Letters: Discoveries

Balancing selection drives maintenance of genetic variation in *Drosophila* antimicrobial peptides

Joanne R. Chapman[1†], Tom Hill[1†], Robert L. Unckless[1*]

[1]Department of Molecular Biosciences, University of Kansas, Lawrence, Kansas, USA

† These authors contributed equally to the work

* unckless@ku.edu

2018-03-19

Keywords: Balancing selection, antimicrobial peptides (AMPs), population genetics, immunity, coevolution, *Drosophila*.

The *Drosophila* sequence data were obtained from John Pool's *Drosophila* Genome Nexus (http://www.johnpool.net/genomes.html).

Abstract

Genes involved in immune defense against pathogens provide some of the most well-known examples of both directional and balancing selection. Antimicrobial peptides (AMPs) are innate immune effector genes, playing a key role in pathogen clearance in many species, including *Drosophila*. Conflicting lines of evidence have suggested AMPs may be under directional, balancing or purifying selection. Here, we use a case-control gene approach to show that balancing selection is an important force shaping AMP diversity in two species of *Drosophila*. In *D. melanogaster*, this is most clearly observed in ancestral African populations. Furthermore, the signature of balancing selection is even clearer once background selection has been accounted for. Balancing selection also acts on AMPs in *D. mauritiana*, an isolated island endemic separated from *D. melanogaster* by about 4 million years of evolution. This suggests that balancing selection may be acting to maintain adaptive diversity in AMPs in insects as it does in other taxa.

1 Introduction

Pathogens exert strong selective pressures on their hosts, both in terms of individual fitness and the evolutionary trajectory of populations and species. Co-evolutionary dynamics of hosts and pathogens results in continual selection for adaptive improvements in both players, often referred to as a co-evolutionary arms race (1, 2, 3). As a consequence, genes involved in immune defense tend to undergo strong positive selection, such that they are among the fastest evolving genes in the genomes of many hosts (4, 5, 6, 7, 8).

However, resistance mutations may not always become fixed. Balancing selection is the process whereby polymorphism is adaptively maintained within genes over extended timescales, sometimes described as trench-warfare dynamics (9). Several processes are thought to contribute to balancing selection (reviewed in (10)). These include heterozygote advantage, whereby individuals heterozygous at a given locus have a fitness advantage over either homozygote; negative frequency dependent selection, whereby the benefit of an allele increases the rarer it is in a population; and selection varying in a context-dependent manner, for example at different spatial or temporal scales, between the sexes, or in the presence or absence of infection. Balancing selection can be detected as an excess of intermediate frequency variants and a region of increased polymorphism around the selected site. The extent to which selection will impact genetic variation within and around immune genes will depend on a number of factors, including: the timescale upon which selection is acting (11); the density, diversity and virulence of pathogens (12); the cost of maintaining resistance alleles in the absence of infection (13); effective population size, mutation and

recombination rates of hosts and pathogens (14); environmental variables (15); and demographic factors such as gene flow and bottlenecks (16).

The dynamic selective pressures exerted by pathogens promote balanced polymorphism of host immune genes in several cases. Perhaps the best documented example is the major histocompatibility complex (MHC) in vertebrates (reviewed in (17, 18, 19, 20)). Individuals tend to be heterozygous at MHC loci, and large numbers of MHC alleles are maintained in populations. Other examples of balancing selection acting on host immune genes in animals include toll-like receptors (TLRs) in humans (21), red deer (22) and birds (23, 24); various cytokine genes (particularly interleukins) in humans (21, 25, 26, 27), birds (28, 29, 30) and voles (31); and viral resistance genes including *Oas1b* in mice (32), *OAS1* in primates (33, 34) and *TRIM5* in humans (35) and primates (36). Balancing selection also appears to play a role in the evolution of antimicrobial peptides (AMPs). AMPs are effectors of innate immunity that are strongly induced upon infection (37, 38). They tend to be membrane active (39, 40), with a direct role in killing and/or impeding the growth of pathogens (41, 42). Balancing selection has been implicated as a driver of AMP evolution in a diverse array of species including birds (43, 44), amphibians (45), fish (46), molluscs (47) and humans (48, 49).

The fruit fly, *Drosophila melanogaster*, is an important model for understanding evolution of the immune system (50, 51, 52, 53, 54). Directional selection on *Drosophila* immune genes appears to be a relatively widespread phenomenon, especially amongst receptor and signaling genes (55, 56, 57, 58, 59, 60). In contrast, evidence for balancing selection acting on *Drosophila* immune genes has been more equivocal. Genome-wide scans by Croze and colleagues (61, 62) found little evidence for balancing selection acting on immune genes in general, and Obbard *et al.* (58) found no evidence for adaptive evolution of AMPs. In contrast, both single gene and genome-wide analyses of selection have indicated that balancing selection (13, 63) or diversifying selection (64) may play an important role in the evolution of AMPs in *Drosophila*. Additionally, recent analyses have shown that seasonal fluctuations in temperate can cause rapid oscillations in *D. melanogaster* allele frequencies (65), particularly in immune genes, including AMPs (66, 67).

Insects and other invertebrates lack an adaptive immune system, so AMPs play a key role in controlling pathogen load and infection outcome (41, 42). Given their direct interaction with pathogens, it is surprising that AMPs do not show signatures of recurrent adaptive substitutions. We hypothesize that AMPs in insects are prone to balancing selection. To test this hypothesis, we examined AMP variation in four populations of *Drosophila melanogaster* and one population of *Drosophila mauritiana*. Using a case-control gene approach, we searched for molecular evolutionary signatures of selection. Our results provide evidence that balancing selection is an important

driver of AMP evolution.

2 Results

2.1 Genetic variation across four *Drosophila melanogaster* populations

To determine whether AMPs show signatures of balancing selection, we examined nucleotide polymorphism data in *D. melanogaster* populations. Coding sequence alignments for 13494 genes (including 35 AMPs and 104 immunity genes) were obtained (68) for four *D. melanogaster* populations: Zambia (ZI), Rwanda (RG), France (FR), and North Carolina (DGRP) (see Materials and Methods, Supplementary Table 1). *D. melanogaster* originated in Sub-Saharan Africa, expanded into Europe approximately 15-16,000 years ago, and subsequently spread to North America less than 200 years ago (69, 70, 71). The ZI and RG lines therefore represent ancestral populations, whereas FR and DGRP are derived populations.

We calculated three population genetic statistics: Watterson's θ (the sample size corrected number of segregating sites), π (pairwise nucleotide diversity) and Tajima's D. Consistent with balancing selection occurring in AMPs, the mean Tajima's D for AMPs is higher than the average across autosomes for Zambia (ZI, -0.713 AMPs versus -1.168 autosome average), Rwanda (RG, -0.358 versus -0.503), France (FR, 0.033 versus -0.021), and the DGRP (-0.171 versus -0.179, Supplementary Table 2). As observed previously (e.g. (72, 73)), the autosome-wide average for Tajima's D is quite negative in *D. melanogaster*, which likely reflects a complex demographic history. In general, a significantly higher proportion of AMPs have a positive Tajima's D when compared to other genes on autosomes (Supplementary Table 3; χ^2 p-value < 0.02 for all populations except France where χ^2 p-value $= 0.36$).

2.2 Case-control tests for balancing selection in *Drosophila*

Given the apparent differences in selection between AMPs and the genome averages described above, we employed a case-control approach to test whether AMPs showed signatures of balancing selection while controlling for local variation in mutation and recombination rates. For each AMP, we randomly sampled genes of similar length (amino acid sequence length $\leq$10 times the size of the AMP) and position (within 100kb on either side), calculated statistics for the AMP and control gene, and then calculated the mean difference over the 35 AMP/control comparisons. We repeated this 10000 times to obtain an empirical distribution of differences (Figure 1). In these instances, a positive difference suggests a higher value for AMPs versus the control gene, and therefore a role for balancing selection. These differences are primarily pos-

itive for both π and Watterson's θ for all populations (Figure 1B-C, Table 1). For Tajima's D, the differences are positive for Zambia and Rwanda (ancestral populations), supporting balancing selection, but close to zero for France and negative for the DGRP (derived populations, Figure 1A, Table 1).

