Microbiomes, Phylogenomics & Heat Shock Proteins of Bee Flies

By

Allan Cabrero

A dissertation submitted in partial satisfaction of the

requirements for the degree of

Doctor of Philosophy

in

Environmental Science, Policy, and Management

in the

Graduate Division

of the

University of California, Berkeley

Committee in charge:

Professor Kipling Will, Chair
Professor Nicholas Mills
Professor Rauri Bowie
Professor Patrick O'Grady

Spring 2023

Abstract

Microbiomes, Phylogenomics & Heat Shock Proteins of Bee Flies

By

Allan Cabrero

Doctor of Philosophy in Environmental Science, Policy, and Management

University of California, Berkeley

Professor Kipling Will, Chair

Bombyliidae or bee flies are a large family of ~5,000 species, these flies have a cosmopolitan distribution and are ubiquitous in arid and semi-arid regions. Bee flies are particularly important as pollinators of many desert wildflowers in Africa, including the unusual Namibian gymnosperm *Welwitschia*. Studies have shown that bee flies are the keystone pollinators of flowers in the Greater Cape Floristic Region. Larval bee flies are parasitoids of other arthropods, attacking a wide variety of hosts, such as grasshoppers, wasps, and even the Tsetse fly. Despite their importance as pollinators and potential use as biological control agents, many groups within bee flies are taxonomically and ecologically poorly known. Research on their natural and evolutionary history is greatly needed to advance our meager understanding of this group. In this dissertation three aspects of bee fly evolutionary biology and ecology were studied:

First, the microbial ecology of bee flies was investigated. Microbial communities play an important role in the life of their hosts, influencing digestion, development, behavior, and even speciation. The microbiome of bee flies was sequenced, and the composition of their microbial communities was identified. Bee flies were found to have a primarily transient microbiome, with some core resident microbes found across all samples. The microbes found in bee flies play an important role in pollen and nectar feeding; additionally, bacteria associated with the processing of chitin were also identified.

For the second chapter of my dissertation, the phylogenetic relationships of the genus *Lordotus* were investigated. Prevalent in arid habitats, *Lordotus* is restricted to the western United States and Mexico, with most species diversity in the Southwest. To date, 29 species are described, all recorded in California. Historically *Lordotus* species have been challenging to delimit, due to their highly variable morphology. Previous authors have attempted to organize *Lordotus* into three species groups, grouping species based on the characteristics of the antennal segments and wing vestiture. Using ultraconserved elements (UCEs), I generated the first phylogenetic hypothesis for *Lordotus*. A UCE dataset consisting of 936 loci from 76 individuals representing all species of the genus was generated using a combination of flies preserved in ethanol and historic, pinned museum specimens The three species groups suggested by previous authors were found to be paraphyletic. These results provide a much-needed foundation for future, detailed revision of *Lordotus*.

The third chapter of my dissertation generated low-coverage genomes for seven bee flies, across different subfamilies. The Dipteran genomes of fruit flies and mosquitoes have long been studied, however, genomes for other groups of flies remain lacking. The assembled bee fly genomes represent a valuable resource, allowing researchers to study a wide range of topics. One

such topic of interest is the identification and evolution of heat shock proteins (HSPs). These are conserved proteins found in virtually all organisms. HSPs play an important role in protecting and stabilizing DNA during periods of heat stress. Desert-adapted organisms have been shown to have higher amounts of HSPs, however, these studies have been limited to select groups. The genomes were used to identify twelve HSPs found in bee flies, with some desert species exhibiting higher amounts of HSPs. The identified HSPs will help future researchers gain a better understanding of the evolution and adaptations of desert flies.

A mi madre, que sobre todo sacrificó tanto para ver sus hijos florecer.
Gracias por todo. Te amo.

Table of Contents

Acknowledgments

There are a lot of people who have supported and encouraged me throughout my academic journey, without them I would not be the person I am today and I would not have finished this dissertation. First, above all, I would like to thank my mother, Elidia Cabrero, who has tirelessly supported and encouraged all endeavors in my life. She sacrificed so much of her own happiness to see her children succeed and for that I will always be eternally grateful.

I am very lucky that I had two kind and supportive advisors throughout my PhD, Kipling Will, and Patrick O'Grady. A very special thank you to Kip for being an amazing advisor, mentor, and of course for adopting me into his lab. I appreciate all his support, guidance, and putting up with all my fly-related jokes. I tried really hard to convince him that bee flies are the better insect, but his love for carabids held out. I am also very grateful for having Patrick O'Grady as an advisor, for accepting me into his lab, and for all his support. Despite the distance, Patrick was always there to help and provided so much encouragement when needed throughout graduate school.

I would like to thank my committee members, Nick Mills and Rauri Bowie, for their valuable feedback and guidance with my dissertation.

Throughout this journey I had an indispensable support system of friends and family. My father, Aaron C. Cabrero and my brother Aaron Cabrero Jr. both provided so much love and support. To all my Tios, Tias and primos, thank you for all the fun times and family gatherings, they are some of my most cherished memories.

I thank my lab mates: Nina Pak (#TeamDiptera), Sean Perez for all their help throughout my dissertation. I also thank my friends: Louis Armenta, Gabriela Ponce, Michael Kong, Erika Garcia, Ben Caro-Lopez, Marseille Mondragon Schrader, Sierra Everly, Andrew Lee, Johannes Henriksson, Marie Rasitera, Charlotte H. E. Alberts(#TeamDiptera), George Alberts, Jake Tracy, Stephanie Kong, Debbie Fong Kong, Anna Holmquist, Lynne Hagelthorn, Jackie Childers, Agustin Cruz, Grace Towers, Juanita More Chelsea Andreozzi, Frankie Soares, Thao Doan, Ardia Kelker and Kedra Davis for all the fun memories, support and encouragement.

I thank Torsten Dikow for facilitating my interest in bee flies and constantly encouraging me to continue my efforts in academia. I thank John Sauvé, who provided mentorship and so much emotional support during my PhD. Who knows where I would be without his help.

In chapter one of my dissertation, I thank Sean Perez and Aman Gill for all the help with the analysis and project design. Sean especially helped guide through each step of the analysis and took the time to teach me how to use R. I thank Patrick O'Grady for funding the field work to collect some of the bee flies. I thank Kip Will for funding the sequencing of the bee fly microbiomes.

In my second chapter I thank Charlotte Alberts, Ziad Khouri and Torsten Dikow for designing the asiloid UCE probes and allowing me to use them for the project. I also thank Torsten for supplying the UCE library prep kit required to do this project. I huge thank you to Lydia Smith for her amazing help during all the UCE lab work, none of the lab work would have been possible without her guidance and patience. I thank Patrick O'Grady for helping during the initial project design and for funding the sequencing of the project.

In my third chapter I thank Michelle Trautwein and Patrick O'Grady for funding the sequencing of the bee fly genomes.

I thank Joyce Gross and Alice Abela for allowing me to use their amazing bee fly photographs.

I thank my funding sources, without their money this would never have happened: NSF Graduate Research Fellowship, Chancellors Fellowship for Graduate Study, Margaret C. Walker Fund for Systematic Entomology, Howie Wier Memorial Conservation Grant, Froland Family Fellowship, Center for African Studies- Rocca Dissertation Fellowship, Williston Diptera Research Fund, Graduate Remote Instruction Innovation Fellowship, Berkeley Connect Mentoring Fellowship, Berkeley Connect Bridge Fellowship and Julius H. Freitag Memorial Award.

I thank the various institutions that provided samples or allowed me to visit to collect bee fly samples: Essig Museum of Entomology, UC Riverside Entomology Research Museum, Gillette Museum of Arthropod Diversity at Colorado State University, California Academy of Sciences, and the San Diego Natural History Museum, Philip L. Boyd Deep Canyon Desert Research System, Anza-Borrego State Park.

Finally, I thank my dog Eros, who I raised as a puppy and who has been on this life journey with me for so long. Patiently waiting for me to return home. He has been an amazing emotional support that lives up to his namesake.

I am sure there are some people I missed, but I am eternally grateful for all the support and love I have received throughout my years as a graduate student. Grad school can be so challenging at times and really grind you down; if it wasn't for my support network, I don't know where I would be. Thank you, everyone.

Chapter 1

The Microbial Ecology of Bee Flies (Diptera: Bombyliidae)

Abstract

A growing interest in the field of evolutionary biology has been the study of host-associated microorganisms. Microbial communities play an important role in insects, influencing diet, development, and in some instances even a host's behavior. The microbiomes of parasitoid insects are under-explored in the microbial environment, especially in non-model systems. Here we collected 34 samples representing eight subfamilies and 18 genera of Bombyliidae. We characterized the bacterial communities of each specimen using 16S rRNA gene sequencing. We found that the composition and diversity of resident microbial communities are primarily transient, with some evidence of a possible resident microbiome. Many of the core bacteria found across all samples were found to be associated with pollen, nectar, and chitin processing. Our results provide a baseline understanding of the resident bacterial communities across the Bombyliidae, which will help with future studies in understanding how bacterial communities influence host selection and diversification within parasitoid flies.

Introduction

Microbes, which include viruses, bacteria, and fungi are some of the most ubiquitous life forms on the planet. Historically, microbes have been challenging to study, experiments were limited by the ability to reliably culture the appropriate microbe or by the cost of sequencing large amounts of microbes. However, advances in sequence technology now allow for cost-effective sequencing of large amounts of microbial DNA, enabling an improved understanding of the depth and magnitude to which microbial communities impact all life. Studies have shown that microbial communities can influence a host's behavior, digestive physiology, development, and even speciation (Engel & Moran 2013; Shropshire & Bordenstein 2016; Silver et al. 2021; Vavre & Kremer 2014; Brooks et al. 2017). Studies of the host-microbe interactions of insects have shown the influence of microbes on host and mate selection in a variety of taxa, such as *Nasonia* wasps, fruit flies (Tephritidae), and *Drosophila* (Shropshire & Bordenstein 2016; Dittmer et al. 2016; Zchori-Fein et al. 2001). In some tephritid fruit flies, the yeast present on fruits influences oviposition choice. Switching the yeast strain can cause the fly to oviposit on different fruit (Jose et al. 2019; Morrow et al. 2015).

These host microbe effects are not limited to plant-insect interactions, some studies have shown that microbial communities can influence insect host selection in parasitoid wasps (Zchori-Fein et al. 2001; Shropshire & Bordenstein 2016). Parasitoids are organisms that develop by drawing nourishment from the host, eventually killing it. Many species of insect parasitoids have been used as biological control agents to regulate and eliminate insect pests of economically important plants (Godfray 1994; Heraty 2017). Parasitoid insects can range from generalists attacking many insects, to specialists, only attacking a specific insect host (Godfray 1994; Heraty 2017). Studies in the parasitoid wasp *Nasonia*, have found that microbial communities can play a role in influencing its host selection, like the fruit flies, changing the microbial community can influence where the females oviposit (Zchori-Fein et al. 2001; Shropshire & Bordenstein 2016; Dittmer et al. 2016; Dicke et al. 2020). Microbial communities

have likely played a role in parasitoid host selection and evolution. Although *Nasonia* is a well-established model organism for studying symbiosis and parasitoid evolution, studying the microbiota of parasitoid flies may provide insight into host selection and speciation within a group that has more diverse parasitic life histories (Feener & Brown 1997). Parasitoid flies attack a wide range of hosts across five phyla and 22 orders (Feener & Brown 1997). Comparatively parasitoid wasps only attack 19 orders in the phylum Arthropoda (Feener & Brown 1997). This has led to the multiple independent evolutions of behaviors (mobile first instar larvae) and structures (piercing ovipositors) utilized by parasitoid flies (Feener & Brown 1997).

Diptera (flies) are one of the most diverse orders of insects (160,000sp +), found in some of the most extreme habitats on the planet (Marshall 2012). From microbial rich anaerobic water to the stomachs of horses and cows, Diptera are known to inhabit most of the world's ecosystems (Marshall 2012). For humans, Diptera are arguably the most important animal group on the planet, their impact ranging from vectoring of major infectious diseases (Dengue, Malaria, Yellow fever, etc.) to causing significant agriculture damage (Marshall 2012). One Mediterranean fruit fly outbreak in the United States was estimated to cost upwards to $58 million (Buck 2017). In general, the species of economic and medical importance have received more attention than the rest of the order (Novakova et al. 2017; Wang et al. 2011; De Cock et al. 2020; Deguenon et al. 2019; Tomberlin et al. 2017). Although many species of synanthropic and economically important Diptera have had their microbial communities extensively studied, very few studies have characterized the microbiomes of Diptera with no obvious health or agricultural importance.

Bombyliidae or bee flies (Figure 1.1) are a large family of ~5,000 species. These flies have a cosmopolitan distribution, with a large diversity found in arid and semi-arid regions (Li et al. 2021; Hull 1973). Adult bee flies can typically be found pollinating flowers. Females often need specific nutrients from the pollen to develop eggs (Hull 1973; Yeates & Greathead 1997). Bee flies are particularly important as pollinators of many desert wildflowers in Africa, including the unusual Namibian gymnosperm *Welwitschia*. Studies have shown that bee flies are the keystone pollinators of flowers in the Greater Cape Floristic Region (de Jager & Ellis 2017). Larval bee flies are parasitoids of other arthropods, attacking a wide variety of hosts, such as grasshoppers, wasps, and even the Tsetse fly (Yeates and Greathead 1997).

Despite their importance as pollinators and potential use as biological control agents, many groups within bee flies are taxonomically and ecologically poorly known. Research on their natural and evolutionary history is greatly needed to advance our meager understanding of this group. Studying the microbial ecology of bee flies could provide unique insights into the microbial communities of a non-model parasitoid insect. In this study, we aim to understand what type of microbiome bee flies have, and what is the composition of the microbial community.

Methods

Sampling

Bee flies were collected at various localities throughout southern California (Figure 1. 2) and stored in 100% ethanol. Collected specimens were rinsed and stored at -80C with fresh

100% ethanol, this was done to prevent contaminants. Specimens were then separated into head, thorax, and abdomen using sterilized forceps, and placed in the same tube; this was done to allow the lysate to penetrate soft tissue throughout the dissected flies. Using each tagma of the dissected fly, gDNA was extracted from the 34 individuals representing seven subfamilies and 18 genera (Table 1.1). Extractions followed the manufacturer's protocol for the Qiagen DNeasy Blood & Tissue Kit. Following the methods of Silver et al. (2021), lysate from the overnight incubation was transferred to sterile 1.5ml O-ring tubes containing 0.25g (+/- 0.02g) of 0.1mm diameter zirconium beads and bead beat at 2000rpm for 3 minutes in a PowerLyzer to lyse bacterial cells.

DNA was extracted from the lysed cells using Solid Phase Reversible Immobilization (SPRI) magnetic beads made by following the method of Rohland (Rohland & Reich 2012). 100µL of lysate was mixed with 180µL of well-mixed, room-temperature SPRI beads, incubated for approximately 5 minutes then transferred to a magnetic rack. After the SPRI beads pelleted, 200µL of 80% ethanol was added. After 30 seconds the supernatant was removed. The ethanol wash was then repeated a second time and the supernatant was removed again. Then, tubes were removed from the magnetic rack and allowed to completely dry. DNA was eluted by adding 50µL TB solution (10mM Tris) directly onto the beads and incubating for 5 minutes. Tubes were then returned to the magnetic rack to pellet the SPRI beads and retrieve the DNA-containing supernatant.

The V4 region of the 16S rRNA gene was PCR amplified following protocol in Silver et al. 2021. For the first round of amplification, previously described primers (Caporaso et al. 2012) 515FB_in (5'-ACA CTCTTT CCC TAC ACG ACG CTC TTC CGA TCT GTG YCA GCM GCC GCG GTA A-3') and 806RB_in (5'-GTG ACT GGA GTT CAG ACG TGT GCT CTT CCG ATC TGG ACT ACH VGG GTW TCT AAT-3'), were adapted to be complementary to the second round primers (Lange et al. 2014), which were then added to the ends of all 16S genes with the following conditions (BioRad thermocycler): initial denaturation at 94℃ for 3 min, followed by 30 cycles of 94℃ for 45 sec, 50℃ for 1 min, 72℃ for 1:30 min, and a final extension step of 72℃ for 10 min. A second round of PCR was performed using unique combinations of barcoded forward (5'-AAT GAT ACG GCG ACC ACC GAG ATC TAC ACX XXX XXX XAC ACT CTT TCC CTA CAC GA-3') and reverse (5'-CAA GCA GAA GAC GGC ATA CGA GAT XXX XXX XXG TGA CTG GAG TTC AGA CGT G-3') primers (Lange et al. 2014) to create a dual-index amplicon library for Illumina sequencing. The conditions for the second PCR reaction were: initial denaturation at 94℃ for 3 min, followed by 10 cycles of 94℃ for 45 sec, 50℃ for 1 min, 72℃ for 1:30 min, and a final extension step of 72℃ for 10 min. All pooled duplicate PCR products were run on a 1% agarose gel for 30 min at 100V and imaged under UV light to verify successful PCR. DNA concentration was quantified using a Qubit fluorometer, and equimolar amounts were pooled. The pooled library was purified (Qiagen Qiaquick PCR Purification Kit) and sent for Illumina MiSeq sequencing at the UC Berkeley Genomics Sequencing Laboratory.

Analysis

Amplicon reads for the V4 region of 16S were de-multiplexed with deML (Renaud 2019) and processed using DADA2 (Callahan 2016), including quality filtering with maxEE = 2. Reads

were de-replicated into unique 16S amplicon sequence variants (ASVs). Paired-end reads were merged and mapped to ASVs to construct a sequence table, chimeric sequences were removed (Silver et al. 2021). Taxonomic assignments for exact matches of ASVs and reference strains were made using the Ribosomal Database Project database (Cole et al. 2014). Sequence tables and taxonomic assignments were imported into R version 3.5 (R Core Team 2018) for downstream analysis and combined into a single phyloseq files (McMurdi et al. 2013; Silver et al. 2021). To account for variation in sequencing effort across samples, samples were scaled according to variance stabilized ASV abundances using DESeq2 (Love et al. 2014; McMurdie et al. 2014; Silver et al. 2021). Alpha diversity measures were calculated using the packages phyloseq (McMurdie & Holmes 2013) and picante (Kembel et al. 2010). Non-metric multidimensional scaling (NMDS) plots of beta diversity were created using phyloseq (McMurdie & Holmes 2013). Bray-Curtis distances were calculated for the subfamily, parasitoid type, and habitat (Silver et al. 2021).

Results

Sequencing Results

Results from Illumina MiSeq sequencing yielded 435,000 high-quality reads that clustered into 1546 OTUs. After quality filtering, the mean number of reads per sample was 12092 and the median number of reads per sample was 11007.

Alpha Diversity

The mean number of unique ASVs per sample was 113.17 and the median number of unique ASVs per sample was 91. A rarefaction curve of the 36 bee fly samples indicates that our sequencing efforts detected most of the microbial diversity found and further sequencing efforts will not statistically uncover additional diversity. (Figure 1.3). The percentage of classified ASVs at each taxonomic level was calculated (Figure 1.4). While no species-level classifications were found; however, Order, Family and Genus were classified in 50-60% of all samples (Figure 1.4). The most abundant ASVs across samples were also categorized. The top 10 operational taxonomic units (OTUs) by sample are shown in (Figure 1.5). Some prevalent OTU families were Commandaceae (OTU 2), Chloroplast (OTU 3), Corynebacteriaceae (OTU 35), Enterobacteriaceae (OTU 60), Entoplasmataceae (OTU 16). When looking at the OTUs by total count (Figure 1.6), we see that OTU_1 (unidentified) and 3 (Chloroplast) had the highest abundance count. The relative abundance of microbial phyla by subfamily of bee fly is visualized (Figure 1.7). As seen across the subfamilies, proteobacteria and cyanobacteria compose most of the phyla. Composition of phyla varied between subfamilies, with proteobacteria and cyanobacteria being the most abundant. When examining microbial phyla by genus, we can see a similar pattern, diversity is dominated by proteobacteria and cyanobacteria (Figure 1.8). Other prominent phyla include the firmicutes, bacterioides, and tenericutes. ASV tables of bacterial species were generated at 50% and 75% appearance across all samples (Tables 1.2 & 1.3).

Beta Diversity

Bray-Curtis distances were calculated for the bee fly microbial community data (Fig 1.9), results indicate that community similarity was not associated by subfamily, parasitoid type or by habitat.

Discussion

This study is the first to characterize the microbiomes in bee flies. In addition, this is the first study to examine the microbial community of a non-hymenopteran parasitoid insect. Bee flies were found to have a diverse assemblage of microbes that have unique characteristics and functions discussed below.

Resident and Transient Microbial Communities

Evidence supports a microbiome that is composed of both resident and transient microbial communities. This can be seen in the bacterial ASVs that were found in 50 to 75% of all samples. Most of the microbes that make up an insect's microbiome tend to be transient in nature, changing over time, and being influenced by their environment and diet (Hammer et al. 2017; Provorov & Onishchuk 2017; Engel & Moran 2013). Some microbes can be considered resident microbes, which remain consistently in the insect's microbiome (Provorov & Onishchuk 2017; Hammer et al. 2017; Engel & Moran 2013). Ascertaining an organism's resident microbiome can be challenging, without rigorous experimentation. However, finding the same microbial species in 75% and 50% of all sample's hints at the possibility that these species represent some sort of core resident microbial community (Larsen et al. 2015; Martinson et al. 2019; Hammer et al. 2017).

Community Association

Results indicate that bee fly microbial community similarity was not influenced by subfamily, parasitoid type or by habitat. The parasitoid type of bee flies (endoparasitoid, ectoparasitoid, egg parasitoid) did not significantly cluster together. This indicates that a bee flies parasitoid type does not influence its microbial community. For example, one might expect all endoparasitoids to share very similar microbial communities. However, results could be biased due to small sample size and the amount of unknown host types- only ~10% of bee fly hosts are known (Kers and Saccenti 2022; Yeates & Greathead 1997). Microbial communities of bee flies did not associate by subfamily or by habitat type; like the parasitoid type these factors likely do not influence bee fly microbial similarity. Even though these three factors of interest did not cluster together, it does not mean that these are not influencing bee fly communities. Studies have shown that beta diversity metrics are sensitive to sample size, with Bray-Curtis metrics being the most sensitive (Kers and Saccenti 2022). With additional sampling, it may be possible to discern if these factors impact bee fly microbial communities.

Microbial Ecology

Predominant ASVs

The most prevalent ASV's found in all 36 bee fly samples was the Proteobacteria, this diverse class of bacteria has a wide range of functions and characteristics. Many Proteobacteria are known to help supplement the host's diet with essential metabolites, some have been found to be pathogenic, and even in some cases are influential in larval development (Tomberlin et al. 2017; Ahmad et al. 2006). For example, the proteobacteria *Acinetobacter*, found in soils has been found to affect the development of sarcophagid larvae and, when removed, the larvae failed to metamorphose properly (Tomberline et al. 2017). In addition, some *Acinetobacter* are known to have antiparasitic qualities and act as a secondary immune system, defending the insect host from potentially harmful microbes (Evans & Armstrong 2006; Douglas 2018). It is unclear exactly what role these proteobacteria are playing in the bee fly microbiome, but it is likely a mixture of some or all the above.

The second most abundant ASV was cyanobacteria. Although typically associated with aquatic environments, cyanobacteria can be found in virtually all environments where some moisture persists (Cohen & Gurevitz 2006). The large amounts of cyanobacteria could have been acquired through pollen and nectar feeding, or perhaps from some nearby source of water (Gawande et al. 2019; Zheng et al. 2021).

Noteworthy ASVs

Besides these large assemblages of their microbial diversity, other ASVs were also discovered in the bee fly microbial community. In large abundance were bacteria in the genus *Staphylococcus*, which were found in over 50% of all samples. A majority of *Staphylococcus* are facultative anaerobic bacteria that are harmless to insects (Douglas 2018). Studies have also found *Staphylococcus* to be commonly found in nectar and in the gut of insects that feed on nectar (Anderson et al. 2013). It is likely that the *Staphylococcus* found in the bee fly samples were a mixture of microbes associated with nectar, and harmless microbes commonly found in the environment (Anderson et al. 2013; Douglas 2018).

Of particular interest was the bacteria *Mesoplasma*, which was found in 75% of all samples. *Mesoplasma* bacteria are known to have a variety of functions like assisting in nutrient uptake, catabolism, or acting as a pathogen in insects. *Mesoplasma* is widespread and has been found in a variety of insects and some arachnids (Seemüller et al. 2002; De Oliveria et al. 2016). Studies have found *Mesoplasma* in army ants and leaf-cutting may be related to the processing of chitin (Sapountzis et al. 2015; De Oliveria et al. 2016). It is possible that the *Mesoplasma* found in bee flies plays a role in the processing of chitin. Digestion of chitin is unlikely to be needed as an adult flower-feeding insect, however, this could be a vertically transferred remnant of the larval microbiome. Most bee fly larvae are ectoparasitoids, which requires eating through the insect cuticle to feed on the internal soft tissue. Having a bacterial species associated with processing chitin would aid them in breaking through the host's cuticle and processing that chitin for nutrients. Given that these ASVs were found across 75% of all samples, it is possible that *Mesoplasma* is not as transient as other bacteria and perhaps evidence of a resident microbiome.

