GBE

Comparative Genomics of the Testacea Group of *Drosophila* Reveals Introgression and Variation in Chromosome Size and Structure

Theresa Erlenbach [1], Ryan M.R. Gawryluk [2], Steve J. Perlman [2], Robert L. Unckless [3], Kelly A. Dyer [1,*]

[1]Department of Genetics, University of Georgia, Athens, GA, USA
[2]Department of Biology, University of Victoria, Victoria, British Columbia, Canada
[3]Department of Molecular Biosciences, The University of Kansas, Lawrence, KS, USA

*Corresponding author: E-mail: kdyer@uga.edu.

Accepted: November 12, 2025

Abstract

Comparative genomic analyses among closely related species provide an opportunity to assess their evolutionary history. The relatedness between species can depend on a variety of factors, including reproductive isolation, introgression, and incomplete lineage sorting, and this can impact divergence across the genome. Here, we use a combination of long- and short-read sequencing and HI-C scaffolding to assemble genomes for each of the four species in the *testacea* species group of *Drosophila*, including *D. testacea*, *D. orientacea*, *D. neotestacea*, and *D. putrida*, and its outgroup, *D. bizonata*. First, among species, we find many structural rearrangements across the genome as well as a large size difference in the dot chromosome that we infer is due to the expansion of repetitive elements. Second, we assess phylogenetic discordance and uncover a difference in the phylogeny inferred from genes on Muller E and the mitogenome relative to the rest of the genome, which may be due to recent hybridization. Lastly, we assess the rate of molecular evolution of genes shared across all species and identify genes evolving at different rates across the phylogeny. Our results present genomic resources for this species group and begin to probe into some of the evolutionary characteristics that contribute to variation in genome structure, while highlighting the need for high-quality genome resources to fully capture and understand the evolutionary history among closely related species.

Key words: phylogenetic discordance, dot chromosome, evolutionary history, hybridization, reproductive isolation.

Significance

Among species, forces such as hybridization and introgression can impact selection and the evolution of genome structure, though a contiguous assembly is required to fully elucidate these factors. Here, we assemble genomes for the *testacea* species group of *Drosophila* and a closely related outgroup species, characterize structural differences among the chromosomes, and identify regions of introgression. Our data provide the first genomic resources for this species group that can be utilized to understand the evolutionary characteristics of these species, as well as provide a baseline for future work on how introgression and hybridization impact the structural evolution of genomes, even among closely related species.

Introduction

As new species form, their genomes may not diverge at a uniform rate (Feder and Nosil 2010; Nosil and Feder 2012; Sendell-Price et al. 2020; Schluter and Rieseberg 2022). This is especially true when radiations occur rapidly. Among closely related species, factors such as incomplete lineage sorting, gene flow, and introgression can cause genomic regions to diverge at different rates and thus impact the interpretations of species trees (Pollard et al. 2006; Pease et al. 2016; Edelman et al. 2019; Dagilis and Matute 2023). For recently diverged groups, different genomic regions can indicate different aspects of the evolutionary history. For instance, these data can be used to identify structural differences or rapidly evolving genes that may promote divergence, as well as regions of similarity due to incomplete lineage sorting or hybridization.

Hybridization occurs when species come into secondary contact (Harrison and Larson 2014; Moran et al. 2021), and it is widespread across plants and animals (Teeter et al. 2008; Payseur and Rieseberg 2016; Schumer et al. 2016; Turissini and Matute 2017; Eberlein et al. 2019). A common outcome of hybridization is introgression of genetic material between species, which can both promote or constrain adaptation. The extent of introgression and the genomic regions that are able to move between species can be limited by selection against hybrids (Svedin et al. 2008; Hedrick 2013; Arnegard et al. 2014; Schumer et al. 2018; Marques et al. 2019). In a hybrid individual, novel combinations of alleles come into contact in the same organism, and the interactions between these alleles can have negative fitness consequences for hybrids (Dobzhansky 1937; Presgraves 2010; Maheshwari and Barbash 2011). These hybrid incompatibilities can restrict gene flow, preventing introgression at the incompatible loci and the regions linked to them (Teeter et al. 2008; Zuellig and Sweigart 2018). Thus, regions of the genome that contribute to species differences are expected to be more differentiated than regions that do not.

The genomic location where hybrid incompatibilities occur shares some patterns across species. For example, loci involved in hybrid incompatibilities tend to arise on X chromosomes, where forces such as faster-X evolution and X-chromosome meiotic drive can also be found (Presgraves 2008; Qvarnström and Bailey 2009; Maheshwari and Barbash 2011; Trier et al. 2014). Further, structural rearrangements can maintain species boundaries in the face of hybridization and gene flow (Noor et al. 2001). In particular, chromosomal inversions can inhibit recombination in heterozygotes (Hoffmann and Rieseberg 2008; Kirkpatrick 2010; Stevison et al. 2011; Zhang et al. 2021) and cause meiosis to fail in hybrids, resulting in hybrid incompatibility (Stebbins 1958; Coyne and Orr 2004). Therefore, as species diverge and their genomes accumulate structural rearrangements between them, it is expected that their genetic incompatibility will also increase.

If hybridization occurs and fertile offspring are produced, genetic exchange can occur between species. Introgression is influenced by a variety of factors, including the genetic relatedness between species (Coyne and Orr 1997), the geographic proximity where they come into physical and reproductive contact (Harrison 2012; Harrison and Larson 2014), and the timing and success of reproductive events (Coyne and Orr 2004; Hamlin et al. 2020). Introgressed alleles can contribute to adaptation, allowing species to evolve more rapidly in new environments (Owens et al. 2016; Khodwekar and Gailing 2017; Edelman and Mallet 2021). There is also evidence of reoccurring gene families commonly introgressing between species, including genes involved in the immune system (Fijarczyk et al. 2018; De Cahsan et al. 2021), pigmentation (Giska et al. 2019; Moest et al. 2020; Semenov et al. 2021), and selfish elements (Hurst and Schilthuizen 1998; Meiklejohn et al. 2018; Svedberg et al. 2021).

Drosophila have been used as a model system to understand genome biology and evolution, as well as the consequences of hybridization and introgression on patterns of molecular evolution (Lohse et al. 2015; Turissini and Matute 2017; Suvorov et al. 2022; Yusuf et al. 2022). The initial 12 *Drosophila* genome project paved the way for comparative genomic analyses (Clark et al. 2007). These studies confirmed that over large phylogenetic distances, there was conservation of genes on chromosomal "Muller' elements despite many structural variants between them, and these analyses identified many groups of genes that were evolving rapidly across the genus (Clark et al. 2007; Hahn et al. 2007; Bhutkar et al. 2008). Across *Drosophila*, there is a conserved X chromosome (Muller A), four large autosomes (Muller B-E), and a small nonrecombining autosomal dot chromosome (Muller F) (Patterson and Stone 1952). Initial comparisons between closely related species were limited to the *melanogaster* group (Ranz et al. 2007; Garrigan et al. 2012). With the advent of inexpensive and efficient genome sequencing, it is now possible to produce high-quality genomes for groups of related non-model species. This work has led to insights about the importance of introgression, structural evolution, and molecular evolution in recent adaptations both in *Drosophila* (Kulathinal et al. 2009; Guillén et al. 2015; Lohse et al. 2015; Turissini and Matute 2017; Mai et al. 2020; Yusuf et al. 2022) as well as other insects that have been models for ecological and evolutionary questions (Crawford et al. 2015; Fallon et al. 2018; Lucek et al. 2019; Martin et al. 2019; Herrig et al. 2024).

In this study, we investigate genome evolution of the *testacea* species group of *Drosophila*, which consists of four species: *D. putrida*, *D. testacea*, *D. orientacea*, and

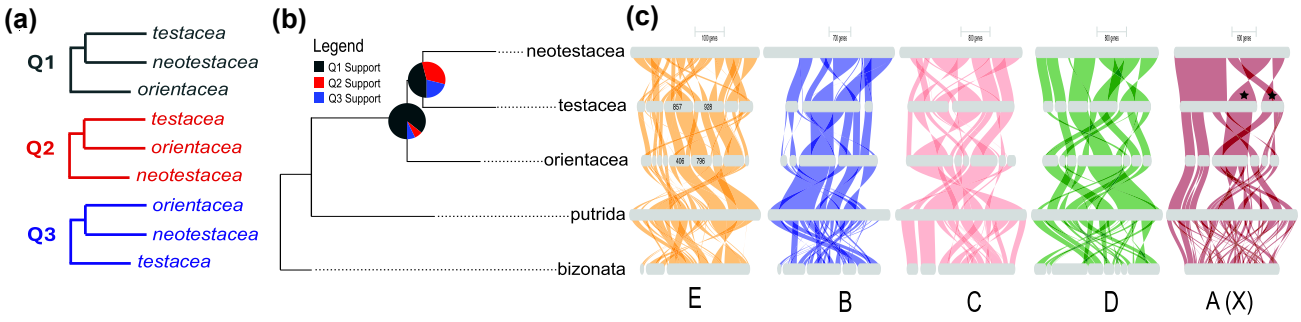

Fig. 1. Syntenic and phylogenetic relationships among *testacea* species group. a) Quartet trees corresponding to the three possible topologies for the three ingroup species. b) Phylogenetic tree representing species relationships to each other. The topology is from the coalescent analysis, but the relationships are the same in the maximum likelihood tree. Quartet support is from the ASTRAL analysis, where Q1 support represents the main topology, Q2 the first alternative, and Q3 the second alternative. c) GENESPACE riparian plots showing syntenic relationship within Muller elements. Colors on braids indicate each Muller element and connecting braids between each species represent homologous segments. Twisted braids suggest inversions between homologous regions. *Drosophila neotestacea* and *D. putrida* Muller elements are full-length scaffolds, whereas the other three assemblies are split up into contigs. These contigs may not be in the correct order, thus inflating the presence of structural variants. Stars denote X-chromosome inversions.

D. neotestacea. We use the closely related species *D. bizonata* as an outgroup (Scott Chialvo et al. 2019; Finet et al. 2021; Erlenbach et al. 2023). These five focal species are part of the *immigrans-tripunctata* radiation, and are common in temperate and boreal forests in the Northern Hemisphere (Grimaldi et al. 1992). *Drosophila putrida* occurs in eastern and central North America, *D. neotestacea* is found across northern North America, *D. testacea* occurs in Europe and continental Asia, and *D. orientacea* and *D. bizonata* are found in Japan (Kikkawa and Peng 1938; Grimaldi et al. 1992). Little is known about the geographic transition between *D. orientacea* and *D. testacea* in Asia and whether these species overlap in range.

All species in the *testacea* group consume rotting mushrooms as a primary food source. Adult flies are attracted to mushrooms, where they mate and females lay eggs. The larvae develop in the decaying mushroom, and then the emergent adults must find a new mushroom. These species are generalists on fleshy Basidiomycota mushrooms, including the toxic *Amanita* species. These species are not commensal with humans, and in the forests where they occur, they can be collected in very large numbers. Species in the *testacea* group have been studied for several evolutionarily and ecologically important traits. Some of these traits include cold tolerance (Kimura 2004), nematode parasite infection (Perlman et al. 2003), microbiome composition (Martinson et al. 2017; Bost et al. 2018), X-chromosome meiotic drivers (Dyer 2012; Keais et al. 2017; Pieper et al. 2018), mycotoxin tolerance (Jaenike et al. 1983; Kokate et al. 2022; Erlenbach et al. 2023), a Y-linked duplication on the dot chromosome in *D. neotestacea*, *D. orientacea*, and *D. testacea* (Dyer et al. 2011), and B chromosomes in *D. putrida* (Patterson and Stone 1952). The group also contains bacterial endosymbionts, including *Wolbachia* and *Spiroplasma* (Jaenike et al. 2010a), and *D. neotestacea* has been a model for studying how endosymbionts defend the host from other parasites (Jaenike et al. 2010b).

The *testacea* group is evolutionarily young, and both pre- and post-mating barriers contribute to reproductive isolation between these species (Grimaldi et al. 1992). In the laboratory, either reciprocal cross of *D. testacea* and *D. orientacea* produces fertile male and female offspring, although there is strong pre-mating isolation. When paired with any other species in the group, *D. neotestacea* and *D. putrida* either do not mate or do not produce viable offspring (Grimaldi et al. 1992). These patterns suggest *D. testacea* and *D. orientacea* are the most closely related of these species. This is inconsistent with previous phylogenetic findings, however, which place *D. testacea* and *D. neotestacea* as most closely related, with *D. orientacea* as the next most closely related and *D. putrida* as the outgroup to the others (as shown in Fig. 1b) (Erlenbach et al. 2023). This pattern occurs across the genome, including for a gene on the Y chromosome as well as for three loci on the mtDNA that often reflect recent introgression (Dyer et al. 2011, Scott Chialvo et al. 2019). Overall, genetic divergence between these species is low, and their effective population sizes may be high based on levels of segregating DNA polymorphism (Pieper et al. 2018), potentially complicating inferences of evolutionary history.

Here, we sequence and assemble five high-quality genomes, including Hi-C scaffolded hybrid assemblies for *D. putrida* and *D. neotestacea* and hybrid assemblies for *D. testacea*, *D. orientacea*, and *D. bizonata*. We then use these assemblies to investigate genome evolution in the *testacea* group. We ask how much discordance there is between the species tree and the patterns of reproductive isolation, and what is the nature of the discordance in terms of the distribution across the genome? And, do regions of

discordance show faster rates of adaptive evolution? To address these questions, first, we assess structural variation across the genome and then examine the small nonrecombining dot chromosome in further detail. Second, we evaluate phylogenetic discordance across the genome and identify regions of introgression. Third, we identify protein-coding genes under strong diversifying selection, and we associate this with introgression data. Lastly, we assemble the mitogenome from each species and assess mitochondrial species relationships. In summary, our data provide a first glimpse at the genetic characteristics of the *testacea* group and provide genome resources for future studies.

