Insect Molecular Biology (2006) **15**(1), 63–68

Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta)

C. Covacin*, R. Shao*†, S. Cameron*‡ and S. C. Barker*

**Parasitology Section, School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, Australia;* †*Faculty of Pharmacy and Pharmaceutical Sciences, Fukuyama University, Fukuyama, Hiroshima, Japan; and* ‡*Department of Integrative Biology, Brigham Young University, Provo, Utah, USA*

Abstract

The arrangement of genes in the mitochondrial (mt) genomes of most insects is the same, or near-identical, to that inferred to be ancestral for insects. We sequenced the entire mt genome of the small pigeon louse, *Campanulotes bidentatus compar*, and part of the mt genomes of nine other species of lice. These species were from six families and the three main suborders of the order Phthiraptera. There was no variation in gene arrangement among species within a family but there was much variation in gene arrangement among the three suborders of lice. There has been an extraordinary number of gene rearrangements in the mitochondrial genomes of lice!

Keywords: Anoplura, Amblycera, *Campanulotes*, mtDNA.

Introduction

The mitochondrial (mt) genome of the wallaby louse, *Heterodoxus macropus*, is extraordinary because all of the 22 tRNA genes, nine of the 13 protein-coding genes and both of its rRNA genes have moved and/or inverted relative to the mt genome that has been inferred to be ancestral for the insects (Shao *et al*., 2001). Here, we address the following questions: are the mt genomes of other lice rearranged, relative to the hypothetical ancestor of the insects and relative to the wallaby louse, *H. macropus*, and if so, at which taxonomic level do species of lice share arrangements of mt genes: suborder, family, genus or groups of species?

doi: 10.1111/j.1365-2583.2005.00608.x

Received 17 May 2005; accepted after revision 4 August 2005. Correspondence: Catherine Covacin, Parasitology Section, School of Molecular and Microbial Sciences, The University of Queensland, Brisbane, 4072, Australia. Tel.: +61 733654881; fax: +61 733654620; e-mail: c.covacin@uq.edu.au

We sequenced the entire mt genome of the small pigeon louse, *Campanulotes bidentatus compar* (suborder Ischnocera: family Goniodidae), and part of the mt genomes of *Pediculus humanus* (Anoplura: Pediculidae), *Coloceras* sp. and *Physconelloides eurysema* (Ischnocera: Goniodidae), *Anaticola crassicornis* (Ischnocera: Philopteridae), *Heterodoxus octoseriatus* and *Macropophila clayae* (Amblycera: Boopiidae), *Gliricola porcelli* (Amblycera: Gyropidae) and *Menopon gallinae* (Amblycera: Menoponidae) (Table 1). Our data indicate that mt gene arrangements are shared by genera and some families in the Phthiraptera. The order Phthiraptera has 293 genera and 25 families (Durden & Musser, 1994 for the Anoplura and Price *et al*., 2003 for the other lice). Therefore, there are doubtless many as yet undiscovered arrangements of genes in the order Phthiraptera. Shared gene arrangements may be powerful phylogenetic markers (Boore & Brown, 1998; Murrell *et al*., 2003; Lavrov *et al*., 2004; Thao *et al*., 2004).

Results and discussion

Mitochondrial genome of Campanulotes bidentatus *compar*

The mt genome of *C. bidentatus* is circular and has 14 804 bp (GenBank accession no. AY968672). This genome has the 37 genes found in almost all studied Metazoa (Fig. 1). One strand, the majority strand (Simon *et al*., 1994), apparently encodes 36 out of the 37 genes: only the tRNA gene for glutamine, *trnQ*, was on the minority strand. This is the first report of a hexapod with a mt genome that has most of the tRNA genes (21 of 22 in this case), all 13 protein-coding genes and both ribosomal genes on the one strand. The only other arthropod with ≥ 36 of the 37 genes on the one strand is the crustacean *Tigriopus japonicus* (Machida *et al*., 2002). The nucleotide composition of the majority strand of the mt genome of *C. bidentatus*, which has 36 of the 37 genes, was: A 26.5% (3918), T 43.7% (6462), C 9.3% (1370) and G 20.6% (3054). The mt genome of *C. bidentatus* has a higher A + T than G + C content; this is typical of the arthropods studied to date (Lavrov *et al*., 2000).

