Heredity **79** (1997) 484–494 *Received 29 October 1996*

Inferences about the origin of a field cricket hybrid zone from a mitochondrial DNA phylogeny

CHRISTOPHER S. WILLETT†, MICHAEL J. FORD‡§ & RICHARD G. HARRISON†*

†Section of Ecology and Systematics, Corson Hall, Cornell University, Ithaca, NY 14853, U.S.A. and ‡Section of Genetics and Development, Cornell University, Ithaca, NY 14853, U.S.A.

Two closely related eastern North American field crickets, *Gryllus firmus* and *G. pennsylvanicus*, hybridize along a zone that extends from Connecticut and the Hudson River Valley, south along the eastern front of the Appalachian Mountains to at least Virginia. Here we use mitochondrial DNA (mtDNA) sequences to construct a population phylogeny for this pair of hybridizing cricket species. Using a phylogenetic approach, we attempt to discriminate between alternative population histories (secondary contact vs. primary intergradation) leading to formation of the hybrid zone. A strict consensus tree, based on >1600 bp of the COI-COII region of the mtDNA genome, reveals four exclusive groups, which correspond to regional groupings of conspecific crickets. Surprisingly, the mtDNA sequence data do not reveal any synapomorphies for either *G. pennsylvanicus* or *G. firmus*. However, the mtDNA data do reveal a clear north–south split within each of the cricket species, a pattern not seen for morphological or other molecular characters. The biogeographical history of the north–south divergence events remains a puzzle. Observed gene genealogies support a model of secondary contact for the southern part of the hybrid zone. Sequence divergence data argue that lineages currently found in New York and New England were already distinct when this region became habitable following the most recent glaciation.

Keywords: crickets, gene genealogy, *Gryllus*, hybrid zone, mitochondrial DNA, secondary contact.

Introduction

A hybrid zone is an interaction between genetically distinct groups of individuals resulting in at least some offspring of mixed ancestry; pure populations of the two genetically distinct groups occur outside of the zone of interaction (Harrison, 1990). Hybrid zones are frequently narrow compared with the geographical range of the pure parental populations and are often recognized in nature as sets of concordant steep clines. Mayr (1942) proposed two alternative explanations for how hybrid zones may originate, which he termed primary intergradation and secondary contact. Primary intergradation implies differentiation within a series of connected populations leading to the establishment of concordant clinal variation for a set of characters. Hybrid zones are a result of secondary contact when genetically distinct populations encounter one another after some period of geographical isolation. Because Mayr believed that genetic differences most often arise in allopatry, he argued that almost all hybrid zones have arisen as a result of secondary contact, especially hybrid zones in which there are many differences between the interacting groups. The proposition that differentiation occurs only in allopatry was challenged by Endler (1977), who showed that it is theoretically possible for parapatric divergence to result in large differences between populations; therefore, he argued that the striking patterns of variation seen in many hybrid zones could result either from secondary contact or primary intergradation. Barton & Hewitt (1985) emphasized that it is important to distinguish between scenarios for the origin of observed genetic differences and scenarios that describe the recent population history leading to the formation of a hybrid zone. Although the history of differentiation will often be inaccessible,

§Current Address: National Marine Fisheries Service, Northwest Fisheries Science Center, Coastal Zone and Estaurine Division, 2725 Montlake Boulevard, East Seattle, WA 98112, U.S.A.
*Correspondence. E-mail: rgh4@cornell.edu

the origin of current patterns of variation may be inferred from analysis of gene genealogies.

Harrison (1990) proposed a phylogenetic approach to discriminate between alternative population histories leading to formation of a hybrid zone. For this approach to be successful, characters used to infer population history should be neutral (at least with respect to the selective forces that create or maintain hybrid zones), because selection can lead to character convergence and thereby obscure history. DNA sequence data are useful for reconstructing phylogenies in closely related taxa and can often be assumed to behave as neutral characters (in the sense defined above). Molecular phylogenetic analysis has been used to infer the history of hybrid zones in the grasshopper *Chorthippus* (Cooper & Hewitt, 1993; Cooper *et al.*, 1995), to examine the origin of parapatric distributions of rodent species along elevation gradients in South America (Patton & Smith, 1992), and to test hypotheses for the origins of geographical structure in spiny rat populations on either side of a river in the Amazon basin (Patton *et al.*, 1994). Here we use mitochondrial DNA (mtDNA) sequences to construct a population phylogeny for a pair of hybridizing cricket species and to infer their recent population history. MtDNA is a particularly useful marker because it is nonrecombining and evolves at a relatively high rate in insects.

We focus on a well-documented hybrid zone between two closely related eastern North American field crickets, *Gryllus firmus* and *G. pennsylvanicus*. The hybrid zone extends from Connecticut and the Hudson River Valley, south along the eastern front of the Appalachian Mountains to at least Virginia. *Gryllus firmus* is found on sandy soils in coastal plain and piedmont areas, whereas *G. pennsylvanicus* is found inland on loam soils (Harrison & Arnold, 1982; Harrison & Rand, 1989; Rand & Harrison, 1989). Although very similar in appearance, the two species can be distinguished on the basis of morphology. *Gryllus firmus* individuals are generally larger, with lighter tegmina, and longer ovipositors in females (Harrison & Arnold, 1982). Viable and fertile offspring are produced when *G. firmus* males are crossed with *G. pennsylvanicus* females but no offspring result from the reciprocal cross (Harrison, 1983). Allozyme studies have uncovered no fixed differences in alleles between the two species, but there are consistent frequency differences at four loci (Harrison & Arnold, 1982). A study of mtDNA restriction fragment length polymorphisms (RFLPs) in Connecticut populations revealed four species-specific polymorphisms (Harrison *et al.*, 1987). Diagnostic anonymous nuclear RFLPs have also been described, and together with mtDNA, suggest that the two cricket species remain distinct despite ample opportunity for genetic exchange (Harrison & Bogdanowicz, 1997). In this paper we analyse sequence data from the cytochrome oxidase I and II (COI-COII) region of the mtDNA molecule to examine relationships among crickets sampled from populations along the length of the hybrid zone.

Materials and methods

We sampled two to four crickets from each of 10 populations (five *G. pennsylvanicus*, five *G. firmus*) (Fig. 1), most of which are found outside of the hybrid zone. We used the congeneric species *Gryllus veletis* (from Sharon, CT) as the outgroup, and also included a single individual of *Gryllus ovisopis* (from Gainesville, FL). These taxa were selected on the basis of a mitochondrial phylogeny produced by Harrison & Bogdanowicz (1995) which showed that *G. ovisopis*, *G. firmus* and *G. pennsylvanicus* form a single clade, with *G. veletis* clearly outside that group. Cricket DNA was isolated as described previously (Harrison *et al.*, 1987), in some cases omitting the diethylpyrocarbonate and subsequent heating of samples.

