

Sorted gene genealogies and species-specific nonsynonymous substitutions point to putative postmating prezygotic isolation genes in *Allonemobius* crickets

Suegene Noh¹ and Jeremy L. Marshall²

¹ Department of Biology, Washington University in St. Louis, St. Louis, MO, United States

² Department of Entomology, Kansas State University, Manhattan, KS, United States

ABSTRACT

In the *Allonemobius socius* complex of crickets, reproductive isolation is primarily accomplished via postmating prezygotic barriers. We tested seven protein-coding genes expressed in the male ejaculate for patterns of evolution consistent with a putative role as postmating prezygotic isolation genes. Our recently diverged species generally lacked sequence variation. As a result, ω -based tests were only mildly successful. Some of our genes showed evidence of elevated ω values on the internal branches of gene trees. In a couple of genes, these internal branches coincided with both species branching events of the species tree, between *A. fasciatus* and the other two species, and between *A. socius* and *A. sp. nov.* Tex. In comparison, more successful approaches were those that took advantage of the varying degrees of lineage sorting and allele sharing among our young species. These approaches were particularly powerful within the contact zone. Among the genes we tested we found genes with genealogies that indicated relatively advanced degrees of lineage sorting across both allopatric and contact zone alleles. Within a contact zone between two members of the species complex, only a subset of genes maintained allelic segregation despite evidence of ongoing gene flow in other genes. The overlap in these analyses was *arginine kinase* (AK) and *apolipoprotein A-1 binding protein* (APBP). These genes represent two of the first examples of sperm maturation, capacitation, and motility proteins with fixed non-synonymous substitutions between species-specific alleles that may lead to postmating prezygotic isolation. Both genes express ejaculate proteins transferred to females during copulation and were previously identified through comparative proteomics. We discuss the potential function of these genes in the context of the specific postmating prezygotic isolation phenotype among our species, namely conspecific sperm precedence and the superior ability of conspecific males to induce oviposition in females.

Subjects Entomology, Evolutionary Studies, Genetics

Keywords Positive selection, Ejaculate proteins, Haplotype networks, Lineage sorting, Postmating prezygotic isolation

INTRODUCTION

Not all genes contribute equally to reproductive isolation during speciation. ‘Speciation’ (*Wu, 2001; Wu & Ting, 2004; Nosil & Schluter, 2011*), ‘isolation’ (*Rieseberg, Church & Morjan, 2004*), or ‘barrier’ (*Noor & Feder, 2006*) genes are expected to show very different

Submitted 11 September 2015

Accepted 14 January 2016

Published 15 February 2016

Corresponding author

Suegene Noh,
suegene.noh@gmail.com

Academic editor

Tanja Schwander

Additional Information and
Declarations can be found on
page 19

DOI 10.7717/peerj.1678

© Copyright
2016 Noh and Marshall

Distributed under
Creative Commons CC-BY 4.0

OPEN ACCESS

patterns of evolution compared to genes that are not directly involved in reproductive isolation when species are still undergoing lineage sorting (Wu, 2001; Feder, Egan & Nosil, 2012). Therefore, we expect to find putative speciation genes among those genes that become fixed for alternative alleles within each incipient species early in the process of divergence, with said alleles rarely crossing the species boundary in sympatry (Ting, Tsaui & Wu, 2000; Dopman et al., 2005).

Rapidly evolving reproductive proteins that can affect fertilization success have an important role in the evolution of postmating prezygotic reproductive isolation. Many reproductive genes are known to evolve rapidly in a variety of organisms (Civetta & Singh, 1998; Swanson & Vacquier, 2002; Clark, Aagaard & Swanson, 2006; Panhuis & Swanson, 2006; Snook et al., 2009). In *Drosophila*, where some of the most extensive work has been done, genes that show male-biased expression evolve faster compared to female-biased and somatically expressed genes (Zhang, Hambuch & Parsch, 2004; Zhang & Parsch, 2005; Metta et al., 2006; Pröschel, Zhang & Parsch, 2006; Haerty et al., 2007). Seminal fluid proteins in particular tend to show an excess of non-synonymous substitutions (Begun et al., 2000; Swanson et al., 2001; Wagstaff & Begun, 2005; Almeida & DeSalle, 2008). Similar patterns have also been observed in mice and primates (Clark & Swanson, 2005; Karn et al., 2008; Ramm et al., 2008; Turner, Chuong & Hoekstra, 2008; Dean et al., 2009). Because they enable us to isolate male reproductive protein-coding genes that can directly interact with their female counterparts, proteomic analyses of insect spermatophores have proved to be particularly effective for narrowing prospects in the search for putative speciation genes (Andrés, Maroja & Harrison, 2008; Marshall et al., 2011; Andrés et al., 2013).

The male ejaculate proteome comprises sperm-expressed proteins and seminal fluid proteins. Sperm not only contribute half of the diploid genome, but are also involved in sperm-egg interactions, including egg activation and the delivery of paternal factors during fertilization (Dorus et al., 2006). The majority of seminal fluid proteins are produced by male accessory glands. These proteins contain conserved functional classes of peptides and pro-hormones that are involved in sperm binding, proteolysis, lipid metabolism, and immune function (Mueller et al., 2004; Chapman & Davies, 2004; Poiani, 2006; Avila et al., 2011). Once transferred into the female reproductive tract, these proteins can initiate a wide-range of physiological functions, including increased egg production and oviposition, decreased receptivity, decreased lifespan, and increased feeding in females (reviewed in Avila et al., 2011). The interacting female counterparts to these ejaculate proteins (EPs) are not as well known (Ram, Ji & Wolfner, 2005; Ram & Wolfner, 2007; Snook et al., 2009) though genomic data is proving to be invaluable for identifying candidates (Baer et al., 2009a; Findlay et al., 2014). The evolution of EPs has been hypothesized to be driven by various processes, including female sperm preference, sperm competition, and sexual conflict (Mueller et al., 2004; Snook et al., 2009).

The *Allonemobius socius* complex of ground crickets, *A. socius*, *A. fasciatus*, and *A. sp. nov.* Tex, represents a powerful system to explore the hypothesized link between EP divergence and reproductive isolation. Members of this complex are primarily isolated from one another by two postmating, prezygotic phenotypes—conspecific sperm precedence (Gregory & Howard, 1994; Howard et al., 1998a; Howard et al., 1998b; Marshall, 2004)

and the superior ability of conspecific males to induce females to lay eggs ([Gregory & Howard, 1993](#); [Howard et al., 1998b](#)). Two other compelling features of this study system are species boundaries that remain intact in sympatry despite some gene flow ([Howard, 1986](#); [Howard & Waring, 1991](#); [Traylor et al., 2008](#)) and the very recent nature of divergence between species (i.e., within the last 30,000 years; [Marshall, 2004](#); [Marshall, 2007](#)). Indeed, divergence is so recent that few species-specific alleles have been identified. Only 2 of 17 allozyme markers ([Howard, 1983](#)), 2 of 5,400 AFLP markers ([Howard et al., 2002](#)), ~21 of 1,660 thorax proteins and ~33 of 922 ejaculate proteins ([Marshall et al., 2011](#)), and 1 of 16 randomly chosen reproductive genes spanning >7,500 bp of coding sequence (J Marshall, 2016, unpublished data) yield evidence of species specificity. Taken together, the above data suggest that while there is sufficient genetic divergence to produce reproductive isolation and maintain species boundaries in sympatry, the vast majority of genes show little evidence of divergence. In all, the *A. socius* complex represents a system whereby speciation is ongoing with potentially relatively few genes contributing to postmating prezygotic reproductive isolation between species. Therefore, if we can identify those ejaculate and female reproductive tract genes that exhibit signatures of positive selection and maintain species-specificity in sympatry, we will gain insight into the postmating prezygotic isolation genes that are ultimately driving speciation in this system.

In this study, we examine EPs in the *A. socius* complex to identify genes that show patterns of sequence evolution and lineage sorting that are consistent with a potential contribution to postmating prezygotic isolation between species. We examined seven EP coding genes, five of which were identified in a comparative proteome study between *A. socius* and *A. fasciatus* ([Marshall et al., 2011](#)), and two additional EP coding genes identified from unpublished EST libraries of *A. socius* accessory glands and testes ([Marshall et al., 2011](#)). The five EP coding genes from the comparative proteome study were previously tested for evidence of positive selection with limited population sampling and sequence fragments. When testing for evidence of selection in recently diverged species, biases can arise when a supposed fixation is in fact a polymorphism, when a true fixation is an ancestral polymorphism rather than a new mutation, and because neutral and adaptive mutations fix at different rates ([Keightley & Eyre-Walker, 2012](#)). We expanded population sampling across the species ranges for the two species that were tested previously, added the third species *A. sp. nov.* Tex, and examined longer sequence lengths to combat the biases that result from insufficient information. A major aim was to test whether our previous conclusions hold up to expanded sampling, at the population and species level and at the sequence level. Finally, we compared how these genes behave by looking at lineage sorting and allele sharing within and across the contact zone of *A. fasciatus* and *A. socius*. These combined analyses point toward an important role for some but not all examined EPs during the evolution of reproductive isolation within the *A. socius* complex of crickets.

METHODS

Background

Striped ground crickets of the *A. socius* complex inhabit moist grasslands across North America and do not show significant habitat isolation (Howard, 1986). The three species *A. socius*, *A. fasciatus*, and *A. sp. nov. Tex* form two contact zones, one between *A. fasciatus* (north) and *A. socius* (south) from Illinois to New Jersey (Howard & Waring, 1991), and one between *A. sp. nov. Tex* (west) and *A. socius* (east) near the Louisiana—Texas state line (Traylor et al., 2008). *A. fasciatus* and *A. socius* seem to have diverged from a common ancestor approximately 30,000 years ago, and *A. sp. nov. Tex* seems to have subsequently diverged from *A. socius* approximately 24,000 years ago (Marshall, 2004; Marshall, 2007). They have previously been shown to be isolated primarily via postmating prezygotic reproductive isolation (Howard et al., 2002; Marshall, 2004; Marshall & DiRienzo, 2012).

Population and gene sampling

Crickets were collected from each population in the summer of 2010 and genotyped in the lab via allozymes (Isocytate dehydrogenase and Hexokinase) to determine species identity (Howard, 1983; Marshall, 2004; Traylor et al., 2008). Sampling localities spanned the range of each species. The seven *A. socius* populations were sampled near Texarkana, AR (AR), Bottom, NC (Bot), Mt. Vernon, IL (IL), Pleasantville, NJ (Mi), Ruston, LA (LA), Gastonia, NC (NC), and Ardmore, OK (OK). The three *A. fasciatus* populations were sampled near Akron, OH (Akn), Frankfort, IL (FF), and New Paltz, NY (NP). The three *A. sp. nov. Tex* populations were sampled near Terrell, TX (Tx20), Royse City, TX (Tx30), and Gainesville, TX (Tx35). Contact zone populations of *A. fasciatus* and *A. socius* were sampled from two habitats at a single location in Kenna, WV. *A. fasciatus* was collected from a hillside habitat, which we call Kenna Hill (KH), and *A. socius* was collected along the base of hill near a creek which we call Kenna Creek (KC). We did not have samples from the contact zone between *A. socius* and *A. sp. nov. Tex*. General maintenance protocols followed Marshall et al. (2009). Briefly, juveniles were reared to maturity in population and sex-specific plastic cages. All crickets were maintained at 27 °C and 14:10 h light to dark photoperiod.

