

SPECIES RELATIONSHIPS AND *WOLBACHIA*-HOST INTERACTIONS IN THE
DROSOPHILA SUBQUINARIA SPECIES COMPLEX

by

PAUL SAMUEL GINSBERG

(Under the Direction of KELLY DYER)

ABSTRACT

Speciation is the process by which a single population can evolve into two or more evolutionarily independent lineages. This dissertation investigates the species relationships and microbe-host dynamics in the *Drosophila subquinaria* species complex, a group of three mushroom-feeding *Drosophila* species. In sexually reproducing organisms, species can be defined as independent lineages that do not produce viable and/or fertile offspring. Importantly, barriers that prevent interbreeding can take multiple forms that can include the effect of vertically inherited endosymbionts. *Wolbachia* is a maternally-inherited endosymbiont that can drastically alter reproduction of its hosts, and is common among arthropods. In this dissertation, I assess species relationships in the *D. subquinaria* species complex, examine *Wolbachia*-induced phenotypes in a native and non-native host species, and generate genome assemblies for *Drosophila* hosts and their *Wolbachia* infections. First, I utilize a combination of phylogenetic and population genetic methods to analyze a multi-locus dataset covering all genomic regions with deep population sampling of all three *D. subquinaria* complex species. I find that despite strong levels of reproductive isolation between *D. subquinaria* and *D. recens*, there is evidence of gene flow and mitochondrial

introgression between these species. A high frequency *Wolbachia* infection in *D. recens* and evidence of gene flow suggests that *D. subquinaria* may be exposed to *Wolbachia* infection. Second, using lab-estimated measurements of *Wolbachia*-induced phenotypes in both *D. recens* and *D. subquinaria*, I predict that *Wolbachia* will likely invade *D. subquinaria* with minimal gene flow from *D. recens*. Third, I generate de novo genome assemblies for all species of the *D. subquinaria* species complex and one outgroup species, *D. quinaria*. I reassess species relationships with broader genomic sampling and find species relationships that are consistent with levels of reproductive isolation among species. Lastly, I generate full genomes for two distinct *Wolbachia* infections in *D. recens*, and show that these two strains are recently diverged, yet possess major structural differences. With preliminary population sampling of *D. recens* populations, I hypothesize that *D. recens* is currently undergoing a *Wolbachia* sweep that may facilitate *Wolbachia* invasion into *D. subquinaria*.

INDEX WORDS: speciation, *Wolbachia*, phylogenetics, gene flow, genomics, endosymbiont, species interactions, reproductive isolation, host-microbe interactions, invasion dynamics

SPECIES RELATIONSHIPS AND *WOLBACHIA*-HOST INTERACTIONS IN THE
DROSOPHILA SUBQUINARIA SPECIES COMPLEX

by

PAUL SAMUEL GINSBERG

BS, University of California, Davis, 2014

A Dissertation Submitted to the Graduate Faculty of The University of Georgia in Partial
Fulfillment of the Requirements for the Degree

DOCTOR OF PHILOSOPHY

ATHENS, GEORGIA

2022

© 2022

Paul Samuel Ginsberg

All Rights Reserved

SPECIES RELATIONSHIPS AND *WOLBACHIA*-HOST INTERACTIONS IN THE
DROSOPHILA SUBQUINARIA SPECIES COMPLEX

by

PAUL SAMUEL GINSBERG

Major Professor:	Kelly Dyer
Committee:	Gaelen Burke
	Michael White
	Andrea Sweigart
	John Wares

Electronic Version Approved:

Ron Walcott
Vice Provost for Graduate Education and Dean of the Graduate School
The University of Georgia
August 2022

DEDICATION

First, I dedicate this dissertation to my parents, Maureen and Howard. Their constant support and guidance throughout my life and especially during my time in graduate school has been integral to my persistence in and completion of this dissertation. Second, I dedicate this dissertation to my brother, who has always been there to help me when I needed someone to talk to, and to provide guidance from his own time in graduate school.

ACKNOWLEDGEMENTS

First, I want to acknowledge my advisor, Dr. Kelly Dyer. Kelly's support and guidance has not only been indispensable to my success as a graduate student, but has made me a better scientist. She has given me the intellectual freedom to explore my interests all the while providing me with the tools I needed to further my academic development. In addition, she made the process of working during the onset of and throughout the COVID-19 pandemic as smooth as possible, while providing support during these difficult times. I would also like to acknowledge my undergraduate advisor, Dr. Michael Turelli. He initially invited to work in his laboratory as an undergraduate, and encouraged my interest in genetics, evolution and *Drosophila* speciation. I am grateful for the opportunities he provided me as a scientist, and his continued encouragement and support during my time as a graduate student. I would additionally like to acknowledge former and current members of the Dyer lab, lab technicians Brooke White, Dylan Ricke, Zach Sanchez and Kevin Moore and fellow graduate students Nick Arthur, Theresa Miorin, and Andrew Duitsman. Their hard work and helpful discussions have helped me become a better scientist and created a stimulating and supportive lab community. I would also like to acknowledge my committee members for providing me with feedback that helped improve my dissertation. Lastly, I would like to acknowledge the numerous undergraduates who have aided me in conducting experiments presented in this dissertation. It has been incredibly rewarding to work with these developing researchers, and to help them discover their enthusiasm for and interests in the sciences. I would like to acknowledge the work of undergraduates Rachel Keener, Madeline Snipes, Shannon Harris, and Jordan Taylor for their help in conducting experiments.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	v
LIST OF TABLES	viii
LIST OF FIGURES	x
CHAPTER	
1 INTRODUCTION	1
Works Cited	9
2 ONGOING HYBRIDIZATION OBSCURES PHYLOGENETIC RELATIONSHIPS IN THE <i>DROSOPHILA SUBQUINARIA</i> SPECIES COMPLEX.....	17
Abstract	18
Introduction.....	18
Methods.....	22
Results.....	28
Discussion.....	33
Works Cited	41
3 <i>WOLBACHIA</i> HOST INTERACTIONS IN A NATIVE AND NON-NATIVE HOST SPECIES OF THE <i>DROSOPHILA SUBQUINARIA</i> SPECIES COMPLEX	58
Introduction.....	58
Methods.....	62
Results.....	70

Discussion	74
Works Cited	81
4 GENOME ASSEMBLIES FOR FOUR SPECIES OF THE QUINARIA SPECIES GROUP AND THEIR <i>WOLBACHIA</i> INFECTIONS.....	108
Introduction.....	108
Methods.....	111
Results.....	117
Discussion.....	122
Works Cited	126
5 CONCLUSIONS AND FUTURE DIRECTIONS	150
Works Cited	153
APPENDICES	
A Supplementary Materials Chapter 2.....	155
B Supplementary Materials Chapter 3.....	215
C Supplementary Materials Chapter 4.....	241

LIST OF TABLES

	Page
Table 2.1: Sampling locations and number of isofemale lines sampled from each location.....	56
Table 2.2: Migration rates and effective populations size as estimated by Migrate-n.	57
Table 3.1: Experimental crosses for cytoplasmic incompatibility in <i>D. recens</i> and <i>D. subquinaria</i> 97	
Table 3.2: Experimental crosses for relative fecundity in <i>D. recens</i> and <i>D. subquinaria</i>	98
Table 3.3: Maternal transmission estimates for <i>D. recens</i> and <i>D. subquinaria</i> genotypes.....	99
Table 3.4: Model parameter and equilibrium frequency estimates.....	100
Table 3.5: Summary of differentially expressed genes in <i>D. recens</i> and <i>D. subquinaria</i>	101
Table 3.6: All differentially expressed genes in <i>D. subquinaria</i> with BLAST annotations.....	102
Table 3.7: Top ten differentially expressed genes in each <i>D. recens</i> treatment with BLAST annotations	105
Table 3.8: Significant GO terms for differentially expressed genes in <i>D. recens</i>	107
Table 4.1: Fly samples used for whole genome sequencing.....	142
Table 4.2: Assembly statistics for fly genome assemblies	143
Table 4.3: Gene count predictions from Braker annotations	144
Table 4.4: Total counts for all TEs identified in genome assemblies by Repeatmasker	145
Table 4.5: Assembly statistics for <i>wRec</i> genomes.....	146
Table 4.6: Gene annotations for genes in <i>wRec</i> inversion.....	147
Table 4.7: Genes in RW <i>wRec</i> missing in Alb <i>wRec</i>	148

Table 4.8: Prevalence of <i>Wolbachia</i> variants from <i>D. recens</i> populations.....	149
Table S2.1: Loci used in this study.....	188
Table S2.2: ABBA-BABA sites per locus.....	190
Table S2.3: Sampling totals for obtained and missing sequence data.....	191
Table S2.4: Divergence and differentiation for each locus and group comparison.....	195
Table S2.5: Fixed differences and shared polymorphisms among groups.....	203
Table S2.6: Chi-square tests comparing the number of fixed differences and polymorphic sites between group comparisons.....	204
Table S2.7: Summary of nucleotide polymorphism and frequency spectrum for each locus and group	205
Table S3.1: qPCR primers for <i>Wolbachia</i> titer estimation.....	216
Table S3.2: RNAseq data used for differential expression analysis.....	217
Table S3.3: Annotations for all differentially expressed genes in <i>D. recens</i>	218
Table S4.1: Oxford Nanopore data for genome assemblies.....	245
Table S4.2: Illumina short read data for genome assemblies.....	246
Table S4.3: <i>Wolbachia</i> isolates used for phylogenetic analysis.....	247
Table S4.4: SNVs between <i>wRec</i> variants.....	248
Table S4.5: Indels distinguishing RW <i>wRec</i> and Alb <i>wRec</i>	250
Table S4.6: Location of BUSCO orthologs in the <i>D. innubila</i> genome that are missing in the Q_QTP assembly.....	251

LIST OF FIGURES

	Page
Figure 2.1: Map of sampling locations	48
Figure 2.2: Gene trees for A) mtDNA based on the <i>COI</i> locus and B) Y-linked loci based on concatenated <i>kl-3</i> and <i>kl-5</i> loci	49
Figure 2.3: Inferred majority rule consensus trees with and without topological constraints from the concatenated Bayesian analyses using all nuclear data (autosomes, X-linked, & Y-linked loci)	50
Figure 2.4: Quartet sampling score on focal branch of three trees generated with nuclear data ...	51
Figure 2.5: Phylogeny generated using the multi-species coalescent model implemented in *BEAST	52
Figure 2.6: Net nucleotide divergence (D_a) for each group comparison	53
Figure 2.7: Genetic differentiation (K_{ST}) between groups, separated by autosomal (blue) and X-linked (green) loci	54
Figure 2.8: The total number of fixed differences (blue) and polymorphic sites (green) summed across autosomal and X-linked loci	55
Figure 3.1: Offspring hatch rate of compatible and incompatible crosses in native host, <i>D. recens</i> (red), and non-native host, <i>D. subquinaria</i> (blue & green)	90
Figure 3.2: Offspring sex ratio of infected (dark shading) and uninfected (light shading) females	

Figure 3.3: Offspring production of infected (darker shading) and uninfected (lighter shading) females	92
Figure 3.4: <i>Wolbachia</i> titer as measured by relative copy number of a host locus (<i>mago</i>) to a <i>Wolbachia</i> locus (<i>wsp</i>).....	93
Figure 3.5: Differential expression of <i>Wolbachia</i> -infected and uninfected <i>D. recens</i> females.....	94
Figure 3.6: Differential expression of <i>Wolbachia</i> -infected and uninfected <i>D. subquinaria</i> female	95
Figure 3.7: Number of shared differentially expressed genes among species and tissues.....	96
Figure 4.1: Assembly pipeline for fly genomes.....	133
Figure 4.2: BUSCO assessment for all fly genome assemblies.....	134
Figure 4.3: Host phylogenies from 100 randomly selected single copy orthologs.....	135
Figure 4.4: Transposable element abundance in all assemblies	136
Figure 4.5: <i>wRec</i> genome alignment from RW and Alb <i>D. recens</i> stocks	137
Figure 4.6: Zoomed in view of aligned regions with deletion in RW <i>wRec</i>	138
Figure 4.7: Inverted region of <i>wRec</i> variants.....	139
Figure 4.8: <i>Wolbachia</i> phylogeny generated from five MLST loci and <i>wsp</i>	140
Figure 4.9: Model of <i>Wolbachia</i> genome evolution from ancestor with <i>wMel</i>	141
Figure S2.1: Individual gene trees from the Bayesian analysis	156
Figure S2.2: Inferred majority rule trees with and without topological constraints from the concatenated Bayesian analyses using autosomal data (Panels A & B) and X-linked data (Panels C & D).....	182
Figure S2.3: Inferred majority rule trees from the concatenated Bayesian analyses using nuclear data with sympatric <i>D. subquinaria</i> samples removed.....	183

Figure S2.4: Synonymous substitution rate (K_s) for each group comparison	184
Figure S2.5: The total number of fixed differences (blue) and polymorphic sites (green) categorized by genomic region (i.e. autosomal and X-linked loci)	185
Figure S2.6: Synonymous nucleotide polymorphism showing A) synonymous pi (π_{syn}) and B) synonymous theta (Θ_{syn})	186
Figure S2.7: Tajima's D for synonymous sites.....	187
Figure S4.1: Gel image for <i>Wolbachia</i> strain PCR.....	242
Figure S4.2: Comparison of genome assembly length by assembly N50.....	243
Figure S4.3: Species tree with estimated gene duplication events	244

CHAPTER 1

INTRODUCTION

Speciation is the process through which a single lineage evolves into two or more independent lineages. In sexually-reproducing organisms, the most commonly used definition of a species is the biological species concept, which defines two populations as separate species when they are reproductively isolated from each other (Mayr, 1942). The biological species concept provides a useful framework to investigate the process of speciation in extant taxa (Butlin & Stankowski, 2020), as it simplifies the question of speciation to: How do populations evolve reproductive isolation? The classical model for understanding this was first articulated by (Bateson, 1909) and later described by Dobzhansky (1937) and Muller (1942). This model proposes that first a single population is split into two geographically isolated subpopulations (e.g. through vicariance, dispersal, etc.), and as populations diverge in allopatry, new alleles accumulate across many loci and fix either through selection (via local adaptation) or genetic drift. Upon secondary contact, and the production of hybrids, these new alleles are exposed to each other for the first time in a single organism. Negative interactions between these alleles (i.e. negative epistasis) lowers fitness in hybrids compared to individuals from either of the two parent populations. Incompatibilities that result in either the inviability or sterility of hybrids maintain species boundaries and permit further divergence of the parent populations.

In many cases, diverging populations that come back together in secondary contact are not completely reproductively isolated. These closely related species that show incomplete pre-

and post-mating isolation allow the process of speciation to be studied (Coyne & Orr, 2004). In particular, the production of hybrid offspring with relatively low fitness can indirectly select for increased levels of premating isolation, and thereby accelerate and/or complete the speciation process (Howard, 1993). This process of reinforcement has been implicated in the completion of the speciation process when other forms of reproductive isolation are incomplete. While historically reinforcement was considered a rare phenomenon, its importance and commonality has been more widely accepted in recent decades (Noor, 1995; Nosil et al., 2003; Ortiz-Barrientos et al., 2004). In their seminal paper, “Patterns of Speciation in *Drosophila*”, (Coyne & Orr, 1989) compared levels of premating isolation (as measured by mate discrimination) and postzygotic reproductive isolation (as measured by hybrid sterility and/or inviability) in 119 pairs of closely-related *Drosophila* species. They found that among allopatric species pairs, premating and postzygotic isolating barriers evolved at comparable rates. However, in sympatric species pairs strong mate discrimination evolved much faster than the severest forms of postzygotic reproductive isolation. Together, these results suggest that in sympatric species, where hybrid production can occur, reinforcement likely contributes to the continued divergence of distinct populations. The conclusion of this study has been supported by more recent analyses (Coyne & Orr, 1997; Yukilevich, 2012) and empirical estimates of reinforcement in a wide variety of taxa (Hoskin et al., 2005; Lukhtanov et al., 2005; Noor, 1995; Pfennig & Rice, 2014; St John & Fuller, 2021; Urbanelli & Porretta, 2008).

At the broader level, resolving species relationships is necessary to understand how reproductive isolating mechanisms generate and maintain species boundaries. Hybridization and subsequent introgression between species as well as incomplete lineage sorting (ILS) due to large ancestral effective population sizes can obscure the phylogenetic signal among groups and

render evolutionary relationships unclear (Degnan & Rosenberg, 2009; Edwards et al., 2016; Hahn & Nakhleh, 2016; Maddison, 1997). Recent phylogenetic methods use genome-wide approaches and incorporate population genetic data in order to resolve species relationships (Edwards, 2009; Wolf & Ellegren, 2017). In addition, these analyses can make use of parts of the genome that are differently affected by evolutionary processes (Pease & Hahn, 2013). For instance, sex chromosomes are often overrepresented in speciation-related processes (i.e. the “large X-effect”) and often show greater divergence among groups than do highly recombining regions of the autosomes (reviewed in (Payseur & Rieseberg, 2016; Presgraves, 2018).

In this dissertation I study the evolutionary relationships of three closely related and incompletely isolated *Drosophila* species: *D. subquinaria*, *D. recens*, and *D. transversa*. All three species are morphologically identical and can only be distinguished via the internal male genitalia (Wheeler, 1960). *D. recens* and *D. subquinaria* occur in eastern and western North America, respectively, and their ranges overlap for approximately 1,500 km in central Canada (Dyer et al., 2018; Dyer et al., 2014; Jaenike et al., 2006). *D. transversa* occurs in central and northern Europe, and is thought to occur eastward to the Kamchatka Peninsula in eastern Russia (Sidorenko, 2009). These species are in the quinaria group of the subgenus *Drosophila*, and all of them are generalists on fleshy basidiomycete mushrooms. Where they co-occur, *D. recens* and *D. subquinaria* can be collected at the same mushroom baits.

In the laboratory, all three species can hybridize and produce fertile F1 females. Crosses between *D. subquinaria* and *D. transversa* also produce fertile F1 males, while hybrid males resulting from crosses between *D. recens* and either *D. subquinaria* or *D. transversa* are sterile (Humphreys et al., 2016; Shoemaker et al., 1999). Additional hybrid death results from a *Wolbachia* infection that is at high frequency in *D. recens* (Jaenike et al., 2006; Shoemaker et al.,

1999; Werren & Jaenike, 1995). When a *Wolbachia*-infected *D. recens* male mates with either a *D. subquinaria* or *D. transversa* female, most of the offspring die as embryos due to cytoplasmic incompatibility (CI) (Humphreys et al., 2016; Shoemaker et al., 1999). *Wolbachia* does not affect offspring production in reciprocal crosses between *D. recens* females and males from the other two species (Humphreys et al., 2016; Shoemaker et al., 1999).

Patterns of premating isolation are partially consistent with the patterns of postzygotic isolation. In the laboratory, *D. recens* females mate at a moderate rate with both *D. subquinaria* and *D. transversa* (Humphreys et al., 2016; Shoemaker et al., 1999). In contrast, *D. subquinaria* females from populations sympatric with *D. recens* east of the Canadian Rockies discriminate strongly against *D. recens* males, whereas *D. subquinaria* females from western regions discriminate less (Dyer et al., 2018; Dyer et al., 2014; Jaenike et al., 2006). Female mate discrimination has evolved such that these “sympatric” *D. subquinaria* females also discriminate against conspecific males from western populations that are allopatric with *D. recens*. These patterns are consistent with both classical reinforcement (against *D. recens*) and cascade reinforcement (against allopatric *D. subquinaria*) given there is no post-mating isolation in conspecific crosses using *D. subquinaria* from different regions. Previous research found no behavioral isolation between *D. transversa* and *D. subquinaria* from the western part of its range, though *D. subquinaria* females from the eastern part of its range discriminate against *D. transversa* males (Humphreys et al., 2016).

Some hybridization occurs between *D. recens* and *D. subquinaria* in the wild. First, roughly 3% of wild-caught *D. subquinaria* harbor a mtDNA haplotype from *D. recens*, indicative of hybridization between *D. recens* females and *D. subquinaria* males and subsequent introgression of the mtDNA into *D. subquinaria* (Jaenike et al., 2006; Shoemaker et al., 1999;

Werren & Jaenike, 1995) Second, microsatellite genotyping also suggests past introgression at nuclear loci, and most of it appears to occur in the geographic regions where these species co-occur (Dyer et al., 2018; Dyer et al., 2014; Jaenike et al., 2006). Genetic differentiation within each species is consistent with variation in behavioral isolation: previous work suggests there is very little genetic differentiation among populations of *D. recens*, while there is moderate differentiation among populations of *D. subquinaria* that are allopatric and sympatric with *D. recens* (Dyer et al., 2018; Dyer et al., 2014; Jaenike et al., 2006).

In spite of extensive study of the patterns of reproductive isolation among these three species, their evolutionary relationships remain unresolved. Based on both the presence of hybrid male sterility of *D. recens* with *D. subquinaria* and *D. transversa* and the lack of pre-mating isolation between *D. subquinaria* and *D. transversa*, the primary hypothesis is that *D. subquinaria* and *D. transversa* are most closely related to each other with *D. recens* as the outgroup. An alternate hypothesis is based on geography, where *D. transversa* is the basal group because it occurs in the old world, while *D. recens* and *D. subquinaria* are most closely related because both are new world species and previous studies suggested they were separated only recently during the Wisconsin glaciation 75 ka – 10 ka (Jaenike et al., 2006; Shoemaker et al., 1999; Werren & Jaenike, 1995). To date, population genetic studies have only included *D. recens* and *D. subquinaria* and not *D. transversa*, and have genotyped microsatellites rather than DNA sequence data, making evolutionary inferences difficult. A phylogenetic analysis of the entire quinaria species group only sampled one to two alleles per species and found no consistent relationship among these three species (Scott Chialvo et al., 2019). The lack of consistency across loci suggested hybridization and introgression or ILS among these species.

In Chapter Two, I use population genetic samples from each of these three species to investigate their evolutionary relationships. We Sanger sequence 29 loci from throughout the genome from a sample of each species, and then we use phylogenetic and population genetic analyses to investigate both species relationships and patterns of introgression among species. We compare patterns of differentiation among different regions of the genome and we interpret our findings in light of known patterns of reproductive isolation. The results suggest that ongoing hybridization between *D. recens* and *D. subquinaria* is common where they co-occur and that this obscures the phylogenetic signal among species.

In Chapter 3, I investigate the phenotypic consequences of *Wolbachia* infection in both naturally infected *D. recens*, and *Wolbachia*-infected *D. subquinaria* generated in the laboratory. *Wolbachia* is a maternally inherited, intracellular alpha-proteobacterium. It is ubiquitous among terrestrial arthropods, with estimates ranging from ~20-60% of insect species harboring infections (Weinert et al., 2015; Werren & Windsor, 2000). *Wolbachia* is well known for its capacity to manipulate the reproduction of its hosts in ways that can enhance its own fitness, despite fitness costs to the host they infect. Reproductive manipulations observed in insects include male-killing, thelytokous parthenogenesis, and the most common, cytoplasmic incompatibility (CI)(Werren et al., 2008). CI results in embryonic death of offspring from crosses between infected males and uninfected females (or females harboring an incompatible *Wolbachia* infection) (Kaur et al., 2021; Yen & Barr, 1971). In addition to manipulating host reproduction, *Wolbachia* can induce a myriad of other phenotypes in the host. The range of *Wolbachia*-induced phenotypes encapsulate the full range of host-endosymbiont relationships from parasite to mutualist and facultative to obligate (Zug & Hammerstein, 2015).

Reproductive manipulations are generally understood as strategies through which *Wolbachia* infections can force their invasion of naive host populations. Long-term phylogenetic patterns support *Wolbachia* movement and invasion across intraspecific populations and interspecific boundaries (Raychoudhury et al., 2009; Schuler et al., 2013; Turelli et al., 2018). In further support of this, *Wolbachia* invasion has been documented in real time among a handful of insect species (Bakovic et al., 2018; Duploux et al., 2010; Kriesner et al., 2013). The conditions under which CI-causing *Wolbachia* infections can invade naive populations have been extensively described (Caspari, 1959; Hoffmann et al., 1990; Hurst, 1996; Prout, 1994; Turelli & Hoffmann, 1995; Turelli et al., 1992).

Contrary to *Wolbachia*'s invasion potential, *Wolbachia*-induced CI itself can serve as an interspecific reproductive boundary if only one species harbors a *Wolbachia* infection or if the two species harbor incompatible *Wolbachia* infections (Bordenstein et al., 2001; Cruz et al., 2021; Jaenike et al., 2006; Shoemaker et al., 1999). This manifests as increased hybrid offspring mortality in incompatible crosses between species, as is seen between *D. recens* and *D. subquinaria*. The asymmetry in levels of postzygotic and premating reproductive isolation suggests that the presence of a CI-causing *Wolbachia* infection in *D. recens* has indirectly selected for increased premating isolation specifically in sympatric *D. subquinaria*. Thus, *Wolbachia* infection in *D. recens* has provided conditions necessary for asymmetrical reinforcement. The resulting mate discrimination observed in sympatric *D. subquinaria* is sufficiently strong that it reduces mating between sympatric and allopatric *D. subquinaria*, potentially driving incipient speciation within *D. subquinaria* (Humphreys et al., 2016). This process of cascading reinforcement (Ortiz-Barrientos et al., 2009) exemplifies the potential for clade-wide consequences of a single *Wolbachia* infection.

Gene flow and mitochondrial introgression from *D. recens* into *D. subquinaria* suggests that *D. subquinaria* is likely exposed to *Wolbachia* infection. As CI-causing *Wolbachia* infections have been observed to traverse species boundaries, should we expect to see introgression of *Wolbachia* from *D. recens* into *D. subquinaria*? Thus far *Wolbachia* has not been observed in *D. subquinaria*, but examination of *Wolbachia*-induced phenotypes in the lab can inform the conditions under which we would expect *Wolbachia* to invade.

Chapter Three addresses the question of why hasn't *Wolbachia* invaded *D. subquinaria*? Research of *Wolbachia* infected arthropod hosts reveal that *Wolbachia*-induced phenotypes depend greatly on the *Wolbachia* strain and the host genetic background (Jaenike, 2007; Lec'h et al., 2012; Zhou & Li, 2016). Previous work in this system showed that when *Wolbachia* from *D. recens* was introgressed into *D. subquinaria* in the lab, the reproductive manipulation shifted from CI in *D. recens* to male-killing (MK) in *D. subquinaria* (Jaenike 2007). In this study, I first introgress *Wolbachia* from *D. recens* into *D. subquinaria*, and find that it causes CI rather than male-killing. I then investigate the conditions that would permit *Wolbachia* invasion into *D. subquinaria*. I find that *Wolbachia* causes CI in both species, with even stronger CI in *D. subquinaria* than in *D. recens*. I parameterize *Wolbachia* invasion models and find that *Wolbachia* should spread within *D. subquinaria*. Given this has not occurred to date, I expand on reasons why this might be. Lastly, I use quantitative PCR and gene expression methods to infer the mechanism of higher CI in *D. subquinaria* than *D. recens*. These analyses reveal no difference in *Wolbachia* titer and very few shared differentially expressed genes between species.

In Chapter Four, I reassess species relationships in a genomic context. Using Oxford-nanopore long-read sequencing and Illumina short-read sequencing I generated de novo genomes

for all species of the *D. subquinaria* species complex and one outgroup species, *D. quinaria*. Surprisingly, phylogenetic analysis of single-copy orthologs displays sympatric and allopatric *D. subquinaria* as sister taxa, consistent with our initial hypothesis and levels of postzygotic reproductive isolation, yet inconsistent with phylogenetic relationships inferred from Chapter Two. I propose that incongruity between inferred phylogenetic relationships of our two studies results from differences in sampling, where greater sampling across the genome results in more accurate delimitation of species as opposed to greater sampling of populations and lower sampling across the genome. Additionally, I observe substantial assembly size variation in these species and show that it is not a consequence of transposable element proliferation. From sequence data from two *D. recens* samples, I pull out *Wolbachia* reads and assemble complete *Wolbachia* genomes. This reveals two distinct *Wolbachia* strains infecting *D. recens*, one from a stock collected from an eastern population *D. recens* in 2001, and the second from a western population of *D. recens* collected around 2010. I show that these two variants are remarkably similar, but differ by two major structural changes, a 33.3 kb inversion and 25.6 kb indel. With a phylogenetic analysis, I show that these *Wolbachia* variants likely evolved after infecting *D. recens*. I discuss phenotypic consequences in light of the observed structural changes, propose a model of their evolutionary history, and discuss clade-level consequences of *Wolbachia* prevalence in *D. recens*, and its potential to invade *D. subquinaria*.

Works Cited

- Bakovic, V., Schebeck, M., Telschow, A., Stauffer, C., & Schuler, H. (2018). Spatial spread of *Wolbachia* in *Rhagoletis cerasi* populations. *Biol Lett*, *14*(5).
<https://doi.org/10.1098/rsbl.2018.0161>

- Bateson, W. (1909). Heredity and variation in modern lights [Chapter]. *Darwin and Medical Science*, 85-101. <https://doi.org/10.1017/CBO9780511693953.007>
- Bordenstein, S. R., O'Hara, F. P., & Werren, J. H. (2001). Wolbachia-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature*, 409(6821), 707-710. <https://doi.org/10.1038/35055543>
- Butlin, R. K., & Stankowski, S. (2020). Is it time to abandon the biological species concept? No. *Natl Sci Rev*, 7(8), 1400-1401. <https://doi.org/10.1093/nsr/nwaa109>
- Caspari, E., Watson, GS. (1959). On the Evolutionary Importance of Cytoplasmic Sterility in Mosquitoes. *Evolution*, 13(4), 568-570. <https://doi.org/10.2307/2406138>
- Coyne, J. A., & Orr, H. A. (1989). Patterns of Speciation in *Drosophila*. *Evolution*, 43(2), 362-381. <https://doi.org/10.1111/j.1558-5646.1989.tb04233.x>
- Coyne, J. A., & Orr, H. A. (1997). "Patterns of Speciation in *Drosophila*" Revisited. *Evolution*, 51(1), 295-303. <https://doi.org/10.1111/j.1558-5646.1997.tb02412.x>
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*. Sinauer Associates Inc.
- Cruz, M. A., Magalhaes, S., Sucena, E., & Zele, F. (2021). Wolbachia and host intrinsic reproductive barriers contribute additively to postmating isolation in spider mites. *Evolution*. <https://doi.org/10.1111/evo.14286>
- Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol*, 24(6), 332-340. <https://doi.org/10.1016/j.tree.2009.01.009>
- Duploux, A., Hurst, G. D., O'Neill, S. L., & Charlat, S. (2010). Rapid spread of male-killing Wolbachia in the butterfly *Hypolimnas bolina*. *J Evol Biol*, 23(1), 231-235. <https://doi.org/10.1111/j.1420-9101.2009.01891.x>

- Dyer, K. A., Bewick, E. R., White, B. E., Bray, M. J., & Humphreys, D. P. (2018). Fine-scale geographic patterns of gene flow and reproductive character displacement in *Drosophila subquinaria* and *Drosophila recens*. *Mol Ecol*. <https://doi.org/10.1111/mec.14825>
- Dyer, K. A., White, B. E., Sztepanacz, J. L., Bewick, E. R., & Rundle, H. D. (2014). Reproductive character displacement of epicuticular compounds and their contribution to mate choice in *Drosophila subquinaria* and *Drosophila recens*. *Evolution*, *68*(4), 1163-1175. <https://doi.org/10.1111/evo.12335>
- Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution*, *63*(1), 1-19. <https://doi.org/10.1111/j.1558-5646.2008.00549.x>
- Edwards, S. V., Potter, S., Schmitt, C. J., Bragg, J. G., & Moritz, C. (2016). Reticulation, divergence, and the phylogeography-phylogenetics continuum. *Proc Natl Acad Sci U S A*, *113*(29), 8025-8032. <https://doi.org/10.1073/pnas.1601066113>
- Hahn, M. W., & Nakhleh, L. (2016). Irrational exuberance for resolved species trees. *Evolution*, *70*(1), 7-17. <https://doi.org/10.1111/evo.12832>
- Hoffmann, A. A., Turelli, M., & Harshman, L. G. (1990). Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*. *Genetics*, *126*(4), 933-948. <https://doi.org/10.1093/genetics/126.4.933>
- Hoskin, C. J., Higgie, M., McDonald, K. R., & Moritz, C. (2005). Reinforcement drives rapid allopatric speciation. *Nature*, *437*(7063), 1353-1356. <https://doi.org/10.1038/nature04004>
- Howard, D. J. (1993). *Reinforcement: Origin, dynamics, and the fate of evolutionary hypothesis*. Oxford University Press.

- Humphreys, D. P., Rundle, H. D., & Dyer, K. A. (2016). Patterns of reproductive isolation in the *Drosophila subquinaria* complex: can reinforced premating isolation cascade to other species? *Curr Zool*, *62*(2), 183-191. <https://doi.org/10.1093/cz/zow005>
- Hurst, L. M., G.T. . (1996). Clade Selection, Reversible Evolution and the Persistence of Selfish elements:: The Evolutionary Dynamics of Cutoplasmic Incompatibility. *Proc. R. Soc. London Ser. B-Biological Sci.*, *263*(104).
- Jaenike, J. (2007). Spontaneous emergence of a new *Wolbachia* phenotype. *Evolution*, *61*(9), 2244-2252. <https://doi.org/10.1111/j.1558-5646.2007.00180.x>
- Jaenike, J., Dyer, K. A., Cornish, C., & Minhas, M. S. (2006). Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. *PLoS Biol*, *4*(10), e325. <https://doi.org/10.1371/journal.pbio.0040325>
- Kaur, R., Shropshire, J. D., Cross, K. L., Leigh, B., Mansueto, A. J., Stewart, V., . . . Bordenstein, S. R. (2021). Living in the endosymbiotic world of *Wolbachia*: A centennial review. *Cell Host Microbe*, *29*(6), 879-893. <https://doi.org/10.1016/j.chom.2021.03.006>
- Kriesner, P., Hoffmann, A. A., Lee, S. F., Turelli, M., & Weeks, A. R. (2013). Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulans*. *PLoS Pathog*, *9*(9), e1003607. <https://doi.org/10.1371/journal.ppat.1003607>
- Le Clec'h, W., Braquart-Varnier, C., Raimond, M., Ferdy, J. B., Bouchon, D., & Sicard, M. (2012). High virulence of *Wolbachia* after host switching: when autophagy hurts. *PLoS Pathog*, *8*(8), e1002844. <https://doi.org/10.1371/journal.ppat.1002844>
- Lukhtanov, V. A., Kandul, N. P., Plotkin, J. B., Dantchenko, A. V., Haig, D., & Pierce, N. E. (2005). Reinforcement of pre-zygotic isolation and karyotype evolution in *Agrodiaetus* butterflies. *Nature*, *436*(7049), 385-389. <https://doi.org/10.1038/nature03704>

- Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, 46(3), 523-536.
[https://doi.org/Doi 10.2307/2413694](https://doi.org/Doi%2010.2307/2413694)
- Mayr, E. (1942). *Systematics and the Origin of Species*. Columbia University Press.
- Noor, M. A. (1995). Speciation driven by natural selection in *Drosophila*. *Nature*, 375(6533), 674-675. <https://doi.org/10.1038/375674a0>
- Nosil, P., Crespi, B. J., & Sandoval, C. P. (2003). Reproductive isolation driven by the combined effects of ecological adaptation and reinforcement. *Proc Biol Sci*, 270(1527), 1911-1918.
<https://doi.org/10.1098/rspb.2003.2457>
- Ortiz-Barrientos, D., Counterman, B. A., & Noor, M. A. (2004). The genetics of speciation by reinforcement. *PLoS Biol*, 2(12), e416. <https://doi.org/10.1371/journal.pbio.0020416>
- Ortiz-Barrientos, D., Greal, A., & Nosil, P. (2009). The genetics and ecology of reinforcement: implications for the evolution of prezygotic isolation in sympatry and beyond. *Ann N Y Acad Sci*, 1168, 156-182. <https://doi.org/10.1111/j.1749-6632.2009.04919.x>
- Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Mol Ecol*, 25(11), 2337-2360. <https://doi.org/10.1111/mec.13557>
- Pease, J. B., & Hahn, M. W. (2013). More Accurate Phylogenies Inferred from Low-Recombination Regions in the Presence of Incomplete Lineage Sorting. *Evolution*, 67(8), 2376-2384. <https://doi.org/10.1111/evo.12118>
- Pfennig, K. S., & Rice, A. M. (2014). Reinforcement generates reproductive isolation between neighbouring conspecific populations of spadefoot toads. *Proc Biol Sci*, 281(1789), 20140949. <https://doi.org/10.1098/rspb.2014.0949>
- Presgraves, D. C. (2018). Evaluating genomic signatures of "the large X-effect" during complex speciation. *Mol Ecol*. <https://doi.org/10.1111/mec.14777>

- Prout, T. (1994). Some Evolutionary Possibilities for a Microbe That Causes Incompatibility in Its Host. *Evolution*, 48(3), 909-911. <https://doi.org/10.1111/j.1558-5646.1994.tb01371.x>
- Raychoudhury, R., Baldo, L., Oliveira, D. C., & Werren, J. H. (2009). Modes of acquisition of Wolbachia: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. *Evolution*, 63(1), 165-183. <https://doi.org/10.1111/j.1558-5646.2008.00533.x>
- Schuler, H., Bertheau, C., Egan, S. P., Feder, J. L., Riegler, M., Schlick-Steiner, B. C., . . . Stauffer, C. (2013). Evidence for a recent horizontal transmission and spatial spread of Wolbachia from endemic *Rhagoletis cerasi* (Diptera: Tephritidae) to invasive *Rhagoletis cingulata* in Europe. *Mol Ecol*, 22(15), 4101-4111. <https://doi.org/10.1111/mec.12362>
- Scott Chialvo, C. H., White, B. E., Reed, L. K., & Dyer, K. A. (2019). A phylogenetic examination of host use evolution in the quinaria and testacea groups of *Drosophila*. *Mol Phylogenet Evol*, 130, 233-243. <https://doi.org/10.1016/j.ympev.2018.10.027>
- Shoemaker, D. D., Katju, V., & Jaenike, J. (1999). Wolbachia and the Evolution of Reproductive Isolation between *Drosophila Recens* and *Drosophila Subquinaria*. *Evolution*, 53(4), 1157-1164. <https://doi.org/10.1111/j.1558-5646.1999.tb04529.x>
- Sidorenko, V. S. (2009). A review of species of the *Drosophila quinaria* Loew species-group (Diptera Drosophilidae) of the Asian part of Russia and adjacent countries. *Entomological Review*, 88(1), 143-149. <https://doi.org/https://doi.org/10.1134/S0013873809010126>
- St John, M. E., & Fuller, R. C. (2021). Asymmetric reinforcement in *Lucania killifish*: assessing reproductive isolation when both sexes choose. *Curr Zool*, 67(2), 215-224. <https://doi.org/10.1093/cz/zoaa049>

- Turelli, M., Cooper, B. S., Richardson, K. M., Ginsberg, P. S., Peckenpaugh, B., Antelope, C. X., . . . Hoffmann, A. A. (2018). Rapid Global Spread of wRi-like Wolbachia across Multiple *Drosophila*. *Curr Biol*, *28*(6), 963-971 e968.
<https://doi.org/10.1016/j.cub.2018.02.015>
- Turelli, M., & Hoffmann, A. A. (1995). Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. *Genetics*, *140*(4), 1319-1338. <https://doi.org/10.1093/genetics/140.4.1319>
- Turelli, M., Hoffmann, A. A., & McKechnie, S. W. (1992). Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics*, *132*(3), 713-723. <https://doi.org/10.1093/genetics/132.3.713>
- Urbanelli, S., & Porretta, D. (2008). Evidence of reinforcement of premating isolation between two species of the genus *Ochthebius* (Coleoptera: Hydraenidae). *Evolution*, *62*(6), 1520-1527. <https://doi.org/10.1111/j.1558-5646.2008.00381.x>
- Weinert, L. A., Araujo-Jnr, E. V., Ahmed, M. Z., & Welch, J. J. (2015). The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc Biol Sci*, *282*(1807), 20150249. <https://doi.org/10.1098/rspb.2015.0249>
- Werren, J. H., Baldo, L., & Clark, M. E. (2008). Wolbachia: master manipulators of invertebrate biology. *Nat Rev Microbiol*, *6*(10), 741-751. <https://doi.org/10.1038/nrmicro1969>
- Werren, J. H., & Jaenike, J. (1995). Wolbachia and cytoplasmic incompatibility in mycophagous *Drosophila* and their relatives. *Heredity (Edinb)*, *75* (Pt 3), 320-326.
- Werren, J. H., & Windsor, D. M. (2000). Wolbachia infection frequencies in insects: evidence of a global equilibrium? *Proc Biol Sci*, *267*(1450), 1277-1285.
<https://doi.org/10.1098/rspb.2000.1139>

- Wheeler, M. R. (1960). New Species of the Quinaria Group of *Drosophila* (Diptera, Drosophilidae). *The Southwestern Naturalist*, 5(3), 160-164.
- Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet*, 18(2), 87-100. <https://doi.org/10.1038/nrg.2016.133>
- Yen, J. H., & Barr, A. R. (1971). New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. *Nature*, 232(5313), 657-658. <https://doi.org/10.1038/232657a0>
- Yukilevich, R. (2012). Asymmetrical patterns of speciation uniquely support reinforcement in *Drosophila*. *Evolution*, 66(5), 1430-1446. <https://doi.org/10.1111/j.1558-5646.2011.01534.x>
- Zhou, X. F., & Li, Z. X. (2016). Establishment of the cytoplasmic incompatibility-inducing *Wolbachia* strain wMel in an important agricultural pest insect. *Sci Rep*, 6, 39200. <https://doi.org/10.1038/srep39200>
- Zug, R., & Hammerstein, P. (2015). Bad guys turned nice? A critical assessment of *Wolbachia* mutualisms in arthropod hosts. *Biol Rev Camb Philos Soc*, 90(1), 89-111. <https://doi.org/10.1111/brv.12098>

CHAPTER 2

ONGOING HYBRIDIZATION OBSCURES SPECIES RELATIONSHIPS IN THE

DROSOPHILA SUBQUINARIA SPECIES COMPLEX¹

¹ Ginsberg, P. S., Humphreys, D. P., & Dyer, K. A. (2019). Ongoing hybridization obscures phylogenetic relationships in the *Drosophila subquinaria* species complex. *J Evol Biol*, 32(10), 1093-1105. doi:10.1111/jeb.13512, Reprinted here with permission of the publisher.

Abstract

Inferring evolutionary relationships among recently diverged lineages is necessary to understand how isolating barriers produce independent lineages. Here we investigate the phylogenetic relationships between three incompletely isolated and closely related mushroom-feeding *Drosophila* species. These species form the *Drosophila subquinaria* species complex and consist of one Eurasian species (*D. transversa*) and two widespread North American species (*D. subquinaria* and *D. recens*) that are sympatric in central Canada. While patterns of pre- and post-mating isolation among these species are well characterized, previous work on their phylogenetic relationships is limited and conflicting. In this study, we generated a multi-locus dataset of 29 loci from across the genome sequenced in a population sample from each species, and then we inferred species relationships and patterns of introgression. We find strong statistical support that *D. subquinaria* is paraphyletic, showing that samples from the geographic region sympatric with *D. recens* are most closely related to *D. recens* whereas samples from the geographic region allopatric with *D. recens* are most closely related to *D. transversa*. We present several lines of evidence that both incomplete lineage sorting and gene flow are causing phylogenetic discordance. We suggest that ongoing gene flow primarily from *D. recens* into *D. subquinaria* in the sympatric part of their ranges causes phylogenetic uncertainty in the evolutionary history of these species. Our results highlight how population genetic data can be used to disentangle the sources of phylogenetic discordance among closely related species.

Introduction

Resolving species relationships is necessary to understand how reproductive isolating mechanisms generate and maintain species boundaries. Closely related species often show

incomplete pre- and post-mating isolation, which allows the process of reproductive isolation to be studied (Coyne & Orr, 2004). However, hybridization and subsequent introgression between species as well as incomplete lineage sorting (ILS) due to large ancestral effective population sizes can obscure the phylogenetic signal among groups and render evolutionary relationships unclear (Degnan & Rosenberg, 2009; Edwards, Potter, Schmitt, Bragg, & Moritz, 2016; Hahn & Nakhleh, 2016; Maddison, 1997). Recent phylogenetic methods use genome-wide approaches and incorporate population genetic data in order to resolve species relationships (Edwards, 2009; Wolf & Ellegren, 2017). In addition, these analyses can make use of parts of the genome that are differently affected by evolutionary processes (Pease & Hahn, 2013). For instance, sex chromosomes are often overrepresented in speciation-related processes (i.e. the “large X-effect”) and often show greater divergence among groups than do highly recombining regions of the autosomes (reviewed in Payseur & Rieseberg, 2016; Presgraves, 2018).

Here we study the evolutionary relationships of three closely related and incompletely isolated *Drosophila* species: *D. subquinaria*, *D. recens*, and *D. transversa*. All three species are morphologically identical and can only be distinguished via the internal male genitalia (Wheeler, 1960). *D. recens* and *D. subquinaria* occur in eastern and western North America, respectively, and their ranges overlap for approximately 1,500 km in central Canada (Figure 2.1; Dyer, Bewick, White, Bray, & Humphreys, 2018; Jaenike, Dyer, Cornish, & Minhas, 2006). *D. transversa* occurs in central and northern Europe, and is thought to occur eastward to the Kamchatka Peninsula in eastern Russia (Sidorenko, 2009). These species are in the quinaria group of the subgenus *Drosophila*, and all of them are generalists on fleshy basidiomycete mushrooms. Where they co-occur, *D. recens* and *D. subquinaria* can be collected at the same mushroom baits.

In the laboratory, all three species are able to hybridize and produce fertile F1 females. Crosses between *D. subquinaria* and *D. transversa* also produce fertile F1 males, while hybrid males resulting from crosses between *D. recens* and either *D. subquinaria* or *D. transversa* are sterile (Humphreys, Rundle, & Dyer, 2016; Shoemaker, Katju, & Jaenike, 1999). Additional hybrid death results from a *Wolbachia* infection that is at high frequency in *D. recens* (Jaenike, Dyer, Cornish, & Minhas, 2006; Shoemaker, Katju, & Jaenike, 1999; Werren & Jaenike, 1995). When a *Wolbachia*-infected *D. recens* male mates with either a *D. subquinaria* or *D. transversa* female, most of the offspring die as embryos due to cytoplasmic incompatibility (CI) (Humphreys, Rundle, & Dyer, 2016; Shoemaker, Katju, & Jaenike, 1999). *Wolbachia* does not affect offspring production in reciprocal crosses between *D. recens* females and males from the other two species (Humphreys, Rundle, & Dyer, 2016).

Patterns of premating isolation are partially consistent with the patterns of postzygotic isolation. In the laboratory, *D. recens* females mate at a moderate rate with both *D. subquinaria* and *D. transversa* (Humphreys, Rundle, & Dyer, 2016; Jaenike, Dyer, Cornish, & Minhas, 2006). In contrast, *D. subquinaria* females from populations sympatric with *D. recens* east of the Canadian Rockies discriminate strongly against mating with *D. recens* males, whereas *D. subquinaria* females from western regions discriminate less (Dyer, Bewick, White, Bray, & Humphreys, 2018; Bewick & Dyer, 2014; Jaenike, Dyer, Cornish, & Minhas, 2006). Female mate discrimination has evolved such that these “sympatric” *D. subquinaria* females also discriminate against conspecific males from western populations that are allopatric with *D. recens*. These patterns are consistent with both classical reinforcement (against *D. recens*) and cascade reinforcement (against allopatric *D. subquinaria*) given there is no post-mating isolation in conspecific crosses using *D. subquinaria* from different regions. Previous research found no

behavioral isolation between *D. transversa* and *D. subquinaria* from the western part of its range, though *D. subquinaria* females from the eastern part of its range discriminate against *D. transversa* males (Humphreys, Rundle, & Dyer, 2016).

Some hybridization occurs between *D. recens* and *D. subquinaria* in the wild. First, roughly 3% of wild-caught *D. subquinaria* harbor a mtDNA haplotype from *D. recens*, indicative of hybridization between *D. recens* females and *D. subquinaria* males and subsequent introgression of the mtDNA into *D. subquinaria* (Jaenike, Dyer, Cornish, & Minhas, 2006; Shoemaker, Katju, & Jaenike, 1999) Second, microsatellite genotyping also suggests past introgression at nuclear loci, and most of it appears to occur in the geographic regions where these species co-occur (Dyer, Bewick, White, Bray, & Humphreys, 2018). Genetic differentiation within each species is consistent with variation in behavioral isolation: previous work suggests there is very little genetic differentiation among populations of *D. recens*, while there is moderate differentiation among populations of *D. subquinaria* that are allopatric and sympatric with *D. recens* (Dyer, Bewick, White, Bray, & Humphreys, 2018).

In spite of extensive study of the patterns of reproductive isolation among these three species, their evolutionary relationships remain unresolved. Based on both the presence of hybrid male sterility of *D. recens* with *D. subquinaria* and *D. transversa* and the lack of pre-mating isolation between *D. subquinaria* and *D. transversa*, we hypothesize that *D. subquinaria* and *D. transversa* are most closely related to each other with *D. recens* as the outgroup. An alternate hypothesis is based on geography, where *D. transversa* is the basal group because it occurs in the old world, while *D. recens* and *D. subquinaria* are most closely related because both are new world species and previous studies suggested they were separated only recently during the Wisconsin glaciation 75 ka – 10 ka (Fulton, 1986; Jaenike, Dyer, Cornish, & Minhas, 2006;

Shoemaker, Katju, & Jaenike, 1999). To date, population genetic studies have only included *D. recens* and *D. subquinaria* and not *D. transversa*, and have genotyped microsatellites rather than DNA sequence data, making evolutionary inferences difficult. A recent phylogenetic analysis of the entire quinaria species group only sampled one to two alleles per species and found no consistent relationship among these three species (Scott Chialvo, White, Reed, & Dyer, 2019). The lack of consistency across loci suggested hybridization and introgression or ILS among these species.

Here we use population genetic samples from each of these three species to investigate their evolutionary relationships. We Sanger sequence 29 loci from throughout the genome from a sample of each species, and then we use phylogenetic and population genetic analyses to investigate both species relationships and patterns of introgression among species. We compare patterns of differentiation among different regions of the genome and we interpret our findings in light of known patterns of reproductive isolation. Our results suggest that ongoing hybridization between *D. recens* and *D. subquinaria* is common where they co-occur and that this obscures the phylogenetic signal among species.

Methods

Drosophila Samples and DNA Sequencing

Laboratory isofemale lines were established from single wild-caught females as described previously (Bewick & Dyer 2014; Humphreys, Rundle, & Dyer, 2016). We used 20 lines of *D. recens* from seven populations and 25 lines of *D. subquinaria* from eight populations (Figure 2.1, Table 2.1). Populations spanned the species ranges and included geographic areas where these species are both sympatric and allopatric with each other, and we hereafter refer to these as

“sympatric” and “allopatric” groups. *D. transversa* was collected from two locations, and included a single line kindly provided by John Jaenike from Lahti, Finland and six lines from Uppsala, Sweden. We also included one line each of two related outgroup quinaria group species, *D. quinaria* and *D. suboccidentalis* (Table 2.1). For one locus, *scully*, we used a single *D. palustris* sample for the outgroup species, as this region failed to amplify for the other outgroup species.

DNA was extracted from single flies using the Puregene DNA extraction kit (Qiagen, Germantown, MD). We sampled 29 loci that span all of the Müller elements based on synteny with *D. melanogaster* (Gramates et al., 2017) and included 19 autosomal loci, seven X-linked loci, two Y-linked loci, and a mtDNA locus (Table S2.1). DNA sequencing was performed at the UGA Georgia Genomic and Bioinformatics Center on a 3730xl DNA Analyzer (Applied Biosystems, Foster City, CA). Sequences were assembled and verified in Sequencher 5.0.1 (Gene Codes, Ann Arbor MI) and aligned using the *D. melanogaster* translated protein sequence (Gramates et al., 2017) as a guide in Geneious 10.1.2 (Kearse et al., 2012). We excluded all intronic sequences from our analyses because they could not always be aligned with confidence. In addition, we excluded a repetitive coding region of the *elav* locus due to alignment issues. DNA sequences have been deposited in Genbank (Accessions MN262233 - MN262437, MN239208 - MN239406, MN267907 - MN268318, MN273787 - MN274430).

In some instances, sites were heterozygous as evidenced by double peaks on the chromatograms. These were left as ambiguous for the phylogenetic analyses. For estimates of divergence and genetic differentiation, sequences were phased into haplotypes using PHASE 2.1.1 (Stephens & Donnelly, 2003; Stephens, Smith, & Donnelly, 2001) and seqPHASE (Flot, 2010), and we randomly sampled one of the two haplotypes per individual.

Phylogenetic inference of Bayesian gene and species trees

We inferred a gene tree for each locus with the exception of the two Y-chromosome loci, for which we generated a single gene tree using their concatenated alignment. To infer gene trees we used MrBayes 3.6.1 (Huelsenbeck & Ronquist, 2001). We used the best-fitting model of nucleotide substitution for each locus based on the lowest delta AIC score, determined using PartitionFinder2 (Guindon et al., 2010; Guindon & Gascuel, 2003; Lanfear, Calcott, Ho, & Guindon, 2012). We ran each locus in two independent chains of three million generations, sampling every 1000 generations, with the average standard deviation of split frequencies calculated every 1000 generations. Each run had three additional heated chains to improve mixing. Acceptable effective sample sizes (greater than 200), proper mixing, and convergence were assessed visually in Tracer 1.6.0 (Rambaut, 2018). The first 500,000 iterations were excluded as burn-in.

We generated Bayesian phylogenies from three concatenated datasets: nuclear loci (autosomal, X-linked, and Y-linked: n=27 loci), autosomal loci (n=18 loci), and X-linked loci (n=7 loci). We ran an additional analysis with the concatenated nuclear dataset that excluded the sympatric *D. subquinaria* samples. Based on our initial results from the gene trees, one locus (*plexA*) produced particularly strong and discordant topological resolution from the others and was excluded from the remaining analyses. The outgroup taxon used only for the *scully* alignment was also removed from the concatenated dataset, and so the outgroup taxon for this partition was treated as missing data. Each analysis was partitioned by locus with a GTR + Γ model of nucleotide substitution with partitions unlinked. Each concatenated dataset was analyzed both with the topology free to vary and with the sympatric *D. subquinaria* samples constrained to be monophyletic. Free and topologically constrained models were compared using

log Bayes factors (LBF), with marginal likelihoods for each model estimated via stepping-stone sampling. All concatenated analyses used two runs of four chains each, and a chain length of 5 million generations, with runs stopping once the average standard deviation of split frequencies fell below 0.01, and a relative burn-in of 25%.

Inference of phylogenetic discordance

Discordance between the most probable unconstrained concatenated trees and our hypothesized species tree (i.e., based on reproductive isolation) prompted us to test whether conflicting phylogenetic signal resulted from sympatric *D. subquinaria*. To investigate support for the conflicting topological branching patterns we used the program quartetsampling (Pease, Brown, Walker, Hinchliff, & Smith, 2018) on our consensus trees. For each analysis we ran 400 replicates per internal branch. Quartet sampling statistics describe the frequency of quartet patterns sampled from the data for all internal branches. We were most interested in QC, the frequency of quartets sampled that are concordant with the consensus tree, and QD, the relative proportion of discordant alternative quartets, where a score of one represents equal proportions of discordant quartets.

As a second method to examine conflicting tree topologies, we used the multi-species coalescent approach implemented in *BEAST 2.5.1 (Bouckaert et al., 2014) to simultaneously estimate gene trees and the species tree. This analysis included all nuclear loci and used an HKY + Γ substitution model for each locus partition. Each partition was assigned a strict clock prior with a Yule process and a constant population size prior. Taxa were grouped as *D. recens*, allopatric *D. subquinaria*, sympatric *D. subquinaria*, and *D. transversa*. The analysis was run with three independent chains for 100 million generations, and then combined using the

LogCombiner function in Beast 2.5.1. Adequate convergence, mixing and ESS values were assessed visually in Tracer 1.6.0 (Rambaut, 2018).

Genetic Divergence and Polymorphism

Population genetic parameter estimates were calculated for each locus in DnaSP 6.11 (Librado & Rozas, 2009). We divided samples into four groups: *D. recens*, sympatric *D. subquinaria*, allopatric *D. subquinaria*, and *D. transversa*. At each locus and for each pairwise group comparison we inferred genetic divergence (synonymous divergence [K_s] and net divergence [D_a]) and genetic differentiation (K_{ST} and the number of fixed and shared differences). For each locus and group we also estimated the segregating synonymous genetic variation (π_{syn} and θ_{syn}) and deviation from the site frequency spectrum (Tajima's D). To test for variation in divergence, genetic differentiation, polymorphism, and the site frequency spectrum we used an analysis of variance with genome region (X-chromosome versus autosome), group comparison, and their interaction as fixed effects in the model, followed by post-hoc Tukey HSD tests. To test for differences in the number of fixed differences vs. shared polymorphisms among groups we used a Pearson's Chi-squared test across all groups, then for each pairwise comparison, and assessed significance with a Bonferroni-corrected P -value. R was used for statistical analyses (R Team, 2008).

Inference of introgression and ongoing gene flow

We tested for introgression between *D. recens* and sympatric *D. subquinaria* using an ABBA-BABA test as implemented in the R package evobiR (Blackmon, 2015). We generated majority rule consensus sequence for each X-linked and autosomal locus from *D. recens*,

sympatric *D. subquinaria*, and allopatric *D. subquinaria*. We used the outgroup species *D. quinaria* as the fourth taxon. In order to fit the conventional ABBA-BABA test model, *D. transversa* was not included in this analysis. We note that the ABBA-BABA assumes no linkage disequilibrium (LD) among sites, which our data violate because some loci have more than one site included in the analyses (Table S2.2). However, LD breaks down within a few hundred bases in *Drosophila* and thus this does not bias our results.

To investigate the strength and direction of introgression between *D. recens* and *D. subquinaria*, we estimated migration rates and effective population sizes using the Bayesian implementation of the program Migrate-n 3.6.11 (Beerli, 2006, 2009; Beerli & Felsenstein, 2001; Beerli & Palczewski, 2010). We included two groups, which consisted of all samples of *D. recens* and *D. subquinaria*, respectively. Using more groups never resulted in model convergence. We analyzed autosomal and X-linked loci separately and tested four models of gene flow between *D. recens* and *D. subquinaria*. In the full model all parameters were free to vary, and nested models included migration from *D. subquinaria* to *D. recens* set to zero, migration from *D. recens* to *D. subquinaria* set to zero, and symmetric migration between *D. subquinaria* and *D. recens*. Each analysis was run with a four-chain heating scheme using the default temperatures of 1.0, 1.2, 3.0, and 1000000.0. Median parameter estimates from the initial run were then used as starting parameters for a second analysis. A total of three subsequent runs were examined for similar posterior distributions of parameter estimates. If convergence was inadequate, chains were extended until consistent results were obtained. Models were compared with Bezier-approximated log marginal-likelihoods and log Bayes factors.

Results

DNA sequences

Our final dataset consisted of 29 loci sequenced from 25 *D. subquinaria*, 20 *D. recens*, and seven *D. transversa* lines as well as from two outgroups, *D. quinaria* and *D. suboccidentalis*. This included 17,916 bp per sample, of which 776 were variable and 348 were parsimony informative. Overall, our dataset contains less than 5% missing data (Table S2.3).

Bayesian analyses of individual gene trees

Nearly all gene trees are highly discordant and rarely depict monophyletic species (Figure S2.1). The only gene tree that shows reciprocal monophyly for all three species is the Y-chromosome tree (Figure 2.2A), where *D. subquinaria* and *D. transversa* are sister species with *D. recens* as the outgroup. The mtDNA *COI* locus has been shown to contain *D. subquinaria* samples in the *D. recens* clade due to presumed hybridization and subsequent mtDNA introgression (Bewick & Dyer, 2014; Jaenike et al., 2006), and one of our *D. subquinaria* samples shows this pattern (Figure 2.2B). Excluding this sample, *D. recens* is monophyletic and basal to the two *D. subquinaria* clades and *D. transversa*, which together form a star phylogeny. The relationship of *D. subquinaria* samples recapitulate previously described patterns with a division at the coast mountains (Bewick & Dyer, 2014; Jaenike et al., 2006).

The lack of reciprocal monophyly at any X-linked or autosomal locus suggests either that individual loci contain too few phylogenetically informative sites or that phylogenetic signal is disrupted by ILS and/or heterospecific gene flow. We note that *plexA* was so discordant that we removed it from all downstream analyses (Figure S2.1R). Specifically, the *D. recens* samples grouped together with a greatly extended basal branch and extensive variation among samples,

whereas *D. subquinaria* and *D. transversa* formed a single clade with very little variation. Even though the branching pattern is consistent with other loci (e.g. Y-linked loci), the long branch to *D. recens* skewed other analyses and thus this locus was removed.

Concatenated Bayesian analyses

In all three Bayesian analyses (nuclear, autosomes, X-linked loci) using concatenated datasets with unconstrained topologies (Figures 2.3A, S2.2A, S2.2C), *D. recens* samples form a derived monophyletic lineage that originates from a paraphyletic clustering of sympatric *D. subquinaria*. In these phylogenies, allopatric *D. subquinaria* is sister to sympatric *D. subquinaria* and *D. recens*, and *D. transversa* is sister to all other ingroup species. For the nuclear dataset, constraining monophyly of the sympatric *D. subquinaria* samples results in a phylogeny where *D. recens* is basal and *D. subquinaria* is paraphyletic with allopatric *D. subquinaria* sister to *D. transversa* (Figure 2.3B), though this model has lower support than the topologically free model (2LBF = 41.9). In contrast, the autosome and X-linked constrained phylogenies suggest sympatric *D. subquinaria* is sister to *D. recens* (Figures S2.2B, S2.2D). When sympatric *D. subquinaria* is removed from the analysis entirely, allopatric *D. subquinaria* and *D. transversa* are sister ingroup species and *D. recens* is more distantly related (Figure S2.3).

Phylogenetic discordance

We conducted a quartet sampling analysis on the most probable phylogeny from the topologically constrained nuclear loci dataset that suggests *D. recens* is outgroup to *D. subquinaria* and *D. transversa* (Figure 2.3B). At our branch of interest, which delimits the relationship between sympatric *D. subquinaria* and the remaining populations/species, there was

very high discordance in quartet topologies (Figure 2.4A, QC = -0.43), with most of the discordant topologies having sympatric *D. subquinaria* sister to *D. recens* (Figure 2.4C, QD = 0.008). This bias suggests that gene flow rather than ILS is the cause of the discordance, as ILS is expected to produce equal proportions of the two discordant topologies (Pease et al., 2018). To investigate how the inclusion of sympatric *D. subquinaria* affects tree discordance we also conducted a quartet analysis using a tree that excludes this group. Here, we find strong support that *D. recens* is the outgroup of allopatric *D. subquinaria* and *D. transversa* (Figure 2.4B, QC = 0.69). Importantly, there was minimal skew in discordant quartet topologies (Figure 2.4D, QD = 0.88), as expected if the discordant topologies are due to ILS (Pease et al., 2018). Taken together, these results suggest that introgression between *D. recens* and sympatric *D. subquinaria* causes discordance in the species phylogeny.

The phylogenetic inference using a multi-species coalescent implemented in *BEAST supports a phylogeny where *D. recens* is the most distantly related ingroup species, with allopatric *D. subquinaria* and *D. transversa* as sister groups and sympatric *D. subquinaria* sister to them both (Figure 2.5). All branches had >0.94 posterior support. This is the same branching pattern as in the constrained topology inferred in the Bayesian analyses for all nuclear loci (Figure 2.3B). Thus, the tree topology resolved with this method supports the hypothesis that *D. recens* is not sister to sympatric *D. subquinaria*, but that incomplete lineage sorting and/or gene flow convolute phylogenetic signal for the “correct” species tree. We note that this method does not incorporate the effect of migration among groups on the species tree.

Estimates of genetic divergence and polymorphism

We calculated divergence and differentiation for each pairwise comparison between sympatric *D. subquinaria*, allopatric *D. subquinaria*, *D. recens*, and *D. transversa*. We combined the sympatric and allopatric populations of *D. recens* because there were no fixed differences between them and estimates of genetic differentiation and divergence were very low (average $D_a = 0.00053$ and $K_{ST} = 0.021$; Table S2.4). Consistent with the patterns observed in the phylogenetic analyses, estimates of divergence and differentiation between sympatric and allopatric *D. subquinaria* were as high as inter-species values (average $D_a = 0.0043$ and $K_{ST} = 0.179$; Table S2.4). Across groups, net divergence (D_a) is higher at the X-linked loci compared to autosomal loci ($F_{1,138} = 98.4$, $P < 0.001$). Within each genome region, D_a between *D. recens* and both *D. transversa* and allopatric *D. subquinaria* is elevated relative to the other comparisons (among groups: $F_{1,138} = 98.4$, $P < 0.001$; Figure 2.6). Synonymous divergence (K_s) is also generally higher among X-linked loci than autosomal loci ($F_{1,138} = 13.9$, $P = 0.0002$; Figure S2.4, Table S2.4), as well as in comparisons between *D. recens* and other groups ($F_{5,138} = 5.6$, $P < 0.0001$). Similar to measures of divergence, genetic differentiation (K_{ST}) was also higher at X-linked loci compared to the autosomal loci ($F_{1,138} = 37.6$, $P < 0.0001$; Figure 2.7, Table S2.4) and higher in group comparisons of *D. recens* with *D. transversa* and allopatric *D. subquinaria* ($F_{5,138} = 3.9$, $P = 0.0026$).

