The Mitochondrial Genome of the Brachiopod *Laqueus rubellus*

Yasuhiro Noguchi,* Kazuyoshi Endo,† Fumio Tajima‡ and Rei Ueshima‡,§

**Institute of Biological Sciences, University of Tsukuba, Tsukuba 305-0006, Japan, †Geological Institute, University of Tokyo, Tokyo 113-0033, Japan, ‡Department of Biological Sciences, University of Tokyo, Tokyo 113-0033, Japan and §PRESTO, Japan Science and Technology Corporation, Kawaguchi, Saitama 332-0012, Japan*

Manuscript received November 3, 1999
Accepted for publication January 7, 2000

ABSTRACT

The complete nucleotide sequence of the 14,017-bp mitochondrial (mt) genome of the articulate brachiopod *Laqueus rubellus* is presented. Being one of the smallest of known mt genomes, it has an extremely compact gene organization. While the same 13 polypeptides, two rRNAs, and 22 tRNAs are encoded as in most other animal mtDNAs, lengthy noncoding regions are absent, with the longest apparent intergenic sequence being 54 bp in length. Gene-end sequence overlaps are prevalent, and several stop codons are abbreviated. The genes are generally shorter, and three of the protein-coding genes are the shortest among known homologues. All of the tRNA genes indicate size reduction in either or both of the putative TΨC and DHU arms compared with standard tRNAs. Possession of a TV (TΨC arm-variable loop) replacement loop is inferred for tRNA(R) and tRNA(L-tag). The DHU arm appears to be unpaired not only in tRNA(S-tct) and tRNA(S-tga), but also in tRNA(C), tRNA(I), and tRNA(T), a novel condition. All the genes are encoded in the same DNA strand, which has a base composition rich in thymine and guanine. The genome has an overall gene arrangement drastically different from that of any other organisms so far reported, but contains several short segments, composed of 2–3 genes, which are found in other mt genomes. Combined cooccurrence of such gene assortments indicates that the Laqueus mt genome is similar to the annelid Lumbricus, the mollusc Katharina, and the octocoral Sarcophyton mt genomes, each with statistical significance. Widely accepted schemes of metazoan phylogeny suggest that the similarity with the octocoral could have arisen through a process of convergent evolution, while it appears likely that the similarities with the annelid and the mollusc reflect phylogenetic relationships.

THE genome organization of animal mitochondrial (mt) DNA has been studied to have insights into regulation mechanisms of the mitochondrial genetic system and to estimate evolutionary processes of the system itself or of the organisms that carry it. To date, complete nucleotide sequences for mt genomes have been reported for 87 species spreading over eight animal phyla (Boore 1999). Animal mt genomes generally consist of a closed circular DNA molecule, 14–17 kbp in length, containing in a compact form the same set of genes for 13 polypeptides for energy pathway proteins plus 2 ribosomal RNAs and 22 transfer RNAs required for their synthesis, as well as a noncoding region of various lengths reserved for transcription and replication controls (Wolstenholme 1992).

In addition to these more or less uniform properties, there are features that mtDNAs of many animals have in common, but there are considerable and rather systematic variations among higher-order taxa. Notable ones include relative gene order, modified genetic code, and variant structures of tRNAs and rRNAs among others (Wolstenholme 1992). These characters have been suggested as potentially useful for unveiling metazoan deep-branch phylogenies, with the relative gene order having been of particular interest due to its apparent stability and consistency over the time span separating major animal lineages (Brown 1985; Moritz *et al.* 1987; Jacobs *et al.* 1988b; Boore and Brown 1994b).

For example, chordates so far examined commonly share the same mt gene order, although minor variations exist (Desjardins and Morais 1990; Pääbo *et al.* 1991; Janke *et al.* 1994; Kumazawa and Nishida 1995; Lee and Kocher 1995). Likewise, gene arrangements are conserved within echinoderms (Jacobs *et al.* 1988a; Cantatore *et al.* 1989; Asakawa *et al.* 1995) and within arthropods (Clary and Wolstenholme 1985a; Beard *et al.* 1993; Crozier and Crozier 1993; Mitchell *et al.* 1993; Valverde *et al.* 1994), both with few variations, which nevertheless provide keys to assess relationships among major groups in each of these phyla (Smith *et al.* 1993; Boore *et al.* 1995, 1998).

Disregarding the tRNA genes, the positions of which are more variable than those of the other genes (Jacobs *et al.* 1988b; Wolstenholme 1992), there are obvious similarities among mt gene orders of chordates, echinoderms, arthropods, a mollusc (*Katharina tunicata*; Boore and Brown 1994a), and an annelid (*Lumbricus terrestris*;

Corresponding author: Kazuyoshi Endo, Geological Institute, University of Tokyo, 7-3-1 Hongo, Tokyo 113-0033, Japan.
E-mail: endo@geol.s.u-tokyo.ac.jp

Boore and Brown 1995). The common chordate mt gene map is most similar to the echinoid map, requiring only two rearrangements to interconvert the maps, and the echinoid map can be converted to the sea star map by a single inversion step. In similar reciprocal comparisons assuming minimum evolution, the arthropod map is united with the chordate map by three transformations. The arthropod map is then connected, by two steps, with the mollusc map, with which the annelid map is linked by a minimum of four consecutive transformations (instead of five simultaneous rearrangements reported by Boore and Brown 1995). Lack of an appropriate outgroup hinders rooting of the tree, but the conserved segment in a coral mt genome (Beaton *et al.* 1998) tends to indicate that the root resides between chordates and arthropods, in effect supporting deuterostomes, protostomes, and the eutrochozoans (Ghiselin 1988) as a coherent group.

The animal mt gene order, however, is not always conserved as exemplified above. The gene maps of cnidarians (Beagley *et al.* 1998; Beaton *et al.* 1998; Pont-Kingdon *et al.* 1998), except for the above-mentioned segment, and of nematodes (Okimoto *et al.* 1991, 1992) are quite different from those of other phyla. There is mounting evidence that gene rearrangements have taken place frequently in molluscan mt genomes (Hoffmann *et al.* 1992; Hatzoglou *et al.* 1995; Terrett *et al.* 1996; Yamazaki *et al.* 1997). Those unconserved cases may be informative for phylogenetic analyses of lower-order taxa (Boore and Brown 1994b; Yamazaki *et al.* 1997) and perhaps also for elucidating the mechanisms of mt gene rearrangements, which still remain unclear.

In this context, we initially hoped to find reliable phylogenetic hallmarks in the mt gene order of the brachiopod *Laqueus rubellus*, of which the complete DNA sequence is presented here. Brachiopods constitute one of the lophophorate phyla, which are of considerable importance in understanding patterns of animal evolution, having both deuterostome and protostome features, thus having occupied particularly controversial positions in schemes of metazoan phylogeny (*cf.* Willmer 1990). In contrast with our expectations, however, the Laqueus mt genome was revealed to have a gene order radically different from that of any known genomes. But a closer inspection of shared gene boundaries suggests that some phylogenetic signatures are still retained. In addition, the genome shows some other unexpected features, such as unusual inferred structures of tRNAs, which are also described. This article represents one of the first reports on a complete mtDNA sequence for the phylum Brachiopoda (*cf.* Stechmann and Schlegel 1999).

MATERIALS AND METHODS

Specimens of *L. rubellus* (Sowerby) were collected by dredging from Sagami Bay, central Japan (35° 07.9′ N; 139° 35.1′ E; 80–85 m in water depth), and kept in an aquarium until subsequent treatment. Total DNA was extracted from whole tissue of a single female individual using the CTAB method as described by Fisher and Skibinski (1990).

A partial sequence for the *cox1* gene was determined first by amplifying the gene fragment using polymerase chain reaction (PCR) with "universal" primers (M. Saito, S. Kojima and K. Endo, unpublished results). Based on this sequence, a pair of primers was designed to amplify, by means of long-and-accurate PCR (LA-PCR; Barnes 1994), the rest of the entire mitochondrial DNA that was presumed and confirmed to have a circular conformation. The primers had the following sequences (5′ to 3′): COI-U (40-mer), GGA TCC GAT ACT TAC TAT GTT ACC GCC CAT TTT CAT TAT G; and COI-R (40-mer), AAG CTT GAA GAA CCG TAT CCA AAG AAG AAT TAG CTA AAA C. The extracted DNA was reconstituted in 20 μl of H_2O, and 1 μl was used in the LA-PCR reaction (50 μl), which also included 0.2 μm of each primer (COI-U and COI-R), 1 × LA-PCR buffer (Takara, Tokyo, Japan), 250 μm dNTP, and 2.5 units of *Ex Taq* DNA polymerase (Takara). PCR was performed using a Cetus DNA Thermal Cycler (Perkin Elmer, Norwalk, CT) with an initial step at 94° for 1 min followed by 30 cycles at 98° for 15 sec, and 68° for 15 min.

The amplified product (*ca.* 14 kbp) was digested with *Hin*dIII (Takara), ligated into pUC18 vector using DNA ligation kit ver. 2 (Takara), and transformed into *Escherichia coli* (strain DH5α) according to the methods described by Sambrook *et al.* (1989). The two fragments containing either end of the PCR product were processed to have phosphorylated blunt ends using the SureClone kit (Pharmacia, Tokyo, Japan), followed by digestion with *Hin*cII and subcloning into the *Sma*I site of pUC18. For the clones with an insert longer than 700 bp, a series of nested deletion mutants were prepared using Exonuclease III and Mungbean nuclease (both from Takara) to facilitate DNA sequence determinations (Henikoff 1984).

