



**Cite this article:** Unckless RL, Lazzaro BP. 2016 The potential for adaptive maintenance of diversity in insect antimicrobial peptides. *Phil. Trans. R. Soc. B* **371**: 20150291. <http://dx.doi.org/10.1098/rstb.2015.0291>

Accepted: 12 February 2016

One contribution of 13 to a theme issue 'Evolutionary ecology of arthropod antimicrobial peptides'.

**Subject Areas:**

evolution, genetics, genomics, immunology

**Keywords:**

insect immunity, antimicrobial peptide, trans-species polymorphism, balancing selection, convergent evolution

**Author for correspondence:**

Robert L. Unckless  
e-mail: [unckless@cornell.edu](mailto:unckless@cornell.edu)

Electronic supplementary material is available at <http://dx.doi.org/10.1098/rstb.2015.0291> or via <http://rstb.royalsocietypublishing.org>.

# The potential for adaptive maintenance of diversity in insect antimicrobial peptides

Robert L. Unckless and Brian P. Lazzaro

Department of Entomology, Cornell University, Ithaca, NY 14853, USA

RLU, 0000-0001-8586-7137; BPL, 0000-0002-3881-0995

Genes involved in immune defence are among the fastest evolving in the genomes of many species. Interestingly, however, genes encoding antimicrobial peptides (AMPs) have shown little evidence for adaptive divergence in arthropods, despite the centrality of these peptides in direct killing of microbial pathogens. This observation, coupled with a failure to detect phenotypic consequence of genetic variation in AMPs, has led to the hypothesis that individual AMPs make minor contributions to overall immune defence and that AMPs instead act as a collective cocktail. Recent data, however, have suggested an alternative explanation for the apparent lack of adaptive divergence in AMP genes. Molecular evolutionary and phenotypic data have begun to suggest that variant AMP alleles may be maintained through balancing selection in invertebrates, a pattern similar to that observed in several vertebrate AMPs. Signatures of balancing selection include high rates of non-synonymous polymorphism, trans-species amino acid polymorphisms, and convergence of amino acid states across the phylogeny. In this review, we revisit published literature on insect AMP genes and analyse newly available population genomic datasets in *Drosophila*, finding enrichment for patterns consistent with adaptive maintenance of polymorphism.

This article is part of the themed issue 'Evolutionary ecology of arthropod antimicrobial peptides'.

## 1. Introduction

Genes of the immune system are among the fastest evolving in the genome. This is often attributed to an intense coevolutionary 'arms race' between hosts and pathogens. In this model, host populations fix a mutation conferring resistance to the pathogen, imposing selective pressure on the pathogen to re-evolve infectivity, causing the host to evolve new resistance, and the cycle repeats *ad infinitum* [1,2]. Despite the clear oversimplification of this model, molecular evolutionary data for many immune genes are consistent with such a dynamic. Immune system genes often show much more rapid adaptive evolutionary divergence than non-immune genes, indicative of recurrent bouts of selection [3–5]. Although the 'arms race' model is a charismatic and readily detectable mode of host–pathogen coevolution, it may not be characteristic of most immune-related genes.

Alternative models of host–pathogen coevolution can also find support in the data. For example, in time-lagged churning allele models, pathogens adapt to the most common host genotypes, driving them down in frequency [6]. Proportionally rarer alleles have a selective advantage in these models by virtue of their rarity, but they are advantageous only until they reach sufficiently high frequency for the pathogen population to adapt to them. At that point, pathogen pressure drives the allele down in frequency, in favour of different rare alleles. This allele churning is readily visible in clonal systems such as *Potamopyrgus* (a snail) infected by a trematode parasite (*Microphallus*) [7,8], but the same principle can also apply in sexually recombinant species. Importantly, such negative frequency-dependent models of evolution do not allow fixation of alleles, but instead promote allelic diversity in the host population and in some cases may result in very long-term maintenance of polymorphism [9].

The two models described above are most relevant to instances where the host is afflicted with a single dominant pathogen species. But in the vast majority of cases, hosts and pathogens are not locked in one-on-one interactions. Instead, each host species may be infected with multiple and varied pathogens, and pathogens may infect a range of hosts. In such general interactions, coevolution is 'diffuse' [10] and the prospects for an arms race are greatly diminished [11,12]. Furthermore, the relative abundance of pathogens or pathogen genotypes may vary in time or space, leading to fluctuation in the adaptive premiums on resistance alleles. If resistance alleles are costly in the absence of infection [13] or if alleles that confer resistance to one pathogen result in susceptibility to another [14–16], host alleles may oscillate in frequency and can be maintained for long evolutionary time [17]. Any of these instances where natural selection adaptively maintains genetic variation over time or space can be termed 'balancing selection' [18]. Far more difficult to detect than adaptive divergence driven by arms races, balancing selection predicts long-term preservation of ancient alleles, with genetic variation in some cases pre-dating the divergence of distinct species [19]. Of critical importance, some components of the immune system may evolve under arms races while others exhibit balancing selection, depending on the specific molecular functions of the encoded genes and the nature of interaction with pathogens.

Antimicrobial peptides (AMPs) are short antibiotic proteins produced by all plants and animals in response to infection. These peptides are secreted extracellularly and they directly kill bacteria and unicellular fungi, typically acting to permeabilize or destabilize microbial membranes [20]. AMPs are generally encoded by suites of multigene families, exhibit broad specificity, and their production is often generically induced by infection and wounding. Bacterial resistance mechanisms to AMPs are similarly generic, including non-specific behaviours like secretion of proteases, efflux of AMPs from outer and inner membranes, and modification of membrane surfaces to make them less recognizable or susceptible to AMPs [21]. Thus, AMP genes may be less prone to specific coevolution with pathogens than are other components of the immune system, especially when pathogen diversity is high. Indeed, the AMP genes of invertebrates typically do not exhibit the rapid evolutionary divergence predicted under the coevolutionary arms race model, even when other components of the immune system appear to be evolving rapidly under positive selection. For example, in the genus *Drosophila*, the immune system as a whole shows a significantly higher rate of amino acid evolution than the remainder of the genome. However, genes encoding AMPs actually evolve more slowly than the genome average at the amino acid level [4,22]. A similar pattern is seen in the evolution of mosquito immune systems [5]. Instead, in both taxa, AMP gene families show high rates of duplication and deletion over short evolutionary timescales [4,5]. These observations led some investigators [23] to conclude that natural selection on insect AMPs is weak at the level of amino acid sequence, assuming AMPs function as efficient broad-spectrum antibiotics with little prospect for specific coevolution, and that the primary selective pressure is to maintain a minimum threshold level of expression, perhaps leading instead to selection on AMP gene expression or copy number [24,25]. This was in contrast to vertebrate AMPs, which have been more often claimed to show signs of strong positive selection or even adaptive maintenance of polymorphism [26–29]. In this article, we revisit the

molecular evolution and population genetics of insect AMP genes to evaluate whether the standing interpretation that insect AMP genes experience little adaptive evolution is correct. Using a combination of new data and previously published reports, we find evidence that balancing selection and adaptive maintenance of amino acid polymorphism may be much more common in insect AMP genes than has been previously appreciated.

