

ORIGINAL ARTICLE

Comparative analysis of five immunity-related genes reveals different levels of adaptive evolution in the *virilis* and *melanogaster* groups of *Drosophila*

R Morales-Hojas, CP Vieira, M Reis and J Vieira

Laboratório de Evolução Molecular, IBMC—Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre, Porto, Portugal

Five immunity-related genes previously reported to be evolving under positive selection in *Drosophila melanogaster* and *D. simulans* have been analysed across the *Drosophila* genus using two types of approaches, random-site and branch-site likelihood models as well as the proportion of synonymous and non-synonymous variation within and between species. Different selective pressures have been detected in the sample of genes, one showing evidence for

adaptive evolution across the phylogeny of *Drosophila* and two showing lineage-specific positive selection. Furthermore, amino-acid sites identified as being under positive selection in the *melanogaster* and the *virilis* groups are different, suggesting that the evolution of the proteins in these two divergent groups may have been shaped by different pathogens. *Heredity* (2009) **102**, 573–578; doi:10.1038/hdy.2009.11; published online 18 February 2009

Keywords: *Drosophila*; immunity; adaptive evolution; positive selection

Introduction

The immune system in *Drosophila* involves both cellular and humoral responses to pathogens (Lemaitre and Hoffmann, 2007). The cellular response consists of phagocytosis and encapsulation of parasites by differentiated hemocytes. The humoral response depends on the recognition of parasitic antigens that triggers signalling cascades that lead to the transcription of immunity peptides that are released into the haemolymph. The most important signalling pathways that induce the transcription of antimicrobial peptides are the Toll and Imd pathways (De Gregorio *et al.*, 2002).

Adaptive changes at the protein level due to the selective pressures that pathogens impose on their hosts should leave a traceable signature of positive selection in their immune-related genes. Several studies have detected these signatures in immunity genes of *Drosophila* (Hughes *et al.*, 1990; Schlenke and Begun, 2003, 2005; Lazzaro, 2005; Jiggins and Kim, 2006; Obbard *et al.*, 2006). Recently, Jiggins and Kim (2007) looked at the amino acid and silent variability of 23 immunity-related genes in both *D. melanogaster* and *D. simulans*. In the McDonald–Kreitman test, 5 out of the 23 genes showed evidence for positive selection. Also, Sackton *et al.* (2007) studied 226 genes using data from the 12 sequenced *Drosophila* species genomes in a likelihood-based framework. About 10% of the genes analysed

showed evidence of positive selection. Of the five genes that were analysed in common in both studies, three showed evidence of adaptive evolution in the Jiggins and Kim (2007) whereas Sackton *et al.* (2007) could not reject neutrality. These results can be due to the different methodologies used in both studies. One of the potential problems with phylogenetic codon-based methods is the effect that alignment of gene sequences from very divergent species might have on the estimation of ω . Homoplasy due to substitution saturation can also be a problem when applying these tests. Furthermore, lineage-specific adaptive substitutions may be difficult to detect using the commonly used likelihood-based random-site model tests as in Sackton *et al.* (2007). This is because the ω -value estimated is averaged over the entire phylogeny and will detect positive selection only if the value is greater than 1 over all branches. On the other hand, McDonald–Kreitman test can be biased by slightly deleterious mutations, which could appear as rare polymorphisms in the sequences (Eyre-Walker, 2002). Furthermore, this test will be more powerful when multiple loci are analysed simultaneously.

The five genes previously described to show evidence of adaptive evolution in *D. melanogaster* and/or *D. simulans* (Jiggins and Kim, 2007) are here analysed in more detail. CG2056 (*spirit*), CG6367 (*persephone*) and CG9631 are serine proteases of the Toll pathway. CG2056 (*spirit*), CG6367 (*persephone*) and CG9631 have been shown by RNAi screen, mutagenesis assays and microarrays, respectively, to be activated in response to bacterial and fungi infection (De Gregorio *et al.*, 2002; Ligoxygakis *et al.*, 2002; Kambris *et al.*, 2006). CG7219 encodes for a serpin, which has been shown by microarray analysis to be upregulated in response to