Large amounts of Enterobacteriaceae were also found in bee flies. Bacterial species from this genus are known to be important symbionts in many other insects (Douglas 2018). Although the identity of these potential symbionts in bee flies is unknown, further study may discover important symbionts in the bee fly microbiome.

Conclusion

Bee flies have a diverse microbiota composed of both transient and resident bacteria. Their microbiota aligns with the life history of bee flies, bacteria associated with pollen and nectar feeding were found in addition to bacteria likely associated with the processing of chitin. A baseline of the bee fly microbial community has been established. Providing more insight into the biology and natural history of these desert-adapted insects. Further study in the form of shotgun sequencing will be required to truly understand the structure and function of bee fly microbiomes.

Subfamily	Genus	Species	State	County	Locality	Lat	Long
Anthracinae	*Anthrax*		California	San Diego	Hollenbeck Canyon Wildlife Area	32.6707	-116.8238
Anthracinae	*Anthrax*		California	San Diego	Carizo Creek	32.8458	-116.2032
Anthracinae	*Anthrax*		California	San Diego	cool canyon	33.0475	-116.4335
Anthracinae	*Anthrax*		California	San Diego	Warner Springs, Pacific Crest Trail	33.2888	-116.6552
Anthracinae	*Aphoebantus*		California	San Diego	Carizo Creek	32.8458	-116.2032
Anthracinae	*Aphoebantus*		California	San Diego	cool canyon	33.0475	-116.4335
Anthracinae	*Aphoebantus*		California	San Diego	Hollenbeck Canyon Wildlife Area	33.6707	-116.8238
Anthracinae	*Aphoebantus*		California	Riverside	Wash off Pinesmoke Road	33.5975	-116.4798
Usiinae	*Apolysis*		California	San Diego	Lawson peak	32.71522	-116.71043
Usiinae	*Apolysis*		California	San Diego	Lizard Wash, Anza Borrego	33.1279	-116.3918
Bombyliinae	*Bombylius*	*albicapillus*	California	San Diego	Wildwood Glen	32.8418	-116.64569
Bombyliinae	*Bombylius*		California	San Diego	Hollenbeck Canyon Wildlife Area	33.6707	-116.8238
Bombyliinae	*Bombylius*		California	San Diego	Kitchen Creek	32.7807	-116.4478
Bombyliinae	*Conophorus*		California	San Diego	SE of El Monte Park	32.8848	-116.822
Bombyliinae	*Conophorus*		California	San Diego	Hollenbeck Canyon Wildlife Area	33.6707	-116.8238
Bombyliinae	*Conophorus*		California	San Diego	Kitchen Creek	32.7807	-116.4478
Anthracinae	*Exoprosopa*		California	San Diego	Black Mtn. Peak Trail	32.98122	-117.11577
Anthracinae	*Hemipenthes*		California	San Diego	cool canyon	33.0475	-116.4335
Anthracinae	*Lepidanthrax*		California	San Diego	Blair Valley, Smugglers/Pictograph Trail	33.03	-116.3969
Anthracinae	*Lepidanthrax*		California	San Diego	Warner Springs, Pacific Crest Trail	33.2888	-116.6552
Lordotinae	*Lordotus*		California	San Diego	Lizard Wash, Anza Borrego	33.1279	-116.3918
Lordotinae	*Lordotus*		California	San Diego	Mortero Wash	32.7902	-116.1094
Lordotinae	*Lordotus*		California	San Diego	cool canyon	33.0475	-116.4335
Lordotinae	*Lordotus*		California	San Diego	Crest Ridge Ecological Reserve	32.8159	-116.8748
Pthiriinae	*Neacreotrichus*		California	San Diego	Crest Ridge Ecological Reserve	32.8159	-116.8748
Cythereinae	*Pantarbes*		California	San Diego	cool canyon	33.0475	-116.4335
Anthracinae	*Paravilla*		California	San Diego	Carizo Creek	32.8458	-116.2032
Ecliminae	*Thevenetimyia*	*tridentata*	California	San Diego	Hollenbeck Canyon Wildlife Area	32.6707	-116.8238
Anthracinae	*Thyridanthrax*		California	San Diego	cool canyon	33.0475	-116.4335
Toxophorinae	*Toxophora*		California	Riverside	Sawmill Trailhead, Cactus Spring Trail Rd.	33.5799	-116.4474
Anthracinae	*Villa*		California	San Diego	Sweeney Pass road	33.8326	-116.1835
Anthracinae	*Villa*		California	San Diego	cool canyon	33.0475	-116.4335
Anthracinae	*Villa*		California	San Diego	Warner Springs, Pacific Crest Trail	33.2888	-116.6552
Anthracinae	*Villa*		California	Riverside	Sawmill Trailhead, Cactus Spring Trail Rd.	33.5799	-116.4474

Table 1.1 : Sample information for bee flies collected.

Order	Family	Genus
Actinomycetales	Corynebacteriaceae	Corynebacterium
Actinomycetales	Dietziaceae	Dietzia
Bacillales	Bacillales_incertae_sedis	Caldalkalibacillus
Lactobacillales	Streptococcaceae	Streptococcus
Bacillales	Staphylococcaceae	Staphylococcus
Bacillales	Bacillaceae_1	Bacillus
Entomoplasmatales	Entomoplasmataceae	Mesoplasma
Chloroplast	Streptophyta	NA
Sphingomonadales	Sphingomonadaceae	NA
Enterobacteriales	Enterobacteriaceae	Serratia
Enterobacteriales	Enterobacteriaceae	NA
Pseudomonadales	Moraxellaceae	Acinetobacter
Burkholderiales	Oxalobacteraceae	Massilia
Burkholderiales	Comamonadaceae	NA
Oceanospirillales	Halomonadaceae	Halomonas
Pseudomonadales	Pseudomonadaceae	Pseudomonas

Table 1.2: ASV table of microbes found in 50% of all samples.

Order	Family	Genus
Actinomycetales	Corynebacteriaceae	Corynebacterium
Bacillales	Bacillales_incertae_sedis	Caldalkalibacillus
Bacillales	Staphylococcaceae	Staphylococcus
Bacillales	Bacillaceae_1	Bacillus
Entomoplasmatales	Entomoplasmataceae	Mesoplasma
Chloroplast	Streptophyta	NA
Enterobacteriales	Enterobacteriaceae	NA
Burkholderiales	Comamonadaceae	NA
Oceanospirillales	Halomonadaceae	Halomonas

Table 1.3: ASV table of microbes found in 75% of all samples.

Figure 1.1: The diversity of Bee flies. Photos on the left by Joyce Gross, photos on the right by Alice Abela

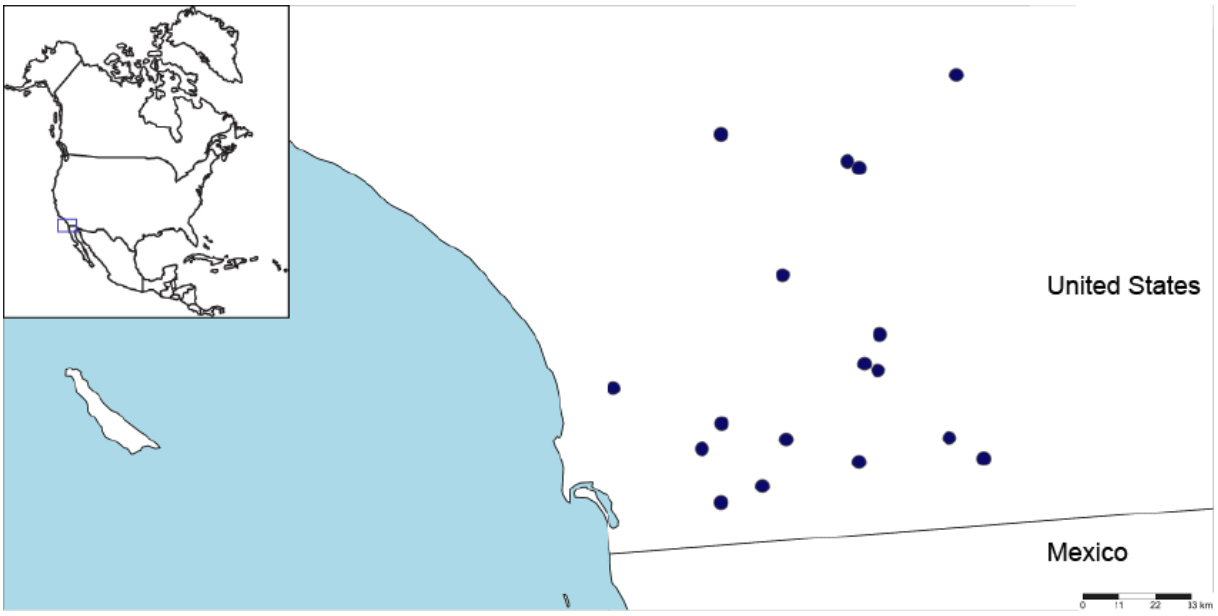


Figure 1.2: Map showing locations of sampled bee flies.

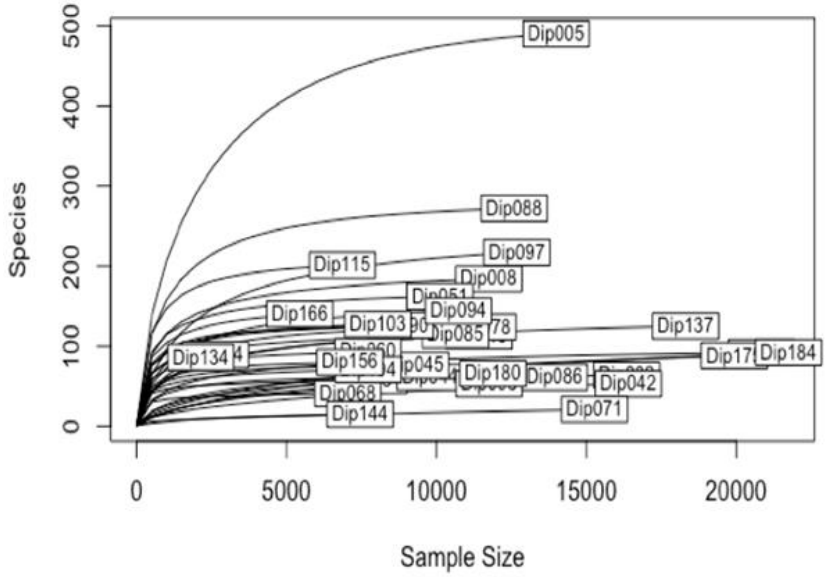


Figure 1.3: Rarefaction Curve of sampled bee fly microbiomes.

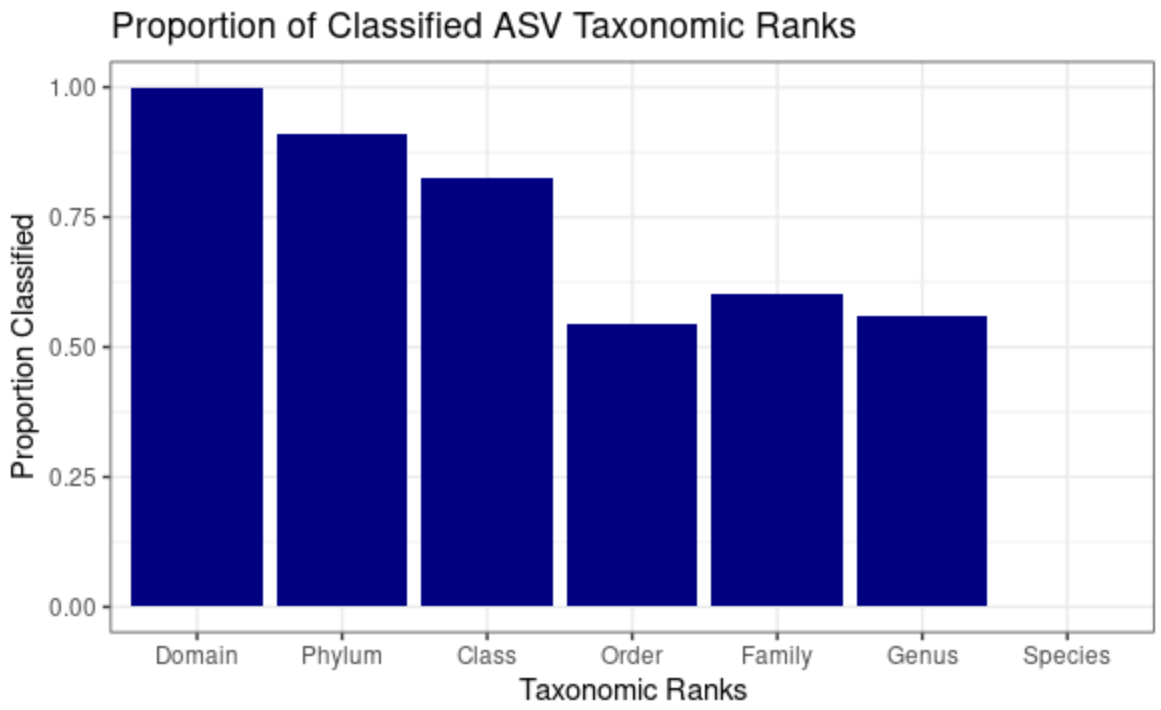


Figure 1.4: The proportion of sample ASVs and which taxonomic ranks were identified.

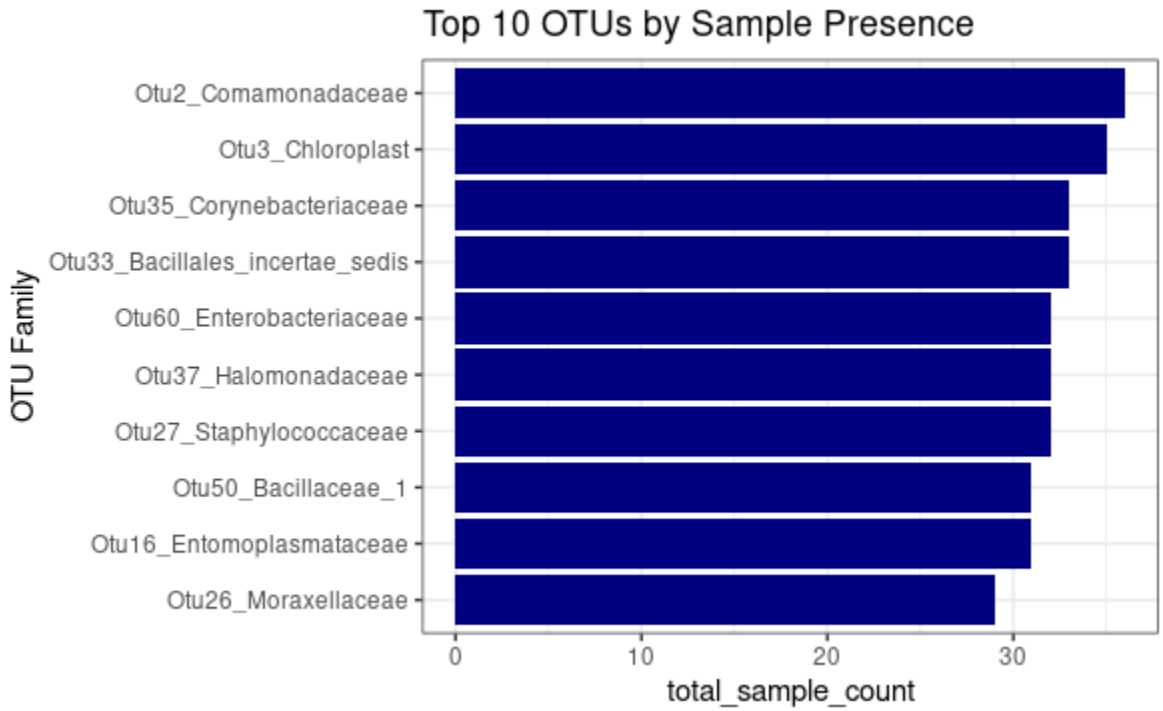


Figure 1.5: The top 10 OTUs found in samples.

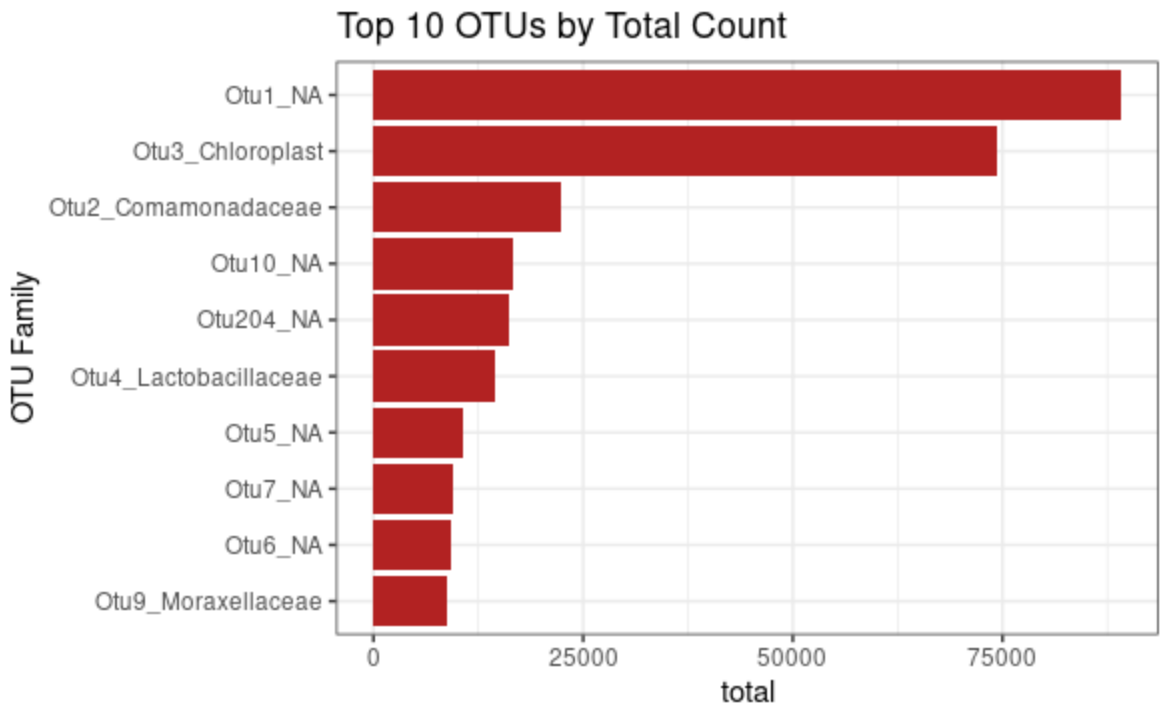


Figure 1.6: The top ten OTUs found in samples by total count.

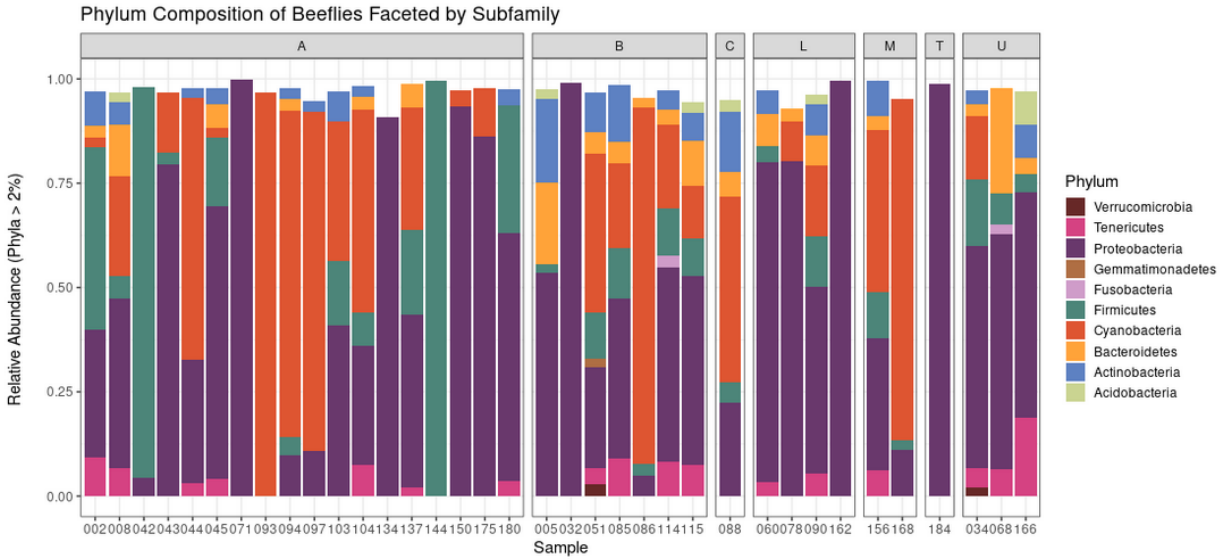


Figure 1.7: Phylum Composition of Bee flies by Subfamily: Categories correspond with subfamily, A: Anthracinae, B: Bombyliinae, C: Cytherinae, L: Lordotinae, M: Mythicomyiinae, T: Toxophorinae, U: Usiinae

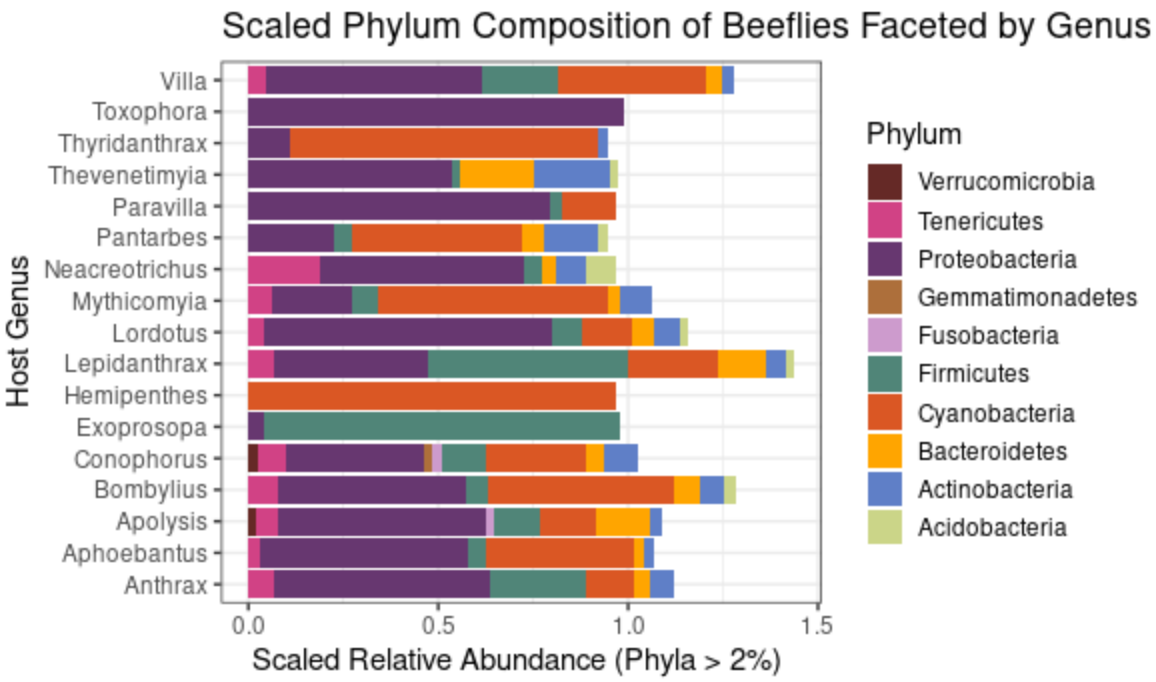


Figure 1.8: Microbial phylum composition of bee flies by genus.

