Antiviral genes are not rapidly evolving in *Drosophila innubila*

Tom Hill[1]*, Boryana S. Koseva[2], Robert L. Unckless[1]

1. 4055 Haworth Hall, The Department of Molecular Biosciences, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045. Email: tom.hill@ku.edu

2. 3005 Haworth Hall, K-INBRE Bioinformatics Core, University of Kansas, 1200 Sunnyside Avenue, Lawrence, KS 66045

* Corresponding author

Abstract

Viruses make up a considerable proportion of the pathogens infecting animals. They can spread rapidly between hosts, and sicken or even kill their hosts to promote their own proliferation. Due to this strong selective pressure, antiviral immune genes are some of the fastest evolving genes across metazoans, as highlighted in mammals and insects. While *Drosophila melanogaster* are frequently exposed to pathogenic RNA viruses, little is known about *D. melanogaster*'s ecology in terms of viral exposure, or if they are representative of other *Drosophila* species. Here, we sequence and assemble the genome of a highly diverged, mushroom-feeding *Drosophila* species, *Drosophila innubila*, a species frequently exposed to a highly pathogenic DNA virus. We investigate the evolution of the immune system and find little evidence for rapid evolution of the antiviral RNAi genes, though we do find rapid evolution of several other pathways, suggesting alternate means of viral resistance. This contrasts with *D. melanogaster*, and suggests that evolution of resistance to DNA viruses differs greatly from that of RNA viruses.

Introduction

Viruses are a substantial pathogenic burden to nearly every species on the planet, providing a strong selective pressure for individuals experiencing infection to evolve resistance. There is considerable evidence for selection acting on genes involved in resistance to this pathogenic burden, as highlighted by several studies across Metazoans (Kimbrell and Beutler 2001; Enard et al. 2016) including the fruit fly model, *Drosophila melanogaster* (Sackton et al. 2007; Obbard et al. 2009). Most work concerning the evolution of the invertebrate antiviral immune system has focused on the *D. melanogaster*, where the antiviral immune response has been characterized mostly after infection with RNA viruses (Merkling and van Rij 2013; Tang and Wu 2013). In *D. melanogaster* and several members of the *Sophophora* subgenus, the canonical antiviral genes are some of the fastest evolving genes in the genome (Obbard et al. 2006; Darren J. Obbard et al. 2009; Palmer et al. 2018a). This suggests that viruses are a major selective pressure, requiring a rapid evolutionary response from the host (Daugherty and Malik 2012). However, recent work suggests that the canonical *D. melanogaster* antiviral immune pathway may not encompass the entire antiviral immune response – particularly concerning response to DNA virus infection (Palmer et al. 2018b; West and Silverman 2018). Therefore, a broader view of antiviral immune gene evolution across *Drosophila* is warranted.

One particular group in the *Drosophila* genus with a rich history of ecological study, and with great potential as a host/virus study system, is the *Drosophila quinaria* group (Jaenike and Perlman 2002; Perlman et al. 2003; Dyer et al. 2005; Jaenike and Dyer 2008; Unckless 2011; Unckless and Jaenike 2011). This species group is mycophagous, found developing and living on the fruiting bodies of several (sometimes toxic) mushrooms. These mushrooms are commonly inhabited by parasitic nematode worms, trypanosomes and a host of parasitic microbes (Dyer et al. 2005; Martinson et al. 2017). While most species in the quinaria group are broadly dispersed across temperate forests, *Drosophila innubila* is limited to the "Sky Islands", forest and woodlands on mountain tops in southwestern USA and Mexico (Patterson and Stone 1949). The flies are restricted to elevations of 900 to 1500m, and are active only during the monsoon season (late July to September) (Jaenike et al. 2003). *Drosophila innubila* is also the only species in the quinaria group frequently (25-46% in females) infected by a male-killing *Wolbachia* strain, leading to female biased sex-ratios (Dyer 2004). This *Wolbachia* is closely related to *wMel*, which infects *D. melanogaster* (but does not kill males) (Dyer and Jaenike 2005; Jaenike and Dyer 2008). Finally, and most importantly, *D. innubila* is also frequently (35-56%, n > 84) infected by Drosophila innubila Nudivirus (DiNV) (Unckless 2011; Hill and Unckless 2017). In contrast, the virus is found at lower frequencies in *D. innubila's* sister species, *D. falleni* (0-3%, n = 95) and undetected in the outgroup species, *D. phalerata* (0%, n = 7) (Unckless 2011). When infected with a similar DNA virus, *Drosophila melanogaster* show a standard antiviral immune response, including the induction of the antiviral siRNA pathway (Palmer et al. 2018a; Palmer et al. 2018b). Thus, despite lacking the genomic resources of *Drosophila melanogaster*, *Drosophila innubila* and *falleni* are potentially useful model systems to understand the evolution of the antiviral immune system, to identify if the signatures of selection are conserved in the canonical antiviral pathways between the highly-diverged *D. melanogaster* and quinaria groups. We predict that, like *D. melanogaster* (Sackton et al. 2007; Obbard et al. 2009), the immune system (especially antiviral pathways) will be rapidly evolving in this species group, particularly in *D. innubila*.

To this end, we survey the evolutionary divergence of genes within a trio of closely related mycophagous *Drosophila* species (*D. innubila, D. falleni* and *D. phalerata*). As a first step, we sequenced and assembled the genome for *D. innubila* resulting in an assembly which is on par with the *Drosophila* benchmark, the *D. melanogaster* release 6. Using short read

alignments of two closely-related species (*D. falleni* and *D. phalerata*) to *D. innubila*, we find evidence of selective constraint on the antiviral RNAi pathway, but rapid adaptation of several other immune pathways, including those involved in the Toll and JAK-STAT pathways.

Results

Genome sequencing and assembly

Drosophila innubila is a mushroom-feeding species found across the sky islands of the South-Western USA and Western Mexico (Patterson and Stone 1949). It is in the quinaria group of the *Drosophila* subgenus, approximately 50 million years diverged from the research workhorse, *D. melanogaster* (Dyer 2004; Markow and O'Grady 2006), has a sister species in northern North America, *D. falleni*, and the pair share an old-world outgroup species, *D. phalerata* (Figure 1B). These species are highly diverged from all previously sequenced *Drosophila* genomes, and represent a grossly genomically understudied group of the *Drosophila* subgenus (Jaenike et al. 2003; Markow and O'Grady 2006).

We sequenced and assembled the genome of *D. innubila* (Figure 1A, Table 1; Submitted to NCBI upon acceptance of manuscript). The genome is 168 Mbp with an N50 of 29.59Mbp, eclipsing release 6 of *D. melanogaster*, which is 25.29Mb (Chakraborty et al. 2017). Using a combined transcriptome assembly of all life stages and protein databases from other species, we annotated the genome and found 12,318 genes (including the mitochondrial genome), with coding sequence making up 11.5% of the genome, at varying gene densities across the Muller elements (Figure 1A blue, Supplementary Table 1-3). Of the annotated genes, 11,925 are shared with other *Drosophila* species (among the 12 genomes available on Flybase.org), 7,094 have orthologs in the human genome, and the annotation recovered 97.2% of the Dipteran BUSCO protein library (Simão et al. 2015). The *D. innubila* genome averages 36.57% GC, varying across the Muller elements (Figure 1A black/white). Using dnaPipeTE (Goubert et al. 2015), RepeatModeler (Smit and Hubley 2008) and RepeatMasker (Smit and Hubley 2015), we estimate 13.53% of the genome consists of transposable elements (TEs, Figure 1A red). Using Mauve (Darling et al. 2004) we next compared our assembly to the best assembled genome within the *Drosophila* subgenus, *D. virilis* and find most the *D. innubila* genome also has some orthology to the *D. virilis* reference genome (Figure 1A green). A summary of the assembled genome including codon bias, TE content, orphan genes, duplications and expression changes

across life stages can be found in the supplementary results (Supplementary Tables 1-18 and Supplementary Figures 2-10).

