J Mol Evol (2007) 65:589–604
DOI 10.1007/s00239-007-9042-8

The Mitochondrial Genome of the Screamer Louse *Bothriometopus* (Phthiraptera: Ischnocera): Effects of Extensive Gene Rearrangements on the Evolution of the Genome

Stephen L. Cameron · Kevin P. Johnson · Michael F. Whiting

Received: 22 April 2007 / Accepted: 18 September 2007 / Published online: 10 October 2007

Abstract Mitochondrial (mt) genome rearrangement has generally been studied with respect to the phenomenon itself, focusing on their phylogenetic distribution and causal mechanisms. Rearrangements have additional significance through effects on substitution, transcription, and mRNA processing. Lice are an ideal group in which to study the interactions between rearrangements and these factors due to the heightened rearrangement rate within this group. The entire mt genome of the screamer louse *Bothriometopus* was sequenced and compared to previously sequenced louse genomes. The mt genome is 15,564 bp, circular, and all genes are encoded on the same strand. The gene arrangement differs radically from both other louse species and the ancestral insect. Nucleotide composition is A+T biased, but there is no skew which may be due to reversal of replication direction or a transcriptional effect. *Bothriometopus* has both tRNA duplication and concerted evolution which has not been observed previously. Eleven of the 13 protein-coding genes have 3′ end stem-loop structures which may allow mRNA processing without flanking tRNAs and so facilitate gene rearrangements. There are five candidate control regions capable of forming stem-loop structures. Two are structurally more similar to the control regions of other insect species than those of other lice. Analyses of *Bothriometopus* demonstrate that louse mt genomes, in addition to being extensively rearranged, differ significantly from most insect species in nucleotide composition biases, tRNA evolution, protein-coding gene structures and putative signaling sites such as the control region. These may be either a cause or a consequence of gene rearrangements.

Keywords Mitochondrial genomics · Lice · Ischnocera · Nucleotide skew · Control region structure

Electronic supplementary material The online version of this article (doi: 10.1007/s00239-007-9042-8) contains supplementary material, which is available to authorized users.

S. L. Cameron (✉)
Australian National Insect Collection and CSIRO Entomology, Black Mountain Laboratories, PO Box 1700, Canberra, ACT 2601, Australia
e-mail: stephen.cameron@csiro.au

K. P. Johnson
Illinois Natural History Survey, 1816 South Oak St, Champaign, IL 61820, USA

S. L. Cameron · M. F. Whiting
Department of Integrative Biology, Brigham Young University, Provo, UT 84602, USA

Introduction

Over the last decade mitochondrial (mt) genomes have become a major data sources for comparative genomics (Boore 1999). These small genomes, typically less than 16,000 bp, in combination with a generally conserved gene complement and rapid rate of nucleotide substitution, provide an ideal system for a wide range of comparative studies that have furthered our understanding of genome evolution. One question has received more attention than any other: the frequency and distribution of mt genome rearrangements. The discovery of novel gene arrangements has been reported with considerable interest (e.g. Shao et al. 2001; Covacin et al. 2006). The relative frequency of rearrangements is known to vary between lineages, and considerable effort has been devoted to documenting which lineages have the highest rates of genome rearrangement. For example, most vertebrate species share a common mt gene order except for groups such as amphibians (Mueller

and Boore 2005; Fonseca et al. 2006) and birds (Mindell et al. 1998). However rearrangements within these groups still appear to be uncommon. In contrast, rearrangements appear much more common in invertebrates and variability in gene order can be found within groups at many taxonomic levels (Dowton et al. 2002; Hassanin et al. 2005).

Beyond the absolute number of genome rearrangements within a taxonomic lineage is the question of whether certain types of gene rearrangement are more common than others. Genome rearrangements can be characterized in several aspects: (1) the types of genes rearranged, tRNAs only (termed minor rearrangements) or also protein-coding and rRNA genes (major rearrangements); (2) whether genes are translocated along the same strand or inverted between coding strands; and (3) the localization of the rearrangement (between local or distant gene blocks) (Dowton et al. 2002). A given gene rearrangement may exhibit one or more of these aspects. In addition, the likelihood of each type of rearrangement would vary according to different mechanistic models of how genome arrangements occur. Accurate knowledge of the frequency of the different types of rearrangements is thus useful in assessing the validity of these models. For example, the incidence of genome inversions has been used as evidence for the occurrence of intra-mitochondrial recombination, because inversions could not be explained by the alternative replication-slippage-based models (Dowton and Campbell 2001). An estimate of the frequency of gene inversions could thus give an indication of the frequency of recombination.

Secondly, there is the influence that a rearrangement itself may play on the evolution and function of the mt genes. The substitution rates of mitochondrial genes have been correlated with their physical position within the genome (Saccone et al. 1999; Faith and Pollock 2003). There is a mutational bias in mt genomes favoring As and Cs on the leading replicative strand. This bias is due to the deamination of As and Cs on the lagging-strand during replication leading to the accumulation of Gs and Ts on the lagging strand and their complementary pairs As and Cs on the leading strand (Reyes et al. 1998). The strength of the mutational bias is related to the length of time each gene spends in a single-stranded state during the replication or transcription cycles (Tanaka and Ozawa 1994; Francino and Ochman 1997; Hassanin et al. 2005). Rearrangements could theoretically alter these patterns by shifting genes to sites with a lower mutational background, or shifting between strands and thus reversing the mutational bias. Another possibility is that the origin of replication itself could be shifted or inverted, thus altering the mutational dynamics of the entire genome. It is unknown over what time scales these mutational effects would persist. Therefore, studies of rearrangements could potentially be informative as to the relative strengths of the various mutational effects operating within the mt genome: gene position versus strand biases, synonymous versus nonsynonymous etc. Rearrangements could also lead genes to violate the models used in phylogenetic or molecular dating approaches; however, such effects have never been examined in groups in which rearrangements are known to occur (Rubinoff et al. 2006).

Functionally, rearrangements would likely have an impact on mt gene transcription and translation. Mitochondrial genomes are transcribed as polycistronic pre-mRNAs which are processed by endonucleases which recognize the specific secondary structures of tRNA genes (Ojala et al. 1980, 1981), or possibly stem-loop structures in the 3′ end of those protein-coding genes which are not flanked by tRNAs (Kim et al. 2006; Fenn et al. 2007). Such mechanisms suggest that to ensure proper translation protein-coding genes would be limited to rearrangements which place them adjacent to tRNAs, otherwise post-transcriptional processing would fail. Genome rearrangements could also potentially be constrained by the signaling regions used to identify initiation and termination of transcription. The traditional model of mt genome transcription has emphasized the transcription of two full genome-sized polycistronic pre-mRNAs (Taanman 1999). However, recently a model was proposed where genes encoded on the same strand, and located between initiation and termination sites, are transcribed as a block considerably smaller than full genome size (Roberti et al. 2003, 2006). Rearrangements that moved genes from within their transcription block would therefore result in a loss or serious reduction in the rate of transcription. Examination of rearranged genomes may yield evidence toward which genome features are important as signaling sites or reveal novel methods of genome function that predisposed them to rearrange by releasing the genome from such constraints.

Comparative evolutionary studies of mt genome rearrangements are therefore necessary to understand their role in genome evolution and to accurately identify the mechanisms responsible for rearrangements. In choosing study groups for additional attention it is necessary to examine examples in which significant genome rearrangements have occurred within the group rather than in the group's common ancestor (genome variability as opposed to genome synapomorphies). Ideally a wide variety of the different rearrangement types (minor and major, translocations and inversions, local and distant) would also occur within the study group to allow the simultaneous examination of multiple classes of rearrangement types. Under these criteria, lice (Insecta: Phthiraptera) appear to be an ideal group for comparative mt genomics. Rearrangements have been identified in only 11 of 29 insect orders examined (Cameron et al. 2006). However, major

rearrangements or those involving inversions and rearrangements between distant gene blocks are limited to the four hemipteroid orders (Hemiptera, Thysanoptera, Psocoptera and Phthiraptera) and appear to be most frequent in lice. Each of the two lice that have been sequenced exhibit unique gene orders, and share only a single gene boundary in common with each other and only three (*Campanulotes*) or two (*Heterodoxus*) gene boundaries in common with the ancestral insect arrangement (Shao et al. 2001; Covacin et al. 2006). Comparisons between the two louse mt genomes are, however, complicated by the fact that each represents a different louse suborder, Amblycera for *Heterodoxus* and Ischnocera for *Campanulotes*. Recent phylogenetic studies also suggest that lice are not monophyletic and that parasitism may have evolved independently in these two suborders (Johnson et al. 2004). In so far as parasitism has been proposed as a potentially significant factor in mt genome rearrangement (Castro et al. 2002), comparisons of additional lice within each suborder are necessary to understand the evolution of mt genome rearrangement in the broadest sense. In this study we sequenced the mt genome of a second species of Ischnocera, *Bothriometopus macrocnemis*, to document the prevalence of additional rearrangements within this suborder and to examine the influence of rearrangements on mt genome evolution.

Materials and Methods

Specimen Collection

Specimens of *Bothriometopus macrocnemis* were collected from the Crested Screamer, *Chauna torquata* (Aves: Ahminidae) in Cordoba, Argentina on 25 April 2001, preserved in 100% ethanol, and stored at –80°C until used for DNA extraction. A voucher consisting of a slide mount of the exoskeleton cleared following DNA extraction was deposited in the Brigham Young University Life Science Museum collection, accession number IGC-PH65.

PCR Amplification and Sequencing

Whole genomic DNA was extracted from an individual louse specimen with the DNeasy Tissue kit (QIAGEN, Valencia, California, USA). Initial polymerase chain reaction (PCR) amplifications included a partial section of the *cox1* gene (approx 400 bp) using the primers L6625 and H7005 (Hafner et al. 1994) and the region *rns* to *rnl* (approx. 3,000 bp) using the degenerate primers GON1 and GON2 designed previously from the *rns* and *rnl* genes of lice (Covacin et al. 2006). These sequences were used to design specific primers to amplify the remaining mt genome by long PCR: *rnl* to *cox1* (approx. 8,000 bp) and *cox1* to *rns* (approx. 6,000 bp). Primer sequence and location for each long PCR is listed in Table 1. Within each long PCR product the full, double-stranded sequence was determined by primer walking (primers available from SLC upon request). Long PCRs were performed using Elongase (Invitrogen, Carlsbad, California) with the following cycling conditions: 92°C for 2 min; 40 cycles of 92°C for 30 sec, 50°C for 30 sec, 68°C for 12 min; and a final extension step of 68°C for 20 min. Sequencing was performed using ABI BigDye ver. 3 dye terminator sequencing technology and run on ABI 3770 or ABI 3740 capillary sequencers. Cycle sequencing conditions were 28 cycles of 94°C for 10 sec, 50°C for 5 sec, 60°C for 4 min.