Protein-coding genes. Blast searches (Altschul *et al*., 1997) of GenBank identified 12 of the 13 protein-coding genes but failed to identify *nad6*. This gene was identified, by

Table 1. Species from the order Phthiraptera (lice) studied

Species	Host and location of collection	Suborder; Family	Collector and reference	Accession numbers
Pediculus humanus	Humans (body or clothes louse – Orlando strain)	Anoplura; Pediculidae	S. C. Barker	DQ054849
Campanulotes bidentatus compar	*Columba livia* (domestic pigeon), Ipswich, Queensland, Australia	Ischnocera; Goniodidae	C. Covacin (B2074)	AY968672
Coloceras sp. (undescribed species)	*Geopelia striata*, Hawaii, USA	Ischnocera; Goniodidae	S. Bush	DQ007340
Physconelloides eurysema	*Columbina passerina*, Utah, USA	Ischnocera; Goniodidae	D. Clayton	DQ007341
Anaticola crassicornis	*Anax sibilatrix*, Rio Negro, Argentina	Ischnocera; Philopteridae	K. Johnson	DQ007339
Heterodoxus macropus	*Macropus agilis* (agile wallaby), Queensland, Australia	Amblycera; Boopiidae	S. C. Barker (B1198)	AF270939
Heterodoxus octoseriatus	*Petrogale herberti* (Herbertís rock wallaby), Mt Sebastapol Station, via Rockhampton, Queensland, Australia	Amblycera; Boopiidae	S. Cameron (B1205)	AY966422
Macropophila clayae	*Thylogale billardierii* (pademelon), Glengarry, Tasmania	Amblycera; Boopiidae	S. Cameron (B904)	AY966423
Gliricola porcelli	*Cavia porcellus* (guinea pig), Queensland, Australia	Amblycera; Gyropidae	A. Murrell (B836)	AY966424
Menopon gallinae	*Gallus gallus* (domestic chicken), Lower Sitesia; Poland	Amblycera; Menoponidae	E. Lonc (B927)	AY966425