Evolution of the immune system and RNAi pathways in D. innubila

Given the wealth of knowledge about host/parasite dynamics in *D. innubila* and its relatives, we were motivated to study the evolution of immune-related genes in the group. Using short reads from *D. falleni* and *D. phalerata* mapped to the *D. innubila* genome, we called species-specific SNPs and generated consensus gene sequences for each species. We then aligned each species version of each gene to the *D. innubila* ortholog (PRANK –codon +F) (Löytynoja 2014). For each ortholog set, we identified the proportion of synonymous (*dS*) substitutions and amino acid changing, non-synonymous substitutions (*dN*) (per possible synonymous or non-synonymous substitution respectively) occurring on each branch of the phylogeny (codeML branch based approach, model 0) (Yang 2007; McKenna et al. 2010; DePristo et al. 2011; Löytynoja 2014). Using these measures, we calculated *dN/dS* to identify genes showing signatures of rapid selection (*dN/dS* > 1, with more potentially functional substitutions than silent substitutions). We found *dN/dS* is highly correlated between *D. melanogaster* and *D. innubila*, additionally (Spearman rank correlation = 0.6856, GLM t = 1.232, p-value = 1.52e-06), most genes are under selective constraint; only 25.6% of genes have *dN/dS* greater than 0.5 (Figure 2A). Though we find no correlation between *dN/dS* and gene length (GLM t = 0.34, p-value = 0.81), we do find a significant negative correlation between *dS* and *dN/dS* (Figure 1A, GLM t = -64.94, p-value < 2e-16). Most genes with high *dN/dS* have lower values of *dS*, possibly due to the short gene branches (and low neutral divergence) between species inflating the proportion of non-synonymous substitutions (Figure 2A). Because of the effect of *dS* on *dN/dS*, we extracted the 166 genes that were in the upper 97.5% *dN/dS* of genes per 0.01 *dS* window and with *dN/dS* greater than 1 and labeled these genes as evolving most rapidly. Contrasting *D. melanogaster* (Obbard et al. 2006), these genes are not enriched for antiviral genes and are instead significantly enriched for the regulation of the Toll pathway (Figure 2A & B, Supplementary Table 5, GOrilla FDR p-value = 0.00015, enrichment = 14.16) (Eden et al. 2009), the pathway responsible for the induction of immune response to gram-positive bacteria and fungi (Janssens and Beyaert 2003; Takeda and Akira 2005). Specifically, we find *spatzle4*, *necrotic*, *spheroide* and *modSP* are the fastest evolving genes in this pathway, and among the fastest in the genome (Figure 2A). Most of

these genes are signaling genes, which were, as a class, not particularly fast evolving in the *D. melanogaster* group (Sackton et al. 2007). Additionally, *Pp1apha-96A* and *Attacin D*, genes involved in Gram-Negative bacterial response, are rapidly evolving (upper 97.5% of genes, *dN/dS* > 1).

Since immune gene evolution is so well characterized in the *D. melanogaster* group, we compared the molecular evolution in genes in each of these immune gene categories for the *melanogaster* trio (*D. melanogaster*, *D. simulans* and *D. yakuba*) with the evolution of those same genes in the *innubila* trio. These trios are of similar levels of divergence (Figure 1B) and *dN*/*dS* is significantly positively correlated between the trios (Spearman rank correlation = 0.6856, *p*-value = 1.52e-06). Contrasting the *melanogaster* trio, only one antiviral RNAi gene appears to be fast-evolving in the *innubila* trio, with most genes in this category close to the median *dN/dS* for the *innubila* genome (Figure 2A & B). Interestingly, *pastrel*, the fast-evolving antiviral gene in *D. innubila* trio is among the slowest evolving antiviral gene in the *D. melanogaster* trio. Also, variation in *pastrel* is associated with survival after *Drosophila* C virus infection (DCV), but is not likely involved with antiviral RNAi (Magwire et al. 2011; Magwire et al. 2012; Barbier 2013).

As *dN/dS* may give false signals of rapid evolution due to multiple nucleotide substitutions (Venkat et al. 2018), we calculated δ (another measurement of the rate of evolution) using a method that can control for multinucleotide substitutions (Pond et al. 2005; Venkat et al. 2018). δ was positively correlated with *dN/dS* in both species (Spearman correlation = 0.17, p = 0.0192). Using this method, we again find that antiviral genes are rapidly evolving exclusively in *D. melanogaster* compared to the background (Supplementary Figure 1, GLM t = 4.398, *p*-value = 1.1e-05), while bacterial response genes are rapidly evolving only in *D. innubila* compared to the background (Supplementary Figure 1, GLM t > 2.682, *p*-value < 0.00731).

Because non-synonymous divergence is elevated in genes with low synonymous divergence in the *D. innubila* trio but not the *D. melanogaster* trio (Figure 2A & B), we attempted to control for its effect. For each focal immune gene, we extracted genes on the same chromosome with *dS* within 0.01 of the focal gene. We then found the differences in the median *dN/dS* of these control genes and the focal genes *dN/dS*, for each branch on the tree, and categorized these differences by immune category based on Flybase gene ontologies. Using this method, we found most categories were slightly positive, suggesting faster evolution than the

background in the immune system (Figure 2B), similar to the *melanogaster* trio (Sackton et al. 2007). Specifically, the Toll signaling, Gram-positive and Gram-negative categories are significantly higher than the background in the *D. innubila* trio (Figure 2C, Supplementary Figure 4, *t*-test $t = 2.39$, *p*-value < 0.05). Again, in contrast to the *D. melanogaster* trio (Obbard et al. 2006), there is no significant elevation of adaptation in antiviral genes in the *innubila* trio.

On some branches of the *D. innubila* trio phylogeny, we find differing signatures from the total tree. Specifically, while the gram-positive bacterial signaling is fast evolving in *D. falleni*, antifungal and gram-negative bacterial responses are fast evolving in *D. innubila* (Figure 2C, Supplementary Table 4, *t*-test $t = 2.11$, *p*-value < 0.05), potentially highlighting differences in the pathogens and environments encountered by the two species.

Low antiviral expression in Drosophila innubila

We also examined immunity and RNAi genes in the context of baseline (e.g. constitutive) gene expression in *D. innubila* compared to *D. melanogaster* gene expression data from ModEncode (Consortium et al. 2011). We note this is not a well-controlled comparison and that differences in expression could be due to different laboratory conditions or other experimental variables as opposed to true baseline expression differences. The antiviral pathway is mostly shut off in all life stages of *D. innubila* (Figure 2C, only 3 of 7 genes show expression greater than 1 read per million counts per kbp of gene), and significantly lower in expression than the rest of the genome across all life stages (both before and after adjusting for gene length, Wilcoxon rank sum *p*-value < 0.02). Comparatively, the piRNA pathway show appreciable levels of expression and no difference from the background genome at any stage (Figure 2C, Wilcoxon rank sum *p*-value > 0.05). This difference in expression between the antiviral RNAi and piRNA pathways may be due to antiviral genes only being expressed upon infection, though AMPs, which are also induced upon infection, here show high levels of constitutive expression compared to the background (Wilcoxon Rank Sum test: *p*-value < 0.05). Additionally, the antiviral genes are significantly more highly expressed compared to background genes in the *D. melanogaster* expression data, even without a known infection, and are further induced upon infection with a DNA virus (Wilcoxon Rank Sum test: $W = 80672$, *p*-value < 0.005) (Obbard et al. 2006; Obbard et al. 2009; Ding 2010; Palmer et al. 2018a; Palmer et al. 2018c). Antiviral RNAi genes are significantly more highly expressed in *D. melanogaster* than in *D. innubila* (Wilcoxon Rank

Sum test: W = 2, *p*-value = 0.0003), while immune signaling and immune recognition genes are more highly expressed in *D. innubila* compared to *D. melanogaster* (Wilcoxon Rank Sum test: W = 178 *p*-value = 0.0002).