Results

We sequenced and assembled genomes for four *testacea* group species (*D. testacea*, *D. orientacea*, *D. neotestacea*, and *D. putrida*) and the outgroup *D. bizonata*. The total length of the genomes is between 152 and 168 Mb for the *testacea* group species, and *D. bizonata* is larger at 212 Mb (Table 1). Based on early cytology, *D. putrida* and *D. neotestacea* are expected to have four chromosomes, including two large metacentric autosomes, a telocentric X chromosome, and a dot chromosome (Patterson and Stone 1952). Our *D. neotestacea* assembly matches this karyotype, whereas our genome for *D. putrida* instead has the Muller elements as separate scaffolds broken at the repetitive centromeric regions. The de novo genomes had N50s of 8.3 Mb for *D. testacea*, 6.6 Mb for *D. orientacea*, and 7.4 Mb for *D. bizonata* (Table 1). Similarly, our auN statistics indicate comparable contiguity for the assemblies (Table 1). As a proxy for annotation quality, we assessed BUSCO scores, where all five genomes had high scores (Table 1), ranging from 95% to 99% with low duplication percentages, and our annotations confirm we captured a majority of gene content, with similarly high BUSCO scores (Table 1). Our assemblies show between 12,129 and 13,193 genes with support from RNA annotation, which is comparable to the ~14,000 genes in *D. melanogaster* (Kaufman 2017).

Between genomes, there is variation in size and repeat content (Table 1). *Drosophila bizonata* has the largest genome (212 Mb) and the greatest percentage of repeats (28.12%), whereas *D. putrida* has the smallest genome (152 Mb) and the lowest percentage of repeats (15.76%). Interspersed repeats, DNA elements, and LINEs have variation corresponding with the total repeat content for each species, with *D. bizonata* having the highest percentage, *D. putrida* the lowest, and *D. orientacea* and *D. testacea* sharing similar amounts (Fig. S1). There is the greatest variation comparing *D. bizonata* to the rest of the group, but this is not surprising given its phylogenetic distance from the other species.

Table 1 Genome assembly statistics, including Benchmark Universal Single-Copy Ortholog (BUSCO) statistic, the total length in bases, number of contigs and scaffolds, N50, L50, and auN for assembly completeness, the percentage of bases masked in RepeatMasker, number of annotated genes that had support in BRAKER, and GC content for the nuclear genome and the mitochondrial genome.

Species	Total length (bases)	Number of contigs	Number of scaffolds	L50	N50 (Mb)	auN	Genome BUSCO score (%) [S; D; F; M]	Annotation BUSCO score (%) [S; D; F; M]	Percentage of repeats	Number of annotated genes	Nuclear GC content (%)	Mitochondrial GC content (%)
D. neotestacea	165,053,033	198	4	2	50.094	56,518,804.20	98.8 [98.1; 0.6; 0.7; 0.6]	96.8 [84.4; 12.4; 1.2; 2.0]	23.06	13,230	37.14	22.18
D. testacea	158,082,396	785	N/A	6	8.325	9,364,287.80	98.8 [97.9; 0.9; 0.5; 0.7]	96.3 [94.4; 1.9; 1.4; 2.3]	19.76	13,026	37.02	22.03
D. orientacea	168,126,735	689	N/A	8	6.636	6,938,249.50	98.9 [98.1; 0.8; 0.3; 0.8]	97.0 [95.7; 1.3; 0.9; 2.1]	19.4	12,139	37.96	23.16
D. putrida	152,076,637	113	6	3	29.442	28,094,135.40	95.6 [94.9; 0.7; 2.4; 2.0]	88.4 [77.8; 10.6; 6.8; 4.8]	15.76	13,306	37.76	25.02
D. bizonata	212,017,377	1,150	N/A	6	7.43	13,158,919.10	98.9 [97.5; 1.4; 0.5; 0.6]	96.3 [94.2; 2.1; 1.0; 2.7]	28.12	13,193	37.23	22.61

All assemblies are derived from females except *D. testacea* which was derived from males. S, Single-Copy; D, Duplicated; F, Fragmented; and M, Missing.

Synteny Analysis

As structural rearrangements can be associated with adaptation, reproductive isolation, and other factors facilitating divergence between species (Zhang et al. 2014; Fuller et al. 2018), we assessed synteny between Muller elements in the *testacea* group. Our *D. neotestacea* and *D. putrida* genomes were scaffolded using Hi-C to aid in this analysis, and we identified contigs in the other species belonging to each Muller element. As expected, Muller elements are largely conserved (Figs. 1a and 3; Fig. S2) (Clark et al. 2007). In *D. neotestacea*, elements E and B form one metacentric autosome, and elements C and D form the other large metacentric autosome. Muller element E is not completely syntenic between *D. bizonata* to the rest of the *testacea* group, as part of the distal end of Contig 1062 shares synteny with regions on Mullers B and E (Fig. S2). We hypothesize that either large translocations have shifted gene content from one Muller to another between *D. bizonata* and the *testacea* group, or *D. bizonata* has a different karyotype.

Within Muller elements, rearrangements and structural differences reduce collinearity between the species. Without having fully scaffolded Muller elements for *D. bizonata*, *D. testacea*, and *D. orientacea*, it becomes difficult to elucidate specifics of autosomal differences; therefore, we broadly assessed structural variation between *D. putrida* and *D. neotestacea*, for which we have fully scaffolded Muller elements. Our results show differences in the minimum number of rearrangements and inversions across Muller elements (Fig. S3). Muller B has at least two inversions and three translocations, whereas Muller E, which is associated with Muller B through a metacentric fusion in *D. neotestacea*, has at least ten inversions (Fig. S3). Muller C has at least four inversions, including one that is nested, and Muller D has at least three inversions, including one that is nested. The dot is completely collinear between the two species (Fig. S3). The X chromosome (Muller A) has at least seven inversions between *D. putrida* and *D. neotestacea* (Fig. S3). We obtained a single contig for the *D. bizonata* X chromosome, and there are many small rearrangements and inversions that differentiate *D. bizonata*'s X chromosome from the other two species (Fig. S3), making it nearly impossible to determine the number of structural variants.

If we expand this structural variant analysis to include our less contiguous assemblies, we identify large blocks of collinearity on most Muller elements. Based on visual inspection, we note that Muller E appears to contain more rearrangements than the other elements (Fig. 1c). In particular, between *D. neotestacea* and either *D. testacea* or *D. orientacea* there are many small rearrangements and inversions, while *D. testacea* and *D. orientacea* contigs share larger blocks of collinearity. For example, Contig 857 and 928 in *D. testacea* share large blocks of synteny with contigs in *D. orientacea*, whereas these contigs are found in many small pieces across most Muller E in *D. neotestacea* (Fig. 1c). We verified several putative inversions using the raw reads. Scaffolding the *D. testacea* and *D. orientacea* genomes is necessary for a more quantitative analysis of chromosome synteny.

Characterization of the Dot Chromosome

The dot chromosome, or Muller element F, in *Drosophila* is distinct from the others with its larger proportion of repetitive sequences, lack of recombination, and extremely long introns, despite it typically being a much smaller chromosome compared to the other Muller elements (Riddle and Elgin 2006). In addition, the dot in the *testacea* group has experienced duplications of Y-linked genes, including the male fertility factor *kl-5* in all species but *D. putrida* (Dyer et al. 2011). Most of the dot is completely syntenic between species (Fig. 2a); however, the *D. neotestacea* and *D. testacea* dots are twice as large as the *D. bizonata* and *D. putrida* dots (4 Mb vs. 2 Mb; Fig. 2a and Fig. S4), and *D. neotestacea* contains a non-syntenic region at the distal end that is 249 kb.

The genes in this non-syntenic region in *D. neotestacea* contain the *kl-5* gene sequence, along with two genes homologous to dynein chains and a gene with piggy-Bac TE-like homology. *Kl-5* also encodes a dynein heavy chain, but these other dynein chain sequences are ~0.8 and ~1.2 kb in length, excluding introns, whereas *kl*-5 is 8.3 kb, and they do not appear to be homologous to *kl-5*. Next, we attempted to identify *kl-5* in the *D. orientacea* and *D. testacea* genome annotations, but we only found short sequences aligning to a small region on the gene. Therefore, we examined homology per exon, of which *kl-5* has 20 in *D. neotestacea*. Three annotated genes had homology to the *kl-5* exons in *D. orientacea* [g92 (contig 4; exons 1 to 9), g4852 (contig 364; exons 10 to 13), g11796 (contig 734; exons 14 to 20)], and *D. testacea* was separated into four annotated genes [g10323 (contig 759; exon 1), g9494 (contig 591; exons 3 to 9), g13107 (contig 951; exon 14), g13111 (contig 951; exons 15 to 20)]; exons 2 and 10 to 13 in *D. testacea* were not found.

The high proportion of repetitive content on the dot could impact the *kl-5* gene annotation. Thus, we assessed if contigs in *D. testacea* and *D. orientacea* contained any homology to *kl-5*, using the full *D. neotestacea kl-5* gene sequence and its exons. These results identified three contigs in *D. testacea:* contig 754 (24,458 bp) was homologous to exon 2, contig 585 (40,441 bp) was homologous to exons 3 to 10, and these exons were in order along the contig, and contig 937 (142,594 bp) was homologous to exons 13 to 19 and in order. In *D. orientacea*, two contigs had

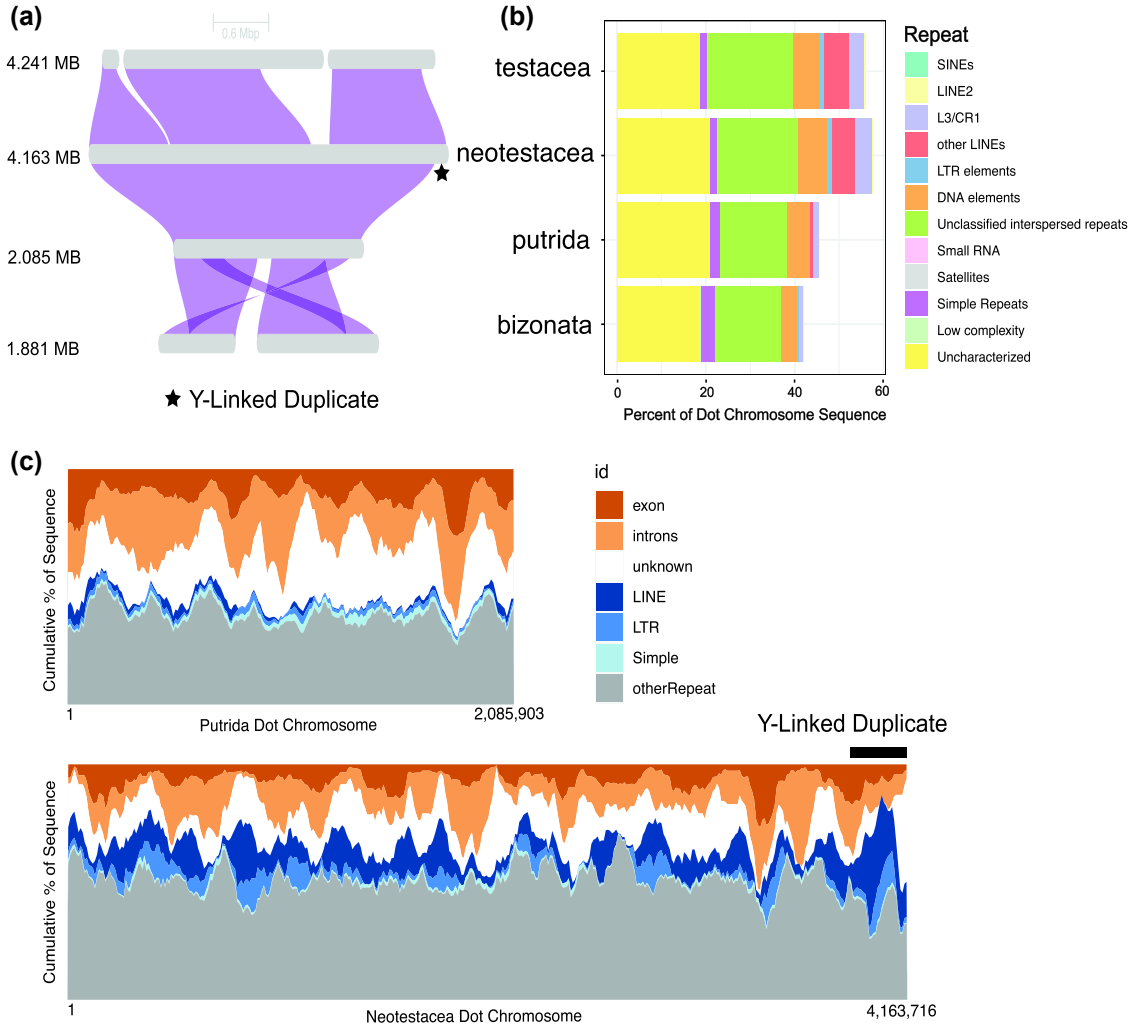

Fig. 2. Synteny and repetitive content characterization for the dot chromosome. a) GENESPACE riparian plot showing the syntenic relationship between each species for the dot chromosome. Each row represents either the scaffolds (*D. neotestacea* and *D. putrida*) or contigs (*D. bizonata* and *D. testacea*) for the dot. Connecting braids between rows indicate syntenic regions, and twisted braids indicate inverted regions. The size of the dot chromosome in megabases is on the left of the plot. b) Stacked barplot of repetitive content on the dot chromosome for each species. Colors represent specific repetitive element classifications based on RepeatMasker output, and the percent is the total percent of the dot chromosome sequence the repeat represents. c) Comparison of exon and intron size to repetitive content along the dot chromosome for *D. putrida* (top) and *D. neotestacea* (bottom). The legend depicts colors denoting sequence type, including the most abundant repetitive elements found on the dot. The X-axis represents the length of the dot chromosome, with values calculated in 1 Mb windows with 100 kb step sizes, and the Y-axis represents the percentage of each sequence type.

homology to the *kl-5* exons: contig 0 (938,640 bp) was homologous to exons 2 to 12, whereas contig 732 (74,276 bp) was homologous to exons 13 to 19; these exons were in sequential order in both contigs. These contigs only contained exons from *kl-5*, consistent with the long introns characteristic of this gene. Because these results indicated different contigs containing homology to *kl-5*, we aligned our male RNA sequencing data for these species to the *D. neotestacea* dot to determine if reads spanned this region, which could impact annotation. *Drosophila orientacea* and *D. testacea* have reads that align to *kl-5*, though the reads do not provide consistent coverage across the gene. Alternatively, in *D. neotestacea*, these reads consistently map across *kl-5* exons. *Drosophila putrida* has a few sporadic reads aligning to this region, but because we aligned the male RNA, these are most likely reads for the Y-chromosome *kl-5* gene sequence.

