Selection and demography shape genomic variation in a 'Sky Island' species

Tom Hill[1]*, 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

* Corresponding author

Keywords: *Drosophila innubila,* local adaptation, phylogeography, inversions

Abstract

Over time, populations of species can expand, contract, and become isolated, creating subpopulations that can adapt to local conditions. Understanding how species adapt following these changes is of great interest, especially as the current climate crisis has caused range shifts for many species. Here, we characterize how *Drosophila innubila* came to inhabit and adapt to its current range: mountain forests in southwestern USA separated by large expanses of desert. Using population genomic data from more than 300 wild-caught individuals, we examine four distinct populations to determine their population history in these mountain-forests, looking for signatures of local adaptation to establish a genomic model for this spatially-distributed system with a well understood ecology. We find *D. innubila* spread northwards during the previous glaciation period (30-100 KYA), and has recently expanded even further (0.2-2 KYA). Surprisingly, *D. innubila* shows little evidence of population structure, though consistent with a recent migration, we find signatures of a population contraction following this migration, and signatures of recent local adaptation and selective sweeps in cuticle development and antifungal immunity. However, we find little support for recurrent selection in these genes suggesting recent local adaptation. In contrast, we find evidence of recurrent positive selection in the Toll-signaling system and the Toll-regulated antimicrobial peptides.

Introduction

In the past 25,000 years, the earth has undergone substantial environmental changes due to both human-mediated events (anthropogenic environment destruction, desert expansion, extreme weather and the growing anthropogenic climate crisis) (CLOUDSLEY-THOMPSON 1978; ROSENZWEIG *et al.* 2008) and events unrelated to humans (glaciation and tectonic shifts) (HEWITT 2000; HOLMGREN *et al.* 2003; SURVEY 2005). These environmental shifts can fundamentally reorganize habitats, influence organism fitness, rates of migration between locations, and population ranges (SMITH *et al.* 1995; ASTANEI *et al.* 2005; ROSENZWEIG *et al.* 2008; SEARLE *et al.* 2009; CINI *et al.* 2012; PORRETTA *et al.* 2012; ANTUNES *et al.* 2015). Signatures of the way organisms adapt to these events are often left in patterns of molecular variation within and between species (CHARLESWORTH *et al.* 2003; WRIGHT *et al.* 2003; EXCOFFIER *et al.* 2009).

When a population migrates to a new location it first goes through a population bottleneck (as only a small proportion of the population will establish in the new location) (CHARLESWORTH *et al.* 2003; EXCOFFIER *et al.* 2009; LI AND DURBIN 2011). These bottlenecks result in the loss of rare alleles in the population (TAJIMA 1989; GILLESPIE 2004). After the bottleneck, the population will grow to fill the carrying capacity of the new niche and adapt to the unique challenges in the new environment, both signaled by an excess of rare alleles (EXCOFFIER *et al.* 2009; WHITE *et al.* 2013). This adaptation can involve selective sweeps from new mutations or standing genetic variation, and signatures of adaptive evolution and local adaptation in genes key to the success of the population in this new location (CHARLESWORTH *et*

al. 2003; HERMISSON AND PENNINGS 2005; MCVEAN 2007; MESSER AND PETROV 2013). However, these signals can confound each other making inference of population history difficult. For example, both population expansions and adaptation lead to an excess of rare alleles, meaning more thorough analysis is required to identify the true cause of the signal (WRIGHT *et al.* 2003).

Signatures of demographic change are frequently detected in species that have recently undergone range expansion due to human introduction (ASTANEI *et al.* 2005; EXCOFFIER *et al.* 2009) or the changing climate (HEWITT 2000; PARMESAN AND YOHE 2003; GUINDON *et al.* 2010; WALSH *et al.* 2011; CINI *et al.* 2012). Other hallmarks of invasive species population genomics include signatures of bottlenecks visible in the site frequency spectrum, and differentiation between populations (CHARLESWORTH *et al.* 2003; LI AND DURBIN 2011). This can be detected by a deficit of rare variants, a decrease in population pairwise diversity and an increase in the statistic, Tajima's D (TAJIMA 1989). Following the establishment and expansion of a population, there is an excess of rare variants and local adaptation results in divergence between the invading population and the original population. These signatures are also frequently utilized in human populations to identify traits which have fixed upon the establishment of a humans in a new location, or to identify how our human ancestors spread globally (LI AND DURBIN 2011).

The Madrean archipelago, located in southwestern USA and northwestern Mexico, contains numerous forested mountains known as 'Sky islands', separated by large expanses of desert (MCCORMACK *et al.* 2009; COE *et al.* 2012). These 'islands' were connected by lush forests during the previous glacial maximum which then retreated, leaving forest habitat separated by hundreds of miles of desert, presumably limiting migration between locations for most species (SURVEY 2005; MCCORMACK *et al.* 2009). The islands are hotbeds of ecological diversity. However, due to the changing climate in the past 100 years, they have become more arid, which may drive migration and adaptation (MCCORMACK *et al.* 2009; COE *et al.* 2012).

Drosophila innubila is a mycophageous *Drosophila* species found throughout these Sky islands and thought to have arrived during the last glacial maximum (DYER AND JAENIKE 2005; DYER *et al.* 2005). Unlike the lab model *D. melanogaster*, *D. innubila* has a well-studied ecology (LACHAISE AND SILVAIN 2004; DYER AND JAENIKE 2005; DYER *et al.* 2005; JAENIKE AND DYER 2008; UNCKLESS 2011a; UNCKLESS AND JAENIKE 2011; COE *et al.* 2012). In fact, in many ways the 'island' endemic, mushroom-feeding ecological model *D. innubila* represent a counterpoint to the human commensal, cosmopolitan, genetic workhorse *D. melanogaster*.

We sought to reconstruct the demographic and migratory history of *D. innubila* inhabiting the Sky islands. Isolated populations with limited migration provide a rare opportunity to observe replicate bouts of evolutionary change and this is particularly interesting regarding the coevolution with pathogens (DYER AND JAENIKE 2005; UNCKLESS 2011a). We also wanted to understand how *D. innubila* adapt to their local

68 climate and if this adaptation is recurrent or recent and specific to local population. We resequenced whole
69 genomes of wild-caught individuals from four populations of *D. innubila* in four different Sky island
70 mountain ranges. Surprisingly, we find little evidence of population structure by location, with structure
71 limited to the mitochondria and a single chromosome (Muller element B which is syntenic with 2L in *D.*
72 *melanogaster)* (MARKOW AND O'GRADY 2006). However, we find some signatures of local adaptation,
73 such as for cuticle development and fungal pathogen resistance, suggesting potentially a difference in fungal
74 pathogens and toxins between locations. We also find evidence of mitochondrial translocations into the
75 nuclear genome, with strong evidence of local adaptation of these translocations, suggesting potential
76 adaptation to changes in metabolic process of the host between location, and possibly even as a means of
77 compensating for reduced efficacy of selection due to *Wolbachia* infection (JAENIKE AND DYER 2008).

78 **Results**

79 *Drosophila innubila has recently expanded its geographic range and shows high levels of gene flow*
80 *between geographically isolated populations*

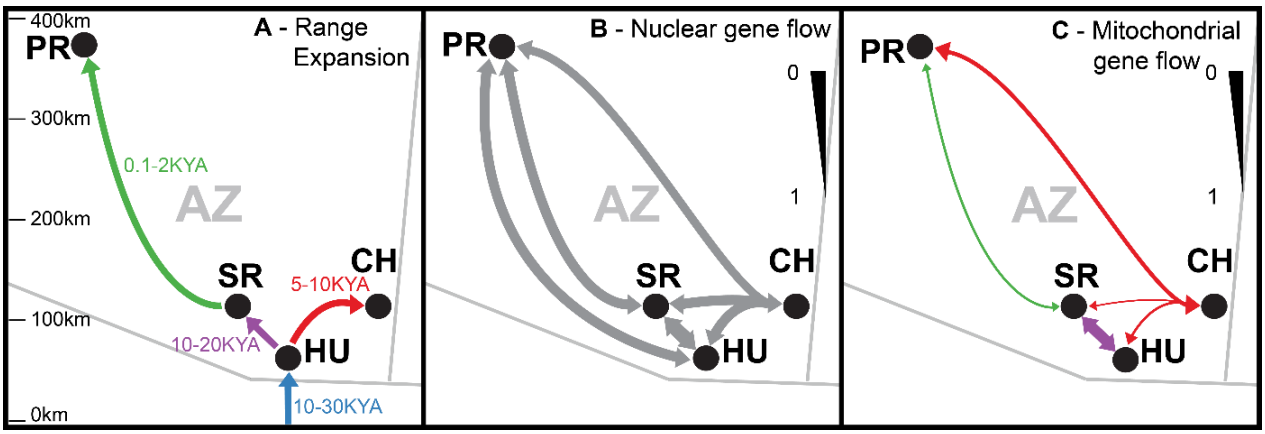
81 To characterize how *D. innubila* came to inhabit its current range, we collected flies from four Sky island
82 locations across Arizona: Chiricahuas (CH, 81 flies), Huachucas (HU, 48 flies), Prescott (PR, 84 flies) and
83 Santa Ritas (SR, 67 flies) (Figure 1). Interestingly, previous surveys mostly failed to collect *D. innubila*
84 north of the Madrean archipelago in Prescott (DYER AND JAENIKE 2005). We easily sampled from that
85 location, suggesting a possible recent invasion (though we were also unable to collect *D. innubila* in the
86 exact locations previously sampled) (DYER AND JAENIKE 2005). If this was a recent colonization event, it
87 could be associated with the changing climate of the area leading to conditions more accommodating to *D.*
88 *innubila*, despite over 300 kilometers of geographic isolation (Figure 1).

89 To determine when *D. innubila* established each population and rates of migration between
90 locations, we isolated and sequenced the DNA from our sampled *D. innubila* populations and characterized
91 genomic variation. We then examined the population structure and changes in demographic history of *D.*
92 *innubila* using silent polymorphism in Structure (FALUSH *et al.* 2003) and StairwayPlot (LIU AND FU 2015).
93 We find all sampled populations have a current estimated effective population size (N_e) of ~1,000,000
94 individuals and an ancestral N_e of ~4,000,000 individuals, though all experience a bottleneck about 100,000
95 years ago to an N_e of 10,000-20,000 (Figure 1, Supplementary Figure 1A & B). This bottleneck coincides
96 with a known glaciation period occurring in Arizona (SURVEY 2005). Each surveyed population then
97 appears to go through separate population expansions between one and thirty thousand years ago, with
98 populations settling from south to north (Figure 1A, Supplementary Figure 1A & B). Specifically, while
99 the HU population appears to have settled first (10-30 thousand years ago), the PR population was settled
100 much more recently (200-2000 years ago). This, in combination of the absence of *D. innubila* in PR until

101 ~2016 sampling suggests recent northern expansion of *D. innubila* (Figure 1). Also note that StairwayPlot
102 (LIU AND FU 2015) has estimated large error windows for PR, meaning the invasion could be more recent
103 or ancient than the 200-2000 year estimate.

104 Given the geographic isolation between populations, we expected to find a corresponding signature
105 of population differentiation among the populations. Using Structure (FALUSH *et al.* 2003), we find
106 surprisingly little population differentiation between locations for the nuclear genome (Supplementary
107 Figure 1C) but some structure by location for the mitochondrial genome (Supplementary Figure 1D),
108 consistent with previous findings (DYER 2004; DYER AND JAENIKE 2005). Together these suggest that there
109 is still consistent gene flow between populations potentially via males.

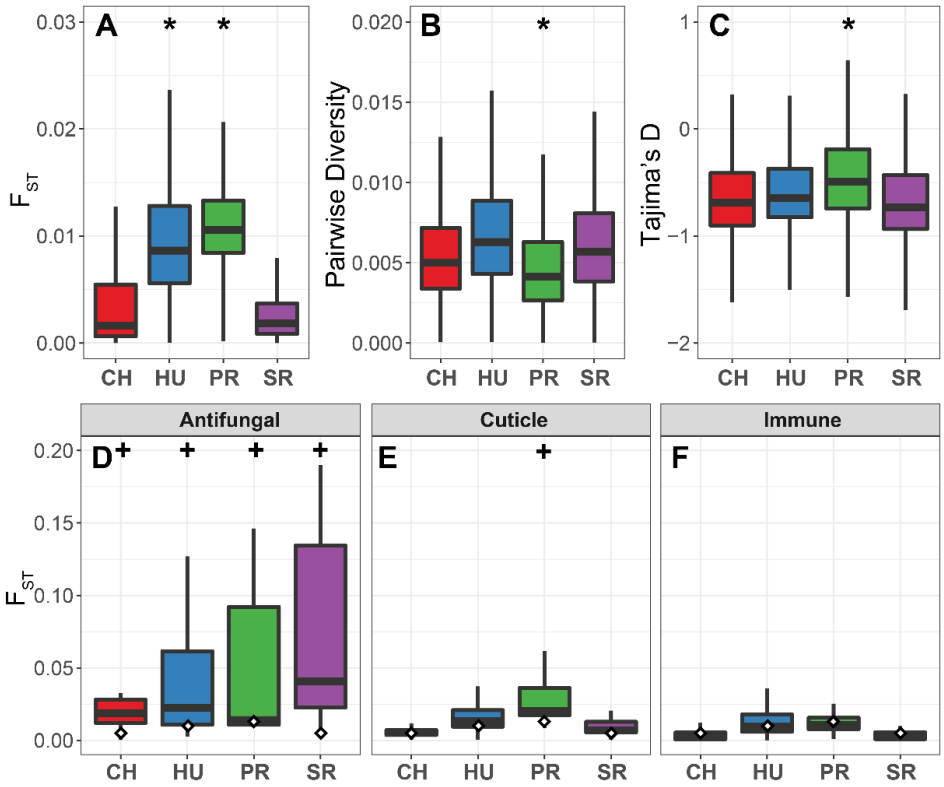
110 **Figure 1: A.** Schematic of the range expansion of *D. innubila* and DiNV based on StairwayPlot results
111 across the four sample locations in Arizona (AZ), Chiracahua's (CH), Huachucas (HU), Prescott (PR) and
112 Santa Ritas (SR). **B** and **C.** Summary of Structure/Fst results for **B.** autosomal polymorphism and **C.**
113 mitochondrial polymorphism. Thickness of arrows in **B** and **C** indicates the median of F_{ST} for genes in
114 each category, with 1 indicating completely isolated populations and 0 indicating complete gene flow.
115 Light grey lines show the Arizona border.

116 

117 We also calculated the amount of differentiation between each population and all other populations,
118 as the fixation index, F_{ST}, using the total polymorphism across the Muller elements (*Drosophila*
119 chromosomes) (WEIR AND COCKERHAM 1984). F_{ST} appears to be generally low across the genes in the *D.*
120 *innubila* genome (Figure 1B, total median = 0.00567), consistent with nuclear gene flow between
121 populations (WEIR AND COCKERHAM 1984). In contrast, there is higher F_{ST} between mitochondrial
122 genomes (Figure 1C). Both nuclear and mitochondrial results are consistent with the Structure/StairwayPlot
123 results. However, consistent with a more recent population contraction upon migration into PR, F_{ST} of the
124 nuclear genome is significantly higher in PR (Figure 2A, GLM t-value = 93.728, *p*-value = 2.73e-102).
125 Though PR F_{ST} is still extremely low genome-wide (PR median = 0.0105), with some outliers on Muller

element B like other populations (Supplementary Figure 2). We also calculated the population genetic statistics pairwise diversity and Tajima's D for each gene using total polymorphism (TAJIMA 1989). As expected with a recent population contraction in PR (suggesting recent migration and establishment in a new location), pairwise diversity is significantly lower (Figure 2B, GLM t-value = -19.728, *p*-value = 2.33e-86, Supplementary Table 2) and Tajima's D is significantly higher than all other populations (Figure 2C, GLM t-value = 4.39, *p*-value = 1.15e-05, Supplementary Table 2). This suggests that there is also a deficit of polymorphism in general in PR, consistent with a more recent population bottleneck, removing rare alleles from the population (Figure 2C, Supplementary Figure 3). Conversely, the other populations show a genome wide negative Tajima's D, consistent with a recent demographic expansion (Supplementary Figure 3).

