Sequence Evolution in Mitochondrial Ribosomal and ND-1 Genes in Lepidoptera: Implications for Phylogenetic Analyses

Dorothy Prowell Pashley and Li Dao Ke
Department of Entomology, Louisiana State University

A 2,256-bp sequence of the mitochondrial genome of a lepidopteran (Spodoptera frugiperda) contains tRNAs for valine and leucine, the 16S rRNA, and three-quarters of the ND-1 presumptive protein-coding gene. A 64-bp stretch of unknown function was located between the rRNA and leucine tRNA. Sequence divergence in the 16S rRNA obtained from alignment with published insect sequences is consistent with phylogenetic hypotheses, in that Diptera and Lepidoptera are more closely related to each other (24% sequence divergence) than either is to Hymenoptera (31%). Within the ND-1 gene, sequences for four additional Lepidoptera were generated for a 314-bp region and contrasted with published sequences for the locust and Drosophila. Sequence divergence in this region was consistent with accepted phylogenetic relationships, but results of parsimony analyses were not. Cladograms consistently recovered accepted higher level relationships (monophyly of Lepidoptera), despite high homoplasy, but were unable to resolve superfamly and family relationships within Lepidoptera, regardless of the outgroup or character subset analyzed. Character analysis indicated that homoplasy was decreased at higher levels when first- and second-codon sites were used exclusively. At the lowest level (families), resolution was enhanced by inclusion of third-codon sites. Inability of molecular data to recover a well-established phylogeny may be rectified by additional characters or taxa, but it is clear that homoplasy is sufficiently high to caution against the acceptance of relationships generated with this molecular region that are not extremely robust.

Introduction

The application of molecular methods to problems in insect systematics has lagged behind that in vertebrates. Only the genus Drosophila has been studied extensively (Solignac et al. 1986; DeSalle et al. 1987; Satta et al. 1987; Monnerot et al. 1990), although several other studies of insect sequences are in progress (Simon 1988). Most studies have focused on the mitochondrial genome, because it was initially the easiest to isolate and characterize, and, because of conservation in gene content, homology among genes could be established readily. The complete sequence of the mitochondrial genome has been determined for the fruit fly, D. yakuba (Clary and Wolstenholme 1985a). Partial sequences are known for the locust, Locusta migratoria (McCracken et al. 1987; Uhlenbusch et al. 1987), a mosquito, Aedes albopictus (HsuChen and Dubin 1984; HsuChen et al. 1984), the honeybee, Apis mellifera (Vlassak et al. 1987; Crozier et al. 1989; Cornuet et al. 1991; Crozier and Crozier 1992), and various

1. Key words: Spodoptera, molecular systematics, mtDNA sequences, 16S ribosome, ND-1 gene.

Address for correspondence and reprints: Dorothy P. Pashley, Entomology Department, Louisiana State University, Baton Rouge, Louisiana 70803.
regions are known for several Drosophila species [D. melanogaster (deBruijn 1983; Satta et al. 1987; Garesse 1988; Kobayashi and Okada 1990), D. teissieri (Monnerot et al. 1990), D. virilis (Clary and Wolstenholme 1987), and Hawaiian Drosophila (DeSalle et al. 1987)].

Comparative studies indicate that mitochondrial DNA (mtDNA) evolution in insects differs from that in vertebrates, in several ways. Insect mtDNA does not exhibit as rapid a rate of change as do nuclear genes; it is only 1-2 times the single-copy nuclear-DNA rate (Powell et al. 1986; Solignac et al. 1986; Vawter and Brown 1986; Caccone et al. 1988; Sharp and Li 1989), compared with 5-10 times the rate in vertebrates (Brown 1983). Insect mtDNA is rich in A and T nucleotides (70%-80%), despite evidence for a kinetic bias favoring transitions (De&he et al. 1987; Satta et al. 1987). It also apparently exhibits an elevated rate of nucleotide substitutions at synonymous codon sites (Moriyama 1987; Sharp and Li 1989), although there is variability in the “neutrality” of these sites within the molecule (DeSalle et al. 1987).

