Mitochondrial DNA Sequence Variation in the Spruce Budworm Species Complex (Choristoneura: Lepidoptera)

Felix A. H. Sperling and Donal A. Hickey
Department of Biology, University of Ottawa

A combination of polymerase-chain-reaction amplification and automated DNA sequencing was used to survey variation in a species complex of pest insects, the spruce budworms (Choristoneura fumiferana species group), and an outgroup species, C. rosaceana. We sequenced an mtDNA region of 1,573 bp that extends from the middle of cytochrome oxidase subunit I (COI) through tRNA leucine (UUR) to the end of cytochrome oxidase subunit II. In addition, we examined levels of intraspecific variation within a 470-bp region of the CO1 gene. Choristoneura fumiferana clearly represented the oldest lineage within its species group, with 2.7%-2.9% sequence divergence from the other species. In contrast, the four remaining species (C. pinus, C. biennis, C. occidentalis, and C. orae) had closely related or identical mtDNA, with <1% divergence among most of their haplotypes. Despite its older lineage and widespread geographic distribution, C. fumiferana showed significantly lower intraspecific genetic diversity than did C. occidentalis. Choristoneura orae shared haplotypes with C. occidentalis and C. biennis, and species-level separation of these three species was not supported. Two divergent, uncommon haplotypes were also found in C. occidentalis and C. biennis. The divergent haplotype in C. biennis had an unusually high number of inferred amino acid replacements, suggesting selective differences between mitochondrial DNA haplotypes. Transition:transversion ratios in Choristoneura paralleled those found in Drosophila; transition:transversion ratios were highest in closely related sequences but decreased with increasing sequence divergence. Nucleotide composition showed an A+T bias that was near the high end of the range known for insects. This work illustrates the potential utility of direct DNA sequencing in assessing population structures, species limits, and phylogenetic relationships among organisms that have not previously been subjected to DNA analysis.

Introduction

By virtue of its simple structure, maternal inheritance, and relatively rapid evolutionary rates, mitochondrial DNA (mtDNA) has become a widely used marker for understanding population structures and phylogenetic relationships of animal species (Harrison 1989; Avise 1991). At the species level and below, surveys of sequence variation have usually been accomplished using restriction-fragment-length polymorphisms. Direct sequence comparisons of mtDNA have been limited to the fast-evolving D-loop (e.g., see Vigilant et al. 1991; Brown et al. 1993), to studies of Drosophila (e.g., DeSalle et al. 1987; Tamura 1992b; Beckenbach et al. 1993), or to relatively distantly related animal species (e.g., see Liu and Beckenbach 1992; but also see Beckenbach et al. 1989; Willis et al. 1992; Vogler et al 1993). As a result, a relatively narrow base of information is available on rates and patterns of DNA sequence evolution in mtDNA coding sequences at the level of sibling species and below. Yet this information is important for an accurate understanding of sequence divergence and homoplasy at higher taxonomic levels.

In this study, we used automated sequencing to survey DNA variation in a species complex of pest insects, the spruce budworms (Choristoneura fumiferana species group). We sequenced an mtDNA region of 1,573 bp that extends from the middle of cytochrome oxidase subunit I (COI) through tRNA leucine (UUR) to the end of cytochrome oxidase subunit II (COII). Although protein sequences of COI and COII are relatively conserved within insects (Simon 1991; Liu and Beckenbach 1992; authors' unpublished data), coding sequences contain numerous sites where synonymous substitutions can occur. We reasoned that approximately one-third of sites would potentially provide fine-grained information about evolution at the species level and below. At the same time, alignment problems should be less than that for noncoding regions such as the D-loop of...
mtDNA, allowing our studies to be extended to phylogenetic comparisons among more distantly related species. The aims of our study were (1) to assess population structures, species limits, and phylogenetic relationships among currently recognized species of the C. fumiferana group in Canada and Alaska and (2) to elucidate the earliest stages of mtDNA sequence divergence in a previously uncharacterized insect order, the Lepidoptera.