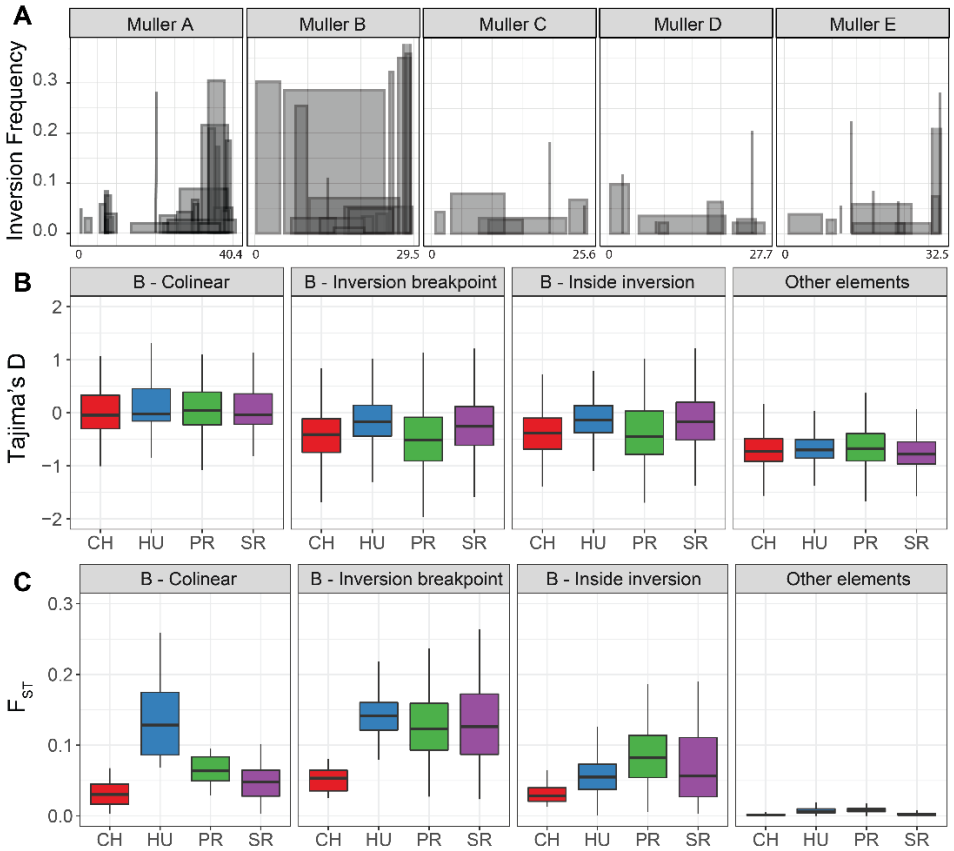
Figure 2: Summary statistics for each population. A. Distribution of F_{ST} across genes for each population versus all other populations. **B.** Distribution of pairwise diversity for each population. **C.** Distribution of Tajima's D for each population. **D.** F_{ST} distribution for Antifungal associated genes for each population. **E.** F_{ST} distribution for cuticular proteins for each population. **F.** F_{ST} distribution for all immune genes (excluding antifungal genes). In A, B & C all cases significant differences from CH are marked with an * and outliers are removed for ease of visualization. In D, E & F, significant differences from the genome background in each population are marked with a + and white diamond mark the whole genome average of F_{ST} for each population.



Population structure in the D. innubila genome is associated with segregating inversions

As mentioned previously, F_{ST} is significantly higher on Muller element B compared to all other elements in all populations (Supplementary Figure 2, GLM t-value = 30.02, *p*-value = 3.567e-56). On Muller element B, regions of elevated F_{ST} are similar in each population (Supplementary Figure 2). Additionally, Muller element B has elevated Tajima's D compared to all other Muller elements (Supplementary Figure 3), suggesting some form of structured population unique to Muller element B (GLM t-value = 10.402, *p*-value = 2.579e-25). We attempted to identify if this elevated structure is due to chromosomal inversions, comparing F_{ST} of a region to the presence or absence of inversions across windows (using only inversions called by both Delly and Pindel (YE *et al.* 2009; RAUSCH *et al.* 2012)). We find several inversions across the genome at appreciable frequencies (89 total above 1% frequency), of which, 37 are found on Muller element B (spread evenly across the entire chromosome) and 22 are found at the telomeric end of Muller element A (Figure 3A). The presence of an inversion over a region of Muller element B is associated with higher F_{ST} in these regions (Figure 3A, Wilcoxon Rank Sum test W = 740510, *p*-value = 0.0129), though these inversions are not unique or even at different frequencies in specific populations ($F_{ST} < 0.22$, χ^2 test for enrichment in a specific population *p*-value > 0.361 for all inversions). Genes within 10kbp of an inversion breakpoint have significantly higher F_{ST} than outside the inverted regions consistent with findings in other species (Figure 3C, GLM t-value = 7.702, *p*-value = 1.36e-14) (MACHADO *et al.* 2007; NOOR *et al.* 2007), while inside inverted regions show no difference from outside (Figure 3C, GLM t-value = -0.178, *p*-value = 0.859). However, all regions of Muller element B have higher F_{ST} than the other Muller elements (Figure 3C, outside inversions Muller element B vs all other chromosomes: GLM t-value = 7.379, *p*-value = 1.614e-13), suggesting some chromosome-wide force drives the higher F_{ST} and Tajima's D. Given that calls for large inversions in short read data are often not well supported (CHAKRABORTY *et al.* 2017) and the apparently complex nature of the Muller element B inversions (Figure 3A), we may not have correctly identified the actual inversions and breakpoints on the chromosome. Despite this, our results do suggest a link between the presence of inversions on Muller element B and elevated differentiation in *D. innubila* and that this may be associated with local adaptation.

Figure 3: Summary of the inversions detected in the *Drosophila innubila* populations. **A.** Location and frequency in the total population of segregating inversions at higher than 1% frequency and greater than 100kbp. **B.** Tajima's D and **C.** F_{ST} for genes across Muller element B, grouped by their presence under an inversion, outside of an inversion, near the inversion breakpoints (within 10kbp) or on a different Muller element.



Evidence for local adaptation in each population

Though F_{ST} is low across most of the genome in each population, there are several genomic regions with elevated F_{ST}. In addition to the entirety of Muller element B, there are narrow chimneys of high F_{ST} on Muller elements D and E (Figure 4, Supplementary Figure 2). We attempted to identify whether any gene ontology groups have significantly higher F_{ST} than the rest of the genome. We find that the genes in the upper 2.5th percentile for F_{ST} are enriched for antifungal genes in all populations, these genes are distributed across the genome and so not all under one peak of elevated F_{ST} (Supplementary Table 3, GO enrichment = 16.414, *p*-value = 1.61e-10). Interestingly, this is the only immune category with elevated F_{ST} (Figure 2F), with most of the immune system showing no divergence between populations (Figure 2F, Supplementary Figure 4). This might suggest that most pathogen pressures are consistent among populations except for fungal pathogen pressure which may be more variable.

Another gene ontology category with significantly higher F_{ST} is cuticle genes (Figure 2E, Supplementary Table 3, GO enrichment = 5.03, *p*-value = 8.68e-08), which could be associated with differences in the environment between locations (toxin exposure, humidity, *etc*.). Consistent with this result, the peak of F_{ST} on Muller element D (Figure 5, Muller element D, 11.56-11.58Mb) is composed of exclusively cuticle development proteins (e.g. Cpr65Au, Cpr65Av, Lcp65Ad) with elevated F_{ST} in these

genes in both the SR and HU populations as well as PR (Figure 5), suggesting that they may be adapting to differing local conditions in those populations.

Two other clear peaks on Muller element E are also composed related genes. Interestingly, there appear to be three regions of the *D. innubila* genome with translocated mitochondrial genes (Figure 5). The first peak (Muller element E, 11.35-11.4Mb) is composed exclusively of one of these translocated mitochondrial regions with 3 mitochondrial genes (including cytochrome oxidase II). The second peak (Muller element E, 23.60-23.62Mb) contains four other mitochondrial genes (including cytochrome oxidase III and ND5) as well as genes associated with nervous system activity (such as *Obp93a* and *Obp99c*). We find no correlation between coverage of these regions and mitochondrial copy number (Supplementary Table 1, Pearsons' correlation t-value = 0.065, *p*-value = 0.861), so this elevated F_{ST} is probably not an artefact of mis-mapping reads. However, we do find these regions have elevated copy number compared to the rest of the genome (Supplementary Figure 5, GLM t-value = 9.245, *p*-value = 3.081e-20), and so this elevated divergence may be due to collapsed paralogs. These insertions of mtDNA are also found in *D. falleni* and are diverged from the mitochondrial genome, suggesting ancient transpositions. The nuclear insertions of mitochondrial genes are also enriched in the 97.5th percentile for F_{ST} in HU and PR, when looking at only autosomal genes (Supplementary Table 3, GO enrichment = 4.53, *p*-value = 3.67e-04). Additionally, several other energy metabolism categories are in the upper 97.5th percentile in CH. Overall these results suggests a potential divergence in the metabolic needs of each population, and that several mitochondrial genes may have found a new function in the *D. innubila* genome and may be diverging due to differences in local conditions. Alternatively, given the male-killing *Wolbachia* parasitizing *D. innubila* (DYER 2004), it is possible the mitochondrial translocations contain functional copies of mitochondrial genes that can efficiently respond to selection unlike their still mtDNA-linked paralogs.

There has been considerable discussion over the last several years about the influence of demographic processes and background selection on inference of local adaptation (CUTTER AND PAYSEUR 2013; CRUICKSHANK AND HAHN 2014; HOBAN *et al.* 2016; MATTHEY-DORET AND WHITLOCK 2018). In contrast to F_{ST} which is a relative measure of population differentiation, D_{XY} is an absolute measure that may be less sensitive to other population-level processes (NEI 1987; CRUICKSHANK AND HAHN 2014). In our data, windows with peaks of elevated F_{ST} also have peaks of D_{XY} in all pairwise comparisons (Figure 4, Supplementary Figure 6), and F_{ST} and D_{XY} are significantly correlated (GLM R^2 = 0.823, t-value = 11.371, *p*-value = 6.33e-30), consistent with local adaptation. The upper 97.5th percentile for D_{XY} is enriched for chorion proteins in all pairwise comparisons and antifungal proteins for all comparisons involving PR (Supplementary Table 5). We also find peaks of elevated pairwise diversity exclusively on the mitochondrial translocations (Supplementary Figure 6), suggesting unaccounted for variation in these genes which is consistent with duplications detected in these genes (RASTOGI AND LIBERLES 2005). This

227 supports the possibility that unaccounted for duplications may be causing the elevated F_{ST}, D_{XY} and pairwise
228 diversity (Supplementary Figure 5 & 6). We find no evidence for duplications in the antifungal, cuticle or
229 chorion proteins, suggesting the elevated F_{ST} and D_{XY} is likely due to local adaptation (Figure 4,
230 Supplementary Figure 5).

231 Recent adaptation often leaves a signature of a selective sweep with reduced polymorphism near
232 the site of the selected variant. We attempted to identify selective sweeps in each population using
233 Sweepfinder2 (HUBER *et al.* 2016). There was no evidence of selective sweeps overlapping with genes with
234 elevated F_{ST} (Supplementary Figure 7A, χ^2 test for overlap of 97.5th percentile windows $\chi^2 = 1.33$ *p*-value
235 = 0 .249) but there was one extreme peak in PR on Muller D (Supplementary Figure 7B, Muller D, 18.75-
236 19Mb). This peak was also found in all other populations though not on the same scale. The center of this
237 peak is just upstream of the cuticle protein *Cpr66D*, in keeping with the suggestion of local adaptation of
238 the cuticle in all populations, with the strongest signal in PR. This sweep is also upstream of four chorion
239 proteins (*Cp15, Cp16, Cp18, Cp19*) and covers (within 10kbp of the sweep center) several cell organization
240 proteins (*Zasp66, Pex7, hairy, Prm, Fhos*). These chorion proteins also have significantly elevated D_{XY}
241 compared to other genes within 50kbp (Wilcoxon Rank Sum W = 45637000, *p*-value = 0.0158) and are
242 under a chimney of elevated D_{XY} in all comparisons (Supplementary Figure 6 & 7), consistent with recent
243 selection of population specific variants. We also find evidence of several selective sweeps in the telomere
244 of the X chromosome (Muller A, 39.5-40.5Mb), among several uncharacterized genes. Given the
245 suppression of recombination in the heterochromatic portions of chromosomes, we would expect evidence
246 for several selective sweeps for even weakly positively selected variants, as is also seen in the non-
247 recombining Muller F (Supplementary Figure 7).

248 **Figure 4:** Comparison of estimated statistics across the *D. innubila* genome for the Prescott (PR)
249 population. Values are as follows: the average pairwise divergence per gene (D_{XY}), the population fixation
250 index per genes (F_{ST}), within population pairwise diversity per genes, Compositive Likelihood Ratio (CLR)
251 per SNP calculated using Sweepfinder2 and within population average Tajima's D per gene.

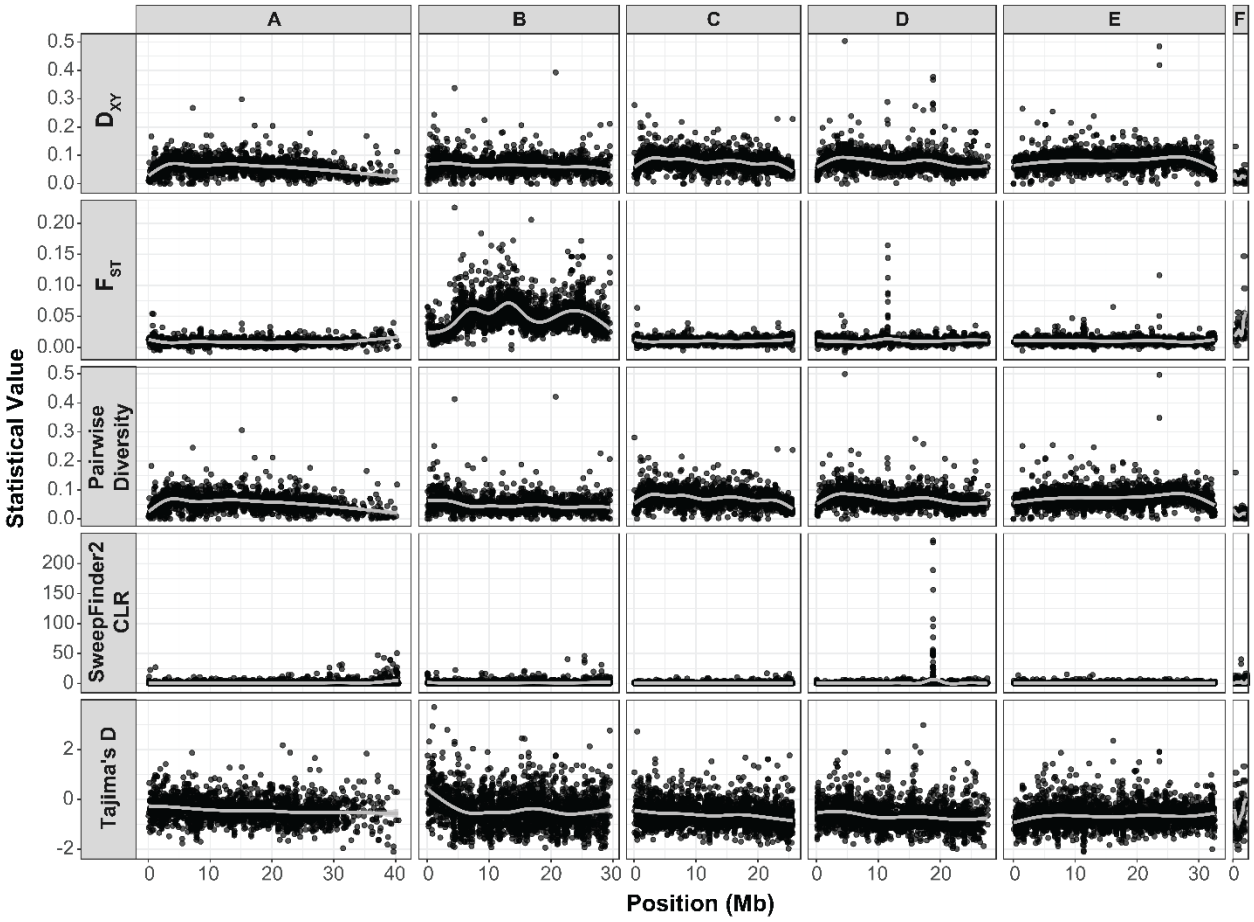
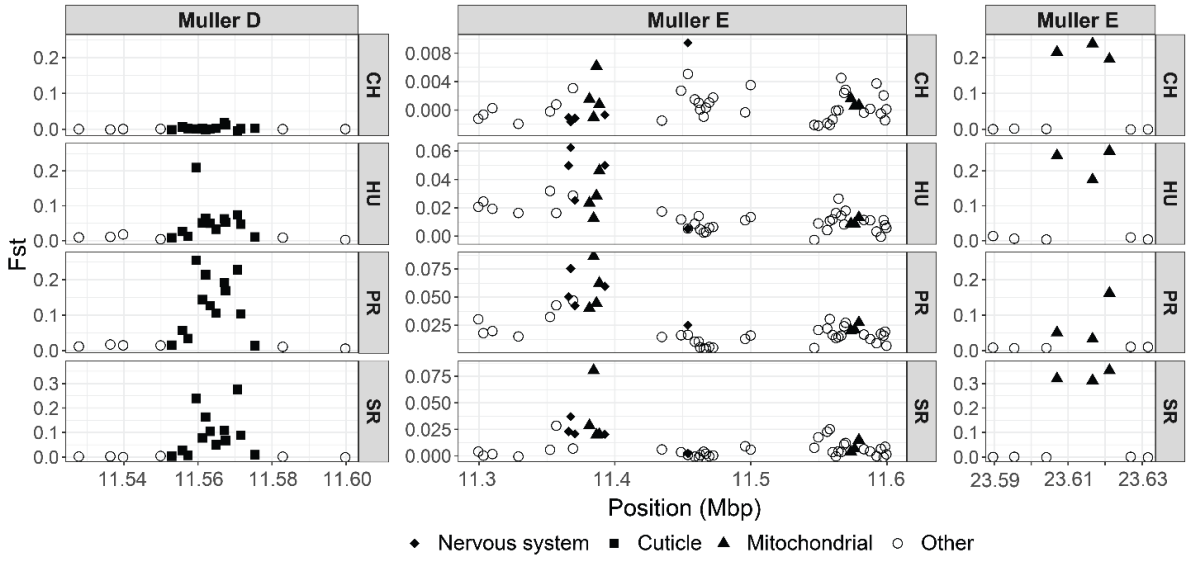


Figure 5: Gene-wise F_{ST} showing regions of elevated divergence between populations for each population. Plot shows F_{ST} for each gene in these regions to identify the causal genes. Genes with noted functions (cuticle development or mitochondrial translocations) are shown by point shape. Note the Y-axes are on different scales for each plot.