Although DNA sequence data have provided insights into evolutionary history in numerous cases (Moritz et al. 1987; Mindell and Honeycutt 1990), evidence is accruing that homoplasy can be high enough at some taxonomic levels to obscure phylogenetic signals (Kraus and Miyamoto 1991; Martin and Pashley 1992). Resolution can be enhanced in some cases by removal of homoplastic character sets. Within mtDNA, protein-coding genes can recover divergence events up to ~30 Mya before multiple substitutions at synonymous codon sites obscure cladogenetic events (Moritz et al. 1987). Removal of these characters and rapidly evolving transitional sites may decrease homoplasy and extend the use of protein-coding regions severalfold (DeSalle et al. 1987). Identification of homoplasy caused by rapidly evolving sites in RNA genes is not so straightforward. For example, regions of secondary structure that were initially thought to be highly constrained and most useful for ancient speciation events (Wheeler and Honeycutt 1988) are not more constrained in all groups (Mindell and Honeycutt 1990). Only through comparative studies on diverse taxa will the evolutionary rates and constraints on molecules become apparent and their use in systematics enhanced. For this reason we undertook this study of lepidopteran insect mtDNA.

**Material and Methods**

**Samples**

*Spodoptera* were obtained from a laboratory colony derived from adults trapped in Crowley, La., and were supplemented with larvae feeding on bermuda grass, corn, and sorghum collected from various Louisiana locations. Two host-plant strains exist that likely represent different species (Pashley 1986, 1989), and cloned mtDNA was determined to be that of the “rice strain” (D. P. Pashley, unpublished data).

**mtDNA Isolation, Cloning, and Mapping**

After their wings had been removed, adult moths were pooled to generate 2 g of material. The sample was homogenized in buffer (220 mM mannitol, 70 mM sucrose, 1 mM ethylenediaminetetraacetate, pH 7.4), and mtDNA was isolated according to a modification of the method of Lansman et al. (1981) (Ke 1990). Purified mtDNA was digested with XbaI and was cloned into plasmid pUC19. *Escherichia coli* strains JM83 and DH5α were transformed with ligated DNA, and recombinant plasmids were isolated. Competent cells were obtained, and selected colonies were prepared.

Three fragments (8.1 kb, 6.2 kb, and 2.0 kb) were cloned, and the *Spodoptera* mtDNA was determined to be ~16.3 kb in size. Recombinant plasmids were obtained
with these three fragments, as well as with several smaller ones (4.3 kb, 1.0 kb, and 0.3 kb). To verify that plasmids contained Spodoptera mtDNA inserts, we cleaved the DNA with XbaI and hybridized it to a Spodoptera mtDNA probe. Smaller clones contained Spodoptera mtDNA and could represent either polymorphisms in sample DNA or cleavage during the cloning procedure. Fragments were positioned relative to each other through restriction-endonuclease mapping using 29 restriction enzymes (Ke 1990). The 0.3-kb clone was determined to be part of the 6.2-kb fragment immediately adjacent to the 2.0-kb fragment.

Sequencing of Clones

The 2.0-kb and 0.3-kb fragments were completely sequenced. The recombinant plasmid containing the 0.3-kb fragment was used as template for double-stranded dideoxy-chain termination sequencing (Sanger et al. 1977) using the Sequenase 2.0 kit (United States Biochemical Corporation 1989). The larger, 2.0-kb fragment was subcloned prior to sequencing, by using a deletion method involving the ExoIII/Mung-bean nuclease kit (Stratagene). Eight subclones were generated and sequenced as above. All clones and subclones were completely sequenced in both directions and were scored two to six times, and a consensus sequence was generated in the few cases when a conflict was evident.

PCR Amplification and Sequencing

Two primers, conserved within insects, were made on the basis of initial results and alignment of the S. frugiperda sequence with Drosophila yakuba (Clary and Wolstenholme 1985a). These were used for PCR amplification (Saiki et al. 1985; Mullis and Faloona 1987) and sequencing (single-stranded sequencing using the Sequenase kit) of a region containing a tRNA and part of the ND-1 gene. Template DNA was isolated (according to the method of Pashley 1989), amplified, and sequenced for the following four lepidoptera: Pseudoplusia includens (Noctuoidea: Noctuidae), Cosmostoma myodora (Noctuoidea: Arctiidae), Symmerista albifrons (Noctuoidea: Notodontidae), and Phoebis sennae (Papilionoidea: Pieridae).