These results make characterizing the Y-linked duplicate region in *D. testacea* and *D. orientacea* unclear, and would require a deeper analysis into this region, including hand annotation and piecing together the contigs with homology, though the large introns and repetitive content of the dot make this difficult. Furthermore, the genomic sequencing for *D. testacea* is from males, and mis-assembly

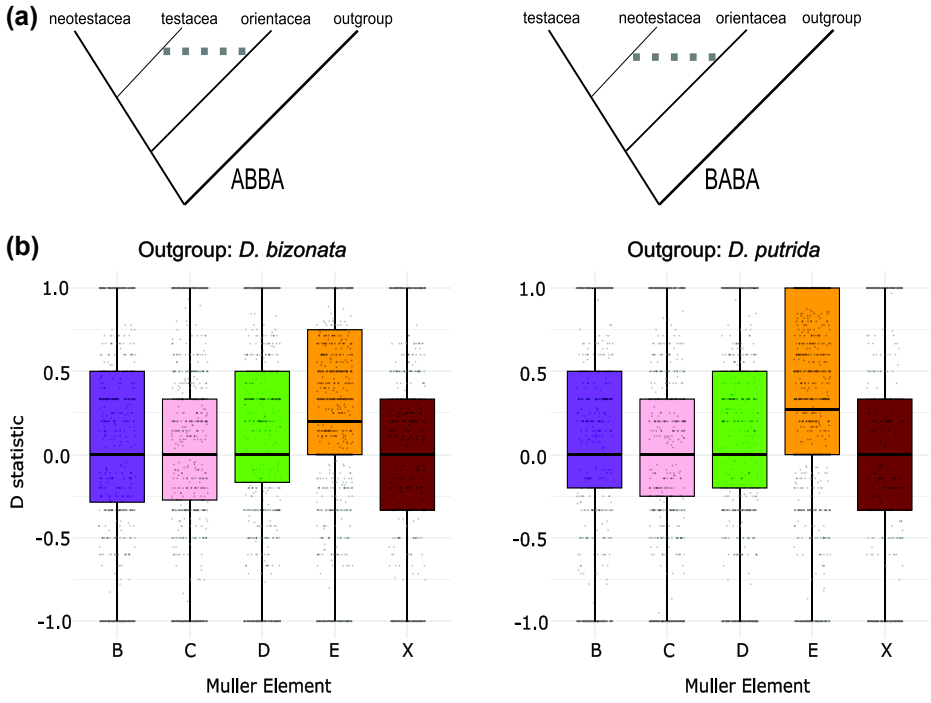

Fig. 3. Introgression within the *testacea* group. a) Topologies representing ABBA (left) and BABA (right), and the corresponding species in which introgression occurs, are indicated by the dotted lines. The outgroup for each test is indicated on the tree, and the test was run with either *D. bizonata* or *D. putrida* as the outgroup species. b) Boxplot of D statistics (Y-axis) calculated per SCO split up by Muller element, with either *D. bizonata* (left) or *D. putrida* (right) as the outgroup for the analysis. Points represent individual D statistics for each SCO.

of Y-chromosome repetitive content with the Y-linked duplicate could have occurred, thus making characterization of this region in *D. testacea* challenging without having a Y-chromosome assembly. Currently, we do not have confidence in identifying the Y-linked duplication, nor the male fertility factor *kl-5*, in *D. testacea* and *D. orientacea*. However, we can show that *D. neotestacea* has a 249 kb region at the distal end of its dot chromosome containing *kl-5*, and this region is not present in *D. putrida*. Between *D. neotestacea* and *D. putrida*, excluding this region, there is gene-by-gene homology and conservation in gene order along the entire rest of the chromosome; however, because it is only 249 kb, it is not the only contributing factor to the 2 Mb size difference.

We assessed repeat content and intron size as factors that could impact the size of the dot chromosome (Muller F). The dot in *D. neotestacea* and *D. testacea* has more repeats than *D. bizonata* and *D. putrida* (Fig. 2b), with L3/CR1, LINEs, LTR elements, and interspersed repeats contributing the most to this difference. When we compare this chromosome-wide between *D. neotestacea* and *D. putrida*, the proportion of LINEs and LTRs is higher across the entire dot (Fig. 2c). Furthermore, this increase in repetitive content in *D. neotestacea* reduces the proportion of "Unknown" content, which is not genic (exons or introns) or repetitive and could be noncoding regions and regulatory elements we do not have annotations for in our assemblies. We also assessed intron size across BUSCOs and SCOs and found that intron size is larger on the dot for all species (Fig. S5). The autosomal intron size in *D. bizonata* is slightly larger than that of other species, but the higher proportion of repetitive content in this genome (Table 1) is most likely impacting its intron size. Taken together, these results indicate that repetitive content, not intron size, is largely contributing to the difference in dot chromosome size between *D. testacea* and *D. neotestacea* and *D. putrida* and *D. bizonata*.

Codon usage can be an indicator of the strength of selection, and the dot chromosome (Muller F) in *Drosophila* often exhibits different patterns in codon usage from the rest of the genome (Vicario et al. 2007; Powell et al. 2011). Thus, we assessed codon usage among Muller elements. When we combine species, GC content and the effective number of codons (Enc) differ across Muller elements (Fig. S6; Table S5), and GC content is higher in coding regions than it is across the entire genome (Table 1). Specifically, Enc is significantly different between the dot and all Muller elements except Muller D, and GC content is significantly different between the dot and the rest of the genome (Table S6). Enc and GC content are also different between the autosomes and the X (Table S6), similar to other *Drosophila* species where codon usage bias has been found to be stronger on the X than the

large autosomes (Vicario et al. 2007). If we separate by species, Enc values do not differ between the autosomes and the dot and are similar to other *Drosophila* species (Kokate et al. 2021). We also find that the Relative Synonymous Codon Usage (RSCU), or the proportion of synonymous codon usage for each amino acid, varies between the dot and the rest of the genome (Table S7), with a higher RSCU for codons ending in A/T, especially for 2-fold degenerate amino acids, though some 3- and 4-fold codons have a higher RSCU for A/T as well (Val and Ile; Table S7). Alternatively, codons ending in G/C tend to have a higher RSCU for the X and autosomes. These results support the lower GC content we find on the dot for all species (Tables S5 and S6).

Comparative Phylogenetics

We identified 4,868 single-copy orthologs (SCOs) that we both filtered for alignment quality and were present in all species, and we built a species tree with these to assess the phylogenetic relationships of the *testacea* group. We used maximum likelihood and coalescent approaches to build trees, and the topology of these two species trees is the same (Fig. 1 and Fig. S6b). The tree puts *D. putrida* as sharing the most distant common ancestor to the clade for *D. neotestacea*, *D. testacea*, and *D. orientacea*, with *D. neotestacea* and *D. testacea* as sisters to each other.

Our species tree indicates that *D. testacea* is more closely related to *D. neotestacea* than *D. orientacea* (Fig. 1b and Fig. S6), but factors such as hybridization, incomplete lineage sorting (ILS), and introgression can impact the topology (Nichols 2001; Pollard et al. 2006). Therefore, we assessed gene tree discordance by analyzing quartet scores in ASTRAL-III (Zhang et al. 2018) and concordance factors in IQ-TREE2 (Minh et al. 2020). The normalized quartet score is 0.702, indicating that nearly one-third of quartet trees are not found in the species tree (Table S4). The strongest amount of discordance is found along the branch leading to *D. neotestacea* and *D. testacea*, with only 46.3% of quartets matching the species tree (Fig. 1b; Q1 Support). Alternative quartet topologies show 33% of quartets supporting the first alternative (Q2 Support), placing *D. orientacea* and *D. testacea* as sisters to each other, and 21% of quartets supporting the second alternative (Q3 Support), putting *D. orientacea* and *D. neotestacea* as sister species (Fig. 1b; Table S3). Our gene (gCF) and site concordance factors (sCF) find similar patterns (Table S3), with the branch to *D. testacea* and *D. neotestacea* having a gCF of 43.71% and sCF of 41%. Conversely, the branch to all three *testacea* species has a gCF of 82.95% and sCF of 74.61%. Despite having high bootstrap support for these branches (100; Fig. S7), these discordance values indicate there are many conflicting signals among the gene trees, potentially due to ILS and introgression. The high bootstrap support despite these factors can be due to low sampling variance resulting from the large number of genes and sites in our analysis, especially with a short branch length (0.2169) leading to *D. testacea* and *D. neotestacea*, which tend to be harder to resolve.

To investigate what could be contributing to this high phylogenetic discordance, we performed a four-taxon ABBA-BABA test on the SCOs to identify putative introgression events. With a nonsymmetric phylogeny, a five-taxon introgression test could not be performed. Our tree has ((*D. neotestacea*, *D. testacea)*, *D. orientacea*) as the ingroup species, with either *D. bizonata* or *D. putrida* as the outgroup (Fig. 3a). Our results indicate stronger shared ancestry between *D. testacea* and *D. orientacea,* which could be the result of introgression. With *D. putrida* as the outgroup, 10 genes are significant for the BABA topology (sharing between *D. neotestacea* and *D. orientacea*) and 52 for the ABBA topology (sharing between *D. testacea* and *D. orientacea*) (Fig. 3a; Table S4). Similarly, with *D. bizonata* as our outgroup, 10 BABA (sharing between *D. neotestacea* and *D. orientacea*) and 66 ABBA topologies (sharing between *D. testacea* and *D. orientacea*) are significant (Fig. 3a; Table S4). There were 31 orthologs shared between significant ABBA genes.

To determine which genes show putative introgression between *D. orientacea* and *D. testacea,* we identified homology to *D. melanogaster.* Some orthologs (OG0011268; OG0007049; OG0006556) included homologs to genes predicted to have zinc ion binding activity. In *Drosophila* oocytes, zinc is required for female fertility and is also the most abundant transition metal in oocytes (Hu et al. 2020). We find homology to *Mettl3* (OG0006359), which aids in female germ-line cyst encapsulation, and *cadherin-99C* (OG0006670), a gene essential for female fertility. Furthermore, there is homology to chitin-binding domains (OG0008100; OG0008099) or regulation of chitin-containing cuticle (OG0006548). Genes with involvement in embryonic development were also found, including *mira* (OG0006554), *sno* (OG0007424), and *side* (OG0007693). Though more work is needed to uncover the role of these genes in regard to fertility between species, homology to female fertility and development could indicate a role in hybridization between *D. orientacea* and *D. testacea*.

An excess of loci significant for the ABBA topology are on Muller E (Fig. 3b; 57.7%, $\chi^2 = 89$, df = 4 with *D. bizonata,* P-value < 0.0001; 63.6%, $\chi^2 = 107$, df = 4 with *D. putrida*, P-value < 0.0001; all pairwise tests with element E were highly significant with a P-value < 0.0001, while no other pairs had $P < 0.01$). However, the significant loci are not clustered on the chromosome regardless of the outgroup (Fig. S8). Thus, we sought to determine if Muller elements have different species tree topologies. The trees for Muller B, C, D, and A (X) share the same topology as

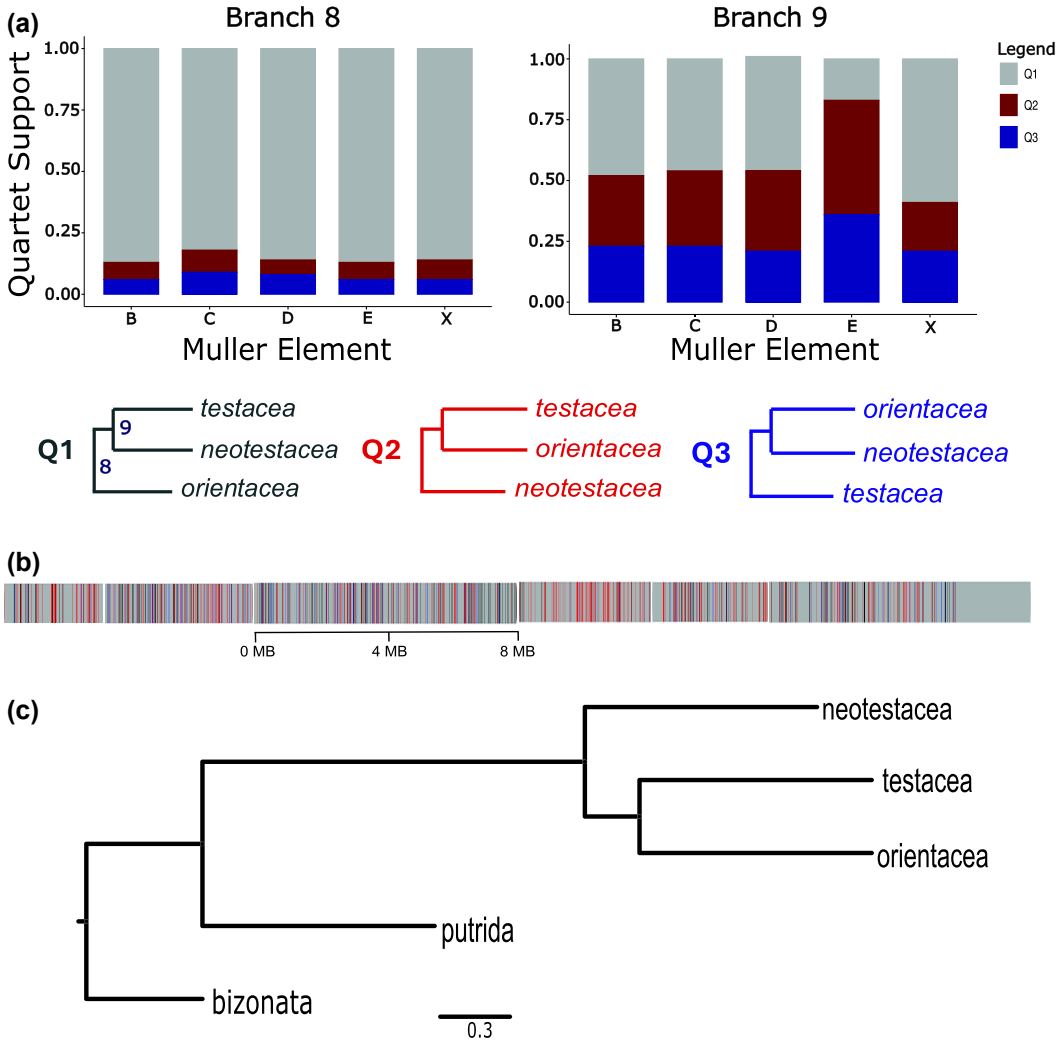

Fig. 4. In contrast to the rest of the genome, in which *D. neotestacea* and *D. testacea* are sister species, the tree for Muller element E places *D. orientacea* and *D. testacea* as sisters to each other. a) Quartet support (Y-axis) across each Muller element (X-axis) for Branch 8 (left) and Branch 9 (right) of the *testacea* group species tree. Quartet trees are shown below the bar plots, and branches 8 and 9 are indicated in the Q1 tree. b) Location of individual genes on Muller element E contigs *in D. testacea*. Genes are colored based on their gene tree topology, with genes supporting the Muller E topology (Q2) in red, genes supporting Q1 in black, and genes supporting Q3 in blue. Contigs are to scale and ordered as in Fig. 1. The distal region without any genes highlighted is most likely the centromere based on synteny to other Muller elements. c) Coalescent phylogeny of Muller element E.