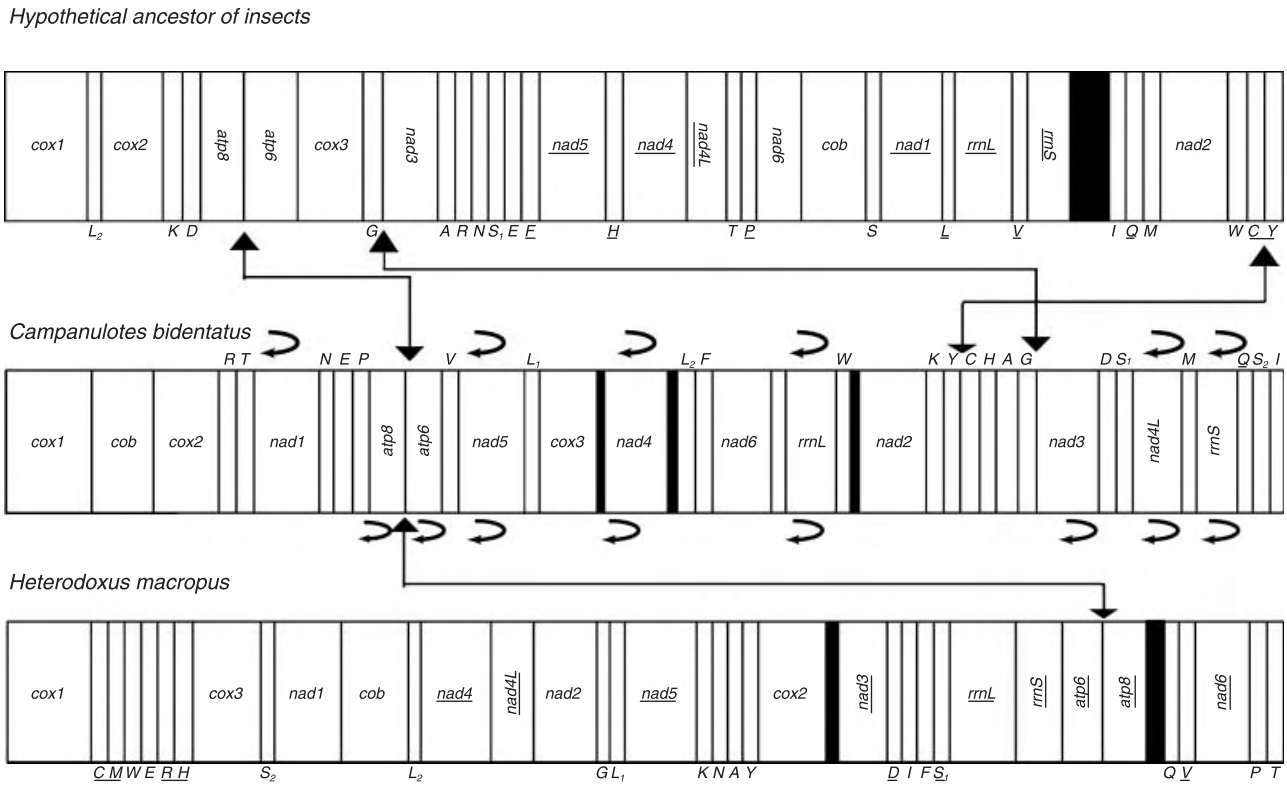

Figure 1. Arrangement of genes in the mitochondrial (mt) genomes of the hypothetical ancestor of insects, *Campanulotes bidentatus* (the small pigeon louse, suborder Ischnocera), and *Heterodoxus macropus* (the wallaby louse, suborder Amblycera). Genes are transcribed from left to right except those with underlined names, which are transcribed from right to left. tRNA genes have their single-letter amino acid abbreviations except for those that encode leucine and serine, which are labelled L_1 (anticodon tag), L_2 (taa), S_1 (tct), and S_2 (tga). Abbreviations of protein-coding genes and rRNA genes are: *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. Noncoding regions are shown as black boxes. The circling arrows indicate inversions of protein-coding genes and rRNA genes relative to the hypothetical ancestor of the insects and to the wallaby louse, *H. macropus*. Many of the tRNA genes have been inverted, but this is not shown. The straight arrows indicate the only three gene boundaries common to the mt genomes of *C. bidentatus* and the hypothetical ancestor of insects (*atp6–atp8*; *trnC–trnY*; *trnG–nad3*); and the single gene boundary common to the mt genomes of *C. bidentatus* and *H. macropus* (*atp6–atp8*).

comparisons of the hydrophilicity profiles of the putative NAD6 protein of *C. bidentatus* with the hydrophilicity profiles of NAD6 of *Drosophila yakuba* (Clary & Wolstenholme, 1985) and the wallaby louse, *H. macropus* (Shao *et al.*, 2001). Three start codons, ATA (for *cob*, *cox2*, *cox3*, *nad4*, *nad6*, *atp6*), ATT (for *cox1*, *nad1*, *nad5*, *nad4L*), ATG (for *atp8*, *nad2*, *nad3*), and the two stop codons TAA (for *cox1*, *cox2*, *cox3*, *cob*, *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*,

nad5, *nad6*, *atp8*) and TAG (for *atp6*), were apparent in the mt genome of *C. bidentatus*. Although incomplete (T or TA) stop codons are common in mt genomes it appears that the small pigeon louse, *C. bidentatus*, like the wallaby louse, *H. macropus*, has only complete stop codons in its protein-coding genes.

