Balancing Selection Drives the Maintenance of Genetic Variation in *Drosophila* Antimicrobial Peptides

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

Department of Molecular Biosciences, University of Kansas

[†]These authors contributed equally to this work.

*Corresponding author: E-mail: unckless@ku.edu.

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 that AMPs may be under directional, balancing, or purifying selection. Here, we use both a linear model and control-gene-based approach to show that balancing selection is an important force shaping AMP diversity in *Drosophila*. In *Drosophila melanogaster*, this is most clearly observed in ancestral African populations. Furthermore, the signature of balancing selection is even more striking once background selection has been accounted for. Balancing selection also acts on AMPs in *Drosophila mauritiana*, an isolated island endemic separated from *D. melanogaster* by about 4 Myr of evolution. This suggests that balancing selection may be broadly acting to maintain adaptive diversity in *Drosophila* AMPs, as has been found in other taxa.

Key words: antimicrobial peptides (AMPs), host-defense peptides (HDPs), population genetics, immunity, coevolution, *Sophophora*.

Introduction

Pathogens exert strong selective pressures on their hosts, both in terms of individual fitness and the evolutionary trajectory of populations and species (Anderson and May 1981). Coevolutionary dynamics of hosts and pathogens results in continual selection for adaptive improvements in both players, often referred to as a coevolutionary arms race (Paterson et al. 2010; Schulte et al. 2010; Thrall et al. 2012). 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 (Nielsen et al. 2005; Kosiol et al. 2008; Downing et al. 2009a; Ekblom et al. 2010; McTaggart et al. 2012).

However, resistance mutations may not always become fixed. Balancing selection is the process whereby polymorphism is adaptively maintained over extended timescales, sometimes described as trench-warfare dynamics (Stahl et al. 1999). Several processes are thought to contribute to balancing selection (reviewed by Llaurens et al. [2017]). 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. These include the form and strength of selection (Kingsolver and Pfennig 2007); the genetic architecture (e.g., dominance, epistasis) of immune traits (Mackay 2001); the timescale upon which selection is acting (Charlesworth 2006); the density, diversity, and virulence of pathogens (Lambrechts et al. 2006); the cost of maintaining resistance alleles in the absence of infection (Unckless and Lazzaro 2016); effective population size (Charlesworth 2009); the mutation and recombination rates of hosts and pathogens

(Gandon and Michalakis 2002); environmental variables (Wolinska and King 2009); and demographic factors such as gene flow and bottlenecks (Brockhurst et al. 2003).

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 Hughes and Yeager [1997], Edwards and Hedrick [1998], Hedrick [1998], and Bernatchez and Landry [2003]). 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 in humans (Ferrer-Admetlla et al. 2008), deer (Quéméré et al. 2015, 2018), bank voles (Kloch et al. 2018), and birds (Alcaide and Edwards 2011; Gilroy et al. 2017; Velová et al. 2018); various cytokine genes (particularly interleukins) in humans (Hughes et al. 2005; Wilson et al. 2006; Ferrer-Admetlla et al. 2008; Fumagalli et al. 2009), birds (Downing et al. 2009a, 2009b, 2010), and field voles (Turner et al. 2012); and viral resistance genes including Oas1b in mice (Ferguson et al. 2008), OAS1 in primates (Ferguson et al. 2012; Fish and Boissinot 2015), and TRIM5 in humans (Cagliani et al. 2010) and primates (Newman et al. 2006).

Balancing selection also appears to play a role in the evolution of antimicrobial peptides (AMPs) in some taxa. AMPs are effectors of innate immunity that are strongly induced upon infection (Lemaitre et al. 1997; Tzou et al. 2000). They are often membrane active (Shai 2002; Brogden 2005), with a direct role in killing and/or impeding the growth of pathogens (De Gregorio et al. 2002; Lemaitre and Hoffmann 2007). Balancing selection has been implicated as a driver of AMP evolution in a diverse array of species including birds (Hellgren and Sheldon 2011; Chapman et al. 2016), amphibians (Tennessen and Blouin 2008), fish (Halldórsdóttir and Árnason 2015), mollusks (Gosset et al. 2014), and humans (Cagliani et al. 2008; Hollox and Armour 2008).

The fruit fly, Drosophila melanogaster, is an important model for understanding evolution of the immune system (Hultmark 1993; Hoffmann 2003; Janssens and Beyaert 2003; Dostert et al. 2005; Lamiable et al. 2016). Directional selection on Drosophila immune genes appears to be a relatively widespread phenomenon, especially among antiviral, receptor, and signaling genes (Schlenke and Begun 2003; Obbard et al. 2006, 2009; Clark et al. 2007; Heger and Ponting 2007; Sackton et al. 2007; Hill et al. 2019). In contrast, evidence for balancing selection acting on Drosophila immune genes has been more equivocal. Genome-wide scans by Croze et al. (2016, 2017) found little evidence for balancing selection acting on immune genes. In contrast, both single gene and genome-wide analyses of selection have indicated that balancing selection may play an important role in the evolution of AMPs in Drosophila (Hanson et al. 2016; Unckless and Lazzaro 2016; Unckless et al. 2016). One

striking example is the AMP Diptericin. This AMP is subject to balancing selection in D. melanogaster, likely driven by a tradeoff between immune defense and another life-history trait (Unckless et al 2016). The same balanced polymorphism, achieved via a different mutation to the derived allele, is found in Drosophila simulans (Unckless et al 2016). Additionally, recent analyses have shown that both spatial and temporal fluctuations are associated with variation in D. melanogaster allele frequencies (Bergland et al. 2014), particularly in immune genes, including AMPs (Bergland et al. 2014; Early et al. 2017; Behrman et al. 2018).

AMPs play a key role in controlling pathogen load and infection outcome (De Gregorio et al. 2002; Lemaitre and Hoffmann 2007), which may be particularly important for insects and other invertebrates that lack an adaptive immune system. Given their direct interaction with pathogens, it is surprising that insect AMPs often do not show signatures of recurrent adaptive substitutions. We hypothesize that AMPs in Drosophila are prone to balancing selection. To test this hypothesis, we examined patterns of nucleotide variation at AMP, immune-, and control-gene loci in four populations of D. melanogaster and one population of Drosophila mauritiana. Using both a linear model and a matched control-gene-based approach, with standard population genetic statistics, we searched for molecular evolutionary signatures of selection on AMPs and immune genes. Our results provide evidence that balancing selection is an important driver of AMP evolution.