our species tree, with *D. neotestacea* and *D. testacea* as sisters to each other. These Muller elements also have similar patterns of gene tree discordance among the two branches to the three ingroup species (Fig. 4a). However, Muller E places *D. testacea* and *D. orientacea* as sisters to each other (Fig. 4c) and has the most gene tree support for the topology placing these two species as the most closely related (Fig. 4a), which is expected if this element has been introgressed in their history. We still find discordance for this topology and the other Muller elements; however, the majority of gene trees place *D. testacea* and *D. orientacea* together (Table S3). The genes placing *D. testacea* and *D. orientacea* as sisters to each other are also distributed along the contigs for Muller E (Fig. 4b). These results suggest Muller E is important for understanding the history of introgression and hybridization between these two species.

Molecular Evolution

To identify if certain genes or gene groups may be under selection in the *testacea* group, we inferred the rate of evolution (dN/dS) for null and branch models on the single copy orthologs (SCOs). We assessed the best-fitting model for each gene, comparing if a branch model where the *testacea* species clade (*D. neotestacea*, *D. orientacea*, and

D. testacea) can evolve differently from the outgroups (*D. putrida* and *D. bizonata*) has a higher likelihood than the null where, the gene evolves at the same rate across the tree. We found 683 orthologs (14%) evolving at a different rate in the *D. neotestacea*, *D. orientacea*, and *D. testacea* clade compared to *D. putrida* and *D. bizonata* ($P < 0.05$; Table S4). Of these, 381 orthologs (56%) had a higher dN/dS in the testacea clade than the outgroup species. We note that a higher dN/dS does not indicate whether this is due to a change in positive or relaxed selection between the groups. In contrast to our ABBA-BABA results, there were more genes on Mullers C and D significant for the branch model than expected based on size ($\chi^2 = 13.702$, $P = 0.0083$, df = 4), with no significant difference among chromosomes in the proportion of loci with higher dN/dS in the testacea clade versus the outgroup species ($\chi^2 = 5.321$, $P = 0.26$, df = 4). We performed a GO analysis to determine if biological processes are enriched in these orthologs, and we found two terms are enriched in the branch model, including G-protein coupled photoreceptor activity (10.61, $P = 0.015$) and cell migration (3.09, $P = 0.025$).

We then compared the genes evolving at a different rate between the testacea clade and the outgroups to those showing evidence of putative introgression between *D. testacea* and *D. orientacea*. Of the genes with evidence for introgression, there are 13 orthologs with a significantly different dN/dS among groups with the branch model using *D. bizonata* as the outgroup, of which 10 (77%) have a higher dN/dS in the testacea clade than the outgroup. Likewise, of the introgressed genes, there are 17 orthologs with a significantly different dN/dS among groups with the branch model using *D. putrida* as the outgroup, of which 12 (71%) have a higher dN/dS in the testacea clade than the outgroups. There are eight genes with a significantly different dN/dS among groups with the branch model that are shared between both ABBA-BABA topologies, and 6 of these orthologs (75%) have a higher dN/dS in the testacea clade than the outgroup. Included in these six genes is *Cadherin-99C* (OG0006670), a female fertility gene that contributes to eggshell formation and is expressed in follicle cells during oogenesis (D'Alterio et al. 2005). This suggests female fertility might act to facilitate hybridization between *D. orientacea* and *D. testacea*, but much more work is required to confirm this.

Mitogenome Assembly and Phylogenetics

Mitochondrial genomes can vary in their evolutionary patterns compared to nuclear DNA due to maternal inheritance (Moritz et al. 1987; Meiklejohn et al. 2007; Dowling and Wolff 2023); therefore, we assembled the mitogenome for each species and assessed their phylogenetic relationship. The mitogenome assemblies ranged from 14 kb to 15 kb in length with GC content ranging between 22.03% and 25.02% (Table 1). In comparison, the *D. innubila* mitogenome is 16 kb, with a GC content of 22.66% (CM015047.1). We annotated the 13 mitochondrial genes found in *D. melanogaster* (YP_009047266.1 to YP_009047278.1) for each species. Visualization of the annotations identified that the order and orientation of these genes are consistent across all five species and in the same order as *D. innubila* (Fig. S9a). The species tree for mitochondrial DNA places *D. testacea* and *D. orientacea* as sisters to each other (Fig. S9b), similar to our Muller E species tree (Fig. 4), further supporting the entangled history of introgression between these species.

Discussion

Species in the *testacea* group of *Drosophila* have been studied for their evolutionary and ecological characteristics, and here, we assemble genomes for these species and begin to probe their genomic differences. We find general conservation of Muller elements, with many rearrangements within all elements (Fig. 1a and Fig. S2). These results are not surprising, as across *Drosophila* species, Muller elements generally share gene content, but the gene order is usually scrambled due to rearrangements (Bhutkar et al. 2008; Chakraborty et al. 2021). Due to our less contiguous assemblies, it is difficult to completely characterize synteny, but we do find many putative rearrangements that differ among the *testacea* group species, especially on Muller E (Fig. 1; Figs. S2 and S3). Whole-chromosome scaffolding for all species would provide greater resolution on the number of rearrangements differing between species, but our results show that *D. bizonata* is structurally very different from the rest of the *testacea* group, and that, despite being very closely related, many rearrangements differ between *D. neotestacea*, *D. testacea*, and *D. orientacea*. Our study complements other recent analyses of structural evolution in closely related *Drosophila* species that are enabled by high-quality genomes (Reis et al. 2018; Capinteyro-Ponce and Macahdo 2024; Poikela et al. 2024).

Across a variety of taxa, inversions contain genes involved in reproductive isolation (Jones et al. 2012; Fishman et al. 2013; Lohse et al. 2015) and local adaptation (Connallon et al. 2018; Sinclair-Waters et al. 2018; Coughlan and Willis 2019) between closely related species. We find an excess of rearrangements on Muller E between *D. neotestacea* and *D. putrida* (Fig. S3). Further, if we compare the three ingroup *testacea* species, there are more structural rearrangements on Muller E in *D. neotestacea* compared to *D. testacea* and *D. orientacea*, which are mostly collinear (Fig. 1a and Fig. S2). This is consistent with recent or ongoing introgression between *D. testacea* and *D. orientacea*. The X chromosome is often the site of incompatibilities between species (Presgraves 2008;

Qvarnström and Bailey 2009; Meisel and Connallon 2013; Charlesworth et al. 2018), and we find two inversions that differentiate *D. testacea* and *D. neotestacea* on the X chromosome that are collinear between *D. testacea* and *D. orientacea* (Fig. 1). The presence of multiple inversions may prevent gene flow from occurring between *D. neotestacea* and the rest of the *testacea* group, ultimately increasing its divergence from them. Because *D. testacea* and *D. orientacea* do not have these structural rearrangements and contain more syntenic regions, this may allow hybridization, and thus gene flow and introgression, to occur between them.

We uncovered a large size difference for the dot chromosome (Muller F), where the dot is twice as large in *D. neotestacea* and *D. testacea* as in *D. putrida* and *D. bizonata* (Fig. 2a and Fig. S4). A major contributor to this size difference is repetitive elements (Fig. 2b and c). In *D. ananassae*, whose dot chromosome is 18.7 Mb, the largest contributors to its increased size are transposons, specifically LTR and LINE retrotransposons (Leung et al. 2017). Similarly, our results show an increase in LINEs and LTR elements for *D. neotestacea* and *D. testacea*, as well as L3/CR1 and interspersed repeats (Fig. 2b). Similar methods of expansion could be occurring in these species as they do in *D. ananassae*.

The dot chromosome (Muller F) also contains a duplicated Y-linked region in the three ingroup *testacea* species, but not *D. putrida* (Dyer et al. 2011). Using homology to *kl-5,* the translocated Y-chromosome gene, we identified the 249 kb duplicated region in *D. neotestacea* at the distal end of the dot. This region, despite being present in the species found to have a larger dot, does not seem to be a major contributor to the size difference (Fig. 2a). Consistent with previous work (Dyer et al. 2011), it also appears to be syntenic between *D. testacea* and *D. neotestacea* (Fig. 2a), suggesting it duplicated once in the common ancestor of these species. However, we could not confidently tease apart the rest of this duplicated region in *D. testacea*, which is a consequence of having male genomic data, where repetitive content between the Y chromosome and dot in this region could lead to misassembly. *Drosophila orientacea* has copies of *kl-5* on the dot and the Y chromosome (Dyer et al. 2011), and a more contiguous genome assembly for *D. orientacea* will be necessary to assemble this genomic region. In addition, a scaffolded dot chromosome in *D. testacea* would confirm if this region is ancestral and provide a better understanding of the evolutionary history of *kl-5*. Additionally, we find a gene with homology to *kl-5* in *D. bizonata*, though it is on a contig homologous to Muller C (Contig 598), highlighting the recurrent movement of genes off the Y chromosome.

The Y chromosome in *Drosophila* can contribute to incompatibilities between species due to the presence of essential male fertility factors (Vigneault and Zouros 1986; Lamnissou et al. 1996) and species-specific repetitive content (Heikkinen and Lumme 1998). *Drosophila testacea* and *D. orientacea* are the only species we studied here that produce fertile offspring when crossed, and both male and female F1 hybrids between them are fertile (Grimaldi et al. 1992; Dyer et al. 2011). However, the gene tree of *kl-3*, a gene on the Y chromosome of all these species (Dyer et al. 2011), as well as genes on the dot chromosome, have gene trees with *D. neotestacea* and *D. testacea* most closely related to each other, consistent with the species tree (Fig. 1b), suggesting hybrid incompatibilities may not reside in these regions.

Our species tree contained phylogenetic discordance in the clade leading to the three ingroup *testacea* species (Fig. 1b), with many trees supporting a topology that placed *D. orientacea* and *D. testacea* as sisters to each other. We tested for introgression to determine what could be contributing to this discordance and identified strong shared ancestry between *D. testacea* and *D. orientacea* on Muller E (Fig. 3b). This was further supported by species trees generated for each Muller element, where the topology for Muller E places *D. orientacea* and *D. testacea* as sisters to each other (Fig. 4a). The mitochondrial DNA tree also shares this topology (Fig. S9b) and has shown evidence of introgression between other species (Llopart et al. 2014; Mastrantonio et al. 2016; Makhov et al. 2021). Homologs to fertility and development genes are introgressed between *D. testacea* and *D. orientacea* and have a higher dN/dS within the species group (Table S4). A consequence of hybridization between species is the reduction in fertility and viability of hybrid offspring, leading to selection on isolating barriers that reduce the risk of hybridization (Coyne and Orr 1998). Hybridization can occur between *D. testacea* and *D. orientacea* in lab crosses, but not to *D. neotestacea* (Grimaldi et al. 1992). With more rearrangements on Muller E increasing divergence between *D. neotestacea* to *D. orientacea* and *D. testacea*, it is not unsurprising to find that the majority of introgressed genes are on this Muller element and between *D. testacea* and *D. orientacea*. Ultimately, it is unclear how rearrangements and introgressed genes on Muller E could be involved, but they suggest a complex history of hybridization and gene flow among these three species. We suggest that incomplete lineage sorting of ancestral polymorphism is less likely, as it would require the entire Muller E to segregate as a different inversion karyotype in the common ancestor and then fix in a different pattern than the rest of the genome. Given this study is limited in that we only used one strain per species, further studies involving the resequencing of multiple strains from each species will start to disentangle this.

