MOLECULAR ECOLOGY

Molecular Ecology (2014) **23**, 1668–1678 doi: 10.1111/mec.12601

FROM THE COVER

Gene flow and the maintenance of species boundaries

ERICA L. LARSON,*† THOMAS A. WHITE,*‡ CHARLES L. ROSS§ and RICHARD G. HARRISON*

*Department of Ecology and Evolutionary Biology, Cornell University, Ithaca, NY 14853, USA, †Division of Biological Sciences, University of Montana, Missoula, MT 59812, USA, ‡CMPG Lab, Institute of Ecology and Evolution, University of Bern, Baltzerstrasse 6, 3012 Bern, Switzerland, §School of Natural Science, Hampshire College, Amherst, MA 01002, USA

Abstract

Hybrid zones are regions where individuals from genetically differentiated populations meet and mate, resulting in at least some offspring of mixed ancestry. Patterns of gene flow (introgression) in hybrid zones vary across the genome, allowing assessment of the role of individual genes or genome regions in reproductive isolation. Here, we document patterns of introgression between two recently diverged species of field crickets. We sampled at a very fine spatial scale and genotyped crickets for 110 highly differentiated single nucleotide polymorphisms (SNPs) identified through transcriptome scans. Using both genomic and geographic cline analysis, we document remarkably abrupt transitions (<100 m) in allele frequencies for 50 loci, despite high levels of gene flow at other loci. These are among the steepest clines documented for any hybridizing taxa. Furthermore, the cricket hybrid zone provides one of the clearest examples of the semi-permeability of species boundaries. Comparisons between data from the fine-scale transect and data (for the same set of markers) from sampling a much larger area in a different region of the cricket hybrid zone reveal consistent patterns of introgression for individual loci. The consistency in patterns of introgression between these two distant and distinct regions of the hybrid zone suggests that strong selection is acting to maintain abrupt discontinuities within the hybrid zone and that genomic regions with restricted introgression likely include genes that contribute to nonecological prezygotic barriers.

Keywords: genomic clines, *Gryllus*, habitat isolation, introgression, spatial scale, speciation

Received 12 October 2013; revision received 14 November 2013; accepted 16 November 2013

Introduction

Species are often viewed as cohesive entities, groups of populations connected by gene flow and reproductively isolated from other such groups (Mayr 1963). Alternatively, species can be defined as diagnosably distinct (Nixon & Wheeler 1990) or exclusive groups (Baum & Shaw 1995) of individuals. In fact, cohesion, diagnosability and exclusivity are properties of individual genes or genome regions and not of whole organisms (or lineages) (Barton & Hewitt 1985; Harrison 1990; Wu 2001; Nosil *et al.* 2009). Because of independent assortment and recombination, each genome region may have a unique evolutionary history, and boundaries between species will be semi-permeable, with permeability depending on the genome region. In the face of ongoing gene flow, some regions of the genome can become (or remain) strongly differentiated, if these regions contain genes that contribute to reproductive isolation or local adaptation.

Hybrid zones provide a unique opportunity to identify the role of individual genes in reproductive isolation. Hybridization and introgression over many generations result in the shuffling of divergent genomes. For a given gene, the strength of selection, the role in reproductive isolation, together with genetic linkage relationships, will determine the extent of introgression. Alleles that are advantageous in the genomic background of either parental species will be easily exchanged via hybridization, whereas genomic regions that contribute to barriers between species will have low gene flow and characteristically steep clines (Barton & Hewitt 1985; Harrison 1990; Payseur 2010). Allele

Correspondence: Erica L. Larson, Fax: (406) 243-4184;
E-mail: erica.larson@mso.umt.edu

frequencies are most often plotted along a geographic transect (indeed, a cline is defined as a gradual change in frequency over space); however, introgression can also be characterized by plotting allele frequency at a single locus against a hybrid index that reflects the proportion of ancestry assigned to a designated parental type (Szymura & Barton 1986; Gompert & Buerkle 2009). These nongeographic clines provide measures of introgression relative to the average extent of introgression across all sampled loci, and analysis of cline width and position allows identification of genome regions for which introgression is significantly restricted. These regions presumably harbour genes that contribute to reproductive isolation or local adaptation.

Hybrid zones often extend across large geographic areas, but interactions between species can occur at very fine spatial scales. For example, species associated with different habitats often have a patchy distribution in heterogeneous environments; local introgression can occur across patch boundaries on scales of only tens to hundreds of metres, even when hybrid zone width and length are measured in tens or hundreds of kilometres (Harrison & Rand 1989). In complex or mosaic hybrid zones, patch boundaries reflect independent contacts between diverging lineages. Genes or genome regions that consistently exhibit steep clinal variation (limited introgression) in different geographic and ecological contexts are of particular interest, because these genome regions may be the most important determinants of the species boundary.

Here, we compare patterns of introgression in different environmental contexts and at different spatial scales within a hybrid zone between two recently diverged species of field crickets, *Gryllus pennsylvanicus* and *Gryllus firmus* (~ 200 000 ya; Willett *et al.* 1997; Broughton & Harrison 2003; Maroja *et al.* 2009a). This well-characterized hybrid zone is maintained by multiple barriers to gene exchange. Some barriers involve adaptation to different environments (temporal isolation: Harrison 1985; habitat association: Rand & Harrison 1989a; Ross & Harrison 2006; Larson *et al.* 2013b), whereas other barriers reflect differences in behaviour (mate choice: Maroja *et al.* 2009b) and sperm–egg interactions (Harrison 1983; Larson *et al.* 2012b). There is no evidence for postzygotic barriers; hybrid offspring between the two species are viable, fertile and able to produce F2 offspring and backcross to the parental species.

We characterize patterns of introgression for 110 highly differentiated single nucleotide polymorphisms (SNPs) along a single fine-scale transect (500 m) that samples a transition between loam and sand soils in Connecticut (see Ross & Harrison 2002). In this region of the cricket mosaic hybrid zone, *G. pennsylvanicus* is associated with loam soils and *G. firmus* with sand. For many genes, we find abrupt changes in allele frequencies across a distance of <60 m, a pattern that can only be maintained by strong selection. These genes define the species boundary, which remains intact despite evidence for high levels of gene flow for some genome regions. We then compare data from the fine-scale transect with patterns of introgression in Pennsylvania, where populations were sampled in a different ecological context and on a much broader spatial scale (Larson *et al.* 2013a). This comparison reveals remarkably consistent patterns of introgression for individual genes. Genes that exhibit reduced introgression at both localities are strong candidates to mark genomic regions that contribute to reproductive barriers between species.