Analyses

The Spodoptera sequence was aligned manually with the four lepidopteran sequences and with the following entire or partial insect mtDNA sequences obtained from the literature or GenBank: D. yakuba (Diptera) (Clary and Wolstenholme 1985a); Aedes albopictus (Diptera) (Hsu-Chen et al. 1984), Apis mellifera (Hymenoptera) (Vlasak et al. 1987), and Locusta migratoria (Orthoptera) (McCracken et al. 1987; Uhlenbusch et al. 1987). A phylogenetic analysis was conducted on the 314-bp region of the ND-1 gene, a region common to all species except Aedes and Apis. Sequence divergence (% divergence = 100 × fraction of compared sites that do not match) was estimated for various taxonomic levels and molecular regions. Relative-rate tests were applied to the data, to assess homogeneity of evolution of certain regions. PAUP (version 3.0; Swofford 1985) was used to generate minimum-length trees by using the maximum-parsimony criterion. Accelerated transformations were used to optimize characters. An exhaustive search of all trees was performed using assigned outgroups to root trees. All most-parsimonious trees were examined, and a strict consensus tree occasionally was generated. Data were constrained to the accepted phylogeny, and tree length was compared with most-parsimonious solutions. Outgroup effects were assessed by reanalysis at four hierarchical levels by using the most closely
related taxon as the outgroup. Effects of homoplasy in character sets were examined by analysis of the same taxa by using all characters, first- and second-codon sites, and amino acid sequences. A bootstrap analysis with 1,000 replications was performed on the entire data set.

Results and Discussion
Sequence Composition

The two sequenced fragments from *Spodoptera* were determined to be 232 bp and 2,020 bp in length and were separated by four bases representing part of an *Xba*I recognition site. The total sequence length is 2,256 bp (fig. 1). By alignment with *Drosophila yakuba* (Clary and Wolstenholme 1985a), the *Spodoptera* sequence was determined to contain the tRNA for valine (tRNA_{Val}), the entire 16S rRNA subunit, the tRNA for leucine (tRNA_{Leu}), and the beginning three-quarters of the ND-1 gene (fig. 2). mtDNA of *Spodoptera* is 82% A/T rich for the entire sequence, a finding consistent with proportions reported for other insects. This richness is present in all regions but is most extreme in the tRNA_{Leu} (A+T = 95%), followed by the 16S rRNA (84%), tRNA_{Val} (83%), and ND-1 (77% overall and 95% in third-codon positions).

Gene order for *Drosophila* and *Spodoptera* is identical, except for the presence, in *Spodoptera*, of a large insert (64 bp) that is of unknown function and between the 16S rRNA and tRNA_{Leu}. This insert is composed almost entirely of A and T nucleotides (one C and one G are present), with an AT repeat occurring 20 times. Although absent in the locust, this insert is highly variable within the lepidopterans sequenced. The species most closely related to *Spodoptera*, *Pseudoplusia*, contains only a 9-bp insert and no repeating AT unit, whereas both *Cosmosoma* and *Symmerista* contain intermediate-length inserts (32 bp and 41 bp, respectively), with the AT repeat occurring nine and seven times, respectively. A similar region of much greater length (>200 bp) and associated with the same tRNA exists between the COI and COII genes in honeybee mtDNA (Cornuet et al. 1991). These sites are highly variable within bee congeners and are hypothesized to function as control regions.

In the dipteran order (*Drosophila* and *Aedes*), sequence conservation is greatest in tRNA genes (table 1). A comparison of Lepidoptera and Diptera reveals that the ND-1 gene and tRNA_{Val} are the most conserved sequences (table 1). This result differs from those of vertebrate studies that indicate that the slowest rate of change occurs in mitochondrial rRNA genes (Moritz et al. 1987).

Transfer RNA

The tRNA_{Val} was easily verified by folding the region into the characteristic cloverleaf structure. The *Spodoptera* tRNA is 7 bp shorter (two indels) than *Drosophila* and contains 13 nucleotide differences (6 of which are compensatory). Almost all changes are located in the two lateral stem/loop structures (the dihydrouridine and T/C arms). The anticodon stem and loop exhibit typical tRNA conservation in size and sequence (Clary and Wolstenholme 1985a).