We used a PCR-based approach to sequence a 1600 bp region of cricket mtDNA including part of both cytochrome oxidase I and II and the intervening $tRNA^{Leu}$. The primers 1751 (5′-GGATCACCT-GATATAGCATTCCC-3′) and 3772 (5′-GAGAC-CATTACTTGCTTTCAGTCATCT-3′) (see Simon *et al.*, 1994) were used to amplify a double-stranded fragment of about 2 kb (numbers correspond to position of the 3′ end of the primer in the *Drosophila yakuba* mtDNA sequence (Clary & Wolstenholme, 1985)). We set up 100 μL PCR reactions with 3 mM $MgCl_2$, 0.2 mM dNTPs, 50 mM KCl, 20 mM Tris (pH 8.4), 2.5 ng of each primer, two units of Taq polymerase (Gibco-BRL), and 1 μL of DNA. Amplifications involved 35 cycles, each consisting of a 30 s denaturing step at 95°C, a 60 s reannealing step at 47°C, and a 90 s extension step at 72°C in a Perkin Elmer Cetus thermocycler.

To generate single-stranded DNA for sequencing we used two different methods. The first method was asymmetric PCR, for which a single internal primer was added to a new PCR reaction using 5 μL of the double-stranded PCR product as the template DNA. Conditions for this reaction were the same as for initial double-stranded PCR except that only 25 cycles were used and the reannealing temperature was raised to 54°C. The resulting single-stranded

Fig. 1 Collecting localities for *Gryllus firmus* and *G. pennsylvanicus* in the eastern United States. Sharon, CT is SH1 and Guilford, CT is GU2 in Harrison & Arnold (1982). These populations were also sampled in Harrison (1986) and Harrison *et al.* (1987). The Blue Ridge Parkway locality is only a few miles from BR5 of Harrison & Arnold (1982).

DNA was sequenced using Sequenase 2.0 (United States Biochemical) following established protocols. A second method that proved to be more reliable was the use of lambda exonuclease to digest away one strand of the double-stranded PCR product (Higuchi & Ochman, 1989). For this method one primer in the initial PCR reaction (3772) was kinased so that the antisense strand was preferentially digested by lambda exonuclease. This single-stranded product was then sequenced by the same procedure as that used previously. The primers used for sequencing included three primers designed to work for most insects (see Simon *et al.*, 1994): 2191 (5′-CCCCGGTAAAATTAAAATATAAACTTC-3′), 3014 (5′-TCCAATGCACTAATCTGCCATATTA-3′) and 3383 (5′-CATATCTTCARTATCATTGATGTCC-3′), as well as additional primers designed from the cricket sequence: 1981 (5′-TAGTGATATTCCTGGTGCTCG-3′), 2389 (5′-ACTGCAATAATTATAGTTGC-3′), 2587 (5′-AACGTAATGAAAATGGGCAAC-3′), 2797 (5′-TCATGATGTGTAAGCGTCTGG-3′) and 3235 (5′-CTTCTAATAAAAATCGATTG-3′). All of these sequencing primers are on the antisense strand. Although we only sequenced one DNA strand for each cricket, gels were loaded with all A lanes adjacent, all T lanes adjacent, etc., a procedure advocated by Nachman *et al.* (1994) which minimizes scoring errors and makes detection of polymorphisms very straightforward. A representative *G. pennsylvanicus* sequence and the *G. veletis* and *G. ovisopis* sequences have been submitted to GenBank (accession numbers U88332–U88334). The sequences were entered into the computer program MACCLADE (Maddison & Maddison, 1992) and aligned by eye. Phylogeny reconstruction was carried out using the computer program PAUP 3.1.1 (Swofford, 1993). To find the most parsimonious trees we used a heuristic search with 10 replicates of starting trees, with random sequence addition and the tree bisection–reconnection swapping method. *Gryllus veletis* sequence was used to root trees at the completion of the search. PAUP was also used to output a distance matrix for all pairs of sequences. We averaged these pairwise distances across regional groups of populations to obtain average sequence divergence between groups.

Results

We sequenced >1600 bp of mtDNA for 16 *Gryllus pennsylvanicus*, 12 *G. firmus* and one individual from each of two congeneric species, *G. veletis* and *G. ovisopis*. The complete sequence for one individual (a *G. pennsylvanicus* from Ithaca, NY) is given in Fig. 2. The cricket mtDNA sequence is rich in A+T (70 per cent), a compositional bias that is common in insects (Liu & Beckenbach, 1992). In *G. pennsylvanicus* and *G. firmus*, there is a total of 42 polymorphic sites in the sample of 28 individuals (Fig.

3); of these 39 are silent substitutions. The vast majority of the observed differences (37/42) are transitions. At only one polymorphic site are there more than two variants segregating in the populations we have sampled. All of the replacement substitutions and transversional changes are autapomorphic and therefore do not provide any information about relationships among the crickets. When the two congeners are compared to *G. firmus and G. pennsylvanicus*, an additional 74 sites vary, with the vast majority of the changes occurring along the branch joining *G. veletis* to the three other crickets (Table 1). These changes are also largely transitions and silent substitutions; only one change is a replacement substitution and eight are transversions.

When all 30 sequences were included in a parsimony analysis, with *G. veletis* as the designated outgroup, 1253 equally parsimonious trees of 139 steps were obtained. *Gryllus ovisopis* falls outside of the *G. firmus* + *G. pennsylvanicus* clade in all of the shortest trees, but only two sites (460, 963) are synapomorphies for *G. firmus* + *G. pennsylvanicus* (Figs 3 and 4). In an earlier restriction site survey of cricket mtDNA (Harrison & Bogdanowicz, 1995), *G. ovisopis* haplotypes fell within the *G. pennsylvanicus/G. firmus* clade. The mtDNA sequence data strongly support a close relationship among the three species, all of which are univoltine, egg-diapausing crickets.

Within *G. firmus* + *G. pennsylvanicus*, none of the polymorphic sites is diagnostic at the species level. The strict consensus tree (Fig. 4) reveals four exclusive groups, which correspond to regional groupings of conspecific crickets. An exclusive group is one in which all individuals within the group are more closely related to each other than to any individuals outside of the group. We refer to the four groups of crickets as (1) northern *G. pennsylvanicus*, (2) southern *G. pennsylvanicus*, (3) northern *G. firmus* and (4) southern *G. firmus*. In both species, crickets from Pennsylvania, which are in the centre of the sampled range, fall out with the southern representatives of their respective species. Two crickets appear to be exceptions to the general pattern. BRP 1, from the Blue Ridge in Virginia, clusters with northern *G. pennsylvanicus*. WTV 2, which was identified as a *G. firmus* on the basis of morphological characteristics and several anonymous nuclear gene markers, exhibits a mtDNA haplotype that is clearly characteristic of *G. pennsylvanicus*. Relationships among the four exclusive groups are not resolved in the strict consensus tree, but examination of the 1253 trees reveals five equally parsimonious tree topologies for these groups (Fig. 5). In all cases, either the northern *G. pennsylvanicus* clade or the southern *G. pennsylvanicus* clade is the basal group. A neighbour-joining tree reveals the same four groups, with northern *G. pennsylvanicus* as the basal group (tree topology 4 in Fig. 5).