258 *Evidence for divergence in the X chromosome over time and between sexes*
259 We next compared the samples from the 35 CH males in 2017 to those we sequenced from a CH collection
260 of 38 males in 2001 to identify changes over time between populations (due to elevated F_{ST} because of
261 differences in allele frequencies between populations). We find little differentiation between the two
262 timepoints (median F_{ST} = 0.0004, 99th percentile = 0.0143), and find no significant enrichments (GO *p*-
263 value < 0.05) in the upper 97.5th percentile of F_{ST}. However, we do find divergence in the genes at the
264 telomere of the X chromosome (Supplementary Figure 8A, Muller element A, 35-40.5Mb, median F_{ST} =
265 0.0029). Looking at actual allele frequency differences between time points, the minor allele frequency
266 increases between 2001 and 2017 at the X telomere while most other chromosomes appear to show little
267 change. Interestingly, there is also evidence of recent selective sweeps in the telomere of X (Supplementary
268 Figure 7A). The minor allele frequency has decreased on average on Muller element B between 2001 and
269 2017 (Supplementary Figure 9A). This suggests something else may be influencing allele frequency change
270 on Muller B compared to other autosomes.
271 We also compared the allele frequencies between 2017 male samples versus 2017 female samples.
272 Again, F_{ST} is extremely low genome wide (median F_{ST} = 0.0004, 99th percentile = 0.0501), but we again
273 find a peak of F_{ST} at the telomere of the X chromosome (Supplementary Figure 8B), which we find when
274 comparing all populations sexes and in a total population male versus female comparison. Again, we find
275 no significant enrichments in the 97.5th percentile for F_{ST}, as most of the divergent genes currently have no
276 functional annotation. We also compared the raw allele frequency change of synonymous variants.
277 Strangely, the population minor allele frequency of euchromatic SNPs on the X chromosome are found at
278 higher frequencies in females (Supplementary Figure 9B), while the X telomere SNPs are overrepresented
279 in male samples. These results are consistent when examining each population separately, suggesting sex
280 specific biases in the X chromosome are found in every populations. It is possible that this signal is caused
281 by an ascertainment bias for SNP calling in females, resulting in more accurate SNP calls in one of the
282 sexes in the euchromatin which is not seen in the heterochromatin. Alternatively, the region of the X
283 chromosome with multiple overlapping inversions could be female-biased due to a female driver, resulting
284 in its overrepresentation in females (and an overrepresentation of the alternate variants in males) (BURT
285 AND TRIVERS 2006). Finally, the X chromosome may be adapting to the skewed sex-ratio associated with
286 *D. innubila*'s male-killing *Wolbachia* (KAGEYAMA *et al.* 2009; UNCKLESS 2011b).

287 *Toll-related immune genes are evolving recurrently in D. innubila likely due to strong pathogen pressures*
288 We next sought to identify genes and functional categories showing strong signatures of adaptive evolution,
289 suggesting recurrent evolution as opposed to recent local adaptation. We reasoned that if the population
290 differentiation seen in antifungal genes and cuticle development proteins (Figure 2 & 3, Supplementary

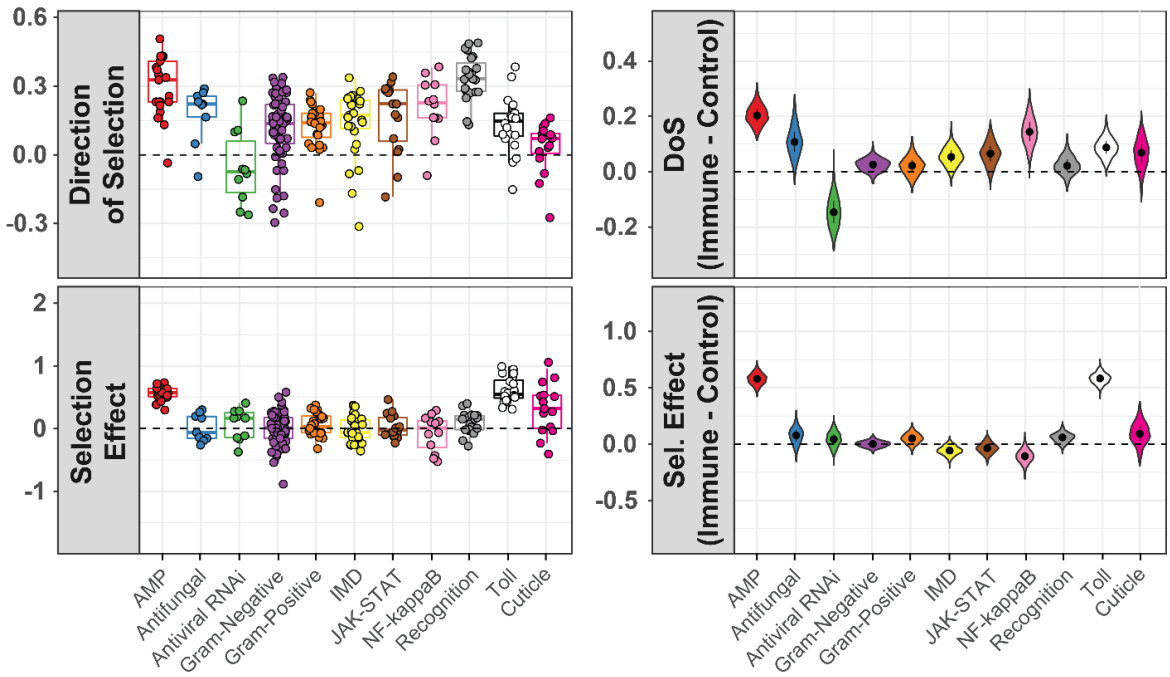
291 Figure 4) was due to local adaptation also acting over longer time periods, we would expect to see signatures
292 of adaptation in those categories. Furthermore, Hill *et al.* used dN/dS-based statistics to show that genes
293 involved in some immune defense pathways were among the fastest evolving genes in the *D. innubila*
294 genome (HILL *et al.* 2019). We also sought to identify what genes are evolving due to recurrent positive
295 selection in *D. innubila* in one or all populations, and if this is associated with environmental factors. To
296 this end we calculated the McDonald-Kreitman based statistic direction of selection (DoS) (Stoletzki and
297 Eyre-Walker, 2011) and SnIPRE selection effect (Eilertson et al., 2012) to identify an excess of selection.
298 We then fit a linear model to identify gene ontology groups with significantly higher DoS or selection effect
299 than expected. In this survey we find cuticle genes and antifungal genes did have some signatures of
300 adaptive evolution (DoS > 0 and selection effect > 0 for 80% of genes in these categories) but as a group
301 showed no significant differences from the background (GLM t-value = 1.128, *p*-value = 0.259,
302 Supplementary Table 4). In fact, we only found two functional groups significantly higher than the
303 background, Toll signaling proteins (GLM t-value = 2.581 *p*-value = 0.00986, Supplementary Table 3) and
304 antimicrobial peptides (AMPs, GLM t-value = 3.66 *p*-value = 0.00025, Supplementary Table 3). In a
305 previous survey we found that these categories were also the only functional groups to have significantly
306 elevated rates of amino acid divergence (HILL *et al.* 2019). These results suggest that this divergence is
307 indeed adaptive.

308 Interestingly, *D. innubila* is burdened by Drosophila innubila nudivirus (DiNV), a Nudivirus that
309 infects 40-50% of individuals in the wild (UNCKLESS 2011a; HILL AND UNCKLESS 2020). A close relative
310 of the virus suppresses Toll-regulated AMPs in *D. melanogaster* (PALMER *et al.* 2018; HILL AND UNCKLESS
311 2020), which might explain why the Toll pathway and AMPs are fast evolving in *D. innubila.* Five AMPs
312 showed consistently positive DoS and selection effect values (which are also among the highest in the
313 genome): four *Bomanins* and *Listericin*. All are AMPs regulated by Toll signaling (and additionally JAK-
314 STAT in the case of *Listericin*) (HOFFMANN 2003; TAKEDA AND AKIRA 2005). *Listericin* has been
315 implicated in the response to viral infection due to its expression upon viral infection (DOSTERT *et al.* 2005;
316 ZAMBON *et al.* 2005; IMLER AND ELFTHERIANOS 2009; MERKLING AND VAN RIJ 2013). For all immune
317 categories, as well as cuticle proteins and antifungal proteins, we find no significant differences between
318 populations for either MK-based statistics, and no significant differences in the distribution of these
319 statistics between populations (GLM t-value < 0.211, p-value > 0.34 for all populations, Supplementary
320 Table 3). Thus, perhaps selection at these loci is ubiquitous and genes flow between populations
321 homogenizes that signature.

322 Mutation rates, efficacy of selection and population structure can vary across the genome, which
323 can confound scans for selection (CHARLESWORTH *et al.* 2003; STAJICH AND HAHN 2005).To work around
324 this, we employed a control-gene resampling approach to identify the average difference from the

325 background for each immune category (CHAPMAN *et al.* 2019). Consistent with our results previous results,
326 we find no signatures of recurrent positive selection in antifungal genes (Supplementary Figure 10, 61%
327 resamples > 0) or cuticle genes (Figure 6, 54% resamples > 0) but do again find extremely high levels of
328 positive selection in AMPs (Figure 6, 100% resamples > 0) and Toll signaling genes (Figure 6, 99.1%
329 resamples > 0). Segregating slightly deleterious mutations can bias inference of selection using McDonald-
330 Kreitman based tests (MESSER AND PETROV 2012). To account for this bias, we also calculated asymptotic
331 α for all functional categories across the genome (HALLER AND MESSER 2017). To this end we calculated
332 the asymptotic α for all functional categories across the genome (HALLER AND MESSER 2017). As before,
333 while we find signals for adaptation in antifungal and cuticle proteins (asymptotic α > 0), we find no
334 evidence of higher rates of adaptation than the background (Supplementary Figure 10, permutation test
335 Antifungal *p*-value = 0.243, Cuticle *p*-value = 0.137). Again, the only categories significantly higher than
336 the background are Toll signaling genes (Permutation test *p*-value = 0.033) and AMPs (Permutation test *p*-
337 value = 0.035). Together these results suggest that while genes involved in antifungal resistance and cuticle
338 development are evolving adaptively, it is not recurrent across the whole functional category, instead only
339 occurring in one or two specific genes. Alternatively, the adaptation may be too recent to detect signal using
340 these metrics. Long-term recurrent adaptation appears to be driven by host-pathogen interactions (likely
341 with DiNV (HILL *et al.* 2019)) as opposed to local adaptation.

342 **Figure 6:** McDonald-Kreitman based statistics for immune categories in *D. innubila* and cuticle
343 development. The left two plots show estimated statistics (Direction of Selection and Selection Effect) for
344 each gene. The right two plots show the difference in average statistic (Direction of Selection and Selection
345 Effect) for each gene and a randomly sampled nearby gene.



Discussion

Migration and environmental change can drive adaptation (RANKIN AND BURCHSTED 1992; CHARLESWORTH *et al.* 2003; GILLESPIE 2004; EXCOFFIER *et al.* 2009; PORRETTA *et al.* 2012; WHITE *et al.* 2013). Species with somewhat isolated or divided populations are likely to adapt to their differing local environments. Migration can both facilitate and hinder such adaptation, allowing new variation (including potentially beneficial variants) to be spread between populations and preventing inbreeding depression. Strong migration can also import locally nonadaptive variants and prevent the fixation of the most fit variants in local populations. We sought to examine the extent that these processes take place in a species of *Drosophila* found across four forests separated by large expanses of desert.

We characterized the phylogeographic history of four populations of *Drosophila innubila*, a mycophagous species endemic to the Arizonan Sky islands using whole genome resequencing of wild-caught individuals. *D. innubila* expanded into its current range during or following the previous glacial maximum (Figure 1, Supplementary Figure 1). We find some evidence of local adaptation, primarily in the cuticle development genes and antifungal immune genes (Figure 2, Supplementary Figure 2). Interestingly, there is very little support for population structure across the nuclear genome (Figures 1 & 2, Supplementary Figures 1 & 2), including in the repetitive content (Supplementary Figure 11), but some evidence of population structure in the mitochondria, as found previously in *D. innubila* (DYER AND JAENIKE 2005). This suggests that if gene flow is occurring, it could be primarily males migrating, as is seen in other non-*Drosophila* species (RANKIN AND BURCHSTED 1992; SEARLE *et al.* 2009; MA *et al.* 2013; AVGAR AND FRYXELL 2014). Based on the polymorphism data available, coalescent times are not deep, and given our

367 estimated population history, this suggests that variants aren't ancestrally maintained and are instead
368 transmitted through migration between locations (CHARLESWORTH *et al.* 2003).

369 Segregating inversions are often associated with population structure and could explain the
370 abnormalities seen on Muller element B here (Supplementary Figures 2-5). Our detection of several putative
371 segregating inversions on Muller element B relative to all other chromosomes (Figure 3A) supports this
372 assertion. However, few of the putative inversions support this hypothesis, in that all are large and common
373 inversions characterized in all populations, suggesting the inversions are not driving the elevated F_{ST}. We
374 suspect that the actual causal inversions may not have been characterized due to the limitations of detecting
375 inversions in repetitive regions with short read data (MARZO *et al.* 2008; CHAKRABORTY *et al.* 2017). The
376 elevated F_{ST} could also be caused by other factors, such as extensive duplication and divergence on Muller
377 element B being misanalysed as just divergence. In fact, the broken and split read pairs used to detect
378 inversions are very similar to the signal used to detect duplications (YE *et al.* 2009; RAUSCH *et al.* 2012;
379 CHEN *et al.* 2016), suggesting some misidentification may have occurred. If a large proportion of Muller B
380 was duplicated, we would see elevated mean coverage of Muller element B in all strains compared to other
381 autosomes, which is not the case (Supplementary Table 1). Further study is necessary to disentangle if
382 inversions or other factors are causing this elevated F_{ST} and the selective and/or demographic pressures
383 driving this differentiation. However, it is worth noting that *D. pseudoobscura* segregates for inversions on
384 Muller element C and these segregate by population in the same Sky island populations (and beyond) as
385 the populations described here (DOBZHANSKY AND STURTEVANT 1937; DOBZHANSKY *et al.* 1963; FULLER
386 *et al.* 2016). Thus, the inversion polymorphism among populations is a plausible area for local adaptation
387 and may provide an interesting contrast to the well-studied *D. pseudoobscura* inversions.