Using data combined from the X-linked and autosomal loci, the ratio of shared polymorphisms to fixed differences was significantly different among groups ($\chi^2_5 = 62.2$, $P < 0.001$; Figure 2.8). These differences remain significant when testing X-linked and autosomal loci separately (X-linked: $\chi^2_5 = 31.0$, $P < 0.001$; Autosomes: $\chi^2_5 = 23.7$, $P < 0.001$, Figure S2.5). These differences are driven by the greater number of fixed differences in the comparisons of *D.*

recens with either *D. transversa* and allopatric *D. subquinaria* (Figure 2.8, Tables S2.5 & S2.6). In contrast, the ratio of fixed differences to shared polymorphisms between *D. recens* and sympatric *D. subquinaria* did not differ from the levels seen between *D. transversa* and *D. subquinaria*. The reduced number of fixed differences and increased number of shared polymorphisms between *D. recens* and sympatric *D. subquinaria* is consistent with introgression between these groups.

Synonymous polymorphism ranges from 0.03-0.08 within groups (Figure S2.6A), and is highest for *D. recens* and sympatric *D. subquinaria* and lowest for allopatric *D. subquinaria* and *D. transversa*. These values suggest high effective population sizes (N_e) of these species. Pairwise nucleotide diversity (π_{syn}) is significantly different among the four groups ($F_{3,96} = 3.5$, $P = 0.018$), though using a post-hoc Tukey test does not result in any significant pairwise differences. No group deviates from the neutral expectation of a 25% reduction in π_{syn} for the X-chromosome compared to the autosomes (t -test, all $P > 0.05$). Watterson's θ_{syn} also varies significantly among the four groups ($F_{3,74} = 9.3$, $P < 0.0001$), with *D. recens* having a higher θ_{syn} than the others (Figure S2.6B). In *D. transversa*, loci on the X-chromosome have a higher than expected θ_{syn} relative to the autosomes (t -test, *D. transversa* $t_5 = -8.9$, $P = 0.0003$; others $P > 0.05$). The values of Tajima's D are generally negative, indicating an overall excess of rare variants (Figure S2.7), though very few individual loci deviate significantly from the neutral expectation (Table S2.6). While there is not significant variation in Tajima's D between X-linked and autosomal loci ($F_{1,93} = 1.8$, $P = 0.18$) there is variation among groups ($F_{3,93} = 15$, $P < 0.0001$), with *D. recens* having a four-fold lower Tajima's D than the other groups (Mean *D. recens*: -1.2, range of the other three groups is -0.2 to -0.4; Figure S2.7).

Inference of introgression and ongoing gene flow

Due to reduced divergence between *D. recens* and sympatric *D. subquinaria* we used an ABBA-BABA test to infer the presence of introgression between these groups (Durand, Patterson, Reich, & Slatkin, 2011). We find an overrepresentation of shared derived alleles among *D. recens* and sympatric *D. subquinaria* compared to shared alleles between *D. recens* and allopatric *D. subquinaria* ($D = 0.474$, $ABBA = 28$, $BABA = 10$; $P = <0.001$). This suggests recent gene flow between *D. recens* and sympatric *D. subquinaria*.

To test for the directionality of this gene flow we inferred migration rates between *D. recens* and *D. subquinaria* using Migrate-n 3.6.11 (Beerli, 2006, 2009; Beerli & Felsenstein, 2001; Beerli & Palczewski, 2010). Using only autosomal loci, the most supported model had asymmetric migration ($R \leftrightarrow S$), with three-fold higher migration from *D. recens* into *D. subquinaria* than the reverse (Table 2.2). However, using only X-linked loci suggests the opposite pattern, where model selection supported gene flow only from *D. subquinaria* into *D. recens* (Table 2.2). Overall, these results suggest that introgression at the autosomes is greater from *D. recens* into *D. subquinaria*, while at the X-chromosome introgression is higher from *D. subquinaria* into *D. recens*.

Discussion

Inferring the evolutionary relationships of closely related species is necessary to interpret patterns of reproductive isolation among them. However, processes such as incomplete lineage sorting (ILS) and hybridization can obscure phylogenetic relationships and even lead to statistically supported yet incorrect phylogenies (Degnan & Rosenberg, 2006, 2009; Edwards, 2009; Phillips, Delsuc, & Penny, 2004). Here we study three closely related and incompletely

isolated species, *D. recens*, *D. subquinaria*, and *D. transversa*. Based on patterns of pre- and post-mating isolation, we hypothesize that *D. subquinaria* and *D. transversa* are most closely related with *D. recens* the more distantly related species. Previous phylogenetic studies of these species included limited samples and few loci, and have been inconclusive (Scott Chialvo et al., 2019). Because of their large effective sizes and known incidence of hybridization, we use a multi-locus approach that combines phylogenetic and population genetic analyses to investigate patterns of divergence among these three species. Overall, we find that our hypothesis of the species relationships based on patterns of reproductive isolation with *D. recens* as most basal has limited support in the phylogenetic analysis. We use several lines of evidence to suggest that historical and ongoing gene flow between *D. recens* and *D. subquinaria* in the region of geographic sympatry has reduced genetic divergence to result in incorrect phylogenetic trees.

Inferences from individual gene trees

Analyses of individual gene trees from non-recombining regions were more informative than loci from highly recombining regions. First, our results are consistent with previous findings that the mtDNA is largely concordant with patterns of reproductive isolation with the exception of clear instances of introgression from *D. recens* into *D. subquinaria* (Figure 2.2B; Bewick & Dyer, 2014; Jaenike et al., 2006; Shoemaker et al., 1999). Second, because of hybrid male sterility between *D. recens* and either *D. subquinaria* or *D. transversa*, the Y-chromosome should be impervious to introgression. Indeed, the Y-chromosome gene tree is consistent with the hypothesized species tree where *D. recens* is basal to *D. subquinaria* and *D. transversa*, which each form a monophyletic clade (Figure 2.2A). Notably, this is the only gene tree in our analyses where allopatric and sympatric *D. subquinaria* form a monophyletic clade. Third,

nuclear loci in reduced recombination regions of the dot and X-chromosome were more informative than autosomal loci. For instance, in eight of the 26 nuclear and X-linked gene trees *D. recens* formed a monophyletic group, and of these six were dot or X-linked loci (Figure S2.1). The lack of resolution among gene trees suggests incomplete lineage sorting is pervasive among all three species in the complex.

Inference of species relationships

In our concatenated Bayesian analyses we find the natural clustering of several groups, including *D. recens*, sympatric *D. subquinaria*, allopatric *D. subquinaria*, and *D. transversa*. Allopatric and sympatric samples of *D. recens* are not divergent in any of our phylogenetic or population genetic analyses, consistent with previous inferences (Dyer et al., 2018; Jaenike et al., 2006). All of our Bayesian phylogenetic inferences depict *D. recens* as derived from sympatric *D. subquinaria* (Figures 2.3A, S2.2). The relative placement of allopatric *D. subquinaria* and *D. transversa* is inconsistent across datasets, but in the nuclear loci phylogeny allopatric *D. subquinaria* are more closely related to sympatric *D. subquinaria* and *D. recens*, with *D. transversa* as the outgroup. However, when sympatric *D. subquinaria* is constrained to be a monophyletic group these patterns change, and *D. recens* is most basal, with *D. subquinaria* as paraphyletic because allopatric *D. subquinaria* is sister to *D. transversa* in the ingroup (Figure 2.3B). This constrained phylogeny that depicts *D. recens* as the most basal ingroup species is also recovered in the *BEAST phylogeny (Figure 2.5), which is the only analysis that explicitly accounts for ILS.

Given the strong intrinsic post-zygotic isolation between *D. recens* and the other two species, it is surprising that *D. recens* is derived from sympatric *D. subquinaria* in our

concatenated phylogenies. It is well known that concatenation can lead to high confidence in incorrect trees (Degnan & Rosenberg, 2006, 2009; Edwards, 2009; Phillips et al., 2004), and this may play a role in the observed phylogenetic discordance. We suggest this is not simply due to ILS, which would be expected to produce a polytomy of *D. recens* with the other groups. In support of this, excluding sympatric *D. recens* from the analyses results in a strongly supported phylogeny where *D. recens* is basal to both allopatric *D. subquinaria* and *D. transversa* (Figure S2.3). Instead, we suggest that the concatenated phylogeny does not depict the true phylogenetic relationships among the groups, and that the *BEAST and constrained topologies are more likely to reflect the true evolutionary histories.

Evolutionarily recent hybridization between *D. recens* and sympatric *D. subquinaria* could result in these groups being more similar to each other than to either allopatric *D. subquinaria* or *D. transversa*. Indeed, our population genetic analysis shows decreased divergence and differentiation of *D. recens* with sympatric *D. subquinaria* than to either allopatric *D. subquinaria* or *D. transversa*. For instance, K_{ST} between *D. recens* and sympatric *D. subquinaria* is about the same as between allopatric and sympatric *D. subquinaria* or between allopatric *D. subquinaria* and *D. transversa*. This decreased divergence is also observed in the number of shared polymorphisms, where *D. recens* and sympatric *D. subquinaria* have more shared polymorphisms than any other comparison, including between conspecific sympatric and allopatric *D. subquinaria*. The ABBA-BABA test supports the inference of gene flow between *D. recens* and sympatric *D. subquinaria*. We acknowledge that our application of the ABBA-BABA test was limited given the quantity of data and violation of the assumption of no linkage among sites, but nevertheless the results appear to be robust to these violations.

Quartet sampling provides another line of evidence for gene flow between *D. recens* and sympatric *D. subquinaria*. With all taxa included, the majority of quartets are discordant with the Bayesian consensus tree topology, and the frequency of discordant topologies is heavily skewed toward trees that place *D. recens* and sympatric *D. subquinaria* together (Figures 2.4A, 2.4C). The low concordance and skewedness of discordant topologies should be resolved by the removal of the taxa group with conflicted phylogenetic relationships to the remaining taxa (Aberer, Krompass, & Stamatakis, 2013; Wilkinson, 1996). Indeed, removing sympatric *D. subquinaria* from the analyses results in high concordance and near complete resolution of skewed frequencies of the discordant topologies (Figures 2.4, S2.3). This suggests that conflicting phylogenetic signal is uniquely enriched in the sympatric *D. subquinaria* population with respect to its close relatedness to *D. recens*, which can be caused by gene flow between these groups.

X-linked versus autosomal loci

Theoretical and empirical findings from a variety of taxa support a model where the sex-chromosome diverges faster than the autosomes (Payseur, Presgraves, & Filatov, 2018; Payseur & Rieseberg, 2016; Presgraves, 2018). Consistent with this body of work, we find that X-linked loci have higher differentiation and divergence than do the autosomes (Figure 2.6, S2.4, Table S2.4). X-linked sites contribute to the excess of ABBA sites (Table S2.2) as well as to the relative reduction of fixed differences and increase in shared polymorphisms between *D. recens* and sympatric *D. subquinaria* (Figures 2.8, S2.5). However, X-linked loci are not entirely resistant to introgression, as is indicated by the paraphyly of sympatric *D. subquinaria* in the X-

linked Bayesian phylogeny (Figure S2.2C) and the inference of some introgression from *D. subquinaria* into *D. recens* in the Migrate-n analyses with the X-linked loci (Table 2.2).

It is especially interesting that the Migrate-n analyses suggest that the autosomal loci experience more introgression from *D. recens* into *D. subquinaria*, whereas the X-linked loci suggest the opposite pattern (Table 2.2). When hybridization occurs in the wild, it is most likely that a *D. recens* female mates with a *D. subquinaria* male, because where *D. recens* are common *D. subquinaria* females discriminate strongly against *D. recens* males (Bewick & Dyer, 2014; Jaenike et al. 2006). The hybrid females are fertile and are expected to be infected with *Wolbachia*, which means that either direction of backcross is compatible and can produce F2 offspring. Our results suggest that introgression at the autosomes is more common from *D. recens* into *D. subquinaria*, suggesting that these F1 females either mate more with *D. subquinaria* males or that genetic incompatibilities are reduced in this direction relative to backcrossing with *D. recens* males. The relatively reduced introgression of the X-chromosome suggests that the *D. recens* X-chromosome is selected against in the *D. subquinaria* genetic background, which is consistent with previous findings that used microsatellite loci (Dyer et al., 2018).

Allopatric and sympatric D. subquinaria

There is strong asymmetric behavioral discrimination between *D. subquinaria* populations that are sympatric and allopatric with *D. recens*, and our study confirms previous findings of substantial genetic differentiation between these groups (Dyer et al., 2018; Bewick & Dyer, 2014). The relatively high degree of divergence between sympatric and allopatric *D. subquinaria* populations is reflected in all of our phylogenetic analyses that utilized multiple loci.

In addition, we find that divergence and differentiation is as high between sympatric and allopatric populations as it is between species (Figures 2.6, 2.7, S2.4). Synonymous divergence between sympatric and allopatric *D. subquinaria* is 0.059 (Table S2.4), which is higher than divergence between other well studied and incompletely isolated *Drosophila* species pairs, including *D. simulans* with either *D. sechellia* or *D. mauritiana* or *D. yakuba* with *D. santomea* ($K_s = 0.04-0.05$) (Turissini, Liu, David, & Matute, 2015). Another species pair in the quinaria group that shows pre-mating but not post-mating isolation is *D. suboccidentalis* and *D. occidentalis*, which have a synonymous divergence of 0.01 across 14 loci (Arthur & Dyer, 2015; Scott Chialvo et al., 2019). It is clear that *D. subquinaria* is in the process of incipient speciation, though we note introgression from *D. recens* into sympatric *D. subquinaria* may inflate levels of within-species genetic divergence. Our current study did not include samples from the geographic region between the allopatric and sympatric regions, and in this middle region the change in both behavioral isolation and genetic differentiation is gradual (Dyer et al., 2018).

D. transversa and *D. subquinaria*

Of these three species, *D. transversa* is the only one that inhabits the old world. It is thought the range of *D. transversa* extends east through Siberia to the Kamchatka Peninsula (Sidorenko, 2009) and the range of *D. subquinaria* extends north through Alaska (Wheeler, 1960). Thus, it is possible that *D. transversa* and *D. subquinaria* are sympatric and hybridize in these unexplored regions, which could explain the low level of divergence between them. *D. transversa* and allopatric *D. subquinaria* exhibit little to no pre- or post-mating reproductive isolation (Humphreys et al., 2016), but as with *D. subquinaria* and *D. recens* these barriers may be stronger in sympatry. Conversely, the low levels of reproductive isolation between *D.*

transversa and *D. subquinaria* may be the result of a recent range expansion from Asia into North America (or vice versa). However, these two species harbor a similar amount of genetic diversity and a similar degree of deviation from the site frequency spectrum (Figure S2.7). If one of them was recently derived from the other we would expect reduced polymorphism and a stronger deviation from the neutral expectation for Tajima's *D*, indicative of a recent population expansion. However, we do not see either of these patterns. Interestingly, the average synonymous divergence (*K*_s) between allopatric *D. subquinaria* and *D. transversa* is 0.049, which is lower than between sympatric and allopatric *D. subquinaria* (*K*_s = 0.059). We suggest that additional geographic sampling and assaying for more subtle forms of reproductive isolation are necessary before the taxonomic status of these species is reconsidered.

Conclusion

Species boundaries can be maintained in the face of ongoing hybridization, especially when intrinsic reproductive barriers are present (Coyne & Orr, 2004; Larson, White, Ross, & Harrison, 2014). Here we disentangle the roles of ILS and hybridization in the inference of phylogenetic relationships of closely related and incompletely isolated species. The lack of resolution among gene trees suggests ILS is pervasive across the complex. Furthermore, we show that there is more ongoing hybridization and subsequent gene flow than was previously appreciated, and that both of these processes as well as statistical issues from locus concatenation can cause incorrect phylogenetic inferences. Overall, our results support a model where *D. recens* is the oldest lineage in the *subquinaria* species complex, and that gene flow between sympatric *D. subquinaria* and *D. recens* creates discrepancies between patterns of reproductive isolation and the resolved phylogenetic topologies. Our results do not disentangle the

evolutionary relationship between *D. subquinaria* and *D. transversa*, and further studies are necessary to infer their demographic history.

Whole genomes with resequencing data will identify specific instances of introgression due to hybridization. Genomic regions that introgress freely are not expected to contain loci that are involved in reproductive isolation or local adaptation (Barton & Bengtsson, 1986; Wolf & Ellegren, 2017; Wu, 2001). Comparisons of *D. recens* with the other species may identify regions containing loci that underlie older reproductive isolating barriers including basal mate discrimination and male sterility. Comparisons of sympatric and allopatric *D. subquinaria* may indicate which genomic regions are diverging due to reinforcement of species boundaries against *D. recens* as well as among conspecific populations. Finally, whole genome data may also help to resolve the complex demographic history of *D. transversa* and *D. subquinaria*, specifically how these groups are related to each other and whether ongoing migration shapes their limited divergence.

Works Cited

- Aberer, A. J., Krompass, D., & Stamatakis, A. (2013). Pruning rogue taxa improves phylogenetic accuracy: an efficient algorithm and webservice. *Syst Biol*, 62(1), 162-166.
doi:10.1093/sysbio/sys078
- Arthur, N. J., & Dyer, K. A. (2015). Asymmetrical sexual isolation but no postmating isolation between the closely related species *Drosophila suboccidentalis* and *Drosophila occidentalis*. *BMC Evol Biol*, 15, 38. doi:10.1186/s12862-015-0328-y
- Barton, N., & Bengtsson, B. O. (1986). The barrier to genetic exchange between hybridising populations. *Heredity (Edinb)*, 57 (Pt 3), 357-376. doi:10.1038/hdy.1986.135

- Beerli, P. (2006). Comparison of Bayesian and maximum-likelihood inference of population genetic parameters. *Bioinformatics*, 22(3), 341-345. doi:10.1093/bioinformatics/bti803
- Beerli, P. (2009). How to use migrate or why are markov chain monte carlo programs difficult to use? *Population Genetics for Animal Conservation*, 17, 14-79. Cambridge, UK: Cambridge University Press.
- Beerli, P., & Felsenstein, J. (2001). Maximum likelihood estimation of a migration matrix and effective population sizes in n subpopulations by using a coalescent approach. *Proc Natl Acad Sci U S A*, 98(8), 4563-4568. doi:10.1073/pnas.081068098
- Beerli, P., & Palczewski, M. (2010). Unified framework to evaluate panmixia and migration direction among multiple sampling locations. *Genetics*, 185(1), 313-326. doi:10.1534/genetics.109.112532
- Bewick, E. R., & Dyer, K. A. (2014). Reinforcement shapes clines in female mate discrimination in *Drosophila subquinaria*. *Evolution*, 68(11), 3082-3094. doi:10.1111/evo.12515
- Blackmon, H. a. R. A. A. (2015). EvobiR: Tool for comparative analyses and teaching evolutionary biology. doi:10.5281/zenodo.30938
- Bouckaert, R., Heled, J., Kuhnert, D., Vaughan, T., Wu, C. H., Xie, D., . . . Drummond, A. J. (2014). BEAST 2: a software platform for Bayesian evolutionary analysis. *PLoS Comput Biol*, 10(4), e1003537. doi:10.1371/journal.pcbi.1003537
- Coyne, J. A., & Orr, H. A. (2004). *Speciation*: Sinauer Associates Inc.
- Degnan, J. H., & Rosenberg, N. A. (2006). Discordance of species trees with their most likely gene trees. *PLoS Genet*, 2(5), e68. doi:10.1371/journal.pgen.0020068

- Degnan, J. H., & Rosenberg, N. A. (2009). Gene tree discordance, phylogenetic inference and the multispecies coalescent. *Trends Ecol Evol*, 24(6), 332-340.
doi:10.1016/j.tree.2009.01.009
- Durand, E. Y., Patterson, N., Reich, D., & Slatkin, M. (2011). Testing for ancient admixture between closely related populations. *Mol Biol Evol*, 28(8), 2239-2252.
doi:10.1093/molbev/msr048
- Dyer, K. A., Bewick, E. R., White, B. E., Bray, M. J., & Humphreys, D. P. (2018). Fine-scale geographic patterns of gene flow and reproductive character displacement in *Drosophila subquinaria* and *Drosophila recens*. *Mol Ecol*. doi:10.1111/mec.14825
- Dyer, K. A., White, B. E., Sztepanacz, J. L., Bewick, E. R., & Rundle, H. D. (2014). Reproductive character displacement of epicuticular compounds and their contribution to mate choice in *Drosophila subquinaria* and *Drosophila recens*. *Evolution*, 68(4), 1163-1175. doi:10.1111/evo.12335
- Edwards, S. V. (2009). Is a new and general theory of molecular systematics emerging? *Evolution*, 63(1), 1-19. doi:10.1111/j.1558-5646.2008.00549.x
- Edwards, S. V., Potter, S., Schmitt, C. J., Bragg, J. G., & Moritz, C. (2016). Reticulation, divergence, and the phylogeography-phylogenetics continuum. *Proc Natl Acad Sci U S A*, 113(29), 8025-8032. doi:10.1073/pnas.1601066113
- Flot, J. F. (2010). seqphase: a web tool for interconverting phase input/output files and fasta sequence alignments. *Mol Ecol Resour*, 10(1), 162-166. doi:10.1111/j.1755-0998.2009.02732.x
- Fulton, R. J. F., M. M.; Rutter N. W. . (1986). Summary of Quarternary stratigraphy and history, Western Canada. *Quarternary Science Reviews*, 5, 229-241.

- Giglio, E. M., & Dyer, K. A. (2013). Divergence of premating behaviors in the closely related species *Drosophila subquinaria* and *D. recens*. *Ecol Evol*, 3(2), 365-374.
doi:10.1002/ece3.477
- Gramates, L. S., Marygold, S. J., Santos, G. D., Urbano, J. M., Antonazzo, G., Matthews, B. B., . . . the FlyBase, C. (2017). FlyBase at 25: looking to the future. *Nucleic Acids Res*, 45(D1), D663-D671. doi:10.1093/nar/gkw1016
- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*, 59(3), 307-321. doi:10.1093/sysbio/syq010
- Guindon, S., & Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst Biol*, 52(5), 696-704.
- Hahn, M. W., & Nakhleh, L. (2016). Irrational exuberance for resolved species trees. *Evolution*, 70(1), 7-17. doi:10.1111/evo.12832
- Huelsenbeck, J. P., & Ronquist, F. (2001). MRBAYES: Bayesian inference of phylogenetic trees. *Bioinformatics*, 17(8), 754-755.
- Humphreys, D. P., Rundle, H. D., & Dyer, K. A. (2016). Patterns of reproductive isolation in the *Drosophila subquinaria* complex: can reinforced premating isolation cascade to other species? *Current Zoology*, 62(2), 183-191. doi:10.1093/cz/zow005
- Jaenike, J., Dyer, K. A., Cornish, C., & Minhas, M. S. (2006). Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. *PLoS Biol*, 4(10), e325.
doi:10.1371/journal.pbio.0040325
- Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S., . . . Drummond, A. (2012). Geneious Basic: an integrated and extendable desktop software platform for

- the organization and analysis of sequence data. *Bioinformatics*, 28(12), 1647-1649.
doi:10.1093/bioinformatics/bts199
- Lanfear, R., Calcott, B., Ho, S. Y., & Guindon, S. (2012). Partitionfinder: combined selection of partitioning schemes and substitution models for phylogenetic analyses. *Mol Biol Evol*, 29(6), 1695-1701. doi:10.1093/molbev/mss020
- Larson, E. L., White, T. A., Ross, C. L., & Harrison, R. G. (2014). Gene flow and the maintenance of species boundaries. *Mol Ecol*, 23(7), 1668-1678.
- Librado, P., & Rozas, J. (2009). DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics*, 25(11), 1451-1452.
doi:10.1093/bioinformatics/btp187
- Maddison, W. P. (1997). Gene trees in species trees. *Systematic Biology*, 46(3), 523-536. doi:10.2307/2413694
- Payseur, B. A., Presgraves, D. C., & Filatov, D. A. (2018). Sex chromosomes and speciation. *Mol Ecol*. doi:10.1111/mec.14828
- Payseur, B. A., & Rieseberg, L. H. (2016). A genomic perspective on hybridization and speciation. *Mol Ecol*, 25(11), 2337-2360. doi:10.1111/mec.13557
- Pease, J. B., Brown, J. W., Walker, J. F., Hinchliff, C. E., & Smith, S. A. (2018). Quartet Sampling distinguishes lack of support from conflicting support in the green plant tree of life. *Am J Bot*, 105(3), 385-403. doi:10.1002/ajb2.1016
- Pease, J. B., & Hahn, M. W. (2013). More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. *Evolution*, 67(8), 2376-2384. doi:10.1111/evo.12118

- Phillips, M. J., Delsuc, F., & Penny, D. (2004). Genome-scale phylogeny and the detection of systematic biases. *Mol Biol Evol*, 21(7), 1455-1458. doi:10.1093/molbev/msh137
- Presgraves, D. C. (2018). Evaluating genomic signatures of "the large X-effect" during complex speciation. *Mol Ecol*. doi:10.1111/mec.14777
- Rambaut A, D. A., Xie D, Baele G & Suchard MA (2018). Tracer v1.7. Retrieved from <http://tree.bio.ed.ac.uk/software/tracer/>
- Scott Chialvo, C. H., White, B. E., Reed, L. K., & Dyer, K. A. (2019). A phylogenetic examination of host use evolution in the quinaria and testacea groups of *Drosophila*. *Mol Phylogenet Evol*, 130, 233-243. doi:10.1016/j.ympev.2018.10.027
- Shoemaker, D. D., Katju, V., & Jaenike, J. (1999). *Wolbachia* and the evolution of reproductive isolation between *Drosophila recens* and *Drosophila subquinaria*. *Evolution*, 53(4), 1157-1164. doi:10.1111/j.1558-5646.1999.tb04529.x
- Sidorenko, V. S. (2009). A review of species of the *Drosophila quinaria* Loew species-group (Diptera Drosophilidae) of the Asian part of Russia and adjacent countries. *Entomological Review*, 88(1), 143-149. doi:<https://doi.org/10.1134/S0013873809010126>
- Stephens, M., & Donnelly, P. (2003). A comparison of bayesian methods for haplotype reconstruction from population genotype data. *Am J Hum Genet*, 73(5), 1162-1169. doi:10.1086/379378
- Stephens, M., Smith, N. J., & Donnelly, P. (2001). A new statistical method for haplotype reconstruction from population data. *Am J Hum Genet*, 68(4), 978-989. doi:10.1086/319501
- Team, R. D. C. (2008). R: A Language and Environment for Statistical Computing *R Foundation for Statistical Computing*. Retrieved from <http://www.R-project.org>

- Turissini, D. A., Liu, G., David, J. R., & Matute, D. R. (2015). The evolution of reproductive isolation in the *Drosophila yakuba* complex of species. *J Evol Biol*, 28(3), 557-575. doi:10.1111/jeb.12588
- Werren, J. H., & Jaenike, J. (1995). *Wolbachia* and cytoplasmic incompatibility in mycophagous *Drosophila* and their relatives. *Heredity (Edinb)*, 75 (Pt 3), 320-326.
- Wheeler, M. R. (1960). New species of the Quinaria group of *Drosophila* (Diptera, Drosophilidae). *The Southwestern Naturalist*, 5(3), 160-164.
- Wilkinson, M. (1996). Majority-rule reduced consensus trees and their use in bootstrapping. *Mol Biol Evol*, 13(3), 437-444. doi:10.1093/oxfordjournals.molbev.a025604
- Wolf, J. B., & Ellegren, H. (2017). Making sense of genomic islands of differentiation in light of speciation. *Nat Rev Genet*, 18(2), 87-100. doi:10.1038/nrg.2016.133
- Wu, C.-I. (2001). The genic view of the process of speciation. *Journal of Evolutionary Biology*, 14(6), 851-865. doi:https://doi.org/10.1046/j.1420-9101.2001.00335.x

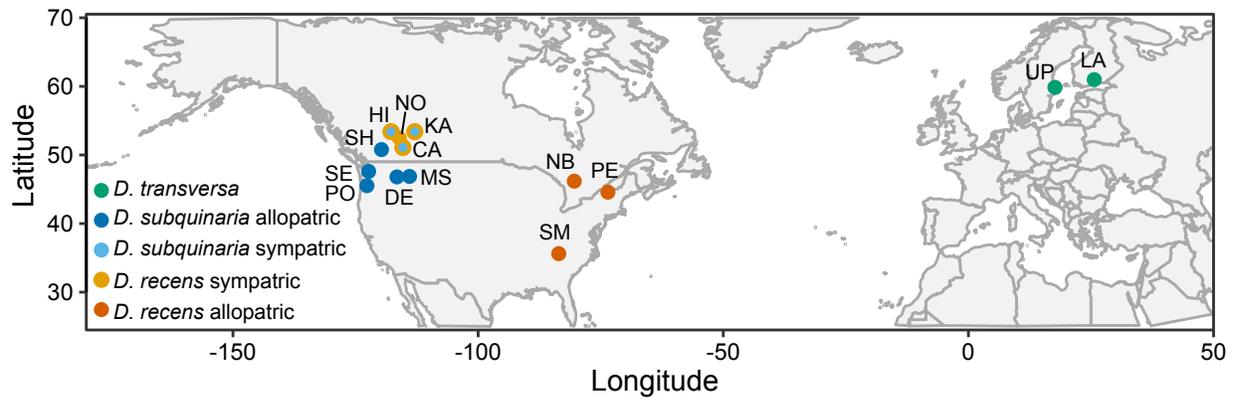


Figure 2.1: Map of sampling locations. Abbreviations are as in Table 2.1.

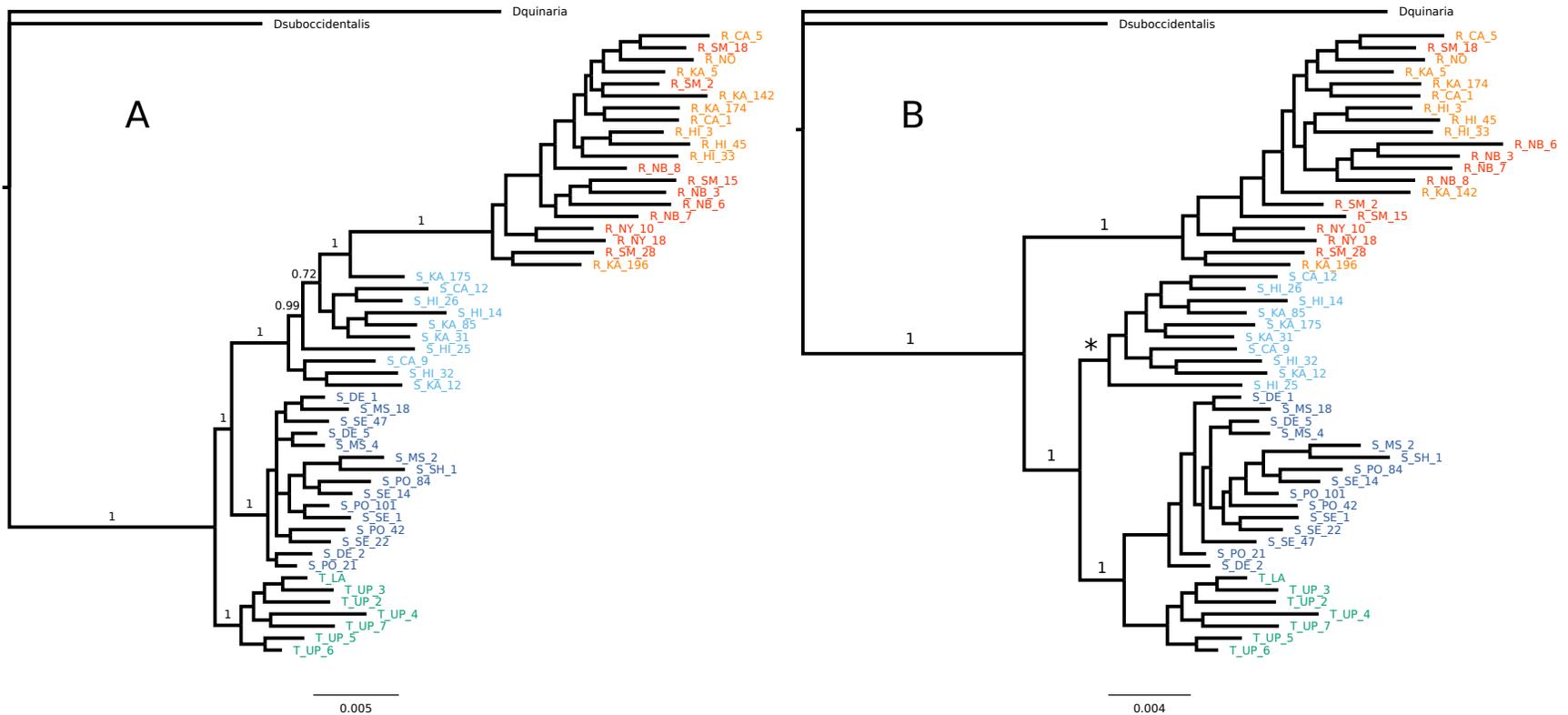


Figure 2.3: Inferred majority rule consensus trees with and without topological constraints from the concatenated Bayesian analyses using all nuclear data (autosomes, X-linked, & Y-linked loci). Panel A shows the results from the analysis with topology free to vary, and Panel B shows the results where sympatric *D. subquinaria* samples are constrained to monophyly, with the * indicating the constrained node. Labels indicate species (S_ for *D. subquinaria*, R_ for *D. recens*, T_ for *D. transversa*) followed by the population abbreviation as in Table 1 and the line number. Labels are colored as in Figure 1, where light blue indicates sympatric *D. subquinaria*, dark blue indicates allopatric *D. subquinaria*, orange indicates sympatric *D. recens*, red indicates allopatric *D. recens*, and green indicates *D. transversa*. Only the support values for the main nodes are indicated.

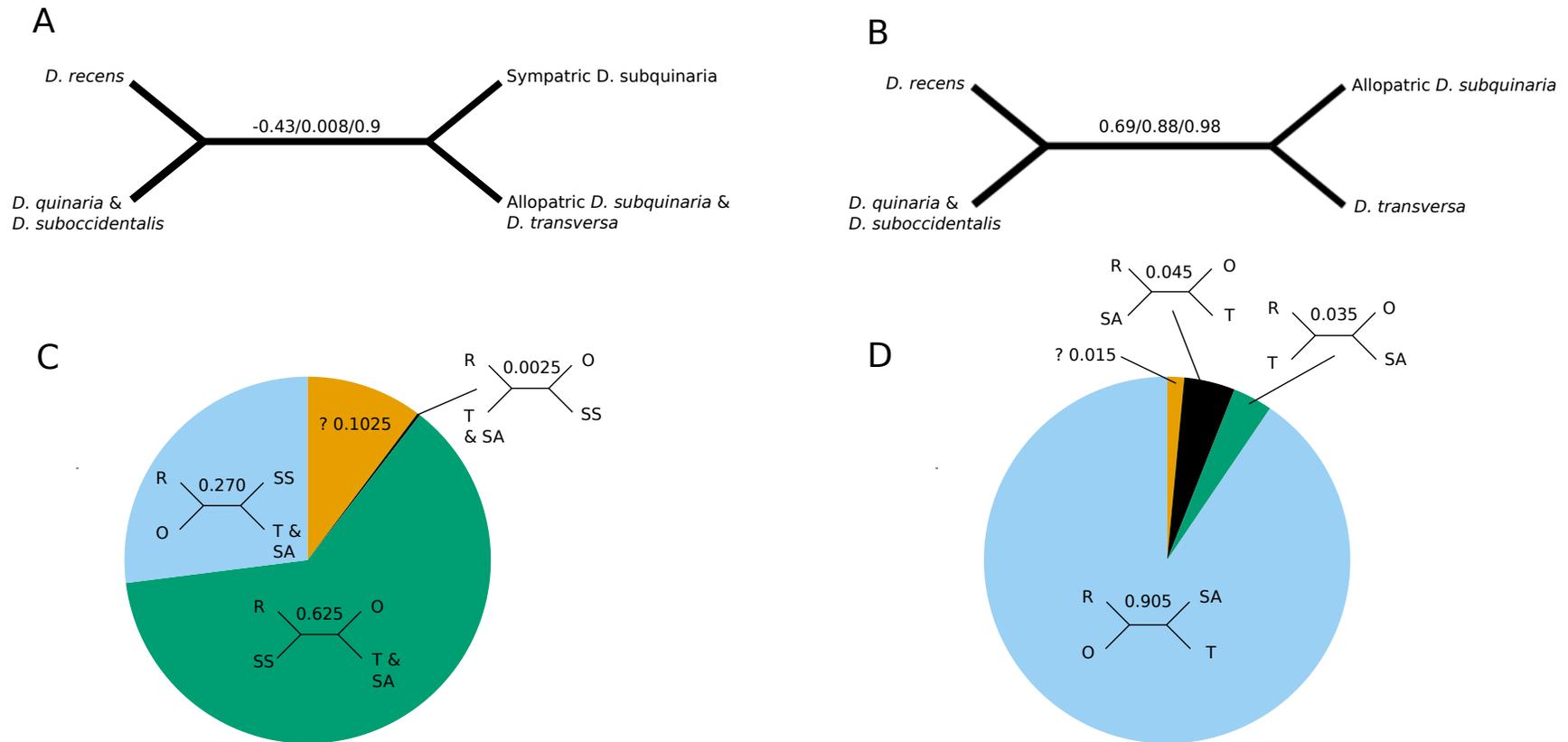


Figure 2.4: Quartet sampling score on focal branch of three trees generated with nuclear data. Scores on the focal branch are in order QC/QD/QI. Pie charts show frequency of each quartet topology, with numbers above focal branch indicating the frequency of the depicted topology. Panel A,C: Analysis including all taxa. Panel B,D: Analysis excluding sympatric *D. subquinaria*. R: *D. recens*; O: *D. quinarian & D. suboccidentalis*; T & SA: *D. transversa* & allopatric *D. subquinaria*. Question mark indicates the frequency of replicates with uncertain quartet topology.

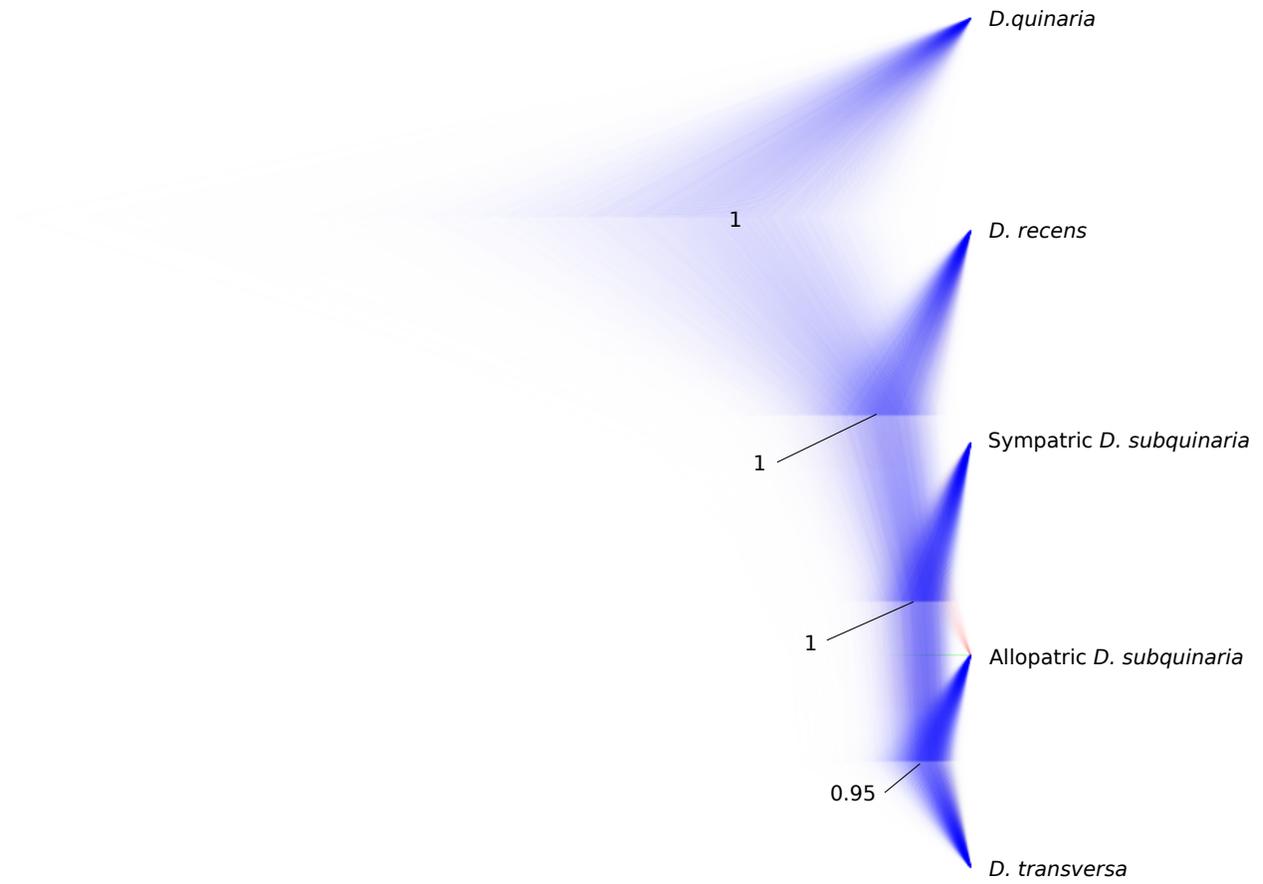


Figure 2.5: Phylogeny generated using the multi-species coalescent model implemented in *BEAST. All nuclear loci were included, and support values are shown for each node.

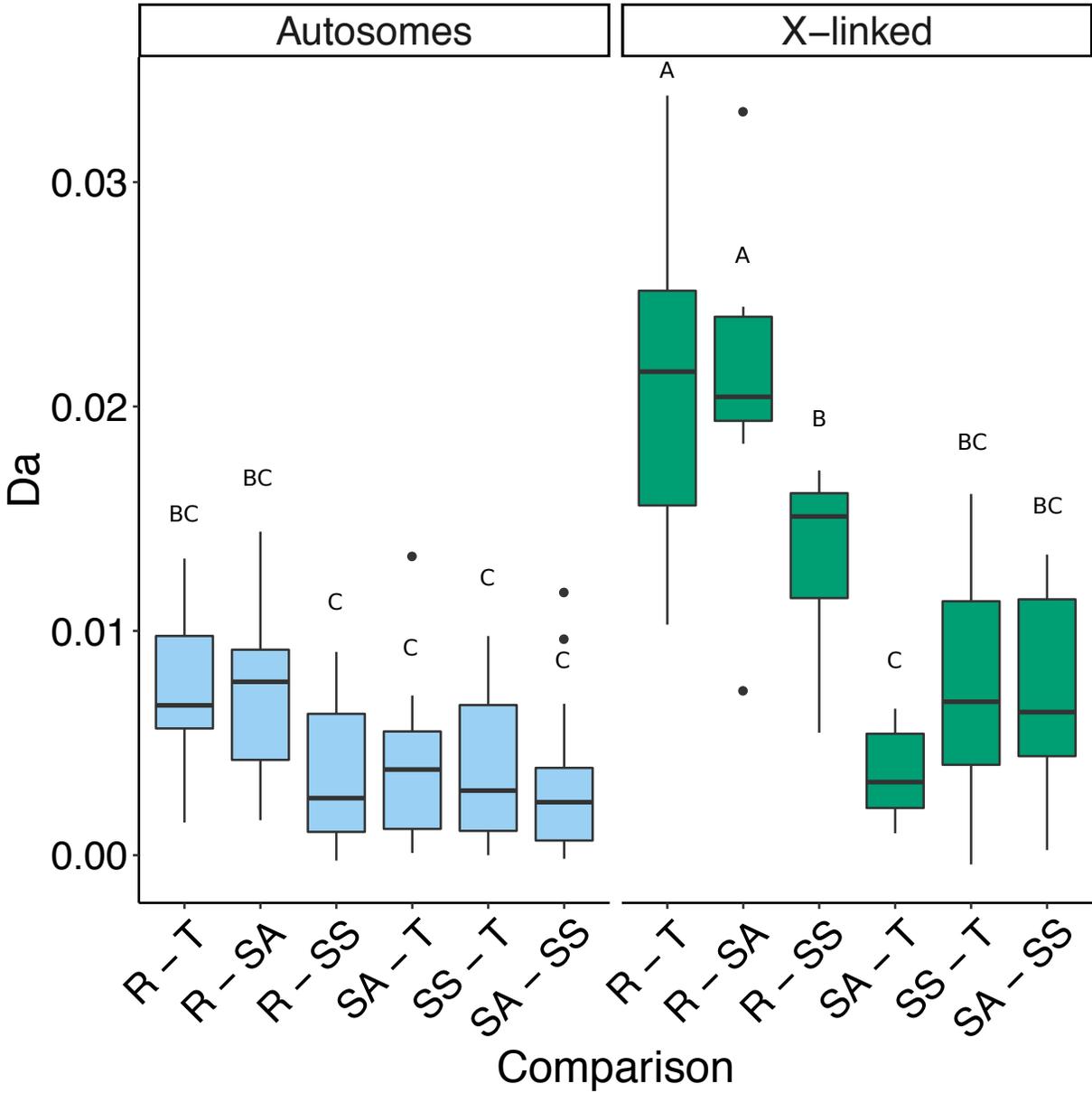


Figure 2.6: Net nucleotide divergence (D_a) for each group comparison. Loci are separated by autosomal (blue) and X-linked (green). Groups in comparisons are abbreviated as *D. recens* (R), *D. transversa* (T), allopatric *D. subquinaria* (SA), and sympatric *D. subquinaria* (SS). Letters above each box indicate post-hoc Tukey groupings, where categories that do not share a letter are significantly different. We find significant variation in D_a between the X-linked and autosomal loci ($F_{1,138} = 98.4$, $P < 0.001$), among group comparisons ($F_{5,138} = 16.1$, $P < 0.001$), and in their interaction ($F_{5,138} = 9.8$, $P < 0.001$).

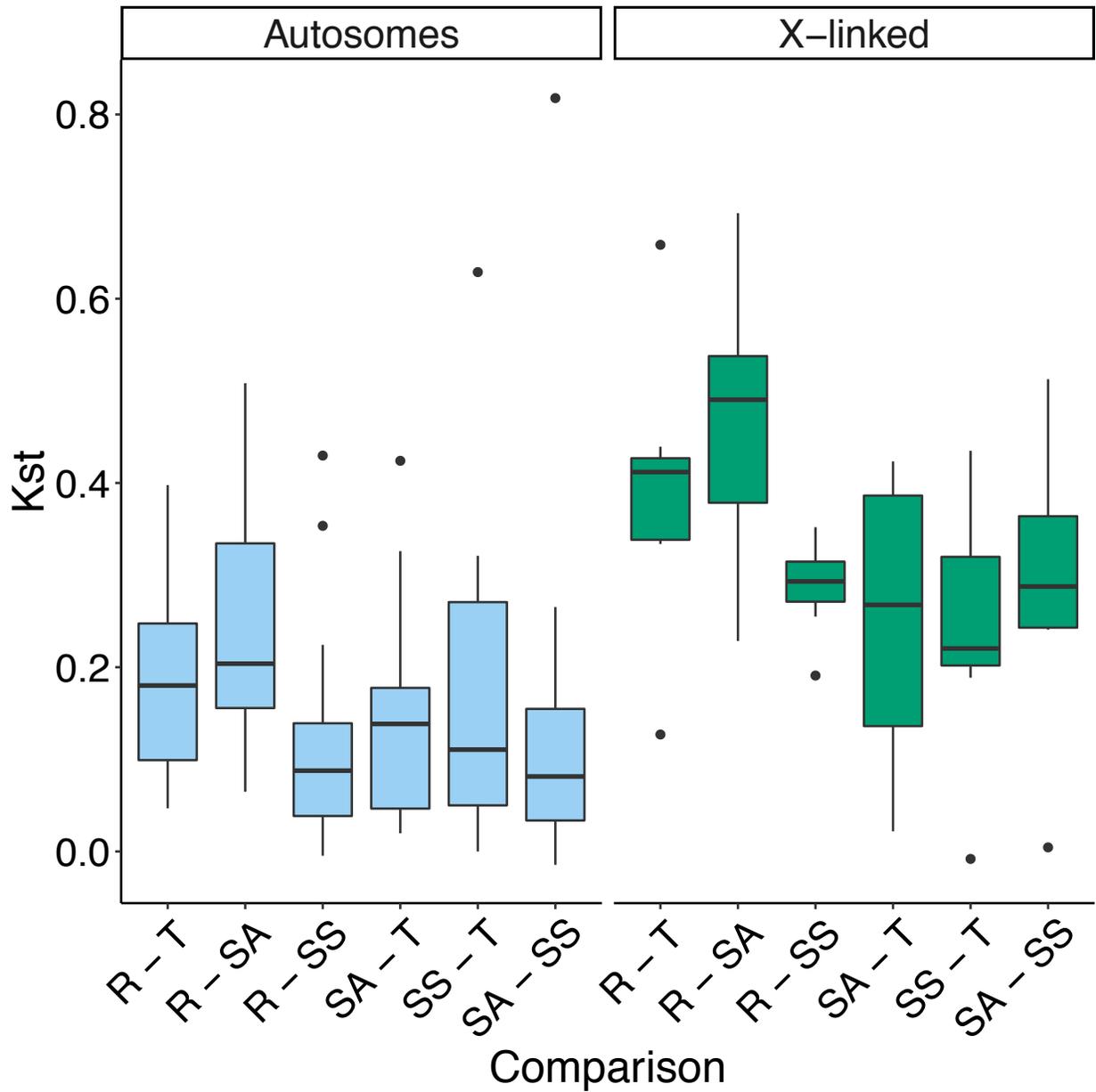


Figure 2.7: Genetic differentiation (K_{ST}) between groups, separated by autosomal (blue) and X-linked (green) loci. Groups in comparisons are abbreviated as *D. recens* (R), *D. transversa* (T), allopatric *D. subquinaria* (SA), and sympatric *D. subquinaria* (SS). No groups were significantly different using a post-hoc Tukey HSD test.

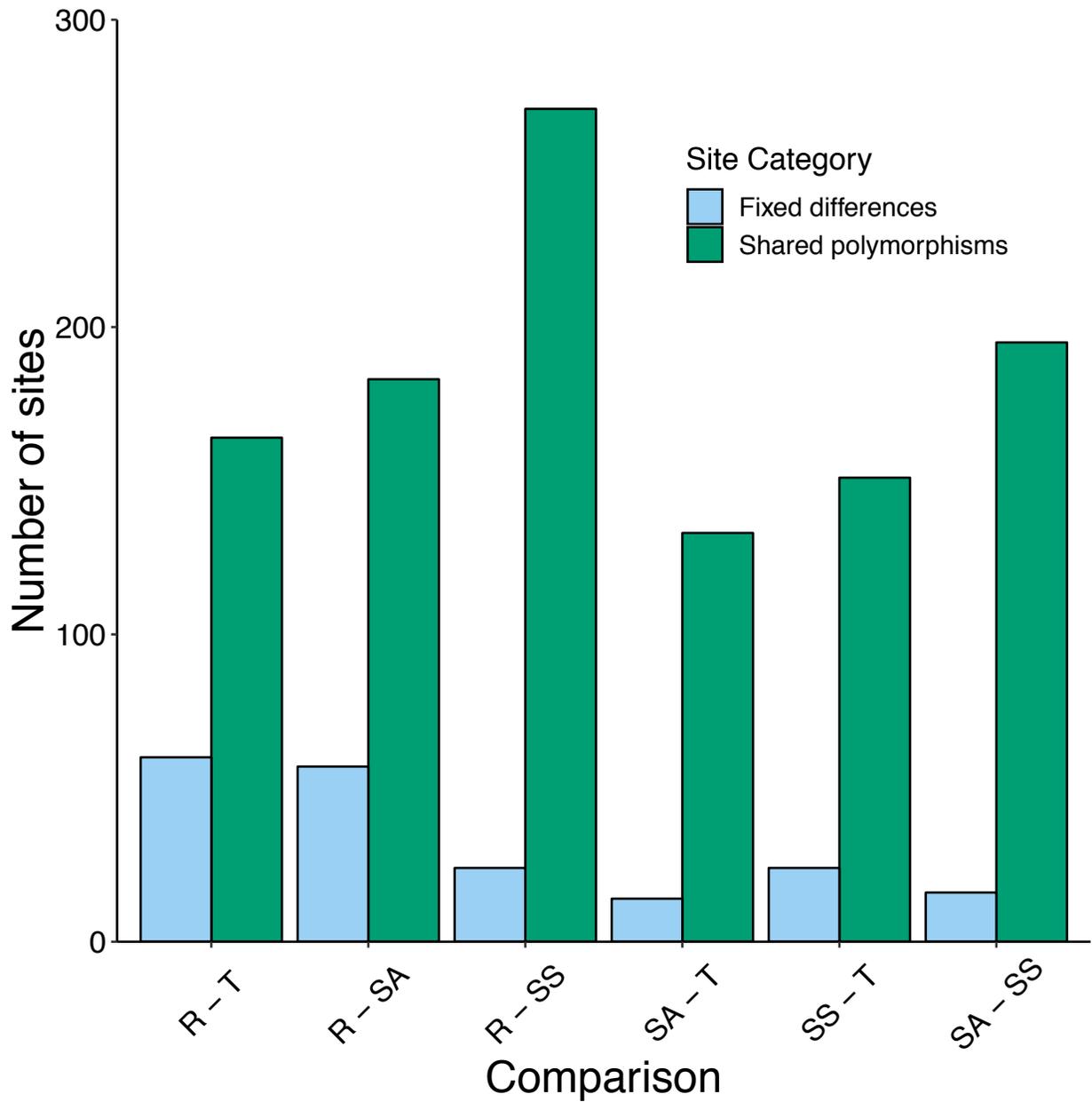


Figure 2.8: The total number of fixed differences (blue) and polymorphic sites (green) summed across autosomal and X-linked loci. Groups in comparisons are abbreviated as *D. recens* (R), *D. transversa* (T), allopatric *D. subquinaria* (SA), and sympatric *D. subquinaria* (SS).

Table 2.1: Sampling locations and number of isofemale lines sampled from each location. Population region refers to range overlap between *D. subquinaria* and *D. recens*; i.e. sympatric where they both occur and allopatric where only one species occurs.

Species / Location	Abbreviation	Population region	N isofemale lines
<i>D. subquinaria</i>			
Portland, OR, USA	PO	Allopatric	4
Seattle, WA, USA	SE	Allopatric	4
Deary, ID, USA	DE	Allopatric	3
Missoula, MT, USA	MS	Allopatric	3
Shuswap, BC, Canada	SH	Allopatric	1
Canmore, AB, Canada	CA	Sympatric	2
Hinton, AB, Canada	HI	Sympatric	4
Kawtikh, AB, Canada	KA	Sympatric	4
<i>D. recens</i>			
Canmore, AB, Canada	CA	Sympatric	2
Hinton, AB, Canada	HI	Sympatric	3
Kawtikh, AB, Canada	KA	Sympatric	4
Nordegg, AB, Canada	NO	Sympatric	1
North Bay, ON, Canada	NB	Allopatric	4
Peru, NY, USA	PE	Allopatric	2
Smoky Mountains, TN, USA	SM	Allopatric	4
<i>D. transversa</i>			
Lahti, Finland	LA		1
Uppsala, Sweden	UP		6
<i>D. quinaria</i>			
Stock center 15130-2011			1
<i>D. suboccidentalis</i>			
Big Sky, Montana, USA			1

Table 2.2: Migration rates and effective population size as estimated by Migrate-n. Inferred estimates include the mutation scaled migration rate ($M = m/\mu$) and the mutation scaled effective population size ($\Theta = xN_e\mu$, where x is the inheritance parameter). Populations are indicated as R = *D. recens*; S = *D. subquinnaria*. Model testing was performed with Bezier-approximated log marginal-likelihoods (Bezier lmL) and log Bayes factors (LBF).

Genomic Region	Model	$M_{R \rightarrow S}$	$M_{S \rightarrow R}$	Θ_R	Θ_S	Bezier lmL	LBF	Model Rank
X-linked	R ↔ S	9	11.3	0.02689	0.03121	-14201.8	-34.54	3
	R → S	32.3	0	0.03116	0.03226	-14189.92	-7.78	2
	R ← S	0	12.7	0.02835	0.03352	-14186.03	0	1
	R = S	15.1	15.1	0.03381	0.03041	-14210.53	-49	4
Autosomes	R ↔ S	16.9	5.6	0.01757	0.00848	-21272.42	0	1
	R → S	48	0	0.0152	0.00797	-21293.38	-41.92	2
	R ← S	0	195.9	0.0918	0.01354	-21296.2	-47.56	3
	R = S	31.6	31.6	0.04574	0.00458	-21318.03	-91.22	4

CHAPTER 3

WOLBACHIA-HOST INTERACTIONS IN A NATIVE AND NON-NATIVE HOST SPECIES OF THE *DROSOPHILA SUBQUINARIA* SPECIES COMPLEX

Introduction

Wolbachia is a maternally inherited, intracellular alpha-proteobacterium that is found in ~20-60% of terrestrial arthropods (Weinert, Araujo-Jnr, Ahmed, & Welch, 2015; Werren & Windsor, 2000). *Wolbachia* is particularly well known for its capacity to manipulate the reproduction of its hosts in ways that promote their spread through a naive host populations. Reproductive manipulations observed in insects include male-killing, thelytokous parthenogenesis, and cytoplasmic incompatibility (Werren, Baldo, & Clark, 2008). Cytoplasmic incompatibility (CI) is the most commonly observed phenotype, and results in embryonic death of offspring from crosses between infected males and uninfected females (or females harboring an incompatible *Wolbachia* infection) (Kaur et al., 2021; Yen & Barr, 1971). In addition to manipulating host reproduction, *Wolbachia* can induce a myriad of other phenotypes in the host, meeting the full range of host-endosymbiont relationships from parasite to obligate mutualisms (Zug & Hammerstein, 2015).

Phylogenetic analyses show incongruence in host and *Wolbachia* phylogenies, indicating that *Wolbachia* can cross interspecific boundaries to invade new host species (Raychoudhury, Baldo, Oliveira, & Werren, 2009; Schuler et al., 2013; Turelli et al., 2018). *Wolbachia* can spread within a species can occur quickly, and has been documented in real time across among a

handful of insect species (Bakovic, Schebeck, Telschow, Stauffer, & Schuler, 2018; Duploux, Hurst, O'Neill, & Charlat, 2010; Kriesner, Hoffmann, Lee, Turelli, & Weeks, 2013). For example, the CI-causing *Wolbachia* strain *wRi* was documented first fixing in western North American populations of *Drosophila simulans*, then invading and fixing in Australian populations of *D. simulans* approximately a decade later (Kriesner et al., 2013; Turelli & Hoffmann, 1995). The conditions under which CI-causing *Wolbachia* infections can invade naive populations has been extensively described (Caspari, 1959; Hoffmann, Turelli, & Harshman, 1990; Hurst, 1996; Prout, 1994; Turelli & Hoffmann, 1995; Turelli, Hoffmann, & McKechnie, 1992). However, the rate at which observed invasions have occurred, and the incidence of CI-causing *Wolbachia* infections among species, suggest that additional *Wolbachia*-induced host phenotypes are required to explain their rapid spread (Turelli, pers. comm.).

In contrast, *Wolbachia*-induced CI can also serve as an interspecific reproductive boundary if only one species harbors a *Wolbachia* infection or the two species harbor incompatible *Wolbachia* infections (Bordenstein, O'Hara, & Werren, 2001; Cruz, Magalhaes, Sucena, & Zele, 2021; Jaenike, Dyer, Cornish, & Minhas, 2006; Shoemaker, Katju, & Jaenike, 1999). This reproductive barrier manifests as increased hybrid offspring mortality in incompatible crosses between species. In this study, we focus on two closely-related and hybridizing sister species in which *Wolbachia* is hypothesized to contribute to reproductive isolation. In natural populations, *D. recens* is nearly always infected with *Wolbachia* (~98% infection frequency) (Shoemaker et al., 1999), and no wild-caught *Wolbachia*-infected *D. subquinaria* have been found despite extensive sampling.

D. recens and *D. subquinaria* occurs in eastern and western North America, respectively, and their ranges overlap in central Canada for approximately 1500 km (Dyer, Bewick, White,

Bray, & Humphreys, 2018; Jaenike et al., 2006; Wheeler, 1960). Where they co-occur, both species can be collected on the same mushroom baits. Crosses between species produce sterile hybrid males (i.e. Haldane's rule) and fertile females. Postzygotic isolation between *D. recens* and *D. subquinaria* is exacerbated by a fixed CI-causing *Wolbachia* infection in *D. recens* (termed *wRec*). In crosses between uninfected *D. subquinaria* females and infected *D. recens* males, CI results in nearly complete offspring mortality (Jaenike et al., 2006). In contrast, all offspring survive in the reciprocal cross between (infected) *D. recens* females and (uninfected) *D. subquinaria* males. This results in an asymmetric level of postzygotic reproductive isolation between *D. recens* and *D. subquinaria* in natural populations, where hybrid matings are more costly for *D. subquinaria* than *D. recens* females.

Asymmetries in premating isolation are also observed between these species. First, female *D. subquinaria* females from the range sympatric with *D. recens* do not mate with *D. recens* males, while *D. subquinaria* females from allopatric regions will mate with *D. recens* males at a moderate rate (Humphreys, Rundle, & Dyer, 2016). *D. recens* females do not show a pattern of reproductive character displacement. Higher mate discrimination by *D. subquinaria* females is consistent with the *Wolbachia* infection in *D. recens* indirectly selecting for increased mate discrimination specifically in sympatric *D. subquinaria*. Furthermore, these *D. subquinaria* sympatric females have diverged in their mate preferences such that they also now show reduced mating with conspecific males from the allopatric part of the range. The resulting mate discrimination observed in sympatric *D. subquinaria* is sufficiently strong that it reduces mating between sympatric and allopatric *D. subquinaria*, potentially driving incipient speciation within *D. subquinaria* (Humphreys, Rundle, & Dyer, 2016). This process exemplifies the potential for the clade-wide consequences of a single *Wolbachia* infection.

Despite the strong reproductive isolation between *D. recens* and *D. subquinaria*, genetic evidence suggests recent and likely ongoing gene flow between them (Jaenike et al. 2006, Bewick and Dyer 2014, Ginsberg et al. 2019, Dyer et al. 2018). Gene flow is predominately in the direction of *D. recens* into *D. subquinaria*, and is observed both in autosomal loci and the maternally inherited mtDNA. For instance, approximately 1-4% of *D. subquinaria* individuals possess a mitochondrial haplotype from *D. recens* (Jaenike et al. 2004, Bewick and Dyer 2018, Ginsberg et. al 2019). This pattern is consistent with the initial hybridization of a *D. recens* females with a *D. subquinaria* male, and then subsequent backcrossing to *D. subquinaria*. None of the *D. subquinaria* individuals with a *D. recens* mtDNA are infected with *Wolbachia*, indicating either the initial *D. recens* female was not infected with *Wolbachia* or the infection was lost after the initial hybridization event(s). Furthermore, *D. subquinaria* with a *D. recens* mtDNA are found throughout the range of *D. subquinaria*, suggesting that this introgression has occurred more than once and these mtDNA haplotypes are able to persist in the *D. subquinaria* host background.

Here we address the question: Why don't we see introgression of *Wolbachia* from *D. recens* into *D. subquinaria*? We expect *D. subquinaria* to be exposed to *Wolbachia* infection at a rate consistent with levels of gene flow. However, thus far there are no instances where we observe *Wolbachia* from *D. recens* infecting *D. subquinaria*. Jaenike (2007) introgressed *Wolbachia* from *D. recens* into allopatric *D. subquinaria* using lab crosses, and found that the reproductive manipulation shifted from CI in *D. recens* to male-killing (MK) in *D. subquinaria*. Furthermore, some *D. subquinaria* lines harbored a suppressor of male-killing, such that no *Wolbachia*-induced phenotype was observed. This shift in reproductive phenotype could prevent *Wolbachia* from invading from *D. recens* into *D. subquinaria*.

In this study, we first introgressed *wRec* from *D. recens* into *D. subquinaria* using genetic backcrosses to create *Wolbachia*-infected lines of *D. subquinaria*. Contrary to the findings of Jaenike (2007), we found that *Wolbachia* causes cytoplasmic incompatibility in *D. subquinaria*, with no evidence of male-killing. To better understand the potential for *Wolbachia* to invade *D. subquinaria*, we next estimated the three parameters that govern CI dynamics in both the native host *D. recens* and the novel host *D. subquinaria*. These parameters include the reduction of offspring hatch rate produced by incompatible vs. compatible crosses (s_h); the relative fecundity of infected vs. uninfected females (F); and maternal transmission rate from infected females to their offspring ($1-\mu$) (Turelli, 1994). In conjunction, we look for evidence of immune system response to *Wolbachia* transfer across a species boundary by quantifying the titer of *Wolbachia* and patterns of differential gene expression between infected and uninfected fly samples of both hosts. We find no evidence for immune system up-regulation in either host, and detect no differences in *Wolbachia* infection titer between species.

Methods

Fly Stocks and maintenance

All fly stocks were kept at 20 °C on a 12-hour light:dark cycle. Unless otherwise noted, fly stocks were maintained on Carolina Biological formula 4-24 instant *Drosophila* medium with a small piece *Agaricus bisporus* mushroom and a cotton dental plug for pupation. Flies used in experiments for *Wolbachia* phenotype were reared in density-controlled vials. Virgin flies were collected with brief CO₂ anesthetization within 24 hours of eclosion and stored at a density of 10-15 flies/vial for 5-7 days before being used in experiments.