We dissected male accessory glands and testes from three individuals per allopatric population and 9 individuals per contact zone population. cDNA was synthesized from each tissue using RNA isolated via an Ambion RNAqueous-4PCR (#AM1914) kit and standard protocols for 1st strand cDNA synthesis. General PCR and sequencing procedures followed Marshall et al. (2011). Our reagent amounts for a 25 µL reactions were: 2.5 10× buffer B, 2.0 MgCl₂ (25 mM), 0.5 dNTP (10 mM), 0.5 for each primer (10 µM), 1 U Taq (Fisher), 0.5 cDNA, 18.3 ddH₂O. We used the following PCR program: 94 °C for 2:00 min, then 30 cycles of 94 °C for 30 s, annealing at 45–55 °C for 30 s, 72 °C for 1:00 min, and final elongation at 72 °C for 7:00 min. Specific annealing temperatures depended on individual primer melting temperatures (primers used are shown in Table S1). Sequencing was done at the Kansas State University Department of Plant Pathology DNA Sequencing and Genotyping Facility using Applied Biosystems BigDye™ chemistry on an Applied Biosystems 3730 DNA Analyzer.

We compared nucleotide sequences of seven EP coding genes: (1) *acg69* (ACG69), a novel protein of unknown function expressed in accessory glands; (2) *arginine kinase* (AK), a phosphagen kinase that catalyzes ATP-regeneration and energy transport in invertebrates and some protozoa (Ellington, 2001; Noguchi, Sawada & Akazawa, 2001; Uda et al., 2006); (3) *apolipoprotein A-1 binding protein* (APBP), a phosphoprotein expressed in sperm that is homologous to a mammalian sperm capacitation gene (Jha et al., 2008); (4) *ejaculate serine protease* (EJAC-SP), an abundant accessory gland-expressed serine protease previously shown to be involved in the induction of egg laying in successfully mated females (Marshall et al., 2009); (5) *serine protease inhibitor* (SPI), a testis-expressed serine-type endopeptidase inhibitor; (6) *aspartate aminotransferase* (GOT), a pyridoxal-phosphate-dependent aminotransferase expressed in the testis that is also an allozyme historically used to diagnose species identity among *A. socius* complex crickets (Howard, 1983; Howard, 1986); (7) *sperm-associated antigen 6* (SPAG6), homologous to a mammalian protein important for sperm flagellar motility and the structural integrity of the central apparatus (Neilson et al., 1999; Sapiro et al., 2002). The first five genes were investigated to a lesser extent in a comparison of male ejaculate proteome profiles in *A. fasciatus* and *A. socius* (Marshall et al., 2011). On 2D-DIGE (differential in-gel electrophoresis) gels ACG69, EJAC-SP and SPI had non species-specific protein spots that indicate similar molecular weights, isoelectric points, and expression levels in the male ejaculate, while AK and APBP had species-specific spots that indicate differences in one or more of these proteomic traits. The latter two genes were EP coding genes present in EST libraries of *A. socius* accessory glands and testes (Marshall et al., 2011) that were identified as potential candidates based on a review of sperm biology literature.

Sequences formatted as haplotypes are available from NCBI GenBank PopSets [372477571](#) (ACG69), [372477483](#) (AK), [372477513](#) (APBP), [372477527](#) (EJAC-SP), [372477535](#) (GOT), [372477555](#) (SPAG6), [372477561](#) (SPI).

Species tree-based analyses

We first applied tests of selection to the species tree of the *A. socius* complex (*A. fasciatus*, *A. socius*, *A. sp. nov.* Tex)). The ratio of non-synonymous to synonymous substitution rates ω is widely used to detect signatures of selection acting upon protein coding genes (Yang & Bielawski, 2000; Nielsen, 2001; Nielsen, 2005; Jensen, Wong & Aquadro, 2007). When ω is larger than 1, positive or balancing selection is inferred. When ω is smaller than 1, negative or purifying selection is inferred. We aligned all sequences in BioEdit v.7.0.5.3 (Hall, 1999). We counted within species polymorphisms and between species fixations and calculated π_s , π_a , and $\theta = 4N_e\mu$ across all populations with DnaSP v.5.10.1 (Librado & Rozas, 2009). Unless stated otherwise, the majority of the remaining analyses were carried out in HyPhy v.2.2.1 (Kosakovsky Pond, Frost & Muse, 2005) or its online server Datamonkey (Delpont et al., 2010). We selected a nucleotide substitution model (NucModelCompare.bf) at a model rejection level of 0.0002, the recommended level based on Bonferroni correction of comparing 203 increasingly parameter-rich nucleotide substitution models (Kosakovsky Pond & Frost, 2005a). Next we tested for evidence of recombination using the Genetic Algorithm Recombination Detection (GARD) method

(SingleBreakpointRecomb.bf). Recombination should be ruled out or accounted for because its presence can mislead inferences of selection due to no single phylogenetic tree being able to accurately describe the evolutionary relationship of recombinant sequences (Anisimova, Nielsen & Yang, 2003; Shriner et al., 2003; Kosakovsky Pond et al., 2006). GARD tests for evidence of recombination by comparing fits of a single tree vs. segment-specific trees to alignments. Segments are potential recombinant fragments, as defined by sequence in between possible breakpoints (variable sites).

Next we fit maximum likelihood models to the species tree to estimate ω at each branching node: the node between *A. fasciatus* and the other two species, and the node between *A. socius* and *A. sp. nov.* Tex. Using the codon substitution model MG94 (Muse & Gaut, 1994) while estimating codon frequencies based on combinations of nucleotide frequencies (called 3x4 in HyPhy), we fit and optimized a maximum likelihood model on the multiple sequence alignment for each gene, and jointly estimated both ω and their fine asymptotic normal CI estimates from the same likelihood model. HyPhy derives CI estimates analytically from the Fisher Information Matrices of the log likelihood surface of each model parameter (Kosakovsky Pond & Muse 2005).

Finally, we compared polymorphism within species to divergence between species at each branching node of the species tree. We compared ω for each gene using McDonald–Kreitman tests (McDonald & Kreitman, 1991) in DnaSP, and across all genes using a multilocus HKA test (Hudson, Kreitman & Aguadé, 1987) with the program HKA (Wang & Hey, 1996). The null model of the McDonald–Kreitman test is that the ratio of non-synonymous to synonymous intraspecific polymorphisms should be no different from the ratio of non-synonymous to synonymous fixations between species. The multilocus HKA test uses coalescent simulations to test the null model that all loci within a species will share the same effective population size, and loci between sister species will evolve under the same neutral mutation rate. The HKA program also reports an outlier observation that deviates the most from its expected value (Wang & Hey, 1996). The outlier corresponds to a specific locus and its level of polymorphism within species or fixation between species.

Gene tree-based analyses

Gene trees show the evolutionary histories of individual genes and are less likely to concur with species trees the shorter the time since divergence due to incomplete lineage sorting and introgression (Pamilo & Nei, 1988; Maddison, 1997). Because genes more directly related to reproductive isolation and speciation should show patterns of evolution more closely resembling the species tree (Wu, 2001; Feder, Egan & Nosil, 2012), comparing gene trees should inform us of, or confirm, which genes are more likely to be the key genes involved in postmating prezygotic isolation, particularly when evidence suggests the action of selection upon internal nodes that separate incipient species.

Using Neighbor-Joining gene trees built using Tamura–Nei distances (Tamura & Nei, 1993) in HyPhy, we tested for evidence of selection on specific internal branches of the gene tree that had non-zero branch length (TestBranchdNdS.bf). This method estimates the non-synonymous rate while assuming a single synonymous substitution rate across a tree, so in effect tests for evidence of elevated ω on specific branches. We used the MG94 and

F81 (*Felsenstein, 1981*) substitution models and fit a single non-synonymous substitution rate across the tree with no within branch variation in κ_s or κ_a . We compared this model to a model where the internal nodes of interest (those whose non-synonymous substitution rate's 95% CI did not overlap with zero based on a first pass run of this method on all internal branches with non-zero branch lengths) were allowed to vary from the global non-synonymous rate using likelihood ratio tests. We used MEGA6 (*Tamura et al., 2013*) to visualize these trees.

We tested for evidence of specific sites evolving under different selection regimes using the single likelihood ancestor counting (SLAC) method (*Kosakovsky Pond & Frost, 2005b*). SLAC (QuickSelectionDetection.bf) tests all sites of each gene for evidence of selection by first taking a tree and fitting a codon model (MG94 \times F81) to obtain a global estimate of ω . It then reconstructs an ancestral sequence for each site across all internal nodes of the tree using joint maximum likelihood estimation, taking into account the estimate of ω obtained in the previous step. Then for each variable site, the method compares expected and observed numbers of synonymous and non-synonymous substitutions to detect selection.

Finally, we tested for evidence of episodic selection anywhere within each gene using a branch-site method. BUSTED (branch-site unrestricted statistical test for episodic diversification) tests whether there is evidence suggesting any single gene is under positive selection, while accounting for site-level variation in selection and variable selection on a subset of branches on a phylogenetic tree (*Murrell et al., 2015*). We partitioned branches into two categories: those of interest (foreground), which were the same branches from the branch method above (TestBranchdNdS) that had non-zero branch lengths, and the remainder (background). BUSTED optimizes the likelihood using a random effects likelihood framework (*Kosakovsky Pond et al., 2011*) and first fits an unconstrained model that proportions sites in the foreground branches into one of three variable rates including those over $\omega = 1$. This alternative model is then compared to a null model in which all rates are constrained to $\omega = <1$, but all sites are still fit and proportioned into three rates. These models are compared to each other using likelihood ratio tests.

Evidence from the contact zone between *A. fasciatus* and *A. socius*

The genealogical sorting index (gsi) reflects the degree of lineage sorting of individual gene genealogies that occurs during speciation, with values ranging from zero (complete polyphyly) to 1 (complete monophyly) (*Cummings, Neel & Shaw, 2008*). We calculated gsi for each gene using the gsi web service (www.molecularevolution.org) with gene trees including both allopatric and contact zone individuals. We generated gene trees for gsi analysis with Neighbor Joining in TreeBeST v.1.9.2 (*Li, 2006*) using the option ntmm that calculates p-distances from codon alignments. We used this method because gsi requires a rooted tree and TreeBeST can use a species tree to root a gene tree using an algorithm (*Zmasek & Eddy, 2011*) that compares potential root positions on an unrooted tree and places the root where the differences between the gene tree and species tree are minimized (*Li, 2006*). The program permutes the labels (in our case, species identity) of the tips of the given tree multiple times (we used the default $n = 10,000$), each time determining the

gsi value of this new tree. The permuted P -value is the probability of randomly observing a gsi value equal to or better (higher) than the gsi value observed from the data. We used MEGA6 to visualize these trees.

Finally, we constructed statistical parsimony haplotype networks ([Templeton, Crandall & Sing, 1992](#)) of alleles from all three species to test for species-specificity of alleles. We used TCS v.1.21 ([Clement, Posada & Crandall, 2000](#)) to generate the haplotype networks using only allopatric individuals. Species-specific alleles were defined as those found only within each respective species. Common or shared alleles were those observed in more than one species. Once alleles were designated common or specific to a species, we looked at nine individuals each within the contact zone of *A. fasciatus* and *A. socius* and determined what types of allele these contact zone individuals possessed. As noted above, these individuals had previously been designated as fully (homozygous) *A. fasciatus* or *A. socius* based on allozymes. To compare the allelic distributions of these genes, we calculated a segregation metric, the dissimilarity index D ([Duncan & Duncan, 1955](#)) using the R package SEG ([Hong, O'Sullivan & Sadahiro, 2014](#)). Duncan and Duncan's D is a measure of segregation in space that ranges from 0 (complete integration) to 1 (complete segregation). For genes for which data was missing (1 individual for the gene SPAG6), we calculated the potential range of D depending on the potential values for the missing observations.