388 We find very few signatures of divergence between samples from 2001 and 2017 (Supplementary
389 Figure 8). Though the environment has changed in the past few decades, there may have been little impact
390 on the habitat of *D. innubila* in the Chiricahuas, resulting in few changes in selection pressures in this short
391 period of time, unlike most bird and mammal's species in the same area (COE *et al.* 2012). Interestingly,
392 there was an extensive forest fire in 2011 which could plausibly have been a strong selective force but we
393 see no genome-wide signature of such (ARECHEDERRA-ROMERO 2012). Alternatively, seasonal
394 fluctuations in allele frequencies may swamp out directional selection. Excessive allele frequency change
395 is limited to a few genes with no known association to each other, and little overlap with the diverging
396 genes between populations. Some of the genes with elevated F_{ST} (and differing in allele frequency) between
397 time points overlap with divergent genes between sexes, primarily at the telomere of the X chromosome
398 (Muller element A, Supplementary Figure 8). In fact, F_{ST} is significantly correlated on Muller element A
399 between the two surveys (Pearson's correlation t = 82.411, p-value = 1.2e-16), even with the 2001-2017
400 survey only considering male samples, supporting an association between the factors driving divergence

401 between sexes and over time. Given the sex bias of SNPs in this region, this could suggest that a selfish
402 factor with differential effects in the sexes is located on the X chromosome near the telomere (BURT AND
403 TRIVERS 2006). Often these selfish elements also accumulate inversions to prevent the breakdown of
404 synergistic genetic components (BURT AND TRIVERS 2006), and the Muller A telomere appears to have
405 accumulated several inversions (Figure 3A). However, populations of *D. innubila* are already female-biased
406 due to the male-killing *Wolbachia* infection found in 30-35% of females (DYER 2004; DYER AND JAENIKE
407 2005; JAENIKE AND DYER 2008). Thus *D. innubila* could be simultaneously parasitized by both the male-
408 killing *Wolbachia* and a selfish X chromosome. Alternatively, the strong signals associated with the
409 telomere of the X could be a signature of selection related to the *Wolbachia* infection (UNCKLESS 2011b).
410 Ours is one of few studies that sequences individual wild-caught *Drosophila* and therefore avoids
411 several generations of inbreeding that would purge recessive deleterious alleles (GILLESPIE 2004; MACKAY
412 *et al.* 2012; POOL *et al.* 2012). The excess of putatively deleterious alleles harkens back to early studies of
413 segregating lethal mutations in populations as well as recent work on humans (DOBZHANSKY *et al.* 1963;
414 MARINKOVIC 1967; DOBZHANSKY AND SPASSKY 1968; WATANABE *et al.* 1974; GAO *et al.* 2015).
415 To date, most of the genomic work concerning the phylogeography and dispersal of different
416 *Drosophila* species has been limited to the *melanogaster* supergroup (POOL *et al.* 2012; POOL AND
417 LANGLEY 2013; BEHRMAN *et al.* 2015; LACK *et al.* 2015; MACHADO *et al.* 2015), with some work in other
418 *Sophophora* species (FULLER *et al.* 2016). This limits our understanding of how non-commensal species
419 disperse and behave, and what factors seem to drive population demography over time. Here we have
420 glimpsed into the dispersal and history of a species of mycophageous *Drosophila* and found evidence of
421 changes in population distributions potentially due to the changing climate (SURVEY 2005) and population
422 structure possibly driven by segregating inversions and selfish elements. Because many species have
423 recently undergone range changes or expansions (EXCOFFIER *et al.* 2009; PORRETTA *et al.* 2012; WHITE *et*
424 *al.* 2013), we believe examining how this has affected genomic variation is important for population
425 modelling and even for future conservation efforts (EXCOFFIER *et al.* 2009; COE *et al.* 2012).

426 **Methods**

427 *Fly collection, DNA isolation and sequencing*

428 We collected wild *Drosophila* at the four mountainous locations across Arizona between the 22nd of August
429 and the 11th of September 2017: the Southwest research station in the Chiricahua mountains (CH, ~5,400
430 feet elevation, 31.871 latitude -109.237 longitude, 96 flies), in Prescott National Forest (PR, ~7,900 feet
431 elevation, 34.586 latitude -112.559 longitude, 96 flies), Madera Canyon in the Santa Rita mountains (SR,
432 ~4,900 feet elevation, 31.729 latitude -110.881 longitude, 96 flies) and Miller Peak in the Huachuca
433 mountains (HU, ~5,900 feet elevation, 31.632 latitude -110.340 longitude, 53 flies) (COE *et al.* 2012). Baits

434 consisted of store-bought white button mushrooms (*Agaricus bisporus*) placed in large piles about 30cm in
435 diameter, with at least 5 baits per location. We used a sweep net to collect flies over the baits in either the
436 early morning or late afternoon between one and three days after the bait was set. We sorted flies by sex
437 and species at the University of Arizona in Tucson, AZ and flash frozen at -80°C before shipping on dry
438 ice to the University of Kansas in Lawrence KS.

439 We sorted 343 flies (172 females and 171 males) which phenotypically matched *D. innubila.* We
440 then homogenized and extracted DNA using the Qiagen Gentra Puregene Tissue kit (USA Qiagen Inc.,
441 Germantown, MD, USA). We also prepared the DNA of 40 *D. innubila* collected in 2001 from CH. We
442 prepared a genomic DNA library of these 383 DNA samples using a modified version of the Nextera DNA
443 library prep kit (~ 350bp insert size) meant to conserve reagents. We sequenced the libraries on four lanes
444 of an Illumina HiSeq 4000 (150bp paired end) (Supplementary Table 1, Data to be deposited in the NCBI
445 SRA).

446 *Sample filtering, mapping and alignment*

447 We removed adapter sequences using Scythe (BUFFALO 2018), trimmed all data using cutadapt to remove
448 barcodes (MARTIN 2011) and removed low quality sequences using Sickle (parameters: -t sanger -q 20 -l
449 50) (JOSHI AND FASS 2011). We masked the *D. innubila* reference genome, using *D. innubila* TE sequences
450 generated previously and RepeatMasker (parameters: -s -gccalc -gff -lib customLibrary) (SMIT AND
451 HUBLEY 2013-2015; HILL *et al.* 2019). We then mapped the short reads to the masked *D. innubila* genome
452 using BWA MEM (LI AND DURBIN 2009), and sorted and indexed using SAMTools (LI *et al.* 2009).
453 Following mapping, we added read groups, marked and removed sequencing and optical duplicates, and
454 realigned around indels in each mapped BAM file using Picard and GATK
455 (HTTP://BROADINSTITUTE.GITHUB.IO/PICARD ; MCKENNA *et al.* 2010; DEPRISTO *et al.* 2011). We then
456 removed individuals with low coverage of the *D. innubila* genome (less than 5x coverage for 80% of the
457 non-repetitive genome), and individuals we suspected of being misidentified as *D. innubila* following
458 collection due to anomalous mapping. This left us with 280 *D. innubila* wild flies (48 - 84 flies per
459 populations) from 2017 and 38 wild flies from 2001 with at least 5x coverage across at least 80% of the
460 euchromatic genome (Supplementary Table 1).

461 *Nucleotide polymorphisms across the population samples*

462 For the 318 sequenced samples with reasonable coverage, we called SNPs using GATK (MCKENNA *et al.*
463 2010; DEPRISTO *et al.* 2011) which generated a multiple strain VCF file. We then used BCFtools
464 (NARASIMHAN *et al.* 2016) to remove sites with a GATK quality score (a composite PHRED score for
465 multiple samples per site) lower than 950 and sites absent (e.g. sites of low quality, or with 0 coverage)
466 from over 5% of individuals. This filtering left us with 4,522,699 SNPs and small indels across the 168Mbp

467 genome of *D. innubila*. We then removed SNPs found as a singleton in a single population (as possible
468 errors), leaving us with 3,240,198 SNPs. We used the annotation of *D. innubila* and SNPeff (CINGOLANI
469 *et al.* 2012) to identify SNPs as synonymous, non-synonymous, non-coding or another annotation.
470 Simultaneous to the *D. innubila* population samples, we also mapped genomic information from outgroup
471 species *D. falleni* (SRA: SRR8651761) and *D. phalerata* (SRA: SRR8651760) to the *D. innubila* genome
472 and called divergence using the GATK variation calling pipeline to identify derived polymorphisms and
473 fixed differences in *D. innubila*.

474 *Population genetic summary statistics and structure*

475 Using the generated total VCF file with SNPeff annotations, we created a second VCF containing only
476 synonymous polymorphism using BCFtools (NARASIMHAN *et al.* 2016). We calculated pairwise diversity
477 per base, Watterson's theta, Tajima's D (TAJIMA 1989) and F_{ST} (WEIR AND COCKERHAM 1984) (versus all
478 other populations) across the genome for each gene in each population using VCFtools (DANECEK *et al.*
479 2011) and the VCF containing all variants. Using ANGSD to parse the synonymous polymorphism VCF
480 (KORNELIUSSEN *et al.* 2014), we generated synonymous unfolded site frequency spectra for the *D. innubila*
481 autosomes for each population, using the *D. falleni* and *D. phalerata* genomes as outgroups to the *D.*
482 *innubila* genome (HILL *et al.* 2019).

483 We used the population silent SFS with previously estimated mutation rates of *Drosophila*
484 (SCHRIDER et al. 2013), as inputs in StairwayPlot (LIU AND FU 2015), to estimate the effective population
485 size backwards in time for each location.

486 We also estimated the extent of population structure across samples using Structure (FALUSH *et al.*
487 2003), repeating the population assignment for each chromosome separately using only silent
488 polymorphism, for between one and ten populations (k = 1-10, 100000 iterations burn-in, 400000 iterations
489 sampling). Following (FRICHOT *et al.* 2014), we manually assessed which number of subpopulations best
490 fits the data for each *D. innubila* chromosome and DiNV to minimize entropy.

491 *Signatures of local adaptive divergence across D. innubila populations*

492 We downloaded gene ontology groups from Flybase (GRAMATES *et al.* 2017). We then used a gene
493 enrichment analysis to identify enrichments for particular gene categories among genes in the 97.5th
494 percentile and 2.5th percentile for F_{ST}, Tajima's D and Pairwise Diversity versus all other genes
495 (SUBRAMANIAN *et al.* 2005). Due to differences on the chromosomes Muller A and B versus other
496 chromosomes in some cases, we also repeated this analysis chromosome by chromosome, taking the upper
497 97.5th percentile of each chromosome.

498 We next attempted to look for selective sweeps in each population using Sweepfinder2 (HUBER *et*
499 *al.* 2016). We reformatted the polarized VCF file to a folded allele frequency file, showing allele counts for

500 each base. We then used Sweepfinder2 on the total called polymorphism in each population to detect
501 selective sweeps in 1kbp windows (HUBER *et al.* 2016). We reformatted the results and looked for genes
502 neighboring or overlapping with regions where selective sweeps have occurred with a high confidence,
503 shown as peaks above the genomic background. We surveyed for peaks by identifying 1kbp windows in
504 the 97.5th percentile for composite likelihood ratio per chromosome.

505 Using the total VCF with outgroup information, we next calculated Dxy per SNP for all pairwise
506 population comparisons (NEI AND MILLER 1990), as well as within population pairwise diversity and dS
507 from the outgroups, using a custom python script. We then found the average Dxy and dS per gene and
508 looked for gene enrichments in the upper 97.5th percentile, versus all other genes.

509 *Inversions*

510 For each sample, we used Delly (RAUSCH *et al.* 2012) to generate a multiple sample VCF file identifying
511 regions in the genome which are potentially duplicated, deleted or inverted compared to the reference
512 genome. Then we filtered and removed inversions found in fewer than 1% of individuals and with a GATK
513 VCF quality score lower than 200. We also called inversions using Pindel (YE *et al.* 2009) in these same
514 samples and again removed low quality inversion calls. We next manually filtered samples and merged
515 inversions with breakpoints within 1000bp at both ends and significantly overlapping in the
516 presence/absence of these inversions across strains (using a χ^2 test, p-value < 0.05). We also filtered and
517 removed large inversions which were only found with one of the two tools. Using the remaining filtered
518 and merged inversions we estimated the frequency of each inversion within the total population.

519 *Signatures of recurrent selection*

520 We filtered the total VCF with annotations by SNPeff and retained only non-synonymous (replacement) or
521 synonymous (silent) SNPs. We then compared these polymorphisms to the differences identified to *D.*
522 *falleni* and *D. phalerata* to polarize changes to specific branches. Specifically, we sought to determine sites
523 which are polymorphic in our *D. innubila* populations or are substitutions which fixed along the *D. innubila*
524 branch of the phylogeny. We used the counts of fixed and polymorphic silent and replacement sites per
525 gene to estimate McDonald-Kreitman-based statistics, specifically direction of selection (DoS)
526 (MCDONALD AND KREITMAN 1991; SMITH AND EYRE-WALKER 2002; STOLETZKI AND EYRE-WALKER
527 2011). We also used these values in SnIPRE (EILERTSON *et al.* 2012), which reframes McDonald-Kreitman
528 based statistics as a linear model, taking into account the total number of non-synonymous and synonymous
529 mutations occurring in user defined categories to predict the expected number of these substitutions and
530 calculate a selection effect relative to the observed and expected number of mutations (EILERTSON *et al.*
531 2012). We calculated the SnIPRE selection effect for each gene using the total number of mutations on the
532 chromosome of the focal gene. Using FlyBase gene ontologies (GRAMATES *et al.* 2017), we sorted each

533 gene into a category of immune gene or classed it as a background gene, allowing a gene to be classed in
534 multiple immune categories. We fit a GLM to identify functional categories with excessively high estimates
535 of adaptation, considering multiple covariates:

536 $$Statistic \sim Population + Gene\ group + (Gene\ group * Population) + Chromosome$$
537 $$+ Chromosome{:}Position$$

538 We then calculated the difference in each statistic between our focal immune genes and a randomly sampled
539 nearby (within 100kbp) background gene, finding the average of these differences for each immune
540 category over 10000 replicates, based on (CHAPMAN *et al.* 2019).

541 To confirm these results, we also used AsymptoticMK (HALLER AND MESSER 2017) to calculate
542 asymptotic α for each gene category. We generated the non-synonymous and synonymous site frequency
543 spectrum for each gene category, which we then used in AsymptoticMK to calculate asymptotic α and a
544 95% confidence interval. We then used a permutation test to assess if functional categories of interest
545 showed a significant difference in asymptotic α from the rest of categories.

546 **Acknowledgements**
547 This work was completed with helpful discussion from Justin Blumensteil, Joanne Chapman, Richard Glor,
548 Stuart MacDonald, Maria Orive and Caroyln Wessinger. We would especially like to thank Kelly Dyer and
549 Paul Guinsberg for proposing the idea of the manuscript and providing feedback on early sections of the
550 manuscript. We greatly appreciate help provided by John Kelly, for providing scripts to calculate D_{XY} as
551 well as advice on population genetic inference and comments on the manuscript. Collections were
552 completed with assistance from Todd Schlenke and the Southwest Research Station. We thank Brittny Smith
553 and the KU CMADP Genome Sequencing Core (NIH Grant P20 GM103638) for assistance in genome
554 isolation, library preparation and sequencing. This work was supported by a K-INBRE postdoctoral grant
555 to TH (NIH Grant P20 GM103418). This work was also funded by NIH Grants R00 GM114714 and R01
556 AI139154 to RLU.

557 **Supplementary Methods**

558 We used dnaPipeTE (GOUBERT *et al.* 2015) to quantify the extent that repetitive element content differed
559 across the populations. Our approach assumed a genome size of 168Mbp, with the number of randomly
560 sampled reads equal to 1-fold coverage of the genome, resampling each strain 2 times to get the average
561 estimate of each strains TE content. Following TE identification, we grouped sequences by known super-
562 families and compared the proportion of the genome composed of each superfamily across strains in the
563 populations. We also used a reciprocal blast (e-value < 0.00000001) (ALTSCHUL *et al.* 1990) to identify TE
564 families present in each strain.

565 We confirmed TE families shared between the previous RepeatModeler (SMIT AND HUBLEY 2008)
566 annotation of the *D. innubila* reference genome and the dnaPipeTE annotation using blast (e-value < 10e-
567 08) (ALTSCHUL *et al.* 1990). After confirming that they did not differ in content, we called TE insertions in
568 each strain across the genome using PopoolationTE2 (KOFLER *et al.* 2016), then merged the output and
569 calculated the frequency of insertions, grouping by TE order, population and if the insertion was exonic,
570 intronic, non-coding or flanking a gene (500bp up or downstream of start or end). When considering
571 individual TE families in *D. innubila*, we used the RepBase TE names and identifications (BAO *et al.* 2015).