Correspondence: Dr R Morales-Hojas, Laboratório de Evolução Molecular, IBMC—Instituto de Biologia Molecular e Celular, Universidade do Porto, Rua do Campo Alegre 823, 4150-180, Porto, Portugal.
E-mail: rmhojas@ibmc.up.pt

Received 10 July 2008; revised 13 January 2009; accepted 19 January 2009; published online 18 February 2009

infection (De Gregorio *et al.*, 2002). Furthermore, it is a homologue of the *Anopheles* mosquitoes *SRNP6* gene that has been shown to be involved in the defence against *Plasmodium* (Abraham *et al.*, 2005). Finally, *CG12297* (*dFADD*) encodes a death domain protein and has been shown by RNAi to be involved in the Imd pathway during the antibacterial response of *Drosophila* (Leulier *et al.*, 2002). The pattern of evolution of these genes is here analysed adding additional taxa using likelihood-based models. Furthermore, the pattern of non-synonymous vs synonymous variation at the intraspecific level has been also analysed by sequencing alleles of one of these genes from a species of the *virilis* group. This allows testing if the patterns observed in the *melanogaster* group can be generalized to other *Drosophilids*. Results suggest different levels of adaptive evolution in these immunity-related genes.

Materials and methods

Genes and samples

Five genes were analysed in this study: *spirit* (CG2056), *persephone* (*psh*), CG7219, CG9631 and CG12297. These were chosen based on the evidence for adaptive evolution observed in *D. melanogaster* and/or *D. simulans* in a previous study (Jiggins and Kim, 2007) and because of their proven involvement in the immune system. It should be noted that Sackton *et al.* (2007) could not reject neutrality in genes *spirit*, *psh* and CG12297 (CG7219 and CG9631 were not included in that study). These genes are either part of the Toll pathway (*spirit* and *persephone*), upregulated by it (CG7219 and CG9631) or part of the Imd pathway (CG12297) (De Gregorio *et al.*, 2002; Ligoxygakis *et al.*, 2002; Kambris *et al.*, 2006). These pathways have been described as major regulators of the immune system in *Drosophila*.

Gene sequences for *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. mojavensis* and *D. virilis* were obtained from their respective sequenced genomes at <http://rana.lbl.gov/drosophila/>. Given the significant test results for CG9631 and CG7219 (see Results), additional species were sequenced for these two genes: *D. orena* (strain 0245.01), *D. santomea* (0271.00), *D. teissieri* (0257.01) and *D. mauritiana* (0241.01) belonging to the *melanogaster* group, and *D. americana* (both the *texana* (ML97.5, Monroe, LA, USA) and *americana* (W14, Wappapello, MO, USA) forms), *D. novamexicana* (1031.00), *D. lummei* (200, Russia), *D. ezoana* (E20, Kemi, Finland), *D. kanekoi* (1061.00), *D. montana* (Mo1, Kemi, Finland), *D. borealis* (0961.03), *D. flavomontana* (0981.00), *D. littoralis* (BP41, Bragança, Portugal) and *D. lacicola* (0991.00) belonging to the *virilis* group.

DNA extraction, gene amplification and sequencing

DNA was extracted from individual flies using QIAamp DNA Mini Kit from QIAGEN (Izasa Portugal, Lda.). Specific primers for the genes CG9631 and CG7219 were designed using Oligo v1.4 (National Biosciences Inc., Plymouth, MN, USA) and based on the gene alignments of the close relatives available. Details of gene amplifications and primers used to amplify and sequence the genes are shown in Supplementary Table 1. PCR

products were extracted from the gel using the QIAEX II agarose gel extraction kit (QIAGEN) and then cloned with the TA cloning kit from Invitrogen (Barcelona, Spain). Positive colonies were picked randomly, grown in 5 ml of LB with Ampicillin and plasmids were extracted using QIAprep Spin Miniprep Kit from QIAGEN. Three clones were sequenced for each sample to account for PCR misincorporations. Cycle sequencing was performed using ABI Big Dye v1.1 (Applied Biosystems Europe, Madrid, Spain) chemistry and reactions consisted of a first denaturation step at 96 °C for 2 min and 30 s followed by 25 cycles of 30 s at 96 °C, 15 s at 50 °C and 4 min at 60 °C. Sequencing products were run at StabVida Inc. (Lisbon, Portugal). Sequences obtained in this study have been deposited in GenBank and have accession numbers FJ608558–FJ608573, FJ615543–FJ615551 and FJ006536–FJ006551. DNA sequences were checked for reading errors with BioEdit v5.0.9 (Hall, 1999) before being aligned by eye in Proseq v2.9 (Filatov, 2002) based on the translated amino-acid sequences.