The assignment of tRNA_{Leu} to the other tRNA-like region is speculative, but the high similarity of the 3' half to those of other insect orders justifies this preliminary designation. A cloverleaf structure with the correct anticodon can be constructed, although the presence, in *Spodoptera*, of an AT insert near the 5' end of the anticodon loop is not consistent with anticodon stem/loop structures. Comparison of structures of four lepidopterans supports the tRNA_{Leu} designation. *Pseudoplusia* contains an
extra base at the same position within the anticodon loop, but *Cosmosoma* and *Symmerista* exhibit typical tRNA conformation.

As with RNA Val, most of the divergence between orders is in the two lateral stem/loop arms. However, within Lepidoptera it is dispersed throughout the molecule, with the same positions tending to show variation in multiple taxa.

Large-Subunit Ribosomal RNA

The entire 16S ribosomal region of *Spodoptera* was aligned with those of *Drosophila*, *Aedes*, and *Apis*. Only the second half of the molecule was aligned with the locust, because extensive divergence in the first half resulted in long stretches of uncertain alignment. Alignment difficulties also arose in certain regions of the three holometabolous insect orders but could be improved, and homology could be verified using secondary-structure configurations for *Drosophila* (Clary and Wolstenholme 1985b). Although numerous highly conserved regions occur throughout the molecule, sequence variation is dispersed fairly evenly, with slightly greater divergence at the 5' end. Changes do not appear to be biased toward stem or loop regions, as nearly equal rates of change exist in each.

Nucleotide divergence in the 16S region is consistent with phylogenetic hypotheses (Kristensen 1981). Diptera and Lepidoptera are more closely related to each other (24% divergence) than to Hymenoptera (31%). Identical rates of change between pairwise taxa in ordinal comparisons (i.e., relative-rate test) suggest a uniform rate of evolution in this molecule. Although too few taxa have been sequenced in insects to determine the use of this molecule for phylogenetic studies, it provides a wealth of characters at the ordinal level. However, alignment difficulties in the first half of the molecule compromise its use in the most ancient insect divergences.

ND-1 Gene

The entire region of the ND-1 gene was easily aligned with those of *Drosophila* and *Locusta*. The methionine start codon (ATA) in *Spodoptera* is probably homologous to the locust start codon, because they are both located one amino acid residue in front of the start codon in *Drosophila*. With the exception of this and a three-base deletion in the locust, there are no other differences in length. Between pairs of insect orders, divergence in the amino acid sequence relative to the nucleotide sequence was approximately the same, in the entire molecule (% amino acid sequence divergence and % nucleotide sequence divergence were as follows: between *Spodoptera* and *Drosophila*, 18 and 19, respectively; between *Spodoptera* and *Locusta*, 22 and 24, respectively; and, between *Drosophila* and *Locusta*, 22 and 24, respectively). As with the rRNA gene, the rate of change along the Lepidoptera and Diptera lines, relative to Orthoptera, is uniform.

Phylogenetic Analysis Using the ND-1 Gene Region

Sequences were generated for four additional Lepidoptera, to determine the use of the ND-1 gene for phylogenetic analyses at various taxonomic levels within insects. A 314 bp region (fig. 1, sites 1556–1869) was common to all Lepidoptera, *Drosophila*, and the locust and served as the character data set for analyses.

Phylogenetic relationships among these seven taxa are not controversial (Forbes 1923; Brock 1971; Kristensen 1981; Kitching 1984; Miller 1991), although divergence times are difficult to estimate, because of a poor insect fossil record (fig. 2). Orthopterans exhibit incomplete metamorphosis (referred to as a "hemimetabolous condition"),
Fig. 1.—DNA sequence (5' to 3') for a 2,256-bp portion of the mitochondrial genome of *Spodoptera*, indicating positions of inferred genes. Transfer RNA genes are underlined, with the anticodon being double underlined. The ND-1 gene amino acid sequence, predicted on the basis of the *Drosophila* code, is shown below the nucleotide sequence.
Fig. 2.—Accepted phylogeny and taxonomic classifications based on morphology for the taxa included in the ND-1 sequence analysis. Divergence times are rough approximations [in millions of years before present (MYBP)] based on the fossil record.