## 2. Material and methods

### (a) *Drosophila* sequence data

To assess population genetic patterns, we used two different strategies. First, to compare levels of polymorphism in AMPs to the rest of the genome, we used previously available estimates of polymorphism from *D. simulans* and compared AMPs to random draws from the rest of the genome. Alternatively, to examine trans-species polymorphisms (TSPs), we compared genes encoding AMPs to a set of control genes selected to match protein length and genomic position (see the electronic supplementary material, tables S1 and S2). Control genes were selected by walking out along the *D. melanogaster* genome in both directions until reaching a gene with between 100 and 800 bp of coding sequence (the approximate size of an AMP gene) that is present in at least three of the four species *D. melanogaster*, *D. simulans*, *D. mauritiana* and *D. yakuba*. Sequences for each AMP gene were obtained from publicly available datasets of *Drosophila simulans* ( $n = 20$  [30]) and *D. yakuba* ( $n = 20$  [30]), *D. mauritiana* ( $n = 10$  [31]) and African *D. melanogaster* (DPGP3 [32]). To achieve equivalent sample sizes, we downsampled the *D. melanogaster* DPGP3 sequences by randomly selecting 20 sequences from the larger sample of 197. Of the four species, only *D. simulans* and *D. mauritiana* are thought to interbreed to any extent [33]. Sequences were aligned using ClustalW [34], then checked, and alignments were corrected by eye as necessary. Simple neighbour joining trees were constructed in Geneious [35] using a Jukes–Cantor distance method. DNA sequences were translated to amino acids and amino acid sequences were realigned. DNA sequence alignments were deposited in Dryad, accession no. 10.5061/dryad.g8n25.

### (b) DNA polymorphism and divergence

To assess levels of non-synonymous polymorphism and divergence in *D. simulans* AMPs compared to the rest of the genome, we used estimates from publicly available whole-genome sequencing of seven *D. simulans* inbred lines [36]. Per site non-synonymous polymorphism ( $\pi_n$ ) and unpolarized non-synonymous divergence ( $d_n$ ) were recorded for each AMP, and the mean for AMPs was compared to a null distribution of 10 000 random draws of size 28 (the number of AMP genes evaluated) from all genes. In this analysis, we employed whole-genome data, not a control set of genes. We also compared the ratio of non-synonymous polymorphism to non-synonymous divergence ( $\pi_n/d_n$ ) for AMPs to the distribution of random draws. Non-synonymous divergence was increased by a factor of 0.0001 (the minimum non-zero value) for all genes to avoid zeros in the denominator. We do not report analyses based on synonymous polymorphism linked to putatively selected sites because we suspect that high rates of effective recombination may in many cases erode signatures of linked selection [37].

### (c) Trans-species polymorphisms

TSPs were identified by visually inspecting amino acid alignments. Any site segregating for the same two amino acids in more than one species was designated a TSP. Given the small sample size, we did not employ a minor allele frequency cut-off.

Differences in the number of TSPs between AMPs and our control set were assessed individually for each species pair using logistic regression with state (co-segregating or non-co-segregating) as a binary response variable and class (AMP or control) as a factor. This analysis was conducted using the `glm` function in R [38].

TSPs in AMP genes from non-*Drosophila* insect species were identified from sequences deposited into Genbank after alignment and translation as described in §2a.

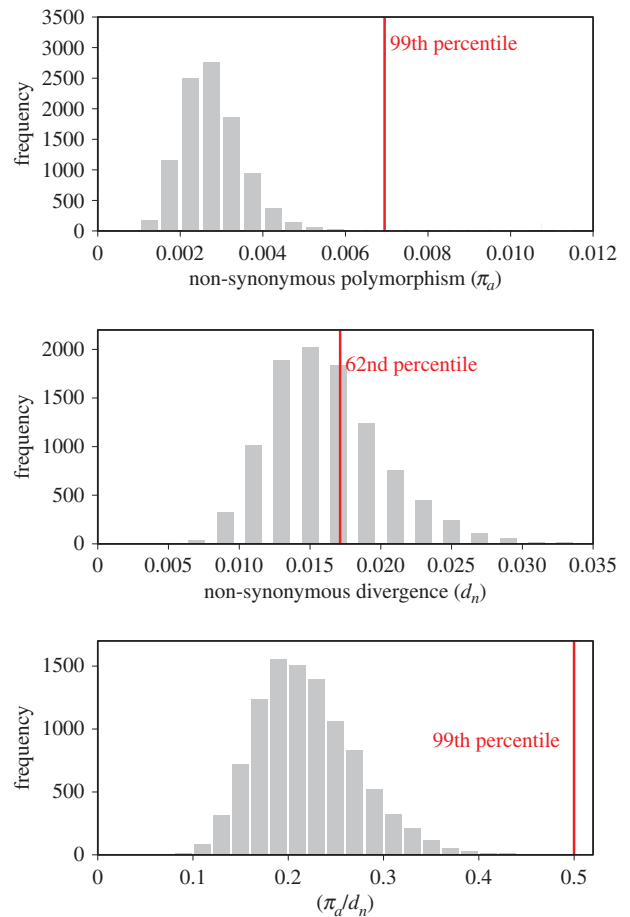
#### (d) Molecular evolutionary convergence

We examined amino acid convergence across the genus *Drosophila* as a means of identifying potential instances of parallel evolution by natural selection. Evolutionary convergence in the genome reference sequences for 22 species was assessed using alignments pulled from the UCSC genome browser [39]. We considered polymorphic sites in *D. melanogaster*, *D. simulans*, *D. mauritiana* and *D. yakuba*, but all other species were represented by a single haploid sequence from an inbred line. Sequences were compiled and aligned as described in §2a. For each amino acid variant in the alignment, a simple convergence score was defined as the minimum number of independent mutations on the phylogeny at that position, minus the number of observed amino acid states, plus one. For example, if two non-sister species have serine at a given residue and all other species have glycine, there are two independent mutations (on the lineages leading to each of the non-sister species) and two states (serine and glycine), so we infer one ( $2 - 2 + 1$ ) instance of convergence over the phylogeny. We assessed enrichment of convergences at sites with TSPs by comparing the number of convergences at those sites to sites with polymorphisms private to *D. melanogaster*. Statistical significance was assessed using a Mann–Whitney *U*-test implemented in R [38].

### 3. Results

Our analysis was motivated by a recent study in which we characterized balancing selection acting on the gene encoding the *Drosophila* AMP, Diptericin [37]. In that work, we noted that a naturally occurring non-synonymous mutation from serine to arginine at the 69th position of the Diptericin mature peptide resulted in a substantial decrease in resistance to bacterial infection in both *D. melanogaster* and *D. simulans*. These two species are not sister taxa, and they have independently derived the arginine state via different mutations of the 69th codon. At the site of the serine/arginine polymorphism, we observed five independent transitions to arginine across 24 *Drosophila* species, suggesting that a common selective pressure might promote the use of arginine at this position in diverse species. Some *D. simulans* chromosomes carry a tandem duplication of the *Diptericin* gene and the serine/arginine polymorphism segregates in both paralogues. Based on the entire body of evidence, we argued that the serine/arginine polymorphism in *Diptericin* is likely to be adaptively maintained through fluctuation in environment and/or pathogen pressure across time and/or space. In this work, we ask whether the qualitative pattern observed in the *Diptericin* locus is generic to AMP genes, and whether insect AMPs may evolve under balancing selection more often than is generally appreciated.