Likelihood tests of selection

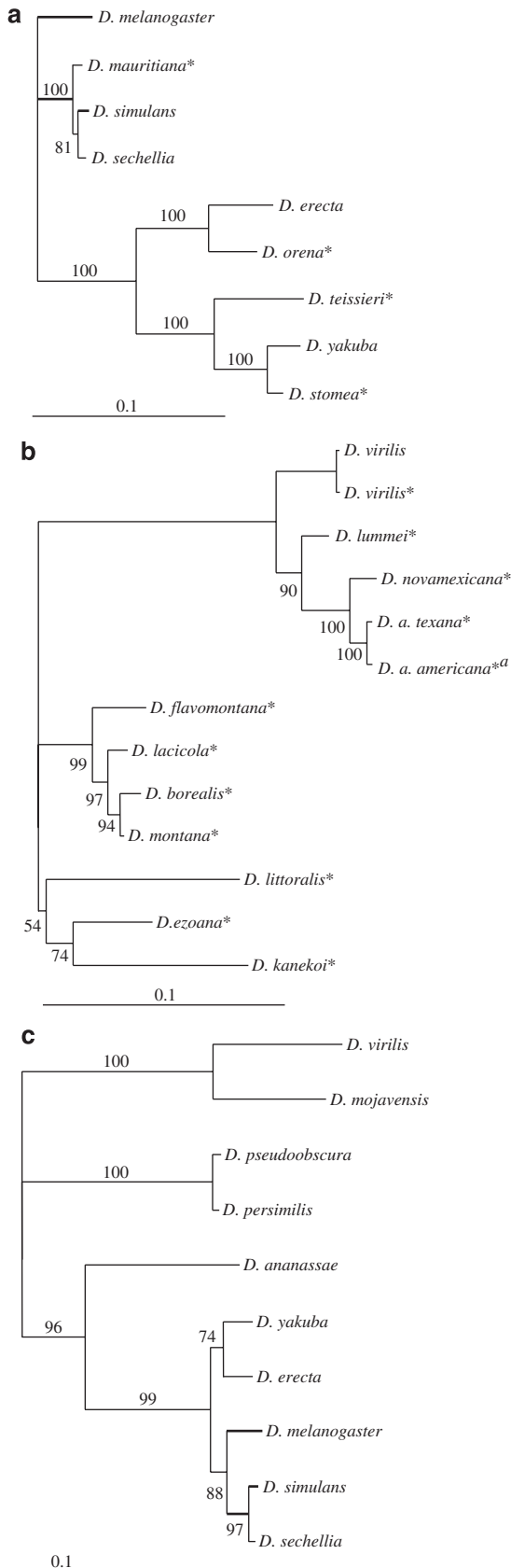
Identification of the amino-acid sites under adaptive evolution has been performed with the codeml software of the PAML 3.15 package (Yang, 1997) using random-site models. The likelihoods estimated using neutral and positive selection models were compared using a likelihood ratio test (LRT): M0 (one rate) vs M3 (discrete), M1a (nearly neutral) vs M2a (positive selection) and M7 (β ; 10 categories) vs M8 (β and $\omega > 1$; 11 categories). Branch-site models were also used to determine if genes were evolving under positive selection in specific lineages (called foreground lineages), the branch leading to *D. melanogaster*, that of *D. simulans* and the ancestral lineage of the clade (*D. simulans*, *D. sechellia*, *D. mauritiana*). Two models were compared using an LRT, the null model in which the ω of the foreground lineage is fixed to 1 against an alternative model in which the foreground ω is allowed to be greater than 1.