**Material and Methods**

**The Choristoneura fumiferana Group**

The eight species recognized in the group constitute a monophyletic complex of conifer-feeding moths endemic to North America (Harvey 1985; Dang 1992b). Two species (C. pinus and C. fumiferana) are broadly sympatric across the boreal zone of Canada and the northern United States. Three species are common in western Canada. One of these three species, C. occidentalis, ranges from southern British Columbia to Oregon and New Mexico. The other two species have narrower distributions: C. biennis is restricted to central British Columbia, while C. orae ranges from northwestern British Columbia to Alaska. Three additional species occur in the western United States, but only one of these, C. lambertiana, is occasionally found in southern British Columbia. The present study focuses on populations of the five species of the C. fumiferana group that occur commonly in Canada and Alaska (Harvey 1985).

Choristoneura rosaceana, which we use here for outgroup comparisons, is one of 16 species of Choristoneura recognized in addition to the species in the C. fumiferana group (Dang 1992a, 1992b). Choristoneura rosaceana primarily feeds on deciduous forest trees and shrubs but is also known as an occasional pest on nursery pines (Otvos 1991).

Species of the C. fumiferana group are primarily distinguished by ecological characteristics such as larval host plant and voltinism. Only differences in the frequency of characters, rather than complete character substitutions, have been found in studies of larval and adult coloration, morphology, and allozymes of species of the C. fumiferana group (Castrovillo 1982; Harvey 1985; Dang 1992b). The reliability of reports of pheromone differences between species has not been determined, since only a few populations per species have been characterized (Silk and Kuenen 1988). Several unnamed populations with unusual pheromone characteristics are known (Gray and Slessor 1989; T. Gray, personal communication). Even the ecological differences reported to distinguish C. occidentalis, C. orae, and C. biennis in British Columbia may be clinal (Harvey 1985). Clearly, there is a need for independent tests of species limits in the C. fumiferana group, as well as for diagnostic markers to distinguish these species.

**Samples, Sequencing, and Analysis**

Sample localities are listed in table 1; these samples were collected by the people and agencies listed in the Acknowledgments. In most cases, DNA was extracted from adult moths that were reared from wild-collected larvae and then kept frozen at -70°C until use. The abdomen and wings of each specimen used in the survey have been deposited in the Canadian National Collection, Ottawa. Specimen identification was based on locality, adult wing pattern, larval host plant, or pheromones.

Collections were made across the ranges of C. fumiferana (eight localities) and C. biennis (three localities). Choristoneura orae is represented by one locality, and C. pinus is represented by collections from two localities separated by 24 km. Four collections of C. occidentalis were obtained within 200 km of each other in southern British Columbia. One population, from McKinney Creek near Bridesville, was unusual in that a proportion of the larvae were feeding on a host, Lodgepole pine (Pinus contorta), that is atypical for C. occidentalis. Population samples from Alaska, unlike other collections, were attracted, as adults, to pheromone traps and were shipped in 95% ethanol. These collections were made on three different dates, with each of two different pheromone attractants: 95: 5 E-Z-11-14 Al for C. fumiferana and 82: 9 9 E11-14 Ac: Z11-14 Ac: E11-14 OH for C. orae (Silk and Kuenen 1988).

Genomic DNA was extracted from individual Choristoneura thoraces by using the protocol of Harrison et al. (1987). Genomic DNA was used as template for amplification of mitochondrial fragments by the polymerase chain reaction (PCR) (Saiki et al. 1988). Initially, we used general insect mtDNA primers developed by Liu and Beckenbach (1992) and the lab of R. Harrison (Cornell University) (primers A and B in fig. 1). Primers for the intervening region of mtDNA, including primer C of figure 1, were designed directly from Choristoneura sequence and were positioned approximately every 300 bp. Double-stranded PCR product was cleaned with Centricon 100 microconcentrators (Amicon) and was sequenced directly by using the Taq DyeDeoxy Terminator Cycle Sequencing system (Applied Biosystems). In all cases, sequence was confirmed from both sense and antisense strands.

The mtDNA of 10 specimens was sequenced over 1,573 bp (fig. 1), which correspond to bp 2188-3771 in the sequence of D. yakuba (Clary and Wolstenholme 1985). The 10 specimens were chosen to represent different species, different parts of the geographic range of species, or unusual variants detected on the basis of shorter sequences. In order to assess variation within populations and species, an additional 37 specimens,
representing 2–4 specimens per population, were sequenced over a 470-bp region of COI.

Haplotype and nucleotide diversity were calculated with equations (8.4) and (10.5) of Nei (1987). Sequence divergence is corrected with the methods of Jukes and Cantor (1969) and Tamura (1992a). In order to compare mtDNA with a published phenogram based on alloenzymes, UPGMA (unweighted pair group method using arithmetic averages) phenograms (Sneath and Sokal 1973) were constructed using the relatively conservative Jukes-Cantor correction.

