Rapid Divergence Of The Copulation Proteins In The Drosophila Dunni Group Is Associated With Hybrid Post-Mating-Prezygotic Incompatibilities

Tom Hill (✉️ tom.hill@ku.edu)  
Kansas University

Hazel-Lynn Rosales-Stephens  
Kansas City University of Medicine and Biosciences

Robert L. Unckless  
University of Kansas

Research article

Keywords: Drosophila, hybrid incompatibilities, genome evolution, seminal fluid proteins.

DOI: https://doi.org/10.21203/rs.3.rs-41472/v1

License: ☑️ This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

Background: Proteins involved in post-copulatory interactions between males and females are among the fastest evolving genes in many species and this has been attributed to reproductive conflict. Likely as a result, these proteins are frequently involved in cases of post-mating-prezygotic isolation between species. The *Drosophila dunni* subgroup consists of a dozen recently diverged species found across the Caribbean islands with varying levels of hybrid incompatibility.

Results: We performed experimental crosses between species in the *dunni* group and find evidence of hybrid inviability likely due to post-mating-prezygotic incompatibilities. We next assessed rates of evolution between these species genomes and find evidence of rapid evolution and divergence of some reproductive proteins, specifically the seminal fluid proteins.

Conclusions: This work suggests the rapid evolution of seminal fluid proteins can lead to post-mating-prezygotic isolation, which acts as a barrier for gene flow between even the most closely related species.

Background

Numerous groups of recently diverged species have been used to study speciation across multicellular taxa (1–5). These and other studies find an array of complex relationships between species caused by varying levels of divergence across genomes, incomplete isolation and differing forms of reinforcement (6–12). Recently diverged species with incomplete reproductive barriers prove to be more useful for understanding how new species can evolve (5, 13, 14). These species groups can be used in QTL studies to identify loci which contribute to the reduced fitness of hybrids (3, 15, 16), or to identify genes which may be involved in the early stages of speciation, such as those causing inviability or sterility in the heterogametic sex (a phenomenon known as Haldane’s Rule) (5, 7, 17–20).

Several studies have also highlighted that proteins transmitted in the seminal fluid to the female reproductive tract may also drive isolation as a post-mating-prezygotic mating barrier for incompletely separated species, either caused by, or resulting in, reinforcement (7, 14, 20–26). Barriers to hybridization have also been examined in different *Drosophila* species groups, finding varying levels of divergence, and in some cases the mechanisms for isolation between species (5, 8, 21, 22, 25, 27–30). Some studies, focusing on the effects of heterospecific matings on females, have found drastic changes in the females, including the swelling of the reproductive tract (28) and the activation of stress response pathways (31), likely due to antagonistic interactions between male seminal fluid proteins that the heterospecific female tract (32). These responses likely result in reinforcement of diverging reproductive behavior to prevent such matings (7, 22).

The *Drosophila dunni* subgroup is found within the *cardini* group in the *Drosophila* subgenus (Supplementary Fig. 1) (33). This species group diverged across the Caribbean islands thousands of years ago creating endemic populations, each on a different island or set of islands (33–35). Despite their extended isolation from each other, species are still able to hybridize (to varying levels of success)
and are a useful species group for understanding several traits, such as the evolution of pigmentation or reproductive isolation (34–37). In some cases, these hybrid offspring show evidence of Haldane’s rule (18, 19), with crosses producing only female offspring, or sterile male offspring (33).

Here we perform experimental crosses in the dunni group and find that in some crosses, heterospecific matings reduces female survival compared to conspecific matings, potentially caused by an insemination reaction-like effect (28). Using a combination of long-read and short-read sequencing, we assembled the genomes of four species in the dunni group to identify proteins driving this incompatibility. We find these genomes are of similar quality and composition as other higher quality genomes in the Drosophila subgenus (38–41). We also estimate rates of evolution across these genomes and identify several pathways of groups of genes of interest diverging between species (particularly between D. nigrodunni and D. arawakana), such as a divergence in immune pathways and in seminal fluid proteins.

**Results**

*The Drosophila dunni group shows varying levels of hybrid compatibility*

The Drosophila dunni group is a species group endemic to islands in the Caribbean, with each island inhabited by a different complement of species (35, 36, 42). These species have varying levels of hybrid incompatibilities, with some crosses producing viable offspring (e.g. D. dunni x D. similis) and others producing sterile offspring (e.g. D. arawakana x D. dunni) or no offspring (e.g. D. nigrodunni x D. similis). In keeping with Haldane’s rule (19), some produce sterile males, or no males at all (Figure 1, Supplementary Table 2, e.g. D. nigrodunni x D. arawakana). Despite divergence on levels comparable to the D. melanogaster subgroup (Supplementary Figure 1, Supplementary Table 3), there are no characterized inversions between species (36, 43), allowing differences across the species group to be investigated with a higher resolution than the D. melanogaster group allows.

Given the variety in levels of divergence and isolation between species, we examined the differences in this species group and identify patterns of divergence between species that could be associated with the reproductive isolation. Our focus is on the two hybrid crosses which produce some compatible offspring, such as with D. nigrodunni and D. arawakana, in which one direction of the heterospecific cross produces only female offspring (Figure 1).

*Drosophila arawakana males reduce the lifespan of D. nigrodunni females*

We next determined if there was evidence of further effects, beyond offspring viability (precopulatory, prezygotic and postzygotic) on hybridization. To do this, we established crosses between species, focusing on crosses that produced some fertile offspring (D. nigrodunni ♀ x D. arawakana♂, D. arawakana ♀ x D. nigrodunni♂, D. similis ♀ x D. dunnii♂, D. dunnii ♀ x D. similis♂, Supplementary Table 2). We also established matched crosses within species, and a matched control of virgin females. For each cross we recorded the survival of females following 5 days of mating.
In all cases, and consistent with studies in *D. melanogaster* (44, 45), virgins generally survive longer than mated females, though not significantly in some cases (Figure 2, Cox Hazard Ratio z-value = 3.868, \( p \)-value = 0.00011). The heterospecific crosses showed no difference from the conspecific crosses for *D. similis* and *D. dunni* (Figure 2, Cox Hazard Ratio z-value = -0.488, \( p \)-value = 0.62545), though *D. similis* heterospecifically mated females lived longer than conspecifically mated females (Figure 2, Cox Hazard Ratio z-value = 2.153, \( p \)-value = 0.03134). In contrast, when *D. nigrodunni* females are crossed to *D. arawakana* males, females have significantly decreased survival compared to conspecific crosses and virgin females (Figure 2, Cox Hazard Ratio z-value = -3.360, \( p \)-value = 0.00078), the same cross which also produced only female offspring (Figure 1).

As the *D. arawakana* strain examined was infected with *Wolbachia* and the *D. nigrodunni* was not, we cured all strains of bacteria using tetracycline-hydrochloride and repeated the survival assays. All females in this second block have reduced survival compared to the original survival assay, (Supplementary Figure 2, Cox Hazard Ratio z-value = -5.654, \( p \)-value = 1.56e-08), suggesting a difference in the two experiments that could be attributed to Tetracycline-Hydrochloride exposure. In the tetracycline exposed flies, we again find reduced survival in the *D. nigrodunni* x *D. arawakana* cross compared to the conspecific crosses (Supplementary Figure 3, Cox Hazard Ratio z-value = -3.815, \( p \)-value = 0.000136).

