The Complete Mitochondrial Genome of the Articulate Brachiopod *Terebratalia transversa*

Kevin G. Helfenbein[1,2], Wesley M. Brown[1], and Jeffrey L. Boore[1,2]

[1] Department of Biology, University of Michigan, 830 N. University Ave., Ann Arbor, MI 48109 USA

[2] DOE Joint Genome Institute and Lawrence Berkeley National Laboratory, 2800 Mitchell Drive, Walnut Creek, CA 94598 USA

Corresponding author:

Kevin G. Helfenbein, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598 USA

Phone: (925)296-5662

FAX: (925) 296-5666

E-mail: kgh@umich.edu

Keywords: *Terebratalia transversa*, mitochondria, genome, evolution, codon usage, nucleotide skew

Running title: *Terebratalia* Mitochondrial Genome

Abstract

We have sequenced the complete mitochondrial DNA (mtDNA) of the articulate brachiopod *Terebratalia transversa*. The circular genome is 14,291 bp in size, relatively small compared to other published metazoan mtDNAs. The 37 genes commonly found in animal mtDNA are present; the size decrease is due to the truncation of several tRNA, rRNA, and protein genes, to some nucleotide overlaps, and to a paucity of non-coding nucleotides. Although the gene arrangement differs radically from those reported for other metazoans, some gene junctions are shared with two other articulate brachiopods, *Laqueus rubellus* and *Terebratulina retusa*. All genes in the *T. transversa* mtDNA, unlike those in most metazoan mtDNAs reported, are encoded by the same strand. The A+T content (59.1%) is low for a metazoan mtDNA, and there is a high propensity for homopolymer runs and a strong base-compositional strand bias. The coding strand is quite G+T-rich, a skew that is shared by the confamilial (laqueid) species *L. rubellus*, but opposite to that found in *T. retusa*, a cancellothyridid. These compositional skews are strongly reflected in the codon usage patterns and the amino acid compositions of the mitochondrial proteins, with markedly different usage observed between *T. retusa* and the two laqueids. This observation, plus the similarity of the laqueid non-coding regions to the reverse complement of the non-coding region of the cancellothyridid, suggest that an inversion that resulted in a reversal in the direction of first-strand replication has occurred in one of the two lineages. In addition to the presence of one non-coding region in *T. transversa* that is comparable to those in the other brachiopod mtDNAs, there are two others with the potential to form secondary structures; one or both of these may be involved in the process of transcript cleavage.

Introduction

The lophophorates are a group of animals that includes three phyla: Brachiopoda, Phoronida and Ectoprocta (see Brusca and Brusca 1990). Their collective name is derived from the most conspicuous morphological characteristic shared among the three groups, a ciliated, tentacular feeding apparatus called a lophophore. One lophophorate phylum, the brachiopods, has a rich fossil record that dates back 600 million years and contains more than 12,000 species; today, about 335 species are known (Brusca and Brusca 1990). Of these, most are in the class Articulata and have two valves (shells) connected by a tooth and socket hinge; the remainder are in the class Inarticulata and have unhinged valves that are connected by muscles alone.

Complete mitochondrial genome sequences are available for 127 animals (see Boore 1999 and the Mitochondrial Genomics link at http://www.jgi.doe.gov). Comparisons among them have demonstrated variation among lineages in many genome features, including gene arrangement, nucleotide composition and skew, and replication and transcription signaling elements. The density of sampling within each phylum varies greatly; Chordata is greatest (80, comprised primarily by 77 vertebrate genomes) followed by Arthropoda (16 genomes), with the remaining 31 genomes representing only 8 of the more than 30 other phyla. Sampling within the lophophorate phyla has begun with two articulate brachiopods, *Terebratulina retusa* (Stechmann and Schlegel 1999) and *Laqueus rubellus* (Noguchi et al. 2000).

The differences between the two previously sequenced articulate brachiopod mtDNAs—gene rearrangements, differences in gene lengths, disparity in number of non-coding nucleotides, differing nucleotide biases—are radical and warrant the examination

of another articulate brachiopod mtDNA. We have determined the mtDNA sequence of the articulate brachiopod *Terebratalia transversa* and have compared it to both the closely related *L. rubellus* and the more distantly related *T. retusa* in order to make a comprehensive analysis of the evolution of mtDNA features within the articulate brachiopod clade. The *T. transversa* genome is relatively small, with size reductions in tRNA, rRNA and protein-encoding genes. The genome contains the 37 genes commonly found in metazoan mtDNA, all encoded by one G+T-rich strand. Its gene arrangement, when compared to those of *T. retusa* and *L. rubellus*, suggests that further comparisons of gene arrangements might provide useful characters for resolving articulate brachiopod relationships.

Materials and Methods

Total DNA of the articulate brachiopod *Terebratalia transversa* was a gift from James M. Turbeville. Pairs of PCR primers were first used to amplify short (500-700 nt) regions within *rrnL*, *cox1,* and *cox3*. The primers used were: (*rrnL*) 16SARL—CGC CTG TTT ATC AAA AAC AT, 16SBRH—CCG GTC TGA ACT CAG ATC ACG T; (*cox1)* LCO1490— GGT CAA CAA ATC ATA AAG ATA TTG G, HCO2198— TAA ACT TCA GGG TGA CCA AAA AAT CA; (*cox3)* CO3-DL1— TGG TGG CGA GAT GTK KTN CGN CGN GA, CO3-DL2— ACW ACG TCK ACG AAG TGT CAR TAT CA. Amplifications employed 40 cycles of 94º for 15s, 50º for 45s, 72º for 120s, and each yielded a single band when visualized with ethidium bromide staining and UV irradiation on a 1.2% agarose gel. The products were purified by three serial passages of ultrafiltration (Millipore Ultrafree™ spin columns, 30,000 NMWL), subjected to cycle sequencing reactions according to the supplier's (Perkin-Elmer) instructions, and analyzed on an ABI 377 automated DNA sequencer. From these sequences, primers were designed that were used to amplify the *T. transversa* mitochondrial genome in three pieces with the GeneAmp® XL PCR Kit. These six primers were tried in all possible pairwise combinations. We initially used reaction conditions of 94º for 45s, followed by 37 cycles of 94º for 15s, 63º for 20s, 68º for 8 min, then a final extension of 72º for 12 min. This generated single fragments of approximately 6.4 kb and 3.0 kb with the primer pairs 16S forward/COI reverse and COI forward/COIII reverse. The primer pairs 16S reverse/COIII forward generated a fragment of approximately 5.4 kb but with two smaller, faint bands even when the annealing temperature was raised to 63º. Purification and sequence determinations were as above, with additional primers used to "primer

walk" through each fragment. These three long-PCR products, together with the three shorter ones, jointly comprise the entire *T. transversa* mitochondrial genome in overlapping segments. Both strands of all amplification products were sequenced.

Sequences were assembled and analyzed using Sequence Analysis and Sequence Navigator™ (ABI) and MacVector™ 6.5 (Oxford Molecular Group). Protein and ribosomal RNA genes were identified by their sequence similarity to the *Lumbricus terrestris* mtDNA homologues (Boore and Brown 1995). tRNA genes were identified either by using tRNAscan-SE (version 1.1, http://www.genetics.wustl.edu/eddy/tRNAscan-SE; Lowe and Eddy 1997) with a cove cutoff score of 0.1, or, in many cases where tRNAs were not found using this program, by recognizing potential secondary structures by eye.

The 5' ends of protein genes were inferred to be at the first legitimate, in-frame start codon (ATN, GTG, TTG, GTT; Wolstenholme 1992) that does not overlap the preceding gene, except that overlap with an upstream tRNA gene was limited to the most 3' nucleotide of the tRNA.