Phylogenetic analysis was performed using PAUP (Swofford 1993). Variable nucleotide positions were treated as unordered characters with one state for each nucleotide. *Choristoneura rosaceana* sequences were used to root trees.

**Results**

**Relationships Within and Among Species**

A total of 75 nucleotide sites varied over a 1,573-bp region sequenced in nine specimens of the *Choristoneura fumiferana* group (figs. 2 and 3). Seven different haplotypes were distinguished, with the bl haplotype occurring in one specimen of *C. biennis* and in one of *C. orae* and with the bβ haplotype occurring in two specimens of *C. biennis* from localities separated by 400 km. Uncorrected mtDNA sequence divergences were 0.1%–2.9% within the *C. fumiferana* group. One *C. rosaceana* was also sequenced over the 1,573-bp region and showed 5.2%–5.6% sequence divergence from members of the *C. fumiferana* group.

A single most-parsimonious tree was obtained from an exhaustive search of the eight haplotypes defined by the 1,573-bp sequences (fig. 4), by using PAUP (Swofford 1993). Its reliability was assessed with 500 replications of the bootstrap algorithm in PAUP, using a branch and bound search. A minimum of 85% bootstrap support was obtained for each node. Thus mtDNA of *C. fumiferana* clearly represents the oldest lineage within its species group, with 2.7%–2.9% divergence from the other species. In contrast, the four remaining species (*C. pinus*, *C. biennis*, *C. occidentalis*, and *C. orae*) had closely related or identical mtDNA, with <1% divergence among

**Table 1**

<table>
<thead>
<tr>
<th>Identification</th>
<th>Haplotype (no. of individuals)</th>
<th>Locality; Larval Host Plant or Adult Pheromone</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. <em>C. fumiferana</em></td>
<td>f1(2)</td>
<td>Blanch River, Newfoundland; balsam fir</td>
</tr>
<tr>
<td>2. <em>C. fumiferana</em></td>
<td>f1(1), f3(1)</td>
<td>Holmes Lake, New Brunswick; balsam fir</td>
</tr>
<tr>
<td>3. <em>C. fumiferana</em></td>
<td>f1(1), f2(1)*</td>
<td>Ignace district, Ontario; white spruce</td>
</tr>
<tr>
<td>4. <em>C. fumiferana</em></td>
<td>f1(2)</td>
<td>Carberry, Manitoba; white spruce</td>
</tr>
<tr>
<td>5. <em>C. fumiferana</em></td>
<td>f2(2), f4(1), o1(1)</td>
<td>Cypress Hills, Alberta; white spruce</td>
</tr>
<tr>
<td>6. <em>C. fumiferana</em></td>
<td>f2(2), b1(1)</td>
<td>Red Lodge, Alberta; white spruce</td>
</tr>
<tr>
<td>7. <em>C. fumiferana</em></td>
<td>f1(2)*</td>
<td>Hawk Hills, Alberta; white spruce</td>
</tr>
<tr>
<td>8. <em>C. fumiferana</em></td>
<td>f3(3)</td>
<td>Fairbanks, Alaska; 95: 5 E: Z-11-14 AL</td>
</tr>
<tr>
<td>9. <em>C. orae</em></td>
<td>b1(1)*, o1(2)</td>
<td>Fairbanks, Alaska; 82: 9 E: Z AC: OH</td>
</tr>
<tr>
<td>10. <em>C. biennis</em></td>
<td>b1(2)<em>, bR(1)</em></td>
<td>Morrisey Creek, British Columbia; alpine fir</td>
</tr>
<tr>
<td>11. <em>C. biennis</em></td>
<td>b1(1), b1(1)*</td>
<td>McBride, British Columbia; alpine fir</td>
</tr>
<tr>
<td>12. <em>C. biennis</em></td>
<td>b1(1), b2(1), b3(1)</td>
<td>Numa Falls, British Columbia; Engleman spruce/alpine fir</td>
</tr>
<tr>
<td>13. <em>C. occidentalis</em></td>
<td>o1(1)*, o3(1)</td>
<td>Monte Creek, British Columbia; Douglas fir</td>
</tr>
<tr>
<td>14. <em>C. occidentalis</em></td>
<td>o2(1), o4(1)</td>
<td>Winfield, British Columbia; Douglas fir</td>
</tr>
<tr>
<td>15. <em>C. occidentalis</em></td>
<td>o5(1), o6(1), o8(1)*</td>
<td>Bridesville, British Columbia; Douglas fir/lodgepole pine</td>
</tr>
<tr>
<td>16. <em>C. occidentalis</em></td>
<td>o1(1), o2(1), o7(1)</td>
<td>Greenwood, British Columbia; Douglas fir</td>
</tr>
<tr>
<td>17. <em>C. pinus</em></td>
<td>p1(2)*</td>
<td>northwest of Parry Sound, Ontario; jack pine</td>
</tr>
<tr>
<td>18. <em>C. pinus</em></td>
<td>p1(1), p2(1)</td>
<td>northwest of Parry sound, Ontario; jack pine</td>
</tr>
<tr>
<td>19. <em>C. rosaceana</em></td>
<td>r2(1)</td>
<td>northwest of Parry Sound, Ontario; jack pine</td>
</tr>
<tr>
<td>20. <em>C. rosaceana</em></td>
<td>r1(1)*</td>
<td>Ste. Agathe, Quebec; at light</td>
</tr>
</tbody>
</table>