The phylogenetic trees used as input in the codeml analyses were those reconstructed with each of the genes. These were estimated as not all genes necessarily have to show the same phylogenetic relationships and a wrong input phylogeny would invalidate the results. Nevertheless, the topologies of the trees of each gene were all identical and in agreement with the traditional view of the group systematics (Figure 1). Phylogenetic analyses were run using maximum likelihood (ML) as optimality criterion in PAUP* v4.0b10 (Swofford, 2002). The evolutionary models employed in the ML analyses were found using the Akaike's information criterion as implemented in Modeltest 3.7 (Posada and Crandall, 1998). Heuristic searches were run with the starting tree obtained via stepwise addition and random addition of sequences with 10, 50 or 100 replicates depending on the computational weight. Tree bisection and reconnection was used as the branch-swapping algorithm.

Intraspecific analyses

Ten alleles of gene CG9631 have been sequenced from nine strains of *D. americana* (ML97-5, H1, H9, H18, LA11, LA12, LA25, LA26 and LA35; for details on the strains

see Reis *et al.*, 2008). McDonald–Kreitman tests were run in DnaSP 4.50.3 (Rozas *et al.*, 2003). The sequence of *D. virilis* was used as an outgroup.



Results

Random-site models tests

Analyses for positive selection using random-site models were run for the five genes (*spirit*, *psh*, CG7219, CG9631 and CG12297) using the sequences from the sequenced genomes of *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. mojavensis* and *D. virilis*. Only CG7219 and CG9631 showed evidence for sites under positive selection when M7 (β) and M8 ($\beta + \omega$) models were compared (Table 1). Furthermore, the lack of significant test results in *spirit*, *psh* and CG12297 is in agreement with Sackton *et al.* (2007) results (they did not include CG7219 and CG9631 in their analysis). Despite detection of positive selection in these two genes, the number of sites identified with a Bayesian posterior probability (BPP) higher than 95% was small (Table 1). These two genes were then sequenced for additional species of the *melanogaster* and the *virilis* group. Each *Drosophila* group was then analysed separately. Results are shown in Table 2. When the additional sequences were included in the analyses, CG7219 showed now no evidence of adaptive evolution whereas CG9631 showed a stronger signal for adaptive evolution when contrasting both type of models, M1a vs M2a and M7 vs M8, in both the *melanogaster* and *virilis* groups. The sites identified to be under positive selection by the M8 model are shown in Supplementary Figure 1. It should be noted that different sites have been identified for each group of *Drosophila*. Whereas in the *melanogaster* group sites identified as being under positive selection are all in exon 3, in the *virilis* group there are sites in exon 3 and also in exons 1

Table 1 Summary of random-site tests for positive selection in the sequenced species of *Drosophila*

Genes	N	LRT statistics		Parameter estimates	
		M1a vs M2a	M7 vs M8	No. of sites (M8)	ω (M8)
<i>Spirit</i>	10	0	3.9598	0	—
<i>psh</i>	10	0	0	0	—
CG7219	10	2.069	16.128***	1	2.72663
CG9631	10	0	34.158***	1	2.48944
CG12297	10	0	1.7278	0	—

N is the number of species analysed. LRT statistics shows the likelihood ratio test estimates for model comparisons. Parameter estimates shows the number of sites identified as under positive selection with a Bayesian posterior probability greater than 95% according to the Bayes empirical Bayes approach and the estimated average value of ω using the M8 model.

***Significance with $P < 0.001$.

Figure 1 Unrooted phylogenies used as input in the likelihood-based model tests. (a and b) Phylogeny obtained with CG9631 for the *melanogaster* (a) and *virilis* (b) groups. (c) Phylogeny obtained with CG12297 used in the 10 sequence species analysis. Branches in bold in (a) and (c) show the specific lineages analysed in the branch-site model tests. Numbers on branches indicate bootstrap values. Those species for which sequences were obtained in this study are indicated with an *; ^a in (b) indicates the species for which 10 alleles were sequenced in this study.

