JOURNAL OF Evolutionary Biology

doi: 10.1111/jeb.12574

Molecular evolution of candidate genes involved in post-mating-prezygotic reproductive isolation

J. M. BONO*[1], L. M. MATZKIN†‡[1], K. HOANG*[2] & L. BRANDSMEIER†

*Biology Department, University of Colorado Colorado Springs, Colorado Springs, CO, USA
†Department of Biological Sciences, University of Alabama in Huntsville, Huntsville, AL, USA
‡HudsonAlpha Institute for Biotechnology, Huntsville, AL, USA

Keywords:

accessory gland proteins;
divergent selection;
Drosophila arizonae;
Drosophila mojavensis;
fertility;
reproductive incompatibilities;
reproductive tract;
sexual conflict;
sexual selection;
speciation.

Abstract

Traits involved in post-copulatory interactions between the sexes may evolve rapidly as a result of sexual selection and/or sexual conflict, leading to post-mating-prezygotic (PMPZ) reproductive isolating barriers between diverging lineages. Although the importance of PMPZ isolation is recognized, the molecular basis of such incompatibilities is not well understood. Here, we investigate molecular evolution of a subset of *Drosophila mojavensis* and *Drosophila arizonae* reproductive tract genes. These include genes that are transcriptionally regulated by conspecific mating in females, many of which are misregulated in heterospecific crosses, and a set of male genes whose transcripts are transferred to females during mating. As a group, misregulated female genes are not more divergent and do not appear to evolve under different selection pressures than other female reproductive genes. Male transferred genes evolve at a higher rate than testis-expressed genes, and at a similar rate compared to accessory gland protein genes, which are known to evolve rapidly. Four of the individual male transferred genes show patterns of divergent positive selection between *D. mojavensis* and *D. arizonae*. Three of the four genes belong to the sperm-coating protein-like family, including an ortholog of *antares*, which influences female fertility and receptivity in *Drosophila melanogaster*. Synthesis of these molecular evolutionary analyses with transcriptomics and predicted functional information makes these genes candidates for involvement in PMPZ reproductive incompatibilities between *D. mojavensis* and *D. arizonae*.

Introduction

Speciation results from the accumulation of reproductive isolating barriers that prevent gene flow between diverging populations (Dobzhansky, 1951; Coyne & Orr, 2004). These barriers arise when genetic differences lead to incompatible interactions that prevent the formation of hybrid zygotes (prezygotic) or produce hybrids with low or no fitness (post-zygotic). Although reproductive isolating barriers have been studied in detail at the phenotypic level, identifying the underlying genetic basis of incompatibilities has proven much more challenging, particularly for prezygotic incompatibilities (Coyne & Orr, 2004).

For internally fertilizing organisms with promiscuous mating systems, post-mating-prezygotic (PMPZ) isolating barriers may evolve early in the speciation process because sexual selection and/or sexual conflict are expected to drive rapid divergence of reproductive traits (Birkhead & Pizzari, 2002; Arnqvist & Rowe, 2005; Howard *et al.*, 2009). In fact, considerable evidence demonstrates that reproductive proteins are among the

Correspondence: Jeremy M. Bono, Biology Department, University of Colorado Colorado Springs, 1420 Austin Bluffs Parkway, Colorado Springs, CO 80918, USA.
Tel.: +1 719 255 3668; fax: +1 719 255 3047; e-mail: jbono@uccs.edu
Luciano M. Matzkin, Department of Biological Sciences, University of Alabama in Huntsville, 301 Sparkman Drive, Huntsville, AL 35899, USA.
Tel.: +1 256 824 4326; fax: +1 256 824 6305; e-mail: lmm0015@uah.edu
[1]These authors contributed equally.
[2]Present address: Department of Biology, Emory University, 1510 Clifton Road NE, Atlanta, GA 30322, USA

most rapidly evolving proteins in a wide array of species (Swanson & Vacquier, 2002). In *Drosophila*, males transfer hundreds of proteins in their seminal fluid during mating, including accessory gland proteins (ACPs), which are known to modulate a variety of physiological processes in females (Wolfner, 2007; Pitnick *et al.*, 2009). The best studied of these, sex peptide (SP), interacts within a network of other proteins to affect female fertility and receptivity (Chen *et al.*, 1988; Yapici *et al.*, 2008; Ravi Ram & Wolfner, 2009; Findlay *et al.*, 2014). Female reproductive tract proteins also evolve rapidly, suggesting coevolution between the sexes (Swanson *et al.*, 2001, 2004; Kelleher *et al.*, 2007; Kelleher & Markow, 2009; Prokupek *et al.*, 2010). Although the functional significance of reproductive proteins and their propensity for rapid evolution suggest a likely role in the evolution of reproductive incompatibilities between diverging populations, we still know relatively little about the genetic basis of PMPZ isolation.

Drosophila mojavensis and *Drosophila arizonae* are sister species with partially overlapping distributions in arid regions of the south-western United States and north-western Mexico. Both species have highly promiscuous mating systems, suggesting that sexual selection and/or sexual conflict may drive the rapid evolution of reproductive traits. Reproductive isolation between these species has been studied extensively, with both prezygotic and post-zygotic barriers contributing to isolation (Markow *et al.*, 1983, 2007; Markow & Hocutt, 1998; Kelleher & Markow, 2007; Bono & Markow, 2009). Although the magnitude of most isolating barriers between *D. mojavensis* and *D. arizonae* varies depending on the source population of males and females (Markow, 1981; Reed & Markow, 2004; Massie & Markow, 2005), PMPZ isolation is strong in crosses involving *D. mojavensis* females from all populations (Kelleher & Markow, 2007), suggesting an early origin during the process of divergence. Incompatible interactions within the female reproductive tract invariably result in the production of few hybrid offspring due to problems with storage and transfer of sperm and with the degradation of the insemination reaction, which forms in the female reproductive tract immediately after mating (Kelleher & Markow, 2007). Conspecifically mated females breakdown the insemination reaction over the course of several hours, but the reaction can persist for much longer in heterospecific crosses, impeding oviposition and potentially permanently sterilizing females (Knowles & Markow, 2001; Kelleher & Markow, 2007).

In a previous study, we compared the identity and abundance of transcripts in the lower reproductive tracts of *D. mojavensis* females that were virgin or mated to either *D. mojavensis* or *D. arizonae* males (Bono *et al.*, 2011). We found 15 genes that were differentially expressed in conspecifically mated females relative to virgins, which we consider to represent the 'normal' post-mating transcriptional response. Of these genes, nine were misregulated when females mated with *D. arizonae* males (heretofore 'misregulated genes'), whereas expression of the other six was normal (heretofore 'normally regulated genes'). Separately from the misexpression analysis, we detected 12 transcripts that were transferred from males to females during mating, a number of which were previously identified candidate *ACPs* (heretofore 'male genes'). Transcript abundance of all of these genes differed in con- vs. heterospecifically mated females. Although the mechanism of RNA transfer is unknown, recent findings in *Drosophila melanogaster* show that males transfer exosomes to females during mating (Corrigan *et al.*, 2014). Although the contents of these vesicles were not reported, other studies have shown that exosomes can carry a variety of materials including mRNA (Valadi *et al.*, 2007). In *D. melanogaster*, the exosomes fuse with sperm and female epithelial cells and are required for full inhibition of female remating behaviour (Corrigan *et al.*, 2014). These findings, coupled with increasing recognition that extracellular RNA may play diverse signalling roles (Dinger *et al.*, 2008), suggest that regardless of the mechanism of transfer, these transcripts may be important functionally.

In this study, we examine molecular evolution of genes belonging to each of the three gene sets (normally regulated, misregulated and male transferred) from Bono *et al.* (2011). We first investigate whether genes in each set differ in average coding sequence divergence and/or evolve under different selective pressures compared to other female and male reproductive tract genes. We then analyse genes individually for evidence of positive selection, reasoning that mating genes under divergent selection pressures between *D. mojavensis* and *D. arizonae* are candidates for involvement in PMPZ incompatibilities.