Cloned fragments were sequenced using the ABI (Urayasu, Japan) Dye Primer CS^+ or FS kit and an ABI Prism 373A automated sequencer. Contiguous sequences were assembled and the consensus sequence analyzed using GENETYX-MAC (Software Development, Tokyo, Japan). Protein and rRNA genes were identified by comparisons with corresponding known sequences of other metazoan taxa, including an annelid (*L. terrestris*; Boore and Brown 1995), a chiton (*K. tunicata*; Boore and Brown 1994a), *Drosophila yakuba* (Clary and Wolstenholme 1985a), a sea urchin (*Strongylocentrotus purpuratus*; Jacobs *et al.* 1988a), and human (Anderson *et al.* 1981). Transfer RNA genes were identified in the remaining sequences by a search of appropriate predicted secondary structures. The complete nucleotide sequence for *L. rubellus* mtDNA reported in this article is available from DDBJ/EMBL/GenBank database (accession no. AB035869).

RESULTS AND DISCUSSION

Genome organization: The gene content and organization of Laqueus mtDNA are summarized in Table 1. The genome size (14,017 bp) represents one of the smallest of known metazoan mt genomes. Comparably small mt genomes include those of nematodes *Caenorhabditis elegans* (13,794 bp) and *Ascaris suum* (14,284 bp; Okimoto *et al.* 1992), and of land snails *Cepaea nemoralis* (14,100 bp; Terrett *et al.* 1996), *Albinaria coerulea* (14,130 bp; Hatzoglou *et al.* 1995), and *Euhadra herklotsi* (14.5 kbp; Yamazaki *et al.* 1997). Size reductions in both coding and noncoding sequences account for the small size of the Laqueus mt genome, as in the land snail *A. coerulea* mt genome (Hatzoglou *et al.*

TABLE 1

Inferred organization of the *L. rubellus* mt genome

Gene	Position	Size	Start	Stop	3′ spacer
cox1	1–1539	1539	CTG	TAG	1
trnV	1541–1598	58	—	—	−1
cob	1598–2705	1108	ATG	T	0
trnH	2706–2765	60	—	—	6
atp6	2772–3452	681	ATG	TAA	1
trnQ	3454–3516	63	—	—	−1
trnW	3516–3576	61	—	—	−1
nad5	3576–5255	1680	GTG	TAG	−2
trnA	5254–5315	62	—	—	−1
nad6	5315–5780	466	ATG	T	0
trnI	5781–5834	54	—	—	−1
nad3	5834–6170	337	GTG	T	0
trnT	6171–6227	57	—	—	−1
trnR	6227–6280	54	—	—	−1
trnF	6280–6340	61	—	—	−3
trnE	6338–6396	59	—	—	−2
trnK	6395–6454	60	—	—	1
trnS(tga)	6456–6506	51	—	—	0
nad4L	6507–6761	255	ATT	TAA	5
trnG	6767–6825	59	—	—	−1
cox3	6825–7604	780	ATG	TAA	−2
trnD	7603–7662	60	—	—	0
atp8	7663–7818	156	ATT	TAG	−2
trnS(tct)	7817–7872	56	—	—	8
nad2	7881–8729	849	ATT	TAA	3
cox2	8733–9402	670	ATG	T	0
trnL(tag)	9403–9459	57	—	—	0
rnl	9460–10601	1142	—	—	0
rns	10602–11369	768	—	—	0
trnM	11370–11432	63	—	—	0
trnL(taa)	11433–11493	61	—	—	0
trnP	11494–11553	60	—	—	−1
nad1	11553–12441	889	GTG	T	0
trnY	12442–12500	59	—	—	−1
nad4	12500–13847	1348	ATG	T	0
trnC	13848–13905	58	—	—	54
trnN	13960–1	59	—	—	−1

The sequence is numbered to start from the first nucleotide of the *cox1* gene. All the genes are encoded in the same DNA strand. Gene identities: *atp6* and *atp8*, ATP synthase subunits 6 and 8; *cox1-3*, cytochrome *c* oxidase subunits I-III; *cob*, cytochrome *b*; *nad1-6* and *nad4L*, NADH dehydrogenase subunits 1-6 and 4L; *rns* and *rnl*, small and large subunits of ribosomal RNA; *trnX*, transfer RNA genes with corresponding amino acids denoted by one-letter codes (*X*). The two tRNA genes for either serine and leucine are further differentiated by the anticodon.

1995), but unlike the nematode mt genomes, in which noncoding sequences are not particularly truncated (Table 2). Intergenic regions in Laqueus mtDNA are few and short, and a total of 16 gene boundaries indicate sequence overlap, obviously contributing to the small genome size, but the total number of overlapped nucleotides is in fact smaller than that of many other animal mt genomes (Table 2).

Laqueus mtDNA contains genes for 13 polypeptides [*cox1–3*, *cob*, *atp6*, *atp8*, *nad1–6*, and *nad4L*: for the standard genetic nomenclature of mitochondrial genes, see Commission on Plant Gene Nomenclature (1994); see also Boore (1999)], two ribosomal RNAs (*rns* and *rnl*), and 22 tRNAs typical of animal mtDNAs. All these genes are transcribed from the same strand of the DNA molecule. This is also the case in sea anemone (Beagley *et al.* 1998), nematode (Okimoto *et al.* 1991, 1992), the mollusc Mytilus (Hoffmann *et al.* 1992), and annelid (Boore and Brown 1995) mtDNAs (Table 2).

The overall gene order of the Laqueus mt genome is unique, so that there is no simple solution to interconvert the Laqueus mt gene map to any known maps of other animals (Figure 1). However, there exist some local gene arrangements in Laqueus mtDNA that are shared with other mtDNAs, and those instances are summarized in Table 3. Disregarding the tRNA genes and noncoding sequences, the gene arrangement *nad6-nad3-nad4L*, where the genes are transcribed in this order, is shared with the octocoral cnidarian *Sarcophyton glaucum* mtDNA, which also has the arrangement *rns-nad1* found in Laqueus mtDNA. The assortment of the three genes *cob-atp6-nad5* is shared by Laqueus and the annelid *L. terrestris* mtDNAs. A couple of two-gene segments are shared with mtDNAs of the nematode *Meloidogyne javanica* (*atp6-nad5*; *rns-nad1*), *C. elegans* and *A. suum* (*rns-nad1*; *nad4-cox1*), and the bivalve mollusc *Mytilus edulis* (*nad5-nad6*; *nad1-nad4*). The gene arrangement *atp6-nad5* is also found in the sea anemone *Metridium senile* mtDNA, and thus is shared by mt genomes of organisms from four phyla, *i.e.*, a brachiopod, an annelid, a nematode, and a cnidarian. The segment *nad3-nad4L* is also found in human and other vertebrate mtDNAs and hence is shared by mt genomes of a brachiopod, a cnidarian, and chordates. The segment *rns-nad1* is shared by mtDNAs of a brachiopod, a cnidarian, and nematodes.

Positions of tRNA genes in different mt genomes are much more variable than those of other genes (Jacobs *et al.* 1988b; Wolstenholme 1992), but even including tRNA genes, several gene boundaries are shared between mtDNAs of Laqueus and other animals. Most noteworthy is the annelid Lumbricus mtDNA, which shares with Laqueus as many as five identical two-gene segments, *i.e.*, *trnF-trnE*, *trnG-cox3*, *trnD-atp8*, *trnS(tct)-nad2*, and *nad4-trnC*, in each of which genes are transcribed from left to right. The chiton Katharina shares three identical two-gene assortments [*trnS(tct)-nad2*, *trnQ-trnW*, and *trnI-nad3*] with Laqueus. The mtDNAs of the land snails *C. nemoralis*, *A. coerulea*, and *E. herklotsi* have two gene boundaries (*cox1-trnV*, *trnA-nad6*) in common with Laqueus mtDNA, while those of nematodes (*C. elegans* and *A. suum*) and human (and most other vertebrates) share a single two-gene segment (*nad5-trnA* and *trnC-trnN*, respectively) with Laqueus mtDNA. The gene arrangements *trnD-atp8* and *trnS(tct)-nad2*, which are shared by Laqueus and Lumbricus mtDNAs, are also found in mtDNAs of *D. yakuba* (and most other arthropods) and the mollusc *K. tunicata*, respectively.

