doi:10.1111/j.1558-5646.2009.00767.x

GENEALOGICAL DISCORDANCE AND PATTERNS OF INTROGRESSION AND SELECTION ACROSS A CRICKET HYBRID ZONE

Luana S. Maroja,[1,2,3] Jose A. Andrés,[1,4] and Richard G. Harrison[1]

[1]*Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, New York 14853*

[3]*E-mail: marojaL@si.edu*

Received August 22, 2008
Accepted May 25, 2009

In recently diverged species, ancestral polymorphism and introgression can cause incongruence between gene and species trees. In the face of hybridization, few genomic regions may exhibit reciprocal monophyly, and these regions, usually evolving rapidly under selection, may be important for the maintenance of species boundaries. In animals with internal fertilization, genes encoding seminal protein are candidate barrier genes. Recently diverged hybridizing species such as the field crickets *Gryllus firmus* and *G. pennsylvanicus*, offer excellent opportunities to investigate the origins of barriers to gene exchange. These recently diverged species form a well-characterized hybrid zone, and share ancestral polymorphisms across the genome. We analyzed DNA sequence divergence for seminal protein loci, housekeeping loci, and mtDNA, using a combination of analytical approaches and extensive sampling across both species and the hybrid zone. We report discordant genealogical patterns and differential introgression rates across the genome. The most dramatic outliers, showing near-zero introgression and more structured species trees, are also the only two seminal protein loci under selection. These are candidate barrier genes with possible reproductive functions. We also use genealogical data to examine the demographic history of the field crickets and the current structure of the hybrid zone.

KEY WORDS: **Accessory gland, barrier genes, *Gryllus firmus*, *Gryllus pennsylvanicus*, isolation with migration, male-limited expression.**

Introgressive hybridization is now recognized as an important process in evolution (Arnold 1997) and has been documented in a variety of animal species (Wang et al. 1997; Machado et al. 2002; Besansky et al. 2003; Grant et al. 2004; Seehausen 2004; Putnam et al. 2007; Kronforst 2008), as well as in plants and prokaryotes (Grant 1981; Rieseberg 1997; Jain et al. 2002). However, alleles at some loci are not free to move across species boundaries. In these genomic regions, introgression will be limited or prevented by incompatibilities, resulting in a semipermeable species boundary (Barton and Hewitt 1981; Harrison 1990; Wu 2001). Such regions might be few in number, but they are essential for the maintenance of species boundaries in the face of hybridization (e.g., Noor et al. 2001; Machado et al. 2002; Machado and Hey 2003).

Allelic introgression violates assumptions of the basic bifurcating model of species divergence and, together with shared ancestral polymorphism, can cause incongruence between gene trees and species trees (Neigel and Avise 1986; Hudson 1992; Nichols 2001). Thus, the genome of recently diverged/hybridizing species will be a mosaic of different genealogical histories (Ting et al. 2000; Hudson and Coyne 2002; Broughton and Harrison 2003; Machado and Hey 2003; Dopman et al. 2005; Andrés et al. 2008). However, genomic regions that cannot cross species boundaries

[2]Present address: Department of Zoology, Cambridge University, Downing Street, Cambridge CB2 3EJ, United Kingdom

[4]Present address: Department of Biology, University of Saskatchewan, 112 WP Thompson Building, 112 Science Place, Sasakatoon, SK S7N 5E2, Canada

or that have experienced recent selective sweeps will exhibit reciprocal monophyly (exclusivity). Thus, searching the genome for regions showing a lack of introgression and/or species monophyly can potentially reveal so-called "speciation" or "barrier" genes (Rieseberg et al. 1999; Wu 2001; Dopman et al. 2005; Payseur and Nachman 2005; Noor and Feder 2006).

Barrier genes are involved in reproductive incompatibilities and may be evolving rapidly under selection. Recently, evolutionary geneticists have made great progress in the identification of barrier/speciation genes in model organisms. Several major effect genes, most of them under selection, have been described: *Xmrk-2* causes inviability in hybrid platyfish (Wittbrodt et al. 1989), *OdsH*, *JYAlpha*, and *Overdrive* cause hybrid male sterility in *Drosophila* (Ting et al. 1998; Wu and Ting 2004; Masly et al. 2006; Phadnis and Orr 2009) and *Hmr*, *Nup96*, and *Lhr* cause hybrid inviability in *Drosophila* (Barbash et al. 2003; Presgraves et al. 2003; Brideau et al. 2006). Analyzing patterns of introgression across the *Mus musculus* and *M. domesticus* hybrid zone, Payseur and Nachman (2005) identified seven candidate barrier genes that showed high rates of protein evolution and male-limited expression. More recently Mihola et al. (2009) identified *Prdm9*, a histone gene responsible for hybrid sterility in the house mouse. In organisms with more limited genetic resources, finding barrier genes has been more challenging. Nonetheless, candidate barrier genes have been identified using a statistical analysis of hybrid zones (e.g., Riesenberg et al. 1999; Grahame et al. 2006), analysis of gene genealogies (e.g., Dopman et al. 2005; Andrés et al. 2008), population genetics (e.g., Vasemagi et al. 2005; Nosil et al. 2008), and coalescent-based approaches (e.g., Putnam et al. 2007).

In animals with internal fertilization, genes encoding seminal proteins represent a class of rapidly evolving and often positively selected genes that are potential candidate barrier genes (Swanson and Vacquier 2002). Seminal proteins are transferred to females along with sperm during copulation and play an important role in reproductive interactions and potentially in the evolution of reproductive isolation. For example, in *Drosophila*, seminal proteins have been shown to influence female physiology and behavior, including oogenesis, ovulation, oviposition, sperm storage, and remating rates (e.g., Harshman and Prout 1994; Herndon and Wolfner 1995; Wolfner 1997; Neubaum and Wolfner 1999; Tram and Wolfner 1999). In both insects and primates, some of these proteins exhibit a clear signature of positive selection (Clark et al. 2006) and may be an important component of reproductive isolation during the early stages of the speciation process (Andrés and Arnqvist 2001).

Recently diverged species that continue to hybridize offer excellent opportunities to investigate the origins of barriers to gene exchange. The field crickets *Gryllus firmus* and *G. pennsylvanicus* are very closely related (<0.5% mtDNA divergence—Willett et al. 1997), come into contact in a well-characterized hybrid zone in eastern North America (Harrison and Bogdanowicz 1997; Ross and Harrison 2002), do not form monophyletic groups, and share ancestral polymorphisms at many loci across the genome (Harrison 1979; Broughton and Harrison 2003). Although morphologically similar (Alexander 1957), these crickets have clearly diverged in ecology (Rand and Harrison 1989; Ross and Harrison 2002, 2006) and behavior (Harrison and Rand 1989; Doherty and Storz 1992; Maroja et al. 2009). Furthermore, there is a striking unidirectional reproductive incompatibility: when mated to *G. pennsylvanicus* males, *G. firmus* females produce many fewer eggs than females mated to conspecifics, and the eggs produced (indistinguishable in size and color from unfertilized eggs) fail to develop (Harrison 1983; Maroja et al. 2008). If seminal proteins are involved in this reproductive barrier, they may show the signatures of restricted introgression in one or both directions.

In an effort to identify proteins that might be responsible for reproductive isolation, Andrés et al. (2006) characterized accessory gland genes from *G. firmus* and *G. pennsylvanicus*, many of which are rapidly evolving and under selection (see also Braswell et al. 2006). Subsequent proteomic analyses provided unambiguous identification of seminal proteins (Andrés et al. 2008). Here, we generate genealogies for six seminal protein loci, three "housekeeping" loci, and mtDNA, using extensive population sampling across both field cricket species and the hybrid zone. Using a combination of analytical approaches, we show that introgression varies strikingly across the genome. Furthermore, two nuclear loci that show a pattern consistent with absence of introgression encode seminal proteins that are under positive selection. We also use the genealogical data to interpret the demographic history of the field crickets and the current structure of the hybrid zone.

Materials and Methods

POPULATION SAMPLING

We collected *G. firmus* and *G. pennsylvanicus* from 14 populations (Fig. 1, Table 1). Six of these populations represent "pure" species: Guilford, CT (GUI, $n = 6$), Tom's River, NJ (TRI, $n = 5$), and Parksley, VA (PAR, $n = 6$) represent "pure" *G. firmus* populations whereas Ithaca, NY (ITH, $n = 6$), Scranton, PA (SCR, $n = 4$), and State College, PA (SCO, $n = 5$) represent "pure" *G. pennsylvanicus* populations. We confirmed the nonhybrid population status with phenotypic measurements. In addition to *G. firmus* and *G. pennsylvanicus*, we also sampled *G. rubens* from Durham, NC ($n = 5$) and from Roanoke, VA ($n = 2$) and *G. bimaculatus* from a colony maintained by the Hoy lab at Cornell University ($n = 2$).

GENE SEQUENCING AND ALLELE INFERENCE

In this article, we focus on one mitochondrial DNA gene (*mtDNA*), cytochrome oxidase I (including part of the adjacent tRNA), and

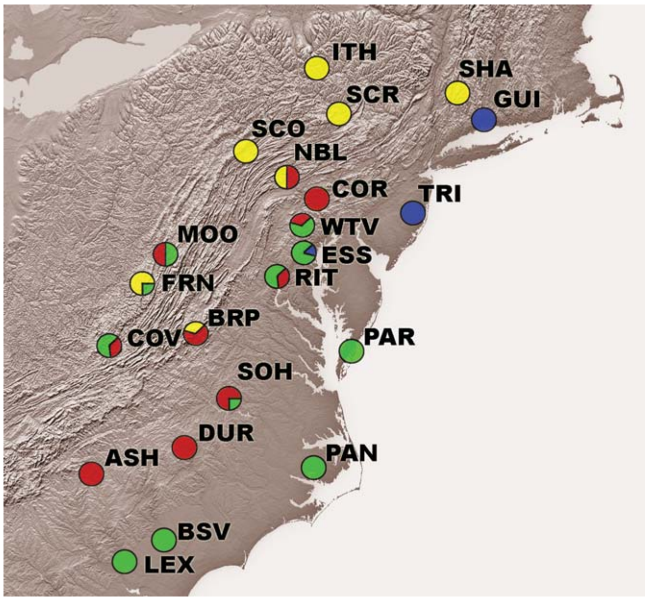

Figure 1. Collection localities. Seven populations from Willett et al. (1997) were included in the mtDNA analysis (ASH, BRP, BSV, COR, LEX, PAN, SHA, and WTV). Population colors represent clade affiliation based on the mtDNA phylogeny. Yellow and red represent the Northern and Southern *G. pennsylvanicus* clades; blue and green represent the Northern and Southern *G. firmus* clades. We used colored ovals to designate individuals in our collections that come from apparently pure species populations (blue and yellow for *G. firmus* and *G. pennsylvanicus*, respectively) and open squares to represent mixed populations.

nine nuclear autosomal genes, all of which were isolated from a field cricket male accessory gland cDNA library (Andrés et al. 2006). The nuclear genes include *Hexokinase* (*Hex*), *Elongation Factor* 1-α (*EF*1-α), *Guanylate Kinase*-1 (*GuKc*), and six anonymous loci (*AG-0005F*, *AG-0032F*, *AG-0090F*, *AG-0211F*, *AG-0254P*, and *AG-0334P*). *Hex* encodes an enzyme that phosphorylates hexose, participating in the first step of the glycolytic pathway. The protein product of *EF1*-α is an essential component of the eukaryotic transcriptional apparatus catalyzing the transfer of aminoacyl-transfer RNA to the ribosome. *GuKc* encodes an enzyme that catalyzes the ATP-dependent phosphorylation of guanosine monophosphate (GMP) into guanosine diphosphate (GDP). These genes do not show a male-biased expression, their products are not secreted (based on absence of a signal peptide), and they likely do not function in cricket reproduction. The genes encoding these proteins are therefore considered "housekeeping genes." The unidentified protein-coding loci were chosen from a pool of 39 loci, based on ability to sequence from genomic DNA combined with male-limited expression and/or the presence of signal peptide and/or presence of the protein in the male spermatophore (see Andrés et al. 2006, 2008). Of the six unidentified proteins, *AG-0254P* is secreted (has a signal peptide) and possibly has an unknown binding function, because it exhibits similarity to olfactory segment-D (OS-D) chemosensory protein. The other five seminal proteins are biochemically uncharacterized and are secreted and/or show a male-biased expression (see Table 2 and Andrés et al. 2006). Through proteomic analysis, the proteins encoded by *AG-0005F*, *AG-0090F*, and *AG-0334P* are known to

Table 1. Sampled populations; pure *G. pennsylvanicus* in bold and pure *G. firmus* in italics. Regular font represents mixed populations.