Protein gene termini were inferred to be at the first in-frame stop codon, unless that codon was located within the sequence of a downstream gene. Otherwise, a truncated stop codon (T or TA) adjacent to the beginning of the downstream gene (with the exception of *atp8*) was designated as the termination codon, and assumed to be completed by polyadenylation after transcript cleavage. The 5' and 3' ends of the *rrnL* and *rrnS* were assumed to be adjacent to the ends of bordering tRNA genes.

Results and Discussion

Gene Content and Arrangement

The *Terebratalia transversa* mtDNA encodes the 37 genes (those for 13 proteins, 22 tRNAs and two rRNAs) most commonly found in animal mitochondrial genomes (Boore 1999). Figure 1 shows a circular map of these genes and the largest non-coding region.

There are seven cases where genes appear to overlap; however, there is an alternative possibility. Among the seven apparent overlaps in the genome, six involve only the discriminator nucleotide of an upstream tRNA and the first nucleotide of the adjacent gene, suggesting that the tRNA gene may not actually encode this nucleotide. The seventh case is of a three-nucleotide overlap between *trnQ* and *trnW*. However, it is possible that *trnQ* is actually shorter by these three nucleotides and that the structure of tRNA(Q) is completed by polyadenylation, which would provide As to match the Ts at the 5' end of the acceptor stem, as has been described for other systems (Yokobori and Pääbo 1997).

Start and stop codons were inferred for each of *T. transversa*'s 13 protein-encoding genes. There are five start codons used in the *T. transversa* mtDNA: ATG (*cox2)*, ATT (*nad4, nad4L, nad5, atp8)*, GTG (*nad6, cox1, cox3, cob),* GTT (*nad2*), and TTG (*nad1, nad3, atp6)*. Five genes end on a complete stop codon, either TAA (*nad2, atp6)* or TAG (*nad4L, cox1, cox3)*; the other eight genes end on an abbreviated stop codon, in all cases T—, except one case of TA—(*nad6)*. It is notable that for six of the eight genes that we interpret as ending at abbreviated stop codons (*cob*, *nad5*, *nad6*, *nad1*, *cox2*, and *nad3*), there is a complete stop codon further downstream that, if used, would cause overlap of

the downstream tRNA gene by one to 26 nts. As we have previously suggested (Boore and Brown 2000), the downstream stop codons could possibly act as "backups" in cases of improper transcript cleavage.

All of the genes are transcribed from the same strand, a relatively uncommon state among the 127 described animal mtDNAs (see Boore 1999 and from the Mitochondrial Genomics link at http://www.jgi.doe.gov), previously found only in the two other articulate brachiopods (Stechmann and Schlegel 1999; Noguchi et al. 2000), the blue mussel *Mytilus edulis* (Hoffmann, Boore, and Brown 1992), two annelids (Boore and Brown 1995, 2000; GenBank record NC_000931), four nematodes (Okimoto et al. 1991, 1992; Keddie et al. 1998), a tunicate (Yokobori et al. 1999), six parasitic flatworms (Le at al. 2000) and the hexacoral *Metridium senile* (Beagley, Okimoto, and Wolstenholme 1998).

Among the three brachiopod species for which complete mtDNA sequences are available, the most closely related pair is *T. transversa* and *L. rubellus*, both members of the family Laqueidae. Only a few small blocks of genes are identically arranged between these two mtDNAs, and even fewer between either of these and the mtDNA of the cancellothyridid *T. retusa* (fig. 2). This latter genome retains a number of features inferred to be primitive for Brachiopoda, since they are also found in the distantly related polyplacophoran mollusk *Katharina tunicata*. The large number of gene rearrangements among these taxa suggests that gene arrangement data will be useful for inferring phylogeny at lower taxonomic levels within the brachiopod clade Articulata.

The genes *atp8* and *atp6* are adjacent in all arthropod and deuterostome mtDNAs, and in that of the yeast *Saccharomyces cerevisiae* (Foury et al. 1999), sometimes in

overlapping reading frames. It has been demonstrated that *atp8* and *atp6* are translated from a bicistronic mRNA in mammals (Fearnley and Walker 1986), and this is also a possible, although undemonstrated, phenomenon in other animals having *atp8*/*atp6* adjacency. These genes are not adjacent in the mtDNAs of annelids (Boore and Brown 2000), of many mollusks and of two laqueid brachiopods. As has been suggested (Boore 1999), this loss of adjacency allowed by the loss of the corresponding mode of translation may be derived features that characterize a clade containing several phyla, a suggestion that awaits further phylogenetic and experimental analysis. However, if this should prove to be true, then the primitive state of gene arrangement (but presumably not of translation) must have been retained in the squid *Loligo bleekeri* (Sasuga et al. 1999), the chiton *K. tunicata* (Boore and Brown 1995), and in *T. retusa* (Stechmann and Schlegel 1999).

Genome Size

The size range reported for completely sequenced animal mtDNAs ranges from 13.5 kb (*Taenia crassiceps*; Le et al. 2000) to 19.5 kb (*Drosophila melanogaster*; Lewis, Farr, and Kaguni 1995). *T. transversa* mtDNA, at 14, 291 nucleotides, is near the low end of this range, and is comparable in size to that of *L. rubellus* (14,017 nts; Noguchi et al. 2000), or to those of some nematodes (Okimoto et al. 1991, 1992; Keddie et al. 1998), of gastropods (Hatzoglou, Rodakis, and Lecanidou 1995; Terrett, Miles, and Thomas 1996; Yamazaki et al. 1997), and of parasitic flatworms (Le et al. 2000).

Each of the two laqueid mtDNAs is reduced in size in similar ways from that of the larger *T. retusa* mtDNA (table 1). There are fewer nucleotides in non-coding regions, the

rRNA genes appear to be shorter, the tRNA genes are somewhat smaller, [due especially to a decreased number of nucleotides in the TΨC arms (fig. 3)] and protein gene sizes have been drastically reduced. There is uncertainty regarding rRNA gene sizes, and precise end assignment is not possible from sequence information alone (e.g., the 45 nucleotides at the 3' end of *rrnL* of *L. rubellus* do not align well to *rrnL* of *T. transversa* [data not shown], suggesting that some nucleotides assigned to the *L. rubellus rrnL* may actually be non-coding). The size reduction of the *T. transversa* protein-encoding genes is striking; there are ca. 1,500 fewer nucleotides (500 fewer codons) in *L. rubellus* and *T. transversa* than in *T. retusa* protein-encoding genes, a surprisingly high reduction.

Nucleotide Composition

T. transversa mtDNA has an A+T composition of 59.1%, a value that is low compared to many other invertebrate mtDNAs. The base composition of the individual strands is biased, and can be described by skewness (Perna and Kocher 1995), which measures the relative number of As to Ts (AT skew = [A-T]/[A+T]) and Gs to Cs (GC skew= [G-C]/[G+C]) on a strand; for the coding strand of *T. transversa* (table 1), AT skew= -0.33 and GC skew= +0.34. The skewness is even more extreme in protein-encoding sequences (AT skew=-0.41, GC skew=+0.37). AT skew is virtually absent outside of protein-coding genes, although GC skew remains positive. Among tRNAs and rRNAs (and in functionally important paired secondary structures in non-coding regions) a similar number of As and Ts is required for stem structure formation, and perhaps accounts for the lower magnitude of AT skew; however, this requirement probably has less effect on GC skew, due to the lack of appreciable helix-destabilization by GT pairs.

These skews may be generated during either transcription or replication, processes that expose Cs and As on the displaced (unpaired) strand to deamination (C → U and A → hypoxanthine, resulting in C → T and A → G transitions), whereas those on the other strand are protected by pairing with the bases of the nascent RNA or DNA strand (Francino and Ochman 1997; Reyes et al. 1998). Thus, one strand (the leading strand of replication and the non-template strand of transcription) may become GT rich, while the other becomes AC rich. MtDNA replication is highly asymmetric in mammals and in fruit flies (Clayton 1982), and this asymmetry may be general; if so, this model for the generation of a strand-biased nucleotide composition is especially important. Although the relative importance of the two contributing process (i.e., transcription *vs*. replication) remains to be assessed, our nucleotide skew data suggests that transcription does not play a role in the *T. transversa* system (discussed below).

