Nucleotide sequences from a variety of sources display compositional bias (e.g., Bernardi et al. 1985; Jukes and Bhushan 1986; Sueoka 1988; Crozier and Crozier 1993; Dowton and Austin 1995). This bias in the proportion of the four nucleotides in a particular DNA fragment was originally thought to arise from mutational bias in the interconversion of adenine-thymine (AT) and guanine-cytosine (GC) base pairs (Sueoka 1962). However, the mechanism whereby compositional bias develops must be more complex than this. Two distinct types of compositional bias are observed: strand-specific and strand-nonspecific. Strand-specific compositional bias occurs when one strand contains higher levels of certain nucleotides over others, e.g., A+C or A+G bias. It is thought to develop when the rate of nucleotide substitution (e.g., A→C) differs from that in the opposite direction (e.g., C→A) (Jermiin et al. 1996). Strand-nonspecific compositional bias occurs when both strands contain a high level of A+T or G+C.

Trends in compositional bias may follow phylogenetic lineages (Sueoka 1962; Hori and Osawa 1987; Jermin and Crozier 1994; Jermin et al. 1994; Dowton and Austin 1995). This suggests that a phylogenetic analysis of compositional bias may lead to an understanding of how this bias develops. In the Hymenoptera (wasps, bees, ants), broad phylogenetic relationships have been consistently suggested from fossil (Rasnitsyn 1980, 1988), morphological (Gibson 1985), and molecular (Dowton and Austin 1994) data (see fig. 1). We recently traced an increase in compositional bias to one particular hymenopteran lineage, the Apocrita (Dowton and Austin 1995). The Hymenoptera thus represent an ideal model group in which to study the evolution of compositional bias.

In the present study, we determined the individual nucleotide contents of a fragment of the 16S rRNA gene, homologous to nucleotides 13474–13895 of the honeybee mitochondrial genome (Crozier and Crozier 1993). We used all available hymenopteran homologues at the time of writing, except for one wasp, Schlettererius cinctipes. This taxon was omitted because of its statistical analysis described here avoids these potentially confounding effects.

We previously reported an increased AT-content in one hymenopteran lineage (the Apocrita) compared with its sister group (the Symphyta). Comparison of individual nucleotide contents suggested that only the A-content was increased in the Apocrita; the T-content appeared unchanged (table 1). Although both the G- and C-contents were quite low in the Symphyta, these appeared to decrease further in the Apocrita. However, to assess the significance of nucleotide content variation between the Diptera, Symphyta, and various apocritan clades, sequence data from multiple taxa within each
FIG. 1.—Broad phylogenetic relationships in the Hymenoptera. Representatives from apocritan clades that had both an increased A- and decreased G-content at variable sites (compared with a symphytan representative) are shown with hatched branches. Unambiguous changes between A and G on the branches of this tree are also shown; these were identified by character state reconstructions using PAUP (Swofford 1993) and MacClade (Maddison and Maddison 1992).

clade cannot be pooled due to the nonindependence of samples. Instead, we chose a representative taxon from each of the clades that appears in table 1. The choice of representative was based on the taxon whose sequence most closely resembled the average nucleotide composition of that clade, determined by the sum of least squares. The sequence of a representative taxon was aligned with the symphytan representative, and the nucleotide contents at the variable sites were determined. Nucleotides that aligned opposite gaps were not included in the calculation. The significance of changes was assessed by the chi-square test; e.g., Hₐ: A-content at variable sites of the symphytan representative ≠ A-content at variable sites of the braconid representative; H₀: A-content at variable sites of the symphytan representative = A-content at variable sites of the braconid representative.