Alternative antiviral pathways are rapidly evolving in the Drosophila innubila trio

Recent work has highlighted that the canonical antiviral genes are not the only pathways induced upon DNA virus infection, in *D. melanogaster* (Palmer et al. 2018b). Additionally, the JAK-STAT pathway likely plays a more prominent role in DNA virus survival, including the induction of Turandot genes, a set of poorly described, rapidly evolving, secreted proteins involved in stress-response (Ekengren and Hultmark 2001; Merkling and van Rij 2013; West and Silverman 2018). Finally, there are also several genes annotated as associated with viral capsids found in the genomes of both *D. melanogaster* and the quinaria group. Using these three lines of evidence, we decided to examine the evolution of the JAK-STAT cascade and the viral capsid associated genes (but not Turandots as we were unable to identify any during genome annotation). The JAK-STAT genes are fast evolving across the *D. innubila trio*, and several genes show consistent signatures in *D. innubila* and *D. melanogaster* (Figure 3, Supplementary Table 4). Overlapping rapidly evolving genes included the JAK-STAT cytokines *upd2* and *upd3*, and JAK-STAT regulatory genes *CG30423* and *asrij*. We repeated this analysis for genes associated with host-viral capsid interaction and found these genes to also be rapidly evolving in all species except *D. phalerata* (Figure 3, $dN/dS > 1$ for 6 of 8 genes, *t*-test *p*-value < 0.05 for difference from background), suggesting that the hosts are also adapting to escape viral cell entry, while genes within the GO term 'Defense response to virus' (and related terms) show no significant difference from the background (*t*-test *p*-value > 0.12). After controlling for multiple nucleotide substitutions with δ (Pond et al. 2005; Venkat et al. 2018), we find that both JAK-STAT and viral capsid associated genes are rapidly evolving in both trios (Supplementary Figure 1, GLM $t > 3.22$, *p*-value > 0.00128), however viral capsid associated genes are even more rapidly evolving in the *innubila* trio (Supplementary Figure 1, GLM $t = 4.124$, *p*-value = 3.73e-05).

Discussion

Host-parasite coevolution is ubiquitous across the tree of life (Dawkins and Krebs 1979; Lively 1996). It is expected to result in rapid evolution in both the host and the parasite, as both organisms are adapting to escape the selective pressure exerted by the other. In support of this, the immune genes in *Drosophila* and other organisms evolve more rapidly than most other gene categories (Kimbrell and Beutler 2001; Sackton et al. 2007; Obbard et al. 2009; Enard et al. 2016), the fastest among these being the antiviral genes (Obbard et al. 2006; Obbard et al. 2009; Enard et al. 2016; Palmer et al. 2018a). We expected the known antiviral genes to also be rapidly evolving in *D. innubila*, driven by a high frequency DNA virus infection compared to its sister species *D. falleni* and *D. phalerata*. We therefore created a high quality, highly contiguous genome for *D. innubila* and assessed the rates of evolution of genes across the genome with a focus on genes involved in the immune response. Surprisingly, we find few of the canonical antiviral genes to be fast evolving in this species, and these genes are significantly under expressed in *D. innubila* (Figure 2). Interestingly, we find the JAK-STAT pathway and associated genes (previously shown to be involved in DNA virus infection response (West and Silverman 2018) are fast evolving in the *D. innubila* trio (Figure 3).

There are several explanations for the observed differences in immune evolution between the *D. melanogaster* and *D. innubila* trios. The most obvious explanation is that the immune response to DNA virus infection has diverged in the approximately 50 million years since the *Drosophila* quinaria group and the *Drosophila melanogaster* group last shared a common ancestor. They may fundamentally differ in the immune response to virus infection, possibly due to the divergence of the small RNA pathways across the *Drosophila* groups (Lewis et al. 2018). A second hypothesis is that the DNA virus response in *Drosophila* may include additional pathways as well as the siRNA pathway used against RNA viruses. This is consistent with previous studies of the DNA virus immune response in *D. melanogaster* (Palmer et al. 2018a; Palmer et al. 2018b; West and Silverman 2018). As *D. innubila* seem to be less frequently exposed to RNA viruses than DNA viruses (Unckless 2011), it makes sense that the canonical RNA virus immune pathway is not fast evolving, while components of the JAK-STAT and IMD pathway implicated in DNA virus defense are (Palmer et al. 2018a; West and Silverman 2018). Furthermore, as DNA viruses still produce dsRNA due to overlapping transcripts their infection still induces the siRNA pathway irrelevant of the source (Webster et al. 2015; Palmer et al. 2018c). The siRNA pathway in *D. melanogaster* may still be active against DNA viruses and

bolsters further response pathways (Palmer et al. 2018a; Palmer et al. 2018b; West and Silverman 2018).

Beyond the siRNA pathway, there are several components of the immune system that may play a role in the response to DNA virus infection. Previous work in *D. melanogaster* has highlighted the role that Gram-negative commensal bacteria play in priming the antiviral immune system (Sansone et al. 2015). As both the Gram-positive and Gram-negative signaling pathways are rapidly evolving across the *innubila* trio, it is conceivable that this or a similar pathway in conjunction with the JAK-STAT pathway, commensal bacteria and (currently unidentified) Turandots may be the primary antiviral immune response in these groups. Studies have also identified an interaction between *Wolbachia* infection and resistance or susceptibility to viral infection (Teixeira et al. 2008; Martinez et al. 2014). The high frequency of *Wolbachia* infection in *D. innubila* (Dyer and Jaenike 2005), may therefore provide some tolerance to viral infection or may be involved in immune system priming.

Overall, we have worked to bring mycophagous *Drosophila* to the table as a modern genomic model for the study of immune system evolution and have highlighted that the evolution of the antiviral immune system among closely related trios of species may differ drastically across the *Drosophila* genera. We find that across the *D. innubila* genome, while the immune system is rapidly evolving, the canonical antiviral pathways are not evolving as if in an arms race with viruses, though several alternative immune pathways may be evolving in response to the high virus infection frequency.

Methods

DNA/RNA isolation, library preparation, genome sequencing and assembly

We extracted DNA following the protocol described in (Chakraborty et al. 2017) for *D. innubila, D. falleni* and *D. phalerata* as further described in the supplementary materials. We prepared the *D. innubila* DNA as a sequencing library using the Oxford Nanopore Technologies Rapid 48-hour (SQK-RAD002) protocol, which was then sequenced using a MinION (Oxford Nanopore Technologies, Oxford, UK; submitted to NCBI SRA upon acceptance) (Jain et al. 2016). The same DNA was also used to construct a fragment library with insert sizes of ~180bp, ~3000bp and ~7000bp, we sequenced this library on a MiSeq (300bp paired-end, Illumina, San Diego, CA, USA; submitted to NCBI SRA upon acceptance). We prepared the *D. falleni* and *D.*

phalerata samples as Illumina libraries like *D. innubila* but with a 300bp insert size. We sequenced the *D. falleni* fragment library on one half of a MiSeq (300bp paired-end, Illumina, San Diego, CA, USA; submitted to NCBI SRA upon acceptance) by the KU CMADP Genome Sequencing Core. We sequenced the *D. phalerata* fragment library on a fraction of an Illumina HiSeq 4000 run (150bp paired end, Illumina, San Diego, CA; submitted to NCBI SRA upon acceptance).

For gene expression analyses, we obtained two replicate samples of female and male heads and whole bodies (including heads), embryos, larvae (pooled across all three instar stages) and pupae (all non-adults were unsexed). RNA was extracted using a standard Trizol procedure (Simms et al. 1993) with a DNAse step. RNA-sequencing libraries were constructed using the standard TruSeq protocol (McCoy et al. 2014) with ½ volume reactions to conserve reagents. Individually indexed RNA libraries (2 replicates from each tissue/sex) were sequenced on one lane of an Illumina "Rapid" run with 100bp single-end reads (submitted to NCBI SRA upon acceptance).

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. Raw reads were assembled using CANU version 1.6 (Koren et al. 2016) with an estimated genome size of 150 million bases and the "nanopore-raw" flag. We then used Pilon to polish the genome with our Illumina fragment library (Walker et al. 2014). The resulting assembly was submitted to PhaseGenomics (Seattle, WA, USA) for scaffolding using Hi-C and further polished with Pilon for seven iterations. With each iteration, we evaluated the quality of the genome and the extent of improvement in quality, by calculating the N50 and using BUSCO to identify the presence of conserved genes in the genome, from a database of 2799 single copy Dipteran genes.

Repetitive regions were identified *de novo* using RepeatModeler (engine = NCBI) (Smit and Hubley 2008) and RepeatMasker (-gff –gcalc –s) (Smit and Hubley 2015).

These sequencing and assembly steps are further described in the supplementary methods, alongside additional steps taken to verify genes, identify additional contigs and genes, and find genes retained across all species.

