The Lepidopteran Mitochondrial Control Region: Structure and Evolution[1]

Martin F. J. Taylor,[2] *Stephen W. McKechnie,*[3] *Naomi Pierce,*[4] *and Martin Kreitman*[5]

Department of Ecology and Evolutionary Biology, Princeton University

For several species of lepidoptera, most of the ~350-bp mitochondrial control-region sequences were determined. Six of these species are in one genus, *Jalmenus;* are closely related; and are believed to have undergone recent rapid speciation. Recent speciation was supported by the observation of low interspecific sequence divergence. Thus, no useful phylogeny could be constructed for the genus. Despite a surprising conservation of control-region length, there was little conservation of primary sequences either among the three lepidopteran genera or between lepidoptera and *Drosophila.* Analysis of secondary structure indicated only one possible feature in common—inferred stem loops with higher-than-random folding energies—although the positions of the structures in different species were unrelated to regions of primary sequence similarity. We suggest that the conserved, short length of control regions is related to the observed lack of heteroplasmy in lepidopteran mitochondrial genomes. In addition, determination of flanking sequences for one *Jalmenus* species indicated (i) only weak support for the available model of insect 12S rRNA structure and (ii) that tRNA translocation is a frequent event in the evolution of insect mitochondrial genomes.

Introduction

The noncoding control region of animal mitochondrial genomes is sometimes called the "D-loop" region for mammals or the "A+T-rich" region for invertebrates. From evidence to date, insect control regions are highly variable, providing a promising source of polymorphic markers for population genetics and phylogenetic reconstruction of closely related taxa. Our present knowledge of control-region function and evolution is hampered by a lack of comparative data, and control-region sequence data for a variety of taxa at different levels of phylogenetic separation would be useful.

Some authors have proposed the existence of highly conserved features common to all mitochondrial control regions. The presence of several replication origins in the *Petunia hybrida* mitochondrial genome has been demonstrated by DeHaas et al.

1. Key words: lepidoptera, control region, mitochondrial, evolution.

2. Present address: Department of Entomology, University of Arizona, Tucson, Arizona 85721.

3. Present address: Department of Genetics and Developmental Biology, Monash University, Clayton, Victoria 3168, Australia.

4. Present address: Museum of Comparative Zoology, Harvard University, Cambridge, Massachusetts 02138.

5. Present address: Department of Ecology and Evolutionary Biology, The University of Chicago, Chicago, Illinois 60637.

Address for correspondence and reprints: Martin F. J. Taylor, Department of Entomology, University of Arizona, Tucson, Arizona 85721.

Mol. Biol. Evol. 10(6):1259–1272. 1993.

0737-4038/93/1006-0009$02.00

(1991). They also claim the presence of primary-sequence homologues of yeast/*Escherichia coli* gyrase recognition sites and of transcription initiation sites, and, on that basis, they infer "structural" (i.e., "gene" order) homologies to yeast and mammalian control regions.

Control regions of mammalian genomes have been extensively sequenced, and, while there is substantial divergence and rearrangement, including heteroplasmy and polymorphism for large repeats, there are blocks of sequence conservation clearly identifiable across mammalian sequences. These include a potential open reading frame (Hoelzel et al. 1991; Saccone et al. 1991). However, with one possible exception for *Drosophila,* no similar sequences appear in invertebrates.

Bark-weevil mitochondrial control regions are very large and variable in size (9–13 kb; Boyce et al. 1989). In these species heteroplasmy is common, and the order of the major genes in the coding regions, based on restriction-site maps, appears to be the same as for *Drosophila.* Although this observation is inconsistent with the idea that smaller size of mitochondrial molecules is selectively favorable, it may be the consequence in these species of an unusually high rate of mutational rearrangements. This could result from misalignment of control-region repeats during replication, as is found for sturgeon and macaques (Buroker et al. 1990; Hayasaka et al. 1991).

There are considerable length differences between control regions of *Drosophila* species, varying from ~1 kb in *D. virilis* and *D. yakuba* to 5.1 kb in *D. melanogaster* (Clary and Wolstenholme 1987). There is extensive intraspecific polymorphism for length of control regions within *D. melanogaster,* with mitochondrial genome size of 18.1–19.9 kb (Hale and Singh 1986). Although of similar size, the control regions of congeners *D. virilis* and *D. yakuba* are very different, with only two small blocks of 49 bp and 276 bp sharing any sequence similarity (78% and 84%, respectively) (Clary and Wolstenholme 1987). The larger of these regions contain inferred stem loops which Clary and Wolstenholme (1987) suggest are homologous to stem loops in the vertebrate light-strand origin of replication.

Invertebrate mitochondrial genomes remain poorly studied relative to vertebrate genomes. Among insects, the only published control-region sequences are those of *D. yakuba* and *D. virilis* (Clary and Wolstenholme 1987). We report here, for the first time, several Lepidopteran control-region sequences, which make possible some useful structural comparisons among the known control regions. Sequences were determined for a number of closely related species in the Australian butterfly genus *Jalmenus* (Lepidoptera: superfamily Papilionoidea: family Lycaenidae: subfamily Theclinae: tribe Zeziini), a New World species *Strymon melinus* from the same subfamily (Theclinae: Eumaeini; classification follows Eliot 1973), and an Australian moth, *Helicoverpa punctigera* (Lepidoptera: superfamily Noctuoidea: family Noctuidae; McKechnie et al. 1993). There are 10 described species of *Jalmenus* (Common and Waterhouse 1987, pp. 324–332; D. Yeates, personal communication). Four species—*J. daemeli, J. icilius, J. ictinus,* and *J. evagoras* (two subspecies)—are distributed widely over the eastern half of Australia. The remaining species are either rare or occur only in Western Australia. A rapid radiation of the lycaenid butterflies is thought to have resulted from their associations with ants (Pierce 1984). The highly variable mitochondrial control region should be an ideal source of DNA sequence characters with which to reconstruct the phylogeny of such closely related species.

