

ABSTRACT

Within the last few decades, it has become increasingly apparent that microbiome composition is impactful to many aspects of a host's life. People may commonly think of the human microbiome, but a microbiome can be defined as all the microbial species present in any defined location.

In this thesis, I have explored the microbiome composition and diversity of two mosquito species – *Aedes aegypti* and *Anopheles gambiae*. Mosquitoes are not only ecologically important, but they also pose a major risk to human health. Due to their ability to act as vectors for human pathogens, they are the deadliest animal on the planet. As such, a complete understanding of how the mosquito microbiome interacts with these human pathogens is critical to achieve some increased level of pathogen transmission suppression.

In this thesis, I first used genomic data collected during a study on 172 female *Ae. aegypti* collected in Cairns, Australia and second, I used data from a study of 711 female *An. gambiae* mosquitos collected across sub-Saharan Africa. *Ae. aegypti* and *An. gambiae* are two of the most dangerous mosquito species, acting as the primary vectors for the dengue virus and malaria, respectively. It is often suggested that without the addition of a biological control strategy, the spread of vector-borne human pathogens will not dramatically decline. To develop biological control strategies, it is necessary to have a deep understanding of the microbial profile of a given mosquito species and explore relationships between microbial species within the microbiome. Here, I have characterized the natural microbiomes of both *Ae. aegypti* and *An. gambiae*. Furthermore, I have identified evidence that distinct core microbiomes exist in each of these species. In *Ae. aegypti*, I define the relationship between the microbiome and artificial

Wolbachia infection. And finally, in *An. gambiae* I have explored differences in microbiome composition between intra-specific molecular forms.

The work within this thesis enhances our understanding of mosquito biology and tri-partite relationship between a host, its microbiome, and pathogens.

CHARACTERIZING MICROBIOME VARIATION IN
WILD POPULATIONS OF MOSQUITO SPECIES

Jane Pascar

B.Sc. State University of New York at Oswego, 2017

Thesis submitted in partial fulfillment of the requirements
for the degree of Master of Science in Biology.

Syracuse University

August 2022

Copyright © Jane Pascar, 2022

All Rights Reserved

ACKNOWLEDGEMENTS

First and foremost, I would like to express my sincerest appreciation and gratitude for my advisor Dr. Steve Dorus. He challenged me to be a better scientist and without his advice, mentorship, and encouragement this thesis would not have been possible. Dr. Dorus' support helped make it possible for me to pursue a line of research I personally found interesting, and he provided guidance as I explored my career and personal interests.

Thank you to my committee members, Dr. Scott Pitnick, Dr. Yasir Ahmed-Braimah, Dr. Jason Fridley and Dr. John Belote for providing useful critiques and feedback on my ideas and projects. I appreciate your mentorship and guidance during my time at Syracuse.

I also want to thank my undergraduate advisor Dr. Chris Chandler. His class, Sexual Diversity and Evolution, is really what sparked my interest in research. Joining his lab at Oswego and then working on my undergraduate thesis with his guidance confirmed that I wanted to pursue graduate school. I greatly appreciate his mentorship.

Finally, I would like to thank my friends and family for their support, I'm incredibly lucky to be surrounded by such amazing people. I especially want to thank my parents, Lewis and Tracy, and my siblings, Jaca and Daniel, for encouraging me to follow my passions. Thank you to my partner, Alec, you've been a source of positivity and encouragement during the final year of this degree. You not only believe in me, but you remind me of the reasons that I should believe in myself. My most heartfelt thanks to my best friends, Bridget, Carolyn, and Julianna. You all have been by my side since the very beginning of my career in research. Your unconditional support means the world to me, and you are each a source of inspiration for me. Thank you for providing me with countless happy memories, hours of laughter but also a shoulder to cry on when I need to be picked up.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	v
TABLE OF CONTENTS	vi
LIST OF ILLUSTRATIVE MATERIALS	viii
CHAPTER 1: INTRODUCTION	
1.1 Small but Mighty: The Impact of the Microbiome	1
1.2 A Matter of Time: How Technology has Improved Microbiome Research	3
1.3 One Man's Trash is Another Man's Treasure: Using Publicly Available Sequencing Data to Explore the Microbiome	5
1.4 Casting a Broad Net: Choosing a Mosquito Model	7
1.5 Figures	10
1.6 References	11
CHAPTER 2: <i>AEDES AEGYPTI</i> MICROBIOME COMPOSITION COVARIATES WITH <i>WOLBACHIA</i> INFECTION	
2.1 Abstract	16
2.2 Introduction	17
2.3 Methods	20
2.4 Results	22
2.5 Discussion	28
2.6 Conclusion	32
2.7 Figures	33
2.8 References	39
CHAPTER 3: MICROBIOME VARIATION IN WILD <i>ANOPHELES GAMBIAE</i> COLLECTED ACROSS SUB- SAHARAN AFRICA	

3.1 Abstract.....	48
3.2 Introduction.....	49
3.3 Methods.....	52
3.4 Results.....	55
3.5 Discussion.....	62
3.6 Conclusion.....	66
3.7 Tables and Figures	67
3.8 References	78

CHAPTER 4: CONCLUSION

4.1 Conclusion.....	87
4.2 Future Directions	88
4.3 References	89
CURRICULUM VITAE.....	91

LIST OF ILLUSTRATIVE MATERIALS

CHAPTER 1:

Fig. 1 Logarithmic increase in nucleotide bases stored in the NCBI WGS database 10

Fig. 2 Logarithmic growth of microbial taxa stored in the NCBI RefSeq database 10

CHAPTER 2:

Fig. 1 Population-level *Ae. aegypti* microbiome characterization 33

Fig. 2 *Ae. aegypti* core microbiome 34

Fig. 3 Variation in wMel abundance across populations and individuals 35

Fig. 4 *Ae. aegypti* microbiome interaction network 36

Fig. 5 Covariance between microbiome composition and wMel abundance 38

CHAPTER 3:

Table 1. *An. gambiae* s.s. molecular form identification demographics 67

Fig. 1 Map of *An. gambiae* s.s. collection sites in Africa 68

Fig. 2 *An. gambiae* s.s. phylum-level microbiome characterization 69

Fig. 3 Variation in bacterial microbiome diversity in *An. gambiae* s.s. molecular forms 70

Fig. 4 Covariance between microbiome composition and *An. gambiae* s.s. molecular form 71

Fig. 5 *An. gambiae* s.s. core microbiome 73

Fig. 6 *P. falciparum* abundance across sampled female *An. gambiae* s.s. 75

Fig. S1 *An. gambiae* s.s. microbiome diversity as a function of geographic location 76

Fig. S2 *An. gambiae* s.s. microbiome composition as a function of geographic location 77

ADDITIONAL FILES – AVAILABLE AT: https://github.com/janepascar/masters_thesis

Additional File 1. Kraken2 and Bracken species identifications and average abundance for fungi, viruses, archaea, and protozoa identified from *Ae. aegypti* samples.

Additional File 2. Kraken2 and Bracken species identifications for the 844 bacterial species (representing 23 phyla and 379 genera) meeting at least one of these conservative inclusion criteria: (1) >1% of the total bacterial microbiome in at least one individual and (2) present in at least 25% of individuals identified as *Ae. aegypti*.

Additional File 3. Core microbiome of 164 bacterial species which were present in 95% or more of individuals across the entire population of *Ae. aegypti* samples surveyed. Average percent contribution to microbiome composition and presence in the strict core microbiome present in all individuals is also provided. Species that belong to a genus that was previously unidentified in *Ae. aegypti* is noted with an “*”.

Additional File 4. Conet network characteristics are provided for the *Ae. aegypti* microbiome interaction network, including the 159 significant edge interactions between all 71 species and the strength of each interaction (Pearson’s Correlation Coefficient).

Additional File 5. Krakenuniq species identifications and average abundance for fungi, viruses, archaea, and protozoa identified in *An. gambiae s.s.* samples.

Additional File 6. Krakenuniq species identifications for the 269 bacterial species (representing 7 phyla and 114 genera) meeting at least one of these conservative inclusion criteria: (1) >1% of the total bacterial microbiome in at least one individual and (2) present in at least 25% of individuals identified from *An. gambiae s.s.*

Additional File 7. Core microbiome of 46 bacterial species which were present in 95% or more of individuals across the entire population of *An. gambiae s.s.* samples surveyed. Average

percent contribution to microbiome composition and presence in the strict core microbiome present in all individuals is also provided.

CHAPTER 1: INTRODUCTION

1.1 *Small but Mighty: The Impact of the Microbiome*

Ed Yong, a renowned science journalist, in his 2016 book *I Contain Multitudes: The Microbes Within Us and a Grander View of Life*, wrote, “even when we are alone, we are never alone.” He was referring to the bevy of microorganisms that inhabit almost every corner of our world. For example, the human gut contains 10-100 trillion microorganisms; comparatively the average human body contains roughly 30 trillion cells¹. The human gut is just one example of a microbiome, which is simply defined as all the microorganisms present in a given location. While the human gut microbiome may be the most familiar example, microbiomes have been broadly sampled and characterized.

An immense amount of research has contributed to our knowledge of microbiomes, particularly the human microbiome. In 2007 the National Institutes of Health (NIH) launched the Human Microbiome Project (HMP), which had two phases: first, identifying and characterizing the human microbiome, and second, beginning in 2014, characterizing the function of the microbiome with a particular focus on the role of microbes in health and disease states^{1,2}. These studies revealed that in humans, microbiome composition is incredibly diverse at the species-level, yet broad patterns of microbiome composition still exist. For example, typically the human gut microbiome is dominated by three phyla, Firmicutes, Bacteroidetes, and Actinobacteria³, regardless of variation at the species level. The variability at the species-level is partially attributed to how the human gut microbiome is formed. Environment, age, diet, sex, pharmaceuticals, and genetics have all been shown to at least partially contribute to human microbiome composition¹⁻⁵. In other organisms this holds true – the composition of a

microbiome is determined by the complex interactions between many factors, likely including factors still unknown.

Furthermore, these studies aimed to identify if a human core microbiome exists. A core microbiome is defined as a group of microbial taxa or genes that are shared by all or most individuals. It is thought that any core species identified may be particularly important to host function^{6,7}. In humans, a core microbiome is largely nonexistent^{1,2}. But this is not the case for other organisms like coral⁸, bees⁹, baboons¹⁰, and *Arabidopsis*¹¹ for example.

While characterizing a host's microbial profile provides some biological insight in host characteristics, another important line of inquiry is exploring how total microbiome composition, or the presence/absence of certain taxa can impact a host's fitness. For example, in humans, microbiome composition has been linked to certain disease states; patients with irritable bowel disease (IBD) have a significantly different microbiome composition relative to healthy individuals. Some of these differences include the enrichment of *Blautia* and *Lachnospiraceae* species, while *Faecalibacterium*, *Subdoligranulum*, and *Alistipes* species are depleted during active inflammation². The link between IBD and microbiome composition is a simple example to illustrate the more broadly observed pattern that microbiome composition often impacts the host.

In model organisms, fundamental links between the microbiome and aspects of host's biology have also been discovered, including links to host fitness¹², fertility¹³, phenotypic plasticity¹⁴, and adaptive potential¹⁵. Yet, despite our contemporary understanding of the important relationship between a microbial species and a host, much of our knowledge has come from a limited number of common biomedical models in a controlled lab environment (e.g., zebrafish, mice, *Drosophila melanogaster* and *Caenorhabditis elegans*¹⁶). This is likely

underestimating natural microbiome diversity since numerous studies have shown that species maintained in lab environments have been shown to contain microbiomes with decreased diversity relative to their wild counterparts¹⁷⁻¹⁹. In addition, the literature is biased towards understanding host-pathogen dynamics, yet most species in the microbiome are non-pathogenic²⁰. For these reasons, in this thesis I have prioritized using data from wild-caught samples, with the goal of capturing a wider diversity of microbial species that better represents natural microbiome composition.

1.2 A Matter of Time: How Technology has Improved Microbiome Research

The last two decades have seen numerous technological advancements that have greatly facilitated studies on microbiomes. These advancements generally fall into two categories: (1) DNA sequencing technology and (2) computational software and analytical tools. The availability and improvements of sequencing technology, combined with innovative computational and analytical tools have revealed previously underappreciated host microbiome complexity.

DNA sequencing methods are continuously improving. Prior to sequencing technology, scientists were limited to isolating and culturing microbes, thus a great deal of diversity was simply overlooked²¹. Depending on the goal, whether it be characterization of a microbiome, a full functional analysis or other objectives, various sequencing approaches are readily available and optimized to produce quality data.

Traditionally, the most commonly used sequencing method in microbiome studies has been amplicon sequencing. By exploiting the fact that all bacterial species contain the 16S rRNA gene, amplicon sequencing is useful for characterizing microbiome composition. While

all bacterial species contain the 16S rRNA gene, in different taxa the nucleotides within the gene are dissimilar enough that a phylogeny can be produced (like in Carl Woese's groundbreaking work²²) or, in the case of microbiome research, a taxonomic identification can potentially be made.

Alternatively, whole-genome sequencing (WGS) is becoming increasingly popular for studies on microbiomes. Without even considering analytical tools, the fact that WGS captures the entire genomic content of a sample (not just the 16S rRNA gene) allows for the possibility of analysis beyond a gross microbiome characterization. WGS has always been more expensive than amplicon sequencing, but as WGS technology has become more efficient this gap has lessened. Again, considering that WGS provides the full genomic content rather than a single universal gene sequence, this additional cost may be worthwhile so a gene-centric or functional analysis can be performed.

Equally important to the field of microbiome research, are the innovations to computational software and analytical tools. With the increasing popularity of WGS to characterize a microbial profile, it is necessary for computational software to simultaneously advance. Once sequencing data are obtained, to garner any biological meaning, the microbial sequences must be taxonomically classified. For amplicon sequencing data, this is accomplished by clustering reads into Operational Taxonomic Units (OTUs) based on nucleotide similarity. Once they are grouped, an identification can *potentially* be assigned to an OTU cluster, often at the family or genus level but rarely at the species level, by comparing the consensus OTU cluster sequence to one or more of the available 16S rRNA gene databases.

Alternatively, when using WGS data, a k -mer classification approach can be used to compare an unknown query sequence to a database of genome sequences belonging to known

species. This has several advantages compared to OTU clustering. First, given that genomic information beyond a single gene is being compared to a database, classifications to the species or even the strain, level of bacteria, viruses, archaea, and prokaryotes are possible²³. This is possible because the additional genomic sequences act as supplemental lines of evidence when a classifier is attempting to make an identification compared to known species sequences. Any genome can be added to the database for comparison as long as the species of origin is known; it is completely customizable depending on your objective. Finally, *k*-mer classification programs may more accurately represent the diversity in a sample due to PCR biases that are potentially introduced during the amplicon sequencing procedure²⁴. Several taxonomic classifier program packages are available^{25–27}, but the Kraken suite^{28–30} has performed exceptionally well in both speed and accuracy²³.

1.3 One Man’s Trash is Another Man’s Treasure: Using Publicly Available Sequencing Data to Explore the Microbiome

Technological innovation, while essential to the advancement of microbiome research, is not solely responsible for the expansion of the field – an important component is data availability. Today, publicly releasing sequencing data into a centralized repository (e.g., the suite of NCBI databases) is highly encouraged upon project completion or publication, if not required by most peer-reviewed journals. Additionally, depositing sequencing data is considered best practice by the scientific community.

Since its inception in late 1982, NCBI has reported that the number of nucleotide bases available doubles every 18 months³¹. To sustain that rate of growth, not only does the sequencing technology need to be available but researchers need to be releasing their data.

Importantly, not only has the sheer amount of data grown exponentially, but the types of data useful in microbiome studies has become more readily available and accessible. As previously described, there are two key types of data that are essential to a microbiome study: (1) a sample of interest with some unknown microbial content and (2) a comprehensive database to which the sample will be compared. Conventionally, the latter is easily obtained via a data repository, however, the former is typically acquired through novel data collection. An example of a typical data collection pipeline could include sampling a tissue/location (e.g., a *D. melanogaster* gut), isolating DNA, preparing a sequencing library, and then sequencing the sample. This is a costly investment both in time and financially. Alternatively, with the use of the ever-growing resource of publicly available WGS data, these steps can be bypassed. A proportion of the WGS data that has been deposited was likely collected for purposes other than microbiome-related lines of inquiry, for example host genome assembly, population demographics, etc. As such, the removal of any microbial “contamination” prior to sequencing may not have been necessary (e.g., tissue sterilization, a course of antibiotics) since microbial reads are typically removed *in silico* during post-sequencing data processing. Note that the NCBI SRA has the specific purpose of housing raw reads that have not been filtered or processed; in other words, the data that could still include microbial “contamination.” After identifying a dataset of interest, by reading any published articles associated with those data, usually the lack of microbial removal or disruption before sequencing can be confirmed. Therefore, once WGS data has been produced, deposited, and vetted, it is free to be upcycled in microbiome-related studies.

In 2002, when NCBI began tracking accumulation statistics in their WGS database, there were only 172,768 nucleotide bases available across all records. As of April 2022, there are 16,071,520,702,170 WGS-derived nucleotides available in that database alone (Fig. 1). Not all

sample accessions that contain these sequenced nucleotides are necessarily appropriate for the analytical model I've described but a non-negligible amount are. For example, if one is interested in working with an arthropod species as a potential model, the NCBI SRA could be searched for accessions that meet the following criteria: sampled from a species within NCBI Taxonomy ID: 6656 (phylum: Arthropoda), is publicly available genomic DNA paired-end sequencing data that was processed on an Illumina platform and is in fastq format. This search alone yields 43,994 results. Again, not all of these may be appropriate, but that is a large resource to begin to draw from for microbiome studies.

Not only has the amount of host sequencing data increased dramatically over the past couple of decades, but the availability of microbial reference genomes has also increased which fulfills the second requirement of having a comprehensive database to which a sample will be compared. Each major microbial taxa (Archaea, Bacteria, Fungi, Protozoa, and Viruses) has seen logarithmic growth in the number of species genome sequences available in the NCBI RefSeq database since the early 2000's (Fig. 2).

1.4 Casting a Broad Net: Choosing a Mosquito Model

Like humans, arthropods fitness is impacted by host relationships with their microbiomes³²⁻³⁶. Using an arthropod model has several appealing qualities when attempting to study host-microbiome relationships. First, they are relatively easy to collect and not strictly regulated, meaning you can typically get permission to collect samples. Second, they are small so they do not take up much physical space in a lab, but they are large enough that a single individual can provide enough DNA for WGS sequencing without needing to pool individuals. Third, arthropods are ecologically important; many carry pathogenic microbes that impact not

only human health but livestock and crops. Finally, they commonly have simple microbiomes compared to humans and other large mammals^{37,38}. This makes them ideal for identifying clear relationships both between microbial species and between a microbial species and the host.

In this thesis, I work with two mosquito species, *Aedes aegypti* and *Anopheles gambiae*, as model organisms. Mosquitoes are the deadliest animal on the planet; they are vectors for human pathogens spanning the protozoan, bacterial, and viral clades and pose a great threat to public health. As such, there is an abundance of publicly available sequencing data for numerous species that have been collected globally. Furthermore, microbiome composition has been shown to impact multiple facets of mosquito biology, including, reproduction, development, survival, biting rate, vector competency, and vector density^{39,40}. Additionally, the microbiome of different tissues in mosquitoes have been identified as distinctly impactful to these host health factors. For example, the microbes in the mosquito salivary glands and gut are directly in contact with human pathogens, while microbial species located in the reproductive tissue may not physically interact with human pathogens or other gut microbes, but they may physiologically affect the host which in turn does impact other tissue-specific microbiomes⁴⁰.

While some components of the mosquito microbiome have already been identified, there is much to be explored. The natural microbiome composition of mosquitoes is largely unknown since most studies use lab-reared samples and pooled individuals for sequencing. This has led to a reduced view of the diversity that may be present in natural populations⁴¹. When studies do collect wild-caught individuals, they are often from a small geographic region. Given that the mosquito microbiome is largely dependent on their immediate environment,⁴¹ studies that only sample from closely related sites are more likely to identify common microbial species. This has caused debate on whether a core microbiome in mosquito species truly exists⁴².

In this thesis, I have characterized the microbiome composition of two mosquito models to explicitly define the relationships within an individual's natural microbial community and enhance understanding of the impact microbes have on a host and the nature of their relationships. Furthermore, I have explored the existence of a core microbiome in each species to identify those species that may have a large biologically functional role. To accomplish this, I performed a comprehensive analysis of microbes present in two data sets—*Ae. aegypti* collected from Cairns, Australia, and *An. gambiae* mosquitoes collected across sub-Saharan Africa.

By capitalizing on the rich abundance of publicly available data, I have analyzed high quality data using alternative methods to confidently make bacterial species identifications that are not possible with commonly employed amplicon sequencing. Ultimately, these results have led to a more in depth understanding of the dynamics of the mosquito microbiome, specifically which species are present in the *Ae. aegypti* and *An. gambiae* microbiome and the relationship of those species with human pathogens.

1.5 Figures

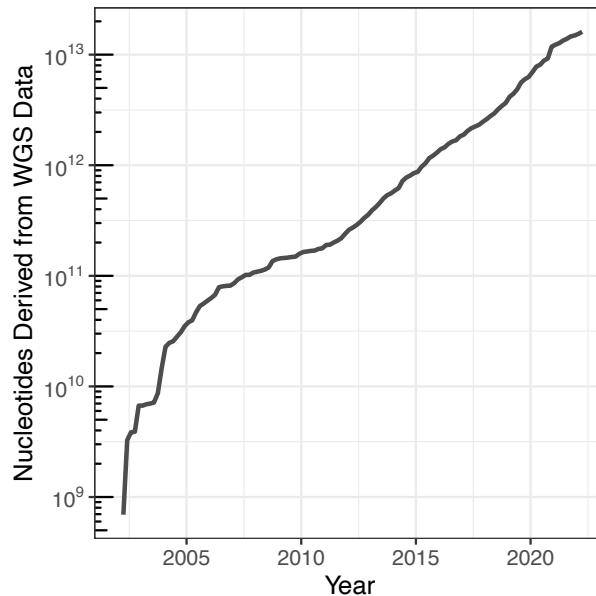


Figure 1. Logarithmic increase in nucleotide bases stored in the NCBI WGS database from April 2002 to April 2022. Data provided at: <https://www.ncbi.nlm.nih.gov/genbank/statistics/>.

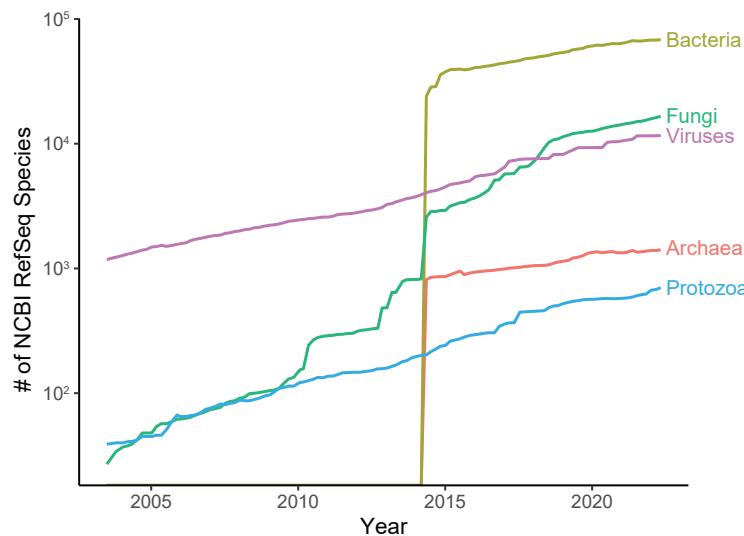


Figure 2. Logarithmic growth of microbial taxa stored in the NCBI RefSeq database from June 2003 to May 2022. Data provided at: <https://ftp.ncbi.nlm.nih.gov/refseq/release/release-statistics/>.