TABLE 2

A comparison of genome structures among representative animal mtDNAs

Species (phylum)	Total length	Protein genes	rRNA genes	tRNA genes	Noncoding nucleotides	Sequence overlap	Coding strand	Gene order
Laqueus rubellus (B)	14017	10758	1910	1292	79 (54)	−22 (16)	S	>6
Caenorhabditis elegans (N)	13794	10292	1650	1239	613 (466)	0	S	>6
Ascaris suum (N)	14284	10279	1661	1252	1092 (886)	0	S	>6
Cepaea nemoralis (M)	14100	10728	1925	1207	404 (158)	−164 (13)	B	>6
Albinaria coerulea (M)	14130	10792	1794	1451	141 (42)	−48 (9)	B	>6
Mytilus edulis (M)	*ca.* 17100	*ca.* 11749	2189	1519	1643 (1158)	0	S	>6
Katharina tunicata (M)	15532	11202	2101	1446	817 (424)	−34 (10)	B	2
Lumbricus terrestris (An)	14998	11127	2030	1395	457 (384)	−11 (3)	S	6
Drosophila yakuba (Ar)	16019	11218	2115	1465	1258 (1077)	−37 (6)	B	0
Strongylocentrotus purpuratus (E)	15650	11504	2408	1545	240 (121)	−47 (8)	B	5
Homo sapiens (Ch)	16569	11397	2513	1512	1213 (1122)	−66 (10)	B	3
Metridium senile (Cn)	17443	12600	3271	141	1431 (324)	0	S	>6
Sarcophyton glaucum (Cn)	18453	15045	2956	71	420 (111)	−39 (3)	B	>6

For each mt genome, total lengths (in base pairs) of the genome, protein genes, rRNA genes, tRNA genes, noncoding sequences (the longest one in parentheses), and overlapped nucleotides (in minus value; total number of overlapped gene boundaries in parentheses) are shown. Genomes in which all the genes are encoded in the same DNA strand are noted as S, and cases where genes are encoded in both DNA strands are noted as B. Gene order is expressed by the minimum number of rearrangements to interconvert the gene map (excluding tRNA genes) to that of Drosophila. References: *L. rubellus*, this study (databank accession no. AB035869); *C. elegans*, Okimoto *et al.* (1992) (X54252); *A. suum*, Okimoto *et al.* (1992) (X54253); *C. nemoralis*, Terrett *et al.* (1996) (U23045); *A. coerulea*, Hatzoglou *et al.* (1995) (X83390); *M. edulis*, Hoffmann *et al.* (1992) (M83756-62); *K. tunicata*, Boore and Brown (1994a) (U09810); *L terrestris*, Boore and Brown (1995) (U24570); *D. yakuba*, Clary and Wolstenholme (1985a) (X03240); *S. purpuratus*, Jacobs *et al.* (1988a) (X12631); *H. sapiens*, Anderson *et al.* (1981) (J01415); *M. senile*, Beagley *et al.* (1998) (AF000023); *S. glaucum*, Pont-Kingdon *et al.* (1998); Beaton *et al.* (1998) (AF063191, AF064823). Abbreviations for the phyla: B, Brachiopoda; N, Nematoda; M, Mollusca; An, Annelida; Ar, Arthropoda; E, Echinodermata; Ch, Chordata; Cn, Cnidaria.

Statistical significance of those shared gene arrangements is discussed later.

The sense strand of Laqueus mtDNA is 20.8% adenine, 15.2% cytosine, 26.5% guanine, and 37.6% thymine. The A + T content (58.4%) is within the range reported for other animal mt genomes, but is the lowest among invertebrate mt genomes (the A + T content ranges in chordates from 55.6 to 63.2%; echinoderms, 58.9–61.3%; molluscs, 59.8–70.7%; annelid, 61.6%; cnidarians, 62.5–64.5%; nematodes, 72.0–76.2%; and arthropods 77.4–84.9%). The G + T content of the sense strands (64.1%) is also within the reported range, but is closer to the higher end (70.2%; *A. suum*). The base composition in codon third positions (A, 17.7%; C, 9.4%; G, 30.2%; and T, 42.7%) clearly indicates a bias toward a high relative frequency of G + T, a condition that could be related to the unique mechanism of asymmetric replication in animal mtDNAs (Asakawa *et al.* 1991). It might also help secondary structure formation for the transcripts from the sense strand via guanine-uracil pairing, as postulated for the *A. suum* mtDNA (Wolstenholme *et al.* 1987). The sense strand base composition of the protein genes (A, 18.3%; C, 14.8%; G, 27.1%; and T, 39.9%) is similar to the overall sense strand composition.

Protein genes: The genes for 13 polypeptides (*cox1–3*, *cob*, *atp6*, *atp8*, *nad1–6*, and *nad4L*) of Laqueus mtDNA were identified by comparison of the inferred amino acid sequence and size similarities to those of known homologues. Based on comparisons of nucleotide and amino acid sequences of the *cox1* gene, brachiopod mtDNAs, including that of *L. rubellus*, have been inferred to employ the same modified mt genetic codes as in nematodes, arthropods, molluscs, and an annelid (M. Saito, S. Kojima and K. Endo, unpublished results); namely, AGA and AGG code for serine, TGA for tryptophan, and ATA for methionine. This inference is supported by the complete nucleotide sequence determined in this study.

Six protein-coding genes start with the orthodox translation initiation codon ATG, three genes (*nad4L*, *atp8*, *nad2*) with ATT, three (*nad5*, *nad3*, *nad1*) with GTG, and the remaining one with CTG (*cox1*). Among the genes for which the GTG translation initiation codon is inferred, *nad3* could alternatively start with ATT immediately after the GTG codon, or with ATG four codons downstream. An in-frame ATG codon exists 9 codons and 16 codons downstream of the GTG codon of putative *nad1* and *nad5* genes, respectively, but if it is taken as the initiation site for each gene, then in each case, a segment containing what appear to be conserved amino acid residues in comparison with Lumbricus, Katharina, and Drosophila needs to be left out. A similar argument applies to the *cox1* gene, which has an in-

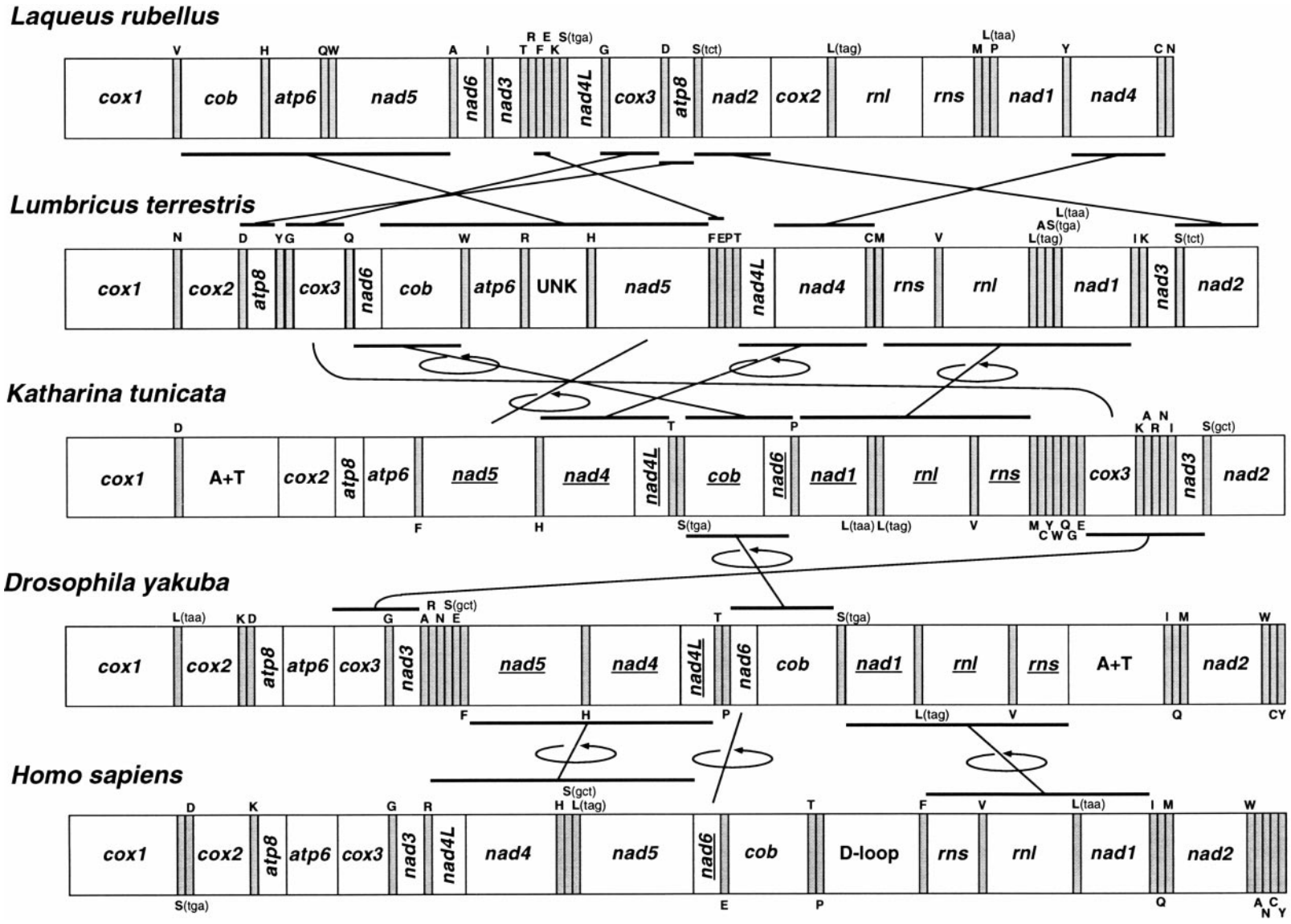

Figure 1.—A comparison of gene arrangements among Laqueus, Lumbricus, Katharina, Drosophila, and human mtDNAs. Each gene map is liberalized to commence from the *cox1* gene with the orientation in which the gene is transcribed from left to right. Abbreviations: A + T, the large adenine- and thymine-rich region; D-loop, displacement loop; UNK, the largest unassigned region of Lumbricus mtDNA. Transfer RNA genes are depicted by one-letter amino acid codes, either on top (when encoded in the same strand as *cox1*) or under the bottom (encoded in the opposite strand) of the columns. Other abbreviations are as in Table 1. Only the protein and rRNA genes that are transcribed in the opposite direction as in the *cox1* gene are underlined. The rearrangements that are needed to interconvert the pair of maps (disregarding tRNA genes) are shown, except for the comparison between Laqueus and Lumbricus, in which gene arrangements that are shared by both genomes (both including and excluding tRNA genes) are shown.

