

# Numerous Gene Rearrangements in the Mitochondrial Genome of the Wallaby Louse, *Heterodoxus macropus* (Phthiraptera)

Renfu Shao, Nick J. H. Campbell, and Stephen C. Barker

Department of Microbiology and Parasitology, and Institute for Molecular Bioscience, University of Queensland, Brisbane, Queensland, Australia

The complete arrangement of genes in the mitochondrial (mt) genome is known for 12 species of insects, and part of the gene arrangement in the mt genome is known for over 300 other species of insects. The arrangement of genes in the mt genome is very conserved in insects studied, since all of the protein-coding and rRNA genes and most of the tRNA genes are arranged in the same way. We sequenced the entire mt genome of the wallaby louse, *Heterodoxus macropus*, which is 14,670 bp long and has the 37 genes typical of animals and some noncoding regions. The largest noncoding region is 73 bp long (93% A+T), and the second largest is 47 bp long (92% A+T). Both of these noncoding regions seem to be able to form stem-loop structures. The arrangement of genes in the mt genome of this louse is unlike that of any other animal studied. All tRNA genes have moved and/or inverted relative to the ancestral gene arrangement of insects, which is present in the fruit fly *Drosophila yakuba*. At least nine protein-coding genes (*atp6*, *atp8*, *cox2*, *cob*, *nad1-nad3*, *nad5*, and *nad6*) have moved; moreover, four of these genes (*atp6*, *atp8*, *nad1*, and *nad3*) have inverted. The large number of gene rearrangements in the mt genome of *H. macropus* is unprecedented for an arthropod.

## Introduction

Mitochondrial (mt) genomes of most animals are circular DNA molecules about 16 kb in size and have one major noncoding region and 37 genes: 13 protein-coding genes (ATP synthase subunits 6 and 8 [*atp6*, *atp8*], cytochrome oxidase subunits 1–3 [*cox1-cox3*], cytochrome *b* [*cob*], and NADH dehydrogenase subunits 1–6 and 4L [*nad1-nad6*, *nad4L*]), 2 rRNA genes (large- and small-ribosomal-subunit RNAs [*rrnL*, *rrnS*]), and 22 tRNA genes (one for each amino acid except for leucine and serine, which have two genes: *trnL*<sub>1</sub> [anticodon uag], *trnL*<sub>2</sub> [uaa], *trnS*<sub>1</sub> [gcu], and *trnS*<sub>2</sub> [uga]) (Wolstenholme 1992; the gene nomenclature is that of J. Boore, available on <http://www.biology.lsa.umich.edu/~jboore>).

Complete mt gene arrangements of over 100 animal species have been determined, and partial arrangements are known for several hundred species (Boore 1999; see "Mitochondrial Gene Arrangement Source Guide," version 5.0, on Boore's web page). The arrangement of genes in the mt genome, particularly the arrangement of the large genes that encode proteins and rRNAs, is generally conserved in an animal phylum but varies substantially among phyla (Boore et al. 1995; Boore and Brown 1998). However, some exceptions have been found so far: (1) the ascidian *Halocynthia roretxi* (Chordata, Urochordata) has a very different gene arrangement from those of other chordates (Yokobori et al. 1999); (2) two brachiopods, *Terebratulina retusa* and *Laqueus rubellus*, have very different gene arrangements from each other (Stechmann and Schlegel 1999; Noguchi et al. 2000); (3) species from four different classes of the Mollusca are very different in gene

arrangement (Boore 1999); and (4) species from the same class (Secernentea) of the Nematoda are very different in gene arrangement (Boore 1999).

The gene content of mt genomes may also vary, since (1) a homolog of the bacterial *mutS* gene, which is involved in mismatch repair, was found in the mt genome of the soft coral *Sarcophyton glaucum* (Pont-Kingdon et al. 1998); (2) an extra *trnM*, but no *atp8*, was found in the mollusk *Mytilus edulis* (Hoffmann, Boore, and Brown 1992); (3) *atp8* is absent in some flatworms (Le et al. 2000; Fukunaga, unpublished but available in GenBank, accession number AB018440), nematodes (Okimoto et al. 1991, 1992; Keddie, Higazi, and Unnasch 1998), and the ascidian *H. roretxi* (Chordata, Urochordata; Yokobori et al. 1999); and (4) some cnidarians lack all but one or two tRNA genes (Beagley, Okimoto, and Wolstenholme 1998; Beaten, Roger, and Cavalier-Smith 1998).

More is known about the arrangement of mt genes for the class Insecta than for any other invertebrates. The mt genomes of 12 insects have been completely sequenced. These include (1) seven flies (Diptera): *Drosophila yakuba* (Clary and Wolstenholme 1985), *Drosophila melanogaster* (Lewis, Farr, and Kaguni 1995), *Drosophila mauritiana*, *Drosophila sechellia*, *Drosophila simulans* (last three in Ballard 2000), *Ceratitis capitata* (Spanos et al. 2000), and *Cochliomyia hominivorax* (Lessinger et al. 2000); (2) two mosquitoes (Diptera): *Anopheles quadrimaculatus* (Mitchell, Cockburn, and Seawright 1993) and *Anopheles gambiae* (Beard, Hamm, and Collins 1993); (3) the domestic silkworm (Lepidoptera) *Bombyx mori* (Lee et al., unpublished but available in GenBank, accession number AF149768); (4) the honeybee (Hymenoptera) *Apis mellifera* (Crozier and Crozier 1993); and (5) the locust (Orthoptera) *Locusta migratoria* (Flook, Rowell, and Gellissen 1995). Moreover, partial gene arrangements of over 300 species of insects are known. Full or partial mt gene arrangements are thus available from 10 orders of insects, but the vast majority of the species studied are from three

Key words: *Heterodoxus macropus*, insect, hemipteroid assemblage, Paraneoptera, mtDNA, gene order.