whereas Diptera and Lepidoptera exhibit the derived developmental pathway involving complete metamorphosis (holometabolous). Fossil evidence indicates that insects evolved ≥400 Mya and that the orthopterans (and other hemimetabolous insect orders) were present by ≥300 Mya (Riek 1970). Holometabolous orders appear in the fossil record 270 Mya, and most extant orders were present by ~250 Mya. Within the 11 holometabolous orders, a panorpoid monophyletic group is composed of two sister clades one of which contains Diptera, Mecoptera, and Siphonaptera and the other of which contains Lepidoptera and Tricoptera (Kristensen 1981). Tricoptera and Lepidoptera probably diverged ~180 Mya, suggesting that their split from other panorpoid orders took place 200–250 Mya (Whalley 1986). Within Lepidoptera, butterflies (Papilionoidea) and noctuid moths are not particularly closely related, according to most systematists (Nielson 1991). The fossil record is scantly, but butterflies were present at least by the Eocene (65 Mya) (Durden and Rose 1978; Whalley 1986) and were probably a distinct superfamily by 100 Mya. The oldest fossil records for Noctuoida are in the Miocene (5–25 Mya; Whalley 1986), but some evidence suggests that modern families existed 65 Mya (Gall and Tiffney 1983). It is reasonable to conclude that the Noctuoidea also were present 100 Mya and that speciation within the superfamily occurred throughout the Tertiary (65 Mya to the present), although no specific dates are available for lower taxonomic levels. Within noctuid moths, notodontids are considered to be basal to Arctiidae and Noctuidae, the latter two of which are referred to herein as “derived noctuoids” (Forbes 1923; Brock 1971). *Spodoptera* and *Pseudoplusia* are members of different subfamilies within the same family, Noctuidae.

Sequence divergence based on nucleotide characters for the ND-1 region (table 2) is consistent with this phylogeny. A UPGMA phenogram of these data is topologically identical to the tree in figure 2. With amino acid data (table 2), only two taxa switch positions; the notodontid (*Symmerista*) is more closely related to the noctuids

<table>
<thead>
<tr>
<th>Order</th>
<th>Superfamily</th>
<th>Family</th>
</tr>
</thead>
<tbody>
<tr>
<td><em>Spodoptera</em></td>
<td>Lepidoptera</td>
<td>Noctuoidea</td>
</tr>
<tr>
<td><em>Pseudoplusia</em></td>
<td>&quot;</td>
<td>&quot;</td>
</tr>
<tr>
<td><em>Cosmosoma</em></td>
<td>&quot;</td>
<td>Arctiidae</td>
</tr>
<tr>
<td><em>Symmerista</em></td>
<td>&quot;</td>
<td>Notodontida</td>
</tr>
<tr>
<td><em>Phoebia</em></td>
<td>&quot;</td>
<td>Papilionoidea</td>
</tr>
<tr>
<td><em>Drosophila</em></td>
<td>Diptera</td>
<td>&quot;</td>
</tr>
<tr>
<td><em>Locusta</em></td>
<td>Orthoptera</td>
<td>&quot;</td>
</tr>
</tbody>
</table>

---

Estimated Divergence Time, MYBP

300 200 100 0
Table 1
Percent Divergence between Insect Orders and Families

<table>
<thead>
<tr>
<th>GENE</th>
<th>Spodoptera and Drosophila</th>
<th>Drosophila and Aedes</th>
</tr>
</thead>
<tbody>
<tr>
<td>tRNA^{Val}</td>
<td>20</td>
<td>(9)^{c}</td>
</tr>
<tr>
<td>16S rRNA</td>
<td>23</td>
<td>16</td>
</tr>
<tr>
<td>tRNA^{Leu}</td>
<td>35</td>
<td>9</td>
</tr>
<tr>
<td>ND-1</td>
<td>19</td>
<td>(17)^{d}</td>
</tr>
</tbody>
</table>

**NOTE.**—Gaps are excluded from all calculations.

^a Represent order comparisons.
^b Represent family comparisons. Values in parentheses are portions of genes.
^c 56-bp region.
^d 177-bp region.