**The insemination reaction may be associated with the reduced female survival and reduced number of hybrid offspring**

In several other hybrid crosses between species in the *Drosophila* subgenus of *Drosophila*, other studies have highlighted a reaction between the seminal fluid of one species with the environment of the reproductive tract in the other species, called the insemination reaction (28, 29, 46). In the hours following mating, the reproductive tract swells, and, in some cases, proteins in the seminal fluid cause the formation of a “reaction mass”, a large dark mass which can burst through the wall of the tract (28).

Given the reduced survival of *D. nigrodunni* females following mating with *D. arawakana* males and the reduced number of hybrid offspring, we hypothesized that an incompatibility between the diverged seminal fluid proteins and the heterospecific reproductive tract could cause an abnormally deleterious reaction mass which reduces female survival.

We established experimental crosses within and between *D. arawakana* and *D. nigrodunni*. Then, 24 and 48 hours after crossing we dissected the females to identify whether sperm was present in the female reproductive tract (Figure 3A and B), and score for abnormal reproductive tracts consistent with the insemination reaction (Figure 3C and D). Interestingly, there was no significant differences between the number of mated females 24 and 48 hours after establishing crosses (Logistic regression: sperm presence ~ collection date: z-value = 1.285, \( p \)-value = 0.198873), but did score significantly fewer mated females in heterospecific crosses versus conspecific crosses (Logistic regression: sperm presence ~ cross type: z-value = -2.948, \( p \)-value = 0.00319). In several mated females when compared to virgin females, we find a swelling of the reproductive tract consistent with the insemination reaction (Figure 3C). Exclusively in several heterospecifically crossed females, we also saw damaged and destroyed
reproductive tracts (Figure 3D). We find a significant excess of swollen/damaged tracts in heterospecifically mated *D. nigrodunni* compared to conspecific controls (Figure 3E, Logistic regression: swollen tract ~ *D. nigrodunni* cross type: z-value = 4.723, p-value = 2.32e-06). While we do find swollen tracts in *D. arawakana* females we find no difference between heterospecific and conspecific females (Figure 3E, Logistic regression: swollen tract ~ *D. arawakana* cross type: z-value = 0.493, p-value = 0.622162).

**Genes involved in copulation and immune defense have high rates of divergence between species**

We reasoned that these incompatibilities between species could be caused by a divergence in copulation proteins. Previous work has suggested that females may be susceptible to bad reactions following hybrid matings due to no protection from the other species accessory gland proteins (32, 46). Specifically, that there is an arms race between sexes to block/unblock the female reproductive tract and that females of other species have not evolved to suppress these reactions. Based on this, we sought to examine the levels of divergence and identify rapidly evolving genes between species. We sequenced, assembled and annotated the genomes of each species involved (see Materials and Methods), producing two high quality genomes with high synteny to each other and to *D. innubila* (Supplementary Tables 1 & 4 and Supplementary Figure 4A), and two assemblies derived from these *de novo* assemblies. The two *de novo* assemblies had high BUSCO scores (*D. dunni* scored 93.9%: 2627 complete, 79 fragmented and 93 missing out of 2799 total; *D. nigrodunni* scored 97.3%: 2721 complete, 37 fragmented and 41 missing out of 2799 total). Consistent with previous findings we find no large structural rearrangements between genomes, and no evidence of fixed inversions between species in the *dunni* group (33, 43), though we do find several inversions between the next closest whole genome available, *D. innubila* on Muller elements B, C and D (*D. nigrodunni* shown in Supplementary Figure 4B). We annotated the *dunni* group genomes using a transcriptome from *D. innubila* in MAKER (47) and found between 10752 and 11581 genes in each species, most of which show orthology to previously identified genes in *D. virilis*, *D. melanogaster* or *D. innubila* (Supplementary Table 5) (40).

When examining the repetitive content of each species, we see an expansion of Helitrons and LTRs along the *D. dunni/D. similis* branch, resulting in higher TE content in these two species compared to *D. nigrodunni/D. arawakana* (Supplementary Figure 5). We also find species-specific expansions of satellites, particularly in *D. arawakana* and *D. nigrodunni*, where ~4% of the genome appears to be satellite sequences exclusive to that species (Supplementary Figure 5).

We identified orthologous genes across species using BLAST (48) with *D. innubila* as an outgroup when possible. For each group of orthologous genes, we identified the proportion of synonymous (dS) substitutions and amino acid changing, nonsynonymous substitutions (dN) (per possible synonymous or nonsynonymous substitution, respectively) occurring on each branch of the phylogeny using codeML (branch-based approach, model 0) (49). We also estimated these substitution rates across the entire *dunni* group phylogeny (sites-based approach, model 7 & 8) (49). This allowed us to calculate dN/dS to identify genes showing signatures of rapid or unconstrained evolution on any branch of the phylogeny,
or across the entire tree. For the dN/dS estimates on each branch, we identified genes in the upper 97.5\textsuperscript{th} percentile for dN/dS in windows of 0.01 dS. dN/dS in \textit{D. nigrodunni} is significantly correlated with dN/dS in \textit{D. arawakana} (Figure 2B), as well as in all other pairwise species comparisons (Supplementary Table 6, Pearson's correlation coefficient = 0.844, t = 7.3774, df = 7569, \textit{p}-value = 1.786e-13), and that similar proteins are rapidly evolving across the entire group. Copulation proteins (specifically seminal fluid proteins) are overrepresented among the most rapidly evolving genes on every branch of the \textit{dunni} group phylogeny (Supplementary Table 6, \textit{p}-value < 0.05 after multiple testing correction). This is consistent with rapid evolution occurring in genes involved in the reproductive conflict between the sexes (Figure 4) (50). While not significant outliers, we also find that immune recognition proteins, antiviral RNA and piRNA pathways are also rapidly evolving in some species, consistent with arms races between the species and their parasites (Supplementary Table 6).

Rapidly evolving genes may provide clues into the selective forces acting on species since their divergence. For the main species pairs of interest (e.g. \textit{D. nigrodunni} and \textit{D. arawakana}) we identified genes in the upper 97.5\textsuperscript{th} percentile for windows of dN/dS in the other species, to find genes rapidly evolving in one species but not the other (Figure 2B). As expected, copulation-associated proteins were in the upper 97.5\textsuperscript{th} percentile for both species, while genes in the Toll immune pathway are rapidly evolving in \textit{D arawakana} but not \textit{D. nigrodunni}, conversely the JAK-STAT immune pathway is rapidly evolving in \textit{D. nigrodunni} but not \textit{D. arawakana} (Supplementary Table 6, Figure 4B). These results suggest each species may differ in their primary pathogen, resulting in context dependent immune evolution, as seen elsewhere in the \textit{Drosophila} subgenus (40, 51).