Genome annotation

As further described in the supplementary methods, we assembled the transcriptome, using all Illumina RNA reads following quality filtering, using Trinity (version 2.4.0) (Haas et al. 2013), Oases (velvetg parameters: -exp_cov 100 -max_coverage 500 -min_contig_lgth 50 -read_trkg yes) (Schulz et al. 2012), and SOAP*denovo* Trans (127mer with parameters: SOAPdenovo-Trans-127mer -p 28 -e 4 and the following kmers: 95, 85, 75, 65, 55, 45, 35, 29, 25, 21) (Xie et al. 2014) which we combined using EvidentialGene (Gilbert 2013; http://eugenes.org/EvidentialGene/). We used the *D. innubila* transcriptome as well as protein databases from *M. domestica*, *D. melanogaster*, and *D. virilis*, in MAKER2 (Holt and Yandell 2011) to annotate the *D. innubila* genome, including using *RepeatModeler* (Smit and Hubley 2008) to not mis-assign repetitive regions. This was repeated for three iterations to generate a GFF file containing gene evidence generated by MAKER2 (Holt and Yandell 2011) (submitted to NCBI with genome upon acceptance).

Drosophila quinaria group species on the Drosophila phylogeny

To build a phylogeny for the *Drosophila* species including *D. innubila, D. falleni* and *D. phalerata*, we identified genes conserved across all *Drosophila* and humans and found in the Dipteran BUSCO gene set (Simão et al. 2015). We then randomly sampled 100 of these genes, extracted their coding sequence from our three focal species and 9 of the 12 *Drosophila* genomes (Limited to nine due to our focus on the *Drosophila* subgenus and the close relation of several species, Clark *et al*. 2007). We also searched for genomes in the *Drosophila* subgenus with easily identifiable copies of these 100 conserved genes, settling on *D. busckii* (Zhou and Bachtrog 2015), *D. neotestactea* (Hamilton et al. 2014), *D. immigrans* and *Scaptodrosophila lebanonensis* (Zhou et al. 2012). We aligned these genes using MAFFT (--auto) (Katoh et al. 2002), concatenated all alignments and generated a phylogeny using PhyML with 500 bootstraps (-M GTR, -Gamma 8, -B 500) (Guindon et al. 2010).

Signatures of adaptive molecular evolution among species

We mapped short read sequencing data of *D. innubila, D. falleni* and *D. phalerata* to the repeat-masked *D. innubila* reference genome using BWA MEM (Li and Durbin 2009). As similar proportions of reads mapped to the genome (97.6% for *D. innubila*, 96.1% for *D. falleni* and 94.3% for *D. phalerata*), covering a similar proportion of the reference genome (99.1% for *D.*

innubila, 98.5% for *D. falleni* and 97.1% for *D. phalerata*), we considered the *D. innubila* genome to be of similar enough to these species to reliably call single nucleotide polymorphisms and indels. We realigned around indels using GATK IndelRealigner then called variants using HaplotypeCaller (default parameters) (McKenna et al. 2010; DePristo et al. 2011). We then used GATK FastaReferenceMaker (default parameters) to generate an alternate, reference genome for each of these species (McKenna et al. 2010; DePristo et al. 2011). We extracted the coding sequence for each gene found in the genomes of *D. innubila, D. phalerata* and *D. falleni* and aligned orthologs using PRANK (-codon +F –f=paml) (Löytynoja 2014). For each PRANK generated gene alignment and gene tree, we used codeML (Yang 2007) for the branches model (M0 model), to identify genes with signatures of rapid evolution on the *D. innubila, D. falleni, D. phalerata* branches. For genes involved in RNAi or the immune system pathways, we attempted to rescale for synonymous divergence. For each focal gene, we found genes with *dS* within 0.01 of the focal gene. We then found the difference in *dN/dS* between the focal gene and the median of the control gene group.

For an independent contrast, we downloaded the coding sequences for *D. melanogaster, D. simulans* and *D. yakuba* from Flybase.org and aligned orthologous genes using PRANK (-codon +F –f=paml) (Löytynoja 2014). Following the generation of a gene alignment and gene tree, we used codeML (Yang 2007) to identify genes with adaptive molecular signatures on each branch of the phylogeny (using the branch based model, M0). Again, we found the difference in *dN/dS* from background genes of similar *dS* (with 0.01) for all immune genes.

Finally, to control for possibly multiple nucleotide substitutions creating false signals of rapid evolution (Venkat et al. 2018), we calculated δ using HyPhy (Pond et al. 2005) based on the method presented in (Venkat et al. 2018). δ was calculated under both the null and alternative models, with the best model selected based on the result of a Chi-squared test. We then compared δ between each species and across immune gene categories.

RNA differential expression analysis

We used GSNAP (-sam –N 1) (Wu and Nacu 2010) to map each set of *D. innubila* RNA sequencing short reads to the masked *D. innubila* genome with the TE sequences concatenated at the end. We then counted the number of reads mapped to each gene per kb of gene using HTSeq (Anders et al. 2015) for all mapped RNA sequencing data and normalized by counts per million

per dataset. Mapped RNA sequencing information for *D. melanogaster* across all life stages was downloaded from ModEncode (modencode.org) (Chen et al. 2014). We compared *D. melanogaster* data to *D. innubila* data using EdgeR (Robinson et al. 2009) to identify differentially expressed genes, and also compared RNAseq reads per million reads per 1kbp of exon (FPKM) between the immune genes of *D. innubila* and *D. melanogaster*.

Statistical analysis

All statistical analyses were performed using R (R Core Team 2013). We used the R packages EdgeR (Robinson et al. 2009), RCircos (Zhang et al. 2013) and ggplot2 (Wickham 2009) for statistical analysis and plot generation/visualization.

Acknowledgements

We are thankful to Justin Blumenstiel, Jamie Walters, John Kelly, Carolyn Wessinger and Joanne Chapman for helpful discussion and advice concerning the manuscript. We thank Brittny Smith and the KU CMADP Genome Sequencing Core (NIH Grant P20 GM103638) for assistance in genome isolation, library preparation and sequencing. Genome annotation was performed in the K-INBRE Bioinformatics Core supported by P20 GM103418. This work was supported by a postdoctoral fellowship from the Max Kade foundation (Austria) and a K-INBRE postdoctoral grant to TH (NIH Grant P20 GM103418). This work was also funded by NIH Grant R00 GM114714 to RLU.