RESULTS

Species tree-based analyses

We fit the nucleotide substitution model F81 to all genes, and found no evidence of recombination in any of our genes at any of the variable sites. We found relatively low levels of both synonymous and non-synonymous nucleotide variation within and among the *A. socius* complex species. The Watterson estimator θ ranged from 0.001 to 0.011, κ_s ranged from 0.004 to 0.054, and κ_a ranged from 0 to 0.009 ([Tables 1 and 2](#)). Next we estimated $\omega = \kappa_a/\kappa_s$ at each branching event of the species tree. In the older split between *A. fasciatus* and the other two species, the maximum likelihood estimates of ω exceeded 1, indicating evidence of positive or balancing selection in the genes AK ($\omega = 24.841$, 95% CI [12.988–36.694]), APBP ($\omega = 13.495$, 95% CI [0–35.670]), and SPI ($\omega = 1.356$, 95% CI [0–Inf]), while EJAC-SP approached $\omega = 1$ ($\omega = 0.993$, 95% CI [0–16.673]). In the younger split between *A. socius* and *A. sp. nov. Tex*, ω did not exceed 1 in any of the genes ([Table 2](#)). Because of the combination of low π_s , π_a , and θ and high ω for these genes, we will henceforth interpret $\omega > 1$ as evidence of positive selection as balancing selection is more likely to be accompanied by higher nucleotide diversity.

Not all genes had fixed non-synonymous substitutions between species and in these cases we were unable to apply the McDonald–Kreitman test ([Table 2](#)). For those genes that were testable, we did not find significant differences in D_N/D_S compared to P_N/P_S at either branching event (Fisher's exact test $P = 0.07$ – 1). We were unable to detect a significant departure from neutral expectations for the first branching event between *A. fasciatus* and the two other species using the multilocus HKA test ($X^2P = 0.916$). We did detect a significant departure from neutral expectations for the second branching

Table 1 Nucleotide variation within each *A. socius* complex species.

Gene	Length	<i>A. fasciatus</i>				<i>A. socius</i>				<i>A. sp. nov. Tex</i>			
		<i>n</i> (pop)	π_s	π_a	θ_{fas}	<i>n</i> (pop)	π_s	π_a	θ_{soc}	<i>n</i> (pop)	π_s	π_a	θ_{Tex}
ACG69	414	9 (3)	0.005	0.004	0.007	14 (6)	0.021	0.009	0.011	7 (3)	0	0	0
AK	1,173	9 (3)	0.002	<0.001	0.001	15 (6)	0.004	0.001	0.002	6 (3)	0.003	0.001	0.002
APBP	705	9 (3)	0.001	0	0.001	15 (5)	0.005	0	0.001	8 (3)	0	0	0
EJAC-SP	726	9 (3)	0	0	0	16 (6)	0.001	<0.001	0.001	9 (3)	0.003	0	0.001
GOT	1,122	9 (3)	0.002	0	<0.001	17 (7)	0	0	0	9 (3)	0.005	0.001	0.002
SPAG6	426	9 (3)	0	0	0	17 (6)	0	0	0	8 (3)	0.005	0	0.001
SPI	315	9 (3)	0.007	0	0.002	16 (6)	0	0	0	9 (3)	0.008	0.001	0.002

Notes.

The number of individuals sampled for each gene are shown with the number of populations they came from (*n* (pop)). π_s , the average number of synonymous nucleotide differences per site for any random pair of sequences; π_a , the average number of non-synonymous nucleotide differences per site for any random pair of sequences; θ , a metric of the population substitution rate; all metrics were calculated across populations.

event between *A. socius* and *A. sp. nov. Tex.* ($X^2P = 0.012$). The outlier that diverged most from null expectations was polymorphism in GOT in *A. sp. nov. Tex.*, but coalescent simulations were unable to determine that this was a significant difference ($P = 0.06$). Both the McDonald–Kreitman and HKA tests are likely to be underpowered for our species given the age and amount of variation present in the *Drosophila* systems that both tests were originally developed for [Hudson, Kreitman & Aguadé \(1987\)](#); [McDonald & Kreitman \(1991\)](#); [Wang & Hey, \(1996\)](#).

Gene tree-based analyses

We tested for evidence that any of the internal branches of each gene tree were evolving at a higher non-synonymous rate than the other branches ([Fig. 1](#)). We found evidence that the model in which two internal nodes in the gene tree of AK were allowed to evolve at a variable non-synonymous rate was a better fit than the single rate class model ($P = 0.004$). These nodes evolved at non-synonymous substitution rates 26.331 (95% CI [4.377–81.378]) and 0.147 (95% CI [0.008–0.649]), while the shared non-synonymous substitution rate was estimated to be 0. The first node (5) is the node that separates *A. fasciatus* and the two other species on the AK gene tree. The second node (43) is the node that separates *A. socius* and *A. sp. nov. Tex.* on the AK gene tree. We also found evidence that two internal nodes in the gene tree of APBP were evolving at a different rate compared to the background branches ($P = 0.015$). These nodes were evolving at rates 43.793 (95% CI [2.515–194.273]) and 19.008 (95% CI [1.088–84.035]), while the shared non-synonymous substitution rate was estimated to be 0. The first node (11) is the node between *A. fasciatus* and the two other species on the APBP gene tree. The second node (45) is the node between *A. socius* and *A. sp. nov. Tex.* on the APBP gene tree. Despite evidence that some internal nodes were evolving at a higher rate compared to background nodes in the other genes, we did not have sufficient evidence to suggest branches in any of the other gene trees were evolving at more than one non-synonymous rate class (ACG69 (6 tested) $P = 0.171$, EJAC-SP (1 tested) $P = 0.277$, GOT (3 tested) $P = 0.218$, SPI (1 tested) $P = 0.265$, SPAG6 (2 tested) $P = 1$).

Table 2 Nucleotide variation at each branching node of the *A. socius* complex species tree.

Gene	Length	Between <i>A. fasciatus</i> & (<i>A. socius</i> + <i>A. sp. nov. Tex</i>)							Between <i>A. socius</i> & <i>A. sp. nov. Tex</i>						
		P_N	P_S	D_N	D_S	κ_s	κ_a	ω (95% CI)	P_N	P_S	D_N	D_S	κ_s	κ_a	ω (95 % CI)
ACG69	414	7	6	0	0	0.024	0.009	0.572 (0, 1.366)	7	6	0	0	0.021	0.007	0.142 (0, 0.420)
AK	1,173	3	12	2	0	0.006	0.003	24.841 (12.988, 36.694)	2	9	1	2	0.011	0.002	0.147 (0, 0.436)
APBP	705	1	4	1	0	0.005	0.003	13.495 (0, 35.670)	0	3	1	0	0.007	0.002	0.285 (0, 0.843)
EJAC-SP	726	1	3	0	0	0.008	0.001	0.993 (0, 16.673)	1	2	0	1	0.014	0.002	0.158 (0, 0.467)
GOT	1,122	3	7	1	1	0.011	0.002	0.274 (0, 0.811)	3	4	0	2	0.011	0.001	0.136 (0, 0.402)
SPAG6	426	0	1	0	2	0.029	0	0 (0, 0.001)	0	1	0	0	0.004	0	0.001 (0.001, 0.002)
SPI	315	3	4	0	0	0.02	0.003	1.356 (0, inf)	1	1	2	3	0.054	0.009	0.178 (0, 0.425)

Notes.

P_N , Non-synonymous polymorphisms; P_S , synonymous polymorphisms; D_N , Non-synonymous fixations; D_S , synonymous fixations; κ_s , Rate of non-synonymous substitutions per non-synonymous site; κ_a , Rate of synonymous substitutions per synonymous site; $\omega = \kappa_a/\kappa_s$, Estimate from maximum likelihood model fit in HyPhy, along with 95% asymptotic normal confidence intervals (CI).

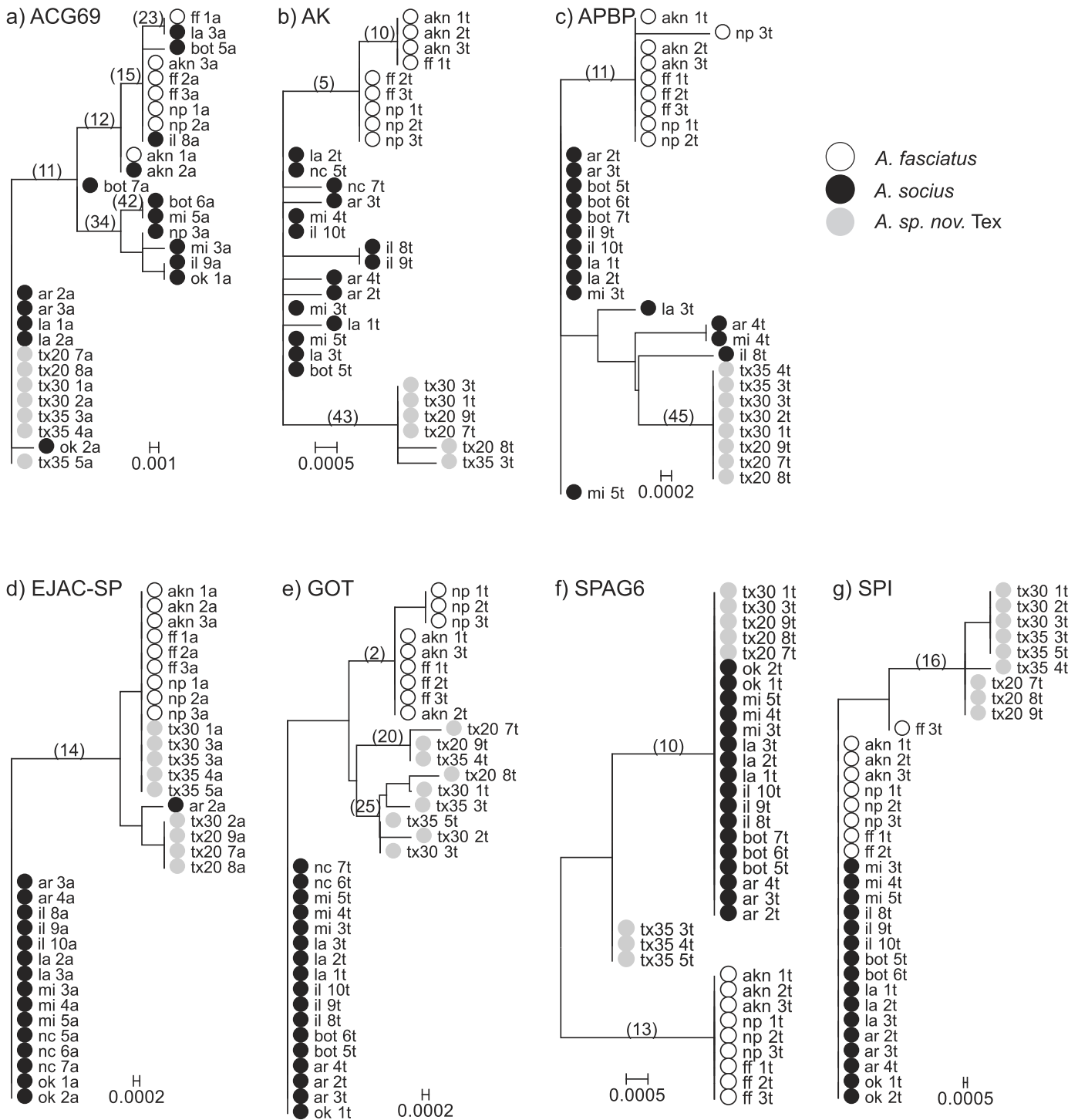


Figure 1 Neighbor-Joining gene trees that include allopatric individuals only. Internal branches whose non-synonymous substitution rate confidence intervals did not overlap with zero are marked. These internal branches were tested for evidence of elevated rates of non-synonymous substitutions (TestBranchdNdS) and for episodic selection among sites and branches (BUSTED).

Table 3 Genealogical sorting index values based on individual rooted gene trees. Values range from zero (complete polyphyly) to one (complete monophyly). P -values are obtained from 10,000 permutations that randomly reassign tip labels on the tree and represent the probability of observing a tree with a gsi value more extreme than the current topology.

