	0.000	0.857	1.000	0.000	0.857	0.571

References

Africa, Charlene W. J., Janske Nel, and Megan Stemmet. 2014. "Anaerobes and Bacterial Vaginosis in Pregnancy: Virulence Factors Contributing to Vaginal Colonisation." *International Journal of Environmental Research and Public Health* 11, no. 7 (July): 6979–7000. https://doi.org/10.3390/ijerph110706979.

Afridi, Ome Kalsoom, Johar Ali, and Jeong Ho Chang. 2021. "Fecal Microbiome and Resistome Profiling of Healthy and Diseased Pakistani Individuals Using Next-Generation Sequencing." *Microorganisms* 9, no. 3 (March): 616. https://doi.org/10.3390/microorganisms9030616.

Babraham Bioinformatics. n.d. "FastQC." Accessed April 5, 2023. https://www.bioinformatics.babraham.ac.uk/projects/fastqc/.

Bolger, Anthony M., Marc Lohse, and Bjoern Usadel. "Trimmomatic: A Flexible Trimmer for Illumina Sequence Data." *Bioinformatics* 30, no. 15 (August 1, 2014): 2114–20. https://doi.org/10.1093/bioinformatics/btu170.

Bookstaver, P. Brandon, Christopher M. Bland, Brooke Griffin, Kayla R. Stover, Lea S. Eiland, and Milena McLaughlin. 2015. "A Review of Antibiotic Use in Pregnancy." *Pharmacotherapy* 35, no. 11 (November): 1052–62. https://doi.org/10.1002/phar.1649.

Buckner, Michelle M. C., Howard T. H. Saw, Rachael N. Osagie, Alan McNally, Vito Ricci, Matthew E. Wand, Neil Woodford, Alasdair Ivens, Mark A. Webber, and Laura J. V. Piddock. 2018. "Clinically Relevant Plasmid-Host Interactions Indicate That Transcriptional and Not Genomic Modifications Ameliorate Fitness Costs of *Klebsiella Pneumoniae* Carbapenemase-Carrying Plasmids." *MBio* 9, no. 2 (May): e02303-17. https://doi.org/10.1128/mBio.02303-17.

Corwin, Elizabeth J., Carol J. Hogue, Bradley Pearce, Cherie C. Hill, Timothy D. Read, Jennifer Mulle, and Anne L. Dunlop. 2017. "Protocol for the Emory University African American Vaginal, Oral, and Gut Microbiome in Pregnancy Cohort Study." *BMC Pregnancy and Childbirth* 17, no. 1 (December): 161. https://doi.org/10.1186/s12884-017-1357-x.

DeLong, Kevin, Fareeha Zulfiqar, Diane E. Hoffmann, Anita J. Tarzian, and Laura M. Ensign. 2019. "Vaginal Microbiota Transplantation: The Next Frontier." *Journal of Law, Medicine & Ethics* 47, no. 4: 555–67. https://doi.org/10.1177/1073110519897731.

Dixon, Philip. 2003. "VEGAN, a package of R functions for community ecology." *Journal of Vegetation Science* 14: 927-30. http://www.pelagicos.net/MARS6910_spring2015/papers/Dixon_2003.pdf.

Focarelli, Francesca, Andrea Giachino, and Kevin John Waldron. 2022. "Copper Microenvironments in the Human Body Define Patterns of Copper Adaptation in

Pathogenic Bacteria." *PLOS Pathogens* 18, no. 7 (July): e1010617. https://doi.org/10.1371/journal.ppat.1010617.

He, Weihua, Yue Jin, Haibin Zhu, Yan Zheng, and Jianhua Qian. 2020. "Effect of Chlamydia Trachomatis on Adverse Pregnancy Outcomes: A Meta-Analysis." *Archives of Gynecology and Obstetrics* 302, no. 3 (September): 553–67. https://doi.org/10.1007/s00404-020-05664-6.

Huttenhower Lab. n.d. "Kneaddata." Accessed April 5, 2023. https://huttenhower.sph.harvard.edu/kneaddata/.

Li, Dinghua, Chi-Man Liu, Ruibang Luo, Kunihiko Sadakane, and Tak-Wah Lam. 2015. "MEGAHIT: An Ultra-Fast Single-Node Solution for Large and Complex Metagenomics Assembly via Succinct *de Bruijn* Graph." *Bioinformatics* 31, no. 10 (May): 1674–76. https://doi.org/10.1093/bioinformatics/btv033.

