EVOLUTION OF MITOCHONDRIAL RIBOSOMAL RNA IN INSECTS AS SHOWN BY THE POLYMERASE CHAIN REACTION

Chris Simon
Graduate Faculty of Zoology, University of Hawaii
Honolulu, HI 96822

Svante Pääbo, Thomas D. Kocher, and Allan C. Wilson
Department of Biochemistry, University of California
Berkeley, CA 94720

INTRODUCTION

In the last several years many evolutionary geneticists have used nucleotide sequence information inferred from restriction site mapping of mitochondrial DNA (mtDNA) to produce phylogenetic hypotheses. The advent of the polymerase chain reaction (PCR) has made direct sequencing, in which 300-600 sequential base pairs are read at once from specific amplified regions, possible on the scale necessary for population studies because it simplifies the preparation of DNA for sequencing. Data from direct sequencing are less ambiguous and supply more information for a given level of effort than data from restriction mapping (1). In this study, PCR and chain termination sequencing were used to examine one domain of the small ribosomal RNA gene (12S, Domain III) in cicadas. The cicada sequence presented here for Magicicada tredecim, the large 13-year periodical cicada, represents the first for this domain in a protostome invertebrate outside the genus Drosophila.

CONSERVED PRIMERS

Highly conserved primers, designated 12SA and 12SB (1), were used to amplify and sequence a 326-bp segment of the mitochondrial small ribosomal subunit. All but 60 bp of this region lie within what has been called Section 3 (2) or Domain III (3). The 12SA primer has an 11-bp core which is conserved across eukaryotic, bacterial, plastid, and

1This work was supported by NSF and NIH grants to A.C. Wilson and NSF BSR 85-09164 to C. Simon.
mitochondrial small ribosomal RNAs in 103 out of 106 species examined. The 12SA primer has a 14-bp 3' end which is conserved across 96 of the same 106 species, with the one exception of the A in the sixth position of the fly, *Drosophila yakuba* (Fig. 1) (4). The conservation of 12SB is almost certainly related to the fact that this area has been identified as the site of tRNA attachment (5).

<table>
<thead>
<tr>
<th>12SA Primer</th>
<th>12SB Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ec C . . A C A . . . . . . . . . . T G G . . . . . . . .</td>
<td>Ec T G . T . . . . . . . . . . . . . . . . . . . .</td>
</tr>
</tbody>
</table>

**FIGURE 1.** Sequences of 12SA and 12SB primers, based on conserved regions (4), modeled after *Homo sapiens* mtDNA (Hs), and compared to homologous positions in *Drosophila yakuba* mtDNA (Dy) and *Escherichia coli* (Ec).

Although there are substitutional differences between human and fly in the 12SA and 12SB primers, we have found that the 12SA-B primer set that matches humans will amplify insect mtDNA and the 12SA-B primer set that matches the fly sequence will amplify human mtDNA (despite the substitutions near the critical 3' end of 12SA).

The extreme conservation of these two primers is important because it allows the examination of mtDNA sequences from species for which we have no previous sequence information such as the cicadas examined here. Furthermore, once one section of the mitochondrial genome of a species is sequenced, it can be used as a gateway to sequencing the rest of the genome by employing the inverse PCR technique (6).

**RESULTS**

Figure 2 gives the nucleotide sequence for a representative periodical cicada species, *Magicicada tredecim*, compared to *Drosophila yakuba* and *Homo sapiens*. 
Figure 2. Alignment and secondary structure of the region sequenced in *Magicicada tredecim* (DECI.M 13) with the homologous region in *Drosophila yakuba* (YAKUBA) and *Homo sapiens* (HOMO). In comparison to HOMO, dots (.) = identity, stars (*) = transitions, and dashes (-) = deletions. Apparent identities and transitions in regions of uncertain alignment are not identified (see text). Numbered boxes represent stem regions corresponding to those of Hixon and Brown (3) for HOMO and that of Clary and Wolstenholme (2) for *Drosophila yakuba*. Stem numbers are our own.
In total, six species of periodical cicada and one species in the presumed sister genus of *Magicicada*, *Okanagana vanhueseii*, were sequenced. At least two individuals of each species were examined and each was sequenced in both directions. No attempt was made in this investigation to obtain an estimate of within-species variation but restriction studies of periodical cicada mtDNA (7) suggest that there will be very little. Because of space limitations, comparisons of the six other cicada species will be presented elsewhere.