Materials and methods

Reference populations

The *Gryllus* hybrid zone stretches from Massachusetts to Virginia along the eastern edge of the Appalachian Mountains. *Gryllus pennsylvanicus* occupies the inland/upland areas to the west and north, and *G. firmus* occupies the lowland/coastal areas to the east and south (Harrison & Arnold 1982; Maroja *et al.* 2009a; Larson *et al.* 2013b). We used a previously published data set of cricket genotypes from three pure allopatric populations of each species to estimate allele frequencies within each species (Larson *et al.* 2013a). The allopatric populations include *G. pennsylvanicus* from Ithaca, NY (N = 12); Scranton, PA (N = 11); and State College, PA (N = 12) and *G. firmus* from Guilford, CT (N = 12); Tom's River, NJ (N = 12); and Parksley, MD (N = 12).

Connecticut

The hybrid zone in Connecticut is a mosaic of pure and admixed populations structured by variation in soil type. The advance and re-treat of glacial ice has left a patchy distribution of soil types; ridges running perpendicular to the coast are characterized by loam soil occupied by *G. pennsylvanicus*, whereas the intervening river drainages are primarily sandy and inhabited by *G. firmus*. Out-pockets of upland habitat and loam soils stretch to the coast while sandy soils extend out from the river valleys and species interact at the boundaries of these patches (Harrison 1986; Harrison & Rand 1989; Rand & Harrison 1989a; Ross & Harrison 2002). Crickets from the Connecticut hybrid zone were collected in September of 1996 and 1997 along a 500 m stretch of River Road (near the Connecticut River southeast of Middleton, CT) that spans a boundary between loam (0 m, *G. pennsylvanicus*) and sand (500 m, *G. firmus*) soil patches (N: 41°33′30″; W: 72°35′18″) (Ross & Harrison

2002). Crickets were collected from adjacent habitat within a few metres of both sides of the road (N = 146) and at both ends of the transect from 10 to 30 m^2 areas (loam, N = 68; sand, N = 46).

Pennsylvania

In Pennsylvania, there is also a patchy species distribution; natural habitat along the mountain ridges is occupied by *G. pennsylvanicus*, whereas the primarily agricultural and suburban lowlands are occupied by *G. firmus*. There are corridors of disturbed habitat through the mountains and out-pockets of natural habitat in the valleys where the two species come into contact (Larson *et al.* 2013b). We reanalysed a previously published data set of 301 cricket genotypes from the Pennsylvania hybrid zone described in Larson *et al.* (2013a). Crickets were collected in August and September of 2008 and 2010 from 36 localities (spaced approximately every 2–5 km) in a 200 km^2 area that spanned the transition from the Appalachian Mountains to the coastal plain.

Gene identification and genotyping

Our panel of markers represents a subset of genes with elevated divergence between *G. pennsylvanicus* and *G. firmus*, identified in a transcriptome scan of the male accessory gland comparing two focal populations (Andrés *et al.* 2013). We genotyped a single SNP per gene in a panel of 71 crickets from three pure allopatric populations of each species (see *Reference populations*), and from these, we selected 110 highly differentiated SNPs, each of which has an absolute allele frequency difference between species >0.80 (Larson *et al.* 2013a). These SNPs were used to estimate admixture for 260 crickets sampled along the 500 m transect in the Connecticut hybrid zone (Ross & Harrison 2002). We extracted genomic DNA from crickets using single adult femurs, thoraces or heads using a DNeasy Tissue Kit (QIAGEN Inc., Valencia, CA, USA), although for a subset of samples DNA was extracted using phenol–chloroform (Ross & Harrison 2002). DNA was diluted to 10 ng/ul and SNPs genotyped using previously designed assays (Larson *et al.* 2013a) on the MassARRAY platform (Sequenom Inc., San Diego, CA, USA) with iPLEX Gold chemistry at the Cornell Life Sciences Core Laboratories Center for Genomics. We called SNP genotypes using the Sequenom MassARRAY Typer v4.0 Analysis software and confirmed genotypes by eye (Dryad doi:10.5061/dryad.258 h4).

Admixture analyses

We used the R-package *introgress* (Gompert & Buerkle 2009, 2010) to quantify genomic admixture. We quantified admixture for each cricket by calculating a hybrid index, defined as the proportion of alleles from all 110 genes that were inherited from *G. firmus* (0 = *G. pennsylvanicus*, 1 = *G. firmus*). We estimated the interspecific heterozygosity for each cricket, defined as the proportion of genes that are heterozygous for the parental alleles (0 = all homozygous genotypes, 1.0 = all heterozygous genotypes). We compared each cricket's interspecific heterozygosity with its hybrid index to estimate the likely ancestry of each cricket. Following Milne and Abbott (2008), we considered crickets with intermediate hybrid indices (≥0.25 and ≤0.75) and high heterozygosity (≥0.30) to be recent-generation hybrids or highly admixed individuals and crickets with low hybrid indices (<0.25 or >0.75) and heterozygosity (<0.30) to be backcrosses into *G. pennsylvanicus* and *G. firmus* (respectively). Although individuals with complex ancestry cannot be easily classified, this approach provides a rough summary of the distribution of genomic admixture.

Genomic clines (multinomial logistic regression)

We used a multinomial regression to estimate individual clines in genotype frequency for each locus along a gradient of genomic admixture (hybrid index) using Gompert & Buerkle's (2009) genomic cline model implemented in the R-package *introgress* (Gompert & Buerkle 2010, see Lexer *et al.* 2007). For significance testing, we compared the likelihood of our regression model to a null model generated using the parametric procedure described in Gompert & Buerkle (2009). We generated an expected distribution of genotypes given the allele frequencies in each parental population, the hybrid index and individual heterozygosity. We used this distribution to construct a simulated admixed population (2000 simulations) with individual hybrid index and interclass heterozygosity values equal to observed data and a large population size (to exclude sampling error). We adjusted all significance thresholds using the false discovery rate procedure (Benjamini & Hochberg 1995). Evidence for an excess (+) or deficiency (−) of either homozygous or heterozygous genotypes was based on the proportion of simulations that yield a model with higher total probability of a given genotype than the model based on the observed data. We summarized these deviations as either 1) gradual clines (excess or deficit of homozygotes and/or excess of heterozygotes) or 2) abrupt clines (deficit of heterozygotes consistent with assortative mating, disruptive selection or underdominance) using a combination of the *introgress* output and visual inspection of cline shape. To test whether a greater number of SNPs have similar cline shapes in the two hybrid zones than would be expected by

chance, we performed a sample label permutation test (n = 10 000).

