Molecular divergence between *Gryllus rubens* and *Gryllus texensis*, sister species of field crickets (Orthoptera: Gryllidae)

D.A. Gray

Department of Biology, California State University, 18111 Nordhoff Street, Northridge, California 91330-8303, United States of America

P. Barnfield, M. Seifried, M.H. Richards

Department of Biological Sciences, Brock University, St. Catharines, Ontario, Canada L2S 3A1

Abstract—We assess the degree of sequence divergence in the maternally inherited mitochondrial cytochrome c oxidase I (COI) and cytochrome b (CytB) genes between two sister species of field crickets, *Gryllus rubens* Scudder, 1902 and *Gryllus texensis* Cade and Otte, 2000. We analyzed 1460 base pairs from 10 individuals of each species; individuals were sampled from areas of both allopatry and sympatry. Overall average pairwise mitochondrial sequence divergence between species was 1.4% ± 0.1% (mean ± SD); however, there was almost an order of magnitude more divergence in COI (2.59% ± 2.25%) than in CytB (0.35% ± 0.24%). *Gryllus texensis* appears to harbor a much greater level of genetic variation than does *G. rubens*. Phylogenetic trees constructed from these sequences show reasonable separation of species; however, sequences are not reciprocally monophyletic. Gene tree polyphyly may reflect recent species-level divergence and (or) interspecific gene flow. The pattern of sequence divergence and genetic variation in these taxa is consistent with allopatric or peripatric speciation in Pleistocene glacial refugia in the southeastern (*G. rubens* ancestral lineage) and southcentral United States (*G. texensis* ancestral lineage).

Résumé—Nous évaluons le degré de divergence des séquences dans les gènes mitochondriaux d’origine maternelle, cytochrome c oxydase I (COI) et cytochrome b (CytB), chez les espèces soeurs de grillons des champs *Gryllus rubens* Scudder, 1902 et *Gryllus texensis* Cade et Otte, 2000. Nous avons analysé 1460 paires de bases chez 10 individus de chaque espèce, prélevés dans des zones d’allopatrie et de sympatrie. La divergence globale des séquences mitochondriales, paire par paire, entre les espèces est de 1,4 % ± 0,1 % (moyenne ± ET); cependant, la divergence de COI (2,59 % ± 2,25 %) est d’un ordre de grandeur plus importante que celle de CytB (0,35 % ± 0,24 %). *Gryllus texensis* semble posséder un niveau beaucoup plus élevé de variation génétique que *G. rubens*. Les arbres phylogénétiques élaborés à partir de ces séquences montrent une séparation adéquate des espèces, mais les séquences ne sont pas réciproquement monophylétiques. La polyphylie des arbres génétiques peut indiquer une divergence récente au niveau des espèces et (ou) un flux génétique interspécifique. Les patrons de divergence des séquences et de variation génétique chez ces taxons sont compatibles avec une spéciation allopatrique ou peripatrique dans les refuges glaciaires du pléistocène dans le sud-est ( lignée ancestrale de *G. rubens*) et le centre-sud ( lignée ancestrale de *G. texensis*) des États-Unis.

[Traduit par la Rédaction]

Introduction

Despite ongoing debate regarding suitable and operationally useful species definitions (e.g., four separate chapters in Howard and Berlocher 1998), it is widely agreed that speciation is a process of increasing divergence contingent upon minimal or nonexistent gene flow (Coyne

Received 21 April 2005. Accepted 7 April 2006.

1Corresponding author (e-mail: dave.gray@csun.edu).
2Present address: Program in Developmental Biology, Room 5128, Elm Wing, The Hospital for Sick Children, 555 University Avenue, Toronto, Ontario, Canada M5G 1X8.
have elected to use maternally inherited mitochondrial DNA (hereafter mtDNA). The characterization of the molecular divergence between species pairs (Shaw 2002). Our investigation of mtDNA therefore represents only a subset of the molecular divergence between these species and should be treated as a first approximation.

A recent molecular phylogeny of North American field crickets, based on mtDNA sequence data for the complete cytochrome b (CytB) gene (1036 base pairs, bp) and a 498-bp fragment of the 16S rRNA gene, showed that G. rubens and G. texensis are in fact closely related (Huang et al. 2000). That conclusion was based on analysis of two G. rubens individuals from Gainesville, Florida, and two G. texensis individuals from Austin, Texas, both allopatric sampling localities.