NOTE.—All specimens were sequenced over a 470-bp region of COI.

* Based on adult wing coloration and pheromone attractant or larval host plant.
* As in fig. 3.
* One specimen was sequenced over an additional 1,103 bp of COI, tRNA leu, and COII.
most of their haplotypes. However, putative populations of both *C. biennis* and *C. occidentalis* included uncommon mtDNA genotypes (ββ and αβ) that showed basal relationships to the o1+b1+p1 lineage. The ββ and αβ sequences were diverged from each other by 1.4%, and by 2.0% from the mtDNAs of specimens that may be conspecific. The wings of moths with β mtDNA genotypes showed no evident differences from those with more typical haplotypes from the same localities.

An additional 37 specimens were sequenced over a 470-bp region of COI. These 37 specimens revealed 12 more mtDNA haplotypes but only two new variable nucleotide sites; most of the 12 new haplotypes were distinguished by combinations of nucleotides at sites that also varied among more distantly related haplotypes.

A total of 20 different mtDNA haplotypes were evident among the 47 specimens sequenced over a 470-bp region of COI (table 1 and fig. 3). Six equally parsimonious trees were obtained from a branch-and-bound search of the 20 haplotypes by using PAUP. Two of the trees were consistent with the analysis of longer sequences, and a consensus of these two has been used to illustrate relationships among the short sequences (fig. 5). The four other trees differed both in rearrangements among the f haplotypes and in placement of the bβ haplotype as the sister group of the o+b+p lineage rather than of the oβ lineage.

*Choristoneura fumiferana* haplotypes formed a monophyletic group, as did those of *C. pinus*. In contrast, there were no haplotypes unique to *C. orae*, which shared haplotypes with *C. occidentalis* and *C. biennis*. The genetic distinctness of *C. orae* and *C. fumiferana* in Alaska was confirmed by the fact that the two different types of pheromone traps each attracted moths with different mtDNA. Two haplotypes of *C. biennis* were part of the lineage leading to most haplotypes in *C. occidentalis*, but the *C. biennis* haplotypes had more basal relationships.

Two collections of putative *C. fumiferana* were made on the western edge of the range of this species in Alberta, and these showed a mixture of either *C. fumiferana* and *C. occidentalis* haplotypes (Cypress Hills) or *C. fumiferana* and *C. biennis* haplotypes (Red Lodge). The wings of moths with non-f haplotypes showed no apparent differences from those with f haplotypes (but single specimens of all three species are normally difficult to distinguish on the basis of wing coloration).

Even when β haplotypes were excluded, there was greater genetic diversity in *C. occidentalis* than in *C. fumiferana* (table 2). No two *C. occidentalis* specimens from the same locality had the same mtDNA, and the localities were separated by only 200 km. In contrast, 15 of the 18 mtDNAs in the *C. fumiferana* lineage were the f1 haplotype, including specimens ranging from Alaska to Newfoundland. The difference in diversity was statistically significant when measured as haplotype diversity but not when measured as nucleotide diversity. The other species had intermediate amounts of genetic diversity.
Pattern and Rate of Nucleotide Substitution

Sequence variation was scattered across most of the region of 1,573 bp (fig. 6). The main exception was an unvaried region of 145 bp that included the 3' end of the COI gene, the entire tRNA leucine gene, and the 5' beginning of the COII gene. Most variation (81%) was found to vary. A total of 19 first-base positions, while two nucleotide differences were found at second-base positions were both transitions. Nineteen transitions and two transversions were found at first-base positions, while two nucleotide differences found at second-base positions were both transitions.