To identify if these signatures of balancing selection are unique to AMPs, or consistent across all immunity genes, we repeated all tests, this time for all non-AMP immunity genes. We found very little evidence of balancing or directional selection across the remaining immunity genes, with differences closer to zero (Supplementary Tables 2 and 4, Figure S1). This result is in general concordance with those of Croze *et al* (61, 62).

It is possible that the observed signature of balancing selection amongst AMPs is due to various sampling artifacts. First, AMP families tend to occur in clusters throughout the genome, so it is possible that including all AMPs in the analyses effectively counts the same selective event multiple times. To account for this, we subsampled 10 unlinked (>5kb apart) AMPs and repeated our analyses. This did not qualitatively change our results (Supplementary Figure 2). Second, the presence of the selfish genetic element *Segregation Distorter* (*SD*), a low-frequency autosomal meiotic drive element (74) on the second chromosome, in some lines (4% in both Zambia and France) may influence our results. However, removing these lines did not qualitatively change our results (Supplementary Table 5, Supplementary Figure 3). We therefore consider that the observed patterns reflect true underlying evolutionary processes rather than sampling artifacts.

2.3 Accounting for background selection strengthens the signature of balancing selection on *Drosophila* AMPs

Background selection, the removal neutral variation due to selection against linked deleterious alleles, can influence levels of polymorphism across the genome. Comeron (75) calculated the observed amount of background selection across the genome in 1000 base pair (bp) windows in the Rwanda population. He then correlated silent polymorphism against this measure. Regions with positive residuals (more silent polymorphism than expected based on background selection) were deemed to be under balancing selection, while those with negative residuals (less silent polymorphism than expected based on background selection) were deemed to be under directional selection. Two regions that contain AMPs (IM4 and Cecropin) were among the handful of outliers discussed by Comeron as being under balancing selection. We identified all AMP-containing windows and replotted Comeron's data. This revealed that AMPs tend to fall in regions well above the trend-line (red points, Figure 2A), indicating they are evolving in a manner consistent with balancing selection. To further ascertain whether AMPs as a group show signatures of balancing selection, we used Comeron's

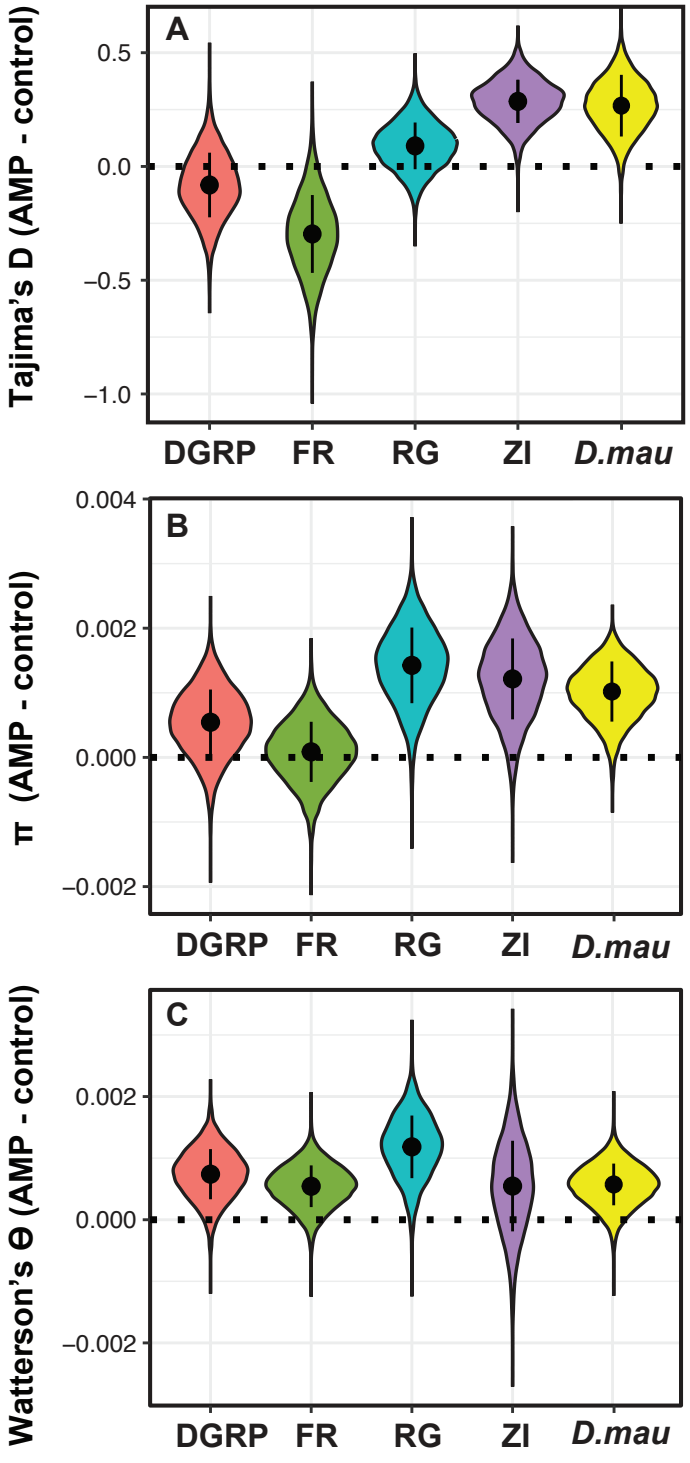


Figure 1: Difference in means between 35 AMPs and randomly chosen control genes, resampled 10000 times, separated by population (DGRP = Drosophila Genetics Reference Panel from North Carolina, USA; FR = France; RG = Rwanda; ZI = Zambia). A) Tajima's D, B) π, C) Watterson's θ. The black dot within each plot shows the median for that population, and the black bar around the dot visualizes the interquartile range of the distribution.

AMP - control statistics	**DGRP**	**FR**	**RG**	**ZI**	***D. mauritiana***
Tajima's D differences > 0 (%)	28.7	4.1	81.4	99.7	98.1
Tajima's D differences mean	-0.084	-0.295	0.092	0.289	0.26
Tajima's D differences std. dev.	0.142	0.171	0.102	0.092	0.12
π differences > 0 (%)	85.9	58.4	98.9	96.9	100
π differences mean	9.6e-5	9.6e-5	1.4e-3	1.2e-3	1.2e-5
π differences std. dev.	5.5e-4	4.8e-5	5.5e-3	6.1e-4	2.1e-6
Watterson's θ differences > 0 (%)	96.2	93.7	98.5	77.4	99.9
Watterson's θ differences mean	7.5e-4	5.5e-4	1.2e-3	5.6e-4	1.7e-5
Watterson's θ differences std. dev.	4.1e-4	3.4e-4	5.1e-4	7.4e-4	1.5e-6

Table 1: AMP minus control gene differences for three statistical measures of selection in four *D. melanogaster* populations and one *D. mauritiana* population. First row per statistic: percentage (%) of 10000 replicates in which the AMP minus control difference was positive (>0), suggestive of balancing selection; second row: mean AMP minus control difference across 10000 replicates; third row: standard deviation (std. dev.) of the mean (DGRP = Drosophila Genetics Reference Panel from North Carolina, USA; FR = France; RG = Rwanda; ZI = Zambia).