Materials and methods

Gene annotation and function

As *Drosophila* gene annotations and information on gene function are continually revised, we provide an updated annotation of genes identified by Bono *et al.* (2011). We first identified *D. melanogaster* orthologs using OrthoDB (Waterhouse *et al.*, 2013) and then obtained gene ontology terms from FlyBase (Marygold *et al.*, 2013). In addition, we used BLASTP against the Conserved Domain Database (CDD) to identify conserved protein domains in each *D. mojavensis* gene (Marchler-Bauer *et al.*, 2013).

PCR amplification and gene sequencing

We obtained gene sequences from one individual from each of the four *D. mojavensis* subspecies: *D. mojavensis mojavensis*, *D. mojavensis baja*, *D. mojavensis sonorensis*

and *D. mojavensis wrigleyi* (Fig. 1), which are geographically isolated and highly genetically differentiated (Machado *et al.*, 2007; Reed *et al.*, 2007; Pfeiler *et al.*, 2009). The *D. mojavensis wrigleyi* sequences were extracted from the published *D. mojavensis* genome sequence (*Drosophila 12 Genomes Consortium*, 2007), whereas sequences for the other three subspecies were obtained from isofemale lines (*D. mojavensis baja*: MJBC-216, collected in La Paz, Baja California Sur, Mexico, in 2000; *D. mojavensis sonorensis*: OPNM-2, collected in Organ Pipe National Monument, Arizona, USA, in 2007; *D. mojavensis mojavensis*: WC-9, collected in Whitman Canyon, Arizona, USA, in 2002). We also sequenced genes from two *D. arizonae* lines representing northern (isofemale: AZTUS-31, collected in Tucson, Arizona, USA, in 2000) and southern (*Drosophila* Species Stock Center id: 15081-1271.14) portions of its range (Fig. 1). Northern and southern *D. arizonae* populations are geographically separated by the Sierra Madre Occidental mountain range and the Trans-Mexican Volcanic belt. Previous studies have shown that they are genetically differentiated (Machado *et al.*, 2007; Reed *et al.*, 2007) and show some evidence of reproductive isolation (Massie, 2006). We also attempted to amplify each gene from *Drosophila navojoa* (*Drosophila* Species Stock Center id: 15081-1374.12), the closest relative of *D. mojavensis* and *D. arizonae*. Amplified DNA was sequenced either at a Genewiz (Plainfield, New Jersey) or Eurofins MWG Operon (Huntsville, Alabama) sequencing facility. Primer sequences are listed in Table S1. Repeated attempts at sequencing two genes (*GI14846* from the misregulated set and *GI23381* from the male set) were unsuccessful, so they were excluded from subsequent analyses leaving a total of 25 genes. All sequences generated in the study are archived under GenBank accession numbers: KP221204–KP221254 and KP236458–KP236543.

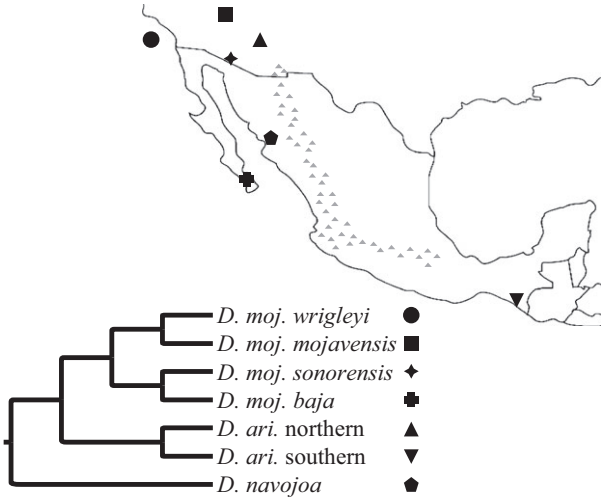

Fig. 1 Map showing geographic locations of *Drosophila mojavensis*, *Drosophila arizonae* and *Drosophila navojoa* strains used in molecular evolutionary analyses and phylogenetic relationships based on Machado *et al.* (2007). Grey triangles denote the Sierra Madre Occidental mountain range and the Trans-Mexican Volcanic belt.

Molecular evolutionary analyses

We were first interested in analysing whether genes in our data set evolved at a different rate relative to other male and female reproductive tract genes. To accomplish this, we utilized the *D. mojavensis wrigleyi* genome sequence and sequences available from two previously published studies on *D. arizonae* male (Wagstaff & Begun, 2005) and female (Kelleher *et al.*, 2007) reproductive tract genes. This yielded a total of 27 male reproductive genes (13 *ACPs* and 14 genes expressed in the testes originally sequenced from *Drosophila* Species Stock Center line: 15081-1271.00) and 607 female reproductive tract genes, which we used as a basis for comparison with our data set (6 normally regulated, 8 misregulated and 11 male). For each gene, we generated pairwise alignments between sequences from *D. mojavensis wrigleyi* and northern *D. arizonae* (sequences were either obtained from this study or from the two previously published studies) using CLUSTALW. We used the KAKS CALCULATOR software package (Zhang *et al.*, 2006) to calculate two metrics of evolutionary rate: (i) nonsynonymous divergence (number of nonsynonymous substitutions/number of nonsynonymous sites) and (ii) d*N*/d*S* (rate of nonsynonymous substitution/rate of synonymous substitution). The first metric allows us to compare overall divergence at potentially functional sites, whereas the second allows us to examine possible differences in selection pressures between gene sets. We separately analysed female (misregulated vs. normally regulated vs. female background) and male (male transferred vs. *ACPs* vs. testes expressed) gene sets using Kruskal–Wallis tests. Genes with no synonymous substitutions were excluded from the d*N*/d*S* analyses, because the rate is undefined.

Our other main objective was to analyse genes individually to identify those with evidence of divergent selection between *D. mojavensis* and *D. arizonae*, as these would be particularly strong candidates for involvement in reproductive incompatibilities. We aligned all *D. mojavensis, D. arizonae* and *D. navojoa* sequences using the CLUSTALW (Larkin *et al.*, 2007) algorithm implemented in GENEIOUS v.5. (Kearse *et al.*, 2012). Alignments were further evaluated and adjusted manually when necessary. The appropriate nucleotide substitution model for each gene was determined using the codon model selection (CMS) algorithm available on the HyPhy Datamonkey web server (www.datamonkey.org; Delport *et al.*, 2010), and the selected models were used for all tests described below. Although limited gene flow among geographically isolated populations should reduce the

likelihood of recombination between alleles of the different populations, we investigated whether there was evidence for recombination in our data set using the Genetic Algorithm for Recombination Detection (GARD; Kosakovsky Pond *et al.*, 2006) on the Datamonkey web server. Break points identified by GARD (multiple tree model preferred over single tree model by an evidence ratio > 100) were further investigated using the Kishino–Hasegawa (K–H) test, which helps rule out spatial rate variation and/or heterotachy as explanations for topological incongruence (Kishino & Hasegawa, 1989).

Comparative phylogenetic tests of selection on individual genes were based on the ratio of the rate of nonsynonymous to synonymous substitutions (dN/dS). We performed several analyses available on the Datamonkey web server aimed at detecting signals of either pervasive or episodic positive selection on individual branches and/or codons (Table 1). For all tests, we supplied a phylogenetic tree based on established relationships between subspecies and species (Machado *et al.*, 2007). To identify positive selection on specific branches of the phylogenetic tree, we performed the Genetic Algorithm Branch (GA-Branch) test (Kosakovsky Pond & Frost, 2005) and the Branch-site Random Effects Likelihood (BS-REL) test (Kosakovsky Pond *et al.*, 2011). The GA-Branch test assigns branches to dN/dS rate classes (assuming branch variation but no site to site variation) to detect lineages experiencing pervasive positive selection (dN > dS averaging over sites). In addition to selecting the model that best fits the data, GA-Branch also examines other models within a 95% confidence set to calculate model-averaged probabilities that dN > dS for every branch on the tree. In contrast, BS-REL permits dN/dS rate variation across sites and lineages to identify branches where selection is episodic, meaning that only a subset of sites evolves under positive selection rather than on average over all sites. The test computes P-values for the probability that individual branches experienced positive selection and also returns adjusted P-values that account for multiple testing using the Holm–Bonferroni method. We considered support for positive selection rigorous when adjusted P-values were ≤ 0.05. We were primarily interested in genes with evidence of divergent selection in *D. mojavensis* and *D. arizonae*, as these are the strongest candidates for involvement in reproductive incompatibilities between the species. For example, a gene evolving under positive selection along the branch leading to all *D. mojavensis* subspecies or the branch leading to *D. arizonae* populations would be a strong candidate, whereas a gene with evidence of positive selection in only one *D. mojavensis* subspecies or one *D. arizonae* population would not be a strong candidate.