Population	Abbrev.	Latitude (N)	Longitude (W)	Elevation (m)	*n*[1]	Total bp[2]	π±SD[3]	
Ithaca, NY	ITH	42°26′01″	76°29′59″	250	10.4	7162	0.0063	0.0039
Scranton, PA	SCR	41°24′25″	75°35′46″	397	7.4	7301	0.0060	0.0039
Guilford, CT	GUI	41°16′48″	72°42′02″	0	10	6956	0.0074	0.0059
State College, PA	SCO	40°47′59″	77°52′05″	371	9.3	6675	0.0064	0.0047
New Bloomfield, PA	NBL	40°28′24″	77°07′50″	379	10.2	6938	0.0073	0.0047
Tom's River, NJ	TRI	39°45′00″	74°11′33″	0	7.8	7133	0.0064	0.0036
Essex, MD	ESS	39°18′20″	76°28′46″	0	8.5	7074	0.0084	0.0046
Moorefield, WV	MOO	39°04′09″	78°55′58″	285	8.1	7058	0.0091	0.0046
Ritchie, MD	RIT	38°52′07″	76°51′01″	0	9.8	7095	0.0073	0.0041
Franklin, WV	FRN	38°39′20″	79°19′59″	551	7.2	7548	0.0080	0.0045
Covington, VA	COV	38°00′50″	78°28′21″	354	9.6	7083	0.0088	0.0056
Parksley, VA	PAR	37°45′58″	75°36′00″	0	10.8	7118	0.0078	0.0054
South Hill, VA	SOH	36°45′07″	78°06′09″	116	7.4	6713	0.0079	0.0068
Durham, NC	DUR	36°03′23″	79°04′45″	159	0.67[4]	2717	0.0025	0.0036

[1] Average number of haplotypes sequenced.

[2] Total number of base pairs sequenced across all loci.

[3] Average π and SD across all 10 loci.

[4] Only 4 haplotypes available for AG0005F and COI.

be present in the spermatophore (see Table 2 and Andrés et al. 2008). Locus-specific information on primer sequence, number of sequenced base pairs, total number of coding nucleotides, total number of variable sites, male expression bias, and whether the protein is secreted can be found in Table 2. All sequences have been deposited in GenBank (GQ226136–GQ227267).

Genomic DNA was isolated from the leg muscle tissue using the DNeasy tissue kit (QIAGEN, Valencia, CA). Locus-specific primers (Table 2) were used to polymerase chain reaction (PCR) amplify each of the 10 loci. PCR reactions (10 μl volume) contained 3 mM $MgCl_2$, 0.2 mM dNTPs, 50 mM KCl, 20 mM Tris (pH 8.4), 2.5 ng of each primer, and 1 U of *Taq* DNA polymerase (Gibco-BRL, Invitrogen, Carlsbad, CA) and 1 μL DNA. PCR amplifications were performed under the following touchdown conditions: 10 cycles of 50 sec at 95°C, 60 sec at 65–55°C (decreasing 1°C per cycle), and 90 sec at 72°C followed by 30 cycles of 50 sec at 95°C, 60 sec at 55°C, and 90 sec at 72°C. All genes were sequenced in both directions. Sequences were aligned in SeqMan (DNASTAR, Madison, WI), and SNPs were identified by a visual inspection; only high-quality traces were considered. Individual haplotypes were reconstructed using the PHASE algorithm (Stephens et al. 2001). Excluding autapomorphies, all haplotype identifications had posterior probabilities greater than 0.8.

We sequenced at least four individuals (eight haplotypes) from each population for each locus (average number of haplotypes per locus/per population is shown in Table 1). The Durham, NC sample (DUR) included only two *G. pennsylvanicus* crickets and was sequenced only for *mtDNA* and *AG-0005F*.

PHYLOGENETIC ANALYSES

For the *mtDNA* locus, phylogeny reconstruction was carried out using MrBayes version 3.1.2 (Huelsenbeck and Ronquist 2001). Searches were run for five million generations, sampling every 100 generations and discarding trees from the first 1,000,000 generations (burn-in time). To generate trees, we used two complex models—the general time reversible model with invariant sites, gamma rates, and default priors (GTR + I + G), allowing the rate at each site to change over evolutionary history; and the model GTR + I + G using site-specific rates (SSR), with sites at each codon position and in the tRNA following a gamma distribution and allowing a proportion of sites to be invariant. Because there were no differences between topologies inferred by the two models, we only show results for the SSR model. In addition to our own sequences, we also included 27 sequences from Willett et al. (1997). The phylogenetic tree was rooted using seven *G. rubens* sequences.

We reconstructed nuclear gene trees using the neighbor-joining algorithm; all methods and results of these analyses are included in Supporting Information (Figs. S1–S10; Table S1). We tested the homogeneity of the phylogenetic signal among

Table 2. Locus-specific information, including primers used, whether gene products are secreted and/or have male-biased gene expression.

Loci	5′-3′Primers forward	5′-3′Primers reverse	Secr[1]	♂ bias[2]	Total[3]	Cod[4]	n[5]	Var[6]
mtDNA	ACCCCCATCATTAACCCTTTTA	GAGACCATTACTTGCTTTCAGTCATCT*	No	No	1889	1182	70**	187
EF1-α	CGAAATCGCCTAACAAACATAACA	AATCCTTTCCTCTTGCGTGTG	No	No	727	372	132	61
GuKc	GCTGCTAATCGCGGAAGTGC	GCTGCCTTGCTTGTGCCATAC	No	No	434	156	126	17
Hex	AATGGGGAGCTTTCGGAGAT	CATTGGCACAGTTTTGGTCAG	No	No	460	282	126	18
AG-0005F	GATGAGGCTGCTGGTCGTCGTG	GTGGTTAGCAGGGGCGTGATGGTT	Sp	Yes	911	911	140	94
AG-0032F	GGCACTGGCCAGTTGGACAC	AAATTAATAAAACACATTTGAGTGTTAATAATAC	No	Yes	471	–	120	18
AG-0090F	AGGAATAATCGCTTTTGCCACTG	CCTCTTGATATGTCTTGCGAAATG	Sp	No	654	–	130	62
AG-0211F	TCGAGTTGGACGAGAGCTGTTACG	ATTTGTGCTATTCGTTTGTCACTG	No	Yes	412	–	130	81
AG-0254P	GTCACCGAGCTACAAAACAACACG	TCTCTTGATATGCTCGCCTTTCTC	SgP	Yes	740	147	132	94
AG-0334P	TGCTGCGAATATGGAGGAG	CATGGTGCTTTTCGTGCTCTT	Sp	Yes	1057	927	118	126

*Primer "3372" from Simon et al. (1994).

**including 27 haplotypes from Willett et al. 1997.

[1] If protein is likely to be secreted. Sp, present in spermatophore; SgP, signal peptide present.

[2] If expression is male biased (i.e., expressed in male accessory gland tissues but not in female abdomen tissues).

[3] Total number of nucleotides (bp) sequenced.

[4] Length of coding region (for some loci only intronic or other noncoding regions were sequenced).

[5] Total number of haplotypes sequenced.

[6] Number of variable sites across all samples.

the nine nuclear loci using the partition homogeneity test (Farris et al. 1995). We combined data from the 47 individuals that were sequenced for all nine loci and performed the test with 1000 replicates.

MOLECULAR POPULATION GENETICS

We used DNAsp version 4.20.2 (Rozas et al. 2003) for a basic polymorphism analysis. Indels were not included in these analyses. Analyses of molecular variance (AMOVA, Excoffier et al. 1992) for pure species populations (i.e., GUI, ITH, PAR, SCO, SCR, TRI) were conducted using Arlequin version 2.000 (Schneider et al. 2000). Tajima's D (Tajima 1989) was calculated to test for departures from neutrality with DNAsp version 4.20.2. This test is based on the expectation that under mutation–drift equilibrium θ and π should be the same parameter (i.e., $4N_e\mu$). Tajima's D can detect signatures of recent demographic events, such as population expansion (excess of low frequency polymorphisms leading to negative Tajima's D values), and/or selective events (selective sweeps, negative Tajima's D values). However, by testing many loci it is possible to distinguish between the demographic and selective scenarios because a population expansion is expected to affect the entire genome, whereas selection should only affect the selected locus and adjacent (hitchhiked) regions.

TEST OF SELECTION

The relative rate of fixation of nonsynonymous (d_N) and synonymous (d_S) substitutions provides an estimate of selection pressures acting on a given protein. For any set of amino acid residues, when $d_N/d_S = \omega = 1$, a neutral model of evolution cannot be rejected, whereas $\omega < 1$ indicates purifying selection, and $\omega > 1$ indicates positive selection. Although the selection parameter ω is commonly calculated using phylogenetic likelihood methods (Goldman and Yang 1994), these methods are unreliable in the presence of recombination because this process leads to not one, but multiple evolutionary trees along the gene sequence (Anisinova et al. 2003; Wilson and McVean 2006). In this article, we used the method recently developed by Wilson and McVean (2006) to calculate ω in the presence of recombination. This method relaxes the assumption of a single common history for all codons, and performs Bayesian inferences of ω using a population genetics approximation to the coalescent with recombination (Hudson 1983; Li and Stephens 2003). One disadvantage of this method is that it does not provide estimates of d_N and d_S. Using OmegaMap (Wilson and McVean 2006), we estimated the selection parameter (ω), recombination rate (ρ), transition-transversion ratio (κ), and the rate of mutation μ for each gene for which we sequenced all or part of the coding region (*EF1*-α, *GuKc*, *Hex*, *AG-0005F*, *AG-0254P*, and *AG-0334P*). The number of coding (and variable) sites analyzed for each locus was 372 (13) for *EF1*-α, 156 (5) for *GuKc*, 282 (7) for *Hex*, 816 (80) for *AG-0005F*, 147 (8) for *AG-0254P*, and 924 (78) for *AG-0334P*. For the other three genes, only intronic regions were sequenced (see Table 3). We used improper inverse prior distributions for all parameters with means $\omega = 1$, $\rho = 0.07$, $\kappa = 3.6$, $\mu = 0.3$. Both ω and ρ were modeled as constant (i.e., all sites are assumed to share common values). The frequency of codons was assumed to be equal, and the number of alignment orderings was set to 10. We ran at least 250,000 iterations with a 10,000 burn-in and a thinning of 100. For each gene, two independent convergent runs were merged to provide the posterior distributions of the estimated parameters. The effective sample size (ESS) for the estimated parameters was always >100, suggesting that the MCMC chains were run long enough to obtain accurate estimates. Swanson et al. (2004) showed that statistical evidence for adaptive evolution at some codons could be found in most genes having overall gene $d_N/d_S > 0.5$ (see also Almeida and DeSalle 2008). Thus, for each distribution of ω values we calculated the mode and posterior probability of selection with a cutoff at $\omega > 0.5$. In addition to ω, we also calculated the d_N/d_S ratio using the Nei and Gojobori (1986) equation as implemented in DNAsp version 4.20.2 (Rozas et al. 2003).