We investigated *Wolbachia*-related phenotypes using eight different fly stocks. The *Wolbachia*-infected *D. recens* stock is referred to as AlbW+, which was created by combining eleven *D. recens* isofemale lines collected from Alberta, Canada, from 2009-2011. The *D. subquinaria* fly stocks are SE70, an isofemale *D. subquinaria* stock collected in 2010 in Seattle, Washington, and MT1, an isofemale line collected in 2011 in Missoula, Montana. *Wolbachia*-infected *D. subquinaria* stocks were generated by crossing approximately ten virgin, *Wolbachia*-infected female *D. recens* AlbW+ to ten *Wolbachia*-uninfected males from each *D. subquinaria* stock (SE70 and MT1). Virgin females from the hybrid cross were backcrossed to males from their respective *D. subquinaria* line for ten generations, after which stocks were maintained normally. During these backcrosses, we did not observe any deviation from a 50:50 offspring sex ratio, as Jaenike (2007) did, suggesting either that this strain of *Wolbachia* did not cause male-killing in *D. subquinaria* or there was a suppressor of male-killing in the *D. subquinaria* lines we used. These *Wolbachia*-infected, *D. subquinaria* stocks are referred to as SE70BC, and MT1BC.

To create genetically matched *Wolbachia*-free stocks, *D. recens* AlbW+ and *D. subquinaria* SE70BC and MT1BC were treated on tetracycline-soaked mushrooms to remove *Wolbachia* infection. These stocks are referred to as Albtet, SE70BCtet, and MT1BCtet, respectively. To remove *Wolbachia* infection, flies were placed on ~2 mm slices of mushrooms that were soaked in 1mg/mL tetracycline for 24 hours. Ten inseminated females were allowed to oviposit for six days, and offspring were collected. After one generation on normal food media, all lines were screened for *Wolbachia* using a PCR assay described below.

Estimates of Wolbachia-induced phenotypes

To estimate CI in each species, we set up the crosses listed in Table 3.1 with a single female and 2-3 males on standard food vials. Flies were observed for three hours and mating pairs were kept. After 24 hours, flies were transferred to randomized experimental vials. Experimental vials consisted of a piece of mushroom sliced flat and placed over a water-soaked piece of cotton. Flies were transferred to a new experimental vials every 24 hours for three days. After the third day, adults were removed and frozen for later *Wolbachia* infection confirmation. Each experimental vial was scored for number hatched and unhatched eggs 48 hours after the adults were removed. This procedure was repeated for three experimental blocks. Each block included two genotypes (Block1: Alb & SE70, block 2: Alb & MT1, and block 3: SE70 & MT1).

To estimate the effect of *Wolbachia* on female fecundity within each species, we set up the crosses listed in Table 3.2. with one female and 2-3 males. For the *D. recens* genotype, crosses were between uninfected or infected females and uninfected males. In *D. subquinaria*, we included crosses between infected females and uninfected males, and crosses with uninfected females from all uninfected stocks (MT1, MT1BCtet, SE70, SE70BCtet) and uninfected males to account for *Wolbachia* infection and possible mito-nuclear interactions between *D. recens* and *D. subquinaria*. Crosses were observed for 3 hours and those where mating was observed were recorded, and only mated flies were kept for later analysis. Twenty-four hours later, flies were transferred to randomized experimental vials on standard food. Flies were transferred to a new vial every five days for 20 days. Vials were checked daily for fly death; crosses where females died during the experiment were excluded from analysis, and dead males were replaced. No substantial differences were found among fly death rates for infected and uninfected females. After the 20 days, the adults were removed and frozen for later *Wolbachia* infection

confirmation. All offspring were counted and frozen. This procedure was repeated for three experimental blocks, each of which consisted of 20-30 replicates for each cross listed in Table 3.2. During the first and second experimental block, offspring sex was also recorded to estimate offspring sex ratio of infected and uninfected females.

Finally, we estimate the rate of maternal transmission within each host species. Because infected *D. subquinaria* have not been observed in nature, maternal transmission had to be estimated from lab-generated *Wolbachia* infected fly stocks. Experimental flies were collected from density-controlled vials. Single virgin female flies from either MT1BC or SE70BC were placed in standard food vials for five to seven days with two uninfected males from the same line, and then frozen. All offspring were collected and frozen. *D. recens* maternal transmission rate was determined from both lab reared flies and offspring of wild-caught females collected from Fern Lake, New York, in 2014. Wild-caught females were placed in a vial of standard fly media and allowed to oviposit. Offspring were collected and frozen upon emergence. *Wolbachia* infection status was determined for each fly with PCR using WSP primers and mtDNA COI primers (as a positive control).

Wolbachia Equilibrium Frequency Prediction

We estimated the predicted unstable and stable equilibrium frequencies for CI-causing microbes based on the model described in (Prout, 1994) and generalized in (Turelli, 1994). This is a discrete generation model that assumes random mating with respect to *Wolbachia* infection state and no paternal transmission of *Wolbachia*. Equilibrium *Wolbachia* frequencies are predicted with the following equation:

$$\hat{p} = \frac{s_h + 1 - F \pm \sqrt{(s_h + 1 - F)^2 - 4s_h[F(1 - \mu) - 1](1 - F\mu)}}{2s_h(1 - F\mu)}$$

(Turelli, 1994)

Where F is the relative fecundity of infected vs. uninfected females, μ is the proportion of uninfected ova produced by infected females (or maternal transmission rate $1 - \mu$) and s_h is the strength of CI ($1 - H$, where H equals the relative hatch rate for incompatible vs. compatible crosses). Stable equilibrium infection frequencies are the values obtained when taking the positive root in the numerator, whereas the unstable equilibrium infection is estimated by taking the negative root. Here, we produce a point estimate for equilibrium frequencies using experimental point estimate for the described parameters and upper/lower bounds from unstable and stable equilibrium frequencies using 95% confidence interval bounds of our three parameters.

Molecular methods: DNA Preparation and PCR

DNA extractions were performed using two techniques: Squish buffer DNA extraction protocol (modified from (Gloor, Nassif, Johnson-Schlitz, Preston, & Engels, 1991) or Genra Puregene Tissue Kit (QIAGEN, LLC). Extractions using the Genra Puregene tissue Kit were performed using 1-10 whole flies following the manufacturer's protocol, then 1:50 dilutions were used for subsequent analysis. To screen for *Wolbachia* infection, we used *Wolbachia*-specific PCR of a 658 bp region of *Wolbachia surface protein* (WSP) and 1432 bp region of the host mitochondrial locus *cytochrome oxidase subunit 1* (COI) as a positive control. For WSP amplification we used *Wolbachia*-specific typing primers 81F (TGGTCCAATAAGTGATGAAGAAAC) and 691R (AAAAATTAACGCTACTCCA)

(Braig, Zhou, Dobson, & O'Neill, 1998) and for COI amplification we used the forward primer TY-J-1460 (TACAATTTATCGCCTAAACTTCAGCC) and reverse primer TL-2N (TCCAATGCACTAATCTGCCATATTA) (Simon C, 1994). For species confirmation, we amplified three loci: A 631 bp region of the autosomal *plexA* locus with species-specific primers (Forward: CMCAATTTACCAGYATACAAGC, Reverse: TCCTGCATATTGCGAAGA), a 1,389 bp region of the Y-chromosome locus K13 (Forward primer: GGWAGCGTTGARCTTTGG, Reverse: CRTGSCGCACCAGTGATG) and a 1,432 bp region of COI as described above. Amplification of *plexA* was done using custom primers designed using Primer3 (Koressaar & Remm, 2007; Untergasser et al., 2012) and a *plexA* consensus sequence from population sampling of *D. subquinaria* (genbank accession: MN274119.1 - MN274095.1). As *plexA* is highly divergent (Ginsberg, Humphreys, & Dyer, 2019), these primers only amplify with *D. subquinaria* DNA samples.

Wolbachia Titer Assay

To test for a difference in *Wolbachia* density within and between species, *Wolbachia* titer was measured with relative quantification of the *Wolbachia* locus, *wsp*, and the host autosomal locus *mago nashi* (*mago*). I collected virgin females from all *Wolbachia*-infected fly stocks and aged for five days (Genotypes: AlbW+, SE70BC, MT1BC). All primers were designed using Primer3Plus (Koressaar & Remm, 2007; Untergasser et al., 2012). Host qPCR primers were designed using a *D. recens* and *D. subquinaria* *mago* consensus sequence (Genbank accessions: MN268207.1 - MN268170.1; Forward primer: GTCATGGAGGAGCTGAAGC; Reverse primer: ACACGATCTGGTGGCGGC). *Wolbachia* primers were designed using a published *wRec wsp* sequence (Genbank accession: AY154399.1; Forward primer:

GCTGGAGCTCGTTATTTTCGG; Reverse primer: GCATCAGCAACCTGTCCGAT). PCR reactions with both primer sets were first optimized with an annealing temperature of 60° C and samples were visually examined for a single amplicon of the expected size. Primer efficiencies were calculated using amplifications of ten-fold dilution series with both *D. recens* and *D. subquinaria* samples. Primers for both *mago* and *wsp* produced identical amplicon size (83 bp) with comparable efficiencies in *D. subquinaria* and *D. recens* (Table S3.1). All reactions were performed in triplicate using single-fly DNA squish preps (see DNA preparation) with the PowerUp SYBR Green Master Mix (Cat. number: A25741, Thermofisher Scientific, MA., USA) following the manufacturer's protocol. Reactions were performed on a BioRad CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, California, USA) using a two-pass reaction protocol with the following parameters: 95°C for 3 min., then 95°C 10 sec., 60°C 30 sec. for 40 cycles. Reactions were only kept for analysis if all three replicates were within 0.5 cycles (In several instances, individual outliers were removed). Relative *Wolbachia* titer was calculated as relative copy number ($2^{-\Delta Cq}$), where Cq is the number cycles at which the amplification curve crossed the threshold fluorescence and ΔCq is *wsp* Cq – *mago* Cq using the mean Cq value of technical replicates.

Differential Expression Analyses

To detect immune system response to *Wolbachia* infection in *D. recens* and *D. subquinaria*, we measured differential expression of *Wolbachia*-infected and uninfected females. We did this using the Alberta *D. recens* (AlbW+ & AlbTet) and SE70 (SE70BC & SE70BCtet) *D. subquinaria* stocks. To increase the ability to detect differences in host response across different tissues, we performed this experiment using dissected ovaries, and the remaining

carcass tissue. *Whole* RNA was extracted using TRIzol Reagent (Life Technologies Corp, Carlsbad, CA, USA) and following the manufacturer's protocol. For each sample, ten age-controlled (five days old) and density-controlled (ten flies per vial) female flies were briefly anesthetized on CO₂ and ovaries were dissected in ice-cold phosphate buffer saline (PBS). After dissection, tissue samples were flash frozen in liquid nitrogen and stored at -80 °C until RNA isolation. RNA QC was analyzed with Nanodrop and Qubit to ensure sufficient quality and quantity before library preparation and sequencing. Library preparation and sequencing was done independently for *D. recens* and *D. subquinaria*. For *D. recens*, library preparation and sequencing was done by the Georgia Genomics and Bioinformatics Core (GGBC). In brief, isolated RNA was quality checked with an Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA). Libraries were prepared with KAPA biosystems stranded RNA kit and sequenced on a single Nextseq 75 Paired-end Mid output flow cell. *D. subquinaria* samples were prepared by the Schmitz lab at UGA from 1.3 µg input RNA with the Illumina TruSeq mRNA Stranded Library Kit v2 (Illumina Inc.) according to the manufacturer's instructions. All volumes were reduced to one-third of the recommended quantity. Samples were sequenced on an Illumina NovaSeq 6000 150 Paired-end flow cell (Illumina Inc.) at the GGBC.

RNAseq reads for each sample were mapped to the same stock's assembled genome (See Chapter 4) using HISAT2 (Kim, Paggi, Park, Bennett, & Salzberg, 2019). Gene expression was then quantified with the htseq-count command of HTSeq 2.0 (Putri, Anders, Pyl, Pimanda, & Zanini, 2022). To identify differentially expressed genes, samples were analyzed independently for each species, using the glmTreat (Chen, Lun, & Smyth, 2016) function in the edgeR package (Robinson, McCarthy, & Smyth, 2010) with $\log\text{-fold-change}=\log_2(1.2)$. All differentially expressed transcripts were annotated with InterProScan v5.55 (Jones et al., 2014) and BLAST+

v2.10.1 (Camacho et al., 2009; Sayers et al., 2022) using the nonredundant (nr) database and the *D. melanogaster* genome r6.46 from Flybase.org (Gramates et al., 2022).

GO term enrichment analysis was performed using the R package topGO (Adrian & Rahnenfuhrer, 2022). First, differentially expressed genes were BLAST searched against the *D. melanogaster* genome release 6.46 from Flybase.org (Gramates et al., 2022). Flybase gene IDs from BLAST hits were then used to assign GO terms from the gene_association.fb GO term database from Flybase.org (Gramates et al., 2022). Go term enrichment was estimated using topGO's "weight01" algorithm and Fisher's exact test statistic.

Results

Estimates of Wolbachia phenotypes

Consistent with CI, wRec infection in *D. recens* displayed higher offspring hatch rate in the compatible cross relative to the incompatible crosses (Figure 3.1, Wilcoxon rank-sum test; $P < 0.0001$). The hatch rate reduction of $s_h = 0.816$ (95% CI: 0.738, 0.875) was comparable, yet slightly weaker, than previous estimates ($s_h \approx 0.948$) (Werren & Jaenike, 1995). When in *D. subquinaria*, wRec causes very strong CI (Figure 3.1, Wilcoxon rank-sum test; $P < 0.0001$). There were no detectable differences in CI severity between the two *D. subquinaria* genotypes (Fig 3.1), thus we combined the data. The hatch rate reduction observed in *D. subquinaria* was $s_h = 0.987$ (95% CI: 0.975, 0.993).

Wolbachia infection was not associated with a deviation in offspring sex ratio from 50:50 in either *D. recens* (Figure 3.2; Infected female sex ratio = 0.510, uninfected female sex ratio = 0.528, Fisher's Exact test, $F_{1,38} = 0$, $P = 0.986$) or *D. subquinaria* (Figure 3.2; Infected female sex ratio = 0.498, uninfected female sex ratio = 0.496, Fisher's Exact test, $F_{1,129} = 1.442$, $P = 0.232$),

confirming that no male-killing is present. In *D. recens* there was no significant difference in fecundity between infected and uninfected females (Figure 3.3; Welch's t-test, $t = -1.10$, $P = 0.27$). Similarly, there was no observed effect of *Wolbachia* infection on offspring production in *D. subquinaria* (Figure 3.3, Welch's t-test, $t = -0.69$, $P = 0.49$). However, in *D. subquinaria* there was a significant effect of genotype (Figure 3.3B, ANOVA, $F_{4,218} = 11.106$, $P < 0.0001$). Tukey groupings revealed a significantly lower offspring production in the MT1 genotype (Figure 3.3B). However, in both species average fecundity of infected females was slightly higher than in uninfected females.

Both *D. recens* and *D. subquinaria* had nearly perfect maternal transmission of *Wolbachia* (Table 3.3). In *D. recens*, the only uninfected individual of 408 surveyed was detected in the offspring of a wild-caught female ($\mu = 0.0025$). All lab reared females displayed perfect maternal transmission. In *D. subquinaria*, three uninfected offspring were detected of 779 total offspring ($\mu = 0.0039$).

Equilibrium Frequency Prediction

Equilibrium frequency estimates with parameter estimates and upper and lower bounds are listed in Table 3.4. Upper and lower bounds were estimated using the 95% confidence intervals for the three parameters. Equilibrium point estimates for *D. recens* and *D. subquinaria* were remarkably similar, with unstable equilibrium point estimate equal to zero and the stable equilibrium frequency equal to 0.99 for both species. This predicts that *Wolbachia* will spread from very low infection frequencies and remain nearly fixed in both species. However, the unstable equilibrium frequency upper bound in *D. subquinaria* is 0.10, or a ten percent infection frequency. The low unstable equilibrium frequencies can be explained by the combination of

high relative fecundity and maternal transmission (low μ). At very low infection frequencies, cytoplasmic incompatibility (CI) does not significantly affect *Wolbachia* frequency dynamics because there are few infected males in the population to induce CI. At low frequencies, the ability of *Wolbachia* to invade a population is given by the joint parameter $F(1 - \mu)$. If $F(1 - \mu) > 1$, we expect *Wolbachia* invade the population. This is the case for both *D. recens* ($F(1 - \mu) = 1.06$) and *D. subquinaria* ($F(1 - \mu) = 1.04$).

Wolbachia Titer

The observed difference in CI strength between *D. recens* and *D. subquinaria* is likely not due to a difference in *Wolbachia* titer. Observation of technical replicates displayed high reproducibility (Figure 3.4A). Relative copy number for *Wolbachia*/host genome was not significantly different between *D. recens* and *D. subquinaria* (Figure 3.4B; Welch's t-test, $t = -0.294$, $P=0.775$). However, sample size for both *D. recens* and *D. subquinaria* was low (*D. recens*: $n=6$ flies, *D. subquinaria*: $n=8$ flies), and so power was lacking to detect small changes in *Wolbachia* titer between species. This is particularly true here due to the high levels of variation in *Wolbachia* titer between individuals in each species. In *D. recens*, relative copy number ranged from 1.67 - 15.10. In *D. subquinaria*, relative copy number ranged from 1.22 – 16.26.

Differential Expression

Immune system upregulation is commonly found when naturally uninfected hosts are exposed to *Wolbachia* infection. To determine whether this occurs in either host species and to look for additional biological processes effected by *Wolbachia* infection, we examined

differential expression in infected and uninfected *D. recens* and *D. subquinaria*. Full data for sequencing results are in Table S3.2. PCA analysis of all samples depicted grouping by treatment (Figures 3.5A & 3.6A). For both *D. recens* and *D. subquinaria*, the majority of variation is explained by differences in carcass vs. ovary tissue (Figures 3.5A and 3.6A, PC1, X-axis). PC1 accounts for 96.1% of the variance in *D. recens*, and 96.4% of the variance in *D. subquinaria*. PC2 distinguishes between *Wolbachia*-infected and uninfected samples (Figures 3.5A and 3.6A, Y-axis). In *D. recens*, PC2 accounts for 1.21% of the variance, and in *D. subquinaria*, PC2 accounts for 0.78%. Notably, uninfected *D. subquinaria* carcass tissue displayed less defined grouping than the remaining samples (Figures 3.6A, red circles).

Overall, we found far more differentially expressed genes in *D. recens* than in *D. subquinaria* (Table 3.5, Figure 3.5B). In *D. recens* carcass tissue we found 520 and 210 upregulated and downregulated genes, respectively. No other samples revealed greater than 30 differentially expressed genes (Table 3.5).

D. subquinaria differentially expressed genes with BLAST annotations are listed in Table 3.6. We identified two down regulated genes in ovary tissue, one which had a BLAST hit, but was an uncharacterized protein. In carcass tissue, *D. subquinaria* did not upregulate any of the major immune response genes commonly associated with *Wolbachia* infection.

DE genes with blast annotations were compared among treatments to identify shared DE genes between species and tissues. Few genes were upregulated in both *D. recens* and *D. subquinaria* (Figure 3.7A). In carcass tissue, three upregulated genes were shared between *D. recens* and *D. subquinaria*. Only one gene had an informative annotation as a “probable U2 small nuclear ribonucleoprotein A” (BLAST accession: XP_034478387). Between *D. recens* carcass tissue and *D. subquinaria* ovary tissue, two genes were upregulated, with only one

annotated as an *exd1* homolog (BLAST accession: XP_034483523). Between *D. subquinaria* carcass and ovary tissue, three genes were upregulated. A “succinate--hydroxymethylglutarate CoA-transferase” (BLAST accession: XP_034485706), “glutamyl-tRNA(Gln) amidotransferase subunit B, mitochondrial” (BLAST accession: XP_034488326), and “respirasome Complex Assembly Factor 1” (BLAST accession: XP_034478388). The most shared upregulated genes were between *D. recens* carcass and ovary tissue, with 18 shared genes (Figure 3.7A). No down regulated genes were shared between species (Figure 3.7B). This is not surprising due to the small number of differentially expressed genes discovered in *D. subquinaria*, especially in ovary tissue (Table 3.5). The only down regulated gene shared between treatments was a “general odorant-binding protein 56d-like” which was down regulated in both ovary and carcass tissue of *D. recens*. The only indication of immune-related function was several key immune response genes identified in down regulated carcass tissue of *D. recens* (*attacin-A*, *attacin-C*, *diptericiin-A*, Table S3.3).

To identify biological functions affected in *Wolbachia*-infected *D. recens* and *D. subquinaria*, we performed a GO term analysis of differentially expressed genes using topGO (Adrian & Rahnenfuhrer, 2022). Zero enriched GO terms were identified among differentially expressed genes in the *D. subquinaria*. Most GO terms in *D. recens* were associated with transcriptional/translation regulation and/or metabolic processes (Table 3.8). In down regulated carcass tissue, there was one significant GO term associated with larval locomotory behavior.

Discussion

How a *Wolbachia* infection spreads to a new host is a key question to understand the distribution and consequences of these endosymbionts for biodiversity. In this study we

investigate the phenotypic consequences of a *Wolbachia* infection in both a native host, *D. recens*, and closely related non-native host, *D. subquinaria*, to understand *Wolbachia*'s potential to invade *D. subquinaria*. Given the high infection prevalence in *D. recens* and the ongoing hybridization and subsequent introgression from *D. recens* into *D. subquinaria*, we are particularly interested in why *Wolbachia* has not spread into *D. subquinaria*. We introgressed $wRec$ from *D. recens* into *D. subquinaria* using genetic backcrosses, and then estimated the three parameters that govern CI dynamics, including the reduction of offspring hatch rate produced by incompatible vs. compatible crosses (s_h); the relative fecundity of infected vs. uninfected females (F); and the proportion of uninfected ova produced by infected females (μ , $(1-\mu)$ being the maternal transmission rate from infected females to their offspring)(Turelli, 1994). CI theory predicts two stable and one unstable equilibria. The two stable equilibria are at zero and a high frequency analogous to mutation-selection balance, where rare uninfected individuals produced by imperfect maternal transmission are quickly selected out of the population by CI. The unstable equilibrium frequency is the inflection point; an initial infection below this frequency will be selected out of the population, whereas an infection prevalence above this frequency will be quickly driven to the high stable equilibrium frequency through the effect of CI. Across a population's range, this leads to "waves of infection" (Barton & Turelli, 2011), where local infection frequencies stochastically surpass the unstable equilibrium, drives infection frequency to near-fixation, then stochastically surpasses the unstable equilibrium in adjacent populations through dispersal. The end result is a high equilibrium frequency throughout the range of the host.

Our phenotype estimates suggest that interspecific *Wolbachia* invasion is likely given sufficient time and minimal interspecific gene flow. In both *D. recens* and *D. subquinaria*, we

found higher point estimates of average fecundity in infected females (although this did not reach statistical significance), high rates of maternal transmission of *Wolbachia*, and very strong CI. Together, these conditions suggest that *Wolbachia* should spread quickly to high prevalence. Despite this, *Wolbachia* has thus far failed to invade *D. subquinaria*.

Previous investigations in these species suggests that interspecific gene flow does occur, and in the direction that would transfer *Wolbachia* from *D. recens* into *D. subquinaria* (Dyer et al., 2018; Ginsberg et al., 2019). The presence of *D. recens* mtDNA in *D. subquinaria* is further evidence of hybridizations between *D. recens* females with *D. subquinaria* and subsequent backcrossing with *D. subquinaria*. Then why does *D. subquinaria* remain uninfected? Below we propose five hypotheses, though we note this list is not exhaustive.

First, it is possible that only hybridizations between uninfected *D. recens* females and *D. subquinaria* males persist and introgress into *D. subquinaria*. While I have examined effects of infected *D. subquinaria*, I did not examine *Wolbachia* effects in early generation hybrids. A combination of incompatible alleles and *Wolbachia* infection may introduce unique hybrid defects that this study did not examine.

Second, all of the phenotypes we measured were done so in a laboratory setting under optimal growing conditions. Fecundity, CI strength, and maternal transmission may differ within and outside of a lab setting. Thus, we did not capture environmental effects, nor did we capture interactions with other organisms. Environmental effects are particularly important for the estimation of maternal transmission rates (Hague, Mavengere, Matute, & Cooper, 2020; Turelli & Hoffmann, 1995). In both *D. recens* and *D. subquinaria*, we estimated maternal transmission to be nearly perfect (Table 3.3). Previous studies have shown lab environments can drive maternal transmission rates to one within a single generation as compared to maternal

transmission rates of *Wolbachia* from wild-caught females (Turelli & Hoffmann, 1995). Notably, environmental effects on maternal transmission rates offer a simple explanation of global infection frequencies in wMel-infected *D. melanogaster* (Hague et al., 2022). In *D. recens*, we estimated maternal transmission rates in the both wild-caught females and lab maintained fly stocks, where we found no difference in transmission rates. In *D. subquinaria*, we are limited to estimating maternal transmission rates to laboratory stocks.

Third, *Wolbachia* may induce secondary phenotypes that we did not assay. *Wolbachia* possesses the capacity to induce a wide range of phenotypes in their host beyond the well-known reproductive manipulations, for instance protection against viral pathogens (Chrostek, Hurst, & McGraw, 2020; Hoffmann et al., 2011; Ndi, Allingham, Hickson, & Glass, 2016; Pinto et al., 2021; Ye et al., 2015) and nutrient provisioning (Balvin, Roth, Talbot, & Reinhardt, 2018; Brownlie et al., 2009; Driscoll et al., 2020; Zug & Hammerstein, 2015). These phenotypes may affect *Wolbachia* invasion and dynamics either positively or negatively, and there may be additional host-*Wolbachia* genetic interactions that affect these phenotypes. For instance, the capacity for *Wolbachia* to reduce viral loads in the host is largely dependent on the *Wolbachia* strain and host genotype, where different *Wolbachia* infections have shown opposite effects on viral loads in the same host species (Chrostek et al., 2013; Martinez et al., 2014). Nutrient provisioning is generally associated with strong mutualist relationships, where *Wolbachia* infections have become obligate or nearly so (e.g. bedbugs and cat fleas; (Balvin et al., 2018; Driscoll et al., 2020; Hosokawa, Koga, Kikuchi, Meng, & Fukatsu, 2010). However, metabolic provisioning in iron-poor environments has also been found to increase fitness in facultative infections, such as in wMel-infected *D. melanogaster* (Brownlie et al., 2009). The effects of both viral pathogen protection and nutrient provisioning are temperature-sensitive and so likely

relevant only under certain environmental conditions (Chrostek, Martins, Marialva, & Teixeira, 2021), making generalizations difficult.

Fourth, we only sampled two lines of *D. subquinaria* and there may be host-*Wolbachia* interactions we did not capture. There is substantial genetic differentiation across populations of *D. subquinaria*, especially between regions that are sympatric and allopatric with *D. recens*, with divergence estimates on the order of that observed between interspecific taxa (Ginsberg et al., 2019). Hybridization can only occur in regions of sympatry, and it is possible that *Wolbachia*-host dynamics are different in this part of the range than in the allopatric populations we used here. If *Wolbachia* faces a strong barrier restricted to the sympatric populations of *D. subquinaria* this could prevent it from spreading throughout the species range.

Finally, it is possible that this particular *Wolbachia* strain has not had sufficient time to invade *D. subquinaria*. The *Wolbachia* infection in *D. recens* is likely more recent than the last time *D. recens* shared common ancestry with *D. subquinaria* ~0.6 million years ago (Shoemaker et al., 1999). Evidence for this comes from the genetic linkage of the mtDNA and *Wolbachia*, as every time a new *Wolbachia* strain sweeps through a host population the associated mtDNA haplotype is expected to sweep with it. Indeed, the mtDNA of *D. recens* shows low diversity and a high rate of nonsynonymous substitutions (Shoemaker et al. 2004), concurrent with a *Wolbachia* sweep in this species. This suggests that *Wolbachia* infection in *D. recens* is not recent on an ecological time scale. In contrast, the mtDNA of *D. subquinaria* shows high levels of nucleotide diversity, concurrent with no recent *Wolbachia* infection in this species.

The co-evolutionary interaction between *Wolbachia* and the host is highly dynamic. There is a general expectation that *Wolbachia* and hosts will evolve towards a more mutualistic relationship (Zug & Hammerstein, 2015), which has been observed to occur in several systems

over a timespan of decades (Carrington, Lipkowitz, Hoffmann, & Turelli, 2011). Less parasitic *Wolbachia* variants are expected to replace older variants within species, for instance a strain with a weaker level of CI is predicted to replace a strain with a stronger level of CI within a host (Hurst, 1996; Turelli, 1994). Likewise, shifts in *Wolbachia*-induced phenotypes can occur, and be due to host-evolved suppression, *Wolbachia*-evolved reduction in parasitism, or a combination of both.

Consistent with this, our recent unpublished data indicate that *D. recens* likely harbors two very closely related strains of *Wolbachia* (See Chapter 4), and may be in the middle of a strain replacement. In a separate experiment, *Wolbachia* from a different stock of *D. recens* was introgressed into the same stocks of *D. subquinaria* as presented here, and within seven generations, this other *Wolbachia* strain causes male-killing in the same host strains where we described strong CI. This indicates that there are (at least) two distinct strains of *Wolbachia* segregating within *D. recens*. The genome sequences of these male-killing and CI strains are remarkably similar, and suggest that the CI-causing *Wolbachia* is recently derived from the male-killing strain (See Chapter 4). This makes our findings consistent with the earlier findings of Jaenike (2007), who found wRec caused male-killing when introgressed into *D. subquinaria*. This finding raises several interesting questions that are beyond the scope of the current study. Most relevant for our findings, it raises the possibility that the *Wolbachia* strain we used in this study is very recently evolved within *D. recens*. In this case, it may not have had the opportunity to invade into *D. subquinaria*. If this is the case, based on our experiments in this study and previous estimates of introgression between species we predict it will begin to spread through *D. subquinaria* over the next several decades.

CI phenotypes in the native versus novel host

In our study we observe that wRec causes stronger CI in the novel host *D. subquinaria* than in the native host *D. recens*. We used two methods to investigate the functional differences of *Wolbachia* infection that may result in this phenotypic difference. First we assayed the titer, or density of *Wolbachia* in each host, which in other species has been shown to be related to strength of the *Wolbachia*-induced phenotypes (Chrostek et al., 2013; Lopez-Madrigal & Duarte, 2019; Min & Benzer, 1997). While our assay was coarse, we found no significant difference in *Wolbachia* titer between *D. recens* and *D. subquinaria*, though we did find substantial variation across flies within each species. A finer-scale study is necessary to characterize variation in *Wolbachia* titer between species, strains within species, fly sex, and fly age.

Next, we assayed differential expression of host genes across species and tissues. Surprisingly, very little was differentially expressed between infected and uninfected tissues in *D. subquinaria* compared to *D. recens*. And there was little signal for biological processes identified with GO term enrichment analysis. The lack of signal for biological processes affected by *Wolbachia* infection was equally surprising as it is clear that *Wolbachia* infection induces strong effects in both hosts (i.e. cytoplasmic incompatibility). It is likely that, at least in *D. subquinaria* carcass tissue, the ability to detect differential expression was hampered by variability in biological replicates of infected tissue samples (Figure 3.6A). This analysis was also limited to flies reared under ideal laboratory conditions. It is widely recognized that many host responses to *Wolbachia* and *Wolbachia*-host phenotypes are environmentally dependent (Brownlie et al., 2009; Hague et al., 2022), which may have prevented identification of important *Wolbachia*-host dynamics in natural populations.

Despite inconclusive results regarding differential expression, phenotyping results from this study reveal strong CI in *D. subquinaria*, which is different than previously measured *Wolbachia*-induced phenotypes. This result, in conjunction with high rates of maternal transmission and a lack of detectable fecundity costs, indicate that *Wolbachia* may indeed cross the species boundary from *D. recens* into *D. subquinaria* with minimal gene flow. Results presented in chapter four support the hypothesis that the *Wolbachia* variant examined here may be a recently derived *Wolbachia* variant in *D. recens*, lending credence to the possibility of *Wolbachia* invasion into *D. subquinaria* in the near future.

Works Cited

- Adrian, A., & Rahnenfuhrer, J. (2022). topGO: Enrichment Analysis for Gene Ontology.
- Bakovic, V., Schebeck, M., Telschow, A., Stauffer, C., & Schuler, H. (2018). Spatial spread of *Wolbachia* in *Rhagoletis cerasi* populations. *Biol Lett*, *14*(5). doi:10.1098/rsbl.2018.0161
- Balvin, O., Roth, S., Talbot, B., & Reinhardt, K. (2018). Co-speciation in bedbug *Wolbachia* parallel the pattern in nematode hosts. *Sci Rep*, *8*(1), 8797. doi:10.1038/s41598-018-25545-y
- Barton, N. H., & Turelli, M. (2011). Spatial waves of advance with bistable dynamics: cytoplasmic and genetic analogues of Allee effects. *Am Nat*, *178*(3), E48-75. doi:10.1086/661246
- Bordenstein, S. R., O'Hara, F. P., & Werren, J. H. (2001). *Wolbachia*-induced incompatibility precedes other hybrid incompatibilities in *Nasonia*. *Nature*, *409*(6821), 707-710. doi:10.1038/35055543

- Braig, H. R., Zhou, W., Dobson, S. L., & O'Neill, S. L. (1998). Cloning and characterization of a gene encoding the major surface protein of the bacterial endosymbiont *Wolbachia pipientis*. *J Bacteriol*, *180*(9), 2373-2378. doi:10.1128/JB.180.9.2373-2378.1998
- Brownlie, J. C., Cass, B. N., Riegler, M., Witsenburg, J. J., Iturbe-Ormaetxe, I., McGraw, E. A., & O'Neill, S. L. (2009). Evidence for metabolic provisioning by a common invertebrate endosymbiont, *Wolbachia pipientis*, during periods of nutritional stress. *PLoS Pathog*, *5*(4), e1000368. doi:10.1371/journal.ppat.1000368
- Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421. doi:10.1186/1471-2105-10-421
- Carrington, L. B., Lipkowitz, J. R., Hoffmann, A. A., & Turelli, M. (2011). A re-examination of *Wolbachia*-induced cytoplasmic incompatibility in California *Drosophila simulans*. *PLoS One*, *6*(7), e22565. doi:10.1371/journal.pone.0022565
- Caspari, E., Watson, G.S. (1959). On the Evolutionary Importance of Cytoplasmic Sterility in Mosquitoes. *Evolution*, *13*(4), 568-570. doi:10.2307/2406138
- Chen, Y., Lun, A. T., & Smyth, G. K. (2016). From reads to genes to pathways: differential expression analysis of RNA-Seq experiments using Rsubread and the edgeR quasi-likelihood pipeline. *F1000Res*, *5*, 1438. doi:10.12688/f1000research.8987.2
- Chrostek, E., Hurst, G. D. D., & McGraw, E. A. (2020). Infectious Diseases: Antiviral *Wolbachia* Limits Dengue in Malaysia. *Curr Biol*, *30*(1), R30-R32. doi:10.1016/j.cub.2019.11.046
- Chrostek, E., Marialva, M. S., Esteves, S. S., Weinert, L. A., Martinez, J., Jiggins, F. M., & Teixeira, L. (2013). *Wolbachia* variants induce differential protection to viruses in

- Drosophila melanogaster*: a phenotypic and phylogenomic analysis. *PLoS Genet*, 9(12), e1003896. doi:10.1371/journal.pgen.1003896
- Chrostek, E., Martins, N., Marialva, M. S., & Teixeira, L. (2021). Wolbachia-Conferred Antiviral Protection Is Determined by Developmental Temperature. *mBio*, 12(5), e0292320. doi:10.1128/mBio.02923-20
- Cruz, M. A., Magalhaes, S., Sucena, E., & Zele, F. (2021). Wolbachia and host intrinsic reproductive barriers contribute additively to postmating isolation in spider mites. *Evolution*. doi:10.1111/evo.14286
- Driscoll, T. P., Verhoeve, V. I., Brockway, C., Shrewsbury, D. L., Plumer, M., Sevdalis, S. E., . . . Gillespie, J. J. (2020). Evolution of Wolbachia mutualism and reproductive parasitism: insight from two novel strains that co-infect cat fleas. *PeerJ*, 8, e10646. doi:10.7717/peerj.10646
- Duploux, A., Hurst, G. D., O'Neill, S. L., & Charlat, S. (2010). Rapid spread of male-killing Wolbachia in the butterfly *Hypolimnas bolina*. *J Evol Biol*, 23(1), 231-235. doi:10.1111/j.1420-9101.2009.01891.x
- Dyer, K. A., Bewick, E. R., White, B. E., Bray, M. J., & Humphreys, D. P. (2018). Fine-scale geographic patterns of gene flow and reproductive character displacement in *Drosophila subquinaria* and *Drosophila recens*. *Mol Ecol*. doi:10.1111/mec.14825
- Ginsberg, P. S., Humphreys, D. P., & Dyer, K. A. (2019). Ongoing hybridization obscures phylogenetic relationships in the *Drosophila subquinaria* species complex. *J Evol Biol*, 32(10), 1093-1105. doi:10.1111/jeb.13512

- Gloor, G. B., Nassif, N. A., Johnson-Schlitz, D. M., Preston, C. R., & Engels, W. R. (1991). Targeted gene replacement in *Drosophila* via P element-induced gap repair. *Science*, 253(5024), 1110-1117. doi:10.1126/science.1653452
- Gramates, L. S., Agapite, J., Attrill, H., Calvi, B. R., Crosby, M. A., Dos Santos, G., . . . Strelets, V. B. (2022). FlyBase: a guided tour of highlighted features. *Genetics*, 220(4). doi:10.1093/genetics/iyac035
- Hague, M. T. J., Mavengere, H., Matute, D. R., & Cooper, B. S. (2020). Environmental and Genetic Contributions to Imperfect wMel-Like *Wolbachia* Transmission and Frequency Variation. *Genetics*, 215(4), 1117-1132. doi:10.1534/genetics.120.303330
- Hague, M. T. J., Shropshire, J. D., Caldwell, C. N., Statz, J. P., Stanek, K. A., Conner, W. R., & Cooper, B. S. (2022). Temperature effects on cellular host-microbe interactions explain continent-wide endosymbiont prevalence. *Curr Biol*, 32(4), 878-888 e878. doi:10.1016/j.cub.2021.11.065
- Hoffmann, A. A., Montgomery, B. L., Popovici, J., Iturbe-Ormaetxe, I., Johnson, P. H., Muzzi, F., . . . O'Neill, S. L. (2011). Successful establishment of *Wolbachia* in *Aedes* populations to suppress dengue transmission. *Nature*, 476(7361), 454-457. doi:10.1038/nature10356
- Hoffmann, A. A., Turelli, M., & Harshman, L. G. (1990). Factors affecting the distribution of cytoplasmic incompatibility in *Drosophila simulans*. *Genetics*, 126(4), 933-948. doi:10.1093/genetics/126.4.933
- Hosokawa, T., Koga, R., Kikuchi, Y., Meng, X. Y., & Fukatsu, T. (2010). *Wolbachia* as a bacteriocyte-associated nutritional mutualist. *Proc Natl Acad Sci U S A*, 107(2), 769-774. doi:10.1073/pnas.0911476107

- Humphreys, D. P., Rundle, H. D., & Dyer, K. A. (2016). Patterns of reproductive isolation in the *Drosophila subquinaria* complex: can reinforced premating isolation cascade to other species? *Curr Zool*, *62*(2), 183-191. doi:10.1093/cz/zow005
- Hurst, L. M., G.T. . (1996). Clade Selection, Reversible Evolution and the Persistence of Selfish elements:: The Evolutionary Dynamics of Cutoplasmic Incompatibility. *Proc. R. Soc. London Ser. B-Biological Sci.*, *263*(104).
- Jaenike, J., Dyer, K. A., Cornish, C., & Minhas, M. S. (2006). Asymmetrical reinforcement and Wolbachia infection in *Drosophila*. *PLoS Biol*, *4*(10), e325. doi:10.1371/journal.pbio.0040325
- Jaenike, J. (2007). Spontaneous emergence of a new Wolbachia phenotype. *Evolution*, *61*(9), 2244-2252. doi:10.1111/j.1558-5646.2007.00180.x
- Jones, P., Binns, D., Chang, H. Y., Fraser, M., Li, W., McAnulla, C., . . . Hunter, S. (2014). InterProScan 5: genome-scale protein function classification. *Bioinformatics*, *30*(9), 1236-1240. doi:10.1093/bioinformatics/btu031
- Kaur, R., Shropshire, J. D., Cross, K. L., Leigh, B., Mansueto, A. J., Stewart, V., . . . Bordenstein, S. R. (2021). Living in the endosymbiotic world of Wolbachia: A centennial review. *Cell Host Microbe*, *29*(6), 879-893. doi:10.1016/j.chom.2021.03.006
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*, *37*(8), 907-915. doi:10.1038/s41587-019-0201-4
- Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, *23*(10), 1289-1291. doi:10.1093/bioinformatics/btm091

- Kriesner, P., Hoffmann, A. A., Lee, S. F., Turelli, M., & Weeks, A. R. (2013). Rapid sequential spread of two *Wolbachia* variants in *Drosophila simulans*. *PLoS Pathog*, *9*(9), e1003607. doi:10.1371/journal.ppat.1003607
- Lopez-Madrigal, S., & Duarte, E. H. (2019). Titer regulation in arthropod-*Wolbachia* symbioses. *FEMS Microbiol Lett*, *366*(23). doi:10.1093/femsle/fnz232
- Martinez, J., Longdon, B., Bauer, S., Chan, Y. S., Miller, W. J., Bourtzis, K., . . . Jiggins, F. M. (2014). Symbionts commonly provide broad spectrum resistance to viruses in insects: a comparative analysis of *Wolbachia* strains. *PLoS Pathog*, *10*(9), e1004369. doi:10.1371/journal.ppat.1004369
- Min, K. T., & Benzer, S. (1997). *Wolbachia*, normally a symbiont of *Drosophila*, can be virulent, causing degeneration and early death. *Proc Natl Acad Sci U S A*, *94*(20), 10792-10796. doi:10.1073/pnas.94.20.10792
- Ndii, M. Z., Allingham, D., Hickson, R. I., & Glass, K. (2016). The effect of *Wolbachia* on dengue outbreaks when dengue is repeatedly introduced. *Theor Popul Biol*, *111*, 9-15. doi:10.1016/j.tpb.2016.05.003
- Pinto, S. B., Riback, T. I. S., Sylvestre, G., Costa, G., Peixoto, J., Dias, F. B. S., . . . Moreira, L. A. (2021). Effectiveness of *Wolbachia*-infected mosquito deployments in reducing the incidence of dengue and other *Aedes*-borne diseases in Niteroi, Brazil: A quasi-experimental study. *PLoS Negl Trop Dis*, *15*(7), e0009556. doi:10.1371/journal.pntd.0009556
- Prout, T. (1994). Some Evolutionary Possibilities for a Microbe That Causes Incompatibility in Its Host. *Evolution*, *48*(3), 909-911. doi:10.1111/j.1558-5646.1994.tb01371.x

- Putri, G. H., Anders, S., Pyl, P. T., Pimanda, J. E., & Zanini, F. (2022). Analysing high-throughput sequencing data in Python with HTSeq 2.0. *Bioinformatics*, 38(10), 2943-2945. doi:10.1093/bioinformatics/btac166
- Raychoudhury, R., Baldo, L., Oliveira, D. C., & Werren, J. H. (2009). Modes of acquisition of Wolbachia: horizontal transfer, hybrid introgression, and codivergence in the *Nasonia* species complex. *Evolution*, 63(1), 165-183. doi:10.1111/j.1558-5646.2008.00533.x
- Robinson, M. D., McCarthy, D. J., & Smyth, G. K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics*, 26(1), 139-140. doi:10.1093/bioinformatics/btp616
- Sayers, E. W., Bolton, E. E., Brister, J. R., Canese, K., Chan, J., Comeau, D. C., . . . Sherry, S. T. (2022). Database resources of the national center for biotechnology information. *Nucleic Acids Res*, 50(D1), D20-D26. doi:10.1093/nar/gkab1112
- Schuler, H., Bertheau, C., Egan, S. P., Feder, J. L., Riegler, M., Schlick-Steiner, B. C., . . . Stauffer, C. (2013). Evidence for a recent horizontal transmission and spatial spread of Wolbachia from endemic *Rhagoletis cerasi* (Diptera: Tephritidae) to invasive *Rhagoletis cingulata* in Europe. *Mol Ecol*, 22(15), 4101-4111. doi:10.1111/mec.12362
- Shoemaker, D. D., Katju, V., & Jaenike, J. (1999). Wolbachia and the Evolution of Reproductive Isolation between *Drosophila Recens* and *Drosophila Subquinaria*. *Evolution*, 53(4), 1157-1164. doi:10.1111/j.1558-5646.1999.tb04529.x
- Simon C, F. F., Beckenbach A, Crespi B, Liu H, Flook P (1994). Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann Entomol Soc Am* 81, 651-701.

- Turelli, M. (1994). Evolution of Incompatibility-Inducing Microbes and Their Hosts. *Evolution*, 48(5), 1500-1513. doi:10.1111/j.1558-5646.1994.tb02192.x
- Turelli, M., Cooper, B. S., Richardson, K. M., Ginsberg, P. S., Peckenpaugh, B., Antelope, C. X., . . . Hoffmann, A. A. (2018). Rapid Global Spread of wRi-like Wolbachia across Multiple *Drosophila*. *Curr Biol*, 28(6), 963-971 e968. doi:10.1016/j.cub.2018.02.015
- Turelli, M., & Hoffmann, A. A. (1995). Cytoplasmic incompatibility in *Drosophila simulans*: dynamics and parameter estimates from natural populations. *Genetics*, 140(4), 1319-1338. doi:10.1093/genetics/140.4.1319
- Turelli, M., Hoffmann, A. A., & McKechnie, S. W. (1992). Dynamics of cytoplasmic incompatibility and mtDNA variation in natural *Drosophila simulans* populations. *Genetics*, 132(3), 713-723. doi:10.1093/genetics/132.3.713
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res*, 40(15), e115. doi:10.1093/nar/gks596
- Weinert, L. A., Araujo-Jnr, E. V., Ahmed, M. Z., & Welch, J. J. (2015). The incidence of bacterial endosymbionts in terrestrial arthropods. *Proc Biol Sci*, 282(1807), 20150249. doi:10.1098/rspb.2015.0249
- Werren, J. H., Baldo, L., & Clark, M. E. (2008). Wolbachia: master manipulators of invertebrate biology. *Nat Rev Microbiol*, 6(10), 741-751. doi:10.1038/nrmicro1969
- Werren, J. H., & Jaenike, J. (1995). Wolbachia and cytoplasmic incompatibility in mycophagous *Drosophila* and their relatives. *Heredity (Edinb)*, 75 (Pt 3), 320-326. doi:10.1038/hdy.1995.140

- Werren, J. H., & Windsor, D. M. (2000). Wolbachia infection frequencies in insects: evidence of a global equilibrium? *Proc Biol Sci*, 267(1450), 1277-1285. doi:10.1098/rspb.2000.1139
- Wheeler, M. R. (1960). NEW SPECIES OF THE QUINARIA GROUP OF DROSOPHILA (DIPTERA, DROSOPHILIDAE). *Southwestern Naturalist*, 5, 160.
- Ye, Y. H., Carrasco, A. M., Frentiu, F. D., Chenoweth, S. F., Beebe, N. W., van den Hurk, A. F., . . . McGraw, E. A. (2015). Wolbachia Reduces the Transmission Potential of Dengue-Infected *Aedes aegypti*. *PLoS Negl Trop Dis*, 9(6), e0003894. doi:10.1371/journal.pntd.0003894
- Yen, J. H., & Barr, A. R. (1971). New hypothesis of the cause of cytoplasmic incompatibility in *Culex pipiens* L. *Nature*, 232(5313), 657-658. doi:10.1038/232657a0
- Zug, R., & Hammerstein, P. (2015). Bad guys turned nice? A critical assessment of Wolbachia mutualisms in arthropod hosts. *Biol Rev Camb Philos Soc*, 90(1), 89-111. doi:10.1111/brv.12098

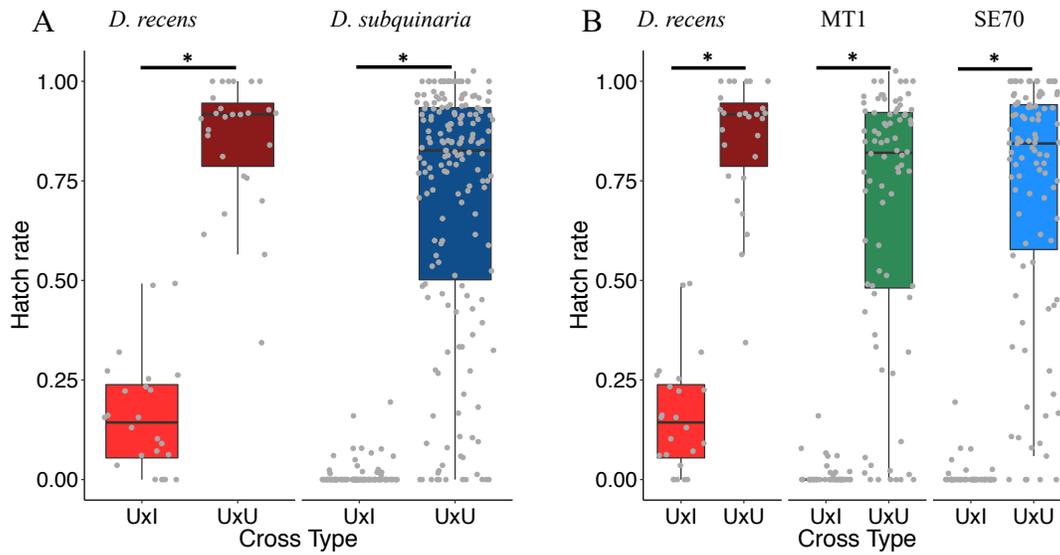


Figure 3.1: Offspring hatch rate of compatible and incompatible crosses in native host, *D. recens* (red), and non-native host, *D. subquinaria* (blue & green). Boxplots show median and 25th and 75th percentiles with whiskers to +/- 1.5 the interquartile range. Grey dots indicate hatch rate of all offspring from each individual female. MT1: Missoula stock, SE70: Seattle stock, *BC: backcrossed from *D. recens* (*Wolbachia* infected), *BCtet: backcrossed from *D. recens*, tetracycline-treated (*Wolbachia* uninfected). All crosses are written “female X male”, with “U” for uninfected flies and “I” for infected flies. Incompatible crosses are indicated as “UxI” and compatible crosses indicated “UxU”. “*” indicates significant difference ($P < 0.001$, Wilcoxon rank-sum test). A: Hatch rate of compatible and incompatible crosses for all *D. recens* and pooled *D. subquinaria* genotypes. B: Same as A, but with *D. subquinaria* split by genotype.

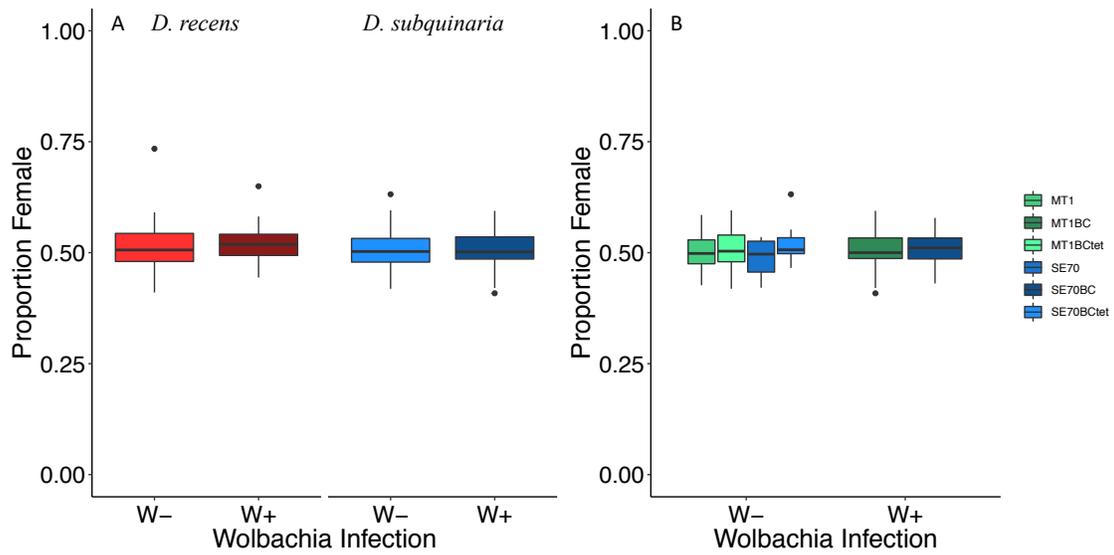


Figure 3.2: Offspring sex ratio of infected (dark shading) and uninfected (light shading) females. Boxplots show 25th and 75th percentiles with whiskers to +/- 1.5 the interquartile range. A: Offspring sex ratio of infected and uninfected females of *D.recens* (red) and *D. subquinaria* (blue & green). Grey points represent individual females. Black points are outliers. B: Offspring sex ratio among *D. subquinaria* genotypes and treatments. MT1: Missoula stock, SE70: Seattle stock, *BC: backcrossed from *D. recens* (*Wolbachia* infected), *BCtet: backcrossed from *D. recens*, tetracycline-treated (*Wolbachia* uninfected).

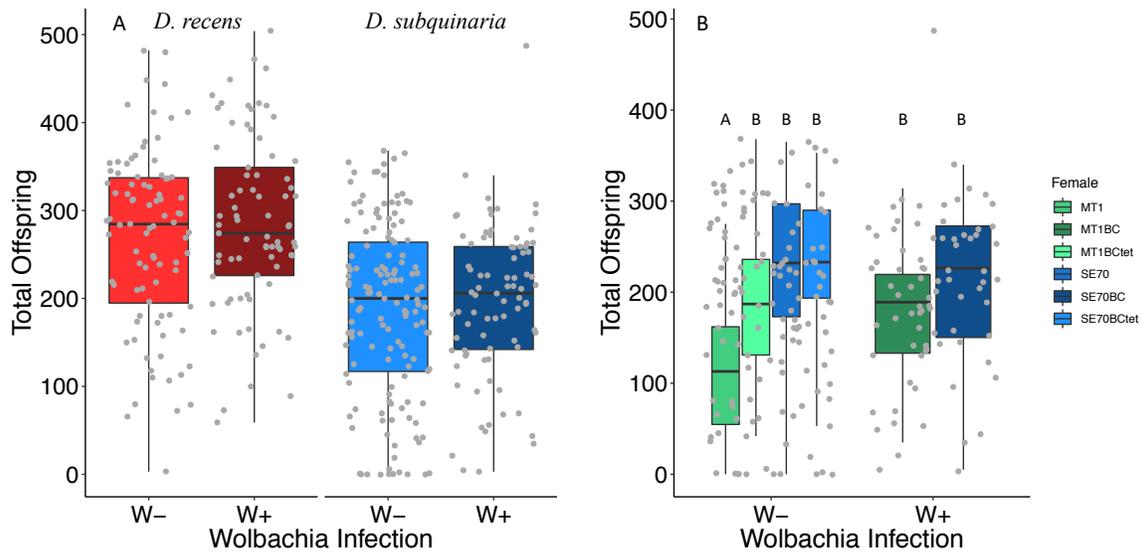


Figure 3.3: Offspring production of infected (darker shading) and uninfected (lighter shading) females. Boxplots show 25th and 75th percentiles with whiskers to ± 1.5 the interquartile range. A: Offspring Production of infected and uninfected females of *D.recens* (red) and *D. subquinaria* (blue & green). Grey points represent fecundity of individual females. B: Offspring Production among *D. subquinaria* genotypes and treatments. MT1: Missoula stock, SE70: Seattle stock, *BC: backcrossed from *D. recens* (*Wolbachia* infected), *BCtet: backcrossed from *D. recens*, tetracycline-treated (*Wolbachia* uninfected). Letters above boxplot indicate Tukey groupings.

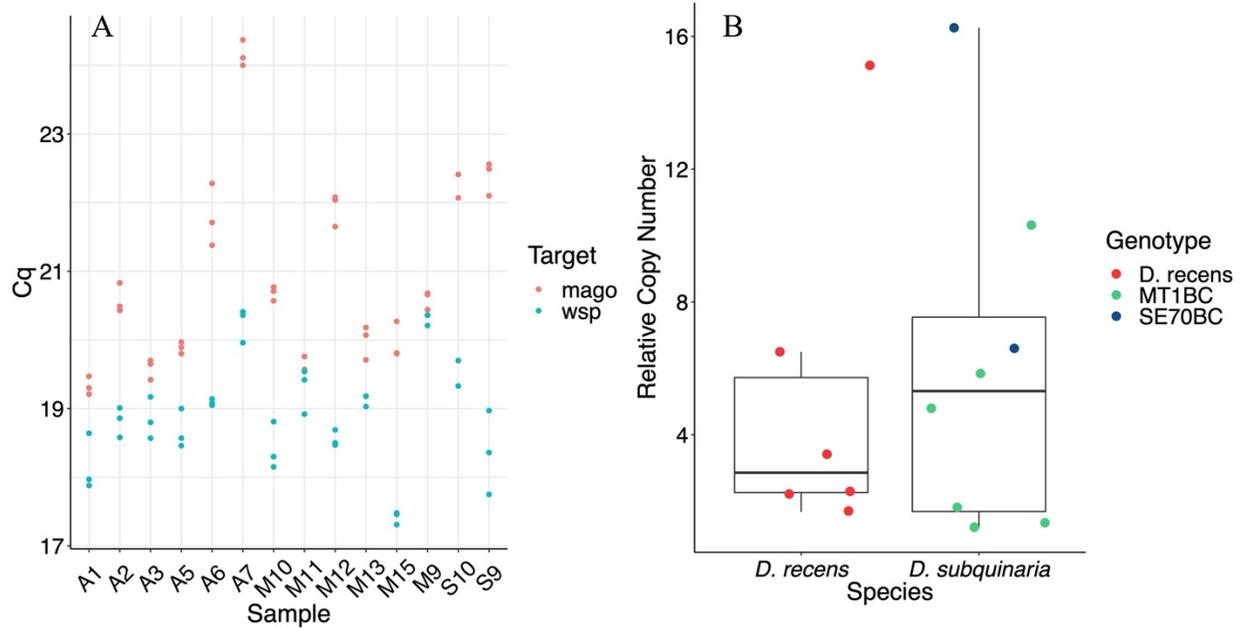


Figure 3.4: *Wolbachia* titer as measured by relative copy number of a host locus (*mago*) to a *Wolbachia* locus (*wsp*). A: Individual samples with technical replicate Cq values. Red: Cq Values of host locus. Blue: Cq values of *Wolbachia* locus. Sample A#: *D. recens*. Sample M#: MT1BC *D. subquinaria*. Sample S#: SE70BC *D. subquinaria*. B: Relative copy number of *Wolbachia* to host gene copies ($2^{-\Delta Cq}$) for *D. recens* and *D. subquinaria* samples. Colors indicate genotype of sample. Red: *D. recens*. Green: MT1BC *D. subquinaria*. Blue: SE70BC *D. subquinaria*.

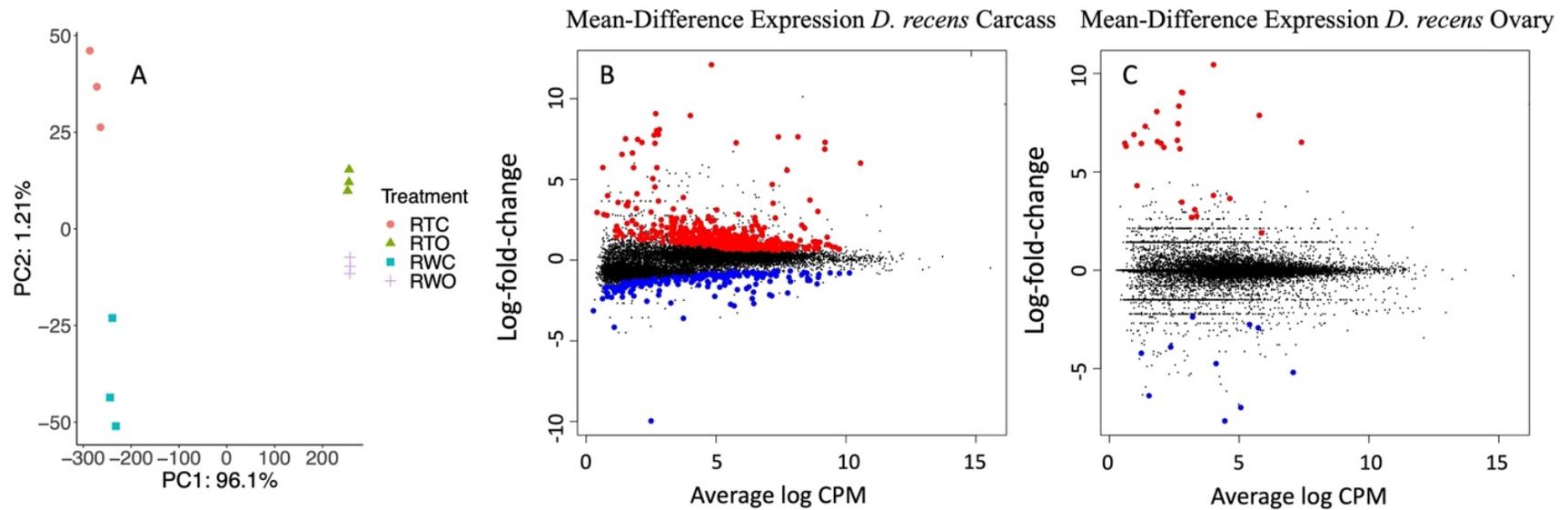


Figure 3.5: Differential expression of *Wolbachia*-infected and uninfected *D. recens* females. A: PCA analysis of count data for all genes. RTC: Uninfected carcass samples, RTO: Uninfected ovary samples, RWC: Infected carcass samples, RWO: Infected ovary samples. B: Mean-Difference of Expression Plot for carcass tissue samples. C: Mean-Difference Expression Plot for ovary tissue samples. For both B and C: Red dots indicate genes significantly upregulated in *Wolbachia*-infected samples. Blue dots indicate genes significantly downregulated in *Wolbachia*-infected samples. Black dots indicate genes that are not significant.

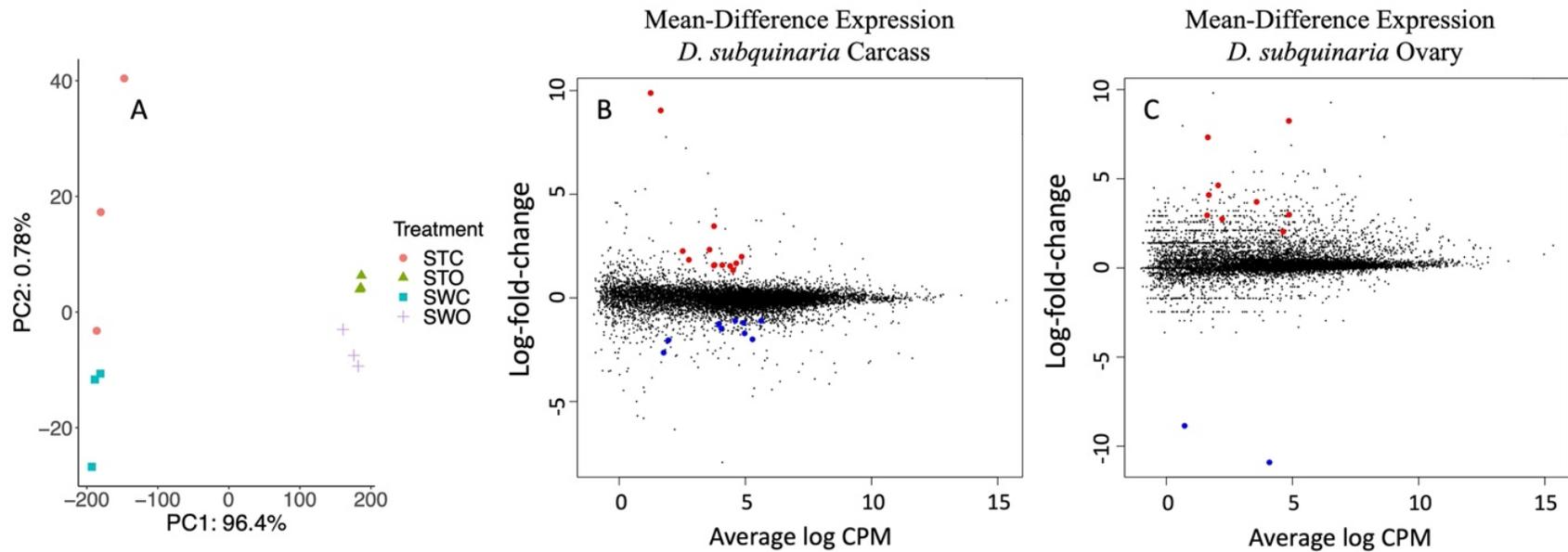


Figure 3.6: Differential expression of *Wolbachia*-infected and uninfected *D. subquinaria* females. A: PCA analysis of count data for all genes. STC: Uninfected carcass samples, STO: Uninfected ovary samples, SWC: Infected carcass samples, SWO: Infected ovary samples. B: Mean-Difference of Expression Plot for carcass tissue samples. C: Mean-Difference Expression Plot for ovary tissue samples. For both B and C: Red dots indicate genes significantly upregulated in *Wolbachia*-infected samples. Blue dots indicate genes significantly downregulated in *Wolbachia*-infected samples. Black dots indicate genes that are not significant.