Concordance clines (logit-logistic model)

The relationship between mean hybrid index and allele frequency at each individual locus was calculated using Fitzpatrick's (2013) logit-logistic model (see also Szymura & Barton 1986). The predicted allele frequency P in deme i is given by

$$\rho_i = \frac{S^{v_j}}{S^{v_j} + (1 - S)^{v_j} e^{u_i}}$$

where S is the mean hybrid index over all loci, u gives the relative difference in cline position and v gives the relative difference in slope. Perfect concordance between a focal locus and the mean hybrid index would result in $u = 0$ and $v = 1$. Parameters u and v were fitted by maximum-likelihood estimation, using the function 'mle2' in the R-package *bbmle* (Bolker 2012). Two further models were also fitted where either u or v was constrained to be 0 or 1, respectively, and these were compared with the unconstrained model using likelihood ratio tests.

Geographic clines

Allele frequencies for each locus were fit to a tanh model of cline shape by maximum-likelihood estimation (Szymura & Barton 1986, 1991). Sample sizes were corrected following Phillips *et al.* (2004) and Raufaste *et al.* (2005) as

$$N_e = \frac{2N}{2N * F_{ST} + F_{IS} + 1}$$

where N is the number of individuals sampled in a deme, F_{IS} is the deficit of heterozygotes (zero if not positive) and F_{ST} is the variation in allele frequencies between loci, after accounting for differences in their cline shapes. F_{ST} is calculated from the residual variation around the regression line fitted during the concordance analysis.

Results

Admixture analyses reveal few recent-generation hybrids, but extensive introgression at a fine spatial scale

For each cricket, we calculated a hybrid index as the proportion of alleles inherited from *G. firmus* (0 = *G. pennsylvanicus*, 1 = *G. firmus*) (Fig. S1, Supporting information) and interspecific heterozygosity as the proportion of loci that are heterozygous for the parental alleles (0 = all homozygous genotypes, 1.0 = all heterozygous genotypes). A plot of the hybrid index against interspecific heterozygosity for each cricket is shown in Figure 1. We found only one cricket that could be characterized as pure *G. pennsylvanicus* and fifteen crickets that were pure *G. firmus*. The majority of crickets were either backcrosses into *G. pennsylvanicus* (N = 156) or *G. firmus* (N = 81); seven crickets were identified as recent-generation hybrids. Two crickets with high interspecific heterozygosity (≥0.75) and an intermediate hybrid index (~ 0.5) could be considered F1 hybrids, given the variance associated with estimating these parameters using markers that are not completely fixed between species. Crickets were collected along a 500-m transect that spans a transition between loamy (0 m, *G. pennsylvanicus*) and sandy (500 m, *G. firmus*) soils (Ross & Harrison 2002). On the loam soil patch (0 m), we found predominately backcrosses into *G. pennsylvanicus*, a few backcrosses into *G. firmus* and two highly admixed individuals (most likely recent-generation hybrids), while on the sand patch (500 m) we found only backcrosses into *G. firmus* and a few pure *G. firmus* (Fig. 2). There is an abrupt transition between *G. pennsylvanicus*-like crickets and

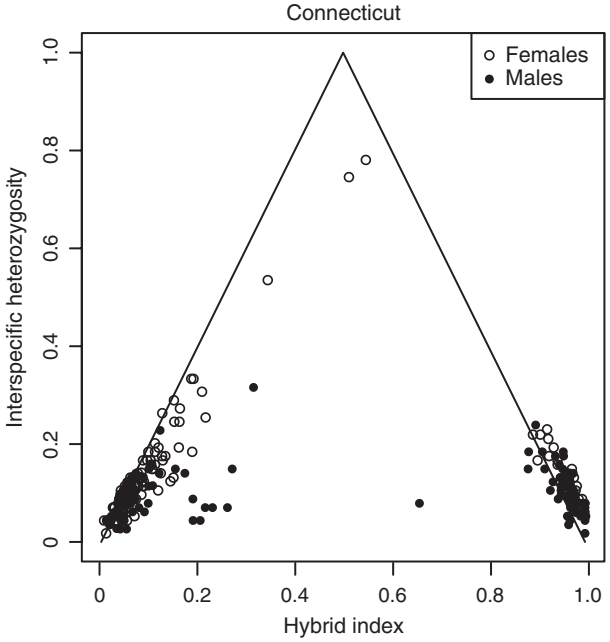

Fig. 1 Interspecific heterozygosity plotted against hybrid index for 260 crickets from the Connecticut hybrid zone. The hybrid index is measured as the proportion of alleles with *Gryllus firmus* ancestry (0 = *Gp*, 1 = *Gf*). Crickets with interspecific heterozygosity ≥85% and hybrid index = 0.5 are likely F1 hybrids, while crickets with hybrid indices ≥0.25 and ≤0.75 are classified as recent-generation hybrids. Crickets with hybrid indices <0.25 or >0.75 are considered backcrosses into *Gryllus pennsylvanicus* and *G. firmus* (respectively).

G. firmus-like crickets at approximately 300–320 m. This transition is more abrupt than the change in soil type along the transect, which changes gradually from lower sand (70%) and higher organic matter (5%) at the loam patch (0 m) to higher sand (85%) and lower organic matter (0%) on the sand patch (500 m) (Ross & Harrison 2002). Recent-generation hybrids are found at the centre of the transect and on the *G. pennsylvanicus* side.

Cline analyses reveal steep, concordant clines

The extent of introgression varies dramatically among markers. For many genes, alleles of one species are present in crickets with overall hybrid indices characteristic of the other species; other genes, however, exhibited little evidence of introgression (Fig. 3). For each gene, we estimated clines using three approaches, genomic clines (Gompert & Buerkle 2009) (Fig. 4A), concordance clines (Fitzpatrick 2013) (Fig. 4B) and geographic clines (Szymura & Barton 1986) (Fig. 4C, Fig. S2, Supporting information). We found very consistent results among these methods. The genomic cline approach identified 50 SNPs with abrupt clines. The concordance method identified 45 SNPs with steep clines, and all of these were among the 50 SNPs identified using the genomic cline method. The two methods also identified similar numbers of clines as gradual (53 and 46) (Table S1, Supporting information). There was considerable variance in the steepness of the geographic clines; widths ranged from 162.2 to 861.1 m (mean $w = 347.0 \pm 174.5$ m) with centres between 153.9 and 519.3 m (mean $c = 318.0 \pm 52.18$ m) (Table S1, Supporting information). Yet, SNPs with narrow genomic or concordance clines also had very steep geographic clines (width = 162.2–309.4 m, mean $w = 226.5 \pm 40.73$ m), all of which were centred between 298.5 and 347.1 m (mean $c = 321.3 \pm 13.14$ m) (Fig. 4C). Abrupt clines indicate restricted introgression, consistent with markers that are associated with assortative mating, disruptive selection or selection against hybrids.