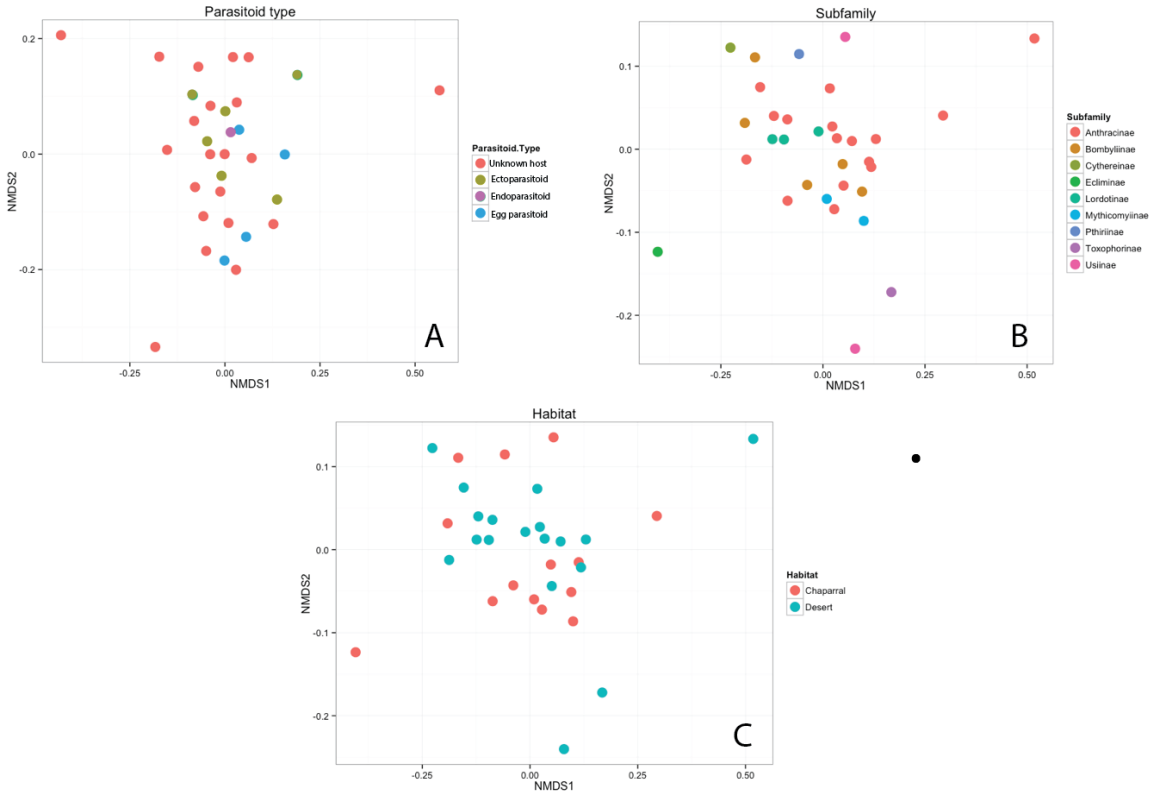


Figure 1.9. Bray-Curtis ordination of microbiome beta diversity using non-metric dimensional scaling. Microbiome data did not cluster by parasitoid type (A), by subfamily (B), or by habitat (C).

Chapter 2

Phylogenomics of the bee fly genus *Lordotus* Loew, 1863 (Diptera: Bombyliidae)

Abstract

Bombyliidae or bee flies are one of the most diverse groups of insects found in deserts across the world. Despite a diversity of over 5,000 species, most are poorly studied and in need of taxonomic and phylogenetic study. The genus *Lordotus* described in 1863 by Hermann Loew, exemplifies this problem. *Lordotus* is commonly found throughout the arid and semi-arid regions in the western United States and northern Mexico, these 29 species of *Lordotus* are classified into three species groups. However, due to the morphological variability in species, *Lordotus* species have been challenging to group or to establish distinct species delimitations. Beyond what might be implied by the species-group arrangement, the phylogenetic relationships within the genus have not been explored. Here, the first phylogenetic hypothesis of the bee fly genus *Lordotus*, generated using ultra-conserved elements (UCEs), is presented. A UCE dataset consisting of 936 loci from 76 individuals representing all species of the genus was generated using a combination of flies preserved in ethanol and historic, pinned museum specimens. Maximum likelihood analysis generated a well-resolved phylogeny and a monophyletic genus *Lordotus*. The three species groups suggested by previous authors were all found to be paraphyletic. These results provide a much-needed foundation for further detailed revision of *Lordotus*.

Introduction

Historically, authors have relied on morphology to delimit species, typically using a variety of diagnostic somatic and genital characters (Wiens 2007). However, the exclusive use of morphology has its shortcomings as it can lead to an underestimation of species diversity if the putative species are morphologically similar or "cryptic species" (Benda et al. 2021; Li and Wiens 2022). Morphology can also lead to an overestimate of species diversity if the species is morphologically phenotypically variable, but genetically similar (Gruber et al. 2013). Advances in molecular sequencing methods have allowed authors to address phylogenetic relationships in morphologically challenging groups. These methods generate large amounts of genomic data that have allowed researchers to resolve both ancient and recent relationships of many diverse and species-rich groups (Homziak et al. 2019; Crawford et al. 2012; Blaimer et al. 2015).

One such method is the use of ultraconserved elements (UCEs). UCEs are highly conserved genomic regions found across all of life (Faircloth et al 2014; Bejerano et al. 2004; Smith et al. 2014). The conserved sequences of UCEs are flanked by variable regions. These flanking regions are useful for estimating phylogenetic relationships, including at the species level (McCormack et al 2016; Newman and Austin 2016; Blaimer et al, 2016 a,b). UCEs allow researchers to capture hundreds of loci, significantly more data than traditional methods such as PCR and Sanger sequencing (Faircloth et al. 2012; Bossert and Danforth 2018). A significant benefit of UCEs is the ability to generate large amounts of sequence data from highly fragmented DNA, such as is found in preserved museum specimens (Buenaventura 2021; McCormack et al 2016; Blaimer et al. 2016a; Derkarabetian et al. 2019). DNA sequence data has been recovered from a variety of historic museum samples including samples from birds, reptiles, and pinned insects (Buenaventura 2021; McCormack et al. 2016; Blaimer et al. 2016a). The use of museum

samples for UCE studies has been dubbed ‘Museomics' and has opened the possibility of gathering data from historically collected samples (Raxworthy et al. 2021). Studies have shown it is possible to harvest UCE loci from samples collected as far back as the 1800s (Raxworthy et al. 2021; Derkarabetian et al. 2019; Buenaventura 2021). This new data source provides enormous benefits to studies, allowing researchers to fill in gaps by including samples of species not readily available through field collection.

Diptera (Flies) are one of the most diverse orders of insects on the planet and play a key role in almost every terrestrial and aquatic ecosystem (Marshall 2012; Marshall and Kirk-Spriggs 2017). Unfortunately, many groups of flies are taxonomically and ecologically poorly known, and research focusing on their natural and evolutionary history is needed. Bombyliidae, or bee flies are a large fly family of ~5,000 species, these flies have a cosmopolitan distribution, with a significant portion of their diversity found in arid and semi-arid regions (Li et al. 2021; Hull 1973). Larvae of bee flies are parasitoids of a wide variety of hosts and have been shown to attack spider eggs, beetles, other flies, butterflies, and wasps (Yeates and Greathead 1997). Of the known host records, the larvae of most species have been found to be ectoparasitoids, while some are endoparasitoids, or hyperparasitoids (Yeates and Greathead 1997). Due to challenges in rearing bee fly larvae, host records are known for only ~10% of all described species (Yeates and Greathead 1997). Adult bee flies can typically be found pollinating a wide variety of flowers, with females searching for pollen that provides nutrients needed to develop their eggs. Although many bee flies are obligate pollinators of some flowers, the pollinator relationships of this group are almost entirely unknown (Larson et al. 2001). However, despite realizing that there is a large diversity of species and life histories, many of the genera remain poorly studied and few researchers have attempted to resolve the phylogenetic relationships of these diverse flies.

The bee fly genus *Lordotus*, which was first described by Hermann Loew in 1863, includes species that are known for their bright yellow or silver pile (Hull 1973, Figure 1). These flies range from 4 to 16 mm and some species are known for being quite robust hairy flies while others are small and slender, with a long proboscis they use for feeding on nectar and pollen (Hull 1973). *Lordotus* is restricted to the western United States and Mexico, with the most species diversity in the Southwest (Hull 1973). All species of *Lordotus* have been recorded in California (Hall and Evenhuis 1982). The first revision of the group was published by Jack C. Hall (1954), in which he redescribed known species and described three new species. *Lordotus* species were distinguished by the color of the pile and tomentum and the color of the sternite and leg cuticle (Hall 1954). In addition, the proportions of the antennal segments were thought to be valuable in distinguishing species, however, these were reported as averages of all specimens observed (Hall 1954). Delimiting a species based on the average of a highly variable trait does not lead to consistent identification and can be heavily influenced by the individual who measures the trait. Hall noted that all these characteristics were highly variable across *Lordotus*, even supposed male-only characteristics were also occasionally found on female flies. Hall (1954) also found the genitalia of males and females to not be useful for delimiting the species of *Lordotus*.

The most recent revision of this group was published by Johnson and Johnson (1959). These authors separated the 29 species of *Lordotus* into three species groups (Table 2.1). Given the large amount of morphological variability found across the genus, the authors attempted to organize the species into these three groups using characteristics of the antennal segments and

wing vestiture (Johnson and Johnson 1959). The groups were distinguished by the presence of the antennal style (*miscellus* group), smooth wing costa of the male (*apicula* group) and the absence of the antennal style and denticulate costa of the male (*gibbus* group) (Table 2.2) (Hall and Evenhuis 1982; Johnson and Johnson 1959). The *apicula* group contains 14 species. Of the *apicula* group species, there appears to be a variety of forms that are best considered as members of a species complex around *L. apicula* and *L.sororculus* (Hall and Evenhuis 1982; Johnson and Johnson 1959). It is obvious that the morphological variation has made the species in *Lordotus* a challenge to delimit. In this study, UCE data generated from newly collected and historical museum specimens are used to generate the first phylogenetic hypothesis for the bee fly genus *Lordotus*.

Methods

Taxon Sampling

Taxon sampling included 77 specimens of *Lordotus*, representing all species in the genus. Sampling included live collected (N=28), then EtOH preserved specimens and pinned, museum specimens (N=49) (Table 2.3). The sampling includes specimens from throughout the known range of the genus in the US and Mexico. One outgroup species, *Bombylius major*, which represents a distantly related species in Bombyliidae (Cohen et al. 2021), was included to allow for rooting of the phylogeny.

Specimens of *Lordotus* that were collected throughout Southern California and Arizona were preserved in 100% EtOH and stored in a -20C freezer until DNA extraction (Figure 2.2). Pinned specimens were borrowed from the Essig Museum of Entomology, UC Berkeley (EMEC), UC Riverside Entomology Research Museum (UCRC ENT), Gillette Museum of Arthropod Diversity at Colorado State University (CSUC), California Academy of Sciences (CAS), and the San Diego Natural History Museum (SDNHM) (Table 2.3).

DNA extraction and sonication

Genomic DNA was extracted from EtOH-preserved specimens using the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA) using the manufacturer's protocol. The head, thorax, abdomen, and legs were all separated to allow the lysis buffer to access the soft tissue without additionally damaging the external cuticle, thus preserving the cuticle for future morphological study. DNA was extracted non-destructively from pinned museum specimens using the QIAamp DNA Micro Kit (Qiagen, Valencia, CA, USA). Pinned specimens were first cleaned with a fine brush, removing dust, pollen, and other debris that had accumulated. Specimens were carefully removed from the pin then the head, thorax, and abdomen were separated and placed in the lysis buffer. To maximize DNA yield, both EtOH preserved and museum specimens were left overnight in the Proteinase K digestion. Extracted genomic DNA was quantified on a Qubit fluorometer using the high-sensitivity kit. Genomic DNA was sheared using qSonica sonicator, EtOH preserved specimen extracts were sonicated for 2 mins and museum specimen extracts were sonicated for 1 min. Post-sonication double-sided, size-selection was performed using a 0.45x/0.65x ratio to achieve an average fragment size of 400–600bp (Valderrama et al. 2022; Cohen et al. 2021).

Library preparation, target enrichment and sequencing

Libraries were prepared using a modified version of the KAPA Hyper prep kit (see appendix A) using the modified protocol of Branstetter et al. 2017 for use with the iTru dual-indexing adapter system (Glenn et al., 2019, Faircloth et al. 2015). A stub adapter with amplify on-unique dual indexes of 8bp was done using 15 cycles of PCR. Volumes of reactions were reduced to 1/5th of the manufacturer's protocol, which helped reduce cost of the library preparation without affecting effectiveness of the library preparation (Valderrama et al. 2022; Cohen et al. 2021).

Library enrichment was performed following the myBaits Hybridization Capture for Targeted NGS protocol provided by Arbor Biosciences (https://arborbiosci.com/wp-content/uploads/2018/04/myBaits-Manual-v4.pdf). A custom-developed Asiloid probe set was used during enrichment, these were designed to be more specific than the general Diptera probe set available through Arbo Biosciences (Faircloth et al. 2014). After hybridization, the DNA concentration of enriched libraries was determined using Qubit to verify that hybridization was fulfilled. Enriched libraries were sequenced using Novaseq PE 150 (165G) at Novagene.

Processing of UCE Data

Raw reads were processed using the PHYLUCE pipeline (Faircloth, 2016). Adapter sequences were trimmed from raw sequencing data using Illumiprocessor, (Faircloth, 2013). Following adapter trimming, remaining undesired non-target sequences such as Illumina primers or artificial Poly G tailing, were identified using sequence summary reports for each sample using FastQC (Andrews, 2010). This Poly G tailing issue is common in Illumina machines that rely on a two-color chemistry system (Chen et al., 2018). Summary reports were generated for base sequence quality, sequence quality scores, base N content, and an indication of overrepresented sequences. Single FASTA files were generated that contained the mentioned undesired sequences that matched Illumina primers, and the sequences that contained more than ten guanine repeats. These contaminant sequences were removed using *fastp* (Chen et al., 2018). Using metaSPAdes assembler, the cleaned reads were assembled *de novo*, default parameters were used (Nurk et al. 2017).

UCE loci were identified from resultant contigs using the Asiloid UCE sequence probes and the *match_contigs_to_probes* function within PHYLUCE. Recovered UCE loci were aligned using MAFFT (Katoh and Standley, 2013) and were trimmed internally using TrimAI (Capella-Gutiérrez et al., 2009) with a gap-threshold value of 0.2 as recommended by Portik and Wiens (2021). To minimize the effect of missing data, two samples (*L. hurdi_11* and *L. MX_sp4*) that rendered less than 300 loci were not considered for downstream analyses. The final data matrix for this study was generated using alignments with ≥ 75% locus occupancy attained using the function *phyluce_align_get_only_loci_with_min_taxa*. Finally, resultant alignments were concatenated using the *phyluce_align_concantenate_alignments* function.

Phylogenetic Estimation

Maximum likelihood (ML) analyses utilizing the concatenated 75% locus occupancy matrix was performed in IQ-TREE 2 (Minh et al., 2020) with ultrafast bootstrap approximation and best-fit models of evolution were identified using ModelFinder (Kalyaanamoorthy et al., 2017).

Results

Sequencing and UCE Capture Results

Extractions of fresh and historic samples were successful and yielded sufficient DNA required for enrichment and sequencing (Table 2.4). 78 samples were enriched using the UCE probes and sequenced on the Novaseq PE 150. Samples derived from both museum and fresh specimens were successfully enriched.

Library sequencing produced an average of 12360285 raw paired-end reads per sample. MetaSPAdes assembled reads into 4145-302091 contigs with an average of 38297 contigs per sample. From the total assembled contigs, a total of 67-1277 UCE loci out of 2384 UCE targets were recovered across all samples with an average of 936 UCE loci per sample.

Phylogenetic Results

Phylogenetic relationships for *Lordotus* were inferred from the 75% matrix using an ML approach in IQ-TREE (Minh et al., 2020)(Figure 2.3). IQ-Tree returned a topology that does not support any of the previously hypothesized species groups. Within the putative *apicula* group, L. *apicula*, *L. hurdi*, and *L. sororculus* were not recovered as monophyletic. *Lordotus schlingeri*, *L. junceus* and *L. divisus* were not recovered in the *apicula* group. Unidentified specimens from Mexico formed a monophyletic group, potentially representing a new species.

Discussion

This study is the first to examine the species relationships within *Lordotus* using phylogenetics. The phylogenetic hypothesis was generated using freshly collected specimens, in addition to the inclusion of historic, pinned specimens, representing unique populations, and rarely collected species. Previous studies used morphology to delimit species and proposed species groups, however, these groups had not been previously tested in a phylogenetic framework (Table 2.1) (Hall and Evenhuis 1982; Johnson and Johnson 1959). Results relative to each of these groups are discussed below:

apicula group

The *apicula* group suggested by Johnson and Johnson (1959), was not supported, and found to be paraphyletic (Figure 2). The *apicula* group was distinguished from other *Lordotus* groups by the presence of a smooth wing costa in males (Hall and Evenhuis 1982; Johnson and Johnson 1959). The *apicula* group contains 14 species, with a further complex of species found around *apicula* and *soroculus* (Hall and Evenhuis 1982; Johnson and Johnson 1959). Of the 14 described species in this putative group, *L. junceus*, *L. divisus* and *L. schlingeri* were not recovered as members and found to be paraphyletic to the *apicular group*. These species are

more closely related to the *gibbus* and *miscellus* group species. The two specimens of *L. schlingeri* included in the study were paratypes and were found to be unrelated to each other. It is possible that one or both of the samples was misidentified or they are actually different species. Close examination of the type and other paratypes will be needed to determine the identity of these samples.

Lordotus ermae was found to be sister to the clade consisting of (*L. bipartitus (L. lutescens* + *L. perplexus*). These species were separated from each other based on their body size and black pile found on the occipital and basal antennal segments (Hall and Evenhuis 1982). A clade consisting of *L. planus* and *L. puella* was found to be sister to the *L. apicula* complex. *Lordotus puella* and *L. ermae* are very similar morphologically and some authors suggested that upon closer examination they would need to be synonymized (Hall and Evenhuis 1982), however, the phylogenetic results suggest that they are actually only distantly related.

The tree places *L. apicula*, *L. arnaudi*, *L. sororculus*, *L. abdominalis* and *L. hurdi* in a species complex, this is consistent with previous hypotheses (Hall and Evenhuis 1982). The species in this group are often described as "nearly inseparable" from each other, often delimited by the presence or abundance of black or white pile on the abdomen and the color of their legs (Hall and Evenhuis 1982). However, these traits were found to be extremely variable throughout the range of the species. Most putative species in the complex were not found to be monophyletic. Only *L. abdominalis* was recovered as monophyletic. Further sampling of all species throughout their ranges will likely be needed to confidently delimit any species in the *apicula* complex. It is possible that this species complex represents a rapid radiation. Studies have found that relatively short branches of cryptic species may be an indication of a rapid radiation (Ješovnik et al. 2017; Longo et al. 2017; Whitfield and Kjer 2008; Barrera et al. 2022). However further testing of diversification rates and phylogenetic dating will be needed to make this determination. With the placement of *L. junceus*, *L. divisus* and *L. schilingeri* in the other *Lordotus* groups, we redefine the *apicula* group to consist of the remaining species.

miscellus and gibbus groups

The *miscellus* and *gibbus* groups were distinguished from the *apicula* group by the presence of denticulate costa and were thought to be distinguished from each other by the absence or presence of a terminal antennal style (Hall and Evenhuis 1982). However, the terminal antennal style is a variable trait, and some species in the *gibbus* group have an antenna style present. *Lordotus junceus*, *L. divisus* and *L. schilingeri* were found to be more closely related to species in the *miscellus* and *gibbus* groups, instead of the *apicula* group, where they have been historically placed. *Lordotus diversus* was closely related to a paraphyletic *L. divisus*. Hall and Evenhuis (1982). The authors note that *L. divisus* is a rarely collected species and that the male genitalia are similar to *L. diversus,* having slight differences like a larger process on the epandrium that is more heavily sclerotized (Hall and Evenhuis 1982). However, given the paraphyly of *L. divisus,* further examination of specimens and sequencing will be needed to delimit these species. The remaining *miscellus* group species were found to be paraphyletic, with *L. rufotibialis* sister to a paraphyletic *L. lineatus*. *Lordotus cingulatus* is also shown to be closely related to a paraphyletic *L. miscellus. Lordotus diplasus* was found to be sister to a paraphyletic *L. miscellus*. The species in the *miscellus* group may represent another species complex within

Lordotus. *Lordotus miscellus* and *L. cingulatus* have were described by Hall and Evenhuis (1982) as being nearly identical, distinguished by the highly variable color of their dark brown to black pile (Hall and Evenhuis 1982). *Lordotus rufotibialis* males are described as being similar in appearance to *L. cingulatus* and *L. miscellus*, only reliably distinguished from these species by close association with the females, which are separated from the other species by the reddish yellow tibia (Hall and Evenhuis 1982). *Lordotus diplasus* was originally described as a subspecies of *L. diversus*, and are separated by the proportions of antennal segments, and black pile on the fourth abdominal tergite (Hall and Evenhuis 1982). Both species also have a sympatric distribution (Hall and Evenhuis 1982). The morphological ambiguity between these species will require additional sequencing and examination of type material to delimit the boundaries within this species complex.

Lordotus nevadensis is strongly supported as the sister species to a clade consisting of *L. albidus*, *L. luteolus* and a potentially new species. These unidentified flies are from Baja California Sur, Mexico, collected around Bahia Tortugas, Guerrero and the Island of Cedros, which is off the coast of these cities. It is possible that these samples represent a new species, but further examination of the specimens is needed to make a final determination. *Lordotus zona* is a widespread species in the Southwest and is very similar in appearance to *L. pulchrissimus*, neither species was found to be monophyletic. It is possible that these species represent another complex that is in need of thorough sampling throughout their ranges to delimit these species. *Lordotus striatus* was originally described as a subspecies of *Lordotus gibbus*, however, it was later elevated to species rank. The placement of *L. striatus* as sister species to *L. gibbus* is supported here and so given the morphological similarities and the paraphyly of both species groups, we suggest that the two species groups should be considered one: the *miscelleus* group, which includes the former *gibbus* members and *L. junceus*, *L. divisus* and *L. schlingeri* from the *apicula* group.

Conclusion

The use of historic, pinned material and freshly collected specimens led to the first phylogenetic hypothesis for a diverse bee fly group. Despite the fly's extraordinary species diversity, conspicuousness, and ease of collecting, *Lordotus* species are morphologically very similar, indicating that these evidently cryptic species may be best thought of as being the result of a rapid radiation. Further sequencing and additional lines of evidence such as biogeography and morphometrics will aid in delimiting the species in this complex group. This study will serve as a foundation to guide future research that further tests these species boundaries. Studying this diverse genus may provide insight into diversification of desert-adapted insects.

apicula	***miscellus***	***gibbus***
abdominalis	*cingulatus*	*albidus*
apicula	*diplasus*	*arizonensis*
arnaudi	*diversus*	*bucerus*
bipartitus	*lineatus*	*gibbus*
divisus	*miscellus*	*luteolus*
ermae	*rufotibialus*	*nevadensis*
hurdi		*pulchrissimus*
junceus		*striatus*
lutescens		*zona*
perplexus		
planus		
puella		
schlingeri		
sororculus		

Table 2.1: Species of *Lordotus* organized by hypothesized species groups.

	Antennal Style	Costa
apicula	absent	smooth
miscellus	present	smooth
gibbus	absent	denticulate

Table 2.2: Character state of *Lordotus* species groups.