Relationships among the 28 *Gryllus firmus* and *G. pennsylvanicus* can also be represented as a haplo-

CO I →

```
   1 CCCACGAATA AATAATATAA GATTTTGACT TCTACCCCCG TCATTAACCC
  51 TTTTATTAAC CAGAAGNATA GTTGAAAATG GTGCAGGAAC AGGATGAACA
 101 GTTTATCCAC CTTTATCAAC AGGAATTGCT CATGCAGGAG CATCTGTTGA
 151 TTTAGCTATT TTCTCGCTAC ATTTAGCAGG AATTTCCTCA ATTTTAGGAG
 201 CTGTAAATTT TATTACTACC ATAATTAATA TACGAGCACC AGGAATATCA
 251 CTAGATCAAA CGCCATTATT TGTTTGAGCA GTTGGGATTA CAGCTCTTCT
 301 ATTATTATTA TCACTACCAG TTTTAGCTGG TGCTATTACA ATATTACTTA
 351 CAGATCGAAA TTTAAATACA TCATTTTTTG ATCCAGCAGG AGGAGGAGAT
 401 CCAATTTTAT ATCAACATTT ATTTTGATTT TTTGGTCATC CTGAAGTTTA
 451 TATTCTTATC CTACCAGGAT TTGGAATAAT TTCTCATATT ATTAGTCAAG
 501 AAAGAGGAAA AAAGGAAGCC TTTGGTACAC TAGGAATAAT TTACGCAATA
 551 TTAGCCATTG GACTATTAGG ATTTGTTGTA TGAGCACATC ATATATTTAC
 601 GGTTGGAATA GATGTTGATA CCCGAGCATA TTTCACTTCA GCAACTATAA
 651 TTATTGCAGT ACCAACAGGT ATTAAAATTT TTAGATGACT TGCCACTCTT
 701 CATGGATCTC AATTATCATA TAACCCTTCA TTATTATGAT CATTAGGCTT
 751 CGTATTTTTA TTTACTATTG GAGGTTTAAC AGGTATTGTT CTTGCTAACT
 801 CATCAATTGA TATTATTTTA CACGATACTT ATTATGTAGT TGCCCATTTT
 851 CATTACGTTT TATCAATAGG GGCAGTATTT GCTATTATAG CAGGATTTAT
 901 TCATTGATAT CCTTTATTCA CAGGATTAAC AATAAATCCT AAATGATTAA
 951 AAATACAATT TATAGTTATA TTTATTGGAG TAAATTTAAC ATTCTTTCCA
1001 CAACACTTTC TTGGATTAGC AGGAATACCA CGACGATATT CCGATTATCC
1051 AGACGCTTAT ACATCATGAA ATATTTTATC TTCACTAGGA TCTACCATTT
1101 CATTAATTGG TATTATTATA TTAATTTTCA TTCTGTGAGA AAGAATAATT
1151 TCAAATCGAA AGCTTATATT TCCCTTAAAT TTAAATAGAT CTCTAGAATG
1201 ATACCAAAAT CTTCCTCCAG CAGAACACTC CTATTCAGAA TTACCTATTC
                    tRNA Leu →
1251 TATCTAATTA CTAATATGGC AGAAAAGTGC GATGAATTTA AGCTTCATTT
                              CO II →
1301 ATAAAGAATT CCTTTTTTTA GTATTATGGC TACATGATCT AACTTAAATT
1351 TACAAAACAG TTCATCACCC TTAATAGAAC AACTCATCTT TTTTCATGAT
1401 CATACATTAA TAATTCTATT AATAATTACA ATTCTCGTTT CATATATTAT
1451 AACAATATTA TTCTTTAATT CCTATACCAA TCGATTTTTA TTAGAAGGAC
1501 AAACCATTGA AATCATCTGA ACTATTCTAC CTGCAATCAC ATTAATTTTT
1551 ATTGCATTAC CATCATTACG ACTATTATAC TTATTAGATG AATCTATAAA
1601 CCCACTCATT ACA
```

Fig. 2 Mitochondrial DNA sequence from one of the sampled field crickets (ITH 1). The sequence spans part of both the cytochrome oxidase I and II genes and the intervening $tRNA^{Leu}$. The asterisks above the sequence indicate positions of the 42 variable sites (see Fig. 3). The sequence corresponds to positions 1749–3360 of the *Drosophila yakuba* mtDNA sequence (as numbered in Clary & Wolstenholme, 1985).

Table 1 Nucleotide sites at which *Gryllus veletis* and/or *G. ovisopis* differ from *G. pennsylvanicus* and *G. firmus*. The numbers along the top correspond to the numbers in Fig. 2 and the site in bold is a replacement change

	16	40	103	113	133	136	160	166	178	194	256	283	299	302	314	344	385	406	439	451
G. firm./penn.	T	G	T	T	T	A	T	G	A	T	T	T	C	T	C	T	A	T	T	T
G. ovisopis	T	G	T	T	T	A	T	G	G	T	T	T	T	T	C	T	A	T	T	T
G. veletis	C	A	C	C	C	T	C	A	A	C	C	A	C	C	T	C	T	C	C	C
	460	461	466	472	478	487	530	551	562	563	566	580	592	601	625	645	670	697	718	799
G. firm./penn.	C	C	A	T	A	T	C	T	A	C	T	A	T	G	A	T	T	T	A	C
G. ovisopis	T	C	A	T	G	T	C	T	A	C	T	A	T	G	A	T	T	T	A	C
G. veletis	T	T	G	C	A	C	T	C	T	T	C	G	C	T	C	A	A	C	T	T
	802	820	844	850	856	919	940	961	**963**	997	1042	1075	1090	1093	1133	1150	1165	1174	1186	1192
G. firm./penn.	A	A	C	T	C	C	T	T	**T**	T	C	T	A	T	T	T	T	C	T	T
G. ovisopis	A	A	C	T	C	C	T	T	**C**	T	C	T	A	T	T	T	T	C	C	T
G. veletis	T	G	T	C	T	T	C	C	**C**	C	T	C	G	A	C	C	C	T	T	C
	1193	1231	1300	1343	1358	1371	1397	1448	1466	1487	1517	1547	1584	1589	1607	1610				
G. firm./penn.	C	C	T	C	C	T	T	T	T	T	C	T	T	T	C	T				
G. ovisopis	C	C	T	C	C	T	T	T	C	T	C	T	T	T	C	T				
G. veletis	T	T	A	T	T	C	C	C	T	C	T	C	C	C	T	C				

type network (Fig. 6). The four exclusive groups are evident. Local populations tend to have unique sets of haplotypes, and only one haplotype is found in more than a single population. Crandall & Templeton (1993) distinguished between 'tip' and 'interior' haplotypes, the former having only a single mutational connection to other haplotypes and the latter having more than one connection. In the cricket mtDNA haplotype network, interior haplotypes are missing in three of the four groups.