In this paper, we describe these Lepidopteran mitochondrial control regions, compare them with the published *Drosophila* sequences, and test the proposal that there is a stem-loop motif conserved among animal control regions (Clary and Wol-

stenholme 1987). We also look for any other sequences that may resemble control-region functional motifs that have been proposed (Cherry and Blackburn 1985; Morin and Cech 1986; Okimoto et al. 1990; DeHaas et al. 1991; Hoelzel et al. 1991). We report on sequences adjacent to the control region. Expected tRNA order is transposed in Lepidoptera, and part of the small subunit (12S) rRNA region was not as conserved as was expected from available structural models.

Material and Methods

Collection of Species and gDNA

Larvae and pupae were collected from several populations each of *Jalmenus evagoras evagoras, J. evagoras eubulus, J. daemeli, J. icilius, J. ictinus,* and *J. pseudictinus* and *J. lithochroa,* ranging all the way from Townsville in the northeastern tropics of Australia to Adelaide in the semiarid south-central coastal regions (sites are detailed in the legend of fig. 3). Dried adults of a North American lycaenid, *Strymon melinus,* and live adults of a moth, *Helicoverpa punctigera,* were also used. Larvae were reared to adulthood for positive species identification.

To prepare genomic DNA, whole bodies were homogenized in a sodium dodecyl sulfate (SDS) buffer [100 mM Tris-HCl pH 8.0, 100 mM NaCl, 2 mM ethylenediaminetetraacetate (EDTA), 1% SDS]. The homogenate was extracted both with redistilled phenol equilibrated with TE (10 mM Tris, 0.1 mM EDTA, pH 8.0) and with chloroform. Finally, nucleic acids were precipitated with potassium acetate and ethanol, washed, and resuspended in 100 μl of TE. For dried specimens, the homogenate was incubated with proteinase K at 37°C overnight, prior to extraction.

Primer Design, Polymerase Chain Reaction (PCR), and Sequencing

PCR primers were made for conserved sequences in two genes on either side of the control region in the *Drosophila yakuba* mitochondrial genome. These primers are (*a*) *D. yakuba* mitochondrial sequences 14592–to 14611 (primer 12S 332+; 5′ TAGGG TATCT AATCC TAGTT) in the 12S rRNA gene and (*b*) the complement of mitochondrial sequences 212–193 (primer Met 20−; −5′ TGGGG TATGA ACCCA GTAGC) in the methionine (Met) transfer RNA gene (Clary and Wolstenholme 1985). The 12S 332+ primer differs by only three bases, and Met 20− differs by only four bases, from the homologous human sequences (Anderson et al. 1981).

PCR reactions contained 0.1–1 μg of gDNA, 100 ng of one primer, 100 ng of the other kinased primer, 0.2 mM of each deoxynucleotide, 10 mM Tris pH 8.3, 50 mM KCl, 2 mM $MgCl_2$, 0.001% gelatin, and 2 units of *Taq* polymerase (Bethseda Research Labs) in a total 100-μl volume. Reactions consisted of 35 cycles, each of 50 s at 94°C, 2 min at 50°C, and 1.5 min at 72°C. The strand incorporating the kinased primer was digested with 10 units of λ-exonuclease for 45 min at 37°C. This reaction was extracted twice with equal volumes of TE-equilibrated, phenol:chloroform:isoamyl alcohol (25:24:1) and once with chloroform. Single-stranded DNA was separated from unused primers and nucleotides by adding 0.5 volume of 7.5 M ammonium acetate and 1 volume of cold ethanol, incubated at room temperature for 5 min, and centrifuged at 12,000 *g* for 15 min. These resulting single-stranded PCR products were sequenced by using the Sequenase™ kit (United States Biochemical) and the appropriate primer.

Sequence Analysis

Sequences were analyzed by using the DNA analysis programs of the University of Wisconsin Genetics Computer Group (GCG). Control-region sequences were aligned to comparison sequences by using the program Bestfit. A crude test of significance of an alignment was made by (1) repeatedly randomizing one of the sequences and (2) recalculating the quality (Q) of the best alignment found. The null hypothesis of independence between sequences was rejected if the quality of the alignment was more than two standard deviations (SD) above the mean quality Q_R of a sample of best alignments found to each of 10 randomized comparison sequences. Q is calculated as (no. of matches) − 0.9 × (no. of mismatches) − (gap weight) × (no. of gaps) − (gap length weight) × (total gap length). Gap weights and length weights were usually set to the default values of 5.0 and 0.30. Progressively smaller-length weights were used in control-region alignments to allow for the possibility of large insertion/deletion (indel) differences. Comparisons were made among sequences in both orientations, to detect possible inversions.

Results and Discussion

Lepidopteran Control Regions Differ from Known Control Regions

Short stretches of TA repeats were found in all mitochondrial control regions examined (fig. 1), but few other sequence similarities were apparent. Apart from the regions of weak similarity to *Drosophila yakuba* sequences, no significant match to any of the other published control region or telomeric conserved motifs listed in table 1 could be found in lepidopteran sequences. These results put in doubt efforts to erect general models of control-region function such as that proposed by DeHaas et al. (1991) and suggest that there may be no single model applicable to all mitochondria.