The genes for NADH dehydrogenase subunits 4 and 4L (*nad4* and *nad4L*) are not adjacent to one another in *C. bidentatus*; rather they are about 4 kb apart (Fig. 1). These two genes generally overlap or are adjacent to one another in mt genomes. Berthier *et al*. (1986) showed that *nad4* and *nad4L* were transcribed together and translated as a bicistonic mRNA in *D. melanogaster.* Although mt genomes with *nad4* and *nad4L* that are not adjacent have not been previously recorded in hexapods, this arrangement has been found in a crustacean, *Tigriopus japonicus* (Machida *et al*., 2002) and the chigger mite (Acari), *Leptotrombidium pallidum* (Shao *et al*., 2005).

rRNA genes and tRNA genes. The rRNA genes of *C. bidentatus* are not adjacent or even close to one another. Indeed, they are 2.3 kb apart (Fig. 1). Only one other insect, *Thrips imaginis*, is known to have rRNA genes that are not adjacent or close to one another (Shao & Barker, 2003). The mt genome of *C. bidentatus* has 22 tRNA genes, 63–73 bp long. The program tRNA scan-SE identified 20 out of the 22 tRNA genes. The other two tRNA genes, alanine (*trnA*) and serine 1 [$trnS_1$ (anticodon tct)] were identified by eye. *trnA* has three mismatches in the amino acid acceptor (AA) arm whereas $trnS_1$, histidine (*trnH*) and phenylalanine (*trnF*) have one mismatch in their anticodon (AC) arm (Fig. 2). The variable loop of $trnS_1$ has six nucleotides instead of the usual four and does not have a dihydrouridine (DHU or D) arm. The existence of two nucleotides instead of one between the (AC) arm, and the D loop in tRNA cysteine (*trnC*), is unusual (Fig. 2).

Noncoding regions. There are three sections of apparently noncoding nucleotides in the mt genome of *C. bidentatus*: (i) 56 bp between *cox3* and *nad4* (T 46.4%; A 28.6%; G 21.4%; C 3.6%) (ii) 66 bp between *nad4* and $trnL_2$ (T 28.8%; A 30.3%; G 40.9%; C 0%) and (iii) 66 bp between *trnW* and *nad2* (T 47%; A 27.3%; G 21.2%; C 4.5%). The 66 bp between *trnW* and *nad2* and the 56 bp between *cox3* and *nad4* can form stem-loops (Fig. 3): stem-loops are common in the large noncoding regions of the mt genomes of arthropods. These three sections of noncoding nucleotides account for 188 bp of the 322 noncoding nucleotides in this genome; the rest are located between genes and vary in size from 1 to 15 bp.

Gene arrangement. The mt genome of *C. bidentatus* has had an extraordinary number of translocations and inversions relative to the arrangement of genes that has been inferred to be ancestral for the insects. Indeed, *C. bidentatus* and the hypothetical ancestral arrangement have only three gene boundaries in common, *atp8–atp6*, *trnC–trnY* and *trnG–nad3* (Fig. 1).

Arrangement of genes downstream of cox1 *in 10 species of lice*

The arrangement of mt genes downstream of *cox1* varies substantially among the 10 species of lice studied so far (Fig. 4). Gene arrangements were shared by: (i) groups of species from a genus, the two *Heterodoxus* species; (ii) groups of genera from families, *C. bidentatus*, *Coloceras* sp. and *P. eurysema* (family Goniodidae), and *H. macropus*, *H. octoseriatus* and *M. clayae* (family Boopiidae); and (iii) groups of families from a suborder, Gyropidae and Menopodidae (suborder Amblycera). The Phthiraptera has 293 genera and 25 families (Durden & Musser, 1994 for the Anoplura and Price *et al*., 2003 for the other lice). So there are doubtless many as yet undiscovered arrangements of mt genes in the order Phthiraptera.