On average, there is <1 per cent sequence divergence between any pair of the four groups of northern and southern *G. pennsylvanicus* and *G. firmus* (Table 2). Comparisons between groups within species do not show smaller divergences than comparisons between species. Average sequence divergence among crickets within groups is small, ranging from 0.1 to 0.32 per cent; the greatest diversity of mtDNA haplotypes is clearly found in southern *G. pennsylvanicus* (Table 2, Fig. 6). *Gryllus ovisopis* averages just >1 per cent sequence divergence when compared with either *G. firmus* or *G. pennsylvanicus*, whereas the *G. veletis* sequence is >5 per cent different from any other sequence (only slightly less than estimates of 6.2–7.1 per cent sequence divergence between *G. veletis* and *G. pennsylvanicus* or *G. firmus* based on restriction site comparisons across the entire mtDNA molecule (Harrison & Bogdanowicz, 1995)).

Discussion

Within the well-defined clade of three North American, univoltine, egg-diapausing field crickets (*G. pennsylvanicus*, *G. firmus* and *G. ovisopis*; see Harrison & Bogdanowicz, 1995), the maximum sequence divergence in the mtDNA COI-COII region is 1.3 per cent, suggesting a relatively recent diversification of this group. Other insects, in which a comparable region of the mtDNA molecule has been sequenced, often show more variation among conspecific populations (Brown *et al.*, 1994; Brower, 1996). Despite low levels of variation, analysis of haplotype diversity in eastern North American *G. pennsylvanicus* and *G. firmus* reveals substantial genetic structuring of populations within and between species. For each of the cricket species, northern and southern groups of populations form distinct clades, with the boundary between 'north' and 'south' occurring between central Pennsylvania and New York/New England. Relationships among these four exclusive groups remain unresolved.

With two exceptions (BRP 1 and WTV 2), the four exclusive groups are composed of individuals

from the same species and the same geographical area. BRP 1 comes from the Blue Ridge Parkway in Virginia, but has a mtDNA haplotype most closely related to haplotypes found in the northern *G. pennsylvanicus* populations. Either gene flow or ancestral polymorphism could explain this result. The basal position of the BRP 1 haplotype within the northern *G. pennsylvanicus* clade (Figs 4 and 6) and its distinctness from haplotypes sampled from New York and Connecticut argue against a recent migration event and favour historical gene flow or ancestral polymorphism as the explanation. If additional sampling were to uncover a similar or identical haplotype in the north, then current gene flow would become a more attractive explanation.

WTV 2 has the morphological characteristics of *G. firmus*, yet it has a mtDNA haplotype most closely related to *G. pennsylvanicus* haplotypes from a neighbouring population in Pennsylvania. Because Wrightsville is close to sites where *G. pennsylvanicus* occurs, introgression of mtDNA from *G. pennsylvanicus* into *G. firmus* is the most likely explanation for this observation. In the hybrid zone in Connecticut, crickets that are identified as *G. firmus* on the basis of morphology often carry *G. pennsylvanicus* mtDNA (Harrison *et al.*, 1987).

Most of the cricket mtDNA haplotypes are restricted to single populations. Even with the relatively small sample sizes, it is clear that historical and current gene flow have not been sufficient to prevent or erase population differentiation. Moreover, not only do crickets from the same collection site tend to be more closely related to each other than to crickets from other collection sites, but also interior haplotypes are missing in both of the *G. pennsylvanicus* clades and in northern *G. firmus*. Absence of interior haplotypes suggests that, subsequent to the divergence of these lineages from a common ancestor, sufficient time has elapsed for base substitutions to have accumulated or lineage sorting to have occurred. The four cricket lineages have presumably been independent over several cycles of glacial advance and retreat (see below). Of course, our sample sizes are small, and missing haplotypes could reflect limited sampling. The only mtDNA haplotype found at more than one site is the interior haplotype in the southern *G. firmus* clade, which is found in coastal North Carolina and