Genus	Species	Specimin Code	Collection	Collected	Country	State	County	Locality	Lat	Long	Preperation
Lordotus	gibbus	*CAS842 4175*	*CAS*	1980	USA	CA	Plumas	Beckworth Pass			Pinned
Lordotus	zona	*CAS8424162*	*CAS*	1970	USA	AZ	Pima	Route 85 26mi NW Sells			Pinned
Lordotus	*gibbus*	*CSUC_1*	*CSUC*	1998	USA	TX	Sterling Co	N Sterling HWY 163			Pinned
Lordotus	apicula	*CSUC_2*	*CSUC*	1992	USA	CO	Montezuma	E of Battle Rock			Pinned
Lordotus	miscellus	*CSUC_3*	*CSUC*	2004	USA	AZ	Maricopa	Coon Bluffs NF			Pinned
Lordotus	pulchrissimu	*CSUC_4*	*CSUC*	2001	USA	AZ	Yavapai	9mi W Congress			Pinned
Lordotus	pulchrissimu	*CSUC_5*	*CSUC*	2005	USA	AZ	Maricopa	Salome rd exit 81 off I-10			Pinned
Lordotus	apicula	*CSUC_6*	*CSUC*	1992	USA	CO	Montezuma	E of Battle Rock			Pinned
Lordotus	gibbus	*CSUC_7*	*CSUC*	2018	USA	CO	Ortero	Swink	N38 00.179	W103 37.71	Pinned
Lordotus	gibbus	*CSUC_8*	*CSUC*	2004	USA	WY	Fremont	A10-D3			Pinned
Lordotus	*junceus*	Dip140		2018	USA	CA	San Diego	Warner Sprir	33.289	-116.7	Ethanol
Lordotus	*planus_2*	Dip163		2018	USA	CA	San Diego	Crest Ridge E	32.816	-116.9	Ethanol
Lordotus	*zona_2*	Dip216		2018	USA	AZ	Pima	Cienega Cree	32.014	-110.6	Ethanol
Lordotus	*zona_3*	Dip217		2018	USA	AZ	Pima	Cienega Cree	32.014	-110.6	Ethanol
Lordotus	*zona_4*	Dip218		2018	USA	AZ	Pima	Cienega Cree	32.014	-110.6	Ethanol
Lordotus	abdominalis	Dip33		2016	USA	CA	San Diego	Meadowbroo	32.964	-117.1	Ethanol
Lordotus	*hurdi_3*	Dip368		2019	USA	CA	San Diego	Anza Borrego	33.131	-116.3	Ethanol
Lordotus	*apicula_6*	Dip37		2016	USA	CA	San Diego	Sycuan Peak,	32.754	-116.8	Ethanol
Lordotus	*hurdi_5*	Dip384		2019	USA	CA	Riverside	Road to Coyo	33.65	-116.4	Ethanol
Lordotus	*hurdi_6*	Dip385		2019	USA	CA	Riverside	Road to Coyo	33.65	-116.4	Ethanol
Lordotus	*hurdi_10*	Dip387		2019	USA	CA	Riverside	Road to Coyo	33.65	-116.4	Ethanol
Lordotus	*hurdi_7*	Dip389		2019	USA	CA	Riverside	Road to Coyo	33.65	-116.4	Ethanol
Lordotus	*hurdi_8*	Dip393		2019	USA	CA	Riverside	Road to Coyo	33.65	-116.4	Ethanol
Lordotus	*hurdi_9*	Dip401		2019	USA	CA	Riverside	Road to Coyo	33.65	-116.4	Ethanol
Lordotus	*sororculus_3*	Dip416		2019	USA	CA	Riverside	Wash N of D	33.944	-116.4	Ethanol
Lordotus	*apicula_3*	Dip441		2019	USA	CA	San Bernardi	~6mi S of I-4	34.781	-116.8	Ethanol
Lordotus	*apicula_8*	Dip443		2019	USA	CA	San Bernardi	~6mi S of I-4	34.781	-116.8	Ethanol
Lordotus	*apicula_4*	Dip444		2019	USA	CA	San Bernardi	~6mi S of I-4	34.781	-116.8	Ethanol
Lordotus	*apicula_5*	Dip469		2019	USA	CA	Imperial	Algodones D	33.061	-115.2	Ethanol
Lordotus	*apicula_7*	Dip49		2018	USA	CA	San Diego	Blair Valley, S	33.02	-116.3	Ethanol
Lordotus	*apicula_9*	Dip58		2018	USA	CA	San Diego	Lizard Wash,	33.128	-116.4	Ethanol
Lordotus	*apicula_10*	Dip61		2018	USA	CA	San Diego	Lizard Wash,	33.128	-116.4	Ethanol
Lordotus	*apicula_11*	Dip64		2018	USA	CA	San Diego	Lizard Wash,	33.128	-116.4	Ethanol
Lordotus	*apicula_12*	Dip65		2018	USA	CA	San Diego	Lizard Wash,	33.128	-116.4	Ethanol
Lordotus	*sororculus_4*	Dip73		2018	USA	CA	San Diego	Sweeney Pas	33.833	-116.2	Ethanol
Lordotus	*hurdi_4*	Dip77		2018	USA	CA	San Diego	Mortero Wa	32.79	-116.1	Ethanol
Lordotus	*hurdi*	Dip79		2018	USA	CA	San Diego	Mortero Wa	32.79	-116.1	Ethanol
Lordotus	*hurdi_2*	Dip91		2017	USA	CA	San Diego	cool canyon	33.048	-116.4	Ethanol
Lordotus	*striatus*	*Essig_1*	*EMEC*	1964	USA	CA	Los Angeles	Lancaster			Pinned
Lordotus	planus	*Essig_10*	*EMEC*	1958	USA	CA	San Luis Obis	Santa Margarita 6mi NE			Pinned
Lordotus	bipartitus	*Essig_11*	*EMEC*	1958	USA	AZ	Cochise	Apache 5mi NE			Pinned
Lordotus	striatus	*Essig_12*	*EMEC*	1957	USA	CA	Lassen	Janesville 1 mi N			Pinned
Lordotus	diversus	*Essig_13*	*EMEC*	1950	USA	CA	Los Angeles	Tanbark Flat			Pinned
Lordotus	ermae	*Essig_14*	*EMEC*	1958	USA	CA	San Luis Obis	Santa Margarita 6mi NE			Pinned
Lordotus	luteolus	*Essig_16*	*EMEC*	1974	USA	CA	Riverside	Palm Springs			Pinned
Lordotus	zona_2	*Essig_17*	*EMEC*	1948	USA	CA	Santa Clara	San Antonio Valley			Pinned
Lordotus	miscellus_2	*Essig_18*	*EMEC*	1957	USA	CA	Inyo	Olancha			Pinned
Lordotus	albidus	*Essig_19*	*EMEC*	1960	USA	CA	Kern	Boron			Pinned
Lordotus	*sp4*	*Essig_2*	*EMEC*	1964	USA	CA	San Bernardi	Lucerne Valley			Pinned
Lordotus	*sororculus_2*	*Essig_3*	*EMEC*	1963	USA	CA	Riverside	Joshua Tree Nat Mon			Pinned
Lordotus	divisus	*Essig_4*	*EMEC*	1956	USA	CA	Los Angeles	1mi W Little Rock			Pinned
Lordotus	bucerus	*Essig_5*	*EMEC*	1951	USA	CA	San Diego	Borego			Pinned
Lordotus	rufotibialis	*Essig_6*	*EMEC*	1958	USA	CA	Riverside	9mi W Blythe			Pinned
Lordotus	gibbus	*Essig_8*	*EMEC*	1956	USA	CA	Inyo	Westgard Pass			Pinned
Lordotus	sororculus	*Essig_9*	*EMEC*	1958	USA	CA	San Bernardi	Vidal 27mi N			Pinned
Lordotus	striatus_2	*SDNHM 1*	*SDNHM*	1982	MX	MX	Baja CA	6-9mi N of Ojos Negros			Pinned
Lordotus	zona_6	*SDNHM 2*	*SDNHM*	1981	MX	MX	Baja CA	.5mi S El Condor			Pinned
Lordotus	MX_sp4	*SDNHM 3*	*SDNHM*	1983	MX	MX	Baja CA	Isla De Cedros, canyon W of Punta Nort			Pinned
Lordotus	MX_sp1	*SDNHM 4*	*SDNHM*	1981	MX	MX	Baja Sur	9.2 mi SE Guerrero			Pinned
Lordotus	ermae	*SDNHM 5*	*SDNHM*	1980	MX	MX	Baja Norte	6mi N of Laguna Hanson, Sierra Juarez			Pinned
Lordotus	MX_sp2	*SDNHM 6*	*SDNHM*	1981	MX	MX	Baja Sur	40.4 mi. W Mexico 1 on Bahia Tortugas			Pinned
Lordotus	divisus_2	*SDNHM 7*	*SDNHM*	1981	MX	MX	Baja Sur	San Bortolo			Pinned
Lordotus	MX_sp3	*SDNHM 8*	*SDNHM*	1983	MX	MX	Baja Norte	Isla De Cedros, canyon W of Punta Nort			Pinned
Lordotus	pulchrissimu	*SDNHM 9*	*SDNHM*	1997	MX	MX	Baja Sur	San Francisco de la Sierra			Pinned
Lordotus	schlingeri_2	*UCR 294522*	*UCRCENT*	1963	USA	CA	Riverside	Deep Canyon			Pinned
Lordotus	abdominalis	*UCR 510372*	*UCRCENT*	1964	MX	MX	Baja CA	Arroyo Grande, 39mi SW La Ventana			Pinned
Lordotus	arizoenensis	*UCR 510395*	*UCRCENT*	1979	USA	AZ	Pinal	PicachoPass			Pinned
Lordotus	arnaudi	*UCR 510399*	*UCRCENT*	1964	MX	MX	Baja CA	3mi N San Felipe			Pinned
Lordotus	cingulatus	*UCR 510426*	*UCRCENT*	1979	USA	CA	San Bernardi	3mi Rice			Pinned
Lordotus	diplasus	*UCR 510437*	*UCRCENT*	1983	USA	AZ	Maricopa	Wittman			Pinned
Lordotus	hurdi_11	*UCR 510454*	*UCRCENT*	1964	USA	CA	San Diego	Borego			pinned
Lordotus	junceus	*UCR 510455*	*UCRCENT*	1964	MX	MX	Baja CA	3mi N San Felipe			Pinned
Lordotus	lineatus	*UCR 510463*	*UCRCENT*	1979	USA	CA	San Bernadir	2.3mi S needles CA			Pinned
Lordotus	lineatus_2	*UCR 510465*	*UCRCENT*	1975	USA	AZ	Maricopa	39mi E of Gila Bend			Pinned
Lordotus	lutecens	*UCR 510468*	*UCRCENT*	1971	USA	AZ	Cochise	10mi SW wilcox Playa			Pinned
Lordotus	nevadensis	*UCR 510478*	*UCRCENT*	1981	USA	CA	Inyo	Emigrant Canyon			Pinned
Lordotus	puella	*UCR 510485*	*UCRCENT*	1961	USA	CA	San Bernadir	Cajon Jct			Pinned
Lordotus	perplexus	*UCR 510486*	*UCRCENT*	1967	USA	AZ	Cochise	17mi W Douglas			Pinned
Lordotus	schlingeri	*UCR 510504*	*UCRCENT*	1963	USA	CA	Riverside	Deep Canyon			Pinned
Bombylius	major										Transcriptome

Table 2.3: Sample information on specimens used for UCE sequencing.

Genus	Species	Extraction ng/ul	Contigs Total	Contigs Total (bp)	Mean length (bp)	UCE loci Total	UCE loci (bp)	UCE Loci Mean length (bp)
Lordotus	gibbus	12.4	28119	8841335	314.4256553	1179	725544	615.389313
Lordotus	zona	15.6	12640	3850748	304.6477848	391	115668	295.826087
Lordotus	*gibbus*	30.1	14717	5185420	352.3421893	1142	811852	710.9036778
Lordotus	apicula	34.5	48717	15543556	319.0581522	1206	792467	657.1036484
Lordotus	miscellus	48.1	180480	34744190	192.509918	942	341151	362.156051
Lordotus	pulchrissimus	56.5	92058	31925579	346.7985292	1182	1107122	936.6514382
Lordotus	pulchrissimus	39.9	17147	5899620	344.0613518	1091	740455	678.6938588
Lordotus	apicula	23.1	10939	3914441	357.842673	1177	777309	660.415463
Lordotus	gibbus	83.5	302091	47651200	157.7379002	782	593826	759.3682864
Lordotus	gibbus	45.5	48379	14976312	309.5622481	1277	741151	580.3844949
Lordotus	*junceus*	146.3	5775	3464799	599.9651948	942	1232021	1307.877919
Lordotus	*planus_2*	169.3	5271	3471381	658.5811042	957	1271149	1328.264368
Lordotus	*zona_2*	104.6	4145	2916651	703.6552473	913	1215174	1330.968237
Lordotus	*zona_3*	150	4868	3269994	671.732539	941	1264894	1344.201913
Lordotus	*zona_4*	113.9	5461	3465706	634.6284563	919	1241118	1350.509249
Lordotus	abdominalis_2	93.6	4567	3051267	668.1118896	952	1240585	1303.135504
Lordotus	*hurdi_3*	151.7	5730	3801565	663.4493892	945	1308419	1384.57037
Lordotus	*apicula_6*	102	6149	3151442	512.5129289	972	971277	999.2561728
Lordotus	*hurdi_5*	172.6	6095	3861761	633.5949139	968	1286370	1328.894628
Lordotus	*hurdi_6*	152.7	5551	3674069	661.8751576	964	1300194	1348.748963
Lordotus	*hurdi_10*	191.7	4965	3256833	655.9583082	954	1194165	1251.745283
Lordotus	*hurdi_7*	111.5	5718	3868962	676.6285414	931	1273223	1367.586466
Lordotus	*hurdi_8*	96.3	6804	4335444	637.1904762	916	1328258	1450.063319
Lordotus	*hurdi_9*	173.7	5177	3573593	690.2825961	937	1319331	1408.037353
Lordotus	*sororculus_3*	103.1	5587	3778301	676.2665115	948	1323159	1395.737342
Lordotus	*apicula_3*	93.7	4434	3402603	767.3890392	977	1404202	1437.258956
Lordotus	*apicula_8*	134.6	10713	4546432	424.3845795	915	1212679	1325.33224
Lordotus	*apicula_4*	100.1	5346	3698623	691.8486719	946	1325432	1401.090909
Lordotus	*apicula_5*	99.2	5295	3664269	692.0243626	971	1381497	1422.756952
Lordotus	*apicula_7*	72	5662	3472329	613.2689862	939	1198811	1276.689031
Lordotus	*apicula_9*	106.7	4508	3010371	667.7841615	899	1145369	1274.047831
Lordotus	*apicula_10*	124.3	4826	3108684	644.1533361	917	1153658	1258.078517
Lordotus	*apicula_11*	156.8	5002	3394610	678.6505398	952	1258157	1321.593487
Lordotus	*apicula_12*	167.8	5550	3814324	687.2655856	946	1354098	1431.393235
Lordotus	*sororculus_4*	164.5	4990	3079235	617.0811623	961	1177197	1224.970864
Lordotus	*hurdi_4*	120.1	4768	2991062	627.3200503	914	1109645	1214.053611
Lordotus	*hurdi*	168.2	5954	3501950	588.1676184	892	1126196	1262.55157
Lordotus	*hurdi_2*	180.3	6840	3209924	469.2871345	903	1057675	1171.290144
Lordotus	*striatus*	11.7	137513	21838511	158.8105197	809	445373	550.5228677
Lordotus	planus	13.4	38540	11933274	309.6334717	1212	741425	611.7367987
Lordotus	bipartitus	8.15	76216	24124252	316.5247717	893	294386	329.6595745
Lordotus	striatus	10.3	12640	3850748	304.6477848	1139	522933	459.1158911
Lordotus	diversus	6.75	65843	16849035	255.8971341	681	208918	306.7812041
Lordotus	ermae	11.9	18242	5454070	298.9842123	1139	478628	420.2177349
Lordotus	luteolus	10.2	217285	39327773	180.996263	897	358467	399.6287625
Lordotus	zona_2	21.4	19985	5221007	261.2462847	556	159183	286.3003597
Lordotus	miscellus_2	12.3	178011	23233408	130.5166984	774	619452	800.3255814
Lordotus	albidus	12	10832	2995986	276.5865953	477	133866	280.6415094
Lordotus	*sp4*	7.6	32778	10448512	318.7660016	813	402988	495.6801968
Lordotus	*sororculus_2*	12.7	45460	14325157	315.1156401	1210	567423	468.9446281
Lordotus	divisus	39.1	163420	28981559	177.3440154	988	430116	435.340081
Lordotus	bucerus	5.55	71739	21760381	303.3270745	328	92884	283.1829268
Lordotus	rufotibialis	11.1	51484	15956433	309.9299394	1271	607692	478.1211644
Lordotus	gibbus	10.1	43755	12962950	296.2621415	1182	451426	381.9170897
Lordotus	sororculus	17.2	67494	21719084	321.7928112	1203	591547	491.726517
Lordotus	striatus_2	24.1	10170	3007508	295.7235005	1111	399143	359.2646265
Lordotus	zona_6	10.7	60120	18928955	314.8528776	570	164885	289.2719298
Lordotus	MX_sp1	9.8	22981	6926220	301.3889735	1182	525246	444.3705584
Lordotus	ermae	12.7	81497	24038289	294.9591887	1015	428035	421.7093596
Lordotus	MX_sp2	11	96443	28051080	290.8565681	802	246128	306.8927681
Lordotus	divisus_2	20.1	14737	4441559	301.3882744	879	359460	408.9419795
Lordotus	MX_sp3	10.6	17459	5156435	295.3453806	1135	463597	408.4555066
Lordotus	pulchrissimus	15.5	8392	2594244	309.1329838	1101	487054	442.3742053
Lordotus	schlingeri_2	20.7	50279	16217951	322.55914	916	661817	722.5076419
Lordotus	abdominalis	9.7	25201	7763829	308.0762271	1126	659835	585.9991119
Lordotus	arizoenensis	42.4	6271	1951904	311.2588104	1126	466335	414.151865
Lordotus	arnaudi	11.2	52070	15665582	300.8561936	925	316793	342.4789189
Lordotus	cingulatus	15.4	41502	11643618	280.5555877	1163	515308	443.0851247
Lordotus	diplasus	10.8	21973	6494655	295.5743412	1241	533555	429.9395649
Lordotus	junceus	6.2	29053	8356054	287.6141534	995	339578	341.2844221
Lordotus	lineatus	12.4	19478	8121455	416.9552829	1131	515668	455.9398762
Lordotus	lineatus_2	10.3	22369	6757875	302.1089454	1223	625751	511.6524939
Lordotus	lutecens	36	10402	3138782	301.7479331	1123	464353	413.4933215
Lordotus	nevadensis	17.9	19773	8363061	422.9535731	958	415066	433.263048
Lordotus	puella	14.9	59662	17762048	297.71124	1110	384686	346.563964
Lordotus	perplexus	19.2	5780	1800668	311.5342561	1095	421485	384.9178082
Lordotus	schlingeri	11.3	124356	40418277	325.0207228	365	103627	283.909589
Bombylius	major					558	478095	856.8010753

Table 2.4: Listing quantity of extracted DNA per each sample, in addition to total contigs, total contigs (bp), mean length of each UCE, UCE loci total, UCE total(bp), and UCE mean length (bp).

Figure 2.1: Diversity of *Lordotus* – they range from large yellow flies to small narrow bodied, silver flies. Top right photo by Alice Abela, all other photos by Joyce Gross

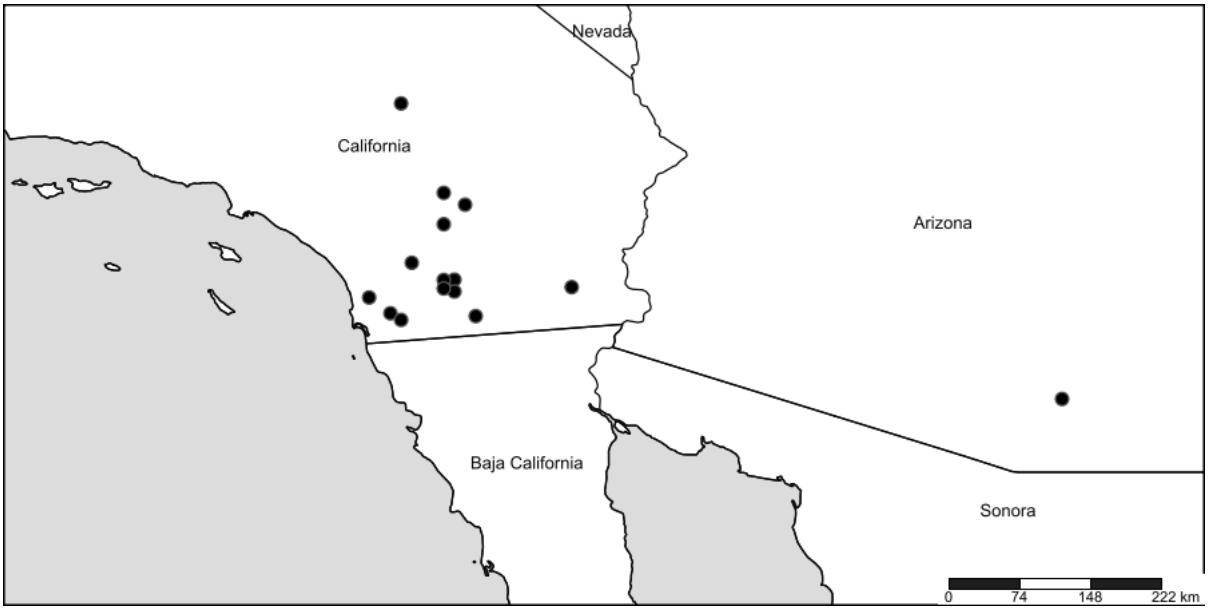


Figure 2.2: Map of live collected *Lordotus* samples.

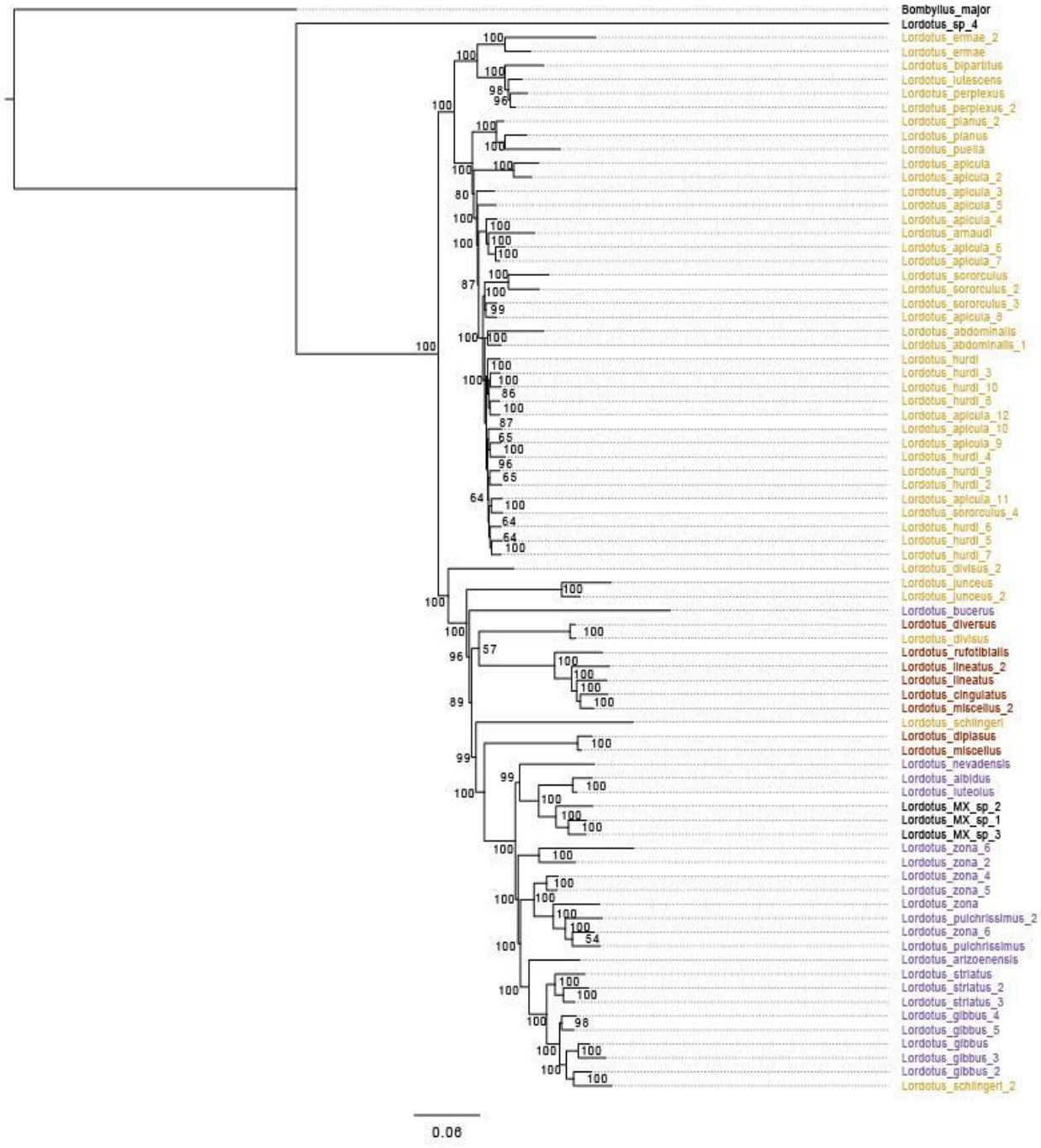


Figure 2.3: ML Phylogeny of *Lordotus*, color-coded by putative species group. Orange= *apicula*, Brown= *miscellus*, Purple= *gibbus*.