The results of these tests appear in table 1, and are mapped onto the evolutionary tree depicted in figure 1. No change in nucleotide contents were observed when the Symphyta were compared with the Diptera. A significant increase in A-content was observed with the Apocrita/Symphyta comparison, suggesting that the Apocrita have acquired a higher A-content than have the Symphyta during their evolution from a common ancestor. No significant change in T-content was observed in this or any other comparison, indicating that the increase in AT-content reported earlier (Dowton and Austin 1995) has proceeded entirely through an increase in A-content. The evolution of a high T-content in this gene fragment must have predated the origin of the Hymenoptera, possibly during the evolution of the Endopterygota, as an increase in the AT-content of another mitochondrial gene has been traced to this time (Jermiin

Table 1
Average Nucleotide Composition of a 16S rRNA Gene Fragment from the Diptera: Nematocera and Various Hymenopteran Groups

<table>
<thead>
<tr>
<th>Taxonomic Group</th>
<th>% A</th>
<th>% T</th>
<th>% G</th>
<th>% C</th>
<th>Number of Variable Sites</th>
</tr>
</thead>
<tbody>
<tr>
<td>Diptera: Nematoceracum (n = 9)</td>
<td>35.2</td>
<td>38.9</td>
<td>16.2</td>
<td>9.7</td>
<td>60</td>
</tr>
<tr>
<td>Hymenoptera: Symphytace (n = 6)</td>
<td>37.2</td>
<td>40.1</td>
<td>14.3</td>
<td>8.6</td>
<td></td>
</tr>
<tr>
<td>Hymenoptera: Apocritaed (n = 35)</td>
<td>42.2**</td>
<td>39.7</td>
<td>11.2</td>
<td>6.8</td>
<td>60</td>
</tr>
<tr>
<td>Apocrita: Cynipoideef (n = 2)</td>
<td>41.4**</td>
<td>41.4</td>
<td>11.2**</td>
<td>6.1</td>
<td>87</td>
</tr>
<tr>
<td>Apocrita: Eviomorphafe (n = 8)</td>
<td>41.0</td>
<td>39.8</td>
<td>12.1</td>
<td>7.1</td>
<td>73</td>
</tr>
<tr>
<td>Apocrita: Proctotrupomorphah (n = 14)</td>
<td>42.8**</td>
<td>39.8</td>
<td>10.9**</td>
<td>6.6</td>
<td>72</td>
</tr>
<tr>
<td>Apocrita: Aculeatax (n = 3)</td>
<td>40.6</td>
<td>39.3</td>
<td>13.1</td>
<td>7.0</td>
<td>69</td>
</tr>
<tr>
<td>Apocrita: Ichneumoniad (n = 3)</td>
<td>39.7</td>
<td>40.1</td>
<td>12.7</td>
<td>7.5</td>
<td>81</td>
</tr>
<tr>
<td>Apocrita: Braconidaee (n = 5)</td>
<td>45.4**</td>
<td>39.0</td>
<td>9.0**</td>
<td>6.6</td>
<td>79</td>
</tr>
</tbody>
</table>

Note.—Sequences (with GenBank accession numbers in parentheses) were from *HsChen, Koin, and Dubin (1984; X01078), † Tang et al. (1995; U17727, U17729, U17731–32, U17735–36, U17738, U17741), ‡ Derr et al. (1992a, 1992b; not deposited), § Crozier and Crozier (1993; X01078), ‡ Dowton and Austin (1994; U06593–56, U06958–75), and the present study (U39948–57). n refers to the number of sequences that were used to estimate the average nucleotide content of a fragment of the 16S rRNA gene from a particular clade (both variable and invariant sites included). Asterisks indicate a significant difference in the nucleotide content at variable sites between a typical representative of a particular clade and a typical representative of the Hymenoptera: Symphyta; sites that aligned opposite inferred gaps were excluded.

** P < 0.01; tests subjected to sequential Bonferroni corrections. 
The extremely low but stable nature of the C-content (Clary and Wolstenholme 1985) marginally significant (P = 0.025) suggests that there is evolutionary pressure toward reduction, but that it has reached a functional minimum.

When comparisons were made between representatives of the Symphyta and those of various apocritan clades, significant variations in nucleotide contents were observed during the evolution of the Proctotrupomorpha, the Cynipoidea, and the Braconidae from a common ancestor with the Symphyta (table 1). In these comparisons, a significant increase in A-content was observed, as was a significant decrease in G-content.