Interestingly, *T. transversa*'s skew values are very similar to those of the other sequenced laqueid, *L. rubellus* (AT skew=-0.29, GC skew=+0.27), but opposite to that of the cancellothyridid *T. retusa* (AT skew=+0.03, GC skew=-0.29) (table 1). Because all genes are transcribed in the same direction in each of these mtDNAs, mutation (deamination) pressure should be the same on the displaced strand and the three mtDNAs should show the same or similar skewness relative to the coding strand if transcription is the most important cause of strand-biased nucleotide composition. That this expectation is not borne out indicates that mutational events during transcription cannot be causative of these differences, and instead suggests that replication is more important. It is possible that the origin of replication might be oppositely oriented in these two families of brachiopods and that deaminations during replication are the primary cause of skewness.

In support of this, the sequence of a portion of the non-coding region of each of the two laqueid mtDNAs is similar to the reverse complement of a portion of the non-coding region of *T. retusa* (fig. 4). The non-coding region in vertebrate and fruit fly mtDNA is known to contain sequence elements that mediate DNA replication, and it is possible that the non-coding region in these brachiopod mtDNAs contains functionally similar sequences. Alternatively, it is possible that the similarities seen here are, themselves, due to the differential biasing of base compositions of the two strands among these mtDNAs.

Codon Usage and Amino Acid Composition

The extent to which synonymous codon usage is determined by selection is not clear, although in some systems it appears that certain codons are used more frequently in highly expressed genes for greater translational efficiency (Ikemura 1981, 1982). The extent to which, or even whether this plays a role in animal mitochondrial systems is unknown, since there is, at present, no evidence for differential protein gene regulation in animal mitochondrial systems. Given that the protein products of these genes all act in concert in coupled oxidative phosphorylation/electron transport reactions, we regard translational regulation as an unlikely possibility.

For the most part, the bias in usage of synonymous codons in the proteins of *T. transversa* mtDNA follows the same pattern of nucleotide frequency (T>G>A>C) as the mtDNA coding strand as a whole (fig. 5 and tables 1 and 2). This bias is evident in both four-fold and two-fold degenerate codon families, suggesting that third codon positions mostly reflect mutational bias. However, six codons are exceptions to this, all of which have an unexpectedly high homodimer frequency at codon positions two and three: TCC,

CCC, GCC, CGG, AGG, and GGG. The most extreme variation is found in the two homotrimeric codons CCC and GGG. The hypothesis that this reflects mutation pressure is bolstered by the observation that homopolymer runs ranging from 2-11 nucleotides in length are more common than expected throughout the mtDNA, given the nucleotide frequencies. We are unable to determine whether this effect increases the frequency of NTT codons, since T is already the most common nucleotide, or of NAA codons, because all of these are in two-fold degenerate codon families, and so can be compared only to NAG codons (all of which are more common than NAA, as expected). For *L. rubellus*, a similar effect is apparent for only four codons, GCC, CGG, AGG, and GGG; all are more common than expected. Otherwise, third codon positions deviate from the expected ranked frequency T>G>A>C only by the infrequency of NCG codons.

Codon usage in *T. retusa* shows no such strikingly consistent biases. However, *T. retusa*'s codon usage supports the hypothesis that selection maintains a high number of hydrophobic, non-polar amino acids in membrane associated proteins (Asakawa et al. 1991). The greater number of Ts than As in *T. retusa*'s protein genes, as indicated by the negative skew value (Table 1), are overrepresented at second codon positions (see figure 5) and the amino acids specified by these codons (phenylalanine, leucine, isoleucine, methionine, and valine) are non-polar and hydrophobic.

To what extent is amino acid composition determined by natural selection and to what extent by mutational bias? Clearly, purifying selection must eliminate certain mutations that would change amino acids at essential sites in mitochondrial proteins, and directional selection may sometimes create novel amino acid identities at some sites. At other, less constrained positions, amino acids with similar physical or chemical traits can

substitute acceptably, with some sites apparently tolerating even quite radical amino acid substitutions; for these, substitutions would be expected to reflect mutational bias.

To address this question, we analyzed the extent to which the base composition at first and second codon positions is similar to that at third codon positions, which are much freer to vary in conformation to mutational bias. The amino acid compositions of the mitochondrial proteins of three brachiopods fall into four physico-chemical groups, and are listed in descending order relative to their change in frequency between *T. transversa* and *T. retusa* (table 3). The laqueids *T. transversa* and *L. rubellus* are very similar in this respect, differing only in their relative usages of the leucine codons TTR or CUN. Both differ greatly, however, from *T. retusa*. Within each of the four amino acid groupings, codons rich in T and G (e.g., TTR, GTN and TTY in the non-polar group) are much more common in the laqueids, and those rich in A and C (e.g., ATR, ATY, CCN, GCN and CTN in the non-polar group) are much rarer, suggesting that mutational bias is a prime factor influencing amino acid composition, at least within each physico-chemical group.

tRNA and rRNA Genes

Proposed secondary structures for all 22 tRNAs are shown in figure 3. Many of the tRNAs have reduced TΨC arms: there is no TΨC arm in tRNA(R), there is potential for only a single base pair in tRNA(K) and tRNA(Y), and of only two base pairs in 10 others. The tRNAs S1, S2 and I lack a paired DHU arm, a condition commonly found among serine, but not isoleucine tRNAs. There is potential for a paired DHU arm in tRNA(T), but only if there are two nucleotides (AG), rather than one, between the anticodon and

DHU stems. Among the other tRNAs, nine have As, five have Gs, three have Ts, and one has a C at this position. Between the DHU and acceptor stems, 14 tRNAs have the dinucleotide TA, two have GA, two have AA, and one has TG. All tRNAs have a "variable arm" with four nucleotides. Thirteen of the anticodons are flanked by T and A, eight by T and G, and one by C and A.

As in all metazoan mitochondrial genomes, that of *T. transversa* encodes two rRNAs. The size of its *rrnL* is 1105 nts, with A+T = 62.6% (the highest of any region in the genome) and AT and GC skews of -0.08 and 0.23, respectively. *rrnS* is 762 nts in size, with A+T = 59.2% and AT and GC skews of -0.06 and 0.24.

Non-coding Nucleotides

T. transversa mtDNA has 202 non-coding nucleotide pairs; 149 are in three non-coding regions of 35 (between *atp8* and *cox3*), 42 (between *nad2* and *cox1*), and 69 (between *cox1* and *trnC*) nucleotides, and 53 are dispersed throughout the genome, either as single nucleotide pairs or in blocks ranging in size from 5-9 pairs.

The enzymatic removal of tRNAs from a polycistronic transcript is necessary to release adjacent, gene-specific messages (Battey and Clayton 1980; Ojala et al. 1980; Rossmanith 1997). When two protein-encoding genes abut directly, in some cases both are translated from the same bicistronic message (Ojala et al. 1980); in some others, sequences with the potential to form stem-loop structures are present at the junctions, and those may mediate transcript cleavage (e.g., Boore and Brown 1994). Three protein-encoding gene pairs abut directly in *T. transversa* mtDNA: *atp8-cox3*, *nad2-cox1*, and *nad4L-cox2*. Both *nad2* and *nad4L* have complete stop codons, so it is possible that each

forms a bicistronic mRNA with *cox1* and *cox2*, respectively. However, *atp8* in the *atp8-cox3* pair lacks a complete stop codon; in the absence of editing or processing, a single product would result from translation of the bicistronic transcript. For the *atp8-cox3* and *nad2-cox1* pairs, adjacent genes are separated by the 35 nt- and 42 nt-long non-coding regions described above; each of the non-coding sequences has the potential to form an RNA stem-loop structure (fig. 6). No such potential could be identified for the *nad4L-cox2* junction or for the 69 nt-long region following *cox1*.