We also performed two tests that identify specific amino acid sites that have evolved under pervasive or episodic positive selection. The Fast Unconstrained Bayesian Approximation (FUBAR) test (Murrell *et al.*, 2012) detects pervasive positive selection on specific sites in an alignment. This test computes Bayesian posterior probabilities that dN > dS for a given site in an alignment. Given the low rate of false positives produced by FUBAR (Murrell *et al.*, 2012), we used the default posterior probability cut-off of ≥ 0.90 to determine significance, and we also report the expected number of false-positive sites in a given gene. The Mixed Effects Model of Evolution (MEME) test (Murrell *et al.*, 2012) is another branch-site model that allows dN/dS to vary among sites and lineages. In contrast to BS-REL, MEME identifies individual sites under selection rather than inferring which branches have experienced selection. The test computes P-values for the probability that a particular site experienced positive selection on at least a subset of branches (we report any sites where $P \leq 0.1$). The test also outputs q-values for each site, which control the false discovery rate using Simes' procedure. Although we interpret results based on q-values, we also note that they are likely to be highly conservative due to the use of a neutral null model (Murrell *et al.*, 2012). We supplied phylogenetic trees for FUBAR and MEME analyses unless there was evidence of recombination, in which case we used GARD-inferred trees. Sites under positive selection identified by FUBAR and/or MEME were further evaluated to determine whether there were fixed differences between *D. mojavensis* and *D. arizonae*, and, if possible, polarized the variants using the outgroup species *D. navojoa*. For some genes with conserved protein domains, information on putative functional sites was also available. We therefore evaluated whether any fixed differences between the species and/or positively selected sites occurred in predicted functional residues.

Table 1 Summary of tests used to detect different signatures of positive selection.

Test	Signal detected	Reference
Genetic Algorithm Branch Test (GA-Branch)	Pervasive positive selection on specific branches	Kosakovsky Pond & Frost (2005)
Branch-site Random Effects Likelihood (BS-REL)	Episodic positive selection on some sites along specific branches	Kosakovsky Pond *et al.* (2011)
Fast Unconstrained Bayesian Approximation (FUBAR)	Specific sites experiencing pervasive positive selection	Murrell *et al.* (2013)
Mixed Effects Model of Evolution (MEME)	Specific sites experiencing episodic positive selection	Murrell *et al.* (2012)

We do note a caveat with regard to the use of d*N*/d*S* to identify genes under selection in our system. Specifically, recent studies have suggested that estimates of d*N*/d*S* are time sensitive, leading to increased risk of Type I or Type II errors when divergence time is short, and/or when samples are from nonindependent populations (Rocha *et al.*, 2006; Kryazhimskiy & Plotkin, 2008; Mugal *et al.*, 2014). Divergence time between *D. mojavensis* and *D. arizonae* is estimated to be approximately *4N* generations (Matzkin & Eanes, 2003; Matzkin, 2004). Whereas a past study suggested divergence of *N* generations as a benchmark for studying independent populations (Rocha *et al.*, 2006), a more recent analysis suggests longer divergence times might be necessary, particularly when mutation rates are low or genes are short in length (Mugal *et al.*, 2014). However, this study focused only on approaches inferring selection across entire genes, and it is not clear what effect, if any, short divergence time has on inferring selection on single codon sites (Mugal *et al.*, 2014), which constitutes the majority of tests we performed here. We also note that alternative tests of selection that consider polymorphism data may be similarly affected by relatively short divergence time (e.g. McDonald–Kreitman test; Mugal *et al.*, 2014) or geared towards detecting selection at a timescale that is too recent for comparisons between *D. mojavensis* and *D. arizonae* (e.g. tests based on the frequency spectrum or F_{ST} outlier tests). Although we thus acknowledge the above caveat, we feel that our approach is reasonable given these considerations.

Results

Gene annotation and functional information

Although the small size of our data set precluded gene ontology overrepresentation tests, an informal examination of gene ontology terms and conserved domains revealed a few interesting patterns. For example, multiple female expressed genes have functions related to each of the following categories: translation initiation, lipid binding/recognition and chitin binding (Table S2). Notably, four of the male transferred genes contain sperm-coating protein (SCP)-like domains (Table S2). One of these genes, *GI23009*, also contains a region with homology to the herpesvirus latent membrane protein 1 domain. Two other male genes have conserved domains indicating lipase activity, whereas an additional two genes are previously identified *Acps* of unknown function.

Molecular evolutionary analyses

There were no statistically supported differences in divergence at nonsynonymous sites (Kruskal–Wallis test, $\chi^2_2 = 2.1$, $P = 0.34$) or d*N*/d*S* ratios (Kruskal–Wallis test, $\chi^2_2 = 1.8$, $P = 0.40$; Fig. S1) for female genes belonging to the three sets (normally regulated, misregulated and female background). For male genes, there was statistical support for differences in nonsynonymous divergence (Kruskal–Wallis test, $\chi^2_2 = 21.4$, $P < 0.001$; Fig. S1) and d*N*/d*S* rates (Kruskal–Wallis test, $\chi^2_2 = 20.1$, $P < 0.001$; Fig. S2) among the three gene sets (*Acps*, testes and male transferred). Subsequent *post hoc* pairwise comparisons using the Dunn–Bonferroni correction for multiple testing revealed that testis-specific genes had both lower nonsynonymous divergence (testis specific vs. *Acps*: $P < 0.001$; testis specific vs. transferred: $P = 0.023$; Fig. S1) and lower d*N*/d*S* ratios (testis specific vs. *Acps*: $P < 0.001$; testis specific vs. transferred: $P < 0.02$) than *Acps* or male transferred genes, which did not differ from one another (Fig. S2). We also performed the same analysis using only the six transferred genes that have been previously confirmed as *Acps* (the origin of the others in male reproductive tracts is unknown) and the conclusions were the same (data not shown). Although there was no evidence that, as a group, the male transferred genes evolved under different selection pressures compared to other *Acps*, we do note that one of these genes (*GI11629*) had the highest pairwise d*N*/d*S* ratio (7.9) of any male or reproductive tract gene in the data set.

For comparative tests, we obtained sequences from all six subspecies/populations of *D. mojavensis* and *D. arizonae* for the 25 genes. However, we excluded *GI20303* (normally regulated set) from all analyses due to the lack of synonymous substitutions, meaning that analyses were carried out on 24 genes in total. We also excluded the southern *D. arizonae* sequence for *GI11629* due to the presence of a frame-shifting 8-bp deletion. This frameshift leads to a premature stop codon resulting in a loss of 171 of the 253 amino acids that make up the full protein, including functional residues. We were successful in sequencing *D. navojoa* for all but six of the genes (*GI10424, GI11600, GI11629, GI20219, Dmoj/Acp2* and *Dmoj/Acp11*).

GARD analysis detected evidence for topological incongruence in five genes, with one break point from *GI11629* supported by the K–H test ($\alpha = 0.05$; Table S3). K–H tests for the remaining four genes were not significant, suggesting that spatial variation and/or heterotachy, rather than recombination, may explain the break point identified by GARD (Table S3). Nevertheless, as a precaution, we used GARD-inferred trees for subsequent analyses for all five of these genes.