The simplest explanation for the above observations is that the reciprocal changes in the G- and A-content of this gene fragment in certain apocritan clades has been predominantly caused by a substitutional bias from G to A. In particular, stem sites that are able to convert G-T or A-T base pairs (Clary and Wolstenholme 1985) might be responsible for the observed changes. Character state reconstructions will identify sites at which changes from G to A can be inferred during the evolution of the Hymenoptera, although, as outlined above, such reconstructions are likely to underestimate changes from G (rare state) to A (common state) in this particular case. With this caveat in mind, we examined the inferred, unambiguous changes in the branches leading to the various apocritan clades and mapped these onto a broad phylogenetic tree (fig. 1). The Cynipoidea were omitted from this analysis as their phylogenetic placement is unclear. Unambiguous changes were identified using PAUP (Swofford 1993) and MacClade (Maddison and Maddison 1992). Most inferred changes were between A and T, but there was no obvious directional trend (e.g., on the branches from the common ancestor of the Hymenoptera to the Proctotrupomorpha, there were 20 T→A and 13 A→T changes). A more detailed description of this apparent AT-transversion bias will be reported elsewhere. Conversely, changes between A and G were extremely unbalanced, with changes from G to A predominating (see fig. 1). Further, the trends observed were consistent with the statistical results presented in table 1. More G→A than A→G changes were inferred on branches leading to clades that had significantly increased A- and decreased G-contents, while the opposite was evident on branches leading to clades that did not show significant changes in nucleotide contents.

As mentioned, we place little confidence in the relative number of these inferred changes, but this examination helps identify some of the secondary structural sites that have changed from G to A or from A to G during the evolution of the Apocrita. These changes did not appear to be restricted to stems. Of the 17 changes from G to A during the evolution of the Apocrita from a common ancestor with the Symphyta, 8 were in stems and 9 were in loops. Of the three changes from A to G, one was in a stem and two were in loops. An examination of the changes in stem structures indicated that only three involved a G-T base pair converting to an A-T base pair.

Oxidative damage is a major source of mutations in mitochondrial genomes generally (e.g., Wagner, Hu, and Ames 1992; Martin 1995), of which the majority are thought to be G→A-T transitions (Wagner et al. 1992). The mechanism is thought to involve oxidative (Wagner et al. 1992) or hydrolytic deamination (Lindahl 1993) of dC to dU, changing the base pairing from dG to dA during the next round of replication. The strand-specific nature of our observations could be explained if the C-content on the coding strand was at a functional minimum, such that only dC residues on the noncoding strand were free to vary. This would lead to reciprocal variations in G- and A-content in the coding strand, as we observed. A related source of mutations in the mitochondria arises after excision repair of bases damaged by oxidation, where dATP is preferentially incorporated opposite the abasic site (Lindahl 1993). However, this mechanism is not consistent with our observations, as oxidation of dG leads to a G→T transversion.

In our previous assessment of compositional bias in the same 16S rRNA gene fragment in the Hymenoptera (Dowton and Austin 1995), we observed fluctuations in AT-content, evidence of strand-nonspecific compositional bias. However, in the present study we observed that these variations arose entirely through fluctuations in A-content, evidence of strand-specific compositional bias. Strand-specific compositional bias is generally measured as fluctuations in A+G- or A+T-content. However, measurement of A+C or A+G would have underestimated the extent of compositional bias in this study, as changes in G would have offset the inverse fluctuations in A. Given that a variety of compositional biases can develop, our findings suggest that assessment of individual nucleotide contents (rather than pairs of nucleotides) may expose additional components of compositional bias.

Acknowledgments

Thanks to Lars Jeremiin and Otto Schmidt whose comments greatly improved an earlier draft of this manuscript. This work was supported by a grant from the Australian Research Council.

LITERATURE CITED


Paul M. Sharp, reviewing editor

Accepted October 1, 1996