Results

Genetic Variation across Four D. melanogaster Populations

To determine whether AMPs show signatures of balancing selection, we obtained coding sequence alignments for 13,494 genes (including 35 AMPs and 288 genes putatively involved in immune defense; hereafter: immune genes) (Lack et al. 2016) for four D. melanogaster populations: Zambia (ZI), Rwanda (RG), France (FR), and North Carolina (DGRP) and quantified nucleotide polymorphism (supplementary table S1, Supplementary Material online). Drosophila melanogaster originated in sub-Saharan Africa, expanded into Europe ~15–16,000 years ago, and subsequently spread to North America <200 years ago (David and Capy 1988; Li and Stephan 2006; Keller 2007). The ZI and RG lines therefore represent ancestral populations, whereas FR and DGRP are derived populations. For each autosomal gene, we calculated three population genetic statistics: Watterson's θ (the number of segregating sites, corrected for sample size), π (pairwise nucleotide diversity), and Tajima's D across all populations, for silent (four-fold degenerate) sites, per after controlling for missing data. We limited our analyses to silent sites to allow us to reduce the possibility that our results were due to relaxed constraint on nonsynonymous variation rather than balancing selection.

Relaxed constraint at the protein level should increase non-synonymous diversity because selection against amino acid changes is lowered (Hartl and Clark 2006; Wang et al. 2016). Excluding nonsynonymous sites therefore allows us to specifically focus on the footprint of balancing selection. We then grouped genes as AMPs, immune genes, and background genes. The mean Tajima's *D* for AMPs is higher than the mean of background genes in all populations (ZI, −0.284 AMPs vs. −0.874 autosomal average; RG, −0.110 vs. −0.232; FR, −0.064 vs. −0.113; DGRP, −0.041 vs. −0.596, supplementary table S2, Supplementary Material online), consistent with relatively more balancing selection occurring in AMPs. As observed previously (e.g., Glinka et al. 2003; Shapiro et al. 2007), the autosome-wide average for Tajima's *D* is quite negative in *D. melanogaster*, which likely reflects a complex demographic history (supplementary fig. S1, Supplementary Material online).

As selection across the genome can be affected by differing levels of mutation and recombination, we next tested for differences in population genetic statistics between AMPs and the autosomal background after controlling for genomic position. We specifically tested whether AMPs have higher values of the three population genetic statistics by employing a linear model with four covariates: gene length, chromosome, chromosomal region (nested in chromosome, explained in more detail in the Materials and Methods), and gene type (AMP or not, nested in chromosomal region and chromosome). This revealed that population genetic measures were elevated for AMPs in ancestral populations (ZI and RG), but not derived populations (DGRP and FR) (table 1 and supplementary table S3 and supplementary fig. S1, Supplementary Material online).

AMP-Control Tests for Balancing Selection in *Drosophila*

Given the apparent differences in selection between AMPs and the background averages described above, we also employed an AMP-control approach to test whether AMPs showed elevated diversity (a signature of balancing selection) in *D. melanogaster* and *D. mauritiana* while controlling for local variation in mutation and recombination rates. This approach also allowed us to visualize the differences found in the linear models above. For each AMP, we randomly sampled genes of similar length (coding sequence length ≤10 times the size of the AMP) and genomic position (within 100,000 bp 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 procedure 10,000 times to obtain an empirical distribution of differences (fig. 1). Each AMP was associated with 8–30 control genes, resulting in each replicate containing a unique set of control genes. In these instances, a positive difference suggests a higher value for AMPs versus the control gene, consistent with balancing selection. Indeed, the differences are

Table 1

Linear Model for Various Population Genetic Statistics (Tajima's *D*, π, and Watterson's θ [θ_w]) Suggests AMPs Are Elevated, Consistent with Balancing Selection in Several *Drosophila* Populations

Pop.	Stat.	AMP (*F/P*)	Region (*F/P*)	Length (*F/P*)	Chr (*F/P*)
	df	12	9	1	2
DGRP	π	1.24/0.254	5.32/<0.001	1.18/0.278	6.35/0.002
	θ_w	3.14/<0.001	5.56/<0.001	0.682/0.410	3.38/0.035
	D	0.88/0.565	1.25/0.267	0.17/0.681	5.37/0.005
FR	π	0.73/0.714	5.50/<0.001	21.17/<0.001	1.05/0.350
	θ_w	3.87/<0.001	4.38/<0.001	28.23/<0.001	1.85/0.16
	D	1.68/0.077	5.31/<0.001	0.05/0.833	1.03/0.357
RG	π	3.56/<0.001	5.90/<0.001	1.07/0.302	3.14/0.045
	θ_w	3.33/<0.001	6.55/<0.001	1.02/0.313	3.43/0.034
	D	1.39/0.169	2.99/0.002	0.11/0.744	0.14/0.871
ZI	π	2.77/0.001	5.52/<0.001	0.93/0.336	3.53/0.030
	θ_w	1.84/0.042	6.31/<0.001	0.77/0.380	4.55/0.011
	D	2.70/0.002	3.15/0.001	0.49/0.484	0.18/0.837
	df	11	8	1	2
D. mau	π	1.55/0.117	3.48/<0.001	0.17/0.682	5.49/0.005
	θ_w	1.71/0.072	4.09/<0.001	0.26/0.608	14.00/<0.001
	D	1.56/0.113	2.12/0.035	0.54/0.463	5.23/0.006

Note.—These linear models include only genes within 100,000 bp and within ten times the size of an antimicrobial peptide. Data is presented as *F*-statistic/*P*-value from the linear model, with degrees of freedom (df) denoted in the second header row. *P* values <0.05 are in bold. AMP refers to AMP nested in region nested in chromosome and region refers to region nested in chromosome. Linear models were run individually for five *Drosophila* populations: four *D. melanogaster* populations (DGRP, *Drosophila* Genetics Reference Panel from North Carolina, USA; FR, France; RG, Rwanda; ZI, Zambia), and one *D. mauritiana* (*D. mau*) population. All three statistics (Tajima's *D*, π and θ_w) were calculated on silent (four-fold degenerate) sites only.

primarily positive for both π and Watterson's θ for all populations (fig. 1*B* and *C* and table 2). For Tajima's *D*, the differences are positive for Zambia and Rwanda (ancestral populations), consistent with balancing selection, but close to 0 for France and negative for DGRP (derived populations, fig. 1*A* and table 2). In line with our previous analyses, we found that AMPs had higher Tajima's *D* in both RG and ZI, but not FR or DGRP. These results were recapitulated when we subsampled AMPs to control for the fact that they often cluster in the genome (see Materials and Methods, supplementary fig. S2, Supplementary Material online). We also examined all non-AMP immune genes using this control gene method and found little evidence of balancing selection in immune genes as a whole, in general concordance with Croze et al. (2017, 2016) (fig. 1 and supplementary table S4, Supplementary Material online).