The GA-Branch test identified 10 genes in which the best-fitting model included a rate class with d*N*/d*S* > 1 for some portion of the tree. To further evaluate the strength of support for positive selection, this test also examines other models in the 95% confidence set to calculate model-averaged probabilities that d*N*/d*S* > 1 for specific branches on the tree. These analyses indicated robust support for d*N*/d*S* > 1 (≥ 95% probability)

along at least one branch for three of the ten genes (*GI11600*, *GI11629*, and *GI20219*). Positive selection was strongly supported on branches separating *D. mojavensis* and *D. arizonae* for *GI11629* and *GI20219* (both from the male set), making these particularly strong candidates for potential involvement in PMPZ incompatibilities (Fig. 2). The best-fitting model for *GI11629* included a d*N*/d*S* rate class of 10 000, which is a default maximum that occurs when there are no inferred synonymous substitutions along a branch. When divergence is low, interpretation of such a result warrants caution because there could be very few nonsynonymous substitutions driving the signal. However, this is clearly not the case here, as sequences are highly divergent at nonsynonymous sites (the fewest nonsynonymous differences between any two sequences is 11). Although the best-fitting model for *GI11600* indicated positive selection on all branches of the tree, only the branch leading to *D. mojavensis sonorensis* was well supported (Fig. S3). The best-fitting model for four other genes (*GI16692*, *GI18586*, *GI22307* and *GI23726*) indicated divergent selection between the two species, but selection was not rigorously supported on any individual branch (Fig. S3). For the remaining three genes (*GI10424*, *GI14885* and *GI23324*), the elevated d*N*/d*S* indicated by the best-fitting model was not found specifically on branches separating *D. mojavensis* and *D. arizonae*, and no individual branches had strong support for positive selection.

The Branch-site REL analyses revealed episodic positive selection on branches in three genes (*GI11629*, *Dmoj/Acp2* and *GI23009*). However, after correcting for multiple testing, only branches identified in *GI11629* and *Dmoj/Acp2* remained significant. Although the test on *GI11629* indicated divergent selection in *D. mojavensis and D. arizonae* (Fig. 3), evidence for selection on *Dmoj/Acp2* was found only on the branch leading to *D. mojavensis baja*. The results for *GI23009* indicated divergent selection between the species ($P = 0.02$ for the branch leading to *D. mojavensis*), but this was not rigorously supported after correcting for multiple testing ($P = 0.21$).

FUBAR identified 18 sites in seven genes that showed evidence of pervasive positive selection (Table 2). We also report the expected number of false positives for each gene; in each case, the number of sites under selection exceeds the expected value. Selected sites in four of these genes showed fixed differences between *D. mojavensis* and *D. arizonae* (*GI11629, GI19546, GI20219* and *GI23009*. MEME identified episodic positive selection on 10 sites in five genes, with sites in two genes (*GI11629* and *GI17858*) involving fixed differences between the species (Table 3). However, no sites

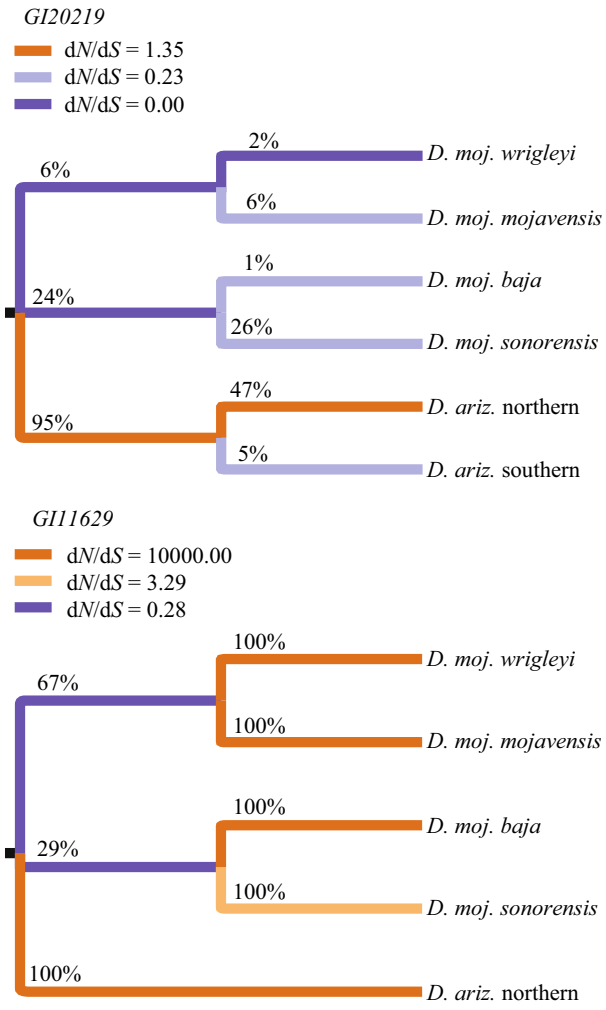

Fig. 2 Results from GA-Branch tests for two genes with the strongest signal of divergent selection between *Drosophila mojavensis* and *Drosophila arizonae*. Colour coding represents d*N*/d*S* rates estimated by the best-fitting model. Percentages are the model-averaged probabilities that d*N* > d*S* for individual branches.

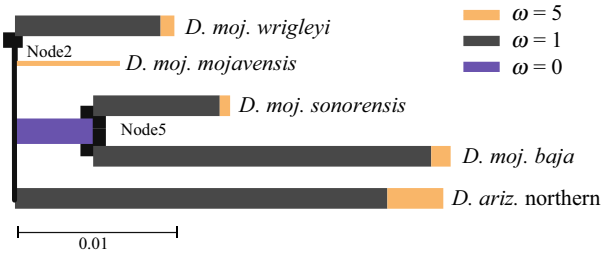

Fig. 3 Branch-site REL results for *GI11629*. Thicker orange lines indicate branches that show evidence of episodic selection after correction for multiple testing using the Holm–Bonferroni procedure (*Drosophila mojavensis wrigleyi*, $P = 0.01$; *D. mojavensis sonorensis*, $P = 0.01$; *D. mojavensis baja*, $P < 0.001$; *Drosophila arizonae*, $P < 0.001$). The length of the orange branch is proportional to the amount of positive selection detected on the branch. The scale bar represents expected substitutions/nucleotide.

Table 2 Sites under positive selection as indicated by Fast Unconstrained Bayesian Approximation. Bayesian posterior probabilities and the expected number of false-positive sites for each gene are given.

Gene	Gene set	Site	Posterior probability $dN > dS$	Expected no. false positives [CI]
GI11600	Misregulated	57	0.95	0.05 [0–1]
GI11629	Male	29*	0.92	0.46 [0–2]
		58*	0.97	
		62	0.96	
		93	0.94	
		156	0.93	
		179	0.96	
		203	0.91	
		214	0.95	
GI16692	Normally regulated	111	0.93	0.14 [0–1]
		118	0.93	
GI17858	Male	161	0.92	0.08 [0–1]
GI19546	Male	245*	0.92	0.08 [0–1]
GI20219	Male	57*	0.90	0.10 [0–1]
GI23009	Male	223*	0.97	0.27 [0–1]
		224	0.90	
		239	0.91	
		346	0.96	

*Fixed differences between *Drosophila mojavensis* and *Drosophila arizonae*.

Table 3 Sites under positive selection as indicated by Mixed Effects Model of Evolution. *P*-values and *q*-values (controlling for false discovery) are given for each site.

Gene	Gene set	Site	*P*-Value $dN > dS$	*q*-Value $dN > dS$
GI11629	Male	58*	0.09	1
		156	0.07	1
GI16692	Normally regulated	58	0.09	1
		118	0.02	1
		181	0.09	1
GI17858	Male	23*	0.02	1
		186	0.05	1
		253	0.04	1
Dmoj/Obp93a	Misregulated	52	0.07	1
Dmoj/Acp2	Male	112	0.02	1

*Fixed differences between *Drosophila mojavensis* and *Drosophila arizonae*.

remained significant after controlling the false discovery rate. Although this correction may be overly conservative (Murrell *et al.*, 2012), we take the cautious approach of assuming a lack of evidence for selection on any of these sites.