Address for correspondence and reprints: Renfu Shao, Department of Microbiology and Parasitology, University of Queensland, Brisbane, Queensland 4072, Australia. E-mail: r.shao@mailbox.uq.edu.au.

Mol. Biol. Evol. 18(5):858–865. 2001

© 2001 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

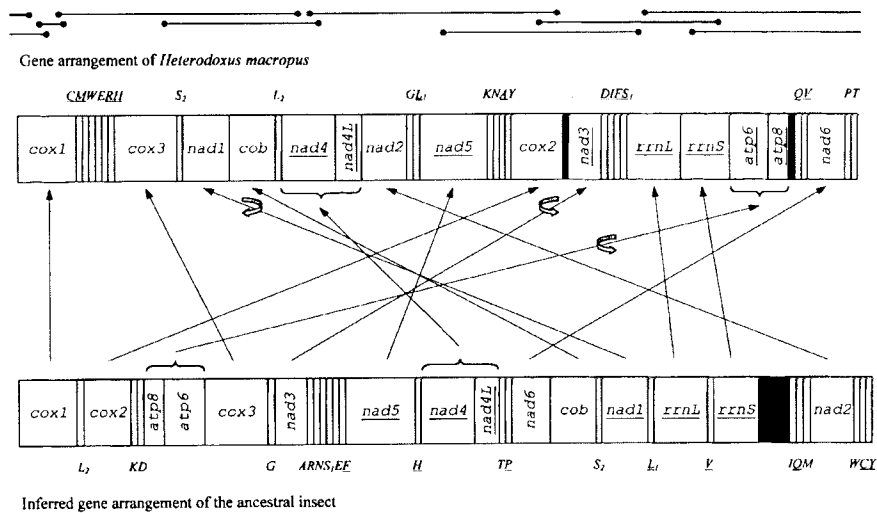


FIG. 1.—Linearized arrangements of genes in the mt genomes of *Heterodoxus macropus* and the ancestral insect. Lines above the mt gene arrangement of *H. macropus* show the extent of the eight overlapping PCR fragments from which this genome was sequenced. Genes are transcribed from left to right except those with underlined names, which are transcribed from right to left. tRNA genes are labeled with single-letter amino acid abbreviations except for those encoding leucine and serine, which are labeled as  $L_1$  (anticodon uag),  $L_2$  (uaa),  $S_1$  (gcu), and  $S_2$  (uga). Abbreviations of protein-coding genes and rRNA genes: *cox1*–*cox3*, cytochrome oxidase subunits 1–3; *cob*, cytochrome *b*; *nad1*–*nad6*, NADH dehydrogenase subunits 1–6; *nad4L*, NADH dehydrogenase subunit 4L; *atp6* and *atp8*, ATP synthase subunits 6 and 8; *rrnL* and *rrnS*, large and small ribosomal subunit RNA. Long arrows are drawn from the location of each protein-coding gene and rRNA gene in the ancestral arrangement to their location in *H. macropus*. The circling arrows indicate inversions. Black regions indicate the largest two noncoding regions of *H. macropus* and the control region of the ancestral insect.

orders: the Diptera, the Lepidoptera, and the Hymenoptera, which are all Endopterygota (flies, fleas, beetles, moths, wasps, etc.).

All the insects studied so far have the same arrangement of protein-coding genes, rRNA genes, and most tRNA genes. Only the positions of a few tRNA genes vary, particularly the positions of those near the control region and in the tRNA gene cluster *trnA-trnR-trnN-trnS<sub>1</sub>-trnE-trnF* (genes underlined are encoded by the minority strand). The most common arrangement of mt genes of insects studied, which is present in *D. yakuba* and many other species, is inferred to be ancestral for insects (Boore, Lavrov, and Brown 1998). In contrast to the Endopterygota, little is known about the arrangements of genes in many other lineages of insects, such as the Apterygota, the Palaeoptera, the hemipteroid assemblage, etc. The hemipteroid assemblage (bugs, thrips, psocopterans, and lice) is particularly interesting because it is thought to be the sister group of the Endopterygota (Kristensen 1991). *Oncopeltus fasciatus*, a milkweed bug (Hemiptera), is the only species so far from the hemipteroid assemblage for which we have any information on the arrangement of mt genes. This species has the arrangement of *trnL<sub>2</sub>-?-cox2-?-trnK*, which is the same as the ancestral arrangement of insects (Liu and Beckenbach 1992; “?” indicates that the amplifying primers in this study annealed in the two tRNA genes, but the sequence data reported are not sufficient to verify that there are no other tRNA genes between these tRNA genes and *cox2*).

Here, we report the complete mt genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera, Amblycera, Boopidae). *Heterodoxus macropus* infests

agile wallabies, *Macropus agilis*, throughout almost all of the geographic range of this marsupial in Australia (unpublished data). Relative to the ancestral gene arrangement of insects, numerous rearrangements of protein-coding and tRNA genes have occurred in the mt genome of this louse. This is the first species from the hemipteroid assemblage for which the entire mt genome has been sequenced, and this paper represents the first report of rearrangements of protein-coding genes in the mt genome of an insect. The number of gene rearrangements required to account for the arrangement of genes in this species is unprecedented for an arthropod.