We focus our new data analysis on the AMP genes of *Drosophila*, as these have been the most extensively studied group of AMPs and *Drosophila* species provide the greatest depth in population genomic data among insects. We (i) compare amino acid polymorphism and divergence in AMPs to the



**Figure 1.** Mean non-synonymous diversity ( $\pi_a$ ), divergence ( $d_n$ ) and the ratio of the two ( $\pi_a/d_n$ ) in AMPs indicated by a vertical red line over the distribution of mean values from 10 000 draws of control gene sets sampled from the genome.

rest of the genome in *D. simulans*, (ii) scan AMPs for TSPs and compare their rate of occurrence to the rate observed in a set of control genes in several species pairs, and (iii) measure the excess rate of evolutionary convergence at TSPs relative to polymorphisms private to *D. melanogaster*. We use the aggregate information to infer whether AMP genes as a class are more likely to evolve under balancing selection than are matched control genes drawn from the background genome.

#### (a) Excess polymorphism in *Drosophila* antimicrobial peptides

Genes under long-term balancing selection are expected to exhibit more synonymous and non-synonymous polymorphism than those evolving neutrally, because the comparatively old age of the selectively balanced mutation (deep coalescent) permits the accumulation of mutations on the genealogical lineage that separates the balanced alleles [19]. We compared non-synonymous nucleotide diversity ( $\pi_n$ ) and unpolarized non-synonymous divergence ( $d_n$ ) using previously published data from *D. simulans* AMPs [36] to an empirical distribution constructed by sampling groups of 28 genes from the genome.

The mean non-synonymous nucleotide diversity for AMPs was in the 99th percentile of randomly sampled sets of genes, indicating that AMPs exhibit exceptional amino acid polymorphism compared to the rest of the genome (figure 1; electronic supplementary material, table S3). High rates of non-synonymous diversity could arise through either balancing

**Table 1.** Trans-species amino acid polymorphisms in AMPs. Reported as A1PA2-Dp, where P is position in alignment, A1 and A2 are the two amino acid alleles, D is the domain (s = signal peptide, p = pro-domain, m = mature peptide; for Attacin D, n = N-terminal, c = C-terminal) and p is the position in the domain. n.d., no comparison due to poor alignments.

AMP	length	mel-sim	mel-mau	mel-yak	sim-mau	sim-yak	mau-yak
Attacin A	236	L21V-p1 A39T-m8	L21V-p1	n.d.	L9M-s9 <sup>b</sup> L21V-p1	n.d.	n.d.
Attacin B	236	L18V-p1 A36T-m8	L18V-p1	n.d.	L18V-p1	n.d.	n.d.
Attacin C	244	—	T28N-p7	—	—	—	—
Attacin D	181	F82I-c16	—	—	A40ST-n40 A58G-n58 M69V-c3 L151V-c85	—	—
Cecropin A <sup>c</sup>	64	N2K-s2	T17S-s17(A2) <sup>a</sup>	A17T-s17(A2)	—	—	—
Cecropin B	63	—	—	A15V-s15	M10V-s10	—	—
Defensin	92	—	M17V-sp17	—	—	H41Q-p20	—
Diptericin A <sup>d</sup>	106	S92R-m69(A1,A2) <sup>a</sup>	—	—	—	—	—
Diptericin B	—	—	A19V-sp19	—	—	—	—
Drosocin	64	A52T-p12	—	—	—	—	—
Metchnikowin	53	P50R-m24	P50R-m24	P50R-m24	P50R-m24	P50R-m24	P50R-m24

<sup>a</sup>Convergent mutations.

<sup>b</sup>Also valine in *D. simulans*.

<sup>c</sup>*D. melanogaster* *CecA1* and *CecA2* considered.

<sup>d</sup>*D. simulans* *DptA1* and *DptA2* considered.

selection or relaxed constraint on protein sequences. If AMP sequences are comparatively unconstrained, we would also predict higher than average non-synonymous divergence between species in AMPs compared with the rest of the genome. Yet AMP non-synonymous divergence is not in the tail of the genome-sampled distribution (figure 1, 62nd percentile). We computed the ratio of non-synonymous polymorphism to divergence for AMPs and the resampled genome sets. Once again, AMPs were exceptional (99th percentile) for the highest ratios of non-synonymous polymorphism to divergence. Thus, we conclude that the excess in AMP genes polymorphism is unlikely to be due to relaxation of selective pressures leading to a generally higher rate of selectively neutral substitution. The data are more consistent with population-level maintenance of variation that does not fix in the species.

## (b) Evidence of trans-species polymorphism in antimicrobial peptides

Ancient balanced polymorphisms can segregate in multiple closely related species if they pre-date the species split [40]. Polymorphisms can also be shared between closely related species via interspecific hybridization and introgression [41]. Even distantly related species can share polymorphism if the same variants arise through independent mutation (e.g. [37]). Finally, independent paralogous gene conversion within two species can result in TSP when both species harbour closely related paralogues.

*Drosophila* AMP genes are rife with TSPs (table 1). Many of these are found in the signal peptide or pro-peptide regions. For example, between *D. melanogaster* and *D. simulans*, we find nine

TSPs spread over six different AMPs (out of 11 examined). This is a species pair with no evidence of ongoing introgression and whose divergence time (approx. 3–5 Ma [42]) far exceeds expected time for which selectively neutral variation from the common ancestral species is expected to persist ( $4N_e$  generations where  $N_e$  is the effective population size, or approximately 400 000 generations or 25 000 years [43]). Other species pairs show similar significant excesses of TSP in AMP genes relative to control genes, although the more divergent *D. yakuba* is less likely to harbour TSPs with the ingroup species (table 1; electronic supplementary material, tables S4 and S5).

A few specific examples are worth noting. We previously reported two instances of convergence at *Dpt* in *D. melanogaster* and *D. simulans* [37]. First, serine (hydrophobic, neutral) and arginine (positively charged) are segregating at the 69th position of the mature peptide via independent mutation in both *D. melanogaster* and *D. simulans*. Second, convergent deletions eliminate the *Diptericin* start codon in some alleles of both species, presumably rendering the gene non-functional. Both of these convergences have measurable phenotypic impact on resistance to bacterial infection [37].