Predicted functional residues (e.g. active site residues and substrate specificity sites) were available for three genes with conserved protein domains (*GI10632*, *GI11629* and *GI17858*). We therefore investigated whether there were differences between *D. mojavensis* and *D. arizonae* in any of these sites. Although we found no differences for *GI10632* and *GI17858*, *GI11629* differed in and around active site residues and substrate specificity sites. Specifically, alignments with other cysteine peptidases in the CDD (Marchler-Bauer *et al.*, 2013) identified four predicted active site residues and six residues involved in substrate specificity. Two active site residues are clustered near one another, and one of these, site 58, is different in four of the five populations sampled, including a radical amino acid difference between *D. arizonae* and all the *D. mojavensis* populations (Fig. 4). Another active site residue (site 201) and adjacent residues exhibited variation between some populations, although there were no fixed differences between species as *D. arizonae* shared the same amino acids as *D. mojavensis baja* (Fig. 4). Two substrate specificity residues were also clustered in this region, one of which was variable between populations. Again, *D. arizonae* shared the same amino acids at this position as *D. mojavensis baja* (Fig. 4). A region surrounding another cluster of two substrate specificity sites also exhibited considerable variability among populations; again *D. arizonae* and *D. mojavensis baja* shared amino acids at these positions. The same was also true for a region surrounding one of the additional substrate specificity residues (Fig. 4).

In summary, positive selection was rigorously supported on branches or sites in eight genes (six male, one misregulated and one normally regulated). The pattern of selection on branches or sites specifically indicated divergent evolution between *D. mojavensis* and *D. arizonae* for four of these genes, all belonging to the male set (Table 4). Evidence for selection is particularly strong for two of the four genes (*GI11629* and *GI20219*), which show evidence from multiple tests. *GI23009* also shows evidence from more than one test, but one of these was not significant after controlling for multiple testing.

Discussion

Functional annotation of genes identified in Bono *et al.* (2011) revealed some interesting patterns of gene expression that help shed light on the unusual female post-mating physiology of *D. mojavensis* and *D. arizonae*. At least four genes, including two female regulated genes (*GI23324* and *GI10632*) and two whose transcripts are transferred by males (*GI17858* and *GI23381*), are involved in lipid recognition/metabolism. One of the female genes has been implicated specifically in haemolymph coagulation (*GI23324*) and was misregulated in heterospecifically mated females (Bono *et al.*, 2011). Although speculative, the potential connection between lipid metabolism and the formation and/or degradation of the insemination reaction mass is intriguing and deserves further investigation. The other most notable finding was that four of the 12 male genes belong to

```
                57                  107        174                196
                |*     *             |  ##      | #               |  # *#
D. arizonae     HL.Q.A......T       Y..TV      T..FNG...Q        K.T.YHSL
D. moj. baja    HT.Q.A......A       Y..TV      T..FNG...Q        K.T.YHSL
D. moj. moj.    YT.Q.A......A  //   S..RA  //  I..ITA...Q  //    N.N.YLTM
D. moj. wrig.   YA.K.R......A       S..HV      I..ITT...Q        K.N.YLSM
D. moj. son.    YS.K.R......A       S..HV      I..ITA...H        K.N.FLSM
```

Fig. 4 Amino acid alignment for regions of *GI11629* with predicted functional residues (*, active site residues; #, substrate specificity residues).

Table 4 Summary of strongest candidates for involvement in post-mating-prezygotic incompatibilities between *Drosophila mojavensis* and *Drosophila arizonae*. Strongest candidates had evidence of divergent selection on branches separating the species or on sites that had fixed differences between the species. Positive tests are denoted by an '×'.

Gene	Gene set	GA-branch	BS-REL	FUBAR	MEME
GI11629	Male	×	×	×	×*
GI19546	Male			×	
GI20219	Male	×		×	
GI23009	Male		×*	×	

BS-REL, Branch-site Random Effects Likelihood; FUBAR, Fast Unconstrained Bayesian Approximation; GA-Branch, Genetic Algorithm Branch; MEME, Mixed Effects Model of Evolution.
*Not significant after correction for multiple comparisons.

the eukaryotic SCP-like gene family. These proteins are members of the cysteine-rich secretory protein (CRISP) superfamily and are linked to a variety of physiological processes in eukaryotes including male fertility, fertilization, immunity and venom production (Hoffman, 1993; Niderman *et al.*, 1995; Haynes *et al.*, 1997; Szyperski *et al.*, 1998; Olson *et al.*, 2001). In *D. melanogaster*, at least 26 SCP-like genes have been identified, with the majority showing biased expression in male reproductive tissues (Kovalick & Griffin, 2005), and at least two functioning in the SP network (Findlay *et al.*, 2014). The transfer of several SCP-like transcripts from males to females suggests that genes from this family may play a particularly important role in mediating reproductive tract interactions between the sexes.

Collective analyses of female mating genes that were misregulated in heterospecifically mated females suggest that nonsynonymous divergence and pairwise d*N*/d*S* ratios are not elevated relative to normally regulated mating genes or female reproductive tract genes in general. Although this conclusion must be tempered somewhat by a lack of statistical power due to small sample sizes, our screen for selection on individual female genes also found little evidence of positive selection. Nevertheless, the fact that these genes are misexpressed in heterospecific matings suggests a possible role, either directly or indirectly, in the observed reproductive incompatibilities between *D. mojavensis* or *D. arizonae*. If these genes are directly involved in incompatible interactions, our results suggest that this may be driven by divergence in regulatory regions, or from divergence in male proteins with which they interact. Alternatively, misexpression in heterospecific crosses may not represent direct involvement in incompatible interactions, but rather be an indirect consequence of incompatible interactions between other proteins acting upstream in a sequence.

Male transferred transcripts were more divergent and appear to evolve under different selection pressures than testis-specific genes, and these metrics were comparable to estimates for the set of previously identified *Acps*. The fact that *Acps* are already known to be among the most rapidly evolving genes in many organisms (Swanson & Vacquier, 2002) suggests that the male transferred genes are notable in this regard, even if they do not stand out when compared to *Acps*. The relatively rapid rate of evolution in the male transferred set also suggests that individual genes within the transferred set could be under divergent selection in *D. mojavensis* and *D. arizonae*, making them interesting candidates for involvement in reproductive incompatibilities. To better address this question, we used comparative molecular evolutionary analyses to investigate patterns of selection on individual male genes. These analyses identified four genes with branches or sites under divergent selection between *D. mojavensis* and *D. arizonae* (Table 4).

It is not clear whether proteins coded by these male genes are also transferred to females, and/or whether transferred transcripts are translated within the female, but the fact that positive selection has occurred on mutations that could affect protein structure suggests that at least one of these scenarios is likely. If transferred RNA instead primarily serves a signalling role, we would not necessarily expect selection on protein structure as reflected by d*N*/d*S* ratios. The possibility that male transcripts could be translated within the female is bolstered by the recent finding in *D. melanogaster* that males transfer exosomes to females during mating, which can then fuse with female epithelial cells (Corrigan *et al.*, 2014). In other organisms, exosomes have been shown to carry mRNA transcripts that can be subsequently translated in recipient cells (Valadi *et al.*, 2007). Such coordination between males and females has been observed at the post-translational level in *D. melanogaster*, where at least one male seminal fluid protein requires female factors for full activity (LaFlamme *et al.*, 2012, 2014). Although the possibility

that transcripts are translated within females is at least plausible in the light of these findings, future studies aimed at identifying the mechanism of transfer and the ultimate fate of the transcripts are clearly necessary.