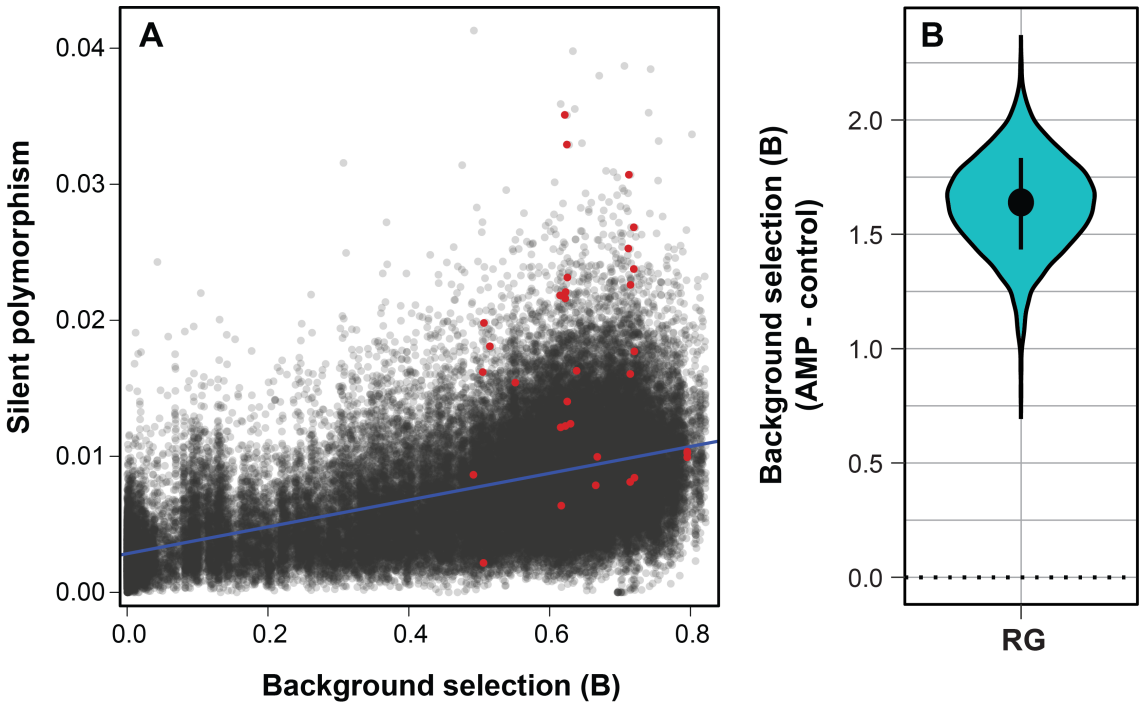


Figure 2: Accounting for background selection strengthens the signal of balancing selection on AMPs. A) Correlation between silent polymorphism and background selection (B) in 1000bp windows for the Rwanda population of *D. melanogaster*. The line of best fit is in blue and regions containing AMPs are indicated by red dots, B) Resampling of mean difference in the background selection statistic between AMPs and control genes.

background selection data (75) to calculate residuals for regions containing AMPs and compared them to residuals for randomly chosen position- and size-controlled genes employing methods similar to those used in the previous analyses. The distribution of differences in residuals was always above zero (Figure 2B, mean = 1.63, std. dev. = 0.20). This supports Comeron's assertion that accounting for background selection improves the ability to detect balancing selection, and also supports our previous results showing that AMPs as a group are subject to balancing selection.

2.4 Balancing selection also acts on *Drosophila mauritiana* AMPs

We also calculated population genetic statistics for 9980 genes in 107 *D. mauritiana* isofemale lines, sequenced as a pool. *D. mauritiana* is an island endemic which diverged from *D. melanogaster* approximately 3-5 million years ago (76, 77). SNP frequencies were called using Popoolation which accounts for low frequency variants and variation in coverage that may influence results from pooled samples (78). As found for *D. melanogaster*, there was a significant excess of AMPs with a positive Tajima's D compared to all other genes ($\chi^2 = 19.96$, p-value < 0.0001), and AMPs have a higher mean Tajima's D (-1.034 versus -1.463). We again resampled the difference in

these statistics between AMPs and neighboring control genes. We found AMPs have consistently higher values for π, Watterson's θ and Tajima's D than their matched controls (Figure 1, Table 1, Supplementary Tables 2 and 3, Supplementary Figure 4). For other immunity genes, the differences from controls are primarily negative for π, Watterson's θ and Tajima's D, suggesting directional selection may be acting on these genes (Supplementary Table 4, Supplementary Figure 4) in *D. mauritiana*.

3 Discussion

We find evidence consistent with balancing selection being an important evolutionary driver of AMP genes in *Drosophila*. This is most clearly observed in ancestral African populations (Zambia and Rwanda). There are several reasons why previous analyses may not have conclusively identified the selective forces acting on AMPs. First, signals of selection can be clouded by background selection. We found that the clearest signal for AMP balancing selection was in the Rwandan population after using Comeron's method (75) to account for background selection. Second, previous studies have tended to group immune genes as a single entity when scanning genomes for footprints of selection. Strong directional selection acting on some receptor and signaling immune genes may swamp a subtler signal of balancing selection acting on antimicrobial peptides. Third, this effect may be exacerbated by the fact that effector genes tend to be smaller (42) than receptor and signaling genes. Fourth, patterns of nucleotide polymorphism are strongly influenced by population demographic history. Our case-control approach should account for the confounding influences of local mutation and recombination rate variation, gene size and demography (79).

As populations establish in new habitats the pathogen pressure will be different, as will prevailing environmental conditions. This could dramatically alter which alleles are selectively advantageous. Loss of disadvantageous alleles (for example alleles resistant to pathogens not present in the new habitat) likely occurs more rapidly than establishment of new, beneficial polymorphisms (for example resistance alleles for newly encountered pathogens). This may explain why we find the strongest evidence for balancing selection on AMPs in ancestral African populations that have been co-evolving with their pathogens, under semi-predictable conditions, for long time-periods.

It is tempting to look to newly developed methods for detecting balancing selection (80, 81), but these statistics were developed for detecting the molecular footprints of selection in human populations. Assumptions about the genomic signatures of a balanced polymorphism that work well in humans are not applicable to *Drosophila*, because the window of linked polymorphism likely to show these signatures is tiny. To state this numerically, DeGiorgio *et al.* (81), based on Gao *et al.* (82), suggest

a window size of $1/\rho$ (where ρ is the population-scaled recombination rate or $4N_e r$) for observing the signature of a linked balanced polymorphism. For humans, ρ is about 0.001 so the window size is about 1000 bp (81). Estimates of ρ in *Drosophila* are highest in the DGRP population and range from 9.6 to 14.8 for the different chromosomes (83). These values correspond to windows of less than 1/10 of a single base in *Drosophila*, rendering these tests unusable in this genus.

We find that, at least in ancestral populations, AMPs tend to evolve in a manner consistent with balancing selection. This is in contrast to other immune genes that show no such pattern. Why are AMPs different than other immune genes? One characteristic of AMPs is that they interact directly with microbes (84), and, in some cases, AMP sequence is directly linked to the efficacy of bacterial membrane interactions (85). If particular AMP alleles encode for peptides that are more effective at fighting infection by particular microbes, a fluctuating suite of pathogens in the environment over time or space could lead to balanced polymorphisms. This "specificity hypothesis" suggests that allele frequencies in AMPs should vary spatially or temporally. There is some evidence for both seasonal (66) and spatial (67) variation in selection pressure on AMPs. However, evidence for AMP specificity against particular pathogens, especially different naturally occurring alleles of the same AMP, is currently rare (but see e.g. (63, 86, 87, 88)).

Alternatively, AMP variation might be maintained because AMP alleles that are more effective against pathogens also carry a higher autoimmune cost. This "autoimmune hypothesis" states that more effective AMP alleles should be common during pathogen epidemics, but decrease in frequency when pathogens are rare. These patterns might also vary spatially and temporally, making the interpretation of these context-dependent patterns more difficult. There is evidence that overexpression of AMPs can have deleterious fitness consequences (89, 90, 91). However, it seems that if autoimmune costs were important in maintaining variation, we would also see signatures of balancing selection in the IMD and toll pathway signaling genes that control expression of AMPs. Most work suggests that these genes are evolving under the arms race model (57, 58, 59). Distinguishing between these two hypotheses for the adaptive maintenance of AMP genetic variation will take careful functional analysis.

4 Methods

4.1 Polymorphism in four populations of *Drosophila melanogaster*

We downloaded chromosome sequences for the Zambia (ZI, n=197), Rwanda (RG, n=27), *Drosophila melanogaster* Genetic Reference Panel (DGRP, n=205) and France (FR, n=96) populations, available as part of the *Drosophila Genome Nexus*, from http://www.johnpool.net/genomes.html (92, 93). We then converted these sequences

into FASTA files, per chromosome, for each population. We also created a second set of FASTA files that excluded chromosomes known to contain the *Segregation Distorter* (*SD*) haplotype (taken from: (74)). The RG and ZI populations are much higher quality data, the average per base coverage of the raw FASTQ data used to generate the FASTA files is much higher, and the number of ambiguous bases is much lower than the DGRP and FR populations (Supplementary Table 1).