1.6 References

1. Turnbaugh, P. J. *et al.* The Human Microbiome Project. *Nature* **449**, 804–810 (2007).
2. Proctor, L. M. *et al.* The Integrative Human Microbiome Project. *Nature* **569**, 641–648 (2019).
3. Rinninella, E. *et al.* What is the Healthy Gut Microbiota Composition? A Changing Ecosystem across Age, Environment, Diet, and Diseases. *Microorganisms* **7**, 14 (2019).
4. Donaldson, G. P., Lee, S. M. & Mazmanian, S. K. Gut biogeography of the bacterial microbiota. *Nat. Rev. Microbiol.* **14**, 20–32 (2015).
5. Jernberg, C., Löfmark, S., Edlund, C. & Jansson, J. K. Long-term ecological impacts of antibiotic administration on the human intestinal microbiota. *ISME J.* **1**, 56–66 (2007).
6. Parfrey, L. W., Moreau, C. S. & Russell, J. A. The host-associated microbiome: Pattern, process and function. *Mol. Ecol.* **27**, 1749–1765 (2018).
7. Shade, A. & Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.* **14**, 4–12 (2012).
8. Ainsworth, T. D. *et al.* The coral core microbiome identifies rare bacterial taxa as ubiquitous endosymbionts. *ISME J.* **9**, 2261–2274 (2015).
9. Sabree, Z. L., Hansen, A. K. & Moran, N. A. Independent Studies Using Deep Sequencing Resolve the Same Set of Core Bacterial Species Dominating Gut Communities of Honey Bees. *PLoS One* **7**, e41250 (2012).
10. Grieneisen, L. E., Livermore, J., Alberts, S., Tung, J. & Archie, E. A. Group Living and Male Dispersal Predict the Core Gut Microbiome in Wild Baboons. *Integr. Comp. Biol.* **57**, 770–785 (2017).
11. Lundberg, D. S. *et al.* Defining the core *Arabidopsis thaliana* root microbiome. *Nature*

488, 86–90 (2012).

12. Bordenstein, S. R. & Theis, K. R. Host Biology in Light of the Microbiome: Ten Principles of Holobionts and Hologenomes. *PLOS Biol.* **13**, e1002226 (2015).
13. Cordaux, R., Bouchon, D. & Grève, P. The impact of endosymbionts on the evolution of host sex-determination mechanisms. *Trends Genet.* **27**, 332–341 (2011).
14. Carrier, T. J. & Reitzel, A. M. Convergent shifts in host-associated microbial communities across environmentally elicited phenotypes. *Nat. Commun.* **9**, 952 (2018).
15. Correa, C. C. & Ballard, J. W. O. Wolbachia Associations with Insects: Winning or Losing Against a Master Manipulator. *Front. Ecol. Evol.* **3**, 153 (2016).
16. Douglas, A. E. Simple animal models for microbiome research. *Nat. Rev. Microbiol.* **17**, 764–775 (2019).
17. Broderick, N. A., Raffa, K. F., Goodman, R. M. & Handelsman, J. Census of the Bacterial Community of the Gypsy Moth Larval Midgut by Using Culturing and Culture-Independent Methods. *Appl. Environ. Microbiol.* **70**, 293–300 (2004).
18. Xiang, H. *et al.* Microbial communities in the larval midgut of laboratory and field populations of cotton bollworm (*Helicoverpa armigera*). *Can. J. Microbiol.* **52**, 1085–1092 (2006).
19. Chandler, J. A., Lang, J., Bhatnagar, S., Eisen, J. A. & Kopp, A. Bacterial communities of diverse *Drosophila* species: Ecological context of a host-microbe model system. *PLoS Genet.* **7**, (2011).
20. Ursell, L. K., Metcalf, J. L., Parfrey, L. W. & Knight, R. Defining the human microbiome. *Nutr. Rev.* **70**, 38–44 (2012).
21. Stewart, E. J. Growing unculturable bacteria. *Journal of Bacteriology* **194**, 4151–4160

- (2012).
22. Woese, C. R. & Fox, G. E. Phylogenetic structure of the prokaryotic domain: The primary kingdoms. *Proc. Natl. Acad. Sci. U.S.A.* **74**, 5088–5090 (1977).
 23. Lindgreen, S., Adair, K. L. & Gardner, P. P. An evaluation of the accuracy and speed of metagenome analysis tools. *Sci. Rep.* **6**, 19233 (2016).
 24. Baker, G. C., Smith, J. J. & Cowan, D. A. Review and re-analysis of domain-specific 16S primers. *Journal of Microbiological Methods* **55**, 541–555 (2003).
 25. Ounit, R., Wanamaker, S., Close, T. J. & Lonardi, S. CLARK: fast and accurate classification of metagenomic and genomic sequences using discriminative k-mers. *BMC Genomics* **16**, 236 (2015).
 26. Huson, D. H., Mitra, S., Ruscheweyh, H. J., Weber, N. & Schuster, S. C. Integrative analysis of environmental sequences using MEGAN4. *Genome Res.* **21**, 1552–1560 (2011).
 27. Ames, S. K. *et al.* Scalable metagenomic taxonomy classification using a reference genome database. *Bioinformatics* **29**, 2253–2260 (2013).
 28. Wood, D. E. & Salzberg, S. L. Kraken: ultrafast metagenomic sequence classification using exact alignments. *Genome Biol.* **15**, R46 (2014).
 29. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
 30. Breitwieser, F. P., Baker, D. N. & Salzberg, S. L. KrakenUniq: Confident and fast metagenomics classification using unique k-mer counts. *Genome Biol.* **19**, 198 (2018).
 31. NCBI. NCBI: GenBank and WGS Statistics.
<https://www.ncbi.nlm.nih.gov/genbank/statistics/>.

32. Engel, P. & Moran, N. A. The gut microbiota of insects – diversity in structure and function. *FEMS Microbiol. Rev.* **37**, 699–735 (2013).
33. Berlanga, M., Llorens, C., Comas, J. & Guerrero, R. Gut Bacterial Community of the Xylophagous Cockroaches *Cryptocercus punctulatus* and *Parasphaeria boleiriana*. *PLoS One* **11**, e0152400 (2016).
34. Nakao, R. *et al.* A novel approach, based on BLSOMs (Batch Learning Self-Organizing Maps), to the microbiome analysis of ticks. *ISME J.* **7**, 1003–1015 (2013).
35. Bolaños, L. M., Rosenblueth, M., Castillo-Ramírez, S., Figuier-Huttin, G. & Martínez-Romero, E. Species-specific diversity of novel bacterial lineages and differential abundance of predicted pathways for toxic compound degradation in scorpion gut microbiota. *Environ. Microbiol.* **18**, 1364–1378 (2016).
36. Esposti, M. D. & Romero, E. M. The functional microbiome of arthropods. *PLoS One* **12**, e0176573 (2017).
37. Bleich, A. & Fox, J. G. The Mammalian Microbiome and Its Importance in Laboratory Animal Research. *ILAR J.* **56**, 153 (2015).
38. Nelson, K. E. An Update on the Status of Current Research on the Mammalian Microbiome. *ILAR J.* **56**, 163–168 (2015).
39. Minard, G., Mavingui, P. & Moro, C. Diversity and function of bacterial microbiota in the mosquito holobiont. *Parasit. Vectors* **6**, 146 (2013).
40. Cansado-Utrilla, C., Zhao, S. Y., McCall, P. J., Coon, K. L. & Hughes, G. L. The microbiome and mosquito vectorial capacity: rich potential for discovery and translation. *Microbiome* **9**, 1–11 (2021).
41. Minard, G., Mavingui, P. & Moro, C. V. Diversity and function of bacterial microbiota in

the mosquito holobiont. *Parasites and Vectors* **6**, 146 (2013).

42. Guégan, M. *et al.* The mosquito holobiont: fresh insight into mosquito-microbiota interactions. *Microbiome* **6**, 49 (2018).

CHAPTER 2: *AEDES AEGYPTI* MICROBIOME COMPOSITION COVARIATES WITH *WOLBACHIA* INFECTION

2.1 Abstract

BACKGROUND: *Wolbachia* is a widespread bacterial endosymbiont that can inhibit vector competency when stably transinfected into the mosquito, *Aedes aegypti*, a primary vector of the dengue virus (DENV) and other arboviruses. Although a complete mechanistic understanding of pathogen blocking is lacking, it is likely to involve host immunity induction and resource competition between *Wolbachia* and DENV, both of which may be impacted by microbiome composition. The potential impact of *Wolbachia* transinfection on host fitness is also of importance given the widespread release of mosquitos infected with the *Drosophila melanogaster* strain of *Wolbachia* (*wMel*) in wild populations. Here, population-level genomic data from *Ae. aegypti* was surveyed to establish the relationship between the density of *wMel* infection and the composition of the host microbiome.

RESULTS: Analysis of genomic data from 172 wild *Ae. aegypti* females across six subpopulations resulted in an expanded and quantitatively refined, species-level characterization of the bacterial, archaeal, and fungal microbiome. This included 844 species of bacteria across 23 phyla, of which 54 species were found to be ubiquitous members of the population core microbiome. The density of *wMel* infection was highly variable between individuals and negatively correlated with microbiome diversity. Network analyses revealed *wMel* as a hub comprised solely of negative interactions with other bacterial species. This contrasted with the large and highly interconnected network of other microbiome species that are likely to represent members of the midgut microbiome community in this population.

CONCLUSION: Our bioinformatic survey provided a species-level characterization of *Ae. aegypti* microbiome composition and variation in a wild population. *wMel* load varied substantially

across populations and individuals and, importantly, *w*Mel was a major hub of a negative interactions across the microbiome. These interactions may be an inherent consequence of heightened pathogen blocking in densely infected individuals or, alternatively, may result from “*Wolbachia*-incompatible” bacteria that could impede the efficacy of *w*Mel as a biological control agent in future applications. The relationship between *w*Mel infection variation and the microbiome warrants further investigation in the context of developing *w*Mel as a multivalent control agent against other arboviruses.

2.2 Introduction

Aedes mosquitoes are a primary vector for dengue (DENV) and other arboviruses, including Zika, chikungunya, and yellow fever¹. Amongst these, DENV transmission poses a particularly severe risk to human health as it is responsible for 50-100 million infections annually across the approximately 125 countries in which it is endemic²⁻⁵. Above and beyond existing strategies for *Aedes* population control (*e.g.*, insecticides and other chemical or physical management mechanisms) alternative approaches are required to achieve sustained disease reduction⁶. Recently, substantial reductions in DENV transmission have been achieved through the release of mosquitoes stably infected with the bacterial symbiont *Wolbachia*⁷⁻¹¹, which confers resistance to DENV and other arboviruses (commonly referred to as “pathogen blocking”)¹²⁻¹⁵.

Wolbachia pipiensis is a maternally inherited gram-negative endosymbiotic bacterium that is estimated to naturally infect at least two-thirds of arthropod species¹⁶. *Wolbachia* facilitates its spread within populations through a variety of mechanisms that manipulate host reproduction, including cytoplasmic incompatibility, male feminization, male killing, and the induction of parthenogenesis¹⁷. The *Wolbachia* strain *w*Mel, which naturally infects *Drosophila*

melanogaster, has been developed to stably transinfect *Ae. aegypti*. In this host, wMel induces cytoplasmic incompatibility, whereby uninfected males can produce viable offspring with wMel-infected females but matings between wMel-infected males with uninfected females result in embryonic death¹⁸. Thus, in the absence of fitness reductions in infected individuals, wMel tends to spread rapidly and become stably maintained in populations^{19,20}.

Two primary mechanisms are believed to form the basis of wMel induced pathogen blocking: (1) induction of innate host immunity and (2) resource competition between microbes²¹. Pathogen blocking due to wMel-induced immunity is supported by several observations. First, wMel confers a partial antiviral effect against RNA viruses in *D. melanogaster* and viral inhibition in mosquito cell lines^{22,23}. Second, wMel infection in *Ae. aegypti* induces the upregulation of the *Toll* and *Ind* immunity pathways that target pathogens for removal through the production of antimicrobial peptides²⁴⁻²⁶. Third, a mechanistic link between immunity induction and pathogen blocking is further supported by the fact that DENV infection in *Ae. aegypti* triggers a *Toll* pathway-mediated response and that *Wolbachia* density itself decreases during DENV infection^{27,28}. The second mechanism implicated in pathogen blocking relates to competition for host resources²⁹⁻³¹. As an obligate endosymbiont, *Wolbachia* (much like viruses) relies on a suite of host resources. For example, neither *Wolbachia* nor DENV are able to independently synthesize cholesterol³⁰. wMel infection in *Ae. aegypti* increases cholesterol storage in lipid droplets and inhibits viral replication³¹. Release of cholesterol back into the cytosol reverses this inhibition, and pathogen blocking is also reduced in other insects when raised on a high-cholesterol diet²⁹. It is also important to note that the strength of pathogen blocking is variable and has been found to positively correlate with variation of *Wolbachia* infection density across strains^{32,33}. For example, the higher-density

infections characteristic of *wMelPop*, relative to *wMel*, is particularly effective in reducing *Ae. aegypti* DENV vectoring capacity across DENV serotypes³⁴.

The density and distribution of *Wolbachia* infection across tissues may also be dependent on interactions with the *Aedes* microbiome, which in turn may impact the efficacy of *Wolbachia* as a biological control agent. In fact, several commensal microbes, such as species in the genera *Serratia* and *Asaia*, have already been demonstrated to inhibit stable *Wolbachia* transinfection^{35–37}. As such, a more wholistic understanding of the tripartite relationships between the *Ae. aegypti* host, *Wolbachia* and the microbiome is a priority. A previous investigation of *wMel* infection on the *Ae. aegypti* microbiome found no decrease in compositional richness, but it did reveal a reduction in a suite of low abundance bacterial taxa³⁸. Furthermore, no changes were observed among species with known *wMel* incompatibilities (e.g., *Serratia* and *Asaia*). It is important to note, however, that this study utilized laboratory-raised *Ae. aegypti* and it is unclear to what extent these results apply to wild populations. Here, we aim to address this uncertainty through the establishment of a refined understanding of *Ae. aegypti* microbiome variation in a wild population subject to previous releases of *wMel* infected individuals.

To assess the influence of *wMel* infection on microbiome composition, we analyzed short-read genomic DNA sequence data from 172 wild-caught *Ae. aegypti* females. This analysis included quantitative estimates of *wMel* infection density across six subpopulations, four of which had historic releases of *wMel*-infected females. Our approach (1) achieved microbiome classifications at the species-level for 94.6% of the nearly 21 million reads that mapped to bacterial genomes and greatly expanded upon previous microbiome characterizations in this species, (2) quantified microbiome variation while also identifying core bacterial species at the population-level, (3) identified substantial variation in the density of *wMel* infection across

individuals and (4) revealed that *wMel* was a central hub of negative interactions within a larger *Ae. aegypti* microbiome interaction network.

2.3 Methods

Ae. aegypti genome sequencing data

Double digest restriction-site associated DNA sequencing data (ddRAD-seq) were utilized from a study of wild *Ae. aegypti* populations from six geographic locations near Cairns, Australia (NCBI BioProject PRJNA412140)²⁰. Releases of *Ae. aegypti* transinfected with the *D. melanogaster* strain of *Wolbachia* (*wMel*) were conducted in four of these locations. Illumina HiSeq2500 (100bp reads) paired-end ddRADseq library sequencing of genomic DNA was conducted from 172 trapped female mosquitoes. Metadata for each sample included information regarding the date and location of collection.

Taxonomic classification *Ae. aegypti* microbiome

The Kraken2 pipeline³⁹ was used to assign taxonomic classifications to all raw ddRAD-seq reads that met standard quality metrics. The Kraken2 search database included all available RefSeq genomes from bacteria (n = 21,333), fungi (n = 60), protozoa (n = 40), viruses (n = 10,388), archaea (n = 390) (NCBI RefSeq Release Number 200 on August 26, 2020) and the recently reannotated *Ae. aegypti* reference genome (AaegL5⁴⁰). The database was built with 512GB of available RAM on 4 threads using the default Kraken2 parameters for minimizer length and spaces, and a conservative k-mer length of 31. Low complexity regions of genomes were masked using the DustMasker program⁴¹. As expected, the vast majority of reads (93.6% average across 172 individuals; SD = 2.6%) mapped to the *Ae. aegypti* genome.

Quantification of Ae. aegypti microbiome composition

We further curated the reads that mapped to bacterial, viral, archaeal, fungal, or protozoan genomes. After taxonomic classification at the species level, Bracken⁴² was used quantify species abundances per individual. The Bracken database was built using default parameters and a read length of 100. Following Bracken quantification, species abundance estimates were standardized to account for variation in the overall sequencing depth per individual. Additionally, we controlled for the number of restriction enzyme cut sites per genome, we used a custom python script that calculated the total number of *NlaIII* (CATG/GTAC) and *MluCI* (AATT/TTAA) enzyme recognition sites in each reference genome. Standardizations were conducted separately in each major taxonomical group (*e.g.*, bacterial, fungi, etc.). For subsequent analyses, rare species were removed from the dataset if they (1) did not constitute >1% of the total bacterial microbiome in any individual and (2) were present in fewer than 25% of individuals. We note that our analysis is based on DNA isolated from whole mosquitos. It therefore reflects the sum of DNA from microbial populations per individual and it does not provide direct information about tissue-specific microbiome composition.

Statistical analysis of bacterial microbiome composition

Microbiome complexity was assessed using two alpha diversity metrics: observed species richness and Shannon's diversity index⁴³. A linear model was used to assess relationships between alpha diversity measures and total *wMel* abundance per individual. An analysis of variance (ANOVA) and Tukey's Post Hoc test was used to compare the density of *wMel* across subpopulations to sites with historical releases. Co-occurrence and mutually-exclusive interactions between bacterial species were assessed using Pearson's correlations as implemented

by CoNet⁴⁴ in Cytoscape⁴⁵ on a standardized bacterial species abundance matrix. The matrix was standardized using the Wisconsin function from the R package vegan (version 2.5-7⁴⁶). Statistical significance of each pairwise comparison was determined using 1000 bootstrap replications and Benjamini-Hochberg multiple testing correction. The network was visualized using the Cytoscape GUI software⁴⁵. Additional analysis of network composition was based on delineating two subnetworks. The first (hereafter referred to as the “wMel subnetwork”) includes those species with a significant edge interaction directly with wMel and the second (hereafter referred to as the “non-wMel subnetwork”) as those species without a direct edge interaction with wMel.

Principal Coordinate Analysis (PCoA) was then used to further characterize bacterial microbiome variation. A distance matrix was calculated by applying the Bray-Curtis dissimilarity index⁴⁷ implemented by the vegdist function from the R package vegan (version 2.5-7⁴⁶) to the same standardized matrix used the CoNet analysis (see above). The PCoA was performed using the pcoa function from the R package ape (version 5.5⁴⁸). Axes of variation were further investigated using linear models to assess the relationship between Principal Coordinate (PC) axis loading values and standardized species abundance estimates.

2.4 Results

Ae. aegypti microbiome composition

The overarching goal of this study was to bioinformatically characterize microbiome variation, including that of wMel, in a wild population of *Ae. aegypti*. To this end, DNA-seq reads from 172 *Ae. aegypti* females were mapped to a database containing the *Ae. aegypti* genome and the full set of RefSeq genomes from bacteria, fungi, viruses, archaea, and protozoans (27,435

genomes in total). Data from each individual resulted in an average of 18.7 million mapped reads ($SD = 220,395$), of which the vast majority mapped to the *Ae. aegypti* reference genome (93.6%; $SD = 2.6\%$). On average, 5.4% ($SD = 1.1\%$) of reads remained unclassified because they were either low quality reads or derived from species not included in our database. The remaining reads (185,811 reads per individual on average) mapped to bacterial, viral, archaeal, fungal, or protozoan genomes. Of these, the vast majority (90.6%; $SD = 5.9\%$) were derived from bacteria (Fig. 1A). Fungi, protozoa, and archaea were consistently identified at low abundances, while viral load was more variable (Additional File 1). As expected, given that our analysis was based on DNA sequencing data, 99.8% of all viral reads identified from bacteriophage (*i.e.*, DNA viruses) and the observed viral load variation was due to the high abundance of *Escherichia* bacteriophages in a limited number of individuals. Full information regarding fungal, viral, archaeal, and protozoan identifications can be found in Additional File 1.

Bacterial microbiome composition

After removing low abundance and rare species, 843 bacterial species remained from 23 phyla and 396 genera (Additional File 2; note that wMel is not included in the following analysis of microbiome composition). These results represent a substantial increase in the size and taxonomical specificity of the characterized *Ae. aegypti* microbiome. Eight phyla comprised >99.0% of total bacterial composition (Fig. 1B). Consistent with previous studies, all of which relied on 16S rRNA sequencing^{38,49–53}, Proteobacteria (49.0% of bacterial reads) was identified as the predominant phylum. Firmicutes (36.6%), Tenericutes (5.9%), Bacteroidetes (4.7%), Actinobacteria (1.5%), Cyanobacteria (0.6%), Spirochaetes (0.5%), and Fusobacteria (0.5%) were also identified as substantive contributors to the microbiome (Fig. 1C). At the genus level,

our analysis confirmed the presence of several previously reported groups including, (1) *Acinetobacter* and *Stenotrophomonas*, which have been identified in reproductive tissues, (2) *Asaia*, *Bacillus*, *Chryseobacterium*, *Enterobacter* and *Klebsiella*, which are commonly identified in the midmidgut, (3) *Burkholderia*, which has been identified specifically in the salivary glands and (4) a suite of six genera with more complex spatial distributions across multiple tissue microbiomes^{38,49–52,54–56}. We also identified several novel species within the *Mycoplasma*, *Salmonella*, and *Mannheimia* genera that were either ubiquitous among individuals or highly abundant. Finally, several taxa (*e.g.*, species in the *Asaia*, *Spiroplasma* and *Serratia* genera) have been demonstrated to negatively influence *Wolbachia* density or transmission^{35,36,57}. Our analysis did not identify any species in the genus *Asaia*, consistent with Audsley *et al.* (2017), but it did identify several species in the genera *Spiroplasma* and *Serratia*.

Towards a core Ae. aegypti bacterial microbiome

Our analysis provided a unique opportunity to leverage intraspecific variation to assess the composition of the core *Ae. aegypti* microbiome at a population level. First, we investigated core microbiome composition as a function of species presence across the population. This analysis revealed that 164 species were present in 95% of individuals and 54 species (hereafter referred to as the “core microbiome”) were present consistently across the entire population (Fig. 2A; Additional File 3). Species present in large numbers of individuals also constituted a high proportion of total microbiome composition (Fig 2B). Specifically, the 517 bacterial species identified in at least 75% of individuals comprised 93.0% of the total bacterial microbiome. The remaining 7.0% of the microbiome was a combination of 326 additional species that were of far lower abundance on average. Among the 54 species comprising the core microbiome, five

species constituted more than 5% of the core microbiome on average (Fig. 2C). *Bacillus subtilis* (42.4%) had the highest average abundance by far, followed by *Salmonella enterica* (7.9%), *Staphylococcus aureus* (6.2%), *Escherichia coli* (6.1%), and *Enterococcus faecium* (5.5%). The high abundance of each of these species was likely due to their contribution to the midgut microbiome (*Bacillus subtilis*, *Salmonella* sp., *Staphylococcus* sp., *Escherichia* sp., and *Enterococcus* sp. have all been previously identified in the *Ae. aegypti* midgut microbiome^{54,58–62}). Species in the genera *Acinetobacter*, *Clostridium*, *Cupriavidus*, *Klebsiella*, *Serratia* were also identified as members of the core microbiome. These genera have all been previously identified in mosquitoes but were not associated with a particular tissue^{63–66}. Unexpectedly, 62.3% (or 34 species) of the core microbiome was comprised of previously unidentified species in the *Ae. aegypti* microbiome (Additional File 3).