DeHaas et al. (1991) claim the presence of primary-sequence homologues of yeast/*Escherichia coli* gyrase recognition sites and transcription initiation sites and, on that basis, infer "structural" (i.e., "gene" order) homologies to yeast and mammalian control regions. However, the gyrase recognition site has only nine bases (four of which are ambiguous), and no better than eight matching bases could be found in the *Petunia* sequence. Similarly, no better than 11 of 16 bases could be matched to the transcription initiation sequence (table 1). Neither of these alignments is better than alignments to randomized sequences (data not shown). Thus the claims of DeHaas et al. for such homologies are not statistically valid.

Several regions of extensive sequence similarity could be found between *Jalmenus* and *D. yakuba* control-region sequences, in either orientation. The optimum alignment for the whole region, however, was not significant [76.4% identity, $Q = 95.3$, $Q_R = 91.2 \pm 5.5$; mean ± SD for 10 randomized sequence alignments (see Material and Methods)]. High, but nonsignificant, percentage identity is to be expected when aligning A+T-rich sequences, which underscores the need for statistical tests of significance. Only regions of significant similarity between *Jalmenus* and *Drosophila* sequences are shown in figure 1. In the control regions of *D. yakuba* and *D. virilis,* the larger of two conserved sequence blocks contains an inferred stem-loop structure which may be related to the vertebrate light-strand origin of replication (Clary and Wolstenholme 1987). The only region of primary sequence similarity between *Jalmenus* and *Drosophila* that lies within these conserved sequence blocks is that in, and adjacent to, the $(TA)_n$ repeat (Clary and Wolstenholme 1985; see fig. 1, bp 15621–15650). However,

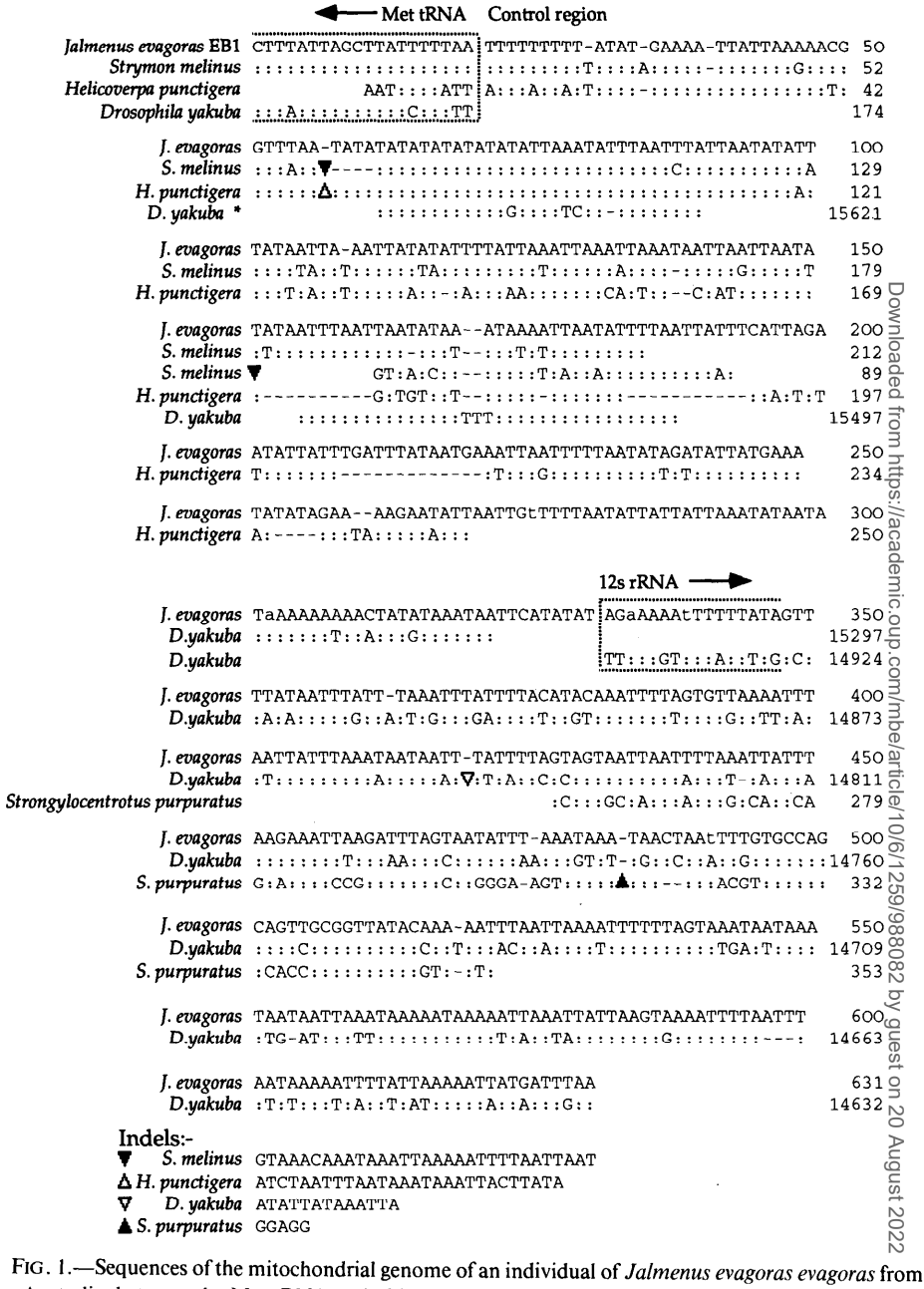

FIG. 1.—Sequences of the mitochondrial genome of an individual of *Jalmenus evagoras evagoras* from Ebor, Australia, between the Met tRNA and 12S rRNA genes (GenBank L16849), compared with similar sequences for *Strymon melinus* from North America (GenBank L16850) *Helicoverpa punctigera* from Australia (GenBank L17343) (McKechnie et al. 1993), *Drosophila yakuba* (Clary and Wolstenholme 1985), and sea urchin *Strongylocentrotus purpuratus* (Jacobs et al. 1988). Variation among, and within, species of *Jalmenus* is detailed further below (fig. 3). A colon (:) indicates identity with *J. evagoras;* and a dash (-) indicates a deduced indel. Lowercase letters in the *J. evagoras* sequence indicate unresolved sequencing ambiguity as to the number of repeated bases at that position. Indels larger than 2 bp relative to the *Jalmenus* sequence are shown at the bottom of the fig. A significant similarity between the *Strymon melinus* indel and a transposed region of the *J. evagoras* sequence is also shown. An asterisk (*) indicates a similar *D. yakuba* sequence which lies within the 276-bp conserved block but outside the conserved stem loop described by Clary and Wolstenholme (1987). The sequence orientation is reverse to that established for *D. yakuba* by Clary and Wolstenholme (1985).