Analysis and Annotation

Raw sequence files were proof read and assembled into contigs in Sequencher version 4.6 and 4.7 (GeneCodes Corporation). Transfer RNA analysis was conducted using tRNAscan-SE (Lowe and Eddy 1997) using invertebrate mitochondrial codon predictors and a cove score cut-off of

Table 1 Primers, sequence and location for long polymerase chain reactions

Region	Primer pair (F and R)	Location[a]	Sequence (5′→3′)
rns → *rnl*	GON1[b]	6180	AAD WGT TGT GCC AGC WCT AGC GG
	GON2[b]	9737	AGA ATC TGA CCT GAC TYR CGT CGG TC
rnl → *cox1*	BOTH5[c]	8509	TGC TGG ATA GTT TTA GAG ATA GG
	BOTH10[c]	1037	ACT ACA CCT GTG AGT CCA CCC AAG G
cox1 → *rns*	BOTH9[c]	899	AGT AGG TAT GGA TAT TGA TAG ACG AGC
	BOTH8[c]	6754	TAT CCT TTT ACG GAG TGA CGG GCG

[a] Location of the 3′ base of primer in the mt genome, relative to 5′ end of cox1

[b] Primers from Covacin et al. (2006)

[c] Primers newly designed for this genome

1. Reading frames between tRNAs were found in Sequencher and identified using translated basic logical alignment search tool (BLAST) searches (blastx) (Altschul et al. 1997) as implemented at the National Centre for Biotechnology Information (NCBI) website (http://www.ncbi.nlm.nih.gov/). Codon usage was calculated using MEGA ver. 3 (Kumar et al. 2004). Homology of tRNAs was determined in the first instance by anticodon sequence and checked by phylogenetic analysis with the tRNAs from *Triatoma* (Hemiptera), lepidopsocid RS-2001 (Psocoptera), and *Campanulotes* (Phthiraptera). All tRNA isotypes for the four species were aligned using their secondary structure (as predicted in tRNAScan-SE) with length-variable loop regions and the anticodon loop removed and phylogenies inferred on the most conserved portions of the genes. The resulting dataset was analyzed using Neighbour Joining in PAUP ver. 4b10 (Swofford 2002) and MrBayes version 3.1 (Ronquist and Huelsenbeck 2003). Alignments of the protein-coding genes were made in MEGA ver. 3 for the three louse species with *Triatoma* and lepidopsocid RS-2001 as outgroups. Hydrophobicity profiles were calculated in Genious version 2.5.4 (Drummond et al. 2006). Secondary structures between adjacent protein-coding genes and within the repetitive nontranslated regions were predicted using Mfold version 3.2 (Zuker 2003).

Results and Discussion

Genome Organization and Structure

The mt genome of *Bothriometopus macrocnemis* is circular and 15,564 bp in size, making it the largest louse mt genome yet sequenced (GenBank accession number EU183542). The genome encodes 40 genes, 36 of the 37 genes common to most metazoan mt genomes plus three additional copies of the tRNA-Trp, and an additional copy of the tRNA-Val. The gene for the tRNA-Ala is lacking. Additionally, there are two putative nontranslated regions each consisting of a 5′ nonrepetitive section plus two complete and one partial repeat units. Each region includes an open reading frame that BLASTs with low significance to the *cob* gene of the pseudococcid bug *Ferrisia* (Baumann and Baumann 2005). There is, however, no significant sequence similarity at either the DNA or amino acid level between these ORFs and the *cob* gene from *Bothriometopus*, suggesting that these regions are not pseudogenes and are unrelated to the full-length *cob* gene. They are the first record of nontranslated, repetitive sequence in a louse mt genome. There are also 25 intergene spacer regions ranging in size from 1 to 83 bp, 14 of which are over 20 bp in size. This is considerably more spacer regions than are found in most insect mitochondrial genomes, particularly the other louse genomes, which are especially compact. Collectively the spacer and nontranslated regions are responsible for increased size of the *Bothriometopus* mt genome, which is approximately 750 bp larger than the next largest louse genome.

All 40 genes of the *Bothriometopus* mt genome are transcribed on the same strand, a condition not typical in arthropods. Similarly, in the ischnoceran louse *Campanulotes*, 36 of 37 genes are transcribed on the same strand (Covacin et al. 2006). The only other arthropod known to have all genes on the same strand is the copepod *Tigriopus* (Machida et al. 2002; Burton et al., unpublished data). More broadly, a mt genome with all genes on the same strand is found in many metazoan phyla including Acanthocephala (Steinauer et al. 2005), Annelida (Boore and Brown 2000), Cnidaria (Medina et al. 2006; Brugler and France 2007), Mollusca (Hoffmann et al. 1992), Nematoda (He et al. 2005; but this is not the ancestral condition for the phylum, see Lavrov and Brown 2001), Platyhelminthes (Littlewood et al. 2006), Terebrantia (Helfenbein et al. 2001) and Urochordata (Yokobori et al. 1999, 2005). Indeed, genes transcribed on both strands only appears to be the ancestral state within Ecdysozoa (Webster et al. 2006), Vertebrata (Boore 1999) and several basal metazoan groups such as Porifera (Lavrov and Lang 2005a) and Placozoa (Signorovitch et al. 2007). It is unclear what, if any, significance the secondary adoption of single-stranded transcription in lice provides. Use of a single polycistronic transcript for mRNA, tRNA, and rRNA production could conceivably be more energetically efficient than using two or more, but this does not explain why such large groups as arthropods and vertebrates almost exclusively use at least two transcripts for genes distributed across both strands.

The mt genome arrangement of *Bothriometopus* differs radically from those of both other louse species and from the inferred ancestral insect arrangement (Fig. 1; Table 2). Only four gene boundaries and three gene blocks are shared between *Bothriometopus* and the ancestral insect: *G-nad3*, *atp8-atp6*, and *nad4-H-nad5*. Two of these boundaries are shared with the pigeon louse *Campanulotes*: *G-nad3* and *atp8-atp6*. However, no derived gene arrangements are shared between these two ischnoceran species. The gene boundary *atp8-atp6* is the only one shared with wallaby louse *Heterodoxus*, an amblyceran. A member of the closest relatives of parasitic lice, Psocoptera (Lepidopsocidae) shares none of its derived gene arrangements with *Bothriometopus* or with any of the other louse mt genomes. Even if the arrangement of the tRNA genes is ignored and only the relative positions of the protein-coding genes and rRNAs are considered, no additional arrangements are shared between *Bothriometopus* and *Campanulotes*. In addition, only a single extra

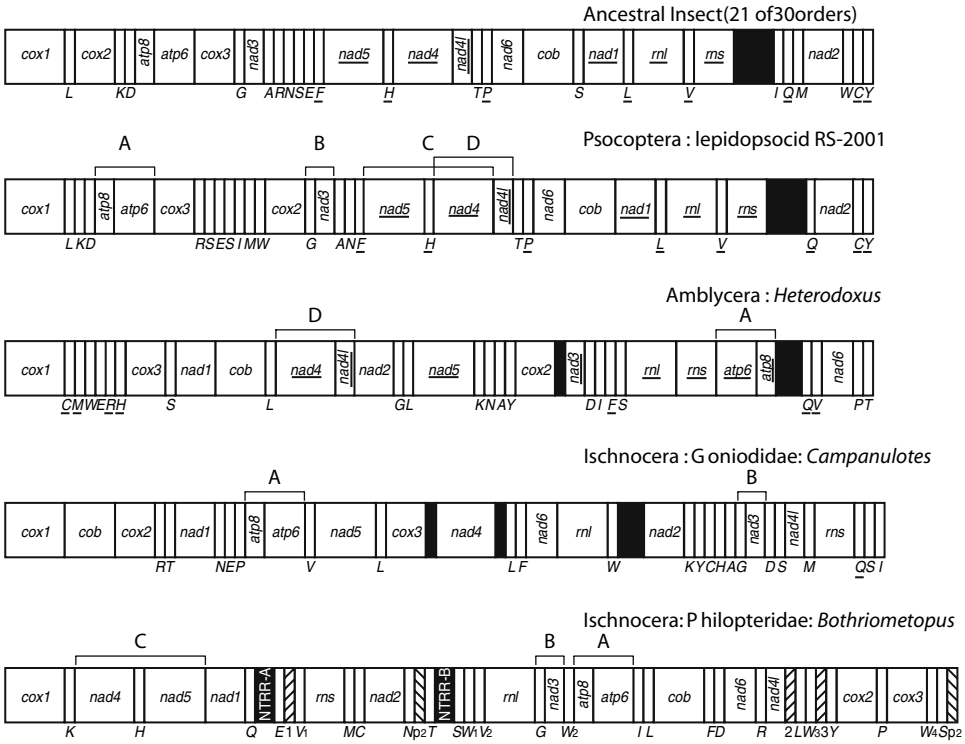

Fig. 1 Maps of the mitochondrial genomes of *Bothriometopus*, the other louse species *Heterodoxus* and *Campanulotes*, a psocopteran, and the ancestral insect. Each genome has been linearized at an arbitrary point, the 5′ end of *cox1*. Gene names are the standard abbreviations used in this paper; tRNA genes are indicated by the single-letter IUPAC-IUB abbreviation for their corresponding amino acid, duplicated tRNA genes are numbered as in the text, pseudo-tRNA genes are hatched and indicated by a lower-case p; intergenic spacers are cross hatched and numbered as in the text, putative control regions are in black; NTRR: nontranslated repeat region. Genome orientation for each species is 5′ to 3′ on the majority strand and genes coded on the minority strand are underlined. Shared gene boundaries are indicated by brackets and letter coded: A: *atp8/atp6*; B: *tRNA-G/nad3*; C: *nad4-tRNA-H-nad5*; D: *nad4L-nad4*

arrangement, *nad3-rnl*, is shared by *Bothriometopus* and *Heterodoxus*, although this gene block is inverted between the two genomes. Such a lack of conserved gene blocks precludes the accurate reconstruction of the rearrangement events that gave rise to the extant mt genomes of lice. Additional louse mt genomes are needed to split up the long evolutionary distances between the current exemplar species and allow more accurate interpretation of their genomic history.