Ma, Bing, Larry J. Forney, and Jacques Ravel. 2012. "Vaginal Microbiome: Rethinking Health and Disease." *Annual Review of Microbiology* 66, no. 1 (October): 371–89. https://doi.org/10.1146/annurev-micro-092611-150157.

Majeroni, Barbara A., and Sreelatha Ukkadam. 2007. "Screening and Treatment for Sexually Transmitted Infections in Pregnancy." *American Family Physician* 76, no. 2 (July): 265–70. https://www.aafp.org/pubs/afp/issues/2007/0715/p265.html.

McGregor, James A. and Janice French. 2000 "Bacterial Vaginosis in Pregnancy." *Obstetrical & Gynecological Survey* 55, no. 5 (May): 1. https://journals.lww.com/obgynsurvey/Abstract/2000/05001/Bacterial_Vaginosis_in_Pregnancy.1.aspx.

Newbury, Arthur, Beth Dawson, Uli Klümper, Elze Hesse, Meaghan Castledine, Colin Fontaine, Angus Buckling, and Dirk Sanders. 2022. "Fitness Effects of Plasmids Shape the Structure of Bacteria–Plasmid Interaction Networks." *Proceedings of the National Academy of Sciences* 119, no. 22 (May): e2118361119. https://doi.org/10.1073/pnas.2118361119.

Patangia, Dhrati V., C. Anthony Ryan, Eugene Dempsey, Catherine Stanton, and R. Paul Ross. 2022. "Vertical Transfer of Antibiotics and Antibiotic Resistant Strains across the Mother/Baby Axis." *Trends in Microbiology* 30, no. 1 (January): 47–56. https://doi.org/10.1016/j.tim.2021.05.006.

Pereira-Dias, Joana, Chau Nguyen Ngoc Minh, Chau Tran Thi Hong, To Nguyen Thi Nguyen, Tuyen Ha Thanh, Caroline Zellmer, Hao Chung The, Lindsay Pike, Ellen E Higginson, and Stephen Baker. 2021. "The Gut Microbiome of Healthy Vietnamese Adults and Children Is a Major Reservoir for Resistance Genes Against Critical Antimicrobials." *The Journal of Infectious Diseases* 224, no. Supplement_7 (December): S840–47. https://doi.org/10.1093/infdis/jiab398.

Public Health Alliance for Genomic Epidemiology. 2022. “hAMRonization.” GitHub. Accessed April 4, 2023. https://github.com/pha4ge/hAMRonization.

Saenkham-Huntsinger, Panatda, Amanda N. Hyre, Braden S. Hanson, George L. Donati, L. Garry Adams, Chanelle Ryan, Alejandra Londoño, Ahmed M. Moustafa, Paul J. Planet, and Sargurunathan Subashchandrabose. 2021. “Copper Resistance Promotes Fitness of Methicillin-Resistant Staphylococcus Aureus during Urinary Tract Infection.” Edited by Sheryl Justice. *MBio* 12, no. 5 (October): e02038-21. https://doi.org/10.1128/mBio.02038-21.

Seedat, Farah, Chris Stinton, Jacoby Patterson, Julia Geppert, Bee Tan, Esther R. Robinson, Noel Denis McCarthy, et al. 2017. “Adverse Events in Women and Children Who Have Received Intrapartum Antibiotic Prophylaxis Treatment: A Systematic Review.” *BMC Pregnancy and Childbirth* 17, no. 1 (December): 247. https://doi.org/10.1186/s12884-017-1432-3.

Sharpe, Donald. 2019. “Chi-Square Test Is Statistically Significant: Now What?” *Practical Assessment, Research, and Evaluation* 20, no. 8. https://doi.org/10.7275/TBFA-X148.

Truong, Duy Tin, Eric A. Franzosa, Timothy L. Tickle, Matthias Scholz, George Weingart, Edoardo Pasolli, Adrian Tett, Curtis Huttenhower, and Nicola Segata. 2015. “MetaPhlAn2 for Enhanced Metagenomic Taxonomic Profiling.” *Nature Methods* 12, no. 10 (October): 902–3. https://doi.org/10.1038/nmeth.3589.