Of the four best candidates, the strongest evidence of positive selection was found for *GI11629*. In fact, this gene had the highest pairwise d*N*/d*S* value of any male or female reproductive tract gene in any of the data sets we examined. Furthermore, tests of both pervasive and episodic positive selection on this gene were consistent with divergent selection between the species, and differences between *D. mojavensis* and *D. arizonae* included changes in or around predicted functional residues. It is unclear what affect these differences have on enzyme activity, but based on the typical catalytic mechanism of C1A cysteine peptidases, it could be substantial. The first step in catalysis involves deprotonation of a thiol in the active site, usually carried out by the adjacent basic amino acid histidine (Storer & Ménard, 1994). In *D. mojavensis,* this residue is changed to leucine, which is neutral and nonpolar. This, coupled with a diversity of radically different amino acids at active site residue 58 in both species, suggests that functional differences between the species may be significant. An additional possibility is that active site mutations actually abolish the catalytic capability of the enzyme altogether. Studies of *D. melanogaster* seminal fluid have revealed that a number of noncatalytic protease homologues, which have mutations in essential active site residues, are under positive selection (LaFlamme & Wolfner, 2013). LaFlamme & Wolfner (2013) hypothesized that these proteins may bind substrates of other proteases without cleaving them, thereby serving as competitive inhibitors, or serve other regulatory functions.

Given the history of positive selection on *GI11629*, it is perhaps surprising that the southern line of *D. arizonae* appears to have a nonfunctional copy of this gene. Further sampling of southern populations is necessary to determine whether this mutation is fixed in southern populations, and what effect, if any, this has on reproductive success. Furthermore, more extensive sampling of all *D. mojavensis* and *D. arizonae* populations will also shed light on other aspects of the evolutionary history of this gene. In particular, the unexpected sequence similarity between *D. arizonae* and *D. mojavensis baja* warrants further investigation. One potential explanation for this pattern is recent introgression, which could also explain why GARD analysis detected recombination between alleles of this gene. Although past studies have failed to detect introgression between *D. mojavensis* and *D. arizonae*, they either excluded *D. mojavensis baja* from the analysis (Counterman & Noor, 2006) or examined putatively neutral genomic regions (Machado *et al.*, 2007). The likelihood of hybridization between *D. mojavensis* and *D. arizonae* may be higher on the Baja California peninsula than in other areas of sympatry (e.g. portions of mainland Mexico), as *D. arizonae* is assumed to have arrived on the peninsula only recently and premating isolation between *D. mojavensis baja* and *D. arizonae* is considerably lower than crosses from mainland Mexico (Wasserman & Koepfer, 1977; Markow, 1981; Massie & Markow, 2005). In addition, recent studies have demonstrated that selectively advantageous alleles are especially likely to move between species (Martin *et al.*, 2006; Pardo-Diaz *et al.*, 2012; Staubach *et al.*, 2012). The history of positive selection on *GI11629* suggests that this might be plausible here. However, if this gene is involved in reproductive incompatibilities between *D. mojavensis* and *D. arizonae*, alleles from one species would be expected to have deleterious effects in the other species, reducing the likelihood of introgression. Thus, if the sequence similarity is explained by introgression, it seems unlikely that this gene is directly involved in reproductive incompatibilities. An alternative possibility is that the pattern of similarity between *D. mojavensis baja* and *D. arizonae* results from incomplete lineage sorting of ancestral polymorphism, which is difficult to distinguish from recombination. Clearly, future work using more extensive population sampling is needed to clarify the complex evolutionary history of this gene.

The remaining three best candidate genes were all members of the SCP gene superfamily. This highlights the potential importance of this gene family in mediating molecular interactions between the sexes, at least in the *D. mojavensis*/*D. arizonae* study system. One of these genes, *GI20219*, is orthologous to *antares* in *D. melanogaster*, which was recently identified as a crucial component of the SP network (Findlay *et al.*, 2014). Knock-down of *antares* in *D. melanogaster* males prevents SP from binding sperm, thereby blocking the release of sperm from storage in mated females. Females mated to knocked down males have near normal fertility initially, but oviposition declines sharply by the second day after mating (Findlay *et al.*, 2014). Interestingly, the low fertility of crosses between *D. mojavensis* females and *D. arizonae* males appears to be partially attributable to problems with storage and/or release of sperm from storage (Kelleher & Markow, 2007). Given this, the fact that multiple tests suggest divergent selection on this gene between *D. mojavensis*/*D. arizonae*, and that *D. mojavensis* and *D. arizonae* males transfer transcripts to females in different quantities is especially intriguing. Whether the function of this gene as part of the SP network is conserved between *D. melanogaster* and *D. mojavensis*/*D. arizonae* is unclear, however, because a SP ortholog has not been identified in *D. mojavensis* (Findlay *et al.*, 2014).

Although we know little about the function of the remaining SCP candidate genes, the analysis of conserved protein domains in *GI23009* is particularly interesting. In addition to a conserved SCP domain, another portion of the gene was similar to the herpesvirus latent membrane protein 1 domain. The LMP1 protein, mostly

known from Epstein–Barr virus, is an oncogene essential for transformation of host cells and is also capable of regulating its own expression and the expression of host genes (Kaye *et al.*, 1993; Lee & Sugden, 2008a,b; Lee *et al.*, 2009). Notably, *D. mojavensis*, *D. arizonae* and *D. navojoa* differ considerably in this portion of the gene. *Drosophila navojoa* lacks 91 amino acid residues compared with *D. mojavensis*, which corresponds to almost the entire region that shows homology to the LMP1 domain. Likewise, this region is only partially intact in *D. arizonae*, with the northern line lacking 45 residues and the southern line lacking 30 residues compared to *D. mojavensis*. At this point, we can only speculate about the potential role this gene might play in reproduction, but its rapid evolution, combined with functional information suggesting possible transformative and self-regulatory properties, makes this an interesting candidate for future functional studies.

In summary, we found little evidence for elevated levels of divergence or positive selection on coding sequences of female genes that were misregulated in heterospecifically mated females. In contrast, whereas as a group, male transferred genes did not stand out as more divergent or under different selection pressures compared to other *Acps*, four individual genes in this set show patterns of divergent evolution between *D. mojavensis* and *D. arizonae*. Synthesizing results here with information on predicted function and with the results of the previous transcriptomic study (Bono *et al.*, 2011) showing differences in the abundance of these transcripts in the reproductive tracts of con- vs. heterospecifically mated females, these genes are candidates for involvement in the observed PMPZ reproductive incompatibilities between the species. Ultimately, functional studies using RNA interference or genome editing tools are required to verify the role of these genes in reproduction and to link them more directly to PMPZ incompatibilities.

Acknowledgments

The authors would like to thank Erin S. Kelleher for providing alignments of *D. arizonae* female reproductive tract sequences. This work was partly supported by a grant from the National Science Foundation (DEB-1219387 to LMM).

References

Arnqvist, G. & Rowe, L. 2005. *Sexual Conflict*. Princeton University Press, Princeton.

Birkhead, T.R. & Pizzari, T. 2002. Postcopulatory sexual selection. *Nat. Rev. Genet.* **3**: 262–273.

Bono, J.M. & Markow, T.A. 2009. Post-zygotic isolation in cactophilic *Drosophila*: larval viability and adult life-history traits of *D. mojavensis/D. arizonae* hybrids. *J. Evol. Biol.* **22**: 1387–1395.

Bono, J.M., Matzkin, L.M., Kelleher, E.S. & Markow, T.A. 2011. Postmating transcriptional changes in reproductive tracts of con- and heterospecifically mated *Drosophila mojavensis* females. *Proc. Natl. Acad. Sci.* **108**: 7878–7883.

Chen, P.S., Stumm-Zollinger, E., Aigaki, T., Balmer, J., Bienz, M. & Böhlen, P. 1988. A male accessory gland peptide that regulates reproductive behavior of female *Drosophila melanogaster*. *Cell* **54**: 291–298.