Notably, the proposed stem-loop structure between *atp8* and *cox3* is identical to the anticodon stem and loop of tRNA(P) at 13 of 17 nucleotide positions; these are marked with asterisks in figure 6. A plausible explanation for the similarity of this stem-loop sequence to that of *trnP* is that this stem-loop is a duplicate of trnP that has mostly been deleted (Moritz, Dowling, and Brown 1987; Boore 2000). In this scenario a duplicate copy of *trnP* could have been selectively maintained to the extent necessary to preserve its function in processing the *atp8-cox3* transcript. An additional observation regarding the non-coding region between *atp8* and *cox3* (fig. 6A) is that the nine nt sequence GAGGCAGCT appears twice (at positions 3008-3016 and 3023-3031); whether this sequence, which appears nowhere else in the *T. transversa* mtDNA, serves any functional role awaits experimental analysis.

The third and longest (69 nts) of the non-coding regions is similar in length to that of *L. rubellus* (54 nts). Each of these is much shorter than the longest non-coding region in *T. retusa* (794 nts), although in *T. retusa* only 287 of the 794 nt are unique, with the remainder being in repeated sequences. Attempts at aligning the non-coding regions to search for possibly conserved functional elements reveal that the “plus” strand of the

non-coding regions of the laqueids aligns best to the “minus” strand of the non-coding region of *T. retusa* (fig. 4). It is possible that this alignment is due to some conserved regulatory function, perhaps for DNA replication. If this is the case, then the opposite strand biases among these mtDNAs might be explained by a reversal in the direction of first strand replication.

The near uniformity of proteins encoded by metazoan mtDNAs makes this DNA an excellent model system for the study of molecular evolution. Comparisons of mtDNAs from closely related species, such as these brachiopods, are of general utility for future studies of genome and protein evolution. Here, for example, we see an example of mutational bias as well as natural selection determining amino acid usage. Further descriptions and comparisons of mtDNAs, along with functional studies, will allow us to address questions such as which amino acids in which of the proteins are selected, what the roles of these amino acids are, and which portions of the proteins encoded by larger mtDNAs, e.g., that of *T. retusa*, have been lost by smaller genomes, e.g., those of the laqueids, and what functions they serve in the larger proteins.

Acknowledgements

We gratefully acknowledge D. Lavrov for his comments that improved this manuscript. This work was supported by grant DEB9807100 from the National Science Foundation. Part of this work under the auspices of the U.S. Department of Energy, Office of Biological and Environmental Research, by the University of California, Lawrence Berkeley National Laboratory under contract No. DE-AC03-76SF00098.

Literature cited

BATTEY J., and D. A. CLAYTON. 1980. The transcription map of human mitochondrial-DNA implicates transfer-RNA excision as a major processing event. J. Biol. Chem. **255:**1599-1606

BEAGLEY, C. T., R. OKIMOTO, and D. R. WOLSTENHOLME. 1998. The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): Introns, a paucity of tRNA genes, and a near-standard genetic code. Genetics **148:**1091-1108.

BOORE, J. L. 1999. Animal mitochondrial genomes. Nucl. Acids Res. **27:**1767-1780.

BOORE, J. L. 2000. The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals, Pp. 133-147, in D. SANKOFF and J. NADEAU, eds. Comparative Genomics, Computational Biology Series vol 1, Kluwer Academic Publishers, Dordrecht, Netherlands.

BOORE, J. L., and W. M. BROWN. 1994. Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*. Genetics **138:**423-443.

BOORE, J. L., and W. M. BROWN. 1995. Complete DNA sequence of the mitochondrial genome of the annelid worm, *Lumbricus terrestris*. Genetics. **141**:305-319.

BOORE, J. L., and W. M. BROWN. 2000. Mitochondrial genomes of *Galatheolinum, Helobdella,* and *Platynereis:* Sequence and gene arrangement comparisons show that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa. Mol. Biol. Evol. **17:**87-106.

BRUSCA R. C., and G. J. BRUSCA. 1990. Invertebrates. Sinauer, Sunderland, Mass.

CLAYTON, D. A. 1982. Replication of animal mitochondrial DNA. Cell **28:**693-705.

FEARNLEY, I. M., and J. E. WALKER. 1986. Two overlapping genes in bovine mitochondrial DNA encode membrane components of ATP synthase. EMBO J **5:**2003-2008.

FOURY, F., T. ROGANTI, N. LECRENIER, and B. PURNELLE. 1999. The complete sequence of the mitochondrial genome of *Saccharomyces cerevisiae*. FEBS Letters **440:**325-331.

FRANCINO, M. P., and H. OCHMAN. 1997. Strand asymmetries in DNA evolution. Trends Genet. **13:**240-245.

HATZOGLOU, E., G. C. RODAKIS, and R. LECANIDOU., 1995. Complete sequence and gene organization of the mitochondrial genome of the land snail *Albinaria coerulea.* Genetics **140:**1353-1366.

HOFFMANN, R. J., J. L. BOORE, and W. M. BROWN. 1992. A novel mitochondrial genome organization for the Blue Mussel, *Mytilus edulis*. Genetics **131:**397-412.

IKEMURA, T. 1981. Correlations between the abundance of *Escherichia coli* transfer RNAs and the occurrence of the respective codons in its protein genes: a proposal for a synonymous codon choice that is optimal for the *E. coli* translational system. J. Mol. Biol. **151:**389-409.

—— 1982. Correlation between the abundance of yeast transfer RNAs and the occurrence of respective codons in protein genes. J. Mol. Biol. **158:**573-597.

KEDDIE E. M., T. HIGAZI, and T. R. UNNASCH. 1998. The mitochondrial genome of *Onchocerca volvulus*: sequence, structure and phylogenetic analysis. Mol. Biochem. Parasitol. **95:**111-127.

LE, T. H., D. BLAIR, T. AGATSUMA, P. F. HUMAIR, N. J. H. CAMPBELL, M. IWAGAMI, D. T. J. LITTLEWOOD, B. PEACOCK, D. A. JOHNSTON, J. BARTLEY, D. ROLLINSON, E. A. HERNIOU, D.S. ZARLENGA, and D. P. MCMANUS. 2000. Phylogenies inferred from mitochondrial gene orders - A cautionary tale from the parasitic flatworms. Mol. Biol. Evol. **17:**1123-1125.

LEWIS, D. L., C. L. FARR, and L. S. KAGUNI. 1995. *Drosophila melanogaster* mitochondrial DNA: completion of the nucleotide sequence and evolutionary comparisons. Insect Mol. Biol. **4:**263-278.

MORITZ, C., T. E. DOWLING, and W. M. BROWN. 1987. Evolution Of Animal Mitochondrial DNA: Relevance For Population Biology And Systematics. Annu. Rev. Ecol. Syst. **18:**269-292.

NOGUCHI Y., K. ENDO, F. TAJIMA, and R.UESHIMA. 2000. The mitochondrial genome of the brachiopod *Laqueus rubellus*. Genetics **155:**245-259.

OJALA D., C. MERKEL, R. GELFAND, and G. ATTARDI. 1980. The transfer RNA genes punctuate the reading of genetic information in human mitochondrial DNA. Cell **22:**393-403.

OKIMOTO, R., H. M. CHAMBERLIN, J. L. MACFARLANE, and D. R. WOLSTENHOLME. 1991. Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (*Meloidogyne*) nucleotide sequences, genome location and potential for host-race identification. Nucl. Acids Res. **19:**1619-1626.

OKIMOTO, R., J. L. MACFARLANE, D. O. CLARY, and D. R. WOLSTENHOLME. 1992. The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics **130:**471-498.

PERNA, N. T, and T. D. KOCHER. 1995. Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes. J. Mol. Evol. **41:**353-358.

REYES, A., C. GISSI, G. PESOLE, and C. SACCONE. 1998. Asymmetrical directional mutation pressure in the mitochondrial genome of mammals. Mol. Biol. Evol. **15:**957-966.