Nucleotide Composition

The nucleotide composition of the coding strand of the *Bothriometopus* mt genome is: A: 32.1% (5,001), C 13.7% (2,135), G 15.5% (2,410) and T 38.7% (5,018). Combined GC content was thus 29.2%, in the middle of the range for insect mt genomes (Fenn et al. 2007) and comparable to the previously sequenced louse mt genomes: *Campanulotes* 29.9% and *Heterodoxus* 20.7%. There is almost no nucleotide skew (*sensu* Perna and Kocher 1995); A-skew is –0.002 and C-skew is –0.06. This is comparable to *Heterodoxus*, A-skew –0.02 and C-skew 0.01, but in marked contrast to the other ischnoceran *Campanulotes*, A-skew –0.25 and C-skew –0.38. Nucleotide frequency and skew statistics calculated for the entire genome do not take into account the distinction between synonymous and nonsynonymous mutations. Mutations at the third codon position of fourfold degenerate sites, those coding for the amino acids A, G, $L^{(CTN)}$, P, R, $S^{(TCN)}$, $S^{(AGN)}$, T and V, are selectively neutral and nucleotide composition at these sites is more indicative of background mutational pressures on nucleotide bias and skew (Kimura 1983). Table 3 shows nucleotide composition, A-skew, and C-skew for *Bothriometopus* and *Campanulotes* in which all 13 protein-coding genes are transcribed from a single strand, plus results for each strand for *Heterodoxus*, a lepidopsocid (Order Psocoptera), and *Triatoma* (Order Hemiptera) in which genes are transcribed on both strands. These results confirm that the mt genome of *Bothriometopus*, while strongly AT-biased, is essentially unskewed; there are approximately equal numbers of each of the complementary nucleotides, A:T, G:C. This finding is unusual because mt genomes typically demonstrate pronounced strand asymmetry in the nucleotide skew (Tanaka and Ozawa 1994; Reyes et al. 1998; Hassanin et al. 2005; Cameron and Whiting 2007).

Table 2 Summary of the mitogenome of *Bothriometopus*

Gene	Location	Size	Anticodon	Start codon	Stop codon
cox1	1–1533	1533		ATT	TAA
tRNA-Lys	1567–1629	63	TTT 1596–1598		
nad4	1630–2934	1305		ATT	TAA
tRNA-His	2938–2997	60	GTG 2966–2968		
nad5	2999–4630	1632		ATT	TAA
nad1	4635–5563	929		ATA	TA
tRNA-Gln	5564–5627	64	TTG 5590–5592		
NNTR-A 5′	5628–5687	60			
NNTR-A repeat 1	5688–5756	69			
NNTR-A repeat 2	5757–5825	69			
NNTR-A repeat 3p	5826–5859	34			
tRNA-Glu	5860–5922	63	TTC 5891–5893		
Intergene spacer 1	5923–5983	62			
tRNA-Val 1	5984–6044	61	TAC 6014–6016		
rns	6045–6828	784			
tRNA-Met	6829–6892	64	CAT 6859–6861		
tRNA-Cys	6915–6975	62	GCA 6944–6946		
nad2	6980–7987	1008		CTG	TAA
tRNA-Asn	8014–8079	66	GTT 8045–8047		
tRNA-Pseudo 1	8079–8152	73	CTAT 8120–8123		
tRNA-Thr	8154–8217	64	TGT 8184–8186		
NNTR-B 5′	8218–8304	87			
NNTR-B repeat 1	8305–8373	69			
NNTR-B repeat 2	8374–8442	69			
NNTR-B repeat 3p	8443–8490	48			
tRNA-Ser$^{(AGN)}$	8495–8568	74	TCT 8528–8530		
tRNA-Trp 1	8583–8642	60	TCA 8613–8615		
tRNA-Val 2	8650–8710	61	TAC 8680–8682		
rnl	8711–9822	1112			
tRNA-Gly	9823–9884	62	TCC 9853–9855		
nad3	9884–10237	354		ATG	TAA
tRNA-Trp 2	10253–10321	69	TCA 10291–10293		
atp8	10322–10495	174		ATT	TAG
atp6	10488–11202	684		ATG	TAA
tRNA-Ile	11203–11268	66	GAT 11232–11234		
tRNA-Leu$^{(CUN)}$	11277–11341	65	TAG 11307–11309		
cob	11342–12439	1098		ATT	TAA
tRNA-Phe	12464–12528	65	GAA 12492–12494		
tRNA-Asp	12529–12593	65	GTC 12559–12561		
nad6	12594–13085	492		ATA	TAG
tRNA-Arg	13106–13173	68	TCG 13136–13138		
nad4l	13174–13419	246		ATA	TAA
Intergene spacer 2	13420–13468	49			
tRNA-Leu$^{(UUR)}$	13469–13535	67	TAA 13500–13502		
tRNA-Trp 3	13558–13617	60	TCA 13588–13590		
Intergene spacer 3	13618–13701	84			
tRNA-Tyr	13702–13767	66	GTA 13731–13733		
cox2	133770–14459	690		ATG	TAA
tRNA-Pro	14475–14538	64	TGG 14505–14507		
cox3	14548–15333	786		TTA	TAA
tRNA-Trp 4	15357–15416	60	TCA 15387–15389		
tRNA-Ser$^{(UCN)}$	15439–15505	67	TGA 15470–15472		
tRNA-Pseudo 2	15508–15563	57	AT 15534–15535		

One strand is typically A and C rich and the other T and G rich, as is clearly demonstrated by the mt genome of *Triatoma*, and to a lesser extent by the lepidopsocid, in which the majority strand shows pronounced A and C skew and the minority strand significant T and G skew. In contrast, each of the three lice differs from this pattern. *Campanulotes* has pronounced reverse skew: strong T and G skew on the majority strand. *Heterodoxus* appears to be largely unskewed, although interestingly both the majority and minority strands have comparable A and G skews. This is unexpected as skews are usually complementary, i.e., an A skew on one strand is balanced by T skew on the other, as nucleotide biases have typically been attributed to strand-specific effects related to replication or transcription (Francino and Ochman 1997). Two models of mt genome replication have been proposed to explain the strand specificity of nucleotide biases. Under the strand-displacement model (Clayton 1982), replication of one strand lags the other. The lagging strand exists in a single-stranded state for much of the replication cycle and is exposed to potential mutations for a longer period than the leading strand, and so is expected to accumulate A and C nucleotides (Sancar and Sancar 1988; Reyes et al. 1998). In the strand-coupled model (Yang et al. 2002), the two strands remain double stranded at all times, and strand-specific nucleotide biases are attributed to the incorporation of RNA intermediates during the replication cycle. However, this model has only been demonstrated in mammals (Bogenhagen and Clayton 2003). These mutational effects are therefore likely to affect nucleotide skew on a strand-wide basis, so the absence of an asymmetric strand effect in *Heterodoxus* and *Bothriometopus* is surprising. Hassanin et al. (2005) attributed this lack of asymmetry in *Heterodoxus* to a recent reversal in the direction of transcription. Following a reversal, strand asymmetry should be eroded and eventually reversed entirely, as was demonstrated for several other invertebrate genomes with reversed skew (e.g., the mollusc *Katharina*). An absence of asymmetry was proposed to be a transient effect and in mt genomes lacking skew, insufficient time had passed to allow opposite asymmetry to become established. However, this does not explain the absence of skew in *Bothriometopus*, while a member of the same suborder, *Campanulotes*, shows such skew.

Other possibilities to explain the absence of skew in the mt genome are easily conceived. Genome rearrangements that include gene inversions are expected to homogenize any general pattern of asymmetry. Flipping a gene's transcription from one strand to another would result in our calculations in Table 3 being based on inappropriate groupings, i.e., some of the genes on the majority strand may retain nucleotide skew accumulated when the gene was located on the minority strand. This would suggest that the organization of the *Bothriometopus* genome, in which all the genes are transcribed on the same strand, is of evolutionarily recent origin. As inversions appear to be common in louse mt genome this is a potentially confounding factor in examinations of nucleotide skew. Calculating skew for each gene for the various louse species (Supplementary Table S1), the genes do not separate into two groups, one A and C skewed and the other T and G skewed. This pattern, if it occurred, could have been interpreted as demonstrating a division between genes historically located on one strand or the other, and the proper strand assignment now obscured by gene inversions and rearrangements (see Fonseca et al. 2006 for an example of this approach in amphibians). Of the *Bothriometopus* genes, only *cob* demonstrates both A and C

Table 3 Strand asymmetry in nucleotide composition at fourfold degenerate sites

	A (%)	G (%)	C (%)	T (%)	A-skew [a]	C-skew [b]
Bothriometopus	641 (41.3)	168 (10.8)	143 (9.2)	600 (38.6)	0.033	–0.08
Campanulotes	419 (27.9)	325 (21.6)	45 (3.0)	713 (47.5)	–0.260	–0.757
Heterodoxus Ma [c]	362 (47.1)	37 (4.8)	25 (3.3)	345 (44.9)	0.024	–0.194
Heterodoxus Mi [d]	220 (49.2)	27 (6.0)	15 (3.4)	185 (41.4)	0.086	–0.286
Lepidopsocid Ma [e]	410 (52.0)	18 (2.3)	45 (5.7)	315 (40.0)	0.131	0.429
Lepidopsocid Mi [f]	188 (43.6)	12 (2.8)	12 (2.8)	219 (50.8)	–0.076	0.0
Triatoma Ma [e]	551 (54.0)	34 (3.3)	182 (17.8)	254 (24.9)	0.369	0.685
Triatoma Mi [f]	87 (15.6)	51 (9.2)	17 (3.1)	401 (72.1)	–0.643	–0.5

[a] A skew = (A–T)/(A+T), where A and T are the numbers of each nucleotide. Negative values denote a skew to T

[b] C skew = (C–G)/(C+G), where G and C are the numbers of each nucleotide. Negative values denote a skew to G

[c] cox1, cox2, cox3, cob, nad1, nad2, nad6

[d] atp6, atp8, nad3, nad4, nad4l, nad5

[e] atp6, atp8, cox1, cox2, cox3, cob, nad2, nad3, nad6

[f] nad1, nad4, nad4l, nad5

skew, although it is extremely weak; *cox1* and *nad1* demonstrate both T and G skew but again this is not strong. In *Heterodoxus*, *nad1* and *nad6* demonstrate A and C skew, and *cox1* and *cob* T and G skew, but the G-skew of *cob* is the only strong skew in this genome. In contrast, in *Campanulotes*, all genes except *cox3* have pronounced T and G skew, and the A skew in *cox3* is marginal and its G skew is very strong. These results suggest that the genome-wide calculations (Table 3) are closely matched by calculations based on single genes (Table S1) and the absence of asymmetry in *Bothriometopus* and *Heterodoxus* is not the result of recent gene inversions that have homogenized any nucleotide biases between the strands.