Using annotation 5.57 of the *D. melanogaster* genome, we extracted the FASTA alignments for each gene. Following this, we used a custom bioperl script, with the the package Bio, to find π, Watterson's θ, Tajima's D and the number of segregating sites for each gene. We categorized each gene using the designations found in Obbard *et al.* (57). We removed non-autosomal genes from all downstream analyses, because the X chromosome does not harbor any AMPs.

For each analysis, (per population, including and excluding SD chromosomes) we then resampled to find the average difference in scores between case and control genes. Case genes were either a) AMPS, or b) immunity genes (using gene ontologies previously described (58)). For each gene in these categories, we randomly sampled a control gene within 100kbp upstream or downstream, that was no more than ten times larger than this gene and not another gene in the given category (AMP or immunity). We then found the average difference ($\bar{\Delta}$) in each measure for the case (AMP/immunity) group and the control group such that:

$$\bar{\Delta} = \frac{1}{n}\sum_{i=1}^{n} X_{Case} - X_{Control}$$

where X_{Case} represents the chosen gene, $X_{Control}$ represents the randomly sampled control gene and n accounts for the number of genes in the group. We then repeated this 10000 times to obtain an empirical distribution of the differences.

We employ this method to control for genomewide variation in recombination rates, mutation rates, and possibly, demographic history. Resampling 10000 times allows for a robust empirical distribution that does not rely on the particular control genes chosen. We therefore present the distribution of differences as violin plots and purposefully do not discuss significance in terms of P-values. Instead, the proportion of resamplings that do not overlap zero is more analogous to a bootstrap value.

4.2 Polymorphism in a population of *Drosophila mauritiana*

We downloaded the reference genome, annotation and mapped BAM file of a population of *D. mauritiana* from http://www.popoolation.at/mauritiana_genome/, and used Popoolation to calculate Tajima's D, π and Watterson's θ for each gene in this population. We then resampled to find the average difference in scores between AMPs and a control set of genes, as described above.

5 Supplementary Material

Supplementary Table 1 - Summary statistics for each dataset, including the average base coverage for each population and the average number of ambiguous bases per 1000 bases in the FASTA files used. Data taken from johnpool.net/genomes.html.

Supplementary Table 2 - Summary statistics for each AMP and immunity gene for each population, also the mean for each statistic for all non-AMP immune genes.

Supplementary Table 3 - χ^2 test contingency tables for each population, showing the number of AMPs and other genes with positive and negative Tajima's D.

Supplementary Table 4 - Summary of resampling results across case (AMPs/immunity) genes and their matched control genes. These statistics include the percentage greater than 0, mean and standard deviation for each resampling set.

Supplementary Table 5 - Summary of resampling results across case (AMPs/immunity) genes and their matched control genes, with all SD containing samples removed. These statistics include the percentage greater than 0, mean and standard deviation for each resampling set.

Supplementary Figure 1 - Summary of resampling results (case - control) for AMPs and other immunity genes for Tajima's D, π, Watterson's θ.

Supplementary Figure 2 - Summary of resampling results (case - control) for AMPs for Tajima's D, π, Watterson's θ, using the subset of non-linked AMPs.

Supplementary Figure 3 - Summary of resampling results (case - control) for AMPs and other immunity genes for Tajima's D, π, Watterson's θ, comparing the results of ZI and FR populations with and without SD chromosomes.

Supplementary Figure 4 - Summary of resampling results (case - control) for AMPs and other immunity genes for Tajima's D, π, Watterson's θ in the *D. mauritiana* population.

6 Acknowledgments

We would like to thank John Kelly and Jamie Walters for thoughtful feedback on a previous version of the manuscript. Josep Comeron graciously provided additional data files from his background selection paper. Michael Degiorgio and Katie Siewart provided guidance on using their software packages for *Drosophila* populations. This work was supported by the National Institutes of Health (P20 GM103420 and R00 GM114714 to RLU) and the Max Kade Foundation (fellowship to TH). The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication.

References

1. S. Paterson, T. Vogwill, A. Buckling, R. Benmayor, A. J. Spiers, N. R. Thomson, M. Quail, F. Smith, D. Walker, B. Libberton, A. Fenton, N. Hall, and M. A. Brockhurst, "Antagonistic coevolution accelerates molecular evolution," *Nature*, vol. 464, no. 7286, pp. 275–278, 2010.

2. R. D. Schulte, C. Makus, B. Hasert, N. K. Michiels, and H. Schulenburg, "Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 107, no. 16, pp. 7359–7364, 2010.

3. P. H. Thrall, A.-L. Laine, M. Ravensdale, A. Nemri, P. N. Dodds, L. G. Barrett, and J. J. Burdon, "Rapid genetic change underpins antagonistic coevolution in a natural host-pathogen metapopulation," *Ecology Letters*, vol. 15, no. 5, pp. 425–435, 2012.

4. T. Downing, P. Cormican, C. O'Farrelly, D. G. Bradley, and A. T. Lloyd, "Evidence of the adaptive evolution of immune genes in chicken," *BMC Research Notes*, vol. 2, p. 254, 2009.

5. C. Kosiol, T. Vinař, R. R. da Fonseca, M. J. Hubisz, C. D. Bustamante, R. Nielsen, and A. Siepel, "Patterns of positive selection in six mammalian genomes," *PLoS Genetics*, vol. 4, no. 8, p. e1000144, 2008.

6. S. J. McTaggart, D. J. Obbard, C. Conlon, and T. J. Little, "Immune genes undergo more adaptive evolution than non-immune system genes in *Daphnia pulex*," *BMC Evolutionary Biology*, vol. 12, no. 63, 2012.

7. R. Nielsen, C. Bustamante, A. G. Clark, S. Glanowski, T. B. Sackton, M. J. Hubisz, A. Fledel-Alon, D. M. Tanenbaum, D. Civello, T. J. White, J. J. Sninsky, M. D. Adams, and M. Cargill, "A scan for positively selected genes in the genomes of humans and chimpanzees," *PLoS Biology*, vol. 3, no. 6, p. e170, 2005.

8. R. Ekblom, L. French, J. Slate, and T. Burke, "Evolutionary analysis and expression profiling of zebra finch immune genes," *Genome Biology and Evolution*, vol. 2, pp. 781–790, 2010.

9. E. A. Stahl, G. Dwyer, R. Mauricio, M. Kreitman, and J. Bergelson, "Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*," *Nature*, vol. 400, no. 6745, pp. 667–671, 1999.

10. V. Llaurens, A. Whibley, and M. Joron, "Genetic architecture and balancing selection: the life and death of differentiated variants," *Molecular Ecology*, vol. 26, no. 9, pp. 2430–2448, 2017.

11. D. Charlesworth, "Balancing selection and its effects on sequences in nearby genome regions," *PLoS Genetics*, vol. 2, no. 4, p. e64, 2006.

12. L. Lambrechts, S. Fellous, and J. C. Koella, "Coevolutionary interactions between host and parasite genotypes," *Trends in Parasitology*, vol. 22, no. 1, pp. 12–16, 2006.

13. R. L. Unckless and B. P. Lazzaro, "The potential for adaptive maintenance of diversity in insect antimicrobial peptides," *Philosophical Transactions of the Royal Society of London. Series B, Biological Sciences*, vol. 371, no. 1695, p. 20150291, 2016.

14. S. Gandon and Y. Michalakis, "Local adaptation, evolutionary potential and host-parasite coevolution: interactions between migration, mutation, population size and generation time," *Journal of Evolutionary Biology*, vol. 15, no. 3, pp. 451–462, 2002.

15. J. Wolinska and K. C. King, "Environment can alter selection in host-parasite interactions," *Trends in Parasitology*, vol. 25, no. 5, pp. 236–244, 2009.

16. M. A. Brockhurst, A. D. Morgan, P. B. Rainey, and A. Buckling, "Population mixing accelerates coevolution," *Ecology Letters*, vol. 6, no. 11, pp. 975–979, 2003.

17. L. Bernatchez and C. Landry, "MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years?," *Journal of Evolutionary Biology*, vol. 16, no. 3, pp. 363–377, 2003.

18. P. W. Hedrick, "Balancing selection and MHC," *Genetica*, vol. 104, no. 3, pp. 207–214, 1998.

19. S. V. Edwards and P. W. Hedrick, "Evolution and ecology of MHC molecules: from genomics to sexual selection," *Trends in Ecology and Evolution*, vol. 13, no. 8, pp. 305–311, 1998.