TABLE 3

Shared gene boundaries between mt genomes of *L. rubellus* and other animals

Genome(s) compared	No. of genes in shared segment	Expected no.		Observed no.	Poisson probability		Gene arrangement in shared segment
		Case I	Case II		Case I	Case II	
			Protein and rRNA genes only				
Sarcophyton glaucum	3	0.018	0.018	1	0.018	0.018	*nad6-nad3-nad4L*
S. glaucum	2	0.500	0.500	1	0.393	0.393	*rns-nad1*
S. glaucum	3 + 2	0.004	0.004	1	0.004	0.004	Combination of the above two
Lumbricus terrestris	3	0.082	0.021	1	0.079	0.021	*cob-atp6-nad5*
Meloidogyne javanica	2	1.000	0.500	2	0.263	0.090	*atp6-nad5; rns-nad1*
Caenorhabditis elegans (and *Ascaris suum*)	2	1.000	0.500	2	0.263	0.090	*nad4-cox1;rns-nad1*
Mytilus edulis	2	1.000	0.500	2	0.263	0.090	*nad5-nad6; nad1-nad4*
Homo sapiens	2	0.536	0.536	1	0.415	0.415	*nad3-nad4L*
Metridium senile	2	1.071	0.536	1	0.657	0.415	*atp6-nad5*
M. senile/*M. javanica*/*L. terrestris*	2	0.005	0.001	1	0.005	0.001	*atp6-nad5*
S. glaucum/*H. sapiens*	2	0.018	0.018	1	0.017	0.017	*nad3-nad4L*
S. glaucum/nematodes	2	0.038	0.019	1	0.037	0.019	*rns-nad1*
			All genes				
L. terrestris	2	1.028	0.514	5	0.004	2.0×10^{-4}	*trnF-trnE; trnG-cox3; trnD-atp8; trnS(tct)-nad2; nad4-trnC*
Katharina tunicata	2	0.514	0.514	3	0.015	0.015	*trnS(tct)-nad2; trnQ-trnW; trnI-nad3*
Cepaea nemoralis (and other snails)	2	0.514	0.514	2	0.094	0.094	*cox1-trnV; trnA-nad6*
Drosophila yakuba	2	0.514	0.514	1	0.402	0.402	*trnD-atp8*
H. sapiens	2	0.514	0.514	1	0.402	0.402	*trnC-trnN*
C. elegans (and *A. suum*)	2	1.000	0.500	1	0.632	0.393	*nad5-trnA*
L. terrestris/*D. yakuba*	2	0.014	0.007	1	0.014	0.014	*trnD-atp8*
L. terrestris/*K. tunicata*	2	0.014	0.007	1	0.014	0.014	*trnS(tct)-nad2*

See text for explanations for expected frequencies and derived Poisson probabilities of occurrence of shared gene boundaries. Abbreviations for the gene identities as in Table 1.

frame ATG codon five codons downstream of the inferred CTG start codon. Seven genes end in a complete termination codon, either TAG or TAA (Table 1). The remaining six genes, each of which is immediately followed by a tRNA gene, are inferred to terminate with an incomplete stop codon, T (for a review on unorthodox translation initiation and termination codons of metazoan mt protein genes, see Wolstenholme 1992).

The proteins of the inferred lengths are generally shorter than most other previously described ones (Table 4), and three of them (Atp8, Cox2, Nad2) have either the same or a shorter size relative to the shortest known homologue. In Laqueus, Nad1, Nad2, Nad4L, and Nad6 are 9, 17, 12, and 11%, respectively, shorter than in Drosophila. Size differences for other proteins are within 5% variation in comparison with Drosophila, Katharina, and Lumbricus.

All codons are used in the 13 protein genes of Laqueus mtDNA (Table 5). In fourfold synonymous codon families, either T or G is the most frequently used nucleotide at the third codon position. In codon groups ending at purine and in those ending at pyrimidine, G and T are always used more frequently at the third codon position than A and C, respectively. In total, 72.9% of codons end at T or G. The most frequently used codon is TTT (Phe), followed by GTT (Val), GGG (Gly), and TTG (Leu), all of which consist exclusively of T and/or G (Table 5). On the other hand, the most frequently used amino acid in the 13 protein genes of Laqueus mtDNA is Leu (14.9%), followed by Val (12.1%), Ser (10.5%), Gly (9.5%), and Phe (8.0%). Among these amino acids, Val and Gly exhibit considerably higher relative frequency values compared with the ranges observed for other animal mt genomes (Val, 4.2–8.3%; Gly, 5.5–7.6%: Wolstenholme 1992). It thus appears that preference for the usage of G and T somewhat affected the amino acid composition of mitochondrially encoded proteins in Laqueus mtDNA.

Comparisons of the percentage of amino acid identity between mt protein genes of Laqueus and those of human, sea urchin (*S. purpuratus*), Drosophila, Katharina, and Lumbricus generally indicate that Laqueus se-

TABLE 4

Comparisons of sequence lengths and similarities among the mitochondrial protein genes of Laqueus, Lumbricus, Katharina, Drosophila, sea urchin, and human

Gene products	Sequence length (in amino acids)						Sequence similarities (%) *vs.*					
	Human	Sea urchin	Drosophila	Katharina	Lumbricus	Laqueus	Human	Sea urchin	Drosophila	Katharina	Lumbricus	
Atp6	226	229	224	230	231	226	21	26	27	38	35	
Atp8	68	55	53	53	53	51	12	25	32	26	33	
Cox1	513	517	512	513	513	512	70	69	71	71	70	
Cox2	227	229	228	229	228	223	50	50	51	55	49	
Cox3	261	260	262	259	259	259	56	57	58	66	63	
Cob	380	380	378	379	379	369	50	51	52	55	51	
Nad1	318	322	324	316	308	296	46	46	44	51	52	
Nad2	347	352	341	338	335	282	18	24	21	24	25	
Nad3	115	116	117	120	117	112	31	39	38	37	43	
Nad4	459	463	446	442	452	449	34	35	38	39	41	
Nad4L	98	97	96	100	98	84	24	34	23	26	38	
Nad5	603	637	573	571	574	559	25	29	37	35	41	
Nad6	174	165	174	166	156	155	18	25	17	21	28	

Sequence similarity was calculated by dividing the number of identical amino acid positions by the total length of the aligned sequences for each pair of sequences.

quences are more similar to protostome homologues than to deuterostome ones (Table 4). The *cox1* gene, the most conserved of all protein genes, of Laqueus shows almost the same extent of similarity to that of the five organisms compared. The *cox3*, *cob*, and *cox2* genes, the next conserved ones, and *atp6* indicate the highest similarities between Katharina and Laqueus. For the remaining eight genes, however, the highest identity is observed between Lumbricus and Laqueus.

Transfer RNA genes: Twenty-two tRNA genes typical of animal mtDNAs have been identified in the Laqueus mt genome. The inferred Laqueus mt-tRNAs have a number of uniform features that are invariant in standard tRNAs, such as possession of a 7-bp amino-acyl arm, a 5-bp anticodon stem, and a 4-bp variable loop [except in tRNA(L-tag) and tRNA(R)]; the nucleotide preceding the anticodon is T, which is preceded by a pyrimidine [except in tRNA(H)]; and the nucleotide after the anticodon is a purine. But they exhibit some notable aberrancy (Figure 2).

In the putative Laqueus mt-tRNAs for Arg and Leu (tag), the TΨC arm and variable loop are replaced by a single loop (TV replacement loop), as found in nematode mt-tRNAs (Wolstenholme *et al.* 1987; Okimoto *et al.* 1992). tRNAs with a TV replacement loop are also reported for land snail and mosquito mtDNAs (Mitchell *et al.* 1993; Terrett *et al.* 1996; Yamazaki *et al.* 1997), but those correspond to other codons. Furthermore, in other Laqueus mt-tRNAs that are inferred to have the TΨC arm and variable loop, the stem in the TΨC arm is generally short, being 1–3 bp in length [except in tRNA(S-tct) and tRNA(I)].

In the Laqueus mt-tRNAs for Cys, Ile, Ser(tct), Ser (tga), and Thr, the DHU arm is replaced by a loop. That tRNA(S-tct) has an unpaired DHU arm is a typical feature of animal mtDNAs (Wolstenholme 1992). Possession of the tRNA(S-tga) with an unpaired DHU arm is also reported for mtDNAs of nematodes (Okimoto *et al.* 1992), a chiton (Boore and Brown 1995), land snails (Yamazaki *et al.* 1997), and an annelid (Boore and Brown 1995). However, tRNAs for Cys, Ile, and Thr with an unpaired DHU arm have not been known among metazoan mt-tRNAs.