ROSSMANITH, W. 1997. Processing of human mitochondrial tRNA-Ser(AGY); A novel pathway in tRNA biosynthesis. J. Mol. Biol. **265:**365-371.

SASUGA, J., S-I. YOKOBORI, M. KAIFU, T. UEDA, K. NISHIKAWA, and K. WATANABE. 1999. Gene contents and organization of a mitochondrial DNA segment of the squid *Loligo bleekeri*. J. Mol. Evol. **48:**692-702.

STECHMANN, A., and M. SCHLEGEL. 1999. Analysis of the complete mitochondrial DNA sequence of the brachiopod *Terebratulina retusa* places Brachiopoda within the protostomes. Proc. R. Soc. Lond. B **266:**2043-2052.

TERRETT, J., S. MILES, and R. THOMAS. 1996. Complete DNA sequence of the mitochondrial genome of *Cepaea nemoralis* (Gastropoda: Pulmonata). J. Mol. Evol. **42:**160-168.

YAMAZAKI, N., R. UESHIMA, J. TERRETT, S-I. YOKOBORI, M. KAIFU, et al. 1997. Evolution of pulmonate gastropod mitochondrial genomes: comparisons of gene organizations of *Euhadra, Cepaea* and *Albinaria* and implications of unusual tRNA secondary structures. Genetics **145:**749-758.

YOKOBORI, S., and S. PÄÄBO. 1997. Polyadenylation creates the discriminator nucleotide of chicken mitochondrial tRNA(Tyr). J. Mol. Biol. **265:**95-99.

YOKOBORI, S., T. UEDA, G. FELDMAIER-FUCHS, S. PÄÄBO, R. UESHIMA, A. KONDOW, K. NISHIKAWA, and K. WATANABE. 1999. Complete DNA sequence of the mitochondrial genome of the ascidian *Halocynthia roretzi* (Chordata, Urochordata). Genetics **153:**1851-1862.

Table 1

Nucleotide Composition

		Nts[a]	A+T	AT skew	GC skew
Whole mtDNA	*T. transversa*	14,291	59.1%	-0.33	0.34
	L. rubellus	14,017	58.3%	-0.29	0.27
	T. retusa	15,451	57.2%	0.03	-0.29
Protein genes[b]	*T. transversa*	10,890	58.8%	-0.41	0.37
	L. rubellus	10,731	58.1%	-0.37	0.29
	T. retusa	12,234	56.3%	-0.11	-0.24
rRNA genes	*T. transversa*	1,867	61.2%	-0.08	0.24
	L. rubellus	1,910	58.5%	-0.01	0.19
	T. retusa	2,057	59.1%	0.28	-0.18
tRNA genes	*T. transversa*	1,318	58.5%	-0.07	0.27
	L. rubellus	1,291	58.8%	0.00	0.21
	T. retusa	1,399	60.8%	0.12	-0.07
Non-coding nts	*T. transversa*	202	56.4%	0.00	0.36
	L. rubellus	79	69.6%	-0.24	0.42
	T. retusa	852	61.8%	0.14	-0.23

[a] minus stop codons

[b] The discrepancy between the sum of the four mtDNA subsections and the whole genome is due to double-counting of overlapping nucleotides and omission of stop codons (some abbreviated) from this breakdown.

Table 2

Codon usage of all mitochondrially encoded genes of three articulate brachiopods[a].

Amino acid	Codon	Ttr[b]	Lru[c]	Tre[d]	Amino acid	Codon	Ttr	Lru	Tre
Phe (F)	TTT	282	252	156	Tyr (Y)	TAT	110	91	54
(GAA)[e]	TTC	31	34	114	(GUA)	TAC	28	34	81
Leu (L2)	TTA	166	135	95					
(UAA)	TTG	239	193	39					
Leu (L1)	CTT	64	88	158	His (H)	CAT	48	51	41
(UAG)	CTC	9	18	131	(GUG)	CAC	22	24	52
	CTA	25	47	181	Gln (Q)	CAA	20	17	65
	CTG	38	51	60	(UUG)	CAG	41	37	19
Ile (I)	ATT	171	153	152	Asn (N)	AAT	62	54	62
(GAU)	ATC	16	35	140	(GUU)	AAC	11	20	60
Met (M)	ATA	63	70	152	Lys (K)	AAA	14	18	72
(CAU)	ATG	93	108	55	(UUU)	AAG	62	60	33
Val (V)	GTT	215	230	101	Asp (D)	GAT	58	43	27
(UAC)	GTC	13	32	52	(GUC)	GAC	20	23	47
	GTA	72	56	71	Glu (E)	GAA	15	23	60
	GTG	136	115	30	(UUC)	GAG	55	55	17

Ser (S2)	TCT	176	150	105	Cys (C)	TGT	59	47	15
(UGA)	TCC	18	17	117	(GCA)	TGC	4	7	26
	TCA	17	37	73	Trp (W)	TGA	24	22	86
	TCG	33	23	10	(UCA)	TGG	81	88	31
Pro (P)	CCT	58	65	46	Arg (R)	CGT	21	18	5
(UGG)	CCC	27	27	103	(UCG)	CGC	5	3	6
	CCA	15	35	55		CGA	8	16	45
	CCG	44	31	27		CGG	25	28	12
Thr (T)	ACT	74	66	72	Ser (S1)	AGT	49	37	9
(UGU)	ACC	9	13	101	(UCU)	AGC	13	7	21
	ACA	14	26	74		AGA	32	40	78
	ACG	21	10	9		AGG	69	62	31
Ala (A)	GCT	101	125	97	Gly (G)	GGT	64	58	22
(UGC)	GCC	18	23	132	(UCC)	GGC	21	19	30
	GCA	18	23	65		GGA	49	68	88
	GCG	35	24	16		GGG	229	194	94

[a] Termination codons were not included in this analysis.

[b] The 3630 codons of *Terebratalia transversa*.

[c] The 3576 codons of *Laqueus rubellus* (Noguchi et al. 2000).

[d] The 4078 codons of *Terebratulina retusa* (Stechmann and Schlegel 1999)

[e] The anticodon of the corresponding tRNA is shown in parentheses below each codon designation.

Table 3

Comparison of the amino acid compositions of the 13 mitochondrially encoded proteins of the articulate brachiopods *Terebratalia transversa* (Ttr), *Laqueus rubellus* (Lru), and *Terebratulina retusa* (Tre).

	Ttr	Δ Tre[a]	Lru	Δ Tre	Tre
Non-polar					
L (TTR)[b]	405	2.02	328	1.45	134
V (GTN)	436	0.72	433	0.70	254
F (TTY)	313	0.16	286	0.06	270
W (TGR)	105	-0.10	110	-0.06	117
M (ATR)	156	-0.25	178	-0.14	207
I (ATY)	187	-0.36	188	-0.36	292
P (CCN)	144	-0.38	158	-0.32	231
A (GCN)	172	-0.45	195	-0.37	310
L (CTN)	136	-0.74	204	-0.62	530
Total	2054	-0.12	2080	-0.11	2345
Polar					
G (GGN)	363	0.55	339	0.45	234
C (TGY)	63	0.54	54	0.32	41
S (AGN)	163	0.17	146	0.05	139
Y (TAY)	138	0.02	125	-0.07	135

S	(TCN)	244	-0.20	227	-0.26	305
Q	(CAR)	61	-0.27	54	-0.36	84
N	(AAY)	73	-0.40	74	-0.39	122
T	(ACN)	118	-0.54	115	-0.55	256
	Total	1223	-0.07	1134	-0.14	1316
Acidic						
D	(GAY)	78	0.05	66	-0.11	74
E	(GAR)	70	-0.09	78	0.01	77
	Total	148	-0.02	144	-0.05	151
Basic						
R	(CGN)	59	-0.13	65	-0.04	68
H	(CAY)	70	-0.25	75	-0.19	93
K	(AAR)	76	-0.28	78	-0.26	105
	Total	205	-0.23	218	-0.18	266

[a] This is the fractional difference from the occurrence of this codon in *Terebratulina retusa*. Within physico-chemical groups, amino acids are listed in descending order according to this value.