Consistent patterns of introgression across different regions and scales

We compared the fine-scale transect in Connecticut with broader sampling of 301 crickets from 36 localities across a 200-km^2 area of the hybrid zone in Pennsylvania (Larson *et al.* 2013a). The sampling localities in Pennsylvania

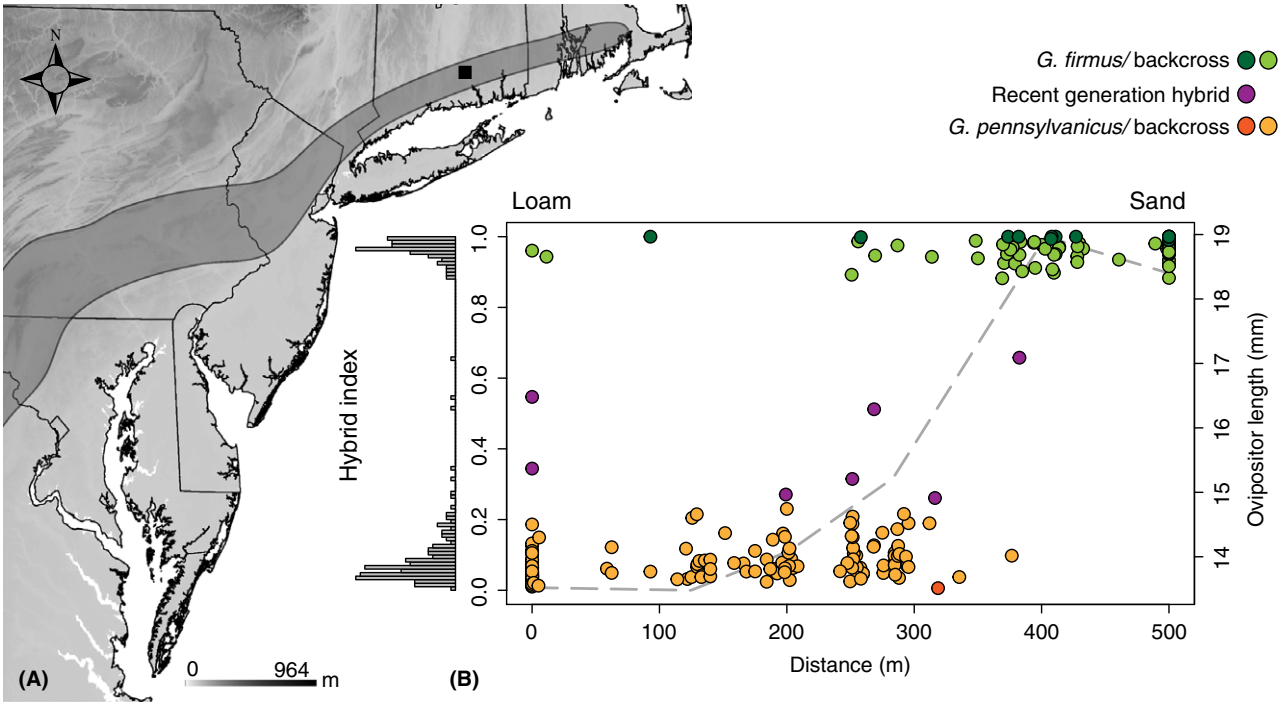

Fig. 2 (A) Map of the eastern United States with the hybrid zone indicated as a dark grey band and the location of the transect in Connecticut as a small black square. (B) The hybrid index (left axis) for each cricket ($N = 260$) plotted against the distance along the 500-m transect from the loam soil patch (0 m, *Gp*) to the sand soil patch (500 m, *Gf*). Each data point is colour-coded to indicate the parental and hybrid class for that individual; *Gryllus pennsylvanicus* (dark orange), *Gryllus firmus* (dark green), backcrosses into *G. pennsylvanicus* (light orange) and *G. firmus* (light green) and recent-generation hybrids (purple). The dashed grey line represents the mean ovipositor length (right axis) of female crickets. Note: the transect runs north to south.

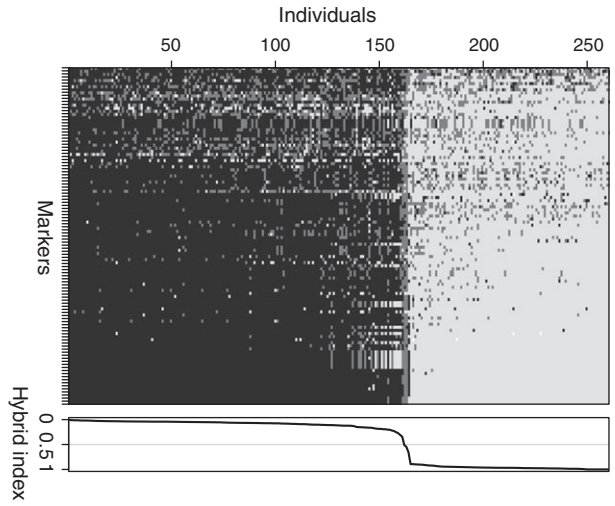

Fig. 3 Overview of patterns of introgression for 110 genes for crickets in the Connecticut hybrid zone. Each row represents a single SNP/gene and each column represents an individual. Markers are ordered by the width of their concordance clines, and individuals are ordered on the basis of hybrid index. The grey scale reflects the genotypes of the individual crickets; dark grey indicates that the individual is homozygous for the *Gryllus pennsylvanicus* allele; medium grey means that the cricket is heterozygous; light grey means that the cricket is homozygous for the *Gryllus firmus* allele, and white represents missing data. Below is a summary plot of the ordered hybrid indices in Connecticut (0 = *G. pennsylvanicus*, 1 = *G. firmus*).

were composed primarily of pure parental types and backcrosses; only eleven localities contained both parental types. Thus, the Pennsylvania sample includes more pure individuals of each species, a result that almost certainly reflects the larger sampling area and sampling fewer habitat patch boundaries. Nonetheless, we found the same proportion of recent-generation hybrids in Pennsylvania as we did in Connecticut and we found only one individual in Pennsylvania that could be considered an F1 hybrid. Patterns of introgression for individual SNPs were consistent between geographic regions and spatial scales. In Connecticut, there were 50 SNPs with significantly restricted introgression (genomic clines = 50, concordance clines = 45). Thirty-three of these markers also had restricted introgression in Pennsylvania (genomic clines = 33, concordance clines = 31) (Fig. 5). The differences between the two regions (and methods) are a result of different significance thresholds. The majority of markers that are 'different' between the two regions have narrow cline shapes but are not significantly different from neutrality ($N = 16$) or concordance ($N = 16$). Only a few loci have patterns inconsistent between the two regions (genomic clines = 1, concordance clines = 3). Overall, both methods revealed clinal patterns between the two transects that were remarkably similar. Patterns were also more similar than expected for all loci (label permutation test, $P < 0.001$), as well as for the subset of markers with abrupt clines (label permutation test, $P < 0.001$).