572 **Supplementary Results**

573 We characterized the repetitive content across our samples using dnaPipeTE (GOUBERT *et al.* 2015) and
574 called TE insertions per line using PopoolationTE2 (KOFLER *et al.* 2016). The reference *D. innubila* genome
575 contains 154 different TE families along with varying satellites and simples repeats, with resequenced
576 individuals varying from 4.4% to 38.4% of reads matching repetitive sequences. Strains varied from 1913
577 to 7479 TE insertions per strain in the non-repetitive portion of the genome. Like nuclear polymorphism,
578 we find little population structure by shared TE insertions, though strains do seem to disperse primarily by
579 the number of insertions (Supplementary Figure 11B).

580 Similar to *D. melanogaster* (CHARLESWORTH AND LANGLEY 1989; CHARLESWORTH *et al.* 1997;
581 PETROV *et al.* 2011; KOFLER *et al.* 2012; KOFLER *et al.* 2015), *D. innubila* harbors a significant excess of
582 low frequency TE insertions compared to the SFS of synonymous variants (Supplementary Figure 11A,
583 GLM Count ~ Frequency * SNP or TE, t-value = -16.401, p-value = 1.889e-60), with no difference in the
584 insertion frequency spectra between populations (GLM Count ~ Frequency * TE order * Population, t-
585 value = -0.341, p-value = 0.733). This implies in every population, TE insertions are on average mildly
586 deleterious and removed via purifying selection.

587 Using dnaPipeTE (GOUBERT *et al.* 2015), we find a significantly higher density of RC & TIR
588 elements compared to other repeat orders (Supplementary Figure 11C, t-value = 3.555 p-value = 3.745e-
589 04), consistent with the reference genome (HILL *et al.* 2019). The density of repetitive content is also higher

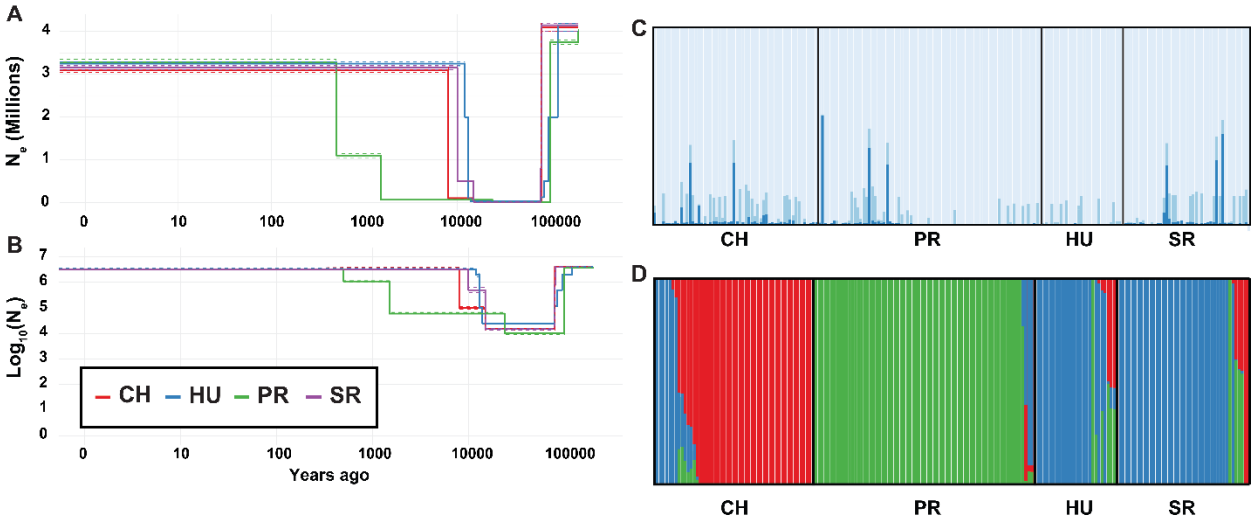
590 genome wide in the CH and PR populations compared to HU and SR (Supplementary Figure 11C & D, t-
591 value = 2.856, p-value = 4.291e-03). This is in keeping with a more recent bottleneck for these species
592 reducing effective population size and efficacy of selection, resulting in bursts of repeat activity with
593 relaxed selection for removal of insertions. These changes are primarily driven by an expansion of simple
594 repeats in the CH population (Supplementary Figure 11D, GLM t-value = 3.978, p-value = 7.31e-05) and
595 an expansion of TIR elements in the PR population (Supplementary Figure 11D, GLM t-value = 3.914, p-
596 value = 9.52e-05). Specifically, we see expansions of the satellite CASAT_HD (GLM t-value = 5.554, p-
597 value = 8.832e-08) and the simple repeat sequences CAACAA, CTC and GTGT in the CH population when
598 compared to all other populations (GLM t-value = 9.204, p-value = 2.555e-17). In the PR population we
599 find significantly higher abundances of a TE families closely related to *Tetris_Dvir* (GLM t-value = 13.641,
600 p-value = 2.889e-32), *Helitron-2N1_DVir* (GLM t-value = 12.381, p-value = 2.789e-28) and *Chapaev3-*
601 *1_PM* (GLM t-value = 11.472, p-value = 1.662e-24) compared to other populations. We do not find any
602 evidence that particular TE orders are more abundant on any one chromosome in *D. innubila* (GLM t-value
603 = 1.854, p-value = 0.633), though do find TEs are at significantly higher insertion densities in the inverted
604 regions of Muller element A than at the regions of the genome (Wilcoxon Rank Sum Test W= 19763, p-
605 value = 0.01488). This suggests the lack of recombination in the inverted region is allowing the
606 accumulation of repetitive content on Muller element A.

607 TE insertions are usually assumed to be at least mildly deleterious (CHARLESWORTH AND LANGLEY
608 1989; PETROV *et al.* 2011). In *D. innubila*, TE density is lower in regions flanking genes or within genes
609 compared to non-coding regions (GLM t-value = -6.538, p-value = 6.23e-11), consistent with the
610 deleterious assumption. However, the frequency of TE insertions was significantly higher in exonic regions
611 compared to introns and UTRs (Supplementary Figure 11A, GLM t-value = 4.040, p-value = 5.34e-05),
612 across all populations, which we may have observed as these are wild caught flies and so may have more
613 recessive deleterious insertions segregating in the population than are seen in inbred samples. Overall the
614 repetitive content in *Drosophila innubila* appears to be mildly deleterious, with TE insertions shared
615 between locations by migration. Despite this there are some major differences in the repeat content of each
616 population, possibly due to the stochastic effect of population bottlenecks.

617 This may have occurred due to a founder effect following the population bottleneck, where a
618 majority of CH founders by chance had a higher proportion of particular satellites or simple repeats
619 (CHARLESWORTH *et al.* 2003), but this is unlikely given the gene flow between populations. Alternatively,
620 the bottleneck could have fixed segregating recessive variation which limits the regulation of repetitive
621 content in the genome, leading to its expansion. However, if this was the case and satellite expansion is
622 even mildly deleterious, we would expect migratory rescue of repeat regulation machinery. A third

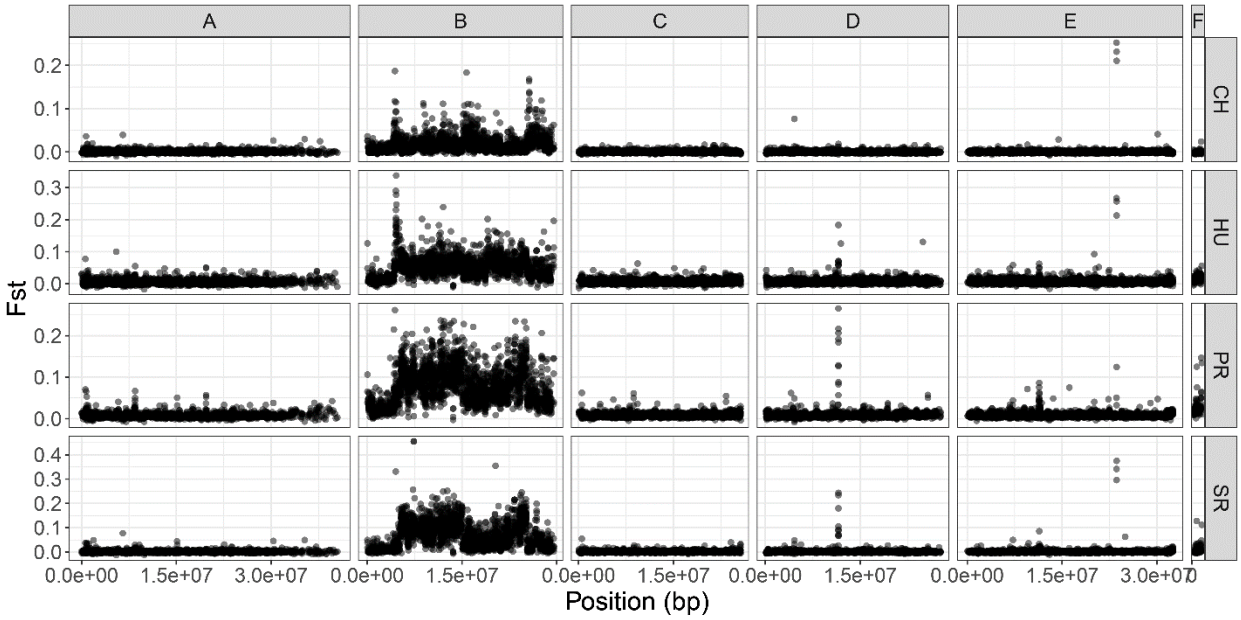
623 possibility is that satellite expansion is associated with local evolutionary dynamics either involved in
624 adaptation or genetic conflict (GARRIDO-RAMOS 2017; LOWER *et al.* 2018).

625 **Supplementary Figure 1: A.** Population size history of *Drosophila innubila* backwards in time for each
626 population. **B.** Population size history on the Log10 scale of *Drosophila innubila* backwards in time for
627 each population. **C.** Results of Structure software (FALUSH *et al.* 2003) for estimating population structure
628 between locations for 100,000 sampled synonymous polymorphisms from all autosomes, with a K=3
629 (estimated optimal K value). Note that this plot summarizes all autosomes (excluding Muller B) and the X
630 chromosome due to very little structure between locations for all chromosomes. **D.** Results of Structure
631 software (FALUSH *et al.* 2003) for estimating population structure between locations for 16
632 polymorphisms on the autosomes, with a K=3 (estimated optimal K value).

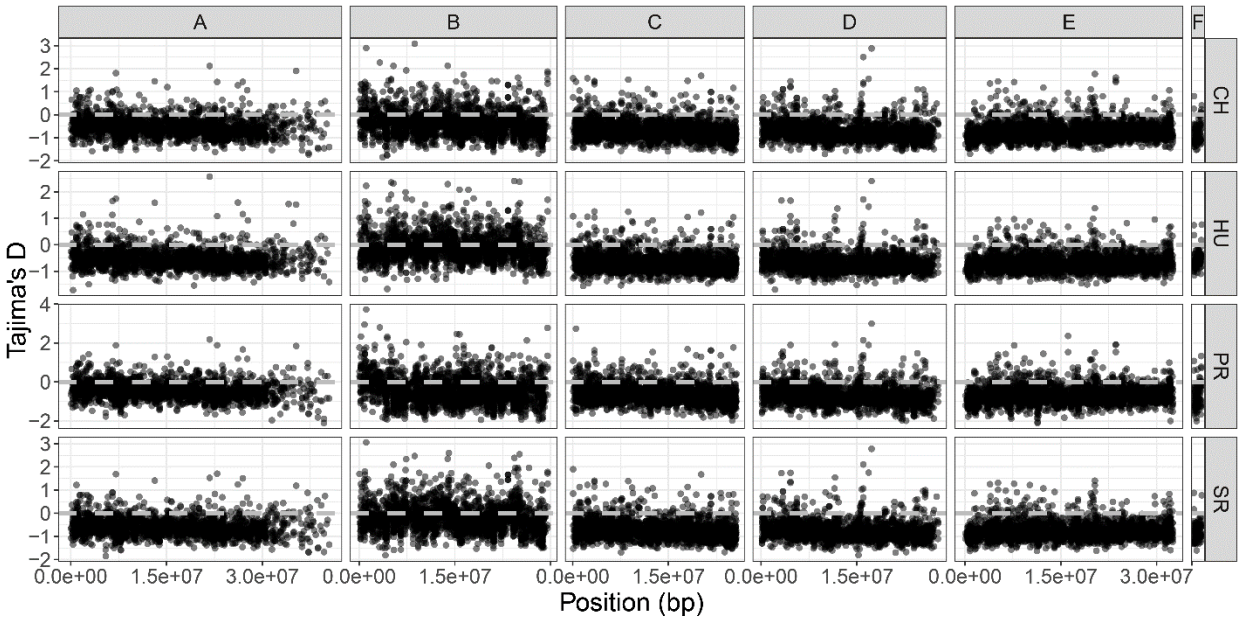


633
634

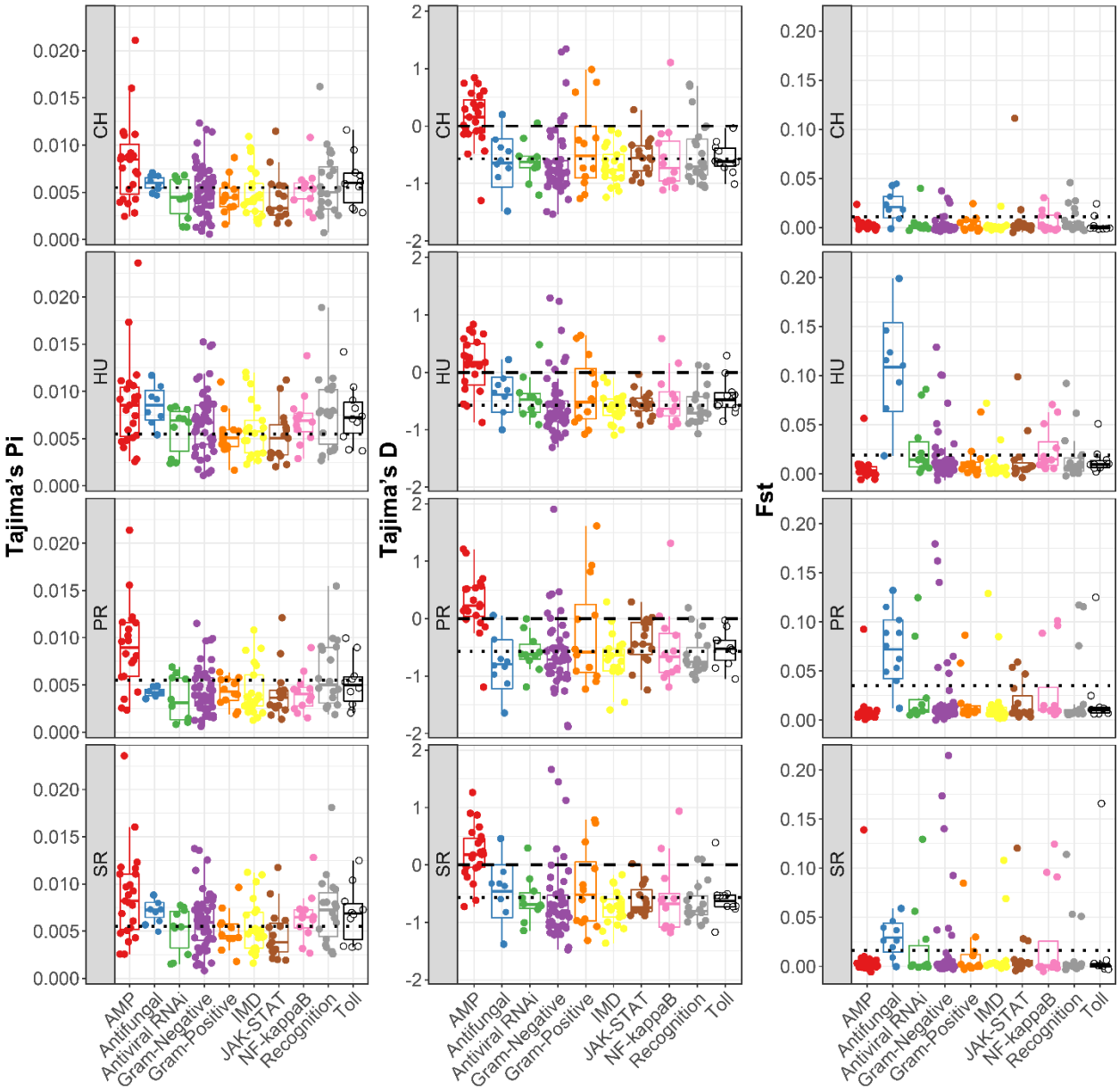
Supplementary Figure 2: Fst by gene across all Muller elements for each population, located by loci (in bp) on the Muller element.