Table 1
Sequences to Which Lepidopteran Control-Region Sequences Were Compared for Primary-Sequence Similarity

Organism	Sequence Description
Saccaromyces cerevisiae[a]	Autonomous replicating site (ARS) consensus (5'WTTTATRTTTW); conserved box 3' to ARS consensus (5'cTtTTAGCWWW); gyrase recognition-site consensus (5'TRTGYTYTR); and transcription-initiation consensus (5'TATTACTTATATATTT)
Tetrahymena thermophila	Mitochondrial telomeric repeat 53 bp (GenBank OR:TETMTTRA)[b] and nuclear telomeric repeat 470 bp (GenBank OR: TETRSTELC)[c]
Caenorhabditis elegans[d]	Mitochondrial control region (CR) 466 bp (GenBank OR: CELMTCE)
Drosophila yakuba[e]	CR 1,078 bp (GenBank OR:DROMTCG)
D. yakuba and *D. virilis*[f]	CR, conserved 49-bp and 276-bp boxes (GenBank OR:DROMTCG and DROMTDVTRN)
Mammals[g]	CR, conserved motif in deduced translation of possible open reading frame (LfS1RAH)

[a] DeHaas et al. (1991).
[b] Morin and Cech (1986).
[c] Cherry and Blackburn (1985).
[d] Okimoto et al. (1990).
[e] Clary and Wolstenholme (1985).
[f] Clary and Wolstenholme (1987).
[g] Hoelzel et al. (1991).

no primary sequence similarities to the region containing the putative stem-loop structure were found in lepidopteran control regions.

Lepidopteran control regions were most similar around the $(TA)_n$ dinucleotide repeat (fig. 1). In the latter half of the control region, however, even the *Strymon melinus* control region had very little sequence similarity, while the *Helicoverpa punctigera* control region could not be aligned with any confidence (fig. 1). Nonetheless, all lepidopteran sequences were ~350 bp, when judged by the uniformity of sizes of the PCR-amplified fragments. The length variation in the $(TA)_n$ repeat region is most likely the result of the misalignments which are characteristic of origins of replication, as in the control region in *D. yakuba* or in telomeric sequences.

Primary sequence similarity may be insignificant even when secondary structures are conserved. To identify possible stem-loop structures similar to those in *Drosophila* species, a window of 80 bp was moved, in 20-bp steps, over a sample of lepidopteran control regions (*J. evagoras, J. daemeli, S. melinus,* and *H. punctigera*). Sequences in these windows were solved for minimum-energy secondary structure by using the GCG-Fold routine. To test for departure from random expectation, structural solutions were also found for 20 80-bp samples, without replacement, from sets of randomized sequences. The folding energies for this sample were between −3.7 and −10.3.

Several structures with folding energies outside the range of the random sample could be found in lepidopteran control regions (fig. 2). One structure lies within the TA repeats of lycaenid sequences. Another region 3' to this region can form strong stem loops in *J. evagoras* and *J. daemeli,* but not in *S. melinus.* Two other, weaker stem-loop possibilities occur only in the *J. evagoras* sequence (fig. 2C and D). Despite the apparent conservation of the TA repeat region in lepidoptera, no such high-energy

FIG. 2.—Stem-loop structures with significantly greater-than-random folding energies (ΔG), in four domains of lepidopteran control regions. A, bp 29–98 of the *Jalmenus evagorus* Eb1 sequence (fig. 1). B, bp 103–163 of the *J. evagorus* Eb1 sequence (fig. 1). C, bp 161–221 of the *J. evagorus* Eb1 sequence (fig. 1). D, bp 250–331 of the *J. evagorus* Eb1 sequence (fig. 1). Also shown are stem loops in *J. daemeli* JC1, *Strymon melinus*, and *Helicoverpa punctigera* control regions which had significant folding energies. Downward-pointing arrows indicate bases that are aligned in the primary sequence comparisons of fig. 1 between *J. evagoras* and *J. daemeli* or either *S. melinus* in domain A or *H. punctigera* in domain B. Energies and optimum structures were calculated by using the GCG-Fold program.

stem loop was found in this region for the *H. punctigera* sequence. One stem loop with a marginally significant folding energy was found in a region of very low primary sequence similarity for *H. punctigera* (fig. 2B).

Length heteroplasmy was not apparent for any of the lepidopteran control regions examined. It could be that PCR, which may favor amplification of shorter templates,

is inappropriate for discovering heteroplasmy. A generous time (1.5 min) was provided in PCR reactions for polymerization steps. Unless heteroplasmic variants were drastically different from those we found, we can conclude that heteroplasmy is rare in the short and simple control regions of lepidoptera. This negative evidence is consistent with the idea that heteroplasmy of mitochondria is a result of both the accumulation of long repeats in the control region and the consequent occurrence of frequent mismatches during replication (Boyce et al. 1989; Buroker et al. 1990; Hayasaka et al. 1991).