## Materials and Methods

One of us (S.C.B.) collected the wallaby louse, *H. macropus*, from an agile wallaby on South Stradbroke Island, Queensland, Australia. The louse was then snap-frozen and stored in liquid nitrogen. A CTAB DNA extraction protocol (Shahjahan et al. 1995) was used to extract DNA from one louse. The entire mt genome was amplified in eight overlapping fragments (fig. 1) by PCR with the primers (1) C1-J-1718 and C1-N-2329, (2) C1-J-2177 and N4-N-8578, (3) C3-J-5476 and N4-N-8924, (4) N4-J-8718 and C2-N-3661, (5) N5-J-7567 and LR-N-12901, (6) C2-J-3400 and SR-N-14594, (7) LR-J-12911 and C1-N-1646, and (8) SR-J-14199 and C1-N-2161 (table 1 and fig. 1). Names of the primers, which indicate the target genes, the strands, and the 3' base positions relative to the corresponding sequences in *D. yakuba*, are after Simon et al. (1994). Elongase (GIBCO: BRL) was used in all PCR reactions under the following conditions: 1 min at 94°C, followed by 35 cycles of 30

**Table 1**  
Primers Used to Amplify the Mitochondrial Genome of *Heterodoxus macropus*

Primer Name	Sequence (5' to 3')
C1-J-1718	GGAGGATTGGAATTGATTAGTTCC
C1-J-2177	TTTTATATCAACATTTATTTTGATT
C1-N-1646	AATTATTAATAATGCATGTG
C1-N-2161	CCAAAAATCAAAATAAATGTTG
C1-N-2329	ACTGTAAATATATGATGAGCTCA
C2-J-3400	ATTGGACATCAATGATATTGA
C2-N-3661	CCACAAATTTCTGAACATTGACCA
C3-J-5476	GCTTCTGCATGATATTGACATTTTGT
N4-J-8718	CTTAATATAAAATTTATGTG
N4-N-8578	ATTTACCCATAGACGAAGAC
N4-N-8924	AAAGCTCATGTTGAAGCTCC
N5-J-7567	AGAAAATGGAAKTTGAGCTCTT
LR-N-12901	AGGATAAATAGTTAGTCTGTTCCGAC
LR-J-12911	GTAAGATTTAAATAGTCGAACAGAC
SR-J-14199	TACTATGTTACGACTTAT
SR-N-14594	AAACTAGGATTAGATACCC

NOTE.—C1 = *cox1*; C2 = *cox2*; C3 = *cox3*; N4 = *nad4*; N5 = *nad5*; LR = *rrnL*; SR = *rrnS*; J = majority strand; N = minority strand.

s at 95°C, 30 s at 40°C, and 1–5 min at 68°C (1 min for each kb), followed by 5–10 min at 68°C.

The amplified PCR fragments were sequenced directly by the dideoxy terminator method. Sequencing products were precipitated with sodium acetate and ethanol and then resolved with an ABI 377 sequencer. Sequence data were compiled and aligned with the Sequencher program, version 3.1.1 (Gene Codes Corporation).

The protein-coding and rRNA genes were identified by BLAST searches (Altschul et al. 1990) of GenBank. The tRNA genes were found with the program tRNAscan-SE (Lowe and Eddy 1997) and by eye from the anticodons and inferred secondary structures of tRNAs.

## Results and Discussion

### Sizes of the Mt Genome and the Mt Genes of *H. macropus*

The mt genome of *H. macropus* is circular and has 14,670 bp; 12,686 bp (86.5%) of the genome and all gene boundaries were sequenced from both strands. All neighboring sequencing fragments overlap by more than 100 bp. The whole sequence of the mt genome of *H. macropus* is in GenBank under accession number AF270939.

For the BLAST searches, the closest matches to the putative genes *atp6*, *cox1–cox3*, *cob*, *nad2*, *rrnL*, *nad1* and *nad6* of *H. macropus* were the corresponding genes

of other insects. The closest matches to the putative genes *atp8*, *nad3–nad5*, *nad4L*, and *rrnS* of *H. macropus* were the corresponding genes of noninsect arthropods or nonarthropods. We identified *trnS<sub>I</sub>* by eye and other tRNA genes by the program tRNAscan-SE search. Thus, all 37 genes were identified putatively. The tRNA genes were 63–69 bp long. The putative sizes of the protein-coding and rRNA genes are as follows: *atp6*, 657 bp; *atp8*, 162 bp; *cox1*, 1,539 bp; *cox2*, 660 bp; *cox3*, 792 bp; *cob*, 1,101 bp; *nad1*, 894 bp; *nad2*, 897 bp; *nad3*, 327 bp; *nad4*, 1,347 bp; *nad4L*, 273 bp; *nad5*, 1,674 bp; *nad6*, 588 bp; *rrnL*, 1,170 bp; *rrnS*, 776 bp. The total size of the 37 genes is 14,299 bp. The majority strand encodes 19 genes—seven protein-coding genes and 12 tRNA genes, 7,263 bp in total; the minority strand encodes 18 genes—six protein-coding genes, two rRNA genes, and 10 tRNA genes, 7,036 bp in total.

Six pairs of adjacent genes apparently overlap in the mt genome of *H. macropus*: (1) 4 bp between *cox1* and *trnC*, (2) 6 bp between *trnL<sub>2</sub>* and *nad4*, (3) 2 bp between *nad4L* and *nad2*, (4) 4 bp between *trnN* and *trnA*, (5) 4 bp between *rrnS* and *atp6*, and (6) 1 bp between *atp6* and *atp8*. For the first four pairs of genes, the overlapping genes are encoded by different strands; for the last two pairs, the overlapping genes are encoded by the same strand.