Supplementary Figure 3: Tajima's D by gene across all Muller elements for each population, located by loci (in bp) on the Muller element. The grey dashed line shows a Tajima's D of 0.

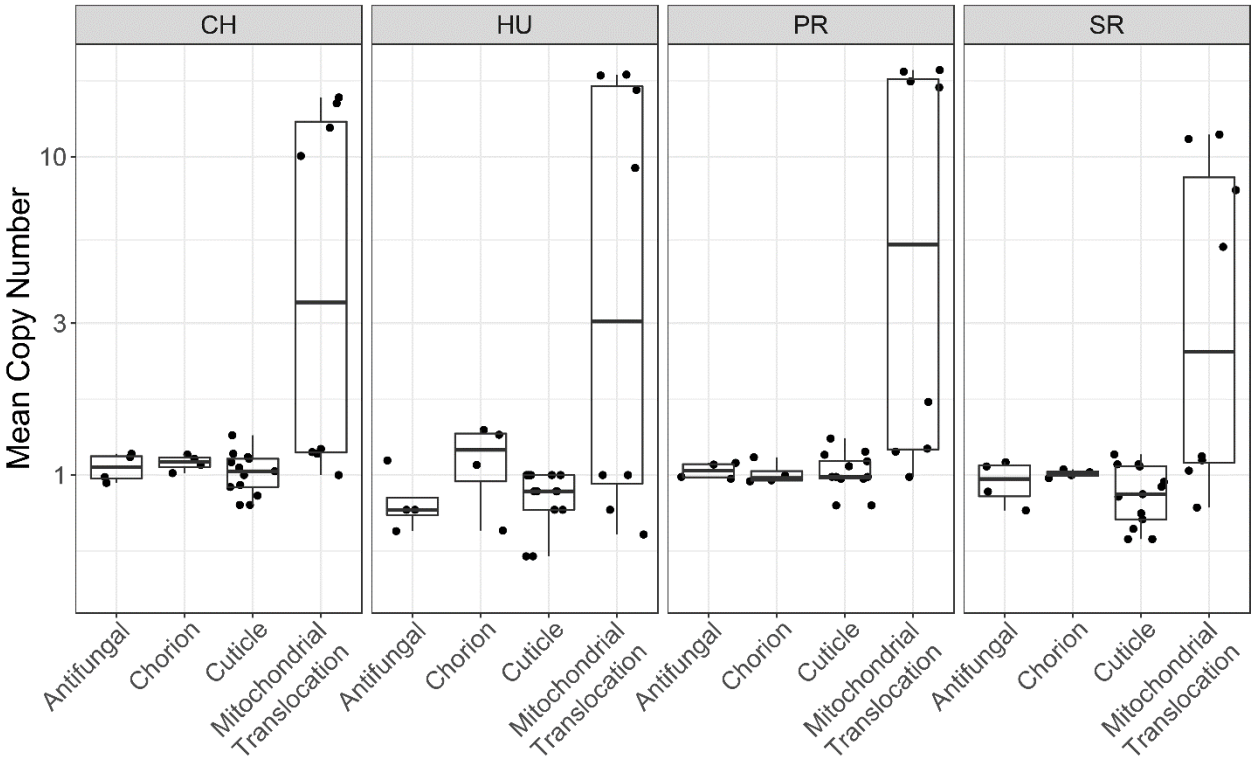


642 **Supplementary Figure 4:** Population genetic statistics (Pairwise diversity, Tajima's D and Fst) for genes
643 in each immune category, for each population. Each plot has a dotted line to show the genomic background
644 statistics for each population. The Tajima's D plot contains a dashed line to show 0.
645



646
647

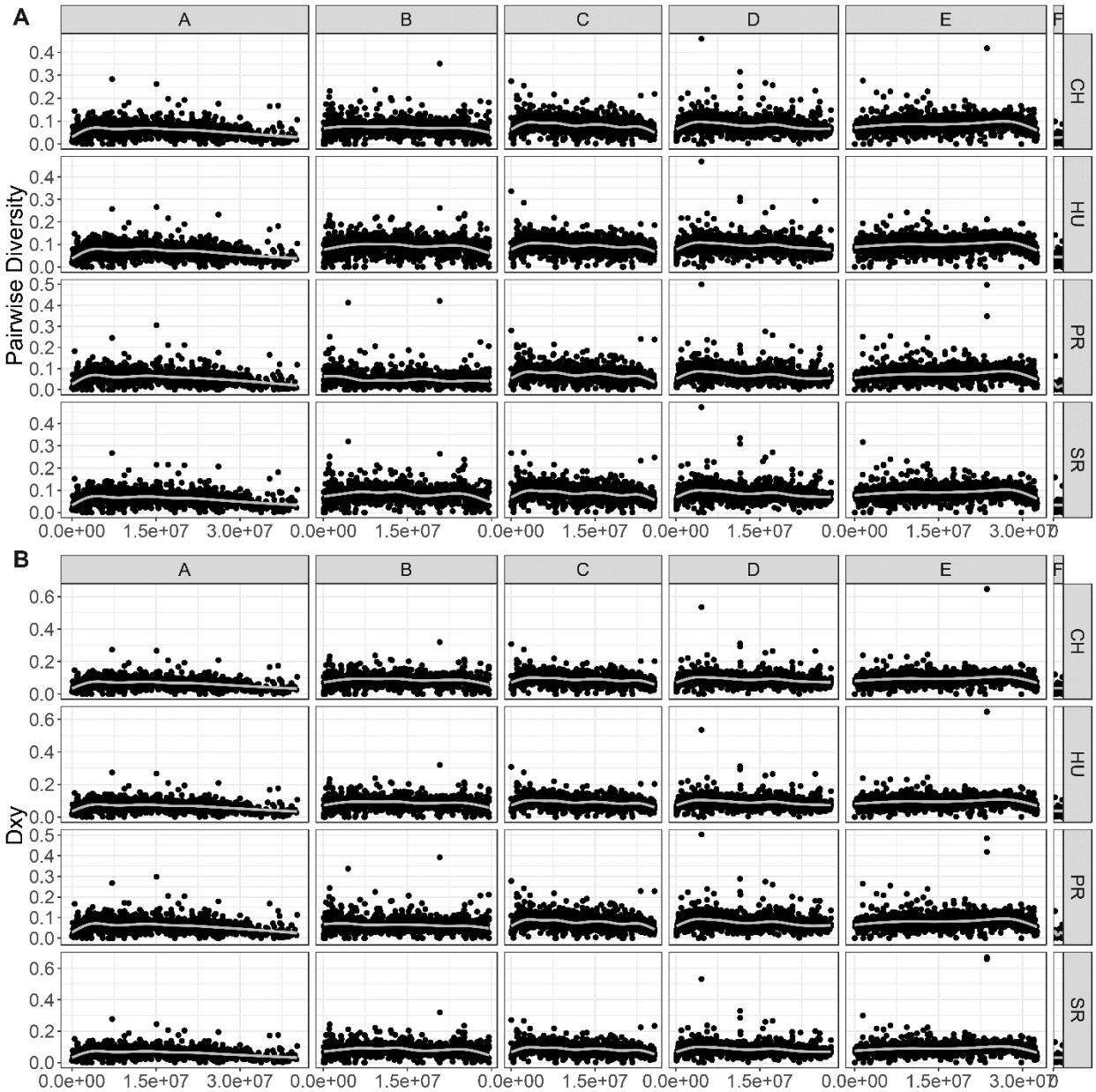
648 **Supplementary Figure 5:** Mean copy number per population for genes of interest in F_{ST} peaks.



649

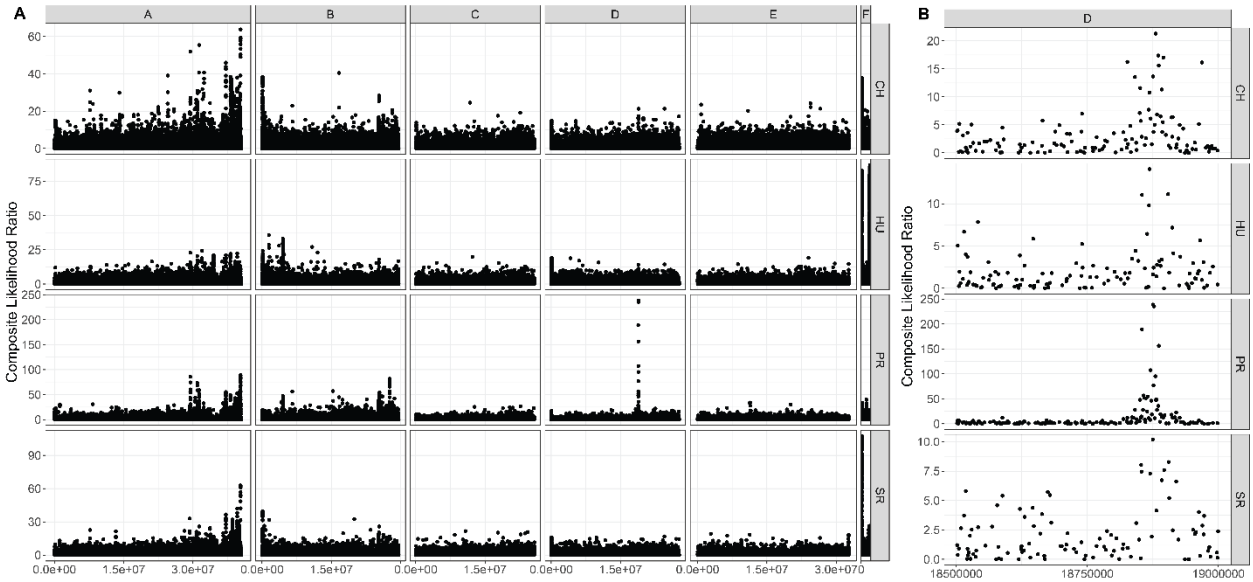
650

651 **Supplementary Figure 6: A.** Within population pairwise diversity per gene across the *D. innubila* genome.
652 **B.** D_{XY} per gene for each population. Instead of showing all pairwise comparisons, we show one randomly
653 chosen comparison for each population, due to no significant differences between comparisons.



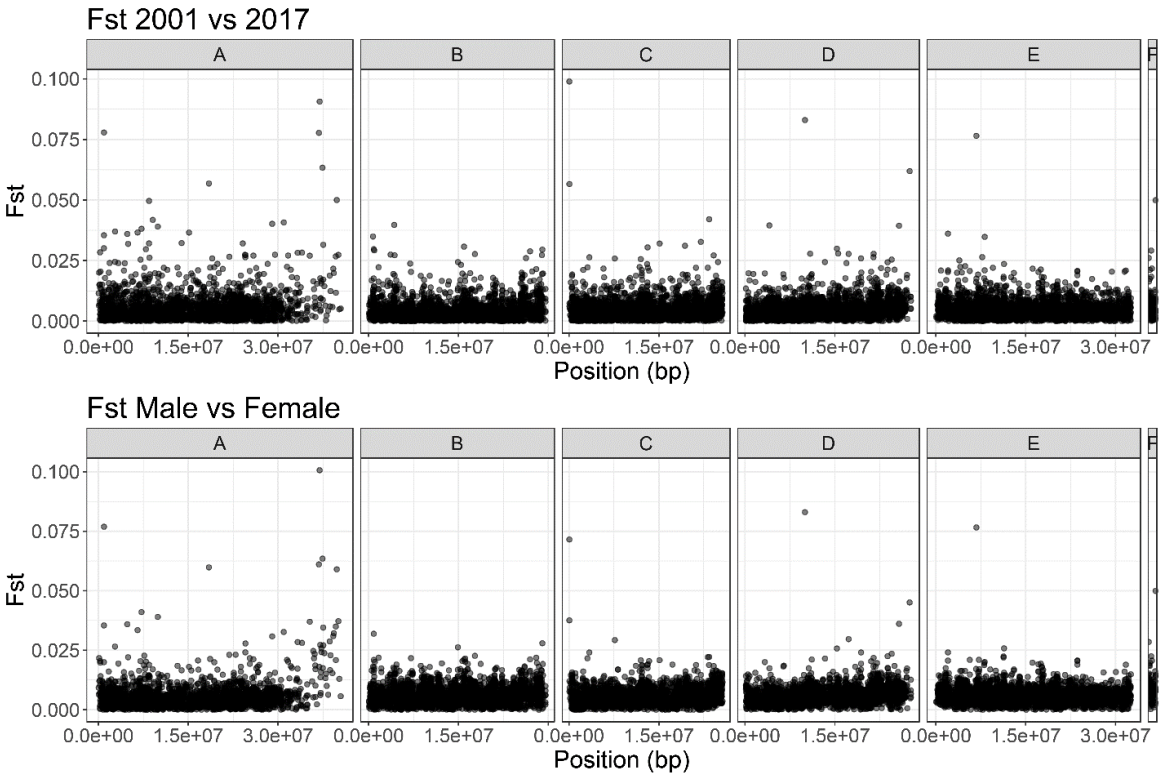
654
655

656 **Supplementary Figure 7:** Composite likelihood score for a selective sweep in 1kbp windows of the
657 genome estimated using Sweepfinder2. Separated by chromosome and population. **A.** Genome wide
658 composite likelihood score. **B.** Focus on 18.5-19Mbp of Muller element D, to show strongest selective
659 sweep in each the PR population.



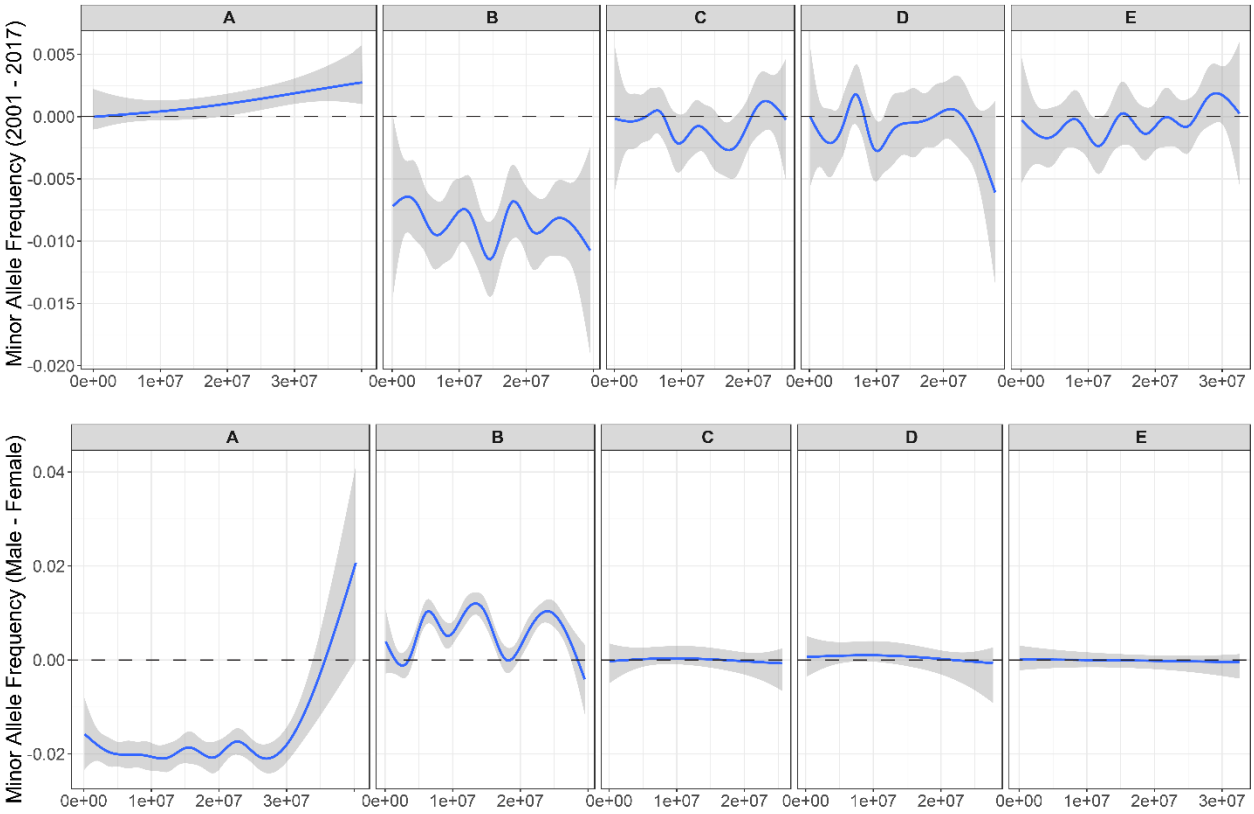
660

661 **Supplementary Figure 8:** Fst of genes between CH samples from 2001 and 2017, by chromosome and
662 position. Also shows Fst between all males and females from 2017, by chromosome and position.