Experimental procedures

DNA was extracted from individual specimens with the DNeasy Tissue Kit (Qiagen, Valencia, CA, USA). PCRs were carried out with an Elongase® Enzyme Mix kit (Invitrogen™, Carlsbad, California, USA). First, insect-specific primers amplified parts of the mt genome of *C. bidentatus*; TM-J-206 (5′-GCTAAATAAGCTAACAGG TTCAT-3′), C1-J-1718 (5′-GGAGGATTTGGAAATTGATTAGTTcc-3′), C1-N-2329 (5′-ACTGTAAATATATGATGAGCTCA-3′), TL-J-3034 (5′-AATATGGCAGATTAGTGCA-3′), C2-N-3661 (5′-CCACAAATT-TCTGAACATTGACCA-3′), and LR-N-13398 (5′-CGCCTGTT-TAACAAAAACAT-3′) (Simon *et al*., 1994). To get the sequences downstream of *cox1* in the Amblycera and Anoplura, the primers C3-N-5460 (5′-TCA ACAAAGTGTCAGTATCA-3′) and LCIFI (5′-TTTTATATCAACATTTATTTTGATT-3′) were used for PCR and sequencing. The sequences of *cox1* for the Goniodidae were from an initial long-PCR with CAM10 (5′-AGAGGGACGAGAAGACCCTAT-AGA TCTTA-3′) and CAM7 (5′-CTTCTTCATAATCTACACCTT-TACTTGC-3′) (16S to 12S); subsequent PCRs were with GON1 (5′-AADWGTTGTGCCAGCWCTAGCGG-3′) and GON2 (5′-AGAA-TCTGACCTGACTYRCGTCGGTC-3′). For *A. crassicornis,* initial PCRs were with GON1 (5′-AADWGTTGTGCCAGCWCTAGCGG-3′) and GON2 (5′-AGAATCTGACCTGACTYRCGTCGGTC-3′) and subsequent PCRs were with CAM10 and N4-J-8944 (Simon *et al*., 1994). PCR conditions were: 92 °C for 2 min for one cycle, 92 °C for 30 s, 45–60 °C for 30 s, 60–68 °C for 2–10 min for 35 cycles followed by 60–68 °C for 3–12 min. Extension times were adjusted to the length of the target fragment; 1 min was allowed for each kb. A Touch Down HYBAID Temperature Cycling System (HYBAID, Middlesex, UK) and an Omni PCR cycler (HYBAID, Middlesex, UK) were used.

A QIAquick PCR Purification Kit (Qiagen) was used to prepare PCR products for sequencing and a QIAquick® Gel Extraction Kit (Qiagen) was used to prepare target bands (DNA) on agarose gels. DNA sequencing was carried out with the ABI BigDye dideoxy chain termination sequencing system (Applied Biosystems, Foster City, CA, USA). Two programs were used for the sequencing reactions: (i) one cycle of 94 °C for 5 min and 29 cycles at 94 °C