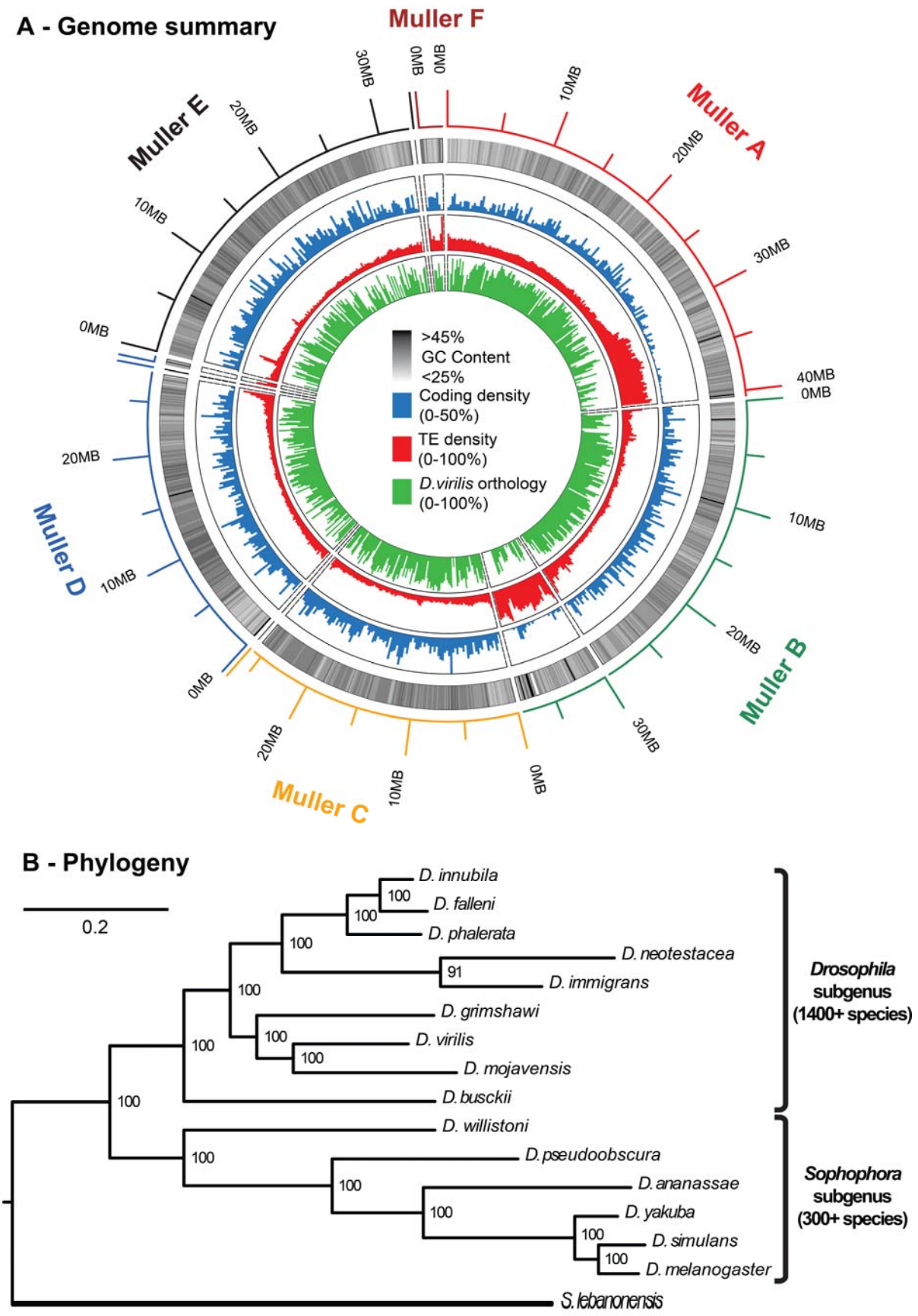
Tables and Figures

Table 1: Summary of the major assembled scaffolds in the *D. innubila* genome, including the length in kilobase pairs (kbp), number of genes, number of orphan genes, gene lengths in base pairs (bp) and proportion of the scaffold that is repetitive content.

Scaffold (Muller element - *D. melanogaster* ortholog - ID)	**Total Length (kbp)**	**Total Gene count**	**Orphan count**	**Mean gene length (bp)**	**Mean repeat content (%)**
A-X-3	40479	2012	73	5090.7	31.7
B-2L-4	29570.5	2249	99	4394.6	13.9
B-2L-5	8037.1	160	13	4679.2	62.9
C-2R-1	25683.3	2518	58	4033.3	12.2
C-2R-27	16.5	0	0	NA	92.4
D-3L-0	27707.2	2345	68	4248.1	12.6
D-3L-29	45.7	0	0	NA	83.7
E-3R-2	32746.5	2863	73	4092.4	11.4
E-3R-11	18.4	1	1	282	75.6
F-4-6	2027.1	106	7	6992	35.5
Unassembled (489 contigs)	1699.4	46	1	1429	39.1
Mitochondria	16.1	18	0	1658.7	0
Total	**168031**	**12318**	**393**	**4350.7**	**13.5**

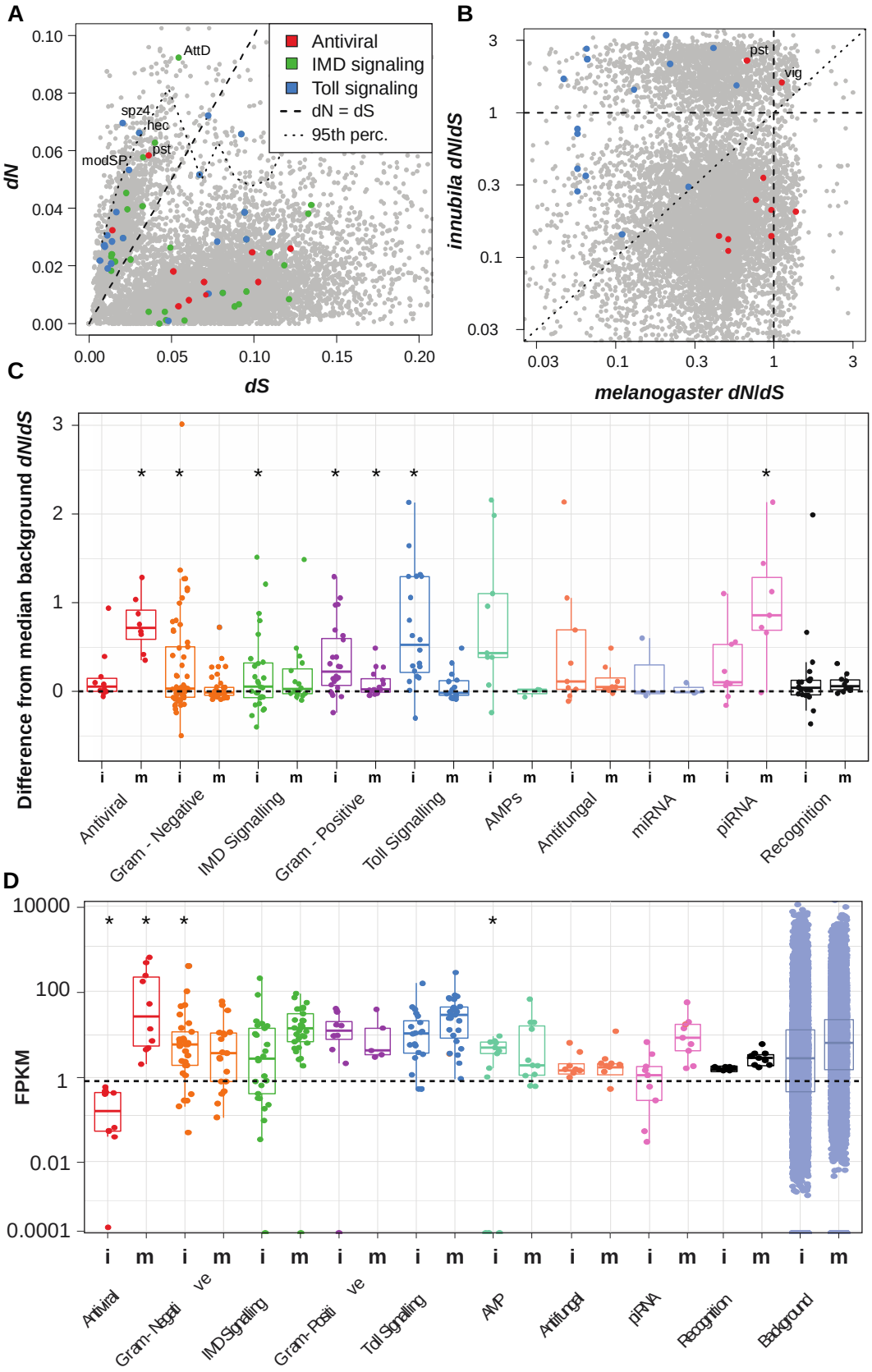
Figure 1. A. A circular summary of the *D. innubila* genome limited to the assembled acrocentric chromosomes on ten major scaffolds. The rings (from the outside in) are the chromosome identity, the length of each segment, the percentage GC content, the coding density, the transposable element density (250 kilobase pairs windows, sliding 250 kilobase pairs), and the percentage of each window that aligns to *D. virilis* (identified with progressiveMauve, 250kbp windows, sliding 250kbp). **B.** Phylogenetic relationships of *D. innubila*, *D. falleni*, *D. phalerata* and the main lineages of *Drosophila*. We randomly sampled 100 genes, conserved across all

species and aligned each gene using MAFFT. We then concatenated these sequences and assembled a maximum-likelihood phylogeny with 500 bootstraps in PhyML (shown as % support on the nodes).