Similar patterns of introgression and hybridization have been identified in the melanogaster species complex,

Table 2 Pairwise divergence among species in the *testacea* species group

Species	*D. bizonata*	*D. putrida*	*D. orientacea*	*D. testacea*	*D. neotestacea*
D. bizonata	…	0.165	0.153	0.152	0.153
D. putrida	0.172	…	0.0687	0.0675	0.0681
D. orientacea	0.160	0.0717	…	0.0277	0.0302
D. testacea	0.158	0.0698	0.0297	…	0.0248
D. neotestacea	0.160	0.0708	0.0321	0.0269	…

K_s values are calculated across single-copy orthologs (SCOs) in PAML and are used as a proxy for genetic distance between species. The lower quadrant of values denotes the average K_s across SCOs, whereas the upper quadrant is the median.

where considerable gene flow despite geographic isolation has occurred between the mainland species *D. simulans* and the island endemic species, *D. mauritiana* and *D. sechellia* (Garrigan et al. 2012; Presgraves and Meiklejohn 2021). The *simulans* clade is a recent polytomy largely due to rapid speciation that occurred ~250,000 years ago (Garrigan et al. 2012; Pease and Hahn 2013), and these three ingroup species experience strong reproductive isolation (Barbash 2010). The genomes of these three species are mostly collinear, with few rearrangements differentiating them, and mostly syntenic to the outgroup *D. melanogaster*, with the majority of rearrangements on the X chromosome (Chakraborty et al. 2021). This contrasts with the larger number of rearrangements identified among the *testacea* group species, though we also find many rearrangements on the X chromosome (Fig. 1c) and note that our assemblies are less contiguous than those used to assess structural variation in the melanogaster complex. Divergence between *D. orientacea* and *D. testacea* has been estimated to be 0.17 MYA (95% CI: 0.35 to 0.06) and the split of the three ingroup *testacea* species to *D. putrida* is 8.02 MYA (95% CI: 11.41 to 5.17) and to *D. bizonata* is 12.02 MYA (95% CI: 16.26 to 8.40) (Izumitani et al. 2016). As a proxy for divergence among these species, we evaluate *Ks* calculated for each pairwise species comparison on the filtered set of SCOs (Table 2). *Ks* between *D. putrida* and the three ingroup species is ~0.07, and among the three ingroup species, *Ks* is between 0.02 and 0.03 (Table 2). These estimates are lower than what is found between closely related species in the *melanogaster* group (0.04 to 0.1)(Turissini et al. 2018) and between *D. recens* and *D. subquinaria* (0.08)(Ginsberg et al. 2019), two hybridizing sister species in the larger adaptive radiation that the *testacea* group is a part of (Erlenbach et al. 2023). These results suggest that reproductive isolation has evolved faster between *D. neotestacea* and either *D. orientacea* or *D. testacea* than in other *Drosophila* species.

Complicating the evolutionary history of the *testacea* group is selfish elements. *Drosophila neotestacea* (James and Jaenike 1990), *D. testacea* (Keais et al. 2017), and *D. orientacea* (K. Dyer, unpublished) all harbor sex-ratio meiotic drivers that distort offspring sex ratios such that males harboring the driving X chromosome only produce female offspring (Jaenike 2008; Helleu et al. 2015). The sex-ratio haplotype of *D. testacea* is thought to predate the split of the three species, which can complicate species relationships if it recombines with the non-driving X chromosome. In addition, all three species harbor *Wolbachia,* a maternally inherited endosymbiont (Jaenike et al. 2010a). The *Wolbachia* infections in these species are closely related, and are more similar between *D. orientacea* and *D. testacea* than with *D. neotestacea* (Jaenike et al. 2010a). *Wolbachia* is maternally inherited along with the mitochondria and is expected to show a similar pattern of divergence (Hurst and Jiggins 2005; Charlat et al. 2009; Jiang et al. 2018). It is unknown if the common ancestor of these species was infected with *Wolbachia* or if it invaded more recently and then introgressed via hybridization. The sexual and intragenomic conflict arising from possessing a meiotic driver, coupled with maternally inherited endosymbionts, present an opportunity to understand the interplay between these factors and how they impact the evolutionary history of the species group, particularly between the three ingroup species.

In conclusion, we present chromosome assemblies for *D. neotestacea* and *D. putrida*, and de novo contig assemblies for *D. orientacea*, *D. testacea*, and *D. bizonata*. High-quality genomes are critical to understand structural differences between species and consequences of hybridization, and our assemblies provide the first resource that can be used for this species group to uncover regions important to the maintenance of reproductive isolation, meiotic drivers, endosymbiont infection, and other characteristics rendering the *testacea* group an exciting evolutionary genomic model. As a next step, genomic studies in this group should complete scaffolded genomes for all the species in the group, especially for Muller E, which would resolve the fixed rearrangements between species. In addition, gene annotations and the resulting alignments can be improved with the use of obtaining RNA from additional life stages, as well as long-read RNA sequencing methods. Obtaining Y-chromosome assemblies for these species will allow a more complete history of the Y-chromosome to the dot chromosome (Muller F) duplication, as well as if any Y-linked genes may underlie hybrid incompatibilities. Future studies that incorporate polymorphism data within each species will show which regions

of the genome are resistant to introgression, shedding light on the frequency and consequences of hybridization and/or incomplete lineage sorting on genetic variation. Finally, these genomes will be useful in broader phylogenetic studies. For instance, all the species in our study have a similar life history of consuming mushrooms; thus, comparisons with more distantly related species will be needed to tease out genes involved in the transition to utilizing mushrooms. Understanding the mechanism of toxin tolerance is also of great interest (Scott Chialvo and Werner 2018; Scott Chialvo et al. 2019; Erlenbach et al. 2023); identifying genetic associations among lines within species that vary in toxin tolerance, as well as comparing between the *testacea* group species and related non-tolerant species, will shed light on this question.

Methods

Genome Assembly

The laboratory stocks used for each species are as follows: *D. neotestacea* (Rochester, New York, 1990 collected by J. Jaenike); *D. orientacea* (collected by M. Kimura in Japan in 2007); *D. testacea* (St. Sulpice, Switzerland, 2012), *D. putrida* (Lawrence, KS, 2016), and *D. bizonata* (FK05-12 Japanese Stock Center). For *D. orientacea* and *D. bizonata*, DNA was isolated from ten females using the GentraPrep Qiagen DNA extraction kit. Long reads were sequenced with Oxford Nanopore's MinION, with one flow cell per species, following the manufacturer's protocol for genomic DNA by ligation using the SQL-LSK109 kit. Bases were called with Guppy v4.2.2 using the high-accuracy model (Wick et al. 2019). Short-read genomic DNA libraries for these species and *D. neotestacea* were sequenced on the Illumina NovaSeq 6000 for PE 150 bp, filtered with fastp (Chen et al. 2018), and quality checked with FastQC (bioinformatics.babraham.ac.uk).

Long- and short-reads for *D. putrida*, long-reads for *D. neotestacea*, and short-reads for *D. bizonata* were sequenced following the methods of Miller et al. (2018). Briefly, DNA from females was extracted using the Qiagen Blood and Cell Culture DNA Mini Kit, and libraries were prepared using the Ligation Sequencing Kit 1D (Oxford Nanopore) following the manufacturer's protocol. Base-calling was performed with Guppy v4.2.2 using the high-accuracy model (Wick et al. 2019). Short-read sequencing was prepared from females using the Qiagen Blood and Tissue Kit and sequenced on an Illumina NextSeq 500 as PE 150 bp reads. The long- and short-reads for *D. testacea* were sequenced as follows. DNA was extracted from the homogenate of forty male flies using two rounds of phenol:chloroform:isoamyl alcohol (25:24:1) extraction and treated with RNase A (Qiagen). Long reads were sequenced with an Oxford Nanopore minION following the manufacturer's protocol for genomic DNA using the SQL-LSK109 kit. Bases were called with Guppy v4.2.2 using the high-accuracy model (Wick et al. 2019). PCR-free TruSeq LT Illumina sequencing libraries (approximately 500 bp inserts) were generated at Génome Québec. Paired-end 125 bp reads were sequenced on a HiSeq2500 instrument. Raw sequencing data for all species are listed in Table S1.

Genomes were assembled following the pipeline established by Kim et al. (2020). This pipeline begins with running Flye v.2.8.1 (Kolmogorov et al. 2019), a long-read draft genome assembler, followed by Minimap2 v2.17 (Li, 2018) to identify overlaps, and Racon v1.4.13 (Vaser et al. 2017) for long-read polishing, each consecutively run twice. Medaka v1.2.1 (Nanoporetech.com) was used to polish the assembly and create consensus sequences before three rounds of short-read polishing with Pilon v1.23 (Walker et al. 2014). Assembly completeness was assessed with the *diptera* database in BUSCO v4.0.6 (Seppey et al. 2019). Contaminated contigs were identified using BLAST + v2.10.1 (Camacho et al. 2009) and removed from the assembly. The remaining contigs were scaffolded with npScarf (Cao et al. 2017), a scaffold program for Oxford Nanopore sequencing. We generated de novo repeat databases for each genome using RepeatModeler v2.0.1 (Flynn et al. 2020) and soft repeat masked each assembly with RepeatMasker v4.0.9 (Smit et al. 2013–2015) using the assembly's de novo repeat database. Final assembly completeness was determined with N50 and auN calculated by QUAST v5.2.0 (Mikheenko et al. 2018) and BUSCO scores.

We obtained HI-C reads to scaffold contigs for *D. putrida* and *D. neotestacea*. Female adult flies and genome assemblies were sent to Phase Genomics, who performed library preparation, quality check, sequencing, and assembly. Genes known to be on each Muller element (Table S2) were blasted to the scaffolds to identify Muller elements. For the species we do not have scaffolded assemblies for (*D. orientacea*, *D. testacea*, and *D. bizonata*), we identified contigs for each Muller element using a blastn search against each Muller element identified in our scaffolded assemblies and visually inspected for homology.

To annotate each genome, total RNA was extracted for each species from ten 7-day-old females and ten 7-day-old males separately using the TRIzol RNA extraction kit by Invitrogen. Libraries were prepared with the Illumina TruSeq mRNA Stranded Library kit, and PE 150 bp reads were sequenced using the NovaSeq 6000. Reads were filtered with fastp (Chen et al. 2018) and quality checked with FastQC (bioinformatics.babraham.ac.uk). We annotated each genome using the BRAKER pipeline for gene prediction (Hoff et al. 2019). First, we indexed the genome and aligned RNA-seq reads to it with Hisat2 v2.1.0 (Kim et al. 2019), generating a SAM file that was sorted using samtools v1.10. This

was used as the input for BRAKER's braker.pl script along with the masked genome assembly to generate a genome annotation. The number of genes predicted was calculated using BRAKER's selectSupportedSubsets.py script.

Synteny and collinearity between Muller elements were visualized with GENESPACE v1.2.3 (Lovell et al. 2022). For this, bed files of annotations were used as input, along with peptide files for these sequences. Peptide files were generated with gffread (Pertea and Pertea 2020) to create a fasta file from the gff files, and then translated to amino acids using seqkit translate (Shen et al. 2016). GENESPACE was run with default parameters to generate a synteny map. We ran GENESPACE on whole genomes and groups of Muller elements to look at sets of chromosome synteny.

Dot Chromosome Characterization

The dot chromosome (Muller element F) in *Drosophila* possesses a unique set of characteristics relative to other autosomes, including a reduction in size, gene content, and recombination (Riddle and Elgin 2006). In addition, a previous study characterized a Y-chromosome-to-dot duplication in *D. neotestacea*, *D. testacea*, and *D. orientacea* (Dyer et al. 2011). Thus, we sought to characterize the dot chromosome in further detail to understand the presence and effect of this duplication. To identify dot contigs in the unscaffolded assemblies, we created a separate BLAST database for the *D. putrida* and *D. neotestacea* unmasked dot scaffolds using makeblastdb. We then performed a blastn search of the *D. bizonata*, *D. testacea*, and *D. orientacea* assemblies against these databases. Contigs were visually inspected for synteny using Mauve (Darling et al. 2004). We were unable to successfully extract dot contigs in *D. orientacea*, as this assembly was not as contiguous, and many contigs with homology contained non-syntenic regions and did not align well to the other dot sequences; therefore, we do not include this species in analyses regarding the synteny of the dot.

The portion of the dot that originated from the Y chromosome contains the male fertility factor gene *kl-5* (Dyer et al. 2011). Thus, to identify this region we performed a blastn search of *kl-5* to the *D. neotestacea* dot scaffold. *Kl-5* was found at the distal end of the dot (bp: 3937061 to 4062246). To determine the length of this duplicated region, pairwise gene homologs between *D. neotestacea* and *D. putrida* dot annotations was identified by blasting the *D. neotestacea* dot-linked genes to a blast database generated from the *D. putrida* dot-linked genes. Synteny in gene order was manually assessed along both dot scaffolds in Geneious Prime (2022.1.1 Biomatters Ltd). This identified a ~ 249 kb region containing seven annotated genes in *D. neotestacea*. Homology to known genes was identified through BLASTN searches using the NT database. To determine if these seven genes were in the other assemblies, they were blasted against either the annotation or the genome databases for each species. Sequence alignments and homology were visualized in Geneious Prime (2022.1.1 Biomatters Ltd). Read coverage for *kl-5* was assessed by aligning the filtered RNA reads for each species to the *D. neotestacea* dot scaffold using STAR (Dobin et al. 2013), filtering for uniquely mapping reads, and visualizing read coverage in IGV (Robinson et al. 2011).

Scaffold sizes and alignments indicated a large size difference between the *D. neotestacea* and *D. testacea* dot, which have the Y-duplication, compared to the *D. putrida* and *D. bizonata* dot, which do not have the Y-duplication. We assessed repeat content and intron size to identify factors that could be contributing to the dot size differences between these species. Repeat content was compared between species by creating a de novo repeat database with all dot sequences using RepeatModeler v2.0.1 (Flynn et al. 2020) and characterizing repeats on the dot in each species with RepeatMasker v4.0.9 (Smit et al. 2013–2015) using the dot de novo repeat database. Intron sizes for single-copy and BUSCO *dipteran* orthologs were extracted from the gff files, and we examined differences in first, second, and third intron size. A Kruskal-Wallis test was used to identify significant differences across species. The relationship between intron size and repetitive content was visualized using GENESPACE (Lovell et al. 2022) and the tools for a sliding window analysis of gene and repeat content for *D. putrida* and *D. neotestacea*. This was done in 1 Mb windows, with a step size of 100 kb.

Codon usage can indicate long-term changes in the strength of selection (Hershberg and Petrov 2008) and can differ on the dot chromosome from the other autosomes (Vicario et al. 2007); therefore, we calculated codon usage bias with coRdon (Elek et al. 2023) on filtered SCO genes and assessed the effective number of codons (Enc) per gene separately in each species. GC content per gene was calculated with seqkit (Shen et al. 2016). Comparisons for GC content or Enc between species on each Muller element were tested for significance with a pairwise Wilcoxon test in R v4.1.0 with a *P*-value adjustment using the BH method. We calculated the Relative Synonymous Codon Usage (RSCU), a measure of the observed synonymous codon usage compared to a uniform synonymous codon usage, using seqinr (Charif and Lobry 2007).