A further factor that has been proposed to account for nucleotide skew is the process of transcription, which also results in the nontranscribed strand spending a significant period in a single-stranded state, with the same expected pattern of A and C biased mutation on the nontranscribed strand (Hassanin et al. 2005). Transcription related effects on one strand should nevertheless be counter-balanced by transcription on the other because the entire mt genome is transcribed in a single polycistronic mRNA (Ojala et al. 1980, 1981). Full transcription of each strand would expose each complementary strand in turn to a similar period in the single-stranded state. An exception would be genomes in which all genes are coded on one strand and have ceased transcription of the other strand as a result, such as *Bothriometopus*. It should be noted that several of the genomes which Hassanin et al. (2005) found to have T and G skew such as *Tigriopus* and *Katharina* and so concluded to have undergone reversals of the direction of replication also have all genes transcribed on a single strand and so the reversed asymmetry could alternatively be attributed to a transcription effect. Transcription in such a genome could either reinforce the strand asymmetry produced by replication, if the transcribed strand is also the leading strand in replication, or counterbalance it, if the transcribed strand is the lagging strand. Thus, the absence of strand asymmetry in *Bothriometopus* could be explained if replication of this mt genome proceeds with the noncoding strand leading. Transcription may also explain the strong T and G skews in the coding strand of *Campanulotes* because transcription-induced asymmetry would produce A and C skew on the nontranscribed strand. However, 36 of the 37 mt genes are transcribed on the same strand in *Campanulotes*. It would seem grossly inefficient to transcribe the entire minority strand of the *Campanulotes* mt genome simply for a single tRNA gene, but the signaling sites used to initiate transcription are unknown for lice and have not been studied in insects for any groups other than *Drosophila* (e.g. Matsushima et al. 2004, 2005; Roberti et al. 2003, 2006).

Under the consensus model of transcription, the initiation of transcription (IT) sites for each strand are located in the control region plus an additional IT located immediately upstream of the *rns* gene which in concert with a termination site (mTERM) located immediately downstream of the *nd1* gene in *Drosophila* or *rnl* in vertebrates (Taanman 1999) allows transcription of the rRNA genes independently of, and at a heightened rate relative to, the rest of the genome. Three transcripts are thus produced, a complete transcript of the entire mt genome for each strand plus one for the ribosomal gene block. Roberti et al. (2003, 2006) have recently proposed an alternative model of transcription in *Drosophila* in which each of the four major blocks of genes coded on the same strand have unique IT sites upstream of their coding region. They suggest that mTERM sites downstream of each coding block act as transcription attenuators, terminating some but not all transcription through these sites, producing a mix of partial and complete genome transcripts. This model accounts for the excess of sense over anti-sense mRNAs (Berthier et al. 1986) within mitochondria and suggests that transcription of just a targeted coding region is possible. A similar mechanism in *Campanulotes* could supply the single tRNA-Gln, which is coded on the minority strand, without the seemingly wasteful production of a transcript for the entire mt genome. The tRNA-Gln in *Campanulotes* is, however, not located adjacent to either of the two noncoding regions identified as potential control regions due to the presence of stem-loop structures (Covacin et al. 2006). This finding indicates that either the partial transcript is much larger than just the tRNA gene, the stem-loop structures are 3,035 and 8,325 bp from tRNA-Gln, respectively, or that other stem-loop structures within adjacent tRNAs or the *rns* gene function as IT sites in this genome.

Nucleotide skew and strand asymmetry in louse mt genomes thus probably reflect the wide range of genomic effects that can alter how the underlying patterns of mutation are reflected in the nucleotide composition of the genome. Reversals of the directions of replication, gene inversion, gene rearrangements and the pattern of transcription likely all contribute to complicated patterns of nucleotide skew. The sequencing of additional louse genomes will allow comparisons between much more closely related taxa and allow these factors to be teased apart. In conjunction with molecular dating analyses it could be possible to begin to determine the rates of substitution induced by each of these strand-specific mutational biases and thus infer which are most responsible for the overall shaping of mt genome nucleotide composition.

Transfer RNA Genes

Twenty-one of the 22 tRNA genes found in most metazoan mt genomes were found in *Bothriometopus* (Fig. 2). The

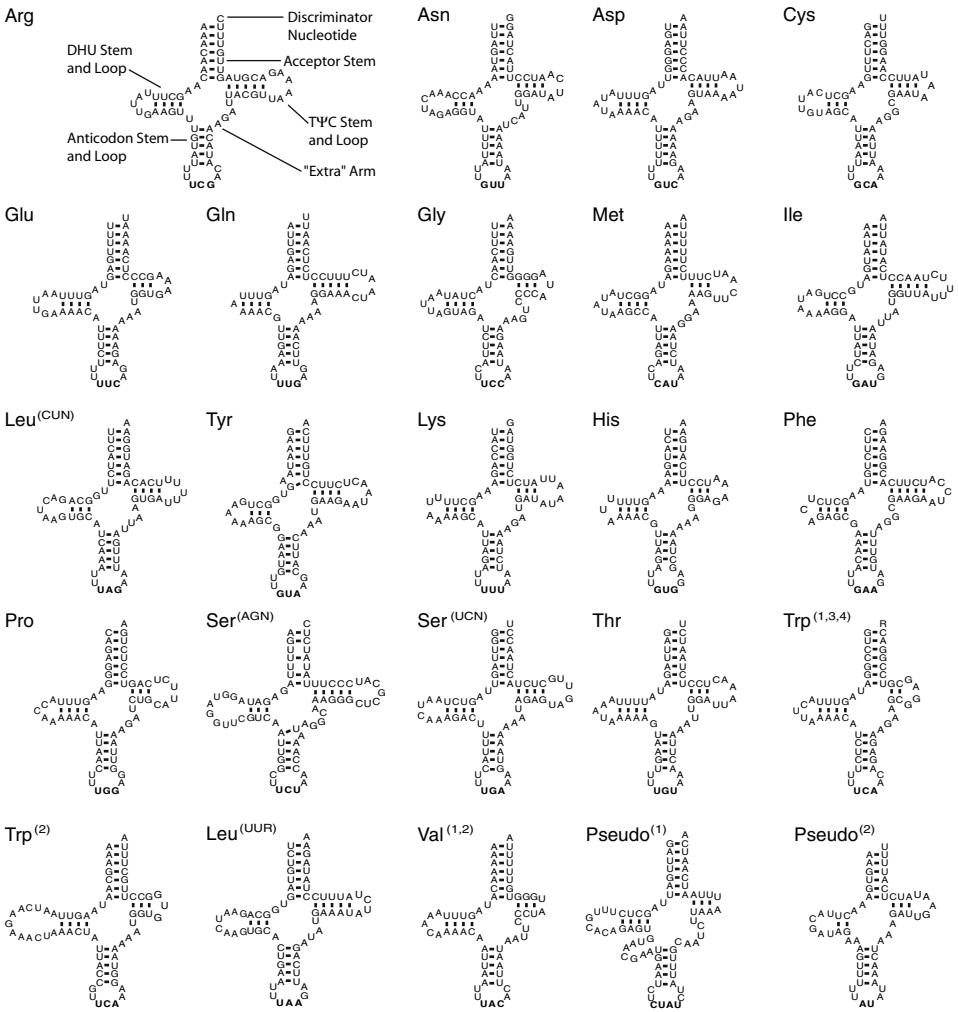

Fig. 2 tRNA secondary structures from *Bothriometopus*. Structural features are listed on tRNA-Arg at the top left. tRNAs are labelled with the abbreviations of their corresponding amino acids. Sequences are given 5′ to 3′ in the direction of transcription; anticodon sequences are in bold; hydrogen bonds indicated by a thick dash. Duplicated and pseudo-tRNAs are labelled as in the order they occur in the genome as for Fig. 1

tRNA-Ala was not found by tRNA-ScanSE and attempts to fold by hand a tRNA that coded for alanine from the nontranslated portions of the genome failed. In addition two tRNA-like pseudogenes were found between tRNA-Asn and tRNA-Thr and between tRNA-Ser$^{(UCN)}$ and *cox1*. In both regions a cruciform secondary structure resembling a tRNA could be found; however, neither could form a functional anticodon loop. The tRNA-Pseudo1 had an additional base in the anticodon loop, seven additional base pairs between the DHU and anticodon loops and two additional bases in the extra arm. The tRNA-Pseudo2 was missing one base from the anticodon loop and had a shortened DHU arm, two bases rather than the normal three or four. Due to the altered anticodon loops it is impossible to assign a tRNA isotype to these genes. Additionally, the stems of each did not show significant sequence similarity to any of the other tRNA genes, so it is impossible to determine homologies for these pseudogenes.

Nineteen of the 21 tRNA genes used the same anticodon sequences as are found in the majority of insect species. The tRNA-Lys and tRNA-Ser$^{(AGN)}$ had the anticodons UUU and UCU rather than the usual CUU and GCU respectively. In *Campanulotes* these same tRNAs have the same divergent anticodon sequences as *Bothriometopus* and the remaining 20 tRNAs possess the same anticodon sequences as most insects. The same divergent anticodons are also found for tRNA-Lys in the *Heterodoxus* and *Pediculus* mt genomes; the remaining tRNAs have the insect ancestral sequence (Shao et al. 2001; Pittendrigh et al., unpublished data). These changes are consistent with the phylogeny of lice: tRNA-Lys (CUU → UUU) occurred in the common ancestor of Amblycera and Ischnocera and tRNA-Ser$^{(AGN)}$ (GCU → UCU) in the common ancestor of Ischnocera (Johnson et al. 2004). Such changes may constitute "rare genomic changes" which are diagnostic for the respective groups (Rokas and Holland 2000).