Corrigan, L., Redhai, S., Leiblich, A., Fan, S.J., Perera, S.M., Patel, R. *et al.* 2014. BMP-regulated exosomes from *Drosophila* male reproductive glands reprogram female behavior. *J. Cell Biol.* **206**: 671–688.

Counterman, B.A. & Noor, M.A.F. 2006. Multilocus test for introgression between the cactophilic species *Drosophila mojavensis* and *Drosophila arizonae*. *Am. Nat.* **168**: 682–696.

Coyne, J.A. & Orr, H.A. 2004. *Speciation*. Sinauer Associates, Inc., Sunderland.

Delport, W., Poon, A.F.Y., Frost, S.D.W. & Kosakovsky Pond, S.L. 2010. Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**: 2455–2457.

Dinger, M.E., Mercer, T.R. & Mattick, J.S. 2008. RNAs as extracellular signaling molecules. *J. Mol. Endocrinol.* **40**: 151–159.

Dobzhansky, T. 1951. *Genetics and the Origin of Species*, 3rd edn. Columbia University Press, New York.

Drosophila 12 Genomes Consortium. 2007. Evolution of genes and genomes on the *Drosophila* phylogeny. *Nature* **450**: 203–218.

Findlay, G.D., Sitnik, J.L., Wang, W., Aquadro, C.F., Clark, N.L. & Wolfner, M.F. 2014. Evolutionary rate covariation identifies new members of a protein network required for *Drosophila melanogaster* female post-mating responses. *PLoS Genet.* **10**: e1004108.

Haynes, S.R., Cooper, M.T., Pype, S. & Stolow, D.T. 1997. Involvement of a tissue-specific RNA recognition motif protein in *Drosophila* spermatogenesis. *Mol. Cell. Biol.* **17**: 2708–2715.

Hoffman, D.R. 1993. Allergens in Hymenoptera venom. XXV: the amino acid sequences of antigen 5 molecules and the structural basis of antigenic cross-reactivity. *J. Allergy Clin. Immunol.* **92**: 707–716.

Howard, D.J., Palumbi, S.R., Birge, L.M. & Manier, M.K. 2009. Sperm and speciation. In: *Sperm Biology: An Evolutionary Perspective* (T.R. Birkhead, D.J. Hosken, S. Pitnick, eds), pp. 367–395. Academic Press, Burlington, MA.

Kaye, K.M., Izumi, K.M. & Kieff, E. 1993. Epstein–Barr virus latent membrane protein 1 is essential for B-lymphocyte growth transformation. *Proc. Natl. Acad. Sci.* **90**: 9150–9154.

Kearse, M., Moir, R., Wilson, A., Stones-Havas, S., Cheung, M., Sturrock, S. *et al.* 2012. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**: 1647–1649.

Kelleher, E.S. & Markow, T.A. 2007. Reproductive tract interactions contribute to isolation in *Drosophila*. *Fly* **1**: 33–37.

Kelleher, E.S. & Markow, T.A. 2009. Duplication, selection and gene conversion in a *Drosophila mojavensis* female reproductive protein family. *Genetics* **181**: 1451–1465.

Kelleher, E.S., Swanson, W.J. & Markow, T.A. 2007. Gene duplication and adaptive evolution of digestive proteases in

Drosophila arizonae female reproductive tracts. *PLoS Genet.* **3**: e148.

Kishino, H. & Hasegawa, M. 1989. Evaluation of the maximum likelihood estimate of the evolutionary tree topologies from DNA sequence data, and the branching order in hominoidea. *J. Mol. Evol.* **29**: 170–179.

Knowles, L.L. & Markow, T.A. 2001. Sexually antagonistic coevolution of a postmating-prezygotic reproductive character in desert *Drosophila*. *Proc. Natl. Acad. Sci.* **98**: 8692–8696.

Kosakovsky Pond, S.L. & Frost, S.D.W. 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol. Biol. Evol.* **22**: 478–485.

Kosakovsky Pond, S.L., Posada, D., Gravenor, M.B., Woelk, C.H. & Frost, S.D.W. 2006. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol. Biol. Evol.* **23**: 1891–1901.

Kosakovsky Pond, S.L., Murrell, B., Fourment, M., Frost, S.D.W., Delport, W. & Scheffler, K. 2011. A random effects branch-site model for detecting episodic diversifying selection. *Mol. Biol. Evol.* **28**: 3033–3043.

Kovalick, G.E. & Griffin, D.L. 2005. Characterization of the SCP/TAPS gene family in *Drosophila melanogaster*. *Insect Biochem. Mol. Biol.* **35**: 825–835.

Kryazhimskiy, S. & Plotkin, J.B. 2008. The population genetics of d*N*/d*S*. *PLoS Genet.* **4**: e1000304.

LaFlamme, B.A. & Wolfner, M.F. 2013. Identification and function of proteolysis regulators in seminal fluid. *Mol. Reprod. Dev.* **80**: 80–101.

LaFlamme, B.A., Ram, K.R. & Wolfner, M.F. 2012. The *Drosophila melanogaster* seminal fluid protease 'seminase' regulates proteolytic and post-mating reproductive processes. *PLoS Genet.* **8**: e1002435.

LaFlamme, B.A., Avila, F.W., Michalski, K. & Wolfner, M.F. 2014. A *Drosophila* protease cascade member, seminal metalloprotease-1, is activated stepwise by male factors and requires female factors for full activity. *Genetics* **196**: 1117–1129.

Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, R., McGettigan, P.A., McWilliam, H. *et al.* 2007. Clustal W and Clustal X version 2.0. *Bioinformatics* **23**: 2947–2948.

Lee, D.Y. & Sugden, B. 2008a. The latent membrane protein 1 oncogene modifies B-cell physiology by regulating autophagy. *Oncogene* **27**: 2833–2842.

Lee, D.Y. & Sugden, B. 2008b. The LMP1 oncogene of EBV activates PERK and the unfolded protein response to drive its own synthesis. *Blood* **111**: 2280–2289.

Lee, D.Y., Lee, J. & Sugden, B. 2009. The unfolded protein response and autophagy: herpesviruses rule!. *J. Virol.* **83**: 1168–1172.

Machado, C.A., Matzkin, L.M., Reed, L.K. & Markow, T.A. 2007. Multilocus nuclear sequences reveal intra- and interspecific relationships among chromosomally polymorphic species of cactophilic *Drosophila*. *Mol. Ecol.* **16**: 3009–3024.

Marchler-Bauer, A., Zheng, C., Chitsaz, F., Derbyshire, M.K., Geer, L.Y., Geer, R.C. *et al.* 2013. CDD: conserved domains and protein three-dimensional structure. *Nucleic Acids Res.* **41**: D348–D352.

Markow, T.A. 1981. Courtship behavior and control of reproductive isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. *Evolution* **35**: 1022–1026.

Markow, T.A. & Hocutt, G.D. 1998. Speciation in Sonoran Desert *Drosophila*: testing the limits of the rules. In: *Endless Forms: Species and Speciation* (D. Howard, S. Berlocher, eds), Oxford University Press, Oxford.

Markow, T.A., Fogleman, J.C. & Heed, W.B. 1983. Reproductive isolation in Sonoran Desert *Drosophila*. *Evolution* **37**: 649–652.

Markow, T.A., Reed, L.K. & Kelleher, E.S. 2007. Sperm fate and function in reproductive isolation in *Drosophila*. *Soc. Reprod. Fertil. Suppl.* **65**: 155–173.

Martin, N.H., Bouck, A.C. & Arnold, M.L. 2006. Detecting adaptive trait introgression between *Iris fulva* and *I. brevicaulis* in highly selective field conditions. *Genetics* **172**: 2481–2489.