Chapter 3

De novo Genome Assembly and Identification of Heat Shock Proteins in Bombyliidae

Abstract

As one of the most important groups of organisms on the planet, it is no surprise that Diptera also has some of the most sequenced genomes compared to other invertebrate groups. However, these fly genomes are limited to species that are medically and agriculturally important. This is because sequencing has always been a costly endeavor. Recent advances in sequencing technology have changed this and allowed for cheaper alternatives like low-coverage genome sequencing. These low-coverage genomes have a high amount of versatility and usability, allowing researchers to investigate a wide range of topics. One such topic of interest is the investigation of heat shock proteins (HSPs). These are conserved proteins found in virtually all organisms. HSPs play an important role in protecting and stabilizing DNA during periods of heat stress. Desert-adapted organisms have been shown to have higher amounts of HSPs compared to non-desert-adapted organisms, however, these studies have been limited to very select groups. Live-caught and then EtOH-preserved bee flies were sequenced and used to assemble low-coverage genomes for seven different subfamilies within bee flies. The assembled genomes represent a valuable resource, allowing researchers to study a wide range of topics. These genomes were used here to identify the 12 HSPs found in these arid-adapted flies. These identified heat shock proteins will help researchers gain a better understanding of the evolution of desert-adapted flies.

Introduction

Advances in sequencing technology have enabled hundreds to thousands of loci to be collected for non-model organisms. A variety of sequencing methods are currently in use by biologists. Some methods in use are comparative transcriptomics, reduced representation sequencing (RadSeq), and hybrid enrichment sequencing (UCE/AHE) (Faircloth et al. 2012; Zhang et al. 2019). These methods have enabled researchers to study aspects of evolutionary biology not possible with traditional Sanger sequencing methods. For example, the study of rapidly diverging, deep phylogenetic relationships, or the ability to cost-effectively gather genomic data from historic museum specimens (Derkarabetian et al. 2019; Buenaventura 2021).

However, each of these methods has some limitations in their utility and potential future usability. Some methods like transcriptomics require high-quality RNA from freshly collected samples (Cron et al. 2012; McCormack et al. 2013). Techniques like AHE/UCEs are not as limited by tissue quality but require specialized baits (probes), which can be challenging to design (Faircloth 2017; Faircloth et al. 2012; Lemon et al., 2012). Additional limitations of AHE/UCE are their narrow application, these methods are primarily used for phylogenetics and have limited utility beyond that (Zhang et al. 2019). As the cost of sequencing has come down, whole genome sequencing (WGS) has become a viable option. Although the overall costs of WGS is typically still more expensive than UCEs or RadSeq, affordable options exist through the sequencing of low-coverage genomes. An advantage that low-coverage genomes have over other types of sequencing is their versatility and future usability/compatibility (Zhang et al.

2019). Many types of data can be extracted from low-coverage genomes like genotyping, evaluating the putative genetic basis of trait association, and extracting phylogenomic markers like UCEs (Homburger et al. 2019; Bowen et al. 2012; Gardner et al. 2016; Clucas et al. 2019; Gilly et al. 2019).

Diptera (Flies) are one of the most ecologically diverse orders of insects on the planet and have evolved to survive in some of the most extreme habitats imaginable (Marshall 2012; Marshall & Kirk-Spriggs 2017). It is no surprise that they have also diversified and flourished in arid habitats across the world. The fly superfamily Asiloidea is one such group of over 15,000 species, with many species found in desert habitats. Asiloidea comprises diverse taxa such as the Asilidae (~7,500 spp.), Bombyliidae (~5,500 spp.), Apioceridae (~143 spp.), Mydidae (~498 spp.), Therevidae (~1,200 spp.) and Scenopinidae (~420 spp.). Despite flies having the most sequenced insect genomes available, to date only one published, complete genome is available from asiloid flies (Dikow et al. 2017). The first aim of this chapter is to sequence and assemble low-coverage genomes of bee flies to generate a reference data set that can be used by future researchers. The second aim is to use the genomes to identify putative heat shock proteins in bee flies.

Heat Shock proteins (HSPs) are ubiquitous, conserved proteins found in the tree of life (Starrett and Waters 2007; Waters 2014). As some of the most conserved cellular and molecular mechanisms, HSPs play an important role in protecting cells during severe stress, typically seen during high temperatures (Garbuz et al. 2008). Heat shock proteins help in preventing the accumulation of denatured proteins, and in refolding denatured proteins (Garbuz et al. 2008; Garbuz and Evgen'ev 2008; Sivan et al. 2017). Heat shock proteins are ubiquitous in cells, but when organisms are heat stressed, they proportionally upregulate the production of HSPs (Waters 2014; Garbuz and Evgen'ev 2008). Overall, HSPs provide organisms with a higher level of cellular heat tolerance. Heat shock proteins can be categorized into six different families, based on their molecular weight (Garbuz and Evgen'ev 2008). These families are HSP 20, HSP40, HSP60, HSP70, HSP 90 and HSP100 (Garbuz and Evgen'ev 2008; Jing and Li 2020) Interestingly, each of these families of HSPs have their own unique evolutionary history and are unrelated to each other (Waters 2014; Garbuz and Evgen'ev 2008). In addition, the HSP families are not consistently found across the tree of life, different HSP families are present in different taxa and may, for example, differ across organismal families (Waters 2014; Garbuz and Evgen'ev 2008; Krebs and Feder 1997).

Many desert-adapted taxa have higher concentrations of HSPs present in their cells and produce larger amounts of HSPs that allow them to tolerate harsh temperatures (Evgen'ev et al. 2007). The heat shock proteins of some Diptera have been studied, however, these studies have primarily focused on *Drosophila*, with few studies on other genera (Garbuz et al 2008; Krebs and Feder 2007; Garbuz et al. 2002; Astakhova et al. 2014). Bombyliidae or bee flies are one of the most diverse taxa of flies found in deserts, acting as important pollinators and parasitoids (de Jager and Ellis 2017; Yeates and Greathead 1997; Hull 1973; Li et al. 2021). Given the diversity of bee fly species (~5,000) and the many species found in arid and semi-arid habitats, the study of their heat shock proteins could yield insight into their adaptation and diversification across desert biomes. The goal of this project is to identify and characterize the heat shock proteins of bee flies to answer these main questions: Can HSPs be identified from a low-coverage genome?

Which heat shock proteins do bee flies have? Do the HSP protein families vary across bee fly subfamilies?

Methods

Taxon Sampling, Extraction and Sequencing

Genomic DNA was extracted from seven EtOH preserved specimens representing different subfamilies in Bombyliidae (*Bombylius albicapillus* Loew, 1872 (Bombyliinae); *Anthrax sp.* (Anthracinae); *Lordotus sp.* (Lordotinae); *Pantarbes sp.* (Cythereinae); *Geron sp.* (Toxophorinae); *Mythicomyia sp.* (Mythicomyiinae); *Apolysis sp.* (Usiinae)) with the DNeasy Blood and Tissue Kit (Qiagen, Valencia, CA, USA) using the manufacturer's protocol (Table 3.1). The head, thorax, abdomen, and legs were all separated to ensure the lysate could reach all the soft tissue. Samples were left overnight in Proteinase K digestion to maximize DNA yield. Samples of *B. albicapillus* and *Anthrax sp.* were sequenced using HiSeq 250 paired-end reads at the University of California, Berkeley Sequencing Facility. The remaining samples were sequenced using Illumina NovaSeq 6000 150 paired-end reads at 10-20x coverage at QB3 Genomics at the University of California, Berkeley.

Genome Assembly

Raw sequences were examined using FastQC (Andrews 2010). FastQC quickly analyzes raw data to determine the quality of the reads. FastQC identifies overrepresented sequences such as adapter contamination, it also provides information about GC content, duplicated sequences, poor quality sequences, and sequence length. Afterward, sequence quality trimming was performed using Trim Galore! (Krueger 2021), which automatically detects and trims adapters, and removes sequence pairs that are shorter than expected (Figure 3.1).

Cleaned sequences were inputted into Velvet Optimizer (Simon 2012) to determine optimal kmer value. The genome of *Bombylius albicapillus* was de novo assembled using Velvet (Zerbino and Birney 2008; Zerbrino et al. 2009), which assembles a genome using de Brujin graphs. Genomes of the other samples were assembled using Minia (https://github.com/GATB/minia) and the GATB Minia pipeline (https://github.com/GATB/gatb-minia-pipeline), which has been found to quickly de novo assemble genomes using de Brujin graphs (Chikhi et al. 2017; Zhang et al. 2019).

Genome Statistics

Contigs generated by Velvet and Minia were uploaded to the Galaxy web platform to analyze the effectiveness of the genome assemblers (https://usegalaxy.org/). Assembled contigs were inputted into FASTA statistics that generated general statistics on the assembled genome FASTA file (Kyran 2021). QUality ASsessment Tool or QUAST was further used to evaluate genome assemblies, QUAST can assess genome quality with or without a reference genome (Gurevich et al. 2013, Mikheenko et al. 2018). Results from QUAST were visualized in Icarus (Mikheenko et al. 2016).

Protein Identification

Assembled contigs were uploaded into the InterPro (Paysan-Lafosse et al. 2022) database to determine what coding sequences were found in the contigs. The InterPro blasts assembled genome contigs across numerous member databases to search sequences and identify protein families, domains, and functional sites (Paysan-Lafosse et al. 2022).

The gene prediction program AUGUSTUS was used on the assembled contigs to generate FASTA files with predicted coding sequences and protein sequences (Stanke & Waack 2003; Keller et al. 2011; Stanke et al. 2008). The generated sequences were then used to search for HSPs. Published HSPs belonging to *Drosophila* were downloaded from the National Center for Biotechnology Information (NCBI) Conserved Domain Database (Table 3.2). Subsequently, these HSPs were used as queries against the assembled protein sequences generated by AUGUSTUS using the BlastP programs on NCBI, an *E* value cut-off (e^{-20}) was used following other authors (Wang et al. 2019; Huang et al. 2018; Chaudhary et al. 2019; Park et al. 2020).

Results and Discussion

Sequencing and Assembly

Extraction of specimens was successful and yielded sufficient DNA required for sequencing (Table 3.1). The quality of sequences was assessed using FastQC, which showed that sequencing was an overall success. FastQC report indicated high per base sequence quality, found no overrepresented sequences, and sequence quality scores passed FactQC quality checkpoint (Andrews, S. 2010). Cleaned sequences were inputted into Velvet Optimiser, which found a kmer value of 41 to be the optimal value for genomic assembly (Simon 2012). Velvet assembler required a significant amount of time and computational power to assemble the genome of *Bombylius albicapillus*, taking over a week on an external server. This could be because it was originally designed for microbial data and sequence data with very short reads (Zerbino and Birney 2008; Zerbrino et al. 2009). In contrast, the GATB Minia pipeline was significantly faster, taking on average, a day per species to assemble each remaining genomes. Low coverage genomes for the seven species of Bombyliidae were successfully assembled. Basic FASTA statistics are summarized in Table 3.3. Results from QUAST quality assessment are summarized in Table 3.4. N50 ranged in size from 144 in *Bombylius albicapillus* to 1252 in *Apolysis sp*.

These sequenced genomes represent an important baseline and set of data that can be used in future studies in a variety of ways. Researchers can extract other genomic regions from the data, like a mitochondrial genome set (Dierckxens et al. 2017; Hahn et al. 2013). Other potential uses include extracting UCE loci from assembled genomes, finding population genomic markers or studying the molecular evolution in regions of interest (Vekemans et al. 2021; Rustagi et al. 2017; Zhang et al. 2019; Gardner et al. 2016; Olofsson et al. 2019).

Protein Identification

The InterPro search of contigs identified numerous signal peptides and some anti-freeze proteins, but no heat shock proteins from the FASTA files. Similar in function to HSPs, anti-freeze proteins (AFPs) are used by organisms to protect their DNA from extreme cold temperatures, they bind to small ice crystals preventing further growth or expansion (Davies et al. 2002; Wen and Laursen 1993; Chou 1992). AFPs have been found in various insects, including midges (Basu et al. 2016). Although typically associated with arid habitats, bee flies can be found in high elevations, especially in high-elevation deserts, it is possible that these AFPs are utilized by bee flies to protect themselves from cold temperatures at night.

Twelve different HSPs were used as queries against the bee fly genomes. Results from the BLASTP are summarized in Table 3.5. Identification of HSPs using low-coverage genomes was successful in all seven of the sequenced bee flies. Using the HSP sequences of various *Drosophila* species, I found 12 HSPs of interest in sequenced bee flies. However, not every sequenced bee fly had all 12 HSPs of interest. The genomes of *Anthrax* and *B. albicapillus* had the fewest identified heat shock proteins, when queried, many of them failed and gave the notice "no significant similarity found". However, this does not necessarily mean that these flies do not have those heat shock proteins. There are various reasons why these proteins may not have been found, these could be from short sequence bias missing data, filtering parameters, sequence, and assembly quality issues (Rhie et al. 2021). Further long-read sequencing may recover additional HSPs in these flies (Rhie et al. 2021; Hirakawa et al. 2019). The use of transcriptome sequencing may be required to adequately capture all HSPs in bee flies (Li et al. 2021). Live bee flies would have to be under severe heat stress before collection, allowing for the HSPs to be expressed in sufficient quantities for sequencing.

Of the queried HSPs, only *Pantarbes sp.*, *Mythicomyia sp.*, and *Apolysis sp.* had all of them. Given that these samples were collected in desert environments, during the hottest part of the day, it is possible that these flies have higher amounts of HSPs. Having numerous HSPs likely plays an important role in protecting bee flies from the intense heat of their environments. Many bee flies can typically be found in higher activity during the warmest part of the day, a time when many animals are less active, it is likely these HSPs protect bee flies during this time (Hull 1973).

Identifying and understanding which heat shock proteins bee flies have is important for understanding the evolution of this group in desert and semi-arid environments. Now that HSPs have been identified, further experimentation into their expression is required to determine how bee flies are utilizing these proteins and to what capacity. Animals that are part of the desert fauna have been found to express larger quantities of HSPs compared to non-desert organisms, so it is possible that desert bee flies also express HSPs in higher amounts compared to non-arid bee flies or other flies (Evgen'ev et al.2007). These types of expression experiments however are only possible with the capture of bee flies while they are being subjected to large amounts of heat stress (Chaudhary et al. 2019; Huang et al. 2018).

Conclusion

This study is the first to assemble genomes for bee flies. These genomes represent a valuable genomic resource for future studies and will allow researchers to investigate many aspects of bee fly biology. Of particular interest has been the diversity and evolution of bee flies across arid habitats. Understanding what HSPs bee flies have, and how they are expressed may provide insight into how these diverse parasitoid insects have diversified across some of the most extreme habitats.

Subfamily	Genuis	Species	Specimen Code	Final Vol	Concentration	Country	State	County	Locality	Latt	Long
Bombyliinae	Bombylius	albicapillus	Dip 31	100ul	unrecorded	USA	California	San Diego	Wildwood Glen	32.8418	-116.64569
Ussinae	Apolysis	sp.	Dip 38	55ul	5.44 ng/uL	USA	California	San Diego	within 5km of Hayde	32.7122	-116.1158
Mythicomyia	Mythicomyia	sp.	Dip 153	55ul	0.708 ng/uL	USA	California	San Diego	Hollenbeck Canyon \	33.6707	-116.8238
Toxophorinae	Geron	sp.	Dip 228	55ul	7.08 ng/uL	USA	Arizona	Santa Cruz	Along Ruby rd. West	31.3888	-111.1206
Lordotinae	Lordotus	sp.	Dip 360	100ul	3.2 ng/uL	USA	California	Riverside	Road to Coyote Was	33.6502	-116.3738
Cythereinae	Pantarbes	sp.	Dip 361	100ul	1.23 ng/uL	USA	California	San Bernardi	~6mi S of I-40, alor	34.7809	-116.8396
Anthracinae	Anthrax	sp.	1021.2	100ul	unrecorded	France	French Poly	Mo'orea	Richard B. Gump Sol	-	-

Table 3.1: Specimen information used for genomic sequencing.

Heat shock Proteins	Accession Number
20	P97541.1
23	P02516.2
26	P02517.2
27	P02518.2
40	P25686.3
60-a	O02649.3
68	O97125.1
70-2	P11146.2
70-4	P11147.3
70-5	P29845.2
83	P04809.2
90	P07900.5

Table 3.2: Heat shock proteins queried.

FASTA Stats	Bombylius al	Anthrax sp.	Geron sp.	Mythicomyi	Apolysis sp.	Pantarbes sp.	Lordotus sp.
Scaffold L50	831930	313741	367851	22950	57862	344955	268737
Scaffold N50	144	268	278	987	1252	220	257
Scaffold L90	2441422	1625048	1691808	236546	507201	1545591	1279500
Scaffold N90	81	48	84	128	69	67	80
Scaffold len_max	4546	40359	72555	224994	28931	56426	48482
Scaffold len_min	81	41	41	41	41	41	41
Scaffold len_mean	143	155	203	291	253	170	186
Scaffold len_median	109	76	136	133	53	107	116
Scaffold len_std	114	269	411	1782	655	385	453
Scaffold num_A	163778490	161408418	203115862	34391963	110600946	141857968	140548357
Scaffold num_T	158663603	147908069	199028724	34070893	105829292	139050917	138349051
Scaffold num_C	52152620	47252387	74095571	29068898	39763592	65823620	44688078
Scaffold num_G	52562689	46740147	73568313	28897180	39105700	64652391	44337370
Scaffold num_N	0	0	0	0	0	0	0
Scaffold num_bp	427157402	403309021	549808470	126428934	295299530	411384896	367922856
Scaffold num_bp_not_N	427157402	403309021	549808470	126428934	295299530	411384896	367922856
Scaffold num_seq	2968776	2587423	2704172	433762	1165344	2418838	1970595
Scaffold GC content overall	24.51	23.31	26.86	45.85	26.71	31.72	24.2
Contig L50	831930	313741	367851	22950	57862	344955	268737
Contig N50	144	268	278	987	1252	220	257
Contig L90	2441422	1625048	1691808	236546	507201	1545591	1279500
Contig N90	81	48	84	128	69	67	80
Contig len_max	4546	40359	72555	224994	28931	56426	48482
Contig len_min	81	41	41	41	41	41	41
Contig len_mean	143	155	203	291	253	170	186
Contig len_median	109	76	136	133	53	107	116
Contig len_std	114	269	411	1782	655	385	453
Contig num_bp	427157402	403309021	549808470	126428934	295299530	411384896	367922856
Contig num_seq	2968776	2587423	2704172	433762	1165344	2418838	1970595
Number of gaps	0	0	0	0	0	0	0

Table 3.3: FASTA stats of assembled genomes.

Statistics without reference	Bombylius albicapillus	Anthrax sp.	Geron sp.	Mythicomyia sp.	Apolysis sp.	Pantarbes sp.	Lordotus sp.
# contigs	2968776	2587423	2704172	433762	1165344	2418838	1970595
# contigs (>= 0 bp)	2968776	2587423	2704172	433762	1165344	2418838	1970595
# contigs (>= 1000 bp)	6436	42522	64202	22590	73223	42545	34933
Largest contig	4546	40359	72555	224994	28931	56426	48482
Total length	427157402	403309021	549808470	126428934	295299530	411384896	367922856
Total length (>= 0 bp)	427157402	403309021	549808470	126428934	295299530	411384896	367922856
Total length (>= 1000 bp)	8291825	71102260	136206962	62857754	164851545	92212156	82233670
N50	144	268	278	987	1252	220	257
N90	81	48	84	128	69	67	80
auN	234.3	622.6	1038	11196	1946.5	1041.9	1289.9
L50	831930	313741	367851	22950	57862	344955	268737
L90	2441422	1625048	1691808	236546	507201	1545591	1279500
GC (%)	24.51	23.31	26.86	45.85	26.71	31.72	24.2
Mismatches							
# N's per 100 kbp	0	0	0	0	0	0	0
# N's	0	0	0	0	0	0	0

Table 3.4: QUAST results for assembled genomes.

Heatshock Proteins	Bombylius albicapillus	Anthrax sp.	Geron sp.	Mythicomyia sp.	Apolysis sp.	Pantarbes sp.	Lordotus sp.
20	not detected	not detected	not detected	present	present	present	not detected
23	not detected	not detected	not detected	present	present	present	not detected
26	not detected	not detected	not detected	present	present	present	not detected
27	not detected	not detected	not detected	present	present	present	not detected
40	not detected	present	present	present	present	present	present
60-a	not detected	not detected	present	present	present	present	present
68	present	present	present	present	present	present	present
70-2	present	present	present	present	present	present	present
70-4	present	present	present	present	present	present	present
70-5	not detected	present	present	present	present	present	present
83	not detected	not detected	present	present	present	present	present
90	not detected	not detected	present	present	present	present	present

Table 3.5: Heat shock proteins found in assembled genomes.

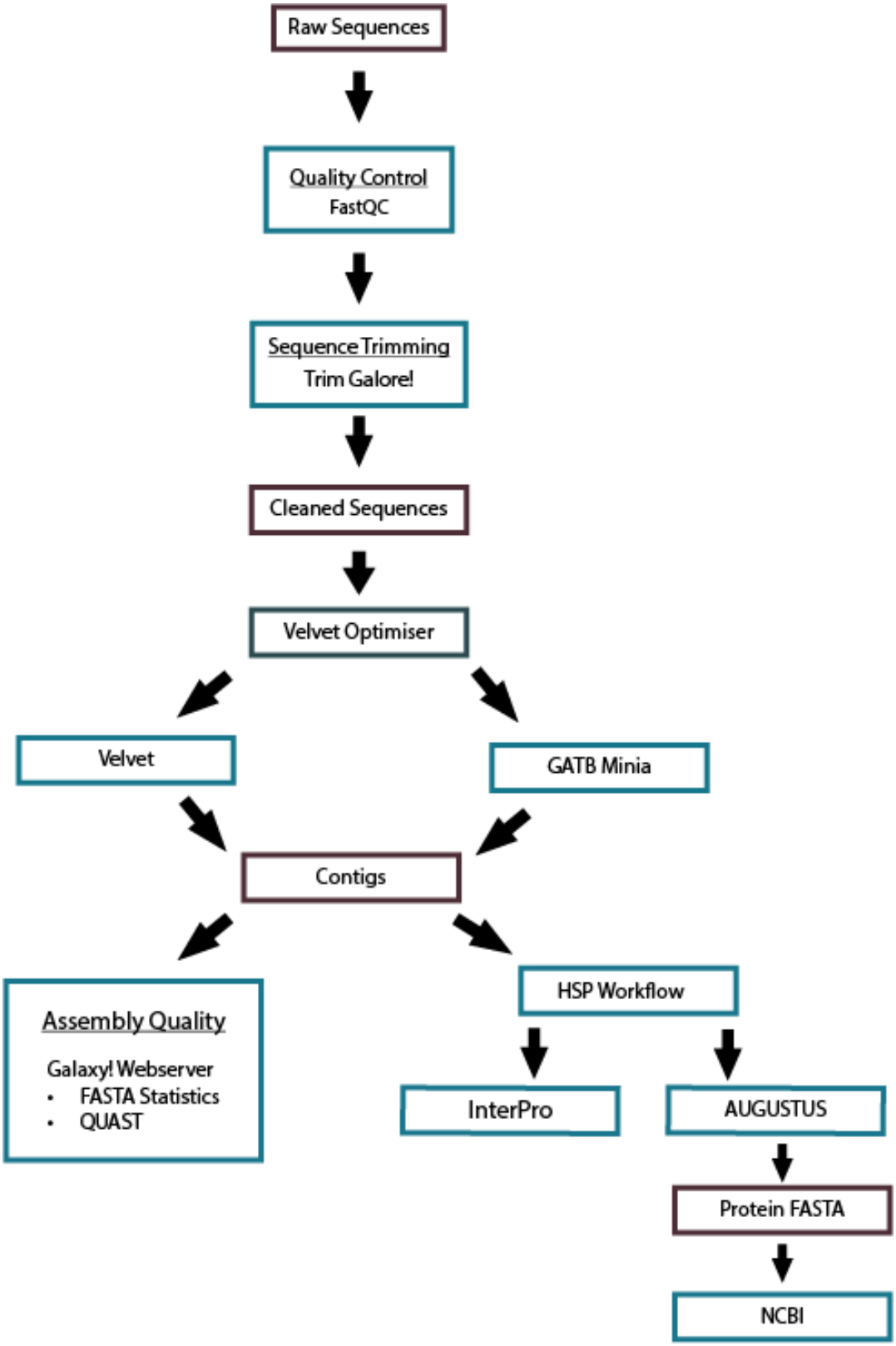


Figure 3.1: Bioinformatic workflow for genome assembly and HSP discovery.

References

Ahmad, A., Broce, A., Zurek, L., 2006. Evaluation of significance of bacteria in larval development of *Cochliomyia macellaria* (Diptera: Calliphoridae). ment 43, 1129–1133.