Gene	gsi-fas	P_{perm}	gsi-soc	P_{perm}	gsi- <i>Tex</i>	P_{perm}
ACG69	0.458	<0.001	0	1	0.836	<0.001
AK	1	<0.001	0.843	<0.001	1	<0.001
APBP	1	<0.001	0.849	<0.001	1	<0.001
EJAC-SP	0.609	<0.001	0.730	<0.001	0.288	0.001
GOT	0.844	<0.001	0.533	<0.001	1	<0.001
SPAG6	1	<0.001	0.667	<0.001	0.191	0.016
SPI	0.189	0.010	0.622	<0.001	1	<0.001

Using SLAC and comparing model fits of nucleotide substitution models and codon models, for most genes (ACG69, AK, APBP, SPI, SPAG6) we found no evidence of any specific sites evolving under positive or negative selection. For EJAC-SP, one site (position 115) in which the amino acid serine was maintained through synonymous substitutions, showed evidence of negative selection ($P = 0.011$). Also for GOT, one site (position 88) in which the amino acid threonine was maintained through synonymous substitutions, showed evidence of negative selection ($P = 0.037$).

We used BUSTED to test for episodic selection anywhere in each gene. We tested whether any sites evolved at a faster ω rate within the internal branches with non-zero branches tested in the previous branch analysis. We found non-significant evidence of episodic selection only in ACG69 using this method (Likelihood Ratio Test $P = 0.062$). The unconstrained model suggests 1.7% of foreground branch sites may be evolving with $\omega = 318.77$ while background branch sites have a maximum $\omega = 0.96$ (100% of the sites were partitioned into this rate class). This branch-site test is likely to be underpowered for our species given that the numbers of parameters that were estimated for the unconstrained model was upwards of 80 (depending on the number of terminal taxa for each gene), while the difference in numbers of parameters between the complex model and null model was only 1. We do not discuss results from this method further because results were non-significant across all genes.

Evidence from the contact zone between *A. fasciatus* and *A. socius*

Comparisons of genealogical sorting index values based on gene trees including all sampled individuals, both allopatric and contact zone, indicated that lineage sorting was ongoing as gsi values ranged from 0 to 1, with a median value of 0.836. If we use the median value as a cutoff, only AK and APBP showed relatively advanced lineage sorting for all three species while GOT showed advanced lineage sorting for *A. fasciatus* and *A. sp. nov.* *Tex* but not *A. socius* (Table 3 and Fig. 2).

The statistical parsimony haplotype networks generated using allopatric individuals of all three species showed that alleles in AK, APBP, and GOT were specific to each species, while the alleles present in other genes included both species-specific alleles and alleles



Figure 2 Rooted Neighbor-Joining gene trees that include both allopatric and contact zone individuals. These trees were used to estimate degrees of lineage sorting for each species.

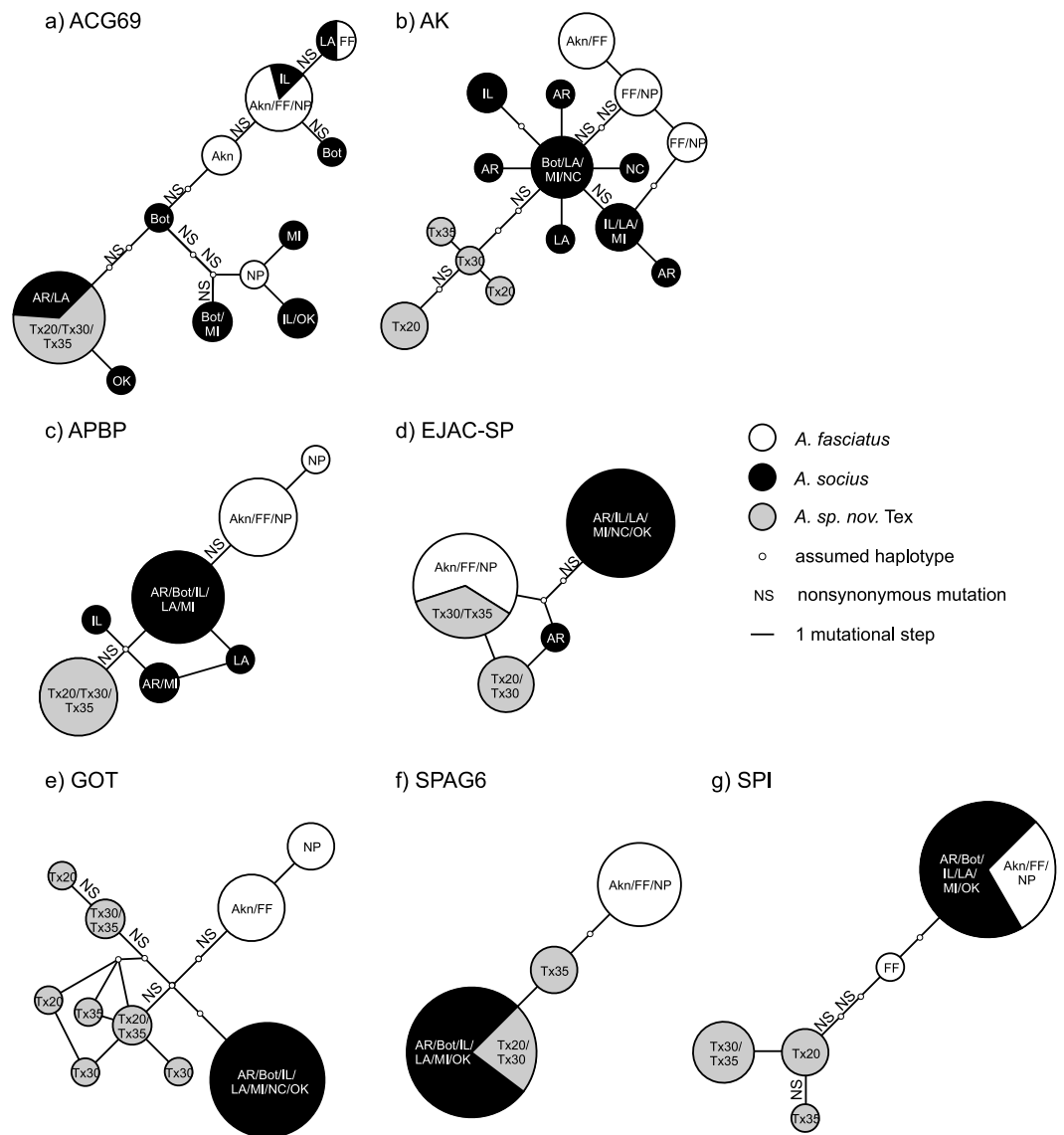


Figure 3 Statistical parsimony haplotype networks including only allopatric individuals, with mutational steps and nonsynonymous substitutions indicated. Population abbreviations are as in the main text. Sizes of each haplotype correspond to the relative number of individuals that possessed each haplotype.

shared between two species (Fig. 3). Within the contact zone between *A. fasciatus* and *A. socius*, all genes except APBP had a mix of alleles that we had already observed in allopatric individuals as well as new alleles that were specific to the contact zone (Fig. 2). We designated these new alleles as species-specific based on their allelic distributions among contact zone individuals and calculated the degree of allelic segregation within the contact zone. The dissimilarity index D values we observed were bimodally distributed and indicated allelic segregation was higher in the contact zone for AK, APBP, EJACSP and SPAG6 ($D = 0.771-1$) and lower in ACG69, GOT and SPI ($D = 0.164-0.353$) (Fig. 4).

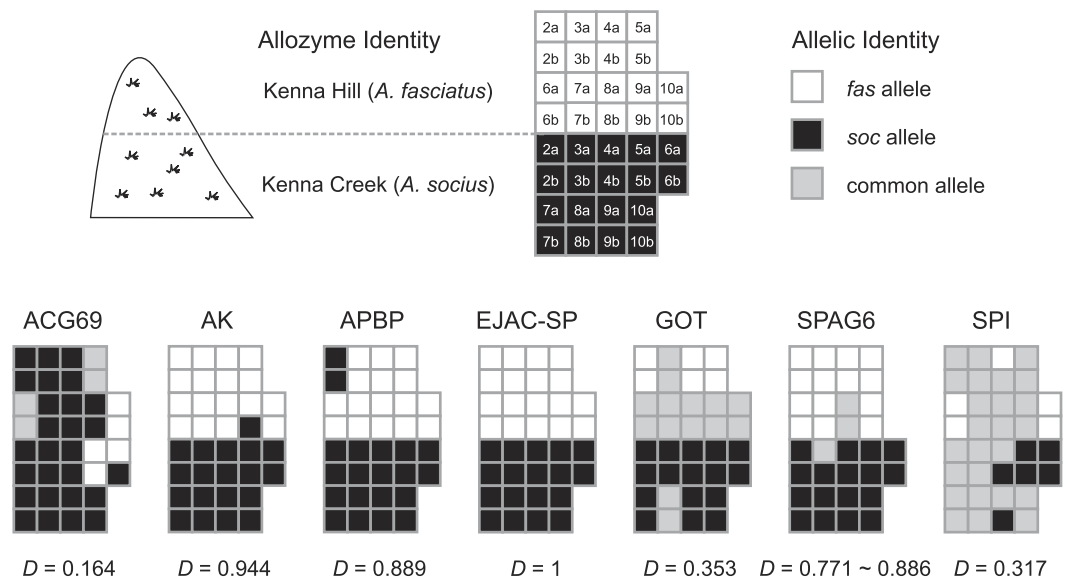


Figure 4 Distribution of species-specific vs. common (shared) alleles within the *A. fasciatus* and *A. socius* contact zone in Kenna, WV. Nine individuals each with allozyme identities of pure (homozygous) *A. fasciatus* and *A. socius* had varying patterns of allelic identities for the seven genes. Numbers (2–9) indicate the sampled individual and letters (A and B) indicate the alleles within each individual. The dissimilarity index D is a metric of the degree of allelic segregation between the two contact zone populations.

DISCUSSION

Studying speciation in recently diverged species is attractive because we can potentially identify variation that is associated with divergence and not accumulated after (Andrés *et al.*, 2013). But using within and between species patterns of sequence variation to detect adaptive evolution is particularly challenging in recently diverged species because less evolutionary time has passed to allow for fixed differences to accumulate between incipient species relative to within species polymorphism (Keightley & Eyre-Walker, 2012). We gathered data from populations across each species distribution so as to accurately differentiate between polymorphisms and fixations and detect new mutations to confirm and expand upon previous results from more limited population sampling (Marshall *et al.*, 2011). Combining evidence from multiple tests allowed us to identify the most likely putative postmating prezygotic isolation genes among our EP coding genes (Table 4). Ultimately, the methodological approaches that we were able to apply most successfully were those that took advantage of evolutionary processes specific to young species: lineage sorting in gene trees and allele sharing in haplotype networks.

Data from the relatively recently (~30,000 years) diverged *A. socius* species complex showed a general lack of both synonymous and non-synonymous nucleotide variation within and among all investigated genes (Tables 1 and 2). Our estimates of sequence variation were also at least an order of magnitude smaller compared to other known estimates from accessory gland protein coding genes in various other species groups (Mueller *et al.*, 2005; Wagstaff & Begun, 2005; Almeida & DeSalle, 2008), including some *Gryllus* crickets whose species are of roughly similar age (Andrés *et al.*, 2006). Since

Table 4 Summary of results found testing for evidence of selection in ejaculate protein coding genes in the *A. socius* complex. The specific method used for each result is shown in parentheses.

Genes	Species tree π_s, π_a, θ	Species tree ω (MLE)	$D_N/D_S >$ P_N/P_S (MK)	Departure from other genes (HKA)	Gene tree evi- dence of branch selection (Test- BranchdNds)	Gene tree evi- dence of site se- lection (SLAC)	Lineage sort- ing (allopatric + contact zone) (gsi)	Allelic segre- gation in con- tact zone (<i>D</i>)
ACG69	Small	All <1	NA	No	No	No	Low	Low
AK	Smaller	Some >1	No	No	Yes	No	High	High
APBP	Smaller	Some >1	No	No	Yes	No	High	High
EJAC-SP	Smaller	Some ~1	NA	No	No	Purifying selection	Low	High
GOT	Smaller	All <1	No	Maybe	No	Purifying selection	Moderate	Low
SPAG6	Smaller	All <1	NA	No	No	No	Low	High
SPI	Smaller	Some >1	No	No	No	No	Low	Low

speciation in the *A. socius* complex is thought to coincide with glaciation history (Marshall, 2004; Marshall, 2007) population bottlenecks may have contributed to sequence variation patterns. Our results would have benefited from increased within-population sampling so that metrics such as Tajima's or Fu and Li's *D* and site frequency spectra, as well as estimates of population size and recombination rates (Nielsen, 2005) could be accurately applied. Unsurprisingly, the tests we used that relied on sequence variation on the known species tree had minor success identifying genes with evidence of positive selection based on ω (Table 4). The maximum likelihood estimates of ω on the species tree indicated that AK, APBP, EJAC-SP and SPI showed evidence of positive selection at the branching event between *A. fasciatus* and the other two species (Table 2). However, McDonald–Kreitman tests were not applicable to several genes and HKA tests were generally inconclusive.