CRICKET	37	55	73	88	97	106	130	220	251	262	286	**296**	**369**	436	475	544	757	796	823	860	871	934	952	979	1006	1036	1060	1135	**1155**	1204	1228	1252	1382	1446	1457	1463	1502	1514	1580	1586	1595	1601
ITH 1	C	A	T	A	A	T	T	C	C	G	G	C	C	T	A	C	T	T	C	T	G	A	A	A	C	A	T	G	A	C	C	A	A	A	A	C	A	C	C	A	T	C
ITH 2	C	A	T	A	A	T	T	C	C	G	G	C	C	T	A	C	T	T	C	T	G	A	A	A	C	A	T	G	A	C	C	A	A	A	A	C	A	C	C	A	T	C
ITH 3	C	A	T	A	A	T	T	C	C	G	G	C	C	T	A	C	T	T	C	T	G	A	A	A	C	A	T	G	A	C	C	A	A	A	A	C	A	C	C	A	T	C
SH 1	C	A	T	A	A	T	T	C	C	G	.	C	C	T	A	C	T	T	C	T	G	A	A	A	T	A	T	G	A	C	C	A	A	A	A	C	A	T	C	A	T	T
SH 2	C	A	T	A	A	T	T	C	C	G	A	C	C	T	A	C	T	T	C	C	G	A	A	A	C	A	T	G	A	C	C	A	A	A	A	C	A	C	C	A	T	C
SH 3	C	A	T	A	A	T	T	C	C	G	.	C	C	T	A	C	T	T	C	T	G	A	A	A	T	A	T	G	A	C	C	A	A	A	A	C	A	C	C	A	T	T
SH 4	C	A	T	A	A	T	T	C	C	G	.	C	C	T	A	C	T	T	C	T	G	A	A	A	T	A	T	G	A	C	C	A	A	A	A	C	A	C	C	A	T	T
BRP 1	C	A	T	A	A	T	T	C	C	G	A	C	C	T	A	C	T	C	C	T	G	A	A	A	C	A	T	G	A	C	C	A	A	A	A	C	A	C	C	A	T	C
BRP 2	C	A	C	G	A	T	T	C	C	A	A	C	C	T	T	C	T	C	C	T	G	A	G	A	T	A	T	A	A	T	C	A	G	A	A	T	A	C	C	A	T	C
BRP 3	C	A	C	G	G	T	T	C	C	A	A	C	G	A	T	T	T	C	C	T	G	G	G	A	T	A	T	G	A	T	C	A	G	A	A	C	A	C	C	A	T	T
ASH 1	C	A	C	A	A	T	T	T	C	A	A	C	C	T	T	C	T	C	C	T	A	A	G	A	T	A	C	G	A	T	C	A	G	A	A	C	A	C	C	A	C	C
ASH 2	C	A	C	A	A	T	T	T	C	A	A	C	C	T	T	C	T	C	C	T	G	A	G	A	T	A	C	G	A	T	C	A	G	A	A	C	A	C	C	A	C	C
ASH 3	T	A	C	A	A	T	T	C	C	A	A	C	C	T	T	C	T	C	C	T	G	A	G	A	T	A	C	G	A	T	C	A	G	A	A	C	A	C	T	A	T	C
COR 1	C	A	C	G	A	T	T	C	C	A	A	C	C	T	T	C	T	C	C	T	G	A	G	A	T	A	T	G	A	T	C	A	G	A	A	C	A	C	C	A	C	C
COR 2	C	A	C	G	A	C	T	C	C	A	A	C	C	T	T	C	T	C	C	T	G	A	G	A	T	A	T	G	A	T	C	A	G	A	A	C	A	C	C	A	T	C
COR 3	C	A	C	G	A	T	T	C	C	A	A	C	C	T	T	C	T	C	C	T	G	A	G	A	T	A	T	G	A	T	T	A	G	A	A	C	A	C	C	G	T	C
WTV 2	C	A	C	G	A	T	T	C	C	A	A	C	C	T	T	C	T	C	C	T	G	A	G	A	T	A	T	G	A	T	C	A	A	A	A	C	A	C	C	A	.	C
WTV 1	C	A	C	A	A	T	T	C	.	G	A	C	C	T	T	C	T	C	C	T	G	A	A	G	C	G	C	G	A	T	C	A	G	G	A	C	G	C	C	A	C	C
WTV 3	C	A	C	A	A	T	T	C	.	G	A	C	C	T	T	C	T	C	C	T	G	A	A	G	C	G	C	G	A	T	C	A	G	A	A	C	A	C	C	A	C	C
BSV 1	C	A	C	A	A	T	T	C	T	G	A	C	C	T	T	C	C	C	C	T	G	A	A	G	C	G	C	G	A	T	C	A	G	A	A	C	A	C	C	A	C	C
BSV 2	C	A	C	A	A	T	T	C	T	G	A	C	C	T	T	C	C	C	C	T	G	A	A	G	C	G	C	G	A	T	C	A	G	A	A	C	A	C	C	A	C	C
LEX 1	C	A	C	A	A	T	T	C	T	G	A	C	C	T	T	C	T	C	C	T	G	A	A	G	C	G	C	G	G	T	C	A	G	A	A	C	A	C	C	A	C	C
PAN 1	C	A	C	A	A	T	T	C	C	G	A	C	C	T	T	C	T	C	C	T	G	A	A	G	C	G	C	G	A	T	C	A	G	A	A	C	A	C	C	A	C	C
PAN 2	C	A	C	A	A	T	T	C	T	G	A	C	C	T	T	C	T	C	C	T	G	A	A	G	C	G	C	G	A	T	C	A	G	A	A	C	A	C	C	A	C	C
PAN 3	C	A	C	A	A	T	T	C	T	G	A	C	C	T	T	C	T	C	C	T	G	A	A	G	C	G	C	G	A	T	C	T	G	A	A	C	A	C	C	A	C	C
GU 1	C	A	C	A	A	T	C	C	C	G	A	C	C	T	T	C	T	C	C	T	G	A	A	A	C	A	C	G	A	T	C	A	G	A	T	C	A	C	C	A	T	C
GU 2	C	G	C	A	A	T	C	C	C	G	A	C	C	T	T	C	T	C	T	T	G	A	A	A	C	A	C	G	A	T	C	A	G	A	A	C	A	C	C	A	T	C
GU 3	C	A	C	A	A	T	C	C	C	G	A	A	C	T	T	C	T	C	C	T	G	A	A	A	C	A	C	G	A	T	C	A	G	A	A	C	A	C	C	A	T	C
G. ovisopis	C	A	T	A	A	C	T	C	C	A	G	C	C	T	T	C	T	C	C	T	A	A	A	A	C	A	C	G	A	T	C	A	G	A	A	C	A	C	C	A	T	T
G. veletis	C	A	T	A	A	T	T	T	C	A	A	C	C	C	T	T	C	C	T	T	G	A	A	A	C	A	C	A	A	T	T	A	A	A	A	C	A	T	T	A	T	T

Fig. 3 Variable sites in *Gryllus firmus and G. pennsylvanicus* populations. Site numbers along the top correspond to the numbers in Fig. 2 and the three sites in bold are replacement changes. Bases in the table are highlighted to emphasize the differences between taxa. Dots indicate missing data. The bold horizontal lines indicate divisions between species.

central Pennsylvania. The presence of identical haplotypes in two geographically distant localities suggests ongoing gene flow, recent range expansion, or a recent selective sweep.

What should be the relationships among cricket populations under the two alternative hypotheses for the origin of hybrid zones? A general model of secondary contact involves the coming together of two allopatric taxa, which have previously diverged in a number of traits. The geographical context in which the initial divergence occurred is not specified. If two genetically distinct lineages existed prior to the formation of a hybrid zone, then individuals within these two lineages may form exclusive groups. Exclusive groups are expected only if the hybridizing taxa evolved independently of each other for a period of time long enough for diagnostic differences to accumulate. In contrast, if a hybrid zone forms as a result of divergence across a selection gradient in the absence of extrinsic barriers (primary intergradation), a gene genealogy may initially reveal that individuals have as their closest relatives other individuals that are close geographically, not necessarily those of the same race, subspecies, or species.