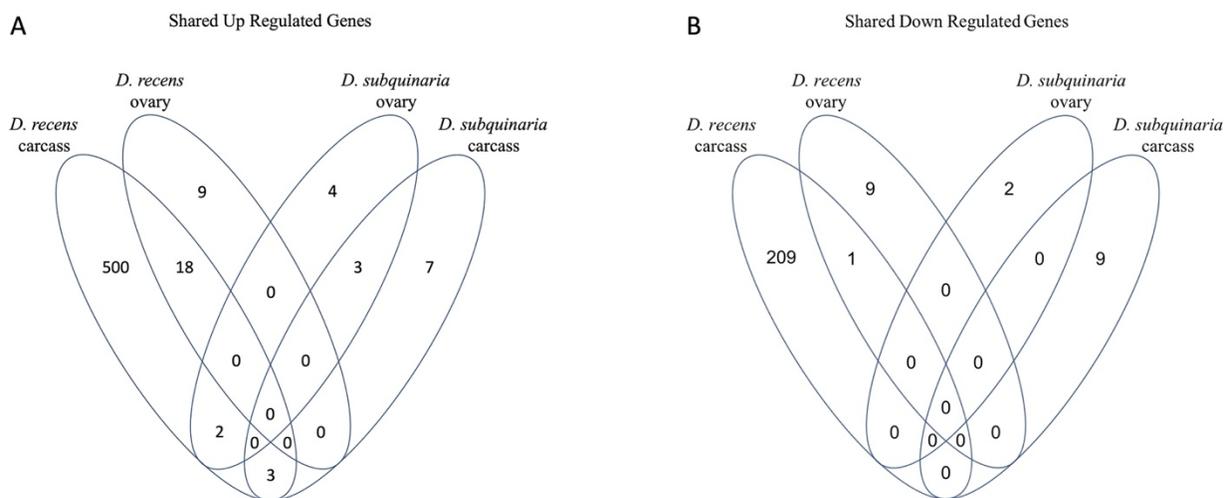


Figure 3.7: Number of shared differentially expressed genes among species and tissues. A: Shared up regulated genes. B: Shared down regulated genes.

Table 3.1: Experimental crosses for cytoplasmic incompatibility in *D. recens* and *D. subquinaria*. UxU: Compatible cross between two uninfected flies. UxI: Incompatible cross between an uninfected female and infected male. N Crosses gives total number of crosses analyzed for each cross.

Species	Female	Male	Cross Type	N Crosses	
<i>D. subquinaria</i>		SE70	UxU	29	
	SE70	SE70BC	UxI	26	
		SE70BCtet	UxU	27	
	SE70BCtet	SE70	UxU	29	
		SE70BC	UxI	28	
		SE70BCtet	UxU	24	
	MT1		MT1	UxU	25
		MT1	MT1BC	UxI	18
			MT1BCtet	UxU	25
		MT1BCtet	MT1	UxU	23
			MT1BC	UxI	12
	MT1BCtet		UxU	21	
<i>D. recens</i>	Albtet	Albtet	UxU	29	
		AlbW+	UxI	23	

Table 3.2: Experimental crosses for relative fecundity in *D. recens* and *D. subquinaria*. All *D. subquinaria* males were from the original, *Wolbachia* uninfected, fly stock. Total number of crosses analyzed are in the “N Crosses” column.

Species	Female	Male	N Crosses
<i>D. subquinaria</i>	SE70BCtet		60
	SE70BC	SE70	62
	SE70		59
	MT1BCtet		40
	MT1BC	MT1	46
	MT1		45
<i>D. recens</i>	AlbW+		77
	Albtet	Albtet	88

Table 3.3: Maternal transmission estimates for *D. recens* and *D. subquinaria* genotypes. N is the number of sample size, μ is the proportion of uninfected offspring produced by infected females. 95% binomial confidence intervals are indicated in parentheses.

Species/Genotype	N Female	N Uninfected	N Offspring	μ (95% CI)
<i>D. subquinaria</i> total	33	3	779	0.0039 (0, 0.0082)
<i>D. subquinaria</i> SE70	18	0	397	0 (0, 0)
<i>D. subquinaria</i> MT1	15	3	382	0.0079 (0, 0.017)
<i>D. recens</i> total	24	1	408	0.0025 (0, 0.0072)
<i>D. recens</i> Alb	16	0	230	0 (0, 0)
<i>D. recens</i> Fern Lake	6	1	178	0.0056 (0, 0.017)

Table 3.4: Model parameter and equilibrium frequency estimates. s_h : Reduction in hatch rate from incompatible cross, strength of cytoplasmic incompatibility, F : Relative fecundity of infected and uninfected females, μ : proportion of uninfected offspring produced by infected females, $\hat{p}_{\text{unstable}}$: unstable equilibrium frequency, \hat{p}_{stable} : stable equilibrium frequency. Ninety-five percent confidence intervals are in parentheses.

Species	S_h (95% CI)	F (95% CI)	μ (95% CI)	$\hat{p}_{\text{unstable}}$	\hat{p}_{stable}
<i>D. recens</i>	0.816 (0.74, 0.87)	1.064 (0.96, 1.19)	0.0025 (0, 0.0072)	0 (0, 0.084)	0.99 (0.99, 1)
<i>D. subquinaria</i>	0.987 (0.97, 0.99)	1.047 (0.92, 1.19)	0.0039 (0, 0.0082)	0 (0, 0.10)	0.99 (0.99, 1)

Table 3.5: Summary of differentially expressed genes in *D. recens* and *D. subquinaria*. The number of differentially expressed genes with BLAST annotations are indicated in parentheses. Direction of regulation is with respect to *Wolbachia*-infected samples (i.e. up regulated genes have higher expression in infected samples).

Species	Tissue	Up regulated	Down regulated
<i>D. subquinaria</i>	Carcass	13 (13)	9 (9)
	Ovary	9 (9)	2 (1)
<i>D. recens</i>	Carcass	520 (402)	210 (198)
	Ovary	27 (25)	10 (10)

Table 3.6: All differentially expressed genes in *D. subquinaria* with BLAST annotations. Up and down regulation are with respect to *Wolbachia*-infected samples. Gene IDs are from Braker annotation. Gene names are taken from homologs *D. melanogaster* reference genome release r6.46. (Flybase.org). Gene descriptions are from best BLAST hit accessions.

Gene ID	Best hit BLAST accession	Gene Name	Length query	Description	Tissue	Expression
g3076	XP_034485706	CG10877-PA	441	succinate--hydroxymethylglutarate CoA-transferase	Carcass	Up
g32050	XP_034484828	CG11893-PA	404	uncharacterized protein LOC117789802	Carcass	Up
g4786	XP_034474438	CG15263-PA	302	uncharacterized protein LOC117781731	Carcass	Up
g37644	XP_034475140	CG30281-PA	212	fibrinogen-like protein 1	Carcass	Up
g37921	XP_034478388	CG12107-PA	134	respirasome Complex Assembly Factor 1	Carcass	Up
g37635	XP_034488326	GatB-PA	443	glutamyl-tRNA(Gln) amidotransferase subunit B, mitochondrial	Carcass	Up
g34194	XP_034479118	Cog6-PB	249	conserved oligomeric Golgi complex subunit 6	Carcass	Up
g18966	XP_034474216	Ugt37A2-PA	484	UDP-glucuronosyltransferase 2B15-like	Carcass	Up
g34193	XP_034479118	Cog6-PA	354	conserved oligomeric Golgi complex subunit 6	Carcass	Up
g34196	XP_034479098	Pkn-PO	302	serine/threonine-protein kinase N isoform X5	Carcass	Up
g34195	XP_034479119	NA	160	SAP30-binding protein	Carcass	Up
g36668	XP_034478122	CG9005-PE	1106	uncharacterized protein LOC117784482	Carcass	Up
g37920	XP_034478387	U2A-PA	264	probable U2 small nuclear ribonucleoprotein A'	Carcass	Up

g3078	XP_034485439	TFAM-PA	258	transcription factor A, mitochondrial isoform X2	Carcass	Down
g3075	XP_034487408	CG15922-PB	66	uncharacterized protein LOC117791701	Carcass	Down
g28143	XP_034481133	nudC-PB	150	nuclear migration protein nudC	Carcass	Down
g37411	XP_034479119	CG2063-PA	320	SAP30-binding protein	Carcass	Down
g34168	XP_034478388	CG12107-PA	136	respirasome Complex Assembly Factor 1	Carcass	Down
g24114	XP_034480406	NA	60	protein-S-isoprenylcysteine O-methyltransferase	Carcass	Down
g3020	XP_034489279	Cys-PA	126	cystatin-like protein	Carcass	Down
g33427	XP_001984469	CG9004-PA	847	nucleolar MIF4G domain-containing protein 1 homolog	Carcass	Down
g26600	XP_034486771	Gnfl-PA	1012	replication factor C subunit 1	Carcass	Down
g3076	XP_034485706	CG10877-PA	441	succinate--hydroxymethylglutarate CoA-transferase	Ovary	Up
g36449	XP_034478769	Fatp3-PA	677	long-chain fatty acid transport protein 4	Ovary	Up
g37635	XP_034488326	GatB-PA	443	glutamyl-tRNA(Gln) amidotransferase subunit B, mitochondrial	Ovary	Up
g26851	XP_034483559	Sse-PA	648	uncharacterized protein LOC117788798 isoform X2	Ovary	Up
g38163	XP_002049625	pnut-PA	529	protein peanut	Ovary	Up
g33379	XP_034483523	CG11263-PB	289	protein Exd1 homolog	Ovary	Up
g4837	XP_034472627	CG11790-PA	301	thioredoxin domain-containing protein	Ovary	Up
g37921	XP_034478388	CG12107-PA	134	respirasome Complex Assembly Factor 1	Ovary	Up

g18541	XP_034474558	CG7860- PB	341	probable isoaspartyl peptidase/L- asparaginase GA20639	Ovary	Up
g36064	XP_030382495	NA	256	uncharacterized protein LOC115630008	Ovary	Down

Table 3.7: Top ten differentially expressed genes in each *D. recens* treatment with BLAST annotations. Up and down regulation are with respect to *Wolbachia*-infected samples. Gene IDs are from Braker annotation. Gene names are taken from homologs in *D. melanogaster* reference genome release r6.46. (Flybase.org). Gene descriptions are from best BLAST hit accessions.

Gene ID	Best hit BLAST accession	Gene Name	Length query	Description	Tissue	Expression
g31432	XP_034480827	RpL18-PB	188	60S ribosomal protein L18	carcass	up
g23671	XP_034482529	Tak1-PA	487	mitogen-activated protein kinase kinase 10-like	carcass	up
g12617	XP_002006628	Ufm1-PA	84	ubiquitin-fold modifier 1	carcass	up
g38585	XP_034482303	Jon66Ci-PA	260	serine protease 1-like	carcass	up
g21330	XP_034483897	CG8562-PA	423	carboxypeptidase B-like	carcass	up
g11813	XP_034477470	CG11414-PA	911	E3 ubiquitin-protein ligase ZNF598	carcass	up
g1313	XP_034488930	CG6204-PA	916	NFX1-type zinc finger-containing protein 1-like	carcass	up
g36874	XP_034111367	CG5157-PA	123	cytochrome b5	carcass	up
g22450	XP_002051316	Fic-PA	280	protein adenylyltransferase Fic	carcass	up
g31425	XP_034482470	D19A-PA	831	zinc finger protein 85-like	carcass	up
g27028	XP_034482300	Jon25Bi-PB	264	serine protease 1-like	carcass	down
g12072	XP_034476695	Mal-A4-PA	578	maltase A3	carcass	down
g26135	XP_034472391	CG4678-PI	527	carboxypeptidase D isoform X2	carcass	down
g34133	XP_034488835	CG31198-PA	937	membrane alanyl aminopeptidase	carcass	down
g2243	XP_034487938	NA	77	vasotab-like isoform X2	carcass	down
g27966	XP_034482043	mag-PA	401	lipase 3	carcass	down
g23610	XP_034482932	CG4835-PB	906	mucin-5AC isoform X2	carcass	down
g37148	XP_034481151	CG32277-PA	257	trypsin II-P29	carcass	down
g32344	XP_002055046	CG12576-PE	215	uncharacterized protein LOC6631944	carcass	down
g29728	XP_034478424	CG31089-PA	376	lipase 3-like	carcass	down
g31432	XP_034480827	RpL18-PB	188	60S ribosomal protein L18	ovary	up
g34753	XP_034490598	tyf-PN	2299	pneumococcal serine-rich repeat protein	ovary	up
g31425	XP_034482470	D19A-PA	831	zinc finger protein 85-like	ovary	up
g31423	XP_034484178	D19A-PA	719	gastrula zinc finger protein XICGF58.1-like	ovary	up
g31429	XP_030569913	CG10274-PB	869	zinc finger protein 540-like	ovary	up
g31422	XP_034484177	D19A-PA	734	zinc finger protein 675-like	ovary	up
g34755	XP_034489569	Nsun2-PA	746	tRNA (cytosine(34)-C(5))-methyltransferase	ovary	up
g31431	XP_034111737	BHD-PA	482	folliculin	ovary	up

g31424	XP_034484174	D19A-PA	765	zinc finger protein 708-like	ovary	up
g31433	XP_034111739	mRpL50-PA	193	39S ribosomal protein L50, mitochondrial	ovary	up
g31594	XP_034476321	NA	433	uncharacterized protein LOC117783168	ovary	down
g35988	XP_034477139	Obp56d-PB	132	general odorant-binding protein 56d-like	ovary	down
g22313	XP_034097411	NA	381	uncharacterized protein LOC117563283	ovary	down
g33510	XP_034486469	NA	508	putative aminopeptidase W07G4.4	ovary	down
g35930	XP_030566462	NA	671	uncharacterized protein LOC115766612	ovary	down
g374	XP_034489491	NA	758	regulator of G-protein signaling 7	ovary	down
g22644	XP_034489021	NA	536	membrane metallo-endopeptidase-like 1	ovary	down
g1208	XP_034490488	NA	726	sodium-dependent neutral amino acid transporter B(0)AT3	ovary	down
g966	XP_034490122	NA	100	protein roadkill-like	ovary	down
g11884	XP_034479242	NA	1429	pikachurin, partial	ovary	down

Table 3.8: Significant GO terms for differentially expressed genes in *D. recens*.

GO term	Annotation	nGenes	Significance
Carcass Up regulated			
GO:0008150	biological_process	409	1.60E-11
GO:0006357	regulation of transcription by RNA polymerase II	48	0.0049
Carcass Down regulated			
GO:0002181	cytoplasmic translation	17	2.40E-06
GO:0006508	proteolysis	51	0.001
GO:0006094	gluconeogenesis	5	0.0047
GO:0008345	larval locomotory behavior	3	0.0061
Ovary Up regulated			
GO:0006355	regulation of transcription, DNA-templated	79	2.50E-05
GO:0036158	outer dynein arm assembly	3	0.0057

CHAPTER 4
GENOME ASSEMBLIES FOR FOUR SPECIES OF THE QUINARIA SPECIES GROUP
AND THEIR *WOLBACHIA* INFECTIONS

Introduction

The *Drosophila quinaria* species group is a group of twenty-six boreal, primarily mushroom-feeding *Drosophila* species, with a distribution spanning the northern hemisphere. This species group has served a model for investigating incipient speciation, reinforcement, viral immunity, meiotic drive, endosymbiont-host interactions, and toxin resistance, among others. The *D. subquinaria* species complex is a young clade of three closely related species within the *quinaria* species group inhabiting North America and Northern Eurasia. One species, *D. transversa*, inhabits northern Europe, and is thought to reside across northern Asia. The other two species, *D. recens* and *D. subquinaria*, are North American species with *D. recens* residing from the Eastern US to the Canadian Rockies, and *D. subquinaria* from ~1,500 km east of the Canadian Rockies to the west coast on North America. All three species are incompletely reproductively isolated, and have been used extensively for studying speciation at its early stages. A *Wolbachia* infection in *D. recens* has opened research into how *Wolbachia* can increase postzygotic reproductive boundaries and drive character displacement in sympatric populations.

Despite their utility for studying speciation, the evolutionary history of this group has proved difficult to ascertain, likely due to signatures of incomplete lineage sorting and interspecific gene flow. In a phylogenetic and population genetic analysis using a Sanger

sequenced twenty-nine locus dataset, I found widespread gene tree-species tree discordance consistent with ILS and gene flow between *D. recens* and sympatric *D. subquinaria*, including mitochondrial introgression from *D. recens* into *D. subquinaria* (See Chapter 2, Ginsberg, Humphreys, & Dyer, 2019; Jaenike, Dyer, Cornish, & Minhas, 2006; Shoemaker, Katju, & Jaenike, 1999). Additionally, I found that low-recombining regions were particularly clear in resolving hypothesized species relationships compared to high recombining regions of the genome (See Chapter 2, Ginsberg et. al 2019). These results provide a foundation for understanding species interactions in incompletely reproductively isolated species in a genomic context.

The advent of long-read, single molecule sequencing has made high contiguity, de novo genome assemblies feasible for a wide variety of organisms. Notably, Kim et al. (2021) generated 101 drosophilid de novo genome assemblies providing an invaluable resource for comparative genomic analysis across *Drosophila* and related genera. At present, only one *quinaria* group species has a published full genome (*Drosophila innubila*; (Hill, Koseva, & Unckless, 2019). Here I present seven de novo genome assemblies across four species: *D. quinaria*, *D. transversa*, three *D. subquinaria* genomes, and two *D. recens* genomes. These samples comprise the whole *D. subquinaria* species complex with *D. quinaria* as an outgroup. I reassess phylogenetic relationships in the *D. subquinaria* species complex with greater sampling across the genome and find that species relationships are resolved in a manner consistent with their levels of reproductive isolation. I show substantial assembly size variation among species in this clade that is not explained by transposable element proliferation in the genome. I discuss possible artefacts of the assemblies that limit additional analyses.

Wolbachia-host dynamics have been a major avenue of research in the *D. subquinaria* species complex. With fly genome sequencing of two *D. recens* stocks, I was able to extract

Wolbachia reads and assemble full *Wolbachia* genomes. These two distinct *Wolbachia* strains infect geographically and temporally distant *D. recens* samples. The *Wolbachia* infection in the *D. recens* (wRec) is present at high frequency (~98%) (Shoemaker, Katju, & Jaenike, 1999) and causes strong cytoplasmic incompatibility (CI) in *D. recens* (see Chapter 3). Previous research has shown that *Wolbachia* introgressed into *D. subquinaria* switches from a CI causing infection to a male-killing infection (Jaenike 2007). Results from the Chapter 3 indicate that this is not the case for at least some wRec variants. In Chapter 3 I show that wRec caused CI in all *D. subquinaria* stocks tested. As *D. subquinaria* is not naturally infected with *Wolbachia*, it is likely that different *Wolbachia* variants are present in *D. recens* populations, or *D. recens* has recently undergone a *Wolbachia* sweep of a younger wRec variant. To test this, we used two *D. recens* samples collected ~10 years apart, and sampled from opposite ends of the species' range. From de novo fly genome assemblies, we extract *Wolbachia* contigs and reassemble full length *Wolbachia* genomes from each of our *D. recens* collections. The resulting *Wolbachia* genomes revealed distinct strains in our *D. recens* collections. I show that the two *Wolbachia* strains are remarkably similar, with only forty-five SNVs across the genome. Phylogenetic analysis reveals these strains almost certainly have a common ancestor after infecting *D. recens*. The primary differences between these *Wolbachia* variants are two major structural changes, a ~25.6kb indel, and a 33.3kb inversion. Notably, both wRec variants have identical genes previously shown to cause both CI and male-killing.

Methods

Fly Collections

Full details of fly collections are presented in Table 4.1. Fly collections were from wild-caught females established as isofemale lines in the laboratory. The exception being *D. recens* Alberta, which was a *D. recens* fly stock generated from mixing of several isofemale lines (see Chapter 3 for full details). We sampled one stock each from *D. quinaria* and *D. transversa*. For *D. recens* and *D. subquinaria*, we sampled stocks from each part of the species' ranges. For *D. recens* we particularly focused on sampling temporally separated *D. recens* collections as to capture variation in both the host and *Wolbachia* infection over time. For *D. subquinaria*, we sampled from across the species range, as previous research has indicated strong population structure and varying levels of mate discrimination (see Chapter 2). All assemblies and descriptions will use the collection name in Table 1 prepended by the species name abbreviation (R = *D. recens*, S = *D. subquinaria*, T = *D. transversa*, Q = *D. quinaria*).

DNA Isolation, Library Preparation and Sequencing

DNA extractions were performed using the Gentra Puregene Tissue Kit (QIAGEN, LLC) following the manufacturer's protocol. Extractions were performed using 10 whole female flies collected as virgins and aged for 3-5 days then placed live on ice for DNA extraction. DNA extractions were checked for purity using a Nanodrop spectrophotometer (Thermo Fischer Scientific, Inc.) to ensure a A260/A280 around 1.8 and a A260/A230 ratio around 2. DNA concentration was assessed using a Qubit fluorometer with the Qubit dsDNA HS Assay kit (Thermo Fischer Scientific, Inc.).

Long-read Oxford Nanopore (ONT) libraries were prepared using the SQK-LSK109 or SQK-LSK110 Genomic DNA by Ligation Library Prep Kit (Oxford Nanopore Technologies Plc.) following the manufacturer's protocol and including purification and clean up steps using the Agencourt AMPure XP beads (Beckman Coulter Inc.) and NEBNext FFPE DNA repair mix, Ultra II End repair and dA-tailing module reagents (New England Biolabs Inc.). Each library sample was sequenced independently on a MinION (Oxford Nanopore Technologies Plc.) using a R9.4.1 or R10.4 flow cell (Oxford Nanopore Technologies Plc.). Raw data output is shown in Table S4.1.

Short read library preparation and sequencing was done at the UGA GGBC using the same DNA isolates when possible, or generated from different samples of the same fly stock using the same protocol as described above. Genomic DNA libraries for 150 bp PE sequencing were prepared by the Schmitz lab at UGA. Briefly, Genomic DNA was sonicated to 500 bp with a Covaris S-series focused ultrasonicator and end-repaired using the End-it DNA end-repair kit (Biosearch Technologies). Then repaired DNA was A-tailed using the Klenow 3'-5' exo- (New England Biolabs) and ligated to methylated adaptors with T4 DNA ligase (New England Biolabs). DNA was then amplified and sequenced on a NovaSeq 6000 (Illumina Inc.). Raw data fastq files from short-read sequencing were trimmed and quality filtered with fastp v0.23.2 (Chen, Zhou, Chen, & Gu, 2018) with default parameters.

RNA Isolation and Sequencing

Whole RNA was extracted using TRIzol Reagent (Life Technologies Corp, Carlsbad, CA, USA) and following the manufacturer's protocol. For each sample ten age-controlled (five days old) and density-controlled (ten flies per vial) female or male flies were briefly anesthetized

on CO₂ and flash frozen in liquid nitrogen, then stored at -80 °C until RNA isolation. RNA isolates were analyzed with a Nanodrop spectrophotometer (Thermo Fischer Scientific, Inc.) to ensure a A160/A180 and a A250/A230 ratio around 2. RNA concentration was assessed using a Qubit fluorometer with the Qubit RNA BR Assay kit (Thermo Fischer Scientific, Inc.). Sequencing was done by the Georgia Genomics and Bioinformatics Core (GGBC). Library preparation for *D. recens* Alb stock was done by the GGBC (Table S4.2). For these samples, isolated RNA was quality checked with an Agilent Bioanalyzer (Agilent, Santa Clara, CA, USA). Libraries were prepared with KAPA biosystems stranded RNA kit and sequenced on a single Nextseq 150 Paired-end Mid output flow cell. All other samples (Table S4.2) were prepared by the Schmitz lab at UGA from 1.3 µg input RNA with the Illumina TruSeq mRNA Stranded Library Kit v2 (Illumina) according to the manufacturer's instructions. All volumes were reduced to 1/3 of the recommended quantity. Samples were sequenced on a Illumina NovaSeq 6000 at the GGBC. One sample, *D. recens* Alb, reused RNAseq data generated for differential expression (See Chapter 2 for details).

Fly Genome Assembly

The genome assembly pipeline was modeled on (B. Y. Kim et al., 2021). The full assembly pipeline for the genomes presented here is shown in Figure 4.1. First, raw nanopore reads were used to generate a de novo assembly using the de novo long-read assembler flye v2.8.1 (Freire, Ladra, & Parama, 2021) with default parameters. The de novo assembly was polished with two rounds of long read polishing with racon v1.4.13 (<https://github.com/isovic/racon>) with parameters -m 8 -x -6 -g -8, and then polished with medaka v1.2.1 (<https://github.com/nanoporetech/medaka>). Draft assembly completeness and

contiguity was then assessed with BUSCO v4.0.6 (Manni, Berkeley, Seppey, & Zdobnov, 2021) using the diptera database (diptera_odb10) and Quast v5.0.2 (Gurevich, Saveliev, Vyahhi, & Tesler, 2013), respectively. After draft assembly assessment, high BUSCO duplication rates warranted haplotig purging using the purge_haplotigs pipeline (Roach, Schmidt, & Borneman, 2018) to reduce heterozygosity of the genome. After haplotig purging, contigs were re-scaffolded using npScarf (Cao et al., 2017) and polished with 150bp paired-end Illumina data with Pilon v1.23 (Walker et al., 2014). Finally, non-*Drosophila* contigs were removed by blasting contigs and removing any whose first hit was not *Drosophila*. This was followed by a final round of BUSCO v4.0.6 (Manni et al., 2021) and Quast v5.0.2 (Gurevich et al., 2013) to assess final genome completeness and contiguity.

Fly Genome Annotation

First, repeat content was predicted using Repeatmodeler v2.0.1 (Flynn JM, 2020) followed by genome masking with Repeatmasker v4.0.9 (Smit, 2015). RNAseq data was mapped to the masked genome assemblies using HISAT2 v2.2.1 (D. Kim, Paggi, Park, Bennett, & Salzberg, 2019). Finally, de novo genome annotation was implemented with Braker v2.1.5 (Hoff, Lomsadze, Borodovsky, & Stanke, 2019). All programs were run with default parameters unless noted otherwise.

Wolbachia Genome Assembly and Annotation

All contigs from draft host assemblies were searched for homology with *Wolbachia* with blastn megablast against the NCBI nt database (2021/11/11 release). All contigs with a best hit to *Wolbachia* were saved. All ONT long read data was mapped back to *Wolbachia* contigs using

minimap2 v2.17 (Li, 2018). Illumina short read data was mapped to *Wolbachia* contigs with Bowtie2 v2.4.5 (Langmead & Salzberg, 2012). Mapped long and short reads were then used for a hybrid assembly with Unicycler (Wick, Judd, Gorrie, & Holt, 2017). The final assemblies were annotated with Prokka v1.14.5 (Seemann, 2014), using the reference *wMel* assembly (RefSeq: GCF_000008025.1) to guide annotations. *wRec* genomes were aligned to the *wMel* reference and published *wRec* genome (RefSeq: GCF_000742435.1) and visualized in Mauve (Darling, Mau, Blattner, & Perna, 2004).

Phylogenetic Analysis

Fly phylogenetic analysis was conducted with two methodologies. Both were implemented using a random subset of 100 single copy orthologs retrieved from the final assemblies' BUSCO output. For the first method, gene sequences were aligned using MUSCLE v5.1 (Madeira et al., 2022), then concatenated in the program Trifusion (<https://github.com/ODiogoSilva/TriFusion>). I generated a maximum-likelihood phylogeny with all loci in RAxML v8.2.12 (Stamatakis, Ludwig, & Meier, 2005) with 1000 rapid bootstrap analyses (-f a -N 1000). As concatenation assumes a shared topology among all loci, I also implemented a consensus approach from gene trees. First, gene trees were generated in PhyML v3.0 (Guindon et al., 2010) under a GTR substitution model. From these gene trees a consensus tree was generated in the program ASTRAL-III (Zhang, Rabiee, Sayyari, & Mirarab, 2018) with default parameters.

We generated a *Wolbachia* phylogeny based on the five Multi-Locus Sequence Typing (MLST) loci (Baldo et al., 2006) and the *Wolbachia* surface protein (*wsp*). To determine the origin of our distinct *Wolbachia* variants, we sampled eleven *Wolbachia* strains infecting a

variety of insect species, including some known to be closely related to wRec (e.g. wMel, infection of *D. melanogaster*), and some more distantly related (e.g. wPip, infection of *Culex pipiens*). MLST sequences for *Wolbachia* strains were downloaded from the PubMLST database (<https://pubmlst.org/organisms/wolbachia-spp>, Isolate ID's in Table S4.3). Sequence alignments were concatenated manually, and a phylogeny was generated in PhyML v3.0 (Guindon et al., 2010) with a GTR substitution model and 100 bootstrap replicates.

Wolbachia Strain PCR

To assess prevalence of *Wolbachia* variants in *D. recens* populations, we designed PCR primers that to amplify *Wolbachia* variant-specific amplicons. All primers were designed with Primer3 (Koressaar & Remm, 2007; Untergasser et al., 2012). Primers were designed to work with three primers per reaction. In each reaction we included forward and reverse primers flanking the 25 kb indel (Forward primer: GCAGCAAACCTTTACGCCTTG, reverse primer: GACGGAGCTGCAGATAGTGA , which should only amplify in Alb wRec, and a unique forward primer (TGGGAGGTGCTCTTGCTACT) that is within the indel and so with the reverse primer should only amplify in RW wRec. Unfortunately, during designing of the primers, it was missed that just within the flanking regions of the 5' end of the indel, there is high homology of transposases and surrounding sequence. This resulted in two bands being amplified with RW wRec (Figure S4.1), one with the indel-specific forward primer and reverse flanking primer, and one with the flanking forward primer and the flanking reverse primer (but this reverse primer annealed within the indel). The variant-specific amplicon size was additionally verified in silico. Initial testing produced clear and reproducible results using these primers. all primers are written 5' – 3'.

Results

Fly Genome Assemblies

All fly genomes were assembled into fewer than 550 contigs (Table 4.2). Genome assembly N50s ranged for *D. subquinaria* were 2,301,551 bp and 4,492,352 bp in *D. subquinaria* S_Cal241 and *D. subquinaria* S_SE70, respectively. The third *D. subquinaria* sample, S_MT1, was closer to S_SE70 with N50 = 2,643,564 bp. Of my two *D. recens* samples, R_AlB had an N50 = 5,462,940 and R_RW had an N50 = 3,525,288. Assemblies for *D. quinaria* and *D. transversa* both had substantially higher N50s, indicating greater contiguity of these two genomes. The *D. transversa* T_T6 N50 = 13,123,973 bp and the *D. quinaria* Q_QTP N50 = 20,712,610 bp. Assembly sizes from all samples were largely similar with exception of one outgroup sample, Q_QTP, which was substantially smaller than the remaining genomes at 148.7Mb. By comparison, *D. recens* genomes for R_AlB and R_RW were 227.8Mb and 234.4Mb, respectively. The assembly for *D. transversa* T_T6 was similar at 230.4Mb. *D. subquinaria* had a much larger range in assembly sizes than interspecific comparisons in *D. recens*. *D. subquinaria* S_MT1 had the smallest assembly at 209,903,681bp and S_Cal241 had the largest assembly with 250,861,281bp. S_SE70 was in between with an assembly size of 230,599,954bp.

Of samples in the *D. subquinaria* species complex, BUSCO analysis showed nearly complete genomes for all assemblies, with only one sample, S_SE70 having a complete BUSCO score below 99.0% (S_SE70 = 98.7%) using the diptera_odb10 database (Figure 4.2). However, the *D. quinaria* Q_QTP assembly had a complete BUSCO score of 95.8%. This combined with the far smaller assembly size suggests that a portion of the genome is not being represented in this assembly. Of complete BUSCOS, the vast majority were represented in a single copy.

However, BUSCO duplication rates were all above 1%. This is not surprising for *D. subquinaria* and *D. recens* samples as they have been shown to exhibit high genetic diversity (see Chapter 2). This is especially true of *D. recens* and sympatric *D. subquinaria*, which are estimated to have the highest genetic diversity among these species (Chapter 2, figure S2.6), and indeed have the highest BUSCO duplication rates (Table 4.2), suggesting that multiple haplotypes are failing to be purged during the assembly pipelines haplotig_purging step of the assembly pipeline.

There was high variation in the gene count estimates across assemblies, from a high gene count of 50,139 in S_Cal241 to a low of 15,356 in Q_QTP (Table 4.3). Only counting genes with partial or complete external evidence reduced the gene counts significantly, with the highest of 19,757 in S_SE70 and lowest of 13,342 in Q_QTP. These numbers, while still high, are more in line with genome assemblies of other *Drosophila* species (e.g. *Drosophila melanogaster* release 6 plus ISO1 MT gene count = 17,864).

Fly Phylogenetic Analysis

The full alignment of 100 random selected loci resulted in a dataset of 168,224 sites with 15,542 variable sites, of which 5,178 were parsimony informative. The concatenated phylogenetic analysis in RAxML produced the species shown in Figure 4.3A. Here, *D. subquinaria* is resolved as monophyletic species, with the sympatric *D. subquinaria* S_Cal241 being sister to the two allopatric *D. subquinaria* samples, S_SE70 and S_MT1. *D. transversa* T_T6 is sister to the *D. subquinaria* clade. *D. recens* R_Alba and R_RW represent a monophyletic clade for *D. recens* and are outgroup to all other species in the *D. subquinaria* species complex. *D. quinaria* is outgroup to the entire species complex.

To account for possible effects of incomplete lineage sorting, I also analyzed gene trees in ASTRAL-III (Zhang et al., 2018), a quartet-based consensus method. Gene trees were first generated in PhyML v3.0 (Guindon et al., 2010) under a GTR model then used as input for ASTRAL-III. The topology of the tree generated with this analysis is identical to that using the concatenated dataset (Figures 4.3A and 4.3B). The phylogenies generated from this dataset depicted *D. subquinaria* as a monophyletic species, with sympatric *D. subquinaria* (S_Cal241) as sister to allopatric *D. subquinaria* (S_MT1 and S_SE70). This result is consistent with the hypothesized species relationships based on levels of pre- and postmating reproductive isolation, and are concordant with phylogenetic relationships produced with low-recombining regions such as Y-chromosome loci and, in part, the mitochondrial locus from my analysis in Chapter 2.

TE abundance

One explanation for differences in the fly assembly size is the abundance of transposable elements (TEs). TE proliferation has been associated with genome size evolution across a variety of taxa (Talla et al., 2017; Wong et al., 2019; Zuccolo et al., 2007), including *Drosophila* (Sessegolo, Bulet, & Haudry, 2016). Repeatmasker estimates of TE abundance show that unclassified repetitive elements represent the vast majority of identified elements in all assemblies, followed by DNA elements, LTRs and LINEs, and finally SINEs were the least abundant elements in all assemblies (Figure 4.4A). TEs represent 17.58 – 23.85 percent of the assemblies (including all unclassified elements, Table 4.4). TE abundance did not explain observed differences in assembly size (Figure 4.4B, $R^2 = 0.0174$, $P = 0.78$).

Wolbachia Genome Assemblies

Both *Wolbachia* assemblies were resolved as single, circularized contigs (Table 4.5). These two *wRec* strains are remarkably similar. They are different by only 46 single nucleotides variants and thirteen indels (Tables S4.4 & S4.5). They are primarily distinguishable by two major structural variants, a ~25.6kb indel, and a ~33.3kb inversion (Figures 4.5, 4.6 and 4.7). The 25.6kb region is present in the RW *wRec*, making a total assembly length of 1,209,303 bp. For simplicity, I will refer to the ~25.6kb indel as a deletion in Alb *wRec*. The Alb *wRec* stock assembly has a total length of 1,184,195 bp. Alignment with a published *wRec* genome (Metcalf, Jo, Bordenstein, Jaenike, & Bordenstein, 2014) collected at the approximate time as the Alberta *D. recens* stock and from the opposite end of the species range, indicate that the published *wRec* genome shares both the 25.6 kb deletion and the 33.3 kb inversion with the Alb *wRec* strain (Figures 4.6 and 4.7).

Both the inversion and the deletion are flanked by transposases, suggesting these rearrangements are TE initiated (IS4 family transposases flank inversion, IS5 family transposases flank deletion). Gene content within the inversion is identical, with zero SNVs (All loci located within inversion in Table 4.6). *wMel* and Alb *wRec* share this inversion with respect to RW *wRec*, but is missing an IS4 transposase on one end (WR_ALB_00076, WR_ALB_00111). Several prophage WO related proteins fall directly upstream of this inversion, but this does not fall within prophage WO-A (WD0250 – WO300) or prophage WO-B (WD0581-WD0650) of *wMel*.

Neither RW *wRec* nor Alb *wRec* possess the full complement of prophage WO-B genes. RW *wRec* is missing WD0581-WD0596, WD607-WD0625, and WD0633-WD0635 from *wMel*, which primarily code for phage head proteins. Alb *wRec* is missing the additional 30 loci listed

in Table 4.7. One annotated region in RW *wRec*, WR_RW_00580, is a hypothetical protein unique to the RW *wRec* strain, and falls within this same WO-B prophage region.

Of the 46 single nucleotide variants, 22 fall in intergenic regions, 11 fall within coding regions but are synonymous, and 13 are nonsynonymous mutations (Table S4.4). Of the nonsynonymous mutations, three are in hypothetical proteins, and two are Proline \leftrightarrow Leucine mutations in transposases (WR_RW_00026, WR_ALB_00026 and WR_RW_01042, WR_ALB_10012, respectively). Eight fall in non-prophage bacterial genes (Table S4.4). Importantly, *cifA* and *cifB*, genes responsible for the induction of cytoplasmic incompatibility (Beckmann, Ronau, & Hochstrasser, 2017; LePage et al., 2017), and *wmk*, the gene proposed to be responsible for the induction of male-killing (Perlmutter et al., 2019) are identical between RW and Alb *wRec* strains.

Wolbachia Phylogeny

Concatenated sequence alignments for *Wolbachia* MLST loci produced a 2,600 bp alignment with 337 variable sites, and 294 were parsimony informative. This analysis showed a clearly resolved groupings of *Wolbachia* clades (i.e. supergroups) with *wRec* and other supergroup A *Wolbachia* *wMel*, *wAu*, and *wInn* (Figure 4.8). This is consistent with previous analysis of *Wolbachia* strain relationships to *wRec* (Metcalf et al., 2014). It is clear from this analysis that the two *Wolbachia* strains from *wRec* are sister taxa, and the inclusion of the published *wRec* genome produces a polytomy. These results suggest that the *wRec* variants likely arose after infecting *D. recens*. In fact, across MLST loci these two *wRec* strains are identical.

Wolbachia strain prevalence in *D. recens*

PCR assay for strain prevalence indicated that the *wRec* variants were highly polymorphic in *D. recens* populations (Table 4.8). The numbers presented here include the two whose sequence data was used for genome assembly. The published *wRec* assembly (Metcalf et al., 2014) is infected with the Alb *wRec* variant (not tested with PCR assay, see Results above).

Discussion

Here I present seven de novo genome assemblies sampling the three members of the *D. subquinaria* species complex and *D. quinaria* as an outgroup species. Compared to assemblies in Kim et. al (2021) these assemblies fall within expected ranges for genome contiguity (Figure S4.2). Additionally, I present two distinct *Wolbachia* variants infecting the two *D. recens* samples. I find substantial variation in fly genome size that cannot be explained by transposable element proliferation, as has been reported for other *Drosophila* species (Sessegolo et al., 2016). Preliminary analyses suggest that the majority of size variation observed in the fly genome assemblies are artefacts of the assembly process, and do not reflect biological variation. First, high heterozygosity among the species sampled and inability for direct inbreeding result in highly heterozygous draft assemblies. To remove diploid contigs, I implemented the haplotig_purging pipeline (Roach et al., 2018). Initial overview of this step reveals substantial levels of high heterozygosity across samples, as expected, but also invariably removes genetic content from the final assemblies. This is particularly clear in the Q_QTP *D. quinaria* assembly, where BUSCO scores from before and after the haplotig purging step went from 99.0% to 95.8%, a difference of 105 single copy orthologs. Additionally, 110 of the total 129 BUSCO orthologs missing from the final Q_QTP assembly fall within a 25 Mb region of chromosome 2L

in *D. innubila* (the most closely related species for which there is a published, chromosome-level assembly, Table S4.6) (Hill et al., 2019). Second, a preliminary analysis for gene duplication rates also suggest high levels of heterozygosity remains in the assemblies as they currently stand. Analysis of Braker annotated genes across all assemblies in Orthofinder (Emms & Kelly, 2019) predicted the number of gene duplications from 74 between S_SE70 and S_MT1, to 4,153 between clades (*D. transversa*: T_T6, (*D. subquinaria*: S_Cal241, (S_SE70, S_MT1))) and (*D. recens*: R_Alb, R_RW) (Figure S4.3). While perhaps not impossible if major chromosomal blocks have been duplicated, these rates of gene duplication far exceed known rates for other *Drosophila* species with similar divergence times (Osada & Innan, 2008). It is worth noting, however, that BUSCO duplication rates for the final genome assemblies were within acceptable ranges of published haploid genomes (Feron & Waterhouse, 2022), at ~1-3% (Table 4.2).

Avoiding the issues discussed above, I reassessed phylogenetic history of these species using 100 randomly sampled single copy orthologs from BUSCO output. The phylogenies generated here are consistent with levels of reproductive isolation and phylogenetic topologies resolved for regions impervious to gene flow and low-recombining regions (see Chapter Two). They are, however, inconsistent in their topological placement of sympatric *D. subquinaria* analyzed with autosomal and X-chromosome loci in my analysis from Chapter Two. The datasets utilized for analyses differ in several important ways. First, the analysis presented here has only one sample of sympatric *D. subquinaria* (Cal241), whereas in the analyses in earlier analyses there are ten samples. It is possible that the sample used for genome sequencing happens to be less admixed than sympatric *D. subquinaria* are on average. Variation in levels of admixture among sympatric *D. subquinaria* is supported by quartet sampling in Chapter 2. In that analysis randomly samples individual taxa from sympatric *D. subquinaria* had different

topological relationships to *D. recens*, suggesting that taxa sampling from this clade can have drastic effects on the topological relationships resolved (Chapter 2, figures 2.4A, 2.4C). Second, the dataset presented here includes approximately 3.5X more loci than the analyses in Chapter 2. The dataset presented here had 15,542 variable sites, whereas the dataset from Chapter 2 had 776 variable sites. This increase in genetic information likely increased ability to strongly resolve species relationships, despite the continued presence of gene tree – species tree discordance observed here. Re-sequencing data with population sampling from sympatric *D. subquinaria* will allow for these species relationships to be disentangled from the shortcomings of the two datasets discussed here.

Wolbachia assemblies reveal that the two *D. recens* samples used for genome sequencing harbor distinct *Wolbachia* variants. One of these variants, Alb *w*Rec, generates cytoplasmic incompatibility in both *D. recens* and when introgressed into *D. subquinaria* (See Chapter 3). Preliminary evidence shows that the other variant, RW *w*Rec, produces cytoplasmic incompatibility in *D. recens*, but when introgressed into the same *D. subquinaria* genotypes from Chapter 3, produces male-killing (unpublished data, K. Dyer). Controlling for the *D. subquinaria* genetic background suggests that the different *Wolbachia*-induced phenotypes are a consequence of differences in the *Wolbachia* strain rather than host-evolved suppressors. The genes underlying both *Wolbachia*-induced phenotypes are well described.

Cytoplasmic incompatibility is caused by two prophage genes, *cifA* and *cifB* (Beckmann et al., 2017; LePage et al., 2017). The most accepted model of how the function of these two genes produce the cytoplasmic incompatibility phenotype is a “Two-by-one model” (Shropshire & Bordenstein, 2019), where *cifA* and *cifB* (toxic effect from *cifB*) expressed during spermatogenesis is rescued by expression of *cifA* in the female. This model suggests that

sequence similarity in *cifA* and *cifB* should predict cross compatibility between strains (Namias, Sicard, Weill, & Charlat, 2022). If this model is correct, then the *Wolbachia* variants in *D. recens* should be perfectly cross compatible, as these two genes (WR_RW_00548, WR_RW_00549 and WR_RW_ALB_00549, WR_ALB_00550) are identical. This also suggests that the reduced severity of cytoplasmic compatibility in *D. recens* ((Shoemaker et al., 1999), see Chapter 3) is more likely due to host suppression of cytoplasmic incompatibility. Modifier alleles may also reduce the severity of cytoplasmic incompatibility, yet there are few differentiating these strains and none of them are found genes known to affect CI strength.

Male-killing, another common *Wolbachia*-induced phenotype, is putatively caused by the *wmk* gene of prophage WO-B (Perlmutter et al., 2019). While the mechanism of male-killing's induction is less well understood, transgenic expression of this gene induces male-killing in *D. melanogaster*, and this gene is shared among all *Wolbachia* strains that can induce this phenotype, to the exclusion of other candidate *Wolbachia* genes (Perlmutter et al., 2019; Perlmutter, Meyers, & Bordenstein, 2020). The results presented here conflict, in part, with previous research. Similar to cytoplasmic incompatibility genes *cifA* and *cifB*, the *wmk* homologs in RW *wRec* and Alb *wRec* are identical (WR_RW_00543, WR_ALB_00544, respectively) yet only RW *wRec* induces male-killing when introgressed into *D. subquinaria* (data not shown for RW *wRec* male-killing; Alb *wRec* does not cause male-killing, Chapter 3, figure 3.2). This suggests an additional factor is necessary for the induction of male-killing in *D. subquinaria*. It seems likely that any additional factors involved in the induction of male-killing are prophage-related, as all reproductive manipulations with understood mechanisms are prophage-related, and the majority of differences in gene content between these two *wRec* strains are part of the 25.6kb deletion of prophage WO-B components in Alb *wRec*, the non-male-killing strain.

The genomic content alterations in RW *w*Rec and Alb *w*Rec as compared to *w*Mel depict an initial reduction of prophage WO-B, emergence of an inversion that was segregating in *w*Rec, followed by a further reduction in prophage WO-B specific to the Alb *w*Rec strain (Figure 4.9). This model explains the WO-B gene content shared between *w*Mel and RW *w*Rec, and the shared orientation of the inverted region between *w*Mel and Alb *w*Rec. It is also supported by preliminary PCR screening for these two *Wolbachia* strains in *D. recens* populations. Ongoing screening has thus far revealed that these two strains are currently segregating in *D. recens* populations. However, this initial analysis is limited in that it is likely not capturing additional variation among *w*Rec strains, and sample sizes are low (Table 4.8). The published *w*Rec genome was generated using eastern (Pittsford, New York) *D. recens* collections (Metcalf et al., 2014), and shares the deletion and orientation of the inversion with Alb *w*Rec (Figures 4.6 and 4.7, not included in Table 4.6). Together this suggests that *D. recens* may be undergoing a *Wolbachia* sweep of the Alb *w*Rec strain that has not yet fixed in *D. recens* populations. Further sampling of *D. recens* populations will provide a clearer picture of current population dynamics of segregating *Wolbachia* variants.

Works Cited

Baldo, L., Dunning Hotopp, J. C., Jolley, K. A., Bordenstein, S. R., Biber, S. A., Choudhury, R. R., . . . Werren, J. H. (2006). Multilocus sequence typing system for the endosymbiont *Wolbachia pipientis*. *Appl Environ Microbiol*, 72(11), 7098-7110.
doi:10.1128/AEM.00731-06

- Beckmann, J. F., Ronau, J. A., & Hochstrasser, M. (2017). A Wolbachia deubiquitylating enzyme induces cytoplasmic incompatibility. *Nat Microbiol*, 2, 17007.
doi:10.1038/nmicrobiol.2017.7
- Cao, M. D., Nguyen, S. H., Ganesamoorthy, D., Elliott, A. G., Cooper, M. A., & Coin, L. J. (2017). Scaffolding and completing genome assemblies in real-time with nanopore sequencing. *Nat Commun*, 8, 14515. doi:10.1038/ncomms14515
- Chen, S., Zhou, Y., Chen, Y., & Gu, J. (2018). fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*, 34(17), i884-i890. doi:10.1093/bioinformatics/bty560
- Darling, A. C., Mau, B., Blattner, F. R., & Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res*, 14(7), 1394-1403.
doi:10.1101/gr.2289704
- Emms, D. M., & Kelly, S. (2019). OrthoFinder: phylogenetic orthology inference for comparative genomics. *Genome Biol*, 20(1), 238. doi:10.1186/s13059-019-1832-y
- Feron, R., & Waterhouse, R. M. (2022). Assessing species coverage and assembly quality of rapidly accumulating sequenced genomes. *Gigascience*, 11.
doi:10.1093/gigascience/giac006
- Flynn JM, H. R., Goubert C, Rosen J, Clark AG, Feschotte C, Smit AF. (2020). RepeatModeler2 for automated genomic discovery of transposable element families. *Proc Natl Acad Sci U S A*, 117(17), 9451-9457. doi:10.1073/pnas.1921046117
- Freire, B., Ladra, S., & Parama, J. R. (2021). Memory-Efficient Assembly using Flye. *IEEE/ACM Trans Comput Biol Bioinform*, PP. doi:10.1109/TCBB.2021.3108843

- Guindon, S., Dufayard, J. F., Lefort, V., Anisimova, M., Hordijk, W., & Gascuel, O. (2010). New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol*, *59*(3), 307-321. doi:10.1093/sysbio/syq010
- Gurevich, A., Saveliev, V., Vyahhi, N., & Tesler, G. (2013). QUASt: quality assessment tool for genome assemblies. *Bioinformatics*, *29*(8), 1072-1075. doi:10.1093/bioinformatics/btt086
- Hill, T., Koseva, B. S., & Unckless, R. L. (2019). The Genome of *Drosophila innubila* Reveals Lineage-Specific Patterns of Selection in Immune Genes. *Mol Biol Evol*, *36*(7), 1405-1417. doi:10.1093/molbev/msz059
- Hoff, K. J., Lomsadze, A., Borodovsky, M., & Stanke, M. (2019). Whole-Genome Annotation with BRAKER. *Methods Mol Biol*, *1962*, 65-95. doi:10.1007/978-1-4939-9173-0_5
- Jaenike, J. (2007). Spontaneous emergence of a new *Wolbachia* phenotype. *Evolution*, *61*(9), 2244-2252. doi:10.1111/j.1558-5646.2007.00180.x
- Kim, B. Y., Wang, J. R., Miller, D. E., Barmina, O., Delaney, E., Thompson, A., . . . Petrov, D. A. (2021). Highly contiguous assemblies of 101 drosophilid genomes. *Elife*, *10*. doi:10.7554/eLife.66405
- Kim, D., Paggi, J. M., Park, C., Bennett, C., & Salzberg, S. L. (2019). Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. *Nat Biotechnol*, *37*(8), 907-915. doi:10.1038/s41587-019-0201-4
- Koressaar, T., & Remm, M. (2007). Enhancements and modifications of primer design program Primer3. *Bioinformatics*, *23*(10), 1289-1291. doi:10.1093/bioinformatics/btm091
- Langmead, B., & Salzberg, S. L. (2012). Fast gapped-read alignment with Bowtie 2. *Nat Methods*, *9*(4), 357-359. doi:10.1038/nmeth.1923

- LePage, D. P., Metcalf, J. A., Bordenstein, S. R., On, J., Perlmutter, J. I., Shropshire, J. D., . . .
- Bordenstein, S. R. (2017). Prophage WO genes recapitulate and enhance Wolbachia-induced cytoplasmic incompatibility. *Nature*, *543*(7644), 243-247.
doi:10.1038/nature21391
- Li, H. (2018). Minimap2: pairwise alignment for nucleotide sequences. *Bioinformatics*, *34*(18), 3094-3100. doi:10.1093/bioinformatics/bty191
- Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., . . . Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Res.*
doi:10.1093/nar/gkac240
- Manni, M., Berkeley, M. R., Seppey, M., & Zdobnov, E. M. (2021). BUSCO: Assessing Genomic Data Quality and Beyond. *Curr Protoc*, *1*(12), e323. doi:10.1002/cpz1.323
- Metcalf, J. A., Jo, M., Bordenstein, S. R., Jaenike, J., & Bordenstein, S. R. (2014). Recent genome reduction of Wolbachia in *Drosophila recens* targets phage WO and narrows candidates for reproductive parasitism. *PeerJ*, *2*, e529. doi:10.7717/peerj.529
- Namias, A., Sicard, M., Weill, M., & Charlat, S. (2022). From Wolbachia genomics to phenotype: molecular models of cytoplasmic incompatibility must account for the multiplicity of compatibility types. *Curr Opin Insect Sci*, *49*, 78-84.
doi:10.1016/j.cois.2021.12.005
- Osada, N., & Innan, H. (2008). Duplication and gene conversion in the *Drosophila melanogaster* genome. *PLoS Genet*, *4*(12), e1000305. doi:10.1371/journal.pgen.1000305
- Perlmutter, J. I., Bordenstein, S. R., Unckless, R. L., LePage, D. P., Metcalf, J. A., Hill, T., . . .
- Bordenstein, S. R. (2019). The phage gene *wmk* is a candidate for male killing by a

- bacterial endosymbiont. *PLoS Pathog*, 15(9), e1007936.
doi:10.1371/journal.ppat.1007936
- Perlmutter, J. I., Meyers, J. E., & Bordenstein, S. R. (2020). Transgenic Testing Does Not Support a Role for Additional Candidate Genes in Wolbachia Male Killing or Cytoplasmic Incompatibility. *mSystems*, 5(1). doi:10.1128/mSystems.00658-19
- Roach, M. J., Schmidt, S. A., & Borneman, A. R. (2018). Purge Haplotigs: allelic contig reassignment for third-gen diploid genome assemblies. *BMC Bioinformatics*, 19(1), 460. doi:10.1186/s12859-018-2485-7
- Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, 30(14), 2068-2069. doi:10.1093/bioinformatics/btu153
- Sessegolo, C., Bulet, N., & Haudry, A. (2016). Strong phylogenetic inertia on genome size and transposable element content among 26 species of flies. *Biol Lett*, 12(8). doi:10.1098/rsbl.2016.0407
- Shoemaker, D. D., Katju, V., & Jaenike, J. (1999). Wolbachia and the Evolution of Reproductive Isolation between *Drosophila Recens* and *Drosophila Subquinaria*. *Evolution*, 53(4), 1157-1164. doi:10.1111/j.1558-5646.1999.tb04529.x
- Shropshire, J. D., & Bordenstein, S. R. (2019). Two-By-One model of cytoplasmic incompatibility: Synthetic recapitulation by transgenic expression of cifA and cifB in *Drosophila*. *PLoS Genet*, 15(6), e1008221. doi:10.1371/journal.pgen.1008221
- Smit, A., Hubley, R & Green, P. (2015). RepeatMasker Open-4.0. Retrieved from <http://www.repeatmasker.org>

- Stamatakis, A., Ludwig, T., & Meier, H. (2005). RAxML-III: a fast program for maximum likelihood-based inference of large phylogenetic trees. *Bioinformatics*, *21*(4), 456-463. doi:10.1093/bioinformatics/bti191
- Talla, V., Suh, A., Kalsoom, F., Dinca, V., Vila, R., Friberg, M., . . . Backstrom, N. (2017). Rapid Increase in Genome Size as a Consequence of Transposable Element Hyperactivity in Wood-White (Leptidea) Butterflies. *Genome Biol Evol*, *9*(10), 2491-2505. doi:10.1093/gbe/evx163
- Untergasser, A., Cutcutache, I., Koressaar, T., Ye, J., Faircloth, B. C., Remm, M., & Rozen, S. G. (2012). Primer3--new capabilities and interfaces. *Nucleic Acids Res*, *40*(15), e115. doi:10.1093/nar/gks596
- Walker, B. J., Abeel, T., Shea, T., Priest, M., Abouelliel, A., Sakthikumar, S., . . . Earl, A. M. (2014). Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*, *9*(11), e112963. doi:10.1371/journal.pone.0112963
- Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial genome assemblies from short and long sequencing reads. *PLoS Comput Biol*, *13*(6), e1005595. doi:10.1371/journal.pcbi.1005595
- Wong, W. Y., Simakov, O., Bridge, D. M., Cartwright, P., Bellantuono, A. J., Kuhn, A., . . . Martinez, D. E. (2019). Expansion of a single transposable element family is associated with genome-size increase and radiation in the genus *Hydra*. *Proc Natl Acad Sci U S A*, *116*(46), 22915-22917. doi:10.1073/pnas.1910106116

Zhang, C., Rabiee, M., Sayyari, E., & Mirarab, S. (2018). ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. *BMC Bioinformatics*, *19*(Suppl 6), 153. doi:10.1186/s12859-018-2129-y

Zuccolo, A., Sebastian, A., Talag, J., Yu, Y., Kim, H., Collura, K., . . . Wing, R. A. (2007). Transposable element distribution, abundance and role in genome size variation in the genus *Oryza*. *BMC Evol Biol*, *7*, 152. doi:10.1186/1471-2148-7-152

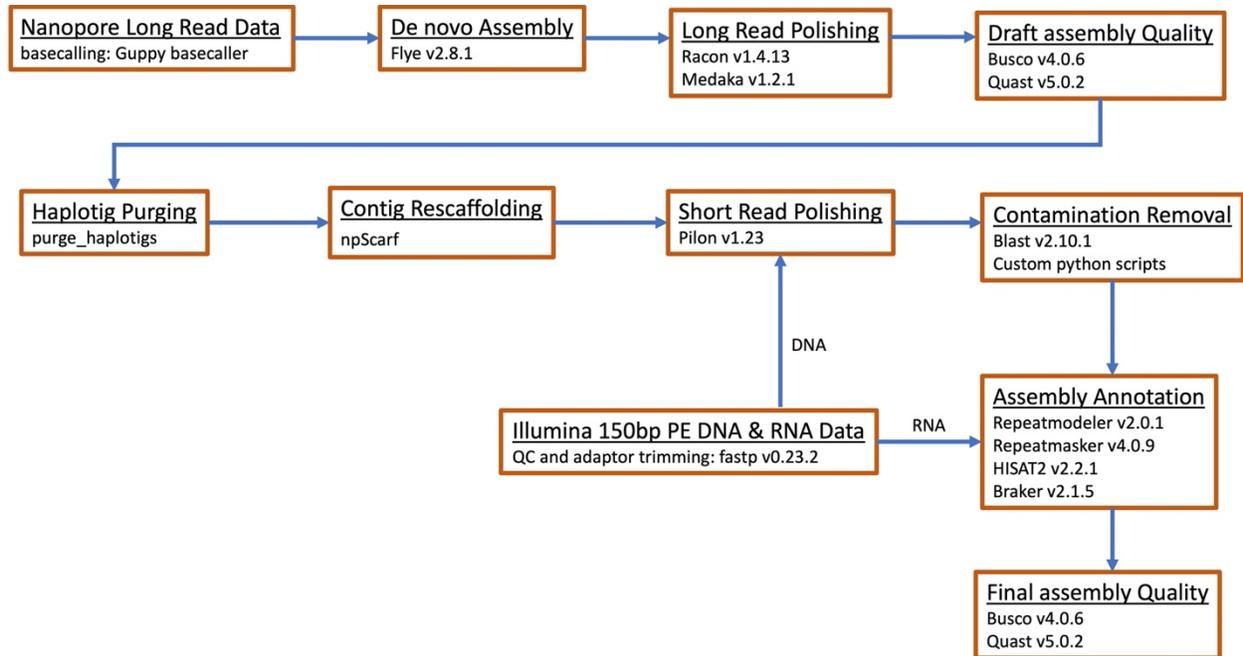


Figure 4.1: Assembly pipeline for fly genomes. Programs implemented (with versions where applicable) at each step are included in orange boxes.

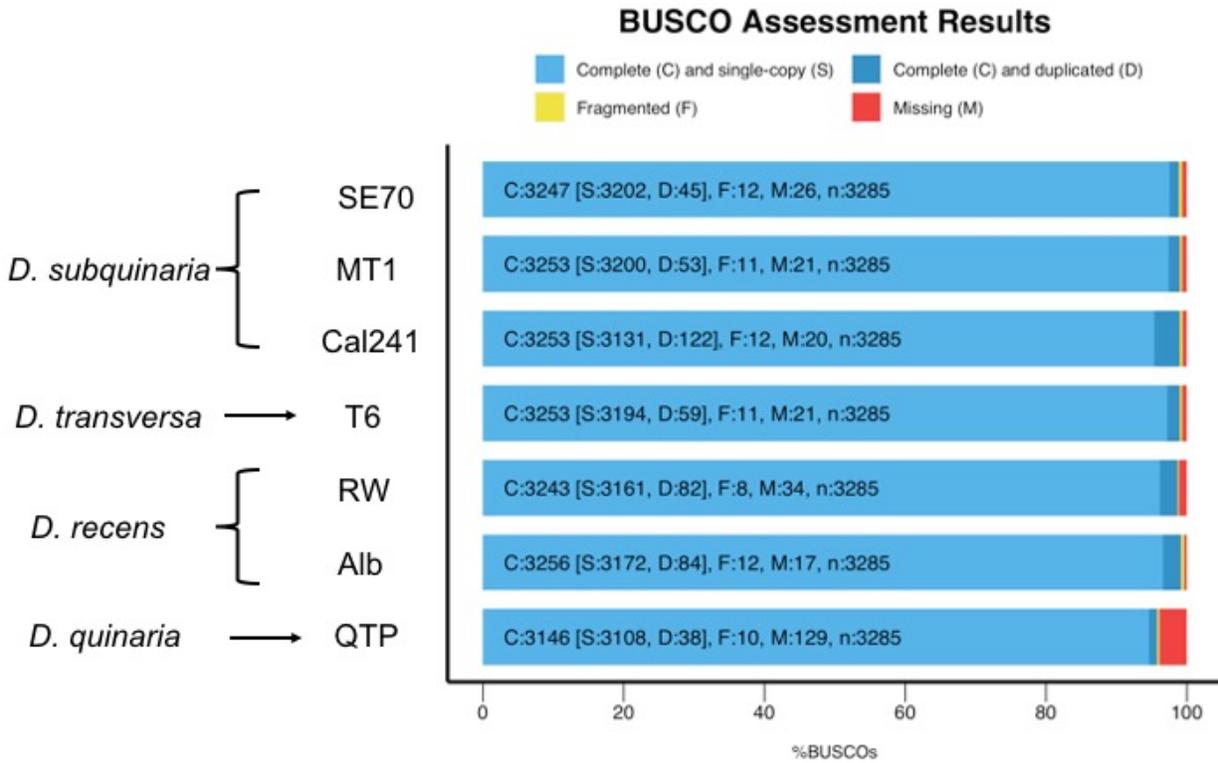


Figure 4.2: BUSCO assessment for all fly genome assemblies. The Y axis indicates the genome assembly. Species designation for sample abbreviations are *D. quinaria*: QTP, *D. transversa*: T6, *D. recens*: Alb, RW, and *D. subquinaria*: Cal241, SE70, and MT1.

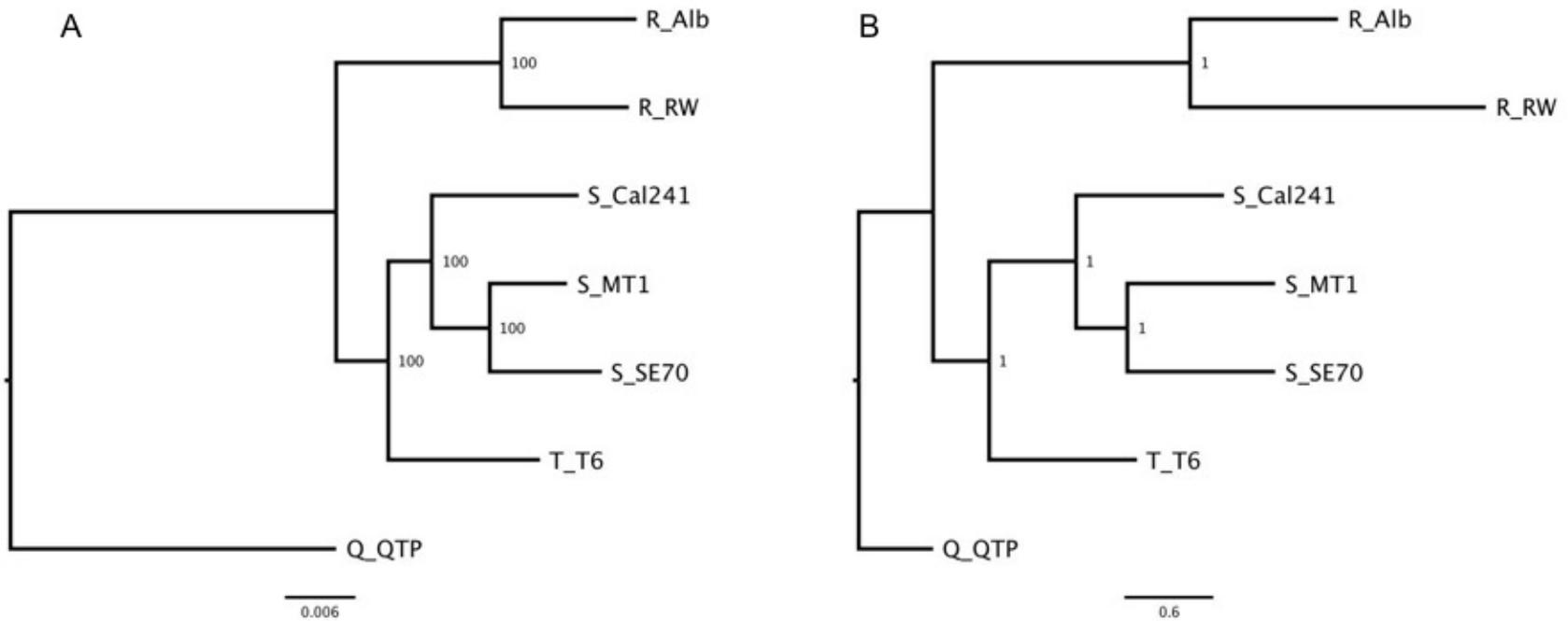


Figure 4.3: Host phylogenies from 100 randomly selected single copy orthologs. Sample names are prepended by species abbreviations: R = *D. recens*, S = *D. subquinaria*, T = *D. transversa*, Q = *D. quinaria*. A: Concatenated phylogeny with 500 bootstrap replicates indicated at nodes. Bootstrap values indicated at branch nodes. B: Astral consensus tree from 100 gene trees. Posterior probabilities are indicated at nodes. Scale bars indicate branch length in average number of substitutions per site (A) and coalescent units (B).