In this paper we report the results of our analyses of another portion of mtDNA, the cytochrome c oxidase I (COI) gene, in addition to a portion of the CytB gene for a larger sample of crickets from both allopatric and sympatric collection localities. One goal of our research was to provide another estimate of divergence between these species at the molecular level, with the expectation that this may enable us to better estimate the timing of speciation. It has been previously shown that for relatively recent divergence times mtDNA evolves in a fairly clocklike manner, despite nonlinearity due to saturation at longer time scales (Ho et al. 2005). Geologically calibrated COI divergence rate estimates in arthropods show considerable variation, but 2% per million years is a typical value; estimates often range from 1.4% to 2.6% (Brower 1994; Juan et al. 1995; Caccone and Sbordoni 2001; Farrell 2001; Ho et al. 2005). A second goal of the current research was to examine molecular variation for a substantially larger sample of crickets from a broader geographic area than that tested by Huang et al. (2000), generate a gene tree, and assess monophyly of the gene sequences.

**Methods**

**Sampling**

We extracted, amplified, and sequenced DNA from 10 individuals of each species (see Table 1).
In all cases species identity was confirmed via behavioral assay of the male calling song or female response to song, as well as behavioral assay of siblings (Gray and Cade 2000).

DNA extraction, PCR amplification, and sequencing

Total genomic DNA was isolated from thoraces of individual specimens using the DNeasy Tissue Kit (QIAGEN Inc., catalog No. 69504) after initially freezing the tissue in liquid nitrogen. We amplified portions of two mitochondrial genes, those encoding cytochrome c oxidase I (COI) and cytochrome b (CytB), using primers (see Simon et al. 1994) obtained from the University of British Columbia Biotechnology Laboratory. COI primers were mtD-8 (aliases C1-J-2183, Jerry; 5′ to 3′ sequence, CAA CAT TTA TTT TGA TTT TTT GG) and mt-D12 (aliases L2-N-3014, Pat; 5′ to 3′ sequence, TCC AAT GCA CTA ATC TGC CAT ATT A). CytB primers were mtD-18 (aliases C2-N-3661, Barbara; 5′ to 3′ sequence, CCA CAA ATT TCT GAA CAT TGA CCA) and mtD-25 (aliases CB-J-10612, CB1L; 5′ to 3′ sequence, CCA TCC AAC ATC TCA GCA TGA TGA AA).

Both genes were amplified in 20 µL reactions (30 cycles of separation at 94 °C, primer annealing at 52–54 °C, and extension at 72 °C). Negative controls using sterile water as a template were included with all reactions to assess contamination. Amplification products of the COI and CytB genes were run out in 1.5% agarose gels and purified using the QIAEX Gel Extraction Kit (QIAGEN Inc., catalog No. 20021). Specimens were sequenced using the same primers used for PCR amplification, in both directions, to detect and correct sequencing errors. Automated sequencing was performed by the York University Core Molecular Biology Facility in Toronto, Ontario. All sequences have been submitted to GenBank (accession Nos. AY234789–AY234808).

Sequence analysis

Sequences were edited and aligned using the computer programs BioEdit (Hall 1999) and CLUSTAL W (Thompson et al. 1994). Alignments and phylogenies were constructed with reference to outgroup sequences from three other Gryllus species, G. veletis (Alexander and Bigelow, 1960) (GenBank accession Nos. GVU8834, AF248678), G. pennsylvanicus Burmeister, 1838 (GPV88332, AF248675), and G. ovisopis T.J. Walker, 1974 (GOU88333, AF248673, AF248674). The consensus alignment for COI was 716 nucleotides (nt) long, and that for CytB was 744 nt long, the total alignment length being 1460 nt. The alignment included 170 variable characters, of which 148 were parsimony-informative.

To compare differences in variability between COI and CytB, we used ANOVA to examine the mean genetic divergence among all possible pairs of sequences of each gene. To compare population genetic differentiation between the two cricket species, we used AMOVA as implemented in the computer program Arlequin version 2.000 (Schneider et al. 2000).