Figure 3 gives a pictorial representation of the secondary structure shown in Figure 2 for *Magicicada tredecim*. Bonds shown with dotted lines across loops are possible but do not correspond to the conformation seen in HOMO and YAKURA (with the exception of stem, 6 which will be discussed in a later paper).

![Secondary Structure Diagram](image)

Figure 3. Proposed secondary structure for Domain III of the 12S rRNA gene of *Magicicada tredecim*. 

---

238 Simon et al.
SELECTIVE CONSTRAINTS

The value of DNA sequences for estimating evolutionary rates can be enhanced by understanding the selective constraints on the molecule. In particular, the secondary structure of ribosomal RNA influences the rate and pattern of nucleotide substitution. In the sequences for Domain III, the existence of selective constraints is inferred from four observations: 1) compensatory mutations in helical stems; 2) higher levels of conservation in stems with long-versus short-range interactions; 3) high observed transition/transversion ratios in the most conserved regions; and 4) higher percent of A-T nucleotides in the least conserved regions. These selective constraints must be understood in order to calculate accurate rates of evolution and improve existing corrections which are designed to estimate number of multiple substitutions per site (8,9). Manske and Chapman (10) address the problem of nonuniformity of substitution rates in 5S rRNA. The problem is more complex for the larger rRNA molecules.

Elucidating patterns of nucleotide conservation is also valuable for phylogenetic analysis. Conserved areas can be used for comparing distantly related taxa and for PCR primers for initiating sequencing of unstudied taxa. Highly variable regions can be used to compare only closely related taxa. In distantly related taxa, alignments of these variable regions are uncertain. Furthermore, these regions would be high in multiple substitutions. Because of these two problems, we have not identified apparent homologies or transitions in hypervariable regions in Figure 2. Areas of uncertain alignment such as these, should only be used with extreme caution, if at all, for phylogenetic analysis (11). Using alternative alignments may yield substantially different results.

Conservation in Stems and Loops.

As expected (3), cicada long-range stems 3, 4, and 5 exhibit few differences from their Drosophila and human homologues. All of these substitutions are compensated in their pairing partner (allowing G-T as well as G-C bonds), and most, 90%, are transitions. From this we infer that these stem regions are slowly evolving and that multiple substitutions have not obscured the record of transitions, which are the replication mutations most common in mtDNA due to tautomeric base pairing possibilities (12,13).
From Figure 2, it can also be noted that one of the short-range stems, stem 7, and the loop it contains are highly conserved. Brimacombe et al. (5) suggested no explicit functional role for this stem but noted that Domain III is involved in binding to ribosomal proteins S7, S9, S10, S13, and S19. We suggest that loop 7 could be an important protein-binding site. Stem 7 may be involved in this interaction as well but if so, Noller and Woese's (14) conjecture that stems held together by proteins are freer to vary would be violated.

Base Composition and Extent of Conservation.

In insects, the degree of bias in base composition (expressed as % AT base pairs) is inversely proportional to the level of conservation. DeSalle et al. (15) point out that AT-richness in Drosophila mtDNA is high in general (77% AT compared to 56% AT in humans [15]) but is highest in silent sites (94% in D. yakuba) while humans show no such bias at silent sites in mtDNA (49% AT). Similarly, Hancock et al. (17) point out that the most variable regions in D. melanogaster nuclear DNA are highest in AT. High % AT is also found in mammalian pseudogenes (18). The % AT for M. tredecim for the 326-bp region sequenced was 78% compared to 77% in YAKUBA and 53% in HOMO for this same region.