Vornhagen, Jay, Kristina M. Adams Waldorf, and Lakshmi Rajagopal. 2017. “Perinatal Group B Streptococcal Infections: Virulence Factors, Immunity, and Prevention Strategies.” *Trends in Microbiology* 25, no. 11 (November): 919–31. https://doi.org/10.1016/j.tim.2017.05.013.

Wissel, Emily F., Brooke M. Talbot, Noriko A. B. Toyosato, Robert A. Petit, Vicki Hertzberg, Anne Dunlop, and Timothy D. Read. 2023. “HAMRoaster: A Tool for Comparing Performance of AMR Gene Detection Software.” *bioRxiv* (January). https://doi.org/10.1101/2022.01.13.476279.

Chapter 5

Closing Remarks

Introduction

The purpose of this dissertation was to (1) understand how bacterial species change in the microbiome during pregnancy with and without urogenital infections, and (2) quantify the AMR genes in the gut and vaginal microbiome over time in pregnancy in those with and without urogenital infections and antimicrobial exposure. Chapter two presents the results from an analysis of the species in the gut and microbiome from metagenome data and replicates the analysis in 16S rRNA data for the same samples. Chapter 3 describes a novel bioinformatics tool called hAMRoaster which allows researchers to compare the results across different analytical pipelines for detecting AMR genes in metagenomic data and presents results from two simulated metagenomic datasets. Chapter 4 analyzes the distribution of AMR genes in those with urogenital infections, those given antimicrobial medications for other infections, and those without any antimicrobial exposure during their pregnancy. A summary of each chapter follows.

Chapter 2: The Impact of Urogenital Infections on the Gut and Vaginal Microbiome in Pregnancy

This manuscript explores how urogenital infections may impact the microbiome during pregnancy. It implements a linear decomposition model is beneficial because it can test the global and individual associations in a unified approach while adjusting for the false discovery rate, controlling for continuous and discrete confounding variables, and it can handle samples which are not independent of each other (as is the case with multiple timepoints and body sites per patient included in the study). This study hypothesized that the gut and vaginal microbiome would become less diverse as a pregnancy progresses (H1.1) and that any significant differences

in the microbiome detected could be explained by urogenital infection status (H1.2; UTIs, BV, and Chlamydia). Since 16S sequencing data was available for the sample samples, this study also hypothesized that findings from the metagenome data would be replicated by repeating the same analysis in the 16S data (H1.3).

This study found no significant difference in the microbiome at either body site over time in the metagenome data, though the vaginal microbiome was significantly different over time in the 16S data. Some specific species were associated with urogenital infections, and these species were in line with previous literature on the effect of urogenital infections on the microbiome. There were taxa found to be significantly associated with the urogenital infections from the metagenome data and the 16S data, likely due to differences in how each of these sequencing methods work. Clinicians will need to weigh which microbes are important to detect when deciding whether to use 16S rRNA or shotgun metagenome sequencing for their patients as microbiome information becomes integrated into healthcare systems.

Chapter 3: hAMRoaster: a tool for comparing performance of AMR gene detection software

The second study in this paper presents a newly developed, open-source bioinformatics tool called hAMRoaster (Harmonized AMR Output compAriSon Tool ER). In this study, two datasets were simulated–highly resistant mock community and several low resistance metagenomic short read (Illumina) samples based on sequenced strains with known phenotypes. This data was then processed through nine open-source bioinformatics tools for detecting AMR genes that 1) were conda or Docker installable, 2) had been actively maintained, 3) had an open-source license, and 4) took FASTA or FASTQ files as input. hAMRoaster calculated sensitivity,

specificity, precision, and accuracy for each tool, comparing detected AMR genes to AMR phenotypes from susceptibility testing.

Overall, all tools were precise and accurate at all genome coverage levels tested (5x, 50x, 100x sequenced bases / genome length) in the highly resistant mock community with more variability in the low resistance community (1x coverage). This study demonstrated that different bioinformatic tools yield differences in AMR gene identification across drug classes, and that these differences become important if researchers are interested in resistance to specific drug classes. hAMRoaster is open-source and conda installable so that researchers can easily analyze results across different bioinformatic tools in their own pretesting.

Chapter 4: Frequency of Antimicrobial Resistance Genes in a Pregnancy Cohort for those with and without Urogenital Infections during Pregnancy

The third study of this dissertation aimed to quantify the distribution of AMR genes across different drug classes from the gut and vaginal microbiome over time during pregnancy. This study hypothesized that those who receive antibiotics for urogenital infections (chlamydia, UTI, and BV) would have a more diverse vaginal microbiome and less diverse gut microbiome than those who did not have any exposure to antimicrobials during their pregnancy (H2.1). It also hypothesized that those exposed to antibiotic therapy for these three urogenital infections would have more AMR genes overall in their gut and vaginal microbiome.