Phylogenies were constructed using neighbour-joining, maximum likelihood (ML),
and parsimony methods as implemented in PAUP* 4.0b10 (Swofford 2002) and Bayesian analysis as implemented in MrBayes version 3,0b4 (Huelsenbeck and Ronquist 2001). For the parsimony tree calculations in PAUP*, initial trees were found by stepwise addition followed by branch-and-bound search with and without the constraint that *G. rubens* and *G. texensis* sequences represent separate monophyletic clades. For the ML tree calculations, initial trees were calculated by neighbour-joining. The ML and Bayesian trees were calculated under the HKY85+G model, with nucleotide frequencies and initial transition/transversion ratios estimated from the empirical frequencies. In the ML calculations, the transition/transversion ratio and the shape parameter (α parameter of the gamma distribution) were carried forward to the succeeding iteration. For the Bayesian analyses, the following additional parameters were used: number of generations = 100,000, burn-in for phylogeny calculation = 50,000 generations, temperature = 0.20. Since chain transition probabilities were lower than 10% for several starting seed integers, lower temperatures were also tried, but these had no apparent influence on the likelihood value on which the model converged after burn-in. The final Bayesian tree was the 50% majority rule consensus tree calculated over the last 50,000 generations of the Markov chain. Trees produced by each of the four methods were compared statistically based on their log likelihoods using the Kishino–Hasegawa test as implemented in PAUP*.

**Results**

**Sequence divergence within and between species**

Examination of both COI and CytB revealed no instances of complete haplotype sharing (i.e., shared COI and CytB sequences) between *G. rubens* and *G. texensis*. Within *G. rubens*, individuals 14 and 16, both from Pensacola/Milton, Florida, and individuals 3 and 9, both from Marianna, Florida, had identical haplotypes. *Gryllus rubens* individuals 14 and 16 and *G. texensis* individual 10 shared the same CytB sequence but differed in their COI sequences. COI sequences were more variable than CytB sequences in *G. texensis* (COI, mean distance 0.026 ± 0.022 vs. CytB, mean distance 0.004 ± 0.003; ANOVA, $F_{1,88} = 50.23, P < 0.0001$) but not in *G. rubens* (COI, mean distance 0.003 ± 0.003 vs. CytB, mean distance 0.002 ± 0.002; ANOVA, $F_{1,88} = 2.10, not significant$). These distances suggested greater levels of genetic variation in *G. texensis* than within *G. rubens*. We confirmed this by comparing average pairwise intraspecific sequence divergence: *G. texensis* sequences showed much higher levels of intraspecific variation than did *G. rubens* sequences (COI, $F_{1,88} = 56.73, P < 0.0001$; CytB, $F_{1,88} = 19.34, P < 0.001$).

Based on the 20 individuals that we analyzed, levels of differentiation between species were not significant when the analysis was based solely on CytB sequences, but interspecific divergence was significant for the analyses based on either COI sequences alone or both genes combined (AMOVA, Table 2).

**Phylogenetic analyses**

Five sets of trees were calculated based on parsimony ($n = 63$ trees), parsimony with an imposed constraint ($n = 20$), neighbour-joining ($n = 1$), maximum likelihood ($n = 1$), and Bayesian analysis ($n = 1$), for a total of 86 trees calculated by the various methods. All 86 trees were compared using the Kishino–Hasegawa and Shimodaira–Hasegawa tests, which compare differences in likelihood scores ($-ln L$) between the “best” tree (lowest score) and every other tree. Both tests indicated that the neighbour-joining and parsimony-with-constraint trees were significantly longer than the “best” tree, which was that produced by maximum likelihood ($P < 0.01$ in all cases). However, inspection of Table 3 indicates that the likelihood score of the Bayesian consensus tree was only marginally larger than that of the ML tree and exactly the same as that of the best parsimony trees. Moreover, the Bayesian consensus tree was five steps shorter than the ML tree and as short as the best parsimony trees. Therefore, the Bayesian consensus tree (Fig. 1) was probably the “best” tree overall, although the Bayesian, parsimony, and ML trees were all very similar in topology (Fig. 2). These consensus gene trees suggest that *G. texensis* and *G. rubens* have not reached reciprocal monophony, and that *G. rubens* may in fact be derived from *G. texensis*.