The anticodons of Laqueus mt-tRNA genes are identical with those of the annelid Lumbricus (Boore and Brown 1995) and the honeybee Apis (Crozier and Crozier 1993), which have the TTT anticodon for tRNA(K) and the TCT anticodon for the tRNA that corresponds to AGN codons of serine, giving evidence for the sharing of the same variant mt genetic codes between Laqueus and Lumbricus (and other protostomes as discussed in Boore and Brown 1995). The anticodon for the tRNA that corresponds to AGN codons of serine is also TCT in nematode mtDNAs, but in mtDNAs of most other animals, including platyhelminthes, molluscs, Drosophila, and echinoderms, the

TABLE 5

Codon usage in the 13 protein genes of Laqueus mtDNA

Amino acid (anticodon)	Codon group	Usage of codon ending in: A	C	G	T	Total	%
Ala (UGC)	GCN	23	23	24	125	195	5.5
Arg (UCG)	CGN	16	3	28	18	65	1.8
Asn (GUU)	AAY	—	20	—	54	74	2.1
Asp (GUC)	GAY	—	23	—	43	66	1.8
Cys (GCA)	TGY	—	7	—	47	54	1.5
Gln (UUG)	CAR	17	—	37	—	54	1.5
Glu (UUC)	GAR	23	—	55	—	78	2.2
Gly (UCC)	GGN	68	19	194	58	339	9.5
His (GUG)	CAY	—	24	—	51	75	2.1
Ile (GAU)	ATY	—	35	—	153	188	5.3
Leu (UAG)	CTN	47	18	51	88	204	5.7
(UAA)	TTR	135	—	193	—	328	9.2
Lys (UUU)	AAR	18	—	60	—	78	2.2
Met (CAU)	ATR	70	—	108	—	178	5.0
Phe (GAA)	TTY	—	34	—	252	286	8.0
Pro (UGG)	CCN	35	27	31	65	158	4.4
Ser (UGA)	TCN	37	17	23	150	227	6.4
(UCU)	AGN	40	7	62	37	146	4.1
Thr (UGU)	ACN	26	13	10	66	115	3.2
Trp (UCA)	TGR	22	—	88	—	110	3.1
Tyr (GUA)	TAY	—	34	—	91	125	3.5
Val (UAC)	GTN	56	32	115	230	433	12.2
Total		633	336	1079	1528	3576	100.1
%		17.7	9.4	30.2	42.7	100.0	

anticodon is GCT, rather than TCT (Wolstenholme 1992; Boore and Brown 1995).

Ribosomal RNA genes: The two mt-rRNA genes of Laqueus mtDNA, identified by comparisons of nucleotide similarities with those of other animals, are arranged side by side (*rnl-rns* in the encoded direction) in the mt genome without apparent coding sequences between them. In other known animal mt genomes, the two rRNA genes are intervened by at least one gene, which in many cases is *trnV*.

As is the case in many other mtDNAs, the precise boundaries of these genes remain uncertain. Figure 3 shows a comparison of the 5′ and 3′ regions with known homologues, for both genes aligned to the nucleotide sequence for the region containing the two rRNAs and the flanking upstream [*trnL(tag)*] and downstream (*trnM*) gene segments of Laqueus mtDNA. Close to the 3′ end of an *rns* gene is an 18-bp conserved region that is followed by an inverted repeat (underlines in Figure 3) corresponding to the final stem and loop structure of Drosophila SrRNA (Clary and Wolstenholme 1985b; Hatzoglou *et al.* 1995). In Laqueus, this composite region is entirely conserved and is immediately followed by the sequence for *trnM* (Figure 3); thus we consider that *rns* directly abuts *trnM* in Laqueus mtDNA. Similarly, close to the 5′ end of an *rns* gene is a 9-bp conserved region (GTTTGGTTC in Laqueus), which is preceded by three to five nucleotides from the 5′ end in other compared genes (Figure 3). From the sequence similarity with Drosophila in the preceding segments, we tentatively assign five nucleotides before the first conserved region for the Laqueus *rns* gene.

Comparisons of the 5′ regions of *rnl* genes among various animals reveal that the first conserved region, 19 bp in length, is observed after 122, 132, and 178 nucleotides from the 5′ end in Lumbricus, Katharina, and Drosophila *rnl* genes, respectively (Figure 4). In Laqueus, there are only 61 nucleotides separating this conserved region and the preceding sequence for *trnL(tag)*, and we assume that the 5′ region of the Laqueus *rnl* gene occupies all this available space.

The 3′ end of the Laqueus *rnl* gene leaves us with considerable ambiguity. The lengths of the nucleotides after the last widely conserved sequence of 13 bp (TAG TACGAAAGGA in Laqueus) in the 3′ regions of *rnl* genes in Lumbricus, Katharina, and Drosophila are 72, 29, and 54, respectively (Figure 3). Whereas, assuming that our interpretation of the 5′ end of *rns* is correct, the available number of nucleotides after that conserved region and before the next gene (*rns*) is 95. The 3′ end of Laqueus *rnl* appears to be at least downstream of the point corresponding to the 3′ end of Drosophila *rnl* because there is a 9-bp conserved sequence (ATTAA TATA) in this region that corresponds to the final stem

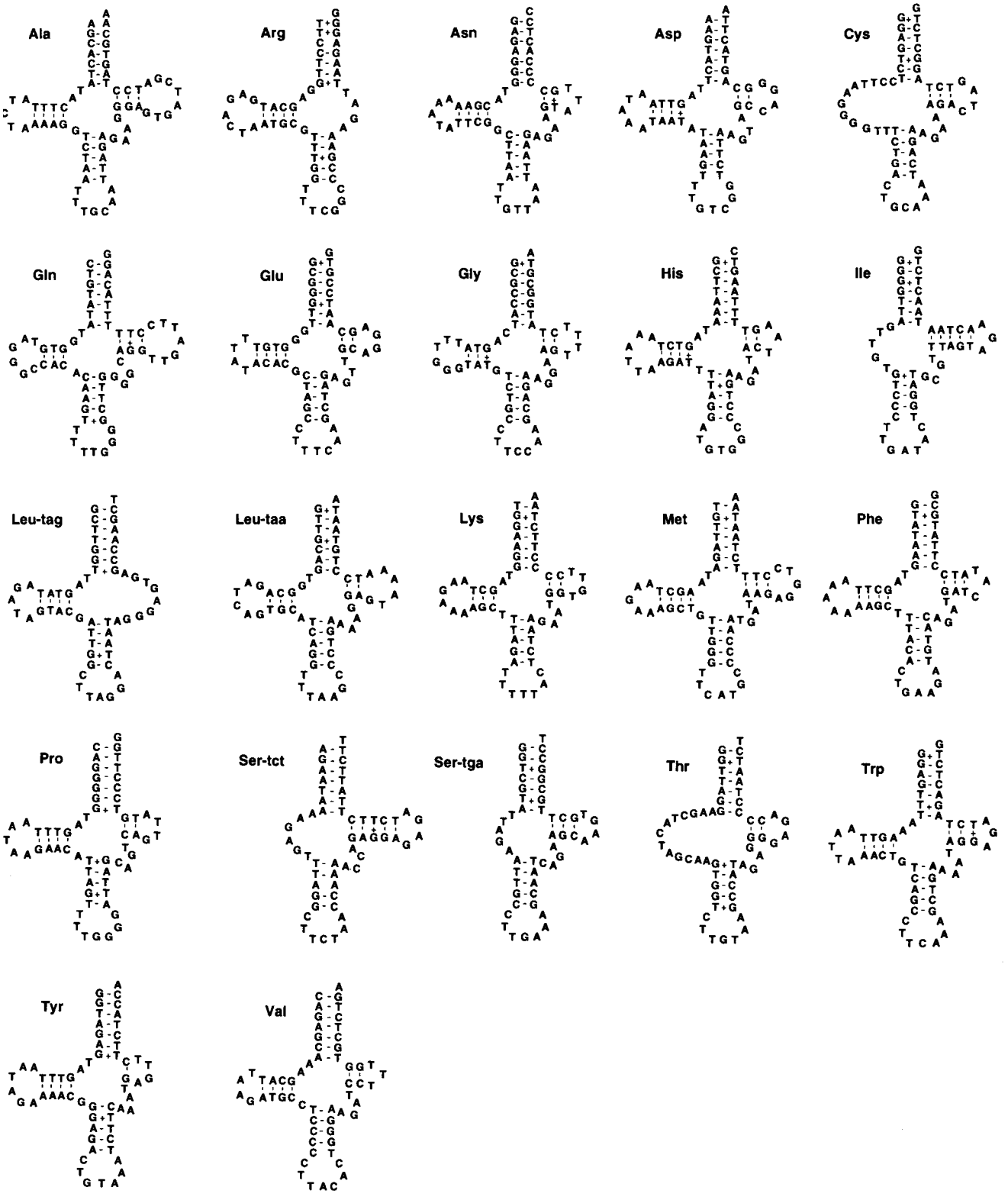

Figure 2.—Predicted secondary structures of the 22 mt-tRNAs of *L. rubellus.* Bars indicate standard base pairings. Nonstandard (G-T) pairings are indicated by crosses.

and loop structure of Drosophila LrRNA (Figure 3; Clary and Wolstenholme 1985b). Since there are no reliable reasons to invoke otherwise, we interpret that the *rnl* and *rns* genes abut directly in Laqueus mtDNA, making the 3′ region of Laqueus *rnl* gene considerably longer even than that in Lumbricus. A comparable length for this region is known, for example, for human *rnl.* However, considering the fact that other regions of