### Nucleotide Composition and Codon Usage

The nucleotide composition of the majority strand of the mt genome of *H. macropus* is as follows: A, 5,680, 38.7%; T, 5,950, 40.6%; C, 1,539, 10.5% G, 1,501, 10.2%. The A+T-richness of the mt genome of *H. macropus* is reflected further in the codon usage. There is a strong bias against C+G (A, 44.6%; T, 46.6%; C, 4.2%; G, 5.1%) at the third codon positions of the protein-coding genes (table 2), as observed in the mt genomes of other insects (Crease 1999). Six A+T-rich codons (encoding amino acids Asn, Ile, Lys, Met, Phe, and Tyr) appear 1,558 times (42.8%), whereas four C+G-rich codons (encoding amino acids Ala, Arg, Gly, and Pro) are present 398 times (10.9%) (table 3). The overall ratio of C+G-rich codons to A+T-rich codons is 0.26. All codons are present in the protein-coding genes of this genome except for CGC and CCG. Four different codons apparently initiate translation of protein-coding genes: ATT for *nad3* and *nad4*, ATC for *cob* and *nad6*, ATG for *cox1* and *cox3*, and ATA for the other seven genes. Nine protein-coding genes have complete termination codons: TAG for *cox3* and TAA for *cox2*, *cob*, *nad1–nad3*, *nad4L*, *nad5*, and *nad6*. The oth-

**Table 2**  
Nucleotide Composition of Protein-Coding Genes in the Mitochondrial Genome of *Heterodoxus macropus*

CODON POSITIONS	MAJORITY STRAND				MINORITY STRAND				BOTH STRANDS			
	A	T	C	G	A	T	C	G	A	T	C	G
1	33.2	37.6	11.1	18.1	36.0	40.8	8.4	14.8	34.3	38.9	10.0	16.8
2	21.1	49.1	18.3	11.5	21.4	52.8	13.8	12.0	21.2	50.6	16.5	11.7
3	43.4	46.7	4.8	5.1	45.1	46.5	3.3	5.1	44.1	46.6	4.2	5.1
Overall	32.6	44.5	11.4	11.6	34.2	46.7	8.5	10.6	33.2	45.4	10.2	11.2

**Table 3**  
Codon Usage of Protein-Coding Genes in the  
Mitochondrial Genome of *Heterodoxus macropus*

AMINO ACID	CODON	MAJORITY STRAND		MINORITY STRAND	
		Count	%	Count	%
Phe	TTT	202	9.4	169	11.4
	TTC	31	1.4	8	0.5
Leu	TTA	236	10.9	191	12.9
	TTG	21	1.0	19	1.3
	CTT	37	1.7	25	1.7
	CTC	4	0.2	1	0.1
	CTA	23	1.1	16	1.1
	CTG	2	0.1	2	0.1
Ile	ATT	201	9.3	141	9.5
	ATC	14	0.6	5	0.3
Met	ATA	134	6.2	118	8.0
	ATG	24	1.1	17	1.1
Val	GTT	60	2.8	31	2.1
	GTC	5	0.2	0	0
	GTA	61	2.8	35	2.4
	GTG	4	0.2	4	0.3
Ser	TCT	88	4.1	44	3.0
	TCC	1	0	1	0
	TCA	47	2.2	44	2.8
	TCG	4	0.2	3	0.2
Pro	CCT	42	1.9	22	1.5
	CCC	2	0.1	1	0.1
	CCA	27	1.3	12	0.8
	CCG	0	0	0	0
Thr	ACT	46	2.1	23	1.6
	ACC	1	0	3	0.2
	ACA	32	1.5	16	1.1
Ala	ACG	0	0	2	0.1
	GCT	25	1.2	20	1.4
	GCC	3	0.1	5	0.3
	GCA	21	1.0	10	0.7
Tyr	GCG	2	0.1	1	0.1
	TAT	90	4.2	62	4.2
	TAC	10	0.5	11	0.7
Stop	TAA	6	0.3	6	0.4
	TAG	1	0	0	0
His	CAT	38	1.8	19	1.3
	CAC	9	0.4	2	0.1
Gln	CAA	27	1.3	8	0.5
	CAG	0	0	2	0.1
Asn	AAT	88	4.1	67	4.5
	AAC	9	0.4	6	0.4
Lys	AAA	71	3.3	64	4.3
	AAG	11	0.5	5	0.3
Asp	GAT	29	1.3	25	1.7
	GAC	9	0.4	0	0
Glu	GAA	46	2.1	34	2.3
	GAG	12	0.6	5	0.3
Cys	TGT	14	0.6	20	1.4
	TGC	3	0.1	1	0.1
Trp	TGA	54	2.5	27	1.8
	TGG	4	0.2	1	0.1
Arg	CGT	3	0.1	1	0.1
	CGC	0	0	0	0
	CGA	22	1.0	11	0.7
	CGG	3	0.1	2	0.1
Ser	AGT	11	0.5	10	0.7
	AGC	0	0	2	0.1
	AGA	64	3.0	50	3.4
	AGG	9	0.4	4	0.3

**Table 3**  
Continued

AMINO ACID	CODON	MAJORITY STRAND		MINORITY STRAND	
		Count	%	Count	%
Gly	GGT	33	1.5	9	0.6
	GGC	3	0.1	2	0.1
	GGA	65	3.0	29	2.0
	GGG	13	0.6	9	0.6
A + T-rich codons (1,558, 42.8% overall)		885	41.0	673	45.5
C + G-rich codons (398, 10.9% overall)		264	12.2	134	9.0
Codon ratio (0.26 overall) . . . . .		0.30		0.20	

NOTE.—A + T-rich codons are those encoding amino acids Asn, Ile, Lys, Met, Phe, and Tyr. C+G-rich codons are those encoding amino acids Ala, Arg, Gly, and Pro. The codon ratio is C + G-rich codons/A+T-rich codons.

er four genes, *cox1*, *nad4*, *atp6*, and *atp8*, could have the complete termination codon TAA if we accept gene overlaps 1, 2, 5, and 6 described above.