**Discussion**

We think our results are noteworthy in several ways. First, despite the lack of reciprocal monophony, there is fairly clear interspecific
divergence in mtDNA between these taxa. Polyphyly of gene trees is to be expected for recently diverged species pairs, with gene trees progressing from polyphyly to paraphyly to monophyly via lineage sorting (Avise 2000). This predictable lack of reciprocal monophyly in closely related species is likely to be problematic for recent attempts to use the COI gene sequence as a unique, species-identifying “DNA barcode” (Hebert et al. 2003). Although DNA barcoding shows a great deal of promise for many, or perhaps even most, species (Hebert et al. 2004; Monaghan et al. 2005; Hajibabaei et al. 2006), even its proponents concede that it may fail in the instance of very closely related species. For example, Hajibabaei et al. (2006) used the COI gene to correctly identify approximately 98% of 521 species of tropical Lepidoptera in the families Hesperiidae, Sphingidae, and Saturniidae, but noted that the failure of DNA barcodes to separate approximately 2% of species most likely involved very closely related taxa due to either recent speciation or hybridization.

The relatively low level of molecular divergence between species, especially compared with the levels of intraspecific variation in *G. texensis*, is our second noteworthy result. The overall 2.59% divergence in COI observed between species suggests separation well within the Pleistocene. Based on typical estimates of a molecular clock divergence rate of approximately 2% per million years for insect COI genes, our data suggest a speciation date approximately 1.3 million years before the present. However, it is clear from inspection of Figure 1 that the *G. rubens* sequences appear clustered within a subset of the *G. texensis* sequences. Thus, the average interspecific divergence (2.6%) may be inflated relative to the true timing of speciation. From Figure 1, it appears that *G. texensis* sequences Tc18, Tc1b, Tc10, Tc20, and Tc13 are as distant from *G. texensis* sequences Tc19, Tc11b, Tc4, Tc5, and Tc7 as they are from the *G. rubens* sequences (note that this separation within *G. texensis* does not coincide with the population origin of the cricket samples, either allopatric or sympatric). Because studies of recent molecular divergence necessarily span both population genetics and phylogenetics (Arbogast et al. 2002), concordance between gene trees and species’ histories decreases (Nichols 2001). Speciation between these taxa may therefore have been considerably more recent.

<table>
<thead>
<tr>
<th>Gene</th>
<th>COI (716 nt)</th>
<th>CytB (744 nt)</th>
<th>Both genes (1460 nt)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nucleotide diversity within <em>G. rubens</em> (mean ± SD)</td>
<td>0.0025 ± 0.0018</td>
<td>0.0043 ± 0.0028</td>
<td>0.0024 ± 0.0020</td>
</tr>
<tr>
<td>Nucleotide diversity within <em>G. texensis</em> (mean ± SD)</td>
<td>0.0256 ± 0.0141</td>
<td>0.0048 ± 0.0081</td>
<td>0.0048 ± 0.0034</td>
</tr>
<tr>
<td>Corrected average pairwise distance between <em>G. rubens</em> and <em>G. texensis</em></td>
<td>0.0107</td>
<td>0.0095</td>
<td>0.0098</td>
</tr>
<tr>
<td><em>F</em>&lt;sub&gt;ST&lt;/sub&gt;, probability</td>
<td>0.0500</td>
<td>0.062</td>
<td>0.062</td>
</tr>
</tbody>
</table>

Note: Probabilities are based on an AMOVA model comparing sequence variation between and within species.

**Table 2.** Measures of DNA sequence diversity within and between *G. rubens* and *G. texensis*.
than the overall 2.6% divergence implies. Perhaps
the relevant degree of molecular divergence is
between the \textit{G. rubens} sequences and the
closely related subset of \textit{G. texensis} sequences
(see also Avise 2000 and Hewitt 2001 for dis-
cussion of genetic divergence preceding spe-
cies’ divergence). Future work with these
species involving coalescent simulations will
address this issue.