20. A. L. Hughes and M. Yeager, "Natural selection at major histocompatibility complex loci of vertebrates," *Annual Review of Genetics*, vol. 32, no. 1, pp. 415–435, 1998.

21. A. Ferrer-Admetlla, E. Bosch, M. Sikora, T. Marquès-Bonet, A. Ramírez-Soriano, A. Muntasell, A. Navarro, R. Lazarus, F. Calafell, J. Bertranpetit, and F. Casals, "Balancing selection is the main force shaping the evolution of innate immunity genes," *Journal of Immunology*, vol. 181, no. 2, pp. 1315–1322, 2008.

22. E. Quéméré, M. Galan, J. F. Cosson, F. Klein, S. Aulagnier, E. Gilot-Fromont, J. Merlet, M. Bonhomme, A. J. M. Hewison, and N. Charbonnel, "Immunogenetic

heterogeneity in a widespread ungulate: the European roe deer (*Capreolus capreolus*)," *Molecular Ecology*, vol. 24, no. 15, pp. 3873–3887, 2015.

23. M. Alcaide and S. V. Edwards, "Molecular evolution of the Toll-like receptor multigene family in birds," *Molecular Biology and Evolution*, vol. 28, no. 5, pp. 1703–1715, 2011.

24. D. L. Gilroy, K. P. Phillips, D. S. Richardson, and C. van Oosterhout, "Toll-like receptor variation in the bottlenecked population of the Seychelles warbler: computer simulations see the 'ghost of selection past' and quantify the 'drift debt'," *Journal of Evolutionary Biology*, vol. 30, no. 7, pp. 1276–1287, 2017.

25. M. Fumagalli, U. Pozzoli, R. Cagliani, G. P. Comi, S. Riva, M. Clerici, N. Bresolin, and M. Sironi, "Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions," *Journal of Experimental Medicine*, vol. 206, no. 6, pp. 1395–1408, 2009.

26. A. L. Hughes, B. Packer, R. Welch, S. J. Chanock, and M. Yeager, "High level of functional polymorphism indicates a unique role of natural selection at human immune system loci," *Immunogenetics*, vol. 57, no. 11, pp. 821–827, 2005.

27. J. N. Wilson, K. Rockett, B. Keating, M. Jallow, M. Pinder, F. Sisay-Joof, M. Newport, and D. Kwiatkowski, "A hallmark of balancing selection is present at the promoter region of interleukin 10," *Genes and Immunity*, vol. 7, no. 8, pp. 680–683, 2006.

28. T. Downing, A. T. Lloyd, C. O'Farrelly, and D. G. Bradley, "The differential evolutionary dynamics of avian cytokine and TLR gene classes," *Journal of Immunology*, vol. 184, no. 12, pp. 6993–7000, 2010.

29. T. Downing, D. J. Lynn, S. Connell, A. T. Lloyd, A. K. Bhuiyan, P. Silva, A. N. Naqvi, R. Sanfo, R.-S. Sow, B. Podisi, O. Hanotte, C. O'Farrelly, and D. G. Bradley, "Evidence of balanced diversity at the chicken interleukin 4 receptor alpha chain locus," *BMC Evolutionary Biology*, vol. 9, p. 136, 2009.

30. T. Downing, D. J. Lynn, S. Connell, A. T. Lloyd, A. K. F. H. Bhuiyan, P. Silva, A. N. Naqvi, R. Sanfo, R.-S. Sow, B. Podisi, C. O'Farrelly, O. Hanotte, and D. G. Bradley, "Contrasting evolution of diversity at two disease-associated chicken genes," *Immunogenetics*, vol. 61, no. 4, pp. 303–314, 2009.

31. A. K. Turner, M. Begon, J. A. Jackson, and S. Paterson, "Evidence for selection at cytokine loci in a natural population of field voles (*Microtus agrestis*)," *Molecular Ecology*, vol. 21, no. 7, pp. 1632–1646, 2012.

412 32. W. Ferguson, S. Dvora, J. Gallo, A. Orth, and S. Boissinot, "Long-term balancing
413 selection at the West Nile virus resistance gene, *Oas1b*, maintains transspecific
414 polymorphisms in the house mouse," *Molecular Biology and Evolution*, vol. 25,
415 no. 8, pp. 1609–1618, 2008.

416 33. W. Ferguson, S. Dvora, R. W. Fikes, A. C. Stone, and S. Boissinot, "Long-term
417 balancing selection at the antiviral gene *OAS1* in Central African chimpanzees,"
418 *Molecular Biology and Evolution*, vol. 29, no. 4, pp. 1093–1103, 2012.

419 34. I. Fish and S. Boissinot, "Contrasted patterns of variation and evolutionary con-
420 vergence at the antiviral *OAS1* gene in old world primates," *Immunogenetics*,
421 vol. 67, no. 9, pp. 487–499, 2015.

422 35. R. Cagliani, M. Fumagalli, M. Biasin, L. Piacentini, S. Riva, U. Pozzoli, M. C.
423 Bonaglia, N. Bresolin, M. Clerici, and M. Sironi, "Long-term balancing selection
424 maintains *trans*-specific polymorphisms in the human *TRIM5* gene," *Human Ge-*
425 *netics*, vol. 128, no. 6, pp. 577–588, 2010.

426 36. R. M. Newman, L. Hall, M. Connole, G.-L. Chen, S. Sato, E. Yuste, W. Diehl,
427 E. Hunter, A. Kaur, G. M. Miller, and W. E. Johnson, "Balancing selection and
428 the evolution of functional polymorphism in Old World monkey TRIM5α," *Pro-*
429 *ceedings of the National Academy of Sciences of the United States of America*,
430 vol. 103, no. 50, pp. 19134–19139, 2006.

431 37. B. Lemaitre, E. Nicolas, L. Michaut, J.-M. Reichhart, and J. A. Hoffmann, "The
432 dorsoventral regulatory gene cassette *spätzle/Toll/cactus* controls the potent an-
433 tifungal response in Drosophila adults," *Cell*, vol. 86, no. 6, pp. 973–983, 1996.

434 38. P. Tzou, S. Ohresser, D. Ferrandon, M. Capovilla, J.-M. Reichhart, B. Lemaitre,
435 J. A. Hoffmann, and J.-L. Imler, "Tissue-specific inducible expression of antimi-
436 crobial peptide genes in *Drosophila* surface epithelia," *Immunity*, vol. 13, no. 5,
437 pp. 737–748, 2000.

438 39. K. A. Brogden, "Antimicrobial peptides: pore formers or metabolic inhibitors in
439 bacteria?," *Nature Reviews Microbiology*, vol. 3, no. 3, pp. 238–250, 2005.

440 40. Y. Shai, "Mode of action of membrane active antimicrobial peptides," *Peptide*
441 *Science*, vol. 66, no. 4, pp. 236–248, 2002.

442 41. E. De Gregorio, P. T. Spellman, P. Tzou, G. M. Rubin, and B. Lemaitre, "The Toll
443 and Imd pathways are the major regulators of the immune response in *Drosophila*,"
444 *The EMBO Journal*, vol. 21, no. 11, pp. 2568–2579, 2002.

445 42. B. Lemaitre and J. Hoffmann, "The host defense of *Drosophila melanogaster*,"
446 *Annual Review of Immunology*, vol. 25, pp. 697–743, 2007.

43. J. R. Chapman, O. Hellgren, A. S. Helin, R. H. S. Kraus, R. L. Cromie, and J. Waldenström, "The evolution of innate immune genes: purifying and balancing selection on β-defensins in waterfowl," *Molecular Biology and Evolution*, vol. 33, no. 12, pp. 3075–3087, 2016.

44. O. Hellgren and B. C. Sheldon, "Locus-specific protocol for nine different innate immune genes (antimicrobial peptides: β-defensins) across passerine bird species reveals within-species coding variation and a case of trans-species polymorphisms," *Molecular Ecology Resources*, vol. 11, no. 4, pp. 686–692, 2011.

45. J. A. Tennessen and M. S. Blouin, "Balancing selection at a frog antimicrobial peptide locus: fluctuating immune effector alleles?," *Molecular Biology and Evolution*, vol. 25, no. 12, pp. 2669–2680, 2008.

46. K. Halldórsdóttir and E. Árnason, "Trans-species polymorphism at antimicrobial innate immunity cathelicidin genes of Atlantic cod and related species," *PeerJ*, vol. 3, p. e976, 2015.