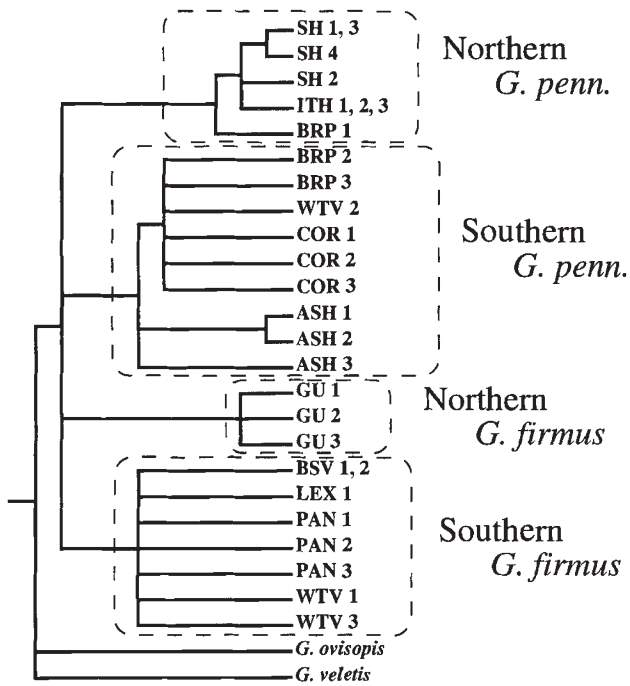

Fig. 4 Strict consensus tree for all taxa. The four exclusive groups are indicated by dashed lines and named by the region and species of most of the individuals that fall within them. The *Gryllus veletis* haplotype was used to root the tree.

According to these arguments, if the current cricket hybrid zone is a consequence of secondary contact, all individuals of *G. pennsylvanicus* should belong to a single clade, as should all *G. firmus* individuals, with the root of the tree falling between the two species clades (Fig. 7). However, if the hybrid zone originated by primary intergradation (differentiation *in situ*), we expect a tree like the one depicted on the right in Fig. 7, in which each population of *G. pennsylvanicus* groups together with the adjacent *G. firmus* population.

The gene genealogies obtained from mtDNA sequence data (Fig. 5) do not coincide with either of the trees shown in Fig. 7. However, the mtDNA data are more consistent with a model of secondary contact. Populations that are geographically very close to one another but of different species are not similar genetically (for example the Connecticut *G. pennsylvanicus* and *G. firmus*). Furthermore, if we exclude from the phylogenetic analysis crickets from

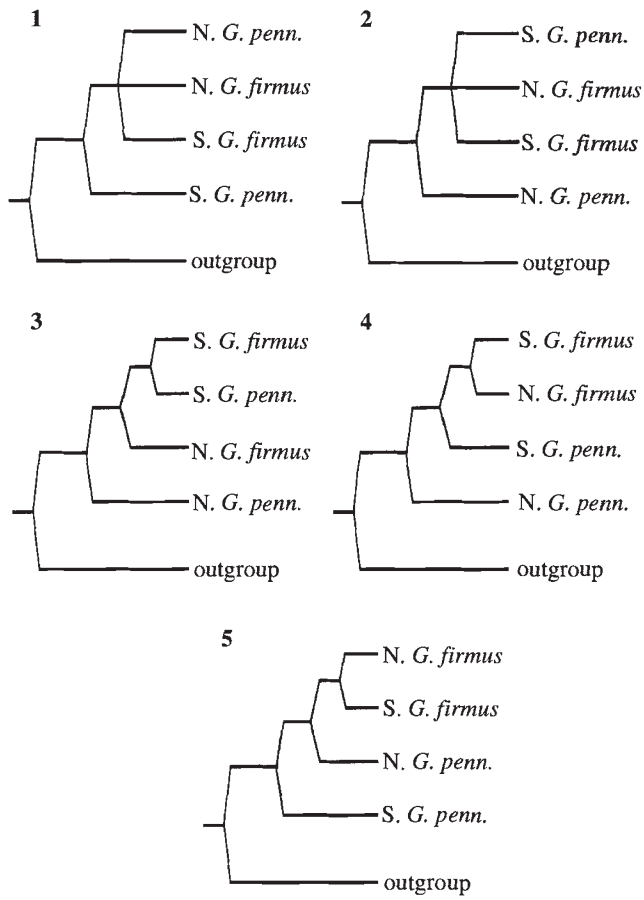

Fig. 5 The five alternative tree topologies found among the 1253 equally parsimonious trees, showing possible relationships among the four exclusive groups of crickets.

New York and New England, both *G. pennsylvanicus* and *G. firmus* do form exclusive groups. Thus, patterns of relationships for crickets from Pennsylvania and further south are exactly what we would predict based on a model of recent secondary contact (Fig. 7).

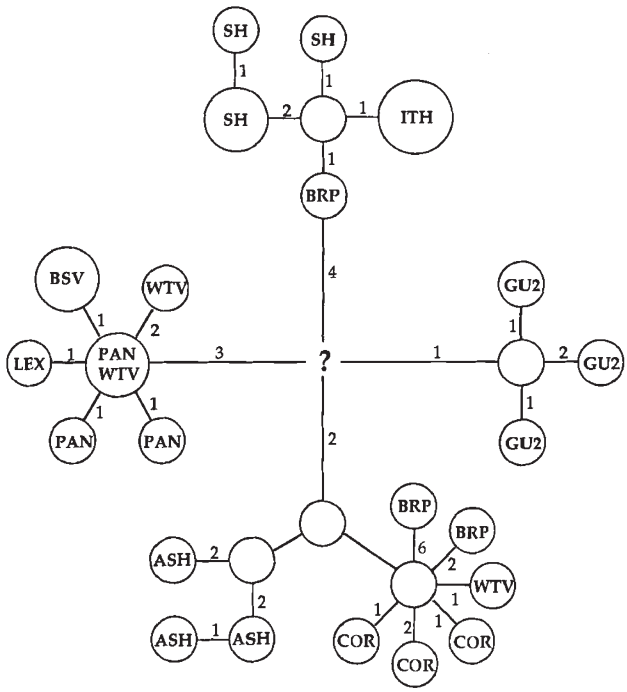

Fig. 6 Mitochondrial DNA haplotype network for *Gryllus pennsylvanicus* and *G. firmus*. The size of each circle reflects the number of individuals of a given haplotype (n = 1,2,3). Empty circles are putative missing haplotypes. Numbers along lines connecting haplotypes are numbers of base substitutions in the 1611 bp segment of cricket mtDNA. The relationships among the four exclusive groups are unresolved.