The COI and COII genes had somewhat different rates of nucleotide substitution. Phylogenetic analysis of the 1.6-kb sequences within the C. fumiferana group showed 50 substitutions in COI and 30 in COII, indicated by an asterisk were only characterized over bp 2-471. Vertical numbers at tops of columns refer to nucleotide position.

Pattern and Rate of Nucleotide Substitution

Sequence variation was scattered across most of the region of 1,573 bp (fig. 6). The main exception was an unvaried region of 145 bp that included the 3' end of the COI gene, the entire tRNA leucine gene, and the 5' beginning of the COII gene. Most variation (81%) was restricted to the third-base position of triplet codons. Only two second-base positions (bp 617 and 752) were found to vary. A total of 21 first-base positions showed variation, and 13 of these were part of a leucine codon.

There was a considerable decline in transition-transversion ratios with increasing sequence divergence. Character optimization of nucleotide substitutions gave a total of 26 transitions and no transversions, among haplotypes that differed by <1%. This number is based on changes that have occurred among both the 1,573- and 470-bp sequences. Haplotypes with divergences of 1.4%-2.9% had transition:transversion ratios of 33:1-38:4, with a total of 71-75 transitions and 5 transversions (6%-7% transversions) among them. Comparisons with C. rosaceana (5.2%-5.6% sequence divergence) gave ratios ranging from 54:25 to 61:24 (=28%-32% transversions). Nineteen transitions and two transversions were found at first-base positions, while two nucleotide differences found at second-base positions were both transitions.

The COI and COII genes had somewhat different rates of nucleotide substitution. Phylogenetic analysis of the 1.6-kb sequences within the C. fumiferana group showed 50 substitutions in COI and 30 in COII, indicated by an asterisk were only characterized over bp 2-471. Vertical numbers at tops of columns refer to nucleotide position.

Pattern and Rate of Nucleotide Substitution

Sequence variation was scattered across most of the region of 1,573 bp (fig. 6). The main exception was an unvaried region of 145 bp that included the 3' end of the COI gene, the entire tRNA leucine gene, and the 5' beginning of the COII gene. Most variation (81%) was restricted to the third-base position of triplet codons. Only two second-base positions (bp 617 and 752) were found to vary. A total of 21 first-base positions showed variation, and 13 of these were part of a leucine codon.

There was a considerable decline in transition-transversion ratios with increasing sequence divergence. Character optimization of nucleotide substitutions gave a total of 26 transitions and no transversions, among haplotypes that differed by <1%. This number is based on changes that have occurred among both the 1,573- and 470-bp sequences. Haplotypes with divergences of 1.4%-2.9% had transition:transversion ratios of 33:1-38:4, with a total of 71-75 transitions and 5 transversions (6%-7% transversions) among them. Comparisons with C. rosaceana (5.2%-5.6% sequence divergence) gave ratios ranging from 54:25 to 61:24 (=28%-32% transversions). Nineteen transitions and two transversions were found at first-base positions, while two nucleotide differences found at second-base positions were both transitions.

The COI and COII genes had somewhat different rates of nucleotide substitution. Phylogenetic analysis of the 1.6-kb sequences within the C. fumiferana group showed 50 substitutions in COI and 30 in COII, indicated by an asterisk were only characterized over bp 2-471. Vertical numbers at tops of columns refer to nucleotide position.
cating a 39% higher rate of substitution per nucleotide site in COI than in COII. Substitutions at third-base positions constituted 88% of the total in COI and 73% in COII.

There was a total of 19 transversions in COI and 8 transversions in COII. The frequency of codons with fourfold-degenerate third-base positions (excluding leucine) was also higher in COI (38.2%–39.0%) than in COII (28.1%). This should give COI a 36% greater opportunity for synonymous third-base transversions. In fact, COI had 17 transversions at fourfold-redundant positions, while COII had 7. Even when the smaller number of COII codons was taken into account, COI had more than twice as many transversions as did COII.