Discussion

Clear species boundaries at a fine spatial scale despite high gene flow

Species can remain distinct for some portion of their genome, even when hybridization and introgression cause allele frequencies in other genome regions to be homogenized. Loci that remain distinct in allele frequency in the face of hybridization and introgression constitute the species boundary. In our sample of crickets from Connecticut, we found an abrupt transition from *G. pennsylvanicus*-like to *G. firmus*-like crickets in a distance of <60 m (Fig. 2). The distinct discontinuity between *G. pennsylvanicus* and *G. firmus* is evident from morphological comparisons (ovipositor length) (Ross & Harrison 2002), the bimodal distribution of the hybrid index (Fig. S1, Supporting information) and from the steep genomic and geographic clines for 50 of the markers we surveyed (Fig. 4). This discontinuity is striking given that nearly all of the crickets, even those collected from within the habitat patches at each end of transect, have alleles characteristic of the other species (Fig. 3). Because all of the markers surveyed have large allele frequency differences (≥0.80) between allopatric *G. pennsylvanicus* and *G. firmus*, shared alleles are evidence of extensive gene flow (as opposed to incomplete lineage sorting) between the two species. Indeed, more than half of the SNPs assayed exhibited substantial introgression (i.e. gradual clines). This transect represents one of the clearest examples of the semi-permeability of species boundaries. There are clearly two distinct entities that maintain their integrity despite gene flow, and this occurs at an extraordinarily fine scale.

The 50 markers that have restricted gene flow (narrow clines) are representative of the genomic regions that maintain species boundaries. These markers have widths of only a few hundred metres, all centred at approximately 310 m along the transect. These are remarkably steep clines, among the steepest clines documented for any hybridizing taxa (see Singhal & Moritz 2012). For a field cricket, 60 m is easily traversed in a single generation; mark and recapture studies of flightless *G. pennsylvanicus* suggest that crickets often move at least 10–20 m within favourable habitat (Harrison & Rand 1989). Moreover, both species produce occasional long-winged individuals that can disperse over longer distances (Harrison 1979). Given the opportunities for

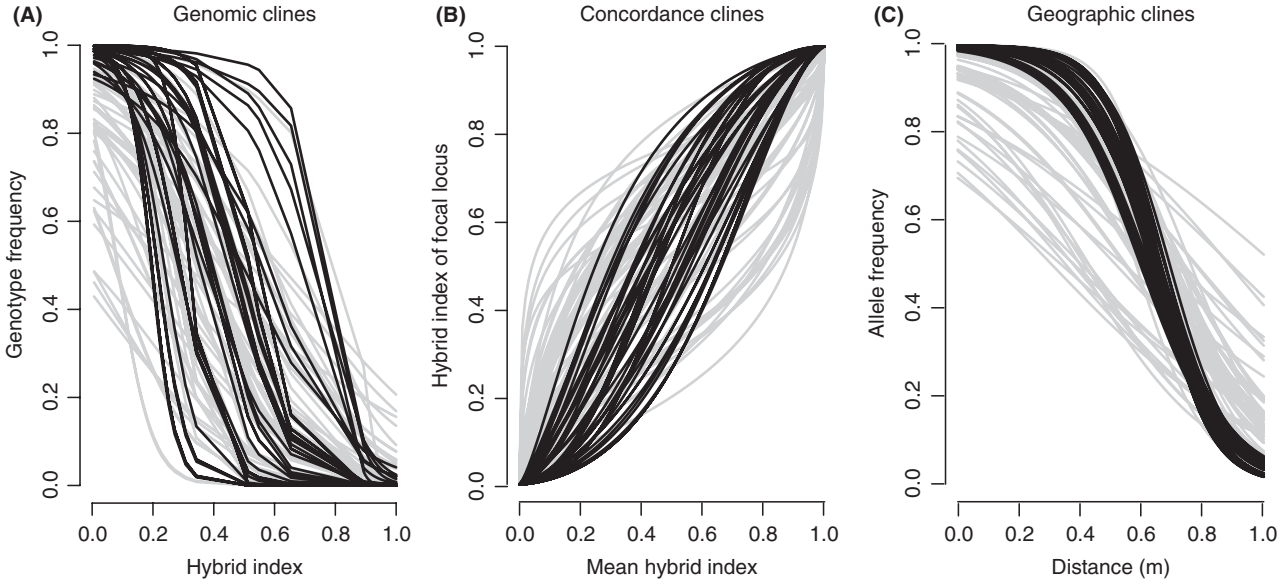

Fig. 4 Clines estimated from 110 SNPs in the Connecticut hybrid zone. (A) Genomic clines (Gompert & Buerkle 2009) plotted as the observed frequency of the *Gryllus pennsylvanicus* homozygote genotype (0 = *Gf*; 1 = *Gp*) against the hybrid index, calculated as the proportion of *G. pennsylvanicus* alleles across all loci (0 = *Gp*, 1 = *Gf*). Narrow clines, with abrupt changes in genotype frequency, represent genes with a deficit of heterozygotes. (B) Concordance clines (Fitzpatrick 2013) plotted as the hybrid index at the focal locus (0 = *Gp*, 1 = *Gf*) against the mean hybrid index calculated over all loci (0 = *Gp*, 1 = *Gf*). The diagonal represents 'concordance' (i.e. introgression is equal to expectation), and loci may have more or less abrupt change (v) or deviate towards either genomic background (u). (C) Geographic clines plotted as the frequency of the *G. pennsylvanicus* allele (0 = *Gf*, 1 = *Gp*) against distance along the transect (0 m = *Gp*, loam soil, 500 m = *Gf*, sand soil). Black lines represent genes with significantly restricted introgression (using the genomic clines method) (N = 50), and grey lines represent genes that have gradual or nonsignificant clines (N = 60).