In a previous study, Lazzaro & Clark [44] noted that alanine and threonine (both hydrophobic) are segregating at the 12 position in the pro-peptide sequence of *Drosocin* in *D. melanogaster*. Lazzaro & Clark [44] noted a significant increase in the frequency of the ancestral alanine allele between 1998 and 2001 with markedly reduced diversity in chromosomes carrying the alanine state, suggesting a partial selective sweep favouring the ancestral allele. In a subsequent survey of haphazard natural collections taken between 2008 and 2014, the allele frequencies continue to match those observed in 2001 (SM Rottschaefer &

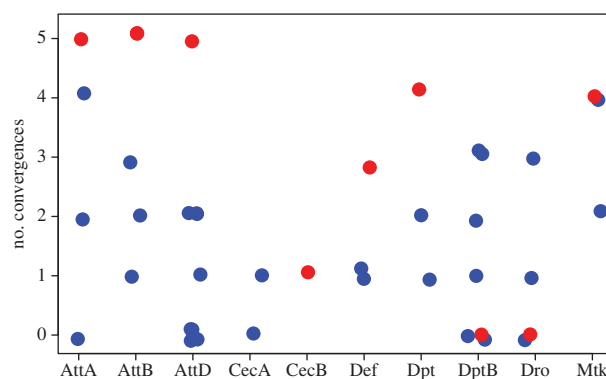


BP Lazzaro 2015, unpublished data). The fact that the derived threonine allele also segregates in *D. simulans* (employing the same codon sequence) is intriguing, although the threonine state is rare in *D. simulans* (0.05% observed frequency).

In the Attacin A/B locus, a lack of reciprocal monophyly between *D. simulans* and *D. mauritiana* is suggestive of introgression through hybridization between species (electronic supplementary material, figure S1). However, a systematic study of introgression patterns between the two species [33] did not report any of the Attacin genes as having a signature of recent introgression, indicating that, if these regions are introgressed, either the introgression was ancient and this polymorphism has persisted or the introgressed region was too small to be classified as such in the previous study. Lysine (positive charge) and valine (hydrophobic/aliphatic) are segregating in both *Attacin A* and *Attacin B* in *D. melanogaster*, *D. simulans* and *D. mauritiana* at the first residue of the propeptide (position 21 in *Atta* and position 18 in *AttB* since a three amino acid insertion in *Atta* shifts the alignment). Lazzaro & Clark reported gene conversion between paralogues in *D. melanogaster* [45], and it stands to reason that paralogous gene conversion may occur in other species as well. For non-adaptive processes to be responsible for this sharing of polymorphism between paralogues in three species, we have to posit that the same mutations arose, were shared between paralogues through independent gene conversion in three species, and have drifted to intermediate population frequency in all of them. This seems unlikely. Similarly, we find alanine and threonine segregating in both *Attacin A* and *Attacin B* at the eighth position of the mature peptide in both *D. melanogaster* and *D. simulans*. Paralogous gene conversion may facilitate maintenance of conditionally beneficial genetic variation since that variation can be regenerated from the paralogous copy.

Finally, and perhaps most interestingly, proline (uncharged) and arginine (positively charged) segregate in the third-to-last position of the *Metchnikowin* mature peptide of *D. melanogaster* (80% proline,  $n = 20$ ), *D. simulans* (52% proline,  $n = 21$ ), *D. mauritiana* (20% proline,  $n = 10$ ) and *D. yakuba* (90% proline,  $n = 20$ ). In all species, the same mutation is responsible for the amino acid change (C to G in the middle position of the codon, the only single mutation that would result in a change from arginine to proline) so we cannot unambiguously determine whether these mutations are convergent or have been maintained since before the divergence of these four species.

To systematically assess whether *Drosophila* AMPs harbour an excess of TSPs relative to the genome expectation, we contrasted the relative incidence of TSPs in AMP genes relative to control genes using logistic regression (see §2c). AMPs harboured significantly more TSPs than control genes in one of six pairwise comparisons (*D. melanogaster*/*D. simulans*, *D. melanogaster*/*D. mauritiana* and *D. melanogaster*/*D. yakuba* were nearly significant) and harboured a non-significant excess in two of the remaining comparisons (electronic supplementary material, table S5). Using Fisher's method for combining  $p$ -values, we find a nearly significant excess of TSPs in AMPs compared to our control set ( $p = 0.055$ ). The AMPs have a similar number of TSPs per polymorphic amino acid residues as the control genes do (electronic supplementary material, table S1), indicating that the excess of TSPs in AMPs is at least partially related to their overall increased amino acid diversity. A similar pattern was observed genome-wide by Begun *et al.* [33], who reported a higher rate of transpecific amino acid polymorphisms when amino acid diversity was the highest.



**Figure 2.** Number of convergences on the *Drosophila* phylogeny for TSPs (red) and amino acid polymorphisms private to *D. melanogaster* (blue).

### (c) Convergence on the *Drosophila* phylogeny

An excess of transitions between two amino acid states in multiple independent species lineages could indicate that selection favours both alleles in alternative environments. To assess whether TSPs show more instances of convergence across the genus *Drosophila*, we tested whether TSPs are more likely to convergently re-evolve elsewhere on the tree than are amino acid polymorphisms private to *D. melanogaster*. TSPs showed higher rates of convergence than polymorphisms that were private to *D. melanogaster* (Mann–Whitney  $U = 78$ ;  $p = 0.014$ ; figure 2). The median number of convergences for TSPs was four (range from zero to five) compared to a median of one (range from zero to four) for *D. melanogaster* private polymorphisms. For most of the species on the tree, we have only a single reference genome sequence. It thus remains possible that with deeper sampling, the convergences could be revealed to be polymorphic in additional species. This excess of convergence at sites with TSPs is consistent with either bias towards neutral substitution of the same amino acids at TSP positions, or an adaptive value that maintains the two or more alternative states at each position. Given that the one mutation we have best characterized, the serine/arginine polymorphism in the mature peptide of *Diptericin*, has a large phenotypic effect on resistance [37], we have a hard time envisioning that these recurrent convergent substitutions are adaptively neutral. Instead, we favour the hypothesis that geographical and/or temporal variation in selective forces such as infectious pathogen pressure leads to adaptive maintenance of polymorphism within species and recurrent convergent evolution across species.

We have identified multiple new and potentially interesting amino acid substitutions that not only segregate in multiple species but also arise by convergence elsewhere on the phylogeny (electronic supplementary material, figure S2). For example, *Mtk* is segregating for arginine and proline at position 50 of the mature peptide in *D. mauritiana*, *D. simulans*, *D. melanogaster* and *D. yakuba*. Across the phylogeny, we find two additional mutations to arginine at this position: one in *D. erecta* and another in *D. kikkawai*. There are also two mutations to glycine at the same position (one in the branch leading to *D. biarmipes*, *D. suzukii* and *D. takahashii* and another in *D. elegans*). The derived arginine allele segregating in *Diptericin A* position 69 in *D. melanogaster* and *D. simulans* is also found in the reference sequences for *D. oreana*, *D. ficusphila* and *D. willistoni* [37]. Additionally, there are two mutations to glutamine at the same site. Other striking instances of convergence include the two polymorphisms in *Defensin*, those in

*Attacin D* and one in the mature peptide of *Attacin A/Attacin B* (electronic supplementary material, figure S1).

#### (d) Revisiting antimicrobial peptide evolution in other insect taxa

Few previous studies in non-*Drosophila* insects have generated data that can be screened for the presence of TSPs, convergence over a phylogeny, or even balancing selection on AMPs. Below we revisit the subset of existing studies from non-*Drosophila* insects with the most appropriate data for addressing these questions, and we find several instances of TSP in other taxa.