Jalmenus Control Regions Are Highly Conserved

The sequences available for inter- and intraspecific comparisons cover almost the entire control region (figs. 1 and 3). Not including the highly variable TA repeat region, there were 15 nucleotide sites that varied among *Jalmenus* species over the ~250 bp of comparable sequence, many of which were also polymorphic within species (fig. 3; 13 substitutions and two single-base indels). This amounts to 94% sequence identity among *Jalmenus* species. *S. melinus* and *Jalmenus* species are in the same subfamily (Eliot 1973). There were a total of 30 variable sites in the 159 bp of comparable control-region sequence, amounting to 81% sequence identity between *S. melinus* and *Jalmenus* sequences (fig. 3; 21 substitutions and nine indels, not including the TA repeat).

This result is similar to that of Meyer et al. (1990), who found that 14 species in nine genera of endemic, Lake Victoria, cichlid fish differed at only 15 of 350 bases (96% identity) in their mitochondrial control regions. Nevertheless, within the family Osteicthyes, sturgeon and cichlid control-region sequences have diverged to the extent that there is no significant similarity of primary sequence (Buroker et al. 1990; Meyer et al. 1990). The slight divergence among Lake Victoria cichlids, relative to the considerable divergence between cichlids and sturgeon, is thought to result from recent ancestry rather than selection, Lake Victoria being only ~1 Myr (Meyer et al. 1990).

The comparatively slight divergence among control-region sequences of *Jalmenus* species relative to divergence from the confamilial *S. melinus* suggests either that, like the Lake Victoria cichlids, these species have radiated comparatively recently or that some unknown selective pressure has been acting to slow divergence of the sequences. The latter possibility is unlikely in view of the absence of strongly conserved primary sequence motifs in lepidopteran control regions.

Like most other lycaenid butterflies, *Jalmenus* species associate obligately with various species of ants. Amino acid–rich secretions are supplied by the larvae to the ant workers that guard them from parasites and predators (Pierce et al. 1987). Consequent restrictions on distribution, together with the potential for shifts to novel species of ant associates, have been hypothesized to favor the frequent formation of population isolates and, thus, enhanced speciation rates within the lycaenids (Pierce 1984; Pierce and Elgar 1985). The foregoing results lend support to that hypothesis.

Phylogenetic reconstructions of *Jalmenus* revealed only one synapomorphy. The A at bp 34 (fig. 3) unites all species to the exclusion of the two subspecies of *J. evagoras* (both of which share a G at this position, with the two outgroup taxa; fig. 1). The lack of sufficient divergence among the sequences available, the extent of intraspecific polymorphism, and the difficulty of inferring the evolutionary relationships among the many variants of the TA repeat provide no further synapomorphies with which to resolve the relationships among *Jalmenus* species.

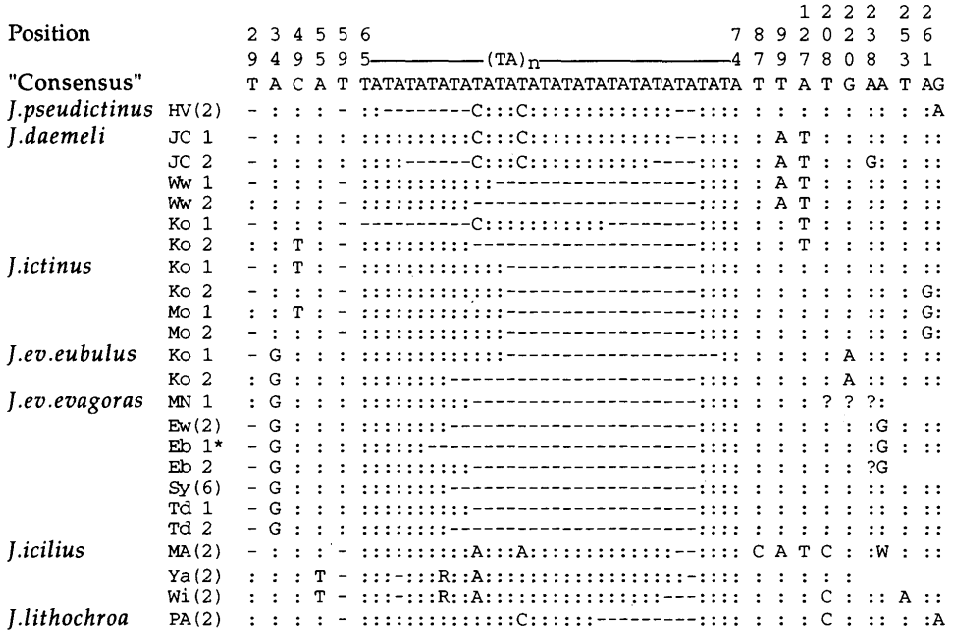

FIG. 3.—All control-region haplotypes found in several populations of six species of *Jalmenus* in Australia. Numbers of individuals sequenced that had identical sequence, if more than one, are placed in parentheses after the location. An asterisk (*) indicates the Eb1 haplotype for which the entire sequence was determined (fig. 1). Symbols are as in fig. 1. Ambiguities were as follows: R = A or G; W = A or T; and a question mark (?) = indeterminate. Locations in approximate north-to-south order are as follows: HV = Hidden Valley, north Queensland; JC = James Cook University campus, Townsville; Ww = Warwick, Queensland; Ko = Kogan, Queensland; Mo = Moonie, Queensland; MN = Mt. Nebo, Queensland; Ew = Eastwood, New South Wales; Eb = Ebor, New South Wales; Sy = Sydney; Td = Tidbinbilla, Australian Capital Territory; MA = Mt. Ainslie, Australian Capital Territory; Ya = Yaroonga, South Australia; Wi = Willunga, South Australia; and PA = Port Augusta, South Australia.

tRNA Transposition in Lepidoptera

Extensive sequence comparisons failed to identify any clear lepidopteran homologues of either the glutamine (Q) or isoleucine (I)—or of any other tRNA genes between the start of the methionine (M) tRNA gene and the control region—as found in *D. yakuba* and *D. virilis*. The best alignment (89%; GCG-Bestfit) between the first 20 bases of the butterfly (*Jalmenus* and *Strymon*) sequences and the entire *D. yakuba* sequence, on either strand, was to the first 20 bases of the met tRNA (fig. 1). For Lepidoptera, relative to *Drosophila,* these two tRNA genes have been either swapped with methionine in the same cluster (MQI becoming either QIM or IQM) or transposed to some other tRNA cluster of the mitochondrial molecule.