An ANOVA revealed that there was no significant difference in alpha diversity over time between groups at either body site, indicating that antibiotics did not significantly impact microbiome diversity during pregnancy. A chi-squared test revealed that there were significantly more AMR genes in the gut microbiome in early pregnancy, specifically copper and silver resistance genes, and significantly more virulence genes in the vaginal microbiome during later

pregnancy in those with urogenital infections compared to those exposed to antimicrobials for other infections and those with no antimicrobial exposure during pregnancy. These results indicate that urogenital infections impact the frequency of AMR genes across drug classes during pregnancy, but the microbiome may stabilize back to a baseline AMR frequency after selective pressure for AMR genes is no longer present.

Discussion

This study helps fill in several knowledge gaps. First, there is not a clear understanding of how the microbiome may be different for those who develop a urogenital infection during pregnancy versus those who do not. Again, this study finds that urogenital infection does not have a significant impact on the microbiome. As pregnant individuals are more susceptible to urogenital infections, and the risks of those infections are greater during pregnancy, it is encouraging to find that contracting one of these infections will not significantly impact microbial health overall. While specific taxa were significantly associated with different urogenital infections, these significant taxa are in line with literature on these infections in nonpregnant populations, so it is not surprising that these same taxa would be significant for the same infections in pregnant patients.

This study also helps to better understand the reservoir of AMR genes in pregnant individuals with and without urogenital infections, and with and without AMR exposure. While those with urogenital infections were significantly more likely to have more AMR genes in their microbiome, the significant effect could be explained by one or two drug classes per body site, and the effects did not last throughout the full pregnancy. While copper and copper/silver resistance genes were significantly elevated in the gut microbiome of those with a urogenital infection in early pregnancy, this effect was no longer observed later in pregnancy. Similarly,

there was no significant difference in the vaginal microbiome in early pregnancy, though later in pregnancy, virulence genes were significantly elevated in those with urogenital infections. These significant findings are biologically plausible as discussed in chapter 4.

Additionally, this dissertation developed a new bioinformatic tool for assessing the performance of different software for detecting AMR genes in metagenomic data. This tool, called hAMRoaster, found that all tools were very sensitive and accurate, and that tool selection will depend on which drug classes researchers are interested in detecting, as opposed to a broad approach in detecting AMR genes, in which all tools performed comparably. hAMRoaster is open-source and conda installable, and these features ensure that the tool will be easily usable by other research groups and will remain stable over time. A common problem with bioinformatic tool developed in academic labs is that they must be compiled from source or don't specify which version of the dependent software they use; these problems are overcome by making the tool conda installable, while being open-source means that others can easily access the source code for hAMRoaster and make adjustments or add additional features for their own use as they see fit.

Collectively, this dissertation finds that the gut and vaginal microbiome are not significantly impacted by urogenital infections or their treatment. Specific microbes and AMR genes tend to be increased in those who developed urogenital infections compared to those who did not, however, these differences do not persist for the entire pregnancy. These findings should reassure most patients that being diagnosed with a urogenital infection and receiving antibiotic therapy for that infection will not have a significant, detrimental impact on their microbiome overall during pregnancy.

Limitations

While this cohort contains a lot of information from their medical chart during their pregnancy, there is not information on when certain diagnoses were received. For example, the urogenital infections explored in chapter 2 and 4 were diagnosed at any time point during their pregnancy, not necessarily when the microbiome was sampled. Additionally, while this dataset is large, it is not large enough to model the effect on all information from the medical charts into the linear decomposition model. As such, targeting hypotheses were explored for this dissertation as opposed to more exploratory analysis.

Another limitation of this data is that relatively few taxa were identified at the species level per sample. This limitation was explored at length in chapter 2, but briefly, an analysis on a small subset of the data comparing three different read-mapping based bioinformatic approaches for identifying taxonomy and showed comparable results across the three methods. This implies that the few taxa identified are not an artifact of the bioinformatic method used in this study, but rather, likely due to the relative abundance threshold (1%). that vaginal samples tend to yield low DNA concentration, and that participants were likely uncomfortable when self-swabbing.