Termicin is a termite Defensin protein. Two studies examined polymorphism in *termicin* in four species [46,47]. Re-examining this polymorphism data, we find valine and arginine segregating at the 13th position of the signal peptide and histidine and arginine segregating at the 14th position of the mature peptide in *Reticulitermes chinensis* and *Odontotermes formosanus*. Neither of these polymorphisms show convergence over the remainder of the termite phylogeny [48].

Alanine and valine segregate in the sixth position of the signal peptide of Cecropin E of the silkworms *Bombyx mori* ( $n = 15$ ) and *Bombyx mandarina* ( $n = 12$ ) [49]. Examination of *Defensin A* and *moricin* from these same samples revealed no non-synonymous TSPs, because there was no amino acid polymorphism in *B. mori* in either of these genes. No closely related species have been sequenced to allow a test of convergence across the *Bombyx* phylogeny.

Glutamine and arginine segregate at the 86th position of the mature peptide in Hymenoptera of the ant, *Monomorium cyaneum* [50] and the bee, *Apis cerana* [51]. Comparisons of other AMPs in hymenoptera were not possible due to poor alignments or only single-species polymorphism data.

AMPs in *Anopheles* mosquitoes harbour frequent TSPs, although this may be due in part to ongoing gene flow between named species [52–54]. Both phenylalanine and isoleucine are segregating at the eighth position of the signal peptide of Cecropin A in *Anopheles gambiae* and *Anopheles melas*—two species unlikely to have experienced recent introgression. Several AMPs show TSP between *A. gambiae* and *A. arabiensis*, although these species are very recently diverged, still sharing polymorphism in many genes and probably hybridizing in the field [55]. In Defensin, both alanine and serine segregate at the 15th position of the signal peptide in *A. gambiae* and *A. arabiensis*, and both alanine and serine segregate at the 12th position of the mature in *A. gambiae* and *A. arabiensis*. In Gambicin, the fourth position of the signal peptide segregates for methionine and valine in *A. gambiae* and *A. arabiensis* and the seventh position of the signal peptide segregates for leucine and isoleucine in the same two species. In both *A. arabiensis* and the more distantly related *A. merus*, the 37th position of the Gambicin mature peptide segregates for alanine and threonine. Phenylalanine and valine segregate at the 55th position of the mature peptide in *A. arabiensis*, *A. gambiae* and the equally closely related *A. quadriannulatus*.

Evaluation of the available data from multiple insects indicates that the observation of TSP in AMP genes is not unique to *Drosophila*. In the absence of population genomic sequencing to yield a set of control genes, we cannot assess whether TSP is more common in AMPs than in other genes in these species. Nevertheless, the existing data suggest that TSP in AMP genes may be widespread, and we predict that molecular

convergence in more distantly related species is probably also common. These predictions can be tested as high-throughput sequencing enables population genomic data collection from a wide variety of insect species.

## 4. Discussion

Prevailing wisdom has been that insect AMPs function as an antibiotic cocktail, where the component proteins have distinct but partially redundant mechanisms of attacking the microbe. This functional hypothesis combined with little evidence of adaptive diversification in insect AMPs led to a general theory that collective AMP dosage is the primary determinant of resistance, and that individual AMP genes are under little natural selective pressure (e.g. [23]). This interpretation was in contrast to observations from vertebrates, where adaptive maintenance of polymorphism has been suggested for AMPs in humans [29,56], birds (including TSP) [28] and frogs [27]. More recently, TSP and adaptive maintenance of polymorphism have been additionally observed in marine mussels [57,58].

Also in contrast to the prevailing wisdom about insect AMPs, two recent studies in *Drosophila* have shown that deletion of a single or small number of AMP genes can have major impact on organism-level resistance to bacterial infection [59,60]. At the *Diptericin* gene, a single amino acid polymorphism is strongly predictive of resistance [37], suggesting that natural selection could act effectively on individual AMP genes. Although there is little evidence in insect AMPs of adaptive divergence in the sense of ‘arms races’, a re-reading of previous literature shows many patterns consistent with potential balancing selection (table 2), including several instances that went undiscussed by the authors in early *Drosophila* papers [44,45,61,62]. We propose a few explanations for this absence. First, many studies were conducted prior to 2005, and balancing selection was generally given less consideration by population geneticists because few formal tests existed. Second, AMPs are small and therefore most standard tests are underpowered—especially if recombination has broken down haplotypes associated with the maintained alleles. Finally, it is only recently that genome-scale analyses and very large sample sizes have become accessible to most researchers, and these may be very important for detection of possible balancing selection. With the present analysis, we suggest that adaptive maintenance of polymorphism and convergent evolution may be more common in insect AMP genes than was previously recognized. Both balancing selection and convergent evolution can be driven by fluctuation in natural selective pressure over time and/or geographical space, and may be mediated by shifting diversity of pathogens as well as by correlated life-history costs of overactive immune systems [19,70]. Patterns of AMP evolution may be further influenced by synergistic interactions among distinct AMPs (e.g. [71,72]), especially if the synergism is quantitatively altered by amino acid changes to the interacting AMPs. We strongly advocate more targeted research to investigate the frequency of adaptive AMP polymorphism in insects that are amenable to experimental study.

We emphasize that TSP and evolutionary convergence do not definitively prove maintenance of polymorphism by balancing or fluctuating selection. However, we do see several patterns that are consistent with balancing selection and that are more common in AMP genes than in the background genome.

**Table 2.** Summary of evidence for selection in AMPs. TSPs = trans-species polymorphisms (table 1)—each instance is counted so a polymorphism present in all four species is counted six times for six comparisons. n.d.=Not determined.

taxon	AMP	TSPs	conv. on phylogeny	positive selection <sup>a</sup>	other patterns consistent with balancing selection
<i>Drosophila</i>	<i>Atta/AttB</i>	9	several	no— <i>Atta</i> <i>AttB</i> —negative Tajima's <i>D</i> [45]	gene conversion between paralogues [45]
	<i>AttC</i>	1	very little	no [45]	—
	<i>AttD</i>	5	moderate	n.d.	—
	<i>CecA</i>	3	very little	no [61,62]	high $F_{st}$ between populations [61] positive Tajima's <i>D</i> [62]
	<i>CecB</i>	2	none	no [61,62]	high $F_{st}$ between populations [61]
	<i>Def</i>	2	several	no [44]	—
	<i>Dpt A</i>	1	several	negative $F_u$ and Li's <i>D</i> [61] no [44]	high $F_{st}$ between populations [61] phenotypic evidence [59]
	<i>Dpt B</i>	1	none	no [44]	—
	<i>Dro</i>	1	none	no [44]	—
	<i>Mtk</i>	6	several	no [44]	—
<i>Bombus</i>	<i>Def-1</i>	n.d.	n.d.	high $dN/dS$ [63]	—
	<i>Abaecin</i>	n.d.	n.d.	high $dN/dS$ [63] no [64]	—
	<i>Hymenoptaecin</i>	n.d.	n.d.	high $dN/dS$ [63]	—
<i>Apis mellifera</i>	<i>Def-2</i>	n.d.	n.d.	negative $F_u$ and Li's $D^a$ and negative $F_u$ and Li's $F^a$ [65]	—
	<i>Lys-1</i>	n.d.	n.d.	negative $F_u$ and Li's $D^a$ and negative $F_u$ and Li's $F^a$ [65]	high non-synonymous polymorphism with low divergence
<i>Reticulitermes</i>	<i>Termicin</i>	2	n.d.	significant MK test [47]	—
<i>Anopheles</i>	<i>Gambicin</i>	4	n.d.	PAML—positive selection on single codon [54] no [51]	positively selected site based on PAML analysis is segregating in three species
	<i>Defensin</i>	2	n.d.	purifying selection [54,66]	—
	<i>Cecropin 1</i>	1	n.d.	no [67]	—
<i>ants</i>	<i>Defensin</i>	n.d.	yes [68]	PAML—positive selection acting on two codons [68]	—
<i>Ixodes</i>	<i>Defensin</i>	n.d.	n.d.	negative Tajima's <i>D</i> , negative $F_u$ and Li's $F$ [69]	very polymorphic gene, but polymorphism confined to intron
<i>Bombyx</i>	<i>Defensin A</i>	no	n.d.	negative Tajima's <i>D</i> [49]	high ratio of non-synonymous to synonymous polymorphism in <i>B. mandarina</i>