Similar rearrangements of tRNAs within gene clusters have previously been reported for insects and vertebrates. Locust, a mosquito, and honeybees all differ from *Drosophila* in ordering of various tRNA genes (HsuChen et al. 1984; Haucke and Gellissen 1988; Crozier et al. 1989). The tRNA order around the light-strand replication origin of vertebrates is transposed in marsupials relative to placental mammals (Pääbo et al. 1991). Altogether, this evidence supports the view that, while major gene order evolves quite slowly, tRNA order, like control-region organization, is quite labile. There has been no evidence to date of reorganization of whole gene clusters (i.e., variation in the ordering or size of the clusters, as opposed to that of genes within clusters) within either insects or vertebrates, although such rearrangements have occurred between vertebrates and insects (Clary and Wolstenholme 1985).

The 12S rRNA Gene in *Jalmenus*

The 5′ region of the 12S rRNA gene that has been sequenced for *Jalmenus* had significant similarity to the equivalent region in *D. yakuba* (71.6%, $Q = 185$, $Q_R = 142.3 \pm 4.0$; fig. 1). There was no significant similarity of *D. yakuba* or *Jalmenus* sequences to either sea urchin, *Strongylocentrotus purpuratus,* or nematode, *Caenorhabditis elegans,* control regions (Jacobs et al. 1988; Okimoto et al. 1990). No significant match could be found for nematode, but a small region of sea urchin 12S sequence (Jacobs et al. 1988; bp 256–353) had significant similarity to insect sequences (fig. 1). In this region *D. yakuba* and sea urchin sequences were 66.7% identical ($Q = 21.1$, $Q_R = 11.7 \pm 1.7$), while *J. evagoras* and sea urchin sequences were 63.4% identical ($Q = 17.7$, $Q_R = 9.7 \pm 1.8$). Sequences of *J. evagoras* and *D. yakuba* were 78.6% identical. Percentage divergences of *J. evagoras* and *D. yakuba* from the sea urchin 12S sequences were not significantly different in a χ^2 test of independence ($P < 0.1$). This result indicates no difference in 12s rRNA evolutionary rates between *Drosophila* and *Jalmenus*. In contrast, the rate of divergence of cytochrome oxidase genes away from a common ancestor has been greater for the honeybee than for *Drosophila* (Crozier et al. 1989).

A structure for 12S rRNA has been proposed for *D. yakuba* and *D. virilis* (Clary and Wolstenholme 1987), on the basis of the vertebrate structure proposed by Zweib et al. (1981). The available *J. evagoras* sequence covered only "section one" of this putative structure (fig. 4), and the fit to this vertebrate model was poorer than the fit to the *Drosophila* sequences. Also, the deduced secondary structures for this region of the 12S rRNA gene (not shown here) derived by using the GCG-Fold algorithm were quite different from the hypothesized vertebrate structure, for both *J. evagoras* and *Drosophila* sequences. Deductions about secondary structures from sequence data are limited to the extent that long-range interactions in the tertiary structure are un-