ISOLATION AND INTROGRESSION

We calculated migration rates between pure species populations (GUI, TRI, and PAR for *G. firmus* and ITH, SCR, and SCO for *G. pennsylvanicus*) as a proxy for gene introgression across the hybrid zone. Because genomes are mosaics and both species exclusivity and introgression are locus specific, we initially tested each locus independently. To discriminate between the relative effects of divergence and gene flow, we analyzed the loci under the isolation-with-migration analytic (IMa) model (Nielsen and Wakeley 2001; Hey and Nielsen 2004, 2007). This model assumes that an ancestral population splits into two populations (species) with gene flow possibly continuing between the diverging populations. Because the IMa model assumes no recombination we used only the longest nonrecombining region for each locus. Nonrecombining regions were inferred using the algorithms implemented in the IMgc software package (Woerner et al. 2007). To fit the IMa model to data, we used a Bayesian coalescent method that approximates the integration over the possible genealogies using a Markov chain Monte Carlo (MCMC) simulation. This method estimates marginal and posterior probability distributions for demographic parameters including directional migration rates scaled by a mutation rate for the entire locus ($m1 = \mathrm{m}_1/\mu_i$ and $m2 = \mathrm{m}_2/\mu_i$), divergence time scaled by mutation ($t = t\mu_i$), and effective population sizes of the two species and the ancestral population (Hey and Nielsen 2004, 2007). We obtained asymmetric estimates of migration rates between species (effective number of migrants per generation, $2N_e m_i$) from the product of m_i and $\theta_i/2$. We conducted the analysis using the Hasegawa–Kishino–Yano

Table 3. Polymorphism statistics for each gene, based on only pure populations of each species (*G. pennsylvanicus*: ITH, SCO, SCR; *G. firmus:* GUI, PAR, TRI).

Locus	Species	*n*[1]	L[2]	Cod[2]	S[3]	Syn[4]	Rep[4]	π[5]	Dxy[6]	D[7]	Rm[8]
mtDNA	*firmus*	16	1825	1119	25	10	3	0.00342	0.00576	−0.6971	–
	pennsylvanicus	14	1834	1125	10	3	2	0.00119		−1.1859	–
EF-1α	*firmus*	20	713	369	20	3	0	0.00927	0.00871	1.1396	0
	pennsylvanicus	18	716	372	18	4	0	0.00361		−2.1637*	0
GuKc	*firmus*	30	431	153	7	2	1	0.00415	0.00410	−0.3438	1
	pennsylvanicus	24	432	156	4	1	1	0.00263		0.1699	0
Hex	*firmus*	34	460	282	10	2	2	0.00314	0.00436	−1.2695	0
	pennsylvanicus	26	460	282	2	0	0	0.00121		0.1341	0
AG-0005F	*firmus*	30	875	875	32	9	23	0.00765	0.01671	−0.6217	3
	pennsylvanicus	26	871	871	31	15	16	0.00733		−0.7963	4
AG-0032F	*firmus*	26	444	0	4	–	–	0.00192	0.00483	−0.9905	0
	pennsylvanicus	24	457	0	9	–	–	0.00557		0.1863	2
AG-0090F	*firmus*	22	621	0	27	–	–	0.01196	0.01365	0.0113	5
	pennsylvanicus	30	570	0	22	–	–	0.01250		0.9952	8
AG-0211F	*firmus*	28	409	0	33	–	–	0.01279	0.01158	−1.5466	2
	pennsylvanicus	28	409	0	30	–	–	0.00976		−1.7644	0
AG-0254P	*firmus*	24	543	147	30	1	5	0.01218	0.01336	−0.0665	6
	pennsylvanicus	28	550	144	25	1	4	0.01245		0.0909	6
AG-0334P	*firmus*	28	1049	915	49	9	36	0.00794	0.00912	−1.3263	6
	pennsylvanicus	30	1049	909	35	8	22	0.00595		−1.0753	4

[1] Number of haplotypes analyzed.

[2] Length of sequence and coding region analyzed.

[3] Total number of polymorphic sites.

[4] Number synonymous and replacement changes.

[5] Average number of nucleotide differences per site.

[6] Average number of nucleotide substitutions per site between species.

[7] Tajima's statistic (1989). *$P<0.01$.

[8] Minimum number of recombination events per locus.

(HKY) model and uninformative prior distributions of parameters. To improve mixing, we used a geometric heating scheme with 50–80 parallel chains. At least 25,000 genealogies were sampled from the primary chain after a 2–5 h burn-in. We replicated each analysis at least three times and all replicates yielded nearly identical estimates. Convergence upon the stationary distribution was assessed by estimating the ESS and autocorrelation of parameter values measured over the course of the run. The analysis was considered to have converged upon a stationary distribution if the independent runs generated similar posterior distributions (Hey 2005), with a minimum ESS of 100 (Kuhner and Smith 2007). For credibility intervals, we report the 90% highest posterior density (HPD) interval, which includes 90% of the probability density of a parameter.

To test for differences between $m1$ and $m2$ in nuclear genes, we used a likelihood-ratio test of nested models (L mode option in IMa). We separated loci into those not under selection and those likely under selection (*AG-0005F* and *AG-0334P*—see Results). We ran two analyses, one combining all nuclear loci except for *AG-0005F* and *AG-0334P* and the other combining the two loci under selection (i.e., *AG-0005F* and *AG-0334P*). Our aim was to test if introgression rates for neutral loci vary in directionality and to see how introgression rates of loci encoding reproductive proteins under selection differ from those of neutral loci. Although IMa assumes no selection, our intention was not to calculate actual migration rates but to compare relative introgression rates between different sets of genes. Here, we use migration as an approximation to introgression rates between parts of the genome subjected to different evolutionary pressures.

Results

PHYLOGENETIC ANALYSES

For the *mtDNA* data, the Bayesian analysis, using the GTR + I + G model or the SSR model, produced a tree with six major haplotype groups (Fig. 2). These six clades correspond to the haplotype groups identified by Willett et al. (1997), except that one of the previously identified groups (northern *G. firmus*) is further subdivided into three clades in our topology. We, thus, use the nomenclature of Willett et al. (1997), combining the three basal

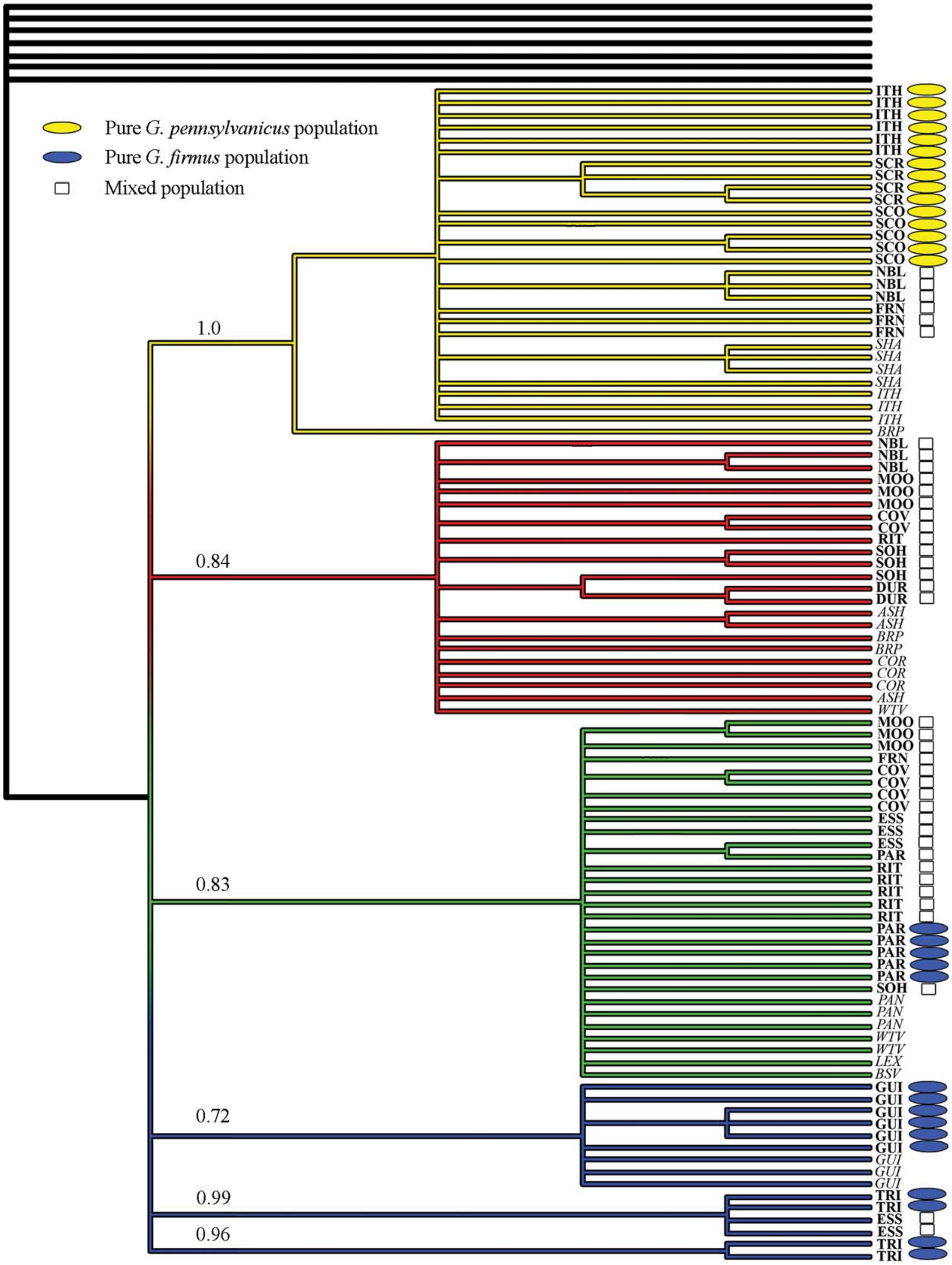

Figure 2. Posterior probability tree of *mtDNA*, partitioned by site-specific rates (SSR). The search with the program MrBayes (Huelsenbeck and Ronquist 2001) was run for five million generations, discarding the first one million generations. We used default priors, a GTR model, invariant sites, and gamma rates. Haplotypes in italics are from Willett et al. (1997). Yellow and red represent northern and southern *G. pennsylvanicus* clades and blue and green represent northern and southern *G. firmus* clades. Yellow and blue ovals represent individuals from pure populations of each species and squares represent mixed hybrid populations.

clades into a northern *G. firmus* group (Fig. 2). We refer to groups as (1) northern *G. pennsylvanicus* (2) southern *G. pennsylvanicus*, (3) northern *G. firmus*, and (4) southern *G. firmus*. The mtDNA tree (Fig. 2) includes sequences produced by Willett et al. (1997), representing additional "pure" southern *G. pennsylvanicus* (COR, ASH), pure northern *G. pennsylvanicus* (SHA), and "pure" southern *G. firmus* (BSV, LEX, PAN) (see Fig. 1). In the *mtDNA* phylogeny, we color each of the four major groups (green and blue for *G. firmus* and red and yellow for *G. pennsylvanicus*) and use these colors in Figure 1 to represent the percentage of crickets belonging to each *mtDNA* clade in each of the populations.