We sought to confirm the rapid evolution of reproductive pathways and immune pathways after controlling for the background rate of evolution. We found the difference between dN/dS for each immune and reproductive gene and genes at neighboring loci on the chromosome (within 100kbp), of similar levels of divergence (+- 0.01 dS). We find significantly elevated rates of evolution of antiviral genes, insemination genes and seminal fluid proteins across the entire phylogeny (Figure 5, one-sided T-test \textit{mu} = 0, \textit{p}-value = 0.0434). We also find a significant correlation between differences in \textit{D. arawakana} and \textit{D. nigrodunni} for antiviral genes (Pearson's correlation = 0.795, t-value = 2.163, \textit{p}-value = 0.0288), immune recognition genes (Pearson's correlation = 0.877, t-value = 5.791, \textit{p}-value = 0.000175) and piRNA genes (Pearson's correlation = 0.659, t-value = 3.506, \textit{p}-value = 0.00292). The highest average rate of evolution occurred seminal fluid proteins on the \textit{D. nigrodunni} and \textit{D. arawakana} branches (Figure 5, one-sided T-test, \textit{mu} = 0, \textit{p}-value < 0.05). Consistent with previous results we find elevated rates of evolution of the Toll signaling pathway in \textit{D. arawakana}, and JAK-STAT in \textit{D. nigrodunni}. Interestingly, when comparing the specific genes rapidly evolving between \textit{D. nigrodunni} and \textit{D. arawakana}, the specific insemination and seminal fluid genes are mostly evolving at different rates between species (Figure 5), while the other rapidly evolving genes are consistent between species (Figure 4B). Consistent with this, we find no correlation between measures between \textit{D. arawakana} and \textit{D. nigrodunni} in copulation (Pearson's correlation = 0.187, t-value = 1.417, \textit{p}-value = 0.162), seminal fluid proteins (Pearson's correlation = 0.0341, t-value = 0.224, \textit{p}-value = 0.823), JAK-STAT (Pearson's correlation = 0.185, t-value =
0.625, p-value = 0.545) or Toll-signaling proteins (Pearson's correlation = 0.450, t-value = 1.334, p-value = 0.224). This could suggest a difference in importance of insemination proteins between the species and could even suggest a functional divergence (50).

Using orthology to D. innubila, we also identified duplications relative to these two species in each dunni group genome, and specific to each species. Consistent with the estimates in rates of evolution, we find enrichments of duplications in cell motility and copulation across the entire phylogeny (Supplementary Figure 6, Supplementary Table 7). We also find enrichments of duplications in Toll signaling genes in D. arawakana (p-value = 0.000569, enrichment = 5.44). Overall this suggests that the pathways showing elevated levels of nucleotide divergence (namely Toll and Copulation genes) also have more copy number variation between species than expected.

**Discussion**

*Drosophila* species have served as prominent models in genetics research, including in understanding the divergence between populations and the evolution of species. This is facilitated by the extensive genetic tools available in the species group to identify the genetic basis of reproductive isolation, both prezygotic and postzygotic. Many islands contain endemic species of *Drosophila* with differing levels of isolation. For example, the island endemics in the *Drosophila simulans* complex (27, 52, 53), with *D. mauritiana, D. simulans* and *D. sechellia* have served as a rich system for understanding reproductive isolation (52, 53). Like the *Drosophila simulans* complex, the *Drosophila dunni* species subgroup has radiated across a chain of islands (33), though with easier to define species relationships than is seen in the *simulans* subcomplex (52–54). Due to the recent radiation of this group, many species pairs in the dunni subgroup produce offspring (33, 36), some of which are fertile, and so provide a potentially useful model system for dissecting the genetics of reproductive isolation.

Here, we assessed the extent of hybrid incompatibilities between species of the dunni subgroup, focusing on post-mating-prezygotic incompatibilities. We then sequenced and assembled the species genomes to identify highly divergent and rapidly evolving genes. Between *D. nigrodunni* and *D. arawakana*, we find elevated divergence of several immune system pathways, as well as divergence in genes involved in copulation. This divergence fits with the hybrid male inviability between these two species, as well as the reduced survival of females following insemination by a heterospecific male. Consistent with the divergence in the seminal fluid proteins, we find evidence of an insemination reaction-like swelling of the reproductive tract (32), and a decrease in hybrid mating compared to within species.

Strangely, in this study most of the striking differences appear when comparing *D. nigrodunni* and *D. arawakana* (Figs. 1–5). This pair is slightly less diverged than other pairings within the group (Supplementary Fig. 1) and are allopatrically separated (33, 35), allowing for the neutral accumulation of substitutions with a reduced chance of introgression or reinforcement (5, 7). Due to this reduced divergence and reduced incidence of incompatibilities (17, 55), we may have caught this species pair at
the opportune time where these hybrid incompatible effects are visible, while other species pairs are too far diverged (Fig. 1).

The functional annotation of the more diverged genes may also provide us with clues as to how these species are diverging. As we find premating- behavior proteins are divergent between *D. arawakana* and *D. nigrodunni*, this may result in a divergence in premating behavior, resulting in the reduced rate of hybrid matings scored (Fig. 3). We also see no difference in the proportion of hybrid matings after 24 hours and 48 hours, suggesting that in these cases, if a female has rejected all males, she may not change her mind later (7, 14, 22). Hybridization between island-endemic flies separated by ~ 500 kilometers of ocean may be unlikely (56), but selection against hybridization between our focal species and other *dunni* group species may have led to the evolution of reinforcement against heterospecific mating (14, 22). We also find seminal fluid and copulation proteins are rapidly diverging between species (Fig. 2) and find an increased incidence of swollen and deformed reproductive tracts, consistent with an insemination reaction-like effect and a toxic incompatibility between the SFPs and their environment (Figs. 2 and 3) (32, 46). In fact, studies in other species have also identified post-mating-prezygotic incompatibilities to be a driver of isolation between species, even in cases with gene flow (20, 22, 24–26, 57). A recent study identified the upregulation of the JAK-STAT pathway (a stress response pathway) in *Drosophila* females following heterospecific mating, likely due to the negative effects of the accessory gland proteins (31). The rapid evolution of JAK-STAT proteins in *D. nigrodunni* could also be due to this species requiring a well-adapted stress response pathway, given its negative reaction to heterospecific matings (Figs. 1–3).

Several of the functional gene categories identified in this study as highly divergent between species are also promising regions for future study, particularly when focusing on immune evolution. Our findings are consistent with other studies that find immune proteins are more rapidly evolving than background genes (51, 58, 59), consistent with an arms-race between the host and its pathogens. However, in the species studied here, we find several cases of species-specific rapid evolution of an immune pathway, such as the rapid evolution of JAK-STAT in *D. nigrodunni* (Figs. 4 and 5). As immune pathways are constantly evolving in response to their pathogens, this could be explained by differences in immune pathogens in this species group (40, 58, 60). Hypothetically, the lack of any substantive natural Gram-Negative bacterial pathogens in *D. dunni* would result in a lack of divergence in the IMD pathway, the immune pathway associated with the resisting Gram-Negative bacteria. While a lack of fungal or Gram-Positive bacterial pathogens in *D. nigrodunni* could result in the lack of evolution of the Toll pathway, but rampant evolution in *D. arawakana* (Figs. 4 and 5).