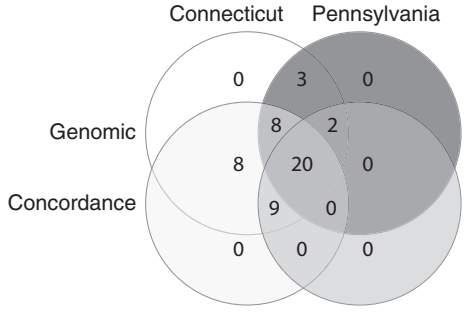

Fig. 5 Venn diagram showing the overlap of markers with significantly restricted introgression for the four comparisons between regions (Connecticut, left; Pennsylvania; right) and cline estimates (genomic, top; concordance, bottom). There were 50 markers that had restricted introgression in Connecticut, and 33 of these also had restricted introgression in Pennsylvania.

dispersal over this spatial scale, strong selection is required to maintain such steep clines (and high linkage disequilibrium). It is possible that the abrupt clines are the transitory result of very recent contact (within the past few generations) and that there simply has not been enough time for hybridization and introgression to homogenize allele frequencies. However, the transect was sampled in two different years, and neighbouring Connecticut populations have been carefully sampled for decades (Harrison & Arnold 1982; Harrison 1986; Harrison *et al.* 1987; Rand & Harrison 1989b; Harrison & Bogdanowicz 1997; Ross & Harrison 2002). These observations suggest that the two cricket species have been interacting across patch boundaries for many generations.

It is tempting to attribute the abrupt changes in allele frequency along this transect to environmental selection imposed by adaptation to different soil types. Throughout the region, crickets are associated with different soil types, but what maintains these soil associations is unclear. Neither differences in oviposition preference nor in overwintering egg viability appear to contribute (Ross & Harrison 2006). In fact, the change in soil type along the transect is much more gradual than the change in allele frequencies (Ross & Harrison 2002). Narrow clines are often maintained by strong selection against hybrids. Although F1 hybrids are rare within the hybrid zone, there is no evidence that they are less fit than parentals (Harrison 1983). In contrast, nonecological prezygotic barriers, such as assortative mating and fertilization barriers, are well documented (Harrison 1983; Rand & Harrison 1989a; Maroja *et al.*

2009b; Larson *et al.* 2012b) and may be sufficiently strong to prevent hybridization. Habitat association may determine the location of boundaries between the two species, but it is likely the coupling of multiple prezygotic barriers that restricts gene flow and maintain steep clines (see Bierne *et al.* 2011).

Nonecological barriers are important in maintaining species boundaries

Hybrid zones extend over large geographic areas, and variation in the relative importance of individual barriers is well documented (Harrison 1990). Variation may be due to intrinsic factors, such as genetic differentiation among conspecific populations, or to extrinsic factors such as variation in the ecological context (e.g. Vines *et al.* 2003). As a result, both the extent of gene flow between species and patterns of introgression for individual genes could vary, depending on sampling locality and spatial scale. Variable patterns of introgression could provide a link between genetic variation and environmental variables that contribute to local adaptation (Gompert *et al.* 2012; Abbott *et al.* 2013), and there is evidence for variable introgression in hybrid zones in house mouse and sculpin (Nolte *et al.* 2009; Teeter *et al.* 2010). However, it is not clear whether these examples reflect differences in the genomic architecture of reproductive isolation; observed variation could result from genetic drift or from sampling complex hybrid zones at intermediate spatial scales (Dufková *et al.* 2011; Macholan *et al.* 2011; Janousek *et al.* 2012). In contrast to these earlier studies, we document many genes that have consistent patterns of introgression at all sites/scales. Patterns of variation for these genes (genome regions) presumably reflect barriers that define global species boundaries and may be more likely to have played a role in the initial isolation of diverging lineages (Buerkle & Rieseberg 2001).

In the field cricket hybrid zone, the determinants of ecological isolation and the mosaic structure of the zone appear to vary among different regions. In Pennsylvania, *G. pennsylvanicus* crickets are associated with more pristine habitat (forest edges and natural clearings), while *G. firmus* is associated with more disturbed habitats (agriculture and suburban lawns) (Larson *et al.* 2013b). In Connecticut, both species occupy disturbed habitat but are associated with distinct soil patches (Rand & Harrison 1989a; Ross & Harrison 2002). Despite variation in ecological context, patterns of introgression for individual genes are consistent between the two geographic regions. What little variation we see in how introgression is classified appears to be due to differences in the significance thresholds (as opposed to cline shape). There were fewer markers with significantly restricted introgression in Pennsylvania (using either genomic or concordance cline estimates), but all of these markers had restricted introgression in Connecticut. It is likely that the greater extent of introgression in the fine-scale sample from Connecticut increases the power to detect deviations in cline shape (Gompert & Buerkle 2009; Payseur 2010). Consistent locus-specific patterns of introgression between two distant and distinct regions of the hybrid zone suggest that the genomic regions with restricted introgression harbour genes that contribute to prezygotic barriers that serve to isolate the species independent of local ecological selection.

A number of SNPs with restricted introgression are in genes with functional roles that are consistent with these genes contributing to prezygotic barriers between the cricket species (Andrés *et al.* 2013; Larson *et al.* 2013a). For example, five genes encode proteins that may be associated with sperm function or sperm–egg interactions and thus might be responsible for the fertilization barrier between *G. firmus* females and *G. pennsylvanicus* males (Andrés *et al.* 2006, 2008; Maroja *et al.* 2009a; Larson *et al.* 2012a,b). However, it must be remembered that the genome regions marked by SNPs may be large and that observed patterns may be driven by linkage relationships and not by selection acting directly on the gene in which the SNP occurs.

We identified 29 markers that are likely on the X chromosome. These loci are never observed to be heterozygous in any of the males genotyped. In some cases, males that have primarily *G. pennsylvanicus* or heterozygote genotypes had homozygote *G. firmus* genotypes at the presumed X-linked markers (Fig. 3), resulting in higher than expected hybrid indices relative to the observed heterozygosity (Fig. 1). The presumed X-linked markers contributed to the generally low heterozygosity observed in males; removing these markers from estimates of hybrid index and interspecific heterozygosity eliminates observed differences between males and females. Twenty-six of the X-linked genes have significantly restricted introgression in Connecticut ($N = 9$) or both regions ($N = 17$). Sex linked loci have been documented to have reduced introgression in hybrid zones (e.g. Payseur & Nachman 2005; Carling *et al.* 2008) and are hypothesized to play an important role in post-zygotic barriers in the heterogametic sex (i.e. Haldane's rule). But, postzygotic barriers between *G. pennsylvanicus* and *G. firmus* (i.e. either reduced fertility and/or viability) are at best weak, whereas strong prezygotic barriers have been well documented. The role of sex-linkage in the evolution of prezygotic barriers is less clear (Qvarnstrom & Bailey 2008).