(Spodoptera and Pseudoplusia) than to the arctiid (Cosmosoma). In contrast to this phenetic measure, a phylogenetic approach employing parsimony did a poor job of recovering the accepted phylogeny below the order level. In the entire character data set of 314 sites, 150 were varied, and 107 were potentially informative, when the locust was used as the outgroup (table 3, col. 1). A single most-parsimonious tree was generated (fig. 3) that supported the monophyly of Lepidoptera, with Diptera placed basal to it. Relationships within Lepidoptera were incorrect. These data were thus able to resolve ordinal relationships but not relationships within an order. The next shortest trees were two steps removed, and the accepted phylogeny was eight steps longer.

Given this result, several approaches were taken to evaluate the nature of changes in this gene region and to determine whether a phylogenetic signal could be extracted by varying both the outgroup and the type of characters included in the analysis. When first and second sites or amino acids were used as characters alone, none of the analyses at the various levels produced the correct tree. When the locust was used as the outgroup, first- and second-site characters produced four trees that retain correct higher taxonomic relationships but little resolution (or incorrect placement) within Lepidoptera [tree length (T1) 117, consistency index (CI) 0.66, rescaled CI (RC) 0.39, and retention index (RI) 0.49]. The accepted tree was two steps longer. Analysis with

Table 2
Percent Divergence for a 314-bp Region of the ND-1 Gene

<table>
<thead>
<tr>
<th></th>
<th>SF</th>
<th>PI</th>
<th>CM</th>
<th>SA</th>
<th>PS</th>
<th>DY</th>
<th>LM</th>
</tr>
</thead>
<tbody>
<tr>
<td>SF</td>
<td>10</td>
<td>20</td>
<td>15</td>
<td>28</td>
<td>32</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>PI</td>
<td>9</td>
<td>17</td>
<td>15</td>
<td>25</td>
<td>31</td>
<td>35</td>
<td></td>
</tr>
<tr>
<td>CM</td>
<td>13</td>
<td>13</td>
<td>22</td>
<td>25</td>
<td>34</td>
<td>37</td>
<td></td>
</tr>
<tr>
<td>SA</td>
<td>13</td>
<td>13</td>
<td>17</td>
<td>27</td>
<td>35</td>
<td>33</td>
<td></td>
</tr>
<tr>
<td>PS</td>
<td>21</td>
<td>19</td>
<td>20</td>
<td>21</td>
<td>38</td>
<td>42</td>
<td></td>
</tr>
<tr>
<td>DY</td>
<td>22</td>
<td>23</td>
<td>24</td>
<td>25</td>
<td>29</td>
<td>38</td>
<td></td>
</tr>
<tr>
<td>LM</td>
<td>25</td>
<td>25</td>
<td>27</td>
<td>26</td>
<td>35</td>
<td>37</td>
<td></td>
</tr>
</tbody>
</table>

**NOTE.**—Data above the diagonal are percent amino acid divergence; data below the diagonal are percent nucleotide divergence. SF = Spodoptera frugiperda; PI = Pseudoplusia includens; CM = Cosmosoma myodora; SA = Symmerista albifrons; PS = Phoebis sennae; DY = Drosophila yakuba; and LM = Locusta migratoria.
Table 3
Summary of Character Information for the ND-1 Region Used in the Phylogenetic Analysis
Assuming the Accepted Phylogeny is Correct