The nuclear gene genealogies are shown in Supporting Information (Figs. S1–S10). Only *EF1*-α and *AG-0005F* exhibited strong bootstrap support (>70%), and only *AG-0005F* revealed a clear separation of pure species *G. firmus* and *G. pennsylvanicus* haplotypes.

MOLECULAR POPULATION GENETICS

Polymorphism analyses for *G. firmus* and *G. pennsylvanicus*, using only pure populations (*G. firmus*: GUI, PAR, and TRI; *G. pennsylvanicus*: ITH, SCO, and SCR), are summarized in Table 3. In general, *G. firmus* has more nucleotide variation suggesting larger population sizes and perhaps more ancient populations (see Discussion). The most noteworthy observation for nuclear genes is the large number of replacement substitutions for *AG-0005F* and *AG-0334P*. Tajima's *D* was not significant for any locus/species or locus/population combination except *EF1*-α/*G. pennsylvanicus*, which had a significant negative value in *G. pennsylvanicus* (Table 3). Signs of demographic expansion were not evident as there were no consistent trends toward positive or negative Tajima's *D* values across loci.

For most nuclear genes, the average nucleotide difference per site (π) for at least one of the species was equal to or greater than the average difference per site between species (*D*xy, Table 3). Only *Hex*, *AG-0005F*, and *AG-0334P* had substantially higher *D*xy values than π values. There were no diagnostic sites at the species level for any of the nuclear loci, although *mtDNA* exhibits fixed differences between clades within and between species.

AMOVA analyses (Excoffier et al. 1992) showed that almost all variation is due to within population variation (all F_{ST} covariance components are significant). Even for *mtDNA*, only 14% of the variation can be attributed to differences between species, and its associated covariance component (F_{CT}) is not significant (Table 4). For nuclear loci, only *Hex*, *AG-0005F*, and *AG-0334P* have more than 20% of the total variation attributed to between species variation; however none of the between species covariance components (F_{CT}) are significant. In general, there is very little population structure within species; most loci have less than 10% of the total variation attributable to among population

Table 4. Analyses of molecular variance (AMOVA) and hierarchical analyses for pure populations of *G. firmus* and *G. pennsylvanicus*.

Source of variation (%)	*mtDNA*	*EF1*-α	*GuKc*	Hex	*AG-0005F*	*AG-0032F*	*AG-0090F*	*AG-0211F*	*AG-0254P*	*AG-0334P*
Among species	14.13	11.94	16.16	47.74	43.20	18.56	5.33	3.93	9.99	22.39
Among populations within species	8.77	14.78	3.75	5.60	6.04	1.62	7.09	4.24	6.10	3.66
Within populations	77.10	73.28	80.09	46.66	50.76	79.82	87.58	91.83	83.91	73.95
Fixation indices										
F_{CT} (species/total)	0.141	0.119	0.162	0.477	0.432	0.186	0.053	0.039	0.100	0.224
F_{SC} (population/species)	0.102	0.168*	0.045	0.107*	0.106**	0.020	0.075*	0.044	0.068*	0.047
F_{ST} (population/total)	0.229*	0.267**	0.199**	0.533***	0.492***	0.202***	0.124**	0.082*	0.161***	0.261***

*$P<0.05$.
**$P<0.01$.
***$P<0.001$.

variation. *EF1-α* is an exception to this pattern with 15% of the total variation attributed to among populations within species variation. This pattern seems to be caused by two very different alleles with very different proportions in northern and southern populations (see Table 4).

TEST OF SELECTION

To estimate selection, we used sequences from all populations, but we could only carry out the test for the six nuclear loci for which sequences from coding regions were available (three housekeeping genes, *EF1-α*, *GuKci* and *Hex*, and three genes encoding seminal fluid proteins, *AG-0005F*, *AG-0254P*, and *AG-0334P*).

The d_N/d_S ratios for the six genes were: *EF1-α* = 0.02, *GuKc* = 0.24, *Hex* = 0.36, *AG-0005F* = 0.65, *AG-0254P* = 0.55, and *AG-0334P* = 1.21. Of our loci, *AG-0334P* and *AG-0005F* had overall d_N/d_S ratios substantially higher than 0.5 and are thus candidates to be under selection (Swanson et al. 2004; Almeida and DeSalle 2008). However, these d_N/d_S ratios are probably inaccurate because the Nei and Gojobori (1986) d_N/d_S calculation does not take into account recombination, and nuclear loci have experienced recombination. Recombination can cause a high number of false positives in d_N/d_S ratios (Anisimova et al. 2003; Shriner et al. 2003), because trees from recombining sequences will have longer terminal branches and smaller time to the most recent common ancestor (Schierup and Hein 2000).

Because ω distributions estimated with OmegaMap (Wilson and McVean 2006) are not normal (Fig. 3), here we report the mode for each locus, which, in this case, is more representative of a maximum-likelihood estimate. Mode values for the six genes were: *EF1-α* = 0.02, *GuKc* = 0.11, *Hex* = 0.24, *AG-0005F* = 0.65, *AG-0254P* = 0.43, and *AG-0334P* = 1.04. Again the only loci with $\omega > 0.5$ are *AG-0005F* and *AG-0334P* (Fig. 3). The probability of selection was greater than 90% only for *AG-0005F* and *AG-0334P* (0.91 and 1.00, respectively). The probabilities of selection for the other loci were zero for *EF1-α*, 0.36 for *Hex*, 0.20 for *GuKc*, and 0.69 for *AG-0254P*.

ISOLATION AND INTROGRESSION

To calculate directional migration rates between pure *G. firmus* and *G. pennsylvanicus* populations, we selected only nonrecombining regions of each gene, using at least 17 haplotypes per species (average of 25 for *G. firmus* and 23 for *G. pennsylvanicus*). The number of sites (and variable sites) analyzed for each locus was 1767 (40) for *mtDNA*, 536 (33) for *EF1-α*, 319 (9) for *GuKc*, 460 (11) for *Hex*, 578 (24) for *AG-0005F*, 443 (8) for *AG-0032F*, 390 (18) for *AG-0090F*, 320 (39) for *AG-0211F*, 284 (23) for *AG-0254P*, and 797 (40) for *AG-0334P*.

The directional migration rate *m*1 represents migration forward in time from *G. firmus* to *G. pennsylvanicus* and *m*2 represents migration forward in time from *G. pennsylvanicus* to *G. firmus*. The overall pattern is one of variation in migration rates

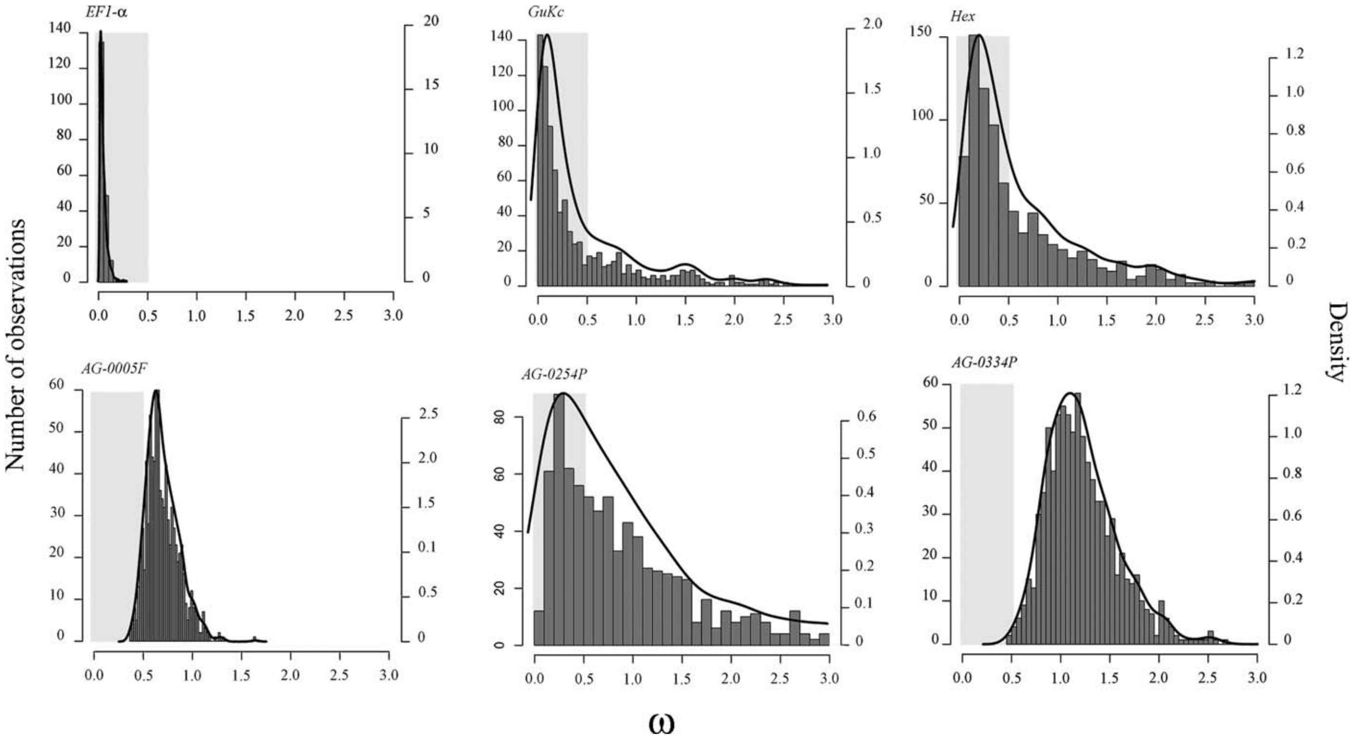

Figure 3. Posterior probability estimates of ω for each locus. Gray shading shows ω values below 0.5 that are unlikely to indicate selection. All loci except for *AG-0005F* and *AG-0334P* have point estimates of ω below 0.5 (see text).