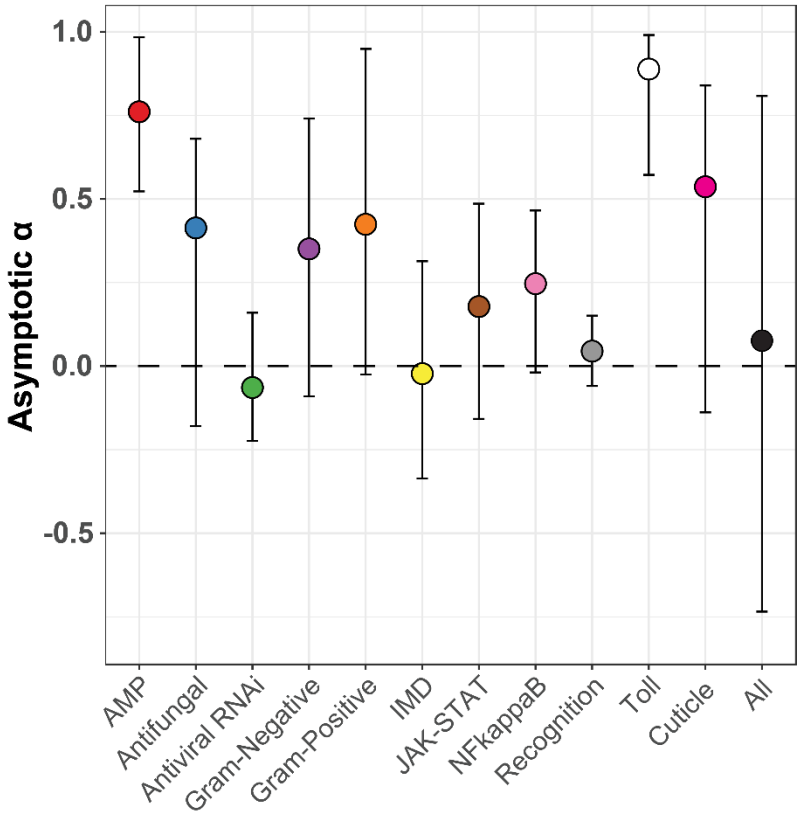


663

Supplementary Figure 9: Minor allele frequency difference curve across the genome (averaged over 2000 SNPs, sliding 1000 SNPs). Shows average difference in the minor allele frequencies (based on total 2017 sample). Comparisons between 2001 Chiricahua and 2017 Chiricahua, and between all 2017 males and 2017 female samples.

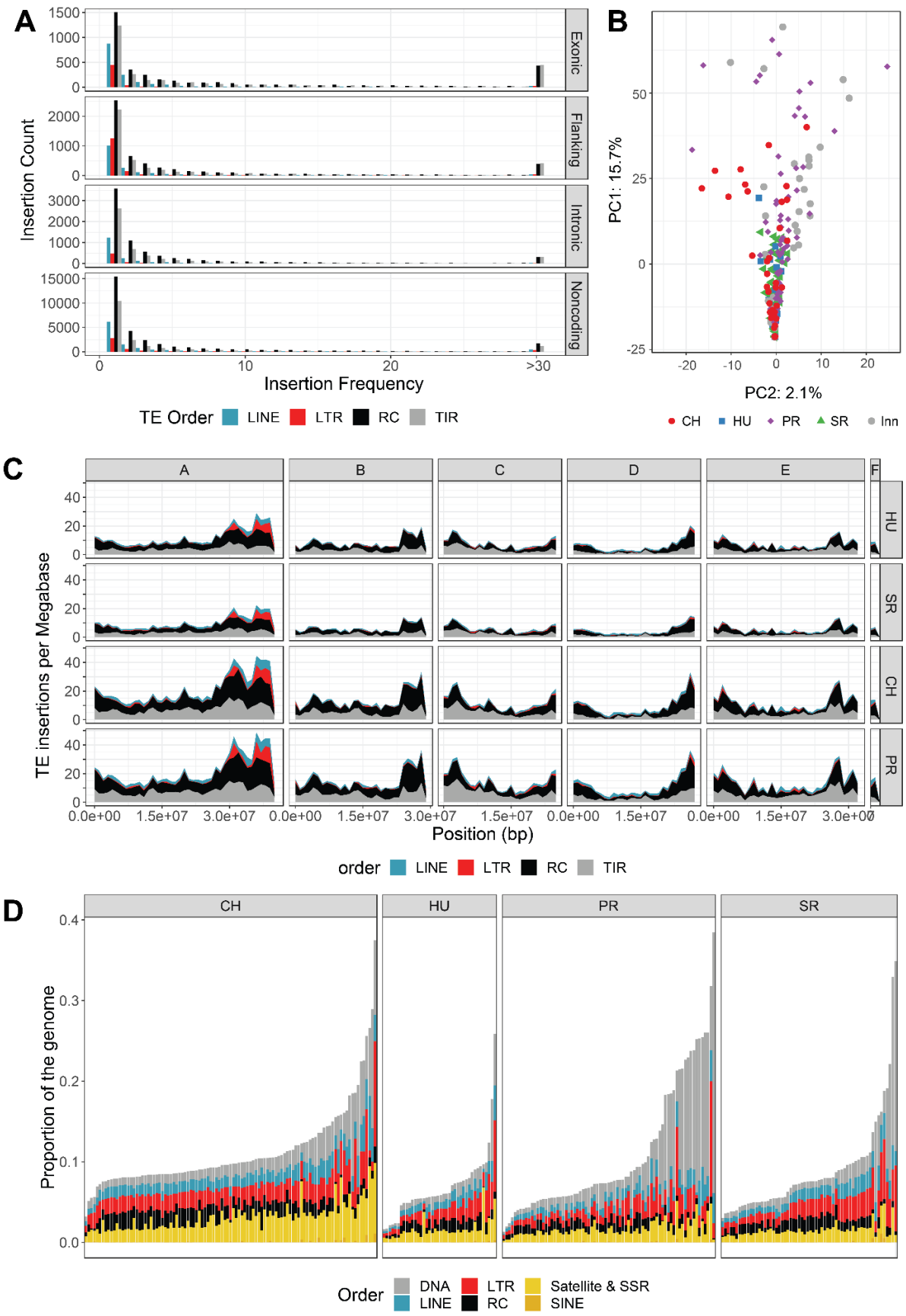


669 **Supplementary Figure 10:** Asymptotic α for immune categories, cuticle development proteins and all
670 proteins, with 95% confidence intervals for categories. Categories marked with a * are significantly higher
671 than the background following a permutation test (<0.05). 0 is marked with a dashed line. In the 'All'
672 category, the median and 95% confidence interval is calculated across all functional categories, while in
673 the specific functional categories the 95% confidence intervals for Asymptotic α are calculated using
674 AsymptoticMK (MESSER AND PETROV 2012; HALLER AND MESSER 2017).



675

676 **Supplementary Figure 11:** The transposable element content of *Drosophila innubila*. **A.** Insertion
677 frequency spectra for TEs in *D. innubila* separated by TE order and the location of insertion (e.g. coding
678 region, non-coding, intronic). **B.** Principle component analysis (showing PCs 1 & 2) of TE insertions
679 across *D. innubila* strains shows little population structure. Strains are labelled by their population (both
680 shape and color). **C.** Mean TE insertion density per 1Mb window (sliding 1Mb) for each population of *D.*
681 *innubila,* identified using PopoolationTE2. TE insertions are colored by their order. **D.** Proportion of the
682 genome made up of repetitive content for each strain, as found with dnaPipeTE. Strains are ordered by
683 total TE content from most to least, with bars colored by TE order.


A
Insertion Count
Exonic
Flanking
Intronic
Noncoding
Insertion Frequency
TE Order LINE LTR RC TIR
B
PC1: 15.7%
PC2: 2.1%
CH HU PR SR Inn
C
A B C D E F
HU SR CH PR
TE insertions per Megabase
Position (bp)
order LINE LTR RC TIR
D
CH HU PR SR
Proportion of the genome
Order DNA LINE LTR RC Satellite & SSR SINE

684

685 **Supplementary Table 1:** Summary of *Drosophila innubila* fly's DNA collected and sequenced for this
686 study, including summary of coverage for X chromosome, autosomes. Also contains SRA accessions for
687 each strain.
688 **Supplementary Table 2:** GLM for population genetic statistics in immune gene categories relative to the
689 background for each population.
690 **Supplementary Table 3:** Summary of gene ontology enrichments for F_{ST} in each population, separated
691 by processes, components and functions.
692 **Supplementary Table 4:** Summary of GLM for elevated McDonald-Krietman statistics GO categories in
693 *D. innubila*.
694 **Supplementary Table 5:** Summary of gene ontology enrichments for D_{XY} in each population.

695 **Supplementary Data 1:** VCF file for SNPs in *D. innubila*, used in estimation of population genetic
696 statistics and in GWAS.
697 **Supplementary Data 2:** Population genetic statistics calculated for each gene in *D. innubila* using
698 VCFtools for each population.
699 **Supplementary Data 3:** McDonald-Kreitman statistics calculated for each gene in *D. innubila* using
700 SnIPRE for each population.
701

702 **Bibliography**

703 Altschul, S. F., W. Gish, W. Miller, E. W. Myers and D. J. Lipman, 1990 Basic local alignment
704 search tool. Journal of Molecular Biology 215**:** 403-410.
705 Antunes, J. T., P. N. Leao and V. M. Vasconcelos, 2015 *Cylindrospermopsis raciborskii:* review
706 of the distribution, phylogeography, and ecophysiology of a global invasive species.
707 Frontiers in Microbiology 6**:** 473.
708 Arechederra-Romero, L., 2012 Southwest Fire Science Consortium Field Trip to the Chiricahua
709 National Monument: Discussion of the Impacts of the 2011 Horseshoe 2 Fire, pp. in
710 *Arizona Geology Magazine*, Arizona Geology Magazine.
711 Astanei, I., E. Gosling, J. Wilson and E. Powell, 2005 Genetic variability and phylogeography of
712 the invasive zebra mussel, *Dreissena polymorpha (Pallas)*. Mol Ecol 14**:** 1655-1666.
713 Avgar, T., G. Street, and J. M. Fryxell, 2014 On the adaptive benefits of mammal migration.
714 Canadian Journal of Zoology 92**:** 481-490.
715 Bao, W., K. K. Kojima and O. Kohany, 2015 Repbase Update, a database of repetitive elements
716 in eukaryotic genomes. Mobile DNA 6**:** 4-9.
717 Behrman, E. L., S. S. Watson, K. R. O'Brien, S. M. Heschel and P. S. Schmidt, 2015 Seasonal
718 variation in life history traits in two *Drosophila* species. Journal of Evolutionary Biology
719 28**:** 1691-1704.
720 Buffalo, V., 2018 *Scythe*.
721 Burt, A., and R. Trivers, 2006 Genes in Conflict.
722 Chakraborty, M., R. Zhao, X. Zhang, S. Kalsow and J. J. Emerson, 2017 Extensive hidden genetic
723 variation shapes the structure of functional elements in *Drosophila*. Doi.Org 50**:** 114967.

Chapman, J. R., T. Hill and R. L. Unckless, 2019 Balancing selection drives maintenance of genetic variation in *Drosophila* antimicrobial peptides. Genome Biology and Evolution 11**:** 2691-2701.

Charlesworth, B., D. Charlesworth and N. H. Barton, 2003 The Effects of Genetic and Geographic Structure on Neutral Variation. Annual Review of Ecology, Evolution, and Systematics 34**:** 99-125.

Charlesworth, B., and C. H. Langley, 1989 The population genetics of *Drosophila* transposable elements. Annual review of genetics 23**:** 251-287.

Charlesworth, B., C. H. Langley and P. D. Sniegowski, 1997 Transposable element distributions in *Drosophila*. Genetics 147**:** 1993-1995.

Chen, X., O. Schulz-Trieglaff, R. Shaw, B. Barnes, F. Schlesinger *et al.*, 2016 Manta: Rapid detection of structural variants and indels for germline and cancer sequencing applications. Bioinformatics 32**:** 1220-1222.

Cingolani, P., A. Platts, L. L. Wang, M. Coon, T. Nguyen *et al.*, 2012 A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly 6**:** 80-92.

Cini, A., C. Ioriatti and G. Anfora, 2012 A review of the invasion of *Drosophila suzukii* in Europe and a draft research agenda for integrated pest management. Bulletin of Insectology 65**:** 149-160.

Cloudsley-Thompson, J. L., 1978 Human Activities and Desert Expansion. The Geographical Journal 144**:** 416-423.

Coe, S. J., D. M. Finch and M. M. Friggens, 2012 An Assessment of Climate Change and the Vulnerability of Wildlife in the Sky Islands of the Southwest, pp. 1-208. United States Department of Agriculture.

Cruickshank, T. E., and M. W. Hahn, 2014 Reanalysis suggests that genomic islands of speciation are due to reduced diversity, not reduced gene flow. Mol Ecol 23**:** 3133-3157.

Cutter, A. D., and B. A. Payseur, 2013 Genomic signatures of selection at linked sites: unifying the disparity among species. Nat Rev Genet 14**:** 262-274.

Danecek, P., A. Auton, G. Abecasis, C. A. Albers, E. Banks *et al.*, 2011 The variant call format and VCFtools. Bioinformatics 27**:** 2156-2158.

DePristo, M. A., E. Banks, R. Poplin, K. V. Garimella, J. R. Maguire *et al.*, 2011 A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature genetics 43**:** 491-498.

Dobzhansky, T., and B. Spassky, 1968 The Genetics of Natural Populations XL: Heterotic and Deleterious Effects of Recessive Lethals in Populations of *Drosophila pseudoobscura*. Genetics 59**:** 411-425.

Dobzhansky, T., and A. H. Sturtevant, 1937 Inversions In Chromosomes of *Drosophila pseudoobscura.* Genetics 23**:** 28-64.

Dobzhansky, T. H., A. S. Hunter, O. Pavlovsky, B. Spassky and B. Wallace, 1963 Genetics of natural populations. XXXI. Genetics of an isolated marginal population of *Drosophila pseudoobscura*. Genetics 48**:** 91-103.

Dostert, C., E. Jouanguy, P. Irving, L. Troxler, D. Galiana-Arnoux *et al.*, 2005 The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila.* Nat Immunol 6**:** 946-953.