Table 2 Summary of random-site tests for positive selection in the *melanogaster* and *virilis* groups

Genes	Group	LRT statistics		Parameter estimates		
		N	M1a vs M2a	M7 vs M8	No. of sites (M8)	ω (M8)
CG7219	<i>melanogaster</i>	9	0	3.19378	0	—
	<i>virilis</i>	12	0	1.160708	0	—
CG9631	<i>melanogaster</i>	9	38.296***	44.883***	7	5.45499
	<i>virilis</i>	12	28.198***	29.772***	9	2.60316

N is the number of species analysed. LRT statistics shows the likelihood ratio test estimates for model comparisons. In the parameter estimates column it is shown the number of sites identified as under positive selection with a Bayesian posterior probability greater than 95% according to the Bayes empirical Bayes approach and the estimated average value of ω using the M8 model. ***Significance with $P < 0.001$.

and 2. Although we cannot be sure that all amino-acid sites under positive selection have been detected, the sites identified in the present analysis are likely the result of differences in selection between the two groups of *Drosophila*. It is also a noteworthy fact that the only site identified with the overall analysis to be under positive selection is not identified in any of the group-specific analyses. Thus, it is possible that the likelihood-based analyses using the *Drosophila*-sequenced genomes can be biased by the alignment of gene sequences from species that have diverged for more than 30 My and possible around 60 My (Powell, 1997).

Branch-site models tests

Drosophila melanogaster has experienced a recent expansion of its population (David and Capy, 1988) most likely involving adaptation to new environments and new pathogens. Thus, it is possible to hypothesize an increased selective pressure in the immunity-related genes in specific branches, for example, that leading to *D. melanogaster*. Furthermore, of these five genes, three have been described to be positively selected in *D. melanogaster* (*spirit*, *psh* and CG12297), one in *D. simulans* (CG7219) and CG9631 was shown to be under positive selection in both species (Jiggins and Kim, 2007). Thus, genes were also tested for the possibility of being evolving adaptively in specific lineages only. Branch-site models were run to test for the hypothesis that those genes were evolving differently in these different branches of the tree (Tables 3 and 4). Tests were significant for *spirit* ($\chi^2 = 8.688$; $P = 0.003$) and nearly significant in CG12297 ($\chi^2 = 3.788$; $P = 0.051$) when it was allowed a $\omega > 1$ in the *D. melanogaster* branch (estimated ω was 228.011 for a 0.005 proportion of sites in *spirit*, and ω of 6.576 for 0.043 proportion of sites in CG12297). In *spirit*, three positively selected sites were detected with the Bayes empirical Bayes (BEB) analysis, although just one had a BPP greater than 99%. In CG12297, seven sites were identified to be under positive selection with the BEB analysis of which just one had a BPP higher than 99%. When the ancestral branch of the (*D. simulans*, *D. sechellia*, *D. mauritiana*) clade was allowed to have a $\omega > 1$, the test was significant in CG12297 ($\chi^2 = 5.248$; $P = 0.022$) and the estimated ω was 848.426 (the proportion of sites obtained with this ω was 0.007). Three sites were identified to be adaptive with the

Table 3 Branch-site model A tests for positive selection in the *D. melanogaster* branch

Genes	LRT statistics		Parameter estimates	
	N	$\omega_2 = 1$ vs $\omega_2 > 1$	No. of Sites	($p_2 + p_3$) and ω_2
<i>spirit</i> ^a	10	8.688**	1	($p_2 + p_3$) = 0.005; $\omega_2 = 228.011$
<i>Psh</i> ^a	10	0	—	—
CG7219 ^b	10	0	—	—
CG9631 ^b	10	0	—	—
CG12297 ^a	10	3.788~	1	($p_2 + p_3$) = 0.043; $\omega_2 = 6.576$

N is the number of species analysed. LRT statistics shows the likelihood ratio test estimates for model A comparisons (with ω_2 fixed to 1 and with $\omega_2 > 1$). In the parameter estimates column it is shown the number of sites identified as under positive selection with a Bayesian posterior probability greater than 95% according to the Bayes empirical Bayes approach and the estimated average value of ω_2 .

^aAnalyses included *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. virilis* and *D. mojavensis*.

^bAnalyses included *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. mauritiana*, *D. yakuba*, *D. erecta*, *D. orena*, *D. santomea* and *D. teissieri*.