A - Genome summary
Muller F
Muller A
Muller B
Muller C
Muller D
Muller E
0MB
10MB
20MB
30MB
40MB
>45%
GC Content
<25%
Coding density
(0-50%)
TE density
(0-100%)
D.virilis orthology
(0-100%)
B - Phylogeny
0.2
D. innubila
D. falleni
D. phalerata
D. neotestacea
D. immigrans
D. grimshawi
D. virilis
D. mojavensis
D. busckii
D. willistoni
D. pseudoobscura
D. ananassae
D. yakuba
D. simulans
D. melanogaster
S. lebanonensis
Drosophila
subgenus
(1400+ species)
Sophophora
subgenus
(300+ species)

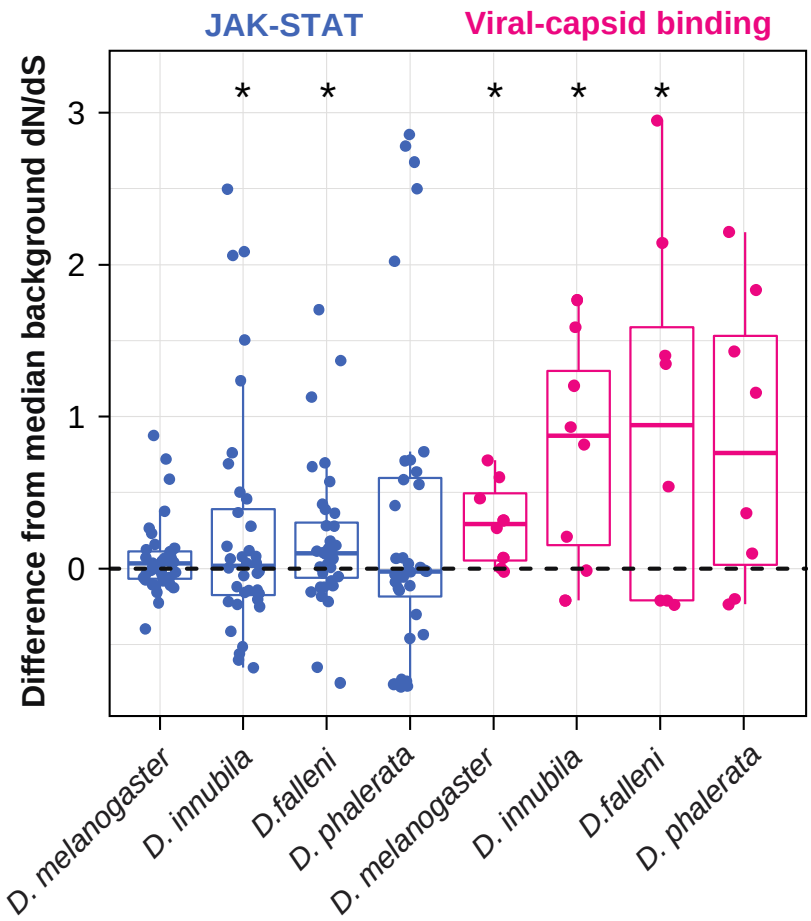
408

Figure 2: A. The non-synonymous (potentially adaptive) divergence (*dN*) for each gene across the *D. innubila* genome, compared to the gene's synonymous (neutral) divergence (*dS*). Genes involved in antiviral and antibacterial (*Toll* and *IMD*) immune signaling pathways are highlighted in color. The upper 97.5^{th} percentile is shown as a dotted line, *dN/dS* = 1 is shown as a dashed line. **B.** Comparison of *dN/dS* for all shared genes in the *D. melanogaster* and *D. innubila* trios, highlighting antiviral and Toll signaling genes in each species. **C.** For each immune gene or RNAi gene, the difference in *dN*/*dS* between the gene and the mean of background genes of similar *dS* (+-0.01*dS*). i refers to *D. innubila* while m refers to *D. melanogaster*. A *p*-value (from a *t*-test looking for significant differences from 0) of 0.05 or lower is designated with *. **D.** Expression by read counts per million read counts per 1kbp of exon for each gene, comparing *D. melanogaster* and *D. innubila* expression for each immune category, as well as compared to the background. i refers to *D. innubila* while m refers to *D. melanogaster*.


A
Antiviral
IMD signaling
Toll signaling
dN = dS
95th perc.
AttD
spz4
hec
pst
modSP
dN
dS
B
pst
vig
innubila dN/dS
melanogaster dN/dS
C
Difference from median background dN/dS
Antiviral
Gram - Negative
IMD Signalling
Gram - Positive
Toll Signalling
AMPs
Antifungal
miRNA
piRNA
Recognition
D
FPKM
Antiviral
Gram-Negative
IMD Signalling
Gram-Positive
Toll Signalling
AMP
Antifungal
piRNA
Recognition
Background

422

Figure 3: Alternative antiviral evolution across the *quinaria* group. Difference between JAK-STAT and viral capsid binding proteins from the background *dN*/*dS* of genes of similar *dS* (+-0.01*dS*) for *D. melanogaster*, *D. innubila*, *D. falleni* and *D. phalerata*. A *p*-value (from a *t*-test looking for significant differences from 0) of 0.05 or lower is designated with *.



Supplementary Tables and Figures

Supplementary Table 1: Summary of reads used for genome sequencing, assembly, annotation and *dN*/*dS* calculation.

Supplementary Table 2: Summary statistics for each iteration of the genome.

Supplementary Table 3: Summary of the genic characteristics of the *D. innubila* genome.

Supplementary Table 4: Summary of *dN*/*dS* statistics for each immune gene category across the total group and on each branch. Additionally, the t-score and p-value for a two-sided *t*-test (μ = 0) for that category is shown. Significant categories are highlighted in bold.

Supplementary Table 5: A table summarizing the DGE shown in Supplementary Tables 6-12, showing the number of genes differentially expressed between *D. innubila* and *D. melanogaster* at differing life stages, with enrichments in gene ontology (GO) categories.

Supplementary Table 6: *dN*/*dS* enrichment for *Drosophila innubila* trio for processes, components and functions, including any enrichments for specific branches.

Supplementary Table 7: *dN*/*dS* enrichment for duplications for processes, components and functions, including any enrichments for specific branches.

Supplementary Table 8: Enrichment for processes, components and functions for expression between *D. melanogaster* and *D. innubila* embryos.

Supplementary Table 9: Enrichment for processes, components and functions for expression between *D. melanogaster* and *D. innubila* larvae.

Supplementary Table 10: Enrichment for processes, components and functions for expression between *D. melanogaster* and *D. innubila* pupae.

Supplementary Table 11: Enrichment for processes, components and functions for expression between *D. melanogaster* and *D. innubila* adults.

Supplementary Table 12: Enrichment for processes, components and functions for expression between *D. melanogaster* and *D. innubila* adult males.

Supplementary Table 13: Enrichment for processes, components and functions for expression between *D. melanogaster* and *D. innubila* adult females.

Supplementary Table 14: Enrichment for processes, components and functions for expression between *D. melanogaster* and *D. innubila* total samples.

Supplementary Table 15: Enrichment for processes, components and functions for expression between *D. innubila* males and females.

Supplementary Table 16: Enrichment for processes, components and functions for expression between *D. innubila* embryos and larvae.

Supplementary Table 17: Enrichment for processes, components and functions for expression between *D. innubila* larvae and pupae.

Supplementary Table 18: Enrichment for processes, components and functions for expression between *D. innubila* pupae and adults.

Supplementary Figure 1: δ (calculated using HyPhy) by immunity category for both *D. innubila* and *D. melanogaster*.

Supplementary Figure 2: Codon bias for each gene across the *Drosophila innubil*a genome, separated by scaffold. CAI = Codon adaptation index. CBI = Codon bias index. Fop = Frequency of optimal codons. GC = Proportion of GC across each gene.

Supplementary Figure 3: A. The proportion of the *D. innubila* genome masked by each type of repeat. LINE = Long interspersed nuclear element RNA transposon, LTR = long terminal repeat RNA transposon, RC = rolling circle DNA transposon, TIR = terminal inverted repeat DNA transposon. **B.** TE content of *D. innubila, falleni* and *phalerata,* **C.** Copy number comparisons between *D. innubila, D. falleni* and *phalerata*. **D.** dnaPipeTE estimates of the genomic proportion of repetitive elements for each species examined here. Other, NA and SINE categories were removed due to small proportions. Though unlabeled, rRNA is shown in yellow and constitutes 1-2% of the genome.

Supplementary Figure 4: Number of TE families found in *D. innubila,* closely related to known TE families (taken from Repbase) in different species group, identified using BLAST, suggesting relatively recent horizontal transfer events.