Population distribution of wMel

Historic releases of *wMel*-infected females were conducted in four of the six subpopulations analyzed here, within which stable *wMel* inheritance has already been documented²⁰. Based on recent observations from controlled releases, which revealed very limited geographic dispersal of transinfected individuals⁸, we predicted a substantively higher prevalence of *wMel* in locations with historic releases. This was confirmed by an ANOVA that showed significant variation between the *wMel* across all sites ($p = 0.0006$), including significantly higher abundance levels in several of the populations with historic releases relative to those without historic releases (Fig. 3A). Nevertheless, a substantial amount of intraspecific variation in *wMel* abundance was also observed in populations with historic releases. Overall, average *wMel* abundance in these populations was 460,458 (coefficient of variance = 1.05) and ranking individuals by *wMel*

abundance revealed a largely continuous gradient of infection density with the highest density infections exceeded 1.5 million reads (Fig. 3B). Given that the strength of pathogen blocking is dependent on the density of *Wolbachia* infection³³, we next sought to assess the relationship between such variation and bacterial microbiome composition.

wMel was a hub of mutually exclusive relationships with other microbiome species

wMel abundance did not correlate with either total bacterial abundance ($p = 0.66$) or observed species richness ($p = 0.89$). However, there was a significant inverse relationship between *wMel* abundance and Shannon's Index ($p = 0.02$). Thus, as *wMel* abundance increased, the relative abundance of other bacterial species in the microbiome became less uniform. A previous investigation of *wMel*-infected *Ae. aegypti* revealed a reduction in several low abundance genera across the microbiome, but it did not report information about *wMel* infection variation³⁸. To specifically explore the effect of *wMel* abundance, we constructed an interaction network to assess patterns of co-occurrence and mutual exclusion within the microbiome. This approach revealed a network of 71 species, which was significantly enriched for core microbiome species (obs. = 42; exp. = 4.6; $\chi^2 = 355.7$; $p = 2.4 \times 10^{-79}$) and comprised two predominant subnetworks (Fig. 4A; Additional File 4). The first subnetwork (hereafter referred to as the “*wMel* subnetwork”) comprised 35 mutually-exclusive relationships specifically with *wMel* and had a low relative level of connectivity (average of 1.5 edges per node). The second, larger subnetwork (hereafter referred to as the “non-*wMel* subnetwork”) comprised 56 species and was notable for a higher average level of connectivity (2.2 edges per node) and both co-occurrence ($n = 66$) and mutually exclusive ($n = 58$) interaction edges. Three additional characteristics further distinguished these subnetworks. First, the *wMel* subnetwork comprised species identified at

significantly lower abundances than those in the non- *w*Mel subnetwork (Fig. 4B) despite the proportion of core microbiome species being indistinguishable between the two ($\chi^2 = 0.45$; $p = 0.50$). Second, interaction edges in the *w*Mel subnetwork were significantly weaker in their absolute magnitude than those in the larger network (Fig. 4C; $p = 8.38 \times 10^{-6}$). Third, the larger network included *B. subtilis*, a confirmed *Ae. aegypti* midgut microbe and the highest abundance species in this study, as well as numerous other high abundance taxa previously identified in the *Ae. aegypti* midgut (e.g., *B. cereus/thuringiensis*, *E. faecium*, *E. coli*, *S. aureus/simulans/cohnii*, *S. enterica*). Thus, the larger subnetwork was more robustly interconnected and contained a diversity of abundant taxa from the midgut microbiome. In contrast, *w*Mel serves as a hub of negative pairwise relationships with relatively low abundance species, even though many of these are part of the core microbiome. We propose that this structure may derive from more direct interactions between *w*Mel and members of the reproductive microbiome.

Microbiome independent and dependent components of wMel variation

To assess overall patterns of variation within the microbiome, we utilized a Principal Coordinate Analysis and observed three main axes that captured 76.34% of the total variation (Fig. 5). The first axis (58.73% of variation) included *w*Mel as the eigenvector with the largest magnitude (Fig. 5A). The contribution of *w*Mel variation to this axis was supported by the fact that Axis 1 loadings exhibited a significant correlation with *w*Mel abundance across individuals ($r = -0.90$; $t = -26.78$, $p = 4.6 \times 10^{-63}$; Fig. 5B). Thus, Axis 1 largely captures variation in *w*Mel and this variation was largely orthogonal to the other main eigenvectors. However, we note that, of the 10 species with the highest positive loadings on Axis 1 (*i.e.*, those with loadings opposite to *w*Mel), five participate in direct negative interactions with *w*Mel in the *w*Mel subnetwork (Fig.

4A). Axis 2 also identified wMel as a large negative eigenvector but, in contrast to Axis 1, it included two species with large positive eigenvectors (*i.e.*, *M. haemolytica* and *B. subtilis*; Fig. 5C). The relationship with these species was further supported by a significant positive correlation between Axis 2 loadings and the abundance of wMel ($r = -0.24$; $t = -3.21$, $p = 0.0016$), and a significant negative correlation with *M. haemolytica* ($r = 0.72$; $t = 13.49$, $p < 2.2 \times 10^{-16}$) and *B. subtilis* ($r = 0.48$; $t = 7.08$, $p = 3.613 \times 10^{-11}$; Fig. 5D). Notably, both species were present in our network analysis (Fig. 4A) with their interaction connections to wMel were consistent with the identified PCoA relationships. In summary, Axis 1 captured wMel variation and its weak, mutually-exclusive interaction with a suite of species identified in the wMel network, and Axis 2 captured stronger, but indirect, wMel covariation with two of the most abundant species in the microbiome, including the midgut microbe *B. subtilis*.

2.5 Discussion

The release of *Ae. aegypti* mosquitoes transinfected with the wMel strain of *Wolbachia* has proven to be an effective strategy to limit the spread of DENV^{12,67}. However, less is known about the long-term dynamics of infection within populations and how these dynamics might influence the multi-valent utility of wMel more broadly as a biological control agent for arboviruses. Given the mechanisms responsible for pathogen blocking (*i.e.*, immune priming and resource competition), it is reasonable to assume that wMel infection stability and density may be dependent on interactions with the remainder of the microbiome. Here, we explore this possibility through a population-level *Ae. aegypti* microbiome survey using a bioinformatic approach that leverages available genomic sequencing data. In addition to achieving a species-level *Ae. aegypti* microbiome characterization from a wild population, our analyses revealed

substantial intraspecific variation in *wMel* density that covaried with microbiome composition. This variation is of potential relevance to the efficacy of pathogen blocking and the implementation of *Wolbachia*-based pathogen control strategies against DENV and other arboviruses.

Towards an Ae. aegypti population-level core microbiome

Interactions between bacterial components of the microbiome and their mosquito host impact important aspects of fitness, including fertility, longevity, and immunity^{14,35,53}. As such, the identification of a population-level, core microbiome – those bacterial species present across many or all individuals – has the potential to provide unique insights into taxa that support host development and function⁶⁸. Importantly, and unlike previous studies that have largely relied on laboratory populations^{50,56,65,69,70}, our survey provides a unique perspective into microbiome structure at a refined taxonomical level for a wild *Ae. aegypti* population. Despite the considerable microbiome variation attributable to environmental factors, blood meal status and age, our analysis was successful in identifying a relatively large repertoire of core microbiome species. Whereas functional analyses will ultimately be required for a refined understanding of core microbiome-host dynamics⁷¹, several observations are worthy of discussion.

The 54 bacterial species ubiquitously present across all individuals comprised only 24% of the total microbiome. However, when the criteria for inclusion was relaxed (to presence in ≥75% of individuals), this group of common species comprised 93% of the total microbiome. As such, the vast majority of the microbiome is comprised of common species at the population level. This observation cannot be accounted for by a possible identification bias towards highly abundant species. Whereas many core species were indeed highly abundant, such as *B. subtilis*

that comprises 11.21% of the microbiome on average, the abundance of core species varied by over 1,100-fold (Fig. 4C). Thus, many core species were quite low in relative abundance. This variation is almost certainly explained, at least in part, by the absolute size of tissue-specific microbiomes. We therefore predict that the low average abundance of species in the *wMel* subnetwork is likely due to their restricted presence in the reproductive microbiome. Similarly, the high average abundance of species in the non-*wMel* subnetwork, which includes *B. subtilis*, suggests that they may be representative of a much larger midgut microbiome. Overall, the significant enrichment of core microbiome species (of both high and low abundance) in our network analysis supports the presence of concerted functional assemblages within the microbial community. Furthermore, the topology of the network, including the presence of delineated subnetworks, suggests compartmentalization between tissues that may inform tissue-specific microbiome functionality. Some of core species identified here have also been identified in laboratory colonies derived from widespread global populations (*e.g.*, Brazil, Grenada, and India^{56,61,62,72}). Surveys across addition natural and lab populations will further refine our understanding of the functional core microbiome and, additionally, the process by which natural microbiomes experience compositional shifts or diversity-decay when reared in laboratory settings.

Variation in wMel infection density and the microbiome

As a reproductive endosymbiont, *Wolbachia* was conventionally believed to be concentrated within the host's reproductive tissues. However, it is now well-established that *Wolbachia* can be present at variable levels across many non-reproductive tissues^{13,32}. The density and spatiotemporal distribution of *Wolbachia* varies among host species and strains of *Wolbachia*^{32,73},

but there is limited information about the spatiotemporal dynamics of *wMel* in wild *Ae. aegypti* populations despite its relevance to pathogen blocking³³. Despite *wMel* being stably established in the populations surveyed here²⁰, we observed a high level of variation in infection density across individuals. Although establishing the cause of this variation is beyond the scope of the current study, we note that *wMel* density is dependent on temperature during egg and larval development⁷⁴.

Our investigation revealed robust covariation between *wMel* abundance and specific sets of species with the microbiome. Several facets of these relationships are worth highlighting. First, we observed a significant negative relationship between *wMel* density and Shannon's Index, indicating a decrease in the compositional evenness in the microbiome. Previous studies comparing alpha diversity metrics in *Wolbachia*-infected hosts have reported contradicting results^{38,75-77}. Our results, based on a large survey of individuals, suggest that the impact of *wMel* on the *Ae. aegypti* microbiome is more nuanced than a simple reduction in species richness or load, as previously suggested^{38,75,76}. Second, *wMel* served as a hub of consistently mutually-exclusive, but weak, interactions and these interactions were generally with lower abundance species. This pattern is consistent with results reported by Audsley *et al.* (2017). Based on the composition of the subnetworks revealed in our analysis, we speculate that this pattern may reflect a suite of direct interactions with other members of the reproductive microbiome. It is also important to emphasize that, although these species were generally low in abundance, they were significantly enriched for members of the core microbiome. Third, our PCoA analysis revealed robust covariation between *wMel* and several of the most abundant species in the microbiome, including the midgut microbe *B. subtilis*. Given the importance of the midgut

microbiome on immune system function⁷⁸, the effect of *wMel* (either direct or indirect) on the midgut microbiome is worthy of further investigation.

2.6 Conclusion

We demonstrate the underappreciated value of existing whole-organism DNA-seq data in relation to microbiome characterization. Our analysis achieved a high coverage, species-level characterization of the *Ae. aegypti* microbiome, including the delineation of a population-level, core microbiome. Despite the absence of information about tissue-specific microbiome composition, network analyses revealed *wMel* as a hub of interactions with species likely to be of the reproductive microbiome and a small set of robust, indirect interactions with likely members of the midgut microbiome. Intraspecific variation in *wMel* infection density, in conjunction with diverse interactions with the native microbiome, may have the potential to impact vector competency. The specific interactions identified here could be leveraged to potentially enhance *wMel* infection density and identify species that may be “*Wolbachia-incompatible*” or third-party players that mediate pathways underlying pathogen blocking.

2.7 Figures

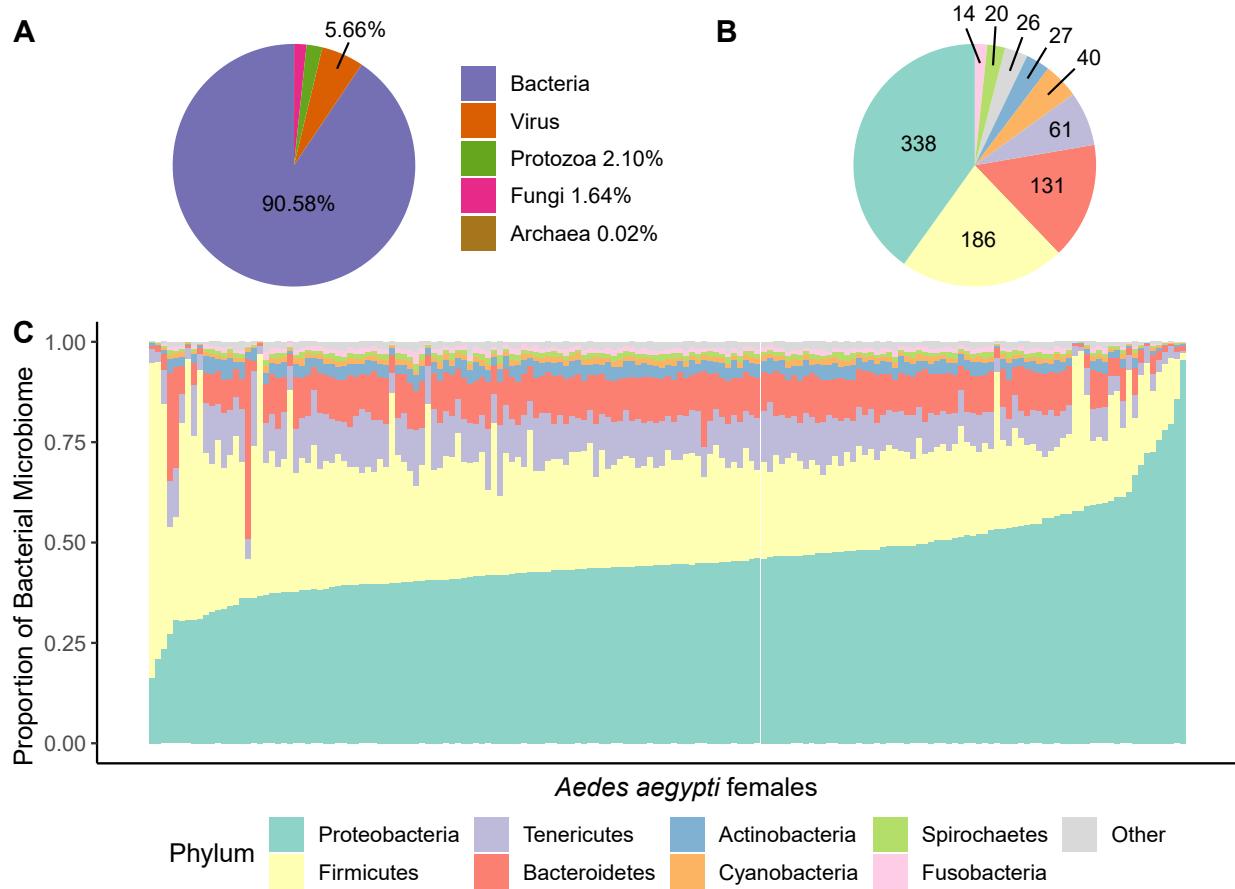


Figure 1. Population-level *Ae. aegypti* microbiome characterization. (A) Taxonomic distribution of classified reads mapping to non-*Ae. aegypti* genomes across the 172 individuals sampled. (B) Species-level distribution of the major phyla in the bacterial microbiome across the 172 individuals sampled. (C) Bacterial microbiome composition by phyla for each individual female sample. The eight most abundant phyla are shown and the remaining are grouped as ‘Other’.

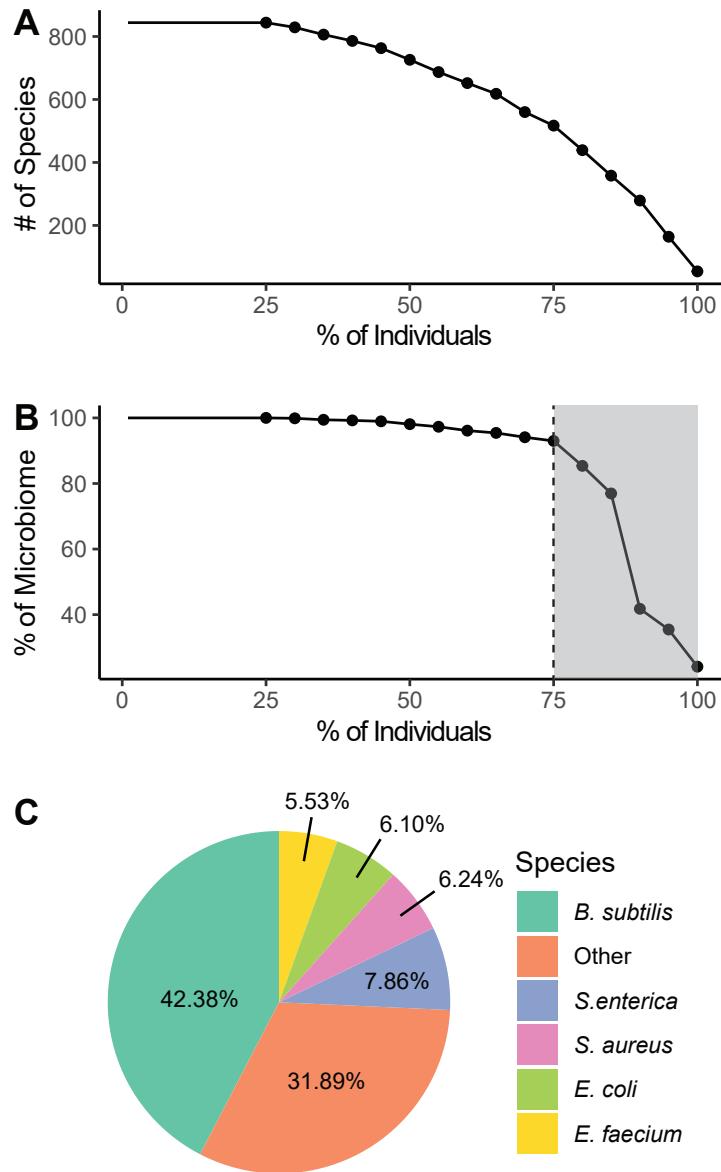


Figure 2. *Ae. aegypti* core microbiome. (A) The number of bacterial species identified is plotted in relation to the proportion of individuals in which each species was identified. (B) The percent of microbiome composition is plotted in relation the proportion of individuals in which each species was identified. Those species present in 75% or more individuals comprised 93% of total microbiome composition. 54 species were identified in all 172 individuals. (C) Contribution of the 54 ubiquitously present species to core microbiome composition. Species that do not constitute >5% of the core microbiome have been grouped as ‘Other.’

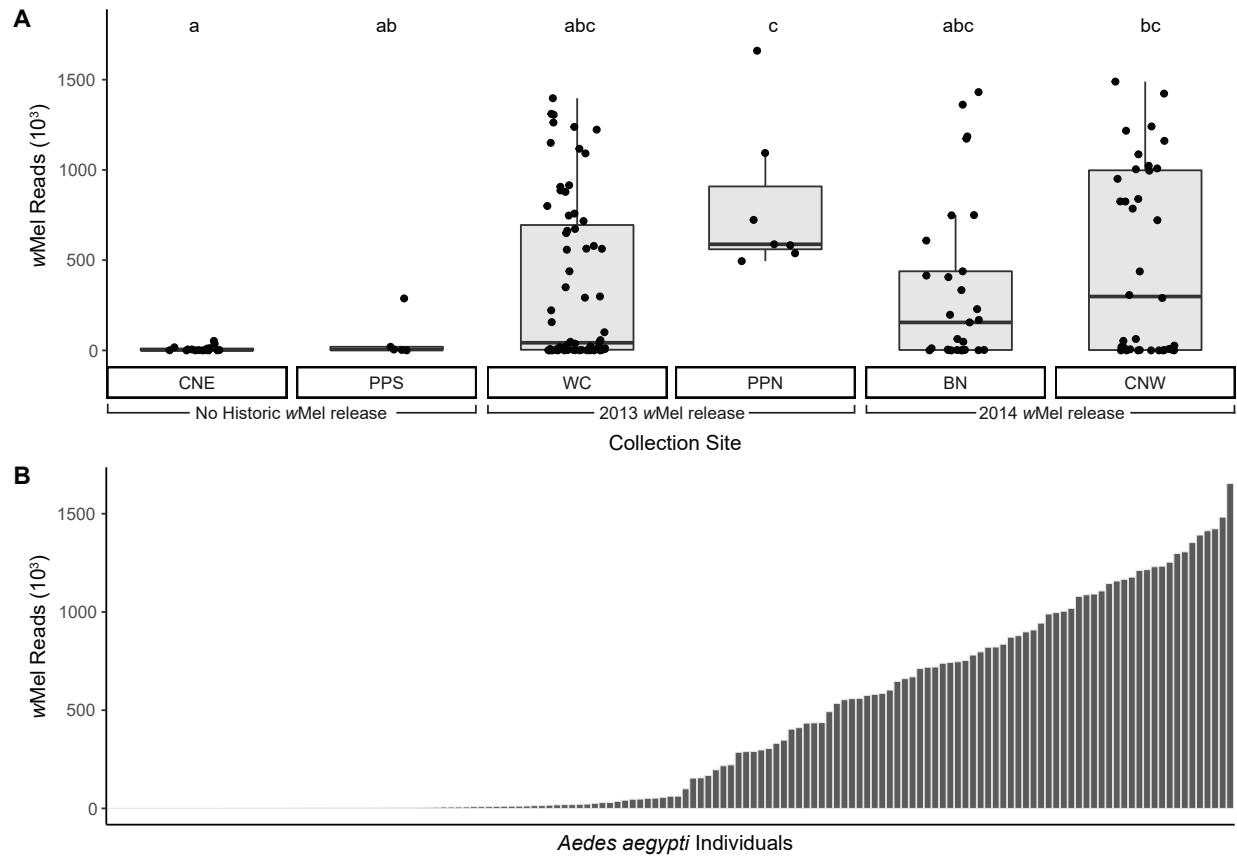


Figure 3. Variation in *wMel* abundance across populations and individuals. (A) *wMel* abundance (normalized read count) per individual for each collection site: Cairns North East (CNE), Parramatta Park South (PPS), Westcourt (WC), Parramatta Park North (PPN), Bungalow (BN), and Cairns North West (CNW)²⁰. An ANOVA indicated significant variation between the *wMel* abundance across sites, $p = 0.0006$. A post-hoc Tukey HSD Test revealed that the average *wMel* abundance in CNE-collected individuals was significantly less than those from PPN ($p = 0.0046$) and CNW ($p = 0.0013$). Additionally, samples collected from PPS had a significantly lower average *wMel* abundance than those from PPN ($p = 0.0490$). (B) Individuals from all populations ranked by total *wMel* abundance.