Comparative Phylogenetics

We generated a species tree for the *testacea* group using maximum likelihood and coalescent-based methods. First, we generated coding sequence fasta files for each species using gffread (Pertea and Pertea 2020) and the gtf annotation file. These fasta files were used to identify single copy

orthologs (SCOs) in an all-by-all blast search in OrthoFinder (Emms and Kelly 2019), and we generated alignments for each SCO using MAFFT (Katoh et al. 2019). We identified 6,411 SCOs. Because these alignments may contain errors due to incorrect gene annotations, we attempted to use only the most robust alignments by removing orthologs where more than 50% of the sequence was gaps. This removed 1,543 (24%) orthologs and left 4,868 (76%) filtered orthologs. This threshold was determined by investigating how alignment quality impacted the calculation of omega for each gene and determining that alignments with less than 50% gaps in the sequence did not have a significant shift in omega value (data not shown). This was identified in the ete3 package (see *Patterns of Molecular Evolution*), where codeml was run with either cleandata = 0 (no removal) or cleandata = 1 (removal of gaps). We built the maximum likelihood tree using a concatenated alignment of all genes with IQ-Tree (Nguyen et al. 2015) using ModelFinder (Kalyaanamoorthy et al. 2017) and the ultrafast bootstrap approximation (Hoang et al. 2018) with 1000 bootstraps. For the coalescent analysis, we generated individual gene trees for SCOs in RAxML v4 (Stamatakis 2014) with the GTRCAT model and 1000 bootstraps. These gene trees were the input for a coalescent analysis with ASTRAL-III (Zhang et al. 2018), and we performed 100 bootstraps on the resulting species tree.

Our species tree indicated that *D. neotestacea* and *D. testacea* are sister to each other; however, hybrids are produced only between *D. testacea* and *D. orientacea*, and they both have strong reproductive isolation with *D. neotestacea* (Grimaldi et al. 1992). Therefore, we assessed gene tree discordance along our phylogeny. We used the coalescent phylogeny as our species tree and calculated the quartet score (-t 1) with ASTRAL-III -q (Zhang et al. 2018) and the gene trees. This score determines the percentage of quartets in the gene tree that support a branch and helps to identify if there is gene tree conflict around it. We also assessed alternative quartet topologies (-t 8), identifying quartet support for the main topology, first alternative, and second alternative. Additionally, we identified gene and site concordance factors (gCF and sCF) in IQ-TREE 2 (Nguyen et al. 2015; Minh et al. 2020). These analyses were done for all SCOs, and separately for SCOs on each Muller element to assess if there were differences in the phylogeny and discordance between individual Muller elements. The location of discordant genes was visualized with chromoMap (Anand and Rodriguez Lopez 2022).

To test for putative introgression between species, we ran a four-taxon ABBA-BABA test using dfoil (Pease and Hahn 2015) and the filtered SCO dataset. Fasta files were provided as input, and the fasta2dfoil.py script was used to first generate counts files for each SCO, followed by dfoil.py, where it was run with the dstat mode to calculate the D statistic per ortholog. The three ingroup species were *D. neotestacea*, *D. testacea*, and *D. orientacea*. This test was run twice, using *D. putrida* or *D. bizonata* as the outgroup, as we could not perform a five-taxon introgression test with the asymmetric phylogeny. Genes were defined as introgressed if they had a significant *P*-value, as calculated in dfoil using a chi-square test. The location of putatively introgressed genes using either outgroup was visualized with chromoMap (Anand and Rodriguez Lopez 2022). D scores were plotted across each Muller element using ggplot (Wickham 2016). We tested if there were significant differences in estimates of D among each of the Muller elements using a Kruskal-Wallis test followed by pairwise Steel–Dwass tests between Muller elements.

Patterns of Molecular Evolution

We identified the rate of molecular evolution of genes in the filtered set of SCOs by assessing patterns of nonsynonymous to synonymous substitutions (dN/dS, also called omega) using ete3 evol (Huerta-Cepas et al. 2016). The phylogeny used was the coalescent tree, which is the same as the ML tree (Fig. S7). For each gene, we ran two models of molecular evolution, including a null model where omega is constant over the entire tree, and a branch model where the (*D. neotestacea*, *D. testacea*, and *D. orientacea*) clade had one omega value, and the outgroups could take on a different omega value. Likelihood ratio tests were used to determine if genes were evolving differently under the branch model versus the null model. Multiple testing correction was done using a false discovery rate (FDR) of 0.05 and calculated in the R package qvalue (Storey et al. 2024). Homology to genes was identified using the *D. melanogaster* Uniprot database (UP000000803). To determine if biological processes are overrepresented in the genes with significant results for the branch model, we performed a GO analysis using the DAVID Database (Sherman et al. 2022).

We evaluated divergence between species pairs using estimates of *Ks* as calculated by PAML in ete3. The results of the null model (M0) were used to calculate this value. We obtained pairwise divergence between each species pair by either calculating the average or median of *Ks* across all filtered SCOs.

Mitogenome Assembly

We assembled the mitochondrial genome for each species using a variety of methods. First, for our scaffolded genomes (*D. putrida* and *D. neotestacea*), we assessed if any unscaffolded sequences contained mitochondrial DNA. The *D. melanogaster* (NC_001709.1) and *D. innubila* (CM015047.1) mitochondrial sequences were obtained from NCBI, and a blast database was generated for each using makeblastdb. We ran a blastn search for both

scaffolded assemblies against each database to identify putative mitochondrial scaffolds. In *D. neotestacea*, one scaffold (Scaffold265__unscaffolded) consisted of mitochondrial DNA and appeared to be looping (ie had multiple copies of the mitochondrial genome assembled). Based on homology to the reference mitochondrial assemblies, we identified the region of this sequence representing one full mitochondrial genome (bp: 8932 to 23963). In *D. putrida*, one scaffold (Scaffold115__unscaffolded) consisted of hits to the mitochondrial references; however, the middle region of the reference mitogenomes appeared to be missing from this scaffold. We expanded this analysis to the other assemblies and identified a contig in *D. bizonata* (Scaffold115) representing a looping mitochondrion, with the consecutive mitochondrial sequence in the middle (bp: 8273 to 23484).

We could not confidently identify mitochondrial sequences from contigs in *D. orientacea*, *D. testacea*, and *D. putrida*, thus, we assembled the mitochondria from our long-read libraries. Protein sequences for each mitochondrial gene in *D. melanogaster* were obtained from NCBI (YP_009047266.1-YP_009047267.1), and a DIAMOND (Buchfink et al. 2021) blastx search was done on each species' raw long-read fastq.gz files. The resulting fastq reads were subsetted with seqtk (https://github.com/lh3/seqtk) and aligned with Flye (Kolmogorov et al. 2019).

The assemblies were visualized for completeness to the *D. innubila* and *D. melanogaster* mitochondrial sequences in Mauve (Darling et al. 2004) and Geneious Prime (2022.1.1 Biomatters Ltd). To annotate the mitochondria, we used MITOS2 (Donath et al. 2019) and the dm6 database, which generated bed files of mitochondrial genes for each species using homology searches. The mitochondrial assemblies were aligned with MAFFT (Katoh et al. 2019), and a maximum likelihood tree was generated with IQ-TREE 2 (Minh et al. 2020), using ModelFinder (Kalyaanamoorthy et al. 2017) to identify the best fit model, after which 1000 ultrafast bootstrap replicates were used to infer tree confidence (Hoang et al. 2018).

Supplementary Material

Supplementary material is available at *Genome Biology and Evolution* online.

Acknowledgments

We are grateful to Paul Ginsberg and Casey Bergman for helpful comments and technical advice.

Funding

This work was supported by the National Institutes of Health (under T32GM007103 to TE), a grant from the Natural Sciences and Engineering Research Council (NSERC) of Canada to SJP, a National Science Foundation CAREER Award (2047052) to RLU, and a grant from the National Science Foundation (DEB1737824) to KAD.

Conflict of Interest

The authors have no conflicts of interest to declare.

Data Availability

Raw sequence reads for genome assembly and annotation are available through NCBI's Short Read Archive under Project PRJNA1112834. Genome assemblies are available on NCBI under Project PRJNA1112834. Scripts for analyses are available on GitHub (https://github.com/trm76056/Genomes).

Literature Cited

Anand L, Rodriguez Lopez CM. ChromoMap: an R package for interactive visualization of multi-omics data and annotation of chromosomes. BMC Bioinformatics. 2022:23:33. https://doi.org/10.1186/s12859-021-04556-z.

Arnegard ME, et al. Genetics of ecological divergence during speciation. Nature. 2014:511:307–311. https://doi.org/10.1038/nature13301.

Barbash DA. Ninety years of *Drosophila melanogaster* hybrids. Genetics. 2010:186:1–8. https://doi.org/10.1534/genetics.110.121459.

Bhutkar A, et al. Chromosomal rearrangement inferred from comparisons of 12 *Drosophila* genomes. Genetics. 2008:179:1657–1680. https://doi.org/10.1534/genetics.107.086108.

Bost A, et al. Functional variation in the gut microbiome of wild *Drosophila* populations. Mol Ecol. 2018:27:2834–2845. https://doi.org/10.1111/mec.14728.

Buchfink B, Reuter K, Drost H-G. Sensitive protein alignments at tree-of-life scale using DIAMOND. Nat Methods. 2021:18:366–368. https://doi.org/10.1038/s41592-021-01101-x.

Camacho C, et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009:10:421. https://doi.org/10.1186/1471-2105-10-421.

Cao MD, et al. Scaffolding and completing genome assemblies in real-time with nanopore sequencing. Nat Commun. 2017:8:14515. https://doi.org/10.1038/ncomms14515.

Capinteyro-Ponce J, Macahdo CA. The complex landscape of structural divergence between the *Drosophila pseudoobscura* and *D. persimilis* genomes. Genome Biol Evol. 2024:16:evae047. https://doi.org/10.1093/gbe/evae047.

Chakraborty M, et al. Evolution of genome structure in the *Drosophila simulans* species complex. Genome Res. 2021:31:380–396. https://doi.org/10.1101/gr.263442.120.

Charif D, Lobry JR. Seqinr 1.0-2: a contributed package to the R project for statistical computing devoted to biological sequences retrieval and analysis. In: Bastolla U, Porto M, Roman HE, Vendruscolo M, editors. Structural approaches to sequence evolution molecules, networks, populations. Springer; 2007. p. 207–232.

Charlat S, et al. The joint evolutionary histories of *Wolbachia* and mitochondria in *Hypolimnas bolina*. BMC Ecol Evol. 2009:9:64. https://doi.org/10.1186/1471-2148-9-64.

Charlesworth B, Campos JL, Jackson BC. Faster-X evolution: theory and evidence from *Drosophila*. Mol Ecol. 2018:27:3753–3771. https://doi.org/10.1111/mec.14534.

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

Clark AG, et al. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature. 2007:450:203–218. https://doi.org/10.1038/nature06341.

Connallon T, et al. Local adaptation and the evolution of inversions on sex chromosomes and autosomes. Philos Trans R Soc Lond B Biol Sci 2018:373:20170423. https://doi.org/10.1098/rstb.2017.0423.

Coughlan JM, Willis JH. Dissecting the role of a large chromosomal inversion in life history divergence throughout the *Mimulus guttatus* species complex. Mol Ecol. 2019:28:1343–1357. https://doi.org/10.1111/mec.14804.

Coyne JA, Orr HA. Patterns of speciation in *Drosophila* revisted. Evolution. 1997:51:295–303. https://doi.org/10.1111/j.1558-5646.1997.tb02412.x.

Coyne JA, Orr HA. The evolutionary genetics of speciation. Philos Trans R Soc Lond B Biol Sci. 1998:353:287–305. https://doi.org/10.1098/rstb.1998.0210.

Coyne JA, Orr HA. Speciation. Sinauer Associations Sunderland; 2004.

Crawford JE, et al. Reticulate speciation and barriers to introgression in the *Anopheles gambiae* species complex. Genome Biol Evol. 2015:7:3116–3131. https://doi.org/10.1093/gbe/evv203.

D'Alterio C, et al. *Drosophila melanogaster* Cad99C, the orthologue of human Usher cadherin PCDH15, regulates the length of microvilli. J Cell Sci. 2005:171:549–558. https://doi.org/10.1083/jcb.200507072.

Dagilis AJ, Matute DR. The fitness of an introgressing haplotype changes over the course of divergence and depends on its size and genomic location. PLoS Biol. 2023:21:e3002185. https://doi.org/10.1371/journal.pbio.3002185.

Darling ACE, Mau B, Blattner FR, Perna NT. Mauve: multiple alignment of conserved genomic sequence with rearrangements. Genome Res. 2004:14:1394–1403. https://doi.org/10.1101/gr.2289704.

De Cahsan B, et al. Southern introgression increases adaptive immune gene variability in northern range margin populations of fire-bellied toad. Ecol Evol. 2021:11:9776–9790. https://doi.org/10.1002/ece3.7805.

Dobin A, et al. STAR: ultrafast universal RNA-Seq aligner. Bioinformatics. 2013:29:15–21. https://doi.org/10.1093/bioinformatics/bts635.

Dobzhansky T. Genetics and the origin of species. Columbia University Press; 1937.

Donath A, et al. Improved annotation of protein-coding genes boundaries in metazoan mitochondrial genomes. Nucleic Acids Res. 2019:47:10543–10552. https://doi.org/10.1093/nar/gkz833.

Dowling DK, Wolff JN. Evolutionary genetics of the mitochondrial genome: insights from *Drosophila*. Genetics. 2023:224:iyad036. https://doi.org/10.1093/genetics/iyad036.

Dyer KA. Local selection underlies the geographic distribution of sex-ratio drive in *Drosophila neotestacea*. Evolution. 2012:66:973–984. https://doi.org/10.1111/j.1558-5646.2011.01497.x.

Dyer KA, White BE, Bray MJ, Pique DG, Betancourt AJ. Molecular evolution of a Y chromosome to autosome gene duplication in *Drosophila*. Mol Biol Evol. 2011:28:1293–1306. https://doi.org/10.1093/molbev/msq334.

Eberlein C, et al. Hybridization is a recurrent evolutionary stimulus in wild yeast speciation. Nat Commun. 2019:10:923. https://doi.org/10.1038/s41467-019-08809-7.