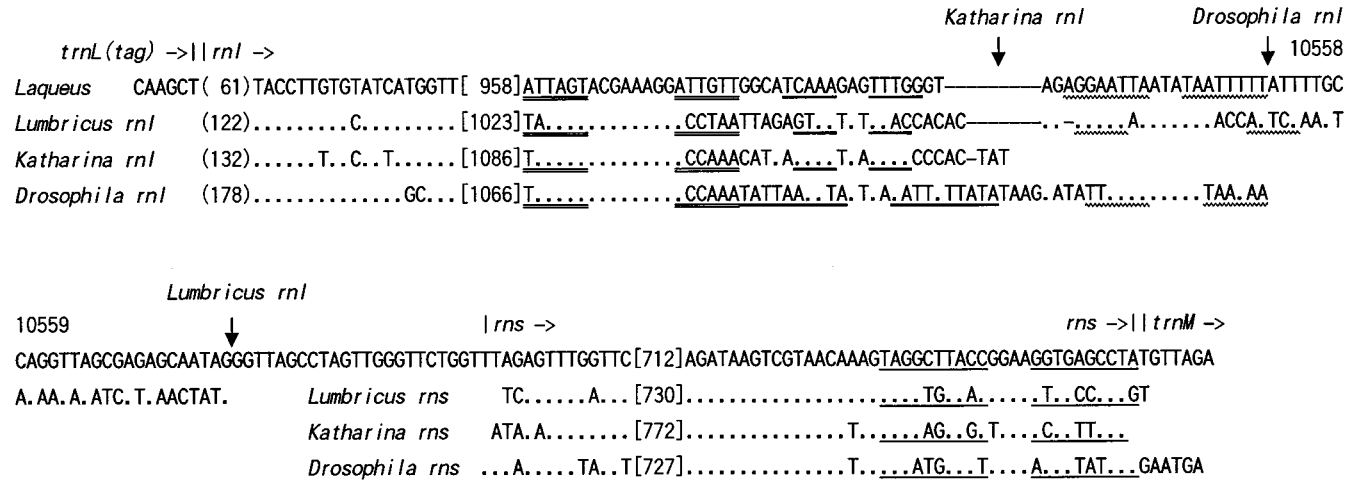

Figure 3.—Alignment of the ends of the two rRNA genes of Laqueus, Lumbricus, Katharina, and Drosophila mtDNAs. For Laqueus mtDNA, the nucleotide sequence for the region containing the two rRNAs and the flanking upstream [*trnL(tag)*] and downstream (*trnM*) gene segments are shown. Numbers in brackets denote the total length of omitted sequences between two conserved regions. Numbers in parentheses denote the total length of omitted regions, of which alignment is ambiguous. Nucleotide identities are depicted by dots. Inverse repeat sequences, corresponding to the last three hairpin structures in the proposed secondary structure of Drosophila LrRNA, are indicated by double, thick, and wavy underlines. Inverse repeat sequences, corresponding to the last hairpin structure in the proposed secondary structure of Drosophila SrRNA, are underlined. The 3′ ends of *rnl* genes of Lumbricus, Katharina, and Drosophila mtDNAs are indicated by arrowheads to the corresponding positions in Laqueus mtDNA.

the human *rnl* gene are significantly longer than the equivalent regions of Laqueus, and that other genes of Laqueus mtDNA indicate shorter sizes in general, it appears possible that the 3′ end of Laqueus *rnl* is actually farther upstream.

Noncoding sequences: The Laqueus mt genome is extremely compact. Out of the 37 gene boundaries of the genome, a total of 16 boundaries indicate sequence overlap, and gene pairs at 13 boundaries directly abut each other (Table 1). Among the remaining 8 boundaries, which accommodate a total of 79 unassigned nucleotides, only the 54-bp region between *trnC* and *trnN* contains a sequence longer than 10 bp.

The A + T content of the 54-bp region is 66.7%, which is higher than the average of the whole genome, but otherwise the region does not have typical features that are often found in the genome's largest noncoding region of other animal mtDNAs, such as a compelling potentiality to form a secondary structure, extensive polypurine and polypyrimidine tracts, certain conserved sequences, and direct repeat motifs (Jacobs *et al.* 1988a; Wolstenholme 1992). This poses a question as to the location of the regions that contain the signals necessary for replication and transcription controls, as addressed for the land snail Albinaria mtDNA, which also lacks lengthy unassigned sequences (Hatzoglou *et al.* 1995). The region between the two rRNA genes as mentioned above provides another possibility, but is an unconvincing candidate as is the 54-bp unassigned region.

Genome features and their interrelationships: The Laqueus mt genome exhibits some unusual features compared with other familiar animal mt genomes. Those include (1) small genome size, (2) absence of lengthy noncoding regions containing possible signals for transcription and replication, (3) truncated tRNA genes with aberrant inferred structures, (4) all the genes are encoded in the same DNA strand, and (5) absence of well-conserved gene arrangements compared to mt genomes of other phyla.

Since combinations of some of these features are also observed in mt genomes of other phyla, notably nematodes and molluscs (*cf.* Table 2), there exist grounds to suspect that at least some of them are interrelated with each other. The feature (1) is obviously not independent of (2) or (3), since the latter two directly contribute to the former, but their interdependence is not self-evident because small mt genomes do not always have the feature (2) as exemplified in nematodes *C. elegans* and *A. suum* (Okimoto *et al.* 1992) or the feature (3) as seen in the land snail *A. coerulea*, which has more or less standard tRNA structures (Hatzoglou *et al.* 1995). However, since the features (2) and (3) are found only in the mt genomes with the feature (1), we may consider that features (2) and (3) are properties of mt genomes with the feature (1), albeit they are not *the* properties of those mt genomes.

Out of the possible 10 combinations of the above-mentioned five features, none exhibits perfect bidirectional correlation among animal mt genomes compared (*cf.* Table 2). However, we note that a small genome of ~14 kbp in size always shows an unconserved gene order (the reverse does not hold as is evident in the case of Mytilus mtDNA). Correlation does not necessarily mean causal link, but it may be possible to argue that genome

size reduction resulted in extensive reorganization of the genome. If this is the case, then there should exist at least another mechanism that has led to the considerable gene rearrangements in the moderately sized Mytilus mtDNA.

Statistical significance of shared gene boundaries: Whatever the underlying mechanisms, it appears evident that much more gene rearrangement has taken place in the lineage leading to the analyzed Laqueus mtDNA than in other eucoelomate mtDNA lineages, such as those of arthropods, chordates, and echinoderms. However, as already described, there are several gene juxtapositions shared between Laqueus and other animal mtDNAs (Table 3). To assess statistical significance of these findings, we calculated probabilities that certain shared gene boundaries between different mt genomes arise purely by chance, on the basis of a similar reasoning applied to the issue of tRNA cluster conservation in echinoderm and other mtDNAs (Jacobs *et al.* 1989; *cf.* appendix).

First, the number of all the possible gene arrangements for a closed circular (cc) mtDNA containing a total of x genes is calculated. This may be given by linearizing the DNA at a given gene end (say 5′ end), fixing the gene at either orientation (say 5′ to 3′), then counting the number of permutations of the remaining genes, which is $(x - 1)!$, multiplied by 2^{x-1}, to take into account that each remaining gene can take two different orientations (designated as case II). If we assume that the genes can be encoded in only one and the same DNA strand due to some constraints (designated as case I), then the equivalent number would simply be $(x - 1)!$ The number of all the possible arrangements for mtDNA containing the usual set of 37 genes as in Laqueus would be $36! \times 2^{36}$, or 2.56×10^{52}. Disregarding the 22 tRNA genes, the number of possible different ways to arrange the remaining 13 protein and two rRNA genes would be 1.43×10^{15}. In case I, the equivalent numbers for the 37 genes and 15 genes become 3.72×10^{41} and 8.72×10^{10}, respectively.

We then calculate the chance that a given segment composed of y genes would be found by chance in a ccDNA composed of x genes. In case II, this is given by the number of ways a segment of y genes may be inserted in either direction into the ccDNA composed of the remaining x-y genes $[(x\text{-}y) \times 2]$, multiplied by the number of possible gene orders in the ccDNA of x-y genes $[(x\text{-}y - 1)! \times 2^{x-y-1}]$, divided by the total number of possible arrangements for the ccDNA containing x genes $[(x - 1)! \times 2^{x-1}]$, which is equivalent to $(x\text{-}y)!/[(x - 1)! \times 2^{y-1}]$. In case I, the equivalent value would be $(x\text{-}y)!/(x - 1)!$

The random probability of finding a given 3-gene segment with a given gene order for the genome of 15 genes would thus be $12!/(14! \times 2^2) = 0.00137$ in case II and $12!/14! = 0.00549$ in case I. For the genome of 16 genes excluding tRNA genes (as in the cnidarian *S. glaucum*, which has an extra protein gene in addition to the usual set of genes), the corresponding figure in case II would equal $13!/(15! \times 2^2) = 0.00119$. The random probability of finding a particular 3-gene segment in both a 15-gene genome (such as Laqueus) and a 16-gene genome (such as Sarcophyton) is therefore calculated as $0.00137 \times 0.00119 = 1.63 \times 10^{-6}$. Similarly, the probabilities that a particular 3-gene segment is found in both Laqueus and Lumbricus mtDNAs, both being 15-gene genomes (excluding tRNA genes) in which genes are encoded in the same DNA strand, are calculated as $0.00137 \times 0.00137 = 1.88 \times 10^{-6}$ in case II and $0.00549 \times 0.00549 = 3.01 \times 10^{-5}$ in case I.

On the other hand, the number of different kinds of arrangements that a y-gene segment in an x-gene genome ($y < x$) can take in case II is calculated as the number of permutations of y genes (${}_xP_y$), multiplied by the number of combinations for the individual gene directions (2^y), divided by the redundancy arisen by counting pairs of the same segments of different directions separately (2), which is equivalent to $x! \times 2^{y-1}/(x\text{-}y)!$ The equivalent number in case I would be ${}_xP_y$, or $x!/(x\text{-}y)!$. Thus there are $15! \times 2^2/12! = 10{,}920$ (case II) and $15!/12! = 2730$ (case I) kinds of 3-gene segments of different gene orders that can occur in a 15-gene genome. This leads to expectations that a total of $1.63 \times 10^{-6} \times 10{,}920 = 0.018$ 3-gene segments of the same gene orders is to be shared between Laqueus and Sarcophyton mtDNAs (excluding tRNA genes) and that a total of $1.88 \times 10^{-6} \times 10{,}920 = 0.021$ (case II), or $3.01 \times 10^{-5} \times 2730 = 0.082$ (case I), 3-gene segments of the same gene orders is to be shared between Laqueus and Lumbricus mtDNAs, solely by random associations.