While there are four copies of tRNA-Trp in the *Bothriometopus* mt genome, there are only two types. Copies 1, 3, and 4 (relative gene order as per Fig. 1) share almost identical sequences differing only at the 3′ discriminator nucleotide (an A in copies 1 and 3 vs a G in copy 4). Copy 2 differs significantly from the other copies. Both copies of

tRNA-Val have identical sequences. A phylogeny of the tRNA genes was constructed to test to homology of the two tRNA-Trp types. This approach has previously been used to identify the homology of duplicated tRNAs in wasp (Dowton and Austin 1999) and sponge mt genomes (Lavrov and Lang 2005b). These trees indicated homology between the *Bothriometopus* tRNA-Trp2 and the tRNA-Trp genes from *Triatoma* and the lepidopsocid and between *Bothriometopus* tRNA-Trp1,3,4 and *Bothriometopus* tRNA-Glu. This suggests that the duplicated copies of tRNA-Trp in the *Bothriometopus* mt genome were ultimately derived from tRNA-Glu by tRNA remolding (Higgs et al. 2003; Rawlings et al. 2003) or tRNA recruitment (Lavrov and Lang 2005b) as substitutions to the anticodon loop which change the apparent identity of a tRNA gene, have been variously termed. This result is somewhat unexpected as it would require, minimally, the anticodon mutations UUC → UCA. Two anticodon substitutions seems excessive given how conservative the anticodon loop is across insect mt tRNAs (i.e., 19 of 22 genes with the identical sequences across insects and the only recorded differences in the first or wobble nucleotide). This suggests that the *Bothriometopus* tRNA-Trp1,3,4 genes are probably nonfunctional but this would be difficult to test.

These are the first instances of tRNA duplications in a louse mt genome although this has been found previously in calliphorid blowflies (Lessinger et al. 2004) and in Hymenoptera (Dowton et al. 2003; Castro et al. 2006). In both these examples, however, the tRNAs are tandemly duplicated and the copies are thus physically closely located. In *Bothriometopus* they are distributed throughout the genome; 4,916 bp separate tRNA-Trp1 and tRNA-Trp3 1,740 bp tRNA-Trp3 and tRNA-Trp4, and 8,731 bp tRNA-Trp4 and tRNA-Trp1. The two copies of tRNA-Val are separated by 2,606 bp. Sequence similarity between the different copies of these tRNAs may be due to their recent duplication or may be maintained over longer time frames by concerted evolution. Under the widely accepted duplication-loss models of genome rearrangement (Moritz and Brown 1987; Macey et al. 1997) it would be unusual to find such large distances between the duplicated genes (cf. the local tRNA duplications in wasps found by Dowton et al. (2003)) as it would be considered unlikely that the loss of the intervening genes would occur faster than point mutations would accumulate within the duplicated tRNAs. The existence of widely separated tRNAs with identical sequences could thus be evidence that the duplication occurred through a recombination mediated process which could move portions of the mt genome to any other point within the genome (Dowton and Campbell 2001). Alternatively sequence similarity between duplicated tRNAs, having arisen through a duplication-loss process, could be maintained through concerted evolution, itself a recombination mediated process. Concerted evolution of nontranslated portions of the mt genome has been frequently recorded (e.g., Shao et al. 2004; Ogoh and Ohmiya 2007) but confined to duplicated control regions. It has been proposed that genomes with two control regions are able to replicate/transcribe faster than those with a single control region. If this is the case, concerted evolution would likely be necessary so that both control regions are recognizable to the replication or transcription machinery (Shao and Barker 2003). Concerted evolution between duplicated tRNAs has not previously been recorded, however tRNAs have been implicated as signaling sites for mt genome replication in vertebrates and such a role could drive their concerted evolution (Taanman 1999).

Protein-Coding Genes

Eleven of the 13 protein-coding genes were identified by comparison with other insect mt genomes on GenBank using Blastx searches. The genes *atp6* and *nad4l* were identified by comparison of amino acid sequence similarity and of hydrophobicity plots between putative open reading frame regions and copies of these genes that have been identified from the mt genomes of related insects. Codon composition of the protein-coding genes was similar to that found in other louse and hemipteroid mt genomes (Table 4). In the *Bothriometopus* mt genome, 14% of codons were G+C rich and 44.8% A+T rich. A+T rich codons are defined as those with an A or a T at the first and second codon position (Asn, Ile, Leu$^{(UUR)}$, Lys, Met, Phe, Tyr); similarly G+C rich codons are those with a G or a C in the first and second positions(Ala, Arg, Gly, Pro) (Foster et al. 1997). This is marginally lower than that of other lice: *Campanulotes* 13.6% G+C rich and 46.4% A+T rich; *Heterodoxus* 11.1% G+C rich and 55.5% A+T rich; and members of closely related orders such as Psocoptera, 13.8% G+C rich and 56% A+T rich, and *Triatoma*, 15.7% G+C rich and 45% A+T rich. This is consistent with the overall nucleotide composition and bias of the *Bothriometopus* mt genome, which was less extreme than in other closely related species.

Recently, Kim et al. (2006) and Fenn et al. (2007) have examined mt protein-coding genes for potential secondary structures that may serve as recognition sites for mRNA cleavage. In the ancestral insect mt genome arrangement, 9 of the 13 protein-coding genes are flanked on their 3′ end by tRNA genes. Ojala et al. (1980, 1981) have proposed that endonucleasis of the polycistronic pre-mRNAs at the tRNA stem liberates mature mRNAs for the protein-coding genes. The remaining four protein-coding genes are flanked at their 3′ end by other protein-coding genes and the reading frames of the two proteins frequently overlaps.

Table 4 Amino acid percentage usage reported for each gene in the *Bothriometopus* (B) mt genome. Total percentage amino acid usage is compared against *Campanulotes* (C), *Heterodoxus* (H), lepidopsocid RS-2001 (L), and *Triatoma* (T)

	B														C	H	L	T
	atp6	*atp8*	*cox1*	*cox2*	*cox3*	*cob*	*nad1*	*nad2*	*nad3*	*nad4*	*nad4l*	*nad5*	*nad6*	Total				
Ala	3.06	3.45	4.89	2.61	2.29	3.28	3.24	4.46	3.39	4.83	1.22	3.68	0.00	3.54	2.65	2.64	3.81	4.92
Arg	0.87	0.00	1.57	2.17	1.91	1.09	2.27	0.30	0.00	0.69	1.22	0.92	0.61	1.15	1.09	1.29	1.49	1.49
Asn	2.62	1.72	3.13	4.78	3.05	4.92	0.97	3.27	2.54	2.07	1.22	2.94	1.83	2.91	2.27	4.48	6.30	4.39
Asp	0.87	3.45	2.35	3.48	2.29	2.46	1.62	0.89	1.69	0.69	0.00	2.57	1.83	1.89	1.94	1.76	1.84	2.06
Cys	2.18	0.00	1.17	1.74	1.91	0.00	0.65	1.19	0.85	1.38	1.22	1.65	0.61	1.21	1.47	0.79	0.53	0.97
Gln	1.75	5.17	0.98	2.61	2.29	2.19	1.94	1.19	0.85	0.92	0.00	1.29	1.22	1.54	1.17	1.25	2.23	1.89
Glu	3.49	5.17	1.57	6.09	3.44	1.64	3.88	2.98	8.47	2.07	2.44	2.76	3.66	3.07	3.17	2.68	2.06	2.55
Gly	4.80	1.72	9.00	4.78	6.11	6.83	6.80	4.76	4.24	4.14	2.44	4.41	5.49	5.63	6.80	5.27	5.17	5.93
His	1.75	0.00	3.13	2.61	3.82	2.46	0.00	1.79	3.39	1.38	1.22	1.29	0.00	1.89	1.67	2.17	2.45	2.77
Ile	9.61	13.79	9.78	11.30	8.02	9.84	11.97	8.93	15.25	12.41	12.20	10.11	12.20	10.62	6.69	9.94	13.22	11.33
Leu	15.72	5.17	13.11	10.00	12.21	13.11	11.97	15.18	11.02	13.56	15.85	11.40	16.46	12.93	12.21	14.93	14.19	13.04
Lys	3.93	0.00	1.76	2.61	1.91	3.01	3.56	4.76	1.69	4.60	2.44	4.78	6.71	3.51	3.90	3.79	2.80	2.72
Met	7.42	10.34	6.46	5.22	4.58	4.92	7.77	9.23	12.71	9.43	12.20	7.35	9.15	7.52	6.31	7.30	6.13	7.02
Phe	8.30	13.79	9.00	5.22	13.36	10.11	10.03	9.23	7.63	9.20	9.76	10.85	10.98	9.69	14.23	10.77	9.63	7.33
Pro	4.80	5.17	4.89	4.78	3.05	4.64	3.56	3.27	2.54	3.45	2.44	2.39	2.44	3.68	3.14	3.28	4.33	4.39
Ser	15.72	12.07	10.37	11.74	11.07	10.11	12.94	11.01	9.32	17.01	13.41	14.15	12.80	12.62	12.48	10.36	8.71	9.09
Thr	2.62	3.45	4.70	3.48	3.05	4.64	2.27	2.68	3.39	2.07	4.88	4.23	4.88	3.54	2.32	3.65	5.39	6.50
Trp	3.06	0.00	3.33	2.61	3.82	3.01	2.27	2.98	3.39	1.38	0.00	2.94	1.83	2.66	2.79	2.68	2.85	3.03
Tyr	1.75	5.17	2.94	3.48	2.67	3.55	2.91	3.27	0.85	3.45	1.22	3.68	1.83	3.02	3.19	4.62	3.15	3.38
Val	5.24	8.62	5.68	8.26	8.78	7.92	9.39	8.33	5.93	5.06	13.41	6.43	4.88	7.05	10.16	6.01	3.33	5.05

Fenn et al. (2007) have demonstrated stem-loop structures at each of these gene boundaries that may act as binding sites for endonucleases. Highly rearranged mt genomes such as those of lice similarly contain gene boundaries between protein-coding genes for which secondary structure-mediated cleavage may play a role in mRNA processing. Several of these gene boundaries are retained from the ancestral insect mt genome arrangement, e.g., *atp8-atp6* found in each of the three louse species, but others are novel gene boundaries not found in other groups, e.g., *nad5-nad1* in *Bothriometopus*. Correct processing of these polycistronic pre-mRNAs would be necessary for the correct translation into mature proteins so we examined the potential of each protein-coding gene to form similar stem-loop structures.