[b] The two leucine and two serine codon families are listed here separately.

Figure legends

FIG 1.—Circular gene map of *Terebratalia transversa* mtDNA (GenBank accession number AF331161). All 37 genes are encoded on the same strand and transcribed clockwise, as indicated by the arrow, although the origin(s) remain unknown. Gene scaling is approximate. NC marks the longest (69 nucleotides) non-coding region. One-letter amino acid abbreviations are used to label the corresponding tRNAs. The anticodon of both leucine tRNAs and both serine tRNAs are indicated in parentheses to distinguish members of each pair.

FIG 2.—Comparison of the mitochondrial gene arrangements of the mollusk *Katharina tunicata* (Boore and Brown 1994) and the three sequenced articulate brachiopods. These circular genomes have been graphically linearized at the 3' end of the arbitrarily chosen *cox3* gene. Thick lines below the *K. tunicata* genome indicate genes encoded by the opposite strand. Illustrated are all the gene junctions shared between *K. tunicata* and *T. retusa* (inferred, therefore, to be primitive for these brachiopods) and between *L. rubellus* and each of the other two brachiopods. Those genes found in a common arrangement in the laqueids, *L. rubellus* and *T. transversa*, but that are shared in a different arrangement between the mollusk and *T. retusa* are inferred to be derived for Laqueidae. Genes are abbreviated as in figure 1.

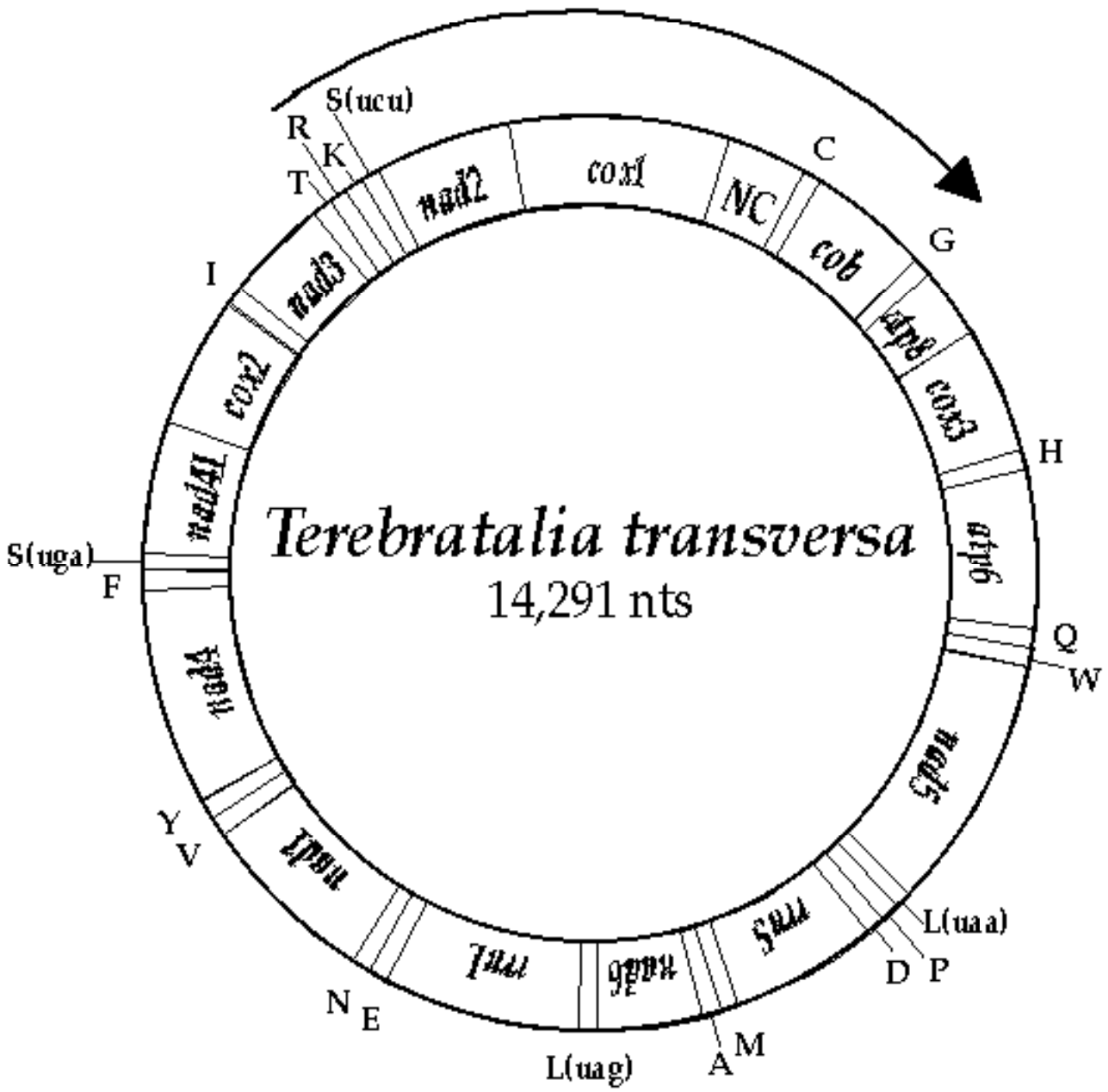
FIG 3.—The potential secondary structures of the 22 inferred tRNAs. The DHU stem and loop of tRNA(T) may form but, if so, there would be two rather than one nucleotide between the DHU arm and the anti-codon arm. The lack of TΨC and DHU arms of

tRNA(R) and tRNA(I), respectively, is an unusual state for these two tRNAs relative to other published animal mtDNAs. We assume a minimum number of three nucleotides in DHU and TΨC loops, and thus do not show potential base pairs that would leave less than three nucleotides in a loop.

FIG 4.—Alignment of the complete sequence of the longest non-coding regions in *T. transversa* (positions 1537-1605), *L. rubellus* (positions 13,906-13,959) and the reverse complement of a segment (positions 7243-7175) of the non-repetitive portion of the long non-coding region in *T. retusa.* Bold indicates conservation of a nucleotide at a particular position in at least two species.

FIG 5.—Frequency of each nucleotide by codon position for all protein encoding genes. Order of bars: *Terebratalia transversa*, *Laqueus rubellus*, *Terebratulina retusa*.

FIG 6.—Inferred secondary structures of two non-coding regions. A) The 35 nucleotides between *atp8* and *cox3* can potentially to form a secondary structure similar to that of a tRNA anti-codon stem-loop. Nucleotides marked with asterisks (*) are the same as those in the corresponding position of the tRNA(P) anti-codon stem-loop. B) The inferred secondary structure of the 45 nucleotides between *nad2* and *cox1*.