Nucleotide frequencies revealed a strong A+T bias. The COI gene showed the same degree of A+T bias (61.2%–63.0%) at first- and second-base positions, while COII showed a slightly higher A+T bias (65.5%–70.5%) at first- and second-base positions. Third-base positions had a much stronger A+T bias (91.6%–96.7%). Choristoneura rosaceana had the strongest A+T bias at most positions, especially when only fourfold-degenerate positions were considered (98.3% and 99.1%).

Nucleotide divergences were considerably higher when raw percent divergences were corrected with a formula that explicitly takes into account strong biases in A+T and transition:transversion frequencies (Tamura 1992a). For Choristoneura mtDNA lineages with an overall divergence >5%, the observed sequence differences at third-base positions accounted for as few as half of the substitutions that may have actually occurred.

Inferred Amino Acid Sequence

The 10 long mtDNA sequences included 273 amino acid codons for COI and 227 for COII. The proportion of leucine codons was almost the same in both genes (11.4%–11.7% in COI and 11.9% in COII). Thus the fact that in COI there were almost half as many nucleotide differences at first-base positions as there were in COII appears to be unrelated to the number of leucine codons. The frequency of leucine codons that began with T rather than with C was 72.9%–81.0%, reflecting the strong A+T bias of Choristoneura mtDNA. As a result, most leucine third-base positions were constrained to A or G, and in fact the frequency of A at this position was 81.4%–84.7%, again reflecting a strong A+T bias.

There was variation in amino acids at 10 locations across the COI and COII genes (fig. 7). Choristoneura rosaceana and C. fumiferana were characterized by four differences and one difference, respectively. The bβ haplotype, which was found in two separate specimens, had at least five amino acid replacements, compared with the haplotypes of other members of the C. fumiferana group. The bβ haplotype did not have unusual nucleo-

### Table 2

<table>
<thead>
<tr>
<th>Sample Size</th>
<th>Haplotype Diversity ± 2 SE</th>
<th>Nucleotide Diversity ± 2 SE</th>
</tr>
</thead>
<tbody>
<tr>
<td>fumiferana (f1–f4)</td>
<td>18 0.305 ± 0.354</td>
<td>0.0007 ± 0.0031</td>
</tr>
<tr>
<td>occidentalis (o1–o7)</td>
<td>10 0.863 ± 0.064</td>
<td>0.0031 ± 0.0031</td>
</tr>
<tr>
<td>orae (o1 and b1)</td>
<td>3 0.533 ± 0.534</td>
<td>0.0014 ± 0.0049</td>
</tr>
<tr>
<td>biennis (b1–b3)</td>
<td>7 0.484 ± 0.360</td>
<td>0.0018 ± 0.0043</td>
</tr>
<tr>
<td>pinus (p1 and p2)</td>
<td>4 0.429 ± 0.571</td>
<td>0.0011 ± 0.0028</td>
</tr>
<tr>
<td>occidentalis β (oβ)</td>
<td>1 0.833 ± 0.053</td>
<td>0.0081 ± 0.0076</td>
</tr>
<tr>
<td>biennis β (ββ)</td>
<td>2 0.654 ± 0.170</td>
<td>0.0096 ± 0.0116</td>
</tr>
<tr>
<td>All biennis</td>
<td>9 0.667 ± 0.577</td>
<td>0.0021 ± 0.0060</td>
</tr>
<tr>
<td>rosaceana (r1 and r2)</td>
<td>2 0.833 ± 0.053</td>
<td>0.0081 ± 0.0076</td>
</tr>
</tbody>
</table>

Note.—Data are based on 470-bp sequences in COI and are calculated with equations (8.4), (8.12), (10.5), and (10.9) of Nei (1987).
tide frequencies or transition:transversion biases, in comparison with other haplotypes. Both the COI and COII genes in *Choristoneura* were terminated by incomplete stop codons (a single T).