The repetitive content also appears to be diverging rapidly across this species complex (Supplementary Fig. 5). This is commonly seen between species, given the elevated mutation rate/transposition of selfish factors compared to the rest of the genome (61–63), and has been implicated in the formation of hybrid incompatibilities for several species (64). Consistent with this we find several TE families unique to specific species in the *dunni* complex. However, we did not find a significant excess of dysgenic ovaries in hybrid females compared to normal females (Fisher’s exact text *p*-value > 0.05 for all cases). Several cases of hybrid incompatibilities caused by differences in TE content results in sterility caused by
maternally inherited factors over paternally inherited (as is usually seen). This may be due to the absence of maternally loaded silencing RNAs against specific TEs (65–67). If this were the case, we would expect the hybrid sterility to be in the opposite direction to what we observe, with sterile females (Fig. 1, Supplementary Fig. 5) (68), and so do not expect the hybrid incompatibilities seen here to be caused by repetitive content. However, this is a simplistic view of the effects of transposon activity on hybrid fertility, given the complex hybrid dysgenesis cases seen in *D. virilis* (69–71), and even the complex cases of tolerance to dysgenesis seen in the supposedly simple case in *D. melanogaster* (72), so may require further study to fully understand if TEs play a role in the divergence of the *dunni* complex.

**Conclusions**

Overall, our findings suggest that the rapid divergence of reproductive genes has led to incompatibilities between species in the *dunni* group, including inviable male offspring and the insemination reaction associated with reduced female survival. We also find multiple areas for further investigation in the *D. dunni* group, either in immune evolution of continuing to investigate the speciation in this species group, suggesting promise in the future of research for this group.

**Methods**

**Drosophila stocks, experimental crosses and survival assays**

We obtained stocks for *Drosophila arawakana* (stock number: 15182-2260.00), *D. dunni* (stock number: 15182-2291.00), *D. nigrodunni* (stock number: 15182-2311.00) and *D. similis* (stock number: 15182-2321.00) from the Cornell *Drosophila* species stock center. Each species was maintained on standard instant fly food (Formula 4-24, Carolina Biological Supply Company, Burlington, NC) in an incubator at 23°C. Before experiments, we inbred for three generations. Specifically, we established 10 single fly crosses for each species and chose a single successful cross per generation. We then repeated this for three generations. We then randomly chose one inbred vial to work with for the remainder of the experiments described.

**Experimental crosses within and between species**

We performed initial crosses in all pairwise combinations of species, for both directions of the cross, as well as within species crosses, to confirm previous assessments of between species viability (33, 35).

For each species we cleared vials of adults at 9:00AM central time and collected any emerged adults in 3-hour intervals following this, separating by sex. We then used these virgin flies to mate all species in pairwise combinations in 3 replicates. For each replicate we mated 10 males with 10 females (all aged 2-3 days) for 5 days (35, 42). We then collecting offspring every day for 30 days following the removal of the parents. After aging virgin F1 offspring for 3 days, we separated these into groups of 10 flies of the same parental species and mated with 10 flies of the opposite sex (5 the paternal species, 5 the maternal species) to assess the fertility of the F1 flies. As *D. arawakana* appeared to be infected with *Wolbachia*,
we sought to cure all species of any bacteria which may affect crosses. We created sublines of each species raised on food containing tetracycline-hydrochloride (0.05mg/ml) for three generations. Following this, we extracted DNA from females of each strain and tested for *Wolbachia* using PCR (wsp-81F (5′-TGGTCCAATAAGTGATGAAGAAAC-3′), wsp-691R (5′-AAAAATTAAACGCTACTCCA-3′), producing a ~600bp product from 10uL reactions, under the following cycling conditions: 94°C for 4 min, followed by 30 cycles of 40 s at 94°C, 40 s at 55°C, 1 min at 72°C and a final extension step of 10 min at 72°C) (73). We then repeated experimental crosses, as described above, with the tetracycline cured strains. We assayed female survival for *D. arawakana*, *D. dunni*, *D. nigrodunni* and *D. similis* in virgins and following mating, in both uncured and tetracycline cured flies. We considered a cross to be conspecific if we mated within species and a cross to be heterospecific if we mated with the most closely related species where fertile hybrids were found in previous crosses (e.g. *D. dunni* to *D. similis* and *D. arawakana* to *D. nigrodunni*). For these crosses we established 5-15 vials of 10 males and 10 females of the given species (with no males when measuring virgin females), all aged 2-3 days. We then recorded the survival of females every day (checking at 10AM Central time) for 30 days, flipping the flies onto new food every 3-4 days and removing males after the first 5 days. We then fit a survival curve across the total data for each cross type using SurvMiner (74) in R (75) and used a Cox’s Hazard Ratio to identify significant differences in survival between sets of crosses. For the initial crosses we used the following model:

\[
\text{Survival (days post mating)} \sim \text{Female species} \times \text{Male species (if any)} + \text{vial}
\]

We set the reference level as the conspecific cross (e.g. *D. arawakana* ♀ x *D. arawakana* ♂) and looked for significant differences from these for interaction terms to determine if unmated females (e.g. *D. arawakana* ♀ not mated) or heterospecifically crossed females (e.g. *D. arawakana* ♀ x *D. nigrodunni* ♂) show significant differences from the conspecific cross. To consider the effect of *Wolbachia* infection on these crosses, we repeated these initial crosses alongside the same crosses with *Wolbachia* cured flies (cured as described above) and a Cox’s Hazard Ratio was used to determine the effect of *Wolbachia* on survival, and to test for differences in survival between sets of crosses after accounting for *Wolbachia*:

\[
\text{Survival (days post mating)} \sim \text{Female species} \times \text{Male species (if any)} + \text{Wolbachia infection} + \text{vial}
\]

**Post-mating dissection of the female reproductive tract**

We collected virgin males and females for tetracycline-cured *D. arawakana* and *D. nigrodunni* as described above and aged them 2-3 days. We then established conspecific and heterospecific experimental crosses for 6 replicates of 10 males and 10 females at 10AM central time, as well as virgin control females for 6 replicates of 10 females. Following 24 hours of cohabitation, for 3 replicates of each cross, we separated the females for each cross and dissected the reproductive tract. Based on previous work describing the insemination reaction (28, 29, 46), we scored the reproductive tract for each
female, identifying if the female had mated (by the presence of sperm), if the reproductive tract appeared to be swollen (relative to the unmated virgin females) or if the reproductive tract was destroyed or damaged (alongside a swollen tract, if possible to tell). We repeated this scoring for the remaining 3 replicates of each cross 24 hours later (48 hours total). We then compared conspecific and heterospecific crosses for rates of mating and rates of insemination reaction occurrence.