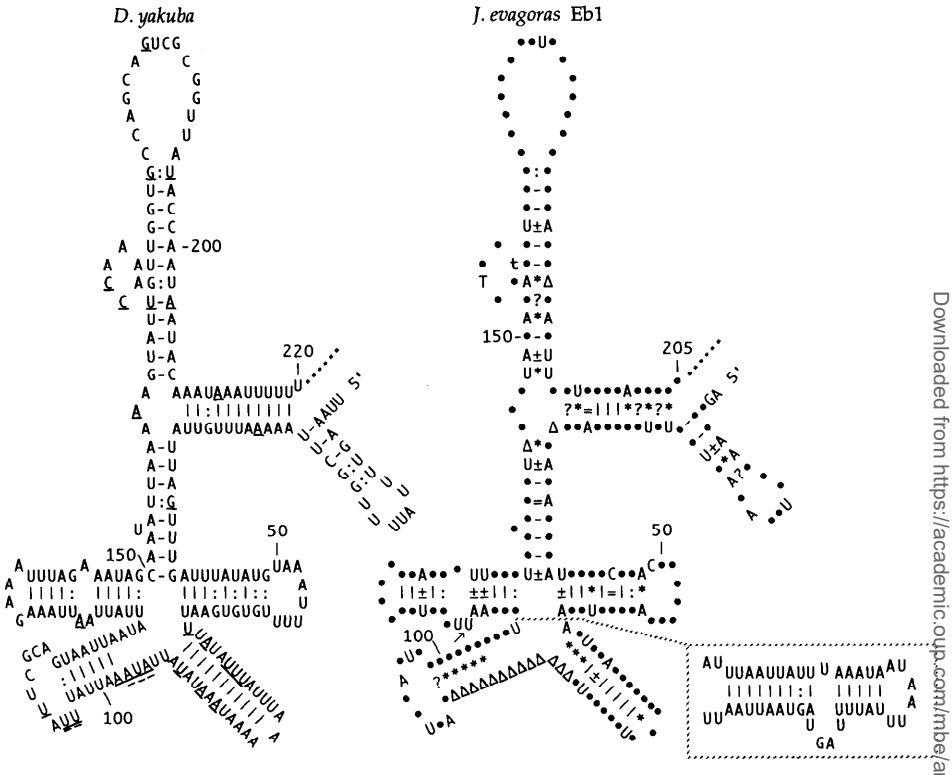

FIG. 4.—Secondary structure of "section one" of the *Drosophila yakuba* 12S rRNA, as proposed by Clary and Wolstenholme (1987), and the *Jalmenus evagoras evagoras* Eb1 sequence (fig. 1) fitted to the same model. Bases that differ from *D. yakuba* in the *D. virilis* sequence are underlined and, where deletions occur, are double-underlined. G-U pairings are denoted by colons. For the *J. evagoras* sequence, a bullet (●) denotes identity with *D. yakuba*, a delta (Δ) denotes deletion, an arrow (→) denotes an insertion, an asterisk (*) denotes a pair-bond lost, an equals sign (=) denotes a change from G-U to either A-U or G-C pairbond, a plus-or-minus sign (±) denotes a double substitution that preserves the pair-bond, and a question mark (?) denotes a base pairing that is flanked by unpaired bases and so is unstable.

known and can be determined only by nuclease analysis of ribosomes, which is beyond the scope of the present study.

There were 45 pairings for *J. evagoras* 12S sequence when fitted to the model structure (fig. 4). Another six base pairings were technically preserved but were also flanked by unpaired bases and could not be included. This compares poorly with the 72 pairings for *D. yakuba* and the 75 pairings for *D. virilis* (fig. 4). Most of the difference derives from one large deletion that removed the sixth stem from the equivalent *J. evagoras* structure (fig. 4). The best structure found for the region 66–108 bp from GCG-Fold is quite different from the model structure (see inset, fig. 4) and brought the *J. evagoras* total to 52 base pairings. Of the 45 base pairings in the exact fit to the model, 13 resulted from substitutional differences that reconstituted *Drosophila* base pairing at the same position or that turned G-U pairs in *Drosophila* into stronger C-G or U-A pairs in *Jalmenus*. Such changes have been regarded as examples of correlated evolution under stem-pairing constraints, despite a lack of appropriate statistical tests (Dunon-Bluteau and Brun 1986; Clary and Wolstenholme 1987).

Tests of goodness of fit to structural models are unavailable, as the statistical

basis is not well understood. A weak, if laborious, test of goodness of fit was achieved by repeatedly randomizing the positions of substitutional differences (indel differences were not included) between sequences while preserving the number of substitutions of various types but also without regard to nearest-neighbor frequencies. Twenty-five such randomized sequences were aligned to the structural model, and the numbers of new base pairings generated or remaining unchanged were counted, to provide a crude sampling distribution for a null hypothesis of "no significance" of sequence order to the fit of the structural model.

The total number of base pairings in the random sample was 25–43. The number of "new" base pairings different from those for *D. yakuba,* but which preserved pairings in the *D. yakuba* model at the same positions, was 1–11 in the sample of randomized alignments. Both the observed total of 45 base pairings in the alignment to the section-one structure and the 13 new base pairings were just outside the respective ranges calculated for the sample of 25 random alignments (fig. 4). We conclude that the vertebrate/*Drosophila* model for the structure of section one of the 12S rRNA fits only weakly to the observed *J. evagoras* sequence and that evidence for compensatory mutations under stem-forming constraints was also slight.

Acknowledgments

CSIRO Australia Division of Tropical Animal Science, Long Pocket Laboratories, Brisbane, very generously provided laboratory space and equipment for extraction of DNA. Don Sands, John Kerr, Ebbe Neilsen, Ted Edwards, Murdoch DeBaar, Mark Elgar, and Graham Wood advised on collection localities. We are grateful to R. H. Fisher, Diane Wagner, and Mark Smith for supplying specimens. Financial assistance (to S.W.M.) from the Australian Research Council and the Australian Cotton Research and Development Corporation is gratefully acknowledged. Thanks are also due to John Baverman and The Australian National Insect Collection.

LITERATURE CITED

ANDERSON, S., A. T. BANKIER, B. G. BARRELL, M. H. L. DEBRUIJN, A. R. COULSON, J. DROUIN, I. C. EPERON, D. P. NIERLICH, B. A. ROE, F. SANGER, P. H. SCHREIER, A. J. H. SMITH, R. STADEN, and I. G. YOUNG. 1981. Sequence and organisation of the human mitochondrial genome. Nature **290**:457–465.

BOYCE, T. M., M. E. ZWICK, and C. F. AQUADRO. 1989. Mitochondrial DNA in the bark weevils: size, structure and heteroplasmy. Genetics **123**:825–836.

BUROKER, N. E., J. R. BROWN, T. A. GILBERT, P. J. O'HARA, A. T. BECKENBACH, W. K. THOMAS, and M. J. SMITH. 1990. Length heteroplasmy of sturgeon mitochondrial DNA: an illegitimate elongation model. Genetics **124**:157–163.

CHERRY, J. M., and E. H. BLACKBURN. 1985. The internally located telomeric sequences in the germ-line chromosomes of *Tetrahymena* are at the ends of transposon-like elements. Cell **43**:747–758.

CLARY, D. O., and D. R. WOLSTENHOLME. 1985. The mitochondrial DNA molecule of *Drosophila yakuba:* nucleotide sequence, gene organisation, and genetic code. J. Mol. Evol. **22**: 252–271.

———. 1987. *Drosophila* mitochondrial DNA: conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J. Mol. Evol. **25**:116–125.