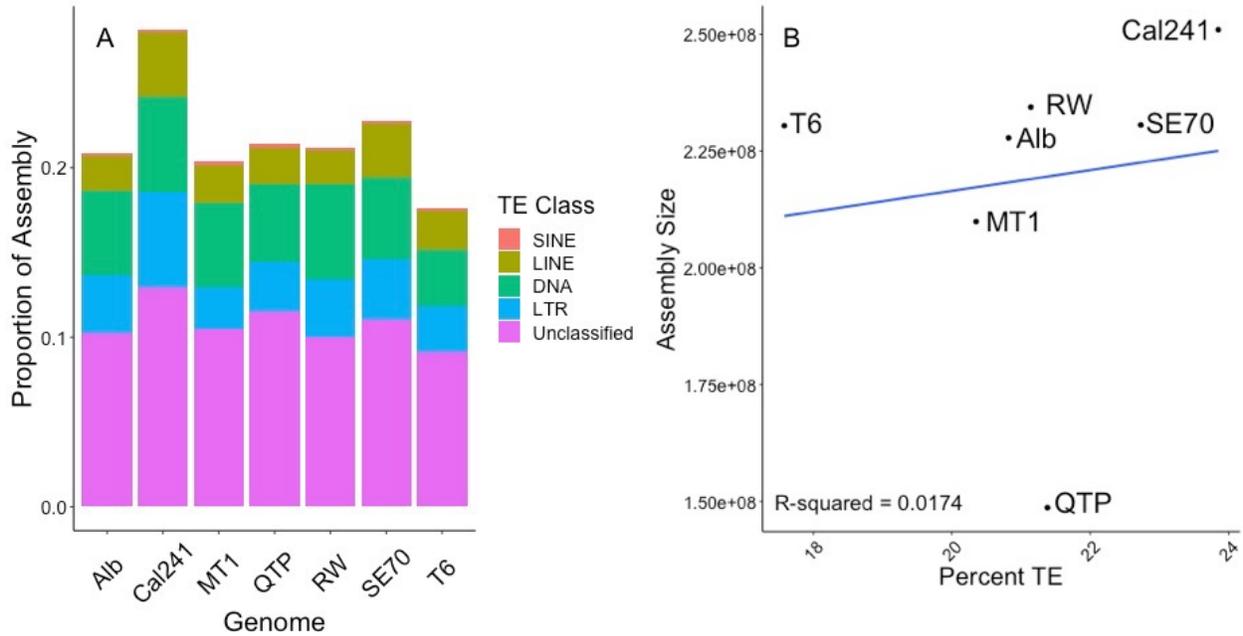


Figure 4.4: Transposable element abundance in all assemblies. A: Bar plots of all major classes of TEs found in genome assemblies, expressed as proportion of genome. B: Linear regression of the percent of the genome comprised of TEs plotted against the assembly size. Linear regression line is in blue and sample names are indicated near its data point.

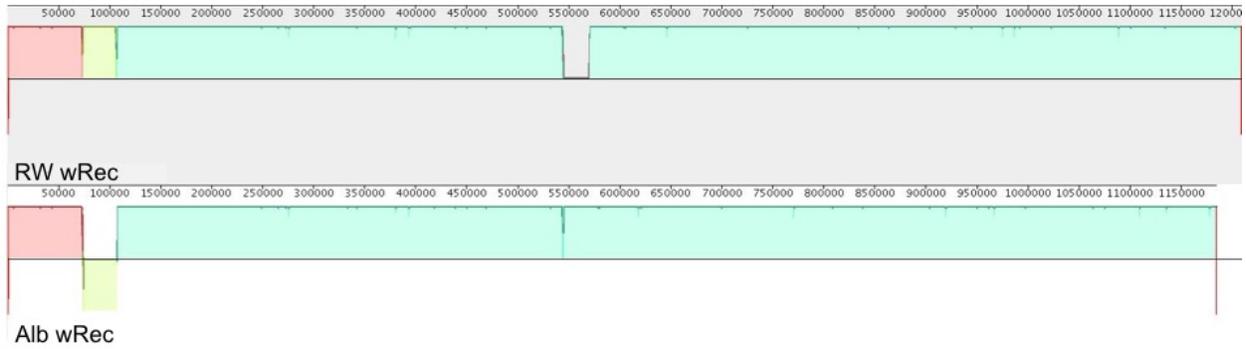


Figure 4.5: *wRec* genome alignment from RW and Alb *D. recens* stocks. Colored blocks represent blocks of homology shared between the two genomes. Above vs. below the center line depicts the sequence orientation. Empty space represents regions lacking homology between genomes. Vertical red lines are contig end points. Here the yellow homology block indicates an inversion between these *Wolbachia* strains and the empty space in RW *wRec* depicts the region deleted from Alb *wRec*.

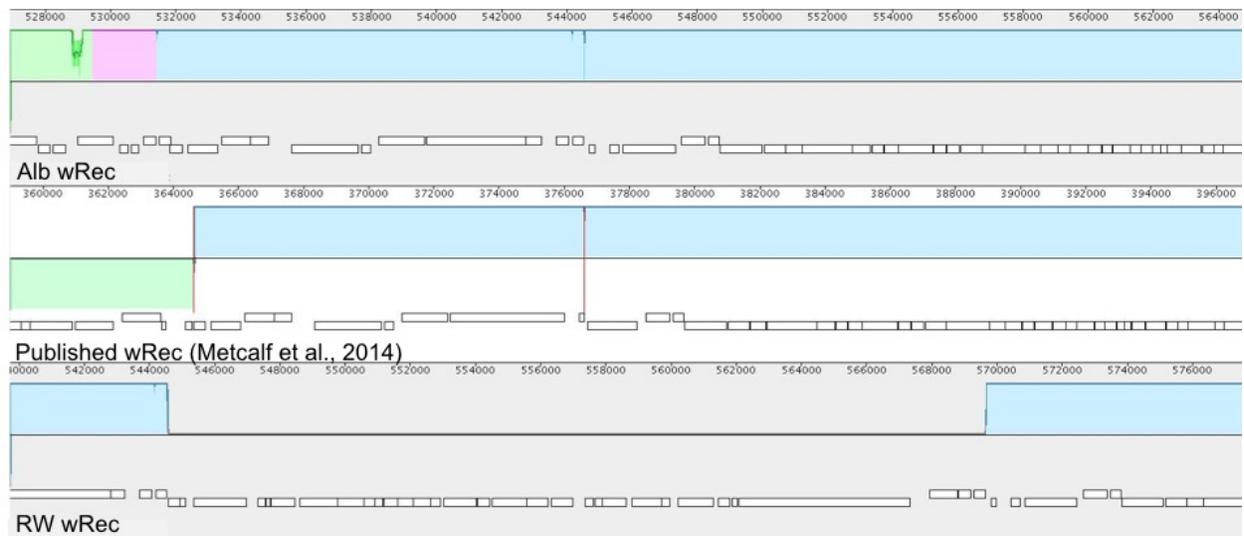


Figure 4.6: Zoomed in view of aligned regions with deletion in Alb *wRec* and the published *wRec*. Different colors depict blocks of homology between *wRec* strains. Empty space represents region present in RW *wRec* and deleted from Alb *wRec* and the published *wRec* genome. Vertical red lines in published *wRec* genome represent contig end points.

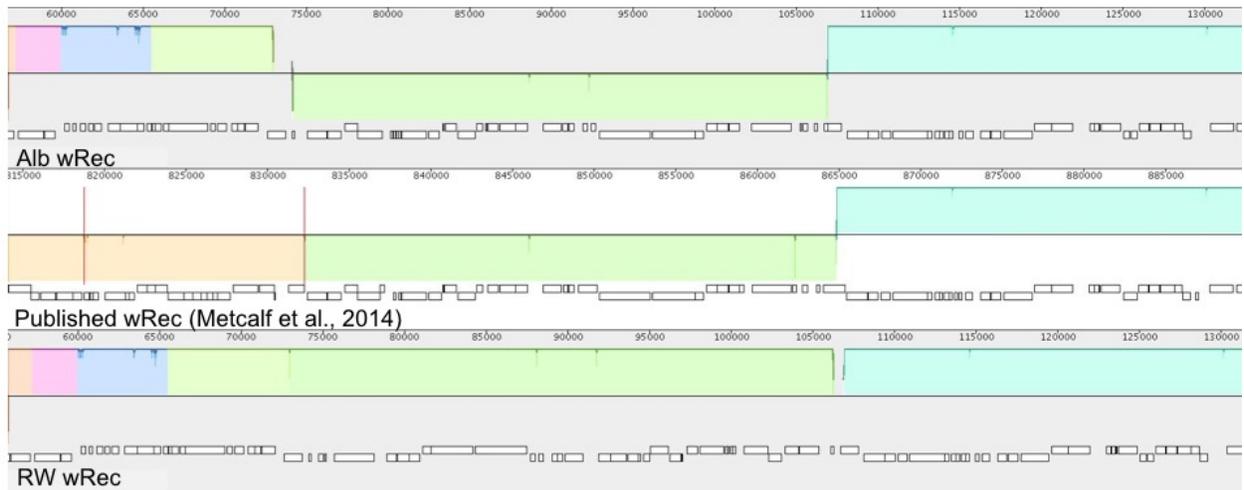


Figure 4.7: Inverted region of *wRec* variants. Green region represents inverted homology block between *wRec* variants. Above and below the center line for each assembly indicate the orientation of sequence with respect to other strains. Vertical red lines in published *wRec* represent contig break points. Regions without coloring represent regions without homology with other variants. Here, these regions are both over an annotation for an IS4 family transposase, and so the lack of homology is more likely an alignment issue than an orphan copy of this gene present in both RW and Alb *wRec*.

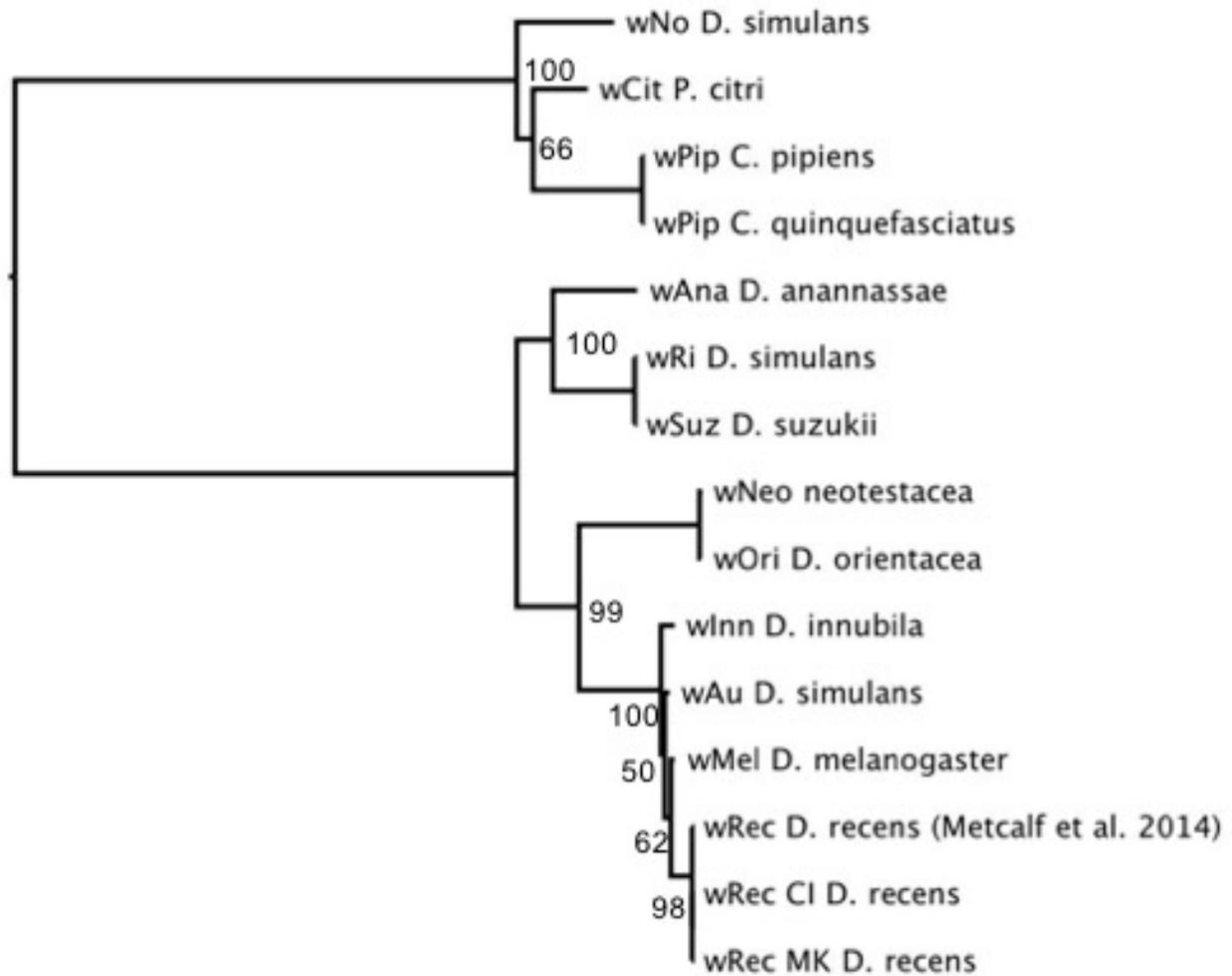


Figure 4.8: *Wolbachia* phylogeny generated from five MLST loci and *wsp*. Tip labels are *Wolbachia* strain names with their host species name appended. Bootstrap values from 100 bootstrap replicates are indicated at nodes. wRec names refer to phenotype induced in *D. subquinaria*. “wRec CI” is the Alb wRec strain and “wRec MK” is the RW wRec strain.

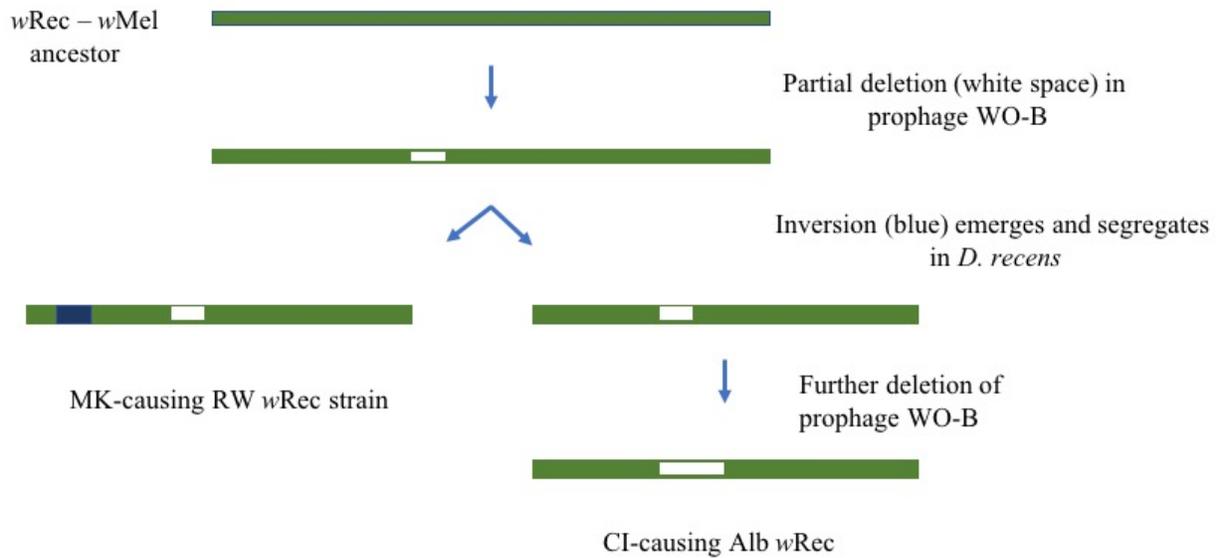


Figure 4.9: Model of *Wolbachia* genome evolution from ancestor with $wMel$. This model focuses on structural variation observed between $wRec$ strains, and so does not incorporate additional structural changes after split with $wMel$. Chromosomes are not drawn to scale and placement of structural changes are approximate.

Table 4.1: Fly samples used for whole genome sequencing. Sympatric populations refers to where *D. subquinaria* and *D. recens* are sympatric with each other. Collection refers to the stock line name.

Species	Population	Collection	Location	Year	Additional Notes
<i>D. subquinaria</i>	Allopatric	SE70	Seattle,	2010	
	Coastal		Washington		
	Allopatric Inland	MT1	Missoula, Montana	2011	
	Sympatric	Cal241	Calgary, Alberta, CA	2019	
<i>D. recens</i>	Sympatric	Alb	Alberta, Canada	2009- 2011	<i>Wolbachia</i> infected, Mixed stock
	Allopatric	RW	New York	2001	<i>Wolbachia</i> infected White Eye stock
<i>D. transversa</i>	NA	T6	Uppsala, Sweden	2013	
<i>D. quinaria</i>	NA	QTP	Pittsford, New York	2004	

Table 4.2: Assembly statistics for fly genome assemblies.

	<i>D. recens</i> Alb	<i>D. recens</i> RW	<i>D. subquinaria</i> Cal241	<i>D. subquinaria</i> SE70	<i>D. subquinaria</i> MT1	<i>D. transversa</i> T6	<i>D. quinaria</i> QTP
Complete BUSCOs I	3,256 (99.2%)	3,243 (98.7%)	3,253 (99.0%)	3,247 (98.7%)	3,253 (99.0%)	3,253 (99.0%)	3,146 (95.8%)
Complete and single-copy BUSCOs (S)	3,172 (96.6%)	3,161 (96.2%)	3,131 (95.3%)	3,202 (97.5%)	3,200 (97.4%)	3,194 (97.2%)	3,108 (94.6%)
Complete and duplicated BUSCOs (D)	84 (2.6%)	82 (2.5%)	122 (3.7%)	45 (1.4%)	53 (1.6%)	59 (1.8%)	38 (1.2%)
Fragmented BUSCOs (F)	12 (0.4)	8 (0.2%)	12 (0.4%)	12 (0.4%)	11 (0.3%)	11 (0.3%)	10 (0.3%)
Missing BUSCOs (M)	17 (0.4)	34 (1.1%)	20 (0.6%)	26 (0.7%)	21 (0.7%)	21 (0.7%)	129 (3.9%)
Total BUSCO groups searched	3,285	3,285	3,285	3,285	3,285	3,285	3,285
# contigs (>= 0 bp)	296	370	483	312	524	184	141
# contigs (>= 25000 bp)	296	370	483	312	520	184	141
# contigs (>= 1000 bp)	230	296	391	274	359	145	115
# contigs (>= 5000 bp)	168	210	319	214	290	101	99
# contigs (>= 10000 bp)	137	160	277	168	258	71	89
# contigs (>= 50000 bp)	116	133	237	146	216	63	77
Total length (>= 0 bp)	227,823,419	234,403,209	250,961,281	230,599,954	209,903,969	230,418,477	148,690,375
Total length (>= 1000 bp)	227,823,419	234,403,209	250,961,281	230,599,954	209,901,451	230,418,477	148,690,375
Total length (>= 5000 bp)	227,587,030	234,125,918	250,647,068	230,450,447	209,400,759	230,277,203	148,599,117
Total length (>= 10000 bp)	227,143,982	233,544,855	250,171,600	230,008,033	208,930,508	229,958,787	148,494,676
Total length (>= 25000 bp)	226,690,083	232,777,302	249,541,919	229,342,366	208,377,221	229,496,850	148,316,689
Total length (>= 50000 bp)	225,959,907	231,744,821	248,129,117	228,491,273	206,822,489	229,200,844	147,893,019
Largest contig	17,700,468	20,780,488	10,374,723	38,486,798	20,180,598	38,347,010	37,761,907
Total length	227,823,419	234,403,209	250,961,281	230,599,954	209,903,681	230,418,477	148,690,375
GC (%)	40.43	40.33	38.91	39.38	38.91	41.04	37.3
N50	5,462,940	3,525,288	2,301,551	4,492,352	2,643,564	13,123,973	20,712,610
N75	1,819,282	1,643,056	941,491	1,865,734	1,067,760	4,834,794	2,386,792
L50	12	16	28	11	21	6	3
L75	32	41	76	29	52	14	10
# N's per 100 kbp	0	0	0	0	0	0	0
Mean Depth of Coverage	39	34	67	45	86	44	91

Table 4.3: Gene count predictions from Braker annotations.

	<i>D. recens</i> Alb	<i>D. recens</i> RW	<i>D. subquinaria</i> Cal241	<i>D.</i> <i>subquinaria</i> SE70	<i>D.</i> <i>subquinaria</i> MT1	<i>D.</i> <i>transversa</i> T6	<i>D.</i> <i>quinaria</i> QTP
Gene count:	38,622	41,474	50,139	38,233	27,427	46,259	15,356
Single-exon genes:	18,476	20,576	25,648	17,026	10,723	24,471	3,647
Multi-exon genes	20,146	20,898	24,491	21,207	16,704	21,788	11,709
Introns per gene:	1.51	1.49	1.35	1.62	2	1.33	2.63
Introns per multi-exon gene:	2.89	2.96	2.77	2.92	3.29	2.81	3.45
Genes fully supported by external evidence:	6,800 (17.61%)	7,531 (18.16%)	7,268 (14.5%)	7,972 (20.85%)	7,570 (27.6%)	7,942 (17.17%)	5,264 (34.28%)
Genes partially supported by external evidence:	10,360 (26.82%)	11,565 (27.88%)	11,297 (22.53%)	11,785 (30.82%)	11,327 (41.3%)	11,238 (24.29%)	8,078 (52.6%)
Genes unsupported by any external evidence:	28,262 (73.18%)	29,909 (72.12%)	38,842 (77.47%)	26,448 (69.18%)	16,100 (58.7%)	35,021 (75.71%)	7,278 (47.4%)
Complete genes:	38,535 (99.77%)	41,380 (99.77%)	49,570 (98.87%)	38,148 (99.78%)	27,295 (99.52%)	46,189 (99.85%)	15,321 (99.77%)
Partial genes:	87 (0.23%)	94 (0.23%)	569 (1.13%)	85 (0.22%)	132 (0.48%)	70 (0.15%)	35 (0.23%)

Table 4.4: Total counts for all TEs identified in genome assemblies by Repeatmasker. Percent assembly is the total length in base pairs of all identified TEs, including unclassified TEs, divided by the total assembly length. Species abbreviations are subq = *D. subquiniaria*, recens = *D. recens*, transversa = *D. transversa*, and quinaria = *D. quinaria*.

Species	Collection	nSINE	nLINE	nLTR	nDNA	nUnclassified	Percent Assembly
subq	MT1	1,145	11,171	8,158	34,263	103,334	20.35
subq	SE70	729	12,566	10,717	33,619	111,825	22.73
subq	Cal241	1,292	15,599	17,436	43,148	131,257	23.85
recens	Alb	1,006	8,365	12,830	35,286	109,717	20.82
recens	RW	1,038	7,755	16,101	41,391	103,831	21.14
trans- versa	T6	1,061	13,839	9,207	26,227	104,097	17.58
quinaria	QTP	981	8,402	6,813	28,935	83,103	21.38

Table 4.5: Assembly statistics for *wRec* genomes.

Sample	Alb <i>wRec</i>	RW <i>wRec</i>
ONT reads mapped	33,833	16,898
Illumina reads mapped	172,236 (1.13%)	406,907 (0.6%)
Contigs	1	1
Mean Depth of Coverage (ONT)	216	109
Genome Size	1,184,152	1,209,303
GC content	34.68	34.69
CDS	1216	1246
Average CDS length	812 bp	811 bp
Transfer RNA's	35	35

Table 4.6: Gene annotations for genes in *wRec* inversion. Gene annotations are from the *wMel* reference genome (Refseq accession: NC_002978.6).

RW <i>wRec</i> locus tag	Alb <i>wRec</i> locus tag	Length bp	Annotation
WR_RW_00077	WR_ALB_00110	165	hypothetical protein
WR_RW_00078	WR_ALB_00109	303	hypothetical protein
WR_RW_00079	WR_ALB_00108	96	hypothetical protein
WR_RW_00080	WR_ALB_00107	2430	DNA gyrase, B subunit
WR_RW_00081	WR_ALB_00106	609	SCO1/SenC family protein
WR_RW_00082	WR_ALB_00105	696	DNA polymerase III, epsilon subunit
WR_RW_00083	WR_ALB_00104	651	hypothetical protein
WR_RW_00084	WR_ALB_00103	504	protein-export protein SecB
WR_RW_00085	WR_ALB_00102	2586	aconitate hydratase
WR_RW_00086	WR_ALB_00101	3141	proline dehydrogenase/delta-1-pyrroline-5-carboxylate dehydrogenase
WR_RW_00087	WR_ALB_00100	282	IS4 family transposase ISWosp8
WR_RW_00088	WR_ALB_00099	285	IS4 family transposase ISWosp8
WR_RW_00089	WR_ALB_00098	345	sugE protein
WR_RW_00090	WR_ALB_00097	315	multidrug resistance protein
WR_RW_00091	WR_ALB_00096	1107	D-alanyl-D-alanine carboxypeptidase
WR_RW_00092	WR_ALB_00095	759	cell division protein FtsQ, putative
WR_RW_00093	WR_ALB_00094	954	D-alanine—D-alanine ligase
WR_RW_00094	WR_ALB_00093	648	hypothetical protein
WR_RW_00095	WR_ALB_00092	105	hypothetical protein
WR_RW_00096	WR_ALB_00091	375	ferredoxin, 4Fe-4S
WR_RW_00097	WR_ALB_00090	1089	DNA processing chain A
WR_RW_00098	WR_ALB_00089	699	triosephosphate isomerase
WR_RW_00099	WR_ALB_00088	74	tRNA-Thr(tgt)
WR_RW_00100	WR_ALB_00087	654	dimethyladenosine transferase
WR_RW_00101	WR_ALB_00086	1494	inosine-5'-monophosphate dehydrogenase
WR_RW_00102	WR_ALB_00085	135	hypothetical protein
WR_RW_00103	WR_ALB_00084	132	hypothetical protein
WR_RW_00104	WR_ALB_00083	147	hypothetical protein
WR_RW_00105	WR_ALB_00082	153	hypothetical protein
WR_RW_00106	WR_ALB_00081	1506	protein-export membrane protein SecD
WR_RW_00107	WR_ALB_00080	786	enoyl-(acyl-carrier-protein) reductase
WR_RW_00108	WR_ALB_00079	816	hypothetical protein
WR_RW_00109	WR_ALB_00078	1173	hypothetical protein
WR_RW_00110	WR_ALB_00077	165	hypothetical protein

Table 4.7: Genes in RW *w*Rec missing in Alb *w*Rec. Gene annotations and *w*Mel locus tags are from the *w*Mel reference genome (Refseq accession: NC_002978.6).

RW <i>w</i> Rec locus tag	<i>w</i> Mel locus tag	Length bp	Annotation
WR_RW_00553	WD_635	339	Hypothetical protein
WR_RW_00554	WD_636	159	Prophage Lambda W5, ankyrin repeat domain protein
WR_RW_00555	NA	1623	Reverse transcriptase, putative
WR_RW_00556	WD_636	222	Prophage Lambda W5, ankyrin repeat domain protein
WR_RW_00557	WD_636	117	Prophage Lambda W5, ankyrin repeat domain protein
WR_RW_00558	WD_637	735	Prophage Lambda W5, ankyrin repeat domain protein
WR_RW_00559	WD_638	1161	Hypothetical protein
WR_RW_00560	WD_639	792	Prophage Lambda W5, baseplate assembly protein J
WR_RW_00561	WD_640	336	Prophage Lambda W5, baseplate assembly protein W
WR_RW_00562	WD_641	255	hypothetical protein
WR_RW_00563	WD_642	465	Prophage Lambda W5, phage baseplate assembly protein V
WR_RW_00564	WD_643	477	hypothetical protein
WR_RW_00565	WD_644	522	Prophage Lambda W5, minor tail protein Z
WR_RW_00566	WD_605	306	Hypothetical protein
WR_RW_00567	WD_604	1005	Hypothetical protein
WR_RW_00568	WD_603	372	Hypothetical protein
WR_RW_00569	WD_602	1062	Prophage Lambda W4, minor capsid protein C
WR_RW_00570	NA	660	Hypothetical protein
WR_RW_00571	NA	663	Hypothetical protein
WR_RW_00572	WD_599	261	Hypothetical protein
WR_RW_00573	WD_598	234	Hypothetical protein
WR_RW_00574	WD_597	732	Prophage Lambda W4, terminase large subunit
WR_RW_00575	WD_597	924	Prophage Lambda W4, terminase large subunit
WR_RW_00576	WD_596	249	Prophage Lambda W4, ankyrin repeat domain protein
WR_RW_00577	NA	1098	IS4 family transposase
WR_RW_00578	NA	357	hypothetical protein
WR_RW_00579	NA	156	hypothetical protein
WR_RW_00580	NA	5265	hypothetical protein
WR_RW_00581	NA	864	hypothetical protein
WR_RW_00582	WD_646	381	Transposase IS5 family, OrfA
WR_RW_00583	WD_647	381	Transposase, IS5 family, OrfB

Table 4.8: Prevalence of *Wolbachia* variants from *D. recens* populations. Numbers represent the number of *D. recens* stocks with *Wolbachia* identified and includes those which were used for genome sequencing.

<i>Wolbachia</i> Strain	N Sample	Fly Stock	Location	Collection Year
RW <i>w</i> Rec	8	Recens White	New York	2001
		Ed245	Edmonton, Canada	2016
		Ed19	Edmonton, Canada	2019
		Cal104	Calgary, Canada	2016
		Cal62	Calgary, Canada	2019
		Cal7	Calgary, Canada	2019
		Can44	Canmore, Canada	2019
		PR79	Peace River, Canada	2016
Alb <i>w</i> Rec	5	AlbW+	Alberta, Canada	2009-2011
		Ed69	Edmonton, Canada	2016
		Ed1	Edmonton, Canada	2019
		Ed30	Edmonton, Canada	2019
		NY24	New York, USA	2014

CHAPTER 5

CONCLUSIONS AND FUTURE DIRECTIONS

In this dissertation I investigate evolutionary relationships in the *D. subquinaria* species complex and assess *Wolbachia*-host dynamics in both native and non-native host species. In Chapter 4 I interpret these findings in light of whole genome sequencing of all host species and two *Wolbachia* variants infecting *D. recens*. In Chapter 2 I show, using a Sanger-sequenced multi-locus dataset, that phylogenetic relationships are obscured among species. Using a combination of phylogenetic and population genetic analyses, I conclude that the most strongly supported interpretation for these patterns is pervasive incomplete lineage sorting among all species, and gene flow between *D. recens* and *D. subquinaria* drives these patterns. While phylogenetic topologies produced are inconsistent with levels of post-mating reproductive isolation (Humphreys, Rundle, & Dyer, 2016), the invocation of gene flow is consistent with previous research in this species complex (Dyer, Bewick, White, Bray, & Humphreys, 2018; Jaenike, Dyer, Cornish, & Minhas, 2006). With the observation mitochondrial introgression into *D. subquinaria*, I ask whether we should expect *Wolbachia* to invade from *D. recens* in Chapter 3. With genetic backcrossing, I estimate the phenotypic consequences of *Wolbachia* infection in *D. subquinaria* and its native host, *D. recens*. I find strong cytoplasmic incompatibility, high maternal transmission, and slightly higher fecundity in *Wolbachia*-infected females (not statistically significant) in both species. Leveraging an established model of CI-causing *Wolbachia* dynamics (Turelli, 1994), I predict that minimal levels of gene flow from *D. recens*

into *D. subquinaria* will likely lead to *Wolbachia* invasion of *D. subquinaria*. In Chapter Four I generate full genome assemblies for all members of the *D. subquinaria* species complex and two *Wolbachia* infections in *D. recens*, one of which was used to conduct experiments from Chapter 3. Here I find that greater sampling of loci from the genome produces a phylogenetic topology consistent with levels of post-mating reproductive isolation. I also find substantial variation in assembly size (both within and between species) and conclude that this variation cannot be explained by transposable element proliferation and is likely at least in part an artifact of the assembly process. *Wolbachia* genome assemblies from the two sequenced *D. recens* collections reveal distinct *Wolbachia* variants differing by two major structural changes. One of these changes, a ~25.6 kb indel, is a likely causal factor in the induction of cytoplasmic incompatibility vs. male-killing induced in the same *D. subquinaria* genetic backgrounds.

Several open questions remain concerning this dissertation. First, discordance among phylogenetic results in Chapter Two and Chapter Four reveal how disparity in datasets can result in uncertainty of species relationships. Population resequencing will help distinguish phylogenetic patterns across the genome and remove limitations in both datasets analyzed here. Additionally, this will help identify regions of high divergence between species, and thereby generate candidate regions for incompatibility loci responsible for post-mating reproductive barriers.

From analyses in Chapter Three I predict *Wolbachia* invasion of *D. subquinaria*, however this is making several assumptions that were not directly addressed in this dissertation. First, future research should assess *Wolbachia* effects in sympatric *D. subquinaria* populations. While we observe mitochondrial introgression from *D. recens* into all *D. subquinaria* populations, this has only occurred without concomitant introgression of *Wolbachia*. It is possible that *Wolbachia*

infection is lost (e.g. by low maternal transmission) during the introgression process through sympatric *D. subquinaria* populations, or that only the few naturally uninfected *D. recens* females produce sufficiently fit hybrids with *D. subquinaria* to permit introgression. Future projects should test this directly by examining *Wolbachia*-infected vs. uninfected hybrids in early generation backcrosses with sympatric *D. subquinaria*.

The results presented Chapter Four suggest that a potential reason *Wolbachia* has not invaded *D. subquinaria* is that the *Wolbachia* strain used to predict invasion is a recent *Wolbachia* variant polymorphic in *D. recens*. In support of this hypothesis, ongoing work in the lab has shown that these two strains induce drastically different phenotypes in *D. subquinaria* (CI vs. male-killing). Further investigation of the differences between these two *Wolbachia* strains is needed to identify necessary genetic factors responsible for the induction of male-killing. Currently, research on the topic implicates a single gene (*wmk*) responsible for the male-killing phenotype (Perlmutter et al., 2019), yet the two strains described here have an identical copy of *wmk*. This suggests that current explanations for induction of male-killing are incomplete, and this system offers an unique opportunity for further investigation. Perlmutter (2019) posited that *Wolbachia* gene expression may be important for the induction of male-killing, and provides a hypothesis for the induction of male-killing on only one of the two *Wolbachia* strains described here.

At this time, we do not have broad sampling of *Wolbachia* variants across the range of *D. recens*. Data presented in Chapter 4 is consistent with an ongoing sweep of the Alb *wRec* variant that emerged in eastern *D. recens* populations. This is, however, based on limited sampling, and a PCR assay that can only distinguished *wRec* variants on the basis of the ~25.6 kb indel. Further work is required to confirm that these variants are consistently different from one another among

D. recens samples (i.e. no recombination among *Wolbachia* variants), and that their identity with respect to the indel covaries with the induction of male-killing in *D. subquinaria*.

There is much left to learn from the fly genome assemblies presented in Chapter 4. All assemblies are high quality by current standards in the field. These genomes provide an information-rich resource to investigate chromosomal and protein evolution across a whole species complex showing varying levels of reproductive isolation. Unfortunately, I was unable to complete many of these analyses due to time constraints. First, with additional investigation into genome heterozygosity I aim to reduce diploid contigs in the assemblies without also losing unique genetic content. Second, I plan to investigate chromosomal evolution in the species complex. Studies of genome evolution in *Drosophila* shown strong conservation of chromosome arm content, but extensive intra-chromosomal rearrangements (Bhutkar et al., 2008; Chakraborty et al., 2021). And in several systems, chromosomal rearrangements have been implicated in the evolution of reproductive isolation (Brown, Burk, Henagan, & Noor, 2004). Future analysis of the genome assemblies presented here will investigate whether chromosomal rearrangements play a role in the diversification of this species complex.

Works Cited

Bhutkar, A., Schaeffer, S. W., Russo, S. M., Xu, M., Smith, T. F., & Gelbart, W. M. (2008).

Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes.

Genetics, 179(3), 1657-1680. doi:10.1534/genetics.107.086108

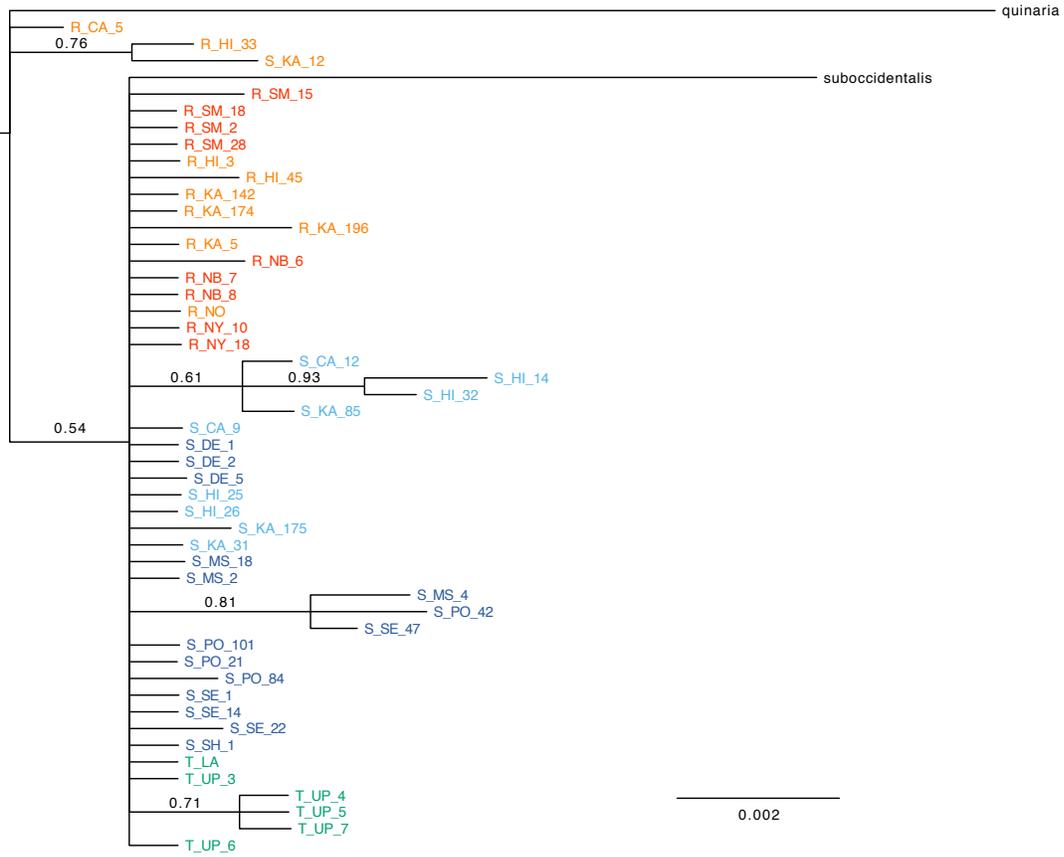
Brown, K. M., Burk, L. M., Henagan, L. M., & Noor, M. A. (2004). A test of the chromosomal rearrangement model of speciation in *Drosophila pseudoobscura*. *Evolution*, 58(8), 1856-1860. doi:10.1111/j.0014-3820.2004.tb00469.x

- Chakraborty, M., Chang, C. H., Khost, D. E., Vedanayagam, J., Adrion, J. R., Liao, Y., . . . Emerson, J. J. (2021). Evolution of genome structure in the *Drosophila simulans* species complex. *Genome Res*, *31*(3), 380-396. doi:10.1101/gr.263442.120
- Dyer, K. A., Bewick, E. R., White, B. E., Bray, M. J., & Humphreys, D. P. (2018). Fine-scale geographic patterns of gene flow and reproductive character displacement in *Drosophila subquinaria* and *Drosophila recens*. *Mol Ecol*. doi:10.1111/mec.14825
- Humphreys, D. P., Rundle, H. D., & Dyer, K. A. (2016). Patterns of reproductive isolation in the *Drosophila subquinaria* complex: can reinforced premating isolation cascade to other species? *Curr Zool*, *62*(2), 183-191. doi:10.1093/cz/zow005
- Jaenike, J., Dyer, K. A., Cornish, C., & Minhas, M. S. (2006). Asymmetrical reinforcement and *Wolbachia* infection in *Drosophila*. *PLoS Biol*, *4*(10), e325. doi:10.1371/journal.pbio.0040325
- Perlmutter, J. I., Bordenstein, S. R., Unckless, R. L., LePage, D. P., Metcalf, J. A., Hill, T., . . . Bordenstein, S. R. (2019). The phage gene *wmk* is a candidate for male killing by a bacterial endosymbiont. *PLoS Pathog*, *15*(9), e1007936. doi:10.1371/journal.ppat.1007936
- Turelli, M. (1994). Evolution of Incompatibility-Inducing Microbes and Their Hosts. *Evolution*, *48*(5), 1500-1513. doi:10.1111/j.1558-5646.1994.tb02192.x