Genome sequencing, assembly and annotation

We extracted DNA following the protocol described in (Chakraborty et al. 2017) for *D. arawakana*, *D. dunni*, *D. nigrodunni* and *D. similis* females. We prepared the *D. dunni* and *D. nigrodunni* DNA as a sequencing library using the Oxford Nanopore Technologies Rapid 48-hour (SQK-RAD002) protocol, which we then sequenced separately using a MinION (Oxford Nanopore Technologies, Oxford, UK) (76) (Supplementary Table 1). We also prepared the *D. arawakana*, *D. dunni*, *D. nigrodunni* and *D. similis* samples as Illumina libraries with a 300bp insert size which we sequenced on an Illumina HiSeq4000 to produce 150bp paired-end reads (Supplementary Table 1). We removed Illumina adapters using Sickle (77) and trimmed the Illumina sequences using Scythe (78). For the two MinION genomes, bases were called *post hoc* using the built in read_fast5_basecaller.exe program with options: –f FLO-MIN106 –k SQK-RAD002 –r–t 4. For *D. dunni*, raw reads were assembled using Minimap2 and Miniasm (parameters: -x ava -o nt -t 8) (79). We then polished using Racon with Oxford Nanopore Technology reads for three iterations and Pilon with Illumina fragment library reads for three iterations (80). For the *D. nigrodunni* genome, we first used wtdbg2 to assemble the genome (parameters: -t 4 -L 1000) (81). We then created a second assembly using Minimap2. For each, we ran Racon and Pilon for three iterations as described for *D. dunni*, then merged the two *D. nigrodunni* assemblies using Quickmerge (82). Following this, we polished this merged genome using Pilon for four more iterations. Both assemblies were benchmarked using BUSCO (v 3.0.2) and the *Diptera* database (83).

For *D. similis*, we mapped data to the *D. dunni* genome before Pilon polishing and polished the *D. dunni* genome using *D. similis* data in Pilon for three iterations, to insert *D. similis* variants into the genome. Following this we mapped *D. similis* data to this genome using BWA (84) and SAMtools (85), and called variants using Picard (86) and GATK Haplotypecaller (87, 88). We then used BCFtools (89) to filter these variants, removing calls below a quality threshold of 200 and inserted them into the polished genome. This was repeated for two more iterations to create a *D. similis* alternate genome. The same pipeline was followed for *D. arawakana* mapped to the *D. nigrodunni* genome.

We used the *D. innubila* transcriptome (40) as well as protein databases from *D. innubila*, *D. virilis*, *D. melanogaster*, and *M. domestica* in MAKER2 (47) to annotate each genome, including using RepeatModeler (90) in an attempt to correctly assign repetitive regions and retraining a HMM using SNAP following each iteration (91). This was repeated for three iterations to generate a GFF file containing gene evidence generated by MAKER2 (47).

Finally, we identified orthologous genomic regions pairwise for each of the four species examined here to each other and to the *D. innubila* genome using progressiveMauve (92). We visualized orthologous
regions using rCircos (93). We attempted to confirm any apparent structural differences based on progressiveMauve by mapping short reads for each species to a different genome and calling copy number differences using Delly (94) and dudeML (95), taking the consensus of the two tools, but favoring the absence of a copy number variant when we found discrepancies between the two tools.

**Assessing the repetitive content across the dunni group**

For each genome, we identified the repetitive content *de novo* using RepeatModeler to call the repeats (engine = NCBI) (90) and RepeatMasker (-gff -gcalc -s) to identify the repetitive regions (96). We also used dnaPipeTE (genome coverage = 1, sample number = 2, cpu = 4, genome size = 168000000) (97) to identify the repetitive content in the short-read data for each species, which we used to make a second map of reference genome repetitive regions using RepeatMasker. For both sets of repeat content assemblies we identified which TE families were shared between species and which were unique to species using blastn (e-value < 10e-5, hsps = 1, alignments = 1). We then identified what proportion of the genome each TE family constituted across species.

**Placing the dunni group in the Drosophila phylogeny**

To find the consensus species tree despite the differing evolutionary histories of different genes (98), we randomly sampled 100 genes conserved across *Drosophila* and humans from and extracted these from our four focal species, as well as from several other *Drosophila* species, taken from Flybase (38) and the NCBI genomes database (39-41, 99-101). We then aligned each gene group separately using MAFFT (--auto) (102) and created a multiple gene super-tree based on the consensus of each gene tree, following 100 bootstraps with PhyML (-b 100 -N 100 -GTR -gamma 8) (103, 104). We also generated gene trees for each of the 100 genes independently, following the same protocol. In this case 66 of the 100 trees gave the same topology of the *dunni* group as the total tree, while 7 trees had distinct topologies and 27 trees gave the topology of *D.similis* as an outgroup to the other three species, with *D. dunni* a sister to the *D. nigrodnunni-D.arawakana* complex.

**Estimating rates of evolution across the dunni group**

For each gene in the genomes of our four focal species, we identified orthology to each other and to genes in *D. innubila* using blastp (e-value < 0.00001, hsp = 1 alignment = 1) (48). For each set of orthologs, we aligned using PRANK to generate a codon alignment and gene-tree (105), as subtle differences between the species tree and gene trees can result in false estimates of divergence (98). We then estimated rates of both non-synonymous and synonymous substitutions using codeML (49), we estimated specific rates of evolution along each branch of the *dunni* group and leading into the *dunni* group using *D. innubila* as an outgroup (model 0) (49). Specifically, we estimated synonymous divergence (dS), non-synonymous divergence (dN) and the proportion of the two values (dN/dS). Finally, we also estimated rates of evolution across the entire *dunni* group phylogeny using codeML (models 7 & 8) (49), choosing the best fitting model using a likelihood ratio test (p-value < 0.05).
Using the estimated rates of evolution, we then compared the rates of evolution across the entire phylogeny and on specific branches to each species, for genes of similar levels of synonymous divergence (dS, windows of 0.001 dS, e.g. all genes within 0.001 dS of each other) we found the 97.5\textsuperscript{th} upper percentile for dN/dS. For the closely related species pairs (\textit{D. nigrodunni} and \textit{D. arawakana}, \textit{D. dunni} and \textit{D. similis}) we compared measures of dN/dS between species and found the 97.5\textsuperscript{th} upper percentile for dN/dS per species per window of dN/dS for the paired species (0.001, sliding 0.001).

We then took outlier genes (e.g. genes above the 97.5\textsuperscript{th} percentile in each category) and looked for enrichments in gene ontology categories compared to non-outlier genes using GOrilla (106). For GO categories of interest, such as those enriched for duplications or for high levels of dN/dS, we compared dN/dS of genes in these categories to the nearby genomic background. For each gene we extracted nearby genes (within 100kbp up or downstream on the same chromosome), of similar divergence levels on each branch (within 0.01 dS), we then found the difference in dN/dS between the median of the background genes and the focal gene. We then used a Wilcoxon-Rank Sum test to identify GO categories on each branch with significantly higher (or lower) dN/dS than the background.

Using the annotations of all species and \textit{D. innubila}, we identified genes with more than one copy in one species, relative to all other species. We confirmed this by estimating copy numbers of genes in each species using short read information and dudeML (following the tutorial pipeline for N = 1) with the short read information mapped to the genome of the sister species (95). We then used GOrilla (106) to identify Gene ontology categories that are enriched for duplicates on specific branches, which we confirmed using PANTHER (107).

\textit{Statistics}

We used R for all statistics in this analysis (75), and ggplot2 for data visualization and figure production (108).

\textit{Declarations}

\textit{Acknowledgements}

This work was completed thanks to feedback from Joanne Chapman and Mason Wilkinson. Mason Wilkinson and Jenna Lea assisted in the library preparation of these samples. Cornell stock center provided the \textit{Drosophila} species stocks used in this study. K-INBRE Bioinformatics core provided us with the Nanopore for use in sequencing.