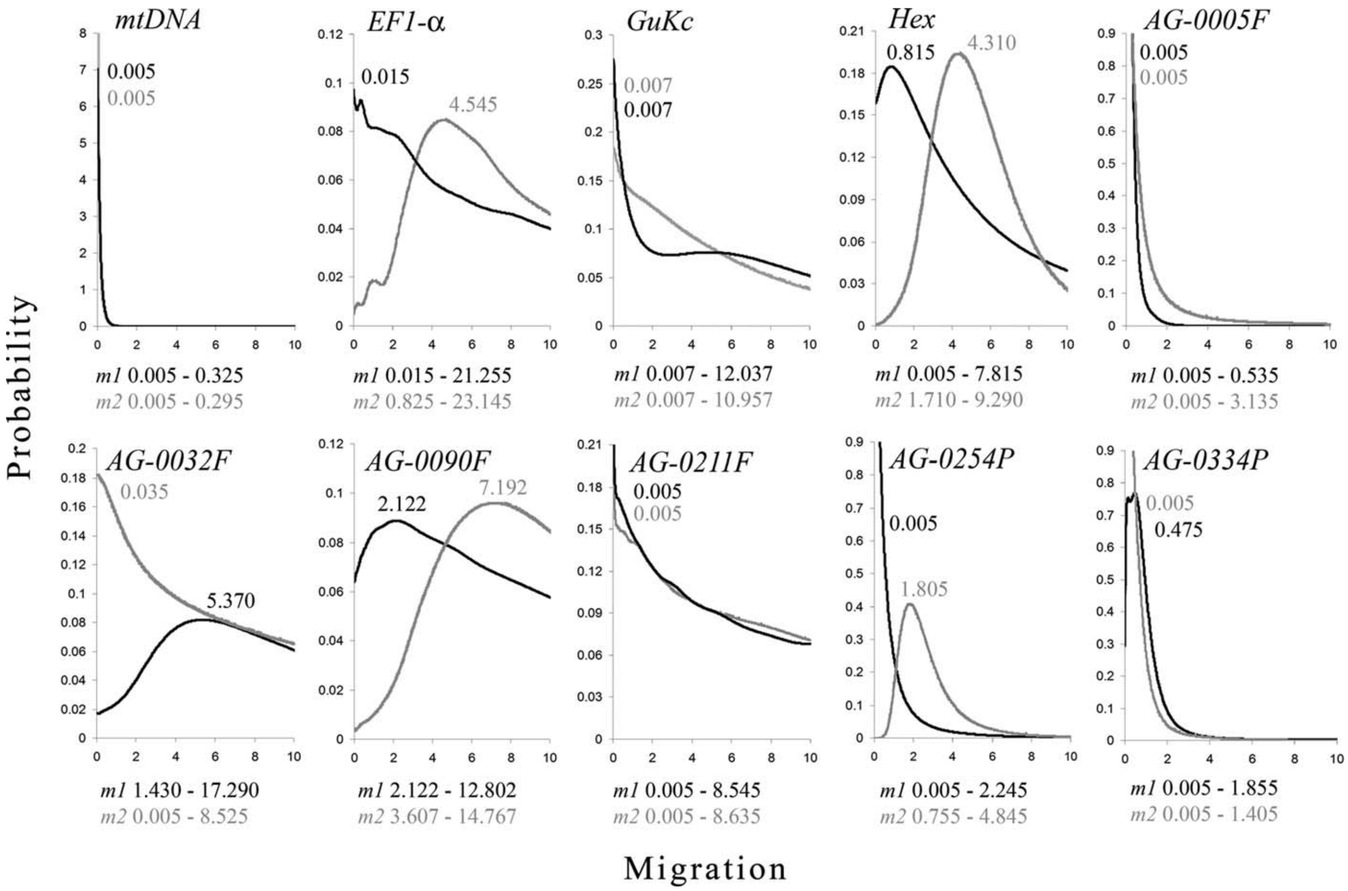

Figure 4. Posterior probability estimates of migration parameters (scaled by mutation rate) between *G. firmus* and *G. pennsylvanicus*. Black line shows *m*1 (migration forward in time from *G. firmus* to *G. pennsylvanicus*) and gray line shows reverse migration (*m*2) for each of the analyzed loci (in *mtDNA* the lines are superimposed). Numbers above lines indicates the maximum-likelihood value. Numbers below each graph show the 90% highest posterior density (HPD) intervals of migration rates *m*1 and *m*2.

across loci (Fig. 4). The two loci that appear to be under selection, *AG-0005F* and *AG-0334P*, have near-zero migration rates in both directions (Fig. 4 and Table 5). For most other nuclear loci, both *m*1 and *m*2 are positive, with *m*2, the migration from *G. pennsylvanicus* into *G. firmus*, higher than *m*1 (Table 5). For *mtDNA*, there was effectively no migration in either direction, although clear evidence of asymmetric introgression of *G. pennsylvanicus* mtDNA into *G. firmus* has been documented in populations within and immediately adjacent to the hybrid zone (Harrison et al. 1987; Harrison and Bogdanowicz 1997; Ross and Harrison 2002). In contrast to these earlier studies, the spatial scale examined here is much larger, and most tested populations are located far from the hybrid zone.

We used the nested model likelihood-ratio statistics (Hey and Nielsen 2007) to test for differences between *m*1 and *m*2. To do this, we combined all nuclear loci excluding *AG-0005F* and *AG-0334P* (Fig. 5). We also tested migration rates using only *AG-0005F* and *AG-0334P* (Fig. 5), to estimate "realized gene flow" for these loci. For neutral nuclear loci, the model with identical migration rates $m1 = m2$ was significantly rejected ($-2\Lambda = 15.39$, df = 1, $P < 0.001$), implying that *m*2 is actually higher than *m*1. For the loci under selection, there was no significant difference between migration rates ($-2\Lambda = 3.82$, df = 1, $P > 0.05$) and their migration maximum-likelihood estimates were very close to zero (Fig. 5 and Table 5).

To get an estimate of effective population sizes, we calculated θ for the neutral loci using IMa (Figs. S12, S13). Using a rough mutation estimate for nuclear loci of 10^{-9} per site/generation, the estimated effective population sizes were large; 2.6 million for *G. firmus* ($\theta = 3.85$) and 1.5 million for *G. pennsylvanicus* ($\theta = 2.25$) (Fig. S11). To get an estimate of time since divergence, we used the mtDNA data, assumed 1.2% divergence per million years per lineage (Brower 1994), and calculated time since split (*t*) with IMa. The estimated time since divergence was 202,320 years, assuming one generation per year ($t/\mu = 4.29$).

Discussion

Individuals within a species are thought to share defining properties that are not easily disturbed by hybridization and gene

Table 5. Effective migration rate and maximum-likelihood estimates of theta for *G. pennsylvanicus* and *G. firmus,* 90% highest posterior density estimates are shown in parenthesis.

Loci	$2n_1m_1$[1]	$2n_2m_2$[2]	θ_1[3]	θ_2[4]
mtDNA	0 (0–2.712)	0 (0–6.810)	16.69	46.17
EF1-α	0.007 (0.007–9.235)	292.406 (53.077–1489.049)	0.869	128.67
GuKc	0.003 (0.003–4.271)	0.001 (0.001–1.864)	0.710	0.340
Hex	0.048 (0–0.461)	29.33 (12.351–63.218)	0.118	13.610
AG-0005F	0 (0–1.285)	0 (0–0.907)	4.805	0.579
AG-0032F	5.205 (1.386–16.759)	0.005 (0–1.901)	1.939	0.446
AG-0090F	0.718 (0.718–4.334)	21.909 (10.987–44.982)	0.677	6.092
AG-0211F	0 (0–57.354)	0 (0–35.256)	13.424	8.166
AG-0254P	0 (0–1.876)	47.186 (19.737–126.658)	1.671	52.834
AG-0334P	1.197 (0–4.667)	0 (0–3.217)	5.042	4.580
Selected loci	0.085 (0–1.229)	0.331 (0–0.932)	4.869	2.323
Neutral loci	0 (0–0.775)	4.077 (2.111–6.563)	2.212	3.855

[1] Effective rate at which genes come into *G. pennsylvanicus*, per generation.
[2] Effective rate at which genes come into *G. firmus*, per generation.
[3] Estimate of θ (4Nμ) for *G. pennsylvanicus*.
[4] Estimate of θ (4Nμ) for *G. firmus*.

introgression (Templeton 1994; Coyne and Orr 2004). However, random sorting of ancestral polymorphism and differential introgression will cause recently diverged species to be mosaics with respect to molecular genealogies (Ting et al. 2000). These species will share alleles throughout much of their genomes. The apparent conflict between a unique species identity and widespread allele sharing disappears when we consider speciation models in which relatively few loci are responsible for the barriers to gene exchange and for species divergence. Because so-called "speciation genes" or "barrier genes" may often experience strong natural selection and are unable to cross species boundaries, they will become fixed or almost fixed in each species. It is thus expected that, across the genome of closely related species, genes will show different patterns of variation depending on their contribution to reproductive barriers, the nature of selection, and linkage relationships and recombination rates.

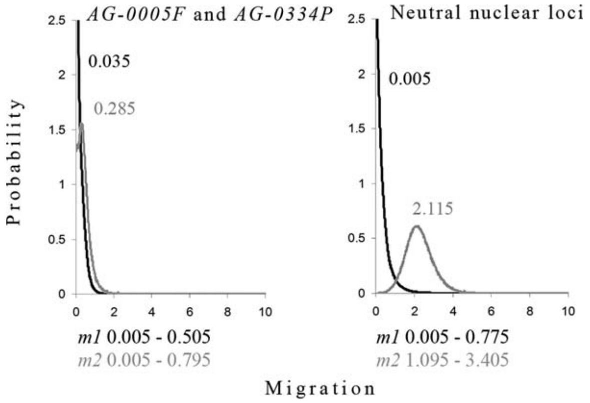

Figure 5. Joint posterior probability estimates of migration parameters (scaled by mutation rate) between *G. firmus* and *G. pennsylvanicus* for the two loci under selection (*AG-0005F* and *AG-0334P*) and for all other nuclear loci combined. Black line shows *m*1 (migration forward in time from *G. firmus* to *G. pennsylvanicus*) and gray line shows reverse migration (*m*2) for each of the analyzed loci. Numbers above lines indicates the maximum-likelihood value. Numbers below each graph show the 90% highest posterior density (HPD) intervals of migration rates *m*1 and *m*2.

PATTERNS OF INTROGRESSION AND SELECTION

Because alleles under selection may not be able to move freely across species boundaries, introgression rate estimates for alleles at strongly selected loci should be near-zero in one or both directions. The accessory gland expressed genes *AG-0005F* and *AG-0334P* have near-zero introgression estimates with narrow 90% highest posterior densities (Fig. 4). Both *AG-0005F* and *AG-0334P* also have ω values substantially greater than 0.5 and probability of selection greater than 90%.