<table>
<thead>
<tr>
<th>OUTGROUP (TAXONOMIC LEVEL OF INGROUP)</th>
<th>LM (HO)</th>
<th>DY (LO)</th>
<th>PS (NS)</th>
<th>SA (DN)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Taxa in ingroup</td>
<td>6</td>
<td>5</td>
<td>4</td>
<td>3</td>
</tr>
<tr>
<td>Variable characters</td>
<td>150</td>
<td>130</td>
<td>99</td>
<td>69</td>
</tr>
<tr>
<td>Autapomorphic characters</td>
<td>43</td>
<td>45</td>
<td>32</td>
<td>34</td>
</tr>
<tr>
<td>Potentially informative characters</td>
<td>107</td>
<td>85</td>
<td>67</td>
<td>35</td>
</tr>
<tr>
<td>Nonhomoplastic characters (NHC)</td>
<td>48</td>
<td>44</td>
<td>43</td>
<td>27</td>
</tr>
<tr>
<td>Homoplastic characters (HC)</td>
<td>59</td>
<td>41</td>
<td>24</td>
<td>8</td>
</tr>
<tr>
<td>NHC transversions/transitions</td>
<td>39/9</td>
<td>38/6</td>
<td>31/12</td>
<td>23/4</td>
</tr>
<tr>
<td>HC transversions/transitions</td>
<td>51/8</td>
<td>35/6</td>
<td>19/5</td>
<td>6/2</td>
</tr>
<tr>
<td>NHC 1 + 2 site/3 site changes</td>
<td>35/13</td>
<td>27/17</td>
<td>20/23</td>
<td>9/18</td>
</tr>
<tr>
<td>HC 1 + 2 site/3 site changes</td>
<td>21/38</td>
<td>14/27</td>
<td>10/14</td>
<td>2/6</td>
</tr>
<tr>
<td>Synapomorphic for Holometabolous orders</td>
<td>23</td>
<td>...</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Synapomorphic for Lepidoptera order</td>
<td>19</td>
<td>32</td>
<td>...</td>
<td>...</td>
</tr>
<tr>
<td>Synapomorphic for Noctuoidea superfamly</td>
<td>3</td>
<td>9</td>
<td>31</td>
<td>...</td>
</tr>
<tr>
<td>Synapomorphic for derived Noctuoidea</td>
<td>1</td>
<td>3</td>
<td>6</td>
<td>18</td>
</tr>
<tr>
<td>Synapomorphic for Noctuidae family</td>
<td>0</td>
<td>0</td>
<td>6</td>
<td>9</td>
</tr>
</tbody>
</table>

* LM = Locusta migratoria; DY = Drosophila yakuba; PS = Phoebis sennaq and SA = Symmerista albifrons.
* HO = Holometabolous orders; LO = Lepidoptera order; NS = Noctuoidea superfamly; and DN = derived Noctuoidea.

Amino acid sequences produced four trees, one of which was close to the correct tree (Symmerista and Cosmosoma were switched), but the consensus left no resolution within Lepidoptera (T1 117, CI 0.75, RC 0.39, and RI 0.44). The correct tree was one step longer.

Resolution was enhanced slightly within Lepidoptera by using both Drosophila as the outgroup and first and second codon sites only. When all the data were used, one tree was produced, which shared little similarity to the correct one (T1 192, CI 0.58, RC 0.27, and RI 0.34). The correct tree was two steps longer. Analysis with first and second sites only produced four trees (T1 88, CI 0.63, RC 0.29, and RI 0.35), one of which was the correct tree. Amino acid data produced nine trees and a consensus with no resolution (T1 88, CI 0.67, RC 0.32, and RI 0.36). The correct tree was one step longer.

Within Lepidoptera, when the pierid butterfly was used as the outgroup, one tree was found that misplaced only Cosmosoma and Symmerista (T1 134, CI 0.63, RC 0.35, and RI 0.30), and the correct tree was one step longer. The same single tree was recovered by using first- and second-site data (T1 58, CI 0.77, R 0.59, and RI 0.63), and the correct tree was two steps longer. The amino acid sequences produced a single poor-quality tree.

Finally, within the derived Noctuoidea, when Symmerista was used as the outgroup, the most-parsimonious tree was the correct one (T1 90, CI 0.65, RC 0.43, and RI 0.47). Analysis of first and second codon sites and amino acid sequences both produced two trees, one being the correct one. At this level, the inclusion of third-site changes enhanced resolution slightly.

Using the accepted phylogeny, an analysis of the information content of the characters was performed to determine how various types of characters were evolving (table 3). Characters were considered nonhomoplastic if they supported (were syna-
Several points are evident from this character analysis. First, as expected, homoplaspy increases with time as more rapidly evolving characters become relatively more common in the data (table 4). With any character subset, the greatest amount of homoplasy occurs at the supraordinal level with the locust as the outgroup (because of the presence of multiple-hit characters). Second, there is a strong apparent transversion bias. In the entire data there are 96 transversions (84 of which are A/T events), 25 transitions, and 29 multistate (transversions and transitions) characters. In pairwise comparisons at all levels, 24%-30% of the substitutions are transitions. There is no indication that homoplastic characters are biased toward transitions, and resolution would not be enhanced by eliminating transitions (table 4; percent homoplasy in transversion-only data is equal to that in all characters). Third, relative to homoplastic ones, nonhomoplastic characters are consistently biased toward first- and second-site changes. Some 3%-17% of the homoplasy is removed at various levels by using only these types of characters, with the greatest reduction in homoplasy occurring at more ancient divergence times. Finally, below the highest level (locust as the outgroup), the distribution of nonhomoplastic characters is concentrated on the branch between the outgroup and the ingroup (table 3). This provides some insight into the poor resolution at lower taxonomic levels. With the locust as the outgroup, the vast majority of nonhomoplastic characters are functioning at the higher level (44 of 48 characters are synapomorphic for Holometabola and Lepidoptera), with the 59 homoplastic ones obscuring resolution at lower levels. This result is expected when molecular data are used to span a broad evolutionary time period, but the surprising result is that, by choosing more appropriate outgroups at each level, resolution within the ingroup is improved only slightly, if at all. For example, with Drosophila as the outgroup, there