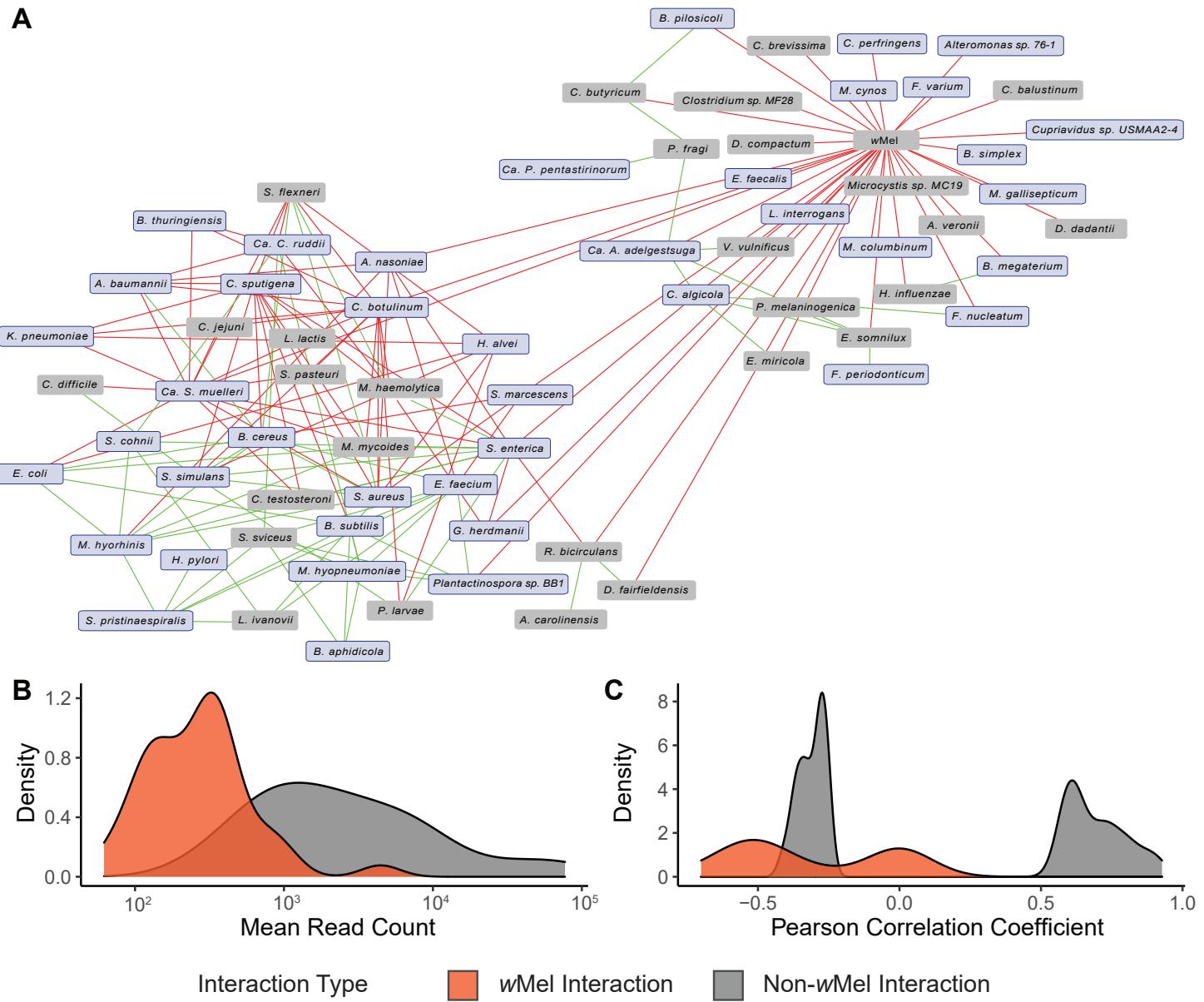


Figure 4. *Ae. aegypti* microbiome interaction network. (A) Microbiome interaction network analysis using Wisconsin normalized read counts resulted in a network of 71 species and 159 edges. Red edges represent mutual exclusion (negative Pearson's correlation coefficient) and green edges represent co-occurrence (positive Pearson's correlation coefficient) interactions of mosquitos infected with *wMel*. Core microbiome species are highlighted in purple. *wMel* forms a negative interaction hub within the *Ae. aegypti* microbiome network and was involved exclusively in negative pairwise interactions with 35 species. (B) Density plot displaying the mean read count distribution for species included in the network. Species with mutually exclusive interactions with *wMel* are shown in orange and have a significantly lower average read count ($t = 2.43$; $p = 0.02$) than all other species in the network (grey). (C) Density plot displaying distribution of the strength of interactions edges with *wMel* (orange) and the remaining edges in the network not involving *wMel* (grey).

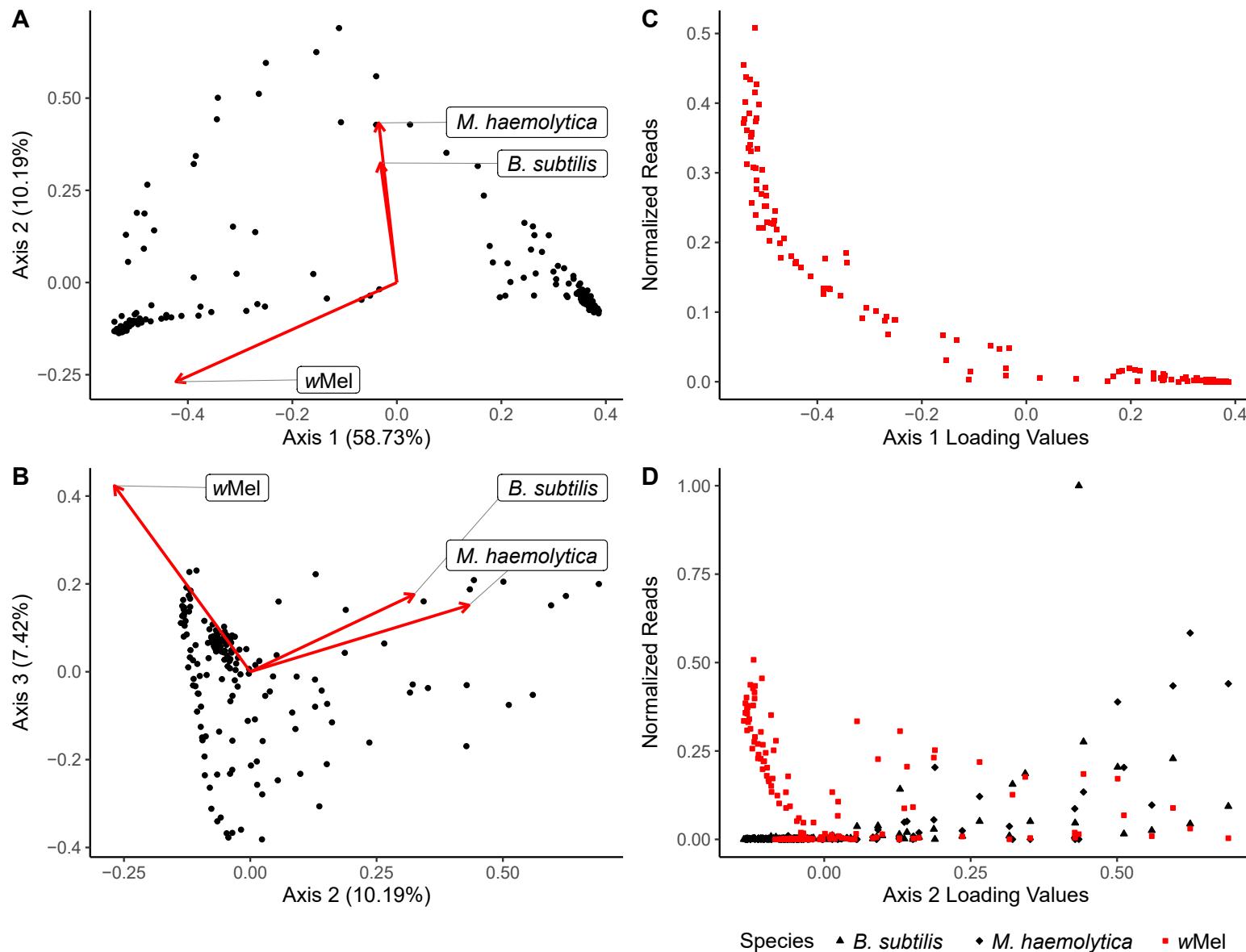


Figure 5. Covariance between microbiome composition and wMel abundance. (A) PCoA of microbiome variation identified a primary axis of variation (PC1: 58.73%) that corresponded with the eigenvector for wMel which was the largest observed vector. Other large eigenvectors representing highly abundant taxa were largely orthogonal to PC1. (B) PC1 loadings were significantly correlated with wMel abundance across individuals but not other microbiome species. (C) PCoA analysis of microbiome variation identified a secondary axis of variation (PC2: 10.19%) which corresponded with the eigenvectors for *B. subtilis*, *M. haemolytica*, and to a lesser extent wMel. (D) PC2 loadings were significantly correlated with *B. subtilis* and *M. haemolytica*.

2.8 References

1. Powell, J. R. Mosquito-borne human viral diseases: Why *aedes aegypti*? *Am. J. Trop. Med. Hyg.* **98**, 1563–1565 (2018).
2. Murray, N. E. A., Quam, M. B. & Wilder-Smith, A. Epidemiology of dengue: Past, present and future prospects. *Clinical Epidemiology* **5**, 299–309 (2013).
3. World Health Organization (WHO) & Special Programme for Research and Tropical Diseases (TDR). Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control. in *Dengue: Guidelines for Diagnosis, Treatment, Prevention and Control: New Edition* (World Health Organization, 2009).
4. Stanaway, J. D. *et al.* The global burden of dengue: an analysis from the Global Burden of Disease Study 2013. *Lancet Infect. Dis.* **16**, 712–723 (2016).
5. Cattarino, L., Rodriguez-Barraquer, I., Imai, N., Cummings, D. A. T. & Ferguson, N. M. Mapping global variation in dengue transmission intensity. *Sci. Transl. Med.* **12**, (2020).

6. Achee, N. L. *et al.* A Critical Assessment of Vector Control for Dengue Prevention. *PLoS Negl. Trop. Dis.* **9**, e0003655 (2015).
7. Anders, K. L. *et al.* Reduced dengue incidence following deployments of Wolbachia-infected *Aedes aegypti* in Yogyakarta, Indonesia: A quasi-experimental trial using controlled interrupted time series analysis. *Gates Open Res.* **4**, (2020).
8. Utarini, A. *et al.* Efficacy of Wolbachia-Infected Mosquito Deployments for the Control of Dengue. *N. Engl. J. Med.* **384**, 2177–2186 (2021).
9. Nazni, W. A. *et al.* Establishment of Wolbachia Strain wAlbB in Malaysian Populations of *Aedes aegypti* for Dengue Control. *Curr. Biol.* **29**, 4241-4248.e5 (2019).
10. O'Neill, S. L. *et al.* Scaled deployment of Wolbachia to protect the community from dengue and other aedes transmitted arboviruses. *Gates Open Res.* **2**, (2018).
11. Ryan, P. A. *et al.* Establishment of wMel Wolbachia in *Aedes aegypti* mosquitoes and reduction of local dengue transmission in Cairns and surrounding locations in northern Queensland, Australia. *Gates Open Res.* **3**, 1547 (2020).
12. Walker, T. *et al.* The wMel Wolbachia strain blocks dengue and invades caged *Aedes aegypti* populations. *Nature* **476**, 450–455 (2011).
13. Moreira, L. A. *et al.* A Wolbachia Symbiont in *Aedes aegypti* Limits Infection with Dengue, Chikungunya, and Plasmodium. *Cell* **139**, 1268–1278 (2009).
14. McMeniman, C. J. *et al.* Stable introduction of a life-shortening Wolbachia infection into the mosquito *Aedes aegypti*. *Science* **323**, 141–144 (2009).
15. Flores, H. A. *et al.* Multiple Wolbachia strains provide comparative levels of protection against dengue virus infection in *Aedes aegypti*. *PLoS Pathog.* **16**, e1008433 (2020).
16. Hilgenboecker, K., Hammerstein, P., Schlattmann, P., Telschow, A. & Werren, J. H. How

many species are infected with Wolbachia? - A statistical analysis of current data. *FEMS Microbiol. Lett.* **281**, 215–20 (2008).

17. Correa, C. C. & Ballard, J. W. O. Wolbachia Associations with Insects: Winning or Losing Against a Master Manipulator. *Front. Ecol. Evol.* **3**, 153 (2016).
18. Turelli, M. & Hoffmann, A. A. Rapid spread of an inherited incompatibility factor in California Drosophila. *Nature* **353**, 440–442 (1991).
19. Hoffmann, A. A. *et al.* Successful establishment of Wolbachia in Aedes populations to suppress dengue transmission. *Nature* **476**, 454–457 (2011).
20. Schmidt, T. L., Filipović, I., Hoffmann, A. A. & Rašić, G. Fine-scale landscape genomics helps explain the slow spatial spread of Wolbachia through the Aedes aegypti population in Cairns, Australia. *Heredity (Edinb)*. **120**, 386–395 (2018).
21. Caragata, E. P., Dutra, H. L. C. & Moreira, L. A. Exploiting Intimate Relationships: Controlling Mosquito-Transmitted Disease with Wolbachia. *Trends in Parasitology* **32**, 207–218 (2016).
22. Hedges, L. M., Brownlie, J. C., O'Neill, S. L. & Johnson, K. N. Wolbachia and Virus Protection in Insects. *Science* **322**, 702–702 (2008).
23. Bhattacharya, T., Newton, I. L. G. & Hardy, R. W. Viral RNA is a target for Wolbachia-mediated pathogen blocking. *PLOS Pathog.* **16**, e1008513 (2020).
24. Rancès, E., Ye, Y. H., Woolfit, M., McGraw, E. A. & O'Neill, S. L. The relative importance of innate immune priming in Wolbachia-mediated dengue interference. *PLoS Pathog.* **8**, e1002548 (2012).
25. Pan, X., Thiem, S. & Xi, Z. Wolbachia-Mediated Immunity Induction in Mosquito Vectors. in *Arthropod Vector: Controller of Disease Transmission* **1**, 35–58 (Elsevier Inc.,

2017).

26. Pan, X. *et al.* Wolbachia induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito Aedes aegypti. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E23–E31 (2012).
27. Ye, Y. H., Woolfit, M., Rancès, E., O'Neill, S. L. & McGraw, E. A. Wolbachia-Associated Bacterial Protection in the Mosquito Aedes aegypti. *PLoS Negl. Trop. Dis.* **7**, e2362 (2013).
28. Xi, Z., Ramirez, J. L. & Dimopoulos, G. The Aedes aegypti Toll Pathway Controls Dengue Virus Infection. *PLoS Pathog.* **4**, e1000098 (2008).
29. Caragata, E. P. *et al.* Dietary Cholesterol Modulates Pathogen Blocking by Wolbachia. *PLoS Pathog.* **9**, e1003459 (2013).
30. Frentiu, F. D. Lipids and Pathogen Blocking by Wolbachia. *Trends in Parasitology* vol. 33 916–917 (2017).
31. Geoghegan, V. *et al.* Perturbed cholesterol and vesicular trafficking associated with dengue blocking in Wolbachia-infected Aedes aegypti cells. *Nat. Commun.* **8**, 1–10 (2017).
32. Osborne, S. E., Leong, Y. S., O'Neill, S. L. & Johnson, K. N. Variation in Antiviral Protection Mediated by Different Wolbachia Strains in *Drosophila simulans*. *PLoS Pathog.* **5**, e1000656 (2009).
33. Lu, P., Bian, G., Pan, X. & Xi, Z. Wolbachia Induces Density-Dependent Inhibition to Dengue Virus in Mosquito Cells. *PLoS Negl. Trop. Dis.* **6**, e1754 (2012).
34. Ferguson, N. M. *et al.* Modeling the impact on virus transmission of Wolbachia-mediated blocking of dengue virus infection of Aedes aegypti. *Sci. Transl. Med.* **7**, 279ra37 (2015).

35. Zink, S. D., van Slyke, G. A., Palumbo, M. J., Kramer, L. D. & Ciota, A. T. Exposure to west nile virus increases bacterial diversity and immune gene expression in culex pipiens. *Viruses* **7**, 5619–5631 (2015).
36. Hughes, G. L. *et al.* Native microbiome impedes vertical transmission of Wolbachia in Anopheles mosquitoes. *Proc. Natl. Acad. Sci. U. S. A.* **111**, 12498–12503 (2014).
37. Rossi, P. *et al.* Mutual exclusion of Asaia and Wolbachia in the reproductive organs of mosquito vectors. *Parasites and Vectors* **8**, 1–10 (2015).
38. Audsley, M. D. *et al.* Wolbachia infection alters the relative abundance of resident bacteria in adult Aedes aegypti mosquitoes, but not larvae. *Mol. Ecol.* **27**, 297–309 (2017).
39. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2. *Genome Biol.* **20**, 257 (2019).
40. Matthews, B. J. *et al.* Improved reference genome of Aedes aegypti informs arbovirus vector control. *Nature* **563**, 501–507 (2018).
41. Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13**, 1028–1040 (2006).
42. Lu, J., Breitwieser, F. P., Thielen, P. & Salzberg, S. L. Bracken: estimating species abundance in metagenomics data. *PeerJ Comput. Sci.* **3**, e104 (2017).
43. Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
44. Faust, K. & Raes, J. CoNet app: inference of biological association networks using Cytoscape. *F1000Research* **5**, 1519 (2016).
45. Shannon, P. *et al.* Cytoscape: a software environment for integrated models of

- biomolecular interaction networks. *Genome Res.* **13**, 2498–504 (2003).
46. Oksanen, J. *et al.* vegan: Community Ecology Package. (2020).
 47. Bray, J. R. & Curtis, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
 48. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in. *Bioinformatics* **35**, 526–528 (2019).
 49. Hegde, S., Rasgon, J. L. & Hughes, G. L. The microbiome modulates arbovirus transmission in mosquitoes. *Current Opinion in Virology* **15**, 97–102 (2015).
 50. Dickson, L. B. *et al.* Diverse laboratory colonies of *Aedes aegypti* harbor the same adult midgut bacterial microbiome. *Parasites and Vectors* **11**, 207 (2018).
 51. Osei-Poku, J., Mbogo, C. M., Palmer, W. J. & Jiggins, F. M. Deep sequencing reveals extensive variation in the gut microbiota of wild mosquitoes from Kenya. *Mol. Ecol.* **21**, 5138–5150 (2012).
 52. David, M. R., Santos, L. M. B. dos, Vicente, A. C. P. & Maciel-de-Freitas, R. Effects of environment, dietary regime and ageing on the dengue vector microbiota: evidence of a core microbiota throughout *Aedes aegypti* lifespan. *Mem. Inst. Oswaldo Cruz* **111**, 577–587 (2016).
 53. Arévalo-Cortés, A. *et al.* The midgut microbiota of colombian *aedes aegypti* populations with different levels of resistance to the insecticide lambda-cyhalothrin. *Insects* **11**, 1–19 (2020).
 54. Yadav, K. K. *et al.* Molecular characterization of midgut microbiota of *Aedes albopictus* and *Aedes aegypti* from Arunachal Pradesh, India. *Parasites and Vectors* **8**, 1–8 (2015).
 55. Ramirez, J. L. *et al.* Reciprocal tripartite interactions between the *Aedes aegypti* midgut

- microbiota, innate immune system and dengue virus influences vector competence. *PLoS Negl. Trop. Dis.* **6**, 1–11 (2012).
56. Gusmão, D. S. *et al.* Culture-dependent and culture-independent characterization of microorganisms associated with *Aedes aegypti* (Diptera: Culicidae) (L.) and dynamics of bacterial colonization in the midgut. *Acta Trop.* **115**, 275–281 (2010).
57. Goto, S., Anbutsu, H. & Fukatsu, T. Asymmetrical Interactions between Wolbachia and Spiroplasma Endosymbionts Coexisting in the Same Insect Host. *Appl. Environ. Microbiol.* **72**, 4805 (2006).
58. Gusmão, D. S. *et al.* First isolation of microorganisms from the gut diverticulum of *Aedes aegypti* (Diptera: Culicidae): new perspectives for an insect-bacteria association. *Mem. Inst. Oswaldo Cruz* **102**, 919–924 (2007).
59. Yadav, K. K. *et al.* Diversity of Cultivable Midgut Microbiota at Different Stages of the Asian Tiger Mosquito, *Aedes albopictus* from Tezpur, India. *PLoS One* **11**, (2016).
60. Apte-Deshpande, A., Paingankar, M., Gokhale, M. D. & Deobagkar, D. N. *Serratia odorifera* a midgut inhabitant of *Aedes aegypti* mosquito enhances its susceptibility to dengue-2 virus. *PLoS One* **7**, (2012).
61. Terenius, O. *et al.* Midgut bacterial dynamics in *Aedes aegypti*. *FEMS Microbiol. Ecol.* **80**, 556–565 (2012).
62. Ramos-Nino, M. E. *et al.* Metagenomic analysis of *Aedes aegypti* and *Culex quinquefasciatus* mosquitoes from Grenada, West Indies. *PLoS One* **15**, e0231047 (2020).
63. Scolari, F., Casiraghi, M. & Bonizzoni, M. *Aedes* spp. and Their Microbiota: A Review. *Frontiers in Microbiology* **10**, (2019).
64. Hegde, S. *et al.* Microbiome Interaction Networks and Community Structure From

- Laboratory-Reared and Field-Collected *Aedes aegypti*, *Aedes albopictus*, and *Culex quinquefasciatus* Mosquito Vectors. *Front. Microbiol.* **9**, 2160 (2018).
65. Audsley, M. D., Ye, Y. H. & McGraw, E. A. The microbiome composition of *Aedes aegypti* is not critical for Wolbachia-mediated inhibition of dengue virus. *PLoS Negl. Trop. Dis.* **11**, e0005426 (2017).
66. Audsley, M. D. *et al.* Wolbachia infection alters the relative abundance of resident bacteria in adult *Aedes aegypti* mosquitoes, but not larvae. *Mol. Ecol.* **27**, 297–309 (2017).
67. Hoffmann, A. A. *et al.* Stability of the wMel Wolbachia Infection following Invasion into *Aedes aegypti* Populations. *PLoS Negl. Trop. Dis.* **8**, e3115 (2014).
68. Shade, A. & Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.* **14**, 4–12 (2012).
69. Coon, K. L., Vogel, K. J., Brown, M. R. & Strand, M. R. Mosquitoes rely on their gut microbiota for development. *Mol. Ecol.* **23**, 2727–2739 (2014).
70. MacLeod, H. J., Dimopoulos, G. & Short, S. M. Larval Diet Abundance Influences Size and Composition of the Midgut Microbiota of *Aedes aegypti* Mosquitoes. *Front. Microbiol.* **12**, 1149 (2021).
71. Risely, A. Applying the core microbiome to understand host–microbe systems. *J. Anim. Ecol.* **89**, 1549–1558 (2020).
72. Yadav, K. K. *et al.* Diversity of Cultivable Midgut Microbiota at Different Stages of the Asian Tiger Mosquito, *Aedes albopictus* from Tezpur, India. *PLoS One* **11**, e0167409 (2016).
73. Osborne, S. E., Iturbe-Ormaetxe, I., Brownlie, J. C., O'Neill, S. L. & Johnson, K. N. Antiviral protection and the importance of Wolbachia density and: Tissue tropism in

- Drosophila simulans*. *Appl. Environ. Microbiol.* **78**, 6922–6929 (2012).
74. Ulrich, J. N., Beier, J. C., Devine, G. J. & Hugo, L. E. Heat Sensitivity of wMel Wolbachia during *Aedes aegypti* Development. *PLoS Negl. Trop. Dis.* **10**, e0004873 (2016).
75. Gupta, V. *et al.* The route of infection determines Wolbachia antibacterial protection in *Drosophila*. *Proc. R. Soc. B Biol. Sci.* **284**, (2017).
76. Ye, Y. H., Seleznev, A., Flores, H. A., Woolfit, M. & McGraw, E. A. Gut microbiota in *Drosophila melanogaster* interacts with Wolbachia but does not contribute to Wolbachia-mediated antiviral protection. *J. Invertebr. Pathol.* **143**, 18–25 (2017).
77. Ourry, M. *et al.* Influential insider: Wolbachia, an intracellular symbiont, manipulates bacterial diversity in its insect host. *Microorganisms* **9**, (2021).
78. Hooper, L. V., Littman, D. R. & Macpherson, A. J. Interactions between the microbiota and the immune system. *Science* **336**, 1268–1273 (2012).