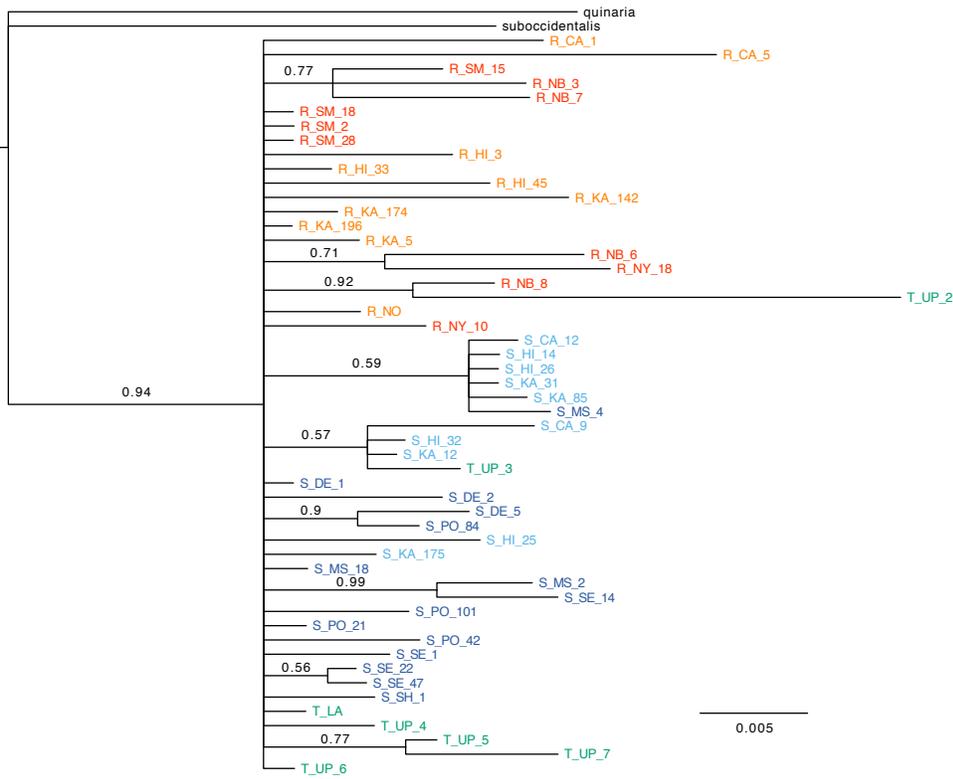
APPENDIX A

Supplementary Materials for Chapter 2

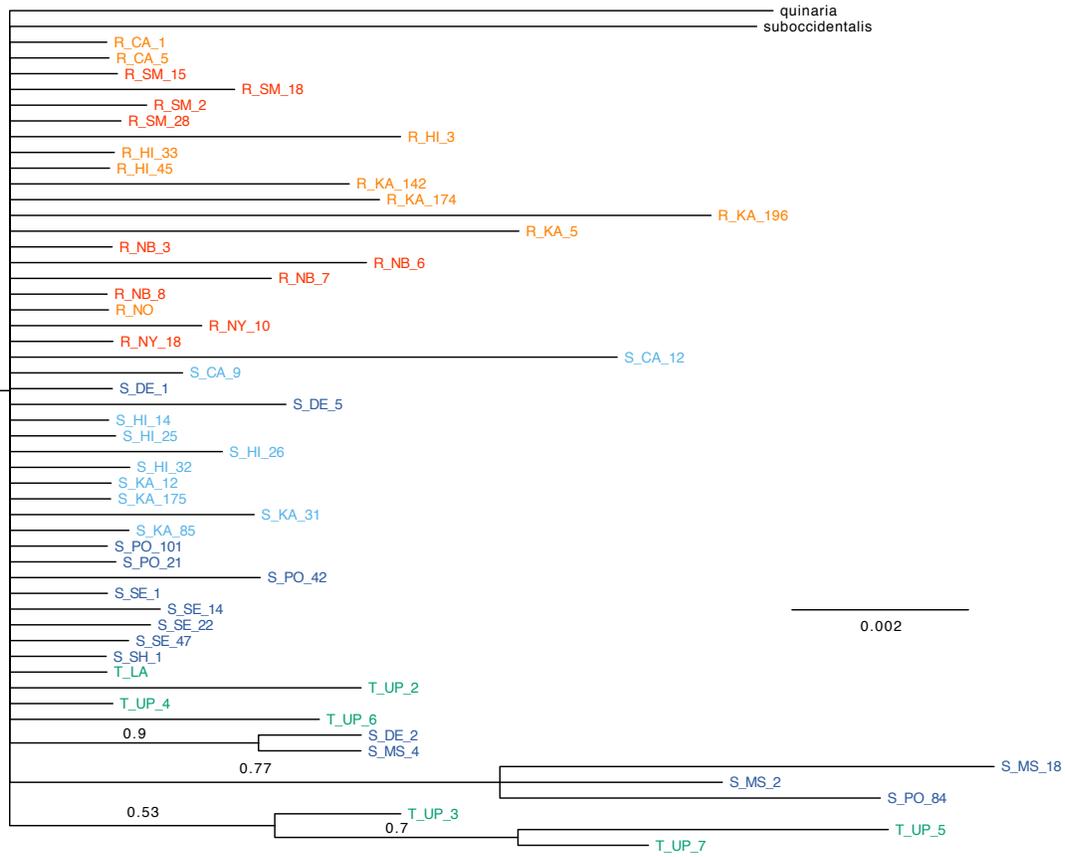
A. abdA



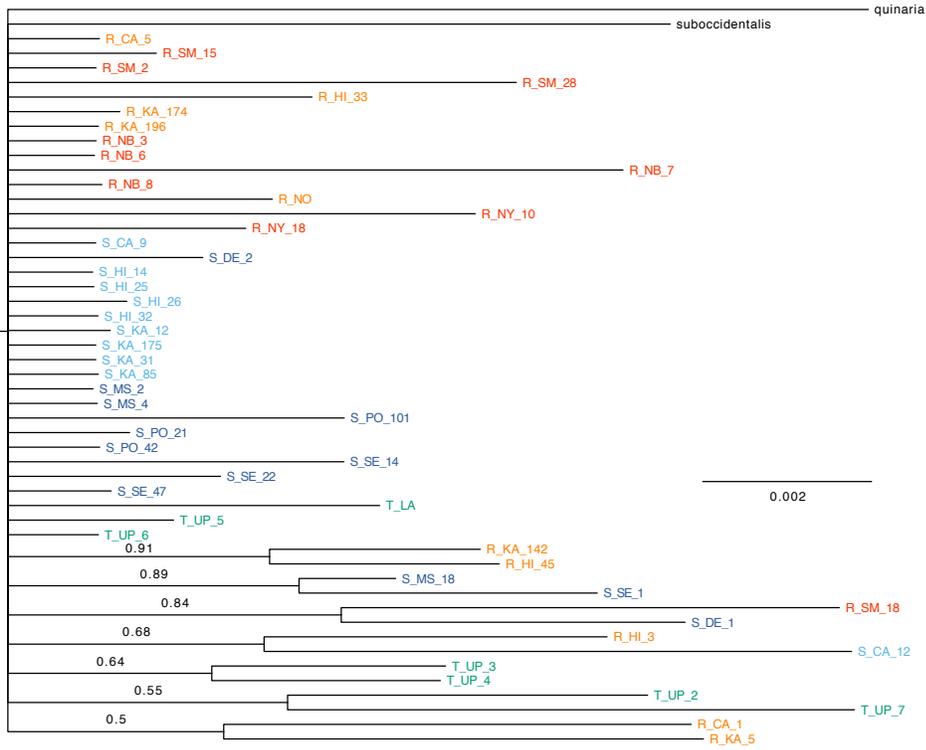
B. adhr



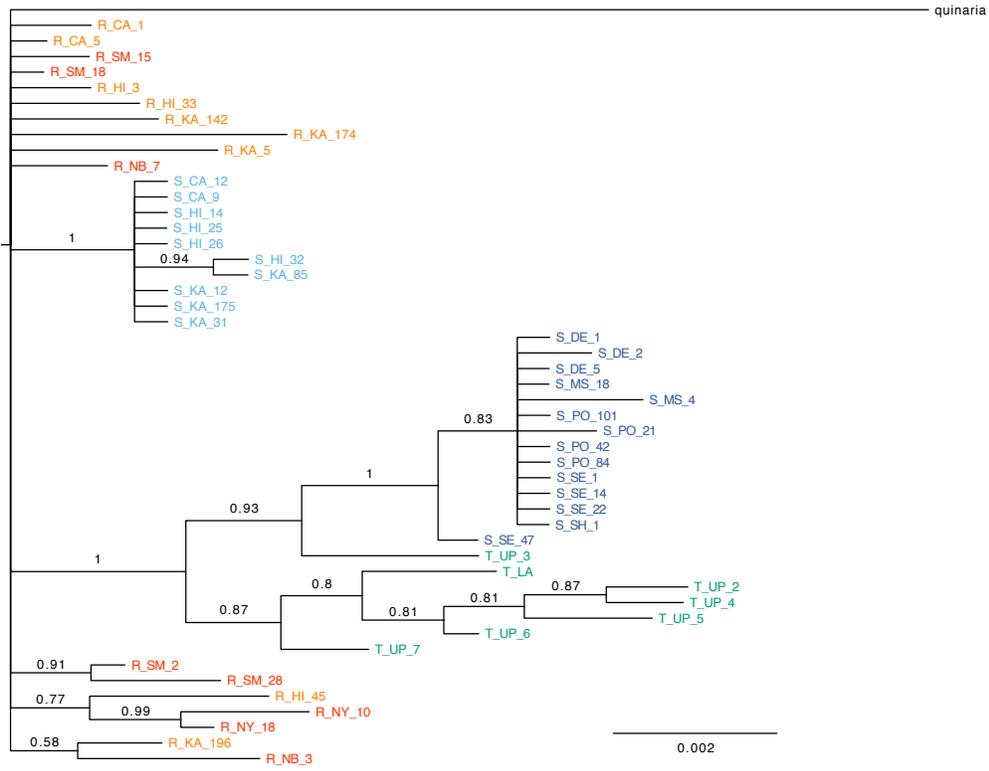
C. bab2



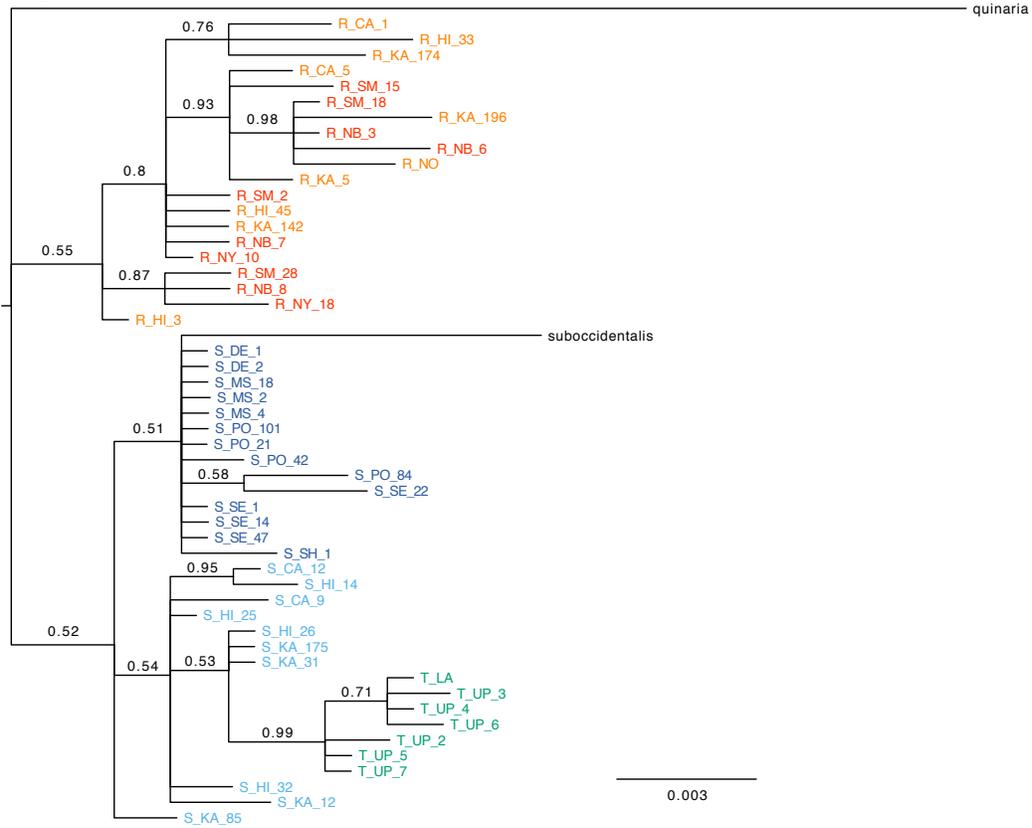
D. black



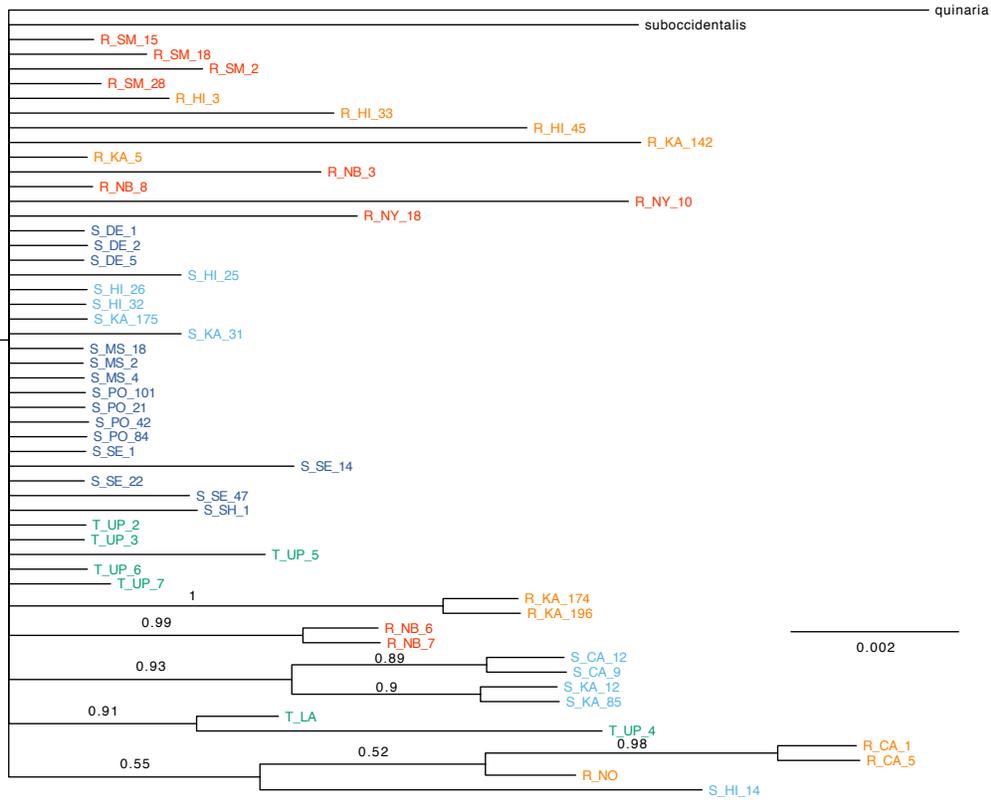
E. bt



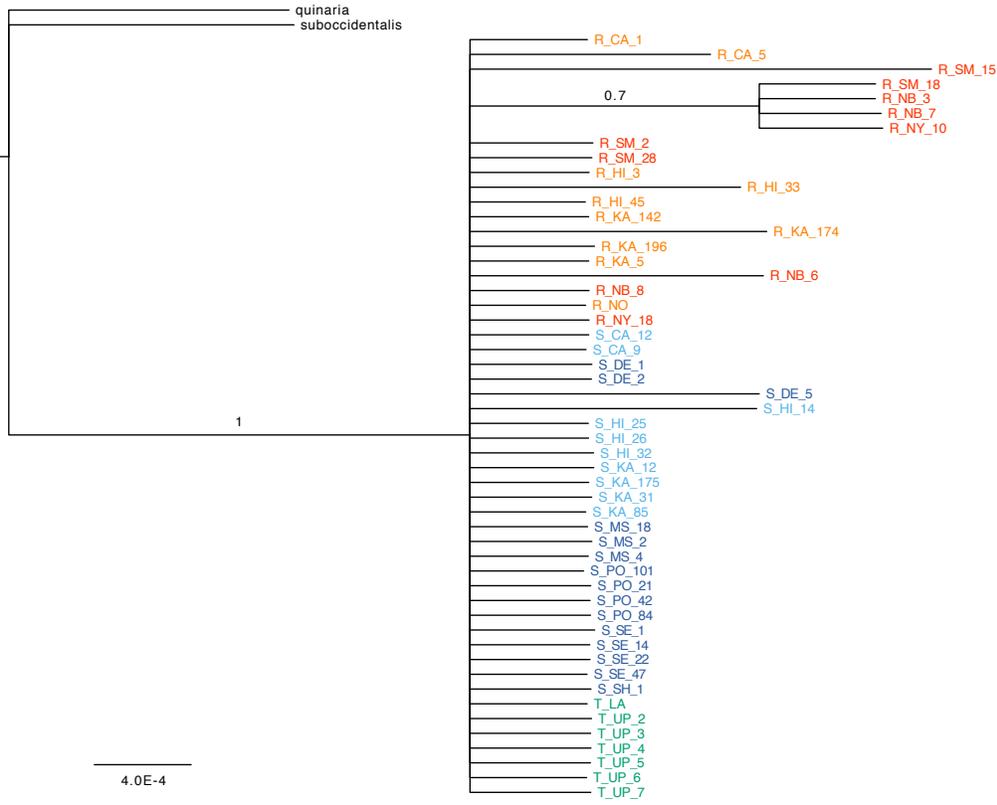
F. cp36



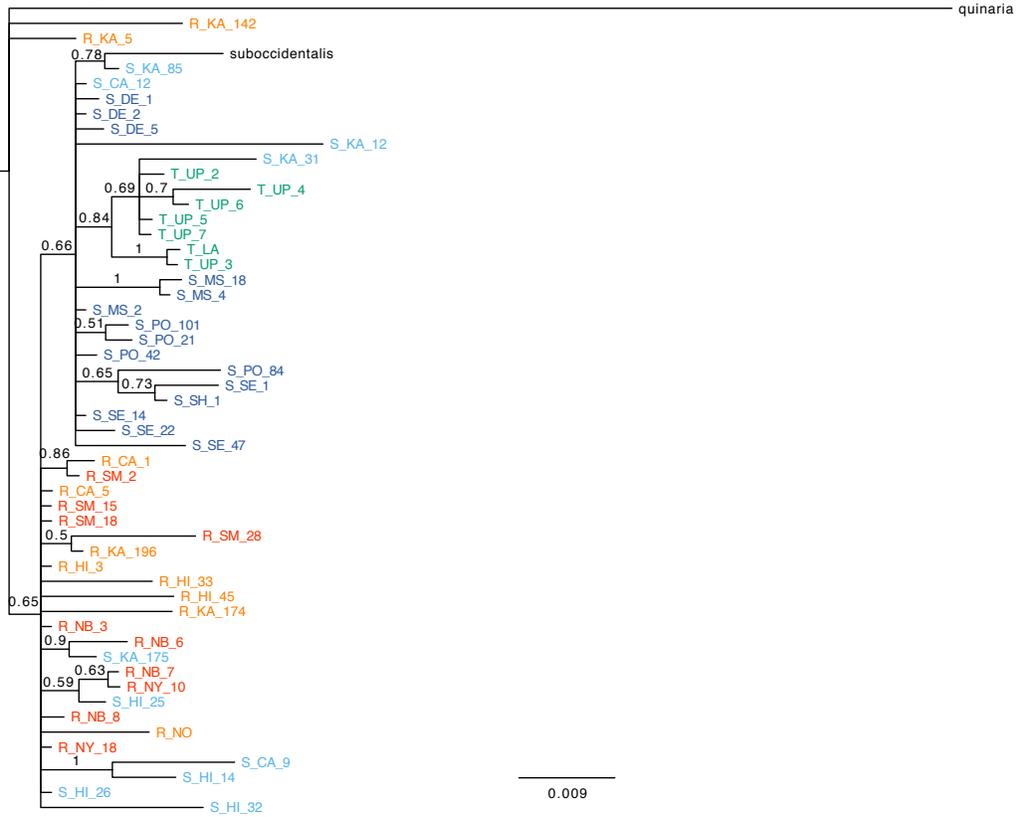
G. desat2



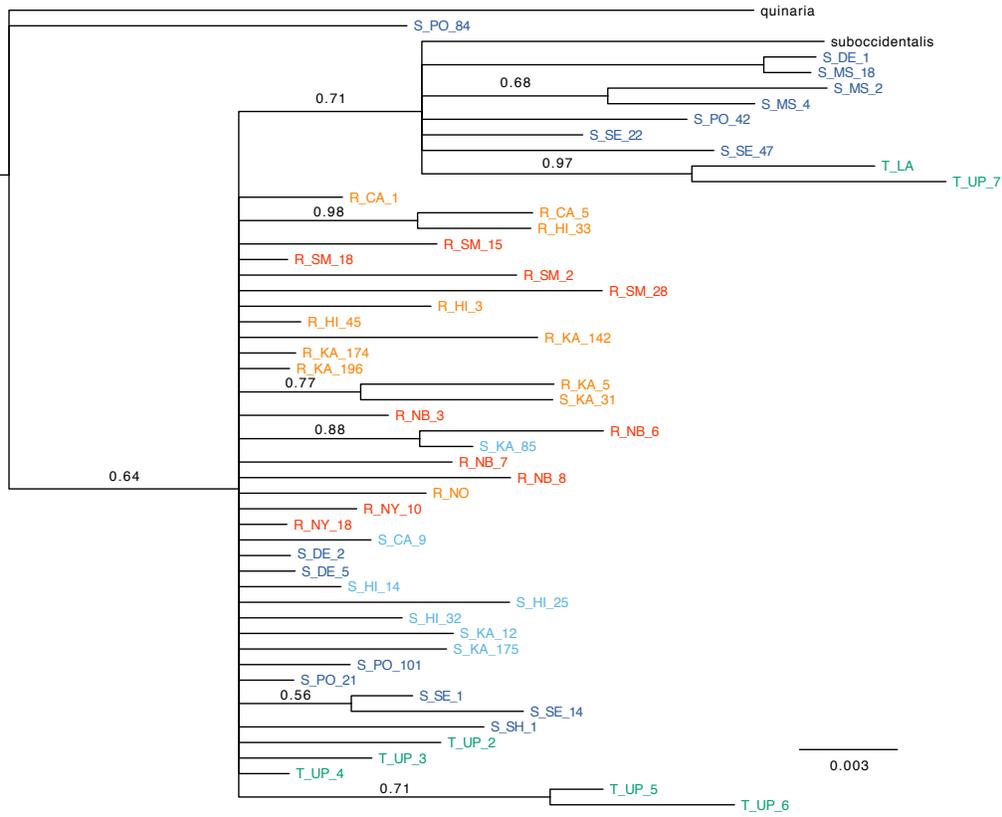
G. dsx



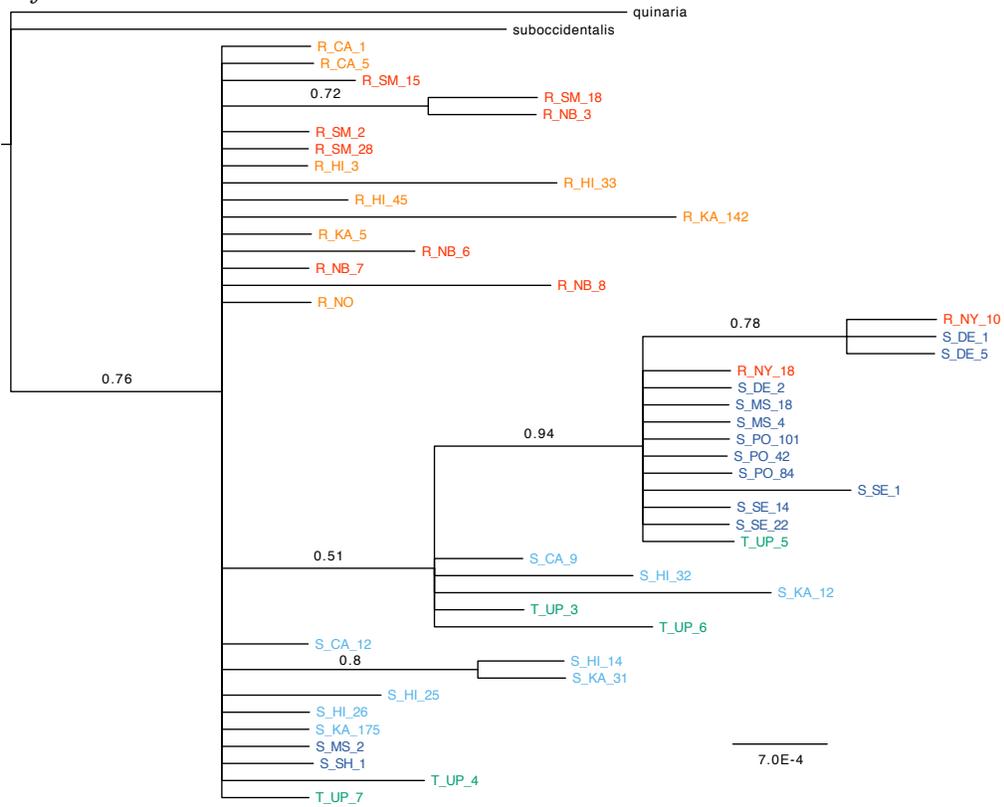
H. ebony



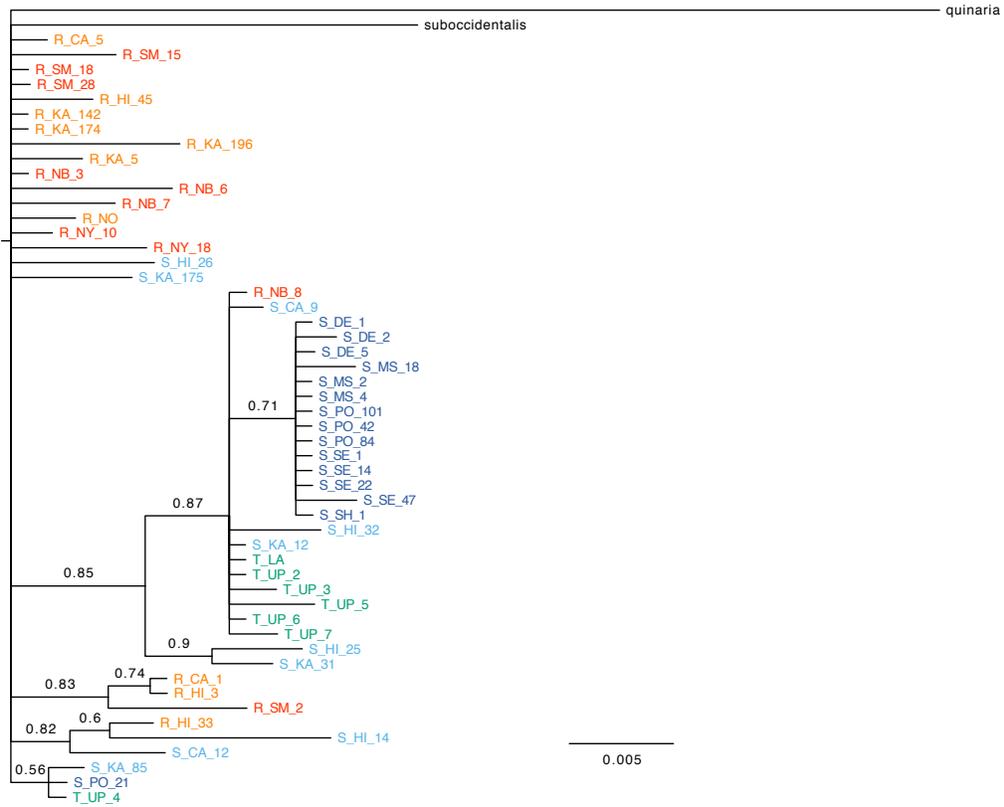
J. esc



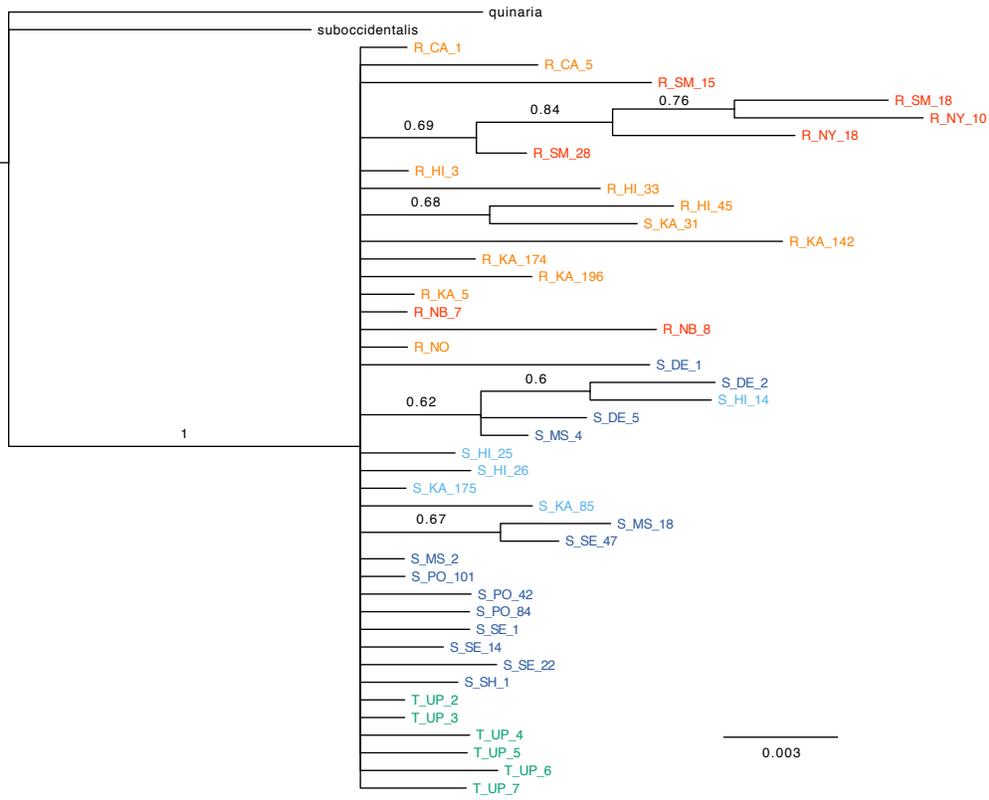
K. fru



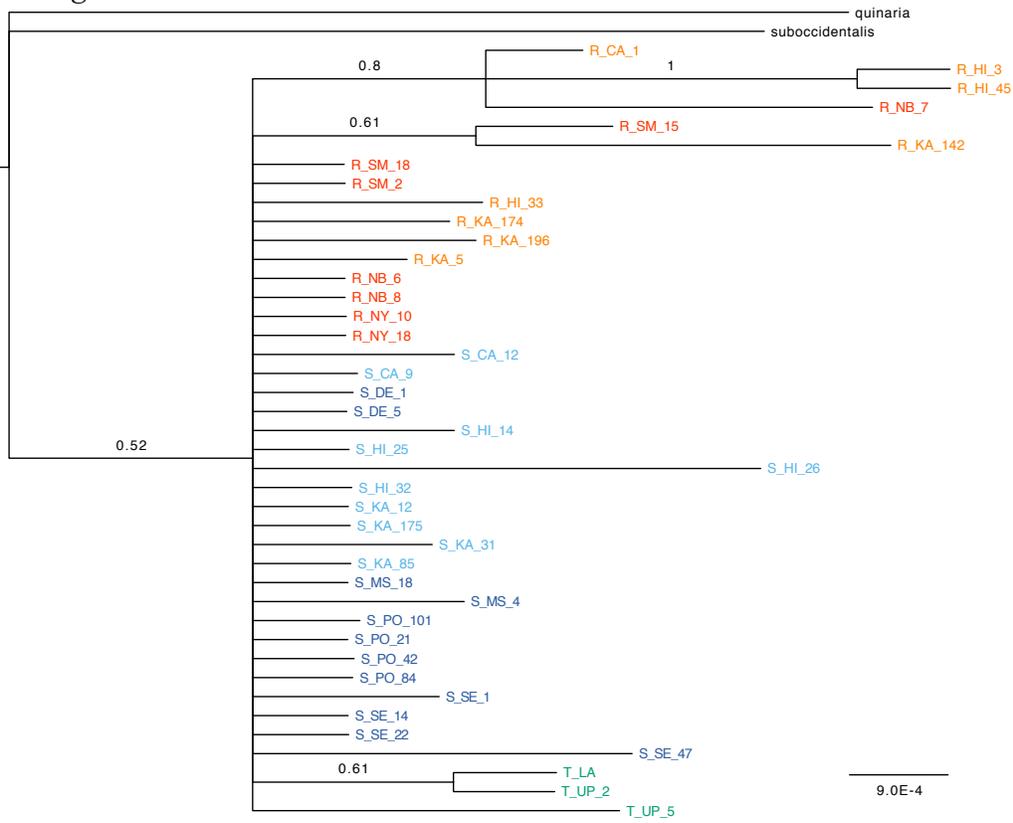
L. hb



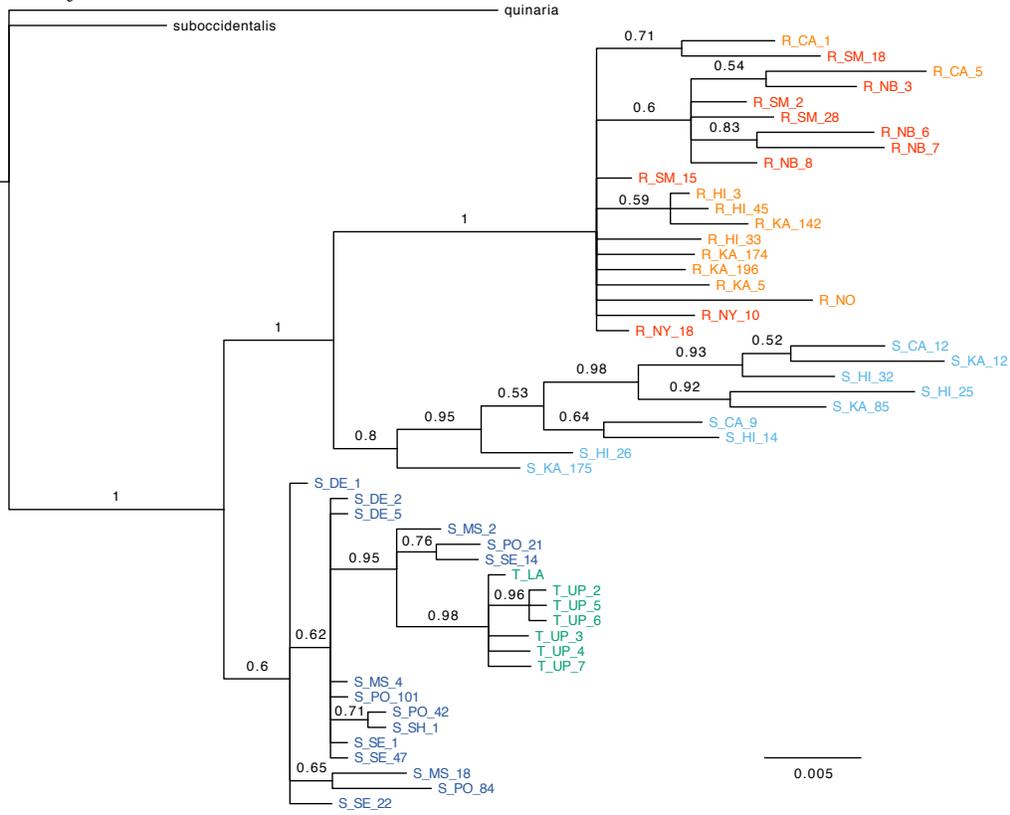
M. ix



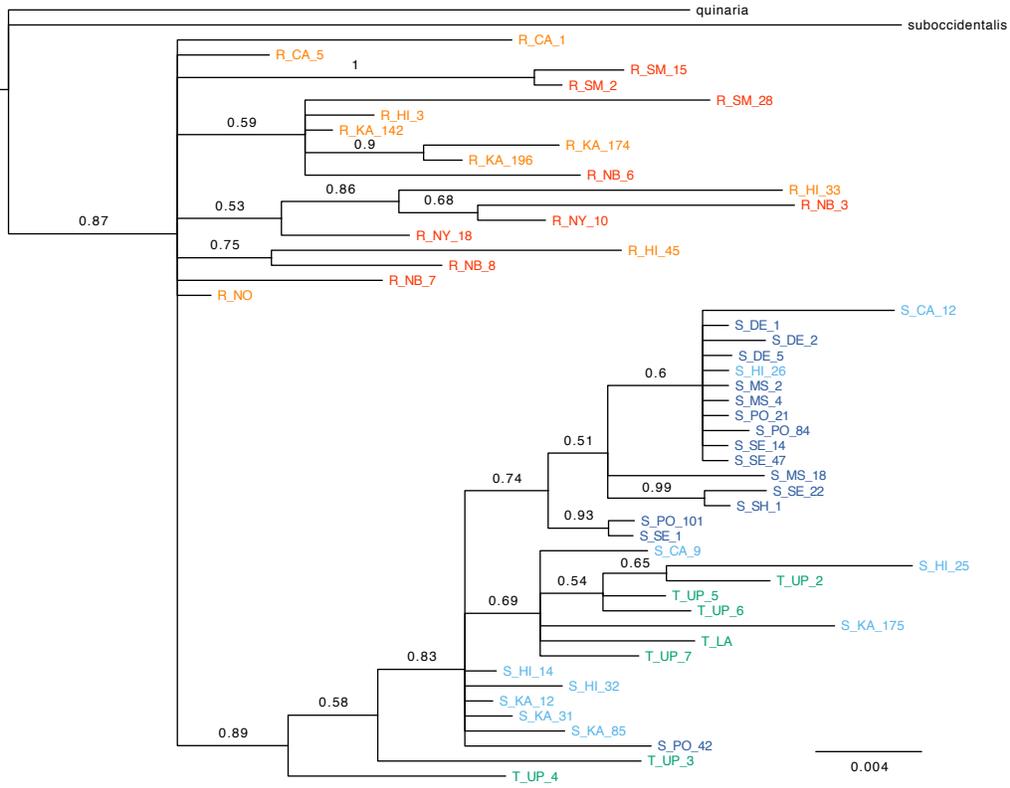
N. mago



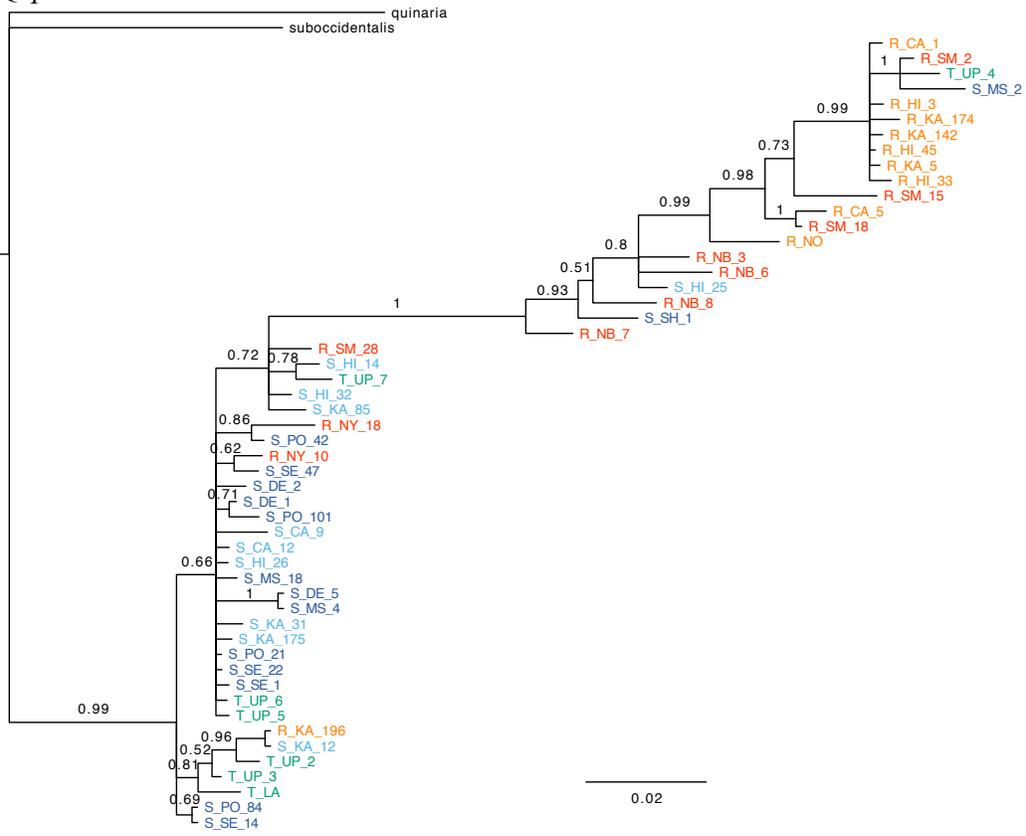
O. mof



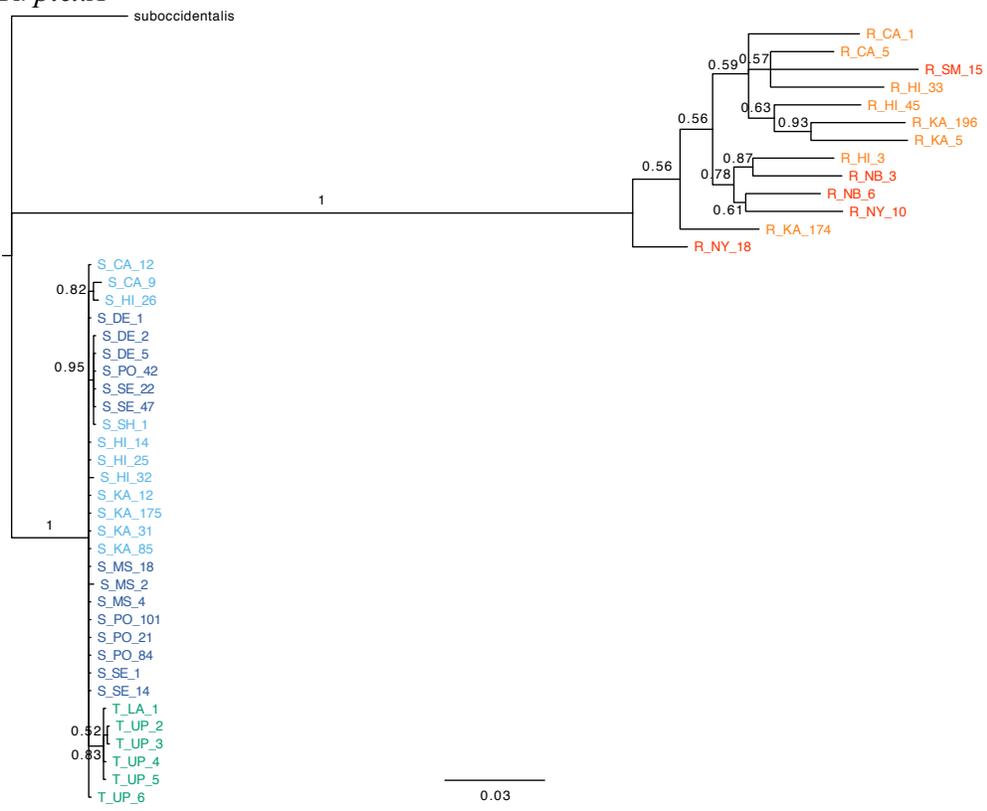
P. ntid



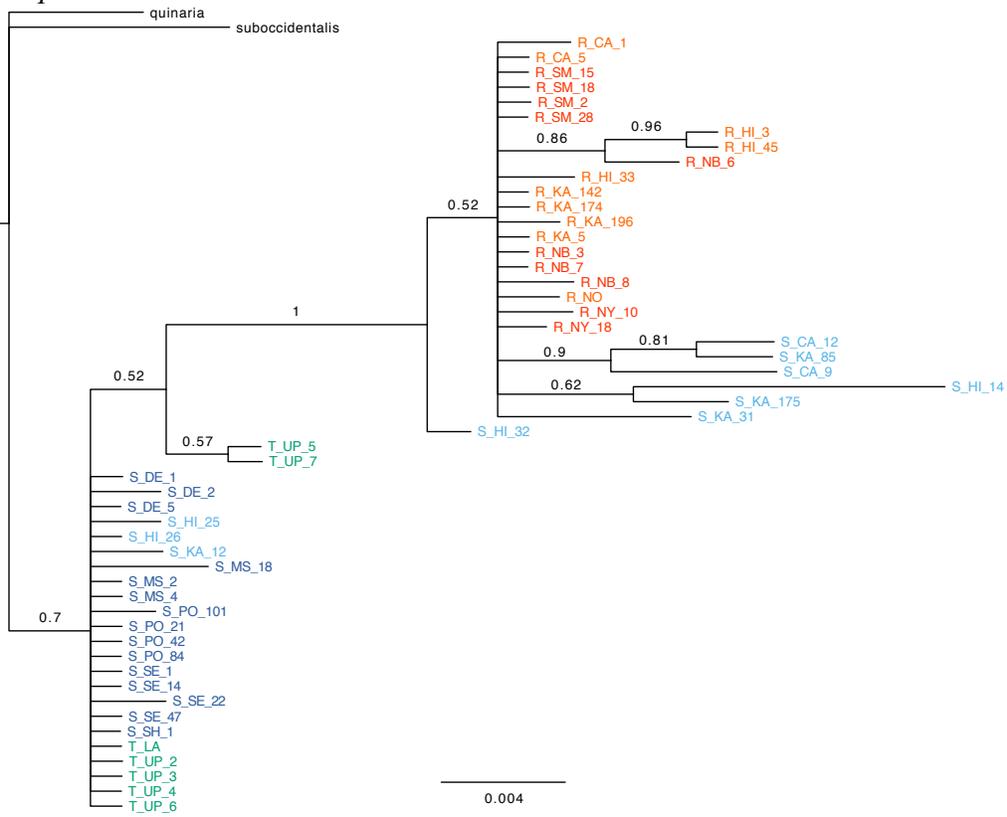
Q. per



R. plexA

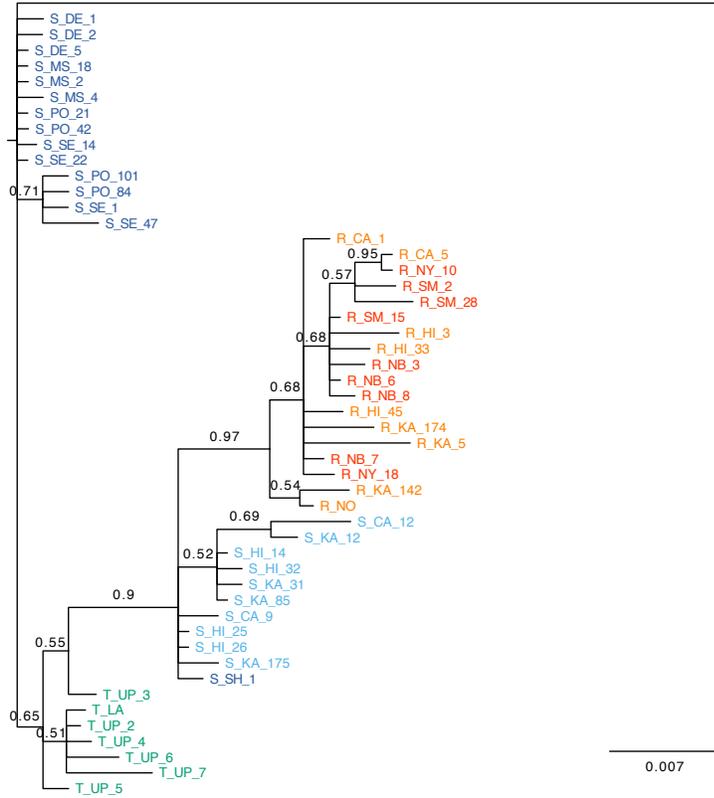


S. rpl36

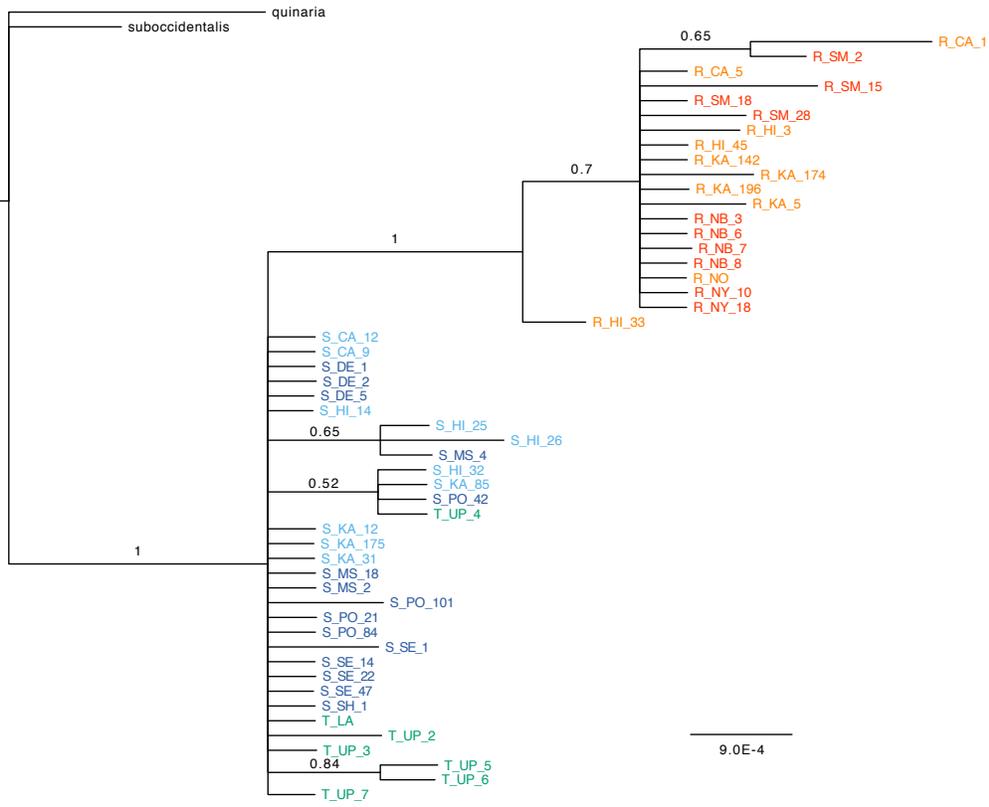


T. scu

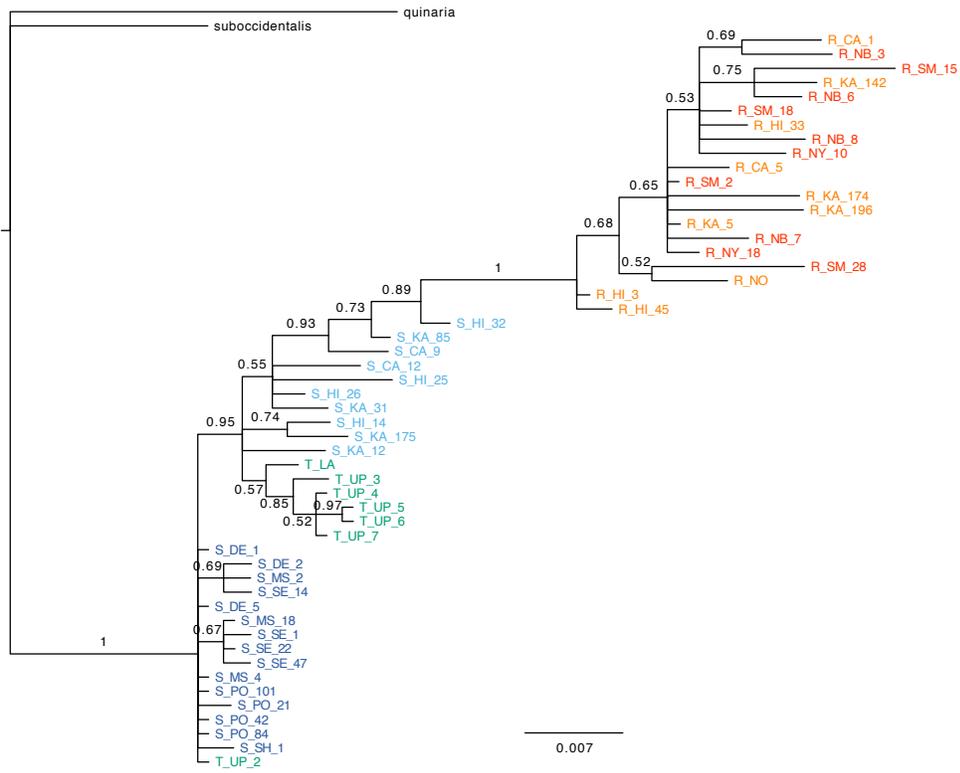
palustris



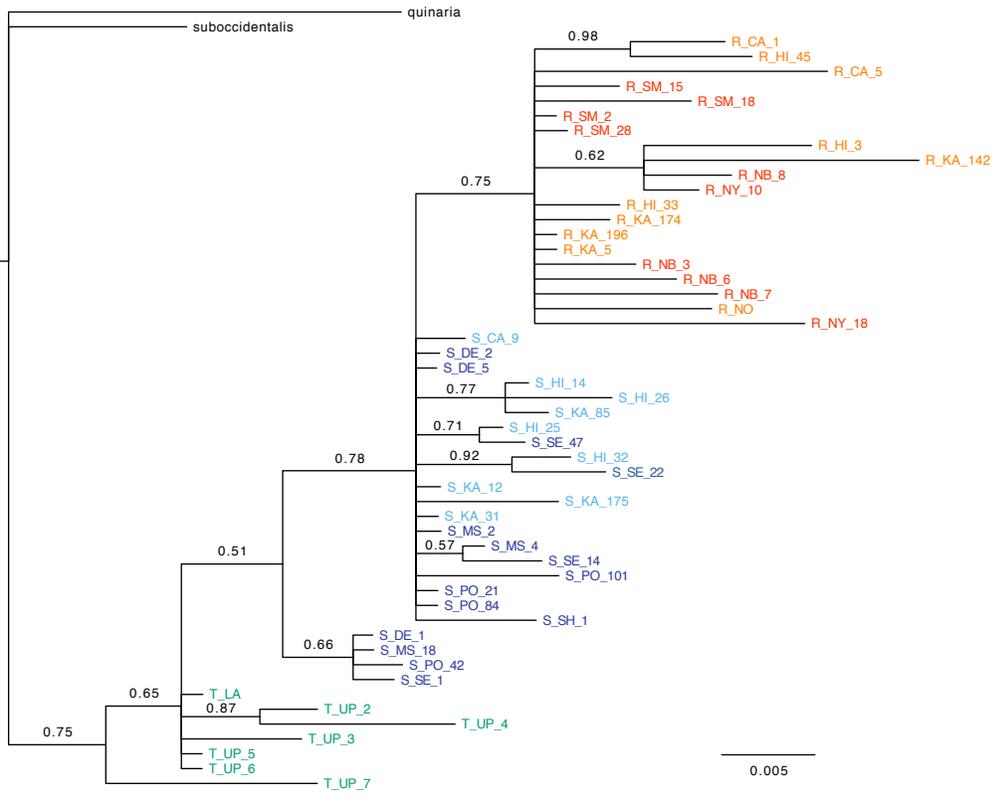
U. sina



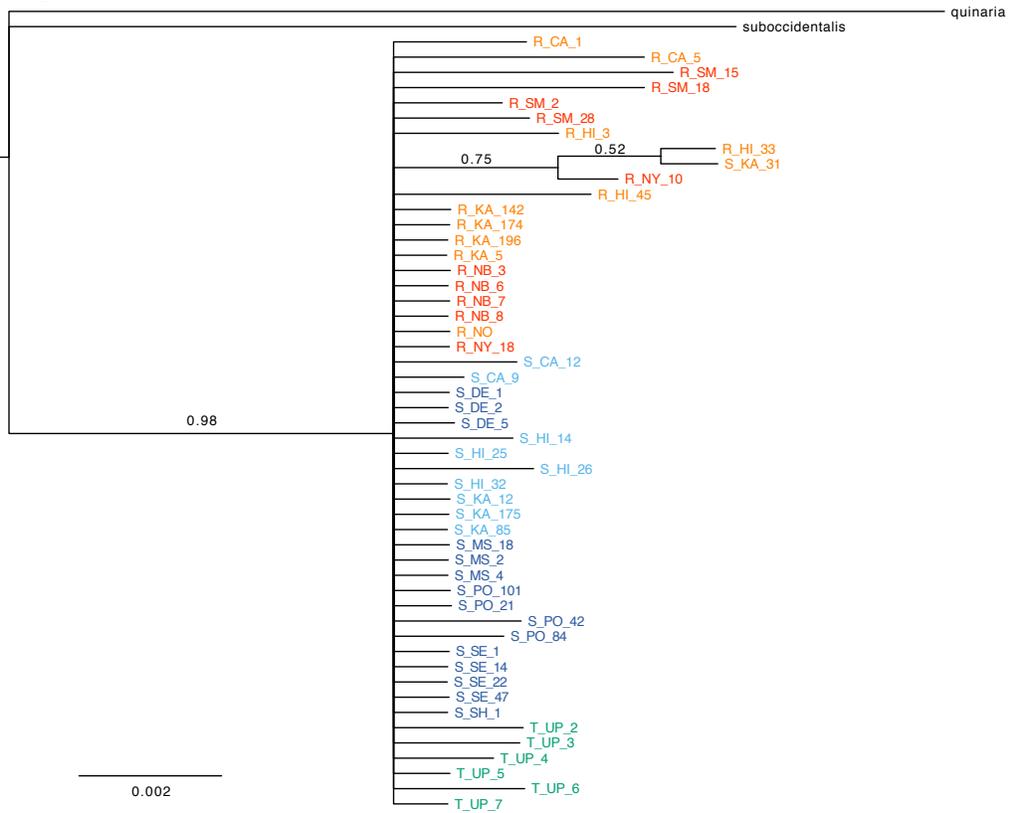
V. svr



W. tim



X. tpi



Y. wee

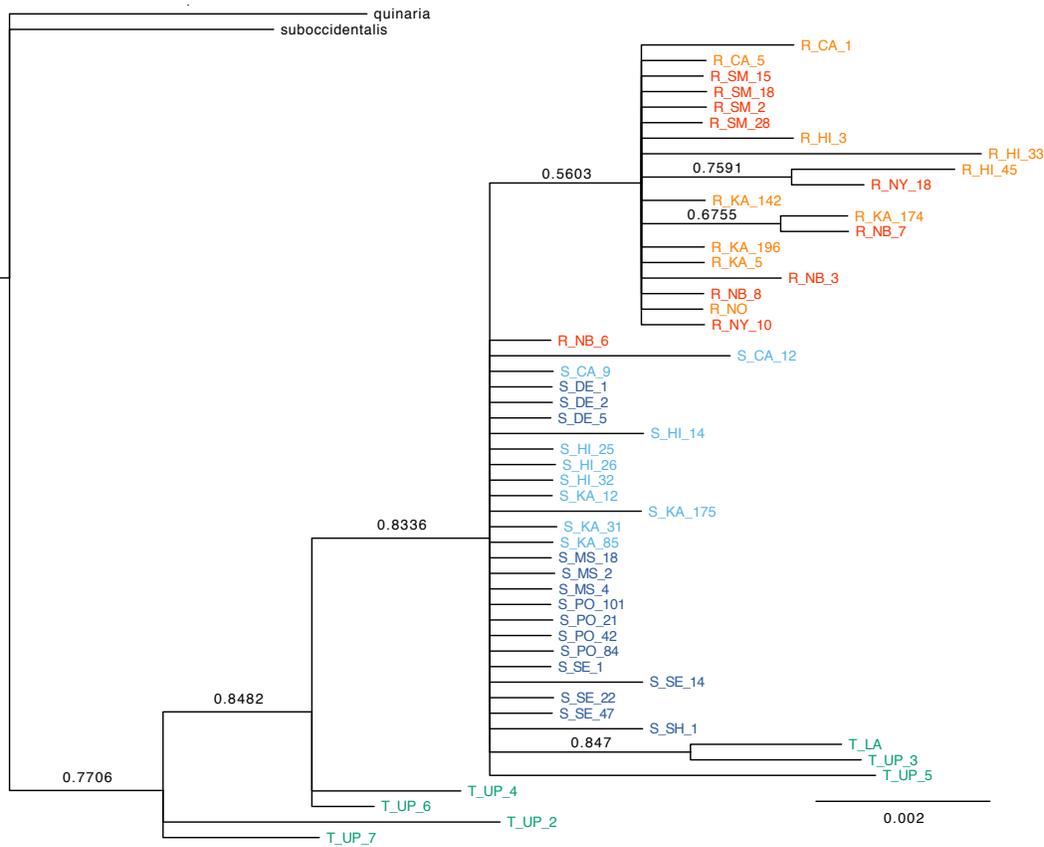


Figure S2.1: Individual gene trees from the Bayesian analysis. Each tree is the majority rule consensus tree and support values greater than 0.50 are shown. Labels indicate species (S_ for *D. subquinaria*, R_ for *D. recens*, T_ for *D. transversa*) followed by the population abbreviation as in Table 1 and the specific line number. Labels are colored as in Figure 1, where light blue indicates sympatric *D. subquinaria*, dark blue indicates allopatric *D. subquinaria*, orange indicates sympatric *D. recens*, red indicates allopatric *D. recens*, and green indicates *D. transversa*.

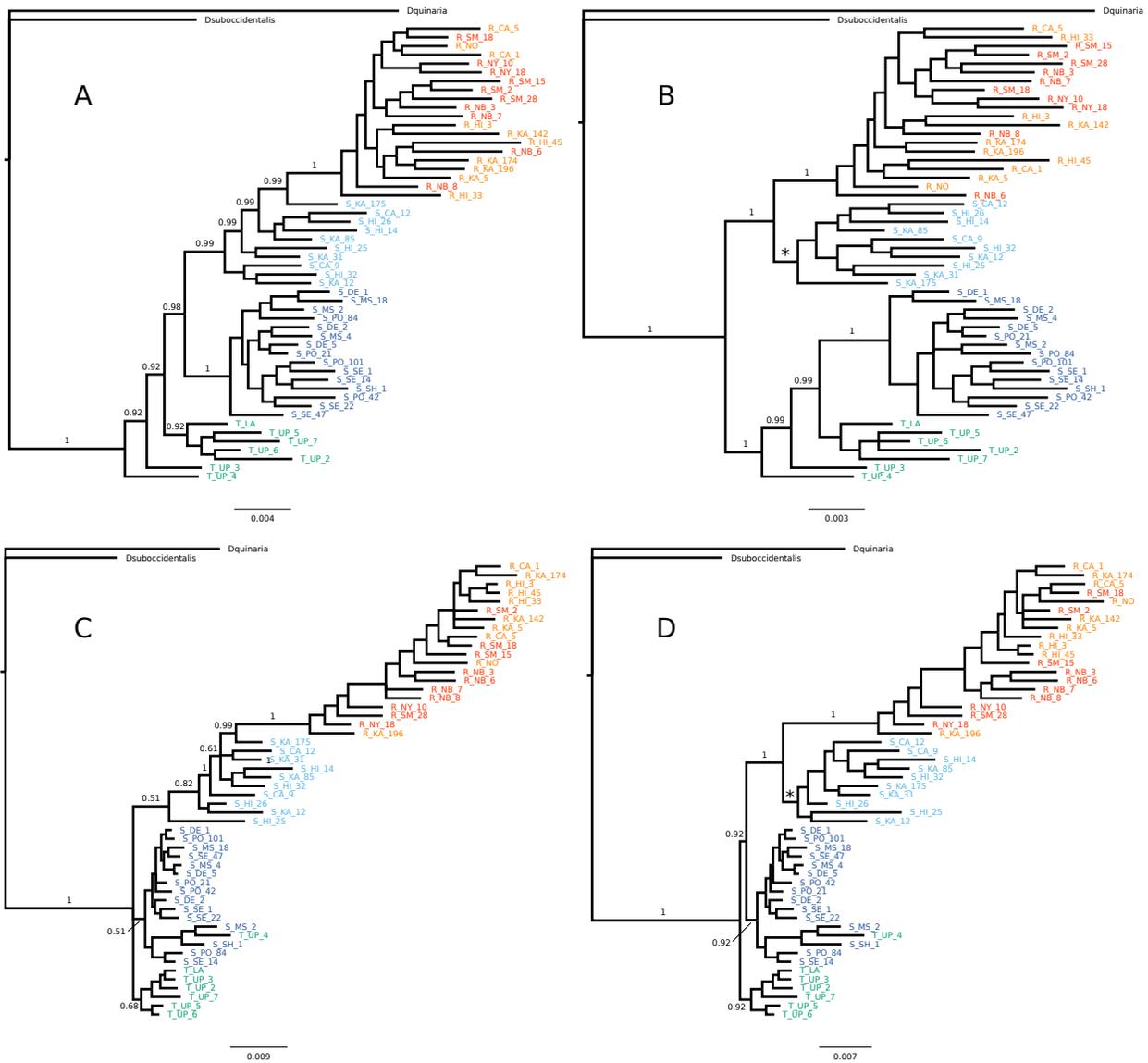


Figure S2.2: Inferred majority rule trees with and without topological constraints from the concatenated Bayesian analyses using autosomal data (Panels A & B) and X-linked data (Panels C & D). Panels A and C show results from analyses with topology free to vary, and Panel B & D show results where sympatric *D. subquinaria* samples are constrained to monophyly, with the * indicating the constrained node. Labels indicate species (S_ for *D. subquinaria*, R_ for *D. recens*, T_ for *D. transversa*) followed by the population abbreviation as in Table 1 and the line number. Labels are colored as in Figure 1, where light blue indicates sympatric *D. subquinaria*, dark blue indicates allopatric *D. subquinaria*, orange indicates sympatric *D. recens*, red indicates allopatric *D. recens*, and green indicates *D. transversa*. Only the support values for the main nodes are indicated.

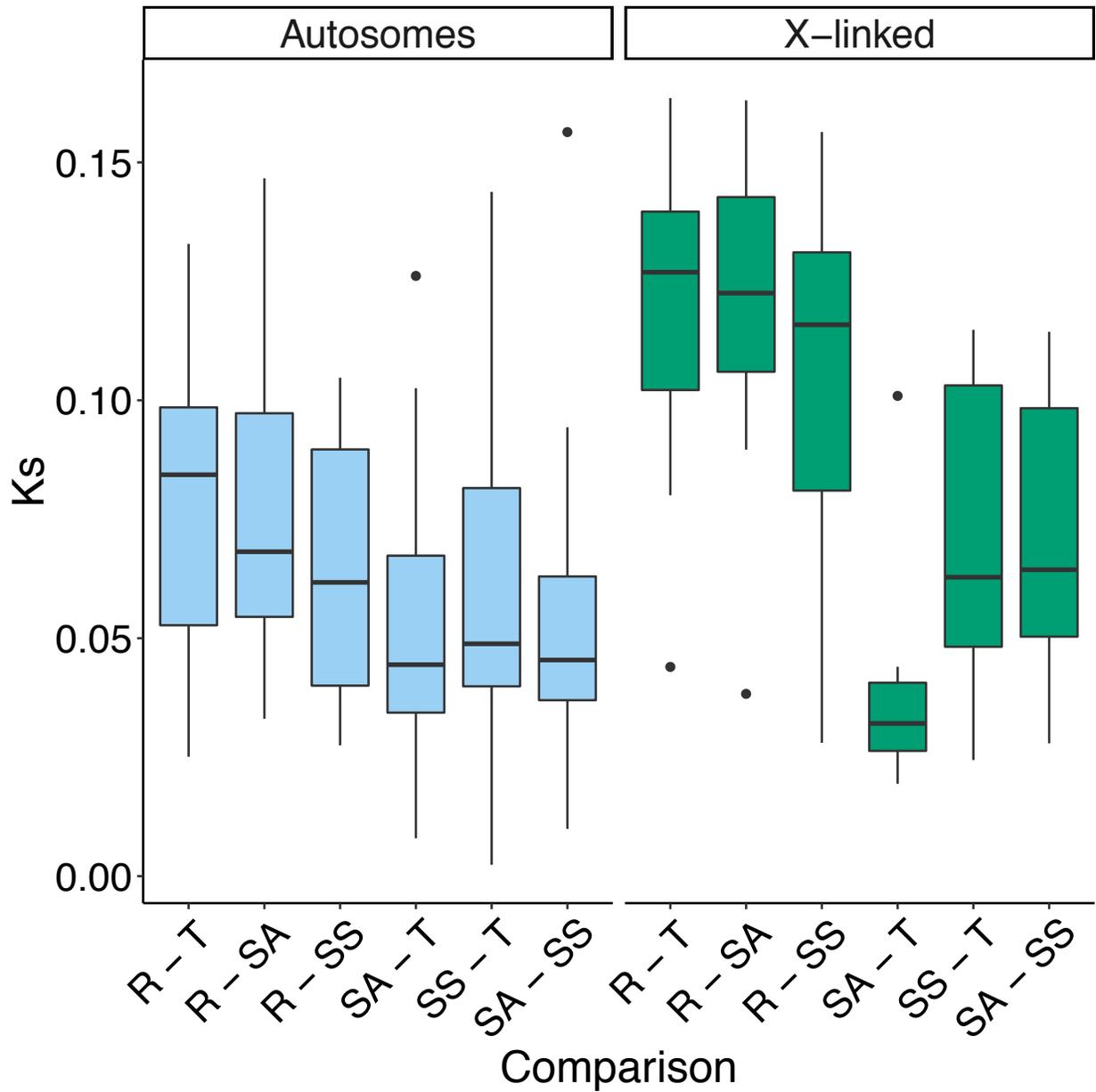


Figure S2.4: Synonymous substitution rate (K_s) for each group comparison. Loci are separated by autosomal (blue) and X-linked (green). Groups in comparisons are abbreviated as *D. recens* (R), *D. transversa* (T), allopatric *D. subquinaria* (SA), and sympatric *D. subquinaria* (SS).

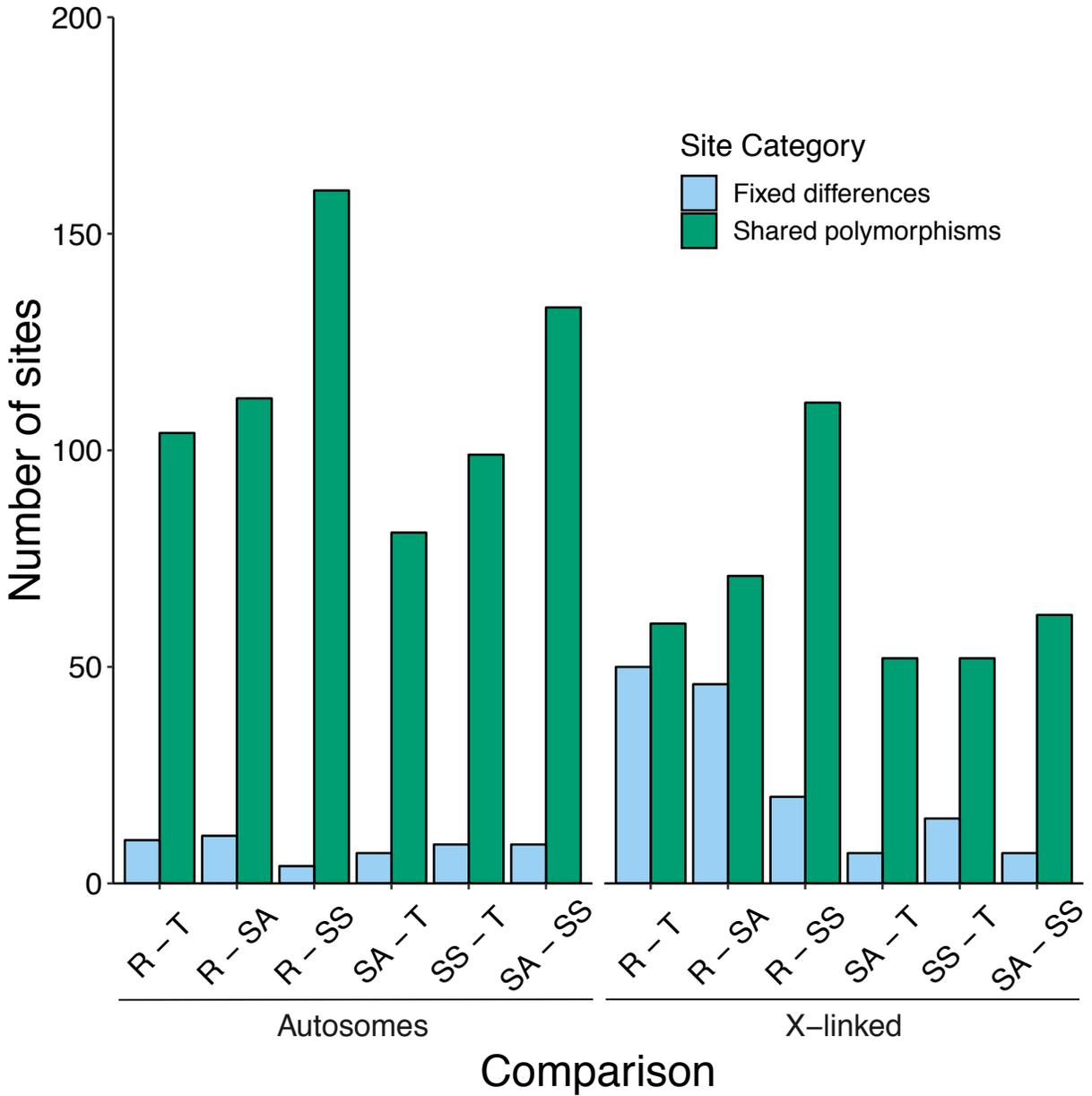


Figure S2.5: The total number of fixed differences (blue) and polymorphic sites (green) categorized by genomic region (i.e. autosomal and X-linked loci). Groups in comparisons are abbreviated as *D. recens* (R), *D. transversa* (T), allopatric *D. subquinaria* (SA), and sympatric *D. subquinaria* (SS).

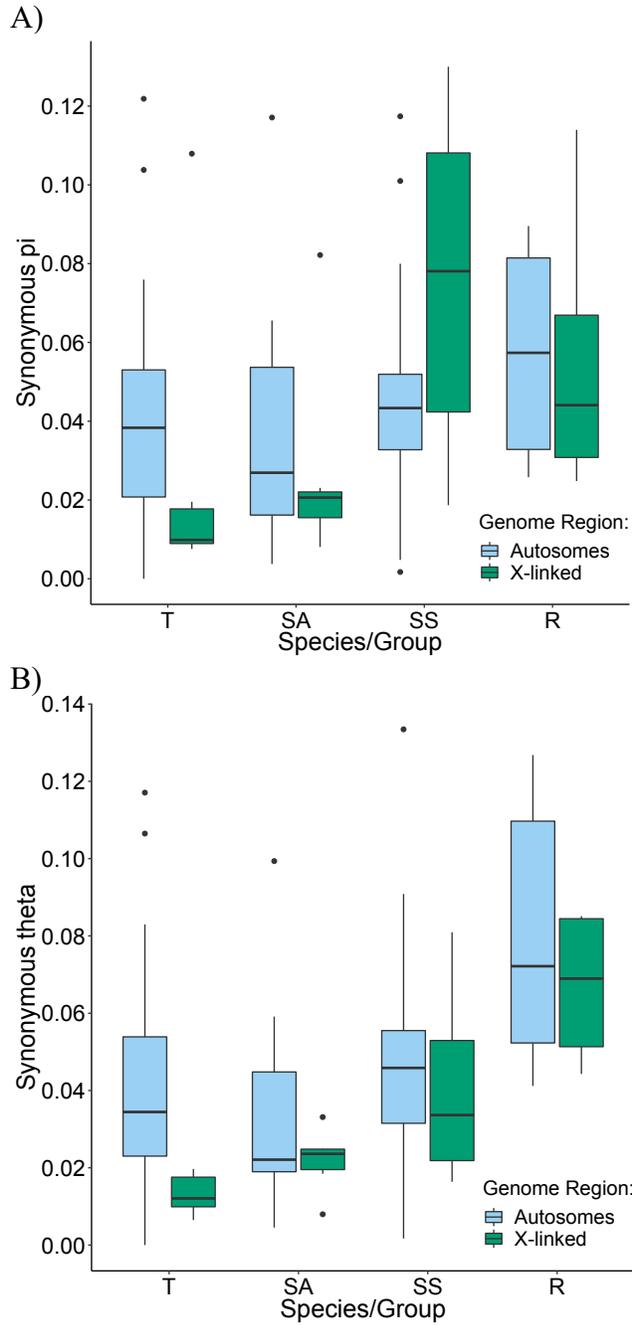


Figure S2.6.: Synonymous nucleotide polymorphism showing A) synonymous π (π_{syn}) and B) synonymous θ (θ_{syn}). Boxplots are shown across loci for each genome region (X-chromosome and autosomes) and by group for *D. transversa* (T), allopatric *D. subquinaria* (SA), sympatric *D. subquinaria* (SS) and *D. recens* (R).

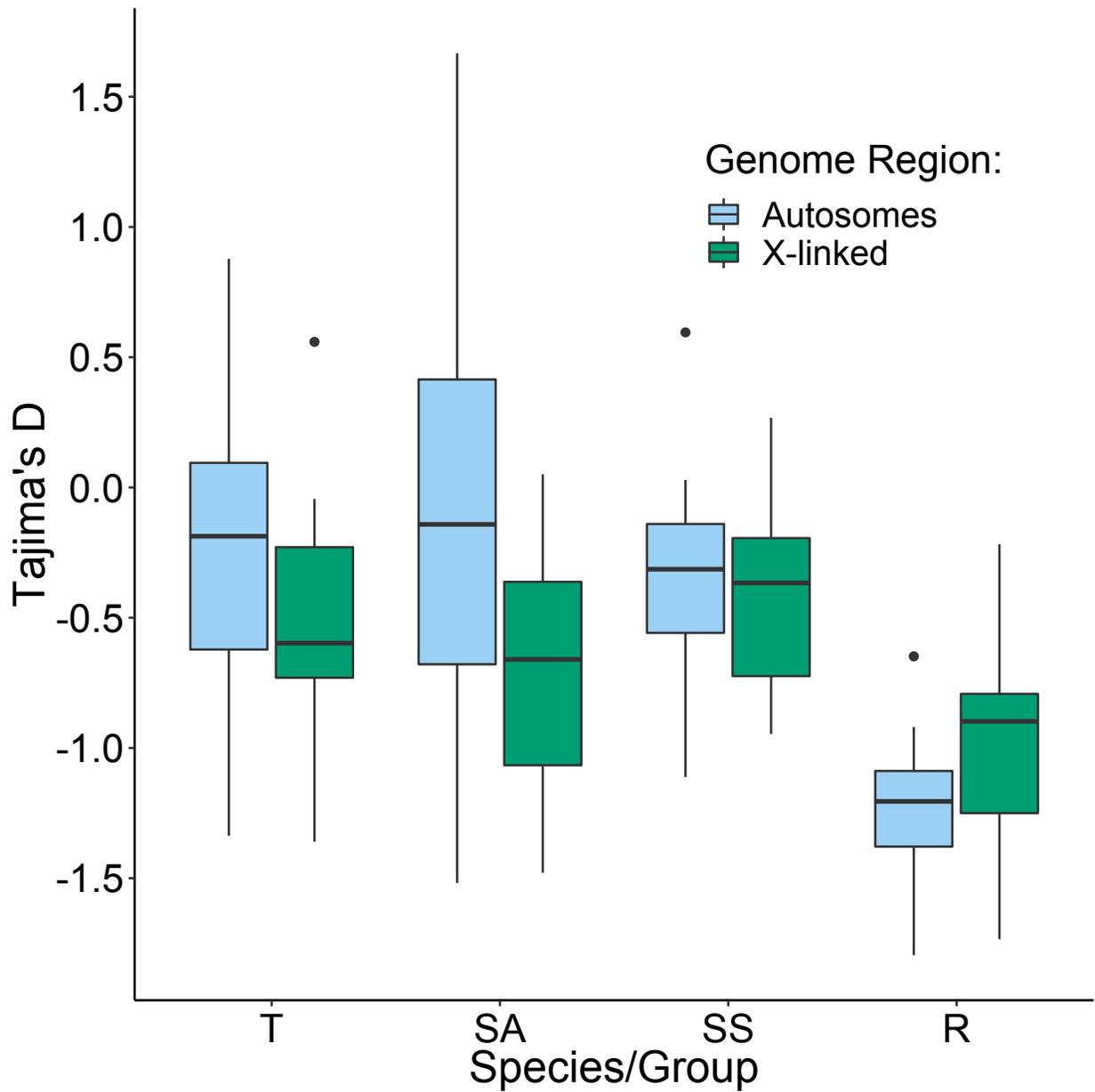


Figure S2.7: Tajima's D for synonymous sites. Boxplots are shown across loci for each genome region (X-chromosome and autosomes) and by group for *D. transversa* (T), allopatric *D. subquinaria* (SA), sympatric *D. subquinaria* (SS) and *D. recens* (R).

Table S2.1: Loci used in this study. Shown are the full locus name, abbreviation, Flybase ID, genome location (Müller element), sequence length, and primers used to amplify and sequence each.

Locus name	Abbreviation	Flybase ID	Genomic location	Total length (bp)	Coding length (bp)	Noncoding length (bp)	Forward primer	Forward primer sequence (5'-3')	Reverse primer	Reverse primer sequence (5'-3')
abdominal-A	abda	FBgn000014	E	551	444	107	abda_f3	GAGTTCCACTTCAA CCACTACT	abdA_r 3	CGCTCATCATTCCCAGA TTG
adh-related	adhr	FBgn000056	B	555	495	60	adhr-D1	ATGGTNCARATGGA YTAYAT	adhr-D4	RTCNGCCATRTGCCART A
bric a brac 2	bab2	FBgn0025525	D	330	330	0	bab2_F4	CCAACCARCAGAAT CTGGTG	bab2_R 3	CTGTCCATTGGCRTACT GATG
black	black	FBgn0000153	B	423	423	0	black_F3	GTCACATGGAATCC ACACAAG	black_R 1	TGATCATCATGGAGCC CTTC
bent	bt	FBgn0005666	F	861	861	0	bt-F4	CYGAYTGGGAYAAR GATCACATYG	bt_R3	AYTCRTATTTTRACRCCY TCAATCAAGTC
cytochrome oxidase subunit I	COI	FBgn0013674	mt	1485	1485	0	TY-J-1460	TACAATCTATCGCCT AAACTTCAGCC	C1-J-2195	TTGATTTTTTGGTCACC CTGAAGT
chorion protein 36	cp36	FBgn0000359	A (X)	769	654	115	cp36_F	TGCAACTYGGTCTC TGGTTTG	cp36_R	TGAGGCTGGCTGTAGA CG
desaturase 2	desat2	FBgn0043043	E	447	374	73	desat2_F1	CGCTCCTACAARGC YAAGTG	desat2_R2	TCACCAGCCAGGTCAT GTTC
doublesex	dsx	FBgn0000504	E	177	177	0	dsx_F1	CRGAGGAGAAAYTGG AAAYAGCGA	dsx_R1	GTCARYCGGCACTTBTC GCA
ebony	eby	FBgn0000527	E	615	615	0	ebony_f1	CAGCTATCGCCAGA TGAACG	ebony_R4	CCCTCCATCTTCAGRTA CATG
embryonic lethal abnormal vision	elav	FBgn00260400	A (X)	1434	1320	114	elav-F	GGCYTTGTTGGTCTT GAAGC	elav_R	GAYACACAGGCRCAGC TAATG
extra sex combs	esc	FBgn0000588	B	369	369	0	esc_F1	GGCCATCAACGAGC TGAARTTYCAYCC	esc_R1	TTCCAGCACACGATGG CRTTYTCRCA
fruitless	fru	FBgn0004652	E	345	345	0	fru_F1	GCCYTGAAGAGCGA RTACCA	fru_R2	GAGGAGTTSAGCTTGA GCAG
hunchback	hb	FBgn0001180	E	630	630	0	hb_F1	GAGCAGCACAAAYGC NTGGTA	hb_R1	GGCCATGTACTTCATRT CYTC
intersex	ix	FBgn0001276	C	375	375	0	ix_F2	GACAACATWTCGAG RGTGAA	ix_R1	ATCGGCCTGWGATATA TTTTGGG
male fertility factor kl3	kl3	FBgn00267432	Y	1389	1335	54	kl3_F11	GGWAGCGTTGARCT TTGG	kl3_R1 1	CRTGSCGCACCAGTGAT G

male fertility factor kl5	kl5	FBgn0267433	Y	1173	1173	0	kl5_F41	ATGAARACRGARTC WCAAGARATG	kl5_R4 3	TGTCTCTRGCYTCSGCY AAAGATTC
mago nashi	mago	FBgn0002736	C	387	300	87	mago_F1	CCACAAGGGCAAGT TCGGNCAYGARTT	mago_R1	CACTTCAGGTCCCTGCAC CARRTARTARAA
males absent on the first neighbor of tid	mof	FBgn0014340	A (X)	639	639	0	mof_F	CAGAAGCGRCGCTA CGA	mof_R	TAKGCCCAATAGCTGC GATA
period	per	FBgn0003068	A (X)	705	705	0	per_F	GGGCCGCATCTTCG ARCAYAARTGG	per_R	TGGAGGGGTAGGTGTT CCARCARTA
plexin A	plexA	FBgn0025741	F	1104	1104	0	plexA_F7	ACAAGGAGAAGTCC AGGAAGAAG	plexA_R3	GAACGTCAACCCAGG CGGAAGG
ribosomal protein L36	rpl36	FBgn0002579	A (X)	348	261	87	rpl_F2	CMRVGSCCACAAGA CCWCSAARRTC	rpl_R2	GAACGTCAACCCAGG CGGAAGG
scully	scu	FBgn0021765	A (X)	669	669	0	scu_F	CMRVTGSCCACAAGA CCWCSAARRTC	scu_R	CRTGRTCTGRGCCTTC C
seven in absentia	sina	FBgn0003410	D	396	396	0	sia_F1	TCGTGTMAAGCAGG TGAACA	sia_R1	TGRAAATGTTTCGTACG GYTC
silver	svr	FBgn0004648	A (X)	724	627	97	sxl_F	TCGAGTGCCCCGTG TGYTTYGAYTA	sxl_R	GAAGTGGAAAGCCGAAG CAGSWYTCATCAT
timeless	tim	FBgn0014396	B	540	540	0	tim_F12	GTCGCGTCGYGGTT ACAATGAT	tim_R1 5	TACAGCTCGCGATCRGT CAT
triose phosphate isomerase	tpi	FBgn0086355	E	381	381	0	tpi_F	AGGATCAAACATTG CGAACC	tpi_R	CCCACATGCTGATCCTT GTAG
wee1 kinase	wee	FBgn0011737	B	294	294	0	wee_F1	CAACTGGAAGATGA AYGGIGACC	wee_R1	TTCTTGGCATAGGCGCA CATYTG
								GCCTGGGCCGAGGA YGAYCAYATG		TCACGTGGCCCAGGTC NCCDATYTT

Table S2.2: ABBA-BABA sites per locus.

Locus	ABBA	BABA	D
<i>abda</i>	0	1	-1
<i>COI</i>	5	4	0.1111
<i>adhr</i>	0	1	-1
<i>bab2</i>	0	0	0
<i>bt</i>	1	0	1
<i>cp36</i>	0	0	0
<i>desat2</i>	0	0	0
<i>dsx</i>	0	0	0
<i>eby</i>	0	1	0
<i>elav</i>	6	2	0.5
<i>esc</i>	6	0	1
<i>fru</i>	1	0	1
<i>hb</i>	1	0	1
<i>ix</i>	2	0	1
<i>kl3</i>	0	0	0
<i>kl5</i>	0	0	0
<i>mago</i>	0	0	0
<i>mof</i>	2	5	-0.4286
<i>ntid</i>	1	0	1
<i>per</i>	0	0	0
<i>plexA</i>	NA	NA	NA
<i>rpl36</i>	5	0	1
<i>scu</i>	NA	NA	NA
<i>sina</i>	0	0	0
<i>svr</i>	1	0	1
<i>tim</i>	1	0	1
<i>tpi</i>	0	0	0
<i>wee</i>	1	0	0
Total (w/o COI)	28	10	0.4737

Table S2.3: Sampling totals for obtained and missing sequence data. A "1" indicates that locus is present in the dataset, whereas a "0" indicates it is missing. Q: *D. quinaria*, SU: *D. suboccidentalis*, S: *D. subquinaria*, T: *D. transversa*, R: *D. recens*. Column headers: Sp.: species, Po.: population, Iso-line number: isofemale line identifier, A: abda, B: adhr, C: bab2, D: black, E: bt, F: COI, G: cp36, H: desat2, I: dsx, J: eby, K: elav, L: esc, M: fru, N: hb, O: ix, P: kl3, Q: kl5, R: mago, S: mof, T: ntid, U: per, V: plexA, W: rpl36, X: scu, Y: sina, Z: svr, AA: tim, AB: tpi, AC: wee, AD: total.