<sup>a</sup>Sackton *et al.* [4] and Obbard *et al.* [22] found no evidence for adaptive evolution in any *Drosophila* AMPs.

These include high rates of non-synonymous polymorphism, TSP, amino acid convergence across phylogenetically distinct taxa and high incidence of loss of function alleles. Although these patterns could also be consistent with relaxed selective constraint, we are wary of that explanation because the amino acid substitutions are often non-conservative and because at least some functionally tested alleles have clear phenotypic effects.

Unfortunately, there are some deficiencies in the available data that preclude a more definitive analysis of whether

insect AMP genes systematically evolve via maintenance of polymorphism. First, we lack well-annotated arthropod genomes outside of *Drosophila*. Even in *Drosophila*, the majority of AMPs are not functionally characterized and whole new families may yet be discovered, exemplified by the novel *Bomanin* family described in 2015 [60]. Poor genome assembly is especially a problem for analysis of multigene families, where inference of shared polymorphism and convergence can be confounded, but ambiguity in orthologous relationships. This problem is exacerbated when the sequence data

are generated by short-read technology. Second, we lack population genomic datasets for pairs or groups of species. Such datasets are beginning to be generated, but most datasets to date involve pairs of species that are so closely related that they continue to hybridize and exchange alleles. This muddies the interpretation of trans-specific polymorphism and retention of ancestral variation. Finally, because AMP genes are short, they harbour little nucleotide diversity, reducing the power of many population genetic tests on individual genes. Thus, empowered analyses rely on pooling the data from multiple AMP genes, which may vary in precise function, activity and tissue of expression. An understanding of the selective forces acting on AMPs also requires a better understanding of the functional consequences of amino acid variation. Both *in vitro* and *in vivo* assays will assist in the functional dissection of natural variation in the mature peptide as well as signal and pro domains of AMPs. This functional understanding may even illuminate drivers for similar patterns of AMP diversity in ecologically diverse host species.

Despite the limitations in the available data, we propose that there is credible and suggestive evidence that balancing

selection and adaptive maintenance of polymorphism may be common in insect AMP genes. We have highlighted some of these lines of evidence, and advocate targeted experimentation in diverse insects to formally test the hypothesis. These experiments should take the form of both population genomic analyses and phenotypic study, and they are well within the reach of the research community. The first 10 years of research on the evolution of insect AMP genes led to the conclusion that individual AMPs are superfluous and redundant, and largely invisible to natural selection. Will the next 10 years overturn that interpretation?

**Data accessibility.** All data were previously published, and where appropriate is present in the electronic supplementary material.

**Authors' contributions.** R.L.U. and B.P.L. conceived of, performed analysis for and wrote this article.

**Competing interests.** We have no competing interests.

**Funding.** This study is supported by NIH grant nos. R01 AI083932 and R01 AI064950 to B.P.L. and NIH grant no. K99-GM114714 to R.L.U.