![PAUP cladogram generated from sequence data for the ND-1 gene region, with the locust as the outgroup (tree length=246, CI 0.59, RC 0.28, and RI 0.38). Branch lengths are indicated above branches. Bootstrap percentages are in circles at nodes.](image-url)
Table 4
Homoplasy Frequency (Homoplastic/Potentially Informative Characters) in the ND-1 Sequence, for Different Character Subsets and Different Outgroups

<table>
<thead>
<tr>
<th>FREQUENCY IN OUTGROUP*</th>
<th>LM</th>
<th>DY</th>
<th>PS</th>
<th>SA</th>
</tr>
</thead>
<tbody>
<tr>
<td>All characters</td>
<td>0.55</td>
<td>0.48</td>
<td>0.36</td>
<td>0.23</td>
</tr>
<tr>
<td>Transversions only</td>
<td>0.57</td>
<td>0.48</td>
<td>0.38</td>
<td>0.21</td>
</tr>
<tr>
<td>1 + 2 Codon sites only</td>
<td>0.38</td>
<td>0.34</td>
<td>0.33</td>
<td>0.18</td>
</tr>
</tbody>
</table>

*Abbreviations are as in table 3.

are 32 nonhomoplastic characters that support the lepidopteran clade but only 12 that resolve relationships among the four Noctuoidea taxa. Furthermore, of the 67 potentially informative characters within this superfamily, only 12 are consistent with the accepted phylogeny.

There are several explanations for poor resolution below the order level. First, the outgroups may be too distantly related to the ingroup. This may be a valid explanation at some levels, but it is hard to imagine a better outgroup for derived noctuids than notodontids. It is also possible that, with more sequence data, a phylogenetic signal would emerge. However, it is disconcerting that such a well-established phylogeny cannot be recovered easily with a fairly high character-to-taxon ratio (even at the lowest taxonomic level there are 35 potentially informative characters for four taxa). Finally, homoplasy in the mtDNA at the subordinal level may simply be so great as to prevent extraction of a phylogenetic signal, irrespective of the amount of sequence data gathered; that is, phylogenetic resolution at particular levels (e.g., superfamilies and families in Lepidoptera) may be difficult because characters tend to fall into one of two classes—(1) those that are highly conserved and resolve ancient splits (e.g., among orders) and (2) those that are unconstrained, evolve rapidly, and provide reliable data only at lower taxonomic levels (among species and genera). This could be a function of both the time of the speciation events and the nature of divergence among taxa. Bursts of speciation resulting in short internodes connecting lineages will present problems for characters with a fairly uniform rate of change (Lanyon 1988). Regardless of whether poor resolution is due to the nature of evolution in the molecule or in the taxa, it is uncertain whether increasing the number of mtDNA sites or taxa will improve resolution. Comparisons using additional molecules are needed to distinguish between these alternatives.

Sequence Availability

The fall armyworm 2,257-bp sequence has been deposited in GenBank under accession number M76713. The sequence and amino acid alignments have been submitted to EMBL.

Acknowledgments

This research was supported by funds from Louisiana Education Quality support fund/NSF-EPSCoR grant 86LBR-48-048-08 and from NSF grant BSR-9006273. Lepidopteran sequence data were gathered with the lab assistance of Julie Constable,
LITERATURE CITED


UHLENBUSCH, I., A. MCCracken, and G. GELLISSEN. 1987. The gene for the large (16S)


**WESLEY M. BROWN,** reviewing editor

Received August 12, 1991; revision received June 5, 1992

Accepted June 5, 1992