COMMON, I. F. B., and D. F. WATERHOUSE. 1981. Butterflies of Australia, 2d ed. Angus & Robertson, Sydney.

CROZIER, R. H., Y. C. CROZIER, and A. G. MACKINLAY. 1989. The CO-I and CO-II region of

honeybee mitochondrial DNA: evidence for variation in insect mitochondrial evolutionary rates. Mol. Biol. Evol. **6**:399–411.

DEHAAS, J. M., J. HILLE, F. KORS, B. VAN DER MEER, A. J. KOOL, O. FOLKERTS, and H. J. J. NIJKAMP. 1991. Two potential *Petunia hybrida* mitochondrial DNA replication origins show structural and in vitro functional homology with the animal heavy and light strand replication origins. Curr. Genet. **20**:503–513.

DUNON-BLUTEAU, D., and G. BRUN. 1986. The secondary structures of the *Xenopus laevis* and human mitochondrial small ribosomal subunit RNA are similar. FEBS Lett. **198**:333–338.

ELIOT, J. N. 1973. The higher classification of the Lycaenidae (Lepidoptera): a tentative arrangement. Bull. Br. Museum Nat. Hist. (Entomol.) **28**:373–506.

HALE, L. R., and R. S. SINGH. 1986. Extensive variation and heteroplasmy in size of mitochondrial DNA among geographic populations of *D. melanogaster.* Proc Natl. Acad. Sci. USA **83**:8813–8817.

HAUCKE, H., and G. GELLISSEN. 1988. Different mitochondrial gene orders among insects: exchanged tRNA gene positions in the COII/COIII region between an orthopteran and a dipteran species. Curr. Genet. **14**:471–476.

HAYASAKA, K., T. ISHIDA, and S. HORAI. 1991. Heteroplasmy and polymorphism in the major noncoding region of mitochondrial DNA in Japanese monkeys: association with tandemly repeated sequences. Mol. Biol. Evol. **8**:399–415.

HOELZEL, R. A., J. M. HANCOCK, and G. A. DOVER. 1991. Evolution of the cetacean mitochondrial D-loop region. Mol. Biol. Evol. **8**:475–493.

HSUCHEN, C. C., R. M. KOTIN, and D. T. DUBIN. 1984. Sequence of the coding and flanking regions of the large ribosomal subunit RNA gene of mosquito mitochondria. Nucleic Acids Res. **12**:7771–7785.

JACOBS, H. T., D. J. ELLIOTT, V. B. MATH, and A. FARQUHARSON. 1988. Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. Mol. Biol. **202**:185–217.

MCCRACKEN, A., I. UHLENBUSCH, and G. GELLISSEN. 1987. Structure of the cloned *Locusta migratoria* mitochondrial genome: restriction mapping and sequence of its ND-1(URF-1) gene. Curr. Genet. **11**:625–630.

MCKECHNIE, S. W., A. A. HOFFMANN, I. V. KOVACS, Z. CACOYIANNI, N. E. NAUGHTON, and S. KATSABANAS. 1993. Genetic variation among Australian populations of native budworm, *Helicoverpa punctigera* (Lepidoptera:Noctuidae). Pp. 428–431 *in* S. A. COREY, D. J. DALL, and W. M. MILNE, eds. Pest control and sustainable agriculture. C.S.I.R.O., Melbourne.

MEYER, A., T. D. KOCHER, P. BASASIBWAKI, and A. C. WILSON. 1990. Monophyletic origin of Lake Victoria cichlid fishes suggested by mitochondrial DNA sequences. Nature **347**:550–553.

MORIN, G. B., and T. R. CECH. 1986. The telomeres of the linear mitochondrial DNA of *Tetrahymena thermophila* consist of 53 bp tandem repeats. Cell **46**:873–883.

OKIMOTO, R., J. L. MACFARLANE, and D. R. WOLSTENHOLME. 1990. Evidence for the frequent use of TTG as the translation initiation codon of mitochondrial protein genes in the nematodes, *Ascaris suum* and *Caenorhabditis elegans.* Nucleic Acids Res. **18**:6113–6118.

PÄÄBO, S., W. K. THOMAS, K. M. WHITFIELD, Y. KUMAZAWA, and A. C. WILSON. 1991. Rearrangements of mitochondrial transfer RNA genes in marsupials. J. Mol. Evol. **33**:426–430.

PIERCE, N. E. 1984. Amplified species diversity: a case study of an Australian lycaenid butterfly and its attendant ants. Pp. 197–200 *in* R. I. VANE-WRIGHT and P. R. ACKERY, eds. Biology of butterflies. Academic Press, London.

PIERCE, N. E., and M. A. ELGAR. 1985. The influence of ants on host plant selection by *Jalmenus evagoras,* a myrmecophilous lycaenid butterfly. Behav. Ecol. Sociobiol. **16**:209–222.

PIERCE, N. E., R. L. KITCHING, R. C. BUCKLEY, M. F. J. TAYLOR, and K. BENBOW. 1987. The costs and benefits of cooperation for the Australian lycaenid butterfly, *Jalmenus evagoras* and its attendant ants. Behav. Ecol. Sociobiol. **21**:237–248.

SACCONE, C., G. PESOLE, and E. SBISÁ. 1991. The main regulatory region of mammalian

mitochondrial DNA: structure-function model and evolutionary pattern. J. Mol. Evol. **33:** 83–91.

ZWIEB, C., C. GLOTZ, and R. BRIMACOMBE. 1981. Secondary structure comparisons between small subunit ribosomal RNA molecules from six different species. Nucleic Acids Res. **9:** 3621–3640.

PAUL SHARP, reviewing editor

Received October 1, 1992; revision received May 26, 1993

Accepted May 26, 1993