If the total numbers of genes in the genomes are large enough compared with the size of the shared segments in question, the frequency of the shared segments of a certain size between two genomes arisen from random associations can be approximated by a Poisson distribution (see appendix). From the expected number of shared segments between genomes (m) as calculated above, the Poisson probabilities $[P(m, k) = m^k \times e^{-m}/k!]$ for each number of occurrences (of shared segments of a certain size), from 0, 1, 2 to k times, can now be computed. The probability that at least one 3-gene segment of the same gene order is shared between Laqueus and Sarcophyton mtDNAs would then be calculated as $1 - 0.018^0 \times e^{-0.018}/0! = 0.018$. The results of similar calculations for other shared segments between Laqueus and other animal mtDNAs are summarized in Table 3.

The results concerning only protein and rRNA genes indicate that although each of the bivalve Mytilus and three nematode mtDNAs has two two-gene segments in common with Laqueus mtDNA, these conditions are within the range expected to arise from random processes ($P = 0.090$), and the probabilities become even less significant assuming the coding strand constraint

($P = 0.263$). Sharing of a single two-gene segment, each with human and Metridium mtDNAs, is not statistically significant either ($P = 0.415$). However, the combined cooccurrence of a two-gene and a three-gene segment in both Laqueus and the octocoral Sarcophyton mtDNAs is highly significant ($P = 0.004$). Also significant is the sharing of a three-gene segment between Laqueus and the annelid Lumbricus mtDNAs ($P = 0.021$), although the probability becomes out of the value normally considered as significant (*i.e.*, $P = 0.05$) when we invoke the coding strand constraint ($P = 0.079$).

It therefore appears highly unlikely that these similarities with Sarcophyton and with Lumbricus (albeit being less significant) arose by chance. Similarities in gene orders are usually attributed to shared common ancestry. However, since it is difficult to envisage common ancestry among the diploblastic cnidarian Sarcophyton and the triploblastic coelomates Lumbricus and Laqueus, to the exclusion of other coelomates, such as arthropods, echinoderms, and chordates, it appears likely that the similarity with Sarcophyton, and quite possibly with Lumbricus as well, did not arise from recent shared common ancestry. Possibilities remain, however, that some ancient gene orders somehow survived in divergent taxa of different phyla and that observed similarities represent shared primitive characters of metazoan mt genomes. Otherwise, if not by chance nor by shared ancestry, the similarities may only be explained by convergent evolution.

The facts that a part of the shared three-gene segment with Sarcophyton (*nad3-nad4L*) is also shared by human mtDNA and that the two-gene segment shared with Sarcophyton (*rns-nad1*) is also shared by nematode mtDNAs, despite the improbable combinations of these taxa as an evolutionary unit, provide support for the interpretation that at least some of those shared gene arrangements arose by a process of convergent evolution. The Poisson probabilities for these segments held in common by three genomes are significant ($P = 0.017$ and 0.019, respectively). A similar argument can be made as to a part of the three-gene segment shared with Lumbricus mtDNA (*atp6-nad5*), which is shared by a total of at least four genomes, including the cnidarian Metridium and the nematode Meloidogyne, of which Poisson probability is also highly significant ($P = 0.001$; Table 3). There is, however, no convincing evidence to indicate functional advantages or constraints for certain local arrangements of protein and rRNA genes in mt genomes.

For the majority of comparisons considering all the genes including tRNA genes, the Poisson probabilities are insignificant. A notable exception is the comparison with Lumbricus mtDNA, which shares as many as five two-gene segments with Laqueus mtDNA, and the Poisson probability is highly significant even estimated under the coding strand constraint ($P = 2.0 \times 10^{-4}$ in case II; $P = 0.004$ in case I: Table 3). Occurrence of three identical two-gene segments in Laqueus and the mollusc Katharina is also significant ($P = 0.015$; Table 3).

One each of the five segments shared by Laqueus and Lumbricus is also shared by Drosophila and Katharina, and these conditions in which each of the two-gene segments is shared by the three genomes are statistically significant ($P = 0.014$ for each comparison), but, unlike in the case with Sarcophyton, the association of Laqueus, Lumbricus, Drosophila, and Katharina does not necessarily conflict with at least some of the phylogenetic schemes proposed to date (*cf.* Willmer 1990). In fact, it accords well with the results of recent molecular phylogenetic studies that demonstrated protostome affinity of brachiopods (Field *et al.* 1988; Halanych *et al.* 1995; Cohen and Gawthrop 1996, 1997; Cohen *et al.* 1998; Stechmann and Schlegel 1999).

Although it is not just shared characters but shared derived characters that count in phylogenetic inferences, the polarities for the shared gene assortments observed in this study are difficult to infer, because diploblastic animals, the only apparent candidates as outgroups, generally lack most of the tRNA genes on their mt genomes, hindering gene order comparisons including tRNA genes. However, considering the variable nature of tRNA gene positions even within various phyla, and the frequent gene rearrangements that the Laqueus mt genome is inferred to have experienced, it can be assumed that most, if not all, of the local gene arrangements of Laqueus shared with Lumbricus or with Katharina represent derived states among the gene arrangements of the coelomate animals compared. It thus appears safe to interpret that the brachiopod mtDNA is closer phylogenetically to the annelid and mollusc mtDNAs than to any known mtDNAs of other animal phyla.

We are grateful to Bernard Cohen, Jeffrey Boore, and Kevin Helfenbein for helpful comments on the manuscript. This work was funded by grants from the Ministry of Education, Science, Sports, and Culture of Japan (K.E. and R.U.) and from Toray Science Foundation (R.U.).

LITERATURE CITED

Anderson, S., A. T. Bankier, B. G. Barrell, M. H. L. De Bruijn, A. R. Coulson *et al.*, 1981 Sequence and organization of the human mitochondrial genome. Nature **290:** 457–465.

Asakawa, S., Y. Kumazawa, T. Araki, H. Himeno, K. Miura *et al.*, 1991 Strand-specific nucleotide composition bias in echinoderm and vertebrate mitochondrial genomes. J. Mol. Evol. **32:** 511–520.

Asakawa, S., H. Himeno, K. Miura and K. Watanabe, 1995 Nucleotide sequence and gene organization of the starfish *Asterina pectinifera* mitochondrial genome. Genetics **140:** 1047–1060.

Barnes, W. M., 1994 PCR amplification of up to 35-kb DNA with high fidelity and high yield from λ bacteriophage templates. Proc. Natl. Acad. Sci. USA **91:** 2216–2220.

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.

Beaton, M. J., A. J. Roger and T. Cavalier-Smith, 1998 Sequence analysis of the mitochondrial genome of *Sarcophyton glaucum*:

conserved gene order among octocorals. J. Mol. Evol. **47:** 697–708.

Beard, C. B., D. M. Hamm and F. H. Collins, 1993 The mitochondrial genome of the mosquito Anopheles gambiae: DNA sequence, genome organization, and comparisons with mitochondrial sequences of other insects. Insect Mol. Biol. **2:** 103–124.

Boore, J. L., 1999 Animal mitochondrial genomes. Nucleic Acids Res. **27:** 1767–1780.

Boore, J. L., and W. M. Brown, 1994a Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*. Genetics **138:** 423–443.

Boore, J. L., and W. M. Brown, 1994b Mitochondrial genomes and the phylogeny of mollusks. Nautilus **108** (Suppl. 2): 61–78.

Boore, J. L., and W. M. Brown, 1995 Complete sequence of the mitochondrial DNA of the annelid worm *Lumbricus terrestris*. Genetics **141:** 305–319.

Boore, J. L., T. M. Collins, D. Stanton, L. L. Dachler and W. M. Brown, 1995 Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements. Nature **376:** 163–165.

Boore, J. L., D. V. Lavrov and W. M. Brown, 1998 Gene translocation links insects and crustaceans. Nature **392:** 667–668.

Brown, W. M., 1985 The mitochondrial genome of animals, pp. 95–130 in *Molecular Evolutionary Genetics*, edited by R. J. MacIntyre. Plenum Press, New York.

Cantatore, P., M. Roberti, G. Rainaldi, M. N. Gadaleta and C. Saccone, 1989 The complete nucleotide sequence, gene order and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J. Biol. Chem. **264:** 10965–10975.

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

Clary, D. O., and D. R. Wolstenholme, 1985b The ribosomal RNA genes of *Drosophila* mitochondrial DNA. Nucleic Acids Res. **31:** 4029–4044.

Cohen, B. L., and A. B. Gawthrop, 1996 Brachiopod molecular phylogeny, pp. 73–80 in *Brachiopods*, edited by P. Copper and J. Jin. A. A. Balkema, Rotterdam, The Netherlands.

Cohen, B. L., and A. B. Gawthrop, 1997 The brachiopod genome, pp. 189–211 in *Treatise on Invertebrate Paleontology, part H Brachiopoda (revised)*, edited by R. L. Kaesler. Geological Society of America and University of Kansas, Boulder, CO/Lawrence, KS.