Accounting for Background Selection Strengthens the Signature of Balancing Selection on *Drosophila* AMPs

Background selection, the removal of neutral variation due to selection against linked deleterious alleles, can influence levels of polymorphism across the genome. Comeron (2014) calculated the observed amount of background selection across the genome in 1,000-bp windows in the *D. melanogaster*
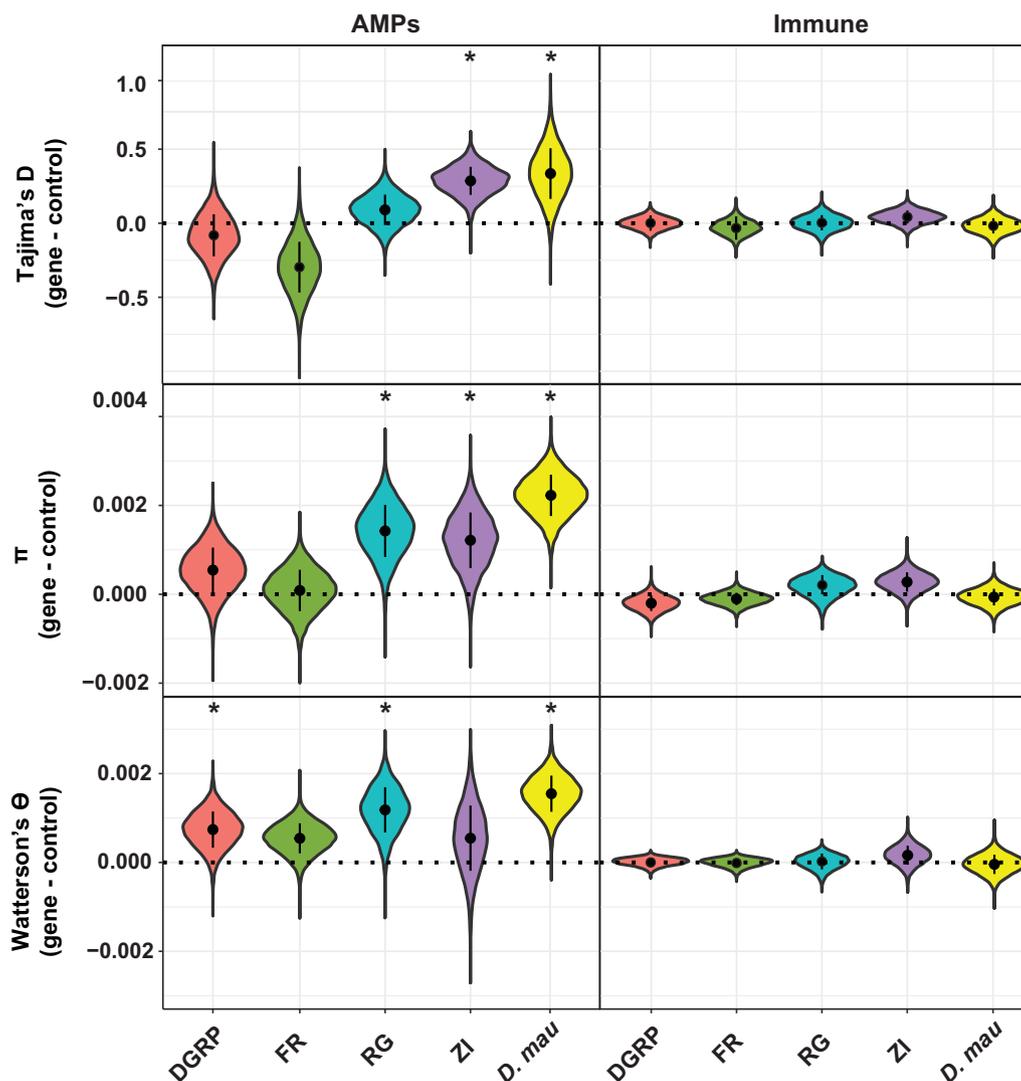
Fig. 1.—Overall, AMPs show more evidence for balancing selection than other immune genes. Difference in means between 35 AMPs and randomly chosen control genes (left-hand side) or 288 immune genes and randomly chosen control genes (right-hand side), resampled 10,000 times, separated by population (DGRP = *Drosophila* Genetics Reference Panel from North Carolina, USA; FR = France; RG = Rwanda; ZI = Zambia; *D. mau*, *D. mauritiana*). Top panel: Tajima's *D*; middle panel: π (nucleotide diversity); bottom panel: Watterson's θ. All three statistics were calculated on silent (four-fold degenerate) sites only. 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. Values above 0 are consistent of balancing selection. Asterisks indicate cases where <5% of resamplings have values <0.

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, whereas 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 (2014) as being under balancing selection, which further motivated us to examine the general pattern for AMPs. 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 (pink points, fig. 2A), indicating they are, in general, evolving in a manner consistent with balancing selection. In contrast, immune genes do not show elevated residuals compared with neighboring genes (teal points, fig. 2A). To further ascertain whether, as a group, AMPs show signatures of balancing selection, we used Comeron's background selection data (Comeron 2014) to fit a linear model as described above but also included Comeron's M1 statistic for background selection for a particular region as a covariate. In this case, AMPs showed significantly elevated silent polymorphism compared with other genes, whether we looked genome wide ($F_{10, 41533} = 16.66$, $P < 0.0001$) or focused on 100,000-bp regions of the genome containing AMPs ($F_{10, 1288} = 10.59$,

Table 2

Control Gene Comparisons Suggest AMPs Are Subject to Balancing Selection, Particularly in Ancestral Populations

AMP − Control Statistics	DGRP	FR	RG	ZI	*D. mau*
Tajima's *D* diff. > 0 (%)	28.7	4.1	81.4	99.7	98.1
Tajima's *D* Mean diff.	−0.084	−0.295	0.092	0.289	0.26
Tajima's *D* diff. std. dev.	0.142	0.171	0.102	0.092	0.12
π diff. > 0 (%)	85.9	58.4	98.9	96.9	100
π mean diff.	9.5×10^{-5}	9.6×10^{-5}	1.4×10^{-3}	1.2×10^{-3}	1.2×10^{-5}
π diff. std. dev.	5.5×10^{-4}	4.8×10^{-5}	5.5×10^{-3}	6.1×10^{-4}	2.1×10^{-6}
θ_{w} diff. > 0 (%)	96.2	93.7	98.5	77.4	99.9
θ_{w} mean diff.	7.5×10^{-4}	5.5×10^{-4}	1.2×10^{-3}	5.6×10^{-4}	1.7×10^{-5}
θ_{w} diff. std. dev.	4.1×10^{-4}	3.4×10^{-4}	5.1×10^{-4}	7.4×10^{-4}	1.5×10^{-6}