\textit{Author Contributions}

TH was involved in experimental design, the genomic and bioinformatic analysis, experimental crosses and analysis of results, and writing the manuscript. HLRS was involved in experimental crosses between species and survival assays. RLU was involved in experimental design and writing the manuscript.
**Conflicts of Interest**

The authors declare no conflicts of interest

**Funding**

This work was supported by a K-INBRE postdoctoral grant to TH (NIH Grant P20 GM103418). This work was also funded by NIH Grant R00 GM114714 to RLU.

**Availability of data and materials**

All data used in the analyses in this study is available in the Supplementary Material. Sequencing information used in this study will be made available on the NCBI SRA upon acceptance for publication. Genomes generated for this study and annotations will be made available on the NCBI genomes server.

**Ethics approval and consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**References**

1. Glor RE, Losos JB, Larson A. Out of Cuba: overwater dispersal and speciation among lizards in the Anolis carolinensis subgroup. Mol Ecol. 2005;14(8):2419-32.

2. Brekke TD, Good JM. Parent-of-origin growth effects and the evolution of hybrid inviability in dwarf hamsters. Evolution. 2014;68(11):3134-48.

3. Kitano J, Ross JA, Mori S, Kume M, Jones FC, Chan YF, et al. A role for a neo-sex chromosome in stickleback speciation. Nature. 2009;461(7267):1079-83.

4. McKinnon JS, Rundle HD. Speciation in nature: the threespine stickleback model systems. Trends in Ecology & Evolution. 2002;17(10):480-8.

5. Coyne JA, Orr HA. Patterns of speciation in Drosophila. Evolution. 1989;43:362-81.

6. Orr HA. Dobzhansky, Bateson, and the genetics of speciation. Genetics. 2004;168:1097-104.

7. Coyne J, Orr HA. Speciation. 2004.

8. Presgraves DC. Speciation genetics: Epistasis, Conflict and the Origin of Species. Current biology : CB. 2007;17:R125-7.

9. Payseur BA, Rieseberg LH. A genomic perspective on hybridization and speciation. Mol Ecol. 2016;25(11):2337-60.

10. Orr HA, Orr LH, Ori HA. Waiting for speciation : The effect of population subdivision on the time to speciation. Evolution. 2013;50:1742-9.
11. Moyle LC, Nakazato T. Hybrid incompatibility "snowballs" between Solanum species. Science (New York, NY). 2010;329:1521-3.

12. Matute DR, Butler Ia, Turissini Da, Coyne JA. A test of the snowball theory for the rate of evolution of hybrid incompatibilities. Science (New York, NY). 2010;329:1518-21.

13. Presgraves DC. Speciation genetics: search for the missing snowball. Current biology : CB. 2010;20:R1073-4.

14. Gourbière S, Mallet J. Are species real? The shape of the species boundary with exponential failure, reinforcement, and the "missing snowball". Evolution; international journal of organic evolution. 2010;64:1-24.

15. Howard DJ, Marshall JL, Hampton DD, Britch SC, Draney ML, Chu J, et al. The genetics of reproductive isolation: a retrospective and prospective look with comments on ground crickets. The American naturalist. 2002;159 Suppl:S8-S21.

16. Noor MaF, Garfield Da, Schaeffer SW, Machado Ca. Divergence between the Drosophila pseudoobscura and D. persimilis genome sequences in relation to chromosomal inversions. Genetics. 2007;177:1417-28.

17. Orr HA. Population genetics of speciation: the evolution of hybrid incompatibilities. Genetics. 1995;139:1805-13.

18. Orr HA. Haldane's Rule. 2013;28:195-218.

19. Haldane J. Sex ratio and unisexual sterility in animal hybrids. J Genet. 1922;12(2):101-9.

20. Gavrilets S, Waxman D. Sympatric speciation by sexual conflict. PNAS. 2002;16(99):10533–8.

21. Matute DR, Comeault AA, Earley E, Serrato-Capuchina A, Peede D, Monroy-Eklund A, et al. Rapid and Predictable Evolution of Admixed Populations Between Two <em>Drosophila</em> Species Pairs. Genetics. 2020;214(1):211-30.

22. Turissini DA, McGirr JA, Patel SS, David JR, Matute DR. The Rate of Evolution of Postmating-Prezygotic Reproductive Isolation in Drosophila. Molecular Biology and Evolution. 2017;35(2):312-34.

23. Andres JA, Maroja LS, Harrison RG. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. Proc Biol Sci. 2008;275(1646):1975-83.

24. Larson EL, Andres JA, Harrison RG. Influence of the male ejaculate on post-mating prezygotic barriers in field crickets. PLoS One. 2012;7(10):e46202.

25. Ahmed-Braimah Y. Multiple Genes Cause Postmating Prezygotic Reproductive Isolation in the Drosophila virilis Group. G3: Genes|Genomes|Genetics. 2016;6:4067-76.

26. Larson EL, Andres JA, Bogdanowicz SM, Harrison RG. Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. Evolution. 2013;67(12):3653-61.

27. Matute DR, Ayroles JF. Hybridization occurs between Drosophila simulans and D. sechellia in the Seychelles archipelago. Journal of evolutionary biology. 2014;27:1057-68.
28. Patterson JT. The insemination reaction and its bearing on the problem of speciation in the mulleri subgroup. Univ Texas Publ. 1947;4720:41-77.

29. Grant B. On the Relationship Between Average Copulation Duration and Insemination Reaction in the Genus Drosophila. Evolution. 1983;37(4):854-6.

30. Miller WJ, Ehrman L, Schneider D. Infectious speciation revisited: Impact of symbiont-depletion on female fitness and mating behavior of Drosophila paulistorum. PLoS Pathogens. 2010;6:e1001214.

31. Ahmed-Braimah YH, Wolfner MF, Clark AG. Differences in post-mating transcriptional responses between conspecific and heterospecific matings in Drosophila. BioRxiv. 2020.

32. Knowles LL, Markow TA. Sexually antagonistic coevolution of a postmating-prezygotic reproductive character in desert Drosophila. PNAS. 2001;98(15).

33. IX. Genetic characteristics of Island Populations, (1962).

34. Hollocher H, Hatcher JL, Dyreson EG. Evolution of abdominal pigmentation differences across species in the Drosophila dunni subgroup. Evolution. 2000;54:2046-56.

35. Wilder JA, Hollocher H. Recent radiation of endemic Caribbean Drosophila of the dunni subgroup inferred from multilocus DNA sequence variation. Evolution. 2003;57:2566-79.

36. Stalker HD, Streisinger D. Taxonomy and Hybridization in the Cardini Group of Drosophila. 1953;46(3):343-58.

37. Studies in the genetics of Drosophila, (1954).

38. Gramates LS, Marygold SJ, Dos Santos G, Urbano JM, Antonazzo G, Matthews BB, et al. FlyBase at 25: Looking to the future. Nucleic Acids Research. 2017;45:D663-D71.

39. Zhou Q, Bachtrog D. Ancestral Chromatin Configuration Constrains Chromatin Evolution on Differentiating Sex Chromosomes in Drosophila. PLoS Genetics. 2015;11:1-21.