Stem-loop structures were inferred at the 3′ end of each of the protein-coding gene to protein-coding gene boundaries in each of the three species of lice (Fig. 3): *nad5–nad1* and *atp8–atp6* in *Bothriometopus*; *nad6–rns*, *cox1–cob–cox2*, *atp8–atp6* and *cox3–nad4* in *Campanulotes*; and *nad1–cob*, *nad4–nad4l* and *atp6–atp8* in *Heterodoxus*. There does not appear to be a consistent pattern to these secondary structures. The stem portion varied from as little as 4 (*nad1* in *Heterodoxus*) to as many as 23 (*atp8* in *Campanulotes*) paired bases. The loop portion varied from as few as 3 (*atp8* in *Bothriometopus*) to 64 (*nad6* in *Campanulotes*) nucleotides in size. Finally the position of the stop codon varied from entirely within the stem (*nad1* in *Heterodoxus*) to two bases down stream of the stem (*atp8* in *Bothriometopus*). There was even large variability between structures inferred for *atp8*, a gene which abuts *atp6* in each of the three louse mt genomes, and which one would expect to demonstrate some degree of similarity due to gene homology. The stems inferred for *atp8* varied from 9 bases long in *Heterodoxus* to 23 in *Campanulotes* and the loops from 3 to 14 bp in size. This variability is however consistent with the low level of amino acid similarity in the 3′ end of the translated *atp8* protein across lice. Such variability confounds attempts to discern what consistent features within the secondary structures are recognized by the endonucleases. The endonucleasis of mt pre-tRNAs is mediated by two enzymes: RNase P which cleaves the 5′ end of the molecule and RNase Z which cleaves the 3′ (Dubrovsky et al. 2004). The consistent clover-leaf secondary structure of tRNAs plays a role in this specificity (Frank and Pace 1998). Because the stem-loop region includes a coding portion of the mRNA, it is probable that RNase P does not play a role in the processing of the mt pre-mRNAs. RNase P is very sensitive to even single base substitutions in the stems of substrate tRNAs (Levinger et al. 1995, 1998), let alone the huge variations in stem length which we found inferred here for the mRNAs, which

is consistent with RNase P not recognizing these structures. RNase Z is less sensitive to substrate variations (Levinger et al. 1995, 1998). However, it is unclear whether this enzyme or another endonuclease is responsible for RNA processing of the protein-coding genes, as enzymatic activity of this type has not previously been investigated for even model insect species such as *Drosophila*, let alone lice.

Messenger RNA processing is also a potential limitation on mt genome rearrangements. The absence of an appropriate stem-loop structure in some of the protein-coding genes could result in mt genome arrangements in which those genes that are not flanked by tRNA genes are impossible to process by the post-transcriptional machinery. Such genome arrangements could therefore be lethal. Under such a scenario, protein-coding genes could only rearrange as units with their flanking tRNA genes. This however is not supported by the genome arrangements of the three louse species in which there is little conservation of the gene boundaries between protein-coding genes and tRNAs. Of the nine protein-coding gene to protein-coding gene boundaries described from lice (Fig. 3), only four are found in other insect groups. The remaining genes have either developed novel stem-loop structures as a compensatory mechanism for the absence of a 3′ tRNA gene or such structures are widespread even in protein-coding genes that are flanked by tRNA genes. Analysis of the 13 protein-coding genes within *Bothriometopus* found stem-loop structures at the 3′ end of all mt genes except *cox1* and *nad4* which were similar to those found in the protein-coding gene flanked genes. In contrast, analysis of the protein-coding genes of *Triatoma*, whose mt genome retains the ancestral insect arrangement, revealed stem-loops in the 4 genes flanked at their 3′end by protein-coding genes as expected, but of those which are flanked by tRNAs only *cox1* and *cob* possessed such structures. These comparisons suggest that the development of 3′ end stem-loops in lice may have provided these genes with the potential to rearrange, because genome arrangements that separated them from 3′ flanking tRNA genes would no longer result in a loss of proper mRNA processing in these genomes. These phenomena, however, require much

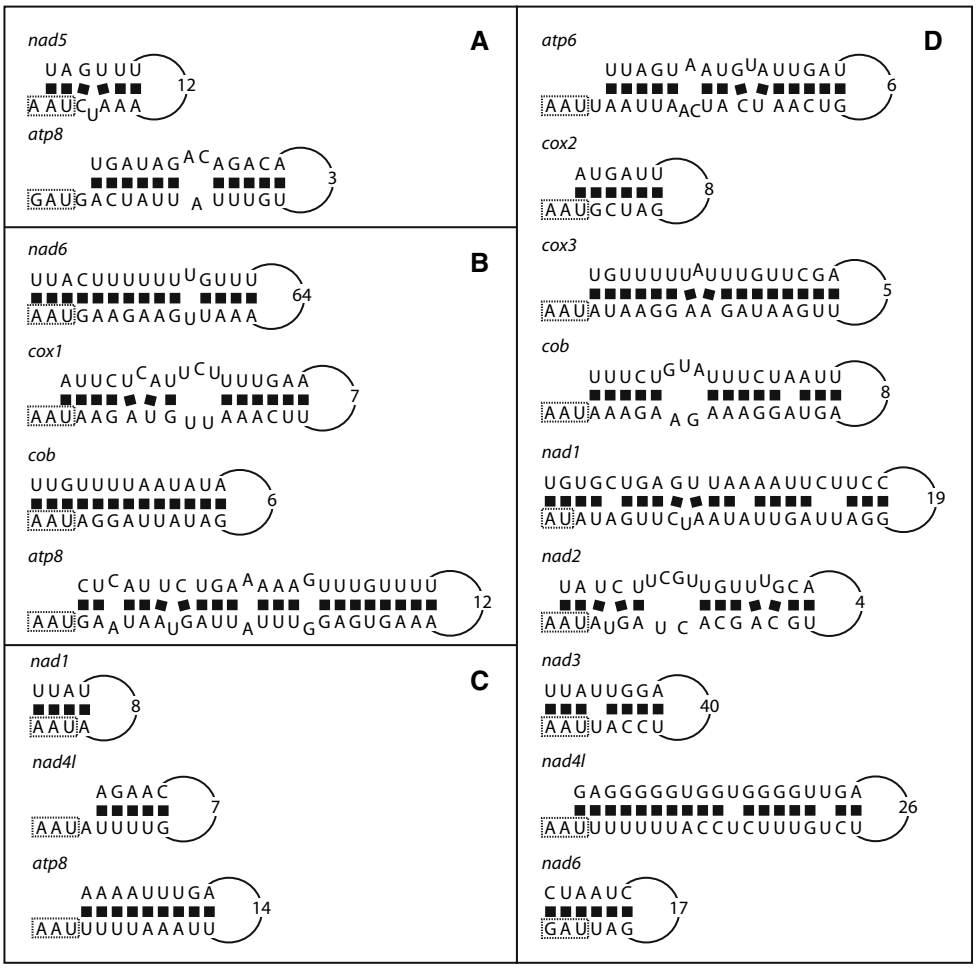

Fig. 3 mRNA secondary structure stem-loops found at the 3′ end of protein-coding genes. A: Protein-coding genes abutting protein-coding genes in *Bothriometopus*. B: Protein-coding genes abutting protein-coding genes in *Campanulotes*. C: Protein-coding genes abutting protein-coding genes in *Heterodoxus*. D: Protein-coding genes abutting tRNA genes in *Bothriometopus*. Sequences are given 5′ to 3′ upper sequence looping to the lower; hydrogen bonds are indicated by a thick dash; the number of bases in the loop region is indicated on the right hand end of each structure; the stop codon is indicated by a dashed box

additional investigation. Until the secondary structures targeted by endonucleases and the specificity of those endonucleases are better understood, it will be impossible to discriminate between genuine structures involved in signaling and random structures without any evolutionary or genomic significance.

Putative Control Regions

The control region of the *Bothriometopus* mt genome is difficult to discern due to the larger than normal number of intergenic spacer regions. Each of the three large intergenic spacer regions, plus each of the nontranslated repeat regions can form the stem-loop structures typically associated with the origins of replication/transcription (Taanman 1999; Saito et al. 2005) (Fig. 4). Stem-loop 1, between tRNA-Glu and tRNA-Val1, is located 10 bp into the spacer region and consists of an 11 bp stem and 8 bp loop. Stem-loop 2, between *nad4l* and tRNA-Leu$^{(UUR)}$ and located 22 bp into the spacer, has an 11 bp stem and 3 bp loop. Two stem-loop structures were found between tRNA-Trp4 and tRNA-Tyr each consisting of an 8 bp stem and 3 bp loop, the first located 5 bp into the spacer and the second 6 bp from the end of the spacer. There is only limited sequence similarity between any of these stem-loop structures and if they do represent duplicated control regions then they are not undergoing concerted evolution.

Additionally each of the nontranslated repeat regions are also capable of forming stem-loop structures. Nontranslated repeat region B, between tRNA-Thr and tRNA-Ser$^{(AGN)}$, is the larger of the two and has a more complete secondary structure. The 5′ nonrepetitive region forms a complicated structure (Fig. 4) with a 7 bp stem between the 5′ and 3′ ends (stem 1), and a large loop comprising most of the region from which two additional stem-loops are formed (stems 2 and 3). The largest portion of the loop, the 13 bp between stem 3 and stem 1, is primarily composed of an 11 bp poly-A stretch. The 5′ nonrepetitive region is 27 bp shorter in nontranslated repeat region A than in nontranslated repeat region B and this corresponds to a loss of the 5′ sides of stems 1 and 2 and stem 3 is the only stem-loop which can be formed. The repeat units within each nontranslated repeat region also form stem-loop structures a long 18 bp stem (stem 4) and short 6 bp stem (stem 5). The two complete repeat units each consists of both stems 4 and 5, the partial repeat unit B3 consists only of stem 4 and the partial repeat unit A3 is 14bp shorter than B3 which corresponds to a loss of all but the uppermost 6 bp of stem 4. The second repeat unit of both nontranslated repeat regions, repeats A2 and B2, has a single substitution A → T in the 3 bp loop of stem 4. This suggests that the sequence identity between corresponding repeat units, A1 to B1 etc, is due to recent duplication of the entire nontranslated repeat region rather than duplication of repeat units independently within each location. Further the