Anderson, K.E., Sheehan, T.H., Mott, B.M., Maes, P., Snyder, L., Schwan, M.R., Walton, A., Jones, B.M., Corby-Harris, V., 2013. Microbial ecology of the hive and pollination landscape: Bacterial associates from floral nectar, the alimentary tract and stored food of honey bees (*Apis mellifera*). PLOS ONE 8, e83125. https://doi.org/10.1371/journal.pone.0083125

Andrews, S., FastQC A quality control tool for high throughput sequence data https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed 4.8.23).

Astakhova, L.N., Zatsepina, O.G., Evgen'ev, M.B., Garbuz, D.G., 2014. Comparative analysis of effectiveness of heat-shock promoters in two Diptera species. Mol Biol 48, 377–383. https://doi.org/10.1134/S0026893314030029

Bahrndorff, S., Alemu, T., Alemneh, T., Lund Nielsen, J., 2016. The microbiome of animals: implications for conservation biology. International Journal of Genomics 2016, e5304028. https://doi.org/10.1155/2016/5304028

Barrera, C.A., Sosa-Calvo, J., Schultz, T.R., Rabeling, C., Bacci Jr, M., 2022. Phylogenomic reconstruction reveals new insights into the evolution and biogeography of *Atta* leaf-cutting ants (Hymenoptera: Formicidae). Systematic Entomology 47, 13–35. https://doi.org/10.1111/syen.12513

Basu, K., Wasserman, S.S., Jeronimo, P.S., Graham, L.A., Davies, P.L., 2016. Intermediate activity of midge antifreeze protein is due to a tyrosine-rich ice-binding site and atypical ice plane affinity. The FEBS Journal 283, 1504–1515. https://doi.org/10.1111/febs.13687

Bejerano, G., Pheasant, M., Makunin, I., Stephen, S., Kent, W.J., Mattick, J.S., Haussler, D., 2004. Ultraconserved elements in the human genome. Science 304, 1321–1325. https://doi.org/10.1126/science.1098119

Benda, D., Votýpková, K., Nakase, Y., Straka, J., 2021. Unexpected cryptic species diversity of parasites of the family Xenidae (Strepsiptera) with a constant diversification rate over time. Systematic Entomology 46, 252–265. https://doi.org/10.1111/syen.12460

Blaimer, B.B., Brady, S.G., Schultz, T.R., Lloyd, M.W., Fisher, B.L., Ward, P.S., 2015. Phylogenomic methods outperform traditional multi-locus approaches in resolving deep evolutionary history: a case study of formicine ants. BMC Evol Biol 15, 271. https://doi.org/10.1186/s12862-015-0552-5

Blaimer, B.B., LaPolla, J.S., Branstetter, M.G., Lloyd, M.W., Brady, S.G., 2016a. Phylogenomics, biogeography and diversification of obligate mealybug-tending ants in the genus *Acropyga*. Molecular Phylogenetics and Evolution 102, 20–29. https://doi.org/10.1016/j.ympev.2016.05.030

Blaimer, B.B., Lloyd, M.W., Guillory, W.X., Brady, S.G., 2016b. Sequence capture and phylogenetic utility of genomic ultraconserved elements obtained from pinned insect specimens. PLOS ONE 11, e0161531. https://doi.org/10.1371/journal.pone.0161531

Bordenstein, S.R., Theis, K.R., 2015. Host biology in light of the microbiome: ten principles of holobionts and hologenomes. PLOS Biology 13, e1002226. https://doi.org/10.1371/journal.pbio.1002226

Bossert, S., Danforth, B.N., 2018. On the universality of target-enrichment baits for phylogenomic research. Methods Ecol Evol 9, 1453–1460. https://doi.org/10.1111/2041-210X.12988

Bowen, M.E., Henke, K., Siegfried, K.R., Warman, M.L., Harris, M.P., 2012. Efficient mapping and cloning of mutations in zebrafish by low-coverage whole-genome sequencing. Genetics 190, 1017–1024. https://doi.org/10.1534/genetics.111.136069

Bowles, D.J., Lillford, P.J., Rees, D.A., Shanks, I.A., Davies, P.L., Baardsnes, J., Kuiper, M.J., Walker, V.K., 2002. Structure and function of antifreeze proteins. Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences 357, 927–935. https://doi.org/10.1098/rstb.2002.1081

Branstetter, M.G., Danforth, B.N., Pitts, J.P., Faircloth, B.C., Ward, P.S., Buffington, M.L., Gates, M.W., Kula, R.R., Brady, S.G., 2017. Phylogenomic insights into the evolution of stinging wasps and the origins of ants and bees. Current Biology 27, 1019–1025. https://doi.org/10.1016/j.cub.2017.03.027

Brooks, A.W., Kohl, K.D., Brucker, R.M., Opstal, E.J. van, Bordenstein, S.R., 2016. Phylosymbiosis: relationships and functional effects of microbial communities across host evolutionary History. PLOS Biology 14, e2000225. https://doi.org/10.1371/journal.pbio.2000225

Buck, B., 2016. Fruit fly outbreak, cost at least $4m in crop damage [WWW Document]. Specialty Crop Industry. URL https://specialtycropindustry.com/fruit-fly-outbreak-cost-at-least-4m-in-crop-damage/

Buenaventura, E., 2021. Museomics and phylogenomics with protein-encoding ultraconserved elements illuminate the evolution of life history and phallic morphology of flesh flies (Diptera: Sarcophagidae). BMC Ecology and Evolution 21, 70. https://doi.org/10.1186/s12862-021-01797-7

Buenaventura, E., Lloyd, M.W., Perilla López, J.M., González, V.L., Thomas-Cabianca, A., Dikow, T., 2021. Protein-encoding ultraconserved elements provide a new phylogenomic perspective of Oestroidea flies (Diptera: Calyptratae). Syst Entomol 46, 5–27. https://doi.org/10.1111/syen.12443

Capella-Gutiérrez, S., Silla-Martínez, J.M., Gabaldón, T., 2009. trimAl: a tool for automated alignment trimming in large-scale phylogenetic analyses. Bioinformatics 25, 1972–1973. https://doi.org/10.1093/bioinformatics/btp348

Caporaso, J.G., Lauber, C.L., Walters, W.A., Berg-Lyons, D., Huntley, J., Fierer, N., Owens, S.M., Betley, J., Fraser, L., Bauer, M., Gormley, N., Gilbert, J.A., Smith, G., Knight, R., 2012. Ultra-high-throughput microbial community analysis on the Illumina HiSeq and MiSeq platforms. ISME J 6, 1621–1624. https://doi.org/10.1038/ismej.2012.8

Chan, K.O., Hutter, C.R., Wood Jr, P.L., Grismer, L.L., Das, I., Brown, R.M., 2020. Gene flow creates a mirage of cryptic species in a Southeast Asian spotted stream frog complex. Molecular Ecology 29, 3970–3987. https://doi.org/10.1111/mec.15603

Chaudhary, R., Baranwal, V.K., Kumar, R., Sircar, D., Chauhan, H., 2019. Genome-wide identification and expression analysis of Hsp70, Hsp90, and Hsp100 heat shock protein genes in barley under stress conditions and reproductive development. Funct Integr Genomics 19, 1007–1022. https://doi.org/10.1007/s10142-019-00695-y

Chen, H., Boutros, P.C., 2011. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics 12, 35. https://doi.org/10.1186/1471-2105-12-35

Chen, S., Zhou, Y., Chen, Y., Gu, J., 2018. fastp: an ultra-fast all-in-one FASTQ preprocessor. Bioinformatics 34, i884–i890. https://doi.org/10.1093/bioinformatics/bty560

Chikhi, Rayan, et al. "High-quality, fast, and memory-efficient assembly of metagenomes and large genomes using Minia-pipeline." Biologie, Informatiqueet Mathématiques: 170.

Chou, K.-C., 1992. Energy-optimized structure of antifreeze protein and its binding mechanism. Journal of Molecular Biology 223, 509–517. https://doi.org/10.1016/0022-2836(92)90666-8

Clucas, G.V., Lou, R.N., Therkildsen, N.O., Kovach, A.I., 2019. Novel signals of adaptive genetic variation in northwestern Atlantic cod revealed by whole-genome sequencing. Evolutionary Applications 12, 1971–1987. https://doi.org/10.1111/eva.12861

Cohen, C.M., Noble, K., Jeffrey Cole, T., Brewer, M.S., 2021. The phylogeny of robber flies (Asilidae) inferred from ultraconserved elements. Systematic Entomology 46, 812–826. https://doi.org/10.1111/syen.12490

Cohen, Y., Gurevitz, M., 2006. The Cyanobacteria—ecology, physiology and molecular genetics, in: Dworkin, M., Falkow, S., Rosenberg, E., Schleifer, K.-H., Stackebrandt, E. (Eds.), The Prokaryotes. Springer US, New York, NY, pp. 1074–1098. https://doi.org/10.1007/0-387-30744-3_39

Cole, J.R., Wang, Q., Fish, J.A., Chai, B., McGarrell, D.M., Sun, Y., Brown, C.T., Porras-Alfaro, A., Kuske, C.R., Tiedje, J.M., 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis. Nucleic Acids Research 42, D633–D642. https://doi.org/10.1093/nar/gkt1244

Colman, D.R., Toolson, E.C., Takacs-Vesbach, C.D., 2012. Do diet and taxonomy influence insect gut bacterial communities? Molecular Ecology 21, 5124–5137. https://doi.org/10.1111/j.1365-294X.2012.05752.x

Crawford, N.G., Faircloth, B.C., McCormack, J.E., Brumfield, R.T., Winker, K., Glenn, T.C., 2012. More than 1000 ultraconserved elements provide evidence that turtles are the sister group of archosaurs. Biology Letters 8, 783–786. https://doi.org/10.1098/rsbl.2012.0331

Cronn, R., Knaus, B.J., Liston, A., Maughan, P.J., Parks, M., Syring, J.V., Udall, J., 2012. Targeted enrichment strategies for next-generation plant biology. American Journal of Botany 99, 291–311. https://doi.org/10.3732/ajb.1100356

De Cock, M., Virgilio, M., Vandamme, P., Bourtzis, K., De Meyer, M., Willems, A., 2020. Comparative microbiomics of tephritid frugivorous pests (diptera: tephritidae) from the field: A tale of high variability across and within species. Frontiers in Microbiology 11.

de Jager, M.L., Ellis, A.G., 2017. Evolutionary history of a keystone pollinator parallels the biome occupancy of angiosperms in the Greater Cape Floristic Region. Molecular Phylogenetics and Evolution 107, 530–537. https://doi.org/10.1016/j.ympev.2016.12.004

Deguenon, J.M., Travanty, N., Zhu, J., Carr, A., Denning, S., Reiskind, M.H., Watson, D.W., Michael Roe, R., Ponnusamy, L., 2019. Exogenous and endogenous microbiomes of wild-caught *Phormia regina* (Diptera: Calliphoridae) flies from a suburban farm by 16S rRNA gene sequencing. Sci Rep 9, 20365. https://doi.org/10.1038/s41598-019-56733-z

Derkarabetian, S., Benavides, L.R., Giribet, G., 2019. Sequence capture phylogenomics of historical ethanol-preserved museum specimens: Unlocking the rest of the vault. Molecular Ecology Resources 19, 1531–1544. https://doi.org/10.1111/1755-0998.13072

Dharampal, P.S., Carlson, C., Currie, C.R., Steffan, S.A., 2019. Pollen-borne microbes shape bee fitness. Proceedings of the Royal Society B: Biological Sciences 286, 20182894. https://doi.org/10.1098/rspb.2018.2894

Di Salvo, M., Calcagnile, M., Talà, A., Tredici, S.M., Maffei, M.E., Schönrogge, K., Barbero, F., Alifano, P., 2019. The microbiome of the *Maculinea-Myrmica* host-parasite interaction. Sci Rep 9, 8048. https://doi.org/10.1038/s41598-019-44514-7

Dicke, M., Cusumano, A., Poelman, E.H., 2020. Microbial symbionts of parasitoids. Annu. Rev. Entomol. 65, 171–190. https://doi.org/10.1146/annurev-ento-011019-024939

Dierckxsens, N., Mardulyn, P., Smits, G., 2016. NOVOPlasty: *de novo* assembly of organelle genomes from whole genome data. Nucleic Acids Res gkw955. https://doi.org/10.1093/nar/gkw955

Dikow, R.B., Frandsen, P.B., Turcatel, M., Dikow, T., 2017. Genomic and transcriptomic resources for assassin flies including the complete genome sequence of *Proctacanthus coquilletti* (Insecta: Diptera: Asilidae) and 16 representative transcriptomes. PeerJ 5, e2951. https://doi.org/10.7717/peerj.2951

Dillon, R., Charnley, K., 2002. Mutualism between the desert locust *Schistocerca gregaria* and its gut microbiota. Research in Microbiology 153, 503–509. https://doi.org/10.1016/S0923-2508(02)01361-X

Dittmer, J., van Opstal, E.J., Shropshire, J.D., Bordenstein, S.R., Hurst, G.D.D., Brucker, R.M., 2016. Disentangling a holobiont – recent advances and perspectives in *Nasonia* wasps. Frontiers in Microbiology 7.

Douglas, A.E., 2018. Fundamentals of microbiome science: how microbes shape animal biology. Princeton University Press.

Engel, P., Moran, N.A., 2013. The gut microbiota of insects – diversity in structure and function. FEMS Microbiology Reviews 37, 699–735. https://doi.org/10.1111/1574-6976.12025

Estes, A.M., Segura, D.F., Jessup, A., Wornoayporn, V., Pierson, E.A., 2014. Effect of the symbiont *Candidatus* Erwinia dacicola on mating success of the olive fly *Bactrocera oleae* (Diptera: Tephritidae). Int. J. Trop. Insect Sci. 34, S123–S131. https://doi.org/10.1017/S1742758414000174

Evans, J.D., Armstrong, T.-N., 2006. Antagonistic interactions between honey bee bacterial symbionts and implications for disease. BMC Ecol 6, 4. https://doi.org/10.1186/1472-6785-6-4

Evgen'ev, M.B., Garbuz, D.G., Shilova, V.Y., Zatsepina, O.G., 2007. Molecular mechanisms underlying thermal adaptation of xeric animals. J Biosci 32, 489–499. https://doi.org/10.1007/s12038-007-0048-6

Faircloth, B.C., 2017. Identifying conserved genomic elements and designing universal bait sets to enrich them. Methods in Ecology and Evolution 8, 1103–1112. https://doi.org/10.1111/2041-210X.12754

Faircloth, B.C., 2016. PHYLUCE is a software package for the analysis of conserved genomic loci. Bioinformatics 32, 786–788. https://doi.org/10.1093/bioinformatics/btv646

Faircloth, B.C., Branstetter, M.G., White, N.D., Brady, S.G., 2015. Target enrichment of ultraconserved elements from arthropods provides a genomic perspective on relationships among Hymenoptera. Mol Ecol Resour 15, 489–501. https://doi.org/10.1111/1755-0998.12328

Faircloth, B.C., Sorenson, L., Santini, F., Alfaro, M.E., 2013. A phylogenomic perspective on the radiation of ray-finned fishes based upon targeted sequencing of ultraconserved elements (UCEs). PLOS ONE 8, e65923. https://doi.org/10.1371/journal.pone.0065923

Feener Jr, D.H., Brown, B.V., 1997. Diptera as parasitoids. Annu. Rev. Entomol. 42, 73–97. https://doi.org/10.1146/annurev.ento.42.1.73

Galaxy Community, 2022. The Galaxy platform for accessible, reproducible and collaborative biomedical analyses: 2022 update. Nucleic Acids Res 50, W345–W351. https://doi.org/10.1093/nar/gkac247

Garbuz, D.G., Evgen'ev, M.B., 2017. The evolution of heat shock genes and expression patterns of heat shock proteins in the species from temperature contrasting habitats. Russ J Genet 53, 21–38. https://doi.org/10.1134/S1022795417010069

Garbuz, D.G., Molodtsov, V.B., Velikodvorskaia, V.V., Evgen'ev, M.B., Zatsepina, O.G., 2002. Evolution of the response to heat shock in genus *Drosophila* 38.

Garbuz, D.G., Zatsepina, O.G., Przhiboro, A.A., Yushenova, I., Guzhova, I.V., Evgen'ev, M.B., 2008. Larvae of related Diptera species from thermally contrasting habitats exhibit continuous up-regulation of heat shock proteins and high thermotolerance. Molecular Ecology 17, 4763–4777. https://doi.org/10.1111/j.1365-294X.2008.03947.x

Gardner, E.M., Johnson, M.G., Ragone, D., Wickett, N.J., Zerega, N.J.C., 2016. Low-coverage, whole-genome sequencing of *Artocarpus camansi* (Moraceae) for phylogenetic marker development and gene discovery. Applications in Plant Sciences 4, 1600017. https://doi.org/10.3732/apps.1600017

Gawande, S.J., Anandhan, S., Ingle, A., Roylawar, P., Khandagale, K., Gawai, T., Jacobson, A., Asokan, R., Singh, M., 2019. Microbiome profiling of the onion thrips, *Thrips tabaci* Lindeman (Thysanoptera: Thripidae). PLOS ONE 14, e0223281. https://doi.org/10.1371/journal.pone.0223281

Geiger, A., Fardeau, M.-L., Grebaut, P., Vatunga, G., Josénando, T., Herder, S., Cuny, G., Truc, P., Ollivier, B., 2009. First isolation of *Enterobacter*, *Enterococcus*, and *Acinetobacter* spp. as inhabitants of the tsetse fly (*Glossina palpalis palpalis*) midgut. Infection, Genetics and Evolution 9, 1364–1370. https://doi.org/10.1016/j.meegid.2009.09.013

Gilly, A., Southam, L., Suveges, D., Kuchenbaecker, K., Moore, R., Melloni, G.E.M., Hatzikotoulas, K., Farmaki, A.-E., Ritchie, G., Schwartzentruber, J., Danecek, P., Kilian, B., Pollard, M.O., Ge, X., Tsafantakis, E., Dedoussis, G., Zeggini, E., 2019. Very low-depth whole-genome sequencing in complex trait association studies. Bioinformatics 35, 2555–2561. https://doi.org/10.1093/bioinformatics/bty1032

Glenn, T.C., Nilsen, R.A., Kieran, T.J., Sanders, J.G., Bayona-Vásquez, N.J., Finger, J.W., Pierson, T.W., Bentley, K.E., Hoffberg, S.L., Louha, S., Leon, F.J.G.-D., Portilla, M.A. del R., Reed, K.D., Anderson, J.L., Meece, J.K., Aggrey, S.E., Rekaya, R., Alabady, M., Belanger, M., Winker, K., Faircloth, B.C., 2019. Adapterama I: universal stubs and primers for 384 unique dual-indexed or 147,456 combinatorially-indexed Illumina libraries (iTru & iNext). PeerJ 7, e7755. https://doi.org/10.7717/peerj.7755

Godfray, H.C.J., 1994. Parasitoids: behavioral and evolutionary ecology. Princeton University Press.

Gruber, K., Schöning, C., Otte, M., Kinuthia, W., Hasselmann, M., 2013. Distinct subspecies or phenotypic plasticity? Genetic and morphological differentiation of mountain honey bees in East Africa. Ecology and Evolution 3, 3204–3218. https://doi.org/10.1002/ece3.711

Gurevich, A., Saveliev, V., Vyahhi, N., Tesler, G., 2013. QUAST: quality assessment tool for genome assemblies. Bioinformatics 29, 1072–1075. https://doi.org/10.1093/bioinformatics/btt086

Hahn, C., Bachmann, L., Chevreux, B., 2013. Reconstructing mitochondrial genomes directly from genomic next-generation sequencing reads—a baiting and iterative mapping approach. Nucleic Acids Research 41, e129. https://doi.org/10.1093/nar/gkt371

Hall J.C. (1954) A revision of the genus *Lordotus* Loew in North America (Diptera: Bombyliidae). University of California Publications in Entomology. Vol.10, No1, pp.1-34.

Hall J.C., Evenhuis N.L. (1982) Genus *Lordotus* Loew (1863) in: Flies of the Nearctic Region V, Part 13, No.3: 186-238

Hammer, T.J., Janzen, D.H., Hallwachs, W., Jaffe, S.P., Fierer, N., 2017. Caterpillars lack a resident gut microbiome. Proc. Natl. Acad. Sci. U.S.A. 114, 9641–9646. https://doi.org/10.1073/pnas.1707186114

Heraty, J., Parasitoid biodiversity and insect pest management - Insect Biodiversity - Wiley Online Library https://onlinelibrary.wiley.com/doi/abs/10.1002/9781118945568.ch19 (accessed 4.8.23).

Hirakawa, H., Sumitomo, K., Hisamatsu, T., Nagano, S., Shirasawa, K., Higuchi, Y., Kusaba, M., Koshioka, M., Nakano, Y., Yagi, M., Yamaguchi, H., Taniguchi, K., Nakano, M., Isobe, S.N., 2019. De novo whole-genome assembly in *Chrysanthemum seticuspe*, a model species of Chrysanthemums, and its application to genetic and gene discovery analysis. DNA Research 26, 195–203. https://doi.org/10.1093/dnares/dsy048

Homburger, J.R., Neben, C.L., Mishne, G., Zhou, A.Y., Kathiresan, S., Khera, A.V., 2019. Low coverage whole genome sequencing enables accurate assessment of common variants and calculation of genome-wide polygenic scores. Genome Medicine 11, 74. https://doi.org/10.1186/s13073-019-0682-2

Homziak, N.T., Breinholt, J.W., Branham, M.A., Storer, C.G., Kawahara, A.Y., 2019. Anchored hybrid enrichment phylogenomics resolves the backbone of erebine moths. Molecular Phylogenetics and Evolution 131, 99–105. https://doi.org/10.1016/j.ympev.2018.10.038

Huang, X., Li, S., Gao, Y., Zhan, A., 2018. Genome-wide identification, characterization and expression analyses of heat shock protein-related genes in a highly invasive ascidian *Ciona savignyi*. Frontiers in Physiology 9.

Hull, F.M., 1901-, 1973. Bee flies of the world: the genera of the family Bombyliidae. Smithsonian Institution Press.

Iancu, L., Angelescu, I.R., Paun, V.I., Henríquez-Castillo, C., Lavin, P., Purcarea, C., 2021. Microbiome pattern of *Lucilia sericata* (Meigen) (Diptera: Calliphoridae) and feeding substrate in the presence of the foodborne pathogen Salmonella enterica. Sci Rep 11, 15296. https://doi.org/10.1038/s41598-021-94761-w

Ješovnik, A., Sosa-Calvo, J., Lloyd, M.W., Branstetter, M.G., Fernández, F., Schultz, T.R., 2017. Phylogenomic species delimitation and host-symbiont coevolution in the fungus-farming ant genus *Sericomyrmex* Mayr (Hymenoptera: Formicidae): ultraconserved elements (UCEs) resolve a recent radiation. Syst Entomol 42, 523–542. https://doi.org/10.1111/syen.12228

Jing, X.-Y., Li, F.-M., 2020. Identifying heat shock protein families from imbalanced data by using combined features. Computational and Mathematical Methods in Medicine 2020, 1–11. https://doi.org/10.1155/2020/8894478

Johnson, D.E., Johnson, L.M., Notes on the genus *Lordotus* Loew, with descriptions of new species (Diptera: Bombyliidae). The Great Basin Naturalist.

Jose, P.A., Ben-Yosef, M., Jurkevitch, E., Yuval, B., 2019. Symbiotic bacteria affect oviposition behavior in the olive fruit fly *Bactrocera oleae*. Journal of Insect Physiology 117, 103917. https://doi.org/10.1016/j.jinsphys.2019.103917

Kaltenpoth, M., Engl, T., 2014. Defensive microbial symbionts in Hymenoptera. Funct Ecol 28, 315–327. https://doi.org/10.1111/1365-2435.12089

Kalyaanamoorthy, S., Minh, B.Q., Wong, T.K.F., von Haeseler, A., Jermiin, L.S., 2017. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods 14, 587–589. https://doi.org/10.1038/nmeth.4285

Katoh, K., Standley, D.M., MAFFT multiple sequence alignment software version 7: improvements in performance and usability | Molecular Biology and Evolution | Oxford Academic https://academic.oup.com/mbe/article/30/4/772/1073398?login=false (accessed 4.8.23).

Keller, O., Kollmar, M., Stanke, M., Waack, S., 2011. A novel hybrid gene prediction method employing protein multiple sequence alignments. Bioinformatics 27, 757–763. https://doi.org/10.1093/bioinformatics/btr010

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Morlon, H., Ackerly, D.D., Blomberg, S.P., Webb, C.O., 2010. Picante: R tools for integrating phylogenies and ecology. Bioinformatics 26, 1463–1464. https://doi.org/10.1093/bioinformatics/btq166

Kers, J.G., Saccenti, E., 2022. The power of microbiome studies: some considerations on which alpha and beta metrics to use and how to report results. Frontiers in Microbiology 12.