Another interesting finding in the data is the
disparity in divergence estimates across genes.
It is widely recognized that different mito-
chondrial genes, and even different functional
domains within single mitochondrial genes,
have different levels of conservation owing to
functional constraint (Simon \textit{et al.} 1994; Lunt
\textit{et al.} 1996; Zhang and Hewitt 1996; Caterino
and Sperling 1999; Lin and Danforth 2004).
Our finding of a 10-fold difference in diver-
gence between the COI and CytB mitochondrial
genes may be an extreme example of such rate
heterogeneity among genes. The portion of the
COI gene that we sequenced includes the
highly variable UEA9/UEA10 region (terminol-
ogy follows Zhang and Hewitt (1996)), thus
potentially resulting in a higher than typical
level of variation in this gene. CytB sequence
variation in insects has been studied by
Simmons and Weller (2001), who found similar
rates of variation in CytB and COI in ctenuchid
moths. However, our analysis of previously
published sequences for \textit{Gryllus} crickets

\begin{table}
\centering
\caption{Summary statistics and comparison of 86 trees calculated by five different methods.}
\begin{tabular}{|c|c|c|c|c|}
\hline
Tree-building method & Neighbour-joining & Parsimony & Parsimony with constraint & Maximum likelihood & Bayesian \\
\hline
Number of trees & 1 & 63 & 20 & 1 & 1$^a$ \\
Tree length & 240 & 234 & 249 & 239 & 234 \\
Maximum likelihood score ($-\ln L$) & 3281.6 & 3230.3 to 3243.0 & 3270.4 to 3281.2 & 3230.0 & 3230.3 \\
\hline
\end{tabular}
\footnote{$^a$Consensus of 50 000 generations.}
\end{table}

\textbf{Fig. 1.} Bayesian consensus tree. Numbers at branch points indicate the posterior probability of clades.
Terminal tip taxa are abbreviated as follows: ovisopis1 and ovisopis 2, pennsyl, and veletis1 and veletis2 are sequences from GenBank for the outgroup taxa \textit{G. ovisopis}, \textit{G. pennsylvanicus}, and \textit{G. veletis}; Tc sequences are from \textit{G. texensis}; and Rc sequences are from \textit{G. rubens} (see also Table 1).
Huang et al. (2000) found rate heterogeneity among genes similar to that reported here. We calculated average pairwise divergence between the four samples of *G. rubens* and *G. texensis* (two per species) in the 16S rRNA and CytB genes and found a 10-fold higher divergence in CytB (2.75%) than in 16S (0.2%). Thus, significant rate heterogeneity among genes is not a unique feature of our results, although we are unable to account for the disparity between Huang et al.’s (2000) CytB divergence of 2.75% and our own CytB divergence, nearly an order of magnitude lower. This may reflect sampling issues, especially because the overall levels of divergence are uniformly low, such that even a few nucleotide changes can produce dramatic differences in average percent divergence, particularly with the inclusion of few samples.

Finally, it appears that *G. rubens* has much less genetic variation than does *G. texensis*. The finding of low levels of genetic variation in *G. rubens* is consistent with previous work (Harrison and Bogdanowicz 1995) and unpublished data (D. Gray). These data, combined with current geographic distribution, suggest a possible scenario for the divergence of these two species. *Gryllus rubens* is currently distributed throughout Florida and the southeastern United States westward across the gulf states to far eastern Texas. *Gryllus texensis* is distributed from west Texas eastward across the gulf states to the far western end of the “panhandle” of Florida. The mtDNA sequence data suggest (i) Pleistocene divergence, (ii) a population genetic bottleneck in *G. rubens*, and (iii) the possibility that *G. rubens* is derived from *G. texensis*. An allopatric or peripatric model of speciation with a formerly widespread *G. texensis* or *texensis*-like ancestor being divided into separate Floridian and western gulf refugia during glacial advance would account for all of these observations. Rapid evolution of mate recognition systems, primarily male pulse rate and female recognition of pulse rate (Gray and Cade 2000; Izzo and Gray 2004; Gray 2005), in either or both descendant populations appears sufficient to maintain species distinctness following range expansion and secondary contact. This scenario is necessarily speculative at the moment and is currently the subject of larger-scale phylogeographic analysis.

Incorporation of molecular divergence data with comparisons of behavioral and morphological divergence is potentially a powerful way to address overall levels of lineage divergence among taxa. Although the current study is limited to mtDNA sequences, the results do show
significant molecular divergence between a sister species pair of crickets recognized as distinct by song. Although further molecular study is clearly needed, the present results are consistent with allopatric or peripatric divergence in Pleistocene glacial refugia.

Acknowledgments

Natural Sciences and Engineering Research Council of Canada grants to W.H. Cade (A6174) and M. Richards (222883-00), as well as an Ontario Premier’s Research Excellence Fund grant to M.R., provided funding for this work.

References