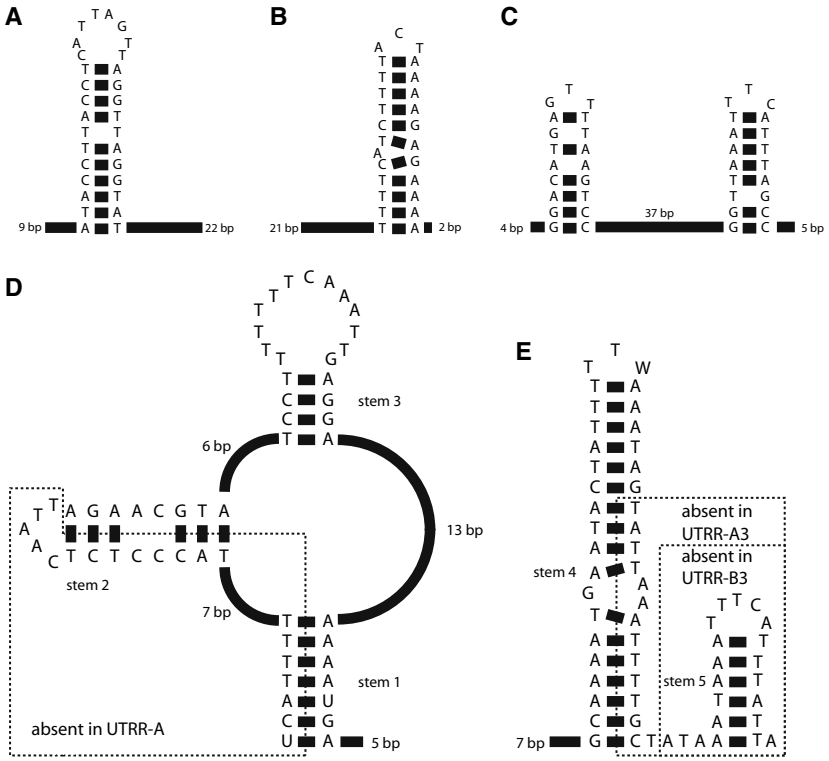

Fig. 4 Putative control regions within the mitochondrial genome of *Bothriometopus*. A: Intergenic spacer 1, between tRNA-Glu and tRNA-Val1. B: Intergenic spacer 2, between *nad4l* and tRNA-Leu$^{(UUR)}$. C: Intergenic spacer 3, between tRNA-Trp3 and tRNA-Tyr. D: 5′ Nonrepetitive section of the nontranslated repetitive regions (NTRR). E: Repeat unit of the nontranslated repeat regions. Unpaired, nonstem regions are indicated by a black line with the size in bp listed. Differences between the partial and complete repeat units, and between the two NTRRs are indicated by box delimiting the regions that are absent in certain copies of the repeats

absence of 27 bp at the 5′ end and 14 bp at the 3′ end in nontranslated repeat region A, which correspond to portions of stem-loops found in nontranslated repeat region B, suggests that the nontranslated repeat region A is an imperfect copy of nontranslated repeat region B, rather than that nontranslated repeat region B was formed by a duplication of nontranslated repeat region A into an existing intergenic spacer region and that the additional bases in nontranslated repeat region B are unrelated to its structure or potential function. The structure of the nontranslated repeat regions is the most similar to that found in the control regions of other insects (Lewis et al. 1994; Zhang and Hewitt 1997) in that it is composed of a stem-loop structure followed by a poly-A or poly-T stretch (the 5′ nonrepetitive region) followed by a series of large repeat units with internal secondary structures (the repetitive region). In contrast, the putative control regions found in other lice consist simply of a stem-loop structure without the poly-A/ T region or repetitive sequence regions.

There are thus five potential control regions in the nontranslated portions of the mt genome, plus the two sets of duplicated tRNAs which appear to be undergoing concerted evolution. Any one of these could potentially be an origin of replication, the initiation or termination sites for transcription or these roles could be independent and occur at different sites in the mt genome, or each of these functions could be simultaneously performed by multiple sites. The mt genome's origin of replication has only been experimentally mapped for a small number of insects (Goddard and Woolstenholme 1980; Saito et al. 2005), but no louse species or close relative has been examined. While stem-loop structures and associated poly-A/T regions are a common feature of the origin of replication, the size of the structures varies considerably and the poly-A/T region, located variously up or down stream of the stem-loop, forms the loop region itself or is absent altogether in Orthoptera (Saito et al. 2005). Given such variability, inferences based on sequence analysis alone cannot be regarded as completely accurate. Previously all major noncoding portions of the louse mt genomes have been interpreted as potential control regions. *Heterodoxus* was posited to have two control regions and *Campanulotes* three (Shao et al. 2001; Covacin et al. 2006). Following these precedents *Bothriometopus* would be interpreted to have five. However, there does not appear to be much justification for such functional interpretations of the potential to form secondary structures, particularly as secondary structures are almost ubiquitous in louse mt genomes occurring in most protein-coding genes, all tRNA and rRNA genes, as well as most nontranslated regions. Additional studies, both functional (Saito et al. 2005) and comparative (Brehm et al. 2001), will be necessary to properly interpret the putative mt control regions of lice.

Conclusion

Louse mt genomes display amazing variability in genome arrangements, nucleotide composition, biases and skew, tRNA duplication and evolution, the mRNA structure of the protein-coding genes, and putative control regions. The degree to which phenomena such as strand-specific nucleotide skew or tRNA anticodon sequences, which over the remainder of the insects do not vary, show variability within lice is remarkable. Lice therefore have the potential to illuminate many aspects of mt genome biology, and the major impediment to this understanding is a lack of mt genomes covering the phylogenetic diversity of lice. Such additional data, coupled with the understanding of the nuclear encoded, mitochondrially functional genes that will be made available by the impending release of the nuclear genome of the human body louse, *Pediuclus humanus*, (Pittendrigh et al. 2006) will make lice a potent model system for understanding the evolution of the mt genome.

Acknowledgments Kim Baer assisted with the analyses and Terri Weese with revision of earlier drafts. This study was supported by NSF grant DEB0444972.

References

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Res 25:3389–3402

Baumann L, Baumann P (2005) Cospeciation between the primary endosymbionts of mealybugs and their hosts. Curr Microbiol 50:84–87

Berthier F, Renaud M, Alziari S, Durand R (1986) RNA mapping on *Drosophila* mitochondrial DNA: Precursors and template strands. Nucleic Acids Res 14:4519–5433

Boore JL (1999) Animal mitochondrial genomes. Nucleic Acids Res 27:1767–1780

Boore JL, Brown WM (2000) Mitochondrial genomes of *Galatheal-inum*, *Helobdella*, and *Platynereis*: Sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol Biol Evol 17:87–106

Brehm A, Harris DJ, Hernández M, Cabrera VM, Larruga JM, Pinto FM, González AM (2001) Structure and evolution of the mitochondrial DNA complete control region in the *Drosophila subobscura* subgroup. Insect Mol Biol 10:573–578

Bogenhagen DF, Clayton DA (2003) The mitochondrial DNA replication bubble has not burst. Trends Biochem Sci 28:357–360

Brugler MR, France SC (2007) The complete mitochondrial genome of the black coral *Chrysopathes Formosa* (Cnidaria: Anthozoa: Antipatharia) supports classification of antipatharians within the subclass Hexacorallia. Mol Phylogen Evol 42:776–788

Castro LR, Austin AD, Dowton M (2002) Contrasting rates of mitochondrial molecular evolution in parasitic Diptera and Hymenoptera. Mol Biol Evol 19:1110–1113

Castro LR, Ruberu K, Dowton M (2006) Mitochondrial genomes of *Vanhornia eucnemidarum* (Apocrita: Vanhorniidae) and

Primeuchroeus spp. (Aculeata: Chrysididae): Evidence of rearranged mitochondrial genomes within the Apocrita (Insecta: Hymenoptera). Genome 49:752–766

Clayton DA (1982) Replication of animal mitochondrial DNA. Cell 28:693–705

Cameron SL, Beckenbach AT, Dowton M, Whiting MF (2006) Evidence from mitochondrial genomics on interordinal relationships in insects. Arthrop System Phylogen 64:27–34

Cameron SL, Whiting MF (2007) Mitochondrial genome comparisons of the subterranean termites from the genus *Reticulitermes* (Insecta: Isoptera: Rhinotermitidae). Genome 50:188–202

Covacin C, Shao R, Cameron S, Barker SC (2006) Extraordinary number of gene rearrangements in the mitochondrial genomes of lice (Phthiraptera: Insecta). Insect Mol Biol 15:63–68

Dowton M, Austin AD (1999) Evolutionary dynamics of a mitochondrial rearrangement "hot spot" in the Hymenoptera. Mol Biol Evol 16:298–309

Dowton M, Campbell NJ (2001) Intramitochondrial recombination – is it why some mitochondrial genes sleep around. Trends Ecol Evol 16:269–271

Dowton M, Castro LR, Austin AD (2002) Mitochondrial gene rearrangements as phylogenetic characters in the invertebrates: The examination of genome "morphology". Invert Syst 16:345–356

Dowton M, Castro LR, Campbell SL, Bargon SD, Austin AD (2003) Frequent mitochondrial gene rearrangements at the hymenoptera nd3-nd5 junction. J Mol Evol 56:517–526

Drummond AJ, Kearse M, Heled J, Moir R, Thierer T, Ashton B, Wilson A, Stones-Havas S (2006) Geneious ver2.5; available at http://www.geneious.com/

Dubrovsky EB, Dubrovskaya VA, Levinger L, Schiffer S, Marchfelder A (2004) *Drosophila* RNase Z processes mitochondrial and nuclear pre-tRNA 3′ ends *in vivo*. Nucleic Acids Res 32:255–262

Faith JJ, Pollock DD (2003) Likelihood analysis of asymmetrical mutation bias gradients in vertebrate mitochondrial genomics. Genetics 165:735–745

Fenn JD, Cameron SL, Whiting MF (2007) The complete mitochondrial genome sequence of the Mormon cricket (*Anabrus simplex*: Tettigoniidae: Orthoptera) and an analysis of control region variability. Insect Mol Biol 16:239–252

Fonseca MM, Froufe E, Harris DJ (2006) Mitochondrial gene rearrangements and partial genome duplications detected by multigene asymmetric compositional bias analysis. J Mol Evol 63:654–661

Foster PG, Jermiin LS, Hickey DA (1997) Nucleotide composition bias effects amino acid content in proteins coded by animal mitochondria. J Mol Evol 44:282–288

Francino MP, Ochman H (1997) Strand asymmetries in DNA evolution. Trends Genetics 13:240–245

Frank DN, Pace NR (1998) Ribonuclease P: Unity and diversity of a tRNA processing ribozyme. Ann Rev Biochem 67:153–180

Goddard JM, Wolstenholme DR (1980) Origin and direction of replication in mitochondrial DNA molecules from genus *Drosophila*. Nucleic Acids Res 8:741–757