Marygold, S.J., Leyland, P.C., Seal, R.L., Goodman, J.L., Thurmond, J., Strelets, V.B. *et al.* 2013. FlyBase: improvements to the bibliography. *Nucleic Acids Res.* **41**: D751–D757. http://arizona.openrepository.com/arizona/bitstream/10150/193326/1/azu_etd_1629_sip1_m.pdf

Massie, K.R. 2006. *Sexual Isolation Between Drosophila mojavensis* and *Drosophila arizonae*. The University of Arizona. http://arizona.openrepository.com/arizona/bitstream/10150/193326/1/azu_etd_1629_sip1_m.pdf

Massie, K.R. & Markow, T.A. 2005. Sympatry, allopatry and sexual isolation between *Drosophila mojavensis* and *D. arizonae*. *Hereditas* **142**: 51–55.

Matzkin, L.M. 2004. Population genetics and geographic variation of alcohol dehydrogenase (Adh) paralogs and glucose-6-phosphate dehydrogenase (G6pd) in *Drosophila mojavensis*. *Mol. Biol. Evol.* **21**: 276–285.

Matzkin, L.M. & Eanes, W.F. 2003. Sequence variation of alcohol dehydrogenase (Adh) paralogs in cactophilic *Drosophila*. *Genetics* **163**: 181–194.

Mugal, C.F., Wolf, J.B.W. & Kaj, I. 2014. Why time matters: codon evolution and the temporal dynamics of d*N*/d*S*. *Mol. Biol. Evol.* **31**: 212–231.

Murrell, B., Wertheim, J.O., Moola, S., Weighill, T., Scheffler, K. & Kosakovsky Pond, S.L. 2012. Detecting individual sites subject to episodic diversifying selection. *PLoS Genet.* **8**: e1002764.

Murrell, B., Moola, S., Mabona, A., Weighill, T., Sheward, D., Kosakovsky Pond, S.L. *et al.* 2013. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for inferring selection. *Mol. Biol. Evol.* **30**: 1196–1205.

Niderman, T., Genetet, I., Bruyere, T., Gees, R., Stintzi, A., Legrand, M. *et al.* 1995. Pathogenesis-related PR-1 proteins are antifungal (isolation and characterization of three 14-kilodalton proteins of tomato and of a basic PR-1 of tobacco with inhibitory activity against *Phytophthora infestans*). *Plant Physiol.* **108**: 17–27.

Olson, J.H., Xiang, X., Ziegert, T., Kittelson, A., Rawls, A., Bieber, A.L. *et al.* 2001. Allurin, a 21-kDa sperm chemoattractant from *Xenopus* egg jelly, is related to mammalian sperm-binding proteins. *Proc. Natl. Acad. Sci.* **98**: 11205–11210.

Pardo-Diaz, C., Salazar, C., Baxter, S.W., Merot, C., Figueiredo-Ready, W., Joron, M. *et al.* 2012. Adaptive introgression across species boundaries in *Heliconius* butterflies. *PLoS Genet.* **8**: e1002752.

Pfeiler, E., Castrezana, S., Reed, L.K. & Markow, T.A. 2009. Genetic, ecological, and morphological differences among populations of cactophilic *Drosophila mojavensis* from southwestern USA and northwestern Mexico, with descriptions of two new subspecies. *J. Nat. Hist.* **43**: 923–938.

Pitnick, S., Wolfner, M.W. & Suarez, S.S. 2009. Ejaculate-female and sperm-female interactions. In: *Sperm Biology: An Evolutionary Perspective* (T.R. Birkhead, D.J. Hosken, eds), pp. 247–304. Academic Press, Burlington, MA.

Prokupek, A.M., Eyun, S.-I., Ko, L., Moriyama, E.N. & Harshman, L.G. 2010. Molecular evolutionary analysis of seminal receptacle sperm storage organ genes of *Drosophila melanogaster*. *J. Evol. Biol.* **23**: 1386–1398.

Ravi Ram, K. & Wolfner, M.F. 2009. A network of interactions among seminal proteins underlies the long-term postmating response in *Drosophila*. *Proc. Natl. Acad. Sci.* **106**: 15384–15389.

Reed, L.K. & Markow, T.A. 2004. Early events in speciation: polymorphism for hybrid male sterility in *Drosophila*. *Proc. Natl. Acad. Sci.* **101**: 9009–9012.

Reed, L.K., Nyboer, M. & Markow, T.A. 2007. Evolutionary relationships of *Drosophila mojavensis* geographic host races and their sister species *Drosophila arizonae*. *Mol. Ecol.* **16**: 1007–1022.

Rocha, E.P.C., Smith, J.M., Hurst, L.D., Holden, M.T.G., Cooper, J.E., Smith, N.H. *et al.* 2006. Comparisons of d*N*/d*S* are time dependent for closely related bacterial genomes. *J. Theor. Biol.* **239**: 226–235.

Staubach, F., Lorenc, A., Messer, P.W., Tang, K., Petrov, D.A. & Tautz, D. 2012. Genome patterns of selection and introgression of haplotypes in natural populations of the house mouse (*Mus musculus*). *PLoS Genet.* **8**: e1002891.

Storer, A.C. & Ménard, R. 1994. Catalytic mechanism in papain family of cysteine peptidases. *Methods Enzymol.* **244**: 486–500.

Swanson, W.J. & Vacquier, V.D. 2002. The rapid evolution of reproductive proteins. *Nat. Rev. Genet.* **3**: 137–144.

Swanson, W.J., Yang, Z., Wolfner, M.F. & Aquadro, C.F. 2001. Positive Darwinian selection drives the evolution of several female reproductive proteins in mammals. *Proc. Natl. Acad. Sci.* **98**: 2509–2514.

Swanson, W.J., Wong, A., Wolfner, M.F. & Aquadro, C.F. 2004. Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* **168**: 1457–1465.

Szyperski, T., Fernández, C., Mumenthaler, C. & Wüthrich, K. 1998. Structure comparison of human glioma pathogenesis-related protein GliPR and the plant pathogenesis-related protein P14a indicates a functional link between the human immune system and a plant defense system. *Proc. Natl. Acad. Sci.* **95**: 2262–2266.

Valadi, H., Ekström, K., Bossios, A., Sjöstrand, M., Lee, J.J. & Lötvall, J.O. 2007. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. *Nat. Cell Biol.* **9**: 654–659.

Wagstaff, B.J. & Begun, D.J. 2005. Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics* **171**: 1083–1101.

Wasserman, M. & Koepfer, H.R. 1977. Character displacement for sexual isolation between *Drosophila mojavensis* and *Drosophila arizonensis*. *Evolution* **31**: 812–823.

Waterhouse, R.M., Tegenfeldt, F., Li, J., Zdobnov, E.M. & Kriventseva, E.V. 2013. OrthoDB: a hierarchical catalog of animal, fungal and bacterial orthologs. *Nucleic Acids Res.* **41**: D358–D365.

Wolfner, M.F. 2007. 'S.P.E.R.M'. (seminal proteins (are) essential reproductive modulators): the view from *Drosophila*. *Soc. Reprod. Fertil. Suppl.* **65**: 183–199.

Yapici, N., Kim, Y.J., Ribeiro, C. & Dickson, B.J. 2008. A receptor that mediates the post-mating switch in *Drosophila* reproductive behaviour. *Nature* **451**: 33–37.

Zhang, Z., Li, J., Zhao, X.Q., Wang, J., Wong, G.K. & Yu, J. 2006. KAKS_CALCULATOR: calculating Ka and Ks through model selection and model averaging. *Genomics Proteomics Bioinformatics* **4**: 259–263.

Supporting information

Additional Supporting Information may be found in the online version of this article:

Table S1 Primers used for PCR amplification.

Table S2 Identity and predicted functional information for genes from the three gene sets analyzed in this study.

Table S3 Results of Kishino–Hasegawa tests for breakpoints identified by GARD.

Figure S1 Boxplot showing median nonsynonymous divergence for (a) female and (b) male gene sets.

Figure S2 Boxplot showing median pairwise d*N*/d*S* for (a) female and (b) male gene sets.

Figure S3 Results from GA-Branch tests for five genes where the best fitting model showed evidence of divergent selection on branches separating *Drosophila mojavensis* and *Drosophila arizonae*, but with model averaged support < 95%.

Received 18 June 2014; revised 11 December 2014; accepted 13 December 2014