We found approaches that used gene trees and haplotype networks were more successful at detecting evidence of positive selection, and these approaches were particularly powerful within the contact zone (Table 4). When allopatric gene trees were tested for evidence of internal branches evolving at different rates compared to the rest of the tree, we found evidence that AK and APBP both had internal branches that were evolving at higher non-synonymous substitution rates (Fig. 1). These internal branches respectively separated *A. fasciatus* from the other two species, and *A. socius* from *A. sp. nov. Tex.* When all allopatric and contact zone individuals were examined, the genealogies of AK and APBP indicated that these genes were relatively advanced in their degrees of lineage sorting in all three species of the *A. socius* complex compared to the other genes (Table 3 and Fig. 2). GOT also showed advanced lineage sorting, but only in the two species *A. fasciatus* and *A. sp. nov. Tex.* Within the contact zone of *A. fasciatus* and *A. socius*, AK, APBP, EJAC-SP, and SPAG6 showed highly segregated allelic distributions (Fig. 4). The two genes with intermediate lineage sorting or allelic segregation patterns, EJAC-SP and GOT, showed evidence of purifying selection acting upon specific sites, as might be expected of most functional protein coding genes (Lawrie et al., 2013), but no evidence of positive selection.

The overlap among these results suggests AK and APBP are the most likely candidates for postmating prezygotic isolation genes (Table 4). While the remaining genes that show intermediate patterns may still contribute to reproductive isolation, they are less likely to be the major contributors to isolation barriers that act in a species-specific manner.

These patterns fit models of ongoing speciation in the face of gene flow, where incomplete lineage sorting and introgression are major confounding factors (Machado & Hey, 2003; Broughton & Harrison, 2003; Payseur, 2010). Speciation genes are more likely to become fixed for species-specific alleles early in the process of speciation, and therefore are expected to be relatively exempt from incomplete sorting and subject to reduced introgression (Wu, 2001; Feder, Egan & Nosil, 2012). Similar patterns have been observed in *Drosophila*, field crickets, and moths (Ting, Tsauro & Wu, 2000; Dopman et al., 2005; Maroja, Andrés & Harrison, 2009; Andrés et al., 2013; Larson et al., 2013). It is possible that these genes are not the direct targets but rather linked to targets of divergent selection as the interaction between linkage and selection makes it challenging to distinguish between recurrent positive selection, background selection, and Hill–Robertson effects (Hill & Robertson, 1966; Charlesworth, 1994; Andolfatto, 2007; Charlesworth et al., 2009). Because both genes were identified as candidates through comparative proteomics (Marshall et al., 2011) this seems relatively unlikely, but the genomic regions around these genes should be investigated for evidence of selective sweeps to rule out this possibility.

Many studies of reproductive proteins report evidence of positive selection acting on a subset of the genes examined, in both males (Begun et al., 2000; Swanson et al., 2001; Clark & Swanson, 2005; Wagstaff & Begun, 2005; Andrés et al., 2006; Karn et al., 2008; Ramm et al., 2008; Almeida & DeSalle, 2008; Walters & Harrison, 2010) and females (Swanson et al., 2004; Panhuis & Swanson, 2006; Lawniczak & Begun, 2007; Prokupek et al., 2008; Kelleher & Markow, 2009; Kelleher, Clark & Markow, 2011). However, there are few examples of adaptive reproductive protein evolution leading to reproductive isolation outside of gamete recognition proteins (e.g., Geyer & Palumbi, 2003; McCartney & Lessios, 2004; Springer & Crespi, 2007). Our putative postmating prezygotic isolation genes AK and APBP are two of the first examples of sperm maturation and capacitation related proteins that show evidence of fixed non-synonymous substitutions between species-specific alleles leading to reproductive isolation (Fig. 3). We had previously observed this pattern between *A. fasciatus* and *A. socius* for both AK and APBP (Marshall et al., 2011), but finding the same pattern in the mutational steps between *A. socius* and *A. sp. nov. Tex* with different species-specific non-synonymous substitutions, with expanded population and sequence sampling, emphasizes the potential importance of these candidates.

Whether there are functional consequences to the species-specific non-synonymous substitutions in AK and APBP needs to be investigated further. Both candidates are homologous to mammalian sperm capacitation proteins. Sperm maturation and capacitation occur inside the female reproductive tract of mammals (Visconti et al., 2011). In insects and nematodes, sperm are capacitated and become motile by serine proteases present in the seminal fluid (LaFlamme & Wolfner, 2013). This process occurs within the spermatophores of Lepidoptera (Osanai & Chen, 1993; Friedländer, Jeshtadi & Reynolds, 2001), and in the seminal vesicles of *Drosophila* who do not make spermatophores (Osanai

& Chen, 1993). In either case, most insect sperm should already be mature and capacitated once spermatophores are transferred to females. Therefore the specific hypothetical function of these genes should be related to a different postcopulatory process that occurs within the female reproductive tract.

Female insects store sperm in specialized organs such as spermathecae and seminal receptacles after copulation, often for prolonged periods of time (Schnakenberg, Siegal & Bloch Qazi, 2012). Sperm lose motility within a day in flies, unless stored by females into these organs (Schnakenberg, Matias & Siegal, 2011). Conspecific sperm precedence among *Drosophila* species has been shown to involve post-copulatory processes that occur within the female reproductive tract that include sperm transfer, the displacement and ejection of less preferred (heterospecific) sperm, and the selective use of preferred (conspecific) sperm from different storage organs for fertilization (Manier et al., 2013a; Manier et al., 2013b). In flies, spermathecal secretory cells are intimately involved in the female driven part of these processes (Schnakenberg, Matias & Siegal, 2011). Endopeptidases produced by these secretory cells are necessary for recruiting sperm to spermathecae and also maintaining sperm motility in the seminal receptacle. The same endopeptidases also affect egg laying, so that females laid fewer eggs when their secretory cells were ablated. More recently, it was shown that the number of secretory cells that produce these endopeptidases determined whether female flies ovulated and laid eggs (Sun & Spradling, 2013). It is yet unknown whether sequence variation in these endopeptidases is related to variation in female fecundity. Other potential roles for the secretions of these cells as related to sperm precedence include a chemotactic function that would attract select sperm to different locations of the female reproductive tract (Wolfner, 2011).

The specific mechanism of conspecific sperm precedence that APBP is involved in may be related to the appropriate phosphorylation state of the phosphoprotein (APBP) depending on the female vs. male species combination. APBP becomes phosphorylated during murine sperm capacitation and co-localizes with cholesterol during this process, but its specific function is unknown (Jha et al., 2008). As noted above, insect sperm do not undergo capacitation as mammalian sperm do. As a putative binding protein with an enzymatic function, hypothetical functions of APBP in the female reproductive tract include that as a signal for sperm to be transferred to the preferred storage organ. It may also be involved in a signaling cascade that induces ovulation and oviposition in females. Between *A. fasciatus* and *A. socius*, when females were mated with only a single male those mated with heterospecific males laid fewer eggs but an equal proportion of fertilized eggs compared to females mated with conspecific males (Gregory & Howard, 1993).

Insects and other ecdysozoans possess AK as their sole phosphagen system for cellular energy metabolism, and accordingly, arginine phosphate and its phosphagen kinase AK are found primarily in muscles, but also in sperm and compound eyes (Strong & Ellington, 1993; Kucharski & Maleszka, 1998; Ellington, 2001). The possible roles of AK as an EP can be related to sperm motility or the acrosome reaction (Strong & Ellington, 1993; Niksirat et al., 2015). As an energy-related phosphagen kinase, AK may confer to sperm an enhanced ability to move toward a sperm storage organ and resist displacement and ejection by females. In honeybees, AK is expressed in both male seminal fluid (Baer et al.,

2009b) and female spermathecal fluid (Baer et al., 2009a). When tested over a two-year period AK enzymatic activity decreased in mated female spermathecal fluid and sperm motility also decreased (Al-Lawati, Kamp & Bienefeld, 2009). The proteomes of freshly ejaculated vs. stored honeybee sperm are also known to differ, particularly in terms of reduced activity of glycolytic enzymes that should be used for energy metabolism (Poland et al., 2011). Poland and colleagues note that most of the same enzymes are also present in the spermathecal fluid (Baer et al., 2009a), suggesting female physiology has an active role in maintaining stored sperm. Two structural loops and several active sites near them are the proposed interaction interface of AK with the guanidinium groups of its substrates (Zhou et al., 1998; Pruett et al., 2003; Azzi et al., 2004; Clark, Davulcu & Chapman, 2012). As might be expected for an integral enzyme, the non-synonymous substitutions we observed do not occur at these specific sites, though they may still influence its activity.

CONCLUSIONS

A. socius complex crickets provide an excellent opportunity to identify patterns of evolution in speciation genes for two major reasons: speciation is incomplete as evidenced by ongoing gene flow in the field, and isolation is through a single type of reproductive isolation barrier, i.e., postmating prezygotic phenotypes. We found that when speciation is ongoing, combining multiple approaches, particularly those that incorporate evidence from gene trees and haplotype networks, was important for identifying putative postmating prezygotic isolation genes. Both AK and APBP have fixed, or nearly fixed, non-synonymous substitutions between both *A. fasciatus* and the other two species, and between *A. socius* and *A. sp. nov.* Tex. Both genes showed advanced lineage sorting across and within the contact zone, and allelic segregation within the contact zone. The next step is to determine the functional consequence of allelic variation in either EP in conspecific sperm precedence and the successful induction of female oviposition in the *A. socius* complex.

ACKNOWLEDGEMENTS

We thank Christopher Garcia for collecting the crickets used in this study and Diana Huestis and Shanda Wheeler for their help in isolating RNA and screening individuals for species status with allozymes. We also thank Daniel J. Howard for his support throughout this work. This is contribution no. 12-015-J from the Kansas Agricultural Experiment Station.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

Funding from the National Science Foundation to JLM (DEB-0746316) supported this work. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Grant Disclosures

The following grant information was disclosed by the authors:
National Science Foundation: DEB-0746316.

Competing Interests

The authors declare there are no competing interests.

Author Contributions

- Suegene Noh performed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, prepared figures and/or tables, reviewed drafts of the paper.
- Jeremy L. Marshall conceived and designed the experiments, analyzed the data, contributed reagents/materials/analysis tools, wrote the paper, reviewed drafts of the paper.

DNA Deposition

The following information was supplied regarding the deposition of DNA sequences:

Sequences formatted as haplotypes are available from NCBI GenBank PopSets [372477571](#) (ACG69), [372477483](#) (AK), [372477513](#) (APBP), [372477527](#) (EJAC-SP), [372477535](#) (GOT), [372477555](#) (SPAG6), [372477561](#) (SPI).

Data Availability

The following information was supplied regarding data availability:

Two archives are included in the submission. The content files can be extracted using the following command in a Unix shell:

```
tar-xvzf archive_file_name.
```

Supplemental Information

Supplemental information for this article can be found online at <http://dx.doi.org/10.7717/peerj.1678#supplemental-information>.