Edelman NB, et al. Genomic architecture and introgression shape a butterfly radiation. Science. 2019:366:594–599. https://doi.org/10.1126/science.aaw2090.

Edelman NB, Mallet J. Prevalence and adaptive impact of introgression. Annu Rev Genet. 2021:55:265–283. https://doi.org/10.1146/annurev-genet-021821-020805.

Elek A, Kuzman M, Vlahovicek K. *coRdon: codon usage analysis and prediction of gene expressivity*. R package version 1.20.0. 2023. https://bioconductor.org/packages/coRdon.

Emms DM, Kelly S. OrthoFinder: phylogenetic orthology inference for comparative genomics. Genome Biol. 2019:20:238. https://doi.org/10.1186/s13059-019-1832-y.

Erlenbach T, et al. Investigating the phylogenetic history of toxin tolerance in mushroom-feeding *Drosophila*. Ecol Evol. 2023:13:e10736. https://doi.org/10.1002/ece3.10736.

Fallon TR, et al. Firefly genomes illuminate parallel origins of bioluminescence in beetles. eLife. 2018:7:e36495. https://doi.org/10.7554/eLife.36495.

Feder JL, Nosil P. The efficacy of divergence hitchhiking in generating genomic islands during ecological speciation. Evolution. 2010:64:1729–1747. https://doi.org/10.1111/j.1558-5646.2009.00943.x.

Fijarczyk A, Dudek K, Niedzicka M, Babik W. Balancing selection and introgression of newt immune-response genes. Proc Natl Acad Sci U S A. 2018:285:20180819. https://doi.org/10.1098/rspb.2018.0819.

Finet C, et al. DrosoPhyla: resources for drosophilid phylogeny and systematics. Genome Biol Evol. 2021:13:evab179. https://doi.org/10.1093/gbe/evab179.

Fishman L, Stathos A, Beardsley PM, Williams CF, Hill JP. Chromosomal rearrangements and the genetics of reproductive barriers in *mimulus* (monkey flowers). Evolution. 2013:67:2547–2560. https://doi.org/10.1111/evo.12154.

Flynn JM, et al. RepeatModeler2 for automated genomic discovery of transposable element families. Proc Natl Acad Sci U S A. 2020:117:9451–9457. https://doi.org/10.1073/pnas.1921046117.

Fuller ZL, Leonard CJ, Young RE, Schaeffer SW, Phadnis N. Ancestral polymorphisms explain the role of chromosomal inversions in speciation. PLoS Genet. 2018:14:e1007526. https://doi.org/10.1371/journal.pgen.1007526.

Garrigan D, et al. Genome sequencing reveals complex speciation in the *Drosophila simulans* clade. Genome Res. 2012:22:1499–1511. https://doi.org/10.1101/gr.130922.111.

Ginsberg P, Humphreys DP, Dyer KA. Ongoing hybridization obscures phylogenetic relationships in the *Drosophila subquinaria* species complex. J Evol Biol. 2019:32:1093–1105. https://doi.org/10.1111/jeb.13512.

Giska I, et al. Introgression drives repeated evolution of winter coat color polymorphism in hares. Proc Natl Acad Sci U S A. 2019:116:24150–24156. https://doi.org/10.1073/pnas.1910471116.

Grimaldi D, James AC, Jaenike J. Systematics and modes of reproductive isolation in the holarctic *Drosophila testacea* species group (Diptera: Drosophilidae). Ann Entomol Soc Am. 1992:85:671–685. https://doi.org/10.1093/aesa/85.6.671.

Guillén Y, et al. Genomics of ecological adaptation in cactophilic *Drosophila*. Genome Biol Evol. 2015:7:349–366. https://doi.org/10.1093/gbe/evu291.

Hahn MW, Han MV, Han S-G. Gene family evolution across 12 *Drosophila* genomes. PLoS Genet. 2007:3:e197. https://doi.org/10.1371/journal.pgen.0030197.

Hamlin JAP, Hibbins MS, Moyle LC. Assessing biological factors affecting postspeciation introgression. Evol Lett. 2020:4:137–154. https://doi.org/10.1002/evl3.159.

Harrison RG. The language of speciation. Evolution. 2012:66:3643–3657. https://doi.org/10.1111/j.1558-5646.2012.01785.x.

Harrison RG, Larson EL. Hybridization, introgression, and the nature of species boundaries. J Hered. 2014:105:795–809. https://doi.org/10.1093/jhered/esu033.

Hedrick PW. Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. Mol Ecol. 2013:22:4606–4618. https://doi.org/10.1111/mec.12415.

Heikkinen E, Lumme J. The Y chromosomes of *Drosophila lummei* and *D. novamexicana* differ in fertility factors. Heredity (Edinb). 1998:81:505–513. https://doi.org/10.1046/j.1365-2540.1998.00422.x.

Helleu Q, Gérard PR, Montchamp-Moreau C. Sex chromosome drive. Cold Spring Harb Perspect Biol. 2015:7:a017616. https://doi.org/10.1101/cshperspect.a017616.

Herrig DK, et al. Whole genomes reveal evolutionary relationships and mechanisms underlying gene-tree discordance in *Neodiprion* sawflies. Syst Biol. 2024:73:839–860. https://doi.org/10.1093/sysbio/syae036.

Hershberg R, Petrov D. Selection on codon bias. Annu Rev Genet. 2008:42:287–299. https://doi.org/10.1146/annurev.genet.42.110807.091442.

Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: improving the ultrafast bootstrap approximation. Mol Biol Evol. 2018:35:518–522. https://doi.org/10.1093/molbev/msx281.

Hoff K, Lomsadze A, Borodovsky M, Stanke M. Whole-genome annotation with BRAKER. Methods Mol Biol. 2019:1962:65–95. https://doi.org/10.1007/978-1-4939-9173-0_5.

Hoffmann AA, Rieseberg LH. Revisiting the impact of inversions in evolution: from population genetic markers to drivers of adaptive shifts and speciation? Annu Rev Ecol Evol Syst. 2008:39:21–42. https://doi.org/10.1146/annurev.ecolsys.39.110707.173532.

Hu Q, et al. Zinc dynamics during *Drosophila* oocyte maturation and egg activation. iScience. 2020:23:101275. https://doi.org/10.1016/j.isci.2020.101275.

Huerta-Cepas J, Serra F, Bork P. ETE 3: reconstruction, analysis, and visualization of phylogenomic data. Mol Biol Evol. 2016:33:1635–1638. https://doi.org/10.1093/molbev/msw046.

Hurst GDD, Jiggins FM. Problems with mitochondrial DNA as a marker in population, phylogeographic and phylogenetic studies: the effects of inherited symbionts. Proc R Soc Lond B Biol Sci. 2005:272:1525–1534. https://doi.org/10.1098/rspb.2005.3056.

Hurst GDD, Schilthuizen M. Selfish genetic elements and speciation. Heredity (Edinb). 1998:80:2–8. https://doi.org/10.1046/j.1365-2540.1998.00337.x.

Izumitani HF, Kusaka Y, Koshikawa S, Toda MJ, Katoh K. Phylogeography of the subgenus *Drosophila* (Diptera: Drosophilidae): evolutionary history of faunal divergence between the old and the new worlds. PLoS One. 2016:11:e0160051. https://doi.org/10.1371/journal.pone.0160051.

Jaenike J. X chromosome drive. Curr Biol. 2008:18:R508–R511. https://doi.org/10.1016/j.cub.2008.03.051.

Jaenike J, Grimaldi DA, Sluder AE, Greenleaf AL. α-Amanitin tolerance in mycophagous *Drosophila*. Science. 1983:221:165–167. https://doi.org/10.1126/science.221.4606.165.

Jaenike J, Stahlhut JK, Boello LM, Unckless RL. Association between Wolbachia and Spiroplasma within *Drosophila neotestacea:* an emerging symbiotic mutualism? Mol Ecol. 2010a:19:414–425. https://doi.org/10.1111/j.1365-294X.2009.04448.x.

Jaenike J, Unckless RL, Cockburn SA, Boelio LM, Perlman SJ. Adaptation via symbiosis: recent spread of a *Drosophila* defensive symbiont. Science. 2010b:329:212–215. https://doi.org/10.1126/science.1188235.

James AC, Jaenike J. "Sex ratio" meiotic drive in *Drosophila testacea*. Genetics. 1990:126:651–656. https://doi.org/10.1093/genetics/126.3.651.

Jiang W, et al. Influence of Wolbachia infection on mitochondrial DNA variation in the genus *Polytremis* (Lepidoptera: Hesperiidae). Mol Phylogenet Evol. 2018:129:158–170. https://doi.org/10.1016/j.ympev.2018.08.001.

Jones FC, et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature. 2012:484:55–61. https://doi.org/10.1038/nature10944.

Kalyaanamoorthy S, Minh BQ, Wong TKF, von Haeseler A, Jermiin LS. ModelFinder: fast model selection for accurate phylogenetic estimates. Nat Methods. 2017:14:587–589. https://doi.org/10.1038/nmeth.4285.

Katoh K, Rozewicki J, Yamada KD. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief Bioinform. 2019:20:1160–1166. https://doi.org/10.1093/bib/bbx108.

Kaufman TC. A short history and description of *Drosophila melanogaster* classical genetics: chromosome aberrations, forward genetic screens, and the nature of mutations. Genetics. 2017:206:665–689. https://doi.org/10.1534/genetics.117.199950.

Keais GL, Hanson MA, Gowen BE, Perlman SJ. X chromosome drive in a widespread palearctic woodland fly, *Drosophila testacea*. J Evol Biol. 2017:30:1185–1194. https://doi.org/10.1111/jeb.13089.

Khodwekar S, Gailing O. Evidence for environment-dependent introgression of adaptive genes between two red oak species with different drought adaptations. Am J Bot. 2017:104:1088–1098. https://doi.org/10.3732/ajb.1700060.

Kikkawa H, Peng FT. *Drosophila* species of Japan and adjacent localities. Jpn J Zool. 1938:7:507–552.

Kim BY, et al. Highly contiguous assemblies of 101 drosophilid genomes. eLife. 2020:10:e66405. https://doi.org/10.1101/2020.12.14.422775.

Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nat Biotechnol. 2019:37:907–915. https://doi.org/10.1038/s41587-019-0201-4.

Kimura MT. Cold and heat tolerance of drosophilid flies with reference to their latitudinal distributions. Oecologia. 2004:140:442–449. https://doi.org/10.1007/s00442-004-1605-4.

Kirkpatrick M. How and why chromosome inversions evolve. PLoS Biol. 2010:8:e1000501. https://doi.org/10.1371/journal.pbio.1000501.

Kokate PP, Smith M, Hall L, Zhang K, Werner T. Inter- and intraspecific variation in mycotoxin tolerance: a study of four *Drosophila* species. Ecol Evol. 2022:12:e9126. https://doi.org/10.1002/ece3.9126.

Kokate PP, Techtmann SM, Werner T. Codon usage bias and dinucleotide preference in 29 *Drosophila* species. G3 (Bethesda). 2021:11:jkab191. https://doi.org/10.1093/g3journal/jkab191.

Kolmogorov M, Yuan J, Lin Y, Pevzner PA. Assembly of long, error-prone reads using repeat graphs. Nat Biotechnol. 2019:37:540–546. https://doi.org/10.1038/s41587-019-0072-8.

Kulathinal RJ, Stevison LS, Noor MAF. The genomics of speciation in *Drosophila*: diversity, divergence, and introgression estimated using low-coverage genome sequencing. PLoS Genet. 2009:5:e1000550. https://doi.org/10.1371/journal.pgen.1000550.

Lamnissou K, Loukas M, Zouros E. Incompatibilities between Y chromosome and autosomes are responsible for male hybrid sterility in crosses between *Drosophila virilis* and *Drosophila texana*. Heredity (Edinb). 1996:76:603–609. https://doi.org/10.1038/hdy.1996.86.

Leung W, et al. Retrotransposons are the major contributors to the expansion of the *Drosophila ananassae* Muller F element. G3 (Bethesda). 2017:7:2439–2460. https://doi.org/10.1534/g3.117.040907.

Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191.

Llopart A, Herrig D, Brud E, Stecklein Z. Sequential adaptive introgression of the mitochondrial genome in *Drosophila yakuba* and *Drosophila santomea*. Mol Ecol. 2014:23:1124–1136. https://doi.org/10.1111/mec.12678.

Lohse K, Clarke M, Ritchie MG, Etges WJ. Genome-wide tests for introgression between cactophilic *Drosophila* implicate a role of inversions during speciation. Evolution. 2015:69:1178–1190. https://doi.org/10.1111/evo.12650.

Lovell JT, et al. GENESPACE tracks regions of interest and gene copy number variation across multiple genomes. eLife. 2022:11: e78526. https://doi.org/10.7554/eLife.78526.

Lucek K, Gompert Z, Nosil P. The role of structural genomic variants in population differentiation and ecotype formation in *Timema cristinae* walking sticks. Mol Ecol. 2019:28:1224–1237. https://doi.org/10.1111/mec.15016.

Maheshwari S, Barbash DA. The genetics of hybrid incompatibilities. Annu Rev Genet. 2011:45:331–355. https://doi.org/10.1146/annurev-genet-110410-132514.

Mai D, Nalley MJ, Bachtrog D. Patterns of genomics differentiation in the *Drosophila nasuta* species complex. Mol Biol Evol. 2020:37: 208–220. https://doi.org/10.1093/molbev/msz215.

Makhov IA, Gorodilova YYU, Lukhtanov VA. Sympatric occurrence of deeply diverged mitochondrial DNA lineages in Siberian geometrid moths (Lepidoptera: Geometridae): cryptic speciation, mitochondrial introgression, secondary admixture or effect of *Wolbachia*? Biol J Linn Soc Lond. 2021:134:342–365. https://doi.org/10.1093/biolinnean/blab089.

Marques DA, Meier JI, Seehausen O. A combinatorial view on speciation and adaptive radiation. Trends Ecol Evol. 2019:34: 531–544. https://doi.org/10.1016/j.tree.2019.02.008.