47. C. C. Gosset, J. Do Nascimento, M.-T. Augé, and N. Bierne, "Evidence for adaptation from standing genetic variation on an antimicrobial peptide gene in the mussel *Mytilus edulis*," *Molecular Ecology*, vol. 23, no. 12, pp. 3000–3012, 2014.

48. E. J. Hollox and J. A. L. Armour, "Directional and balancing selection in human beta-defensins," *BMC Evolutionary Biology*, vol. 8, p. 113, 2008.

49. R. Cagliani, M. Fumagalli, S. Riva, U. Pozzoli, G. P. Comi, G. Menozzi, N. Bresolin, and M. Sironi, "The signature of long-standing balancing selection at the human defensin β-1 promoter," *Genome Biology*, vol. 9, no. 9, p. R143, 2008.

50. S. Janssens and R. Beyaert, "Role of toll-like receptors in pathogen recognition," *Clinical Microbiology Reviews*, vol. 16, no. 4, pp. 637–646, 2003.

51. J. A. Hoffmann, "The immune response of *Drosophila*," *Nature*, vol. 426, no. 6962, pp. 33–38, 2003.

52. O. Lamiable, C. Kellenberger, C. Kemp, L. Troxler, N. Pelte, M. Boutros, J. T. Marques, L. Daeffler, J. A. Hoffmann, A. Roussel, and J.-L. Imler, "Cytokine Diedel and a viral homologue suppress the IMD pathway in *Drosophila*," *Proceedings of the National Academy of Sciences*, vol. 113, no. 3, pp. 698–703, 2016.

53. C. Dostert, E. Jouanguy, P. Irving, L. Troxler, D. Galiana-Arnoux, C. Hetru, J. A. Hoffmann, and J.-L. Imler, "The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila*," *Nature Immunology*, vol. 6, no. 9, pp. 946–953, 2005.

482 54. D. Hultmark, "Immune reactions in *Drosophila* and other insects: a model for
483 innate immunity," *Trends in Genetics*, vol. 9, no. 5, pp. 178–183, 1993.

484 55. A. G. Clark, M. B. Eisen, D. R. Smith, C. M. Bergman, B. Oliver, T. A. Markow,
485 T. C. Kaufman, *et al.*, "Evolution of genes and genomes on the *Drosophila* phy-
486 logeny," *Nature*, vol. 450, no. 7167, pp. 203–218, 2007.

487 56. A. Heger and C. P. Ponting, "Evolutionary rate analyses of orthologs and paralogs
488 from 12 *Drosophila* genomes," *Genome Research*, vol. 17, no. 12, pp. 1837–1849,
489 2007.

490 57. D. J. Obbard, F. M. Jiggins, D. L. Halligan, and T. J. Little, "Natural selec-
491 tion drives extremely rapid evolution in antiviral RNAi genes," *Current Biology*,
492 vol. 16, no. 6, pp. 580–585, 2006.

493 58. D. J. Obbard, J. J. Welch, K.-W. Kim, and F. M. Jiggins, "Quantifying adap-
494 tive evolution in the *Drosophila* immune system," *PLoS Genetics*, vol. 5, no. 10,
495 p. e1000698, 2009.

496 59. T. B. Sackton, B. P. Lazzaro, T. A. Schlenke, J. D. Evans, D. Hultmark, and
497 A. G. Clark, "Dynamic evolution of the innate immune system in *Drosophila*,"
498 *Nature Genetics*, vol. 39, no. 12, pp. 1461–1468, 2007.

499 60. T. A. Schlenke and D. J. Begun, "Natural selection drives *Drosophila* immune
500 system evolution," *Genetics*, vol. 164, no. 4, pp. 1471–1480, 2003.

501 61. M. Croze, A. Wollstein, V. Božičević, D. Živković, W. Stephan, and S. Hutter, "A
502 genome-wide scan for genes under balancing selection in *Drosophila melanogaster*,"
503 *BMC Evolutionary Biology*, vol. 17, p. 15, 2017.

504 62. M. Croze, D. Živković, W. Stephan, and S. Hutter, "Balancing selection on im-
505 munity genes: review of the current literature and new analysis in *Drosophila*
506 *melanogaster*," *Zoology*, vol. 119, no. 4, pp. 322–329, 2016.

507 63. R. L. Unckless, V. M. Howick, and B. P. Lazzaro, "Convergent balancing selec-
508 tion on an antimicrobial peptide in *Drosophila*," *Current Biology*, vol. 26, no. 2,
509 pp. 257–262, 2016.

510 64. M. A. Hanson, P. T. Hamilton, and S. J. Perlman, "Immune genes and divergent
511 antimicrobial peptides in flies of the subgenus Drosophila," *BMC Evolutionary*
512 *Biology*, vol. 16, p. 228, 2016.

513 65. A. O. Bergland, E. L. Behrman, K. R. O'Brien, P. S. Schmidt, and D. A. Petrov,
514 "Genomic evidence of rapid and stable adaptive oscillations over seasonal time
515 scales in Drosophila," *PLoS Genetics*, vol. 10, no. 11, p. e1004775, 2014.

66. E. L. Behrman, V. M. Howick, M. Kapun, F. Staubach, A. O. Bergland, D. A. Petrov, B. P. Lazzaro, and P. S. Schmidt, "Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*," *Proceedings of the Royal Society of London. Series B, Biological Sciences*, vol. 285, no. 1870, p. 20172599, 2018.

67. A. M. Early, J. R. Arguello, M. Cardoso-Moreira, S. Gottipati, J. K. Grenier, and A. G. Clark, "Survey of global genetic diversity within the *Drosophila* immune system," *Genetics*, vol. 205, no. 1, pp. 353–366, 2017.

68. J. B. Lack, J. D. Lange, A. D. Tang, R. B. Corbett-Detig, and J. E. Pool, "A thousand fly genomes: an expanded *Drosophila* Genome Nexus," *Molecular Biology and Evolution*, vol. 33, no. 12, pp. 3308–3313, 2016.

69. A. Keller, "*Drosophila melanogaster*'s history as a human commensal," *Current Biology*, vol. 17, no. 3, pp. R77–R81, 2007.

70. H. Li and W. Stephan, "Inferring the demographic history and rate of adaptive substitution in *Drosophila*," *PLoS Genetics*, vol. 2, no. 10, p. e166, 2006.

71. J. R. David and P. Capy, "Genetic variation of *Drosophila melanogaster* natural populations," *Trends in Genetics*, vol. 4, no. 4, pp. 106–111, 1988.

72. S. Glinka, L. Ometto, S. Mousset, W. Stephan, and D. De Lorenzo, "Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach," *Genetics*, vol. 165, no. 3, pp. 1269–1278, 2003.

73. J. A. Shapiro, W. Huang, C. H. Zhang, M. J. Hubisz, J. Lu, D. A. Turissini, S. Fang, H.-Y. Wang, R. R. Hudson, R. Nielsen, Z. Chen, and C.-I. Wu, "Adaptive genic evolution in the *Drosophila* genomes," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 104, no. 7, pp. 2271–2276, 2007.

74. C. L. Brand, A. M. Larracuente, and D. C. Presgraves, "Origin, evolution, and population genetics of the selfish *Segregation Distorter* gene duplication in European and African populations of *Drosophila melanogaster*," *Evolution*, vol. 69, no. 5, pp. 1271–1283, 2015.

75. J. M. Comeron, "Background selection as baseline for nucleotide variation across the *Drosophila* genome," *PLoS Genetics*, vol. 10, no. 6, p. e1004434, 2014.

76. D. J. Obbard, J. Maclennan, K.-W. Kim, A. Rambaut, P. M. O'Grady, and F. M. Jiggins, "Estimating divergence dates and substitution rates in the *Drosophila* phylogeny," *Molecular Biology and Evolution*, vol. 29, no. 11, pp. 3459–3473, 2012.

77. V. Nolte, R. V. Pandey, R. Kofler, and C. Schlötterer, "Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in *Drosophila mauritiana*," *Genome Research*, vol. 23, no. 1, pp. 99–110, 2013.

78. R. Kofler, P. Orozco-terWengel, N. De Maio, R. V. Pandey, V. Nolte, A. Futschik, C. Kosiol, and C. Schlötterer, "PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals," *PLoS ONE*, vol. 6, no. 1, p. e15925, 2011.