CHAPTER 3: MICROBIOME VARIATION IN WILD *ANOPHELES GAMBIAE* COLLECTED ACROSS SUB-SAHARAN AFRICA

3.1 Abstract

BACKGROUND: As the vector of malaria-causing *Plasmodium* spp., *Anopheles* mosquitos are responsible for over 600,000 deaths each year and are therefore the deadliest animal on the planet. In Africa, where most malaria cases occur, *An. gambiae* s.s. is the primary *Plasmodium* vector, however, substantial intraspecific variation in vector competency exists. This variation is influenced, in part, by a combination of the host's microbiome and an associated priming of the immune response. It is well-established that the microbiome is critical to mosquito development and reproduction, but evidence also suggests a relationship with vector competency. Despite the importance of the microbiome to mosquito biology, the *An. gambiae* s.s. microbiome has not been well characterized, especially in natural populations. Here, we surveyed genomic data from wild-caught *An. gambiae* s.s. collected across sub-Saharan Africa to characterize the natural microbiome and explore variation and diversity.

RESULTS: Our analysis of 711 wild *An. gambiae* s.s., across sub-Saharan Africa provides an expanded characterization of the microbiome, completely at the species-level. This included the identification of 269 bacterial species across 7 phyla, of which 18 species were identified as members of the core microbiome. Furthermore, microbiome diversity was similar between molecular forms of *An. gambiae* s.s., however, M/S hybrids differed with a significantly higher number of observed microbiome species. However, microbiome composition is quite variable amongst *An. gambiae* s.s.. A PCoA revealed covariance between microbiome composition and molecular form, and collection site latitude and longitude. Finally, *P. falciparum* was identified

in 545 individuals, but abundance was not correlated with the collection site coordinates of *An. gambiae* s.s. individuals. *P. falciparum* abundance also did not differ between molecular forms.

CONCLUSION: Our survey is the largest characterization of the natural *An. gambiae* s.s. microbiome, both in terms of sample size and geographic range, to date. Furthermore, it has provided a species-level characterization of *An. gambiae* s.s. microbiome composition. It also refined molecular form-specific differences in bacterial microbiome composition. While there were few differences in the tested metrics of diversity, individuals of different molecular forms differed in their overall microbiome composition. Finally, both latitude and longitude were correlated with microbiome composition. While not explicitly investigated here, the relationship between the *An. gambiae* s.s. microbiome and *Plasmodium* abundance can be explored with this dataset. We provide evidence that this methodological approach can identify both components of the bacterial microbiome and the presence of *Plasmodium* spp.

3.2 Introduction

It is estimated that more than 3,500 species of *Anopheles* mosquitoes exist globally and are present on every continent except Antarctica¹. In Africa one of the most common mosquito groups is the *An. gambiae* species complex (*An. gambiae sensu lato*). The seven sister species within the group are morphologically indistinguishable but differ genetically and behaviorally. Importantly, they also vary in vector competency—their ability to transmit disease²⁻⁴. Anopheline mosquitos carry several human pathogens, but most epidemiologically concerning is their ability to transmit *Plasmodium*, the protozoan clade that causes malaria.

Malaria is a major global health concern. In 2020, 241 million people were infected with malaria across 87 countries. Of these infections, an estimated 627,000 were fatal⁵. Compared to the rest of the world, Africa is disproportionately impacted by malaria; 95% of malaria cases and 96% of malaria-related deaths occur in Africa⁶. Five *Plasmodium* species are known to cause malaria in humans, but the most common species to infect humans is *P. falciparum*⁶. Not only is *P. falciparum* the most common *Plasmodium* species in humans, but it also causes the most severe symptoms and carries the highest mortality rate^{7,8}.

In Africa, the primary vector of *P. falciparum* is *An. gambiae sensu stricto*^{5,9}. *An. gambiae s.s.* is experiencing incipient speciation and is further classified into two molecular forms: the M- and S-form¹⁰. This classification is based on a single nucleotide polymorphism (SNP) in the ribosomal DNA on the X chromosome and an inversion within the sodium channel gene on chromosome II¹¹⁻¹⁵. It was previously thought that hybridization between the two forms was rare; surveys of sympatric populations, did not identify any hybrids¹⁴. This observed general lack of hybrids was attributed to the extent of the pre-mating behavioral isolation¹⁶. However, with improved diagnostic testing and broadened approach to species surveillance, more recent studies suggest that M/S hybridization is fairly common¹⁷⁻¹⁹.

Despite the fact that the *An. gambiae s.s.* molecular forms are morphologically identical and often overlap geographically, which is thought to be the main driver of microbiome composition assembly¹⁵, they still differ in vector competency. Variation in vector competency is likely due to a combination of microbiome composition and differences in host immunological response²⁰⁻²⁶. However, a complete understanding of the mechanism behind how the microbiome impacts vector competency is unknown but is thought to be a combination of two

factors: (1) modulation (priming) of the immune response and (2) direct contact and inhibition pathogens²².

Another important consideration is the potential existence of an *An. gambiae* s.s. core microbiome. In mosquitos, microbiome composition, especially in the midgut, plays a pivotal role in mosquito life history, greatly impacting individual development and reproduction^{27,28}. There is conflicting evidence on whether a core microbiome exists in certain species of mosquitoes but if one does exist, these core species are hypothesized to be particularly important to host function^{29,30}. Any relationship between core species and a pathogen is relevant to overall *An. gambiae* health and vector competency.

As such, a deep investigation into the tripartite relationship between the *An. gambiae* s.s., *Plasmodium* pathogens, and the microbiome is relevant and important for understanding mosquito biology, human epidemiology, and general patterns host-microbe interactions. However, before such an investigation can take place, a full characterization of the *An. gambiae* s.s. microbiome is necessary, including identifying which species are present, broad patterns in composition across *An. gambiae* s.s., any differences in the microbiomes between molecular forms, and how microbiomes differ geographically.

To characterize these aspects of the natural *An. gambiae* microbiome, we bioinformatically analyzed whole-genome sequencing data from 711 wild-caught *An. gambiae* s.s. females collected across sub-Saharan Africa. Through this approach we: (1) achieved microbiome classifications to the species-level for 96.6% of the 38.5 million reads that mapped to bacterial genomes, expanding upon previous *An. gambiae* microbiome characterizations, (2) compared the microbiome composition of *An. gambiae* s.s. molecular forms, (3) identified an *An.*

gambiae s.s. core microbiome and (4) putatively identified the presence *P. falciparum* in a majority of samples.

3.3 Methods

An. gambiae genome sequencing data

We utilized whole-genome sequencing data from wild populations of *An. gambiae* adult females that were collected from 25 sites located in 11 countries across Sub-Saharan Africa³¹ (Fig. 1). A full description of collection and sequencing methods is described in The *Anopheles gambiae* 1000 Genomes Consortium, 2017³² but briefly, samples were collected from a variety of sites including large urban areas, municipalities, and small rural villages spanning diverse ecosystems, for example, rainforests, savannas, and mangroves. Sequencing was performed on the Illumina HiSeq 2000 platform at the Wellcome Trust Sanger Institute and raw sequencing data was deposited to the NCBI SRA (BioProject PRJEB1670). To confirm that the individuals collected were *An. gambiae* s.s. and not one of the seven morphologically identical sister species, a PCR-based assay was used to distinguish between the other members of the *An. gambiae* species complex and *An. gambiae* s.s. individuals that were collected³³. Individuals confirmed to be *An. gambiae* s.s. were subjected to additional diagnostics assays that identify if an individual is an M-form, S-form, or M/S hybrid based on a marker located on the X chromosome^{34,35}.

In total, this dataset contained sequencing data from 1,037 individual female samples. Each sample was sequenced multiple times, containing between 3 and 7 biological replicates, for a total of 4,078 total sequencing runs. Metadata was available both through the NCBI SRA BioProject PRJEB1670 and the Ag1000G Phase 2 AR1 Data Release³¹. This metadata included

information regarding collection coordinates, date, and molecular form assignment (*An. gambiae* s.s. M-form, S-form, or M/S hybrid).

Taxonomic classification of the An. gambiae microbiome

The Krakenuniq pipeline was used to assign taxonomic classifications to sequencing reads that met standard quality metrics³⁶. The Krakenuniq database was built using the default parameters for minimizer length and a k-mer length of 31. The database included all available NCBI RefSeq genomes from bacteria (21,333), fungi (60), protozoa (40), viruses (10,388), archaea (390) (NCBI RefSeq Release Number 200 on August 26, 2020) and the *An. gambiae* reference genome (AgamP3³⁷). The final Krakenuniq database was 1.4TB and used a maximum of 308GB of RAM on 28 threads. Prior to classification, low complexity regions of genomes were masked using the DustMasker program using default parameters³⁸.

Quantification of An. gambiae s.s. microbiome composition

To eliminate potential biases introduced by abnormally low-quality sequencing runs, the data was filtered by several criteria: (1) total number of reads, (2) percentage of unclassified reads, and (3) percentage of reads identified as *An. gambiae*. First, sequencing runs were removed that had total read counts that fell below of Tukey's lower fence (<10,225,291 total reads) and a percentage of reads remaining unclassified that was greater than Tukey's upper fence (>10.0% of reads unclassified). Lastly, in a limited number of samples there was evidence of bacterial contamination, as indicated by the proportion of bacterial reads outnumbering the proportion of reads identified as *An. gambiae*. To eliminate these potentially contaminated

samples, sequencing runs were removed with less than 94.1% of reads mapping to the *An. gambiae* reference genome, again based on Tukey's outlier test. After culling the original dataset, there were 2,694 sequencing runs from 711 unique female samples that were used for all downstream analyses. Each sample still retained between 2-7 biological replicates.

To account for the overall variation in sequencing depth per individual sequencing run, species abundance estimates were standardized. These standardizations were conducted separately in each major taxonomical group (*e.g.*, bacterial, fungi, etc.). For subsequent analyses that focused solely on the bacterial microbiome, rare species were removed from the dataset if they (1) did not constitute >1% of the total bacterial microbiome in any individual and (2) were present in fewer than 25% of individuals.

In general, the biological replicates of the same sample were highly correlated on average ($r > 0.90$). If a biological replicate was not correlated with the other replicates with a Pearson correlation coefficient of at least 0.70 it was removed from further analysis. After culling these relatively less correlated replicates, each sample still retained between 2-7 biological replicates. For each sample, the taxonomic classifications for biological replicates were averaged together to achieve a single microbial profile for each unique sample. This resulted in a single microbiome profile for each of the remaining 711 *An. gambiae* *s.s.* samples. Of these 711 samples, 236 were identified as M-form, 325 were identified as S-form, and 150 were hybrids (Table 1).

Statistical analysis of bacterial microbiome composition

Microbiome complexity and diversity for each of the molecular forms and hybrid samples was compared using three metrics – total microbiome size, Shannon’s diversity index³⁹, and observed species richness (OSR) as diversity metrics. An analysis of variance (ANOVA) and post-hoc Tukey HSD Test was used to assess significant differences in these diversity metrics across samples. Similarly, to determine if collection location impacts these metrics, linear models were used to evaluate any relationship between total microbiome size, Shannon’s diversity index, and OSR and the latitude and longitude of the collection site for each sample. All statistical analyses were performed in R (version 4.0.4)⁴⁰.

Principal Coordinate Analysis (PCoA) was then used to further characterize bacterial microbiome variation. A distance matrix was calculated by applying the Bray-Curtis dissimilarity index⁴¹ implemented by the vegdist function from the R package vegan (version 2.5-7⁴²) to a bacterial species abundance matrix. The PCoA was performed using the pcoa function from the R package ape (version 5.5⁴³). Axes of variation were further investigated using linear models to assess the relationship between Principal Coordinate (PC) axis loading values and species abundance estimates.

3.4 Results

An. gambiae s.s. microbiome composition

We aimed to characterize microbiome variation in *An. gambiae* s.s. collected across Sub-Saharan Africa, with specific attention to the bacterial microbiome. We used the Krakenuniq³⁶ pipeline to map whole genome sequencing data collected from 711 *An. gambiae* females to a database containing the *An. gambiae* reference genome and the full set of available NCBI RefSeq genomes from bacterial, viral, archaeal, fungal, and protozoan taxa (32,212 genomes in total).

An overwhelming majority of reads mapped to the *An. gambiae* reference genome as expected (average = 96.18%; SD = 0.57%). On average, 3.44% (SD = 0.53%) of reads remained unclassified, most likely the result of low complexity or low-quality reads or derived from species not included in our database.

The remaining 0.38% (SD = 0.25%) of reads (average of 322,697 reads per sample post-depth standardization; SD = 209,811) mapped to archaeal, bacterial, viral, fungal, or protozoan genomes. Of these microbially derived reads, the majority were classified as either bacterial (49.6%, SD = 21.0%) or fungal (47.1%, SD = 20.4%) (Fig. 2A). Archaeal and viral species were consistently identified at low abundances (mean archaeal standardized reads = 13,399, SD = 10,054; mean viral standardized reads = 1,036, SD = 2,761). Protozoan species were also identified in low abundances on average (mean protozoan standardized reads = 2,298, SD = 23,394), but abundance was highly variable (CV = 1017.81%) compared to the other identified low abundance taxa (archaeal CV = 75.03%; viral CV = 266.63%). This variability in Protozoan load was primarily due to the presence of *Plasmodium* spp., particularly *P. falciparum*. *P. falciparum* was identified in 76.65% of samples, on average a sample had 2,242 *P. falciparum* reads (SD = 21,411). Full information regarding fungal, viral, archaeal, and protozoan species-level identifications can be found in Additional Table 5.

Bacterial microbiome composition

After removing low abundance species that did not constitute >1% of the total bacterial microbiome in any individual or rare species that were present in fewer than 25% of individuals, 269 bacterial species remained from 114 genera (Additional File 6). In these samples, 7 phyla are represented within the bacterial the microbiome. Proteobacteria is the dominant phyla both

in terms of the number of representative species identified ($N = 163$; Fig. 2B), and in regard to the average proportion of the bacterial microbiome that is made up by a species of *Proteobacteria* (79.79% of bacterial reads; Fig. 2C). The other 20.21% of the microbiome, on average, was composed of species belonging to *Actinobacteria* (7.62%), *Firmicutes* (7.42%), *Bacteroidetes* (4.54%), *Cyanobacteria* (0.54%), *Chloroflexi* (0.06%), and *Fusobacteria* (0.02%) (Fig. 2C).

At the genus level, our analysis was consistent with several previously reported genera in *An. gambiae s.s.* including, *Actinobacter*, *Asaia*, *Bacillus*, *Elizabethkingia*, *Enterobacter*, *Enterococcus*, *Klebsiella*, *Korcuria*, *Methylobacterium*, *Pseudomonas*, *Serratia*, *Sphingomonas*, *Staphylococcus*, *Serratia*, and *Streptococcus*^{24,44-47}.

Additional studies that characterized the microbiome of closely related *Anopheles* species, but not *An. gambiae s.s.* specifically, identified *Acetobacter*, *Aeromonas*, *Burkholderia*, *Cedecea*, *Escherichia*, *Microbacterium*, *Pantoea* as common components of the microbiome⁴⁸⁻⁵². Here, we also identify these genera in a substantial number of individuals but not necessarily at a high abundance: *Acetobacter* ($N = 281$; average reads = 435.32; SD = 4.16), *Aeromonas* ($N = 693$; average reads = 704.21; SD = 13,653.64), *Burkholderia* ($N = 674$; average reads = 4,663.16; SD = 36,402.44), *Cedecea* ($N = 699$; average reads = 250.87; SD = 5,161.16), *Escherichia* ($N = 711$; average reads = 943.69; SD = 2,813.50), *Microbacterium* ($N = 473$; average reads = 8.21; SD = 12.02), *Pantoea* ($N = 711$; average reads = 10,294.16; SD = 69,060.25). In fact, most are relatively rare on average, apart from *Pantoea*. But notably, *Cedecea* (CV = 2,057.31%), *Aeromonas* (CV = 1,938.86%), and *Burkholderia* (CV = 780.64%) are particularly variable and are found in substantially high abundances in a limited number of individuals. In addition to confirming the presence of several previously reported genera, we identified the presence of genera that have not been previously reported in *An. gambiae s.s.* or

related species of mosquito. These genera include *Aeromicrobium*, *Anoxybacillus*, *Limnobacter*, *Massilia*, *Oscillatoria*, and *Roseiflexus*.

Acinetobacter soli was the most abundant species on average that was identified with an average of 10,286.22 reads per individual ($SD = 64,718.96$). It was also present in the majority of individuals sampled (71.45%; Table S2). The genus *Acinetobacter* has been previously identified as part of other mosquito species microbiomes including *An. albimanus*, *An. stephensi*, *An. maculipennis*, and other unspecified anopheline samples^{44,53–55}, but since these studies rely on 16S rRNA data this is the first report of *A. soli* specifically. However, when considering the genus more generally, between all of the *Acinetobacter spp.* that were identified, the genus was represented in all 711 individuals.

An. gambiae intraspecific microbiome diversity and composition

To evaluate if *An. gambiae* s.s. M-form, S-form, and M/S hybrid females differ in microbiome complexity and diversity we used three metrics to characterize the bacterial microbiome – total microbiome size, Shannon’s diversity index, and observed species richness (OSR). First, we assessed whether the three groups of *An. gambiae* s.s. molecular forms, differ based on these three diversity measures. The average total microbiome size was not significantly different between the three groups (Fig. 3A; $M = 119,248$; $S = 91,787$; M/S Hybrid = 85,351; $p = 0.137$). Similarly, Shannon’s diversity index also did not differ between any of the three groups of *An. gambiae* s.s. (Fig. 3B; $M = 3.33$; $S = 3.18$; M/S Hybrid = 3.25; $p = 0.188$). However, when comparing the average OSR of each group, the M/S hybrid individuals did have a significantly greater average OSR than either the M- or S-form (Fig. 3C; $M = 165$; $S = 168$; M/S Hybrid = 185; $p = 1.090\text{e-}10$).

Next, each of these metrics was evaluated as a function of the longitude and latitude from which each sample was collected to identify putative relationships between microbiome diversity and geographic locale. Neither longitude nor latitude significantly correlated with total microbiome size (longitude $p = 0.332$; latitude $p = 0.213$; Fig. S1A-B), Shannon's index (longitude $p = 0.365$; latitude $p = 0.075$; Fig. S1C-D), or OSR (longitude $p = 0.795$; latitude $p = 0.900$; Fig. S1E-F).

Patterns of An. gambiae s.s. microbiome composition

To assess overall patterns of composition and variation within the microbiome we utilized a PCoA and observed the first four main axes which captured 43.57% of total variation (Fig. 4A-B). To evaluate if microbiome composition differs between the M-form, S-form, and M/S hybrids, the average loading values these groups on each axis were compared (Fig. 4C-F). Axis 1 (19.46% of the total variation) and Axis 3 (8.93% of the total variation) both show significantly different average loading values between each of the groups (Axis 1: M = 0.12, S = -0.11, M/S Hybrid = 0.05, $p = 2.24\text{e-}36$, Fig. 4C; Axis 3: M = 0.08, S = -0.06, M/S Hybrid = 0.00, $p = 2.33\text{e-}25$, Fig. 4E). On Axes 2 and 4, individual molecular form is partially distinguished based on average loading value. On Axis 2 (9.00% of the total variation), the average loading value M- and S-form samples are not significantly different from each other, but the M/S hybrid samples loading are significantly different than both M- and S-form ($M = -0.04$, $S = -0.01$, M/S Hybrid = 0.10, $p = 3.37\text{e-}18$; Fig. 4D). Similarly, on Axis 4 (6.18% of the total variation) M-form and M/S hybrid average loading values are not significantly different from each other, but both are significantly different than the S-form individuals ($M = -0.02$, $S = 0.02$, M/S Hybrid = -0.01, $p = 0.002$; Fig. 4F).

Despite the absence of a correlation between either longitude or latitude and metrics of microbiome diversity when comparing *An. gambiae* s.s. molecular forms, when assessing overall microbiome composition, geographic collection location does correlate with the loading values of the first 4 axes of variation (Fig. S2). Both latitude and longitude are significantly correlated with Axes 1 and 4 loading values (Axis 1: latitude $r = -0.14$, $p = 0.0002$; longitude $r = -0.21$, $p = 1.272\text{e-}08$; Axis 4: latitude $r = -0.20$, $p = 3.322\text{e-}07$; longitude $r = 0.23$, $p = 2.552\text{e-}10$). The loading values on Axes 2 and 3 are significantly correlated with latitude (Axis 2: $r = -0.21$; $p = 7.129\text{e-}09$; Axis 3: $r = -0.20$, $p = 1.279\text{e-}07$) but not longitude (Axis 2: $r = -0.03$; $p = 0.399$; Axis 3: $r = -0.10$, $p = 0.006$).

Evidence in favor of an An. gambiae s.s. core microbiome

By capitalizing on this geographically broad dataset, we were able to explore the existence of an intra-specific core microbiome in *An. gambiae* s.s. that is shared amongst the molecular forms. First, when using the most conservative definition of a “core microbiome,” meaning that a species is present in every individual sampled, we identified 18 core species. By relaxing this threshold slightly to include all bacterial species present in at least 95% of samples ($N \geq 675$), we identified 46 core species (Fig. 5A; Additional File 7). Furthermore, the bacterial species that were present in a majority of the individuals sampled, also constituted a relatively high proportion of total microbiome composition (Fig. 5B). Specifically, the 96 distinct bacterial species that were identified in at least 75% of sampled individuals also comprised 75.0% of the total bacterial microbiome. The remaining 25.0% of the microbiome was a combination of 173 additional species that were relatively lower abundance on average. When comparing the 18

core species to the other 251 non-core species, the core species did in fact have a significantly greater abundance ($p = 1.10e-8$).

Of the 18 bacterial species that made up the *An. gambiae* s.s. core microbiome, five species constituted more than 5% of the core microbiome on average (Fig. 5D). *Cutibacterium acnes* (26.0%) had the highest average abundance, followed closely by *Klebsiella pneumoniae* (23.3%), and *Bradyrhizobium* sp. PSBB068 (14.3%), *Xanthomonas translucens* (7.2%), and *Escherichia coli* (5.4%).

Identification of P. falciparum across sub-Saharan Africa

Typically rates and prevalence of malaria are identified via clinical data and often species-level identification is not available. Providing insights into the proportion of mosquitos carrying *Plasmodium*, and which species, is epidemiologically relevant. Our approach allows the identification of microbial species beyond bacterial species. As such, we surveyed the 711 *An. gambiae* s.s. samples for the presence of protozoan reads (Additional File 5). Specifically, *P. falciparum* was identified in 545 of the 711 total samples (76.65%). *P. falciparum* was identified in each country and at each site that was sampled.

Most individuals contained relatively low amounts of *P. falciparum* reads (median read count = 10), however, some individuals contained read abundance counts greater than 100,000. On average a sample had 2,242 *P. falciparum* reads ($SD = 21,411$) (Fig. 6A). Burkina Faso, Gabon, Cameroon, and Uganda were the only countries where the abundance of *P. falciparum* reads exceeded 1,000 in any sample.

Since we know that vector competency in *An. gambiae* s.s. is variable, we questioned if molecular form is related to *P. falciparum* abundance (Fig. 6B). There is no significant

difference between the average *P. falciparum* abundance between individuals identified as M-form, S-form, or M/S hybrids ($F = 0.03$, $p = 0.971$). Furthermore, we were interested in the possibility of a relationship between collection site coordinates and the abundance of *P. falciparum*. Like molecular form, there was no significant relationship between *P. falciparum* abundance and either latitude ($p = 0.45$; Fig. 6C) or longitude ($p = 0.54$; Fig. 6D).