Krebs, R.A., Feder, M.E., 1997. Natural variation in the expression of the heat-shock protein hsp70 in a population of *drosophila melanogaster* and its correlation with tolerance of ecologically relevant thermal stress. Evolution 51, 173–179. https://doi.org/10.1111/j.1558-5646.1997.tb02398.x

Krueger, F., 2023. Trim Galore. https://github.com/FelixKrueger/TrimGalore.

Kwong, W.K., Engel, P., Koch, H., Moran, N.A., 2014. Genomics and host specialization of honey bee and bumble bee gut symbionts. Proc. Natl. Acad. Sci. U.S.A. 111, 11509–11514. https://doi.org/10.1073/pnas.1405838111

Kwong, W.K., Moran, N.A., 2015. Evolution of host specialization in gut microbes: the bee gut as a model. Gut Microbes 6, 214–220. https://doi.org/10.1080/19490976.2015.1047129

Lange, V., Böhme, I., Hofmann, J., Lang, K., Sauter, J., Schöne, B., Paul, P., Albrecht, V., Andreas, J.M., Baier, D.M., Nething, J., Ehninger, U., Schwarzelt, C., Pingel, J., Ehninger, G., Schmidt, A.H., 2014. Cost-efficient high-throughput HLA typing by MiSeq amplicon sequencing. BMC Genomics 15, 63. https://doi.org/10.1186/1471-2164-15-63

Larsen, A.M., Bullard, S.A., Womble, M., Arias, C.R., 2015. Community structure of skin microbiome of gulf killifish, *Fundulus grandis*, is driven by seasonality and not exposure to oiled sediments in a Louisiana salt marsh. Microb Ecol 70, 534–544. https://doi.org/10.1007/s00248-015-0578-7

Larson, B.M.H., Kevan, P.G., Inouye, D.W., 2001. Flies and flowers: taxonomic diversity of anthophiles and pollinators. Can Entomol 133, 439–465. https://doi.org/10.4039/Ent133439-4

Lemmon, A.R., Emme, S.A., Lemmon, E.M., 2012. Anchored hybrid enrichment for massively high-throughput phylogenomics. Systematic Biology 61, 727–744. https://doi.org/10.1093/sysbio/sys049

Leung, T.L.F., Poulin, R., 2008. Parasitism, commensalism, and mutualism: exploring the many shades of symbioses. Vie Milieu.

Li, M.-Y., Huang, Y., Lei, X., Xu, C.-T., Li, B., Chen, D.-X., Liu, S., 2021. Identification of six heat shock protein 70 genes in *Lasioderma serricorne* (Coleoptera: Anobiidae) and their responses to temperature stress. Journal of Asia-Pacific Entomology 24, 597–605. https://doi.org/10.1016/j.aspen.2021.05.005

Li, X., Teasdale, L.C., Bayless, K.M., Ellis, A.G., Wiegmann, B.M., Lamas, C.J.E., Lambkin, C.L., Evenhuis, N.L., Nicholls, J.A., Hartley, D., Shin, S., Trautwein, M., Zwick, A., Lessard, B.D., Yeates, D.K., 2021. Phylogenomics reveals accelerated late Cretaceous diversification of bee flies (Diptera: Bombyliidae). Cladistics 37, 276–297. https://doi.org/10.1111/cla.12436

Li, X., Wiens, J.J., 2022. Estimating global biodiversity: the role of cryptic insect species. Systematic Biology syac069. https://doi.org/10.1093/sysbio/syac069

Loew, H., 1863. Diptera Americae septentrionalis indigena. Mitt. Mus. Nat.kd. Berl., Dtsch entomol. 7, 1–55. https://doi.org/10.1002/mmnd.18630070104

Longo, S.J., Faircloth, B.C., Meyer, A., Westneat, M.W., Alfaro, M.E., Wainwright, P.C., 2017. Phylogenomic analysis of a rapid radiation of misfit fishes (Syngnathiformes) using ultraconserved elements. Molecular Phylogenetics and Evolution 113, 33–48. https://doi.org/10.1016/j.ympev.2017.05.002

Love, M.I., Huber, W., Anders, S., 2014. Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biology 15, 550. https://doi.org/10.1186/s13059-014-0550-8

Marshall, S. A., 2012 Flies: the natural history & diversity of Diptera. No. 595.77 M3.

Marshall, S.A. and Kirk-Spriggs, A.H., 2017. Natural history of Diptera. In: A.H. Kirk-Spriggs and B.J. Sinclair (eds.), Manual of Afrotropical Diptera, vol. 1. Introductory chapters and keys to Diptera families, pp. 135–152. Suricata 4, SANBI, Pretoria.

Martínez-Falcón, A.P., Durbán, A., Latorre, A., Antón, J., Marcos-García, M. de los Á., 2011. Bacteria associated with *Copestylum* (Diptera, Syrphidae) larvae and their cactus host *Isolatocereus dumortieri*. PLOS ONE 6, e27443. https://doi.org/10.1371/journal.pone.0027443

Martinson, J.N.V., Pinkham, N.V., Peters, G.W., Cho, H., Heng, J., Rauch, M., Broadaway, S.C., Walk, S.T., 2019. Rethinking gut microbiome residency and the Enterobacteriaceae in healthy human adults. ISME J 13, 2306–2318. https://doi.org/10.1038/s41396-019-0435-7

Martinson, V.G., Danforth, B.N., Minckley, R.L., Rueppell, O., Tingek, S., Moran, N.A., 2011. A simple and distinctive microbiota associated with honey bees and bumble bees. Molecular Ecology 20, 619–628. https://doi.org/10.1111/j.1365-294X.2010.04959.x

McCabe, R.A., Receveur, J.P., Houtz, J.L., Thomas, K.L., Benbow, M.E., Pechal, J.L., Wallace, J.R., 2020. Characterizing the microbiome of ectoparasitic louse flies feeding on migratory raptors. PLOS ONE 15, e0234050. https://doi.org/10.1371/journal.pone.0234050

McCormack, J.E., Hird, S.M., Zellmer, A.J., Carstens, B.C., Brumfield, R.T., 2013. Applications of next-generation sequencing to phylogeography and phylogenetics. Molecular Phylogenetics and Evolution, Morris Goodman Memorial Symposium 66, 526–538. https://doi.org/10.1016/j.ympev.2011.12.007

McCormack, J.E., Tsai, W.L.E., Faircloth, B.C., 2016. Sequence capture of ultraconserved elements from bird museum specimens. Mol Ecol Resour 16, 1189–1203. https://doi.org/10.1111/1755-0998.12466

McMurdie, P.J., Holmes, S., 2014. Waste not, want not: why rarefying microbiome data is inadmissible. PLOS Computational Biology 10, e1003531. https://doi.org/10.1371/journal.pcbi.1003531

McMurdie, P.J., Holmes, S., 2013. phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8, e61217. https://doi.org/10.1371/journal.pone.0061217

Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D., Gurevich, A., 2018. Versatile genome assembly evaluation with QUAST-LG. Bioinformatics 34, i142–i150. https://doi.org/10.1093/bioinformatics/bty266

Mikheenko, A., Valin, G., Prjibelski, A., Saveliev, V., Gurevich, A., 2016. Icarus: visualizer for de novo assembly evaluation. Bioinformatics 32, 3321–3323. https://doi.org/10.1093/bioinformatics/btw379

Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., Lanfear, R., 2020. IQ-TREE 2: New models and efficient methods for phylogenetic inference in the genomic era. Molecular Biology and Evolution 37, 1530–1534. https://doi.org/10.1093/molbev/msaa015

Morrow, J.L., Frommer, M., Shearman, D.C.A., Riegler, M., 2015. The microbiome of field-caught and laboratory-adapted Australian tephritid fruit fly species with different host plant use and specialisation. Microb Ecol 70, 498–508. https://doi.org/10.1007/s00248-015-0571-1

Newman, C.E., Austin, C.C., 2016. Sequence capture and next-generation sequencing of ultraconserved elements in a large-genome salamander. Molecular Ecology 25, 6162–6174. https://doi.org/10.1111/mec.13909

Novakova, E., Woodhams, D.C., Rodríguez-Ruano, S.M., Brucker, R.M., Leff, J.W., Maharaj, A., Amir, A., Knight, R., Scott, J., 2017. Mosquito microbiome dynamics, a background for prevalence and seasonality of west nile virus. Frontiers in Microbiology 8.

Nurk, S., Meleshko, D., Korobeynikov, A., Pevzner, P.A., 2017. metaSPAdes: a new versatile metagenomic assembler. Genome Res. 27, 824–834. https://doi.org/10.1101/gr.213959.116

Oliveira, T.B. de, Ferro, M., Bacci, M., Souza, D.J. de, Fontana, R., Delabie, J.H.C., Silva, A., 2016. Bacterial communities in the midgut of ponerine ants (Hymenoptera: Formicidae: Ponerinae). Sociobiology 63, 637. https://doi.org/10.13102/sociobiology.v63i1.882

Olofsson, J.K., Cantera, I., Van de Paer, C., Hong-Wa, C., Zedane, L., Dunning, L.T., Alberti, A., Christin, P.-A., Besnard, G., 2019. Phylogenomics using low-depth whole genome sequencing: A case study with the olive tribe. Molecular Ecology Resources 19, 877–892. https://doi.org/10.1111/1755-0998.13016

Ortiz, M.F., Wallau, G.L., Graichen, D.Â.S., Loreto, E.L.S., 2015. An evaluation of the ecological relationship between Drosophila species and their parasitoid wasps as an opportunity for horizontal transposon transfer. Mol Genet Genomics 290, 67–78. https://doi.org/10.1007/s00438-014-0900-y

Pal, S., Karmakar, P., Symbionts associated with insect digestive system and their role in insect nutrition. Journal of Entomology and Zoology Studies.

Paradis, E., Schliep, K., 2019. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. Bioinformatics 35, 526–528. https://doi.org/10.1093/bioinformatics/bty633

Park, J.C., Kim, D.-H., Lee, Y., Lee, M.-C., Kim, T.K., Yim, J.H., Lee, J.-S., 2020. Genome-wide identification and structural analysis of heat shock protein gene families in the marine rotifer *Brachionus* spp.: Potential application in molecular ecotoxicology. Comparative Biochemistry and Physiology Part D: Genomics and Proteomics 36, 100749. https://doi.org/10.1016/j.cbd.2020.100749

Paysan-Lafosse, T., Blum, M., Chuguransky, S., Grego, T., Pinto, B.L., Salazar, G.A., Bileschi, M.L., Bork, P., Bridge, A., Colwell, L., Gough, J., Haft, D.H., Letunić, I., Marchler-Bauer, A., Mi, H., Natale, D.A., Orengo, C.A., Pandurangan, A.P., Rivoire, C., Sigrist, C.J.A., Sillitoe, I., Thanki, N., Thomas, P.D., Tosatto, S.C.E., Wu, C.H., Bateman, A., 2023. InterPro in 2022. Nucleic Acids Research 51, D418–D427. https://doi.org/10.1093/nar/gkac993

Portik, D.M., Wiens, J.J., 2021. Do alignment and trimming methods matter for phylogenomic (UCE) analyses? Systematic Biology 70, 440–462. https://doi.org/10.1093/sysbio/syaa064

Provorov, N.A., Onishchuk, O.P., 2018. Microbial Symbionts of Insects: Genetic Organization, Adaptive Role, and Evolution. Microbiology 87, 151–163. https://doi.org/10.1134/S002626171802011X

Raxworthy, C.J., Smith, B.T., 2021. Mining museums for historical DNA: advances and challenges in museomics. Trends in Ecology & Evolution 36, 1049–1060. https://doi.org/10.1016/j.tree.2021.07.009

Renaud, G., 2023. deML: Maximum likelihood demultiplexing for NGS data.

Rhie, A., McCarthy, S.A., Fedrigo, O., Damas, J., Formenti, G., Koren, S., Uliano-Silva, M., et al., 2021. Towards complete and error-free genome assemblies of all vertebrate species. Nature 592, 737–746. https://doi.org/10.1038/s41586-021-03451-0

Rohland, N., Reich, D., 2012. Cost-effective, high-throughput DNA sequencing libraries for multiplexed target capture. Genome Res. 22, 939–946. https://doi.org/10.1101/gr.128124.111

Rustagi, N., Zhou, A., Watkins, W.S., Gedvilaite, E., Wang, S., Ramesh, N., Muzny, D., Gibbs, R.A., Jorde, L.B., Yu, F., Xing, J., 2017. Extremely low-coverage whole genome sequencing in South Asians captures population genomics information. BMC Genomics 18, 396. https://doi.org/10.1186/s12864-017-3767-6

Sapountzis, P., Zhukova, M., Hansen, L.H., Sørensen, S.J., Schiøtt, M., Boomsma, J.J., 2015. *Acromyrmex* leaf-cutting ants have simple gut microbiota with nitrogen-fixing potential. Appl Environ Microbiol 81, 5527–5537. https://doi.org/10.1128/AEM.00961-15

Seemann, T., 2022. VelvetOptimiser: automate your Velvet assemblies.

Seemüller, E., Garnier, M., Schneider, B., 2002. Mycoplasmas of plants and insects, in: Razin, S., Herrmann, R. (Eds.), Molecular Biology and Pathogenicity of Mycoplasmas. Springer US, Boston, MA, pp. 91–115. https://doi.org/10.1007/0-306-47606-1_5

Shapira, M., 2016. Gut microbiotas and host evolution: scaling up symbiosis. Trends in Ecology & Evolution 31, 539–549. https://doi.org/10.1016/j.tree.2016.03.006

Shropshire, J.D., Bordenstein, S.R., 2016. Speciation by symbiosis: the microbiome and behavior. mBio 7, e01785-15. https://doi.org/10.1128/mBio.01785-15

Silver, A., Perez, S., Gee, M., Xu, B., Garg, S., Will, K., Gill, A., 2021. Persistence of the ground beetle (Coleoptera: Carabidae) microbiome to diet manipulation. PLOS ONE 16, e0241529. https://doi.org/10.1371/journal.pone.0241529

Sivan, A., Shriram, A.N., Muruganandam, N., Thamizhmani, R., 2017. Expression of heat shock proteins (HSPs) in *Aedes aegypti* (L) and *Aedes albopictus* (Skuse) (Diptera: Culicidae) larvae in response to thermal stress. Acta Tropica 167, 121–127. https://doi.org/10.1016/j.actatropica.2016.12.017

Smith, B.T., Harvey, M.G., Faircloth, B.C., Glenn, T.C., Brumfield, R.T., 2014. Target capture and massively parallel sequencing of ultraconserved elements for comparative studies at shallow evolutionary time scales. Systematic Biology 63, 83–95. https://doi.org/10.1093/sysbio/syt061

Stanke, M., Diekhans, M., Baertsch, R., Haussler, D., 2008. Using native and syntenically mapped cDNA alignments to improve de novo gene finding. Bioinformatics 24, 637–644. https://doi.org/10.1093/bioinformatics/btn013

Stanke, M., Waack, S., 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics 19, ii215–ii225. https://doi.org/10.1093/bioinformatics/btg1080

Starrett, J., Waters, E.R., 2007. Positive natural selection has driven the evolution of the Hsp70s in Diguetia spiders. Biology Letters 3, 439–444. https://doi.org/10.1098/rsbl.2007.0159

Tomberlin, J.K., Crippen, T.L., Tarone, A.M., Chaudhury, M.F.B., Singh, B., Cammack, J.A., Meisel, R.P., 2017. A review of bacterial interactions with blow flies (Diptera:

Calliphoridae) of Medical, Veterinary, and Forensic Importance. Ann Entomol Soc Am 110, 19–36. https://doi.org/10.1093/aesa/saw086

Valderrama, E., Landis, J.B., Skinner, D., Maas, P.J.M., Maas-van de Kramer, H., André, T., Grunder, N., Sass, C., Pinilla-Vargas, M., Guan, C.J., Phillips, H.R., Almeida, A.M.R. de, Specht, C.D., 2022. The genetic mechanisms underlying the convergent evolution of pollination syndromes in the Neotropical radiation of *Costus* L. Front. Plant Sci. 13, 874322. https://doi.org/10.3389/fpls.2022.874322

Vavre, F., Kremer, N., 2014. Microbial impacts on insect evolutionary diversification: from patterns to mechanisms. Current Opinion in Insect Science, Molecular physiology/Environmental physiology 4, 29–34. https://doi.org/10.1016/j.cois.2014.08.003

Venu, I., Durisko, Z., Xu, J., Dukas, R., 2014. Social attraction mediated by fruit flies' microbiome. Journal of Experimental Biology 217, 1346–1352. https://doi.org/10.1242/jeb.099648

Wang, X.-R., Wang, C., Ban, F.-X., Zhu, D.-T., Liu, S.-S., Wang, X.-W., 2019. Genome-wide identification and characterization of HSP gene superfamily in whitefly (*Bemisia tabaci*) and expression profiling analysis under temperature stress. Insect Science 26, 44–57. https://doi.org/10.1111/1744-7917.12505

Wang, Y., Iii, T.M.G., Kukutla, P., Yan, G., Xu, J., 2011. Dynamic gut microbiome across life history of the malaria mosquito *Anopheles gambiae* in Kenya. PLOS ONE 6, e24767. https://doi.org/10.1371/journal.pone.0024767

Waters, E.R., 2014. Conservative innovation: The mixed-up evolutionary history of the heat-shock proteins. The Biochemist 36, 9–14. https://doi.org/10.1042/BIO03601009

Wen, D., Laursen, R.A., 1993. Structure-function relationships in an antifreeze polypeptide. The effect of added bulky groups on activity. Journal of Biological Chemistry 268, 16401–16405. https://doi.org/10.1016/S0021-9258(19)85434-0

Whitfield, J.B., Kjer, K.M., 2008. Ancient rapid radiations of insects: Challenges for phylogenetic analysis. Annu. Rev. Entomol. 53, 449–472. https://doi.org/10.1146/annurev.ento.53.103106.093304

Wiens, J.J., 2007. Species delimitation: New approaches for discovering diversity. Systematic Biology 56, 875–878. https://doi.org/10.1080/10635150701748506

Yeates, D.K., Greathead, D., 1997. The evolutionary pattern of host use in the Bombyliidae (Diptera): a diverse family of parasitoid flies. Biological Journal of the Linnean Society 60, 149–185. https://doi.org/10.1111/j.1095-8312.1997.tb01490.x

Zchori-Fein, E., Gottlieb, Y., Kelly, S.E., Brown, J.K., Wilson, J.M., Karr, T.L., Hunter, M.S., 2001. A newly discovered bacterium associated with parthenogenesis and a change in

host selection behavior in parasitoid wasps. Proceedings of the National Academy of Sciences 98, 12555–12560. https://doi.org/10.1073/pnas.221467498

Zerbino, D.R., Birney, E., 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res 18, 821–829. https://doi.org/10.1101/gr.074492.107

Zerbino, D.R., McEwen, G.K., Margulies, E.H., Birney, E., 2009. Pebble and rock band: heuristic resolution of repeats and scaffolding in the velvet short-read de novo assembler. PLOS ONE 4, e8407. https://doi.org/10.1371/journal.pone.0008407

Zhang, F., Ding, Y., Zhu, C.-D., Zhou, X., Orr, M.C., Scheu, S., Luan, Y.-X., 2019. Phylogenomics from low-coverage whole-genome sequencing. Methods in Ecology and Evolution 10, 507–517. https://doi.org/10.1111/2041-210X.13145

Zheng, X., Zhu, Q., Zhou, Z., Wu, F., Chen, L., Cao, Q., Shi, F., 2021. Gut bacterial communities across 12 *Ensifera* (Orthoptera) at different feeding habits and its prediction for the insect with contrasting feeding habits. PLOS ONE 16, e0250675. https://doi.org/10.1371/journal.pone.0250675

Appendix A. UCE Protocol for Chapter 2

Initial notes

Before beginning, please read the entire KAPA Hyper Prep Kit protocol that this document is based on

- What is true for all library preparation methods is just as true here: the vast majority of researcher time and effort should go into obtaining the best quality DNA fragmented to the optimal range. Take your time to get it right here and it will pay off with easier library preparation and better quality data.

- This protocol assumes enough starting material to begin sonication with 1 µg of DNA. However, the Kapa kit is suitable for library construction from 1 ng – 1 µg of double-stranded DNA, so in cases where 1 µg in unrealistic, this same protocol can be used with increased numbers of indexing PCR cycles.

- A refinement of size selection is possible after adapter ligation but it is easiest to get fragments to approximately the correct range before the enzymatic steps of library prep begin.

- The optimal size range of library inserts will vary based on whether data will be de novo assembled, mapped to a reference, or captured. Please consult with the lab manager before making fragmentation decisions.

- SPRI beads and PEG/NaCl solution should only be used at room temperature.

- Fresh 80% ethanol should be diluted each day.

- Always ensure that KAPA Hyper Prep Kit components have been fully thawed and thoroughly mixed before use. The KAPA Hyper Prep End Repair & A-tailing Buffer and Ligation Buffer may contain precipitates when thawed at 4°C. These buffers must be thawed at room temperature and vortexed thoroughly before use.

- Reaction master mixes prepared from the enzymes and buffers for end repair and A-tailing, as well as for ligation, are viscous and require special attention during pipetting. Keep all enzyme components and master mixes on ice as long as possible during handling and preparation.

- Always thaw the adapter stub on ice. Do not warm in hands or leave at room temperature for long periods.

- Safe stopping points: The protocol can be paused after sonication, post-sonication size selection, and post-ligation bead clean-up. <u>Do not stop after end repair.</u>

Sonication

Before using the qSonica instrument for the first time, consult with the lab manager for proper training.

1) In a qSonica 0.2 mL tube, dilute a 100µL aliquot of each sample to 10 ng/µL (by qubit) or 15 ng/µL (by nanodrop) using 1x LTE (10mM Tris, 0.1 mM EDTA), 1x TE or EB(T) (10mM Tris with or without 0.05% tween). Make sure tubes are completely closed. Spin down samples to keep all liquid at the bottom. Keep samples very cold before sonicating. (It is okay to freeze them; then they can be thawed just before beginning sonication.)

[Note: this protocol can use less starting material if 1 µg is not available. It is best to keep the mass of material within each 18 sample batch as similar as possible. The sonication tube volume should always be identical within each batch)

2) Ensure that all tubing is properly connected. Add ~1.5 L of cold deionized water to the qSonica bath compartment (if empty). Turn on the water cooler and wait 5-10 to allow the system to cool to 4°C. (If the bath is filled with room temperature water, it will take 15-20 minutes to cool.)

3) Turn the water adjustment dial to "-" to add water to the reservoir in back. Fill this to ~50% full. The water level inside the bath should be ~2 cm above the titanium horn. If not, add more water to achieve this level.

4) Turn on the power supply ("|" icon on the top right). Ensure that the sonication bath is empty and that the cabinet door is latched closed. Select the "degas" program which will run for 10 minutes.

5) While degassing, load your samples into the blue 18-place tube holder. (If you have < 18 tubes to sonicate, fill empty spaces with blank tubes containing water.) Cover with the white donut and screw on the top section. Leave the assembly on ice or at +4C until ready for use.

6) Sonication times will vary by DNA size, genome size, tube volume, and extraction method. The following is just an example protocol. Before you begin, please consult with the lab manager for the best trial conditions to select based on other recent results. Additional user validated protocols can be found here: https://www.sonicator.com/pages/publications-and-protocols-chromatin-dna-shearing

(Note: in all cases qSonica and users report the "Total Sonication ON Time". A protocol using a pulse such as 15s on/15s off will take twice as long to complete)

7) After degrassing, use the setting on the power supply to set the sonication conditions. Example:
   Timer = 2:00 (total sonication on time)
   Pulse = 15s ON / 15s OFF
   Amplitude = 40%

8) Attach the sample rack to the lid of the sonicator bath. Make sure the lid is plugged in.

9) Assess the bath water level by eye and adjust using the water adjustment dial if it is too high or too low. The water level in the bath should match that of the sample tubes as closely as possible. However, if they are not a perfect match it is better if the water level of the bath is slightly below the water level of the sample tubes. (Otherwise excessive splashing may occur.)

10) Close the cabinet door and use the red start/stop button to start the sonication run. Check that you see only minor splashing during the first ON cycle. (A slight misting is fine, but large droplets are not.) You can adjust the water level while sonication is proceeding if small adjustments need to be made.

11) Halfway through then sonication protocol, stop sonication, remove the tubes rotator from the instrument, and vortex the tubes (easiest to do while they are still in the rotator.) Then remove the tubes from the blue 18-place tube holder and spin down any splashing in your tubes. This will result in more even sonication and very little residual HMW DNA. (Note: this means that if you actually want 5:00 minutes of ON time for sonication, set the instrument for 2:30 and run this program twice.)