Discrepancies between observed mtDNA relationships and the predictions of Fig. 7 might indicate that the 'real' population history does not conform to either of the simple models. Alternatively, the mtDNA phylogeny may not reflect the true population history. Surprisingly, the mtDNA sequence data do not reveal any synapomorphies for either *G. pennsylvanicus* or *G. firmus*. If species differences evolved in allopatry following a vicariance event, and subsequent secondary contact explains the current species distributions and interactions, population subdivision between northern and southern crickets within each species must have occurred very soon after the initial vicariance event. Otherwise, we should be able to identify synapomorphies for each of the species. Alternative scenarios, consistent with the basal position of one of the *G. pennsylvanicus* clades (see Fig. 5), involve the splitting of an ancestral population into two daughter lineages, one of which gave rise to either northern or southern *G. pennsylvanicus* 'populations' and a second that gave rise to the other *G. pennsylvanicus* clade and to *G. firmus*. In these models, the initial divergence is not 'between species', and *G. pennsylvanicus* is expected to be paraphyletic with respect to *G. firmus* (Neigel & Avise, 1986). Furthermore, the mtDNA data suggest that the northern and southern parts of the hybrid zone may represent independent secondary contact events that have since merged into a continuous hybrid zone. If true, there must be a secondary mtDNA cline or discontinuity within each 'species' where the northern and southern haplotype arrays come into contact. The southern *G. pennsylvanicus* clade harbours the most diverse array of mtDNA haplotypes, and the maximum sequence divergence

Table 2 Average percentage sequence divergence within and between species of *Gryllus*. For the four exclusive groups within *G. pennsylvanicus* and *G. firmus*, sequence divergences were calculated by averaging all pairwise comparisons within a group (numbers on the diagonal) or between members of two groups (numbers off the diagonal). The ranges of values for sequence divergences are given in parentheses. The exclusive groups were defined as in Fig. 4. *Gryllus veletis* and *G. ovisopis* are represented by only one sequence and therefore the within-species polymorphism could not be calculated for these two taxa

	N.*G.p.*	S.*G.p.*	N.*G.f.*	S.*G.f.*	*G.o.*
G. veletis	5.30 (5.22–5.35)	5.33 (5.22–5.41)	5.26 (5.22–5.28)	5.40 (5.35–5.48)	5.41
G. ovisopis (*G.o.*)	1.08 (1.06–1.18)	1.11 (0.99–1.31)	1.02 (0.99–1.12)	1.14 (1.06–1.18)	
Southern *G. firmus* (S.*G.f.*)	0.73 (0.50–0.87)	0.64 (0.37–0.99)	0.42 (0.31–0.50)	**0.10** (0–0.19)	
Northern *G. firmus* (N.*G.f.*)	0.60 (0.44–0.75)	0.57 (0.44–0.87)	**0.17** (0.12–0.19)		
Southern *G. pennsylvanicus* (S.*G.p.*)	0.72 (0.44–0.99)	**0.32** (0.06–0.68)			
Northern *G. pennsylvanicus* (N.*G.p.*)	**0.13** (0–0.25)				

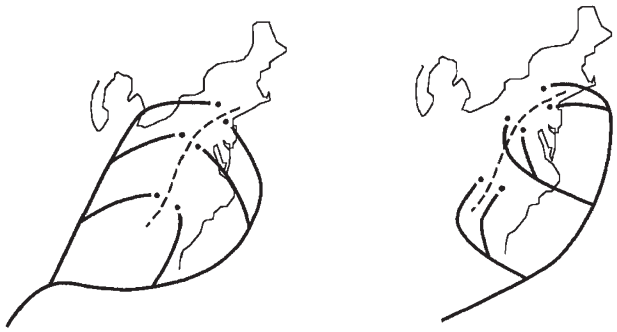

Fig. 7 Expected phylogenetic relationships given two models of hybrid zone origin. The two tree topologies are those expected for models of secondary contact (left) or primary intergradation (right). Modified from Harrison (1990). In both figures, lineage splitting events are assumed to occur from south to north, but other scenarios are possible.

within this clade (0.68 per cent) is approximately equal to the average sequence divergence between groups. Thus, southern *G. pennsylvanicus* may have persisted as a single large population or as a subdivided population during the late Pleistocene. The other three clades are less diverse and have probably arisen more recently or experienced recent population bottlenecks or selective sweeps. For any single genetic marker or set of linked markers, random sorting of ancestral polymorphisms (Avise *et al.*, 1984) could obscure historical relationships among the four major groups of crickets, if (as suggested above) the lineage splitting events giving rise to these four clades occurred in very rapid succession. If ancestral polymorphism is a problem at this level, recovering the 'true' relationships among field cricket populations will require multiple independent phylogenies. Data from allozymes and anonymous nuclear RFLP markers (Harrison & Arnold, 1982; R. G. Harrison & S. M. Bogdanowicz, unpubl. data) reveal relative homogeneity of allele frequencies within each of the cricket species and significant frequency differences between species. There is no evidence of a north–south split. Thus, these data are consistent with a scenario in which an initial lineage splitting event gave rise to *G. pennsylvanicus* and *G. firmus*, with the extensive hybrid zone a result of secondary contact between these species. Alternatively, contemporary gene flow, within but not between species, could account for patterns of allele frequency variation. However, the distinctness of mtDNA haplotypes between the northern and southern clades of each species argues against the gene flow hypothesis, unless we invoke either (i) strong selection against introgression of mtDNA haplotypes or (ii) higher rates of dispersal for males than females, which could homogenize allele frequencies at nuclear gene loci but not for mtDNA.

The northern part of the current ranges of both species was either covered by glaciers or otherwise uninhabitable by these crickets as little as 15 000 years ago (Davis, 1976; Dyke & Prest, 1987). Pollen evidence suggests that when the Wisconsin glaciation was at its maximum, tundra vegetation extended considerably south of the glacial margin (Maxwell & Davis, 1972; Watts, 1979, 1983), and that spruce and pine forests (indicative of a climate that crickets would tolerate) did not return to New England until 12 000–10 000 years ago. Clearly the northern field cricket populations have been established since the glaciers retreated. The estimated net mtDNA sequence divergence between northern *G. pennsylvanicus* and *G. firmus* is 0.45 per cent, where the net divergence is the total divergence minus the average divergence within the two groups. If the rate of evolution of cricket mtDNA is ≈ 1 per cent per million years per lineage (Brower (1994) suggests that the rate for arthropods is 1.2 per cent), then the observed difference would take 187500 years to accumulate. Although molecular clock calibrations are obviously quite crude, these calculations suggest that northern populations of the two cricket species have been distinct for a period of time much longer than the 10 000–15 000 years that they have been able to live in their current locations. Therefore, it appears that the lineages representing northern *G. pennsylvanicus* and *G. firmus* substantially predate the most recent glacial advance and that the northern part of the current hybrid zone is a result of secondary contact.

Conclusions

A mtDNA phylogeny of the hybridizing crickets *G. firmus* and *G. pennsylvanicus* reveals clear genetic structure both within and between species. The phylogeny argues against a model of primary intergradation for the origin of the field cricket hybrid zone; current hybrid zone structure is more probably a result of the coming together of already differentiated lineages. However, genealogical data cannot provide insights into the geographical context in which differentiation originally occurred and therefore do not address the issue of whether speciation is allopatric or nonallopatric. The data also reveal a north–south split within each of the two cricket

species. It would seem that northern and southern populations of both cricket species have been evolutionarily independent lineages for a substantial length of time, but the biogeographical history of the north–south divergence events remains a puzzle.