REFERENCES

- Al-Lawati H, Kamp G, Bienefeld K. 2009.** Characteristics of the spermathecal contents of old and young honeybee queens. *Journal of Insect Physiology* **55**:116–121
[DOI 10.1016/j.jinsphys.2008.10.010](#).
- Almeida FC, DeSalle R. 2008.** Evidence of adaptive evolution of accessory gland proteins in closely related species of the *Drosophila repleta* group. *Molecular Biology and Evolution* **25**:2043–2053 [DOI 10.1093/molbev/msn155](#).
- Andolfatto P. 2007.** Hitchhiking effects of recurrent beneficial amino acid substitutions in the *Drosophila melanogaster* genome. *Genome Research* **17**:1755–1762
[DOI 10.1101/gr.6691007](#).
- Andrés JA, Larson EL, Bogdanowicz SM, Harrison RG. 2013.** Patterns of transcriptome divergence in the male accessory gland of two closely related species of field crickets. *Genetics* **193**:501–513 [DOI 10.1534/genetics.112.142299](#).

- Andrés JA, Maroja LS, Bogdanowicz SM, Swanson WJ, Harrison RG. 2006.** Molecular evolution of seminal proteins in field crickets. *Molecular Biology and Evolution* 23:1574–1584 DOI [10.1093/molbev/msl020](https://doi.org/10.1093/molbev/msl020).
- Andrés JA, Maroja LS, Harrison RG. 2008.** Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. *Proceedings of the Royal Society of London B: Biological Sciences* 275:1975–1983 DOI [10.1098/rspb.2008.0423](https://doi.org/10.1098/rspb.2008.0423).
- Anisimova M, Nielsen R, Yang Z. 2003.** Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. *Genetics* 164:1229–1236.
- Avila FW, Sirot LK, LaFlamme BA, Rubinstein CD, Wolfner MF. 2011.** Insect seminal fluid proteins: identification and function. *Annual Review of Entomology* 56:21–40 DOI [10.1146/annurev-ento-120709-144823](https://doi.org/10.1146/annurev-ento-120709-144823).
- Azzi A, Clark SA, Ellington WR, Chapman MS. 2004.** The role of phosphagen specificity loops in arginine kinase. *Protein Science* 13:575–585 DOI [10.1110/ps.03428304](https://doi.org/10.1110/ps.03428304).
- Baer B, Eubel H, Taylor NL, O'Toole N, Harvey Millar A. 2009a.** Insights into female sperm storage from the spermathecal fluid proteome of the honeybee *Apis mellifera*. *Genome Biology* 10:R67 DOI [10.1186/gb-2009-10-6-r67](https://doi.org/10.1186/gb-2009-10-6-r67).
- Baer B, Heazlewood JL, Taylor NL, Eubel H, Harvey Millar A. 2009b.** The seminal fluid proteome of the honeybee *Apis mellifera*. *Proteomics* 9:2085–2097 DOI [10.1002/pmic.200800708](https://doi.org/10.1002/pmic.200800708).
- Begun DJ, Whitley P, Todd BL, Waldrip-Dail HM, Clark AG. 2000.** Molecular population genetics of male accessory gland proteins in *Drosophila*. *Genetics* 156:1879–1888.
- Broughton RE, Harrison RG. 2003.** Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics* 163:1389–1401.
- Chapman T, Davies SJ. 2004.** Functions and analysis of the seminal fluid proteins of male *Drosophila melanogaster* fruit flies. *Peptides* 25:1477–1490 DOI [10.1016/j.peptides.2003.10.023](https://doi.org/10.1016/j.peptides.2003.10.023).
- Charlesworth B. 1994.** The effect of background selection against deleterious mutations on weakly selected, linked variants. *Genetical Research* 63:213–227 DOI [10.1017/S0016672300032365](https://doi.org/10.1017/S0016672300032365).
- Charlesworth B, Betancourt AJ, Kaiser VB, Gordo I. 2009.** Genetic recombination and molecular evolution. *Cold Spring Harbor Symposium on Quantitative Biology* 74:177–186 DOI [10.1101/sqb.2009.74.015](https://doi.org/10.1101/sqb.2009.74.015).
- Civetta A, Singh RS. 1998.** Sex and speciation: genetic architecture and evolutionary potential of sexual versus nonsexual traits in the sibling species of the *Drosophila melanogaster* complex. *Evolution* 52:1080–1092.
- Clark NL, Aagaard JE, Swanson WJ. 2006.** Evolution of reproductive proteins from animals and plants. *Reproduction* 131:11–22 DOI [10.1530/rep.1.00357](https://doi.org/10.1530/rep.1.00357).
- Clark SA, Davulcu O, Chapman MS. 2012.** Crystal structures of arginine kinase in complex with ADP, nitrate, and various phosphagen analogs. *Biochemical and Biophysical Research Communications* 427:212–217 DOI [10.1016/j.bbrc.2012.09.053](https://doi.org/10.1016/j.bbrc.2012.09.053).

- Clark NL, Swanson WJ. 2005.** Pervasive adaptive evolution in primate seminal proteins. *PLoS Genetics* **1**:e35 DOI [10.1371/journal.pgen.0010035](https://doi.org/10.1371/journal.pgen.0010035).
- Clement M, Posada D, Crandall KA. 2000.** TCS: a computer program to estimate gene genealogies. *Molecular Ecology* **9**:1657–1659 DOI [10.1046/j.1365-294x.2000.01020.x](https://doi.org/10.1046/j.1365-294x.2000.01020.x).
- Cummings MP, Neel MC, Shaw KL. 2008.** A genealogical approach to quantifying lineage divergence. *Evolution* **62**:2411–2422 DOI [10.1111/j.1558-5646.2008.00442.x](https://doi.org/10.1111/j.1558-5646.2008.00442.x).
- Dean MD, Clark NL, Findlay GD, Karn RC, Yi X, Swanson WJ, MacCoss MJ, Nachman MW. 2009.** Proteomics and comparative genomic investigations reveal heterogeneity in evolutionary rate of male reproductive proteins in mice (*Mus domesticus*). *Molecular Biology and Evolution* **26**:1733–1743 DOI [10.1093/molbev/msp094](https://doi.org/10.1093/molbev/msp094).
- Delport W, Poon AF, Frost SD, Pond SLK. 2010.** Datamonkey 2010: a suite of phylogenetic analysis tools for evolutionary biology. *Bioinformatics* **26**:2455–2457 DOI [10.1093/bioinformatics/btq429](https://doi.org/10.1093/bioinformatics/btq429).
- Dopman EB, Pérez L, Bogdanowicz SM, Harrison RG. 2005.** Consequences of reproductive barriers for genealogical discordance in the European corn borer. *Proceedings of the National Academy of Sciences of the United States of America* **102**:14706–14711 DOI [10.1073/pnas.0502054102](https://doi.org/10.1073/pnas.0502054102).
- Dorus S, Busby SA, Gerike U, Shabanowitz J, Hunt DF, Karr TL. 2006.** Genomic and functional evolution of the *Drosophila melanogaster* sperm proteome. *Nature Genetics* **38**:1440–1445 DOI [10.1038/ng1915](https://doi.org/10.1038/ng1915).
- Duncan OD, Duncan B. 1955.** A methodological analysis of segregation indexes. *American Sociological Review* **20**:210–217 DOI [10.2307/2088328](https://doi.org/10.2307/2088328).
- Ellington WR. 2001.** Evolution and the physiological roles of phosphagen systems. *Annual Review of Physiology* **63**:289–325 DOI [10.1146/annurev.physiol.63.1.289](https://doi.org/10.1146/annurev.physiol.63.1.289).
- Feder JL, Egan SP, Nosil P. 2012.** The genomics of speciation-with-gene-flow. *Trends in Genetics* **28**:342–350 DOI [10.1016/j.tig.2012.03.009](https://doi.org/10.1016/j.tig.2012.03.009).
- Felsenstein J. 1981.** Evolutionary trees from DNA-sequences: a maximum-likelihood approach. *Journal of Molecular Evolution* **17**:368–376 DOI [10.1007/BF01734359](https://doi.org/10.1007/BF01734359).
- Findlay GD, Sitnik JL, Wang W, Aquadro CF, Clark NL, Wolfner MF. 2014.** Evolutionary rate covariation identifies new members of a protein network required for *Drosophila melanogaster* female post-mating responses. *PLoS Genetics* **10**:e1004108 DOI [10.1371/journal.pgen.1004108](https://doi.org/10.1371/journal.pgen.1004108).
- Friedländer M, Jeshtadi A, Reynolds S. 2001.** The structural mechanism of trypsin-induced intrinsic motility in *Manduca sexta* spermatozoa *in vitro*. *Journal of Insect Physiology* **47**:245–255 DOI [10.1016/S0022-1910\(00\)00109-8](https://doi.org/10.1016/S0022-1910(00)00109-8).
- Geyer LB, Palumbi SR. 2003.** Reproductive character displacement and the genetics of gamete recognition in tropical sea urchins. *Evolution* **57**:1049–1060 DOI [10.1111/j.0014-3820.2003.tb00315.x](https://doi.org/10.1111/j.0014-3820.2003.tb00315.x).
- Gregory PJ, Howard DJ. 1993.** Laboratory hybridization studies of *Allonemobius fasciatus* and *A. socius* (Orthoptera: Gryllidae). *Annals of the Entomological Society of America* **86**:694–701 DOI [10.1093/aesa/86.6.694](https://doi.org/10.1093/aesa/86.6.694).
- Gregory PJ, Howard DJ. 1994.** A post-insemination barrier to fertilization isolates two closely related ground crickets. *Evolution* **48**:705–710 DOI [10.2307/2410480](https://doi.org/10.2307/2410480).

- Haerty W, Jagadeeshan S, Kulathinal RJ, Wong A, Ram KR, Sirot LK, Levesque L, Artieri CG, Wolfner MF, Civetta A, Singh RS. 2007.** Evolution in the fast lane: rapidly evolving sex-related genes in *Drosophila*. *Genetics* **177**:1321–1335 DOI [10.1534/genetics.107.078865](https://doi.org/10.1534/genetics.107.078865).
- Hall TA. 1999.** BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT. *Nucleic Acids Symposium Series* **41**:95–98.
- Hill WG, Robertson A. 1966.** The effect of linkage on limits to artificial selection. *Genetical Research* **8**:269–294 DOI [10.1017/S0016672300010156](https://doi.org/10.1017/S0016672300010156).
- Hong S-Y, O’Sullivan D, Sadahiro Y. 2014.** Implementing spatial segregation reasures in R. *PLoS ONE* **9**:e113767 DOI [10.1371/journal.pone.0113767](https://doi.org/10.1371/journal.pone.0113767).
- Howard DJ. 1983.** Electrophoretic survey of eastern North American *Allonemobius* (Orthoptera: Gryllidae): evolutionary relationships and the discovery of three new species. *Annals of the Entomological Society of America* **76**:1014–1021 DOI [10.1093/aesa/76.6.1014](https://doi.org/10.1093/aesa/76.6.1014).
- Howard DJ. 1986.** A zone of overlap and hybridization between two ground cricket species. *Evolution* **40**:34–43.
- Howard DJ, Gregory PJ, Chu J, Cain ML. 1998a.** Conspecific sperm precedence is an effective barrier to hybridization between closely related species. *Evolution* **52**:511–516 DOI [10.2307/2411086](https://doi.org/10.2307/2411086).
- Howard DJ, Marshall JL, Hampton DD, Britch SC, Draney ML, Chu J, Cantrell RG. 2002.** The genetics of reproductive isolation: a retrospective and prospective look with comments on ground crickets. *American Naturalist* **159**:S8–S21.
- Howard DJ, Reece PG, Gregory PJ, Chu J, Cain ML. 1998b.** The evolution of barriers to fertilization between closely related organisms. In: Howard DJ, Berlocher SH, eds. *Endless forms: species and speciation*. New York: Oxford University Press, 279–288.
- Howard DJ, Waring GL. 1991.** Topographic diversity, zone width, and the strength of reproductive isolation in a zone of overlap and hybridization. *Evolution* **45**:1120–1135.
- Hudson RR, Kreitman M, Aguadé M. 1987.** A test of neutral molecular evolution based on nucleotide data. *Genetics* **116**:153–159.
- Jensen JD, Wong A, Aquadro CF. 2007.** Approaches for identifying targets of positive selection. *Trends in Genetics* **23**:568–577 DOI [10.1016/j.tig.2007.08.009](https://doi.org/10.1016/j.tig.2007.08.009).
- Jha KN, Shumilin IA, Digilio LC, Chertihin O, Zheng H, Schmitz G, Visconti PE, Flickinger CJ, Minor W, Herr JC. 2008.** Biochemical and structural characterization of apolipoprotein AI binding protein, a novel phosphoprotein with a potential role in sperm capacitation. *Endocrinology* **149**:2108–2120 DOI [10.1210/en.2007-0582](https://doi.org/10.1210/en.2007-0582).
- Karn RC, Clark NL, Nguyen ED, Swanson WJ. 2008.** Adaptive evolution in rodent seminal vesicle secretion proteins. *Molecular Biology and Evolution* **25**:2301–2310 DOI [10.1093/molbev/msn182](https://doi.org/10.1093/molbev/msn182).
- Keightley PD, Eyre-Walker A. 2012.** Estimating the rate of adaptive molecular evolution when the evolutionary divergence between species is small. *Journal of Molecular Evolution* **74**:61–68 DOI [10.1007/s00239-012-9488-1](https://doi.org/10.1007/s00239-012-9488-1).