NOTE.—AMP minus control gene differences for three statistical measures (Tajima's *D*, π, and Watterson's θ [θ_{w}]) of selection in four *D. melanogaster* populations (DGRP, *Drosophila* Genetics Reference Panel from North Carolina, USA; FR, France; RG, Rwanda; ZI, Zambia), and one *D. mauritiana* (*D. mau*) population. First row per statistic: percentage (%) of 10,000 replicates in which the AMP minus control difference (diff.) was positive (>0), suggestive of balancing selection; second row: mean AMP minus control difference across 10,000 replicates; third row: standard deviation (std. dev.) of the mean. All three statistics (Tajima's *D*, π and θ_{w}) were calculated on silent (four-fold degenerate) sites only.
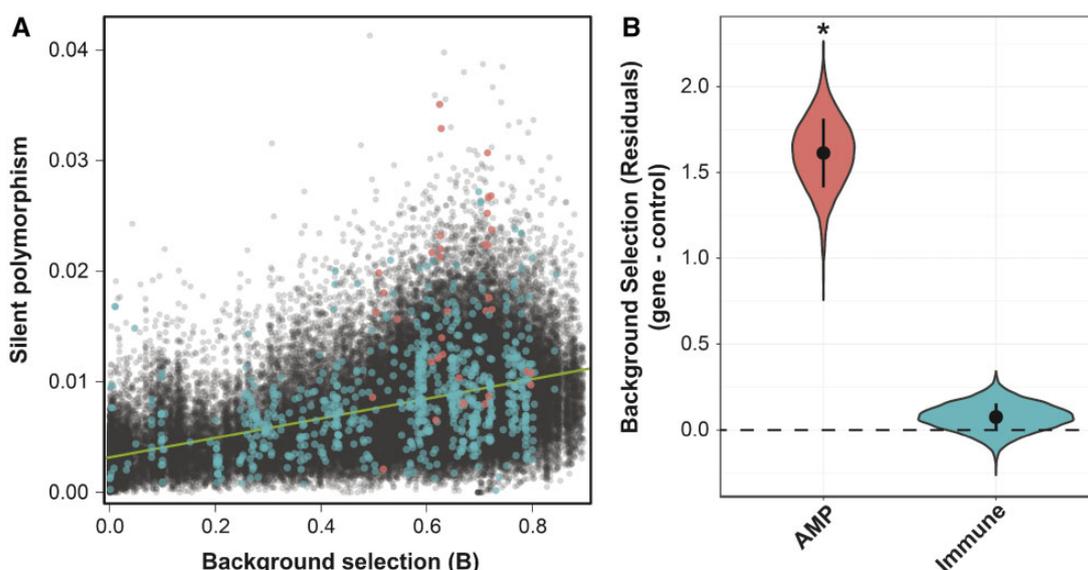

FIG. 2.—Accounting for background selection in the Rwanda (RG) population strengthens the signal of balancing selection on AMPs. (*A*) Correlation between silent polymorphism and the background selection statistic (*B*) in 1,000-bp windows for the Rwanda population of *D. melanogaster*. The line of best fit is in blue and regions containing AMPs are indicated by red dots. (*B*) Resampling of mean difference (AMP/immune minus control) in the residuals after regressing silent polymorphism against the background selection statistic *B*. Values above 0 are consistent of balancing selection. Asterisks indicate cases where <5% of resamplings have values <0.

$P < 0.0001$). We also found that regions containing AMPs, but not immune genes, were significantly elevated for residuals compared with nearby regions using our resampling approach (fig. 2*B*). This supports Comeron's assertion that accounting for background selection improves the ability to detect balancing selection (Comeron 2014) and also supports our previous results showing that AMPs as a group are likely subject to balancing selection.

Balancing Selection Also Acts on *D. mauritiana* AMPs

We also calculated population genetic statistics for 9,980 genes in 107 *D. mauritiana* isofemale lines, sequenced as a pool (Nolte et al. 2013). *D. mauritiana* is an island endemic which diverged from *D. melanogaster* ~3–5 Ma (Obbard et al. 2012; Nolte et al. 2013). SNP frequencies were called using PoPoolation, which accounts for low frequency variants and variation in coverage that may influence results from pooled samples (Kofler et al. 2011). As found for *D. melanogaster*, AMPs have a higher mean Tajima's *D* than background genes (−1.034 vs. −1.463). Linear models, as described above and in the Materials and Methods, revealed elevated Tajima's *D* for AMPs in *D. mauritiana*, whereas Watterson's θ and π were not significantly different when comparing AMPs and other genes in this species (fig. 1, table 1, and supplementary table S3, Supplementary Material online). Additionally, 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 (fig. 1, table 2, and supplementary table S4, Supplementary Material online). For other immune 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 in *D. mauritiana* (fig. 1 and supplementary tables S3 and S4, Supplementary Material online).

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 identified these 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 in the Rwandan population after using Comeron's (2014) method 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 AMPs. Third, this effect may be exacerbated by the fact that effector genes tend to be smaller (Lemaitre and Hoffmann 2007) than receptor and signaling genes, providing fewer sites and therefore less power to detect any population genetic signature. Relatedly, single gene analyses might lack the power to show such signatures. Fourth, patterns of nucleotide polymorphism are strongly influenced by population demographic history. Our AMP-control approach should account for the confounding influences of local mutation and recombination rate variation, gene size, and demography (Garrigan and Hedrick 2003).

The differences between derived (DGRP and France) and ancestral populations (Rwanda and Zambia) were striking. As populations establish in new habitats, they will encounter different pathogen pressures and prevailing environmental conditions. This could dramatically alter which alleles are selectively advantageous. First, bottlenecks may lead to the loss of one or more of the balanced alleles. Furthermore, loss of disadvantageous alleles (e.g., alleles resistant to pathogens not present in the new habitat) likely occurs more rapidly than establishment of new, beneficial polymorphisms (e.g., 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 coevolving with their pathogens, under semipredictable conditions, for long time-periods.

An alternate explanation for these differences could be that the quality of genomic data in the derived populations

is lower than that for the ancestral populations (supplementary table S1, Supplementary Material online). We therefore limited our analysis to a subset of 114 DGRP lines with the highest quality data (DGPR-HQ, being those lines with fewest Ns and highest coverage, see Materials and Methods). Though we did still did not find strong evidence for balancing selection acting on AMPs in this population (DGRP-HQ in supplementary table S2, Supplementary Material online), both measures of diversity (pairwise diversity and Watterson's theta) further increased in AMPs (but not other immune genes) when compared with the background, supporting an increase in diversity in AMPs compared with other genes. This also serves as a cautionary tale, as including lower quality genomes adds noise (increased standard deviation in resampling, supplementary table S2, Supplementary Material online) that could potentially mask population genetic signals.