3.5 Discussion

As the primary vector of malaria-inducing *Plasmodium spp.* in Africa, a wholistic understanding of *An. gambiae s.s.* biology is critical to mitigate the spread of disease. An individual mosquito's vector competency determines its ability to transmit a human pathogen like *Plasmodium*. The microbiome is intimately involved with determining vector competency in two ways: (1) modulation (priming) of the immune response and (2) direct contact and inhibition pathogens²². In a study that compared transcriptional data between aseptic mosquitos and those with a natural microbiome, genes involved in the immune response were generally up-regulated in those individuals with a natural microbiome compared to aseptic individuals⁵⁶. In that same study, when the two sample groups were exposed to *P. falciparum* infection, the aseptic mosquitos were significantly more susceptible to *P. falciparum* colonization and proliferation in the midgut⁵⁶.

In other species of mosquitoes, the presence of bacterial species in the mosquito microbiome has been shown to prime the immune response though the upregulation of the *Toll* and *Ind* immunity pathways⁵⁷. When these pathways are activated, there is an increased production of antimicrobial peptides that target pathogens for removal⁵⁷⁻⁵⁹. Similarly, physical

contact between components of the microbiome and pathogens have been shown to be inhibit vector competency through competition for host resources and physical space within the host⁶⁰⁻⁶².

While half of the *Plasmodium* lifecycle occurs in a human host, the steps that occur within the mosquito are intimately intertwined with the microbiome. When a female mosquito takes a blood meal from a human infected with *Plasmodium*, if the mosquito ingests *Plasmodium* gametocytes, that female in turn becomes infected. Once inside the mosquito, the *Plasmodium* gametocytes are transported to the midgut, where sexual reproduction will begin in as little as 24 hours. This sexual reproduction stage results in a motile zygote that transcends the gut epithelium into the hemolymph and forms an oocyst. Over the next 7 days, the oocyst undergoes mitosis, releasing *Plasmodium* sporozoites into the hemolymph. With the sporozoites freely circulating in the hemolymph, the infection eventually reaches the mosquito salivary glands about 10-14 days post-infected blood meal. At this point the female mosquito is infectious and when the individual takes another blood meal, the saliva transferred to the human will contain *Plasmodium* sporozoites. Importantly, this female will remain infected for the entirety of their life, any subsequent blood meal carries a high likelihood of *Plasmodium* transmission⁶³.

It is critical to note that at any one of these stages, the mosquito immune response and/or physical contact with the natural microbiome can interfere with *Plasmodium* colonization and reproduction thus ultimately limiting the individual's vectorial capacity. *Plasmodium* not only encounters the gut microbiome, but it is well documented that distinct and diverse microbial community exist within the mosquito salivary glands^{44,64}. Additionally, microbial communities in the reproductive tissue of other mosquito species have been shown to also play a role in the modulation of the immune response, thus ultimately impacting vector competency⁶⁵.

To better characterize the natural *An. gambiae* s.s. microbiome and develop a deeper understanding of the relationships between different components of the microbiome, we bioinformatically analyzed whole-genome sequencing data from 711 wild-caught *An. gambiae* s.s. females collected across sub-Saharan Africa. Through this approach we: (1) classified nearly 38.5 million reads present in the sequencing data as bacterial, (2) characterized aspects of microbiome composition separately for each molecular form and hybrids, (3) identified a putative species-wide core bacterial species, and (4) identified the presence *P. falciparum* in more than 75% of the individuals sampled.

Evidence in favor of an An. gambiae s.s. core microbiome

It is well-established that there is an important relationship between the microbiome and aspects of host biology, including fertility, longevity, and immunity⁶⁶⁻⁶⁸. By identifying ubiquitous components of the microbiome, it is thought that we can gain insights into the species that are particularly important to host fitness³⁰. Typically, studies into the *An. gambiae* s.s. core microbiome have relied on populations maintained in the lab, often for many generations, which can diminish the diversity of the natural microbiome variation^{46,69,70}. Our study was able to utilize data from wild-caught individuals to broaden our understand of *An. gambiae* natural microbiome diversity.

The mosquito microbiome, in particular, is heavily shaped by their immediate environment— aquatic larval habitat, blood meal status, and breeding sites are thought to shape microbiome composition⁷¹. Despite this overwhelming impact of a mosquito's direct environment on microbiome composition that could likely cause strong geographic differences,

our analysis successfully identified 18 bacterial species within the *An. gambiae* s.s. core microbiome.

While a more in-depth functional analyses would be necessary to fully elucidate core microbiome-host dynamics⁷², several important points can be made. These 18 bacterial species that were ubiquitously present across all individuals comprised 55.2% of the total microbiome on average. The remaining minority of the total microbiome was a combination of the additional 238 bacterial species identified. Interestingly, the most abundant core microbes have not necessarily been previously identified in *An. gambiae* s.s. It is possible that the high abundance of these species is due to their presence in the midgut, where mosquitos typically carry the highest bacterial load⁷³. However, it is difficult to know definitely, since no species from *Cutibacterium* or *Bradyrhizobium* have been identified in *An. gambiae* or any closely related species of mosquito prior to our analysis. *Escherichia* spp. have been identified in the midgut of *An. gambiae* and other closely related species of mosquito^{52,74-76}. Species of *Klebsiella* have previously been commonly identified in other mosquito species in addition to *An. gambiae* s.s. but not specifically in the midgut⁷⁷⁻⁸⁰. *Xanthomonas* has been identified once in another *Anopheles* species⁵³ – *An. barbumbrosus* – but the genera does not seem to be a common component of the mosquito microbiome. Without additional studies that confirm the presence of these species and in which tissues specifically they are enriched, it is difficult to determine how they are functionally impacting the *An. gambiae* s.s. host and how they may be important to host biology.

Identification and variation of P. falciparum in An. gambiae s.s. individuals

It is not surprising that *P. falciparum* was identified in these samples collected across sub-Saharan Africa given that *An. gambiae s.s.* is the primary vector^{5,9}. However, the number of individuals in which *P. falciparum* was identified combined with the degree of variation across all individuals is substantial. Interestingly, the variation in *P. falciparum* load was not statistically related to any of the factors investigated here, including host molecular form or the geographic coordinates of the sample collection site (Fig. 6). This implies that there are other factors that contribute to this level of variation. Evidence suggests that host genetics or the presence of specific bacterial species may influence vector competency^{65,81,82}. While not specifically investigated in our study, these additional factors could be informed by these results

3.6 Conclusion

In this study we provide the largest survey of *An. gambiae s.s.* microbiome characterization to date. With our bioinformatic approach we were able to take high-quality whole genome sequencing data and determine species-level identifications for the component of the *An. gambiae s.s.* microbiome. It was previously unknown how the molecular forms of *An. gambiae s.s.* differ from each other in terms of microbiome composition. Furthermore, with the number of M/S hybrids greater than traditionally believed, we were able to show that hybrid microbiome composition also differs in a number of ways. Variation in microbiome composition is a critical factor in mosquito life history; it impacts development, reproduction and, importantly, vector competency. Here, we identified *P. falciparum*, a malaria-causing protozoan species, in a significant number of individual females. The identifications made here should be investigated further to fully understand how the microbiome may impact *P. falciparum* abundance, and alternatively, how *P. falciparum* may alter the natural *An. gambiae s.s.* microbiome.

3.7 Tables and Figures

Country	An. gambiae s.s. Molecular Identification	Molecular Form N	Total Country N
The Gambia	M/S Hybrid	37	37
Guinea-Bissau	M/S Hybrid	79	79
Guinea	M-form	4	40
	S-form	35	
	M/S Hybrid	1	
Cote d'Ivoire	M-form	66	66
Burkina Faso	M-form	42	81
	S-form	39	
Ghana	M-form	50	62
	S-form	12	
Gabon	S-form	64	64
Angola	M-form	74	74
Cameroon	S-form	81	81
Uganda	S-form	94	94
Kenya	M/S Hybrid	33	33

Table 1. *An. gambiae s.s.* molecular form identification demographics for each country in which sampling occurred.

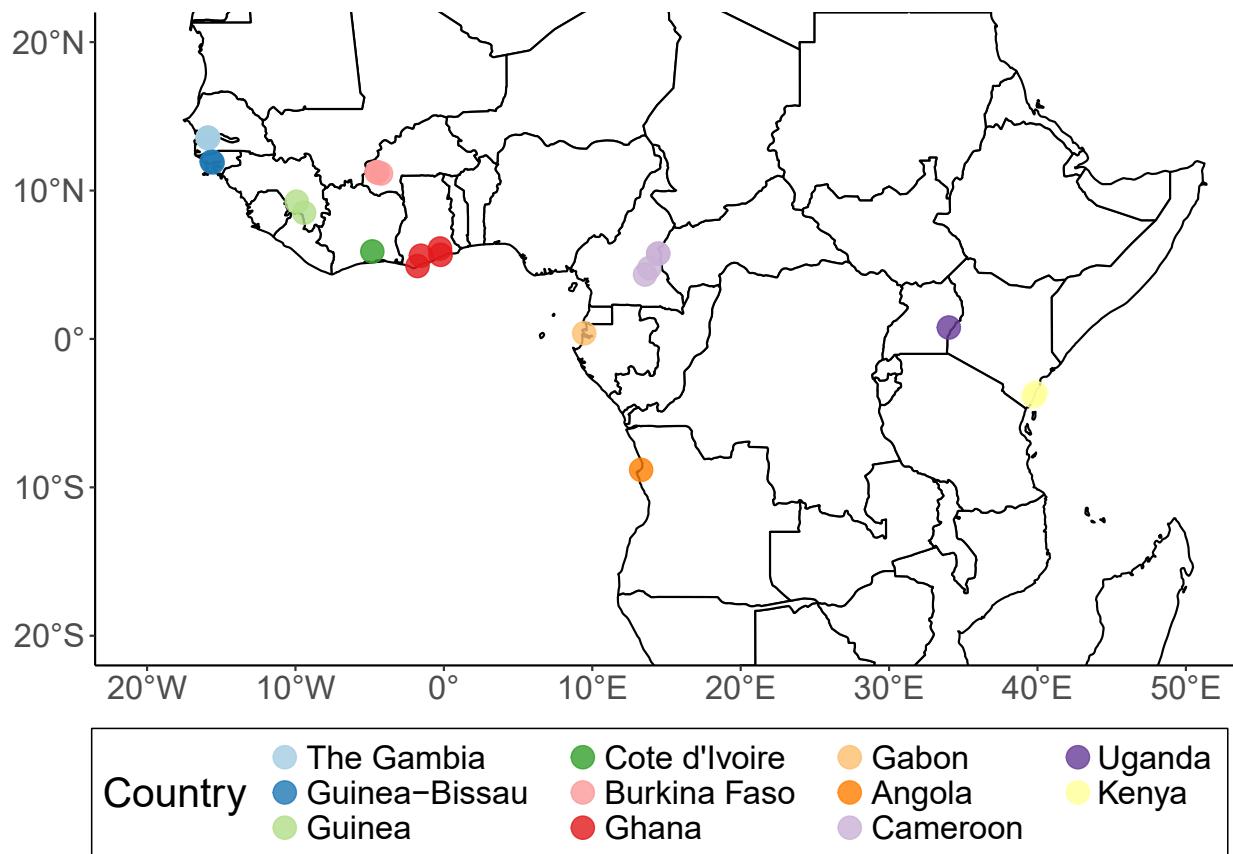


Figure 1. Map of *An. gambiae* s.s. collection sites in Africa colored by the country from which they were collected. All samples collected were confirmed as *An. gambiae* s.s. females. Additional diagnostic testing was used to identify each individual as M-form, S-form, or a M/S hybrid. A full description of sampling procedure and diagnostic testing for species/molecular form identification can be found in The *Anopheles gambiae* 1000 Genomes Consortium, 2017¹⁴⁷.

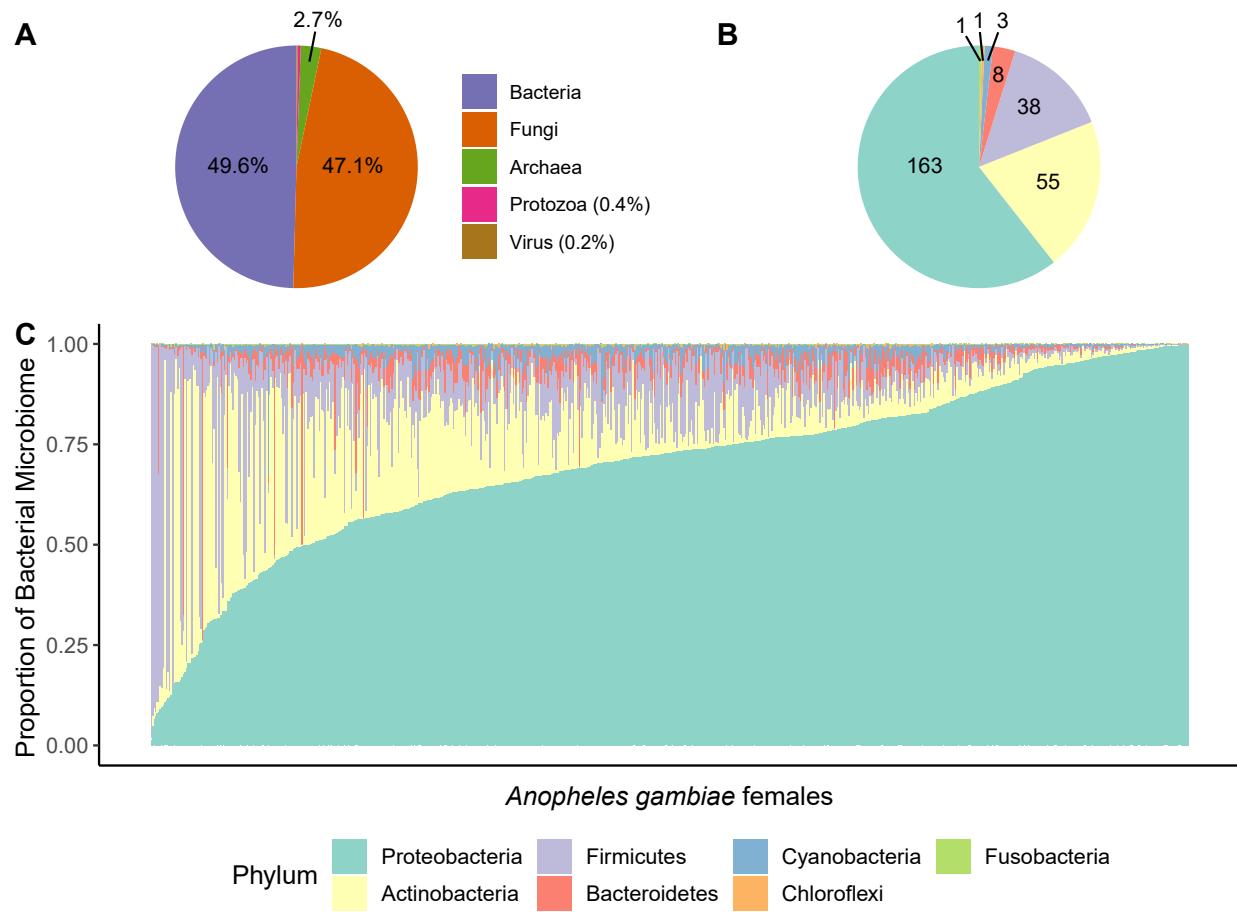


Figure 2. *An. gambiae* s.s. phylum-level microbiome characterization. (A) Taxonomic distribution of classified reads mapping to non-*An. gambiae* genomes across the 711 individuals sampled. (B) Species-level distribution of the major phyla in the bacterial microbiome across the 711 individuals sampled. (C) Bacterial microbiome composition by phyla for each individual female sample. These analyses have not accounted for collection location or molecular form differences.

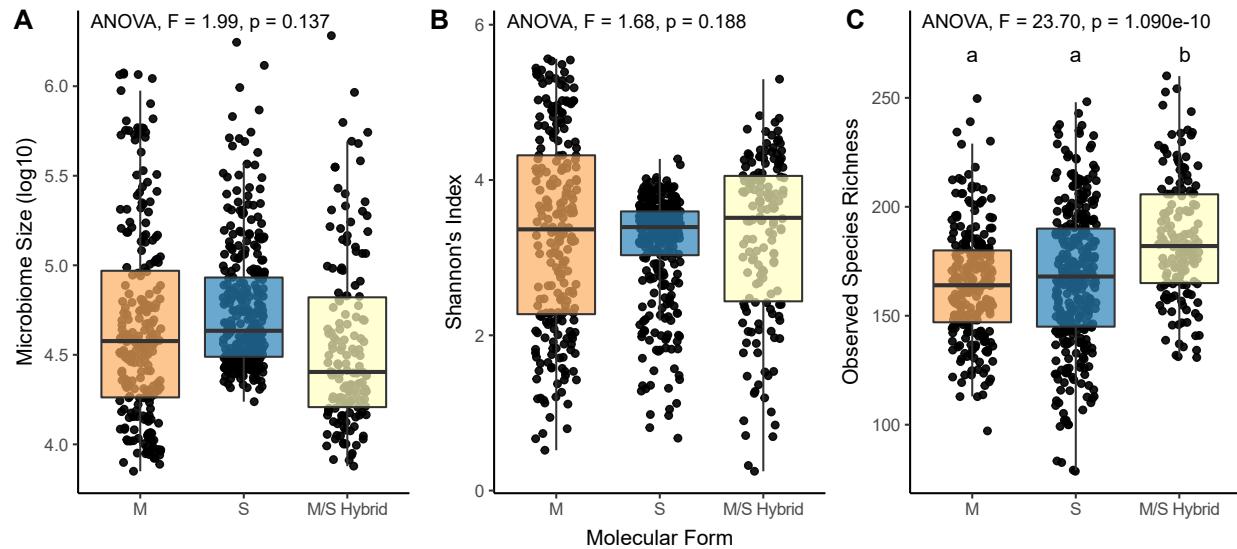


Figure 3. Variation in bacterial microbiome diversity and complexity between *An. gambiae* s.s. molecular forms. (A) There is no significant difference in the average total bacterial microbiome size abundance (normalized read count) between each of the molecular forms. (B) There is no significant difference in the average Shannon's diversity index between each of the molecular forms. (C) When comparing the average OSR between molecular forms, the M/S hybrid samples have a significantly higher number of species present in the microbiome than either the M- or S-form samples.

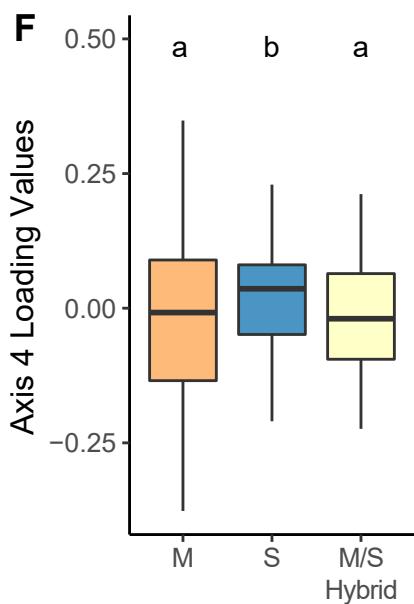
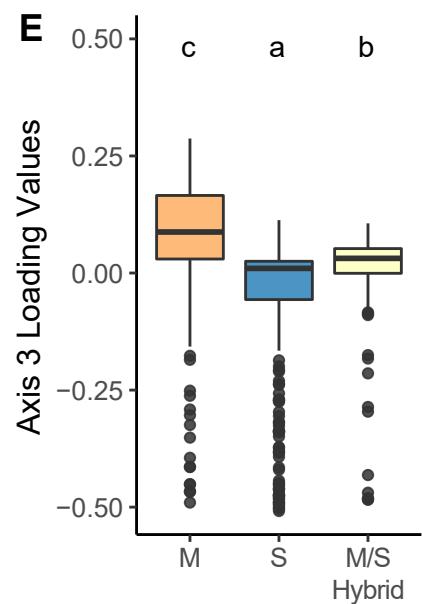
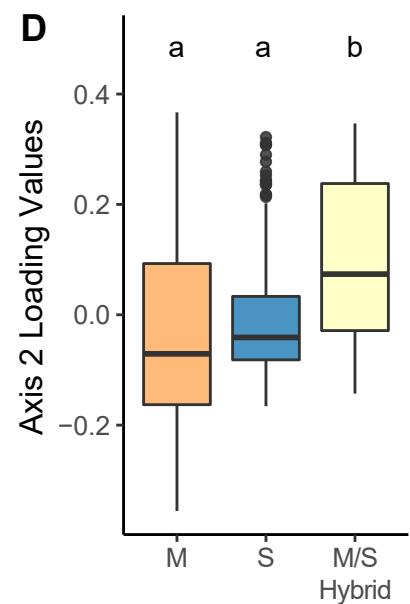
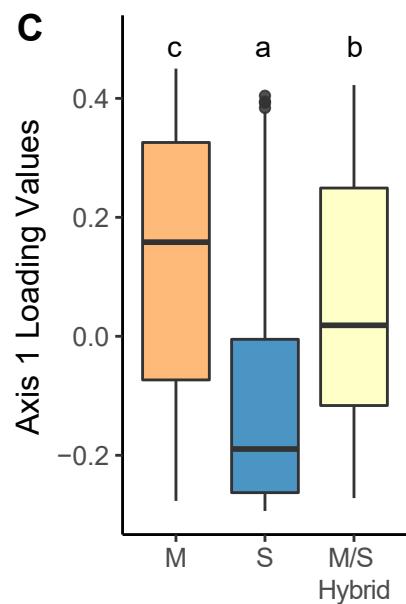
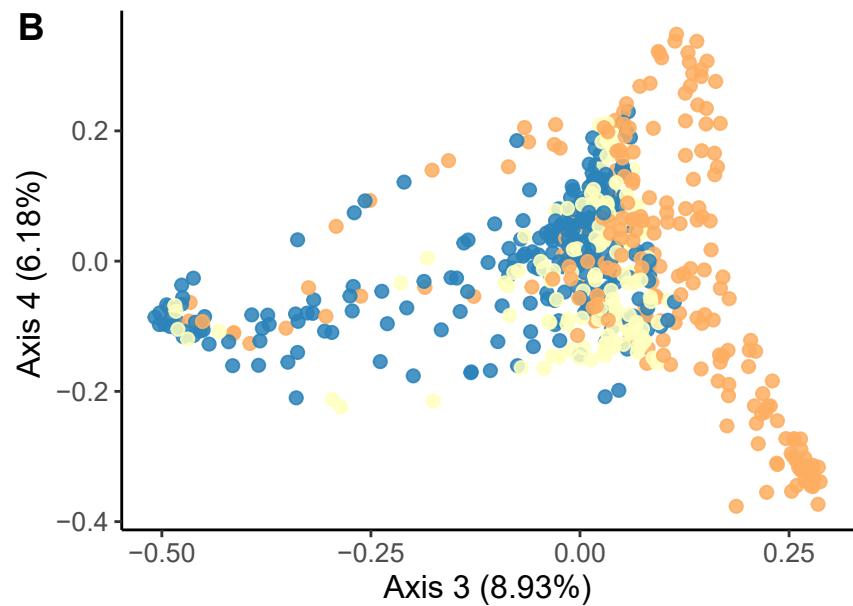
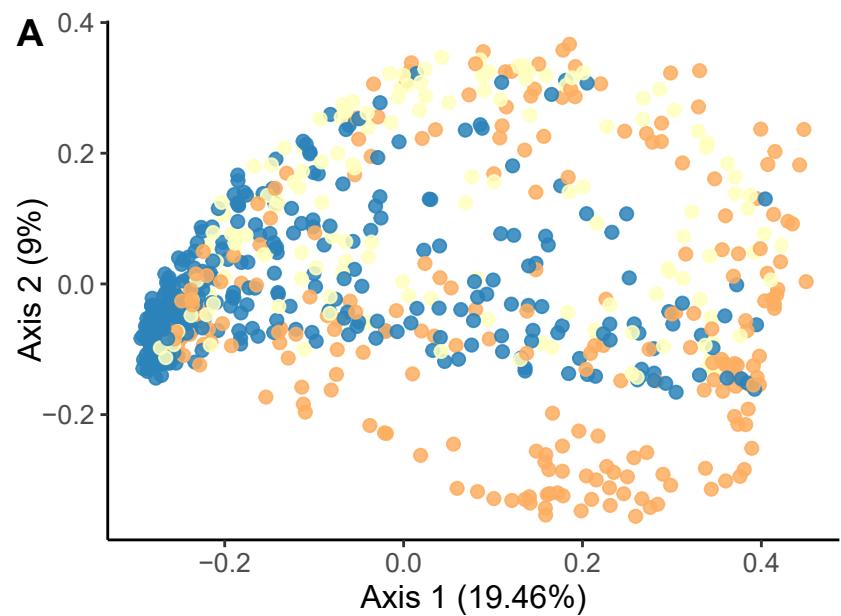


Figure 4. Covariance between microbiome composition and *An. gambiae* s.s. molecular form.