Supplementary Figure 5: Comparison of orphan genes to previously known genes, including: **A.** Codon adaptation index (CAI). **B.** Codon bias index (CBI). **C.** Frequency of optimal codons (Fop). **D.** Gene length (in bp). **E.** Number of introns per gene. **F.** Mean expression across life stages (read counts per million).

Supplementary Figure 6: *dN*/*dS* versus *dS* across paralogs for recently duplicated genes. Metal ion binding, protein metabolism and immunity genes are highlighted.

Supplementary Figure 7: Volcano plots showing differential gene expression between *D. innubila* and *D. melanogaster* at different life stages. Dots are colored by their significance and if a recent duplication or not (duplicates layered on top), the significance cut off is set a 0.05 following multiple testing correction.

Supplementary Figure 8: Volcano plot showing differential gene expression between *D. innubila* male and female samples and significant differences, highlighting if genes are duplicated relative to *D. virilis* or not, and if genes are involved in sperm motility.

Supplementary Figure 9: Inversions identified between *D. innubila and D. falleni, and between D. innubila/falleni and D. phalerata* using Pindel (Ye et al. 2009) and Manta (Chen et al. 2016) (taking the consensus of the two programs). Scaffolds are labelled and colored by the Muller element they belong to.

Supplementary Figure 10: Size and number of each structural variant between *D. innubila* and *D. falleni* identified using Pindel and Manta (taking the consensus of the two programs).

References

Anders S, Pyl PT, Huber W. 2015. HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics 31:166–169.

Barbier V. 2013. Pastrel, a restriction factor fo picornalike-viruses in Drosophila melanogaster.

Chakraborty M, Zhao R, Zhang X, Kalsow S, Emerson JJ. 2017. Extensive hidden genetic variation shapes the structure of functional elements in Drosophila. Doi.Org 50:114967.

Chen X, Schulz-Trieglaff O, Shaw R, Barnes B, Schlesinger F, Källberg M, Cox AJ, Kruglyak S, Saunders CT. 2016. Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32:1220–1222.

Chen Z, Sturgill D, Qu J, Jiang H. 2014. Comparative validation of the D. melanogaster modENCODE transcriptome annotation. Genome …:1209–1223.

Clark AG, Eisen MB, Smith DR, Bergman CM, Oliver B, Markow T a, Kaufman TC, Kellis M, Gelbart W, Iyer VN, et al. 2007. Evolution of genes and genomes on the Drosophila phylogeny. Nature 450:203–218.

Consortium T, Roy S, Ernst J, Kharchenko P V, Kheradpour P, Negre N, Eaton ML, Landolin JM, Bristow C a, Ma L, et al. 2011. Identification of Functional Elements and Regulatory Circuits by Drosophila modENCODE. Science (80-.). [Internet] 330:1787–1797. Available from: http://science.sciencemag.org/content/330/6012/1787.abstract

Darling ACE, Mau B, Blattner FR, Perna NT. 2004. Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements Mauve: Multiple Alignment of Conserved Genomic Sequence With Rearrangements. :1394–1403.

Daugherty MD, Malik HS. 2012. Rules of engagement: molecular insights from host-virus arms races. Annu. Rev. Genet. 46:677–700.

Dawkins R, Krebs JR. 1979. Arms Races between and within Species. Proc. R. Soc. London B Biol. Sci. 205:489–511.

DePristo MA, Banks E, Poplin R, Garimella K V, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and

genotyping using next-generation DNA sequencing data. Nat. Genet. 43:491–498.

Ding SW. 2010. RNA-based antiviral immunity. Nat. Rev. Immunol. 10:632–644.

Dyer K a., Jaenike J. 2005. Evolutionary dynamics of a spatially structured host-parasite association: Drosophila innubila and male-killing Wolbachia. Evolution 59:1518–1528.

Dyer KA. 2004. Evolutionarily Stable Infection by a Male-Killing Endosymbiont in Drosophila innubila: Molecular Evidence From the Host and Parasite Genomes. Genetics 168:1443–1455.

Dyer KA, Minhas MS, Jaenike J. 2005. Expression and modulation of embryonic male-killing in Drosophila innubila: opportunities for multilevel selection. Evolution 59:838–848.

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

Ekengren S, Hultmark D. 2001. A family of Turandot-related genes in the humoral stress response of Drosophila. Biochem. Biophys. Res. Commun. 284:998–1003.

Enard D, Cai L, Gwennap C, Petrov DA. 2016. Viruses are a dominant driver of protein adaptation in mammals. Elife 5:1–25.

Gilbert D. 2013. EvidentialGene: mRNA Transcript Assembly Software. EvidentialGene Evid. Dir. Gene Constr. Eukaryotes.

Goubert C, Modolo L, Vieira C, Moro CV, Mavingui P, Boulesteix M. 2015. De novo assembly and annotation of the Asian tiger mosquito (Aedesalbopictus) repeatome with dnaPipeTE from raw genomic reads and comparative analysis with the yellow fever mosquito (Aedes aegypti). Genome Biol. Evol. 7:1192–1205.

Guindon S, Dufayard J-F, Lefort V, Anisimova M, Hordijk W, Gascuel O. 2010. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. Syst. Biol. 59:307–321.

Haas BJ, Papanicolaou A, Yassour M, Grabherr M, Blood PD, Bowden J, Couger MB, Eccles D, Li B, Lieber M, et al. 2013. De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. Nat. Protoc. 8:1494–1512.

Hamilton PT, Leong JS, Koop BF, Perlman SJ. 2014. Transcriptional responses in a Drosophila defensive symbiosis. Mol. Ecol. 23:1558–1570.

Hill T, Unckless RL. 2017. The dynamic evolution of Drosophila innubila Nudivirus. Infect. Genet. Evol.:1–25.

Holt C, Yandell M. 2011. MAKER2: an annotation pipeline and genome- database management tool for second- generation genome projects. BMC Bioinformatics [Internet] 12:491. Available from: http://www.biomedcentral.com/1471-2105/12/491

Jaenike J, Dyer KA. 2008. No resistance to male-killing Wolbachia after thousands of years of infection. J. Evol. Biol. [Internet] 21:1570–1577. Available from: http://dx.doi.org/10.1111/j.1420-9101.2008.01607.x

Jaenike J, Dyer KA, Reed LK. 2003. Within-population structure of competition and the dynamics of male-killing Wolbachia. Evol. Ecol. Res. 5:1023–1036.

Jaenike J, Perlman SJ. 2002. Ecology and Evolution of Host-Parasite Associations: Mycophagous Drosophila and Their Parasitic Nematodes. Am. Nat. 160:S23–S39.

Jain M, Olsen HE, Paten B, Akeson M. 2016. Erratum to: The Oxford Nanopore MinION: Delivery of nanopore sequencing to the genomics community [Genome Biol. (2016), 17, 239] DOI:10.1186/s13059-016-1103-0. Genome Biol. 17:1–11.

Janssens S, Beyaert R. 2003. Role of Toll-Like Receptors in Pathogen Recognition Role of Toll-Like Receptors in Pathogen Recognition. Clin. Microbiol. Rev. 16:637–646.

Katoh K, Misawa K, Kuma K, Miyata T. 2002. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. Nucleic Acids Res. 30:3059–3066.

Kimbrell DA, Beutler B. 2001. The evolution and genetics of innate immunity. Nat Rev Genet 2:256–267.

Koren S, Walenz BP, Berlin K, Miller JR, Bergman NH, Phillippy AM. 2016. Canu: scalable and accurate long- - - read assembly via adaptive k - - - mer weighting and repeat separation. :1–35.

Lewis SH, Quarles KA, Yang Y, Tanguy M, Frézal L, Smith SA, Sharma PP, Cordaux R, Gilbert C, Giraud I, et al. 2018. Pan-arthropod analysis reveals somatic piRNAs as an ancestral defence against transposable elements. Nat. Ecol. Evol. 2:174–181.

Li H, Durbin R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25:1754–1760.

Lively CMCM. 1996. Host-parasite coevolution and sex. Bioscience [Internet] 46:107–114. Available from: http://www.jstor.org/stable/10.2307/1312813%5Cnhttp://www.jstor.org/stable/1312813

Löytynoja A. 2014. Phylogeny-aware alignment with PRANK. In: Russell DJ, editor. Multiple

Sequence Alignment Methods. Totowa, NJ: Humana Press. p. 155–170. Available from: http://dx.doi.org/10.1007/978-1-62703-646-7_10

Magwire MM, Bayer F, Webster CL, Cao C, Jiggins FM. 2011. Successive increases in the resistance of Drosophila to viral infection through a transposon insertion followed by a duplication. PLoS Genet. 7:1–11.