**Significance with $P < 0.01$ and ~ indicates $P = 0.051$.

Table 4 Branch-site model A tests for positive selection in the ancestral lineage of the (*D. simulans*, *D. sechellia*, *D. mauritiana*) clade

Genes	LRT statistics		Parameter estimates	
	N	$\omega_2 = 1$ vs $\omega_2 > 1$	No. of sites	($p_2 + p_3$) and ω_2
<i>spirit</i> ^a	10	0	—	—
<i>psh</i> ^a	10	0	—	—
CG7219 ^b	10	0	—	—
CG9631 ^b	10	0	—	—
CG12297 ^a	10	5.248*	0	($p_2 + p_3$) = 0.008; $\omega_2 = 848.426$

N is the number of species analysed. LRT statistics shows the likelihood ratio test estimates for model A comparisons (with ω_2 fixed to 1 and with $\omega_2 > 1$). In the parameter estimates column it is shown the number of sites identified as under positive selection with a Bayesian posterior probability greater than 95% according to the Bayes empirical Bayes approach and the estimated average value of ω_2 .

^aAnalyses included *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. yakuba*, *D. erecta*, *D. ananassae*, *D. pseudoobscura*, *D. persimilis*, *D. virilis* and *D. mojavensis*.

^bAnalyses included *D. melanogaster*, *D. simulans*, *D. sechellia*, *D. mauritiana*, *D. yakuba*, *D. erecta*, *D. orena*, *D. santomea* and *D. teissieri*.

*Significance with $P < 0.05$.

BEB analysis, although none of them had a BPP greater than 95%.

Intraspecific tests

CG9631 was reported to show a signature of positive selection in *D. melanogaster* and *D. simulans* (members of the *melanogaster* group) when the McDonald–Kreitman test was run using sequence data from these species (Jiggins and Kim, 2007). As our likelihood-based analysis also showed signature of positive selection in CG9631 in the *virilis* group of species, a randomly chosen species from this group was analysed using the McDonald–Kreitman test to determine if evidence of adaptive

evolution was also detected in this group of species. Ten partial *D. americana* alleles were obtained and aligned with that of *D. virilis*, which was used as an outgroup. The McDonald–Kreitman test was borderline significant ($P=0.055$) with an estimated α -value of 0.689. This indicates that approximately 69% of the observed amino-acid substitutions could be adaptive. The α -value estimated for *D. americana* was very similar to that reported for *D. melanogaster* and *D. simulans*, 68% (Jiggins and Kim, 2007).

Discussion

The five genes studied here are either members of the two major signalling cascades of the immune system of *Drosophila*, the Toll and Imd pathways, or have been described to be upregulated by them (De Gregorio *et al.*, 2002; Leulier *et al.*, 2002; Ligoxygakis *et al.*, 2002; Abraham *et al.*, 2005; Kambris *et al.*, 2006). All five genes have been previously reported to be positively selected in either or both *D. melanogaster* and *D. simulans* by contrasting synonymous and non-synonymous variation within and between species (Jiggins and Kim, 2007). However, with random-site likelihood-based models, M7 and M8, using the 12 sequenced *Drosophila* species, neutrality could not be rejected in *psh*, *spirit* and *CG12297* (the three genes analysed in common in these two studies; Sackton *et al.*, 2007). Results from this study suggest that positive selection is working on specific lineages in the case of *spirit* and *CG12297* because when branch-site model tests were used, evidence of positive selection on specific lineages was found for these two genes. This offers support to the previous results of Jiggins and Kim (2007), who detected positive selection in *D. melanogaster* in *spirit* and *CG12297*.