Implications for future research and practice

Overall, this dissertation reassures that urogenital infections and their treatment in pregnancy do not have a significant impact on the microbiome. This evidence should help clinicians communicate the safety of antibiotic treatment (in reference to the microbiome and collection of AMR genes) for their pregnant patients. Microbiome sequencing is not currently integrated into healthcare systems, in large part due to the cost associated with sequencing. However, costs are going down as more sequencing platforms are available (targeted 16S rRNA sequencing, short read metagenomic sequencing, long read nanopore sequencing), and there are

many biotech companies that offer products centered around helping customer better understand the state of their microbiome (e.g., Tiny Health [Tiny Health, n.d.]). In the future, patients may ask medical providers to help them understand their microbiome or what types of products to buy. Many studies have compared results from different sequencing technologies from the same samples (Wei et al. 2020, Gehrig et al. 2022, and Tamburini et al. 2022), though not always with specific clinical questions. This research is still in its infancy, especially as long read sequencing platforms are relatively new to the market. Overall, clinicians will want to consider the cost and what clinical questions can be answered from each different sequencing technology (e.g., no viral reads from 16S rRNA gene sequencing data).

Future research may include more timepoints in the studies, or specifically sample the microbiome at prenatal visits where urogenital infections are diagnosed if they want to better understand the dynamics of the microbiome during active infection and after a defined recovery period. Further, sampling time points during pregnancy vary between research studies which makes it difficult to understand how the microbiome is changing during pregnancy. For example, it is generally understood that the microbiome becomes less diverse and stabilizes over the course of a pregnancy but becomes more diverse again closer to childbirth. Ideally, samples from each trimester could be collected for a more nuanced understanding of microbiome dynamics. Further, having more time points and metagenomic sequencing can help future studies develop a more refined understanding of the reserve of AMR genes during pregnancy.

Key Findings
The gut and vaginal microbiome does not differ significantly between those who have UTIs BV, or Chlamydia infections during their pregnancy. However, specific taxa from each body site are significantly associated with each of these three infections. Different sequencing technology can help understand the role of different taxa and will be relevant to different clinical contexts.
Those who have any of these three urogenital infections during pregnancy have significantly more AMR genes in the gut and vaginal microbiome than those exposed to antimicrobials for other reasons and those not exposed to antimicrobials during their pregnancy. Specifically, copper and copper/silver resistance genes are elevated in the gut microbiome in early pregnancy, and virulence genes are elevated in late pregnancy, for those who develop a urogenital infection in pregnancy.
Most bioinformatic tools for detecting AMR genes are sensitive and accurate, though differences arise when examining results across different drug classes. A novel bioinformatic tool called hAMRoaster can help researchers analyze the results from different tools to compare performance and select the best processing pipeline for their research questions.

References

Gehrig, Jeanette L., Daniel M. Portik, Mark D. Driscoll, Eric Jackson, Shreyasee Chakraborty, Dawn Gratalo, Meredith Ashby, and Ricardo Valladares. 2022. “Finding the Right Fit: Evaluation of Short-Read and Long-Read Sequencing Approaches to Maximize the Utility of Clinical Microbiome Data.” *Microbial Genomics* 8, no. 3 (March): 000794. https://doi.org/10.1099/mgen.0.000794.

Tamburini, Fiona B., Dylan Maghini, Ovokeraye H. Oduaran, Ryan Brewster, Michaella R. Hulley, Venesa Sahibdeen, Shane A. Norris, et al. 2022. “Short- and Long-Read Metagenomics of Urban and Rural South African Gut Microbiomes Reveal a Transitional Composition and Undescribed Taxa.” *Nature Communications* 13, no. 1 (February): 926. https://doi.org/10.1038/s41467-021-27917-x.

Tiny Health. n.d. “At-Home Gut Health Test for Babies and Moms – Tiny Health.” Accessed April 6, 2023. https://www.tinyhealth.com/.

Wei, Po-Li, Ching-Sheng Hung, Yi-Wei Kao, Ying-Chin Lin, Cheng-Yang Lee, Tzu-Hao Chang, Ben-Chang Shia, and Jung-Chun Lin. 2020. “Characterization of Fecal Microbiota with Clinical Specimen Using Long-Read and Short-Read Sequencing Platform.” *International Journal of Molecular Sciences* 21, no. 19 (January): 7110. https://doi.org/10.3390/ijms21197110.