12) Run the sonicator program again. You may have to make small adjustments to the water level again.

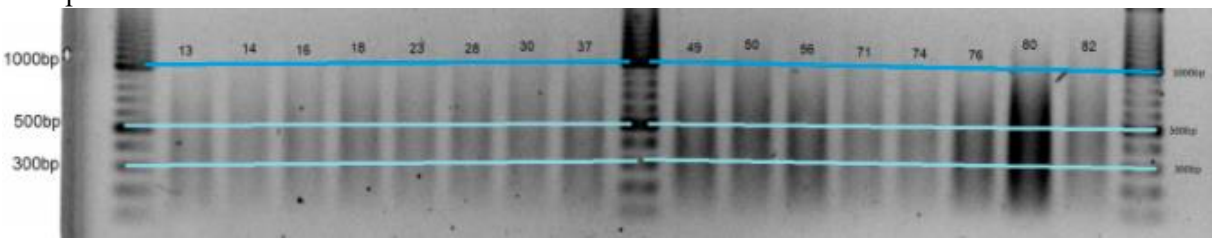
13) After sonication is complete, open the cabinet and remove the sample rack. Spin down tubes and either proceed directly to bead cleaning or store frozen until ready to proceed.

14) When optimizing, after sonicating, take an aliquot (~5 µL) of sample to run on an agarose gel to assess the sizing pattern. This is not necessary once the sonication parameters have been worked out for your project. If you assess some samples in a set but not others, spike in the same amount of liquid you remove (EB or water) to keep the volumes equal.

15) Tips for gel electrophoresis:

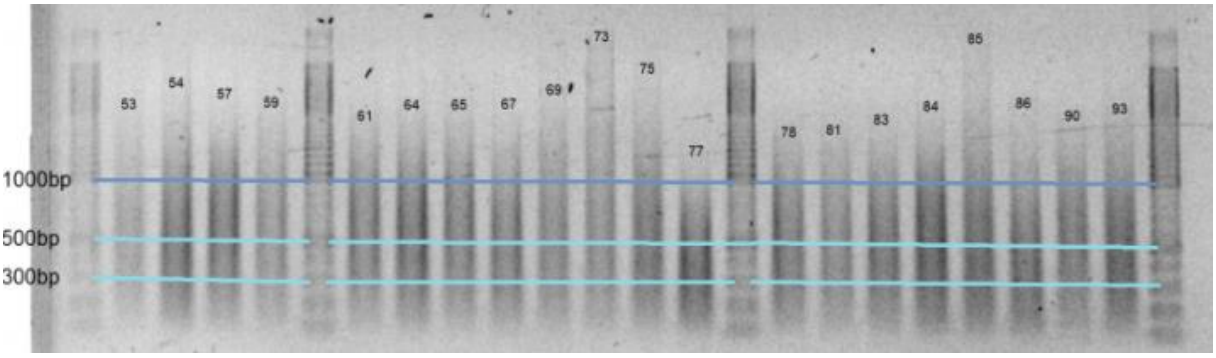
- Pour a 1.5-2.0% gel to better visualize smaller fragments
- Use fresh running buffer rather than recycled (which may contribute background which may be difficult to distinguish from your sonicated material
- Flank your sonicated samples on both sides of the row with ladder (instead of just using one). This will make it easier to assess the samples in the middle of the row.

16) With most library types, an ideal sonication result is to have the brightest part of the smear overlap with the 300-500 bp range and to have only trace amounts of material greater than 1000 bp. A good example:



In this case further sonication would move some of the 500-1000 bp pieces into the 300-500 bp window, but will also make some of the currently 300-500 bp pieces too small. So, in this example, it is better to just proceed rather than to spend too much time on additional sonication and gel checks. The double-sided bead cleaning will further refine the sizing.

However, in a case like this, it would be best to give most of these samples 1-2 minutes of further sonication time due to how much material is > 500 bp:



17) Once general conditions are worked out for a project, continue with the next set of tubes to sonicate. When you are done for the day, turn off the chiller pump and use the "0" button on the power supply to turn it off. Consult with the lab manager about when and how to empty the water from the bath.

18) Optional: assess all project samples on a gel. (note: only those starting with 500 ng+ will be easily visible.) Add extra sonication time to any that are outliers.

19) Completed samples may be stored at -20C while sonication for a project is on-going.

20) When all samples in set are sonicated to the desired size, spin down and use a multichannel to transfer the same volume from each to a plate or set of strip tubes. 100 µL is ideal but other volumes are fine so long as all samples are the same volume. If any are slightly lower than the others, spike-in water or EB to make all the volumes equal before double-sided selection.

Note: when storing sonication tubes before or after sonication, keep them in a box or rack with a lid. Then store this box/rack in a plastic baggie. That way if it is dropped, all your tubes will stick together and not scatter to the far corners of the lab.

Post-Sonication Double-sided Size Selection

The following protocol has been used successfully for a double-sided selection centered around 350bp, using 0.5x for right-side selection and 0.65x for left-side. However, the best bead ratio will be partially dependent on the shearing profile of your samples. Initial testing of size selection ratios on non-essential samples (or with special ladder) is recommended before beginning.

This protocol follows the notation of an Rx/Lx double-sided clean-up, where R is the right-side ratio and L is the left-side ratio.

1) Resuspend your aliquot of **low-ratio formula SPRI** (Sera-Mag) bead suspension by inverting or vortexing. Take the bead tube out to warm up to room temperature (< 30 min)

2) Standardize all reactions/extractions to the same volume (V) by adding water to your samples if necessary

3) Add SPRI bead suspension to samples as follows reaction to start the right-side clean-up:
   A) Add an **R-fold volume** of SPRI bead suspension to each (ex: if V=100µL and R=0.5, add 50µL low-ratio formula SPRI beads to each sample).

   B) Pipette up and down 10 times to mix well and/or seal the tubes and vortex gently. Ensure that the beads are resuspended and homogenized in the liquid.

   C) Let the tubes incubate for **15 minutes** at room temperature.

   D) While the samples are incubating, label a new set of 0.2mL strip tubes for the supernatant in step 5.

4) Briefly (0.5s) spin down the contents of the tubes. Place them in a magnetic plate or stand and let sit for **3 minutes** (or until the supernatant is clear) to separate the beads from solution.

*****Reminder: when you remove the supernatant from the beads in step 5, keep the liquid. Do not discard it!*****

5) Carefully pipette the supernatant without removing or disturbing the beads and **move it to the new, empty strip tubes. Be very careful not to carry over any beads at this step.** (The tubes with the beads can then be discarded.)

6) Add **low-ratio formula** SPRI (Sera-Mag) bead suspension to the sample in the new tube as follows to start the left-side clean-up:
   A) If the second ratio is L, add an **L - R -fold volume** of low-ratio formula SPRI bead suspension to each reaction based on the original volume

   (ex: if the right-side ratio (R) is 0.5x and the left-side ratio (L) is 0.65x, add 0.15 * 100 µL or 15 µL).

   B) Pipette up and down 10 times to mix well and/or seal the tubes and vortex gently. Ensure that the beads are resuspended and homogenized in the liquid.

   C) Let the tubes incubate for **15 minutes** at room temperature.

7) Briefly (0.5s) spin down the contents of the tubes. Place them in a magnetic plate or stand and let sit for **3 minutes** (or until the supernatant is clear) to separate the beads from solution.

8) Pipette off the supernatant and discard without removing or disturbing the beads

9) Leave beads on magnet and wash with **200 µL** of freshly prepared **80% ethanol** (make a new ethanol dilution before every SPRI clean-up).
Let stand for **at least 30 seconds** and discard supernatant.

10) Repeat step 9.

11) Then use a small volume pipette (such as Rainin 20µL LTS pipette) to remove as much residual ethanol as possible without disturbing the beads. A toothpick can be used to soak up alcohol spots. Let the beads air-dry for **3-5 minutes** at room temperature without caps. (For large sets of samples, you may be able to resuspend the first tubes as soon as you are done removing residual ethanol from the final ones.) Avoid overdrying which appears as cracking in the beads

12) Elute as follows:

A) Remove the tubes from the magnetic rack. Add **12.5 µL** of **EB** to the wells

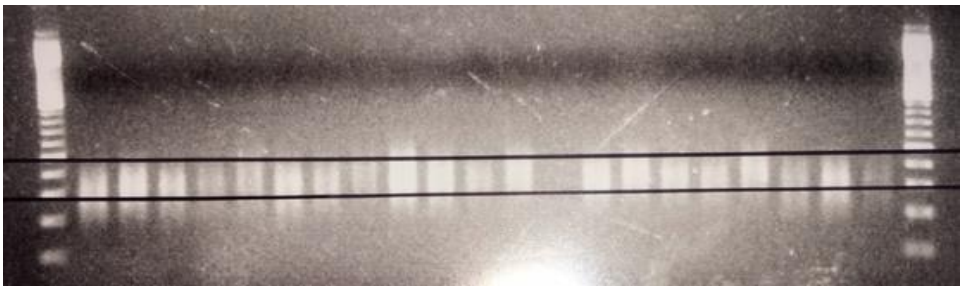
B) Pipette up and down 10 times to mix well and/or seal the tubes and vortex gently. Ensure that the beads are resuspended and homogenized in the liquid.

C) Let the tubes incubate for **10 minutes** at room temperature.

13) Briefly (0.5s) spin down the contents of the tubes. Samples can stay in the bead solution if you do not need to assess them before continuing to end repair.

14) Optional: if concentration assessment is required, place the plate back on the magnetic rack, let stand for **3 minutes** (or until the supernatant is clear) to separate the beads from solution. Then remove 1 µL of clear DNA solution to qubit these samples to know how much DNA is available as input material for End Repair and A-Tailing

15) Optional if sufficient material is available: run an agarose gel with 1 µL of sample in order to check that sizing is correct. (Here an insert size of 300-500 bp was targeted)



17) If optional assessment steps were performed, add water or EB to return the volume of the sample/bead solution to 12.5 µL

End Repair and A-tailing

1) While the double-sided selection is in progress, remove the End Repair and A-Tailing reagents from the yellow Kapa kit on the top shelf of the 4122 freezer. Keep the enzyme on ice, and let the buffer thaw at room temperature.

2) Make a master mix composed of:

End Repair & A-Tailing Buffer:	1.75 µL
End Repair & A-Tailing Enzyme:	0.75 µL

Gently vortex and briefly spin before aliquotting. The master mix can be distributed into a strip tube in order to facilitate adding to samples with a multichannel pipette.

3) Add **2.5 µL of ERAT master mix** to each sample well to make a 15 µL reaction.

4) When adding the ERAT master mix, pipette each well up and down ~2-3 times to mix the eluted bead solution and the ERAT master mix. Pipette gently to avoid introducing excessive bubbles.

5) Seal the tubes, gently vortex*, and briefly spin down in a centrifuge (< 30 s).

> **Note: to gently vortex, start with the vortex on the lowest setting. Then gradually increase the speed until the liquid starts to dance around and mix but doesn't tornado violently or move towards the top of the well.*

6) Incubate* in a thermocycler programmed as outlined below:

20°C: 30 minutes

65°C: 30 minutes

4°C: hold

*A heated lid is required for this incubation. For PTC-200 cyclers, set the temperature of the lid at tracking 10 °C above the block. For all others, set the lid temperature to 75°C

7) Return end repair enzymes to the yellow Kapa gDNA box and retrieve the Ligation Enzyme, adapter stub, and Ligase Buffer. Leave the enzyme and stub on ice, and let the buffer thaw at room temperature.

8) After the End-repair and A-tailing reaction is completed and the sample has cooled to 4°C, proceed to Adapter Ligation

Adapter Ligation

1) Make a master mix composed of the following and keep it on ice:

Ligation Buffer:	7.5 µL
Ligation Enzyme:	2.5 µL
Nuclease-free water	3.5 µL
*Adapter stub (50 µM)	1.5 µL

**Note: do not add the adapter stub to the ligation master mix until immediately before it will be aliquoted. (The ligase will start to make dimer as soon as the adapter stub is introduced into the master mix, even when kept on ice.)*

The master mix can be distributed into a strip tube to facilitate adding to samples with a multichannel pipette

2) Add **15 µL of Ligation master mix** to each sample well to make a 30 µL reaction.

3) When adding the ligation master mix, pipette each well up and down ~2-3 times to mix the ERAT reaction and the ligation master mix. Pipette gently to avoid introducing excessive bubbles.

4) Seal the tubes, gently vortex (just so that the liquid starts to move around), and briefly spin down in a centrifuge (< 30 s).

Note: this is just to mix the eluted DNA and the enzyme. The beads do not need to be resuspended for the ligation reaction. They can remain at the bottom of the tube if they have settled there.

5) Incubate at 20°C for 15-60 min. or overnight at 4°C.

Note: The overnight incubation may result in higher ligation efficiency, but it may also result in more adapter dimers. However, the 0.67X bead cleaning should do a good job removing most dimer. Overnight ligations are recommended when starting with less sonication input material (< 500 ng), but can also be used when the timing is better to proceed with the Post-Libation Bead Clean-up the next day.

6) Proceed to the next step. (Post-Ligation Bead Clean-up)

Post-Ligation Bead Clean-up

1) Before beginning the bead clean-up, warm a tube of PEG/NaCl solution to room temperature for < 30 minutes.

2) In the same tubes as the ligation reaction, perform a 0.67X bead-based cleanup by combining the following:

Adapter ligation reaction product: 30 µL

PEG/NaCl solution at room temperature: **20 µL**

*Note: if you removed the beads from your samples after double-sided cleaning, you will need to add beads again at this step. In place of PEG/NaCl solution, use 20 µL of room-temperature **low-ratio SPRI beads**)*

3) Mix thoroughly by gently vortexing and/or pipetting up and down multiple times.

4) Incubate the tube at room temperature for **15 minutes** to bind DNA to the beads.

5) Place the tubes on a magnet to capture the beads. Incubate **5 minutes** or until the liquid is clear.

6) Carefully remove and discard the supernatant.

7) Keeping the tubes on the magnet, add **200 µL** of freshly prepared **80% ethanol**.

8) Incubate the tubes on the magnet at room temperature for **30 seconds**

9) Carefully remove and discard the ethanol.

10) Keeping the tubes on the magnet, add **200 µL** of freshly prepared **80% ethanol**.

11) Incubate the tubes on the magnet at room temperature for **30 seconds**

12) Carefully remove and discard the ethanol. Try to remove all residual ethanol without disturbing the beads. To accomplish this, use a small volume (10µL) tip to remove ethanol remaining at the bottom of the tube. Then use wooden toothpicks to blot up any dots of ethanol on the sides of the tubes.

13) Dry the beads at room temperature for **3 minutes or less**, until all of the residual ethanol has evaporated from the tube but the beads themselves are still damp and shiny. *Caution: over-drying the beads may result in reduced yield.*

14) Remove the tubes from the magnet. Then thoroughly resuspend the beads: in **22 µL** of elution buffer (10 mM Tris-HCl, pH 8.0 – 8.5) a.k.a. Buffer EB or EBT

15) Incubate the tubes at room temperature for **5-10 minutes** to elute DNA off the beads.

16) Place the tubes on a magnet to capture the beads. Incubate for **3 minutes** or until the liquid is clear.

17) Transfer **20** µL of the clear supernatant to a new set of tubes and seal.

SAFE STOPPING POINT: samples can be stored here until ready to proceed to indexing PCR

Indexing PCR

Before starting indexing PCR, make a detailed plan of which sample will be matched with which well of the indexing oligo plate and which color plate. Then make a very clear map that you can bring with you into the lab to be certain that the correct library will be matched with the chosen index well. Obviously working in sets of 8 and using the multichannel makes it far less likely that mistakes will occur. For larger projects, use the Benchsmart to ensure correct matching of sample well to indexing oligo mix.

> *Note: the GSL indexing oligo plates contain a unique P5 and a unique P7 indexing oligo pre-mixed in each well, both at 5µM.*

1) Take the Kapa 2x Ready Mix and indexing oligo plate out to thaw on top of ice (or leave at +4C for ~30 minutes).

2) Label a new plate for indexing PCR for each sample. If you wish to use up all the library product, set up two parallel reactions for each sample..

3) Add 15 µL Kapa 2x Ready Mix to each well of the empty indexing PCR plate.

4) Then add 10 µL of adapter-ligated product to each reaction well, ideally, add 8 at a time using the LTS 20 µL multichannel. Gently pipette up and down once or twice to mix.

> *Note: if you just used the Benchsmart for post-PCR bead cleaning, re-purpose your final set of tips to load 10 µL of adapter-ligated product into the PCR plate and mix.*

5) Finally, add 5 µL of the corresponding indexing oligo plate., ideally add 8 at a time using the LTS 20 µL multichannel. Gently pipette up and down to mix. Then seal tubes well.

6) Briefly spin down the tubes and place in a cycler **with a heating lid.**

7) Amplify using the following cycling protocol:

Initial Denaturation: 98 °C for 45 seconds

Denaturation: 98 °C for 15 seconds \

Annealing: 60 °C for 30 seconds | x 6 - 9 cycles*

Extension: 72 °C for 60 seconds /

Final Extension: 72 °C for 5 minutes

Hold at 10 °C

**6 cycles is sufficient for most samples starting sonication with 500-1000 ng. (Don't use fewer cycles since 6 is the minimum to ensure that DNA becomes full-length sequence-ready libraries after indexing PCR.) If you start with less input, Table 3 of the Kapa Hyper Prep kit is a useful resource. Generally, about 25% of initial sonication starting material enters the end repair reaction. So, starting with 200 ng of DNA during sonication, we can estimate that there is ~50 ng remaining after size-selection*

Post-PCR Bead Clean-up

1) Before beginning the bead clean up, warm a tube of **low-ratio SPRI bead solution (SeraMag)** to room temperature.

2) In the same tubes as the PCR reaction, perform a 0.6X bead-based cleanup by combining the following:

Indexing PCR product: 30 µL

Water or elution buffer: 70 µL

Low-ratio SPRI bead solution at room temperature: **60 µL**

(Note: if you set up two indexing PCR reactions, they can be combined before bead cleaning. In that case, after combining add 40 µL water/EB and 60 µL low-ratio SPRI beads to preserve the 0.6X ratio)

3) Mix thoroughly by gently vortexing and/or pipetting up and down multiple times.

4) Incubate the tube at room temperature for **15 minutes** to bind DNA to the beads.

5) Place the tubes on a magnet to capture the beads. Incubate **5 minutes** or until the liquid is clear.

6) Carefully remove and discard the supernatant.

7) Keeping the tubes on the magnet, add **200 µL** of freshly prepared **80% ethanol**.

8) Incubate the tubes on the magnet at room temperature for **30 seconds**

9) Carefully remove and discard the ethanol.

10) Keeping the tubes on the magnet, add **200 µL** of freshly prepared **80% ethanol**.

11) Incubate the tubes on the magnet at room temperature for **30 seconds**

12) Carefully remove and discard the ethanol. Try to remove all residual ethanol without disturbing the beads. To accomplish this, use a small volume (10µL) tip to remove ethanol remaining at the bottom of the tube. Then use wooden toothpicks to blot up any dots of ethanol on the sides of the tubes.

13) Dry the beads at room temperature for **3 minutes or less**, until all of the residual ethanol has evaporated from the tube but the beads themselves are still damp and shiny. *Caution: over-drying the beads may result in reduced yield.*

14) Remove the tubes from the magnet. Then thoroughly resuspend the beads: in **27 µL** (or more*) of **elution buffer** (10 mM Tris-HCl, pH 8.0 – 8.5, a.k.a. Buffer EB or EBT) or **nuclease-free water****

**Note1: if you know or suspect your samples will all have high yields, eluting in 50 µL or more will help to keep them in preferred range of the qubit/plate reader, bioanalyzer, and GSL*
***Note2: water is the best choice for most samples that will be captured since for some probe technologies, the concentrated salts could interfere. Use commercially-purified water to ensure that the pH is not too low. If samples will be directly sequenced, EB(T) is best for long-term preservation since it is buffered.*

15) Incubate the tubes at room temperature for **10 minutes** to elute DNA off the beads.

16) Place the tubes on a magnet to capture the beads. Incubate for **3 minutes** or until the liquid is clear.

17) Transfer ~**5 µL** of the clear supernatant to a plate or a set of **low bind/siliconized** tubes and seal. This will be used for quality control assays. (At a minimum you will need 1 µL for qubit or nanodrop, and 2-3 µL for bioanalyzer or agarose gel)

18) The remaining **20 µL+** will be moved to a second set of **low bind/siliconized** tubes or a plate and sealed. This will be where your libraries will be stored until submitting for sequencing. So label these well.

18) The remaining **~5 µL** can be used for quality control. (At a minimum you will need 1 µL for qubit or nanodrop, and 2-3 µL for bioanalyzer or agarose gel)

Quality control, GSL submission and/or pooling:

1) Qubit or nanodrop all samples using 1 µL of final library product. If any sample values are too low, they can be re-amplified using IS5 and IS6 oligos.

- Nanodrop is recommended when library quantities are high (> 50 ng/µL) and researchers are pooling the libraries themselves.
- Use Qubit/plate reader when the GSL is pooling since they require that value with your submission

Very little DNA is required for sequencing. The GSL requests 10 µL at 10 nM which is around 4 ng/µL at an average length of 600 bp. However, it may be difficult for us to accurately assess library quantities and sizing that are so low.

2) To reamplify any samples that are too low:
a) Set up the following master mix:

25 µL 2x Kapa HiFi HotStart ReadyMix
5 µL 10x Kapa primer mix (or 2.5 µL each of IS5 and IS6 at 10 µM)
20 µL final library product

b) Use the same indexing PCR cycler conditions at 2 or more cycles, depending on how many are required to achieve a more robust product but not overly amplify the libraries.
c) Clean using the 0.6X low-ratio SPRI bead protocol

3) If submitting to the GSL for pooling, run all samples on a bioanalyzer DNA 1000 chip. Any samples with an adapter peak around 150 bp may need to be re-cleaned. Use the region table to determine the average size and molarity of the library hill.

If self-pooling, running 50-100 ng of library product on an agarose gel will usually suffice in assessing general sizing and whether libraries are free of dimer.

4) Optional: prepare samples to be pooled for capture or direct sequencing by calculating the amount needed to combine equimolar amounts of each library per pool.

Note: if all libraries have roughly the same size distribution, mass (total ng) can be a proxy for molarity if you assessed by gel rather than bioanalyzer.

5) Chat with the lab manager for recommendations on pooling

Special notes for full-plate reactions

A high-throughput protocol is in development, but until that is prepared, here are tips for how to efficiently do this protocol with a full (or nearly full plate)

Use the Benchsmart for all bead cleaning steps. This instrument is designed to work with 96 tips, but we can manually remove tips from the box when working with samples that are less than a full plate.

In this case, exclusively use the tips to mix during bead cleaning. Do not vortex.

Always be in the habit of orienting plates so that the letters are to the right and the numbers on top. Really, the only big mistake we can make is inverting a plate 180°, so if we are used to always orienting plates in the same direction, we will minimize the chances of that occurring.

Space sonication tubes out in 6 rows on two plates, with an empty row between them. The plates can balance each other in a plate centrifuge for a spin-down. Then use a 200 µL multichannel pipette to transfer 100 µL of each row into the corresponding row of a semi-skirted plate. Samples that have volume below this value can be topped off in the plate to be equal volume with the others. The plate can be frozen until ready to proceed to double-sided selection. Keep the original 0.2 mL tubes in a PCR rack with lid and/or photograph for later reference.

The ERAT mix needs a larger than usual multiplier since volumes are so low. Master mix for a full plate may need a multiplier of 115 or 120. Then distribute 35-36 µL into each well of your strip tube for distribution.

The ligation master mix multiplier can be 105-110 and still have sufficient volume. Distribute 195-200 µL into each well of your strip tube for distribution.

For distributing 2X ReadyMix, add 200 µL into each well of your strip tube. You can return any unused mix to the stock tube if you have excess remaining once it is aliquoted. You can also re-use that same strip tube over multiple plates if its condition remains good.

If setting up the PCR reaction immediately following post-ligation cleaning, use the same tips that just moved 20 µL of your adapter-ligated product off the beads to move 10 µL from there to your PCR plate. Then switch to a clean set of tips to use the Benchsmart to add 5 µL of indexing oligo mix to the PCR reaction plate and mix.

Seal plates with clear or foil sticky mats and use a hard sealing tool to ensure that the seal is thorough. Foil seals can be punctured with a pipette tip which is useful for keeping track of your location when performing QC assays. However, be sure to stretch the puncture hole to be larger than the pipette tip to prevent a vacuum forming that will distort pipetting accuracy.