AG-0005F and *AG-0334P* encode proteins found in the spermatophore and likely transferred from males to females during mating (Andrés et al. 2008). Given the reproductive functions of the proteins encoded by *AG-0005F* and *AG-0334P*, it is unlikely that they play a role in adaptations of crickets to local environments (e.g., adaptation to sand vs. loam soils [Rand and Harrison 1989; Ross and Harrison 2002, 2006]). More likely they play a role in sperm competition, gametic compatibility, and/or the ability of the male to alter female reproductive physiology or behavior. In insects, many accessory gland proteins exhibit a clear signature of selection (Aguadé 1998, 1999; Begun et al. 2000; Swanson et al. 2001; Swanson and Vacquier 2002; Andrés et al. 2006), and some of these proteins have been shown to influence female oogenesis, ovulation, and oviposition (Wolfner 1997; Neubaum and Wolfner 1999; Tram and Wolfner 1999). The proteins encoded by *AG-0005F* and *AG-0334P* show radical amino acid substitutions between species that may contribute to functional differences. The presence of these proteins in the spermatophore suggests a possible role in the *G. firmus* and *G. pennsylvanicus* one-way reproductive incompatibility. Such a role could explain the absence of introgression from *G. pennsylvanicus* into *G. firmus*, because *G. pennsylvanicus* alleles would compromise the ability of *G. firmus* males to produce progeny with conspecific females.

We estimated directional migration rates between pure *G. firmus* and *G. pennsylvanicus* populations as a proxy for gene introgression across the hybrid zone. Although the analytical tool used was not designed for this type of data, and although we violated one of the IMa assumptions (neutrality), the effects of these violations are not yet fully understood. As multilocus data for hybridizing species become more common, we will require a new generation of analytical tools able to provide reliable estimates of per locus gene flow for natural populations. With these caveats in mind, for the 10 loci that we assayed (nine nuclear and one mitochondrial), introgression rates between the two species varied both in magnitude and direction (Fig. 4). The two directional introgression rates for neutral loci are substantially different ($P < 0.0001$), with $m2$, the introgression rate forward in time from *G. pennsylvanicus* into *G. firmus*, significantly greater than $m1$. Thus, *G. pennsylvanicus* alleles are flowing into *G. firmus*, but gene flow in the other direction is low or absent. This asymmetry is expected for *mtDNA*. As a result of the one-way incompatibility, all F1 hybrids carry *G. pennsylvanicus mtDNA*. Furthermore, recent behavioral studies (Maroja et al. 2009), suggest that F1 hybrids prefer to backcross to *G. firmus*. This hybrid mate choice behavior will obviously limit $m1$, the introgression rate from *G. firmus* into *G. pennsylvanicus*, and may explain an apparently genome-wide phenomenon.

Differential and asymmetric introgression between *G. firmus* and *G. pennsylvanicus* has previously been reported for mtDNA (Harrison et al. 1987; Harrison and Bogdanowicz 1997) and allozymes (Harrison and Arnold 1982). Furthermore, in a fine-scale study of the hybrid zone in Connecticut, Ross and Harrison (2002) also observed differential introgression at nuclear loci, with alleles moving from *G. pennsylvanicus* into *G. firmus*. Here, we see no evidence of mtDNA gene flow in either direction, presumably a consequence of the fact that our sampled populations are relatively far from the hybrid zone, whereas previous studies have examined introgression within the hybrid zone. Asymmetries in introgression may be relatively common; for example Kronforst (2008) reported unidirectional introgression between several pairs of hybridizing *Heliconius* butterflies.

The extent of allele sharing between *G. firmus* and *G. pennsylvanicus* at most loci (see Supporting information) suggests that gene flow combined with unsorted ancestral polymorphism likely account for observed patterns of variation. The high levels of genetic variation and lack of significant Tajima's *D* suggest speciation without population bottlenecks. In this scenario, ancestral polymorphism would persist even if species barriers were complete (i.e., no hybridization), because only after many ($>9Ne$) generations are taxa expected to become reciprocally monophyletic for most loci (Tajima 1983; Neigel and Avise 1986; Harrison 1991; Hey 1994; Maddison 1997; Hudson and Coyne 2002). Given that these crickets likely have large effective population sizes (IMa estimate of over 1 million—see Results) and are still exchanging genes, it will be a long time until complete reciprocal monophyly is achieved.

Previous efforts to identify diagnostic differences in nuclear genes between *G. firmus* and *G. pennsylvanicus* have met with mixed success; Harrison and Bogdanowicz (1997) identified four diagnostic restriction fragment length polymorphisms, but intron sequences revealed shared polymorphisms and no evidence of exclusivity (Broughton and Harrison 2003). However, application of a "genealogical sorting index" (Cummings et al. 2008) suggested that *G. firmus* and *G. pennsylvanicus* do show evidence of substantial genealogical differentiation for intron sequences, in spite of the absence of monophyly for any single locus. Here, using a coalescence/population genetics approach, we have found that most nuclear loci, including many genes expressed in male accessory gland, exhibit extensive haplotype sharing and exhibit high levels of introgression. The only exceptions are the two genes from accessory gland that are likely under selection. We, thus, complement the results of Andrés et al. (2008) and reiterate the potential role of *AG-0005F* and *AG-0334P* as barrier genes. However, it should be noted that even these two potential barrier genes are not fully fixed between species. It is possible that they are playing a role in conjunction with other genes. Because of the difficulty of amplifying protein-coding loci from field cricket genomic DNA, of the 39 accessory gland candidate genes we were able to analyze only six; it is thus possible that other genes also play a potential role as barrier genes.

RECENT DEMOGRAPHIC HISTORY AND STRUCTURE OF THE HYBRID ZONE

We identified six major mtDNA clades, three of which represent northern *G. firmus* populations, with the other three corresponding to northern *G. pennsylvanicus*, southern *G. firmus*, and southern *G. pennsylvanicus* populations (Fig. 2). These are the same mtDNA haplotype groups identified by Willett et al. (1997) with two additional clades of northern *G. firmus* individuals (Fig. 2). As in Willett et al. (1997), these clades have strong support (Fig. 2) and distinguish *G. firmus* from *G. pennsylvanicus*. However, the mtDNA data still do not provide resolution at the base of the tree and leave unanswered whether each of the species is an exclusive group with respect to mtDNA.

Based on mtDNA haplotype distributions, we also found evidence for a north/south split in both species, again in an agreement with Willett et al. (1997). With larger sample sizes and broader geographic coverage, it appears that this phylogeographic break runs east–west from the Delmarva Peninsula through northern Maryland and southern Pennsylvania. NBL in southern Pennsylvania contains both clades of *G. pennsylvanicus*, and ESS in northern Maryland contains both clades of *G. firmus*. Although mtDNA clades are geographically well defined, there is evidence of historical or ongoing gene flow between northern and southern *G. pennsylvanicus* populations, e.g., two southern populations (FRN and BRP) include individuals with northern *G. pennsylvanicus* haplotypes (see Fig. 1). Of the nuclear genealogies, only *EF1*-α showed a pattern consistent with a north/south phylogeographic split (see nuclear phylogenetic analyses in Supporting Information and Fig. S2).

The significance of the north/south phylogeographic break remains unclear. It is possible that the pattern reflects the presence of crickets in both northern and southern refugia at some point during the late Pleistocene. A northeastern North American refugium has been invoked for other taxa (e.g., Jaramillo-Correa et al. 2004), and patterns of variation have provided evidence for northern refugia in other regions of Europe and North America (e.g., Kotlík et al. 2006).

Given the current geographic distribution of the two crickets, we expected to see a signal of population expansion. The northern part of the current range of both species became inhabitable only about 15,000 years ago (Davis 1976; Dyke and Prest 1987), which would suggest that populations must have expanded their numbers recently. However, the lack of a significant Tajima's D for most loci indicates that population expansion was not so substantial as to leave a lasting genetic signature. Of the two species, *G. firmus* has higher average nucleotide diversity, an observation consistent both with introgression of *G. pennsylvanicus* alleles and with a phylogeographic history in which the sizes of *G. firmus* populations have been greater during past glaciation cycles, because of its association with sandy soils and coastal habitats. Our divergence estimate suggests that *G. firmus* and *G. pennsylvanicus* divergence predates the most recent glacial advance. If the rate of mtDNA evolution is 1.2% per million years per lineage (Brower 1994), using the time since split ($t/\mu = 4.29$) calculated with IMa gives an estimative of 202,320 years from a common ancestor, which is in agreement with estimates of Broughton and Harrison (2003) (0.1 N_e ~200,000 years) and Willett et al. (1997) (187,500 years).

Based both on morphology and mtDNA variation, it appears that the hybrid zone is wider than once thought (Harrison et al. 1987). In previous studies, the hybrid zone was defined as a long but narrow zone extending from the Blue Ridge Mountains in Virginia to southern Connecticut (Harrison and Arnold 1982; Harrison and Bogdanowicz 1997). However, Harrison and Arnold (1982) reported mixed populations in the Shenandoah Valley and speculated that the hybrid zone might also extend to the west of the Blue Ridge. Indeed, we found mixed populations (COV, FRN, MOO) in the Appalachian Mountains west of the Shenandoah Valley. Individuals from the COV, FRN, and MOO populations had a substantial variation in color and body size (data not shown) and were on average larger than pure *G. pennsylvanicus*, more similar to *G. firmus*. These populations also included crickets with *mtDNA* haplotypes from both *G. firmus* and *G. pennsylvanicus* clades (Figs. 1 and 2).

The zone of overlap between *G. firmus* and *G. pennsylvanicus* is likely a result of a secondary contact between previously isolated forms (Willett et al. 1997). Because both of these cricket species are inhabitants of grassy fields and disturbed open areas, they have presumably benefited from extensive human habitat alterations. The increased amount of a suitable habitat probably has provided avenues for range expansion and increased gene flow/hybridization between the two species. It is thus possible that the hybrid zone has been expanding.

In ground crickets of the genus *Allenomobius*, Howard and Waring (1991) described a mosaic hybrid zone in which altitude determines the relative abundance of two hybridizing species. A northern species, *Allenomobius fasciatus* and a southern species, *A. socius*, meet in the Appalachians. Along a transect through this region, *A. fasciatus* is most abundant at high elevations whereas *A. socius* predominates at lower elevations (Howard and Waring 1991). The *Gryllus* hybrid zone shows similar features, with the two hybridizing species segregated to some extent by altitude. Outside of the hybrid zone, *G. pennsylvanicus* is found primarily in inland/upland situations, with *G. firmus* in coastal/lowland (Harrison and Arnold 1982). The hybrid zone along the eastern front of the Blue Ridge occurs along a steep elevational transect. All of the sites that we sampled that are to the west of the Shenandoah Valley occur at relatively low elevations (COV: 354 m; MOO: 285 m; FRN: 551 m), which may explain why these populations are mixed rather than pure *G. pennsylvanicus*. The Shenandoah Valley might thus have provided a migration route for *G. firmus* individuals to colonize suitable habitats further west, producing a mosaic of pure and mixed populations in the mountain and valley regions of Virginia and West Virginia.

The expansion of the hybrid zone does not imply that species identities will be eventually erased in a hybrid swarm. As in other insect hybrid zones (e.g., Mendelson and Shaw 2002; Bailey et al. 2004), *G. firmus* and *G. pennsylvanicus* have multiple trait differences that restrict gene flow. Some of these barriers operate throughout the zone, whereas others vary geographically.

Temporal isolation (due to differences in development time) is observed in Virginia but not in Connecticut (Harrison 1985). This barrier may be of particular importance in mixed populations along the Blue Ridge and southern Appalachians, because of the interaction between intrinsic differences in development rate between the species and the variation in length of the growing season along elevational gradients.