#### tRNA Genes

All of the mt tRNAs of *H. macropus* have the secondary structures commonly seen in animals studied (fig. 2). Three unusual features were observed: (1) an A-A mismatch in the amino acid acceptor (AA) arm of tRNA-Val and the anticodon (AC) arm of tRNA-His; (2) a U-U mismatch in the AA arm of tRNAs Ala, Arg, Asp, and Thr, in the D arm of tRNA-Lys, and in the AC arm of tRNA-Phe and tRNA-Ser (gcu); and (3) nine tRNAs—Ala, Arg, Asn, Ile, Leu (uag), Phe, Pro, Ser (uga), and Trp—have few or no nucleotides in the loop structure of the T $\psi$ C (T) arm.

#### rRNA Genes

We identified the 5' and 3' ends of *rrnL* and *rrnS* by comparing their nucleotide sequences with those of *D. yakuba* and other insects. There are 16 bp, but no tRNA gene, between *rrnL* and *rrnS* in *H. macropus*. The *trnV*, which is between the two rRNA genes in all arthropods studied so far, has moved to between *trnQ* and *nad6* in *H. macropus*.

#### Noncoding Regions and the Control of Gene Transcription and DNA Replication

There are 392 bp unassigned to genes in the mt genome of *H. macropus*. The largest noncoding region is 73 bp long. It is between *atp8* and *trnQ* and has 93% A+T. The second-largest noncoding region is 47 bp long. It is between *cox2* and *nad3* and has 92% A+T. Both of these regions seem to be able to form stem-loop structures (fig. 3). Besides these two regions, there are 272 bp unassigned to genes, scattered in short runs (1–32 bp) between neighbor genes. In contrast to *H. macropus*, most animal species studied so far have a single large noncoding region in their mt genomes, which is thought to control gene transcription and DNA replica-