For *CG9631* and *CG7219* (not analysed by Sackton *et al.*, 2007), a reliable signal of positive selection was detected only in *CG9631* when additional species of both the *melanogaster* and *virilis* groups were added to the likelihood-based tests. These inconsistent results between the analyses including very divergent species (diverging for at least 30–40 My) and those including only species of the same group are possibly a consequence of alignment bias. Furthermore, saturation makes difficult the inference of synonymous substitutions from a distant past. Using the McDonald–Kreitman test, a nearly significant result was also obtained in *CG9631* in a sample of 10 *D. americana* alleles. This gene was also reported to be positively selected in *D. melanogaster* and *D. simulans* (Jiggins and Kim, 2007), and the reported value of α was very similar to that estimated in this study. Interestingly, branch-site model tests failed to detect positive selection in specific lineages in the *melanogaster* group. Thus, *CG9631* is positively evolving across the entire *Drosophila* genus in contrast to the above-discussed genes that seem to be evolving adaptively in specific lineages.

Estimated rates of adaptive evolution in *CG9631* from *D. americana* were similar to those reported previously for *D. melanogaster* and *D. simulans*. Approximately 68% of the observed amino-acid changes were estimated to be adaptive. This figure is higher than the 45% genomic average estimated for *Drosophila* (Eyre-Walker, 2006). Thus, some of the immunity-related genes are adapting at higher rates than the average, which is not surprising

given the relevance of the immune system for the survival of individuals.

In conclusion, this study reports different types of adaptive evolution in *Drosophila* immunity-related genes. Some genes may be under adaptive selection across the entire genus and even across different arthropod taxa (Little *et al.*, 2004; Jiggins and Kim, 2006), suggesting their relevance in the common pathways of the immune system of flies and other arthropods. There are other genes that show a lineage-specific signature, indicating that particular pathogen environments are also influencing adaptation in some species but not in others. Lineage-specific positive selection has been observed in some immune-related genes across different arthropod taxa (Jiggins and Kim, 2005; Bulmer and Crozier, 2006). Also, lineage-specific adaptive evolution within *Drosophila* has been detected in genes involved in reproductive isolation (Barbash *et al.*, 2004; Presgraves and Stephan, 2007), dosage compensation (Bachtrog, 2008) or spermatogenesis (Llopart and Comeron, 2008), for example. In the case of those immunity-related genes of this study in which *D. melanogaster*-specific positive selection was detected, this could be related to the ‘out of Africa’ population expansion that occurred approximately 10000 years ago (David and Cappy, 1988). As *D. melanogaster* expanded from sub-Saharan Africa and colonized other regions of the world, it may have come into contact with new pathogens to which the species had to adapt. Furthermore, given these different modes of evolution of genes, this study underlines the importance of complementing different approaches to the detection of positive selection. It is also very likely that the suggested 10% figure of immunity-related genes being under positive selection (Sackton *et al.*, 2007) is an underestimation.

Acknowledgements

This work was funded by the projects POCI/BIA-BCM/59399/2004 and PTDC/BIA-BDE/66765/2006, funded by POCI 2010, cofunded by FEDER funds. RM-H. was partially funded by the FCT fellowship SFRH/BPD/24329/2005. The fly collection was funded by APBRF.

References

- Abraham EG, Pinto SB, Ghosh A, Vanlandingham DL, Budd A, Higgs S *et al.* (2005). An immune-responsive serpin, SRPN6, mediates mosquito defense against malaria parasites. *Proc Natl Acad Sci USA* **102**: 16327–16332.
- Bachtrog D (2008). Positive selection at the binding sites of the male-specific lethal complex involved in dosage compensation in *Drosophila*. *Genetics* **180**: 1123–1129.
- Barbash DA, Awadalla P, Tarone AM (2004). Functional divergence caused by ancient positive selection of a *Drosophila* hybrid incompatibility locus. *PLoS Biol* **2**: e142.
- Bulmer MS, Crozier RH (2006). Variation in positive selection in termite GNBPs and Relish. *Mol Biol Evol* **23**: 317–326.
- David JR, Cappy P (1988). Genetic variation of *Drosophila melanogaster* natural populations. *Trends Genet* **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**: 2568–2579.
- Eyre-Walker A (2002). Changing effective population size and the McDonald–Kreitman test. *Genetics* **162**: 2017–2024.
- Eyre-Walker A (2006). The genomic rate of adaptive evolution. *Trends Ecol Evol* **21**: 569–575.