Terebratalia transversa
14,291 nts
S(ucu)
R
K
T
C
G
I
H
S(uga)
F
Q
W
Y
V
N
E
L(uag)
A
M
D
P
L(uaa)
nad2
cox1
NC
cob
atp8
cox3
atp6
nad5
nad6
nad1
rrnL
rrnS
nad4
nad4L
cox2
nad3

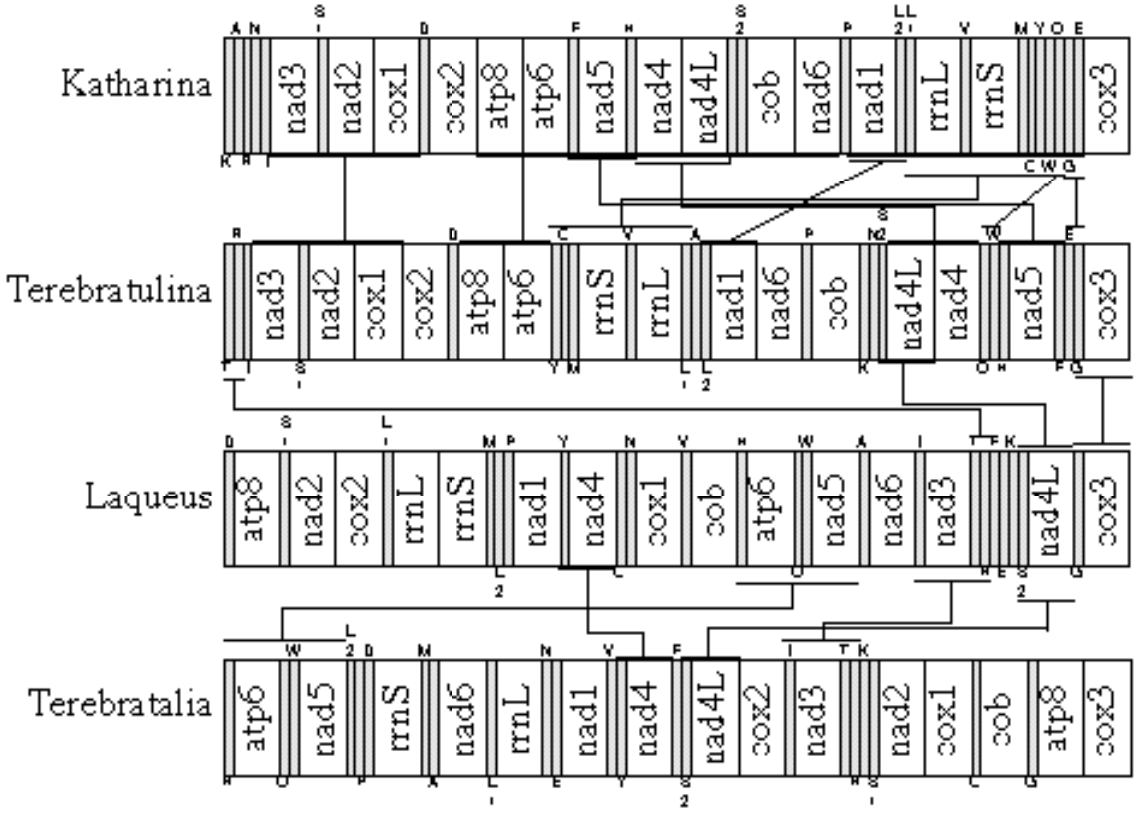
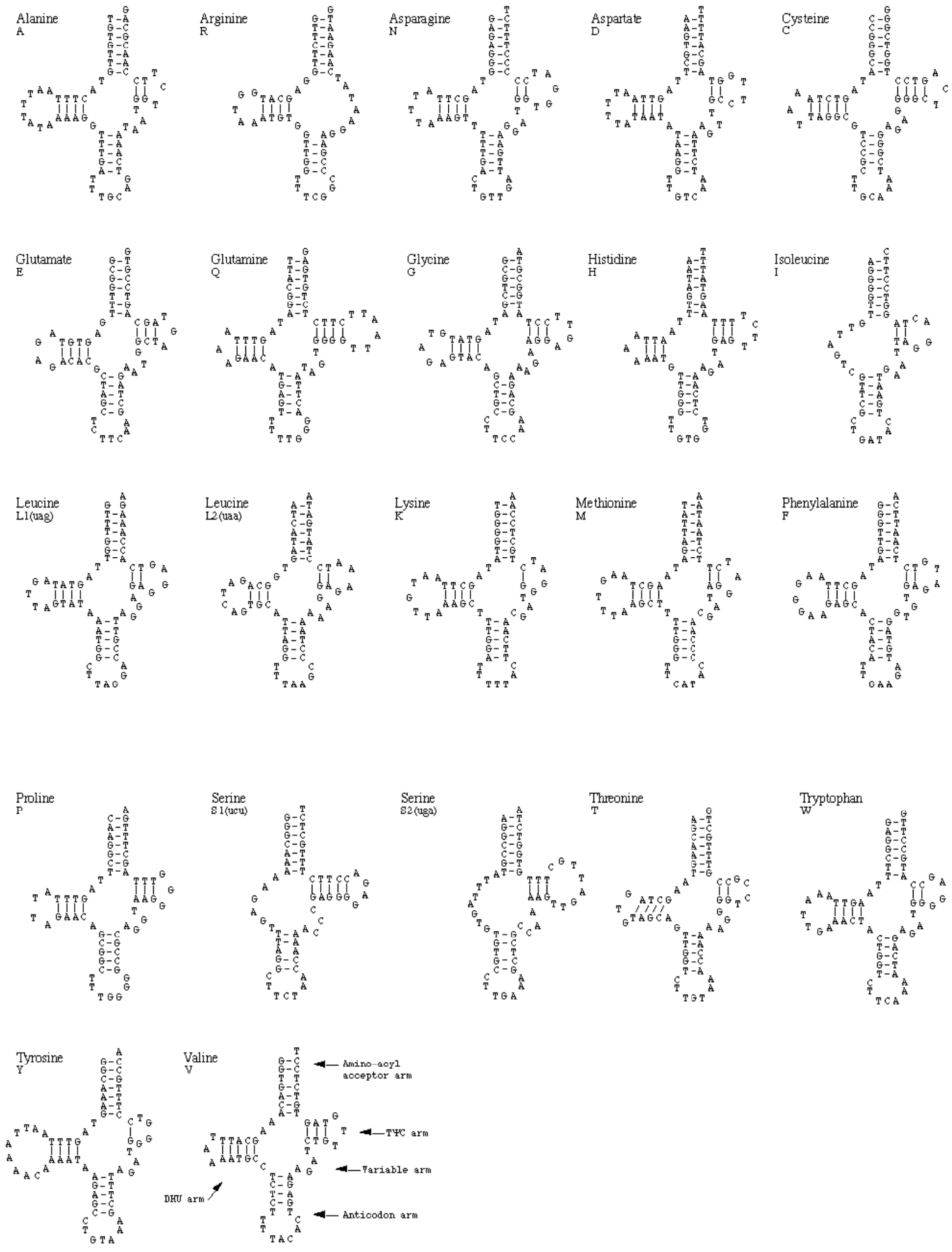
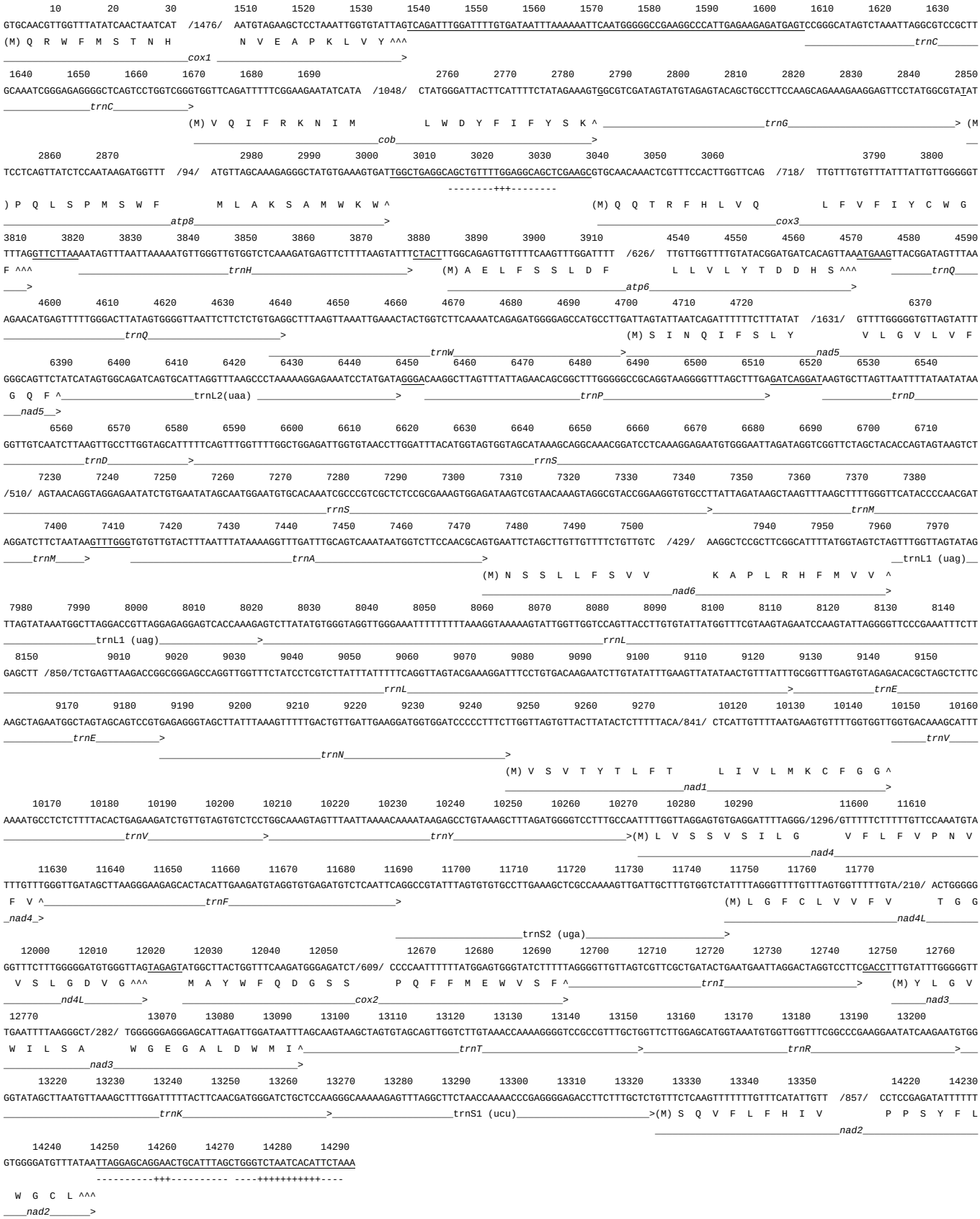