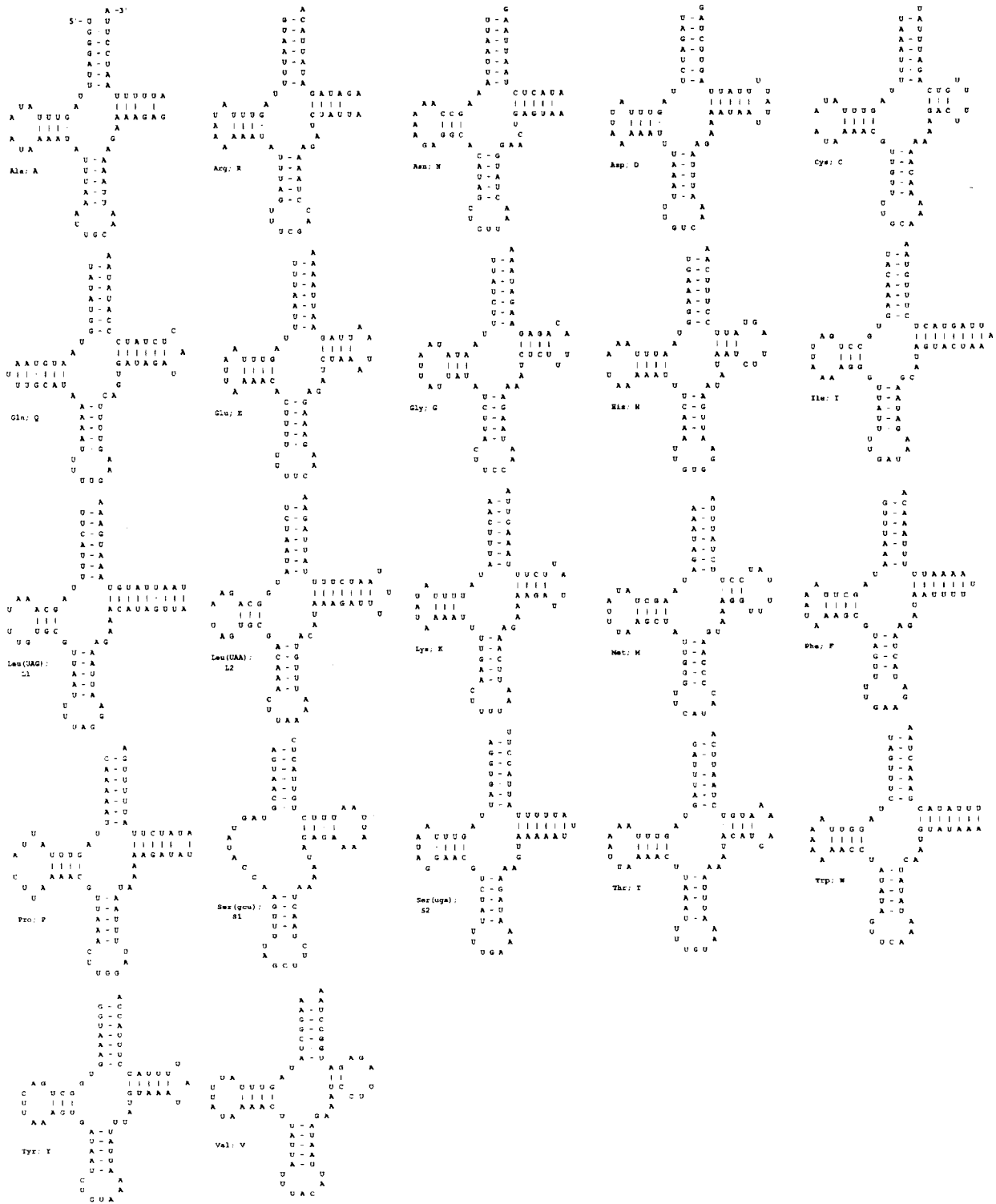


FIG. 2.—Inferred secondary structures of the 22 tRNAs of *Heterodoxus macropus*. tRNAs are labeled with the abbreviations of their corresponding amino acids. Nucleotide sequences are from 5' to 3' as indicated for tRNA-Ala. Dashes indicate Watson-Crick bonds, and centered dots indicate bonds between U and G. Arms of tRNAs (clockwise from top) are the amino acid acceptor (AA) arm, the TψC (T) arm, the anticodon (AC) arm, and the dihydrouridine (DHU or D) arm.

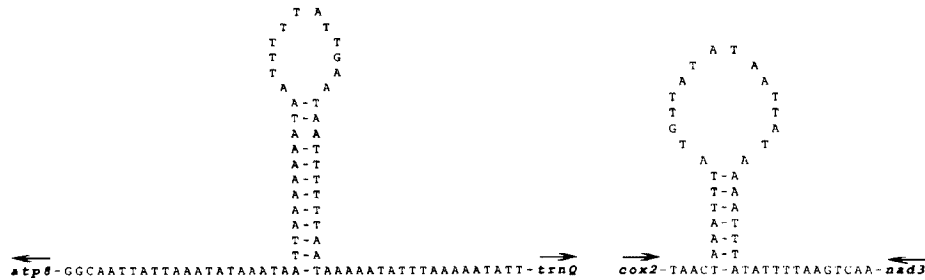


FIG. 3.—Inferred stem-loop structures in the two largest noncoding regions of *Heterodoxus macropus*. A, The stem-loop in the 73-bp noncoding region between *atp8* and *trnQ*. B, The stem-loop in the 47-bp noncoding region between *cox2* and *nad3*. Arrows indicate the direction of transcription of genes that flank the noncoding regions.

tion (Clayton 1982, 1991). Thus, it is called the control region, or the “A+T-rich region,” in insects. This region varies from about 350 bp in some butterflies (Lepidoptera) (Taylor et al. 1993) to 13,000 bp in some weevils (Coleoptera) (Boyce, Zwick, and Aquadro 1989). In insects, conserved stem-loop structures have been found in the control regions of fruit flies (Diptera) and locusts (Orthoptera) (Zhang and Hewitt 1997). In mammals, initiation of DNA replication of the minority strand (commonly called the light strand in vertebrates) occurs in a small (~30-bp) noncoding region which has the potential to form a stem-loop structure (Shadel and Clayton 1997). Macey et al. (1997) proposed that nucleotide mutations to the stem-loop region in vertebrates cause errors during the replication of the minority strand, and these errors might lead to rearrangements of mt genes. Lavrov, Boore, and Brown (2000) suggested that the stem-loop structures in the noncoding regions of the horseshoe crab *Limulus polyphemus* and the prostriate tick *Ixodes hexagonus* might correspond to the origin of DNA replication. They also suggested that the change of position of this origin might have caused gene rearrangements in the metastriate ticks *Rhipicephalus sanguineus* and *Boophilus microplus* (see also Campbell and Barker 1998, 1999).

Both the 73-bp and the 47-bp noncoding regions in the mt genome of *H. macropus* may fold into stem-loop structures (fig. 3). Thus, we propose that these two noncoding regions may be the origins of DNA replication of the two strands and may also play a role in gene transcription. However, experiments are needed to test this hypothesis. Small noncoding regions and very different gene arrangements have also been found in *Cepaea nemoralis* (Mollusca, Gastropoda; Terrett, Miles, and Thomas 1994), *H. roretxi* (Chordata, Urochordata; Yokobori et al. 1999), and *Laqueus rubellus* (Brachiopoda, Articulata; Noguchi et al. 2000). It would be interesting to see how DNA replication and gene transcription are controlled in these compact mt genomes.

#### Gene Rearrangements

Compared with the ancestral gene arrangement of insects, numerous gene rearrangements have apparently occurred in the mt genome of *H. macropus* (fig. 1). All 22 of the tRNA genes have moved and/or inverted relative to their ancestral positions. Of the 13 protein-cod-

ing genes and two rRNA genes, only *cox1*, *cox3*, *nad4*, *nad4L*, *rrnL*, and *rrnS* are in the same positions relative to each other in *H. macropus* and the ancestral insect. If we assume that these six genes are in their ancestral positions, then the other nine protein-coding genes (*atp6*, *atp8*, *cox2*, *cob*, *nad1*–*nad3*, *nad5*, and *nad6*) have moved. Moreover, four of these genes (*atp6*, *atp8*, *nad1*, and *nad3*) have inverted.

With a few notable exceptions, the gene arrangement in the mt genome is very conserved within closely related groups of animals. Before our study, the class Insecta typified this conservatism, since rearrangements of the protein-coding and rRNA genes had not been found in insects despite the fact that the Insecta were the most studied group of invertebrates.