(A-B) A PCoA of bacterial microbiome variation identified the first 4 axes of variation which account for 43.57% of the total variation. (C) Axis 1 loading values significantly differed between each of the *An. gambiae* s.s. molecular forms. (D) M/S hybrid samples had significantly greater average Axis 2 loading values compared to either the M- or S-form samples. (E) On Axis 3, the average loading values for each of the molecular forms significantly differed from each other. (F) Axis 4 loading values between the M-form samples and M/S hybrids are not significantly different, on average, but the loading values of S-form samples is significantly different.

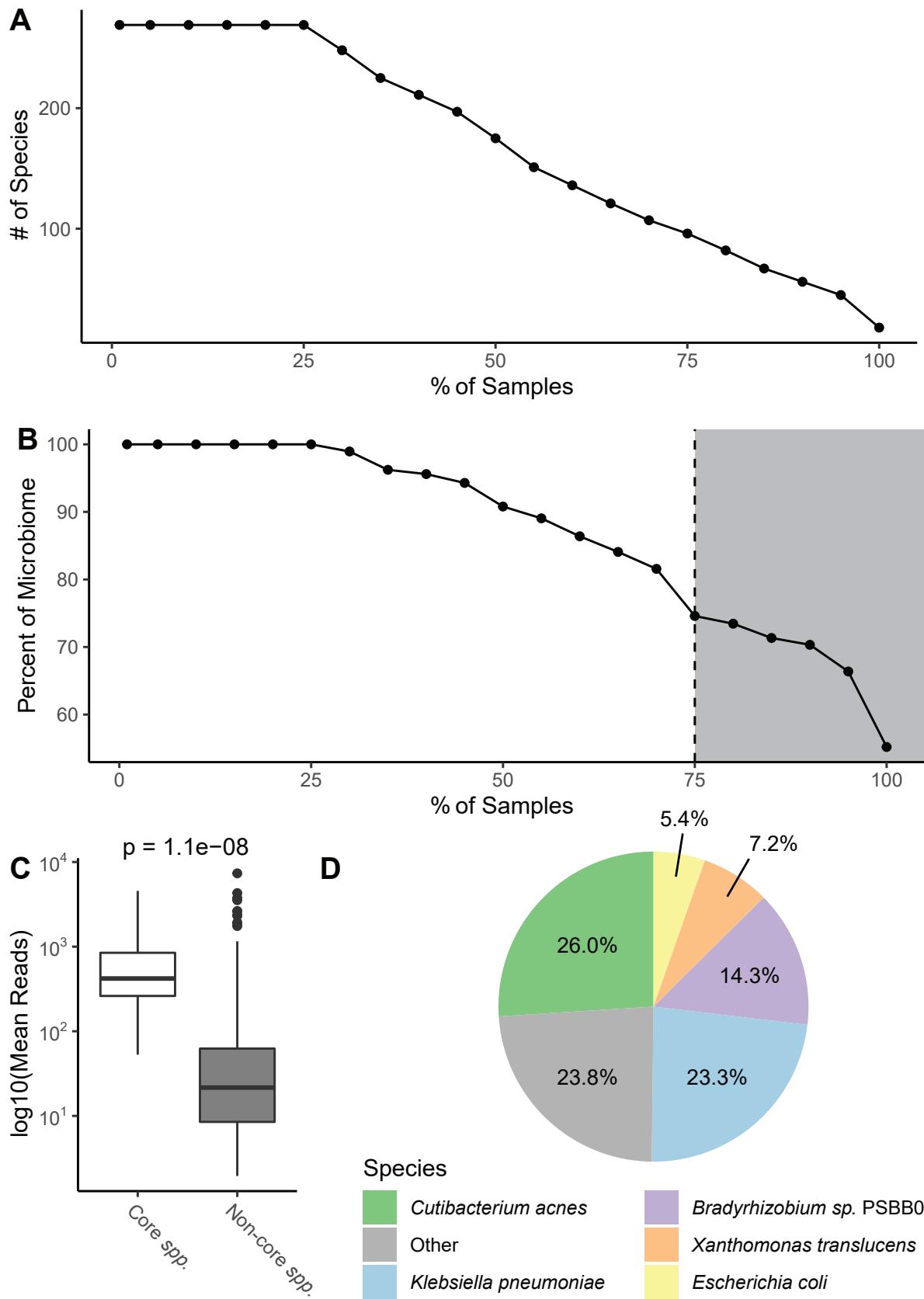


Figure 5. *An. gambiae* s.s. core microbiome. (A) Number of species identified is plotted in relation to the proportion of individuals in which each species was identified. (B) Percent of microbiome composition is plotted in relation the proportion of individuals in which each species was identified. Those species present in 75% or more individuals comprised 75% of total microbiome composition. 18 species were identified in all individuals ($N = 711$). (C) On average, the 18 core species had a significantly greater abundance than the non-core species. (D) Contribution of the 18 core bacterial species to core microbiome composition. Species that do not constitute >5% of the core microbiome have been grouped as ‘Other.’

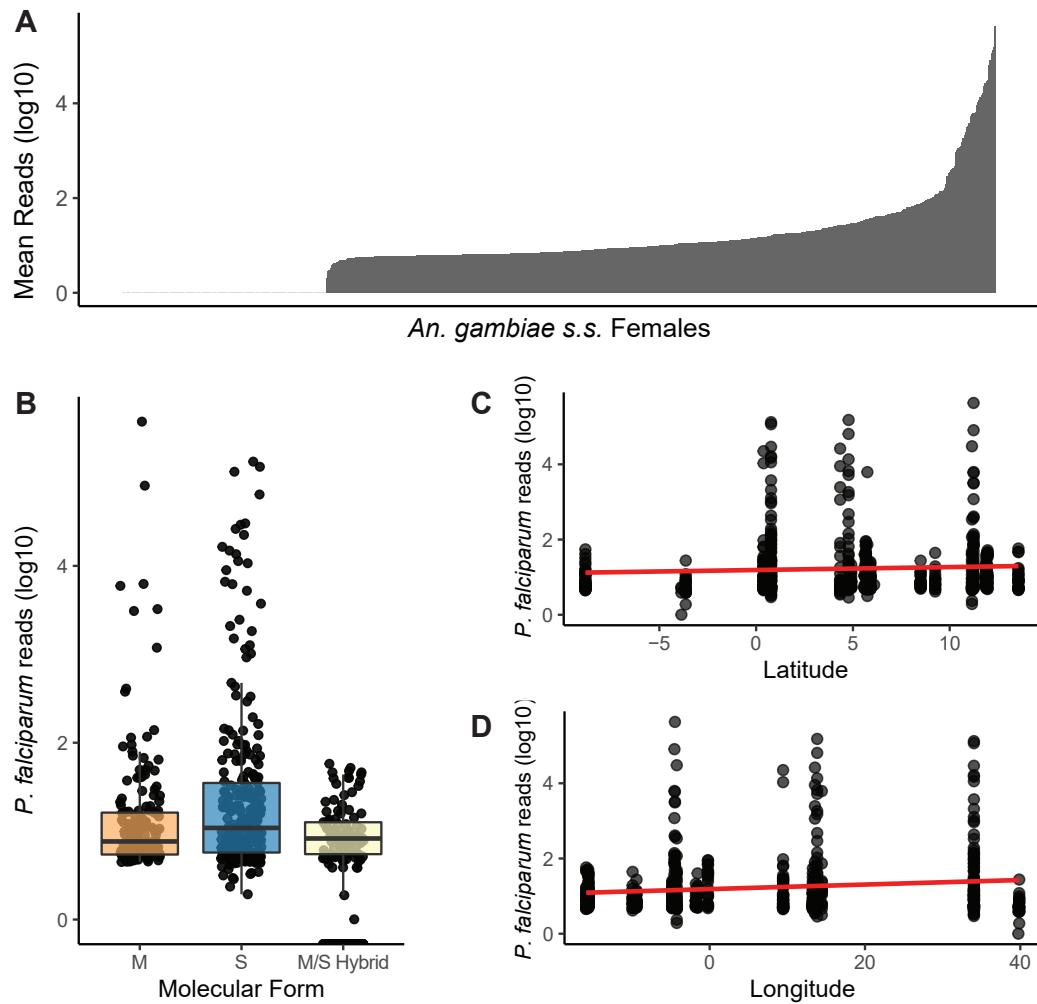


Figure 6. *Plasmodium falciparum* abundance across the 711 female *An. gambiae* s.s. individuals sampled. (A) 545 of the sampled *An. gambiae* s.s. individuals had at least 1 read that was identified as *P. falciparum*. (B) There is no significant difference in the mean *P. falciparum* reads between the molecular forms of *An. gambiae* s.s. ($F = 0.03$, $p = 0.971$). (C) There is no significant relationship between the latitude from which a sample was collected, and the number of *P. falciparum* reads identified ($t = 0.758$, $p = 0.45$). (D) Likewise, there is no significant relationship between the longitude from which a sample was collected, and the number of *P. falciparum* reads identified ($t = 0.608$, $p = 0.54$).

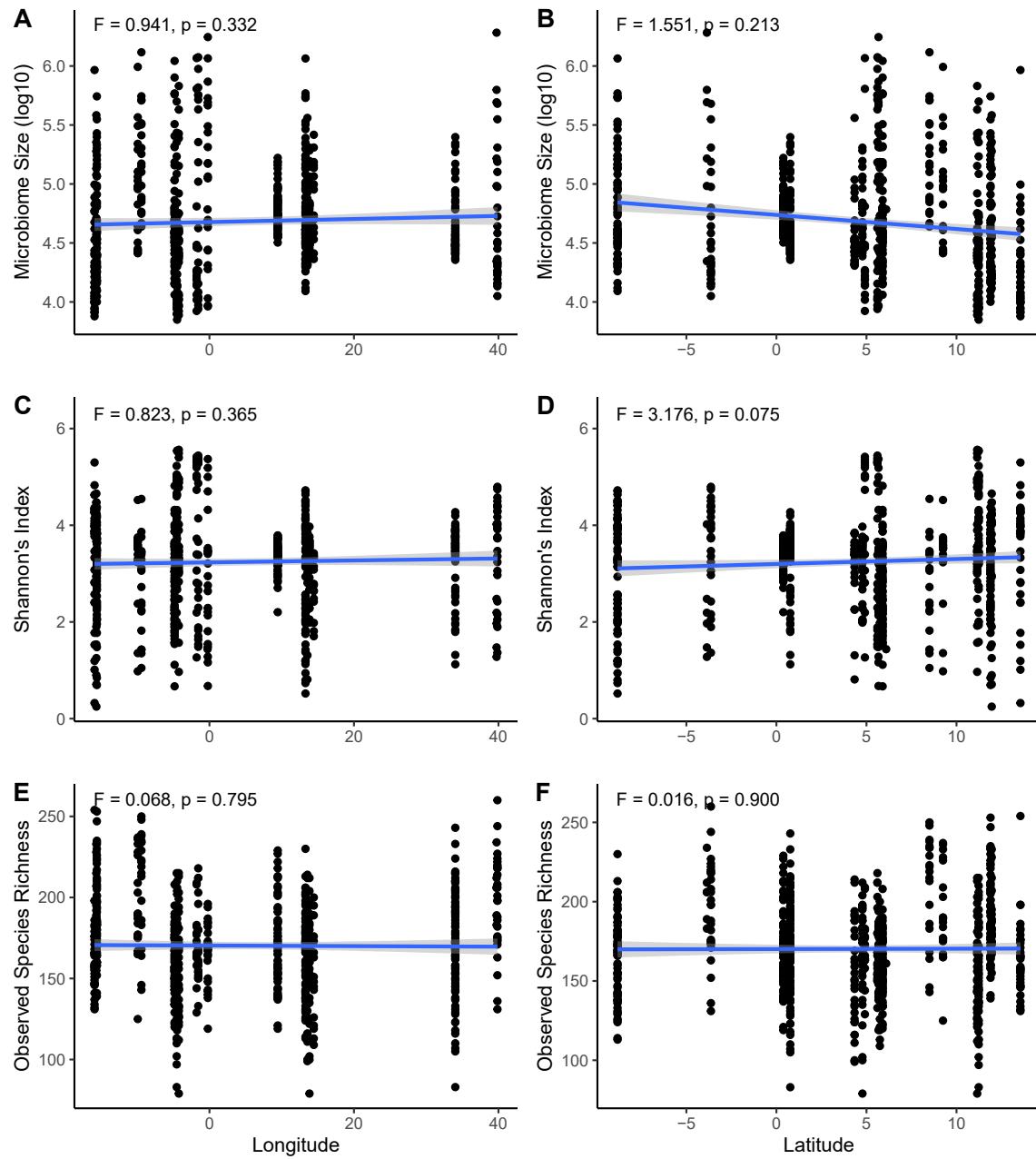


Figure S1. *An. gambiae* s.s. microbiome diversity as a function of geographic location. (A-B) Total microbiome size is not significantly correlated with either longitude ($p = 0.332$) or latitude ($p = 0.213$). (C-D) Similarly, Shannon's diversity index is not significantly correlated with either longitude ($p = 0.365$) or latitude ($p = 0.075$). (E-F) Finally, OSR is not significantly correlated with either longitude ($p = 0.795$) or latitude ($p = 0.900$).

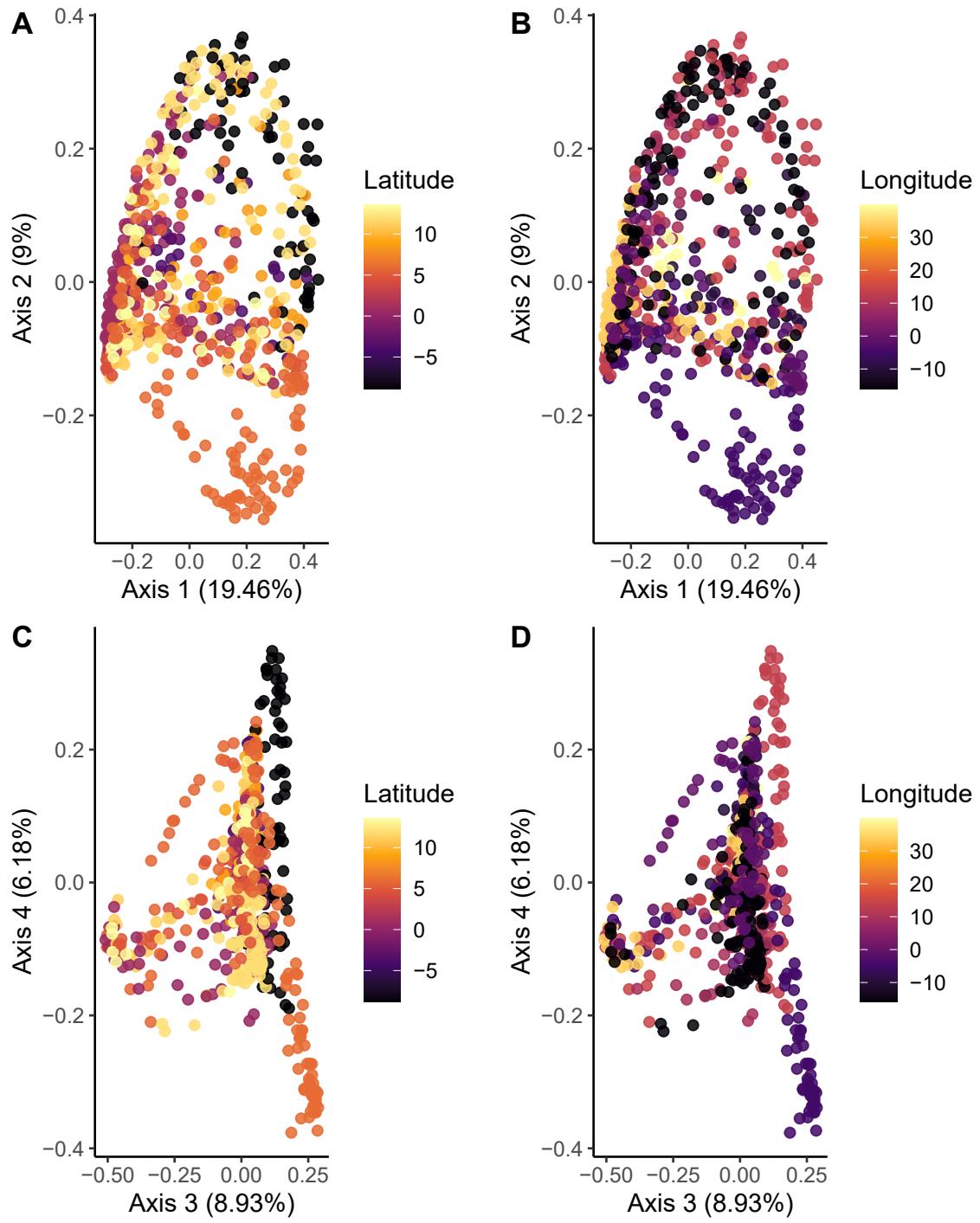


Figure S2. *An. gambiae* s.s. microbiome composition as a function of geographic location. (A) Axis 1 and Axis 2 loading values are both significantly correlated with the latitude of sample collection sites (Axis 1 $p = 0.0002$; Axis 2 $p = 7.129\text{e-}09$). (B) Axis 1 loading values are

significantly correlated with the longitude of sample collection sites, but Axis 2 is not (Axis 1 p = 1.272e-08; Axis 2 p = 0.399). (C) Axes 3 and 4 loading values are both significantly correlated with the latitude of sample collection sites (Axis 3 p = 1.279e-07; Axis 4 p = 3.322e-07). (D) Axis 3 loading values are not significantly correlated with the longitude of sample collection, but Axis 4 loading values are (Axis 3 p = 0.006; Axis 4 p = 2.552e-10).

3.8 References

1. Sinka, M. E. *et al.* A global map of dominant malaria vectors. *Parasites and Vectors* **5**, (2012).
2. Costantini, C., Sagnon, N., della Torre, A. & Coluzzi, M. Mosquito behavioural aspects of vector-human interactions in the *Anopheles gambiae* complex. *Parassitologia* **41**, 209–217 (1999).
3. Loughlin, S. O. The expanding *Anopheles gambiae* species complex. *Pathog. Glob. Health* **114**, 1 (2020).
4. Coluzzi, M., Sabatini, A., Petrarca, V. & Angela Di Deco, M. Behavioural divergences between mosquitoes with different inversion karyotypes in polymorphic populations of the *Anopheles gambiae* complex. *Nat. 1977 2665605* **266**, 832–833 (1977).
5. U.S. Department of Health and Human Services. Malaria. *Centers Dis. Control Prev.* (2022).
6. WHO, W. H. O. World Malaria Report. *World Heal. Organ.* (2021).
7. Basu, S. & Sahi, P. K. Malaria: An Update. *Indian Journal of Pediatrics* vol. 84 521–528 (2017).
8. Sharma, R. & Dutta, A. K. Malaria and National Vector Borne Disease Control

- Programme. *Indian J. Pediatr.* **78**, 1527–1535 (2011).
9. U.S. President's Malaria Initiative. U.S. President's Malaria Initiative: 16th Annual Report to Congress. (2022).
 10. Lehmann, T. & Diabate, A. The molecular forms of *Anopheles gambiae*: A phenotypic perspective. *Infect. Genet. Evol.* **8**, 737 (2008).
 11. Chandre, F. *et al.* Current distribution of a pyrethroid resistance gene (kdr) in *Anopheles gambiae* complex from west Africa and further evidence for reproductive isolation of the Mopti form. *Parassitologia* **41**, 319–322 (1999).
 12. Wang, R., Zheng, L., Touré, Y. T., Dandekar, T. & Kafatos, F. C. When genetic distance matters: Measuring genetic differentiation at microsatellite loci in whole-genome scans of recent and incipient mosquito species. *Proc. Natl. Acad. Sci. U. S. A.* **98**, 10769–10774 (2001).
 13. Stump, A. D. *et al.* Centromere-proximal differentiation and speciation in *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 15930–15935 (2005).
 14. Lehmann, T. *et al.* Population Structure of *Anopheles gambiae* in Africa. *J. Hered.* **94**, 133–147 (2003).
 15. Niang, E. H. A., Konaté, L., Diallo, M., Faye, O. & Dia, I. Reproductive Isolation among Sympatric Molecular Forms of *An. gambiae* from Inland Areas of South-Eastern Senegal. *PLoS One* **9**, e104622 (2014).
 16. Wondji, C., Simard, F. & Fontenille, D. Evidence for genetic differentiation between the molecular forms M and S within the Forest chromosomal form of *Anopheles gambiae* in an area of sympatry. *Insect Mol. Biol.* **11**, 11–19 (2002).
 17. Lee, Y. *et al.* Spatiotemporal dynamics of gene flow and hybrid fitness between the M and

- S forms of the malaria mosquito, *Anopheles gambiae*. *Proc. Natl. Acad. Sci. U. S. A.* **110**, 19854–19859 (2013).
18. Diabaté, A. *et al.* Mixed Swarms of the Molecular M and S Forms of *Anopheles gambiae* (Diptera: Culicidae) in Sympatric Area from Burkina Faso. *J. Med. Entomol.* **43**, 480–483 (2006).
 19. Marsden, C. D. *et al.* Asymmetric introgression between the M and S forms of the malaria vector, *Anopheles gambiae*, maintains divergence despite extensive hybridisation. *Mol. Ecol.* **20**, 4994 (2011).
 20. van Tol, S. & Dimopoulos, G. Influences of the Mosquito Microbiota on Vector Competence. *Adv. Insect Phys.* **51**, 243–291 (2016).
 21. Cirimotich, C. M., Ramirez, J. L. & Dimopoulos, G. Native Microbiota Shape Insect Vector Competence for Human Pathogens. *Cell Host Microbe* **10**, 307–310 (2011).
 22. Dennison, N. J., Jupatanakul, N. & Dimopoulos, G. The mosquito microbiota influences vector competence for human pathogens. *Curr. Opin. Insect Sci.* **3**, 6–13 (2014).
 23. Jupatanakul, N., Sim, S. & Dimopoulos, G. The Insect Microbiome Modulates Vector Competence for Arboviruses. *Viruses* **6**, 4294–4313 (2014).
 24. Weiss, B. & Aksoy, S. Microbiome influences on insect host vector competence. *Trends Parasitol.* **27**, 514–522 (2011).
 25. Romoli, O. & Gendrin, M. The tripartite interactions between the mosquito, its microbiota and Plasmodium. *Parasit. Vectors* **11**, 200 (2018).
 26. Wang, Y., Gilbreath, T. M., Kukutla, P., Yan, G. & Xu, J. Dynamic Gut Microbiome across Life History of the Malaria Mosquito *Anopheles gambiae* in Kenya. *PLoS One* **6**, e24767 (2011).