Conclusions

We sampled the field cricket hybrid zone at a fine spatial scale across a single habitat patch boundary and found abrupt transitions in allele frequencies for 50 loci over a span of <60 m, despite ongoing gene flow between species. This is a textbook example of the semi-permeability of species boundaries; gene flow can occur throughout some parts of the genome, but in regions that contribute to reproductive barriers, gene flow is restricted. The permeability of species boundaries can also vary across different areas of contact, each with its own environmental and historical context and potentially unique evolutionary trajectory. Surprisingly, the 50 loci that showed abrupt clines (restricted introgression) when sampled at a fine spatial scale show the same pattern when crickets are sampled at a very different scale, in a different geographic region, where different environmental variables contribute to hybrid zone structure. Given the potential for variation, the remarkable consistency across different areas of contact must be due to selection. These observations suggest that nonecological prezygotic barriers (i.e. mate preference, post-mating prezygotic barriers) are most important in maintaining species boundaries in the field cricket hybrid zone. The complex association among multiple reproductive barriers is not consistent with the argument that ecological barriers deserve increased attention (e.g. Schluter 2009). Although ecological factors are clearly important in determining the spatial distribution of species, in many taxa nonecological barriers may be key to the long-term maintenance of species boundaries.

Acknowledgements

We thank E.R. Bondra for assistance in many different ways and L. Cote for assistance with genotyping. This manuscript was greatly improved by thoughtful comments from Nick Barton, Kerry Shaw, Brian Lazzaro, Paulo Alves, Jeff Good, members of the Harrison Laboratory and two anonymous reviewers. This work was supported by an NSF Grant (DEB-0639904) to RGH and an AAUW American Fellowship to ELL.

References

Abbott R, Albach D, Ansell S *et al.* (2013) Hybridization and speciation. *Journal of Evolutionary Biology*, **26**, 229–246.

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.

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, Series B: Biological Sciences*, **275**, 1975–1983.

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.

Barton NH, Hewitt GM (1985) Analysis of hybrid zones. *Annual Review of Ecology and Systematics*, **16**, 113–148.

Baum DA, Shaw KL (1995) Genealogical perspectives on the species problem. In: *Experimental and Molecular Approaches to Plant Biosystematics* (edsHoch PC, Stephenson AG), pp. 289–303. Missouri Botanical Garden, St. Louis, MO.

Benjamini Y, Hochberg Y (1995) Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, **57**, 289–300.

Bierne N, Welch J, Loire E, Bonhomme F, David P (2011) The coupling hypothesis: why genome scans may fail to map local adaptation genes. *Molecular Ecology*, **20**, 2044–2072.

Bolker B (2012) *bbmle: Tools for general maximum likelihood estimation*. R package, Version 1.0 4.1.

Broughton RE, Harrison RG (2003) Nuclear gene genealogies reveal historical, demographic and selective factors associated with speciation in field crickets. *Genetics*, **163**, 1389–1401.

Buerkle CA, Rieseberg LH (2001) Low intraspecific variation for genomic isolation between hybridizing sunflower species. *Evolution*, **55**, 684–691.

Carling MD, Brumfield RT, Webster M (2008) Haldane's rule in an avian system: using cline theory and divergence population genetics to test for differential introgression of mitochondrial, autosomal, and sex-linked loci across the *Passerina* bunting hybrid zone. *Evolution*, **62**, 2600–2615.

Dufková P, Macholán M, Piálek J (2011) Inference of selection and stochastic effects in the house mouse hybrid zone. *Evolution*, **65**, 993–1010.

Fitzpatrick BM (2013) Alternative forms for genomic clines. *Ecology and Evolution*, **3**, 1951–1966.

Gompert Z, Buerkle CA (2009) A powerful regression-based method for admixture mapping of isolation across the genome of hybrids. *Molecular Ecology*, **18**, 1207–1224.

Gompert Z, Buerkle CA (2010) INTROGRESS: a software package for mapping components of isolation in hybrids. *Molecular Ecology Resources*, **10**, 378–384.

Gompert Z, Lucas LK, Nice CC *et al.* (2012) Genomic regions with a history of divergent selection affect fitness of hybrids between two butterfly species. *Evolution*, **66**, 2167–2181.

Harrison RG (1979) Flight polymorphism in the field cricket *Gryllus pennsylvanicus*. *Oecologia*, **40**, 125–132.

Harrison RG (1983) Barriers to gene exchange between closely related cricket species. I. laboratory hybridization studies. *Evolution*, **37**, 245–251.

Harrison RG (1985) Barriers to gene exchange between closely related cricket species. II. life-cycle variation and temporal isolation. *Evolution*, **39**, 244–259.

Harrison RG (1986) Pattern and process in a narrow hybrid zone. *Heredity*, **56**, 337–349.

Harrison RG (1990) Hybrid zones: windows on evolutionary process. In: *Oxford Surveys in Evolutionary Biology*(eds Futuyma D, Antonovics J), pp. 69–128. Oxford University Press, New York, NY.

Harrison R, Arnold S (1982) A narrow hybrid zone between closely related cricket species. *Evolution*, **36**, 535–552.

Harrison RG, Bogdanowicz SM (1997) Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution*, **51**, 493–505.

Harrison RG, Rand DM (1989) Mosaic hybrid zones and the nature of species boundaries. In: *Speciation and its Consequences*(eds Otte D, Endler JA), pp. 111–133. Sinauer, Sunderland, MA.

Harrison RG, Rand DM, Wheeler WC (1987) Mitochondrial-DNA variation in field crickets across a narrow hybrid zone. *Molecular Biology and Evolution*, **4**, 144–158.

Janousek V, Wang LY, Luzynski K *et al.* (2012) Genome-wide architecture of reproductive isolation in a naturally occurring hybrid zone between *Mus musculus musculus* and *M. m. domesticus*. *Molecular Ecology*, **21**, 3032–3047.

Larson EL, Andrés JA, Harrison RG (2012a) Influence of the male ejaculate on post-mating prezygotic barriers in field crickets. *PLoS ONE*, **7**, e46202.