40. Hill T, Koseva B, Unckless RL. The genome of Drosophila innubila reveals lineage-specific patterns of selection in immune genes. Molecular Biology and Evolution. 2019;1-36.

41. Zhou Q, Zhu H-m, Huang Q-f, Zhao L, Zhang G-j, Roy SW, et al. Deciphering neo-sex and B chromosome evolution by the draft genome of Drosophila albomicans. BMC Genomics. 2012;13:109.

42. Cenzi De Ré F, Loreto ELS, Robe LJ. Gene and species trees reveal mitochondrial and nuclear discordance in the Drosophila cardini group (Diptera: Drosophilidae). Invertebrate Biology. 2010;129:353-67.

43. Cordeiro J, De Toni DC, de Souza da Silva G, da Silva Valente VL. Chromosomal evolution in the Drosophila cardini group (Diptera: Drosophilidae): photomaps and inversion analysis. Genetica. 2014;142:461-72.

44. Chapman T, Hutchings J, Partridge L. No reduction in the cost of mating for Drosophila melanogaster females mating with spermless males. Proceedings R Soc 1993;253:211-7.

45. Wigby S, Chapman T. Sex peptide causes mating costs in female Drosophila melanogaster. Curr Biol. 2005;15(4):316-21.
46. Markow TA, Ankney PF. Insemination Reaction in Drosophila: Found in Species Whose Males Contribute Material to Oocytes Before Fertilization. Evolution. 1988;42(5):1097-101.

47. Holt C, Yandell M. MAKER2 : an annotation pipeline and genome- database management tool for second- generation genome projects. BMC Bioinformatics. 2011;12:491.

48. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. Basic local alignment search tool. Journal of Molecular Biology. 1990;215(3):403-10.

49. Yang Z. PAML 4: Phylogenetic analysis by maximum likelihood. Molecular Biology and Evolution. 2007;24:1586-91.

50. Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ravi Ram K, Sirot LK, et al. Evolution in the fast lane: rapidly evolving sex-related genes in Drosophila. Genetics. 2007;177(3):1321-35.

51. Obbard DJ, Welch JJ, Kim KW, Jiggins FM. Quantifying adaptive evolution in the Drosophila immune system. PLoS Genetics. 2009;5:e1000698.

52. Cabot L, Davis AW, Johnson NA, Wu C-i. Genetics of reproductive isolation in the Drosophila simulans clade: complex epistasis underlying hybrid male sterility. Genetics. 1994;137:175-89.

53. Kliman RM, Andolfatto P, Coyne JA, Depaulis F, Kreitman M, Berry aJ, et al. The population genetics of the origin and divergence of the Drosophila simulans complex species. Genetics. 2000;156:1913-31.

54. Matute DR, Gavin-Smyth J, Liu G. Variable post-zygotic isolation in Drosophila melanogaster/ D. simulans hybrids. Journal of evolutionary biology. 2014;27:1691-705.

55. Welch JJ. Accumulating Dobzhansky-Muller incompatibilities: reconciling theory and data. Evolution; international journal of organic evolution. 2004;58:1145-56.

56. Coyne JA, Boussy IA, Prout T, Bryant SH, Jones JS, Moore JA. Long Distance Migration of Drosophila. The American Naturalist. 1982;11(4):589-95.

57. Gavrilets S. Perspective: models of speciation: what have we learned in 40 years? Evolution; international journal of organic evolution. 2003;57:2197-215.

58. Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. Dynamic evolution of the innate immune system in Drosophila. Nature Genetics. 2007;39(12):1461-8.

59. Shultz A, Sackton TB. Immune genes are hotspots of shared positive selection across birds and mammals. eLife. 2019;8:e41815.

60. Unckless RL, Howick VM, Lazzaro BP. Convergent Balancing Selection on an Antimicrobial Peptide in Drosophila. Current Biology. 2016;26(2):257-62.

61. Adrion JR, Begun DJ, Hahn MW. Patterns of transposable element variation and clinality in Drosophila. Mol Ecol. 2019;28(6):1523-36.

62. Kofler R, Betancourt AJ, Schlötterer C. Sequencing of pooled DNA Samples ( Pool-Seq ) uncovers complex dynamics of transposable element insertions in Drosophila melanogaster. PloS Genetics. 2012;8:1-16.
63. Köfter R, Nolte V, Schlötterer C. Tempo and mode of transposable element activity in *Drosophila*. PLoS Genet. 2015;11:e1005406.

64. Satyaki PR, Cuykendall TN, Wei KH, Brideau NJ, Kwak H, Aruna S, et al. The Hmr and Lhr hybrid incompatibility genes suppress a broad range of heterochromatic repeats. PLoS Genet. 2014;10(3):e1004240.

65. Bingham PM, Kidwell MG, Rubin GM. The molecular basis of P-M hybrid dysgenesis: the role of the P-element, a P-strain-specific transposon family. Cell. 1982;29:995-1004.

66. Aravin AA, Hannon GJ, Brennecke J. The piwi-piRNA pathway provides an adaptive defense in the transposon arms race. Science. 2007;318:761-4.

67. Brennecke J, Malone CD, Aravin AA, Sachidanandam R, Stark A, Hannon GJ. An epigenetic role for maternally inherited piRNAs in transposon silencing. Science. 2008;322:1387-92.

68. Kidwell JF, Kidwell MG, Sved JA. Hybrid dysgenesis in *Drosophila melanogaster*. A syndrome of aberrant traits including mutation, sterility, and male recombination. Genetics. 1977;86:813-33.

69. Erwin AA, Galdos MA, Wickersheim ML, Harrison CC, Marr KD, Colicchio JM, et al. piRNAs Are Associated with Diverse Transgenerational Effects on Gene and Transposon Expression in a Hybrid Dysgenic Syndrome of *D. virilis*. PLoS Genetics. 2015;11:1-33.

70. Petrov Da, Schutzman JL, Hartl DL, Lozovskaya ER. Diverse transposable elements are mobilized in hybrid dysgenesis in *Drosophila virilis*. Proceedings of the National Academy of Sciences of the United States of America. 1995;92:8050-4.

71. Kelleher ES, Jaweria J, Akoma U, Ortega L, Tang W. QTL mapping of natural variation reveals that the developmental regulator bruno reduces tolerance to P-element transposition in the *Drosophila* female germline. PLoS Biol. 2018;16(10):e2006040.

72. Zhou W, Rousset F, O'Neill S. Phylogeny and PCR-based classification of *Wolbachia* strains using wsp gene sequences. Proc R Soc. 1998;265(265):509-15.

73. Kassambara A, Kosinski M, Biecek P. survminer: Drawing Survival Curves using‘ggplot2’. R package version 03. 2017;1.

74. R-Core-Team. R: A Language and Environment for Statistical Computing. Vienna, Austria: R Foundation for Statistical Computing; 2013.

75. Jain M, Olsen HE, Paten B, Akeson M. The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community. Genome Biology. 2016;17(239):1-11.

76. Joshi N, Fass J. Sickle: A sliding window, adaptive, quality-based trimming tool for fastQ files. 2011;1.33.

77. Buffalo V. Scythe2018.
79. Li H. Minimap and miniasm: fast mapping and de novo assembly for noisy long sequences. Bioinformatics. 2016;32(14):2103-10.

80. Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, et al. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS ONE. 2014;9.

81. Ruan J, Li H. Fast and accurate long-read assembly with wtdbg2. Nat Methods. 2020;17(2):155-8.

82. Liu Y, Yang Y, editors. Quick-merge sort algorithm based on multi-core linux. Proceedings 2013 International Conference on Mechatronic Sciences, Electric Engineering and Computer (MEC); 2013: IEEE.

83. Simão FA, Waterhouse RM, Ioannidis P, Kriventseva EV, Zdobnov EM. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics. 2015;31:3210-2.

84. Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics (Oxford, England). 2009;25:1754-60.

85. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, et al. The sequence alignment/map format and SAMtools. Bioinformatics (Oxford, England). 2009;25:2078-9.

86. Broad-Institute. Picard. http://broadinstitute.github.io/picard2017.

87. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky AM, et al. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Proceedings of the International Conference on Intellectual Capital, Knowledge Management & Organizational Learning. 2010;20:1297-303.

88. DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, et al. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics. 2011;43:491-8.

89. Narasimhan V, Danecek P, Scally A, Xue Y, Tyler-Smith C, Durbin R. BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics. 2016;32:1749-51.

90. Smit AFA, Hubley R. RepeatModeler Open-1.0. 2008.

91. Johnson AD, Handsaker RE, Pulit SL, Nizzari MM, O'Donnell CJ, de Bakker PI. SNAP: a web-based tool for identification and annotation of proxy SNPs using HapMap. Bioinformatics. 2008;24(24):2938-9.

92. Darling ACE, Mau B, Blattner FR, Perna NT. Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements Mauve : Multiple Alignment of Conserved Genomic Sequence With Rearrangements. Genome Res. 2004;14(7):1394-403.

93. Zhang H, Meltzer P, Davis S. RCircos : an R package for Circos 2D track plots. 2013.

94. Rausch T, Zichner T, Schlattl A, Stutz AM, Benes V, Korbel JO. DELLY: structural variant discovery by integrated paired-end and split-read analysis. Bioinformatics. 2012;28(18):i333-i9.

95. Hill T, Unckless RL. A Deep Learning Approach for Detecting Copy Number Variation in Next-Generation Sequencing Data. G3. 2019;9(11).
96. Smit AFA, Hubley R. RepeatMasker Open-4.0. RepeatMasker; 2013-2015.

97. Goubert C, Modolo L, Vieira C, Moro CV, Mavingui P, Boulesteix M. De novo assembly and annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes aegypti*). Genome Biology and Evolution. 2015;7:1192-205.

98. Mendes F, Hahn MW. Gene tree discordance causes apparent substitution rate variation. Systematic Biology. 2016;65:711-21.

99. Kitts PA, Church DM, Thibaud-Nissen F, Choi J, Hem V, Sapojnikov V, et al. Assembly: a resource for assembled genomes at NCBI. Nucleic Acids Res. 2016;44(D1):D73-80.

100. Hamilton PT, Leong JS, Koop BF, Perlman SJ. Transcriptional responses in a *Drosophila* defensive symbiosis. Molecular Ecology. 2014;23:1558-70.

101. Palmieri N, Kosiol C, Schlötterer C. The life cycle of *Drosophila* orphan genes. eLife. 2014;3:1-21.

102. Katoh K, Misawa K, Kuma K-i, Miyata T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic acids research. 2002;30:3059-66.

103. Le SQ, Gascuel O. An improved general amino acid replacement matrix. Molecular biology and evolution. 2008;25:1307-20.

104. Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Systematic biology. 2010;59:307-21.

105. Löytynoja A. Phylogeny-aware alignment with PRANK. In: Russell DJ, editor. Multiple Sequence Alignment Methods. Totowa, NJ: Humana Press; 2014. p. 155-70.

106. Eden E, Navon R, Steinfeld I, Lipson D, Yakhini Z. GOrilla: A tool for discovery and visualization of enriched GO terms in ranked gene lists. BMC Bioinformatics. 2009;10:1-7.

107. Thomas PD, Campbell MJ, Kejariwal A, Mi H, Karlak B, Daverman R, et al. PANTHER: a library of protein families and subfamilies indexed by function. Genome Res. 2003;13(9):2129-41.

108. Wickham H. ggplot2: Elegant Graphics for Data Analysis. 2009.

**Figures**
Figure 1

Points show the mean number of offspring produced by three replicates of 10 females of each species when crossed to males of different species. Points of the same color represent conspecific crosses while dots with a different colored center represent a cross between two different species. Point shape shows the state of fertility of F1 offspring, either both fertile, both sterile or only females fertile. Error bars show the standard deviation of offspring count and sex ratio across replicates. D. ara = D. arawakana, D. dun = D. dunni, D. nig = D. nigrodunni, D. sim = D. similis. While we performed all pairwise heterospecific crosses, only crosses which produced offspring are shown on the plot.
Figure 2

Survival of females postmating. Survival probability of females for each species used in each cross, compared to virgin female survival. Boxes show the survival of females separated by their species, and colored based on the females being virgins (red), conspecific crossed (yellow, crossed to a male of their own species), heterospecific crossed (blue, crossed to a male of a different species). In the case of heterospecific crosses, D. arawakana is only crossed to D. nigrodenuni and D. dunni is only crossed to D. similis.
Abnormal insemination reactions may be responsible for reproductive isolation. A-C. Dissections showing differing conditions of the female reproductive tract. When applicable, arrows label the start and end of same section of the oviduct between dissections. Ovipositors and scale bar also shown for scale. A. Normal oviduct containing sperm. B. Normal oviduct with no sperm. C. Swollen oviduct containing sperm. D. Ruptured oviduct in sample with reaction mass-like phenotype. E. Plots summarizing rate of mating,
and the effect of mating on the reproductive tract in crosses within and between D. arawakana and D. nigrodnunni. Plots are separated by the male involved in the cross (columns) and the female involved in the cross (rows), with plots scoring the number of females with sperm in the reproductive tract, and if the tract was normal or swollen/damaged.

![Figure 4](image)

**Figure 4**

Rates of evolution across the Drosophila dunni phylogeny, showing non-synonymous divergence versus synonymous divergence across A. the whole phylogeny and B. comparing the proportion of non-synonymous to synonymous divergence between D. nigrodnunni and D. arawakana. JAK-STAT, Toll and seminal fluid proteins are highlighted due to their enrichments in one or the other species.
Figure 5

Difference of dN/dS between focal genes in specific functional categories and their nearby background genes. We find different insemination proteins and seminal fluid proteins are rapidly evolving between D. nigrodunni and D. arawakana. A selection of genes in each category are labelled by name in each plot. Plots are labelled with a * if we find a positive correlation between the two axes (p-value < 0.05).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- SupplementaryInformation.docx
- supplementarytables.xlsx
- Supp.Fig1.JPG
- Supp.Fig2.JPG
- Supp.Fig3.JPG
- Supp.Fig4.JPG
- Supp.Fig5.JPG
• Supp.Fig6.JPG