Martin SH, Davey J, Salazar C, Jiggins CD. Recombination rate variation shapes barriers to introgression across butterfly genomes. PLoS Biol. 2019:17:e2006288. https://doi.org/10.1371/journal.pbio.2006288.

Martinson VG, Douglas AE, Jaenike J. Community structure of the gut microbiota in sympatric species of *Drosophila*. Ecol Lett. 2017:20: 629–639. https://doi.org/10.1111/ele.12761.

Mastrantonio V, Porretta D, Urbanelli S, Crasta G, Nascetti G. Dynamics of mtDNA introgression during species range expansion: insights from an experimental longitudinal study. Sci Rep. 2016:6: 30355. https://doi.org/10.1038/srep30355.

Meiklejohn CD, et al. Gene flow mediates the role of sex chromosome meiotic drive during complex speciation. eLife. 2018:7:e35468. https://doi.org/10.7554/eLife.35468.

Meiklejohn CD, Montooth KL, Rand DM. Positive and negative selection on the mitochondrial genome. Trends Genet. 2007:23: 259–263. https://doi.org/10.1016/j.tig.2007.03.008.

Meisel RP, Connallon T. The faster-X effect: integrating theory and data. Trends Genet. 2013:29:537–544. https://doi.org/10.1016/j.tig.2013.05.009.

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

Miller DE, Staber C, Zeitlinger J, Hawley RS. Highly contiguous genome assemblies of 15 *Drosophila* species generated using nanopore sequencing. G3 (Bethesda). 2018:8:3131–3141. https://doi.org/10.1534/g3.118.200160.

Minh BQ, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. Mol Biol Evol. 2020:37:1530–1534. https://doi.org/10.1093/molbev/msaa015.

Moest M, et al. Selective sweeps on novel and introgressed variation shape mimicry loci in a butterfly adaptive radiation. PLoS Biol. 2020:18:e3000597. https://doi.org/10.1371/journal.pbio.3000597.

Moran BM, et al. The genomic consequences of hybridization. eLife. 2021:10:e69016. https://doi.org/10.7554/eLife.69016.

Moritz C, Dowling TE, Brown WM. Evolution of animal mitochondrial DNA: relevance for population biology and systematics. Annu Rev Ecol Syst. 1987:18:269–292. https://doi.org/10.1146/annurev.es.18.110187.001413.

Nguyen L-T, Schmidt HA, von Haeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. Mol Biol Evol. 2015:32:268–274. https://doi.org/10.1093/molbev/msu300.

Nichols R. Gene trees and species trees are not the same. Trends Ecol Evol. 2001:16:358–364. https://doi.org/10.1016/S0169-5347(01)02203-0.

Noor MA, Grams KL, Bertucci LA, Reiland J. Chromosomal inversions and the reproductive isolation of species. Proc Natl Acad Sci U S A. 2001:98:12084–12088. https://doi.org/10.1073/pnas.221274498.

Nosil P, Feder JL. Genomic divergence during speciation: causes and consequences. Philos Trans R Soc Lond B Biol Sci 2012:367: 332–342. https://doi.org/10.1098/rstb.2011.0263.

Owens GL, Baute GJ, Rieseberg LH. Revisiting a classic case of introgression: hybridization and gene flow in Californian sunflowers. Mol Ecol. 2016:25:2630–2643. https://doi.org/10.1111/mec.13569.

Patterson JT, Stone WS. Evolution in the genus Drosophila. The Macmillan Company; 1952.

Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. Mol Ecol. 2016:25:2337–2360. https://doi.org/10.1111/mec.13557.

Pease JB, Haak DC, Hahn MW, Moyle LC. Phylogenomics reveals three sources of adaptive variation during a rapid radiation. PLoS Biol. 2016:14:e1002379. https://doi.org/10.1371/journal.pbio.1002379.

Pease JB, Hahn MW. More accurate phylogenies inferred from low-recombination regions in the presence of incomplete lineage sorting. Evolution. 2013:67:2376–2384. https://doi.org/10.1111/evo.12118.

Pease JB, Hahn MW. Detection and polarization of introgression in a five-taxon phylogeny. Syst Biol. 2015:64:651–662. https://doi.org/10.1093/sysbio/syv023.

Perlman SJ, Spicer GS, Shoemaker DD, Jaenike J. Associations between mycophagous *Drosophila* and their *Howardula* nematode parasites: a worldwide phylogenetic shuffle. Mol Ecol. 2003:12: 237–249. https://doi.org/10.1046/j.1365-294x.2003.01721.x.

Pertea G, Pertea M. GFF utilities: GffRead and GffCompare. F1000Res. 2020:9:304. https://doi.org/10.12688/f1000research.23297.1.

Pieper KE, Unckless RL, Dyer KA. A fast-evolving X-linked duplicate of *importin-α2* is overexpressed in *sex-ratio* drive in *Drosophila neotestacea*. Mol Ecol. 2018:27:5165–5179. https://doi.org/10.1111/mec.14928.

Poikela N, Laetsh DR, Hoikkala V, Lohse K, Kankare M. Chromosomal inversions and the demography of speciation in *Drosophila montana* and *Drosophila flavomontana*. Genome Biol Evol. 2024:16: evae024. https://doi.org/10.1093/gbe/evae024.

Pollard DA, Iyer VN, Moses AM, Eisen MB. Widespread discordance of gene trees with species tree in *Drosophila*: evidence for incomplete lineage sorting. PLoS Genet. 2006:2:e173. https://doi.org/10.1371/journal.pgen.0020173.

Powell JR, et al. Nonrecombining genes in a recombination environment: the *Drosophila* "dot" chromosome. Mol Biol Evol. 2011:28:825–833. https://doi.org/10.1093/molbev/msq258.

Presgraves DC. Sex chromosomes and speciation in *Drosophila*. Trends Genet. 2008:24:336–343. https://doi.org/10.1016/j.tig.2008.04.007.

Presgraves DC. Darwin and the origin of interspecific genetic incompatibilities. Am Nat. 2010:176:S45–S60. https://doi.org/10.1086/657058.

Presgraves DC, Meiklejohn CD. Hybrid sterility, genetic conflict and complex speciation: lessons from the *Drosophila simulans* clade

species. Front Genet. 2021:12:669045. https://doi.org/10.3389/fgene.2021.669045.
Qvarnström A, Bailey RI. Speciation through evolution of sex-linked genes. Heredity (Edinb). 2009:102:4–15. https://doi.org/10.1038/hdy.2008.93.
Ranz JM, et al. Principles of genome evolution in the *Drosophila melanogaster* species group. PLoS Biol. 2007:5:e152. https://doi.org/10.1371/journal.pbio.0050152.
Reis M, Vieira CP, Lata R, Posnein N, Vieira J. Origin and consequences of chromosomal inversions in the *virilis* group of *Drosophila*. Genome Biol Evol. 2018:10:3152–3166. https://doi.org/10.1093/gbe/evy239.
Riddle NC, Elgin SCR. The dot chromosome of *Drosophila*: insights into chromatin states and their change over evolutionary time. Chromosome Res. 2006:14:405–416. https://doi.org/10.1007/s10577-006-1061-6.
Robinson JR, et al. Integrative genomics viewer. Nat Biotechnol. 2011:29:24–26. https://doi.org/10.1038/nbt.1754.
Schluter D, Rieseberg LH. Three problems in the genetics of speciation by selection. Proc Natl Acad Sci U S A. 2022:119:e2122153199. https://doi.org/10.1073/pnas.2122153119.
Schumer M, Cui R, Powell DL, Rosenthal GG, Andolfatto P. Ancient hybridization and genomic stabilization in a swordtail fish. Mol Ecol. 2016:25:2661–2679. https://doi.org/10.1111/mec.13602.
Schumer M, et al. Natural selection interacts with recombination to shape the evolution of hybrid genomes. Science. 2018:360:656–660. https://doi.org/10.1126/science.aar3684.
Scott Chialvo CH, Werner T. *Drosophila*, destroying angels, and deathcaps! Oh my! A review of mycotoxin tolerance in the genus *Drosophila*. Front Biol (Beijing). 2018:13:91–102. https://doi.org/10.1007/s11515-018-1487-1.
Scott Chialvo CH, White BE, Reed LK, Dyer KA. A phylogenetic examination of host use evolution in the *quinaria* and *testacea* groups of *Drosophila*. Mol Phylogenet Evol. 2019:130:233–243. https://doi.org/10.1016/j.ympev.2018.10.027.
Semenov GA, et al. Asymmetric introgression reveals the genetic architecture of a plumage trait. Nat Commun. 2021:12:1019. https://doi.org/10.1038/s41467-021-21340-y.
Sendell-Price AT, et al. The genomic landscape of divergence across the speciation continuum in island-colonising silvereyes (*Zosterops lateralis*). G3 (Bethesda). 2020:10:3147–3163. https://doi.org/10.1534/g3.120.401352.
Seppey M, Manni M, Zdobnov EM. BUSCO: assessing genome assembly and annotation completeness. In: Kollmar M, editor. Gene prediction: methods and protocols. Springer; 2019. p. 227–245.
Shen W, Le S, Li Y, Hu F. SeqKit: a cross-platform and ultrafast toolkit for FASTA/Q file manipulation. PLoS One. 2016:11:e0163962. https://doi.org/10.1371/journal.pone.0163962.
Sherman BT, et al. DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Res. 2022:50:W216–W221. https://doi.org/10.1093/nar/gkac194.
Sinclair-Waters M, et al. Ancient chromosomal rearrangement associated with local adaptation of a postglacially colonized population of Atlantic cod in the northwest Atlantic. Mol Ecol. 2018:27:339–351. https://doi.org/10.1111/mec.14442.
Smit AF, Hubley R, Green P. *RepeatMasker Open-4.0*. 2013-2015. http://www.repeatmasker.org.
Stamatakis A. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics. 2014:30:1312–1313. https://doi.org/10.1093/bioinformatics/btu033.
Stebbins GL. The inviability, weakness, and sterility of interspecific hybrids. In: Advances in genetics. Academic Press; 1958. p. 147–215.
Stevison LS, Hoehn KB, Noor MAF. Effects of inversions on within- and between-species recombination and divergence. Genome Biol Evol. 2011:3:830–841. https://doi.org/10.1093/gbe/evr081.
Storey JD, Bass AJ, Dabney A, Robinson D. qvalue: Q-value estimation for false discovery rate control. Version 2.36.0. 2024. https://bioconductor.org/packages/qvalue.
Suvorov A, et al. Widespread introgression across a phylogeny of 155 *Drosophila* genomes. Curr Biol. 2022:32:111–123.e5. https://doi.org/10.1016/j.cub.2021.10.052.
Svedberg J, et al. An introgressed gene causes meiotic drive in *Neurospora sitophila*. Proc Natl Acad Sci U S A. 2021:118:e2026605118. https://doi.org/10.1073/pnas.2026605118.
Svedin N, Wiley C, Veen T, Gustafsson L, Qvarnström A. Natural and sexual selection against hybrid flycatchers. Proc Biol Sci. 2008:275:735–744. https://doi.org/10.1098/rspb.2007.0967.
Teeter KC, et al. Genome-wide patterns of gene flow across a house mouse hybrid zone. Genome Res. 2008:18:67–76. https://doi.org/10.1101/gr.6757907.
Trier CN, Hermansen JS, Sætre G-P, Bailey RI. Evidence for mito-nuclear and sex-linked reproductive barriers between the hybrid Italian sparrow and its parent species. PLoS Genet. 2014:10:e1004075. https://doi.org/10.1371/journal.pgen.1004075.
Turissini DA, McGirr JA, Patel SS, David JR, Matute DR. The rate of evolution of postmating-prezygotic reproductive isolation in *Drosophila*. Mol Biol Evol. 2018:35:312–334. https://doi.org/10.1093/molbev/msx271.
Turissini DD, Matute DR. Fine scale mapping of genomic introgressions within the *Drosophila yakuba* clade. PLoS Genet. 2017:13:e1006971. https://doi.org/10.1371/journal.pgen.1006971.
Vaser R, Sović I, Nagarajan N, Šikić M. Fast and accurate de novo genome assembly from long uncorrected reads. Genome Res. 2017:27:737–746. https://doi.org/10.1101/gr.214270.116.
Vicario S, Moriyama EN, Powell JR. Codon usage in twelve species of *Drosophila*. BMC Ecol Evol. 2007:7:226. https://doi.org/10.1186/1471-2148-7-226.
Vigneault G, Zouros E. The genetics of asymmetrical male sterility in *Drosophila mojavensis* and *Drosophila arizonensis* hybrids: interactions between the *y*-chromosome and autosomes. Evolution. 1986:40:1160–1170. https://doi.org/10.1111/j.1558-5646.1986.tb05741.x.
Walker BJ, et al. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One. 2014:9:e112963. https://doi.org/10.1371/journal.pone.0112963.
Wick RR, Judd LM, Holt KE. Performance of neural network basecalling tools for Oxford Nanopore sequencing. Genome Biol. 2019:20:129. https://doi.org/10.1186/s13059-019-1727-y.
Wickham H. Ggplot2: elegant graphics for data analysis. Springer; 2016.
Yusuf LH, Tyukmaeva V, Hoikkala A, Ritchie MG. Divergence and introgression among the *virilis* group of *Drosophila*. Evol Lett. 2022:6:537–551. https://doi.org/10.1002/evl3.301.
Zhang C, Rabiee M, Sayyari E, Mirarab S. ASTRAL-III: polynomial time species tree reconstruction from partially resolved gene trees. BMC Bioinformatics. 2018:19:153. https://doi.org/10.1186/s12859-018-2129-y.
Zhang G, et al. Comparative genomics reveals insights into avian genome evolution and adaptation. Science. 2014:346:1311–1320. https://doi.org/10.1126/science.1251385.
Zhang L, Reifová R, Helenková Z, Gombert Z. How important are structural variants for speciation? Genes (Basel). 2021:12:1084. https://doi.org/10.3390/genes12071084.
Zuellig MP, Sweigart AL. A two-locus hybrid incompatibility is widespread, polymorphic, and active in natural populations of *Mimulus*. Evolution. 2018:72:2394–2405. https://doi.org/10.1111/evo.13596.

Associate editor: Carina Mugal