It is tempting to look to newly developed methods for detecting balancing selection (e.g. DeGiorgio et al. 2014; Siewert and Voight 2017), 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. (2014), based on Gao et al. (2015), 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 1,000 bp (DeGiorgio et al. 2014). Estimates of ρ in *D. melanogaster* are highest in the DGRP population and range from 9.6 to 14.8 per kb for the different chromosomes (Chan et al. 2012). These values correspond to windows of 100 bp or less in *D. melanogaster*. Given estimates of nucleotide diversity between 0.001 and 0.01 (Lack et al. 2016), we expect less than one segregating site per window, rendering these tests unusable in this species. Recombination is even higher in *D. mauritiana* (True et al. 1996), and the use of pooled sequencing (Pool-seq) data for this species would further complicate the interpretation of newer selection statistics, due to the fact that low frequency alleles are dropped in Pool-seq data.

We find that, at least in ancestral populations, AMPs tend to evolve in a manner consistent with balancing selection, showing increased diversity but no increase in divergence (Unckless and Lazzaro 2016) relative to other genes. This contrasts with other immune genes which show no such pattern. Why are AMPs different than other immune genes? One characteristic of AMPs is that they interact directly with microbes (Bulet and Stocklin 2005), and, in some cases, AMP sequence is directly linked to the efficacy of bacterial membrane interactions (Schmitt et al. 2016; Franzoi et al. 2017). 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 (Behrman et al. 2018) and spatial (Early et al. 2017) 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., Tennessen et al. 2009; Hellgren et al. 2010; Park et al. 2014; Unckless et al. 2016). Additionally, the patterns of divergence and polymorphism expected after periods of fluctuating selection are not clear-cut and may complicate the detection of balancing selection (see, e.g., Huerta-Sanchez et al. 2008; Miura et al. 2013; Gossmann et al. 2014).

Alternatively, AMP variation might be maintained because AMP alleles that are more effective against pathogens also tend to 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 (Gilliet and Lande 2008; Benachour et al. 2009; Maneerat et al. 2016). 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 directional selection, consistent with an arms race model (Obbard et al. 2006, 2009; Sackton et al. 2007). Distinguishing between these two hypotheses for the adaptive maintenance of AMP genetic variation will take careful functional analysis.

Materials and Methods

Polymorphism in Four Populations of *D. melanogaster*

We downloaded chromosome sequences for the Zambia (ZI, $n = 197$), Rwanda (RG, $n = 27$), *Drosophila* 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; last accessed September 10, 2019. (Mackay et al. 2012; Pool et al. 2012). These data were collected as described elsewhere (Pool et al. 2012). Briefly, short read Illumina data were generated for each individual and mapped to the *D. melanogaster* reference genome version 5.22 using BWA and Stampy. Following mapping and alignment, GATK indel realigner was used to refine short indel alignments and SNPs were called for each individual. A custom reference genome was then generated for each individual based on all SNPs called and the process was repeated. Following a second round of mapping and indel realignment, SNPs called across both rounds were

then inserted to create a second, final custom genome for each individual (Pool et al. 2012).

We converted these sequences into FASTA files, per chromosome, for each population. 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 S1, Supplementary Material online). Additionally, we created a subset of 114 DGRP lines (hereafter DGRP-HQ) with high coverage to calculate statistics in DGRP using only high-quality data (at least 25-fold coverage, and at least 75% of sites called in 1,000-bp windows across the genome).

Using annotation 5.57 of the *D. melanogaster* genome, we extracted FASTA alignments for each gene and extracted silent (four-fold degenerate) sites using a custom Biopython script (personal comm. Yasir Ahmed, May 2018). We then used a custom Python script with the package DendroPy, to find π, Watterson's θ, Tajima's D, and the number of segregating sites for the subset of silent sites per gene (Sukumaran and Holder 2010; Ferretti et al. 2012). We categorized genes involved in immune defense (hereafter immune genes) using the designations given at https://www.epfl.ch/labs/lemaitre-lab/lemaitre-lab/resources/list-of-drosophila-genes-potentially-involved-in-the-immune-response/; last accessed September 10, 2019. (Lemaitre and Hoffmann 2007). We removed nonautosomal genes from all downstream analyses, because the X chromosome does not harbor any AMPs, and has a different effective population size than autosomes.

We employed a linear model in the form of $Y = chromosome + gene\ length + chromosome/region + chromosome/region/AMP$, where Y refers to the value of a particular statistic for that gene, region refers either to the region of the genome in which the gene is found (200,000-bp width) or a window around the focal AMP, and AMP refers to whether or not each gene was an AMP. Slashes such as in *chromosome/region* refer to *region* nested in *chromosome*. For the analysis of AMP silent polymorphism while accounting for background selection, we employed a linear model in the form of $Y = M1 + region/AMP$ (Comeron 2014). All linear models included only genes on chromosomes 2L, 2R, and 3R, as these are the only chromosomes harboring AMPs in *D. melanogaster* and *D. mauritiana*. ANOVA results were analyzed using the *car* package (Fox and Weisberg 2011) in R v 3.5.1 (R Core Team 2017). The results of these models are summarized in supplementary table S3, Supplementary Material online.

To control for missing data, we removed all sequences containing over 25% ambiguous bases and recalculated each statistic using DendroPy v 4.4.0 in Python v 2.7.0, accounting for missing data in our calculation (Sukumaran and Holder 2010; Ferretti et al. 2012). We again fitted a linear model to this data, for each AMP and those control genes within 10,000, 50,000, and 100,000 bp of that AMP.

For each population, we then resampled to find the average difference in scores between AMP/immune genes and control genes. Genes were assigned as AMPs or immune genes based on data from Bruno Lemaitre. For each gene in these categories (AMP or immune), we randomly sampled a control gene in a 200,000-bp window centered on the AMP (100,000-bp upstream or downstream), that was no more than ten times larger than the focal gene (defined as coding sequence length measured in base pairs), and not another gene in the given category. We then found the average difference ($\bar{\Delta}$) in each measure for the focal gene (AMP or immune) group and the control group such that

$$\bar{\Delta} = \frac{1}{n} \sum_{i=1}^{n} (X\ \text{AMP/Immune}_i - X_{\text{Control}_i}),$$

where X AMP/Immune represents the chosen AMP/immune gene, X_{Control} represents the randomly sampled control gene, and n accounts for the number of genes in the group. We then resampled 10,000 times to obtain an empirical distribution of the differences. Each AMP or immune gene was associated with between 8 and 47 control genes. With 10,000 resamplings, we expect only one particular combination of control genes to be chosen twice. We tested a range of window sizes (50,000–150,000 bp) and found qualitatively similar results (data not shown). To account for the fact that many AMPs cluster in the genome which might lead to pseudoreplication, we additionally repeated our analysis with a subsampled data set. To this end, we choose a subset of ten AMPs that were at least 5,000 bp from any other included AMPs and repeated the AMP-control analysis described above, with the specific subset of ten AMPs changing randomly per resampling.