27. Coon, K. L., Vogel, K. J., Brown, M. R. & Strand, M. R. Mosquitoes rely on their gut microbiota for development. *Mol. Ecol.* **23**, 2727–2739 (2014).
28. Chen, S. *et al.* Multiple Factors Determine the Structure of Bacterial Communities Associated With *Aedes albopictus* Under Artificial Rearing Conditions. *Front. Microbiol.* **11**, 605 (2020).
29. Parfrey, L. W., Moreau, C. S. & Russell, J. A. The host-associated microbiome: Pattern, process and function. *Mol. Ecol.* **27**, 1749–1765 (2018).
30. Shade, A. & Handelsman, J. Beyond the Venn diagram: the hunt for a core microbiome. *Environ. Microbiol.* **14**, 4–12 (2012).
31. The Anopheles gambiae 1000 Genomes Consortium. Ag1000G phase 2 AR1 data release. *MalariaGEN*. <http://www.malariaegen.net/data/ag1000g-phase2-ar1>. (2017).
32. The Anopheles gambiae 1000 Genomes Consortium. Genetic diversity of the African malaria vector *Anopheles gambiae*. *Nature* **552**, 96 (2017).
33. Scott, J. A., Brogdon, W. G. & Collins, F. H. Identification of single specimens of the *Anopheles gambiae* complex by the polymerase chain reaction. *Am. J. Trop. Med. Hyg.* **49**, 520–529 (1993).
34. Fanello, C., Santolamazza, F. & Della Torre, A. Simultaneous identification of species and molecular forms of the *Anopheles gambiae* complex by PCR-RFLP. *Med. Vet. Entomol.* **16**, 461–464 (2002).
35. Santolamazza, F. *et al.* Insertion polymorphisms of SINE200 retrotransposons within speciation islands of *Anopheles gambiae* molecular forms. *Malar. J.* **7**, 1–10 (2008).
36. Breitwieser, F. P., Baker, D. N. & Salzberg, S. L. KrakenUniq: Confident and fast metagenomics classification using unique k-mer counts. *Genome Biol.* **19**, 198 (2018).

37. Sharakhova, M. V. *et al.* Update of the *Anopheles gambiae* pest genome assembly. *Genome Biol.* **8**, (2007).
38. Morgulis, A., Gertz, E. M., Schäffer, A. A. & Agarwala, R. A fast and symmetric DUST implementation to mask low-complexity DNA sequences. *J. Comput. Biol.* **13**, 1028–1040 (2006).
39. Shannon, C. E. A Mathematical Theory of Communication. *Bell Syst. Tech. J.* **27**, 379–423 (1948).
40. Team, R. C. R: A Language and Environment for Statistical Computing. (2021).
41. Bray, J. R. & Curtis, J. T. An Ordination of the Upland Forest Communities of Southern Wisconsin. *Ecol. Monogr.* **27**, 325–349 (1957).
42. Oksanen, J. *et al.* vegan: Community Ecology Package. (2020).
43. Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in. *Bioinformatics* **35**, 526–528 (2019).
44. Berhanu, A. *et al.* Isolation and identification of microflora from the midgut and salivary glands of *Anopheles* species in malaria endemic areas of Ethiopia. *BMC Microbiol.* **19**, (2019).
45. Wang, Y., Gilbreath, T. M., Kukutla, P., Yan, G. & Xu, J. Dynamic Gut Microbiome across Life History of the Malaria Mosquito *Anopheles gambiae* in Kenya. *PLoS One* **6**, e24767 (2011).
46. Singh, A. *et al.* The dynamic gut microbiota of zoophilic members of the *Anopheles gambiae* complex (Diptera: Culicidae). *Sci. Rep.* **12**, 1495 (2022).
47. Straif, S. C. *et al.* Midgut Bacteria in *Anopheles gambiae* and *An. funestus* (Diptera: Culicidae) from Kenya and Mali. *J. Med. Entomol.* **35**, 222–226 (1998).

48. Arruda, A. *et al.* Diversity of Culturable Bacteria Isolated From the Feces of Wild Anopheles darlingi (Diptera: Culicidae) Mosquitoes From the Brazilian Amazon. *J. Med. Entomol.* **58**, 1900–1907 (2021).
49. Rocha, E. M. *et al.* Culturable bacteria associated with Anopheles darlingi and their paratransgenesis potential. *Malar. J.* **20**, (2021).
50. Feng, Y. *et al.* The Microbiota of Three Anopheles Species in China. *J. Am. Mosq. Control Assoc.* **37**, 38–40 (2021).
51. Gendrin, M. & Christophides, G. K. The Anopheles mosquito microbiota and their impact on pathogen transmission. in *Anopheles mosquitoes - New insights into malaria vectors* (ed. Manguin, S.) (IntechOpen, 2013). doi:10.5772/55107.
52. Ramos-Nino, M. E. *et al.* Metagenomic analysis of Aedes aegypti and Culex quinquefasciatus mosquitoes from Grenada, West Indies. *PLoS One* **15**, e0231047 (2020).
53. Ngo, C. T., Romano-Bertrand, S., Manguin, S. & Jumas-Bilak, E. Diversity of the bacterial microbiota of Anopheles mosquitoes from binh Phuoc Province, Vietnam. *Front. Microbiol.* **7**, 2095 (2016).
54. Dinparast Djadid, N. *et al.* Identification of the Midgut Microbiota of An. stephensi and An. maculipennis for Their Application as a Paratransgenic Tool against Malaria. *PLoS One* **6**, e28484 (2011).
55. Galeano-Castañeda, Y., Bascuñán, P., Serre, D. & Correa, M. M. Trans-stadial fate of the gut bacterial microbiota in Anopheles albimanus. *Acta Trop.* **201**, 105204 (2020).
56. Dong, Y., Manfredini, F. & Dimopoulos, G. Implication of the mosquito midgut microbiota in the defense against malaria parasites. *PLoS Pathog.* **5**, (2009).
57. Rancès, E., Ye, Y. H., Woolfit, M., McGraw, E. A. & O'Neill, S. L. The relative

- importance of innate immune priming in Wolbachia-mediated dengue interference. *PLoS Pathog.* **8**, e1002548 (2012).
58. Pan, X., Thiem, S. & Xi, Z. Wolbachia-Mediated Immunity Induction in Mosquito Vectors. in *Arthropod Vector: Controller of Disease Transmission* vol. 1 35–58 (Elsevier Inc., 2017).
59. Pan, X. *et al.* Wolbachia induces reactive oxygen species (ROS)-dependent activation of the Toll pathway to control dengue virus in the mosquito *Aedes aegypti*. *Proc. Natl. Acad. Sci. U. S. A.* **109**, E23–E31 (2012).
60. Caragata, E. P. *et al.* Dietary Cholesterol Modulates Pathogen Blocking by Wolbachia. *PLoS Pathog.* **9**, e1003459 (2013).
61. Frentiu, F. D. Lipids and Pathogen Blocking by Wolbachia. *Trends in Parasitology* vol. 33 916–917 (2017).
62. Geoghegan, V. *et al.* Perturbed cholesterol and vesicular trafficking associated with dengue blocking in Wolbachia-infected *Aedes aegypti* cells. *Nat. Commun.* **8**, 1–10 (2017).
63. Tuteja, R. Malaria – An Overview. *FEBS J.* **274**, 4670–4679 (2007).
64. Sharma, P. *et al.* Salivary glands harbor more diverse microbial communities than gut in *Anopheles culicifacies*. *Parasit. Vectors* **7**, 235 (2014).
65. Hegde, S., Rasgon, J. L. & Hughes, G. L. The microbiome modulates arbovirus transmission in mosquitoes. *Current Opinion in Virology* vol. 15 97–102 (2015).
66. Zink, S. D., van Slyke, G. A., Palumbo, M. J., Kramer, L. D. & Ciota, A. T. Exposure to west nile virus increases bacterial diversity and immune gene expression in *culex pipiens*. *Viruses* **7**, 5619–5631 (2015).

67. Arévalo-Cortés, A. *et al.* The midgut microbiota of colombian aedes aegypti populations with different levels of resistance to the insecticide lambda-cyhalothrin. *Insects* **11**, 1–19 (2020).
68. McMeniman, C. J. *et al.* Stable introduction of a life-shortening Wolbachia infection into the mosquito Aedes aegypti. *Science (80-.).* **323**, 141–144 (2009).
69. Saab, S. A. *et al.* The environment and species affect gut bacteria composition in laboratory co-cultured Anopheles gambiae and Aedes albopictus mosquitoes. *Sci. Reports 2020 101* **10**, 1–13 (2020).
70. Sina, H. *et al.* Bacteria Load Determination of the Intestinal Microbiota and Identification of Spiroplasma and Wolbachia in Anopheles gambiae. *Int. J. Zool.* **2022**, 1–9 (2022).
71. David, M. R., Santos, L. M. B. dos, Vicente, A. C. P. & Maciel-de-Freitas, R. Effects of environment, dietary regime and ageing on the dengue vector microbiota: evidence of a core microbiota throughout Aedes aegypti lifespan. *Mem. Inst. Oswaldo Cruz* **111**, 577–587 (2016).
72. Risely, A. Applying the core microbiome to understand host–microbe systems. *J. Anim. Ecol.* **89**, 1549–1558 (2020).
73. Guégan, M. *et al.* The mosquito holobiont: fresh insight into mosquito-microbiota interactions. *Microbiome* **6**, 49 (2018).
74. Yadav, K. K. *et al.* Diversity of Cultivable Midgut Microbiota at Different Stages of the Asian Tiger Mosquito, Aedes albopictus from Tezpur, India. *PLoS One* **11**, (2016).
75. Terenius, O. *et al.* Midgut bacterial dynamics in Aedes aegypti. *FEMS Microbiol. Ecol.* **80**, 556–565 (2012).
76. Ippolito, M. M., Denny, J. E., Langelier, C., Sears, C. L. & Schmidt, N. W. Malaria and

- the Microbiome: A Systematic Review. *Clin. Infect. Dis.* **67**, 1831–1839 (2018).
77. Scolari, F., Casiraghi, M. & Bonizzoni, M. Aedes spp. and Their Microbiota: A Review. *Frontiers in Microbiology* vol. 10 (2019).
78. Hegde, S. *et al.* Microbiome Interaction Networks and Community Structure From Laboratory-Reared and Field-Collected *Aedes aegypti*, *Aedes albopictus*, and *Culex quinquefasciatus* Mosquito Vectors. *Front. Microbiol.* **9**, 2160 (2018).
79. Audsley, M. D., Ye, Y. H. & McGraw, E. A. The microbiome composition of *Aedes aegypti* is not critical for Wolbachia-mediated inhibition of dengue virus. *PLoS Negl. Trop. Dis.* **11**, e0005426 (2017).
80. Audsley, M. D. *et al.* Wolbachia infection alters the relative abundance of resident bacteria in adult *Aedes aegypti* mosquitoes, but not larvae. *Mol. Ecol.* **27**, 297–309 (2017).
81. Cansado-Utrilla, C., Zhao, S. Y., McCall, P. J., Coon, K. L. & Hughes, G. L. The microbiome and mosquito vectorial capacity: rich potential for discovery and translation. *Microbiome* **9**, 1–11 (2021).
82. Dada, N. *et al.* Considerations for mosquito microbiome research from the Mosquito Microbiome Consortium. *Microbiome* **9**, 1–16 (2021).

CHAPTER 4: CONCLUSION

4.1 Conclusion

The objective of this thesis project was to investigate natural mosquito microbiome variation using novel approaches that leverage an abundance of publicly available sequencing data. These studies are not only some of the largest surveys of microbiome variation in wild populations of *Ae. aegypti* and *An. gambiae* s.s. mosquitoes, but they contain high-confidence microbial species identifications that have expanded our understanding of which bacterial species are harbored by mosquitoes in general.

As biological control strategies become an increasingly popular approach to limit the spread of human pathogens, it becomes increasingly essential to understand how the introduction of non-native species to a host interact with the natural microbiome. Microbiome incompatibility is a potential issue when introducing a non-native species; a bacterial species may be introduced to a host but inoculation does not guarantee colonization. If a non-native biological control agent is introduced to a host that harbors an incompatible microbiome, colonization may be unsuccessful and thus the suppression of pathogen transmission may be negatively impacted and even eliminated completely. Releasing artificially inoculated individuals is a costly endeavor—economically, environmentally, and socially. Risking an incompatible microbiome is avoidable but an extensive understanding of the microbiome prior to non-native introduction must be achieved. Additionally, given the important and impactful role the bacterial microbiome has on so many aspects of mosquito fitness and health, alterations to the natural microbiome need to be carefully considered^{171,172}. Otherwise we risk negatively impacting mosquito health inadvertently.

4.2 Future Directions

There are two primary lines of investigation that these projects lay the groundwork for: 1) continuing to explore the mosquito microbiome through a tissue-specific functional lens and 2) applying these methods to additional alternative datasets.

Despite the contributions these projects have made to our understanding of mosquito microbiome dynamics, there remain questions regarding how the microbiome interacts with its mosquito host, especially in a tissue-specific manner and functionally. Here, I have characterized the *Ae. aegypti* and *An. gambiae* bacterial microbiomes in depth on the individual but was limited to whole-body samples. It is well-established that distinct tissues harbor distinct microbiomes. Within the gut, variation in microbial composition exists and diversity differs across the foregut, midgut and hindgut^{173,174}. To further expand these studies, a tissue-specific characterization of the mosquito microbiome would further elucidate microbial interactions in the microbiome. By having knowledge about where microbial species are spatially found in an individual, one can begin to identify species that are physically interacting, which is an important consideration when attempting to classify a relationship as positive, negative, or neutral between species^{175–178}. When considering the mosquito life history, the three most relevant tissues for investigation are the salivary glands, the gut, and the reproductive tissue. Each of these tissues are obviously functionally important to the mosquito itself, but they are also important for the human pathogens life cycles that mosquitos transmit.

Furthermore, a functional understanding of the mosquito microbiome is a logical future objective. Here, I have characterized the microbiome and observed complex relationships between numerous bacterial species. However, a gene-centric metagenomic approach is an alternative way to characterize a microbiome. Rather than identifying microbial species

composition, a gene-centric approach quantifies gene content and then analyzes the functional pathways that are enriched in a microbiome. This approach minimizes the emphasis that is placed on the variation in microbiome diversity and composition and allows the actual functional capacity of a microbiome to be explored. Furthermore, a functional analysis would directly address the question as to why species-level diversity may fluctuate between individuals but there is more consistency at in composition at high taxonomic levels.

Finally, the methods used here are not limited to mosquito genomic data. Other datasets exist that can monopolize on these methods and follow similar analytical pipelines to efficiently investigate other model organisms. Characterizing a microbiome is inherently exploratory, but it is a necessary step to understand more complex relationships both between a host and microbe but also between microbial species in the microbiome. Moreover, an extensive catalog of microbiome variation in a broad range of models will eventually lead to exploring larger patterns of microbiome composition. For example: Do all mosquito species have a core set of microbes or is microbiome composition species-specific? Do different species from the same geographic region have more similar microbiome composition or is microbiome composition independent of geography? These questions have been asked in a limited number of model organisms like humans and insect species, but without a more wholistic view of diverse representative species, these questions will remain unanswered.

4.3 References

1. Benelli, G., Jeffries, C. L. & Walker, T. Biological Control of Mosquito Vectors: Past, Present, and Future. *Insects* **7**, 52 (2016).
2. Huang, Y. J. S., Higgs, S. & Vanlandingham, D. L. Biological Control Strategies for Mosquito Vectors of Arboviruses. *Insects* **8**, 21 (2017).

3. Buchon, N. *et al.* Morphological and Molecular Characterization of Adult Midgut Compartmentalization in Drosophila. *Cell Rep.* **3**, 1725–1738 (2013).
4. Marianes, A. & Spradling, A. C. Physiological and stem cell compartmentalization within the Drosophila midgut. *Elife* **2013**, (2013).
5. Rossbach, S., Cardenas, A., Perna, G., Duarte, C. M. & Voolstra, C. R. Tissue-Specific Microbiomes of the Red Sea Giant Clam *Tridacna maxima* Highlight Differential Abundance of Endozoicomonadaceae. *Front. Microbiol.* **10**, 2661 (2019).
6. Dubé, C. E., Ky, C. L. & Planes, S. Microbiome of the black-lipped pearl oyster *Pinctada margaritifera*, a multi-tissue description with functional profiling. *Front. Microbiol.* **10**, 1548 (2019).
7. Anderson, K. E., Ricigliano, V. A., Copeland, D. C., Mott, B. M. & Maes, P. Social Interaction is Unnecessary for Hindgut Microbiome Transmission in Honey Bees: The Effect of Diet and Social Exposure on Tissue-Specific Microbiome Assembly. *Microb. Ecol.* **1**, 1–16 (2022).
8. Gall, C. A., Scoles, G. A., Magori, K., Mason, K. L. & Brayton, K. A. Laboratory colonization stabilizes the naturally dynamic microbiome composition of field collected *Dermacentor andersoni* ticks. *Microbiome* **5**, 133 (2017).

Jane A. Pascar

Center for Reproductive Evolution
Department of Biology
Syracuse University

Email: janepascar@gmail.com
Web: cre.syr.edu
Github: @janepascar

EDUCATION

- 2022 M.Sc. in Biology
Syracuse University
- 2017 B.Sc. in Zoology, *cum laude* with University Honors
State University of New York at Oswego

RESEARCH EXPERIENCE

Academic Experience

- 2018–22 Graduate Research, Syracuse University, Center for Reproductive Evolution
Advisor: Dr. Steve Dorus
- 2017–18 Research Technician, University of Southern California, Marine and Environmental Biology
PI: Dr. Suzanne Edmands
- 2016–17 Undergraduate Honors Thesis Research, SUNY Oswego, Department of Biological Sciences
Advisor: Dr. Chris Chandler
- 2016 Visiting Scholar, Université de Poitiers, Laboratory of Ecological and Biological Interactions
PI: Dr. Richard Cordaux

Industry Experience

- 2022 Research Scientist, Dairy Farmers of America, Acumen Detection Inc.

GRANTS and AWARDS

Awards and Honors

- 2017 Sigma Xi/SUNY Oswego Office of Research and Sponsored Programs Research Award

Grants and Fellowships

- 2019–22 National Science Foundation Graduate Research Fellowship
- 2019 Syracuse University Dept. of Biology Travel Grant
- 2018–19 Syracuse University STEM Research Fellowship
- 2013–17 SUNY Oswego Presidential Scholarship
- 2013 CCU High School Scholarship Program

PUBLICATIONS

Published Articles

1. **Pascar, J.** and Chandler, C.H. 2018. A bioinformatics approach to identifying *Wolbachia* infections in arthropods. *PeerJ* 6:e5486.
2. Russell, A., Borrelli, S., Fontana, R., Laricchiuta, J. **Pascar, J.**, Becking, T., Giraud, I., Cordaux, R., Chandler, C.H. 2021. Evolutionary transition to XY sex chromosomes associated with Y-linked duplication of a male hormone gene in a terrestrial isopod. *Heredity* 127:266–277.

Submitted Articles

3. Watson, E.T., Flanagan, B.A., **Pascar, J.**, and Edmands, S. Mitochondrial effects on fertility and longevity in *Tigriopus californicus* contradict predictions of the mother's curse hypothesis. *Submitted to Current Biology*. 2022.
4. Watson, E.T., **Pascar, J.**, and Edmands, S. Rapid evolution of sex-limited genes in *Tigriopus californicus*, a marine copepod with polygenic sex determination. *Submitted to Evolution*. 2022.
5. **Pascar, J.**, Middleton, H., and Dorus, S. *Aedes aegypti* microbiome covaries with *Wolbachia* infection. *Submitted to Microbiome*. 2022.

CONFERENCE PRESENTATIONS

* indicates presenter † indicates undergraduate advisee

Talks

- 2020 Russell, A., Borrelli, S., Fontana, R., Laricchiuta, J. **Pascar, J.**, Becking, T., Giraud, I., Cordaux, R., Chandler, C.H.* A transition to XY sex chromosomes associated with Y-linked duplication of a male hormone gene in a terrestrial isopod. Society for Integrative and Comparative Biology (SICB), Austin, TX
- 2017 **Pascar, J.*** and Chandler, C.H. Identification of novel *Wolbachia* infections in arthropods using publicly accessible next-generation sequencing data. Quest, Oswego, NY • Awarded Sigma Xi/ORSP Research Award

Posters

- 2021 **Pascar, J.***, Middleton, H., Dorus, S. *Aedes aegypti* microbiome composition covaries with *Wolbachia* infection. Great Lakes Annual Meeting of Evolutionary Genomics (Virtual)
- 2019 Middleton, H.†*, **Pascar, J.**, Dorus, S. Microbiome patterns in *Aedes aegypti* mosquitoes transinfected with *Wolbachia*. Syracuse University Summer Undergraduate Research Conference, Syracuse, NY
- 2018 **Pascar, J.***, Watson, E.T., Edmands, S. Properties of sex-biased gene expression in the absence of sex chromosomes. Population, Evolution and Quantitative Genetics (PEQG), Madison, WI
- 2017 **Pascar, J.*** and Chandler, C.H. Testing for prevalence of *Wolbachia* in the terrestrial isopod species *Porcellio laevis* and *Trachelipus rathkei*. Society for Integrative and Comparative Biology (SICB), New Orleans, LA

- 2016 **Pascar, J.*** and Chandler, C.H. Testing for prevalence of *Wolbachia* in the terrestrial isopod species *Porcellio laevis* and *Trachelipus rathkei*. RISE Scholarly & Creative Activities Symposium, Oswego, NY

TEACHING EXPERIENCE

SUNY Oswego

- 2016 Microbiology Lab (BIO 310) - One Semester
2015 Molecular and Cellular Biology (BIO 120) - One Semester
2015 Introduction to the Honors Program (HON 150) - One Semester

PROFESSIONAL DEVELOPMENT

Workshops

- 2020 Microbial 'Omics Online Series- University of Chicago, Meren Lab
2019 Genomic Sequence Data Analysis - Cornell University, Institute of Biotechnology

Memberships

- New York Academy of Sciences
Society for the Study of Evolution
Syracuse University Biology Graduate Student Organization

SERVICE

Outreach

- 2021 **Undergraduate Research Festival Judge:** Judged Syracuse University undergraduate research posters in biology, biotechnology, and chemistry.
2019 **Westcott Community Center Kids Club:** As part of the Biology Graduate Student Organization Outreach Committee, I volunteered at a local after school program where we perform interactive science activities.
2019 **Letters to a Pre-Scientist:** LPS pairs middle school students with scientists at all career levels. Over the course of an academic year I exchanged letters with a student discussing my educational experiences, overcoming obstacles, and careers possibilities.
2019 **Frontiers in Science:** I co-led a lab activity for high school students on *Drosophila* genetics. Students studied flies under dissecting microscopes and identified various visible mutations.
2018 **USC McMorrow Neighborhood Academic Initiative:** The NAI is a pre-college enrichment program for low-income students in 6th through 12th grade in the Los Angeles area. I traveled with a group of these students to the Wrigley Marine Science Center on Catalina Island to collect intertidal copepods to conduct short-term observational studies. This introduced the students to evolutionary topics such as the mother's curse hypothesis and maintenance of polygenic sex determination. Throughout the year I helped host lab tours for participating teachers and students.

- 2018 **Darwin Day Workshop:** I helped facilitate a workshop with 50 Los Angeles K-12 teachers on lab activities they can implement in their classrooms to bolster their teaching of evolution.

Mentoring

Henry Middleton - Cornell University, Computer Science, 2 summers full-time research

Peter Wengert - Rochester Institute of Technology, Bioinformatics, 9 months full-time research

Syracuse University Service

- 2020 **Biology Dept. Website Diversity Content Committee:** I was the graduate student representative on the committee tasked with developing the portion of the Biology Department website that highlights our departmental contributions to promoting diversity and providing both on and off campus resources.
- 2020 **Biology Graduate Student Organization Social Committee:** I participated in planning annual departmental events and special events that promote social interaction between graduate students and faculty.

MEDIA COVERAGE

- 2019 *Syracuse University News*. Students Earn 2019 National Science Foundation Awards. Apr 24.
- 2018 *Wrigley Institute Research Blog*. Sex-biased expression in the copepod *Tigriopus*. May 16.