**Acknowledgements.** We thank members of the Lazzaro lab for useful discussion.

## References

- Van Valen L. 1973 A new evolutionary law. *Evol. Theory* **1**, 1–30.
- Dawkins R, Krebs JR. 1979 Arms races between and within species. *Proc. R. Soc. Lond. B* **205**, 489–511. (doi:10.1098/rspb.1979.0081)
- Nielsen R *et al.* 2005 A scan for positively selected genes in the genomes of humans and chimpanzees. *PLoS Biol.* **3**, e170. (doi:10.1371/journal.pbio.0030170)
- Sackton TB, Lazzaro BP, Schlenke TA, Evans JD, Hultmark D, Clark AG. 2007 Dynamic evolution of the innate immune system in *Drosophila*. *Nat. Genet.* **39**, 1461–1468. (doi:10.1038/ng.2007.60)
- Waterhouse RM *et al.* 2007 Evolutionary dynamics of immune-related genes and pathways in disease-vector mosquitoes. *Science* **316**, 1738–1743. (doi:10.1126/science.1139862)
- Dybdahl MF, Lively CM. 1998 Host–parasite coevolution: evidence for rare advantage and time-lagged selection in a natural population. *Evolution* **52**, 1057–1066. (doi:10.2307/2411236)
- Koskella B, Lively CM. 2009 Evidence for negative frequency-dependent selection during experimental coevolution of a freshwater snail and a sterilizing trematode. *Evolution* **63**, 2213–2221. (doi:10.1111/j.1558-5646.2009.00711.x)
- Koskella B, Lively CM. 2007 Advice of the rose: experimental coevolution of a trematode parasite and its snail host. *Evolution* **61**, 152–159. (doi:10.1111/j.1558-5646.2007.00012.x)
- Hamilton WD. 1980 Sex versus non-sex versus parasite. *Oikos* **35**, 282–290. (doi:10.2307/3544435)
- Janzen DH. 1980 When is it coevolution. *Evolution* **34**, 611–612. (doi:10.2307/2408229)
- Strauss SY, Sahli H, Conner JK. 2005 Toward a more trait-centered approach to diffuse (co)evolution. *New Phytol.* **165**, 81–89. (doi:10.1111/j.1469-8137.2004.01228.x)
- Lapchin L. 2002 Host-parasitoid association and diffuse coevolution: when to be a generalist? *Am. Nat.* **160**, 245–254. (doi:10.1086/341020)
- Tian D, Traw MB, Chen JQ, Kreitman M, Bergelson J. 2003 Fitness costs of R-gene-mediated resistance in *Arabidopsis thaliana*. *Nature* **423**, 74–77. (doi:10.1038/nature01588)
- Glazebrook J. 2005 Contrasting mechanisms of defense against biotrophic and necrotrophic pathogens. *Annu. Rev. Phytopathol.* **43**, 205–227. (doi:10.1146/annurev.phyto.43.040204.135923)
- Cotter SC, Kruuk LE, Wilson K. 2004 Costs of resistance: genetic correlations and potential trade-offs in an insect immune system. *J. Evol. Biol.* **17**, 421–429. (doi:10.1046/j.1420-9101.2003.00655.x)
- Wilfert L, Schmid-Hempel P. 2008 The genetic architecture of susceptibility to parasites. *BMC Evol. Biol.* **8**, 187. (doi:10.1186/1471-2148-8-187)
- Stahl EA, Dwyer G, Mauricio R, Kreitman M, Bergelson J. 1999 Dynamics of disease resistance polymorphism at the *Rpm1* locus of *Arabidopsis*. *Nature* **400**, 667–671. (doi:10.1038/23260)
- Charlesworth B, Charlesworth D. 2010 *Elements of evolutionary genetics*. Austin, TX: Roberts & Co.
- Charlesworth D. 2006 Balancing selection and its effects on sequences in nearby genome regions. *PLoS Genet.* **2**, e64. (doi:10.1371/journal.pgen.0020064)
- Yeaman MR, Yount NY. 2003 Mechanisms of antimicrobial peptide action and resistance. *Pharmacol. Rev.* **55**, 27–55. (doi:10.1124/pr.55.1.2)
- Joo H-S, Fu C-I, Otto M. 2016 Bacterial strategies of resistance to antimicrobial peptides. *Phil. Trans. R. Soc. B* **371**, 20150292. (doi:10.1098/rspb.2015.0292)
- Obbard DJ, Welch JJ, Kim KW, Jiggins FM. 2009 Quantifying adaptive evolution in the *Drosophila* immune system. *PLoS Genet.* **5**, e1000698. (doi:10.1371/journal.pgen.1000698)
- Lazzaro BP. 2008 Natural selection on the *Drosophila* antimicrobial immune system. *Curr. Opin. Microbiol.* **11**, 284–289. (doi:10.1016/j.mib.2008.05.001)
- Franzenburg S, Walter J, Kunzel S, Wang J, Baines JF, Bosch TC, Fraune S. 2013 Distinct antimicrobial peptide expression determines host species-specific bacterial associations. *Proc. Natl Acad. Sci. USA* **110**, E3730–E3738. (doi:10.1073/pnas.1304960110)
- Drosophila* 12 Genomes Consortium. 2007 Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**, 203–218. (doi:10.1038/nature06341)
- Tennesen JA, Blouin MS. 2007 Selection for antimicrobial peptide diversity in frogs leads to gene duplication and low allelic variation. *J. Mol. Evol.* **65**, 605–615. (doi:10.1007/s00239-007-9045-5)
- Tennesen JA, Blouin MS. 2008 Balancing selection at a frog antimicrobial peptide locus: fluctuating immune effector alleles? *Mol. Biol. Evol.* **25**, 2669–2680. (doi:10.1093/molbev/msn208)
- Hellgren O, Sheldon BC. 2011 Locus-specific protocol for nine different innate immune genes (antimicrobial peptides: beta-defensins) across passerine bird species reveals within-species coding variation and a case of trans-species polymorphisms. *Mol. Ecol. Resour.* **11**, 686–692. (doi:10.1111/j.1755-0998.2011.02995.x)
- Hollox EJ, Armour JA. 2008 Directional and balancing selection in human beta-defensins. *BMC Evol. Biol.* **8**, 113. (doi:10.1186/1471-2148-8-113)
- Rogers RL, Cridland JM, Shao L, Hu TT, Andolfatto P, Thornton KR. 2014 Landscape of standing variation for tandem duplications in *Drosophila yakuba* and *Drosophila simulans*. *Mol. Biol. Evol.* **31**, 1750–1766. (doi:10.1093/molbev/msu124)
- Garrigan D, Kingan SB, Geneva AJ, Vedanayagam JP, Presgraves DC. 2014 Genome diversity and