We employed this method to control for genome-wide variation in recombination rates, mutation rates. Resampling 10,000 times allows for a robust empirical distribution that does not rely on the particular control genes chosen per iteration. We therefore present the distribution of differences as violin plots with the proportion of resamplings that do not overlap zero, analogous to a bootstrap value.

Polymorphism in a Population of *D. mauritiana*

We downloaded the reference genome, annotation, and mapped BAM file of a population of *D. mauritiana* sequenced as a pool (Pool-seq) from http://www.popoolation.at/mauritiana_genome; last accessed September 10, 2019. (Nolte et al. 2013), and used PoPoolation to calculate π, Watterson's θ, and Tajima's D for each gene in this population. We then resampled to find the average difference in scores between AMPs/immune and a control set of genes, as described above.

Data Availability

All data used in this study are publicly available and freely accessible. The *D. melanogaster* sequence data were obtained from John Pool's *Drosophila* Genome Nexus (http://www.johnpool.net/genomes.html; last accessed September 10, 2019) in FASTA format. All *D. mauritiana* data (BAM file, reference genome and gene annotation) were downloaded from http://www.popoolation.at/mauritiana_genome; last accessed September 10, 2019.

Literature Cited

Alcaide M, Edwards SV. 2011. Molecular evolution of the Toll-like receptor multigene family in birds. Mol Biol Evol. 28(5):1703–1715.

Anderson RM, May RM. 1981. The population dynamics of microparasites and their invertebrate hosts. Philos Trans R Soc Lond B Biol Sci. 291(1054):451–524.

Behrman EL, et al. 2018. Rapid seasonal evolution in innate immunity of wild *Drosophila melanogaster*. Proc R Soc B 285(1870):20172599.

Benachour H, et al. 2009. Association of human cathelicidin (hCAP-18/LL-37) gene expression with cardiovascular disease risk factors. Nutr Metab Cardiovasc Dis. 19(10):720–728.

Bergland AO, Behrman EL, O'Brien KR, Schmidt PS, Petrov DA. 2014. Genomic evidence of rapid and stable adaptive oscillations over seasonal time scales in *Drosophila*. PLoS Genet. 10(11):e1004775.

Bernatchez L, Landry C. 2003. MHC studies in nonmodel vertebrates: what have we learned about natural selection in 15 years? J Evol Biol. 16(3):363–377.

Brockhurst MA, Morgan AD, Rainey PB, Buckling A. 2003. Population mixing accelerates coevolution. Ecol Lett. 6(11):975–979.

Brogden KA. 2005. Antimicrobial peptides: pore formers or metabolic inhibitors in bacteria? Nat Rev Microbiol. 3(3):238–250.

Bulet P, Stocklin R. 2005. Insect antimicrobial peptides: structures, properties and gene regulation. Protein Pept Lett. 12(1):3–11.

Cagliani R, et al. 2008. The signature of long-standing balancing selection at the human defensin beta-1 promoter. Genome Biol. 9(9):R143.

Cagliani R, et al. 2010. Long-term balancing selection maintains *trans*-specific polymorphisms in the human *TRIM5* gene. Hum Genet. 128(6):577–588.

Chan AH, Jenkins PA, Song YS. 2012. Genome-wide fine-scale recombination rate variation in *Drosophila melanogaster*. PLoS Genet. 8(12):e1003090.

Chapman JR, et al. 2016. The evolution of innate immune genes: purifying and balancing selection on *β*-defensins in waterfowl. Mol Biol Evol. 33(12):3075–3087.

Charlesworth B. 2009. Effective population size and patterns of molecular evolution and variation. Nat Rev Genet. 10(3):195–205.

Charlesworth D. 2006. Balancing selection and its effects on sequences in nearby genome regions. PLoS Genet. 2(4):e64.

Clark AG, et al. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. Nature 450(7167):203–218.

Comeron JM. 2014. Background selection as baseline for nucleotide variation across the *Drosophila* genome. PLoS Genet. 10(6): e1004434.

Croze M, Živković D, Stephan W, Hutter S. 2016. Balancing selection on immunity genes: review of the current literature and new analysis in *Drosophila melanogaster*. Zoology 119(4):322–329.

Croze M, et al. 2017. A genome-wide scan for genes under balancing selection in *Drosophila melanogaster*. BMC Evol Biol. 17(1):15.

David JR, Capy P. 1988. Genetic variation of *Drosophila melanogaster* natural populations. Trends Genet. 4(4):106–111.

De Gregorio E, Spellman PT, Tzou P, Rubin GM, Lemaitre B. 2002. The Toll and Imd pathways are the major regulators of the immune response in *Drosophila*. EMBO J. 21(11):2568–2579.

DeGiorgio M, Lohmueller KE, Nielsen R. 2014. A model-based approach for identifying signatures of ancient balancing selection in genetic data. PLoS Genet. 10(8):e1004561.

Dostert C, et al. 2005. The Jak-STAT signaling pathway is required but not sufficient for the antiviral response of *Drosophila*. Nat Immunol. 6(9):946–953.

Downing T, et al. 2009a. Evidence of balanced diversity at the chicken interleukin 4 receptor alpha chain locus. BMC Evol Biol. 9(1):136.

Downing T, et al. 2009b. Contrasting evolution of diversity at two disease-associated chicken genes. Immunogenetics 61(4):303–314.

Downing T, Lloyd AT, O'Farrelly C, Bradley DG. 2010. The differential evolutionary dynamics of avian cytokine and TLR gene classes. J Immunol. 184(12):6993–7000.

Early AM, et al. 2017. Survey of global genetic diversity within the *Drosophila* immune system. Genetics 205(1):353–366.

Edwards SV, Hedrick PW. 1998. Evolution and ecology of MHC molecules: from genomics to sexual selection. Trends Ecol Evol. 13(8):305–311.

Ekblom R, French L, Slate J, Burke T. 2010. Evolutionary analysis and expression profiling of Zebra Finch immune genes. Genome Biol Evol. 2:781–790.

Ferguson W, Dvora S, Fikes RW, Stone AC, Boissinot S. 2012. Long-term balancing selection at the antiviral gene OAS1 in Central African chimpanzees. Mol Biol Evol. 29(4):1093–1103.