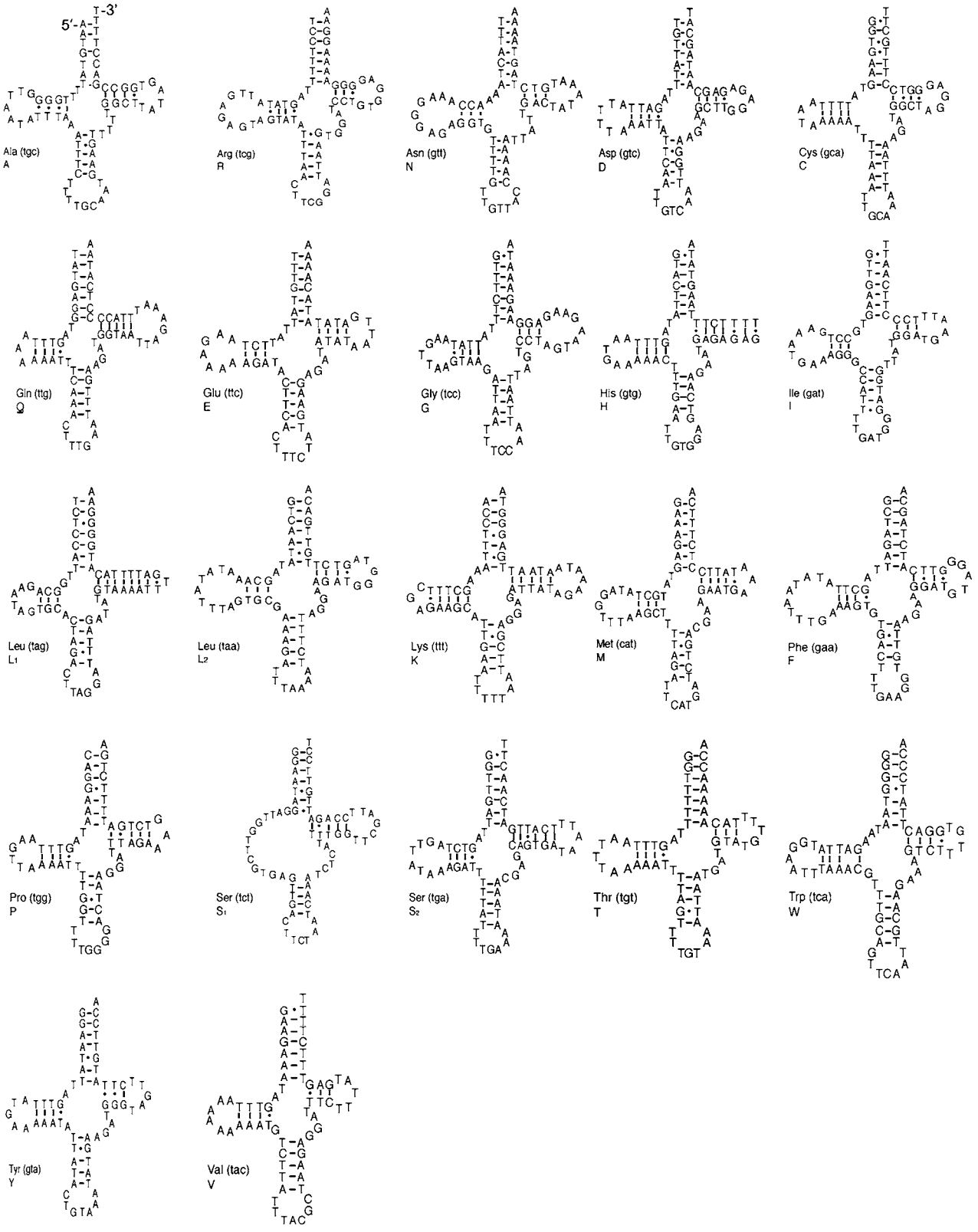

Figure 2. The secondary structures of the 22 tRNAs of *Campanulotes bidentatus*, inferred from nucleotide sequence of the tRNA genes. The tRNAs are labelled with the abbreviations of their corresponding amino acids. Nucleotide sequences are from 5′ to 3′ as indicated for tRNA-Ala. Dashes (–) indicate Watson–Crick bonds, and dots (•) indicate bonds between T 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.

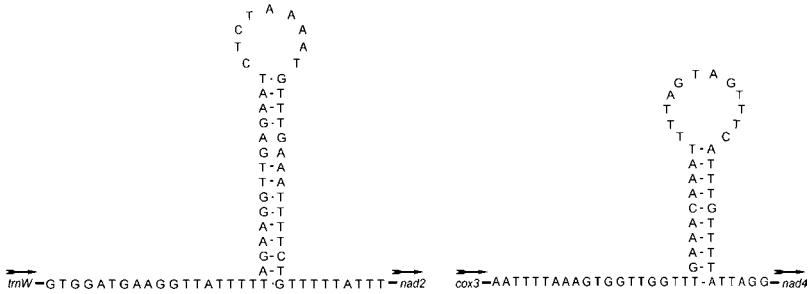

Figure 3. Inferred stem-loop structures in the noncoding regions of the mitochondrial genome of *Campanulotes bidentatus*. Arrows indicate the inferred direction of transcription; dashes (–) in the stem region indicate Watson-Crick bonds; and dots (•) indicate bonds between T and G.