**Discussion**

**Species Relationships and Population Structure**

mtDNA sequences confirm the species-level separation of *Choristoneura fumiferana* from *C. pinus* and the western species. The only taxonomic ambiguities involving *C. fumiferana* occurred in central and southern Alberta. This region lies at the western edge of the range of *C. fumiferana* and at the eastern edge of the range of *C. occidentalis* and *C. biennis*. Mixed collections were obtained only from a narrow region of potential overlap between species; they are thus unlikely to represent retained ancestral polymorphism within *C. fumiferana*. Species of the *C. fumiferana* group are normally uncommon in this region (Harvey 1985), and these collections may represent a mixture of species. In this case, *C. occidentalis* and *C. biennis* would be feeding on a host, white spruce, that is normally only a minor component of the diet of these species (Volney 1989). Alternatively, specimens with o and b haplotypes may represent mtDNA introgression from *C. occidentalis* and *C. biennis* into *C. fumiferana*. Similar ambiguities have been found in the mtDNA of other lepidoptera in this region (e.g., see Sperling 1993; Sperling and Harrison, in press).

Species-level separation of *C. orae*, *C. occidentalis*, and *C. biennis* is not supported by mtDNA. The mtDNAs of *C. occidentalis* and *C. biennis* constitute very similar groups. The mtDNA of *C. orae* forms a transition between them, and this species is also intermediate for ecological characters such as voltinism (Harvey 1985). We suggest that *C. orae*, *C. biennis*, and *C. occidentalis* represent very recently diverged geographic or ecological races. However, since the mtDNAs of *C. occidentalis* and *C. biennis* form separate groups, it will be important to investigate the western *Choristoneura* species along zones of potential genetic interaction.

The higher genetic diversity of *C. occidentalis* compared with *C. fumiferana* may reflect older colonization of the range of *C. occidentalis*, lower rates of gene flow, fewer bottleneck effects, or more even rates of survival among families. A potential practical consequence is suggested by the fact that *C. occidentalis* shows greater genetic diversity over a much smaller geographic range than does *C. fumiferana*. If the extent of variation in mtDNA is representative of the rest of the genome, then any single population-control strategy is less likely to be effective across different parts of the range of *C. occidentalis* than across the range of *C. fumiferana*.

The presence of the bβ haplotype in *C. biennis* and of the oβ haplotype in *C. occidentalis* presents several interesting problems. Specimens with these haplotypes may represent individuals from a different species, *C. lambertiana*. The single specimen with the oβ genotype came from a population in which some larvae were feeding on lodgepole pine, a common host plant of *C. lambertiana*. In contrast, specimens with the bβ haplotype were collected as larvae on alpine fir, which is not a known host of *C. lambertiana* (Volney 1989). None of the specimens with oβ or bβ haplotypes had the diffuse gray or ochraceous wing pattern that generally distinguishes *C. lambertiana* (Harvey 1985).

Other explanations are that the β haplotypes represent (a) retained ancestral polymorphisms within *C. biennis* and *C. occidentalis* (Neigel and Avise 1986), (b) remnant haplotypes that have survived introgressive inundation of mtDNA from invading populations, or (c) rare introgression from related species. The sequence divergences of the β haplotypes from the b and o haplotypes are 2%, which would represent unusually large intrapopulation divergences for insect mtDNA (Simon et al. 1993; Vogler et al. 1993; Sperling and Harrison, in press). Specimens with the β haplotypes may therefore belong to one or more unnamed species, such as Gray and Slessor's (1989) "probable new species" of *Choristoneura* from Scots pine. However, Gray and Slessor (1989) have pointed out that their "new species" occurs near port facilities and may represent an introduced species, whereas we found the β mtDNA haplotypes in populations that were in the interior and on the eastern side of British Columbia.

We are currently investigating nuclear gene markers for species of the *C. fumiferana* group, in an effort to further resolve species limits. We are encouraged, however, by the agreement between our results and studies of allozymes in *Choristoneura* species (fig. 8). The only substantive disagreement between them is that allozymes indicate a more basal relationship for *C. biennis* than does mtDNA.
Patterns and Rates of mtDNA Sequence Divergence

The mtDNA of *Choristoneura* is characterized by an A+T content intermediate between that of *Drosophila* and *Apis* (Crozier et al. 1989). The A+T content in *Choristoneura* mtDNA was 72%–74% for the downstream half of COI, compared with 68% in *Drosophila* and 76% in *Apis*, across the entire COI gene. A+T content for COII in *Choristoneura* was 75%–77%, which includes the value of 77% reported for a pyralid moth (Liu and Beckenbach 1992). Thus A+T content in *Choristoneura* falls at the high range of values reported for insect mtDNA.