- Kelleher ES, Clark NL, Markow TA. 2011.** Diversity-enhancing selection acts on a female reproductive protease family in four subspecies of *Drosophila mojavensis*. *Genetics* **187**:865–876 DOI [10.1534/genetics.110.124743](https://doi.org/10.1534/genetics.110.124743).
- Kelleher ES, Markow TA. 2009.** Duplication, selection and gene conversion in a *Drosophila mojavensis* female reproductive protein family. *Genetics* **181**:1451–1465 DOI [10.1534/genetics.108.099044](https://doi.org/10.1534/genetics.108.099044).
- Kosakovsky Pond SL, Frost SDW. 2005a.** A simple hierarchical approach to modeling distributions of substitution rates. *Molecular Biology and Evolution* **22**:223–234.
- Kosakovsky Pond SL, Frost SDW. 2005b.** Not so different after all: a comparison of methods for detecting amino-acid sites under selection. *Molecular Biology and Evolution* **22**:1208–1222 DOI [10.1093/molbev/msi105](https://doi.org/10.1093/molbev/msi105).
- Kosakovsky Pond SL, Frost SDW, Muse SV. 2005.** HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**:676–679 DOI [10.1093/bioinformatics/bti079](https://doi.org/10.1093/bioinformatics/bti079).
- Kosakovsky Pond SLK, Murrell B, Fourment M, Frost SDW, Delpont W, Scheffler K. 2011.** A random effects branch-site model for detecting episodic diversifying selection. *Molecular Biology and Evolution* **28**:3033–3043 DOI [10.1093/molbev/msr125](https://doi.org/10.1093/molbev/msr125).
- Kosakovsky Pond SL, Muse SV. 2005.** HyPhy: hypothesis testing using phylogenies. In: Nielsen R, ed. *Statistical methods for molecular evolution. Statistics for biology and health*, New York: Springer, 125–181.
- Kosakovsky Pond SL, Posada D, Gravenor MB, Woelk CH, Frost SDW. 2006.** GARD: a genetic algorithm for recombination detection. *Bioinformatics* **22**:3096–3098 DOI [10.1093/bioinformatics/btl474](https://doi.org/10.1093/bioinformatics/btl474).
- Kucharski R, Maleszka R. 1998.** Arginine kinase is highly expressed in the compound eye of the honey-bee, *Apis mellifera*. *Gene* **211**:343–349 DOI [10.1016/S0378-1119\(98\)00114-0](https://doi.org/10.1016/S0378-1119(98)00114-0).
- LaFlamme BA, Wolfner MF. 2013.** Identification and function of proteolysis regulators in seminal fluid. *Molecular Reproductive Development* **80**:80–101 DOI [10.1002/mrd.22130](https://doi.org/10.1002/mrd.22130).
- Larson EL, Andrés JA, Bogdanowicz SM, Harrison RG. 2013.** Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution; International Journal of Organic Evolution* **67**:3653–3661 DOI [10.1111/evo.12205](https://doi.org/10.1111/evo.12205).
- Lawniczak MK, Begun DJ. 2007.** Molecular population genetics of female-expressed mating-induced serine proteases in *Drosophila melanogaster*. *Molecular Biology and Evolution* **24**:1944–1951 DOI [10.1093/molbev/msm122](https://doi.org/10.1093/molbev/msm122).
- Lawrie DS, Messer PW, Hershberg R, Petrov DA. 2013.** Strong purifying selection at synonymous sites in *D. melanogaster*. *PLoS Genetics* **9**:e1003527 DOI [10.1371/journal.pgen.1003527](https://doi.org/10.1371/journal.pgen.1003527).
- Li H. 2006.** Constructing the TreeFam database. PhD thesis, Institute of Theoretical Physics, Chinese Academy of Science.
- Librado P, Rozas J. 2009.** DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. *Bioinformatics* **25**:1451–1452 DOI [10.1093/bioinformatics/btp187](https://doi.org/10.1093/bioinformatics/btp187).

- Machado CA, Hey J. 2003.** The causes of phylogenetic conflict in a classic *Drosophila* species group. *Proceedings of the Royal Society of London B: Biological Sciences* **270**:1193–1202 DOI [10.1098/rspb.2003.2333](https://doi.org/10.1098/rspb.2003.2333).
- Maddison WP. 1997.** Gene trees in species trees. *Systematic Biology* **46**:523–536 DOI [10.1093/sysbio/46.3.523](https://doi.org/10.1093/sysbio/46.3.523).
- Manier MK, Belote JM, Lüpold S, Berben KS, Ala-Honkola O, Collins WF, Pitnick S. 2013b.** Rapid diversification of sperm precedence traits and processes among three sibling *Drosophila* species. *Evolution* **67**:2348–2362 DOI [10.1111/evo.12117](https://doi.org/10.1111/evo.12117).
- Manier MK, Lüpold S, Belote JM, Starmer WT, Berben KS, Ala-Honkola O, Collins WF, Pitnick S. 2013a.** Postcopulatory sexual selection generates speciation phenotypes in *Drosophila*. *Current Biology* **23**:1853–1862 DOI [10.1016/j.cub.2013.07.086](https://doi.org/10.1016/j.cub.2013.07.086).
- Maroja LS, Andrés JA, Harrison RG. 2009.** Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. *Evolution* **63**:2999–3015 DOI [10.1111/j.1558-5646.2009.00767.x](https://doi.org/10.1111/j.1558-5646.2009.00767.x).
- Marshall JL. 2004.** The *Allonemobius-Wolbachia* host-endosymbiont system: evidence for rapid speciation and against reproductive isolation driven by cytoplasmic incompatibility. *Evolution* **58**:2409–2425 DOI [10.1111/j.0014-3820.2004.tb00871.x](https://doi.org/10.1111/j.0014-3820.2004.tb00871.x).
- Marshall JL. 2007.** Rapid evolution of spermathecal duct length in the *Allonemobius socius* complex of crickets: species, population and *Wolbachia* effects. *PLoS ONE* **2**:e720 DOI [10.1371/journal.pone.0000720](https://doi.org/10.1371/journal.pone.0000720).
- Marshall JL, DiRienzo N. 2012.** Noncompetitive gametic isolation between sibling species of a cricket: a hypothesized link between within-population incompatibility and reproductive isolation between species. *International Journal of Evolutionary Biology* **2012**: 593438 DOI [10.1155/2012/593438](https://doi.org/10.1155/2012/593438).
- Marshall JL, Huestis DL, Garcia C, Hiromasa Y, Wheeler S, Noh S, Tomich JM, Howard DJ. 2011.** Comparative proteomics uncovers the signature of natural selection acting on the ejaculate proteomes of two cricket species isolated by postmating, prezygotic phenotypes. *Molecular Biology and Evolution* **28**:423–435 DOI [10.1093/molbev/msq230](https://doi.org/10.1093/molbev/msq230).
- Marshall JL, Huestis DL, Hiromasa Y, Wheeler S, Oppert C, Marshall SA, Tomich JM, Oppert B. 2009.** Identification, RNAi knockdown, and functional analysis of an ejaculate protein that mediates a postmating, prezygotic phenotype in a cricket. *PLoS ONE* **4**:e7537–e7546 DOI [10.1371/journal.pone.0007537](https://doi.org/10.1371/journal.pone.0007537).
- McCartney MA, Lessios HA. 2004.** Adaptive evolution of sperm binding tracks egg incompatibility in neotropical sea urchins of the genus *Echinometra*. *Molecular Biology and Evolution* **21**:732–745 DOI [10.1093/molbev/msh071](https://doi.org/10.1093/molbev/msh071).
- McDonald JH, Kreitman M. 1991.** Adaptive protein evolution at the *Adh* locus in *Drosophila*. *Nature* **351**:652–654 DOI [10.1038/351652a0](https://doi.org/10.1038/351652a0).
- Metta M, Gudavalli R, Gibert J-M, Schlötterer C. 2006.** No accelerated rate of protein evolution in male-biased *Drosophila pseudoobscura* genes. *Genetics* **174**:411–420 DOI [10.1534/genetics.106.057414](https://doi.org/10.1534/genetics.106.057414).

- Mueller JL, Ram KR, McGraw LA, Qazi MB, Siggia ED, Clark AG, Aquadro CF, Wolfner MF. 2005.** Cross-species comparison of *Drosophila* male accessory gland protein genes. *Genetics* **171**:131–143 DOI [10.1534/genetics.105.043844](https://doi.org/10.1534/genetics.105.043844).
- Mueller JL, Ripoll DR, Aquadro CF, Wolfner MF. 2004.** Comparative structural modeling and inference of conserved protein classes in *Drosophila* seminal fluid. *Proceedings of the National Academy of Sciences of the United States of America* **101**:13542–13547 DOI [10.1073/pnas.0405579101](https://doi.org/10.1073/pnas.0405579101).
- Murrell B, Weaver S, Smith MD, Wertheim JO, Murrell S, Aylward A, Eren K, Pollner T, Martin DP, Smith DM, Scheffler K, Kosakovsky Pond SLK. 2015.** Gene-wide identification of episodic selection. *Molecular Biology and Evolution* **32**:1365–1371 DOI [10.1093/molbev/msv035](https://doi.org/10.1093/molbev/msv035).
- Muse SV, Gaut BS. 1994.** A likelihood approach for comparing synonymous and nonsynonymous nucleotide substitution rates, with application to the chloroplast genome. *Molecular Biology and Evolution* **11**:715–724.
- Neilson LI, Schneider PA, Van Deerlin PG, Kiriakidou M, Driscoll DA, Pellegrini MC, Millinder S, Yamamoto KK, French CK, Strauss JF. 1999.** cDNA cloning and characterization of a human sperm antigen (SPAG6) with homology to the product of the *Chlamydomonas PF16* locus. *Genomics* **60**:272–280 DOI [10.1006/geno.1999.5914](https://doi.org/10.1006/geno.1999.5914).
- Nielsen R. 2001.** Statistical tests of selective neutrality in the age of genomics. *Heredity* **86**:641–647 DOI [10.1046/j.1365-2540.2001.00895.x](https://doi.org/10.1046/j.1365-2540.2001.00895.x).
- Nielsen R. 2005.** Molecular signatures of natural selection. *Annual Review of Genetics* **39**:197–218 DOI [10.1146/annurev.genet.39.073003.112420](https://doi.org/10.1146/annurev.genet.39.073003.112420).
- Niksirat H, James P, Andersson L, Kouba A, Kozák P. 2015.** Label-free protein quantification in freshly ejaculated versus post-mating spermatophores of the noble crayfish *Astacus astacus*. *Journal of Proteomics* **123**:70–77 DOI [10.1016/j.jprot.2015.04.004](https://doi.org/10.1016/j.jprot.2015.04.004).
- Noguchi M, Sawada T, Akazawa T. 2001.** ATP-regenerating system in the cilia of *Paramecium caudatum*. *Journal of Experimental Biology* **204**:1063–1071.
- Noor MAF, Feder JL. 2006.** Speciation genetics: evolving approaches. *Nature Reviews Genetics* **7**:851–861 DOI [10.1038/nrg1968](https://doi.org/10.1038/nrg1968).
- Nosil P, Schluter D. 2011.** The genes underlying the process of speciation. *Trends in Ecology & Evolution* **26**:160–167 DOI [10.1016/j.tree.2011.01.001](https://doi.org/10.1016/j.tree.2011.01.001).
- Osanai M, Chen PS. 1993.** A comparative study on the arginine degradation cascade for sperm maturation of *Bombyx mori* and *Drosophila melanogaster*. *Amino Acids* **5**:341–350 DOI [10.1007/BF00806952](https://doi.org/10.1007/BF00806952).
- Pamilo P, Nei M. 1988.** Relationships between gene trees and species trees. *Molecular Biology and Evolution* **5**:568–583.
- Panhuis TM, Swanson WJ. 2006.** Molecular evolution and population genetic analysis of candidate female reproductive genes in *Drosophila*. *Genetics* **173**:2039–2047 DOI [10.1534/genetics.105.053611](https://doi.org/10.1534/genetics.105.053611).
- Payseur BA. 2010.** Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources* **10**:806–820 DOI [10.1111/j.1755-0998.2010.02883.x](https://doi.org/10.1111/j.1755-0998.2010.02883.x).