CONCLUSIONS

Although independent species will ultimately exhibit divergence across their entire genome, persistence of shared ancestral polymorphism and introgression cause recently diverged species to be mosaics with respect to genetic differentiation. Depending on the genetic architecture (Ting et al. 2001) and as long as alleles at barrier genes do not introgress, species integrity can be maintained even in the face of a substantial gene flow. Indeed, multilocus studies of closely related species often report discordant genealogical patterns despite well-defined boundaries based on morphological, behavioral, and ecological characters (Beltran et al. 2002; Broughton and Harrison 2003; Machado and Hey 2003; Dopman et al. 2005; Putnam et al. 2007). In accord with these studies, we report discordant genealogical patterns and differential introgression rates across the genome of the two hybridizing cricket species. The most dramatic outliers are the two accessory gland loci under selection, *AG-0005F* and *AG-0334P,* which showed near-zero introgression and more structured species trees. *AG-0005F* and *AG-0334P* are candidate barrier genes with possible reproductive functions in the field crickets *G firmus* and *G. pennsylvanicus.*

ACKNOWLEDGMENTS

We thank T. M. McCleskey for writing a program that formats sequence data for PHASE analysis and reincorporates the PHASE output into a nexus format. We thank S. Bogdanowicz for technical advice in all aspects of this study. Part of this work was carried out using the resources of the Computational Biology Service Unit (CBSU) at Cornell University, which is partially funded by Microsoft Corporation. Funding was provided by National Science Foundation grants DEB-0111613 (Willie Swanson, PI) and DEB-0639904 to RGH and an AAUW American Fellowship to LSM.

LITERATURE CITED

Aguadé, M. 1998. Different forces drive the evolution of the *Acp*26Aa and *Acp*26Ab accessory gland genes in the *Drosophila melanogaster* species. Genetics 150:1079–1089.

———. 1999. Positive selection drives the evolution of the *Acp*29AB accessory gland protein in *Drosophila*. Genetics 152:543–551.

Alexander, R. D. 1957. The taxonomy of the field crickets of the eastern United States (Orthoptera: Gryllidae: *Acheta*). Ann. Entomol. Soc. Am. 50:584–602.

Almeida, F. C., and R. DeSalle. 2008. Evidence of adaptive evolution of accessory gland proteins in closely related species of the *Drosophila repleta* group. Mol. Biol. Evol. 25:2043–2053.

Andrés, J. A., and G. Arnqvist. 2001. Genetic divergence of the seminal signal-receptor system in houseflies: the footprints of sexually antagonistic coevolution? Proc. R. Soc. Lond. B 268:399–405.

Andrés, J. A., L. S. Maroja, S. M. Bogdanowicz, W. Swanson, and R. G. Harrison. 2006. Molecular evolution of seminal proteins in field crickets. Mol. Biol. Evol. 23:1574–1584.

Andrés, J. A., L. S. Maroja, and R. G. Harrison. 2008. Searching for candidate speciation genes using a proteomic approach: seminal proteins in field crickets. Proc. R. Soc. Lond. B 275:1975–1983.

Anisimova, M., R. Nielsen, and Z. Yang. 2003. Effect of recombination on the accuracy of the likelihood method for detecting positive selection at amino acid sites. Genetics 164:1229–1236.

Arnold, M. L. 1997. Natural hybridization and evolution. Oxford Univ. Press, Oxford.

Bailey, R. I., C. D. Thomas, and R. K. Butlin. 2004. Premating barriers to gene exchange and their implications for the structure of a mosaic zone between *Chorthippus brunneus* and *C. jacobsi* (Orthoptera: Acrididae). J. Evol. Biol. 17:108–119

Barbash, D., D. Sinno, A. Tarone, and J. Roote. 2003. A rapidly evolving MYB-related protein causes species isolation in *Drosophila*. Proc. Natl. Acad. Sci. USA 100:5302–5307.

Barton, N. H., and G. M. Hewitt. 1981. Hybrid zones and speciation. Pp 109–145 *in* W. R. Atchley and D. S. Woodruf, eds. Evolution and speciation. Cambridge Univ. Press, Cambridge, UK.

Begun, D. J., P. Whitley, B. L. Todd, H. M. Waldrip-Dail, and A. G. Clark. 2000. Molecular population genetics of male accessory gland proteins in *Drosophila*. Genetics 156:1879–1888.

Beltran, M., C. D. Jiggins, V. Bull, M. Linares, J. Mallet, W. O. McMillan, and E. Berminghan. 2002. Phylogenetic discordance at the species boundary: comparative gene genealogies among rapidly radiating *Heliconius* butterflies. Mol. Biol. Evol. 19:2176–2190.

Besansky, N. J., J. Krzywinski, T. Lehmann, F. Simard, M. Kern, O. Mukabayire, D. Fontenille, Y. Touré, and N. F. Sagnon. 2003. Semipermeable species boundaries between *Anopheles gambiae* and *Anopheles arabiensis*: evidence from multilocus DNA variation. Proc. Natl. Acad. Sci. USA 100:10818–10823.

Braswell, W. E., J. A. Andrés, L. S. Maroja, R. H. Harrison, D. J. Howard, and W. J. Swanson. 2006. Identification and comparative analysis of accessory gland proteins in Orthoptera. Genome 84:1–13.

Brideau, N. J., H. A. Flores, J. Wang, S. Maheshware, X. Wang, and D. A. Barbash. 2006. Two Dobzhansky-Muller genes interact to cause hybrid lethality in *Drosophila*. Science 314:1292–1295.

Broughton, R. E., and R. G. Harrison. 2003. Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. Genetics 163:1389–1401.

Brower, A. V. Z. 1994. Rapid morphological radiation and convergence among races of the butterfly *Heliconius erato* inferred from patterns of mitochondrial DNA evolution. Proc. Natl. Acad. Sci. USA 91:6491–6495.

Clark, N. L., J. E. Aagaard, and W. J. Swanson. 2006. Evolution of reproductive proteins from animals and plants. J. Reprod. Fert. 131:11–22.

Coyne, J. A., and H. A. Orr. 2004. Speciation. Sinauer Associates, Sunderland, MA.

Cummings, M. P., M. C. Neel, and K. L. Shaw. 2008. A genealogical approach to quantifying lineage divergence. Evolution 62:2411–2422.

Davis, M. B. 1976. Pleistocene biogeography of temperate deciduous forests. Geosci. Man, 13:13–26.

Doherty, J. A., and M. Storz. 1992. Calling song and selective phonotaxis in field crickets, *Gryllus firmus* and *G. pennsylvanicus* (Orthoptera: Gryllidae). J. Insect Behav. 5:555–569.

Dopman, E. B., L. Peréz, S. M. Bogdanowicz, and R. G. Harrison. 2005. Consequences of reproductive barriers for genealogical discordance in the European corn borer. Proc. Natl. Acad. Sci. USA 102:14706–14711.

Dyke, A. S., and V. K. Prest. 1987. Late Wisconsinan and Holocene history of the Laurentide ice sheet. Geogr. Phys. Quaternaire 41:237–264.

Excoffier, L. P., P. E. Smouse, and J. M. Quattro. 1992. Analysis of molecular variance inferred from metric distances among DNA haplotypes: applications to human mitochondrial DNA restriction data. Genetics 131:479–491.

Farris, J. S., M. Källersjö, A. G. Kluge, and C. Bult. 1995. Constructing a significance test for incongruence. Syst. Biol. 44:570–572.

Goldman, N., and Z. Yang. 1994. A codon-based model of nucleotide substitution for protein-coding DNA sequences. Mol. Biol. Evol. 11:725–736.

Grahame, J. W., C. S. Wilding, and R. K. Butlin. 2006. Adaptation to a steep environmental gradient and an associated barrier to gene exchange in *Littorina saxatilis*. Evolution 60:268–278.

Grant, P. R., B. R. Grant, J. A. Markert, L. F. Keller, and K. Petren. 2004. Convergent evolution of Darwin's finches caused by introgressive hybridization and selection. Evolution 58:1588–1599.

Grant, V. 1981. Plant Speciation. Columbia Univ. Press, New York.

Harrison, R. G. 1979. Speciation in North American field crickets: evidence from electrophoretic comparisons. Evolution 33:1009–1023.

———. 1983. Barriers to gene exchange between closely related cricket species. I. Laboratory hybridization studies. Evolution 37:245–251.

———. 1985. Barriers to gene exchange between closely related cricket species. II. Life cycle variation and temporal isolation. Evolution 39:244–259.

———. 1990. Hybrid zones: windows on evolutionary processes. Pp. 69–128 *in* D. Futuyma and J. Antonovics, eds. Oxford surveys in evolutionary biology Vol 7. Oxford Univ. Press, Oxford.

———. 1991. Molecular changes at speciation. Ann. Rev. Ecol. Syst. 22:281–308.

Harrison, R. G., and J. Arnold. 1982. A narrow hybrid zone between closely related cricket species. Evolution 36:355–552.

Harrison, R. G., and S. M. Bogdanowicz. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. Evolution 51:493–505.

Harrison, R. G., and D. M. Rand. 1989. Mosaic hybrid zones and the nature of species boundaries. Pp. 111–133 *in* D. Otte and J. A. Endler, eds. Speciation and its consequences. Sinauer, Sunderland, MA.

Harrison, R. G., D. M. Rand, and W. C. Wheeler. 1987. Mitochondrial DNA variation in field crickets across a narrow hybrid zone. Mol. Biol. Evol. 4:144–158.

Harshman, L. G., and T. Prout. 1994. Sperm displacement without sperm transfer in *Drosophila melanogaster*. Evolution 48:758–766.

Herndon, L. A., and M. F. Wolfner. 1995. A *Drosophila* seminal protein, *Acp*26Aa, stimulates egg laying in females for 1 day after mating. Proc. Natl. Acad. Sci. USA 92:10114–10118.

Hey, J. 1994. Bridging phylogenetics and population genetics with gene tree models. Pp. 435–449 *in* B. Schierwater, B. Streit, G. Wagner, and R. DeSalle, eds. Molecular ecology and evolution: approaches and applications. Birkhauser, Basel, Switzerland.

———. 2005. On the number of new world founders: a population genetic portrait of the peopling of the Americas. PLoS Biol. 3:965–975.

Hey, J., and R. Nielsen. 2004. Multilocus methods for estimating population sizes, migration rates and divergence time, with applications to the divergence of *Drosophila pseudoobscura* and *D. persimilis*. Genetics 167:747–760.

———. 2007. Integration within the Felsenstein equation for improved Markov chain Monte Carlo methods in population genetics. Proc Natl. Acad. Sci. USA 104:2785–2790.

Howard, D. J., and G. L. Waring. 1991. Topographic diversity, zone width and the strength of reproductive isolation in a zone of overlap and hybridization. Evolution 45:1120–1135.

Hudson, R. R. 1983. Properties of a neutral allele model with intragenic recombination. Theor. Popul. Biol. 23:183–201.

———. 1992. Gene trees, species trees and the segregation of ancestral alleles. Genetics 131:509–512.

Hudson, R. R., and J. A. Coyne. 2002. Mathematical consequences of the genealogical species concept. Evolution 56:1557–1565.

Huelsenbeck, J. P., and F. Ronquist. 2001. MRBAYES: Bayesian inference of phylogenetic trees. Bioinformatics 17:754–755.