Ferguson W, Dvora S, Gallo J, Orth A, Boissinot S. 2008. Long-term balancing selection at the West Nile virus resistance gene, *Oas1b*, maintains transspecific polymorphisms in the house mouse. Mol Biol Evol. 25(8):1609–1618.

Ferrer-Admetlla A, et al. 2008. Balancing selection is the main force shaping the evolution of innate immunity genes. J Immunol. 181(2):1315–1322.

Ferretti L, Raineri E, Ramos-Onsins S. 2012. Neutrality tests for sequences with missing data. Genetics 191(4):1397–1401.

Fish I, Boissinot S. 2015. Contrasted patterns of variation and evolutionary convergence at the antiviral OAS1 gene in old world primates. Immunogenetics 67(9):487–499.

Fox J, Weisberg S. 2011. An R Companion to Applied Regression. Thousand Oaks (CA): SAGE Publications.

Franzoi M, et al. 2017. Structural insights into the mode of action of the peptide antibiotic Copsin. Biochemistry 56(37):4992–5001.

Fumagalli M, et al. 2009. Parasites represent a major selective force for interleukin genes and shape the genetic predisposition to autoimmune conditions. J Exp Med. 206(6):1395–1408.

Gandon S, Michalakis Y. 2002. Local adaptation, evolutionary potential and host-parasite coevolution: interactions between migration, mutation, population size and generation time. J Evol Biol. 15(3):451–462.

Gao ZY, Przeworski M, Sella G. 2015. Footprints of ancient-balanced polymorphisms in genetic variation data from closely related species. Evolution 69(2):431–446.

Garrigan D, Hedrick PW. 2003. Perspective: detecting adaptive molecular polymorphism: lessons from the MHC. Evolution 57(8):1707–1722.

Gilliet M, Lande R. 2008. Antimicrobial peptides and self-DNA in autoimmune skin inflammation. Curr Opin Immunol. 20(4):401–407.

Gilroy DL, Phillips KP, Richardson DS, van Oosterhout C. 2017. 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'. J Evol Biol. 30(7):1276–1287.

Glinka S, Ometto L, Mousset S, Stephan W, De Lorenzo D. 2003. Demography and natural selection have shaped genetic variation in *Drosophila melanogaster*: a multi-locus approach. Genetics 165(3):1269–1278.

Gosset CC, Do Nascimento J, Augé MT, Bierne N. 2014. Evidence for adaptation from standing genetic variation on an antimicrobial peptide gene in the mussel *Mytilus edulis*. Mol Ecol. 23(12):3000–3012.

Gossmann TI, Waxman D, Eyre-Walker A. 2014. Fluctuating selection models and McDonald–Kreitman type analyses. PLoS One 9(1):e84540.

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

Hanson MA, Hamilton PT, Perlman SJ. 2016. Immune genes and divergent antimicrobial peptides in flies of the subgenus *Drosophila*. BMC Evol Biol. 16(1):228.

Hartl DL, Clark AG. 2006. Principles of Population Genetics. Sunderland (MA): Sinauer Associates, Inc.

Hedrick PW. 1998. Balancing selection and MHC. Genetica 104(3):207–214.

Heger A, Ponting CP. 2007. Evolutionary rate analyses of orthologs and paralogs from 12 *Drosophila* genomes. Genome Res. 17(12):1837–1849.

Hellgren O, Sheldon BC. 2011. 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. Mol Ecol Resour. 11(4):686–692.

Hellgren O, Sheldon BC, Buckling A. 2010. *In vitro* tests of natural allelic variation of innate immune genes (avian *β*-defensins) reveal functional differences in microbial inhibition. J Evol Biol. 23(12):2726–2730.

Hill T, Koseva B, Unckless RL. 2019. The genome of *Drosophila innubila* reveals lineage-specific patterns of selection in immune genes. Mol Biol Evol. 36(7):1405–1417.

Hoffmann JA. 2003. The immune response of *Drosophila*. Nature 426(6962):33–38.

Hollox EJ, Armour JAL. 2008. Directional and balancing selection in human beta-defensins. BMC Evol Biol. 8(1):113.

Huerta-Sanchez E, Durrefft R, Bustamante CD. 2008. Population genetics of polymorphism and divergence under fluctuating selection. Genetics 178(1):325–337.

Hughes AL, Packer B, Welch R, Chanock SJ, Yeager M. 2005. High level of functional polymorphism indicates a unique role of natural selection at human immune system loci. Immunogenetics 57(11):821–827.

Hughes AL, Yeager M. 1997. Coordinated amino acid changes in the evolution of mammalian defensins. J Mol Evol. 44(6):675–682.

Hultmark D. 1993. Immune reactions in *Drosophila* and other insects: a model for innate immunity. Trends Genet. 9(5):178–183.

Janssens S, Beyaert R. 2003. Role of Toll-like receptors in pathogen recognition. Clin Microbiol Rev. 16(4):637–646.

Keller A. 2007. *Drosophila melanogaster*'s history as a human commensal. Curr Biol. 17(3):R77–R81.

Kingsolver JG, Pfennig DW. 2007. Patterns and power of phenotypic selection in nature. Bioscience 57(7):561–572.

Kloch A, et al. 2018. Signatures of balancing selection in toll-like receptor (TLRs) genes—novel insights from a free-living rodent. Sci Rep. 8(1):8361.

Kofler R, et al. 2011. PoPoolation: a toolbox for population genetic analysis of next generation sequencing data from pooled individuals. PLoS One 6(1):e15925.

Kosiol C, et al. 2008. Patterns of positive selection in six mammalian genomes. PLoS Genet. 4(8):e1000144.

Lack JB, Lange JD, Tang AD, Corbett-Detig RB, Pool JE. 2016. A thousand fly genomes: an expanded *Drosophila* Genome Nexus. Mol Biol Evol. 33(12):3308–3313.

Lambrechts L, Fellous S, Koella JC. 2006. Coevolutionary interactions between host and parasite genotypes. Trends Parasitol. 22(1):12–16.

Lamiable O, et al. 2016. Cytokine Diedel and a viral homologue suppress the IMD pathway in *Drosophila*. Proc Natl Acad Sci U S A. 113(3):698–703.

Lemaitre B, Hoffmann J. 2007. The host defense of *Drosophila melanogaster*. Annu Rev Immunol. 25(1):697–743.

Lemaitre B, Reichhart JM, Hoffmann JA. 1997. *Drosophila* host defense: differential induction of antimicrobial peptide genes after infection by various classes of microorganisms. Proc Natl Acad Sci U S A. 94(26):14614–14619.