- Poiani A. 2006. Complexity of seminal fluid: a review. *Behavioral Ecology and Sociobiology* 60:289–310 DOI 10.1007/s00265-006-0178-0.
- Poland V, Eubel H, King M, Solheim C, Harvey Millar A, Baer B. 2011. Stored sperm differs from ejaculated sperm by proteome alterations associated with energy metabolism in the honeybee *Apis mellifera*. *Molecular Ecology* 20:2643–2654 DOI 10.1111/j.1365-294X.2011.05029.x.
- Prokupek A, Hoffmann F, Eyun S, Moriyama E, Zhou M, Harshman L. 2008. An evolutionary expressed sequence tag analysis of *Drosophila* spermatheca genes. *Evolution* 62:2936–2947 DOI 10.1111/j.1558-5646.2008.00493.x.
- Pröschel M, Zhang Z, Parsch J. 2006. Widespread adaptive evolution of *Drosophila* genes with sex-biased expression. *Genetics* 174:893–900 DOI 10.1534/genetics.106.058008.
- Pruett PS, Azzi A, Clark SA, Yousef MS, Gattis JL, Somasundaram T, Ellington WR, Chapman MS. 2003. The putative catalytic bases have, at most, an accessory role in the mechanism of arginine kinase. *Journal of Biological Chemistry* 278:26952–26957 DOI 10.1074/jbc.M212931200.
- Ram KR, Ji S, Wolfner MF. 2005. Fates and targets of male accessory gland proteins in mated female *Drosophila melanogaster*. *Insect Biochemistry and Molecular Biology* 35:1059–1071 DOI 10.1016/j.ibmb.2005.05.001.
- Ram KR, Wolfner MF. 2007. Seminal influences: *Drosophila* Acps and the molecular interplay between males and females during reproduction. *Integrative and Comparative Biology* 47:427–445 DOI 10.1093/icb/icm046.
- Ramm SA, Oliver PL, Ponting CP, Stockley P, Emes RD. 2008. Sexual selection and the adaptive evolution of mammalian ejaculate proteins. *Molecular Biology and Evolution* 25:207–219.
- Rieseberg LH, Church SA, Morjan CL. 2004. Integration of populations and differentiation of species. *New Phytologist* 161:59–69.
- Sapiro R, Kostetskii I, Olds-Clarke P, Gerton GL, Radice GL, Strauss III JF. 2002. Male infertility, impaired sperm motility, and hydrocephalus in mice deficient in Sperm-Associated Antigen 6. *Molecular and Cellular Biology* 22:6298–6305 DOI 10.1128/MCB.22.17.6298-6305.2002.
- Schnakenberg SL, Matias WR, Siegal ML. 2011. Sperm-storage defects and live birth in *Drosophila* females lacking spermathecal secretory cells. *PLoS Biology* 9:e1001192 DOI 10.1371/journal.pbio.1001192.
- Schnakenberg SL, Siegal ML, Bloch Qazi MC. 2012. Oh, the places they'll go: female sperm storage and sperm precedence in *Drosophila melanogaster*. *Spermatogenesis* 2:224–235 DOI 10.4161/spmg.21655.
- Shriner D, Nickle DC, Jensen MA, Mullins JI. 2003. Potential impact of recombination on sitewise approaches for detecting positive natural selection. *Genetical Research* 81:115–121 DOI 10.1017/S0016672303006128.
- Snook RR, Chapman T, Moore PJ, Wedell N, Crudgington HS. 2009. Interactions between the sexes: new perspectives on sexual selection and reproductive isolation. *Evolutionary Ecology* 23:71–91 DOI 10.1007/s10682-007-9215-3.

- Springer SA, Crespi BJ. 2007.** Adaptive gamete-recognition divergence in a hybridizing *Mytilus* population. *Evolution* **61**:772–783 DOI [10.1111/j.1558-5646.2007.00073.x](https://doi.org/10.1111/j.1558-5646.2007.00073.x).
- Strong SJ, Ellington WR. 1993.** Horseshoe crab sperm contain a unique isoform of arginine kinase that is present in the midpiece and flagellum. *Journal of Experimental Zoology* **267**:563–571 DOI [10.1002/jez.1402670603](https://doi.org/10.1002/jez.1402670603).
- Sun J, Spradling AC. 2013.** Ovulation in *Drosophila* is controlled by secretory cells of the female reproductive tract. *eLife* **2**:e00415 DOI [10.7554/eLife.00415](https://doi.org/10.7554/eLife.00415).
- Swanson WJ, Clark AG, Waldrip-Dail HM, Wolfner MF, Aquadro CF. 2001.** Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. *Proceedings of the National Academy of Sciences of the United States of America* **98**:7375–7379 DOI [10.1073/pnas.131568198](https://doi.org/10.1073/pnas.131568198).
- Swanson WJ, Vacquier VD. 2002.** The rapid evolution of reproductive proteins. *Nature Reviews Genetics* **3**:137–144.
- Swanson WJ, Wong A, Wolfner MF, Aquadro CF. 2004.** Evolutionary expressed sequence tag analysis of *Drosophila* female reproductive tracts identifies genes subjected to positive selection. *Genetics* **168**:1457–1465 DOI [10.1534/genetics.104.030478](https://doi.org/10.1534/genetics.104.030478).
- Tamura K, Nei M. 1993.** Estimation of the number of nucleotide substitutions in the control region of the mitochondrial DNA in humans and chimpanzees. *Molecular Biology and Evolution* **10**:512–526.
- Tamura K, Stecher G, Peterson D, Filipowski A, Kumar S. 2013.** MEGA6: Molecular Evolutionary Genetics Analysis version 6.0. *Molecular Biology and Evolution* **30**:2725–2729 DOI [10.1093/molbev/mst197](https://doi.org/10.1093/molbev/mst197).
- Templeton AR, Crandall KA, Sing CF. 1992.** A cladistic analysis of phenotypic associations with haplotypes inferred from restriction endonuclease mapping and DNA sequence data. III. Cladogram estimation. *Genetics* **132**:619–633.
- Ting C-T, Tsauro S-C, Wu C-I. 2000.** The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. *Proceedings of the National Academy of Sciences of the United States of America* **97**:5313–5316 DOI [10.1073/pnas.090541597](https://doi.org/10.1073/pnas.090541597).
- Traylor T, Birand AC, Marshall JL, Howard DJ. 2008.** A zone of overlap and hybridization between *Allonemobius socius* and a new *Allonemobius* sp. *Annals of the Entomological Society of America* **101**:30–39 DOI [10.1603/0013-8746\(2008\)101\[30:AZOOAH\]2.0.CO;2](https://doi.org/10.1603/0013-8746(2008)101[30:AZOOAH]2.0.CO;2).
- Turner LM, Chuong EB, Hoekstra HE. 2008.** Comparative analysis of testis protein evolution in rodents. *Genetics* **179**:2075–2089 DOI [10.1534/genetics.107.085902](https://doi.org/10.1534/genetics.107.085902).
- Uda K, Fujimoto N, Akiyama Y, Mizuta K, Tanaka K, Ellington WR, Suzuki T. 2006.** Evolution of the arginine kinase gene family. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **1**:209–218.
- Visconti PE, Krapf D, De la Vega-Beltrán JL, Acevedo JJ, Darszon A. 2011.** Ion channels, phosphorylation and mammalian sperm capacitation. *Asian Journal of Andrology* **13**:395–405 DOI [10.1038/aja.2010.69](https://doi.org/10.1038/aja.2010.69).

- Wagstaff BJ, Begun DJ. 2005.** Molecular population genetics of accessory gland protein genes and testis-expressed genes in *Drosophila mojavensis* and *D. arizonae*. *Genetics* 171:1083–1101 DOI [10.1534/genetics.105.043372](https://doi.org/10.1534/genetics.105.043372).
- Walters JR, Harrison RG. 2010.** Combined EST and proteomic analysis identifies rapidly evolving seminal fluid proteins in *Heliconius* butterflies. *Molecular Biology and Evolution* 27:2000–2013 DOI [10.1093/molbev/msq092](https://doi.org/10.1093/molbev/msq092).
- Wang R-L, Hey J. 1996.** The speciation history of *Drosophila pseudoobscura* and close relatives: inferences from DNA sequence variation at the *period* locus. *Genetics* 144:1113–1126.
- Wolfner MF. 2011.** Precious essences: female secretions promote sperm storage in *Drosophila*. *PLoS Biology* 9:e1001191 DOI [10.1371/journal.pbio.1001191](https://doi.org/10.1371/journal.pbio.1001191).
- Wu C-I. 2001.** The genic view of the process of speciation. *Journal of Evolutionary Biology* 14:851–865 DOI [10.1046/j.1420-9101.2001.00335.x](https://doi.org/10.1046/j.1420-9101.2001.00335.x).
- Wu C-I, Ting C-T. 2004.** Genes and speciation. *Nature Reviews Genetics* 5:114–122 DOI [10.1038/nrg1269](https://doi.org/10.1038/nrg1269).
- Yang Z, Bielawski JP. 2000.** Statistical methods for detecting molecular adaptation. *Trends in Ecology & Evolution* 15:496–503 DOI [10.1016/S0169-5347\(00\)01994-7](https://doi.org/10.1016/S0169-5347(00)01994-7).
- Zhang Z, Hambuch TM, Parsch J. 2004.** Molecular evolution of sex-biased genes in *Drosophila*. *Molecular Biology and Evolution* 21:2130–2139 DOI [10.1093/molbev/msh223](https://doi.org/10.1093/molbev/msh223).
- Zhang Z, Parsch J. 2005.** Positive correlation between evolutionary rate and recombination rate in *Drosophila* genes with male-biased expression. *Molecular Biology and Evolution* 22:1945–1947 DOI [10.1093/molbev/msi189](https://doi.org/10.1093/molbev/msi189).
- Zhou G, Somasundaram T, Blanc E, Parthasarathy G, Ellington WR, Chapman MS. 1998.** Transition state structure of arginine kinase: implications for catalysis of bimolecular reactions. *Proceedings of the National Academy of Sciences of the United States of America* 95:8449–8454 DOI [10.1073/pnas.95.15.8449](https://doi.org/10.1073/pnas.95.15.8449).
- Zmasek CM, Eddy SR. 2011.** A simple algorithm to infer gene duplication and speciation events on a gene tree. *Bioinformatics* 17:821–828.