79. D. Garrigan and P. W. Hedrick, "Perspective: detecting adaptive molecular polymorphism: Lessons from the MHC," *Evolution*, vol. 57, no. 8, pp. 1707–1722, 2003.

80. K. M. Siewert and B. F. Voight, "Detecting long-term balancing selection using allele frequency correlation," *Molecular Biology and Evolution*, vol. 34, no. 11.

81. M. DeGiorgio, K. E. Lohmueller, and R. Nielsen, "A model-based approach for identifying signatures of ancient balancing selection in genetic data," *PLoS Genetics*, vol. 10, no. 8, p. e1004561, 2014.

82. Z. Y. Gao, M. Przeworski, and G. Sella, "Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species," *Evolution*, vol. 69, no. 2, pp. 431–446, 2015.

83. A. H. Chan, P. A. Jenkins, and Y. S. Song, "Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*," *PLoS Genetics*, vol. 8, no. 12, p. e1003090, 2012.

84. P. Bulet and R. Stocklin, "Insect antimicrobial peptides: structures, properties and gene regulation," *Protein and Peptide Letters*, vol. 12, no. 1, pp. 3–11, 2005.

85. P. Schmitt, R. D. Rosa, and D. Destoumieux-Garzón, "An intimate link between antimicrobial peptide sequence diversity and binding to essential components of bacterial membranes," *Biochimica Et Biophysica Acta-Biomembranes*, vol. 1858, no. 5, pp. 958–970, 2016.

86. O. Hellgren, B. C. Sheldon, and A. Buckling, "*In vitro* tests of natural allelic variation of innate immune genes (avian β-defensins) reveal functional differences in microbial inhibition," *Journal of Evolutionary Biology*, vol. 23, no. 12, pp. 2726–2730, 2010.

87. D. Park, J. W. Jung, M. O. Lee, S. Y. Lee, B. Kim, H. J. Jin, J. Kim, Y.-J. Ahn, K. W. Lee, Y. S. Song, S. Hong, J. E. Womack, and H. W. Kwon, "Functional characterization of naturally occurring melittin peptide isoforms in two honey bee species, *Apis mellifera* and *Apis cerana*," *Peptides*, vol. 53, pp. 185–193, 2014.

88. J. A. Tennessen, D. C. Woodhams, P. Chaurand, L. K. Reinert, D. Billheimer, Y. Shyr, R. M. Caprioli, M. S. Blouin, and L. A. Rollins-Smith, "Variations in

the expressed antimicrobial peptide repertoire of northern leopard frog (*Rana pipiens*) populations suggest intraspecies differences in resistance to pathogens," *Developmental and Comparative Immunology*, vol. 33, no. 12, pp. 1247–1257, 2009.

89. M. Gilliet and R. Lande, "Antimicrobial peptides and self-DNA in autoimmune skin inflammation," *Current Opinion in Immunology*, vol. 20, no. 4, pp. 401–407, 2008.

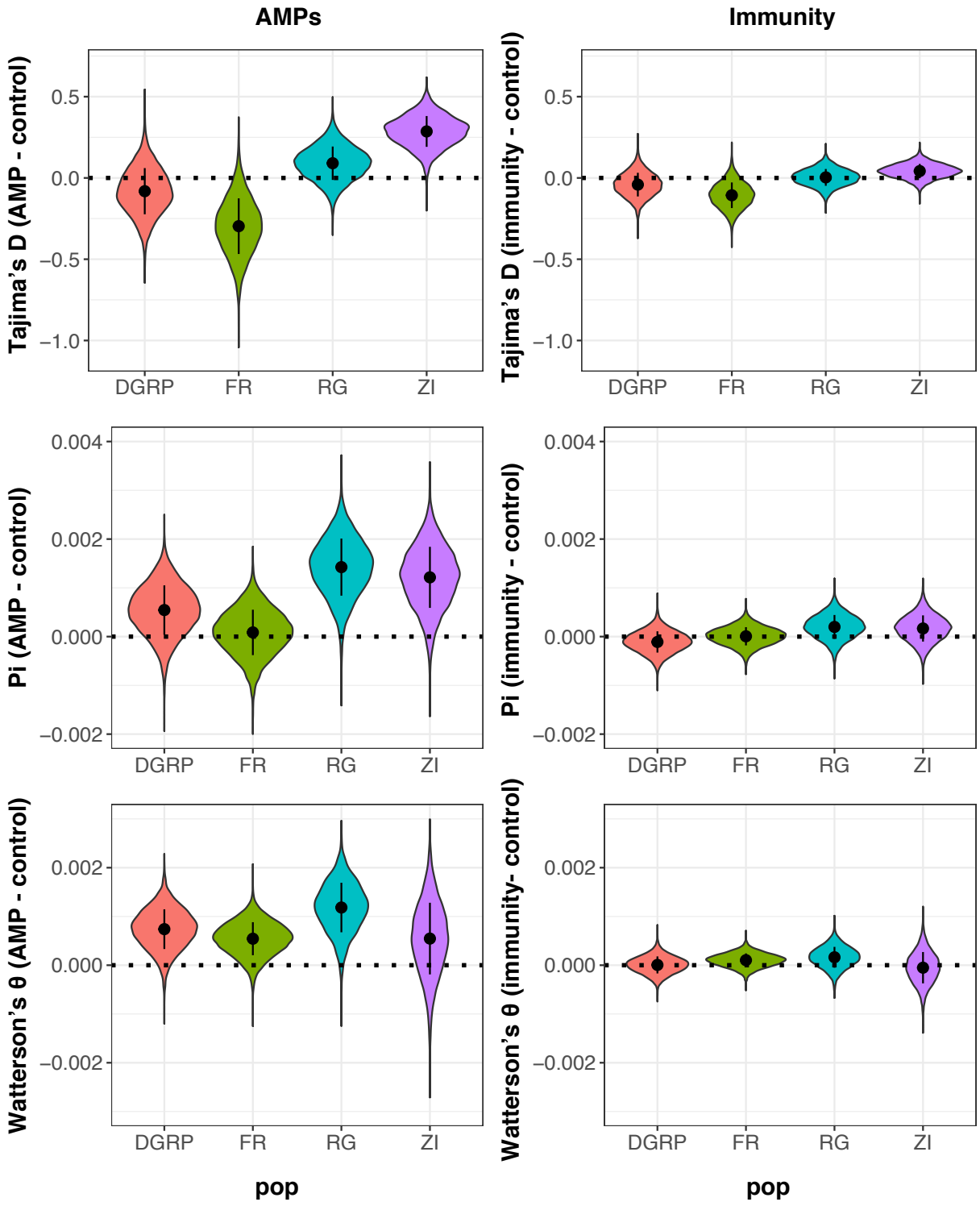
90. H. Benachour, M. Zaiou, A. Samara, B. Herbeth, M. Pfister, D. Lambert, G. Siest, and S. Visvikis-Siest, "Association of human cathelicidin (hCAP-18/LL-37) gene expression with cardiovascular disease risk factors," *Nutrition, Metabolism and Cardiovascular Diseases*, vol. 19, no. 10, pp. 720–728, 2009.

91. Y. Maneerat, K. Prasongsukarn, S. Benjathummarak, W. Dechkhajorn, and U. Chaisri, "Increased alpha-defensin expression is associated with risk of coronary heart disease: a feasible predictive inflammatory biomarker of coronary heart disease in hyperlipidemia patients," *Lipids in Health and Disease*, vol. 15, p. 117, 2016.