Hafner MS, Sudman PD, Villablance FX, Spradling TA, Demaster JW, Nadler SA (1994) Disparate rates of molecular evolution in cospeciating hosts and parasites. Science 365:1087–1090

Hassanin A, Léger N, Deutsch J (2005) Evidence for multiple reversals of asymmetric mutational constraints during the evolution of the mitochondrial genome of Metazoa, and consequences for phylogenetic inferences. Syst Biol 54:277–298

He Y, Jones J, Armstrong M, Lamberti F, Moens M (2005) The mitochondrial genome of *Xiphinema americana* sensu stricto (Nematoda: Enoplea): Considerable economization in the length and structural features of encoded genes. J Mol Evol 61:819–833

Helfenbein K, Brown WM, Boore JL (2001) The complete mitochondrial genome of the articulate brachiopod *Terebratalia transversa*. Mol Biol Evol 18:1734–1744

Higgs PG, Jameson D, Jow H, Rattay M (2003) The evolution of tRNA-Leu genes in animal mitochondrial genomes. J Mol Evol 57:435–445

Hoffmann RJ, Boore JL, Brown WM (1992) A novel mitochondrial genome organization for the blue mussel, *Mytilus edulis*. Genetics 131:397–412

Johnson KP, Yoshizawa K, Smith VS (2004) Multiple origins of parasitism in lice. Proc Roy Soc London B 271:1771–1776

Kim I, Lee EM, Seol KY, Yun EY, Lee YB, Hwang JS, Jin BR (2006) The mitochondrial genome of the Korean hairstreak, *Coreana raphaelis*, (Lepidoptera: Lycaenidae). Insect Mol Biol 15:217–225

Kimura M (1983) The Neutral Theory of Molecular Evolution. Cambridge University Press, Cambridge

Kumar S, Tamura K, Nei M (2004) MEGA3: Integrated software for molecular evolutionary genetics analysis and sequence alignment. Brief Bioinform 5:150–163

Lavrov DV, Brown WM (2001) *Trichinella spiralis* mtDNA: A nematode mitochondrial genome that encodes a putative ATP8 and normally structured tRNAs and has a gene rearrangement relatable to those of coelomate metazoans. Genetics 157:621–637

Lavrov DV, Lang BF (2005a) Poriferan mtDNA and animal phylogeny based on mitochondrial gene arrangements. Syst Biol 54:651–659

Lavrov DV, Lang BF (2005b) Transfer RNA gene recruitment in mitochondrial DNA. Trends Genetics 21:130–133

Lessinger AC, Junqueira ACM, Conte FF, Azeredo-Espin AML (2004) Analysis of a conserved duplicated tRNA gene in the mitochondrial genome of blowflies. Gene 339:1–6

Levinger L, Vasisht V, Greene V, Bourne R, Birk A, Kolla S (1995) Sequence and structure requirements for *Drosophila* tRNA 5′- and 3′- end processing. J Biol Chem 270:18903–18909

Lewis DL, Farr CL, Farquhar AM, Kaguni LS (1994) Sequence, organisation and evolution of the A+T region of *Drosophila melanogaster* mitochondrial DNA. Mol Biol Evol 11:523–538

Levinger L, Bourne R, Kolla S, Cylin E, Russell K, Wang X, Mohan A (1998) Matrices of paired substitutions show the effects of tRNA D/T loop sequence on *Drosophila* RNase P and 3′-tRNase processing. J Biol Chem 273:1015–1025

Littlewood DTJ, Lockyer AE, Webster BL, Johnston DA, Le TH (2006) The complete mitochondrial genomes of *Schistoma haematobium* and *Schistosoma spindale* and the evolutionary history of mitochondrial genome changes among parasitic flatworms. Mol Phylogen Evol 39:452–467

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

Macey JR, Larson A, Ananjeva NB, Papenfuss TJ (1997) Replication slippage may cause parallel evolution in the secondary structures of mitochondrial transfer RNAs. Mol Biol Evol 14:30–39

Machida RJ, Miya MU, Nishida M, Nishida S (2002) Complete mitochondrial DNA sequence of *Tigriopus japonicus* (Crustacea: Copepoda). Mar Biotechnol 4:406–417

Matsushima Y, Garesse R, Kaguni LS (2004) *Drosophila* mitochondrial transcription factor B2 regulates mitochondrial DNA copy number and transcription in Schneider cells. J Biol Chem 279:26900–26905

Matsushima Y, Adán C, Garesse R, Kaguni LS (2005) *Drosophila* mitochondrial transcription factor B1 modulates mitochondrial translation but not transcription or DNA copy number in Schneider cells. J Biol Chem 280:16815–16820

Medina M, Collins AG, Takaoka TL, Kuehl JV, Boore JL (2006) Naked corals: Skeleton loss in Scleractinia. Proc Nat Acad Sci USA 103:9096–9100

Mindell DP, Sorenson MD, Dimcheff DE (1998) Multiple independent origins of mitochondrial gene order in birds. Proc Nat Acad Sci USA 95:10693–10697

Moritz C, Brown WM (1987) Tandem duplications in animal mitochondrial DNAs: Variation in incidence and gene content among lizards. Proc Natl Acad Sci USA 84:7183–7187

Mueller RL, Boore JL (2005) Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders. Mol Biol Evol 22:2104–2112

Ogoh K, Ohmiya Y (2007) Concerted evolution of duplicated control regions within an ostracod mitochondrial genome. Mol Biol Evol 24:74–78

Ojala D, Merkel C, Gelfand R, Attardi G (1980) The tRNA genes punctuate the reading of genetic information in human mitochondrial DNA. Cell 22:393–403

Ojala D, Montoyo J, Attardi G (1981) tRNA punctuation model of RNA processing in human mitochondria. Nature 290:470–474

Perna NT, Kocher TD (1995) Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J Mol Evol 41:353–358

Pittendrigh BR, Clark JM, Johnston JS, Lee SH, Romero-Severson J, Dasch GA (2006) Sequencing of a new target genome: the *Pediculus humanus humanus* (Phthiraptera: Pediculidae) genome project. J Med Entomol 43:1103–1111

Rawlings TA, Collins TM, Bieler R (2003) Changing identities: tRNA duplication and remolding within animal mitochondrial genomes. Proc Natl Acad Sci USA 100:15700–15705

Reyes A, Gissi C, Pesole G, Saccone C (1998) Asymmetrical direction mutations pressure in the mitochondrial genome of mammals. Mol Biol Evol 15:957–966

Roberti M, Polosa PL, Bruni F, Musicco C, Gadaleta MN, Cantatore P (2003) DmTTF, a novel mitochondrial transcription termination factor that recognises two sequences of *Drosophila melanogaster* mitochondrial DNA. Nucleic Acids Res 31:1597–1604

Roberti M, Bruni F, Loguercio Polsa P, Gadaleta MN, Cantatore P (2006) The *Drosophila* termination factor DmTTF regulates *in vivo* mitochondrial transcription. Nucleic Acids Res 34:2109–2116

Rokas A, Holland PWH (2000) Rare genomic changes as a tool for phylogenetics. Trends Ecol Evol 15:454–459

Ronquist F, Huelsenbeck JP (2003) MRBAYES 3: Bayesian phylogenetic inference under mixed models. Bioinformatics 19:1572–1574

Rubinoff D, Cameron SL, Will K (2006) A genomic perspective on the shortcomings of mitochondrial DNA for "barcoding" identification. J Heredity 97:581–594

Saccone C, De Giogi C, Gissi C, Pesole G, Reyes A (1999) Evolutionary genomics in Metazoa: The mitochondrial DNA as a model system. Gene 238:195–209

Saito S, Tamura K, Aotsuka T (2005) Replication origin of mitochondrial DNA in insects. Genetics 171:1695–1705

Sancar A, Sancar GB (1988) DNA repair enzymes. Ann Rev Biochem 57:29–67

Shao R, Barker SC (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 NJH, Barker SC (2001) Numerous gene rearrangements in the mitochondrial genome of the wallaby louse, *Heterodoxus macropus* (Phthiraptera). Mol Biol Evol 18:858–865

Shao R, Baker SC, Mitani H, Aoki Y, Fukunaga M (2004) Evolution of duplicate control regions in the mitochondrial genomes of Metazoa: A case study with Australasian *Ixodes* ticks. Mol Biol Evol 22:620–629

Signorovitch AY, Buss LW, Dellaporta SL (2007) Comparative genomics of the large mitochondria in placozoans. Public Lib Sci, Genetics 3:44–50

Steinauer ML, Nickol BB, Broughton R, Orti G (2005) First sequenced mitochondrial genome from the phylum Acanthocephala (Leptorhynchoides thecatus) and its phylogenetic position within Metazoa. J Mol Evol 60:706–715

Swofford DL (2002) PAUP* Phylogenetic Analysis using Parsimony (*and Other Methods). Ver. 4. Sinauer Associates, Sunderland, MA

Taanman J-W (1999) The mitochondrial genome: structure, transcription, translation and replication. Biochim Biophys Acta 1410:103–123

Tanaka M, Ozawa T (1994) Strand asymmetry in human mitochondrial DNA mutations. Genomics 22:327–335

Webster BL, Copley RR, Jenner RA, Mackenzie-Dodds JA, Bourlat SJ, Rota-Stabelli O, Littlewood DTJ, Telford MJ (2006) Mitogenomics and phylogenomics reveal priapulid worms as extant models of the ancestral ecdysozoan. Evol Develop 8:502–510

Yang MY, Bowmaker M, Reyes A, Vergani L, Angeli P, Gringeri E, Jacobs HT, Holt IJ (2002) Biased incorporation of ribonucleotides on the mitochondrial L-strand accounts for apparent strand-asymmetric DNA replication. Cell 111:495–505

Yokobori S-I, Takuya U, Feldmaier-Fuchs G, Paabo S, Ueshima R, Kondow A, Nishikawa K, Watanabe K (1999) Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). Genetics 153:1851–1862

Yokobori S-I, Oshima T, Wada H (2005) Complete nucleotide sequence of the mitochondrial genome of *Doliolum nationalis* with implications for evolution of urochordates. Mol Phylogenet Evol 34:273–283

Zhang DS, Hewitt FM (1997) Insect mitochondrial control region: A review of its structure, evolution and usefulness in evolutionary studies. Biochem Syst Ecol 25:99–210

Zuker M (2003) Mfold web server for nucleic acid folding and hybridization prediction. Nucleic Acids Res 31:3406–3415