- divergence in *Drosophila mauritiana*: multiple signatures of faster X evolution. *Genome Biol. Evol.* **6**, 2444–2458. (doi:10.1093/gbe/evu198)
32. Lack J, Cardeno C, Crepeau M, Taylor W, Corbett-Detig R, Stevens K, Langley CH, Pool J. 2014 The *Drosophila* Genome Nexus: a population genomic resource of 605 *Drosophila melanogaster* genomes, including 197 genomes from a single ancestral range population. *Genetics* **199**, 1229–1241.
  33. Kingan SB, Geneva AJ, Vedanayagam JP, Garrigan D. In preparation. Genome divergence and gene flow between *Drosophila simulans* and *D. mauritiana*. *bioRxiv*. (doi:10.1101/02471)
  34. Larkin MA *et al.* 2007 Clustal W and Clustal X version 2.0. *Bioinformatics* **23**, 2947–2948. (doi:10.1093/bioinformatics/btm404)
  35. Kearse M *et al.* 2012 Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647–1649. (doi:10.1093/bioinformatics/bts199)
  36. Begun DJ *et al.* 2007 Population genomics: whole-genome analysis of polymorphism and divergence in *Drosophila simulans*. *PLoS Biol.* **5**, e310. (doi:10.1371/journal.pbio.0050310)
  37. Unckless RL, Howick VM, Lazzaro BP. 2016 Convergent balancing selection on an antimicrobial peptide in *Drosophila*. *Curr. Biol.* **26**, 257–262. (doi:10.1016/j.cub.2015.11.063)
  38. R Development Core Team 2015 *R: a language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing.
  39. Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002 The human genome browser at UCSC. *Genome Res.* **12**, 996–1006. (doi:10.1101/gr.229102)
  40. Edwards SV, Hedrick PW. 1998 Evolution and ecology of MHC molecules: from genomics to sexual selection. *Trends Ecol. Evol.* **13**, 305–311. (doi:10.1016/S0169-5347(98)01416-5)
  41. Hedrick PW. 2013 Adaptive introgression in animals: examples and comparison to new mutation and standing variation as sources of adaptive variation. *Mol. Ecol.* **22**, 4606–4618. (doi:10.1111/mec.12415)
  42. Obbard DJ, Maclennan J, Kim KW, Rambaut A, O'Grady PM, Jiggins FM. 2012 Estimating divergence dates and substitution rates in the *Drosophila* phylogeny. *Mol. Biol. Evol.* **29**, 3459–3473. (doi:10.1093/molbev/mss150)
  43. Andolfatto P, Wong KM, Bachtrog D. 2011 Effective population size and the efficacy of selection on the X chromosomes of two closely related *Drosophila* species. *Genome Biol. Evol.* **3**, 114–128. (doi:10.1093/gbe/evq086)
  44. Lazzaro BP, Clark AG. 2003 Molecular population genetics of inducible antibacterial peptide genes in *Drosophila melanogaster*. *Mol. Biol. Evol.* **20**, 914–923. (doi:10.1093/molbev/msg109)
  45. Lazzaro BP, Clark AG. 2001 Evidence for recurrent paralogous gene conversion and exceptional allelic divergence in the Attacin genes of *Drosophila melanogaster*. *Genetics* **159**, 659–671.
  46. Xu P, Shi M, Lai R, Chen XX. 2012 Differences in numbers of termicins expressed in two termite species affected by fungal contamination of their environments. *Genet. Mol. Res.* **11**, 2247–2257. (doi:10.4238/2012.May.10.2)
  47. Bulmer MS, Lay F, Hamilton C. 2010 Adaptive evolution in subterranean termite antifungal peptides. *Insect Mol. Biol.* **19**, 669–674. (doi:10.1111/j.1365-2583.2010.01023.x)
  48. Bulmer MS, Crozier RH. 2004 Duplication and diversifying selection among termite antifungal peptides. *Mol. Biol. Evol.* **21**, 2256–2264. (doi:10.1093/molbev/msh236)
  49. Guo Y, Shen YH, Sun W, Kishino H, Xiang ZH, Zhang Z. 2011 Nucleotide diversity and selection signature in the domesticated silkworm, *Bombyx mori*, and wild silkworm, *Bombyx mandarina*. *J. Insect Sci.* **11**, 155. (doi:10.1673/031.011.15501)
  50. Fave MJ, Johnson RA, Cover S, Handschuh S, Metscher BD, Muller GB, Gopalan S, Abouheif E. 2015 Past climate change on Sky Islands drives novelty in a core developmental gene network and its phenotype. *BMC Evol. Biol.* **15**, 183. (doi:10.1186/s12862-015-0448-4)
  51. Xu P, Shi M, Chen XX. 2009 Antimicrobial peptide evolution in the Asiatic honey bee *Apis cerana*. *PLoS ONE* **4**, e4239. (doi:10.1371/journal.pone.0004239)
  52. Harris C, Lambrechts L, Rousset F, Abate L, Nsango SE, Fontenille D, Morlais I, Cohuet A. 2010 Polymorphisms in *Anopheles gambiae* immune genes associated with natural resistance to *Plasmodium falciparum*. *PLoS Pathog.* **6**, e1001112. (doi:10.1371/journal.ppat.1001112)
  53. Parmakelis A, Slotman MA, Marshall JC, Awono-Ambene PH, Antonio-Nkondjo C, Simard F, Caccione A, Powell JR. 2008 The molecular evolution of four anti-malarial immune genes in the *Anopheles gambiae* species complex. *BMC Evol. Biol.* **8**, 79. (doi:10.1186/1471-2148-8-79)
  54. Lehmann T, Hume JC, Licht M, Burns CS, Wollenberg K, Simard F, Ribeiro JM. 2009 Molecular evolution of immune genes in the malaria mosquito *Anopheles gambiae*. *PLoS ONE* **4**, e4549. (doi:10.1371/journal.pone.0004549)
  55. Weetman D, Steen K, Rippon EJ, Mawejje HD, Donnelly MJ, Wilding CS. 2014 Contemporary gene flow between wild *An. gambiae* s.s. and *An. arabiensis*. *Parasit. Vectors* **7**, 345. (doi:10.1186/1756-3305-7-345)
  56. Selsted ME, Ouellette AJ. 2005 Mammalian defensins in the antimicrobial immune response. *Nat. Immunol.* **6**, 551–557. (doi:10.1038/ni1206)
  57. Padhi A, Verghese B. 2008 Molecular diversity and evolution of myticin-C antimicrobial peptide variants in the Mediterranean mussel, *Mytilus galloprovincialis*. *Peptides* **29**, 1094–1101. (doi:10.1016/j.peptides.2008.03.007)
  58. Gosset CC, Do Nascimento J, Auge MT, Bierne N. 2014 Evidence for adaptation from standing genetic variation on an antimicrobial peptide gene in the mussel *Mytilus edulis*. *Mol. Ecol.* **23**, 3000–3012. (doi:10.1111/mec.12784)
  59. Unckless RL, Rottschaefer SM, Lazzaro BP. 2015 The complex contributions of genetics and nutrition to immunity in *Drosophila melanogaster*. *PLoS Genet.* **11**, e1005030. (doi:10.1371/journal.pgen.1005030)
  60. Clemmons AW, Lindsay SA, Wasserman SA. 2015 An effector peptide family required for *Drosophila* Toll-mediated immunity. *PLoS Pathog.* **11**, e1004876. (doi:10.1371/journal.ppat.1004876)
  61. Clark AG, Wang L. 1997 Molecular population genetics of *Drosophila* immune system genes. *Genetics* **147**, 713–724.
  62. Ramos-Onsins S, Aguadé M. 1998 Molecular evolution of the Cecropin multigene family in *Drosophila*: functional genes vs. pseudogenes. *Genetics* **150**, 157–171.
  63. Erler S, Lhomme P, Rasmont P, Lattorff HMG. 2014 Rapid evolution of antimicrobial peptide genes in an insect host–social parasite system. *Infect. Genet. Evol.* **23**, 129–137. (doi:10.1016/j.meegid.2014.02.002)
  64. Viljakainen L, Evans JD, Hasselmann M, Rueppell O, Tingek S, Pamilo P. 2009 Rapid evolution of immune proteins in social insects. *Mol. Biol. Evol.* **26**, 1791–1801. (doi:10.1093/molbev/msp086)
  65. Harpur BA, Zayed A. 2013 Accelerated evolution of innate immunity proteins in social insects: adaptive evolution or relaxed constraint? *Mol. Biol. Evol.* **30**, 1665–1674. (doi:10.1093/molbev/mst061)
  66. Simard F, Licht M, Besansky NJ, Lehmann T. 2007 Polymorphism at the *defensin* gene in the *Anopheles gambiae* complex: testing different selection hypotheses. *Infect. Genet. Evol.* **7**, 285–292. (doi:10.1016/j.meegid.2006.11.004)
  67. Slotman MA, Parmakelis A, Marshall JC, Awono-Ambene PH, Antonio-Nkondjo C, Simard F, Caccione A, Powell JR. 2007 Patterns of selection in anti-malarial immune genes in malaria vectors: evidence for adaptive evolution in LRIM1 in *Anopheles arabiensis*. *PLoS ONE* **2**, e793. (doi:10.1371/journal.pone.0000793)
  68. Viljakainen L, Pamilo P. 2008 Selection on an antimicrobial peptide defensin in ants. *J. Mol. Evol.* **67**, 643–652. (doi:10.1007/s00239-008-9173-6)
  69. Noureddine R, Chauvin A, Plantard O. 2011 Lack of genetic structure among Eurasian populations of the tick *Ixodes ricinus* contrasts with marked divergence from north-African populations. *Int. J. Parasitol.* **41**, 183–192. (doi:10.1016/j.ijpara.2010.08.010)
  70. O'Huigin C, Sato A, Klein J. 1997 Evidence for convergent evolution of A and B blood group antigens in primates. *Hum. Genet.* **101**, 141–148. (doi:10.1007/s004390050603)
  71. Yu G, Baeder DY, Regoes RR, Rolff J. 2016 The more the better? Combination effects of antimicrobial peptides. *Antimicrob. Agents Chemother.* **60**, 1717–1724. (doi:10.1128/AAC.02434-15)
  72. Marxer M, Vollenweider V, Schmid-Hempel P. 2016 Insect antimicrobial peptides act synergistically to inhibit a trypanosome parasite. *Phil. Trans. R. Soc. B* **371**, 20150302. (doi:10.1098.rsth.2015.0302)