- Filatov D (2002). ProSeq: a software for preparation and evolutionary analysis of DNA sequence data sets. *Mol Ecol Notes* **2**: 621–624.
- Hall T (1999). BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symp Ser* **41**: 95–98.
- Hughes AL, Ota T, Nei M (1990). Positive Darwinian selection promotes charge profile diversity in the antigen-binding cleft of class I major-histocompatibility-complex molecules. *Mol Biol Evol* **7**: 515–524.
- Jiggins FM, Kim KW (2005). The evolution of antifungal peptides in *Drosophila*. *Genetics* **171**: 1847–1859.
- Jiggins FM, Kim KW (2006). Contrasting evolutionary patterns in *Drosophila* immune receptors. *J Mol Evol* **63**: 769–780.
- Jiggins FM, Kim KW (2007). A screen for immunity genes evolving under positive selection in *Drosophila*. *J Evol Biol* **20**: 965–970.
- Kambris Z, Brun S, Jang IH, Nam HJ, Romeo Y, Takahashi K et al. (2006). *Drosophila* immunity: a large-scale *in vivo* RNAi screen identifies five serine proteases required for Toll activation. *Curr Biol* **16**: 808–813.
- Lazzaro BP (2005). Elevated polymorphism and divergence in the class C scavenger receptors of *Drosophila melanogaster* and *D. simulans*. *Genetics* **169**: 2023–2034.
- Lemaitre B, Hoffmann J (2007). The host defense of *Drosophila melanogaster*. *Annu Rev Immunol* **25**: 697–743.
- Leulier F, Vidal S, Saigo K, Ueda R, Lemaitre B (2002). Inducible expression of double-stranded RNA reveals a role for dFADD in the regulation of the antibacterial response in *Drosophila* adults. *Curr Biol* **12**: 996–1000.
- Ligoxygakis P, Pelte N, Hoffmann JA, Reichhart JM (2002). Activation of *Drosophila* Toll during fungal infection by a blood serine protease. *Science* **297**: 114–116.
- Little TJ, Colbourne JK, Crease TJ (2004). Molecular evolution of *Daphnia* immunity genes: polymorphism in a gram-negative binding protein gene and an alpha-2-macroglobulin gene. *J Mol Evol* **59**: 498–506.
- Llopart A, Comeron JM (2008). Recurrent events of positive selection in independent *Drosophila* lineages at the spermatogenesis gene *roughex*. *Genetics* **179**: 1009–1020.
- Obbard DJ, Jiggins FM, Halligan DL, Little TJ (2006). Natural selection drives extremely rapid evolution in antiviral RNAi genes. *Curr Biol* **16**: 580–585.
- Posada D, Crandall KA (1998). MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**: 817–818.
- Powell J (1997). *Progress and Prospects in Evolutionary Biology: The Drosophila Model*. Oxford University Press: New York. 562pp.
- Presgraves DC, Stephan W (2007). Pervasive adaptive evolution among interactors of the *Drosophila* hybrid inviability gene, *Nup96*. *Mol Biol Evol* **24**: 306–314.
- Reis M, Vieira CP, Morales-Hojas R, Vieira J (2008). An old *bilbo*-like non-LTR retroelement insertion provides insight into the relationship of species of the *virilis* group. *Gene* **425**: 48–55.
- Rozas J, Sanchez-DelBarrio JC, Messeguer X, Rozas R (2003). DnaSP, DNA polymorphism analyses by the coalescent and other methods. *Bioinformatics* **19**: 2496–2497.
- 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.
- Schlenke TA, Begun DJ (2003). Natural selection drives *Drosophila* immune system evolution. *Genetics* **164**: 1471–1480.
- Schlenke TA, Begun DJ (2005). Linkage disequilibrium and recent selection at three immunity receptor loci in *Drosophila simulans*. *Genetics* **169**: 2013–2022.
- Swofford DL (2002). *PAUP*: Phylogenetic Analysis Using Parsimony (and other methods)*, ver. 4.0b10. Sinauer: Sunderland.
- Yang Z (1997). PAML: a program package for phylogenetic analysis by maximum likelihood. *Comput Appl Biosci* **13**: 555–556.

Supplementary Information accompanies the paper on Heredity website (<http://www.nature.com/hdy>)