768 Dyer, K. A., 2004 Evolutionarily Stable Infection by a Male-Killing Endosymbiont in *Drosophila*
769 *innubila*: Molecular Evidence From the Host and Parasite Genomes. Genetics 168**:** 1443-
770 1455.
771 Dyer, K. a., and J. Jaenike, 2005 Evolutionary dynamics of a spatially structured host-parasite
772 association: *Drosophila innubila* and male-killing *Wolbachia*. Evolution; international
773 journal of organic evolution 59**:** 1518-1528.
774 Dyer, K. A., M. S. Minhas and J. Jaenike, 2005 Expression and modulation of embryonic male-
775 killing in *Drosophila innubila*: opportunities for multilevel selection. Evolution;
776 international journal of organic evolution 59**:** 838-848.
777 Eilertson, K. E., J. G. Booth and C. D. Bustamante, 2012 SnIPRE: Selection Inference Using a
778 Poisson Random Effects Model. PLoS Computational Biology 8.
779 Excoffier, L., M. Foll and R. J. Petit, 2009 Genetic Consequences of Range Expansions. Annual
780 Review of Ecology, Evolution, and Systematics 40**:** 481-501.
781 Falush, D., M. Stephens and J. K. Pritchard, 2003 Inference of population structure using
782 multilocus genotype data: Linked loci and correlated allele frequencies. Genetics 164**:**
783 1567-1587.
784 Frichot, E., F. Mathieu, T. Trouillon, G. Bouchard and O. François, 2014 Fast and efficient
785 estimation of individual ancestry coefficients. Genetics 196**:** 973-983.
786 Fuller, Z. L., G. D. Haynes, S. Richards and S. W. Schaeffer, 2016 Genomics of Natural
787 Populations: How Differentially Expressed Genes Shape the Evolution of Chromosomal
788 Inversions in. Genetics.
789 Gao, Z., D. Waggoner, M. Stephens, C. Ober and M. Przeworski, 2015 An estimate of the average
790 number of recessive lethal mutations carried by humans. Genetics 199**:** 1243-1254.
791 Garrido-Ramos, M. A., 2017 Satellite DNA: An Evolving Topic. Genes (Basel) 8.
792 Gillespie, J., 2004 Population Genetics: A Concise Guide. 232.
793 Goubert, C., L. Modolo, C. Vieira, C. V. Moro, P. Mavingui *et al.*, 2015 De novo assembly and
794 annotation of the Asian tiger mosquito (*Aedes albopictus*) repeatome with dnaPipeTE from
795 raw genomic reads and comparative analysis with the yellow fever mosquito (*Aedes*
796 *aegypti*). Genome Biology and Evolution 7**:** 1192-1205.
797 Gramates, L. S., S. J. Marygold, G. Dos Santos, J. M. Urbano, G. Antonazzo *et al.*, 2017 FlyBase
798 at 25: Looking to the future. Nucleic Acids Research 45**:** D663-D671.
799 Guindon, S., J.-F. Dufayard, V. Lefort, M. Anisimova, W. Hordijk *et al.*, 2010 New algorithms
800 and methods to estimate maximum-likelihood phylogenies: assessing the performance of
801 PhyML 3.0. Systematic biology 59**:** 307-321.
802 Haller, B. C., and P. W. Messer, 2017 asymptoticMK: A Web-Based Tool for the Asymptotic
803 McDonald–Kreitman Test. G3: Genes, Genomes, Genetics 7**:** 1569-1575.
804 Hermisson, J., and P. S. Pennings, 2005 Soft sweeps: molecular population genetics of adaptation
805 from standing genetic variation. Genetics 169**:** 2335-2352.
806 Hewitt, G., 2000 The genetic legacy of the Quaternary ice ages. Nature 405**:** 907-913.
807 Hill, T., B. Koseva and R. L. Unckless, 2019 The genome of *Drosophila innubila* reveals lineage-
808 specific patterns of selection in immune genes. Molecular Biology and Evolution**:** 1-36.
809 Hill, T., and R. Unckless, 2020 Recurrent evolution of two competing haplotypes in an insect DNA
810 virus. Biorxiv**:** 1-45.
811 Hoban, S., J. L. Kelley, K. E. Lotterhos, M. F. Antolin, G. Bradburd *et al.*, 2016 Finding the
812 Genomic Basis of Local Adaptation: Pitfalls, Practical Solutions, and Future Directions.
813 Am Nat 188**:** 379-397.

814 Hoffmann, J. A., 2003 The immune response of *Drosophila*. Nature 426**:** 33-38.
815 Holmgren, K., J. A. Lee-Thorp, G. R. J. Cooper, K. Lundblad, T. C. Partridge *et al.*, 2003 Persistent
816 millennial-scale climatic variability over the past 25,000 years in Southern Africa.
817 Quaternary Science Reviews 22**:** 2311-2326.
818 Http://broadinstitute.github.io/picard, *Picard*.
819 Huber, C. D., M. DeGiorgio, I. Hellmann and R. Nielsen, 2016 Detecting recent selective sweeps
820 while controlling for mutation rate and background selection. Mol Ecol 25**:** 142-156.
821 Imler, J., and I. Elftherianos, 2009 *Drosophila* as a model for studying antiviral defences. Insect
822 infection and immunity (eds Rolff J., Reyolds SE)**:** 49-68.
823 Jaenike, J., and K. A. Dyer, 2008 No resistance to male-killing *Wolbachia* after thousands of years
824 of infection. Journal of Evolutionary Biology 21**:** 1570-1577.
825 Joshi, N., and J. Fass, 2011 Sickle: A sliding window, adaptive, quality-based trimming tool for
826 fastQ files. 1.33.
827 Kageyama, D., H. Anbutsu, M. Shimada and T. Fukatsu, 2009 Effects of host genotype against
828 the expression of spiroplasma-induced male killing in Drosophila melanogaster. Heredity
829 (Edinb) 102**:** 475-482.
830 Kofler, R., A. J. Betancourt and C. Schlötterer, 2012 Sequencing of pooled DNA Samples (Pool-
831 Seq) uncovers complex dynamics of transposable element insertions in *Drosophila*
832 *melanogaster*. PloS Genetics 8**:** 1-16.
833 Kofler, R., G. Daniel and C. Schlötterer, 2016 PoPoolationTE2 : comparative population genomics
834 of transposable elements using Pool-Seq. Molecular Biology and Evolution**:** 1-12.
835 Kofler, R., V. Nolte and C. Schlötterer, 2015 Tempo and mode of transposable element activity in
836 *Drosophila*. PLoS Genet 11**:** e1005406.
837 Korneliussen, T. S., A. Albrechtsen and R. Nielsen, 2014 ANGSD: Analysis of Next Generation
838 Sequencing Data. BMC Bioinformatics 15**:** 356.
839 Lachaise, D., and J.-F. Silvain, 2004 How two Afrotropical endemics made two cosmopolitan
840 human commensals: the *Drosophila melanogaster-D. simulans* palaeogeographic riddle.
841 Genetica 120**:** 17-39.
842 Lack, J. B., C. M. Cardeno, M. W. Crepeau, W. Taylor, R. B. Corbett-Detig *et al.*, 2015 The
843 *Drosophila* genome nexus: A population genomic resource of 623 *Drosophila*
844 *melanogaster* genomes, including 197 from a single ancestral range population. Genetics
845 199**:** 1229-1241.
846 Li, H., and R. Durbin, 2009 Fast and accurate short read alignment with Burrows-Wheeler
847 transform. Bioinformatics (Oxford, England) 25**:** 1754-1760.
848 Li, H., and R. Durbin, 2011 Inference of human population history from individual whole-genome
849 sequences. Nature 475**:** 493-496.
850 Li, H., B. Handsaker, A. Wysoker, T. Fennell, J. Ruan *et al.*, 2009 The sequence alignment/map
851 format and SAMtools. Bioinformatics (Oxford, England) 25**:** 2078-2079.
852 Liu, X., and Y.-X. Fu, 2015 Exploring population size changes using SNP frequency spectra.
853 Nature genetics 47**:** 555-559.
854 Lower, S. S., M. P. McGurk, A. G. Clark and D. A. Barbash, 2018 Satellite DNA evolution: old
855 ideas, new approaches. Curr Opin Genet Dev 49**:** 70-78.
856 Ma, X., J. L. Kelley, K. Eilertson, S. Musharoff, J. D. Degenhardt *et al.*, 2013 Population Genomic
857 Analysis Reveals a Rich Speciation and Demographic History of Orangutans (*Pongo*
858 *pygmaeus* and *Pongo abelii*). PLoS ONE 8.

Machado, C. a., T. S. Haselkorn and M. a. F. Noor, 2007 Evaluation of the genomic extent of effects of fixed inversion differences on intraspecific variation and interspecific gene flow in *Drosophila pseudoobscura* and *Drosophila persimilis*. Genetics 175**:** 1289-1306.

Machado, H. E., A. O. Bergland, K. R. O'Brien, E. L. Behrman, P. S. Schmidt *et al.*, 2015 Comparative population genomics of latitudinal variation in *D. simulans* and *D. melanogaster*. Molecular Ecology**:** n/a-n/a.

Mackay, T. F. C., S. Richards, E. a. Stone, A. Barbadilla, J. F. Ayroles *et al.*, 2012 The *Drosophila melanogaster* genetic reference panel. Nature 482**:** 173-178.

Marinkovic, D., 1967 Genetic Loads Affecting Fecundity in Natural Populations of *Drosophila pseudoobscura*. Genetics**:** 61-71.

Markow, T. A., and P. O'Grady, 2006 *Drosophila*: a guide to species identification.

Martin, M., 2011 Cutadapt removes adapter sequences from high-throughput sequencing reads. Technical Notes**:** 1-12.

Marzo, M., M. Puig and A. Ruiz, 2008 The Foldback-like element Galileo belongs to the P superfamily of DNA transposons and is widespread within the Drosophila genus. Proceedings of the National Academy of Sciences of the United States of America 105**:** 2957-2962.

Matthey-Doret, R., and M. C. Whitlock, 2018 Background selection and the statistics of population differentiation: consequences for detecting local adaptation. Biorxiv**:** 1-5.

McCormack, J. E., H. Huang and L. L. Knowles, 2009 Sky Islands, pp. 841-843 in *Encyclopedia of islands*.

McDonald, J. H., and M. Kreitman, 1991 Adaptive protein evolution at the Adh locus in *Drosophila*. Nature 351**:** 652-654.

McKenna, A., M. Hanna, E. Banks, A. Sivachenko, K. Cibulskis *et al.*, 2010 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 20**:** 1297-1303.

McVean, G., 2007 The structure of linkage disequilibrium around a selective sweep. Genetics 175**:** 1395-1406.

Merkling, S. H., and R. P. van Rij, 2013 Beyond RNAi: Antiviral defense strategies in *Drosophila* and mosquito. Journal of Insect Physiology 59**:** 159-170.

Messer, P. W., and D. A. Petrov, 2012 The McDonald-Kreitman Test and its Extensions under Frequent Adaptation: Problems and Solutions. Proceedings of the National Academy of Sciences 110**:** 8615-8620.

Messer, P. W., and D. A. Petrov, 2013 Population genomics of rapid adaptation by soft selective sweeps. Trends in Ecology & Evolution 28**:** 659-669.

Narasimhan, V., P. Danecek, A. Scally, Y. Xue, C. Tyler-Smith *et al.*, 2016 BCFtools/RoH: A hidden Markov model approach for detecting autozygosity from next-generation sequencing data. Bioinformatics 32**:** 1749-1751.

Nei, M., 1987 *Molecular evolutionary genetics*. Columbia university press.

Nei, M., and J. Miller, 1990 A Simple Method for Estimating Average Number of Nucleotide Substitutions Within and Between Populations From Restriction Data. Genetics 125**:** 873-879.

Noor, M. a. F., D. a. Garfield, S. W. Schaeffer and C. a. Machado, 2007 Divergence between the *Drosophila pseudoobscura* and *D. persimilis* genome sequences in relation to chromosomal inversions. Genetics 177**:** 1417-1428.

905 Palmer, W. H., J. Joosten, G. J. Overheul, P. W. Jansen, M. Vermeulen *et al.*, 2018 Induction and
906 suppression of NF- κB signalling by a DNA virus of *Drosophila.*
907 Parmesan, C., and G. Yohe, 2003 A globally coherent fingerprint of climate change impacts across
908 natural systems. Nature 421**:** 37-42.
909 Petrov, D. a., A.-S. Fiston-Lavier, M. Lipatov, K. Lenkov and J. González, 2011 Population
910 genomics of transposable elements in *Drosophila melanogaster.* Molecular Biology and
911 Evolution 28**:** 1633-1644.
912 Pool, J., and C. H. Langley, 2013 DPGP3.
913 Pool, J. E., R. B. Corbett-detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno *et al.*, 2012 Population
914 Genomics of Sub-Saharan *Drosophila melanogaster:* African Diversity and Non-African
915 Admixture. PLoS Genetics 8**:** 1-24.
916 Porretta, D., V. Mastrantonio, R. Bellini, P. Somboon and S. Urbanelli, 2012 Glacial history of a
917 modern invader: phylogeography and species distribution modelling of the Asian tiger
918 mosquito Aedes albopictus. PLoS One 7**:** e44515.
919 Rankin, M. A., and J. C. A. Burchsted, 1992 The Cost of Migration in Insects. Annual Review of
920 Entymology 37**:** 533-559.
921 Rastogi, S., and D. a. Liberles, 2005 Subfunctionalization of duplicated genes as a transition state
922 to neofunctionalization. BMC evolutionary biology 5**:** 28.
923 Rausch, T., T. Zichner, A. Schlattl, A. M. Stutz, V. Benes *et al.*, 2012 DELLY: structural variant
924 discovery by integrated paired-end and split-read analysis. Bioinformatics 28**:** i333-i339.
925 Rosenzweig, C., D. Karoly, M. Vicarelli, P. Neofotis, Q. Wu *et al.*, 2008 Attributing physical and
926 biological impacts to anthropogenic climate change. Nature 453**:** 353-357.
927 Schrider, D. R., D. Houle, M. Lynch and M. W. Hahn, 2013 Rates and genomic consequences of
928 spontaneous mutational events in Drosophila melanogaster. Genetics 194**:** 937-954.
929 Searle, J. B., P. Kotlik, R. V. Rambau, S. Markova, J. S. Herman *et al.*, 2009 The Celtic fringe of
930 Britain: insights from small mammal phylogeography. Proc Biol Sci 276**:** 4287-4294.
931 Smit, A. F. A., and R. Hubley, 2008 RepeatModeler Open-1.0.
932 Smit, A. F. A., and R. Hubley, 2013-2015 RepeatMasker Open-4.0, pp. RepeatMasker.
933 Smith, F. A., J. L. Betancourt and J. H. Brown, 1995 Evolution of Body Size in the Woodrat Over
934 the Past 25,000 Years of Climate Change. Science 270**:** 2012-2014.
935 Smith, N. G. C., and A. Eyre-Walker, 2002 Adaptive protein evolution in *Drosophila*. Nature 415**:**
936 1022-1024.
937 Stajich, J. E., and M. W. Hahn, 2005 Disentangling the effects of demography and selection in
938 human history. Mol Biol Evol 22**:** 63-73.
939 Stoletzki, N., and A. Eyre-Walker, 2011 Estimation of the neutrality index. Molecular Biology and
940 Evolution 28**:** 63-70.
941 Subramanian, A., P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert *et al.*, 2005 Gene set
942 enrichment analysis: A knowledge-based approach for interpreting genome-wide
943 expression profiles. PNAS 102**:** 15545–15550.
944 Survey, A. G., 2005 Arizona Geology. Arizona Geology 35**:** 1-6.
945 Tajima, F., 1989 Statistical method for testing the neutral mutation hypothesis by DNA
946 polymorphism. Genetics 123**:** 585-595.
947 Takeda, K., and S. Akira, 2005 Toll-like receptors in innate immunity. International Immunology
948 17**:** 1-14.
949 Unckless, R. L., 2011a A DNA Virus of *Drosophila*. PLoS ONE 6**:** e26564.

Unckless, R. L., 2011b The potential role of the X chromosome in the emergence of male-killing from mutualistic endosymbionts. J Theor Biol 291**:** 99-104.

Unckless, R. L., and J. Jaenike, 2011 Maintenance of a Male-Killing *Wolbachia* in *Drosophila innubila* By Male-Killing Dependent and Male-Killing Independent Mechanisms. Evolution 66**:** 678-689.

Walsh, D. B., M. P. Bolda, R. E. Goodhue, A. J. Dreves, J. Lee *et al.*, 2011 *Drosophila suzukii* (Diptera: *Drosophilidae*): Invasive Pest of Ripening Soft Fruit Expanding its Geographic Range and Damage Potential. Journal of Integrated Pest Management 2**:** 1-7.

Watanabe, T. K., O. Yamaguchi and T. Mukai, 1974 The Genetic Variability of Third Chromosomes in a Local Population of *Drosophila melanogaster*. Genetics 82**:** 63-82.

Weir, B. S., and C. C. Cockerham, 1984 Estimating F-Statistics for the Analysis of Population Structure. Evolution 38**:** 1358-1370.

White, T. A., S. E. Perkins, G. Heckel and J. B. Searle, 2013 Adaptive evolution during an ongoing range expansion: the invasive bank vole (*Myodes glareolus*) in Ireland. Mol Ecol 22**:** 2971-2985.

Wright, S. I., B. Lauga and D. Charlesworth, 2003 Subdivision and haplotype structure in natural populations of *Arabidopsis lyrata*. Molecular Ecology 12**:** 1247–1263.

Ye, K., M. H. Schulz, Q. Long, R. Apweiler and Z. Ning, 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.

Zambon, R. A., M. Nandakumar, V. N. Vakharia and L. P. Wu, 2005 The Toll pathway is important for an antiviral response in *Drosophila*. Proceedings of the National Academy of Sciences 102**:** 7257-7262.