Sp.	Po.	Iso-line number	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R	S	T	U	V	W	X	Y	Z	A	A	A	A	
Q			1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	2
SU			1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	2
R	CA	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
R	CA	5	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	2
R	SM	15	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	2
R	SM	18	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	0	1	0	1	1	1	1	1	1	1	6
R	SM	2	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	7
R	SM	28	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	0	1	1	1	1	1	1	1	1	1	2
R	HI	3	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	0	0	1	1	1	1	1	1	1	1	1	1	1	1	1	6
R	HI	33	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	2
R	HI	45	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	9
R	KA	142	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	2
R	KA	174	1	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	8
R	KA	196	1	1	1	1	1	1	1	1	1	1	0	1	0	1	1	1	1	1	1	1	1	1	1	1	0	1	1	1	1	1	2
																																	6

Table S2.4: Divergence and differentiation for each locus and group comparison. Groups are abbreviated R = *D. recens*, SA = allopatric *D. subquinaria*, SS = sympatric *D. subquinaria*, T = *D. transversa*, RA = allopatric *D. recens*, and RS = sympatric *D. recens*. For Kst, bold values are significant at $0.001 < P < 0.01$, and bold italic values are significant at $P > 0.001$. As explained in text, plexA was highly divergent relative to other loci and was excluded in analyses. SP: Shared polymorphisms.

Comparison:				R - SA						R - SS					
Locus	Location	Genome region	Length (bp)	Fixed diffs	SP	Da	Kst	Ka	Ks	Fixed diffs	SP	Da	Kst	Ka	Ks
abda	E	A	446	0	4	0.00324	0.23506	0.00053	0.04353	0	4	0.00067	0.04052	0.00034	0.03188
adhr	B	A	495	0	14	0.00785	0.17164	0.0013	0.10388	0	16	0.00662	0.1298	0.00094	0.10475
bab2	D	A	330	0	9	0.00203	0.0761	0.00081	0.05666	0	7	0.00007	0.00279	0	0.04405
black	B	A	425	0	11	0.00219	0.06483	0.00249	0.06869	0	12	0.0009	0.02656	0.0021	0.06111
bt	F	A	862	6	0	0.00901	0.46758	0.00104	0.05429	1	0	0.00308	0.22429	0.00027	0.02922
COI	mt	mt	1487	0	12	0.02145	0.57255	0.00076	0.1282	24	5	0.02668	0.71842	0.00088	0.13819
cp36	A (X)	X	655	3	2	0.00734	0.3642	0.00484	0.03831	2	1	0.00546	0.29421	0.0049	0.02802
desat2	E	A	377	0	3	0.00735	0.23445	0.00652	0.06073	0	12	0.00248	0.06558	0.00525	0.06236
dsx	E	A	177	0	0	0.00156	0.10624	0	0.03307	0	0	0.00146	0.09552	0	0.02749
eby	E	A	617	1	18	0.00728	0.15034	0.00278	0.10289	0	24	0.00209	0.03788	0.00272	0.10124
elav	A (X)	X	1320	20	0	0.02355	0.39276	0.00023	0.12234	9	16	0.0151	0.28732	0.00016	0.10302
esc	B	A	369	0	13	0.01104	0.1987	0.00228	0.14665	0	18	-0.00024	-0.00464	0.00063	0.1023
fru	E	A	345	0	8	0.00748	0.26137	0.0109	0.03852	0	6	0.00229	0.0799	0.00636	0.03772
hb	E	A	630	0	2	0.01442	0.35878	0.00234	0.11803	0	21	0.00261	0.05611	0.00279	0.09536
ix	C	A	375	0	7	0.00761	0.20305	0.00041	0.0952	0	6	0.00598	0.12412	0.00041	0.08741
kl3	Y	Y	1336	15	0	0.01123	1	0.00195	0.04233	15	0	0.01177	0.97787	0.00198	0.0452
kl5	Y	Y	1175	12	0	0.01138	0.94256	0.00123	0.04971	12	0	0.01138	0.93732	0.00114	0.0497
mago	C	A	300	1	2	0.00938	0.40024	0.00054	0.07158	1	2	0.00873	0.35349	0.00054	0.07289
mof	A (X)	X	639	7	1	0.02038	0.49037	0.00203	0.12722	2	11	0.01524	0.29314	0.00203	0.13324
ntid	C	A	573	2	2	0.0119	0.20462	0.01714	0.07842	1	10	0.00858	0.10897	0.01429	0.09038
per	A (X)	X	705	0	57	0.01834	0.22852	0.01596	0.15818	0	57	0.01715	0.19098	0.0174	0.15638

plexA	F	NA	1106	80	0	0.10847	0.67604	0.0137	0.54775	79	0	0.10779	0.62835	0.01401	0.54563
rpl36	A (X)	X	262	4	1	0.03315	0.69289	0.00171	0.16304	0	8	0.01199	0.25497	0.00116	0.12894
scu	A (X)	X	669	3	4	0.02043	0.54255	0.00611	0.08964	2	4	0.01093	0.35199	0.00418	0.05904
sina	D	A	396	0	0	0.00921	0.50832	0	0.05509	0	0	0.00906	0.42975	0	0.05793
svr	A (X)	X	628	9	6	0.02445	0.5328	0.00657	0.12253	5	14	0.01703	0.335	0.00573	0.11588
tim	B	A	542	1	14	0.00791	0.18986	0.00014	0.09797	1	12	0.00641	0.14228	0.00013	0.08755
tpi	E	A	381	0	4	0.00256	0.11978	0.00105	0.0437	0	7	0.00025	0.01115	0.00035	0.0387
wee	B	A	294	0	1	0.00884	0.41202	0.00051	0.06767	0	3	0.00395	0.18036	0.00022	0.05819
	Average	autosomal		0.61	6.22	0.00727	0.24239	0.00282	0.07425	0.22	8.89	0.00361	0.11691	0.00207	0.06614
	Average	X-linked		6.57	10.14	0.02109	0.46344	0.00535	0.11732	2.86	15.86	0.01327	0.28680	0.00508	0.10350
	Average	A + X		2.28	7.32	0.01114	0.30428	0.00353	0.08631	0.96	10.84	0.00632	0.16448	0.00292	0.07660
	Sum	autosomal		11	112					4	160				
	Sum	X-linked		46	71					20	111				
	Sum	A + X		57	183					24	271				

Comparison:				R – T						SA – SS					
Locus	Location	Genome region	Length (bp)	Fixed diffs	SP	Da	Kst	Ka	Ks	Fixed diffs	SP	Da	Kst	Ka	Ks
abda	E	A	446	0	0	0.00563	0.31078	0.00033	0.0427	0	2	0.00327	0.24439	0.0002	0.04573
adhr	B	A	495	0	23	0.00608	0.09765	0.00094	0.11679	0	15	0.0055	0.13853	0.00039	0.09169
bab2	D	A	330	1	3	0.00844	0.25424	0.00405	0.0592	0	6	0.00094	0.03287	0.00081	0.0552
black	B	A	425	0	5	0.00274	0.06303	0.00307	0.07018	0	8	0.00056	0.02832	0.0012	0.04499
bt	F	A	862	3	2	0.00571	0.25197	0.00027	0.05116	9	0	0.01172	0.81774	0.00077	0.05072
COI	mt	mt	1487	45	0	0.03487	0.90451	0.00176	0.14835	0	40	0.00443	0.13482	0.00012	0.07701
cp36	A (X)	X	655	3	0	0.01028	0.41186	0.00468	0.04398	1	2	0.00296	0.24078	0.00036	0.02792
desat2	E	A	377	0	3	0.01161	0.22177	0.00652	0.09627	0	3	0.0029	0.15651	0.0024	0.04384
dsx	E	A	177	0	0	0.00146	0.08059	0	0.0251	0	0	0.00011	0.02041	0	0.01036
eby	E	A	617	0	9	0.00581	0.10003	0.00216	0.09244	0	28	0.00275	0.06014	0.00246	0.09434
elav	A (X)	X	1320	20	3	0.02387	0.33373	0.00016	0.12428	2	2	0.00638	0.32159	0.00007	0.04944
esc	B	A	369	0	11	0.00602	0.09889	0.00069	0.13287	0	17	0.00964	0.14957	0.00199	0.15637
fru	E	A	345	0	9	0.00183	0.04693	0.00804	0.03768	0	5	0.0041	0.20183	0.00889	0.02385
hb	E	A	630	0	13	0.0082	0.15284	0.00202	0.10477	0	4	0.00675	0.26537	0.00218	0.07841
ix	C	A	375	1	1	0.01139	0.22719	0.00041	0.09926	0	6	0.00117	0.04903	0	0.04769
kl3	Y	Y	1336	14	0	0.01048	1	0.00195	0.03906	0	0	0.00036	0.54167	0	0.00231
kl5	Y	Y	1175	11	0	0.01048	0.90045	0.00114	0.04556	0	0	0	0	0.00008	0
mago	C	A	300	3	0	0.01322	0.36336	0.00054	0.08172	0	5	-0.00016	-0.01026	0	0.03681
mof	A (X)	X	639	7	0	0.02155	0.41433	0.00203	0.12691	3	1	0.0134	0.40626	0	0.10832
ntid	C	A	573	1	7	0.00994	0.11368	0.01399	0.09354	0	10	0.00269	0.09297	0.0101	0.03748
per	A (X)	X	705	0	53	0.01279	0.12697	0.01677	0.14885	0	46	0.00023	0.00444	0.00765	0.08836
plexA	F	NA	1106	80	2	0.11116	0.56097	0.0165	0.55522	0	99	0.00173	0.10476	0.00024	0.00376
rpl36	A (X)	X	262	5	0	0.03386	0.6585	0	0.1635	0	3	0.01013	0.24513	0.00111	0.1144
scu	A (X)	X	669	6	1	0.01839	0.34293	0.00668	0.08004	0	2	0.01268	0.5126	0.00318	0.0644
sina	D	A	396	0	1	0.00927	0.3977	0	0.05744	0	2	-0.00007	-0.01446	0	0.00993

svr	A (X)	X	628	9	3	0.02645	0.43939	0.00764	0.13044	1	6	0.00587	0.28765	0.00258	0.05123
tim	B	A	542	1	11	0.01203	0.20746	0.00019	0.11908	0	16	0.00204	0.06984	0	0.06558
tpi	E	A	381	0	4	0.00411	0.12592	0.00152	0.05054	0	4	0.00056	0.03601	0.0007	0.03042
wee	B	A	294	0	2	0.00728	0.23407	0.00022	0.087	0	2	0.002	0.11702	0.00032	0.04512
	Average	autosomal		0.56	5.78	0.00727	0.18601	0.00250	0.07876	0.50	7.39	0.00314	0.13644	0.00180	0.05381
	Average	X-linked		7.14	8.57	0.02103	0.38967	0.00542	0.11686	1.00	8.86	0.00738	0.28835	0.00214	0.07201
	Average	A + X		2.40	6.56	0.01112	0.24303	0.00332	0.08943	0.64	7.80	0.00432	0.17897	0.00189	0.05890
	Sum	autosomal		10	104					9	133				
	Sum	X-linked		50	60					7	62				
	Sum	A + X		60	164					16	195				

Comparison:				SA - T						SS - T					
Locus	Location	Genome region	Length (bp)	Fixed diffs	SP	Da	Kst	Ka	Ks	Fixed diffs	SP	Da	Kst	Ka	Ks
abda	E	A	446	0	0	0.00377	0.24248	0.0002	0.03555	1	0	0.00716	0.31841	0.00034	0.03045
adhr	B	A	495	0	16	0.00466	0.09947	0.00038	0.10254	0	16	0.00215	0.04904	0	0.09833
bab2	D	A	330	1	3	0.00616	0.175	0.00486	0.0578	1	2	0.00786	0.27906	0.0041	0.05962
black	B	A	425	0	5	0.00069	0.03068	0.00215	0.04606	0	3	0.00122	0.05899	0.00228	0.04513
bt	F	A	862	2	1	0.0051	0.42409	0.00077	0.03437	5	0	0.00859	0.62891	0	0.04831
COI	mt	mt	1487	7	1	0.01424	0.36196	0.001	0.09412	13	2	0.01785	0.58085	0.00088	0.09793
cp36	A (X)	X	655	2	0	0.00605	0.39491	0.00015	0.03731	1	0	0.00363	0.37657	0.00021	0.02441
desat2	E	A	377	0	2	0.00712	0.32601	0	0.06847	0	3	0.00854	0.26202	0.0024	0.0857
dsx	E	A	177	0	0	0.00011	0.0198	0	0.00797	0	0	0	0	0	0.00239
eby	E	A	617	0	9	0.00631	0.14678	0.00212	0.08423	0	13	0.00343	0.07415	0.00223	0.09132
elav	A (X)	X	1320	2	0	0.00326	0.26767	0.00007	0.02218	3	4	0.00684	0.26307	0	0.05206
esc	B	A	369	0	14	0.00297	0.04885	0.00257	0.12614	0	15	0.00529	0.09507	0.00093	0.14379
fru	E	A	345	0	7	0.00099	0.0458	0.00664	0.02484	0	6	0.00006	0.00284	0.00616	0.02868
hb	E	A	630	0	3	0.00387	0.22654	0.00149	0.04725	0	12	0.00104	0.04176	0.00193	0.06686
ix	C	A	375	0	1	0.00072	0.03661	0	0.03436	1	0	0.00356	0.21429	0	0.04573
kl3	Y	Y	1336	3	0	0.00225	1	0	0.00978	3	0	0.00264	0.85597	0	0.01222
kl5	Y	Y	1175	1	0	0.00085	0.79545	0.00008	0.00395	1	0	0.00085	1	0	0.00395
mago	C	A	300	4	0	0.01333	0.14817	0	0.04203	0	1	0.00407	0.15942	0	0.04887
mof	A (X)	X	639	3	1	0.00653	0.42343	0	0.04399	5	0	0.0161	0.43498	0	0.11482
ntid	C	A	573	0	4	0.00564	0.17844	0.01089	0.04282	0	12	0.00076	0.02216	0.00807	0.0488
per	A (X)	X	705	0	46	0.00129	0.02195	0.00774	0.10093	0	44	-0.00041	-0.00807	0.00875	0.09835
plexA	F	NA	1106	0	0	0.0035	0.60356	0.00275	0.01155	0	0	0.00344	0.54028	0.00299	0.01168
rpl36	A (X)	X	262	0	0	0.00098	0.0931	0	0.01941	0	1	0.01002	0.2152	0.0005	0.10788
scu	A (X)	X	669	0	3	0.00292	0.17906	0.00149	0.0321	6	1	0.01262	0.22036	0.00423	0.06282
sina	D	A	396	0	1	0.0001	0.02186	0	0.00923	0	1	0.00009	0.321	0	0.01265

svr	A (X)	X	628	0	2	0.00478	0.3777	0.00178	0.03048	0	2	0.00443	0.18863	0.0038	0.04437
tim	B	A	542	0	11	0.00515	0.13952	0	0.08523	1	8	0.00977	0.27357	0	0.10149
tpi	E	A	381	0	2	0.00172	0.09155	0.00187	0.03149	0	3	0.00233	0.12608	0.00117	0.03815
wee	B	A	294	0	2	0.00314	0.13739	0.00029	0.06397	0	4	0.00143	0.05314	0	0.0691
	Average	autosomal		0.39	4.50	0.00398	0.14106	0.00190	0.05246	0.50	5.50	0.00374	0.16555	0.00165	0.05919
	Average	X-linked		1.00	7.43	0.00369	0.25112	0.00160	0.04091	2.14	7.43	0.00760	0.24153	0.00250	0.07210
	Average	A + X		0.56	5.32	0.00389	0.17187	0.00182	0.04923	0.96	6.04	0.00482	0.18683	0.00188	0.06280
	Sum	autosomal		7	81					9	99				
	Sum	X-linked		7	52					15	52				
	Sum	A + X		14	133					24	151				

Comparison:				RA-RS				
Locus	Location	Genome region	Length (bp)	Fixed diffs	SP	Da	Kst	Kst P value
abda	E	A	446	0	3	0.00059	0.03245	0.12
adhr	B	A	495	0	16	0.00033	0.00815	0.223
bab2	D	A	330	0	8	-0.00013	-0.00654	0.562
black	B	A	425	0	11	0.00055	0.01597	0.124
bt	F	A	862	0	5	-0.00005	-0.00307	0.524
COI	mt	mt	1487	0	2	0.00022	0.06406	0.049
cp36	A (X)	X	655	0	4	0.00008	0.00581	0.322
desat2	E	A	377	0	9	0.00052	0.01536	0.18
dsx	E	A	177	0	3	0.00138	0.08121	0.042
eby	E	A	617	0	22	0.00032	0.00732	0.249
elav	A (X)	X	1320	0	16	-0.00015	-0.00754	0.599
esc	B	A	369	0	13	-0.00053	-0.01406	0.839
fru	E	A	345	0	4	0.00249	0.09933	0.001
hb	E	A	630	0	18	0.0007	0.01376	0.169
ix	C	A	375	0	10	0.00026	0.0069	0.312
kl3	Y	Y	1336	0	0	0	na	na
kl5	Y	Y	1175	0	0	0.00015	0.15	0.202
mago	C	A	300	0	2	0.00042	0.03226	0.213
mof	A (X)	X	639	0	14	0.00199	0.06692	0.002
ntid	C	A	573	0	21	-0.00062	-0.01054	0.754
per	A (X)	X	705	0	49	0.00635	0.09606	0.013
plexA	F	NA	1106	0	99	0.00173	0.01552	0.103
rpl36	A (X)	X	262	0	5	-0.00041	-0.0243	0.907
scu	A (X)	X	669	0	8	0.00004	0.00307	0.368
sina	D	A	396	0	6	-0.00014	-0.01113	0.647

svr	A (X)	X	628	0	21	-0.00024	-0.00753	0.681
tim	B	A	542	0	19	0.00035	0.00965	0.207
tpi	E	A	381	0	6	-0.00039	-0.01809	0.968
wee	B	A	294	0	4	-0.00039	-0.02876	0.99
	Average	autosomal		0	10.00	0.00031	0.01279	
	Average	X-linked		0	16.71	0.00109	0.01893	
	Average	A + X		0	11.88	0.00053	0.01451	
	Sum	autosomal		0	180			
	Sum	X-linked		0	117			
	Sum	A + X		0	297			

Table S2.5: Fixed differences and shared polymorphisms among groups. Groups are abbreviated R = *D. recens*, SA = allopatric *D. subquinaria*, SS = sympatric *D. subquinaria*, and T = *D. transversa*.

Comparison	Genomic Region	Fixed Differences	Shared polymorphisms	% Fixed
R – T	A	10	104	8.8
	X	50	60	45.5
	Total	60	164	26.8
R – SA	A	11	112	9
	X	46	71	39.3
	Total	57	183	23.8
R – SS	A	4	160	2.4
	X	20	111	15.3
	Total	24	271	8.1
SS – T	A	9	99	8.3
	X	15	52	22.4
	Total	24	151	13.7
SA – T	A	7	81	8
	X	7	52	11.9
	Total	14	133	9.5
SA – SS	A	9	133	6.3
	X	7	62	10.1
	Total	16	195	7.6

Table S2.6: Chi-square tests comparing the number of fixed differences and polymorphic sites between group comparisons. Groups are abbreviated R = *D. recens*, SA = allopatric *D. Subqiunaria*, SS = sympatric *D. subquinaria*, and T = *D. transversa*. Bold values are significant at $P < 0.05$ and bold italic are significant with a Bonferroni correction ($P < 0.003$).

Group 1	Group 2	df	χ^2	<i>P</i>-value
R-T	R-SA	1	0.42	0.52
R-T	R-SS	1	31.3	<i>2.23E-08</i>
R-T	SA-T	1	15.5	<i>8.25E-05</i>
R-T	SS-T	1	9.3	<i>0.002256</i>
R-T	SA-SS	1	26.5	<i>2.68E-07</i>
R-SA	R-SS	1	23.9	<i>1.01E-06</i>
R-SA	SA-T	1	11.4	<i>7.41E-04</i>
R-SA	SS-T	1	5.9	0.01544
R-SA	SA-SS	1	20.5	<i>6.10E-06</i>
R-SS	SA-T	1	0.1	0.7562
R-SS	SS-T	1	3.1	0.07619
R-SS	SA-SS	1	0.004	0.9521
SA-T	SS-T	1	0.97	0.3234
SA-T	SA-SS	1	0.2	0.6469
SS-T	SA-SS	1	3.2	0.07188
All Comparisons		5	62.2	<i>4.30E-12</i>

Table S2.7: Summary of nucleotide polymorphism and frequency spectrum for each locus and group. For Tajima's D, a bold value indicates significance at $P < 0.05$ using 1000 coalescent simulations. Groups are abbreviated R = *D. recens*, SA = allopatric *D. subquinaria*, SS = sympatric *D. subquinaria*, and T = *D. transversa*. Table is continued on following page.

Locus	Location	Region	Group	length bp	n alleles	all seg sites	syn seg sites	nonsyn seg sites
abda	E	A	R	446	18	14	12	2
			SA	446	15	7	6	1
			SS	446	10	9	9	0
			T	446	6	1	1	0
adhr	B	A	R	495	20	61	55	6
			SA	495	15	24	22	2
			SS	495	10	27	27	0
			T	495	7	35	35	0
bab2	D	A	R	330	20	19	19	0
			SA	330	15	14	13	1
			SS	330	10	12	12	0
			T	330	7	7	7	0
black	B	A	R	425	20	44	40	4
			SA	425	12	12	10	2
			SS	425	10	15	14	1
			T	425	7	11	9	2
bt	F	A	R	862	17	41	38	3
			SA	862	14	6	3	3
			SS	862	10	1	1	0
			T	862	7	12	12	0
COI	mt	mt	R	1487	20	17	17	0
			SA	1487	15	100	99	1
			SS	1487	10	61	61	0

			T	1487	7	3	3	0
cp36	A (X)	X	R	655	20	27	25	2
			SA	655	14	13	12	1
			SS	655	10	9	8	1
			T	655	7	4	0	0
desat2	E	A	R	377	20	36	32	4
			SA	377	15	6	6	0
			SS	377	10	14	11	3
			T	377	7	11	11	0
dsx	E	A	R	177	20	7	0	0
			SA	177	15	3	3	0
			SS	177	10	1	1	0
			T	177	7	0	0	0
eby	E	A	R	617	20	71	65	6
			SA	617	15	38	34	4
			SS	617	10	52	48	4
			T	617	7	25	23	2
elav	A (X)	X	R	1320	13	55	53	2
			SA	1320	14	9	8	1
			SS	1320	10	39	39	0
			T	1320	7	10	10	0
esc	B	A	R	369	20	37	36	1
			SA	369	15	27	26	1
			SS	369	8	29	28	1
			T	369	7	22	21	1
fru	E	A	R	345	18	20	16	4
			SA	345	13	9	5	4
			SS	345	9	9	6	3

			T	345	5	10	6	4
hb	E	A	R	630	20	58	47	11
			SA	630	15	9	6	3
			SS	630	10	37	31	6
			T	630	7	16	14	2
ix	C	A	R	375	17	33	31	2
			SA	375	14	15	15	0
			SS	375	6	9	9	0
			T	375	6	5	5	0
kl3	Y	Y	R	1336	18	0	0	0
			SA	1336	13	0	0	0
			SS	1336	10	1	1	0
			T	1336	6	0	0	0
kl5	Y	Y	R	1175	18	2	2	0
			SA	1175	14	1	0	1
			SS	1175	9	0	0	0
			T	1175	5	0	0	0
mago	C	A	R	300	16	8	8	0
			SA	300	12	7	7	0
			SS	300	10	7	7	0
			T	300	3	2	2	0
mof	A (X)	X	R	639	20	44	44	0
			SA	639	16	12	12	0
			SS	639	9	32	32	0
			T	639	7	4	4	0
ntid	C	A	R	573	18	80	51	29
			SA	573	15	17	8	9
			SS	573	10	31	19	12

			T	573	7	24	16	8
per	A (X)	X	R	705	20	85	67	18
			SA	705	15	64	50	14
			SS	705	10	62	45	17
			T	705	7	60	45	15
plexA	F	NA	R	1106	13	185	172	13
			SA	1106	15	2	2	0
			SS	1106	10	5	3	2
			T	1106	6	6	3	3
rpl36	A (X)	X	R	262	20	12	12	0
			SA	262	15	5	5	0
			SS	262	10	22	21	1
			T	262	7	1	1	0
scu	A (X)	X	R	669	18	25	24	1
			SA	669	15	19	17	2
			SS	669	10	14	11	3
			T	669	7	11	8	3
sina	D	A	R	396	20	14	14	0
			SA	396	15	4	4	0
			SS	396	10	5	5	0
			T	396	7	3	3	0
svr	A (X)	X	R	628	20	49	45	4
			SA	628	15	11	9	2
			SS	628	10	25	21	4
			T	628	7	8	7	1
tim	B	A	R	542	20	52	51	1
			SA	542	15	24	24	0
			SS	542	9	19	19	0

			T	542	7	22	22	0
tpi	E	A	R	381	20	29	27	2
			SA	381	15	8	6	2
			SS	381	10	9	9	0
			T	381	6	6	5	1
wee	B	A	R	294	20	11	10	1
			SA	294	15	6	5	1
			SS	294	10	10	10	0
			T	294	7	13	13	0

Locus	Location	Region	Group	pi all sites	pi silent sites	theta all sites	theta syn sites	Tajima's D all sites	Tajima's D synonymous sites
abda	E	A	R	0.00593	0.0258	0.00951	NA	-1.41843	-1.1262
			SA	0.00658	0.02966	0.00496	0.02009	1.17821	1.66675
			SS	0.00716	0.03382	0.00759	0.03587	-0.24964	-0.24964
			T	0.00123	0.0058	0.00101	0.00476	0.85057	0.85057
adhr	B	A	R	0.02132	0.08067	0.03474	0.1268	-1.56588	-1.46936
			SA	0.01616	0.06346	0.01491	NA	0.34809	0.58989
			SS	0.01854	0.07535	0.01928	0.07836	-0.18343	-0.18343
			T	0.02559	0.1038	0.02886	0.11707	-0.64965	-0.64965
bab2	D	A	R	0.0103	0.04095	0.01623	0.0645	-1.3809	-1.3809
			SA	0.01532	0.05691	0.01305	0.04826	0.69189	0.70439
			SS	0.01172	0.04666	0.01285	0.05119	-0.39947	-0.39947
			T	0.00866	0.03444	0.0086	0.03444	0	0
black	B	A	R	0.01809	0.07548	0.02918	0.12525	-1.52079	-1.58306
			SA	0.01055	0.04401	0.01169	NA	-0.22845	-0.02921
			SS	0.00931	0.04164	0.01248	0.05486	-1.16988	-1.10431
			T	0.01098	0.04222	0.01056	0.04071	0.21306	0.19685
bt	F	A	R	0.00795	0.03123	0.01407	0.05401	-1.8005	-1.73966
			SA	0.00198	0.00375	0.00255	0.00453	-0.62078	-0.52939
			SS	0.00041	0.00171	0.00041	0.0017	0.01499	0.01499
			T	0.00575	0.02379	0.00568	0.02353	0.0605	0.0605
COI	mt	mt	R	0.00178	0.00761	0.00322	0.01377	-1.67399	-1.67399
			SA	0.0167	0.06971	0.02068	0.08757	-0.89333	-0.89357
			SS	0.01086	0.04645	0.0145	0.062	-1.23156	-1.23156
			T	0.00058	0.00247	0.00082	0.00353	-1.35841	-1.35841
cp36	A (X)	X	R	0.00712	0.02481	0.0133	NA	-1.82157	-1.7341

			SA	0.00537	0.01969	0.00642	0.02269	-0.6589	-0.52634
			SS	0.00373	0.13	0.00486	0.01638	-1.01429	-0.89013
			T	0.00218	0.00827	0.00249	0.00946	-0.59756	-0.59756
desat2	E	A	R	0.01787	0.07172	0.02692	0.108	-1.33168	-1.32218
			SA	0.0044	0.01989	0.00489	0.02215	-0.3569	-0.3569
			SS	0.01291	0.04479	0.01313	0.04665	-0.07601	-0.17851
			T	0.01162	0.05257	0.01191	0.05388	-0.13112	-0.13112
dsx	E	A	R	0.00895	0.03787	0.01115	0.04717	-0.64802	-0.64802
			SA	0.00355	0.01503	0.00521	0.02205	-0.94808	-0.94808
			SS	0.00113	0.00478	0.002	0.00845	-1.11173	-1.11173
			T	0	0	0	0	NA	NA
eby	E	A	R	0.02306	0.08458	0.03518	NA	-1.50127	-1.37131
			SA	0.01763	0.06556	0.02044	NA	-0.55911	-0.43189
			SS	0.02723	0.10099	0.02979	NA	-0.42087	-0.36144
			T	0.01459	0.05317	0.01654	0.0613	-0.66888	-0.74918
elav	A (X)	X	R	0.01101	0.04408	0.01361	0.05368	-0.85929	-0.80224
			SA	0.00209	0.00808	0.0022	0.00797	-0.18681	0.05072
			SS	0.00944	0.03858	0.01066	0.0436	-0.55801	-0.55801
			T	0.00236	0.00962	0.00316	0.01289	-1.35933	-1.35933
esc	B	A	R	0.01995	0.08955	0.02826	NA	-1.16655	-1.12617
			SA	0.02576	0.1171	0.0225	0.09936	0.605	0.5189
			SS	0.02642	0.1174	0.03031	0.13346	-0.67904	-0.63592
			T	0.02736	0.12186	0.02433	0.10648	0.70058	0.81282
fru	E	A	R	0.01313	0.03763	0.01731	0.05711	-0.94084	-1.30206
			SA	0.00699	0.015	0.00841	0.01885	-0.66553	-0.72839
			SS	0.00998	0.02612	0.0096	0.02595	0.18487	0.02885
			T	0.01333	0.03274	0.01391	0.03367	-0.29817	-0.19092
hb	E	A	R	0.02016	0.08573	0.03032	NA	-1.3448	-1.00955

			SA	0.00363	0.0119	0.00476	0.01538	-0.88546	-0.79055
			SS	0.01873	0.07997	0.02259	0.09086	-0.82719	-0.57642
			T	0.01042	0.04583	0.01116	0.04741	-0.36994	-0.18346
ix	C	A	R	0.01957	0.0815	0.02603	0.10521	-1.01659	-0.91982
			SA	0.01011	0.04343	0.01258	0.05404	-0.80018	-0.80018
			SS	0.00978	0.04188	0.01051	0.04502	-0.41545	-0.41545
			T	0.00444	0.01907	0.00584	0.02506	-1.33698	-1.33698
kl3	Y	Y	R	0	0	0	0	NA	NA
			SA	0	0	0	0	NA	NA
			SS	0.00036	0.00154	0.00027	0.00117	0.8198	0.8198
			T	0	0	0	0	NA	NA
kl5	Y	Y	R	0.00053	0.00244	0.00052	0.00241	0.02839	0.02839
			SA	0.00012	0	0.00027	0	-1.15524	NA
			SS	0	0	0	0	NA	NA
			T	0	0	0	0	NA	NA
mago	C	A	R	0.00689	0.02676	0.01105	NA	-1.07554	-1.07554
			SA	0.00747	0.03328	0.00773	0.0344	-0.12755	-0.12755
			SS	0.00941	0.04179	0.00825	0.03664	0.59539	0.59539
			T	0.00444	0.01975	0.00444	0.01975	NA	NA
mof	A (X)	X	R	0.01562	0.06848	0.01941	0.08511	-0.78199	-0.78199
			SA	0.000468	0.02061	0.00566	0.02491	-0.65989	-0.65989
			SS	0.01778	0.07809	0.01843	0.08093	-0.17632	-0.17632
			T	0.00224	0.00985	0.00256	0.01126	-0.59756	-0.59756
ntid	C	A	R	0.03056	0.0813	0.01466	NA	-1.50867	-1.18567
			SA	0.00983	0.01958	0.01151	NA	-0.47812	-0.11885
			SS	0.01802	0.04767	0.01953	NA	-0.37165	-0.11209
			T	0.01579	0.04415	0.01765	0.04964	-0.59738	-0.61271
per	A (X)	X	R	0.03634	0.11397	0.03734	NA	-0.26003	-0.21781

			SA	0.02396	0.08219	0.02865	NA	-0.71105	-0.67893
			SS	0.02772	0.09244	0.0319	NA	-0.64376	-0.36629
			T	0.03209	0.10792	0.0358	NA	-0.60103	-0.41448
plexA	F	NA	R	0.05708	0.22369	0.06031	NA	-0.18946	-0.06497
			SA	0.00059	0.00252	0.00056	0.00239	0.13916	0.13916
			SS	0.00115	0.00337	0.0016	0.00412	-1.1361	-0.65748
			T	0.00199	0.00391	0.00238	0.00513	-0.93169	-1.23311
rpl36	A (X)	X	R	0.00882	0.03257	0.01291	NA	-1.13759	-1.13759
			SA	0.00553	0.02308	0.00587	0.02452	-0.19766	-0.19766
			SS	0.0307	0.12378	0.02968	NA	0.16278	0.26755
			T	0.00182	0.00756	0.00156	0.00648	0.55902	0.55902
scu	A (X)	X	R	0.00756	0.02903	0.01148	0.04428	-1.35631	-1.36237
			SA	0.00568	0.021	0.00928	0.03312	-1.57934	-1.47901
			SS	0.00608	0.01867	0.00757	0.02368	-0.89865	-0.94624
			T	0.00649	0.01953	0.0068	0.0197	-0.24584	-0.04378
sina	D	A	R	0.00663	0.02741	0.00997	0.04118	-1.22437	-1.22437
			SA	0.00164	0.00675	0.00311	0.01281	-1.51811	-1.51811
			SS	0.00331	0.01365	0.00446	0.0184	-1.03527	-1.03527
			T	0.00265	0.01091	0.00309	0.01276	-0.65405	-0.65405
svr	A (X)	X	R	0.01658	0.06536	0.02199	0.08423	-0.99013	-0.89763
			SA	0.00312	0.01132	0.00539	0.01848	-1.61907	-1.45315
			SS	0.01348	0.0461	0.01407	NA	-0.19958	-0.21234
			T	0.00455	0.01594	0.0052	0.01913	-0.65665	-0.86284
tim	B	A	R	0.01933	0.08207	0.02756	0.11477	-1.17645	-1.14765
			SA	0.01395	0.06058	0.01362	0.05913	0.1017	0.1017
			SS	0.0122	0.05286	0.0129	0.05589	-0.26628	-0.26628
			T	0.01546	0.06447	0.01732	NA	-0.55534	-0.55534
tpi	E	A	R	0.01129	0.043	0.02145	0.07982	-1.85365	-1.79528

			SA	0.00705	0.02418	0.00646	0.01936	0.3373	0.87174
			SS	0.00811	0.0324	0.00835	0.03337	-0.12729	-0.12729
			T	0.00805	0.02664	0.0069	0.02302	0.95607	0.87805
wee	B	A	R	0.00705	0.03091	0.01055	0.04415	-1.17461	-1.04834
			SA	0.00544	0.02298	0.00628	0.02409	-0.46531	-0.15588
			SS	0.01066	0.04909	0.01202	0.05538	-0.50348	-0.50348
			T	0.01652	0.07595	0.01805	0.08297	-0.46339	-0.46339

APPENDIX B

Supplementary Materials Chapter 3

Table S3.1: qPCR primers for *Wolbachia* titer estimation. Start and stop positions refer to within the sequenced used for primer design (See Methods). “Eff.” columns give the primer efficiencies in either *D. subquianria* and *D. recens*.

Organism Target	Locus	bp	start F	start R	forward primer	reverse primer	Tm F	Tm R	GC F	GC R	Eff. <i>D. recens</i>	Eff. <i>D. subquianria</i>
Drosophila	mag0	83	103	185	5'-GTCATGGAG GACTGAAGC – 3'	5'-ACACGATCTG GTGGCGGC-3'	57.9	62.8	57.9	66.7	99.48	97.11
Wolbachia (wRec)	wsp	83	458	540	5'- GCTGGAGCTC GTTATTTCCGG-3'	5'-GCATCAGCA ACCTGTCCGAT- 3'	59.1	60.7	55	55	94.81	99.08

Table S3.2: RNAseq data used for differential expression analysis. “M” means numbers are in millions and “G” means bases are in billions.

Sample	nReads.prefilter(M)	nBases.prefilter(G)	nReads(M)	nBases(G)	Read length
1RTC	17.213	1.301034	16.45794	1.24158	75PE
1RTO	19.468936	1.471357	18.745314	1.414285	75PE
1RWC	16.03374	1.212215	15.158792	1.143918	75PE
1RWO	20.403538	1.542032	19.68113	1.484631	75PE
2RTC	17.520034	1.324292	16.550118	1.248671	75PE
2RTO	19.231882	1.453806	18.164044	1.370745	75PE
2RWC	15.93848	1.204792	15.153234	1.143236	75PE
2RWO	18.383626	1.38946	17.484332	1.319311	75PE
3RTC	13.646616	1.031522	13.027416	0.983088949	75PE
3RTO	13.893734	1.05007	13.333602	1.00605	75PE
3RWC	12.616858	0.95387819	11.909084	0.898787361	75PE
3RWO	14.35631	1.085105	13.808472	1.042079	75PE
STC1	42.27487	6.383505	41.897582	6.193477	150PE
STC2	34.081996	5.146381	33.4614	4.97395	150PE
STC3	33.135868	5.003516	32.22466	4.763533	150PE
STO1	32.829994	4.957329	32.415854	4.816595	150PE
STO2	35.229118	5.319597	34.85856	5.148945	150PE
STO3	41.624552	6.285307	40.753736	5.998461	150PE
SWC1	42.460934	6.411601	42.16573	6.251875	150PE
SWC2	43.445634	6.560291	43.131448	6.41081	150PE
SWC3	46.297124	6.990866	45.857172	6.73953	150PE
SWO1	48.42103	7.311576	48.19125	7.163934	150PE
SWO2	42.513436	6.419529	42.11791	6.244847	150PE
SWO3	44.597962	6.734292	44.362768	6.537218	150PE

Table S3.3: Annotations for all differentially expressed genes in *D. recens*. Gene IDs are from Braker annotations. Gene names are taken from homologs in *D. melanogaster* reference genome release r6.46 (Flybase.org).

Gene ID	Best hit BLAST accession	Gene Name	Length query	Description	Tissue	Expression
g31432	XP_034480827	RpL18-PB	188	60S ribosomal protein L18	carcass	up
g23671	XP_034482529	Tak1-PA	487	mitogen-activated protein kinase kinase kinase 10-like	carcass	up
g12617	XP_002006628	Ufm1-PA	84	ubiquitin-fold modifier 1	carcass	up
g38585	XP_034482303	Jon66Ci-PA	260	serine protease 1-like	carcass	up
g21330	XP_034483897	CG8562-PA	423	carboxypeptidase B-like	carcass	up
g11813	XP_034477470	CG11414-PA	911	E3 ubiquitin-protein ligase ZNF598	carcass	up
g1313	XP_034488930	CG6204-PA	916	NFX1-type zinc finger-containing protein 1-like	carcass	up
g36874	XP_034111367	CG5157-PA	123	cytochrome b5	carcass	up
g22450	XP_002051316	Fic-PA	280	protein adenylyltransferase Fic	carcass	up
g31425	XP_034482470	D19A-PA	831	zinc finger protein 85-like	carcass	up
g53	XP_034479988	CG12384-PA	96	death-associated protein 1	carcass	up
g177	EDV92370	ctp-PF	78	GH24086	carcass	up
g192	XP_034490041	sno-PB	1349	protein strawberry notch	carcass	up
g208	XP_034490747	Ran-PD	216	GTP-binding nuclear protein Ran	carcass	up
g236	XP_017963452	Bap111-PA	627	extensin isoform X2	carcass	up
g238	XP_034489597	Prp16-PA	705	pre-mRNA-splicing factor ATP-dependent RNA helicase PRP16 isoform X1	carcass	up
g244	XP_034118447	Chchd2-PA	167	coiled-coil-helix-coiled-coil-helix domain-containing protein 2	carcass	up
g269	XP_034476296	CG8675-PA	238	uncharacterized protein C9orf85 homolog	carcass	up
g279	XP_032308353	rdgA-PD	165	eye-specific diacylglycerol kinase isoform X18	carcass	up
g292	XP_034488119	CG7744-PA	736	uncharacterized protein LOC117792192	carcass	up
g302	XP_034489070	CG1785-PA	483	ribosome biogenesis protein NOP53	carcass	up
g304	XP_034488400	wgn-PA	301	tumor necrosis factor receptor superfamily member wengen isoform X2	carcass	up
g402	XP_002058986	Tim8-PB	88	mitochondrial import inner membrane translocase subunit Tim8	carcass	up
g406	XP_034475775	CG1737-PC	1052	uncharacterized protein LOC117782831	carcass	up
g430	XP_034490104	waw-PA	651	translation factor waclaw, mitochondrial	carcass	up
g459	XP_034474284	anox-PB	243	acyl-CoA-binding domain-containing protein 6	carcass	up
g461	XP_023169313	NA	834	basic proline-rich protein	carcass	up
g462	XP_034488798	PIG-T-PA	611	GPI transamidase component PIG-T	carcass	up

g492	XP_034489730	Hers-PI	1363	pneumococcal serine-rich repeat protein isoform X1	carcass	up
g494	XP_034489731	Hers-PI	1685	AF4/FMR2 family member lilli isoform X2	carcass	up
g496	XP_034476177	Nep3-PC	785	neprilysin-3 isoform X2	carcass	up
g520	XP_034118392	Ykt6-PA	199	synaptobrevin homolog YKT6	carcass	up
g529	XP_034490391	LPCAT-PB	559	lysophosphatidylcholine acyltransferase	carcass	up
g711	XP_034477207	CG8611-PB	895	probable ATP-dependent RNA helicase CG8611	carcass	up
g849	XP_030569238	CG3638-PH	760	LOW QUALITY PROTEIN: protein tweety	carcass	up
g1066	XP_034479208	Pop1-PA	874	ribonucleases P/MRP protein subunit POP1	carcass	up
g1240	XP_034473127	GIIIsla2-PC	246	acidic phospholipase A2 PA4	carcass	up
g1327	XP_034488957	mRpL1-PA	339	50S ribosomal protein L1	carcass	up
g1452	XP_034486072	Cyp6d4-PA	513	probable cytochrome P450 6d4	carcass	up
g1737	XP_034486485	NK7.1-PD	834	probable serine/threonine-protein kinase cdc7 isoform X2	carcass	up
g1843	XP_034489175	CG2519-PC	1307	uncharacterized protein LOC117792944	carcass	up
g1866	XP_034486660	Alh-PP	690	protein AF-10	carcass	up
g1901	XP_023173847	larp-PH	1079	LOW QUALITY PROTEIN: la-related protein 1	carcass	up
g1924	XP_034489418	CG15534-PA	668	sphingomyelin phosphodiesterase	carcass	up
g1947	XP_034486446	NA	73	formin-1-like	carcass	up
g2028	XP_034484866	Srlp-PA	390	probable enoyl-CoA hydratase echA8	carcass	up
g2035	XP_034486122	CG9795-PE	743	ecdysone-induced protein 74EF isoform X1	carcass	up
g2036	KRG02064	CG9776-PA	1374	uncharacterized protein Dmoj_GI22274, isoform C	carcass	up
g2054	XP_034475220	CG1532-PB	286	glyoxalase 1	carcass	up
g2058	XP_034486231	kar-PA	581	monocarboxylate transporter 10 isoform X1	carcass	up
g2063	XP_034113904	CG4467-PA	1080	endoplasmic reticulum aminopeptidase 2 isoform X3	carcass	up
g2073	XP_034488636	l(3)mbt-PB	1341	uncharacterized protein LOC117792556, partial	carcass	up
g2077	XP_034489108	TfIIA-L-PA	405	transcription initiation factor IIA subunit 1 isoform X3	carcass	up
g2079	XP_034487265	CG12288-PA	357	nucleolar protein 12	carcass	up
g2085	XP_034485723	CG17568-PA	575	zinc finger protein weckle-like isoform X1	carcass	up
g2099	XP_034489113	Efa6-PH	1539	AF4/FMR2 family member lilli isoform X2	carcass	up
g2101	XP_034487036	CG6937-PA	354	MKI67 FHA domain-interacting nucleolar phosphoprotein	carcass	up
g2103	XP_034486492	CSN6-PA	347	COP9 signalosome complex subunit 6	carcass	up
g2104	XP_034485302	Apc-PB	2565	adenomatous polyposis coli protein	carcass	up

g2106	XP_034484989	CG1647-PA	1313	histone-lysine N-methyltransferase, H3 lysine-79 specific isoform X2	carcass	up
g2118	XP_034485439	TFAM-PA	258	transcription factor A, mitochondrial isoform X2	carcass	up
g2119	XP_034485705	CG5191-PF	528	fatty-acid amide hydrolase 2-A	carcass	up
g2121	XP_034487408	CG15922-PB	66	uncharacterized protein LOC117791701	carcass	up
g2125	XP_034486968	MFS9-PA	504	sialin	carcass	up
g2136	XP_034488418	NA	137	LOW QUALITY PROTEIN: uncharacterized protein LOC117792396	carcass	up
g2147	XP_034485939	EMC10-PC	228	ER membrane protein complex subunit 10	carcass	up
g2166	XP_034489279	Cys-PA	118	cystatin-like protein	carcass	up
g2208	XP_034484817	cindr-PC	955	serine-rich adhesin for platelets	carcass	up
g2234	XP_002001128	CG34112-PB	138	uncharacterized protein LOC6575110	carcass	up
g2236	XP_034487979	tub-PC	514	protein Tube	carcass	up
g2238	XP_034486415	Cont-PA	1439	contactin	carcass	up
g2255	XP_034487114	ball-PB	593	nucleosomal histone kinase 1	carcass	up
g2262	XP_034485991	CG31223-PB	394	zinc finger HIT domain-containing protein 2	carcass	up
g2265	XP_001994483	CG17272-PA	149	calmodulin-like protein 4	carcass	up
g2266	XP_034485994	CG17271-PB	135	G-box-binding factor isoform X3	carcass	up
g2268	XP_034485974	Fancd2-PA	1489	Fanconi anemia group D2 protein	carcass	up
g2276	XP_034487812	Cdk2-PC	314	cyclin-dependent kinase 2	carcass	up
g2432	XP_034485189	Mink-PA	543	titin isoform X1	carcass	up
g2475	XP_034487611	Sfxn1-3-PB	321	sideroflexin-1-3-like	carcass	up
g2482	XP_034488364	Cyp9f2-PA	515	uncharacterized protein LOC117792363	carcass	up
g11457	XP_034479023	Egm-PA	643	complex I assembly factor ACAD9, mitochondrial	carcass	up
g11458	XP_034478122	CG9005-PE	1129	uncharacterized protein LOC117784482	carcass	up
g11463	B4KR05	Tret1-1-PA	913	RecName: Full=Facilitated trehalose transporter Tret1	carcass	up
g11466	XP_034476492	cnk-PB	1717	uncharacterized protein LOC117783279	carcass	up
g11490	XP_034477184	Pcl-PB	1150	polycomb protein Pcl	carcass	up
g11494	EDW00737	otk-PA	1641	GH21048	carcass	up
g11501	XP_034476720	CG8321-PA	245	uncharacterized protein LOC117783427	carcass	up
g11511	XP_034478051	NA	118	uncharacterized protein LOC117784424	carcass	up
g11545	XP_034479812	jbug-PL	1625	filamin-C isoform X1	carcass	up
g11563	XP_030570452	Taf7-PB	230	transcription initiation factor TFIID subunit 7	carcass	up
g11564	XP_034478716	CG13516-PB	187	lipopolysaccharide-induced tumor necrosis factor-alpha factor	carcass	up
g11565	XP_034478717	CG30269-PB	160	lipopolysaccharide-induced tumor necrosis factor-alpha factor homolog	carcass	up
g11574	XP_034110358	CG13510-PC	103	lipopolysaccharide-induced tumor necrosis factor-alpha factor homolog	carcass	up

g11575	XP_017867131	CG13510-PC	148	PREDICTED: lipopolysaccharide-induced tumor necrosis factor-alpha factor homolog	carcass	up
g11581	XP_034478050	CG34445-PA	123	uncharacterized protein LOC117784423	carcass	up
g11585	XP_034477543	CG8046-PA	542	proton-coupled folate transporter	carcass	up
g11588	XP_034476631	Prp38-PA	275	pre-mRNA-splicing factor 38	carcass	up
g11589	XP_034476618	shrb-PA	228	charged multivesicular body protein 4b	carcass	up
g11606	XP_034480092	sktl-PC	845	phosphatidylinositol 4-phosphate 5-kinase type-1 alpha	carcass	up
g11639	XP_034477564	rig-PB	658	protein rigor mortis isoform X2	carcass	up
g11641	XP_034477564	rig-PB	533	protein rigor mortis isoform X2	carcass	up
g11653	XP_034479033	Nox-PE	1378	NADPH oxidase 5 isoform X1	carcass	up
g11686	XP_034476138	Camta-PF	375	uncharacterized protein LOC117783070	carcass	up
g11693	XP_034476799	Alp6-PA	519	membrane-bound alkaline phosphatase	carcass	up
g11781	TDG48262	NA	116	hypothetical protein AWZ03_005217	carcass	up
g11782	XP_034107823	M7BP-PN	3932	mucin-2 isoform X1	carcass	up
g11783	XP_034479694	M7BP-PF	291	calphotin isoform X9	carcass	up
g11790	XP_034477709	Rdh1-PA	329	retinol dehydrogenase 12-like	carcass	up
g11795	XP_034478387	U2A-PA	264	probable U2 small nuclear ribonucleoprotein A'	carcass	up
g11800	XP_034476435	tor-PA	919	tyrosine-protein kinase receptor torso	carcass	up
g11802	XP_034478943	CG3548-PA	460	transcription factor SPT20 homolog	carcass	up
g11803	XP_034478170	ATPsynF-PA	107	putative ATP synthase subunit f, mitochondrial	carcass	up
g11812	XP_034476371	uri-PA	705	unconventional prefoldin RPB5 interactor-like protein	carcass	up
g11815	XP_017017017	ITP-PE	115	PREDICTED: ion transport peptide	carcass	up
g11818	XP_034475788	CG4622-PC	540	zinc finger CCHC domain-containing protein 8 homolog	carcass	up
g11820	XP_034475786	NA	796	breast cancer type 2 susceptibility protein homolog isoform X1	carcass	up
g11823	XP_034475790	Nup44A-PA	354	nucleoporin seh1	carcass	up
g11828	XP_034476876	Odc1-PA	392	ornithine decarboxylase 1-like	carcass	up
g11833	XP_034479163	CG12769-PC	768	bromodomain-containing protein DDB_G0270170	carcass	up
g11844	XP_034476491	Asap-PB	1162	arfGAP with SH3 domain, ANK repeat and PH domain-containing protein isoform X2	carcass	up
g11845	XP_034477885	Nup50-PA	575	nuclear pore complex protein Nup50	carcass	up
g11846	XP_034477886	coil-PE	481	coilin	carcass	up
g11853	XP_034479391	kermit-PE	346	PDZ domain-containing protein GIPC3	carcass	up
g11855	XP_034477509	mRpS17-PA	155	28S ribosomal protein S17, mitochondrial	carcass	up
g11856	XP_034477508	CG3328-PD	1511	uncharacterized protein LOC117784022	carcass	up
g11857	EDW02287	spag-PA	480	GH19976	carcass	up

g11858	XP_034476368	DnaJ-60-PC	216	dnaJ-like protein 60	carcass	up
g11860	XP_034475876	Naa35-PA	811	N-alpha-acetyltransferase 35, NatC auxiliary subunit homolog	carcass	up
g11861	XP_034475877	Zfrp8-PB	352	programmed cell death protein 2	carcass	up
g11862	XP_034475874	tamo-PC	847	protein tamozhennic	carcass	up
g11869	XP_034479349	eIF2Bgamma-PA	457	translation initiation factor eIF-2B subunit gamma	carcass	up
g11872	XP_034479348	CG30467-PA	541	uncharacterized protein LOC117785451	carcass	up
g11876	XP_034476245	CG8388-PA	619	transcription factor grauzone	carcass	up
g11877	XP_034476246	CG8389-PA	474	monocarboxylate transporter 13	carcass	up
g11888	XP_034477891	CG8414-PA	768	polynucleotide 5'-hydroxyl-kinase NOL9	carcass	up
g11893	XP_034476180	CG10417-PB	732	probable protein phosphatase CG10417	carcass	up
g12026	XP_034477953	Polr1A-PA	1649	DNA-directed RNA polymerase I subunit RPA1	carcass	up
g12040	XP_034477903	chn-PG	1369	protein charlatan	carcass	up
g12127	XP_032292800	NA	445	glycine-rich cell wall structural protein 1	carcass	up
g12326	XP_034479238	NA	290	uncharacterized protein LOC117785379	carcass	up
g12338	XP_034479020	PGRP-SC2-PA	184	peptidoglycan-recognition protein SC2-like	carcass	up
g12383	XP_034476723	Pepck2-PA	639	phosphoenolpyruvate carboxykinase	carcass	up
g12447	XP_034475706	CG10936-PA	1614	mucin-5AC	carcass	up
g12459	XP_034103994	CG6191-PB	728	CDK5 and ABL1 enzyme substrate 2 isoform X1	carcass	up
g12460	XP_034478940	sced-PB	885	uncharacterized protein LOC117785144	carcass	up
g12461	XP_034478949	Dpit47-PB	396	DNA polymerase interacting tetratricopeptide repeat-containing, protein of 47 kDa	carcass	up
g12463	XP_034478106	CG9143-PA	826	ATP-dependent RNA helicase DDX24	carcass	up
g12465	XP_034478702	CG16868-PA	1478	VWFA and cache domain-containing protein CG16868	carcass	up
g12470	XP_034478043	CG15358-PE	252	pulmonary surfactant-associated protein D-like	carcass	up
g12471	XP_034478776	CG11200-PC	339	dehydrogenase/reductase SDR family member on chromosome X	carcass	up
g12474	XP_034479754	tapas-PB	1198	tudor domain-containing protein 7 isoform X1	carcass	up
g12477	XP_034479758	Pex6-PC	901	peroxisome assembly factor 2 isoform X2	carcass	up
g12478	XP_034477595	Taf5-PA	705	transcription initiation factor TFIID subunit 5	carcass	up
g12479	XP_034477596	nclb-PA	462	periodic tryptophan protein 1 homolog	carcass	up
g12481	XP_034476220	CG30015-PB	1585	serine-rich adhesin for platelets	carcass	up
g12486	XP_034110818	Tmtc3-PB	937	protein O-mannosyl-transferase Tmtc3	carcass	up
g12487	NP_476636	mago-PA	147	mago nashi	carcass	up
g12491	XP_034480183	Hmg-2-PA	407	high mobility group protein 20A	carcass	up

g12493	XP_034480179	CG30389-PE	973	macoilin isoform X1	carcass	up
g12495	XP_034480184	hng1-PA	157	uncharacterized protein LOC117785989	carcass	up
g12502	XP_001986085	Dp-PA	120	transcription factor Dp isoform X1	carcass	up
g12505	XP_034479003	CG11141-PC	699	WD repeat-containing protein CG11141	carcass	up
g12510	XP_034478698	CG30503-PB	173	phospholipase A2 large subunit	carcass	up
g12512	XP_034478418	CG11123-PA	675	nucleolar protein 9	carcass	up
g12516	TDG44657	Eaf-PA	259	hypothetical protein AWZ03_008892	carcass	up
g12519	XP_034478516	Glo1-PA	178	lactoylglutathione lyase	carcass	up
g12524	XP_034107673	NA	194	apidaccins type 73-like	carcass	up
g12525	XP_034478510	Dhx15-PB	730	putative pre-mRNA-splicing factor ATP-dependent RNA helicase PRP1 isoform X1	carcass	up
g12526	XP_034478509	cos-PB	1206	kinesin-like protein costa	carcass	up
g12538	XP_034478560	Psi-PB	836	far upstream element-binding protein 2 isoform X1	carcass	up
g12539	XP_034476829	CG5532-PB	111	transmembrane protein 14 homolog	carcass	up
g12544	XP_034480133	Ssrp-PA	743	FACT complex subunit Ssrp1	carcass	up
g12545	XP_034478794	CG5543-PA	651	gastrulation defective protein 1 homolog	carcass	up
g12547	XP_034479284	jef-PB	774	major facilitator superfamily domain- containing protein 6	carcass	up
g12552	XP_034476751	ND-B15-PA	113	NADH dehydrogenase	carcass	up
g12556	XP_034479678	cnn-PA	1098	centrosomin isoform X1	carcass	up
g12561	XP_034102543	Tbp-PA	355	TATA-box-binding protein	carcass	up
g12562	XP_034477657	eIF3k-PA	222	eukaryotic translation initiation factor 3 subunit K	carcass	up
g12569	XP_034476067	pea-PA	1249	ATP-dependent RNA helicase DHX8	carcass	up
g12570	XP_017069285	pea-PA	834	PREDICTED: LOW QUALITY PROTEIN: ATP-dependent RNA helicase DHX8	carcass	up
g12592	XP_032293145	CG9380-PG	376	uncharacterized protein LOC6627116 isoform X2	carcass	up
g12616	XP_034476357	mute-PC	1830	uncharacterized protein LOC117783197	carcass	up
g12625	XP_034478402	CG6421-PA	161	lysozyme	carcass	up
g12627	XP_034477732	eEF1beta-PC	222	probable elongation factor 1-beta	carcass	up
g12628	XP_034478238	prel-PB	248	protein preli-like	carcass	up
g12631	XP_017865744	dom-PE	2591	PREDICTED: helicase domino isoform X1	carcass	up
g14489	XP_034474928	homer-PE	402	LOW QUALITY PROTEIN: homer protein homolog 1	carcass	up
g14534	XP_034472915	baf-PB	75	barrier-to-autointegration factor-like	carcass	up
g14607	XP_034474312	bru1-PE	594	CUGBP Elav-like family member 2 isoform X2	carcass	up
g14619	XP_034474726	Cdc23-PB	717	cell division cycle protein 23 homolog	carcass	up
g14639	XP_034475311	NA	98	U-scoloptoxin(16)-Ssd1a-like	carcass	up

g14654	XP_034478043	CG15358-PE	252	pulmonary surfactant-associated protein D-like	carcass	up
g14673	XP_034474328	gkt-PA	459	LOW QUALITY PROTEIN: probable tyrosyl-DNA phosphodiesterase	carcass	up
g14832	XP_034472008	smo-PA	1061	protein smoothened isoform X1	carcass	up
g14883	XP_034473161	CG30280-PD	357	ficolin-1-like isoform X1	carcass	up
g14992	XP_034474438	CG15263-PA	301	uncharacterized protein LOC117781731	carcass	up
g15040	XP_034475352	Pld3-PC	492	5'-3' exonuclease PLD3-like	carcass	up
g15042	XP_034484363	rdx-PB	243	protein roadkill-like	carcass	up
g15043	XP_001970188	CG7203-PC	136	adult cuticle protein 1	carcass	up
g15050	XP_034098354	Brd7-9-PA	905	bromodomain-containing protein 7	carcass	up
g15055	XP_034475336	CG5787-PB	886	GATA zinc finger domain-containing protein 14	carcass	up
g15062	XP_034473386	CCT4-PA	533	T-complex protein 1 subunit delta	carcass	up
g15070	XP_034097919	bun-PF	1132	protein bunched, class 2/F/G isoform isoform X6	carcass	up
g15170	XP_034473883	step-PE	303	probable serine/threonine-protein kinase DDB_G0272282	carcass	up
g15171	XP_034473560	CG1416-PC	355	activator of 90 kDa heat shock protein ATPase homolog 1	carcass	up
g15267	XP_034475250	Nup160-PA	1412	nuclear pore complex protein Nup160 homolog	carcass	up
g15329	XP_034475198	esc-PA	423	polycomb protein esc	carcass	up
g21188	XP_034483545	NA	130	division abnormally delayed protein isoform X3	carcass	up
g21189	XP_034481528	CG13308-PB	216	uncharacterized protein LOC117787185	carcass	up
g21206	XP_034482881	Cp16-PA	141	chorion protein S16	carcass	up
g21207	XP_034481847	Cp19-PA	188	chorion protein S19	carcass	up
g21208	XP_034481898	NA	114	chorion protein S15	carcass	up
g21209	XP_034481844	NA	169	chorion protein S18	carcass	up
g21360	XP_034480848	CG16986-PB	143	acyl-coenzyme A thioesterase 13	carcass	up
g21396	XP_034480601	CG7120-PG	721	uncharacterized protein LOC117786446 isoform X1	carcass	up
g22352	XP_034490998	Trpgamma-PD	1143	transient receptor potential-gamma protein	carcass	up
g22353	XP_030557410	NA	276	fibronectin type-III domain-containing protein 3A isoform X1	carcass	up
g22368	XP_034474764	CG5888-PB	460	leucine-rich repeat-containing protein 15	carcass	up
g22449	ALC40399	CG9500-PB	363	CG9500	carcass	up
g22498	XP_034473364	CG17239-PB	128	trypsin alpha-3-like	carcass	up
g22595	XP_034472492	CG33307-PB	179	uncharacterized protein LOC117780175	carcass	up
g22610	XP_034098361	Sec71-PA	1443	brefeldin A-inhibited guanine nucleotide-exchange protein 1	carcass	up
g22627	XP_034474082	CG13083-PB	336	uncharacterized protein LOC117781435	carcass	up
g22629	XP_034472553	MSBP-PC	188	membrane-associated progesterone receptor component 1	carcass	up

g22647	XP_034473214	Hacd2-PA	376	very-long-chain (3R)-3-hydroxyacyl-CoA dehydratase 3	carcass	up
g22687	XP_034472583	PCNA2-PA	255	proliferating cell nuclear antigen	carcass	up
g22692	XP_034473672	CG10462-PA	870	uncharacterized protein LOC117781061	carcass	up
g23490	XP_034482370	pbl-PB	1348	protein ECT2 isoform X1	carcass	up
g23497	XP_034482126	AlaRS-m-PA	1037	alanine--tRNA ligase, mitochondrial	carcass	up
g23500	XP_034481882	Dhrs4-PA	323	dehydrogenase/reductase SDR family member 4	carcass	up
g23502	XP_034481477	Tcs5-PA	241	EKC/KEOPS complex subunit TP53RK	carcass	up
g23505	XP_034482461	Yod1-PC	360	ubiquitin thioesterase OTU1	carcass	up
g23512	XP_034480977	Setx-PB	1783	uncharacterized protein LOC117786729	carcass	up
g23516	XP_034481645	mRpL36-PA	133	uncharacterized protein LOC117787272	carcass	up
g23517	XP_034483002	CG7550-PA	261	2-aminoethanethiol dioxygenase	carcass	up
g23519	XP_034484115	msk-PA	1049	importin-7	carcass	up
g23521	XP_034482651	Nulp1-PC	713	transcription factor 25	carcass	up
g23522	XP_034104695	Mfap1-PA	474	microfibrillar-associated protein 1	carcass	up
g23523	XP_034480538	mu2-PC	1276	uncharacterized protein LOC117786408	carcass	up
g23524	XP_034482724	Sras-PA	264	CAAX prenyl protease 2	carcass	up
g23525	XP_034482722	Msr-110-PB	620	putative uncharacterized protein DDB_G0268364 isoform X2	carcass	up
g23526	XP_034481411	l(3)psg2-PA	1041	serine/arginine repetitive matrix protein 2	carcass	up
g23530	XP_034483559	Sse-PA	650	uncharacterized protein LOC117788798 isoform X2	carcass	up
g23533	XP_017844284	Eaf6-PA	224	chromatin modification-related protein MEAF6	carcass	up
g23536	XP_034484254	Alp9-PB	520	membrane-bound alkaline phosphatase	carcass	up
g23538	XP_034480706	DnaJ-1-PA	352	dnaJ protein homolog 1	carcass	up
g23539	XP_034483109	Usp47-PB	1575	ubiquitin carboxyl-terminal hydrolase 47	carcass	up
g23540	AAV36985	CG10576-PD	399	LD30448p	carcass	up
g23544	XP_034480506	Mul1-PB	339	mitochondrial E3 ubiquitin protein ligase 1	carcass	up
g23555	XP_034482142	CG15877-PA	253	uncharacterized protein C7orf50 homolog isoform X2	carcass	up
g23556	XP_034482139	nSMase-PA	418	putative neutral sphingomyelinase	carcass	up
g23560	XP_034481597	rdx-PB	203	speckle-type POZ protein-like	carcass	up
g23595	XP_034482678	Ack-PA	1084	non-receptor tyrosine-protein kinase TNK1	carcass	up
g23617	XP_034107207	frm-PB	711	nascent polypeptide-associated complex subunit alpha, muscle-specific form isoform X1	carcass	up
g23624	XP_034483586	sls-PU	8703	LOW QUALITY PROTEIN: titin	carcass	up
g23629	XP_034483414	CG15822-PG	1260	SEC14 domain and spectrin repeat-containing protein 1 isoform X1	carcass	up
g23653	XP_034481597	rdx-PB	140	speckle-type POZ protein-like	carcass	up