For 12S Domain III, the relationship between degree of conservation and % AT-richness is shown in the table below. The % AT of bases conserved between insects and HOMO is significantly less than the % AT of bases conserved only across the insects compared ($\chi^2 = 5.2, 1$ df, $p < .05$). Note that cicadas and Drosophila are on the opposite ends of the insect phylogeny with Drosophila being highly derived holometabolous insects and cicadas being an early offshoot among the hemimetabolous orders (19).

TABLE 1. INVERSE RELATION BETWEEN BASE COMPOSITION AT CONSERVED SITES AND THE EXTENT OF CONSERVATION

<table>
<thead>
<tr>
<th>Criterion of Conservation</th>
<th>Degree</th>
<th># Bases</th>
<th>% AT</th>
</tr>
</thead>
<tbody>
<tr>
<td>Identical between Homo &amp; insects</td>
<td>High</td>
<td>97</td>
<td>61</td>
</tr>
<tr>
<td>Among Drosophila and cicadas</td>
<td>Moderate</td>
<td>36</td>
<td>86</td>
</tr>
<tr>
<td>Between cicada sister genera</td>
<td>Weak</td>
<td>49</td>
<td>83</td>
</tr>
<tr>
<td>Conserved only among Magicicada</td>
<td>Very Weak</td>
<td>57</td>
<td>93</td>
</tr>
</tbody>
</table>
The % AT of the bases conserved across the insects examined is not significantly different from the % AT of bases at lower levels of conservation. When more bases are examined, it is possible that a significant difference may be observed for this comparison.

ALTERNATIVE SECONDARY STRUCTURES

Dams et al. [4] provide an alignment and secondary structure for the mitochondrial small ribosomal RNA for 26 species ranging from animals to protists. The secondary structure models they provide for primates and Drosophila do not agree with those previously proposed for these two taxa [2, 3], both of which were based on the structure proposed for E. coli by Glotz and Brimacombe [20]. The difference between the Dams et al. model and the Glotz and Brimacombe model lies in what we have labeled stems 9-12.

There can be at least two explanations for the proposed alternative structures: 1) one structure may be incorrect; or 2) perhaps ribosomal RNA can switch between alternative structures during protein synthesis. Data from the seven cicada species examined may help resolve this question. An examination of the structure of M. tredecim in this region (Figure 4) shows that both structural arrangements can be drawn. The same is true for the other six cicada species examined, including an outgroup species O. vanduzei. The fact that for every species examined both structures can be drawn supports the hypothesis of a structural switch. We are not the first to propose structural switches for the small ribosomal RNA [20, 21]. Glotz and Brimacombe [20] found empirical evidence for two alternative stem pairings via ribonuclease digestion and two dimensional electrophoresis of E. coli 30S subunits. The switch they proposed required stems from Domain II to pair with stems from Domain III. One of the stems involved was the stem we have numbered "9". This stem lies just above the region we propose to be involved in switching. However, the switch we propose is qualitatively different in that it is a local one and it occurs in a region where mitochondrial ribosomes are "drastically eroded" in structure [22], with two stems shortened and one missing entirely, compared to bacterial and eukaryotic small subunits.

Comparative studies, in progress, will determine if both structures can be drawn for known mitochondrial 12S sequences. Experimental evidence (as in [20]) will then be required to see if both structures can be found.
Figure 4. Alternative structures for stems 11 and 12 of Domain III in *Magicicada tredecim*.

ACKNOWLEDGMENTS

We thank J. Archie, R. Cann, S. DeSa, J. Hunt, C. McIntosh, S. Palumbi, A. Sidow & other members of the Cann, Hunt, Palumbi, Simon, & Wilson labs.

REFERENCES