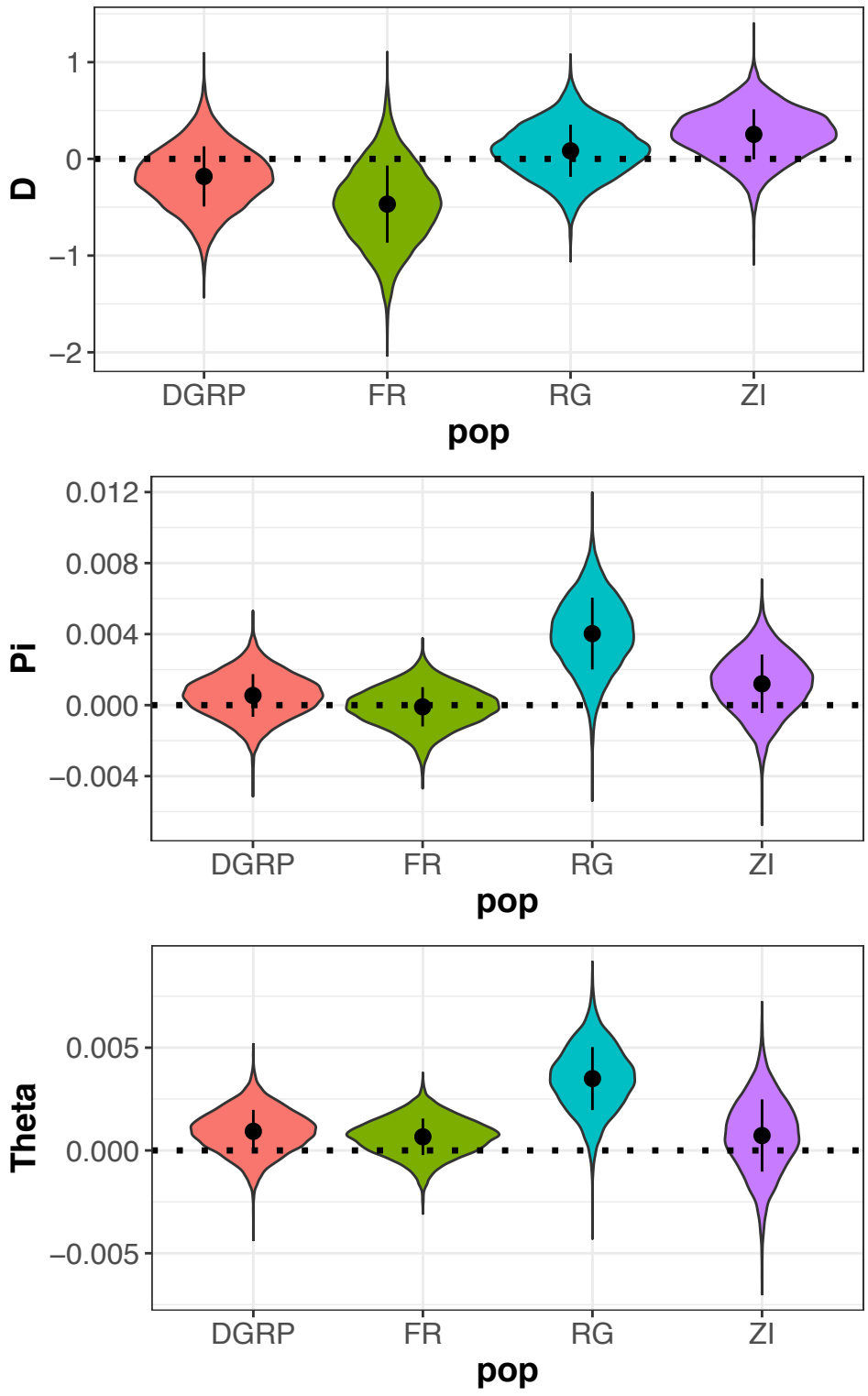
92. J. E. Pool, R. B. Corbett-Detig, R. P. Sugino, K. A. Stevens, C. M. Cardeno, M. W. Crepeau, P. Duchen, J. J. Emerson, P. Saelao, D. J. Begun, and C. H. Langley, "Population genomics of Sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture," *PLoS Genetics*, vol. 8, no. 12, p. e1003080, 2012.

93. T. F. C. Mackay, S. Richards, E. a. Stone, A. Barbadilla, J. F. Ayroles, D. Zhu, *et al.*, "The *Drosophila melanogaster* Genetic Reference Panel," *Nature*, vol. 482, no. 7384, pp. 173–178, 2012.

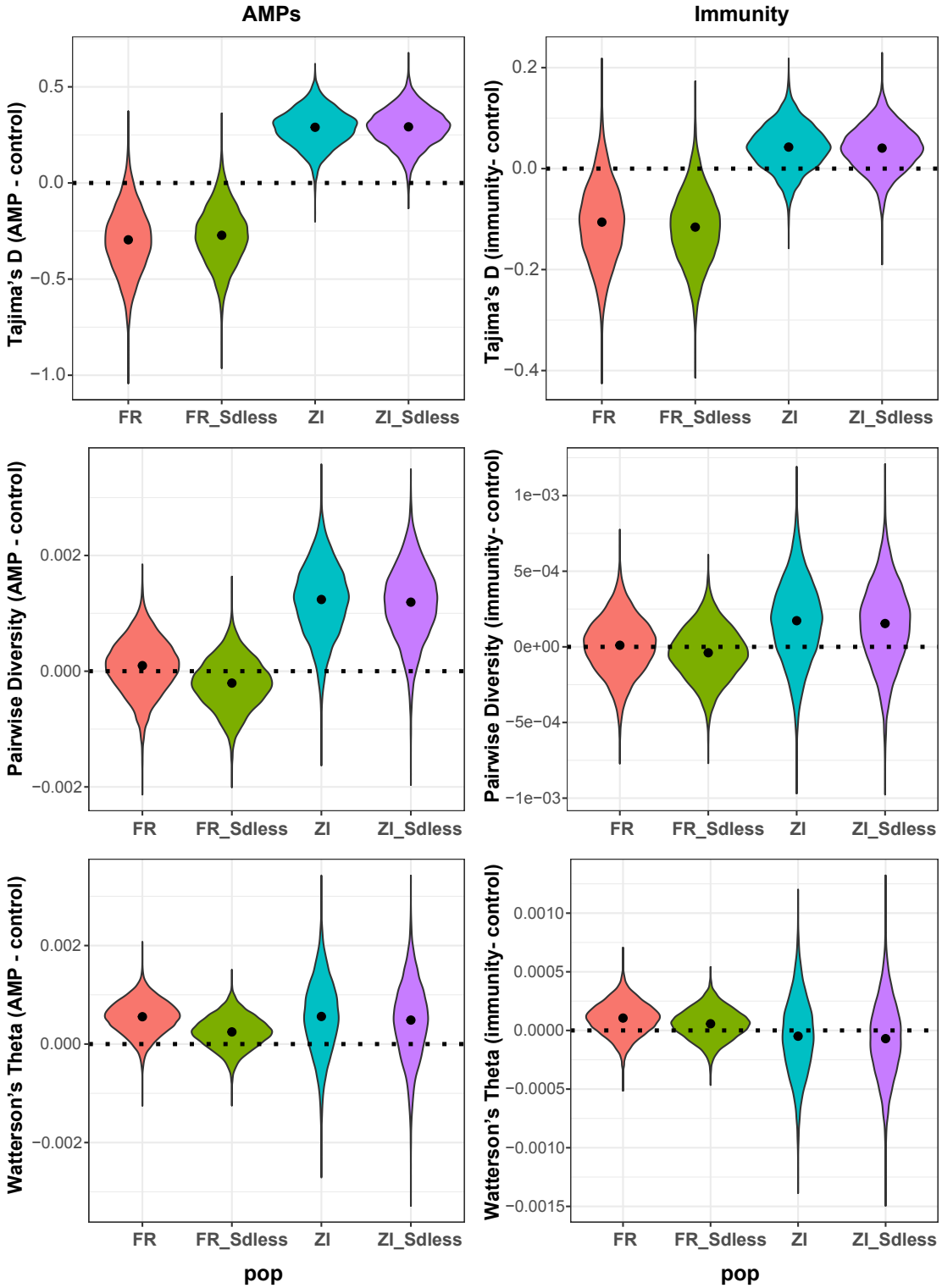
Supplementary Figure 1: Summary of resampling results (case - control) for AMPs and other immunity genes for Tajima's D, π, Watterson's θ.



Supplementary Figure 2: Summary of resampling results (case - control) for AMPs for Tajima's D, π, Watterson's θ, using the subset of non-linked AMPs.



Supplementary Figure 3: Summary of resampling results (case - control) for AMPs and other immunity genes for Tajima's D, π, Watterson's θ, comparing the results of ZI and FR populations with and without SD chromosomes.



Supplementary Figure 4: Summary of resampling results (case - control) for AMPs and other immunity genes for Tajima's D, π, Watterson's θ in the *D. mauritiana* population.