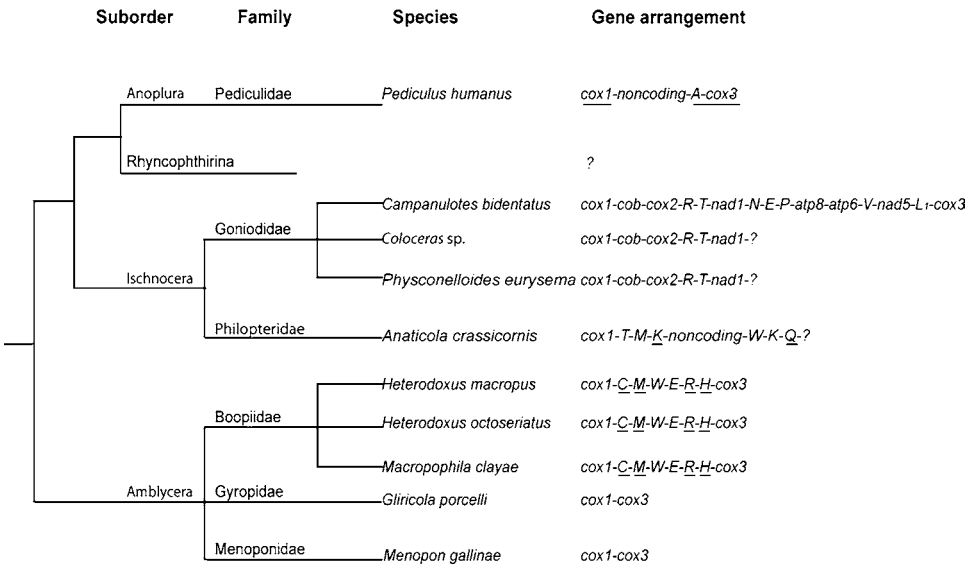

Figure 4. Arrangements of genes immediately downstream of *cox1* in the mitochondrial genomes of 10 species of lice. Genes are transcribed from left to right except those with underlined names, which are transcribed from right to left. The phylogenetic tree of the four suborders is after Barker (1994), only the families studied are listed. See Fig. 1 for the abbreviations of gene names.

for 20 s, 50 °C for 15 s and 60 °C for 4 min (this program was used for most of the sequencing reactions); and (ii) 94 °C for 5 min, 94 °C for 20 s and 60 °C for 4 min. Samples were submitted to the Australian Genome Research Facility (AGRF) for sequencing with ABI 377, 3700 and MD MegaBACE (capillary) automated sequencers.

Sequencing of the three Goniodidae species and *A. crassicornis* was carried out with the ABI BigDye ver. 3, with the following PCR sequencing conditions, 28 cycles of 94 °C for 10 s, 50 °C for 5 s, 60 °C for 4 min. All samples were cleaned with a sephadex column, dried in a 60 °C vacuum centrifuge and sequenced with an ABI 3770 capillary automated sequencer (sequencing primers for the Goniodidae and *A. crassicornis* are available from S.L.C. on request). Products that could not be sequenced directly were cloned with the pGEM®-T Easy Vector System II (Promega, Madison, WI, USA). The TOPO® XL PCR Cloning Kit (Invitrogen, Carlsbad, CA, USA) was used to clone fragments larger than 2 kb (4 kb). Protein-coding and rRNA genes were identified with BLAST searches (Altschul *et al.*, 1997; http://www.ncbi.nlm.nih.gov/BLAST/). Transfer RNA genes were identified with the program tRNAscan-SE 1.21 (Lowe & Eddy, 1997; http://www.genetics.wustl.edu/eddy/tRNAscan-SE/). Transfer RNAs not recognized by this program were found by examining the nucleotide sequences that flank the open reading frames of the protein-coding genes. The program MacVector (version 6.5.3, Accelrys Inc., San Diego, CA, USA) was used to identify open reading frames that could not be identified with the BLAST searches by comparing Hopp/Woods hydrophilicity profiles of putative proteins of *C. bidentatus* with those of *D. yakuba* (Kyte & Doolittle, 1982).