Magwire MM, Fabian DK, Schweyen H, Cao C, Longdon B, Bayer F, Jiggins FM. 2012. Genome-Wide Association Studies Reveal a Simple Genetic Basis of Resistance to Naturally Coevolving Viruses in Drosophila melanogaster. PLoS Genet. 8:1–11.

Markow TA, O'Grady P. 2006. *Drosophila*: a guide to species identification.

Martinez J, Longdon B, Bauer S, Chan YS, Miller WJ, Bourtzis K, Teixeira L, Jiggins FM. 2014. Symbionts commonly provide broad spectrum resistance to viruses in insects: A comparative analysis of *Wolbachia* strains. PLoS Pathog. 10:e1004369.

Martinson VG, Douglas AE, Jaenike J. 2017. Community structure of the gut microbiota in sympatric species of wild Drosophila. Ecol. Lett. 20:629–639.

McCoy RC, Taylor RW, Blauwkamp TA, Kelley JL, Kertesz M, Pushkarev D, Petrov DA, Fiston-Lavier AS. 2014. Illumina TruSeq synthetic long-reads empower de novo assembly and resolve complex, highly-repetitive transposable elements. PLoS One 9.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky AM, Garimella K V, Altshuler D, Gabriel SB, Daly MJ, et al. 2010. The Genome Analysis Toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Proc. Int. Conf. Intellect. Capital, Knowl. Manag. Organ. Learn. 20:1297–1303.

Merkling SH, van Rij RP. 2013. Beyond RNAi: Antiviral defense strategies in Drosophila and mosquito. J. Insect Physiol. 59:159–170.

Obbard DJ, Gordon KHJ, Buck AH, Jiggins FM. 2009. The evolution of RNAi as a defence against viruses and transposable elements. Philos. Trans. R. Soc. Lond. B. Biol. Sci. 364:99–115.

Obbard DJ, Jiggins FM, Halligan DL, Little TJ. 2006. Natural Selection Drives Extremely Rapid Evolution in Antiviral RNAi Genes. Curr. Biol. 16:580–585.

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

Palmer WH, Hadfield JD, Obbard DJ. 2018. RNA Interference Pathways Display High Rates of

Adaptive Protein Evolution in Multiple Invertebrates. Genetics:genetics.300567.2017.

Palmer WH, Joosten J, Overheul GJ, Jansen PW, Vermeulen M, Obbard J, Rij RP Van. 2018. Induction and suppression of NF- κB signalling by a DNA virus of Drosophila.

Palmer WH, Medd N, Beard PM, Obbard DJ. 2018. Isolation of a natural DNA virus of Drosophila melanogaster, and characterisation of host resistance and immune responses. PLoS Pathog. 14:1–26.

Patterson JT, Stone WS. 1949. Studies in the genetics of *Drosophila*. Univ. Texas Publ. 4920:7–17.

Perlman SJ, Spicer GS, Dewayne Shoemaker D, Jaenike J. 2003. Associations between mycophagous Drosophila and their Howardula nematode parasites: A worldwide phylogenetic shuffle. Mol. Ecol. 12:237–249.

Pond S, Frost S, Muse S. 2005. HyPhy: hypothesis testing using phylogenies. Bioinformatics21: 676–679. :1–57. Available from: https://scholar.google.com/scholar?cluster=1114532013043740554&hl=pt-BR&as_sdt=2005&sciodt=0,5

Robinson MD, McCarthy DJ, Smyth GK. 2009. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics 26:139–140.

Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007. Dynamic evolution of the innate immune system in Drosophila. Nat. Genet. 39:1461–1468.

Sansone CL, Cohen J, Yasunaga A, Xu J, Osborn G, Subramanian H, Gold B, Buchon N, Cherry S. 2015. Microbiota-dependent priming of antiviral intestinal immunity in Drosophila. Cell Host Microbe [Internet] 18:571–581. Available from: http://dx.doi.org/10.1016/j.chom.2015.10.010

Schulz MH, Zerbino DR, Vingron M, Birney E. 2012. Oases: Robust de novo RNA-seq assembly across the dynamic range of expression levels. Bioinformatics 28:1086–1092.

Simão FA, Waterhouse RM, Ioannidis P, Kriventseva E V., Zdobnov EM. 2015. BUSCO: Assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31:3210–3212.

Simms D, Cizdziel P, Chomczynski P. 1993. TRIzol: a new reagent for optimal single-step isolation of RNA. Focus (Madison).:99–102.

Smit AFA, Hubley R. 2008. RepeatModeler Open-1.0. Available from: www.repeatmasker.org

Smit AFA, Hubley R. 2015. RepeatMasker Open-4.0.

Takeda K, Akira S. 2005. Toll-like receptors in innate immunity. Int. Immunol. 17:1–14.

Tang J, Wu L. 2013. Identification of genes involved in the antiviral response through genetic screens in *Drosophila*. J. Chem. Inf. Model. 53:1689–1699.

Team RC. 2013. R: A Language and Environment for Statistical Computing. Available from: http://www.r-project.org

Teixeira L, Ferreira A, Ashburner M. 2008. The bacterial symbiont *Wolbachia* induces resistance to RNA viral infections in *Drosophila melanogaster*. PLoS Biol. 6:e2.

Unckless RL. 2011. A DNA Virus of *Drosophila*. PLoS One [Internet] 6:e26564. Available from: http://dx.plos.org/10.1371/journal.pone.0026564

Unckless RL, Jaenike J. 2011. Maintenance of a Male-Killing *Wolbachia* in *Drosophila Innubila* By Male-Killing Dependent and Male-Killing Independent Mechanisms. Evolution (N. Y). 66:678–689.

Venkat A, Hahn MW, Thornton J. 2018. Multinucleotide mutations cause false inferences of lineage-specific positive selection. Biorxiv [Internet] 2018:1–32. Available from: https://scholar.google.com/scholar?cluster=1114532013043740554&hl=pt-BR&as_sdt=2005&sciodt=0,5

Walker BJ, Abeel T, Shea T, Priest M, Abouelliel A, Sakthikumar S, Cuomo CA, Zeng Q, Wortman J, Young SK, et al. 2014. Pilon: An integrated tool for comprehensive microbial variant detection and genome assembly improvement. PLoS One 9.

Webster CL, Waldron FM, Robertson S, Crowson D, Ferrari G, Quintana JF, Brouqui J-M, Bayne EH, Longdon B, Buck AH, et al. 2015. The discovery, distribution, and evolution of viruses associated with Drosophila melanogaster. PLoS Biol. 13:e1002210.

West C, Silverman N. 2018. p38b and JAK-STAT signaling protect against invertebrate iridescent virus 6 infection in Drosophila. Plant Pathol.:1–22.

Wickham H. 2009. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York Available from: http://ggplot2.org

Wu TD, Nacu S. 2010. Fast and SNP-tolerant detection of complex variants and splicing in short reads. Bioinformatics [Internet] 26:873–881. Available from: http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2844994&tool=pmcentrez&rendertype=abstract

Xie Y, Wu G, Tang J, Luo R, Patterson J, Liu S, Huang W, He G, Gu S, Li S, et al. 2014. SOAPdenovo-Trans: De novo transcriptome assembly with short RNA-Seq reads. Bioinformatics 30:1660–1666.

Yang Z. 2007. PAML 4: Phylogenetic analysis by maximum likelihood. Mol. Biol. Evol. 24:1586–1591.

Ye K, Schulz MH, Long Q, Apweiler R, Ning Z. 2009. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25:2865–2871.

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

Zhou Q, Bachtrog D. 2015. Ancestral Chromatin Configuration Constrains Chromatin Evolution on Differentiating Sex Chromosomes in Drosophila. PLoS Genet. 11:1–21.

Zhou Q, Zhu H, Huang Q, Zhao L, Zhang G, Roy SW, Vicoso B, Xuan Z, Ruan J, Zhang Y, et al. 2012. Deciphering neo-sex and B chromosome evolution by the draft genome of [i]Drosophila albomicans[/i]. BMC Genomics 13:109.