Katharina
nad3
nad2
cox1
cox2
atp8
atp6
nad5
nad4
nad4L
cob
nad6
nad1
rrnL
rrnS
cox3
Terebratulina
nad3
nad2
cox1
cox2
atp8
atp6
rrnS
rrnL
nad1
nad6
cob
nad4L
nad4
nad5
cox3
Laqueus
atp8
nad2
cox2
rrnL
rrnS
nad1
nad4
cox1
cob
atp6
nad5
nad6
nad3
nad4L
cox3
Terebratalia
atp6
nad5
rrnS
nad6
rrnL
nad1
nad4
nad4L
cox2
nad3
nad2
cox1
cob
atp8
cox3

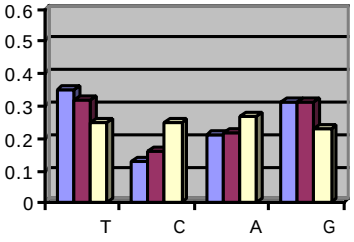
Figure 3


Alanine
A
Arginine
R
Asparagine
N
Aspartate
D
Cysteine
C
Glutamate
E
Glutamine
Q
Glycine
G
Histidine
H
Isoleucine
I
Leucine
L1(uag)
Leucine
L2(uaa)
Lysine
K
Methionine
M
Phenylalanine
F
Proline
P
Serine
S1(ucu)
Serine
S2(uga)
Threonine
T
Tryptophan
W
Tyrosine
Y
Valine
V
Amino-acyl
acceptor arm
TΨC arm
Variable arm
Anticodon arm
DHU arm

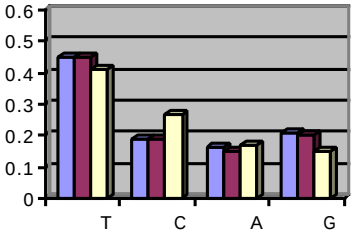


```
Terebratulina     TTGGGACTGATTTCTGAGAAATCTTTGTGGGATTTTTCGGGGATAATCGGCTAATGTGGGGGTTATGG
Terebratalaia     TCAGATTTGGATTTTTGTGATAATTTAAAAAATTCAATGGGGGCCGAAGGCCCATTGAGAAGAGATGAGT
Laqueus              CCTTTTGGTGTTTGAAATTTTTTTTATTAGTGGGGCGGGAGCGAAA-----ATTAAAAAA
```

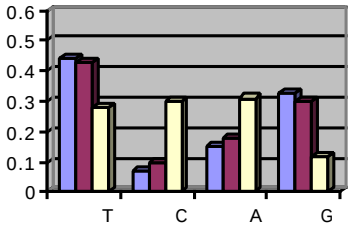
First Position

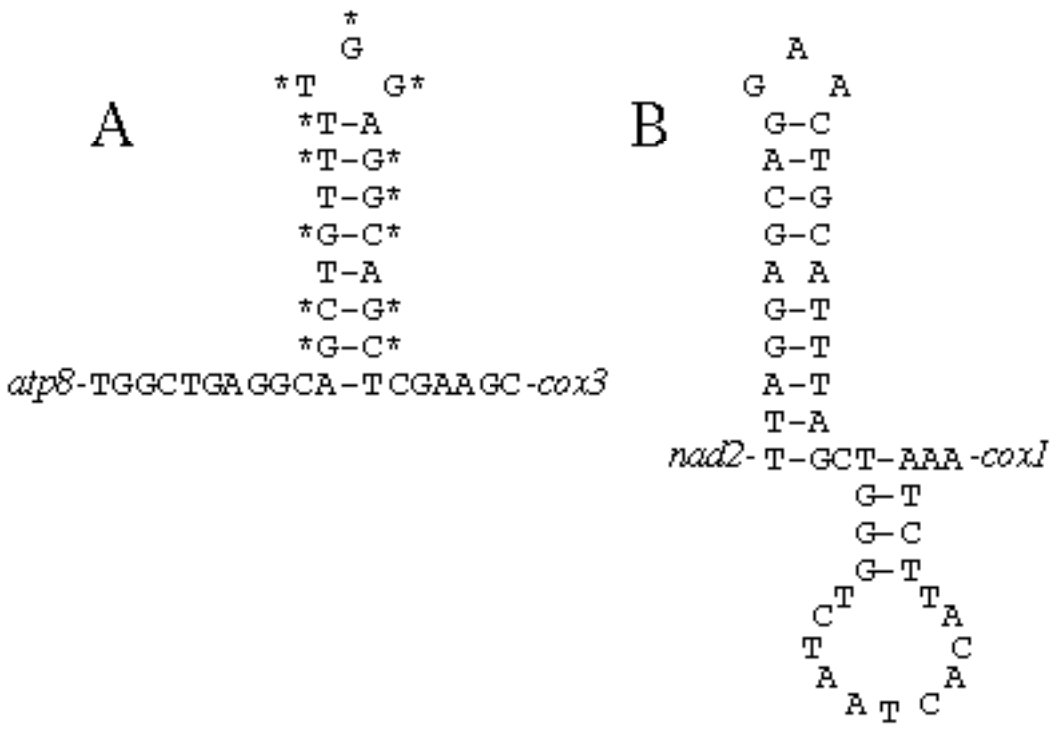


Second Position



Third Position




A
*
G
T G
*T-A
T-G
T-G*
G-C
T-A
C-G
G-C
atp8-TGGCTGAGGCA-TCGAAGC-cox3
B
A
G A
G-C
A-T
C-G
G-C
A A
G-T
G-T
A-T
T-A
nad2-T-GCT-AAA-cox1
G-T
G-C
G-T
C T T A
T C
A A
A T C