Acknowledgements

The Goniodidae and *Anaticola crassicornis* samples were courtesy of K.P. Johnson, D. Clayton and S. Bush. The work completed by Stephen Cameron was supported by NSF grants DEB0120718 and DEB444972.

References

Altschul, S.F., Madden, T.L., Schaffer, A.A., Zhang, J., Zhang, Z., Miller, W. and Lipman, D.J. (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* **25**: 3389–3402.

Barker, S.C. (1994) Phylogeny and classification, origins, and evolution of host associations of lice. *Int J Parasitol* **24**: 1285–1291.

Berthier, F., Renaud, M., Alziari, S. and Durand, R. (1986) RNA mapping on Drosophila mitochondrial-DNA – precursors and template strands. *Nucleic Acids Res* **14**: 4519–4533.

Boore, J.L. and Brown, W.M. (1998) Big trees from little genomes: mitochondrial gene order as a phylogenetic tool. *Curr Opin Genet Devel* **8**: 668–674.

Clary, D.O. and Wolstenholme, D.R. (1985) The mitochondrial DNA molecule of *Drosophila yakuba*: nucleotide sequence, gene organization, and genetic code. *J Mol Evol* **22**: 252–271.

Durden, L.A. and Musser, G.G. (1994) The mammalian hosts of the sucking lice (Anoplura) of the world: a host-parasite list. *Bull Soc Vect Ecol* **19**: 130–168.

Kyte, J. and Doolittle, R.F. (1982) A simple method for displaying the hydropathic character of a protein. *J Mol Biol* **157**: 102–132.

Lavrov, D.V., Boore, J.L. and Brown, W.M. (2000) The complete mitochondrial DNA sequence of the horseshoe crab *Limulus polyphemus*. *Mol Biol Evol* **17**: 813–824.

Lavrov, D.V., Brown, W.M. and Boore, J.L. (2004) Phylogenetic position of the Pentastomida and (pan) crustacean relationships. *Proc Roy Soc Lond B* **271**: 537–544.

Lowe, T., M. and Eddy, R. (1997) tRNAscan-SE: a program for improved detection of transfer RNA genes in genomic sequence. *Nucleic Acids Res* **25**: 955–964.

Machida, R.J., Miya, M.U., Nishida, M. and Nishida, S. (2002) Complete mitochondrial sequence of *Tigriopus japonicus* (Crustacea: Copepoda). *Mar Biotech* **4**: 406–417.

Murrell, A., Campbell, N.J.H. and Barker, S.C. (2003) The value of idiosyncratic markers and changes to conserved tRNA sequences from the mitochondrial genome of hard ticks (Acari: Ixodida: Ixodidae) for phylogenetic inference. *Syst Biol* **52**: 296–310.

Price, R.D., Hellenthal, R.A., Palma, R.L., Johnson, K.P. and Clayton D.A. (2003) *The chewing lice: world checklist and biological overview*. Illinois Natural History Survey Special Publication 24, Illinois.

Shao, R. and Barker, S.C. (2003) The highly rearranged mitochondrial genome of the Plague Thrips, *Thrips imaginis* (Insecta: Thysanoptera): convergence of two novel gene boundaries and an extraordinary arrangement of rRNA genes. *Mol Biol Evol* **20**: 362–370.

Shao, R., Campbell, N.J.H. and Barker, S.C. (2001) Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). *Mol Biol Evol* **18**: 858–865.

Shao, R., Mitani, H., Barker, S.C., Takahashi, M. and Fukunaga, M. (2005) Novel mitochondrial gene content and gene arrangement indicate illegitimate inter-mtDNA recombination in the chigger mite, *Leptotrombidium pallidum*. *J Mol Evol* **60**: 764–773.

Simon, C., Frati, F., Beckenbach, A., Crespi, B., Liu, H. and Flook, P. (1994) Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. *Annal Ento Soc Am* **87**: 651–701.

Thao, M.L., Baumann, L. and Baumann, P. (2004) Organization of the mitochondrial genomes of whiteflies, aphids, and psyllids (Hemiptera, Sternorrhyncha). *BMC Evol Biol* **4**: 25.