Jain, R., M. C. Rivera, J. E. Moore, and J. A. Lake. 2002. Horizontal gene transfer in microbial genome evolution. Theor. Popul. Biol. 61:489–495.

Jaramillo-Correa, J. P., J. Beaulieu, and J. Bousquet. 2004. Variation in mitochondrial DNA reveals multiple distant glacial refugia in black spruce (*Picea mariana*), a transcontinental North American conifer. Mol. Ecol. 13:2735–2747.

Kotlík, P., V. Deffontaine, S. Mascheretti, J. Zima, J. R. Michaux, and J. B. Searle. 2006. A northern glacial refugium for bank voles (*Clethrionomys glareolus*). Proc Natl. Acad. Sci. USA 103:14860–14864.

Kronforst, M. R. 2008. Gene flow persists millions of years after speciation in *Heliconius* butterflies. BMC Evol. Biol. 8:98.

Kuhner, M. K., and L. P Smith. 2007. Comparing likelihood and Bayesian coalescent estimation of population parameters. Genetics 175:155–165.

Li, N., and M. Stephens. 2003. Modeling linkage disequilibrium and identifying recombination hotspots using single-nucleotide polymorphism data. Genetics 165:2213–2233.

Machado, C. A., and J. Hey. 2003. The causes of phylogenetic conflict in a classic *Drosophila* species group. Proc. R. Soc. Lond. B 270:1193–1202.

Machado, C. A., R. M. Kilman, J. A. Markert, and J. Hey. 2002. Inferring the history of speciation from multilocus DNA sequence data: the case of *Drosophila pseudoobscura* and close relatives. Mol. Biol. Evol. 19:472–488.

Maddison, W. P. 1997. Gene trees in species trees. Syst. Biol. 46:523–536.

Maroja, L. S., M. E. Clark, and R. G. Harrison. 2008. *Wolbachia* plays no role in the one-way reproductive incompatibility between the hybridizing field crickets *Gryllus firmus* and *G. pennsylvanicus*. Heredity 101:435–444.

Maroja, L. S., J. Andrés, J. R. Walters, and R. G. Harrison. 2009. Multiple barriers to gene exchange in a field cricket hybrid zone. Biol. J. Linn. Soc. Lond. 97:390–402.

Masly, J. P., C. D. Jones, M. A. F. Noor, and H. A. Orr. 2006. Gene transposition as a cause of hybrid sterility. Science 313:1448–1450.

Mendelson, T. C., and K. L. Shaw. 2002. Genetic and behavioral components of the cryptic species boundary between *Laupala cerasina* and *L. kohalensis* (Orthoptera: Gryllidae). Genetica 116:301–310.

Mihola O., Z. Trachtulec, C. Vlcek, J. C. Schimenti, and J. Forejt. 2009. A mouse speciation gene encodes a meiotic histone H3 methyltransferase. Science 323:373–375.

Nei M., and T. Gojobori. 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3:418–426.

Neigel, J. E., and J. C. Avise. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. Pp. 515–534 *in* S. Karlin and E. Nevo, eds. Evolutionary processes and theory. Academic Press, New York.

Neubaum, D. M., and M. F. Wolfner. 1999. Mated *Drosophila melanogaster* females require a seminal fluid protein, Acp36DE, to store sperm efficiently. Genetics 153:845–857.

Nichols, R. 2001. Gene trees and species trees are not the same. Trend Ecol. Evol. 16:358–364.

Nielsen, R., and J. Wakeley. 2001. Distinguishing migration from isolation: a Markov chain Monte Carlo approach. Genetics 158:885–896.

Noor, M., and J. L. Feder. 2006. Speciation genetics: evolving approaches. Nat. Rev. Genet. 7:851–861.

Noor, M. A., K. L. Grams, L. A. Bertucci, and J. Reiland 2001 Chromosomal inversions and the reproductive isolation of species. Proc. Natl. Acad. Sci. USA 98:12084–12088.

Nosil, P., S. P. Egan, and D. J. Funk. 2008. Heterogeneous genomic differentiation between walking-stick ecotypes: "isolation by adaptation" and multiple roles for divergent selection. Evolution 62:316–336.

Payseur, B., and M. Nachman. 2005. The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. Biol. J. Linn. Soc 84:523–534.

Phadnis, N., and A. H. Orr. 2009. A single gene causes both male sterility and segregation distortion in *Drosophila* hybrids. Science 323:376–379.

Presgraves, D., L. Balagopalan, S. Abmayr, and H. Orr. 2003. Adaptive evolution drives divergence of hybrid incompatibility gene between two species of *Drosophila*. Nature 243:715–719.

Putnam, A. S., J. M. Scriber, and P. Andolfatto. 2007. Dicordant divergence times among Z-chromosome regions between two ecologically distinct swallowtail butterfly species. Evolution 61:912–927.

Rand, D. M., and R. G. Harrison. 1989. Ecological genetics of a mosaic hybrid zone: mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. Evolution 43:432–449.

Rieseberg, L. 1997. Hybrid origins of plant species. Annu. Rev. Ecol. Syst. 28:359–389.

Rieseberg, L. H., J. Whitton, and K. Gardner. 1999. Hybrid zones and the genetic architecture of a barrier to gene flow between two sunflower species. Genetics 152:713–727.

Ross, C. L., and R. G. Harrison. 2002. A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. Evolution 56:2296–2312.

———. 2006. Viability selection on overwintering eggs in a field cricket mosaic hybrid zone. Oikos 115:53–68.

Rozas, J., J. C. Sánchez-Delbarrio, X. Messeguer, and R. Rozas. 2003. DnaSP, DNA polymorphism analyses by the coalescent and other methods. Bioinformatics 19:2496–2497.

Schierup, M. H., and J. Hein. 2000. Consequences of recombination on traditional phylogenetic analysis. Genetics 156:879–891.

Schneider, S., D. Roessli, and L. Excoffier. 2000. Arlequin: a software for population genetics data analysis. Genetics and Biometry Laboratory, Department of Anthropology, Univ. of Geneva.

Seehausen, O. 2004. Hybridization and adaptive radiation. Trends Ecol. Evol. 19:198–207.

Shriner, D., D. C. Nickle, M. A. Jensen, and J. I. Mullins. 2003. Potential impact of recombination on sitewise approaches for detecting positive natural selection. Genet. Res. 81:115–121.

Simon C., F. Frati, A. Beckenbach, B. Crespi, H. Liu, and P. Flook. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Annals of the Entomological Society of America 87:651–700.

Stephens, M., N. J. Smith, and P. Donnelly. 2001 A new statistical method for haplotype reconstruction from population data. Am. J. Hum. Genet. 68:978–989.

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

Swanson, W. J., A. G. Clark, H. M. Waldrip-Dail, M. F. Wolfner, and C. F. Aquadro. 2001. Evolutionary EST analysis identifies rapidly evolving male reproductive proteins in *Drosophila*. Proc. Natl. Acad. Sci. USA 95:4051–4054.

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

Tajima, F. 1983. Evolutionary relationship of DNA sequences in finite populations. Genetics 105:437–460.

———. 1989. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics 123:585–595.

Templeton, A. R. 1994. The role of molecular genetics in speciation studies. Pp. 455–477 *in* B. Schierwater, B. Streit, G. Wagner and R. DeSalle, eds. Molecular ecology and evolution: approaches and applications. Birkhauser, Basel, Switzerland.

Ting, C., S.-C. Tsaur, and C.-I. Wu. 1998. A rapidly evolving homeobox at the site of a hybrid sterility gene. Science 282:1501–1504.

Ting, C.-T., S.-C. Tsaur, and C.-I. Wu. 2000. The phylogeny of closely related species as revealed by the genealogy of a speciation gene, *Odysseus*. Proc. Natl. Acad. Sci. USA. 97:5313–5316.

Ting, C.-T., A. Takahashi, and C.-I. Wu. 2001. Incipient speciation by sexual isolation in *Drosophila*: concurrent evolution at multiple loci. Proc. Natl. Acad. Sci. USA 98:6709–6713.

Tram, U., and M. F. Wolner. 1999. Male fluid proteins are essential for sperm storage in *Drosophila melanogaster*. Genetics 153:837–844.

Vasemagi, A., J. Nilsson, and C. R. Primmer. 2005. Expressed sequence tag-linked microsatellites as a source of gene-associated polymorphisms for detecting signatures of divergent selection in Atlantic salmon (*Salmo salar* L.). Mol. Biol. Evol. 22:1067–1076.

Wang, R. L., J. Wakeley, and J. Hey. 1997. Gene flow and natural selection in the origin of *Drosophila pseudoobscura* and close relatives. Genetics 147:1091–1106.

Willett, C., M. J. Ford, and R. G. Harrison. 1997. Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. Heredity 79:484–494.

Wilson, D. J., and G. McVean. 2006. Estimating diversifying selection and functional constraint in the presence of recombination. Genetics 172:1411–1425.

Wittbrodt, J., D. Adam, B. Malitscheck, W. Maueler, F. Raulf, A. Telling, S. M. Robertson, and M. Schartl. 1989. Novel putative receptor tyrosine kinase encoded by melanoma-inducing *Tu* locus in *Xiphophorus*. Nature 341:415–421.

Woerner, A. E., P. C. Murray, and M. F. Hammer. 2007. Recombination-filtered genomic datasets by information maximization. Bioinformatics 23:1851–1853.

Wolfner, M. F. 1997. Tokens of love: functions and regulations of *Drosophila* accessory male products. Insect. Biochem. Mol. Biol. 27:179–192.

Wu, C.-I. 2001. The genic view of the process of speciation. J. Evol. Biol. 14:851–865.

Wu, C.-I., and C. T. Ting. 2004. Genes and speciation. Nat. Rev. Genet. 5:114–122.

Associate Editor: W. Owen McMillan

Supporting Information

The following supporting information is available for this article:

Figure S1. Neighbor-joining trees for all nuclear loci.
Figure S2. Neighbor-joining tree for *EF1-α* labeled with population names.
Figure S3. Neighbor-joining tree for *GuKc* labeled with population names.
Figure S4. Neighbor-joining tree for *Hex* labeled with population names.
Figure S5. Neighbor-joining tree for *AG-0005F* labeled with population names.
Figure S6. Neighbor-joining tree for *AG-0032F* labeled with population names.
Figure S7. Neighbor-joining tree for *AG-0090F* labeled with population names.
Figure S8. Neighbor-joining tree for *AG-0211F* labeled with population names.
Figure S9. Neighbor-joining tree for *AG-0254P* labeled with population names.
Figure S10. Neighbor-joining tree for *AG-0334P* labeled with population names.
Figure S11. Posterior probability estimates of ancestral population size ($4N_e\mu$) for each locus.
Figure S12. Posterior probability estimates of time since population split (scaled by mutation rate: $t = t\mu$) for each locus.
Figure S13. Joint posterior probability estimates of ancestral population size and time since population split (scaled by mutation rate: $t = t\mu$) for the two loci under selection (*AG-0005F* and *AG-0334P*) and for all other nuclear loci combined.
Table S1. Selected model for each locus (using MODELTEST 3.06).

Supporting Information may be found in the online version of this article.
(This link will take you to the article abstract).

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supporting information supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.