Larson EL, Hume GL, Andrés JA, Harrison RG (2012b) Post-mating prezygotic barriers to gene exchange between hybridizing field crickets. *Journal of Evolutionary Biology*, **25**, 174–186.

Larson EL, Andrés JA, Bogdanowicz SM, Harrison RG (2013a) Differential introgression in a mosaic hybrid zone reveals candidate barrier genes. *Evolution*, **67**, 3653–3661.

Larson EL, Guilherme Becker C, Bondra ER, Harrison RG (2013b) Structure of a mosaic hybrid zone between the field crickets *Gryllus firmus* and *G. pennsylvanicus*. *Ecology and Evolution*, **3**, 985–1002.

Lexer C, Buerkle CA, Joseph JA, Heinze B, Fay MF (2007) Admixture in European *Populus* hybrid zones makes feasible the mapping of loci that contribute to reproductive isolation and trait differences. *Heredity*, **98**, 74–84.

Macholan M, Baird SJE, Dufkova P *et al.* (2011) Assessing multilocus introgression patterns: a case study on the mouse X chromosome in central Europe. *Evolution*, **65**, 1428–1446.

Maroja LS, Andrés JA, Harrison RG (2009a) Genealogical discordance and patterns of introgression and selection across a cricket hybrid zone. *Evolution*, **63**, 2999–3015.

Maroja LS, Andrés JA, Walters JR, Harrison RG (2009b) Multiple barriers to gene exchange in a field cricket hybrid zone. *Biological Journal of the Linnean Society*, **97**, 390–402.

Mayr E (1963) *Animal Species and Evolution*. Belknap Press, Cambridge.

Milne RI, Abbott RJ (2008) Reproductive isolation among two interfertile *Rhododendron* species: low frequency of post-F1 genotypes in alpine hybrid zones. *Molecular Ecology*, **17**, 1108–1121.

Nixon KC, Wheeler QD (1990) An amplification of the phylogenetic species concept. *Cladistics*, **6**, 211–223.

Nolte AW, Gompert Z, Buerkle CA (2009) Variable patterns of introgression in two sculpin hybrid zones suggest that genomic isolation differs among populations. *Molecular Ecology*, **18**, 2615–2627.

Nosil P, Funk DJ, Ortiz-Barrientos D (2009) Divergent selection and heterogeneous genomic divergence. *Molecular Ecology*, **18**, 375–402.

Payseur BA (2010) Using differential introgression in hybrid zones to identify genomic regions involved in speciation. *Molecular Ecology Resources*, **10**, 806–820.

Payseur BA, Nachman MW (2005) The genomics of speciation: investigating the molecular correlates of X chromosome introgression across the hybrid zone between *Mus domesticus* and *Mus musculus*. *Biological Journal of the Linnean Society*, **84**, 523–534.

Phillips BL, Baird SJE, Moritz C, Wiens J (2004) When vicars meet: a narrow contact zone between morphologically cryptic phylogeographic lineages of the rainforest skink, *Carlia rubrigularis*. *Evolution*, **58**, 1536–1548.

Qvarnstrom A, Bailey RI (2008) Speciation through evolution of sex-linked genes. *Heredity*, **102**, 4–15.

Rand DM, Harrison RG (1989a) Ecological genetics of a mosaic hybrid zone—mitochondrial, nuclear, and reproductive differentiation of crickets by soil type. *Evolution*, **43**, 432–449.

Rand DM, Harrison RG (1989b) Molecular Population-Genetics Of Mtdna Size Variation In Crickets. *Genetics*, **121**, 551–569.

Raufaste N, Orth A, Belkhir K *et al.* (2005) Inferences of selection and migration in the Danish house mouse hybrid zone. *Biological Journal of the Linnean Society*, **84**, 593–616.

Ross CL, Harrison RG (2002) A fine-scale spatial analysis of the mosaic hybrid zone between *Gryllus firmus* and *Gryllus pennsylvanicus*. *Evolution*, **56**, 2296–2312.

Ross CL, Harrison RG (2006) Viability selection on overwintering eggs in a field cricket mosaic hybrid zone. *Oikos*, **115**, 53–68.

Schluter D (2009) Evidence for ecological speciation and its alternative. *Science*, **323**, 737–741.

Singhal S, Moritz C (2012) Strong selection against hybrids maintains a narrow contact zone between morphologically cryptic lineages in a rainforest lizard. *Evolution*, **66**, 1474–1489.

Szymura JM, Barton NH (1986) Genetic analysis of a hybrid zone between the fire-bellied toads, *Bombina bombina* and *B. variegata*, near Cracow in southern Poland. *Evolution*, **40**, 1141–1159.

Szymura JM, Barton NH (1991) The genetic-structure of the hybrid zone between the fire-bellied toads *Bombina-bombina* and *B-variegata*—comparisons between transects and between loci. *Evolution*, **45**, 237–261.

Teeter KC, Thibodeau LM, Gompert Z *et al.* (2010) The variable genomic architecture of isolation between hybridizing species of house mice. *Evolution*, **64**, 472–485.

Vines TH, Kohler SC, Thiel A *et al.* (2003) The maintenance of reproductive isolation in a mosaic hybrid zone between the fire-bellied toads *Bombina bombina* and *B-variegata*. *Evolution*, **57**, 1876–1888.

Willett CS, Ford MJ, Harrison RG (1997) Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny. *Heredity*, **79**, 484–494.

Wu CI (2001) The genic view of the process of speciation. *Journal of Evolutionary Biology*, **14**, 851–865.

The study was conceived and designed by E.L.L., C.L.R. and R.G.H. Field work was carried out by C.L.R. and E.L.L. The data were analyzed by E.L.L., T.A.W. and R.G.H. The paper was written by E.L.L. and R.G.H. with contributions from the other authors.

Data accessibility

Sample information, phenotype data, hybrid index, heterozygosity, clines estimates and linkage disequilibrium estimates are available through Dryad (doi:10.5061/dryad.258 h4).

Supporting information

Additional supporting information may be found in the online version of this article.

Fig. S1 Distribution of hybrid indices calculated based on all markers (N = 110), autosomal markers (N = 99) and markers with abrupt (N = 50) and gradual clines (N = 60) (based on genomic cline estimates in Connecticut).

Fig. S2 Individual genomic clines, concordance clines and geographic clines for all 110 genes analysed in Connecticut (N = 260).

Table S1 Results of genomic, concordance and geographic cline analysis for 110 genes genotyped in Pennsylvania and Connecticut regions of the hybrid zone between *Gryllus firmus* and *G. pennsylvanicus*.