g23957	XP_034480902	asf1-PB	224	histone chaperone asf1	carcass	up
g23958	XP_034480874	l(3)76BDm-PB	1313	trafficking protein particle complex subunit 8	carcass	up
g25994	XP_034490197	spirit-PD	367	serine protease snake	carcass	up
g25998	XP_034489467	r-PE	1211	CAD protein	carcass	up
g26108	XP_034490066	Chpf-PB	839	chondroitin sulfate glucuronyltransferase	carcass	up
g26110	XP_002010974	hep-PF	503	dual specificity mitogen-activated protein kinase kinase hemipterous isoform X2	carcass	up
g27034	XP_034106197	Jon66Ci-PA	267	serine protease 1-like	carcass	up
g27038	XP_034480961	Jon65Aiii-PA	271	serine protease 1-like	carcass	up
g27047	XP_034483481	CG10479-PB	402	uncharacterized protein LOC117788747 isoform X1	carcass	up
g27143	XP_034104561	Mvk-PB	395	mevalonate kinase	carcass	up
g27147	XP_034478964	Dera-PD	319	deoxyribose-phosphate aldolase-like	carcass	up
g27191	XP_034479042	Alp7-PA	539	LOW QUALITY PROTEIN: membrane-bound alkaline phosphatase	carcass	up
g27209	XP_034106516	CG34423-PB	85	ATPase inhibitor A, mitochondrial	carcass	up
g27214	XP_034477800	bonsai-PA	279	28S ribosomal protein S15, mitochondrial	carcass	up
g27219	XP_034478910	Vps20-PC	190	charged multivesicular body protein 6-A	carcass	up
g27262	TDG48318	CG18609-PA	439	hypothetical protein AWZ03_005273	carcass	up
g27307	XP_034478720	wcd-PA	527	U3 small nucleolar RNA-associated protein 18 homolog	carcass	up
g27311	XP_034108984	Alp8-PB	527	membrane-bound alkaline phosphatase-like	carcass	up
g27377	XP_034484678	Smyd5-PA	374	SET and MYND domain-containing protein 5	carcass	up
g27396	XP_034486237	CLS-PD	337	probable cardiolipin synthase (CMP-forming)	carcass	up
g27682	AFK29221	msi-PA	685	musashi-PA	carcass	up
g27690	XP_034485328	RpL13A-PB	205	60S ribosomal protein L13a	carcass	up
g27692	XP_034485326	CG12171-PA	257	2-(R)-hydroxypropyl-CoM dehydrogenase-like	carcass	up
g27693	XP_034115472	CG31548-PA	255	uncharacterized oxidoreductase TM_0325-like	carcass	up
g27694	XP_034485321	tacc-PC	1284	putative mediator of RNA polymerase II transcription subunit 26 isoform X4	carcass	up
g27706	XP_034486404	BBS5-PA	401	LOW QUALITY PROTEIN: Bardet-Biedl syndrome 5 protein homolog	carcass	up
g27711	XP_034486771	Gnf1-PA	985	replication factor C subunit 1	carcass	up
g27826	XP_034489153	CG18476-PA	934	zinc finger protein 84	carcass	up
g27828	XP_034489151	CG42327-PE	246	mucin-5AC isoform X1	carcass	up
g27868	XP_034482797	CG32444-PA	375	aldose 1-epimerase isoform X1	carcass	up
g27886	XP_034483197	CG14570-PA	267	pollen-specific leucine-rich repeat extensin-like protein 2	carcass	up

g27913	XP_034483523	CG11263-PB	260	protein Exd1 homolog	carcass	up
g28029	XP_017100372	upSET-PB	3533	PREDICTED: uncharacterized protein LOC108127685	carcass	up
g28046	XP_034480778	IRSp53-PA	1107	probable serine/threonine-protein kinase yakA isoform X1	carcass	up
g28056	XP_034483383	NA	187	uncharacterized protein LOC117788660	carcass	up
g28195	XP_034482598	CG6839-PA	428	uncharacterized protein LOC117788053	carcass	up
g28213	XP_034481237	FRG1-PA	266	protein FRG1 homolog	carcass	up
g28241	XP_034480341	Wnk-PN	916	serine/threonine-protein kinase WNK	carcass	up
g28377	XP_034481892	CTPsyn-PC	590	CTP synthase isoform X1	carcass	up
g29247	XP_034472594	NA	177	uncharacterized protein LOC117780252	carcass	up
g29248	XP_034473411	dbe-PA	351	KRR1 small subunit processome component homolog	carcass	up
g29254	XP_034473804	drongo-PH	619	arf-GAP domain and FG repeat- containing protein 1 isoform X1	carcass	up
g29330	XP_034483451	NA	324	uncharacterized protein LOC117788719	carcass	up
g29335	XP_034480832	Arts-PC	1084	importin-4-like isoform X1	carcass	up
g29361	XP_034480789	MED14-PA	1543	mediator of RNA polymerase II transcription subunit 14 isoform X2	carcass	up
g29400	XP_034483532	Pdk1-PJ	854	3-phosphoinositide-dependent protein kinase 1 isoform X1	carcass	up
g29411	XP_034482088	LysE-PA	140	lysozyme 1-like	carcass	up
g29418	XP_034484050	NHP2-PB	160	H/ACA ribonucleoprotein complex subunit 2-like protein	carcass	up
g29453	XP_034474607	CG31650-PF	338	reticulocalbin-2 isoform X1	carcass	up
g29510	XP_034472247	CG6724-PA	411	ribosome biogenesis protein WDR12 homolog	carcass	up
g29586	XP_034490913	MESR3-PD	279	uncharacterized protein LOC117794437 isoform X1	carcass	up
g29610	XP_034471569	NA	890	pre-mRNA-splicing factor CWC22 homolog isoform X1	carcass	up
g29611	XP_034471573	CG12010-PB	729	spermatogenesis-associated protein 5- like protein 1	carcass	up
g29672	XP_034473262	pths-PA	467	probable ATP-dependent RNA helicase DDX47	carcass	up
g29705	XP_032294955	Gclc-PD	707	glutamate--cysteine ligase	carcass	up
g29720	XP_034481395	NA	77	uncharacterized protein LOC117787056	carcass	up
g29749	XP_034489700	Nmd3-PA	530	60S ribosomal export protein NMD3	carcass	up
g29751	XP_034118251	CG8173-PA	401	lymphokine-activated killer T-cell- originated protein kinase	carcass	up
g29758	XP_034490840	NO66-PC	868	bifunctional lysine-specific demethylase and histidyl-hydroxylase NO66 isoform X2	carcass	up
g29888	XP_034489699	REG-PB	253	proteasome activator complex subunit 3	carcass	up
g30015	XP_034490429	Cdc7-PA	719	cell division cycle 7-related protein kinase	carcass	up
g30037	XP_034474850	NA	114	putative vitellogenin receptor isoform X1	carcass	up

g30223	XP_034489604	dhd-PB	108	thioredoxin-1	carcass	up
g30256	XP_017872668	ovo-PA	709	PREDICTED: protein ovo	carcass	up
g30259	XP_034480637	bou-PA	133	U-scoloptoxin(05)-Sm1a	carcass	up
g30313	XP_034489592	dyw-PA	254	circadian clock-controlled protein	carcass	up
g30355	XP_034479762	sov-PC	3696	uncharacterized protein LOC117785688	carcass	up
g30413	XP_034475125	Naa15-16-PB	880	N-alpha-acetyltransferase 15, NatA auxiliary subunit	carcass	up
g30558	XP_034490071	e(y)3-PE	1478	supporter of activation of yellow protein	carcass	up
g30711	XP_034476750	Ser6-PA	257	serine protease SP24D-like	carcass	up
g30832	XP_034477244	ana3-PA	2012	uncharacterized protein LOC117783807	carcass	up
g30835	XP_034478285	CG4554-PA	2766	small subunit processome component 20 homolog	carcass	up
g30842	XP_034478788	px-PA	543	uncharacterized protein LOC117785021 isoform X2	carcass	up
g30850	XP_034478164	SMC2-PA	1172	structural maintenance of chromosomes protein 2	carcass	up
g30872	XP_034479584	NA	184	probable basic-leucine zipper transcription factor E	carcass	up
g30914	XP_034473813	MED26-PB	706	mediator of RNA polymerase II transcription subunit 26 isoform X2	carcass	up
g31133	XP_034473853	Phb1-PD	275	protein I(2)37Cc	carcass	up
g31142	XP_034474267	Akap200-PG	877	A-kinase anchor protein 200 isoform X1	carcass	up
g31209	XP_034473848	CG17544-PD	695	peroxisomal acyl-coenzyme A oxidase 3	carcass	up
g31259	XP_034475219	CG16972-PA	1073	uncharacterized protein LOC117782271	carcass	up
g31271	XP_034473480	CG10702-PD	846	insulin receptor	carcass	up
g31370	XP_034114694	dpr9-PB	412	serine-rich adhesin for platelets isoform X1	carcass	up
g31373	XP_034487073	Sdr-PB	881	insulin-like peptide receptor isoform X1	carcass	up
g31387	XP_034486057	CG18764-PA	425	transcription factor Ouib-like isoform X1	carcass	up
g31402	XP_034489077	Cbs-PC	518	cystathionine beta-synthase-like protein isoform X2	carcass	up
g31421	XP_034484179	D19B-PA	154	zinc finger protein 431-like	carcass	up
g31422	XP_034484177	D19A-PA	734	zinc finger protein 675-like	carcass	up
g31423	XP_034484178	D19A-PA	719	gastrula zinc finger protein XICGF58.1- like	carcass	up
g31424	XP_034484174	D19A-PA	765	zinc finger protein 708-like	carcass	up
g31429	XP_030569913	CG10274-PB	869	zinc finger protein 540-like	carcass	up
g31431	XP_034111737	BHD-PA	482	folliculin	carcass	up
g31433	XP_034111739	mRpL50-PA	193	39S ribosomal protein L50, mitochondrial	carcass	up
g31434	XP_034480825	Neos-PB	243	uncharacterized protein LOC117786597	carcass	up
g31437	XP_034480478	path-PD	471	proton-coupled amino acid transporter- like protein pathetic	carcass	up
g31438	XP_034107663	CG3408-PA	337	leucine-rich repeat-containing protein 59	carcass	up

g31449	XP_034481492	Cyp40-PD	381	peptidyl-prolyl cis-trans isomerase D	carcass	up
g31450	XP_034102974	fry-PI	1552	protein furry isoform X7	carcass	up
g31452	XP_034102983	Jon25Bi-PB	276	serine protease 1-like	carcass	up
g31453	XP_034102983	Jon25Bi-PB	272	serine protease 1-like	carcass	up
g31470	XP_034483060	CG4022-PC	295	mediator of RNA polymerase II transcription subunit 15	carcass	up
g31472	XP_034482429	Cpr67B-PA	261	uncharacterized protein LOC117787909	carcass	up
g31489	XP_034484095	CG7194-PA	164	uncharacterized protein LOC117789144	carcass	up
g31491	XP_034482823	Nmt-PA	442	glycylpeptide N-tetradecanoyltransferase	carcass	up
g31728	XP_034478117	Gp210-PB	1847	nuclear pore membrane glycoprotein 210	carcass	up
g31740	XP_034477252	CG8545-PA	880	25S rRNA (cytosine-C(5))-methyltransferase nop2	carcass	up
g31747	XP_034476238	l(2)k10201-PA	212	protein lethal(2)k10201	carcass	up
g31770	XP_034478494	CG30377-PB	913	uncharacterized protein LOC117784784	carcass	up
g31999	XP_034478775	prod-PA	344	uncharacterized protein LOC117785013	carcass	up
g32027	XP_034478885	Tfb1-PD	581	general transcription factor IIIH subunit 1	carcass	up
g32080	XP_034478878	cGlr1-PC	416	uncharacterized protein LOC117785089	carcass	up
g32094	XP_034478641	NiPp1-PA	403	nuclear inhibitor of protein phosphatase 1	carcass	up
g32103	XP_034476973	SmydA-7-PB	489	SET domain-containing protein SmydA-8	carcass	up
g32134	XP_034476755	sotv-PB	705	exostosin-2	carcass	up
g32139	XP_034476056	mei-W68-PA	186	meiotic recombination protein W68 isoform X1	carcass	up
g32163	XP_034478194	Lpt-PA	1472	histone-lysine N-methyltransferase 2C	carcass	up
g32220	XP_034476884	Wnt5-PB	993	protein Wnt-5	carcass	up
g32247	XP_034477231	CG6362-PA	608	uncharacterized protein LOC117783795	carcass	up
g32252	XP_034478148	CG5033-PA	784	ribosome biogenesis protein BOP1 homolog	carcass	up
g32283	XP_034478603	CG2790-PA	547	dnaJ homolog subfamily C member 21	carcass	up
g32395	XP_034473308	lok-PC	462	ovarian-specific serine/threonine-protein kinase Lok isoform X2	carcass	up
g32509	XP_030561219	Df31-PB	196	enolase-phosphatase E1	carcass	up
g33025	ALC46975	cno-PI	2059	cno	carcass	up
g33063	XP_034488360	CG5728-PA	1404	uncharacterized protein LOC117792359	carcass	up
g33095	XP_034485674	ana1-PC	1803	uncharacterized protein LOC117790353	carcass	up
g33132	XP_034486256	msps-PE	1004	protein mini spindles isoform X2	carcass	up
g33288	XP_034486198	CG7218-PB	675	protein TAPT1 homolog isoform X2	carcass	up
g33291	XP_034486202	CG7215-PB	125	ubiquitin-like protein 4A	carcass	up
g33466	XP_037884605	Rm62-PA	915	LOW QUALITY PROTEIN: ATP-dependent RNA helicase p62-like	carcass	up

g33562	XP_023163716	CG7518-PF	2648	uncharacterized protein LOC111594587 isoform X4	carcass	up
g33593	XP_034489217	CG31224-PC	1924	AF4/FMR2 family member lilli	carcass	up
g33602	XP_034484752	fray-PJ	735	serine/threonine-protein kinase OSR1 isoform X1	carcass	up
g33617	XP_023161178	Vti1b-PA	123	vesicle transport through interaction with t-SNAREs homolog 1B	carcass	up
g33636	XP_034487567	CG33332-PA	293	GPALPP motifs-containing protein 1	carcass	up
g33638	XP_034487561	Rrp6-PC	913	exosome component 10	carcass	up
g33639	XP_034487565	CG3631-PE	421	glycosaminoglycan xylosylkinase homolog	carcass	up
g33878	XP_034485110	mod-PC	509	DNA-binding protein modulo isoform X2	carcass	up
g33925	XP_034488220	CG10254-PA	1434	(E3-independent) E2 ubiquitin-conjugating enzyme UBE2O	carcass	up
g34018	XP_034475639	Set1-PI	693	histone-lysine N-methyltransferase SETD1	carcass	up
g34033	XP_034475629	NA	193	uncharacterized protein LOC117782712 isoform X1	carcass	up
g34105	ALC45635	spg-PE	142	spg	carcass	up
g34201	XP_034480754	Pex1-PA	950	peroxisome biogenesis factor 1	carcass	up
g34376	XP_034485034	m-cup-PA	691	protein fem-1 homolog C isoform X1	carcass	up
g34425	TDG42578	vas-PC	1500	hypothetical protein AWZ03_010987	carcass	up
g34528	XP_034472177	Acsx3-PA	500	4-coumarate--CoA ligase 2-like	carcass	up
g34638	XP_034489421	Polr1B-PA	1123	DNA-directed RNA polymerase I subunit RPA2	carcass	up
g34641	XP_034473383	CG2794-PA	698	pseudouridine-metabolizing bifunctional protein C1861.05	carcass	up
g34644	XP_034471984	CG11835-PD	736	proteoglycan 4 isoform X1	carcass	up
g34645	XP_034471987	Tspo-PA	175	translocator protein	carcass	up
g34646	XP_034472969	rempA-PB	1548	LOW QUALITY PROTEIN: intraflagellar transport protein 140 homolog	carcass	up
g34651	NP_001259833	lwr-PD	159	lesswright, isoform C	carcass	up
g34652	EDW64995	ush-PE	1356	LOW QUALITY PROTEIN: uncharacterized protein Dvir_GJ17781	carcass	up
g34658	XP_034471931	shv-PA	358	dnaJ homolog shv	carcass	up
g34682	XP_017857007	spen-PH	4047	PREDICTED: protein split ends	carcass	up
g34686	XP_034475030	kis-PC	945	chromodomain-helicase-DNA-binding protein 7 isoform X5	carcass	up
g34750	XP_034489462	GlcAT-I-PA	307	galactosylgalactosylxylosylprotein 3-beta-glucuronosyltransferase I	carcass	up
g34753	XP_034490598	tyf-PN	2299	pneumococcal serine-rich repeat protein	carcass	up
g34755	XP_034489569	Nsun2-PA	746	tRNA (cytosine(34)-C(5))-methyltransferase	carcass	up
g34757	XP_034478801	cib-PE	106	thymosin beta	carcass	up
g34758	XP_034476309	brn-PA	325	beta-1,3-galactosyltransferase brn	carcass	up

g34853	XP_001992589	CG42258-PC	498	probable serine/threonine-protein kinase nek3	carcass	up
g34865	XP_034489727	CG4332-PA	572	cleft lip and palate transmembrane protein 1-like protein	carcass	up
g34875	XP_017055033	CG5921-PD	303	PREDICTED: la-related protein CG11505 isoform X2	carcass	up
g34885	XP_034480738	CG33107-PA	399	uncharacterized protein LOC117786531	carcass	up
g35121	XP_034488712	CG42232-PB	1485	titin	carcass	up
g35133	XP_034486043	CG8301-PA	597	zinc finger protein 2 homolog	carcass	up
g35138	QMU95573	NA	841	hypothetical protein	carcass	up
g35151	XP_034488316	Lk6-PB	601	probable serine/threonine-protein kinase MARK-C isoform X2	carcass	up
g35152	XP_034488315	Lk6-PB	303	probable serine/threonine-protein kinase MARK-C isoform X1	carcass	up
g35161	XP_034486221	CG11980-PA	285	UPF0160 protein C27H6.8	carcass	up
g35196	XP_034484348	alrm-PA	472	toll-like receptor 7	carcass	up
g35207	EDW66699	ns11-PF	1865	uncharacterized protein Dvir_GJ23484	carcass	up
g35213	XP_034479275	CG10911-PA	197	protein TsetseEP isoform X6	carcass	up
g35227	XP_034485444	CG10365-PD	297	glutathione-specific gamma-glutamylcyclotransferase 1	carcass	up
g35294	TDG39571	CG7166-PC	150	hypothetical protein AWZ03_014006	carcass	up
g35375	XP_034487535	NA	48	uncharacterized protein LOC117791768 isoform X1	carcass	up
g35432	XP_034489268	ZIPIC-PA	440	zinc finger protein 771	carcass	up
g35480	XP_034485318	tacc-PA	920	putative mediator of RNA polymerase II transcription subunit 26 isoform X1	carcass	up
g35520	XP_034480612	CG13917-PD	2139	uncharacterized protein LOC117786453	carcass	up
g35697	XP_034103291	MESK2-PE	454	protein NDRG3 isoform X4	carcass	up
g35713	XP_017837483	mRpS16-PA	125	probable 28S ribosomal protein S16, mitochondrial	carcass	up
g35725	XP_034477356	Snrk-PC	765	SNF-related serine/threonine-protein kinase	carcass	up
g35730	XP_034479505	shn-PF	2568	uncharacterized protein LOC117785538	carcass	up
g35745	XP_034478362	CG10927-PA	366	probable inactive tRNA-specific adenosine deaminase-like protein 3	carcass	up
g35756	XP_034477861	NA	244	integumentary mucin C.1	carcass	up
g35772	XP_034478633	tea-PB	647	protein telomere ends associated	carcass	up
g35774	XP_034479287	AspRS-PB	537	aspartate--tRNA ligase, cytoplasmic	carcass	up
g35783	XP_034478753	CG10543-PG	1736	uncharacterized protein LOC117784990 isoform X2	carcass	up
g35811	XP_034479793	babos-PB	202	uncharacterized protein LOC117785708	carcass	up
g35812	XP_034479790	CG13506-PB	503	immunoglobulin superfamily member 10	carcass	up
g35818	XP_034476364	NA	59	uncharacterized protein LOC117783203	carcass	up
g35927	XP_034476123	CG3502-PB	899	aminopeptidase N	carcass	up
g36027	XP_034475737	eEF2-PD	844	elongation factor 2-like	carcass	up

g36072	XP_034478316	NA	175	uncharacterized protein LOC117784631	carcass	up
g36079	XP_034476643	FLASH-PB	276	rho-associated protein kinase 2 isoform X3	carcass	up
g36083	XP_030081923	CG17724-PD	348	transcription factor mef2A	carcass	up
g36092	XP_030243128	NA	285	mucin-5AC-like	carcass	up
g36110	XP_034478325	Mlf-PD	391	myeloid leukemia factor isoform X1	carcass	up
g36114	XP_034478366	etaTry-PA	257	trypsin eta-like	carcass	up
g36133	XP_034479971	CG1827-PD	409	putative N(4)-(beta-N-acetylglucosaminy)-L-asparaginase GA14866	carcass	up
g36135	XP_034479968	clos-PD	1527	uncharacterized protein LOC117785829 isoform X1	carcass	up
g36162	XP_034479148	Yeti-PB	296	craniofacial development protein 1	carcass	up
g36166	XP_034479145	wde-PB	1444	activating transcription factor 7-interacting protein 1	carcass	up
g36167	XP_034479992	CG33144-PB	1223	uncharacterized protein LOC117785848	carcass	up
g36189	XP_034477002	lola-PR	472	longitudinals lacking protein isoform X3	carcass	up
g36193	XP_034106104	CG11777-PA	161	peptidyl-prolyl cis-trans isomerase-like 3	carcass	up
g36217	XP_034476302	RNaseZ-PB	781	ribonuclease Z, mitochondrial	carcass	up
g36241	XP_034479227	MED9-PC	145	mediator of RNA polymerase II transcription subunit 9	carcass	up
g36242	XP_034479228	CG42518-PB	85	protein NCBP2AS2 homolog	carcass	up
g36316	XP_034477037	CG1418-PA	194	prenylated Rab acceptor protein 1	carcass	up
g36323	XP_017866360	CG8320-PA	174	PREDICTED: transmembrane protein 208 isoform X1	carcass	up
g36327	XP_034477235	Strn-Mlck-PS	1013	muscle M-line assembly protein unc-89-like	carcass	up
g36367	XP_034480195	Yp1-PB	432	vitellogenin-1	carcass	up
g36494	XP_034486638	CG18596-PA	464	uncharacterized protein LOC117791092	carcass	up
g36690	XP_034483888	pst-PG	697	uncharacterized protein LOC117789014	carcass	up
g36873	XP_034482897	CG5151-PE	458	myb-like protein P isoform X1	carcass	up
g37109	XP_034101725	dikar-PD	2335	protein split ends	carcass	up
g37112	XP_034472410	Npc2a-PA	144	NPC intracellular cholesterol transporter 2 homolog a	carcass	up
g37134	XP_034481120	Tsen54-PA	395	uncharacterized protein LOC117786822	carcass	up
g37147	XP_034481156	CG32271-PA	236	seminase	carcass	up
g37306	XP_034480817	CG6321-PA	335	uncharacterized protein YER152C	carcass	up
g37307	XP_034480816	CG7888-PA	470	proton-coupled amino acid transporter-like protein CG1139	carcass	up
g37330	XP_034481653	Usp10-PC	981	chromatin modification-related protein eaf-1 isoform X3	carcass	up
g37362	XP_034480759	CG10565-PB	638	dnaJ homolog subfamily C member 2	carcass	up
g37367	XP_034480307	CG16753-PA	180	protein FAM207A	carcass	up

g37669	XP_034482993	CG7369-PA	697	ras-GEF domain-containing family member 1B-B	carcass	up
g37670	XP_034483471	I(3)04053-PA	643	serine-rich adhesin for platelets	carcass	up
g37681	XP_034488254	CG4360-PA	176	zinc finger protein 84 isoform X1	carcass	up
g37695	XP_034480453	CG4042-PB	281	distal membrane-arm assembly complex protein 2	carcass	up
g37722	XP_023170827	CG17732-PB	800	protein lava lamp	carcass	up
g37739	XP_034481803	CG42588-PB	1316	uncharacterized protein LOC117787400	carcass	up
g37752	XP_030569426	CG14120-PA	1362	uncharacterized protein LOC115768859	carcass	up
g37764	XP_022216803	Chmp1-PB	204	charged multivesicular body protein 1b	carcass	up
g37895	XP_034484225	NA	316	extensin isoform X7	carcass	up
g37901	XP_034482926	Grip163-PA	1312	uncharacterized protein LOC117788304	carcass	up
g37913	XP_034481534	CG17147-PB	338	peritrophin-44-like	carcass	up
g37914	XP_034483926	Su(var)3-3-PB	911	possible lysine-specific histone demethylase 1	carcass	up
g37918	XP_034483918	CG17233-PG	1156	nuclear factor of activated T-cells 5 isoform X2	carcass	up
g37939	XP_034483724	AGO2-PB	823	protein argonaute-2 isoform X4	carcass	up
g38081	XP_034487415	NSD-PA	1445	probable histone-lysine N-methyltransferase Mes-4	carcass	up
g38096	XP_034486951	CG12951-PA	265	chymotrypsin-2-like	carcass	up
g38386	XP_034479616	Epac-PH	433	rap guanine nucleotide exchange factor 4 isoform X2	carcass	up
g38574	XP_034473054	Ipk2-PA	256	inositol polyphosphate multikinase	carcass	up
g27028	XP_034482300	Jon25Bi-PB	264	serine protease 1-like	carcass	down
g12072	XP_034476695	Mal-A4-PA	578	maltase A3	carcass	down
g26135	XP_034472391	CG4678-PI	527	carboxypeptidase D isoform X2	carcass	down
g34133	XP_034488835	CG31198-PA	937	membrane alanyl aminopeptidase	carcass	down
g2243	XP_034487938	NA	77	vasotab-like isoform X2	carcass	down
g27966	XP_034482043	mag-PA	401	lipase 3	carcass	down
g23610	XP_034482932	CG4835-PB	906	mucin-5AC isoform X2	carcass	down
g37148	XP_034481151	CG32277-PA	257	trypsin II-P29	carcass	down
g32344	XP_002055046	CG12576-PE	215	uncharacterized protein LOC6631944	carcass	down
g29728	XP_034478424	CG31089-PA	376	lipase 3-like	carcass	down
g565	XP_034489967	Antdh-PA	250	farnesol dehydrogenase-like	carcass	down
g738	XP_034490705	CG15773-PA	455	uncharacterized protein LOC117794238 isoform X1	carcass	down
g924	XP_034476024	Ser6-PA	266	serine protease SP24D-like	carcass	down
g1750	XP_034486839	CG12111-PB	163	C-type lectin 37Db-like	carcass	down
g2242	XP_034485669	CG42822-PA	114	uncharacterized protein LOC117790350	carcass	down
g2365	XP_034487988	CG17224-PD	299	uridine phosphorylase 1	carcass	down
g2413	XP_034485883	Alp13-PB	482	alkaline phosphatase, tissue-nonspecific isozyme isoform X2	carcass	down
g2492	XP_034117629	CG14292-PA	166	uncharacterized protein LOC117576718	carcass	down

g12335	XP_034479020	PGRP-SC2-PA	184	peptidoglycan-recognition protein SC2-like	carcass	down
g12548	XP_034476889	AttA-PA	236	attacin-B-like	carcass	down
g12681	KPI89340	NA	333	hypothetical protein ABL78_1569	carcass	down
g12793	GET86973	RpS4-PC	273	40S ribosomal protein S4, putative	carcass	down
g12912	KAf8287956	NA	415	hypothetical protein TcBrA4_0016180	carcass	down
g12930	XP_015653549	RpS3-PB	217	putative 40S ribosomal protein S3	carcass	down
g12938	KPI83570	NA	369	hypothetical protein ABL78_7393	carcass	down
g12977	KPI89879	Dhc36C-PB	4295	dynein heavy chain putative dynein heavy chain point mutation	carcass	down
g13133	XP_003861143	NA	485	DNAJ domain protein, putative	carcass	down
g13155	CCW71541	His3:CG33866-PA	130	unnamed protein product	carcass	down
g13172	EPY30378	Tpi-PC	250	triosephosphate isomerase	carcass	down
g13257	XP_001682869	CalpA-PA	979	putative calpain-like cysteine peptidase	carcass	down
g13269	EPY19523	NA	509	axoneme central apparatus protein	carcass	down
g13273	XP_001463451	betaTub56D-PB	443	beta tubulin	carcass	down
g13276	EPY27466	betaTub85D-PA	309	tubulin beta	carcass	down
g13372	EPY27681	NA	735	hypothetical protein STCU_05629	carcass	down
g13402	XP_015653859	Pgk-PE	486	putative mitochondrial phosphoglycerate kinase C, glycosomal (PGKC)	carcass	down
g13465	XP_010703706	NA	200	hypothetical protein LPMP_040600	carcass	down
g13508	XP_001561784	CIC-b-PA	978	putative chloride channel protein	carcass	down
g13521	EPY36686	fbp-PA	343	fructose-1,6-bisphosphatase, cytosolic	carcass	down
g13584	EPY25451	NA	564	lysosomal/endosomal membrane protein p67	carcass	down
g13593	XP_001563514	NA	436	conserved hypothetical protein	carcass	down
g13633	EPY14961	Jafrac1-PE	198	cytosolic trypanoxidin peroxidase, trypanosomatid typical 2-Cys peroxidase	carcass	down
g13635	EPY14961	Jafrac1-PE	197	cytosolic trypanoxidin peroxidase, trypanosomatid typical 2-Cys peroxidase	carcass	down
g13640	AGT02644	NA	1023	glutamate dehydrogenase	carcass	down
g13657	KPI88506	NA	465	putative phosphomevalonate kinase protein	carcass	down
g13812	EPY22894	NA	708	hypothetical protein STCU_08065	carcass	down
g13893	XP_015658323	NA	225	putative mitochondrial hypothetical protein	carcass	down
g13923	XP_001564466	AsnRS-PB	484	putative asparaginyl-tRNA synthetase	carcass	down
g13984	XP_009308748	NA	876	hypothetical protein DQ04_01261050	carcass	down
g13996	XP_003878716	NA	349	putative NLI-interacting factor	carcass	down
g14034	XP_015660586	Mdr50-PA	981	putative mitochondrial p-glycoprotein	carcass	down

g14044	ACS87900	NA	183	d-amastin	carcass	down
g14062	CCW60545	RpL13A-PB	222	unnamed protein product	carcass	down
g14063	EPY23003	NA	583	intraflagellar transport protein-like protein	carcass	down
g14197	XP_023174094	CG16704-PB	79	chymotrypsin inhibitor SCI-II-like	carcass	down
g14370	XP_034472516	CG5390-PB	486	phenoloxidase-activating factor 2-like	carcass	down
g14428	XP_034472325	NA	233	uncharacterized protein LOC117780055	carcass	down
g14637	XP_034475312	NA	107	uncharacterized protein LOC117782326	carcass	down
g14889	XP_034473685	CG1791-PC	206	ficolin-1-like	carcass	down
g14891	XP_034473685	CG8642-PB	201	ficolin-1-like	carcass	down
g15619	XP_015664553	RpL5-PG	305	putative 60S ribosomal protein L5	carcass	down
g15620	XP_001467225	RpL15-PI	204	putative ribosomal protein L15	carcass	down
g15630	CCW62052	NA	285	unnamed protein product	carcass	down
g15634	XP_009309792	RpS6-PB	247	40S ribosomal protein S6	carcass	down
g15749	GET92826	Gk2-PF	512	glycerol kinase, glycosomal, putative	carcass	down
g15766	AGT02546	Cth-PB	409	cystathionine gamma-synthase	carcass	down
g15777	KPI85258	Pgd-PB	479	putative 6-phosphogluconate dehydrogenase decarboxylating	carcass	down
g16035	CCW62421	NA	608	unnamed protein product	carcass	down
g16047	KPI89102	NA	501	nucleobase/nucleoside transporter	carcass	down
g16049	KPI89102	NA	353	nucleobase/nucleoside transporter	carcass	down
g16064	CCW68737	flw-PC	226	unnamed protein product	carcass	down
g16096	KPI89147	NA	989	oxidoreductase-like protein	carcass	down
g16100	XP_015663866	NA	240	putative mitochondrial P27 protein, putative (P27)	carcass	down
g16123	EKF29592	Hsc70-3-PE	651	glucose-regulated protein 78, putative	carcass	down
g16137	XP_015663807	Coq2-PA	495	putative mitochondrial Prenyltransferase	carcass	down
g16168	XP_003877442	NA	402	conserved hypothetical protein	carcass	down
g16219	EPY42055	SCOT-PA	487	succinyl-coA:3-ketoacid-coenzyme A transferase,mitochondrial precursor	carcass	down
g16330	GET89841	betaTub56D-PB	1161	hypothetical protein, conserved	carcass	down
g16331	TPP45131	alphaTub84B-PA	451	Tubulin/FtsZ family, GTPase domain protein	carcass	down
g16370	XP_951605	NA	448	hypothetical protein, conserved	carcass	down
g16372	EPY36406	NA	330	hypothetical protein AGDE_06929	carcass	down
g16380	KPI87332	NA	494	transcription elongation factor-like protein	carcass	down
g16412	XP_003862902	NA	492	hypothetical protein, unknown function	carcass	down
g16420	XP_009307693	NA	2689	Hydin	carcass	down
g16426	XP_028884965	NA	425	uncharacterized protein TM35_000073230	carcass	down
g16440	KPI83552	NA	240	hypothetical protein ABL78_7409	carcass	down
g16455	PBJ80770	NA	738	hypothetical protein BCY84_00972	carcass	down

g16485	XP_003877276	Pdxk-PC	301	putative Pyridoxal kinase	carcass	down
g16496	TPP45131	alphaTub84B-PA	451	Tubulin/FtsZ family, GTPase domain protein	carcass	down
g16520	EPY35159	NA	180	amastin-like surface protein-like protein	carcass	down
g16530	CCW61714	dj-1beta-PB	197	unnamed protein product	carcass	down
g16555	XP_015660372	NA	143	hypothetical protein ABB37_04169	carcass	down
g16585	CCW64533	NA	327	unnamed protein product	carcass	down
g16661	WP_116046257	NA	193	galactoside O-acetyltransferase	carcass	down
g16746	XP_001687597	NA	131	putative prefoldin subunit	carcass	down
g16796	KPI88047	NA	144	hypothetical protein ABL78_2873	carcass	down
g16893	KPI86692	NA	832	putative major vault protein	carcass	down
g17106	XP_015664315	NA	443	hypothetical protein ABB37_00198	carcass	down
g17117	XP_001468859	RpS3A-PF	247	putative 40S ribosomal protein S3A	carcass	down
g17128	GET91420	RpL18A-PA	179	60S ribosomal protein L18a, putative	carcass	down
g17142	XP_010702496	NA	148	CBS domain protein, conserved	carcass	down
g17160	XP_003875055	Pdi-PE	459	protein disulfide isomerase	carcass	down
g17162	XP_001569337	CCT8-PA	537	putative T-complex protein 1, theta subunit	carcass	down
g17176	EPY23416	CG8520-PB	444	hypothetical protein STCU_07721	carcass	down
g17243	XP_015664449	NA	1117	putative mitochondrial NADH-dependent fumarate reductase	carcass	down
g17265	XP_003722493	CG10467-PA	374	aldose 1-epimerase-like protein	carcass	down
g17302	XP_003873841	NA	112	conserved hypothetical protein	carcass	down
g17311	CCW64643	NA	599	unnamed protein product	carcass	down
g17336	KPI85553	NA	375	putative cyclophilin putative PPIase putative rotamase putative CYP13	carcass	down
g17340	GET87481	deltaCOP-PB	543	coatomer delta subunit-like protein	carcass	down
g17364	EPY20134	NA	462	sucrose-phosphate synthase-like protein	carcass	down
g17410	TPP52586	NA	342	Phosphorylase family protein	carcass	down
g17429	XP_001562894	Rab1-PA	221	putative rab1 small GTP-binding protein	carcass	down
g17439	EPY37453	NA	323	hypothetical protein AGDE_06481	carcass	down
g17543	XP_015654325	vas-PC	688	putative mitochondrial DEAD/DEAH box helicase	carcass	down
g21450	XP_015652381	NA	671	putative mitochondrial cytoskeleton-associated protein CAP5.5	carcass	down
g21512	XP_009309812	NA	572	receptor-type adenylate cyclase	carcass	down
g21553	XP_015657389	NA	384	hypothetical protein ABB37_06014	carcass	down
g21561	EPY25094	NA	767	vacuolar-type proton translocating pyrophosphatase 1	carcass	down
g21649	CCW69254	NA	162	unnamed protein product	carcass	down
g21650	XP_015657068	RpS27-PB	162	putative 60S ribosomal protein L34	carcass	down
g21651	EPY25925	RpL10Ab-PA	214	large subunit ribosomal protein L10Ae	carcass	down
g21654	XP_015659840	GlyRS-PA	629	putative glycyl tRNA synthetase	carcass	down

g21674	AGT02540	Ahcy-PC	437	adenosylhomocysteinase	carcass	down
g21733	XP_010703375	NA	1792	C2 domain protein, putative	carcass	down
g21821	GET93414	NA	277	hypothetical protein, conserved	carcass	down
g21845	CCW65631	NA	238	unnamed protein product	carcass	down
g21864	XP_009313767	alphaCOP-PA	1208	putative coatomer alpha subunit	carcass	down
g21876	XP_015660990	btv-PD	4243	putative Cytoplasmic dynein 2 heavy chain (DYNC2H1)	carcass	down
g21914	XP_029228657	Pgm2a-PA	576	putative phosphomannomutase-like protein	carcass	down
g21919	TPP46746	Spt-I-PC	487	Aminotransferase class I and II family protein	carcass	down
g21926	KPI88001	Vha68-2-PF	610	putative vacuolar ATP synthase catalytic subunit A	carcass	down
g21927	XP_029235955	NA	148	large subunit ribosomal protein L21e	carcass	down
g21951	EPY35770	NA	389	oxidoreductase-like protein	carcass	down
g22043	KAF8285757	NA	407	hypothetical protein TcYC6_0032640	carcass	down
g22434	XP_001988090	NA	211	uncharacterized protein LOC6562607	carcass	down
g22459	ALC38490	CG9500-PB	454	maker431, partial	carcass	down
g22575	XP_034475432	NimB3-PA	124	uncharacterized protein LOC117782522	carcass	down
g22675	XP_034472549	CG17633-PA	426	zinc carboxypeptidase A 1	carcass	down
g23636	XP_034484118	Arp3-PB	409	actin-related protein 3	carcass	down
g23638	XP_034482746	CG10472-PA	130	plasma kallikrein-like	carcass	down
g23647	XP_034483240	oxt-PB	238	xylosyltransferase oxt isoform X2	carcass	down
g23657	XP_034482746	CG10472-PA	97	plasma kallikrein-like	carcass	down
g23703	XP_034480827	RpL18-PB	188	60S ribosomal protein L18	carcass	down
g23704	XP_034111739	mRpL50-PA	193	39S ribosomal protein L50, mitochondrial	carcass	down
g23710	XP_017840973	RpS9-PE	195	40S ribosomal protein S9	carcass	down
g23737	GET90247	NA	736	sodium/sulphate symporter, putative	carcass	down
g23745	CCW60170	NA	169	unnamed protein product	carcass	down
g23768	GET90196	Pp4-19C-PI	305	protein phosphatase 2A catalytic subunit, putative	carcass	down
g26083	XP_034482770	MgstI-PC	145	microsomal glutathione S-transferase 1 isoform X1	carcass	down
g26246	CCW60515	NA	319	unnamed protein product	carcass	down
g26273	EPY24850	UQCR-C1-PB	428	mitochondrial processing peptidase	carcass	down
g26291	XP_003722515	NA	151	putative short chain dehydrogenase	carcass	down
g26326	KPI86395	NA	415	hypothetical protein ABL78_4546	carcass	down
g26369	XP_003872702	NA	820	conserved hypothetical protein	carcass	down
g26383	GET86244	CG9272-PA	261	endonuclease III, putative	carcass	down
g26424	XP_015658619	NA	502	putative mitochondrial hypothetical protein	carcass	down
g26451	KPI84447	NA	349	putative sterol 24-c-methyltransferase	carcass	down

g26489	XP_001565783	NA	282	conserved hypothetical protein	carcass	down
g26519	EPY23018	NA	666	hypothetical protein STCU_07940	carcass	down
g26588	XP_001466130	Dhc1-PF	4693	putative dynein heavy chain	carcass	down
g26629	XP_009307723	NA	414	putative sphingosine kinase A, B	carcass	down
g26642	XP_001682657	CG9281-PD	611	putative ATP-binding cassette protein subfamily F, member 2	carcass	down
g26686	XP_028886554	Nap1-PC	304	putative nucleosome assembly protein	carcass	down
g26990	EPY20135	NA	921	WD repeat-containing protein 96	carcass	down
g27029	XP_034482300	Jon25Bi-PB	264	serine protease 1-like	carcass	down
g27033	XP_034482303	Jon66Cii-PA	208	serine protease 1-like	carcass	down
g29971	XP_034489697	Ser6-PA	255	serine protease SP24D-like	carcass	down
g30507	XP_034486562	Bace-PB	375	lysosomal aspartic protease-like	carcass	down
g30547	XP_034489914	scu-PA	255	3-hydroxyacyl-CoA dehydrogenase type-2	carcass	down
g30995	XP_034484998	Ho-PA	300	heme oxygenase 2	carcass	down
g31858	XP_034477134	AttC-PB	229	attacin-C	carcass	down
g32004	XP_034477621	NA	94	dipteracin A	carcass	down
g32125	XP_034477538	PCB-PB	1181	pyruvate carboxylase, mitochondrial-like	carcass	down
g32630	KPI84018	NA	489	hypothetical protein ABL78_6927	carcass	down
g32672	EPY16042	NA	553	hypothetical protein AGDE_16857	carcass	down
g32705	CCW64620	NA	1654	unnamed protein product	carcass	down
g33870	XP_034487927	NA	91	uncharacterized protein LOC117792078	carcass	down
g34155	XP_034488603	CG17191-PA	336	phospholipase A1 VesT1.02-like	carcass	down
g34490	XP_034475376	Acbp1-PB	90	acyl-CoA-binding protein homolog	carcass	down
g35003	XP_034134501	CG4364-PA	145	pescadillo homolog	carcass	down
g35186	XP_034488582	CG10562-PA	256	uncharacterized protein LOC117792519	carcass	down
g35668	XP_034484005	ORMDL-PA	106	ORM1-like protein	carcass	down
g35669	XP_034482797	CG32444-PA	375	aldose 1-epimerase isoform X1	carcass	down
g35683	XP_034111367	CG5157-PA	123	cytochrome b5	carcass	down
g35988	XP_034477139	Obp56d-PB	132	general odorant-binding protein 56d-like	carcass	down
g36112	XP_034478334	lambdaTry-PA	267	vitellin-degrading protease-like	carcass	down
g36261	EDW31890	Khc-PA	128	GL11356	carcass	down
g36762	XP_034484199	Hexo1-PF	605	chitooligosaccharidolytic beta-N-acetylglucosaminidase	carcass	down
g36812	XP_034480767	CG4842-PA	255	alcohol dehydrogenase 1	carcass	down
g36817	XP_034484171	Pdh-PC	261	fat body protein 2	carcass	down
g36860	XP_034481663	l(3)72Dr-PA	321	gamma-glutamyl hydrolase isoform X1	carcass	down
g36865	XP_034481535	MED10-PB	133	mediator of RNA polymerase II transcription subunit 10	carcass	down
g36877	XP_034482340	Kaz1-ORFB-PF	109	uncharacterized protein LOC117787829	carcass	down

g37785	XP_034481691	CG5506-PA	127	uncharacterized protein LOC117787310 isoform X3	carcass	down
g37786	XP_034483038	CG16775-PB	186	uncharacterized protein LOC117788383	carcass	down
g37789	XP_034481690	CG5506-PA	181	uncharacterized protein LOC117787310 isoform X2	carcass	down
g37840	XP_034480276	NA	162	uncharacterized protein LOC117786244 isoform X3	carcass	down
g38563	XP_034487611	Sfxn1-3-PB	321	sideroflexin-1-3-like	carcass	down
g31432	XP_034480827	RpL18-PB	188	60S ribosomal protein L18	ovary	up
g34753	XP_034490598	tyf-PN	2299	pneumococcal serine-rich repeat protein	ovary	up
g31425	XP_034482470	D19A-PA	831	zinc finger protein 85-like	ovary	up
g31423	XP_034484178	D19A-PA	719	gastrula zinc finger protein XICGF58.1-like	ovary	up
g31429	XP_030569913	CG10274-PB	869	zinc finger protein 540-like	ovary	up
g31422	XP_034484177	D19A-PA	734	zinc finger protein 675-like	ovary	up
g34755	XP_034489569	Nsun2-PA	746	tRNA (cytosine(34)-C(5))-methyltransferase	ovary	up
g31431	XP_034111737	BHD-PA	482	folliculin	ovary	up
g31424	XP_034484174	D19A-PA	765	zinc finger protein 708-like	ovary	up
g31433	XP_034111739	mRpL50-PA	193	39S ribosomal protein L50, mitochondrial	ovary	up
g53	XP_034479988	CG12384-PA	96	death-associated protein 1	ovary	up
g11476	XP_034476548	CG15701-PA	211	uncharacterized protein LOC117783305	ovary	up
g11477	XP_034476548	CG15701-PA	653	uncharacterized protein LOC117783305	ovary	up
g11481	XP_034477113	NA	389	uncharacterized protein LOC117783693	ovary	up
g12562	XP_034477657	eIF3k-PA	222	eukaryotic translation initiation factor 3 subunit K	ovary	up
g23523	XP_034480538	mu2-PC	1276	uncharacterized protein LOC117786408	ovary	up
g27181	XP_034476077	fest-PA	483	uncharacterized protein LOC117783031	ovary	up
g31421	XP_034484179	D19B-PA	154	zinc finger protein 431-like	ovary	up
g31434	XP_034480825	Neos-PB	243	uncharacterized protein LOC117786597	ovary	up
g34646	XP_034472969	rempA-PB	1548	LOW QUALITY PROTEIN: intraflagellar transport protein 140 homolog	ovary	up
g34752	XP_034479100	Tip60-PC	601	histone acetyltransferase Tip60	ovary	up
g34757	XP_034478801	cib-PE	106	thymosin beta	ovary	up
g34758	XP_034476309	brn-PA	325	beta-1,3-galactosyltransferase brn	ovary	up
g35729	XP_034477128	CG9084-PB	275	phospholipid scramblase 3	ovary	up
g36872	XP_034482901	CG32152-PD	359	protein arginine N-methyltransferase 1	ovary	up
g31594	XP_034476321	NA	433	uncharacterized protein LOC117783168	ovary	down
g35988	XP_034477139	Obp56d-PB	132	general odorant-binding protein 56d-like	ovary	down
g22313	XP_034097411	NA	381	uncharacterized protein LOC117563283	ovary	down
g33510	XP_034486469	NA	508	putative aminopeptidase W07G4.4	ovary	down

g35930	XP_030566462	NA	671	uncharacterized protein LOC115766612	ovary	down
g374	XP_034489491	NA	758	regulator of G-protein signaling 7	ovary	down
g22644	XP_034489021	NA	536	membrane metallo-endopeptidase-like 1	ovary	down
g1208	XP_034490488	NA	726	sodium-dependent neutral amino acid transporter B(0)AT3	ovary	down
g966	XP_034490122	NA	100	protein roadkill-like	ovary	down
g11884	XP_034479242	NA	1429	pikachurin, partial	ovary	down

APPENDIX C

Supplementary Materials Chapter 4

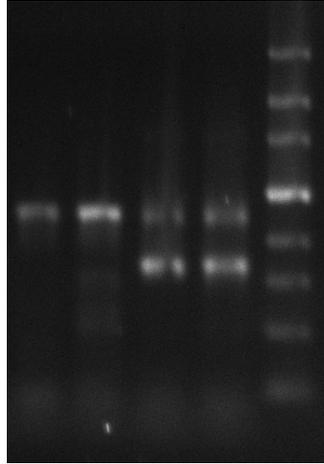


Figure S4.1: Gel image for *Wolbachia* strain PCR. First two samples are independent DNA extractions from the Alb *D. recens* stock and the last two lanes are results from independent extractions of the RW *D. recens* stock.

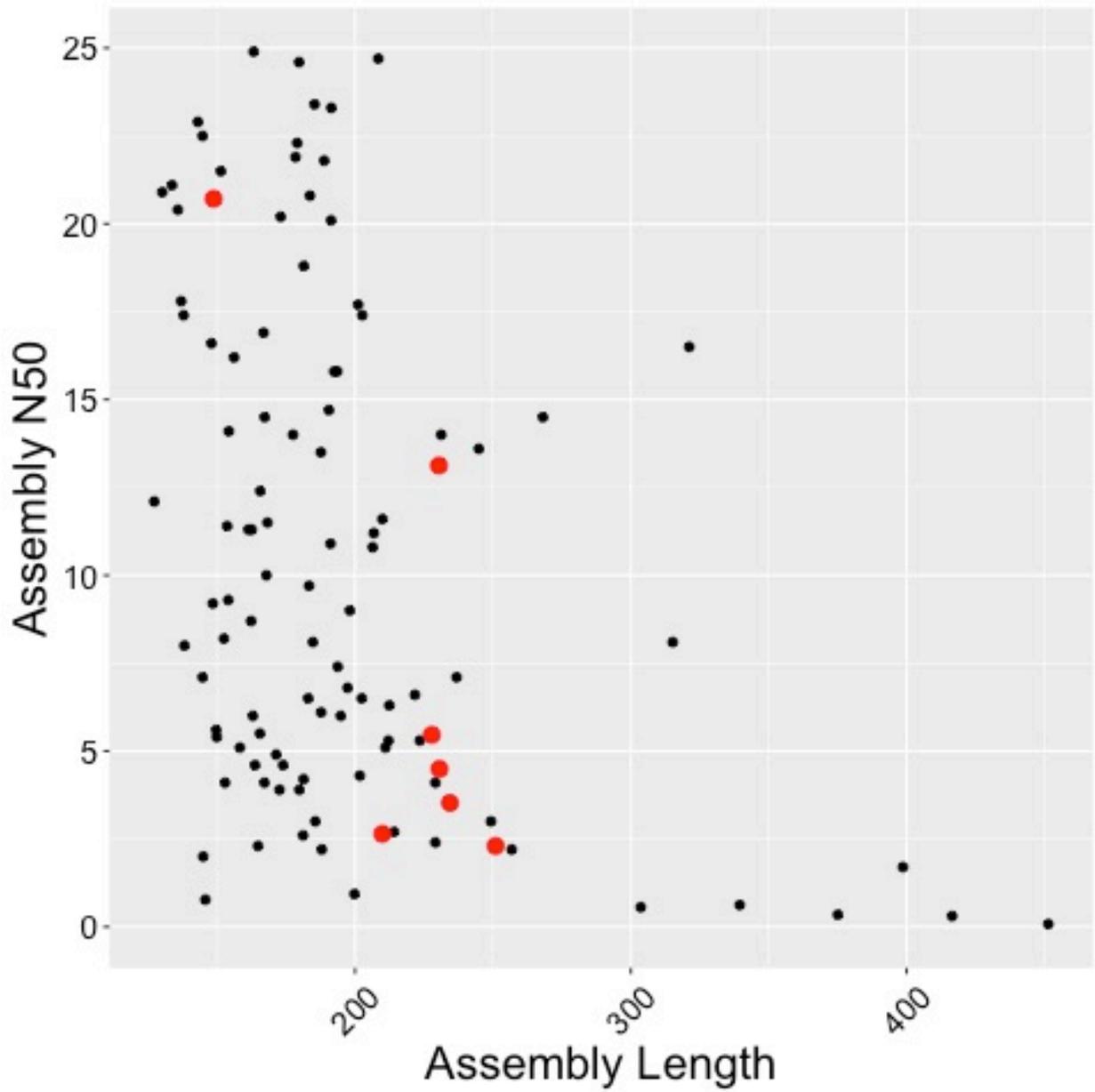


Figure S4.2: Comparison of genome assembly length by assembly N50. Black data points are genomes from Kim et al. (2021). Red data points are genome assemblies presented here.

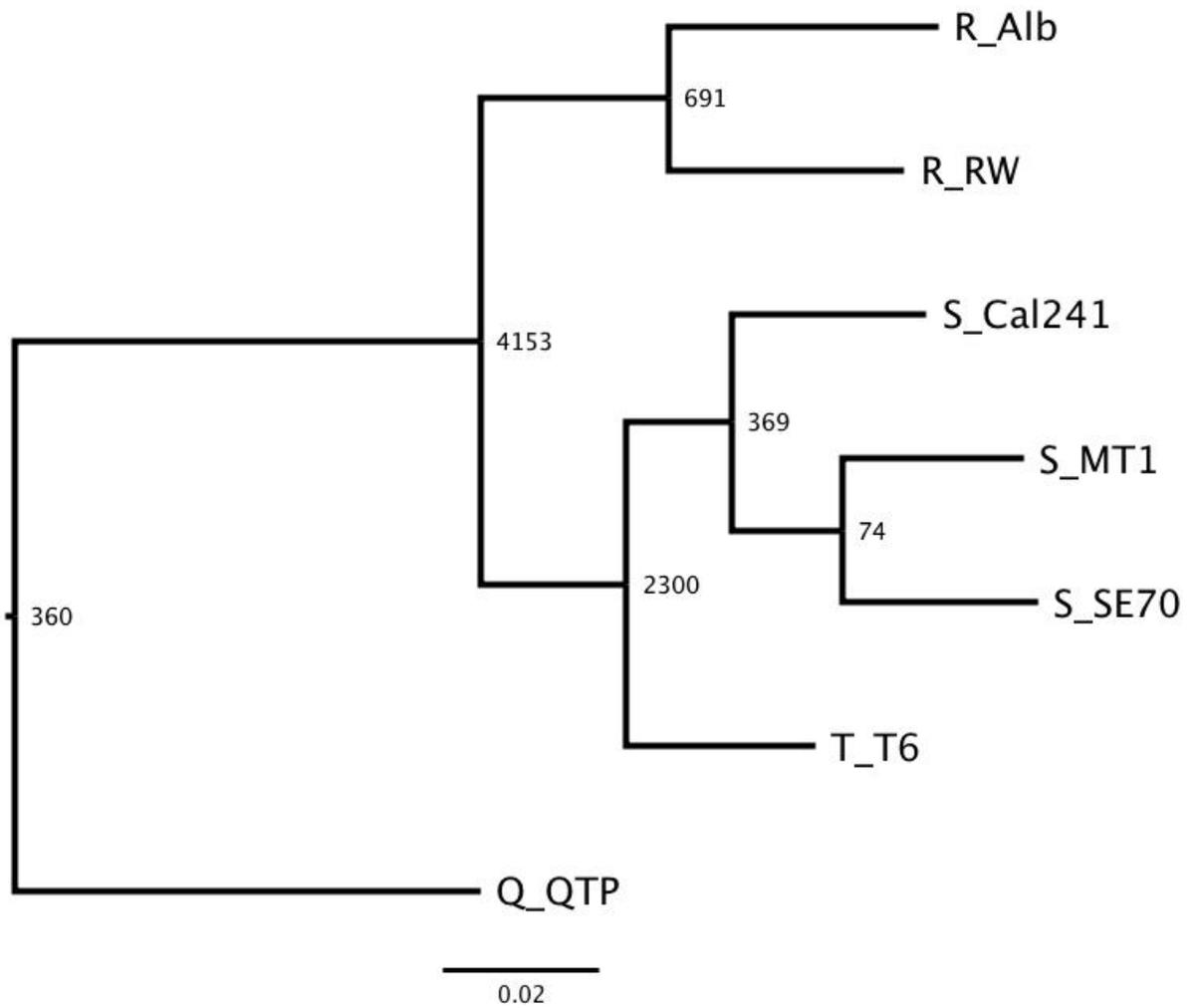


Figure S4.3: Species tree with estimated gene duplication events. Estimates were predicted from Braker assembly annotations in Orthofinder (Emms & Kelly, 2019) run on default parameters.

Table S4.1: Oxford Nanopore data for genome assemblies.

Sample	Platform	type	Num seqs	Sum length	Min length	Avg length	Max length	Q1	Q2	Q3	Sum gap	N50	Q20 (%)	Q30 (%)
Cal241	ONT	DNA	3,869,516	30,386,920,726	1	7,852.9	250,906	1,075	3,458	9,596	0	19,498	69	33
QTP	ONT	DNA	4,686,195	24,213,102,295	1	5,166.9	154,497	830	2,391	6,261	0	11,723	68.98	32.2
RW	ONT	DNA	1,874,020	13,328,410,003	1	7,112.2	184,676	637	2,660	8,565	0	19,515	70.21	35.0 3
T6	ONT	DNA	1,734,646	14,427,370,745	1	8,317.2	228,574	879	3,529	9,370	0	23,272	71.6	35.6 8
SE70	ONT	DNA	1,182,809	6,567,959,630	1	5,552.8	211,315	756	2,295	5,840	0	14,535	64.35	29.3 7
MT1	ONT	DNA	6,557,020	25,196,548,719	1	3,842.7	103,628	974	2,414	5,040	0	6,749	68.15	30.6 1
Alb	ONT	DNA	1,855,957	13,798,089,522	1	7,434.5	226,697	851	2,722	8,240	0	20,911	69.22	34.7 7

Table S4.2: Illumina short read data for genome assemblies.

Sample	Platform	Molecule	nReads.prefilter	nBases.prefilter	nReads	Read length
Cal241 female	Illumina	DNA	37896922	5722435222	37687848	150 bp PE
Cal241 female	Illumina	RNA	14310528	2160889728	14249820	150 bp PE
Cal241 male	Illumina	RNA	15279448	2307196648	15182438	150 bp PE
QTP female	Illumina	DNA	31105568	4696940768	30916426	150 bp PE
QTP female	Illumina	RNA	14317748	2161979948	14211034	150 bp PE
QTP male	Illumina	RNA	10591704	1599347304	10531468	150 bp PE
RW female	Illumina	DNA	137603272	20778094072	134717352	150 bp PE
RW female	Illumina	RNA	37484968	5660230168	37054480	150 bp PE
RW male	Illumina	RNA	52506488	7928479688	52248072	150 bp PE
T6 female	Illumina	DNA	134443622	20300986922	132402652	150 bp PE
T6 female	Illumina	RNA	47461574	7166697674	47222404	150 bp PE
T6 male	Illumina	RNA	51785262	7819574562	25645649	150 bp PE
SE70 female	Illumina	DNA	140894956	21275138356	137056738	150 bp PE
SE70 female	Illumina	RNA	46268736	6986579136	46008150	150 bp PE
SE70 male	Illumina	RNA	48069114	7258436214	47759746	150 bp PE
MT1 female	Illumina	DNA	153228632	23137523432	150051234	150 bp PE
MT1 female	Illumina	RNA	38225534	5772055634	37756132	150 bp PE
MT1 male	Illumina	RNA	41767402	6306877702	41407064	150 bp PE
Alb female	Illumina	DNA	30523684	4609076284	30347872	150 bp PE
Alb female	Illumina	RNA	198706754	14903006550	189476478	75 bp PE

Table S4.3: *Wolbachia* isolates used for phylogenetic analysis. All sequences were downloaded in batch mode from the PubMLST database (<https://pubmlst.org/organisms/wolbachia-spp>)

Host Species	<i>Wolbachia</i> Strain	Isolate ID
<i>Drosophila melanogaster</i>	wMel	1
<i>Drosophila innubila</i>	wInn	6
<i>Drosophila neotestacea</i>	wNeo	7
<i>Drosophila orientacea</i>	wOri	8
<i>Drosophila recens</i>	wRec	9
<i>Drosophila simulans</i>	wAu	10
<i>Drosophila simulans</i>	wRi	11
<i>Drosophila simulans</i>	wNO	27
<i>Culex pipiens</i>	wPip	29
<i>Drosophila ananassae</i>	wAna	250
<i>Drosophila suzukii</i>	wSuz	607
<i>Culex quinquefasciatus</i>	wPip	1808
<i>Diaphorina citri</i>	wCit	1810

Table S4.4: SNVs between *wRec* variants. Gene annotations for *wMel* reference (Refseq: GCF_000008025.1). First nucleotide and amino acid in SNV and Mutation columns are the state in RW *wRec*.

SNV	Locus RW <i>wRec</i>	Position RW <i>wRec</i>	Locus Alb <i>wRec</i>	Position Alb <i>wRec</i>	Gene	Mutation
a <--> g	WR_RW_00006	5225	WR_ALB_00006	5225	type IV secretion system protein VirB9	synonymous
t <--> c	WR_RW_00025	31524	WR_ALB_00025	31524	transposase, IS5 family, OrfB	synonymous
a <--> g	WR_RW_00026	31762	WR_ALB_00026	31762	transposase, IS5 family, OrfA	Leucine <--> Proline
t <--> c	WR_RW_00037	43181	WR_ALB_00037	43181	transposase, IS5 family, OrfB	synonymous
a <--> c	NA	248554	NA	248553	intergenic	NA
g <--> a	WR_RW_00274	264567	WR_ALB_00275	264566	reverse transcriptase, putative	synonymous
t <--> c	WR_RW_00282	271773	WR_ALB_00283	271772	hypothetical protein	Leucine <--> Proline
c <--> t	NA	275128	NA	275127	intergenic	NA
g <--> a	WR_RW_00344	341799	WR_ALB_00345	341799	pyrG CTP synthase	Glutamic acid <--> Lysine
a <--> c	NA	380213	NA	380205	intergenic	NA
c <--> t	NA	380214	NA	380206	intergenic	NA
a <--> g	NA	380215	NA	380207	intergenic	NA
g <--> a	NA	380225	NA	380217	intergenic	NA
g <--> a	NA	380241	NA	380233	intergenic	NA
t <--> c	NA	380271	NA	380263	intergenic	NA
g <--> a	NA	380277	NA	380269	intergenic	NA
t <--> c	NA	380311	NA	380303	intergenic	NA
t <--> c	NA	380314	NA	380306	intergenic	NA
a <--> c	NA	380316	NA	380308	intergenic	NA
a <--> c	NA	380321	NA	380313	intergenic	NA
c <--> t	NA	380342	NA	380334	intergenic	NA
g <--> a	NA	392751	NA	392743	intergenic	NA
a <--> g	NA	392791	NA	392783	intergenic	NA

t <--> c	WR_RW_00449	437779	WR_ALB_00450	437771	dnaB replicative DNA helicase	Leucine <--> Proline
t <--> a	WR_RW_00460	449991	WR_ALB_00461	449983	hypothetical protein	Leucine <--> Stop
t <--> c	NA	468714	NA	468706	intergenic	NA
t <--> c	NA	531422	NA	531414	intergenic	NA
t <--> c	WR_RW_00635	603574	WR_ALB_00605	578451	hypothetical protein	Serine <--> Proline
c <--> t	WR_RW_00636	604933	WR_ALB_00606	579810	reverse transcriptase, putative	synonymous
g <--> a	WR_RW_00681	645648	WR_ALB_651	620526	hypothetical protein	synonymous
c <--> t	WR_RW_00756	724291	WR_ALB_00726	699169	transposase, IS5 family, OrfA	synonymous
c <--> t	NA	833127	NA	808014	intergenic	NA
g <--> a	NA	833155	NA	808042	intergenic	NA
t <--> c	NA	833167	NA	808054	intergenic	NA
c <--> t	WR_RW_00897	862499	WR_ALB_00867	837386	surE stationary phase survival protein SurE	synonymous
g <--> a	WR_RW_00968	928258	WR_ALB_00938	903145	gidA glucose-inhibited division protein A	Glycine <--> Serine
g <--> a	WR_RW_00986	944135	WR_ALB_00956	919022	reverse transcriptase, putative	synonymous
t <--> c	NA	974712	NA	949599	intergenic	NA
c <--> t	WR_RW_01034	986241	WR_ALB_01004	961128	membrane protein, putative	Leucine <--> Phenylalanine
g <--> a	WR_RW_01042	991414	WR_ALB_10012	966301	IS110 family transposase ISWpi12	Proline <--> Leucine
g <--> a	WR_RW_01043	991414	WR_ALB_01013	966301	hypothetical protein	synonymous
a <--> g	WR_RW_01085	1022268	WR_ALB_01055	997155	tyrS tyrosyl-tRNA synthetase	Threonine <--> Alanine
g <--> a	WR_RW_01153	1088300	WR_ALB_01123	1063187	gyrA DNA gyrase, A subunit	synonymous
a <--> g	WR_RW_01171	1100015	WR_ALB_01141	1074902	rcsC Sensor histidine kinase RcsC	isoleucine <--> methionine
c <--> t	WR_RW_1205	1133948	WR_ALB_01175	1108835	glutathione-regulated potassium-efflux system protein	Methionine <--> Isoleucine
t <--> c	WR_RW_01276	1202653	WR_ALB_1246	1177545	nusA M utilisation substance protein A	Isoleucine <--> Valine

Table S4.5: Indels distinguishing RW wRec and Alb wRec.

Indel Length	Locus RW wRec	Position RW wRec	Locus Alb wRec	Position Alb wRec	Gene(s)	Effect
1217	WR_RW_00076	72937	WR_ALB_00076	72937	transposase IS4 family	Alignment issue, CDS is identical
4	WR_RW_00110	106228	WR_ALB_00110	74155	hypothetical protein	Alignment issue, CDS is identical
652	WR_RW_00111	106875	WR_ALB_00111	106875	transposase IS4 family	Alignment issue, CDS is identical
1	NA	332973	NA	332972	intergenic	NA
1	WR_RW_00427	417685	WR_ALB_00428	417676	TrkH Trk system potassium uptake protein	premature stop, protein truncated by 11 aa in Alb wRec
1	WR_RW_00428	417685	WR_ALB_00429	417676	TrkI Trk system potassium uptake protein	7 aa deletion from RW wRec
722	NA	569658	NA	543814	intergenic	NA
1	NA	642931	NA	617808	intergenic	NA
9	WR_RW_00831	795416	WR_ALB_00801	770294	hypothetical protein	3 aa deletion in RW wRec
5	NA	1160708	NA	1135595	intergenic	NA
8	NA	380197	NA	380197	intergenic	NA
1	WR_RW_00400	393907	WR_RW_00401	393907	transposase IS4 family, OrfA	6 aa deletion in Alb wRec
25837	WR_RW_00553 - WR_RW_00583	543813	missing	543813	See table 6	WO-B phage-related loci missing in Alb wRec

Table S4.6: Location of BUSCO orthologs in the *D. innubila* that are missing in the Q_QTP assembly. *D. innubila* Genbank accession: GCA_004354385.2.

Scaffold	Chromosome	N Missing
CM027943.1	3L	6
CM027942.1	2R	3
CM015046.2	X	3
CM027941.1	2L	110
CM027944.1	3R	5
SKCT02000006.1	Unknown	2