Acknowledgements

We thank Steve Bogdanowicz for technical assistance and advice. This work was supported by the Systematic Biology Program at NSF (grant BSR-8906707 to R.G.H.). C.S.W. was supported by an NSF graduate fellowship. M.J.F. was supported by a Howard Hughes Medical Institute predoctoral fellowship.

References

AVISE, J. C., NEIGEL, J. E. AND ARNOLD, J. 1984. Demographic influences on mitochondrial DNA lineage survivorship in animal populations. *J. Mol. Evol.*, **20**, 99–105.

BARTON, N. H. AND HEWITT, G. M. 1985. Analysis of hybrid zones. *Ann. Rev. Ecol. Syst.*, **16**, 113–148.

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. U.S.A.*, **91**, 6491–6495.

BROWER, A. V. Z. 1996. Parallel race formation and the evolution of mimicry in *Heliconius* butterflies: a phylogenetic hypothesis from mitochondrial DNA sequences. *Evolution*, **5**, 195–221.

BROWN, J. M., PELLMYR, O., THOMPSON, J. N. AND HARRISON, R. G. 1994. Phylogeny of *Greya* (Lepidoptera: Prodoxidae), based on nucleotide sequence variation in mitochondrial cytochrome oxidase I and II: congruence with morphological data. *Mol. Biol. Evol.*, **11**, 128–141.

CLARY, D. O. AND WOLSTENHOME, D. R. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.*, **22**, 252–272.

COOPER, S. J. B. AND HEWITT, G. M. 1993. Nuclear DNA sequence divergence between parapatric subspecies of the grasshopper *Chorthippus parallelus*. *Insect Mol. Biol.*, **2**, 185–194.

COOPER, S. J. B., IBRAHIM, K. M. AND HEWITT, G. M. 1995. Postglacial expansion and genome subdivision in the European grasshopper *Chorthippus parallelus*. *Mol. Ecol.*, **4**, 49–60.

CRANDALL, K. A. AND TEMPLETON, A. R. 1993. Empirical tests of some predictions from coalescent theory with applications to intraspecific phylogeny reconstruction. *Genetics*, **134**, 959–969.

DAVIS, M. B. 1976. Pleistocene biogeography of temperate deciduous forests. *Geosci. and Man*, **13**, 13–26.

DYKE, A. S. AND PREST, V. K. 1987. Late Wisconsinan and Holocene history of the Laurentide ice sheet. *Geogr. Phys. Quaternaire*, **41**, 237–264.

ENDLER, J. A. 1977. *Geographic Variation, Speciation, and Clines*. Princeton University Press, Princeton, NJ.

HARRISON, R. G. 1983. Barriers to gene exchange between closely related species. I. Laboratory hybridization studies. *Evolution*, **37**, 245–251.

HARRISON, R. G. 1986. Pattern and process in a narrow hybrid zone. *Heredity*, **56**, 337–349.

HARRISON, R. G. 1990. Hybrid zones: windows on evolutionary process. In Futuyma, D. and Antonovics, J. (eds) *Oxford Surveys in Evolutionary Biology*, **7**, 69–128. Oxford University Press, Oxford.

HARRISON, R. G. AND ARNOLD, J. 1982. A narrow hybrid zone between closely related cricket species. *Evolution*, **36**, 535–552.

HARRISON, R. G. AND RAND, D. M. 1989. Mosaic hybrid zones and the nature of species boundaries. In: Otte, D. and Endler, J. A. (eds) *Speciation and Its Consequences*, pp. 111–133. Sinauer, Sunderland, MA.

HARRISON, R. G. AND BOGDANOWICZ, S. M. 1995. Mitochondrial phylogeny of North American field crickets: perspectives on the evolution of life cycles, songs and habitat associations. *J. Evol. Biol.*, **8**, 209–232.

HARRISON, R. G. AND BOGDANOWICZ, S. M. 1997. Patterns of variation and linkage disequilibrium in a field cricket hybrid zone. *Evolution*, **51**, 493–505.

HARRISON, R. G., RAND, D. M. AND WHEELER, W. C. 1987. Mitochondrial DNA variation in field crickets across a narrow hybrid zone. *Mol. Biol. Evol.*, **4**, 144–158.

HIGUCHI, R. AND OCHMAN, H. 1989. Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. *Nucl. Acids Res*, **17**, 5865.

LIU, H. AND BECKENBACH, A. T. 1992. Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Mol. Phylogen. Evol.*, **1**, 41–52.

MADDISON, W. P. AND MADDISON, D. R. 1992. MACCLADE 3.0. *Analysis of Phylogeny and Character Evolution.* Sinauer Associates, Sunderland, MA.

MAXWELL, J. A. AND DAVIS, M. B. 1972. Pollen evidence of Pleistocene and Holocene vegetation on the Allegheny Plateau, Maryland. *Quarternary Res.*, **2**, 506–530.

MAYR, E. 1942. *Systematics and the Origin of Species*. Columbia University Press, New York.

NACHMAN, M. W., BOYER, S. N., SEARLE, J. B. AND AQUADRO, C. F. 1994. Mitochondrial DNA variation and the evolution of Robertsonian chromosomal races of house mice, *Mus domesticus*. *Genetics*, **136**, 1105–1120.

NEIGEL, J. E. AND AVISE, J. C. 1986. Phylogenetic relationships of mitochondrial DNA under various demographic models of speciation. In: Karlin, S. and Nevo, E. (eds) *Evolutionary Processes and Theory*, pp. 515–534. Academic Press, New York.

PATTON, J. L. AND SMITH, M. F. 1992. MtDNA phylogeny of Andean mice: a test of diversification across ecological gradients. *Evolution*, **46**, 174–183.

PATTON, J. L., DA SILVA, M. N. F. AND MALCOLM, J. R. 1994. Gene genealogy and differentiation among arboreal spiny rats (Rodentia: Echimyidae) of the Amazon Basin: a test of the riverine barrier hypothesis. *Evolution*, **48**, 1314–1323.

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

SIMON, C., FRATI, F., BECKENBACH, A., LIU, H. AND FLOOK, P. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.*, **87**, 651–701.

SWOFFORD, D. L. 1993. PAUP 3.1.1. Illnois Natural History Survey, Champaign, IL.

WATTS, W. A. 1979. Late Quaternary vegetation of central Appalachia and the New Jersey coastal plain. *Ecol. Monogr.*, **49**, 427–469.

WATTS, W. A. 1983. Vegetational history of the eastern United States 25,000 to 10,000 years ago. In: Porter, S. C. (ed.) *Late-Quarternary Environments of the United States*, vol. 1, *The Late Pleistocene*, pp. 294–310. University of Minnesota Press, Minneapolis, MN.