The fact that no transversions were found between haplotypes diverged by <1% suggests that initial bias toward transitions is extremely high. At <3% divergence, 93% of substitutions were still transitions, while >5% divergence there were 68%–72% transitions. The decline in transition:transversion ratios parallels the situation described for *Drosophila* (e.g., see Tamura 1992b; Beckenbach et al. 1993). The likely causes for this phenomenon have been described elsewhere (DeSalle et al. 1987, and references therein).

In combination, the strong A+T content and transition:transversion biases suggest that as much as half of all third-position mutations are obscured by the time that total uncorrected sequence divergences between lineages are >5%. Yet, Crozier et al. (1989), Liu and Beckenbach (1992), and we (authors' unpublished data) found strongly conserved amino acid sequences for insect COI and COII. The conservative protein sequences of COI and COII are thus not entirely due to strong selective constraints at the level of gene products. Even first- and second-base positions showed A+T bias, indicating a compromise between gene function and constraint acting to maintain only codons compatible with forces operating at the DNA level. The strong bias toward T in the first-base position of leucine codons, combined with a strong bias toward A at the third-base position, illustrates the heterogeneous results that may be produced by such compromises.

The differences in rates and patterns of nucleotide substitutions between COI and COII and between *C. fumiferana* and *C. rosaceana* suggest considerable heterogeneity in the end result of substitution processes. Reeves (1992) has noted that such heterogeneity contributes to a poor fit between real data and any currently available model of nucleotide substitution. Fortunately, the apparent similarity between *Choristoneura* and *Drosophila* in initial DNA-level biases suggests that the relevant processes can be disentangled and generalized. Distinguishing these processes will be important to allowing mtDNA coding sequence data to be used more effectively for constructing higher level phylogenies.

Patterns of Adaptive Evolution

The *C. fumiferana* species group has been divided into two or more subgroups or series characterized by their major host groups. Powell (1980) and Harvey (1985) separated *C. pinus* and *C. lambertiana*, which feed primarily on pines (Pinoideae), from *C. fumiferana*. *C. occidentalis*, *C. biennis*, and *C. orae*, which feed primarily on spruces and firs (Abeitiodeae). When primary host groups are considered in light of the mtDNA phylogeny for the *C. fumiferana* group (fig. 9), it is evident that pine feeders and spruce/fir feeders are closely related. There is some uncertainty about the identity and feeding habits of the single specimen with the B β haplotype, provisionally identified as *C. occidentalis*. Nonetheless, pine feeding has probably been recently derived at least once from spruce/fir-feeding habits.
Changes in inferred amino acid composition of COI and COII are also of considerable interest because they may indicate differences in selective constraints acting on different mtDNA lineages (fig. 9). Three of the replacements that have occurred in Choristoneura (one in C. fumiferana mtDNA and two in the bβ lineage) involve nonpolar residues and are unlikely to have major structural consequences. Two of the differences that characterize the mtDNA of C. rosacea (fig. 7; K at 1227 and D at 1278) involve replacements with complementary charges and are located at sites that are only 16 amino acids apart on a hydrophilic loop of COII (Saraste 1990). It is plausible that one of these charged substitutions compensates for a mutation causing the earlier substitution.

The bβ haplotype presents a particularly intriguing case. The ratio of synonymous to nonsynonymous substitutions in the C. biennis β branch is significantly different from that in all other branches with six or more nucleotide substitutions (P<.05, test of independence using G statistic; Sokal and Rohlf 1981). The fact that three of five replacements that characterize the bβ haplotype involve different classes of amino acids suggests the possibility of a selective basis for these differences whether the bβ haplotype represents a cryptic species or polymorphism within C. biennis mtDNA.

Sequence Availability

The sequences of the f1, o1, p1, r1, bβ, and bβ haplotypes have been deposited in GenBank under accession numbers L19094–L1909.

Acknowledgments

We are grateful to T. Gray, J.-F. Landry, D. Langor, J. Volney, R. Werner, and the Canadian Forest Insect and Disease Survey, for providing specimens for this study. D. Langor generously took time to coordinate collections in 1991. Several mtDNA primers were donated by the lab of R. Harrison. We thank C. Simon and two anonymous reviewers for comments on the manuscript. This work was supported by a grant from the Canadian Networks of Centres of Excellence program to Insect Biotech Canada.

LITERATURE CITED


HARVEY, G. T. 1985. The taxonomy of the coniferophagous Choristoneura (Lepidoptera Tortricidae): a review. Pp. 16-