Li H, Stephan W. 2006. Inferring the demographic history and rate of adaptive substitution in *Drosophila*. PLoS Genet. 2(10):e166.

Llaurens V, Whibley A, Joron M. 2017. Genetic architecture and balancing selection: the life and death of differentiated variants. Mol Ecol. 26(9):2430–2448.

Mackay TFC. 2001. The genetic architecture of quantitative traits. Annu Rev Genet. 35(1):303–339.

Mackay TFC, et al. 2012. The *Drosophila melanogaster* Genetic Reference Panel. Nature 482(7384):173–178.

Maneerat Y, Prasongsukarn K, Benjathummarak S, Dechkhajorn W, Chaisri U. 2016. 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 Health Dis. 15(1):117.

McTaggart SJ, Obbard DJ, Conlon C, Little TJ. 2012. Immune genes undergo more adaptive evolution than non-immune system genes in *Daphnia pulex*. BMC Evol Biol. 12(1):63.

Miura S, Zhang ZG, Nei M. 2013. Random fluctuation of selection coefficients and the extent of nucleotide variation in human populations. Proc Natl Acad Sci U S A. 110(26):10676–10681.

Newman RM, et al. 2006. Balancing selection and the evolution of functional polymorphism in Old World monkey TRIM5α. Proc Natl Acad Sci U S A. 103(50):19134–19139.

Nielsen R, et al. 2005. A scan for positively selected genes in the genomes of humans and chimpanzees. PLoS Biol. 3(6):e170.

Nolte V, Pandey RV, Kofler R, Schlötterer C. 2013. Genome-wide patterns of natural variation reveal strong selective sweeps and ongoing genomic conflict in *Drosophila mauritiana*. Genome Res. 23(1):99–110.

Obbard DJ, Jiggins FM, Halligan DL, Little TJ. 2006. Natural selection drives extremely rapid evolution in antiviral RNAi genes. Curr Biol. 16(6):580–585.

Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009. Quantifying adaptive evolution in the *Drosophila* immune system. PLoS Genet. 5(10):e1000698.

Obbard DJ, et al. 2012. Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. Mol Biol Evol. 29(11):3459–3473.

Park D, et al. 2014. Functional characterization of naturally occurring melittin peptide isoforms in two honey bee species, *Apis mellifera* and *Apis cerana*. Peptides 53:185–193.

Paterson S, et al. 2010. Antagonistic coevolution accelerates molecular evolution. Nature 464(7286):275–278.

Pool JE, et al. 2012. Population genomics of sub-Saharan *Drosophila melanogaster*: African diversity and non-African admixture. PLoS Genet. 8(12):e1003080.

Quéméré E, et al. 2015. Immunogenetic heterogeneity in a widespread ungulate: the European roe deer (*Capreolus capreolus*). Mol Ecol. 24(15):3873–3887.

Quéméré E, et al. 2018. Fluctuating pathogen-mediated selection drives the maintenance of innate immune gene polymorphism in a widespread wild ungulate. bioRxiv. doi: 10.1101/458216.

R Core Team. 2017. R: a language and environment for statistical computing. Vienna (Austria): R Foundation for Statistical Computing.

Sackton TB, et al. 2007. Dynamic evolution of the innate immune system in *Drosophila*. Nat Genet. 39(12):1461–1468.

Schlenke TA, Begun DJ. 2003. Natural selection drives *Drosophila* immune system evolution. Genetics 164(4):1471–1480.

Schmitt P, Rosa RD, Destoumieux-Garzón D. 2016. An intimate link between antimicrobial peptide sequence diversity and binding to essential components of bacterial membranes. Biochim Biophys Acta Biomembr. 1858(5):958–970.

Schulte RD, Makus C, Hasert B, Michiels NK, Schulenburg H. 2010. Multiple reciprocal adaptations and rapid genetic change upon experimental coevolution of an animal host and its microbial parasite. Proc Natl Acad Sci U S A. 107(16):7359–7364.

Shai Y. 2002. Mode of action of membrane active antimicrobial peptides. Pept Sci. 66(4):236–248.

Shapiro JA, et al. 2007. Adaptive genic evolution in the *Drosophila* genomes. Proc Natl Acad Sci U S A. 104(7):2271–2276.

Siewert KM, Voight BF. 2017. Detecting long-term balancing selection using allele frequency correlation. Mol Biol Evol. 34(11):2996–3005.

Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. 1999. Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. Nature 400(6745):667–671.

Sukumaran J, Holder MT. 2010. DendroPy: a Python library for phylogenetic computing. Bioinformatics 26(12):1569–1571.

Tennessen JA, Blouin MS. 2008. Balancing selection at a frog antimicrobial peptide locus: fluctuating immune effector alleles? Mol Biol Evol. 25(12):2669–2680.

Tennessen JA, et al. 2009. Variations in the expressed antimicrobial peptide repertoire of northern leopard frog (*Rana pipiens*) populations suggest intraspecies differences in resistance to pathogens. Dev Comp Immunol. 33(12):1247–1257.

Thrall PH, et al. 2012. Rapid genetic change underpins antagonistic coevolution in a natural host–pathogen metapopulation. Ecol Lett. 15(5):425–435.

True JR, Mercer JM, Laurie CC. 1996. Differences in crossover frequency and distribution among three sibling species of *Drosophila*. Genetics 142(2):507–523.

Turner AK, Begon M, Jackson JA, Paterson S. 2012. Evidence for selection at cytokine loci in a natural population of field voles (*Microtus agrestis*). Mol Ecol. 21(7):1632–1646.

Tzou P, et al. 2000. Tissue-specific inducible expression of antimicrobial peptide genes in *Drosophila* surface epithelia. Immunity 13(5):737–748.

Unckless RL, Howick VM, Lazzaro BP. 2016. Convergent balancing selection on an antimicrobial peptide in *Drosophila*. Curr Biol. 26(2):257–262.

Unckless RL, Lazzaro BP. 2016. The potential for adaptive maintenance of diversity in insect antimicrobial peptides. Philos Trans R Soc B 371(1695):20150291.

Velová H, Gutowska-Ding MW, Burt DW, Vinkler M. 2018. Toll-like receptor evolution in birds: gene duplication, pseudogenization, and diversifying selection. Mol Biol Evol. 35(9):2170–2184.

Wang M-S, et al. 2016. Positive selection rather than relaxation of functional constraint drives the evolution of vision during chicken domestication. Cell Res. 26(5):556–573.

Wilson JN, et al. 2006. A hallmark of balancing selection is present at the promoter region of interleukin 10. Genes Immun. 7(8):680–683.

Wolinska J, King KC. 2009. Environment can alter selection in host–parasite interactions. Trends Parasitol. 25(5):236–244.

Associate editor: Marta Wayne