The number of gene rearrangements implied by the gene arrangement in the mt genome of *H. macropus* indicates that the rate of change in the arrangement of mt genes may vary substantially among closely related lineages. Arthropods evolved more than 600 MYA (Bergström 1979). The inferred ancestral arrangement of mt genes of arthropods is the same as that of insects except for the position of one tRNA gene (Staton, Daehler, and Brown 1997; Boore, Lavrov, and Brown 1998). Thus, in the lineage leading to *D. yakuba*, there has been only one change in the arrangement of mt genes in more than 600 Myr. However, in the lineage leading to *H. macropus*, most genes have been rearranged over the course of about 300 Myr since it diverged from the lineage that leads to *D. yakuba* (Kukalová-Peck 1991).

Interestingly, the milkweed bug, *O. fasciatus* (Hemiptera), has the arrangement *trnL<sub>2</sub>*?-*cox2*?-*trnK* (Liu and Beckenbach 1992), which is the same as the ancestral arrangement of insects. Thus, the gene rearrangements in the mt genome of *H. macropus* might have occurred after the lineage leading to lice (Phthiraptera) and the lineage leading to bugs (Hemiptera) diverged about 280 MYA (Kukalová-Peck 1991). We are studying a variety of hemipteroid insects to determine the phylogenetic extent of the various gene rearrangements. We are also studying other species of lice to see when the gene rearrangements in the mt genome of *H. macropus* evolved and whether any intermediate states can be found. Doubtless, informative rearrangements would improve our understanding of the phylogeny and evolution of lice and other hemipteroid insects.

## Acknowledgments

We thank J. Boore for comments on a draft of the manuscript and the anonymous reviewers for comments which greatly improved this manuscript. This work was supported by a grant from the Australian Research Council.

## LITERATURE CITED

- ALTSCHUL, S. F., W. GISH, W. MILLER, E. W. MYERS, and D. J. LIPMAN. 1990. Basic local alignment search tool. *J. Mol. Biol.* **215**:403–410.
- BALLARD, J. W. O. 2000. Comparative genomics of mitochondrial DNA in members of the *Drosophila melanogaster* subgroup. *J. Mol. Evol.* **51**:48–63.
- BEAGLEY, C. T., R. OKIMOTO, and D. R. WOLSTENHOLME. 1998. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): introns, a paucity of tRNA genes, and a near-standard genetic code. *Genetics* **148**:1091–1108.
- BEARD, C. B., D. M. HAMM, and F. H. COLLINS. 1993. The mitochondrial genome of the mosquito *Anopheles gambiae*: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. *Insect Mol. Biol.* **2**:103–104.
- BEATEN, M. J., A. J. ROGER, and T. CAVALIER-SMITH. 1998. Sequence analysis of the mitochondrial genome of *Sarcophyton glaucum*: conserved gene order among octocorals. *J. Mol. Evol.* **47**:697–708.
- BERGSTROM, J. 1979. Morphology of fossil arthropods as a guide to phylogenetic relationships. Pp. 3–56 in A. P. GUP- TA, eds. *Arthropod phylogeny*. Van Nostrand Reinhold, New York.
- BOORE, J. L. 1999. Animal mitochondrial genomes. *Nucleic Acids Res.* **27**:1767–1780.
- BOORE, J. L., and W. M. BROWN. 1998. Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr. Opin. Genet. Dev.* **8**:668–674.
- BOORE, J. L., T. M. COLLINS, D. STANTON, L. L. DAEHLER, and W. M. BROWN. 1995. Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. *Nature* **376**:163–165.
- BOORE, J. L., D. V. LAVROV, and W. M. BROWN. 1998. Gene translocation links insects and crustaceans. *Nature* **392**:667–668.
- 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.
- CAMPBELL, N. J. H., and S. C. BARKER. 1998. An unprecedented major rearrangement in an arthropod mitochondrial genome. *Mol. Biol. Evol.* **15**:1786–1787.
- . 1999. The novel mitochondrial gene arrangement of the cattle tick, *Boophilus microplus*: fivefold tandem repetition of a coding region. *Mol. Biol. Evol.* **16**:732–740.
- CLARY, D. O., and D. R. WOLSTENHOLME. 1985. The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J. Mol. Evol.* **22**:252–271.
- CLAYTON, D. A. 1982. Replication of animal mitochondrial DNA. *Cell* **28**:693–705.
- . 1991. Replication and transcription of vertebrate mitochondrial DNA. *Annu. Rev. Cell Biol.* **7**:453–478.
- CREASE, T. J. 1999. The complete sequence of the mitochondrial genome of *Daphnia pulex* (Cladocera: Crustacea). *Gene* **233**:89–99.
- CROZIER, R. H., and Y. C. CROZIER. 1993. The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. *Genetics* **133**:97–117.
- FLOOK, P. K., C. H. F. ROWELL, and G. GELLISSEN. 1995. The sequence, organization and evolution of the *Locusta migratoria* mitochondrial genome. *J. Mol. Evol.* **41**:928–941.
- HOFFMANN, R. J., J. L. BOORE, and W. M. BROWN. 1992. A novel mitochondrial genome organization for the blue mussel *Mytilus edulis*. *Genetics* **131**:397–412.
- KEDDIE, E. M., T. HIGAZI, and T. R. UNNASCH. 1998. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. *Mol. Biochem. Parasitol.* **95**:111–127.
- KRISTENSEN, N. P. 1991. Phylogeny of extant hexapods. Pp. 125–140 in CSIRO, eds. *The insects of Australia: a textbook for students and research workers*. Melbourne University Press, Carlton, Victoria, Australia.
- KUKALOVÁ-PECK, J. 1991. Fossil history and the evolution of hexapod structures. Pp. 167–171 in CSIRO, eds. *The insects of Australia: a textbook for students and research workers*. Melbourne University Press, Carlton, Victoria, Australia.
- LAVROV, D. V., J. L. BOORE, and W. M. BROWN. 2000. The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. *Mol. Biol. Evol.* **17**:813–824.
- LE, T. H., D. BLAIR, T. AGARSUMA et al. (13 co-authors). 2000. Phylogenies inferred from mitochondrial gene orders—a cautionary tale from the parasitic flatworms. *Mol. Biol. Evol.* **17**:1123–1125.
- LESSINGER, A. C., A. C. M. JUNQUEIRA, T. A. LEMOS, E. L. KEMPER, F. R. DA SILVA, A. L. VETTORE, P. ARRUDA, and A. M. L. AZEREDO-ESPIN. 2000. The mitochondrial genome of the primary screwworm fly *Cochliomyia hominivorax* (Diptera: Calliphoridae). *Insect Mol. Biol.* **9**:521–529.
- LEWIS, D. L., C. L. FARR, and L. S. KAGUNI. 1995. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. *Insect Mol. Biol.* **4**:263–278.
- LIU, H., and A. T. BECKENBACH. 1992. Evolution of the mitochondrial cytochrome oxidase II gene among 10 orders of insects. *Mol. Phylogenet. Evol.* **1**:41–52.
- LOWE, T. M., and S. R. EDDY. 1997. tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res.* **25**:955–964.
- MACEY, J. R., A. LARSON, N. B. ANAJEVA, Z. FANG, and T. J. PAPPENFUSS. 1997. Two novel gene orders and the role of light-strand replication in the rearrangement of the vertebrate mitochondrial genome. *Mol. Biol. Evol.* **14**:91–104.
- MITCHELL, S. E., A. F. COCKBURN, and J. A. SEAWRIGHT. 1993. The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. *Genome* **36**:1058–1073.
- NOGUCHI, Y., K. ENDO, F. TAJIMA, and R. UESHIMA. 2000. The mitochondrial genome of the brachiopod *Laqueus rubellus*. *Genetics* **155**:245–259.
- OKIMOTO, R., H. M. CHAMBERLIN, J. L. MACFARLANE, and D. R. WOLSTENHOLME. 1991. Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (Meloidogyne): nucleotide sequences, genome location and potential for host-race identification. *Nucleic Acids Res.* **19**:1619–1626.
- OKIMOTO, R., J. L. MACFARLANE, D. O. CLARY, and D. R. WOLSTENHOLME. 1992. The mitochondrial genome of 2 nematodes, *Caenorhabditis elegans* and *Ascaris suum*. *Genetics* **130**:471–498.
- PONT-KINGDON, G., N. A. OKADA, J. L. MACFARLANE, C. T. BEAGLEY, C. D. WATKINS-SIMS, T. CAVALIER-SMITH, G. D. CLARK-WALKER, and D. R. WOLSTENHOLME. 1998. Mito-

- chondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial *mutS*: a possible case of gene transfer from the nucleus to the mitochondrion. *J. Mol. Evol.* **46**:419–431.
- SHADEL, G. S., and D. A. CLAYTON. 1997. Mitochondrial DNA maintenance in vertebrates. *Annu. Rev. Biochem.* **66**:409–433.
- SHAHJAHAN, R. M., K. J. H. ROGER, R. A. LEOPOLD, and J. D. DEVAULT. 1995. Lower incubation-temperature increases yield of insect genomic DNA isolated by the CTAB method. *Biotechniques* **19**:332.
- SIMON, C., F. FRATI, A. BECKENBACH, B. CRESPI, H. LIU, and P. FLOOK. 1994. Evolution, weighting and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Ann. Entomol. Soc. Am.* **87**:651–701.
- SPANOS, L., G. KOUTROUMBAS, M. KOTSYFAKIS, and C. LOUIS. 2000. The mitochondrial genome of the Mediterranean fruit fly, *Ceratitidis capitata*. *Insect Mol. Biol.* **9**:139–144.
- STATON, J. L., L. L. DAEHLER, and W. M. BROWN. 1997. Mitochondrial gene arrangement of the horseshoe crab *Limulus polyphemus* L: conservation of major features among arthropod classes. *Mol. Biol. Evol.* **14**:867–874.
- STECHMANN, A., and M. SCHLEGEL. 1999. Analysis of the complete mitochondrial DNA sequence of the brachiopod *Terbratulina retusa* places Brachiopoda within the protostomes. *Proc. R. Soc. Lond. B Biol. Sci.* **266**:2043–2052.
- TAYLOR, M. F. J., S. W. MCKECHNIE, N. PIERCE, and M. KREITMAN. 1993. The lepidopteran mitochondrial control region: structure and evolution. *Mol. Biol. Evol.* **10**:1259–1272.
- TERRETT, J. A., S. MILES, and R. H. THOMAS. 1994. The mitochondrial genome of *Cepaea nemoralis* (Gastropoda: Stylommatophora): gene order, base composition, and heteroplasmy. *Nautilus* **108**(Suppl. 2):79–84.
- WOLSTENHOLME, D. R. 1992. Animal mitochondrial DNA: structure and evolution. *Int. Rev. Cytol.* **141**:173–216.
- YOKOBORI, S., T. UEDA, G. FELDMAIER-FUCHS, S. PAABO, R. UESHIMA, A. KONDOW, K. NISHIKAWA, and K. WATANABE. 1999. Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretxi* (Chordata, Urochordata). *Genetics* **3**:1851–1862.
- ZHANG, D. X., and G. M. HEWITT. 1997. Insect mitochondrial control region: a review of its structure, evolution and usefulness in evolutionary studies. *Biochem. Syst. Ecol.* **25**:99–120.

ROSS CROZIER, reviewing editor

Accepted January 5, 2001