Cohen, B. L., S. Stark, A. B. Gawthrop, M. E. Burke and C. W. Thayer, 1998 Comparison of articulate brachiopod nuclear and mitochondrial gene trees leads to a clade-based redefinition of protostomes (Protostomozoa) and deuterostomes (Deuterostomozoa). Proc. R. Soc. Lond. Ser. B **265:** 1–8.

Commission on Plant Gene Nomenclature, 1994 Mnemonic/Numeric.1: a database of plant-wide designations. Plant Mol. Biol. Rep. Suppl. **12:** S81–S88.

Crozier, R. H., and Y. C. Crozier, 1993 The mitochondrial genome of the honeybee *Apis mellifera*: complete sequence and genome organization. Genetics **133:** 97–117.

Desjardins, P., and R. Morais, 1990 Sequence and organization of the chicken mitochondrial genome: a novel gene order in higher vertebrates. J. Mol. Biol. **212:** 599–634.

Field, K. G., G. J. Olsen, D. J. Lane, S. J. Giovannoni, M. T. Ghiselin *et al.*, 1988 Molecular phylogeny of the animal kingdom. Science **239:** 748–753.

Fisher, C., and D. O. F. Skibinski, 1990 Sex-biased mitochondrial DNA heteroplasmy in the marine mussel *Mytilus*. Proc. R. Soc. Lond. Ser. B **242:** 149–156.

Ghiselin, M. T., 1988 The origin of molluscs in the light of molecular evidence, pp. 66–95 in *Oxford Surveys in Evolutionary Biology*, Vol. 5, edited by P. H. Harvey and L. Partridge. Oxford University Press, Oxford.

Halanych, K. M., J. D. Bacheller, A. M. Aguinaldo, S. M. Liva, D. M. Hillis *et al.*, 1995 Evidence from 18S ribosomal DNA that the lophophorates are protostome animals. Science **267:** 1641–1643.

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.

Henikoff, S., 1984 Unidirectional digestion with exonuclease III creates targeted breakpoints for DNA sequencing. Gene **29:** 351–359.

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.

Jacobs, H. T., D. J. Elliott, V. B. Math and A. Farquharson, 1988a Nucleotide sequence and gene organization of sea urchin mitochondrial DNA. J. Mol. Biol. **202:** 185–217.

Jacobs, H. T., P. Balfe, B. L. Cohen, A. Farquharson and L. Comito, 1988b Phylogenetic implications of genome rearrangement and sequence evolution in echinoderm mitochondrial DNA, pp. 121–137 in *Echinoderm Phylogeny and Evolutionary Biology*, edited by C. R. C. Paul and A. B. Smith. Clarendon Press, Oxford.

Jacobs, H. T., S. Asakawa, T. Araki, K. Miura, M. J. Smith *et al.*, 1989 Conserved tRNA gene cluster in starfish mitochondrial DNA. Curr. Genet. **15:** 193–206.

Janke, A., G. Feldmaier-Fuchs, W. K. Thomas, A. Von Haeseler and S. Pääbo, 1994 The marsupial mitochondrial genome and the evolution of placental mammals. Genetics **137:** 243–256.

Kumazawa, Y., and M. Nishida, 1995 Variations in mitochondrial tRNA gene organization of reptiles as phylogenetic markers. Mol. Biol. Evol. **12:** 759–772.

Lee, W.-J., and T. D. Kocher, 1995 Complete sequence of a sea lamprey (*Petromyzon marinus*) mitochondrial genome: early establishment of the vertebrate genome organization. Genetics **139:** 873–887.

Mitchell, S., A. Cockburn and J. Seawright, 1993 The mitochondrial genome of *Anopheles quadrimaculatus* species A: complete nucleotide sequence and gene organization. Genome **36:** 1058–1073.

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.

Okimoto, R., J. L. Macfarlane, D. O. Clary and D. R. Wolstenholme, 1991 Repeated sequence sets in mitochondrial DNA molecules of root knot nematodes (Meloidogyne): nucleotide sequences, genome location and potential for host-race identification. Nucleic Acids Res. **19:** 1619–1626.

Okimoto, R., J. L. Macfarlane and D. R. Wolstenholme, 1992 The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*. Genetics **130:** 471–498.

Pääbo, S., W. K. Thomas, K. M. Whitfield, Y. Kumazawa and A. C. Wilson, 1991 Rearrangements of mitochondrial transfer RNA genes in marsupials. J. Mol. Evol. **33:** 426–430.

Pont-Kingdon, G. A., N. A. Okada, J. L. Macfarlane, C. T. Beagley, C. D. Watkins-Sims *et al.*, 1998 Mitochondrial DNA of the coral *Sarcophyton glaucum* contains a gene for a homologue of bacterial MutS: a possible case of gene transfer from the nucleus to the mitochondrion. J. Mol. Evol. **46:** 419–431.

Sambrook, J., E. F. Fritsch and T. Maniatis, 1989 *Molecular Cloning: A Laboratory Manual*, Ed. 2. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY.

Smith, M. J., A. Arndt, S. Gorski and E. Fajber, 1993 The phylogeny of echinoderm classes based on mitochondrial gene arrangements. J. Mol. Evol. **36:** 545–554.

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. Ser. B **266:** 2043–2052.

Terrett, J. A., S. Miles and R. H. Thomas, 1996 Complete DNA sequence of the mitochondrial genome of Cepaea nemoralis (Gastropoda: Pulmonata). J. Mol. Evol. **42:** 160–168.

Valverde, J., B. Batuecas, C. Moratilla, R. Marco and R. Garesse, 1994 The complete mitochondrial DNA sequence of the crustacean *Artemia franciscana*. J. Mol. Evol. **39:** 400–408.

Willmer, P., 1990 *Invertebrate Relationships: Patterns in Animal Evolution*. Cambridge University Press, Cambridge, United Kingdom.

Wolstenholme, D. R., 1992 Animal mitochondrial DNA: structure and evolution. Int. Rev. Cyt. **141:** 173–215.

Wolstenholme, D. R., J. L. Macfarlane, R. Okimoto, D. O. Clary and J. A. Wahleithner, 1987 Bizarre tRNAs inferred from DNA sequences of mitochondrial genomes of nematode worms. Proc. Natl. Acad. Sci. USA **84:** 1324–1328.

Yamazaki, N., R. Ueshima, J. A. Terrett, S. Yokobori, M. Kaifu *et al.*, 1997 Evolution of pulmonate gastropod mitochondrial genomes: comparisons of gene organization of Euhadra, Cepaea and Albinaria and implications of unusual tRNA secondary structures. Genetics **145:** 749–758.

Communicating editor: S. Yokoyama

TABLE 6

Probability distribution of S_2

	S_2						
	0	1	2	3	4	5	6
				Case I			
$n = 10$							
Approximation	0.32919	0.36577	0.20321	0.07526	0.02091	0.00465	0.00086
Simulation	0.33160	0.36391	0.20156	0.07558	0.02086	0.00557	0.00059
$n = 15$							
Approximation	0.34252	0.36698	0.19660	0.07021	0.01881	0.00403	0.00072
Simulation	0.34348	0.36615	0.19591	0.07038	0.01902	0.00415	0.00077
$n = 20$							
Approximation	0.34902	0.36739	0.19336	0.06785	0.01785	0.00376	0.00066
Simulation	0.34958	0.36706	0.19302	0.06778	0.01796	0.00381	0.00068
$n = 37$							
Approximation	0.35780	0.36774	0.18898	0.06474	0.01664	0.00342	0.00059
Simulation	0.35810	0.36754	0.18876	0.06480	0.01670	0.00343	0.00057
				Case II			
$n = 10$							
Approximation	0.57375	0.31875	0.08854	0.01640	0.00228	0.00025	0.00002
Simulation	0.57487	0.31721	0.08853	0.01668	0.00239	0.00028	0.00003
$n = 15$							
Approximation	0.58525	0.31353	0.08398	0.01500	0.00201	0.00022	0.00002
Simulation	0.58584	0.31294	0.08385	0.01508	0.00205	0.00022	0.00002
$n = 20$							
Approximation	0.59078	0.31094	0.08183	0.01436	0.00189	0.00020	0.00002
Simulation	0.59108	0.31052	0.08187	0.01441	0.00189	0.00020	0.00002
$n = 37$							
Approximation	0.59816	0.30739	0.07898	0.01353	0.00174	0.00018	0.00002
Simulation	0.59804	0.30733	0.07910	0.01359	0.00174	0.00018	0.00002

Approximation was obtained from the Poisson distribution with mean $E(S_2)$.

APPENDIX: STATISTICAL TEST FOR RANDOM GENE ARRANGEMENTS

Consider two circular DNA sequences on which n genes are randomly distributed. We consider two types of random gene arrangements. One is the case where the direction of transcription is the same among all genes, and we call this case I. The other is the case where the direction of transcription for each gene is also random, and we call this case II.

When we compare two sequences, we can observe the number of shared arrangements with k genes (S_k). For example, when we have two sequences shown in Figure 4, we have $S_2 = 3$ and $S_3 = 1$. Note that when there is one shared arrangement with three genes, we treat this as two shared arrangements with two genes. We can also find the longest shared arrangement, and we denote the number of genes in this arrangement by Lmax. In this example the longest shared arrangement has three genes so that we have Lmax = 3.

The expected number of shared arrangements with k genes, $E(S_k)$, can be obtained as follows. In case I, when one sequence has gene B next to gene A, the probability that the other sequence also has gene B next to gene A is $1/(n-1)$. Since there are n genes, the expected number of shared arrangements with two genes is $E(S_2) = n/(n-1)$. In the same way the probability of having a particular shared arrangement with three genes is $1/\{(n-1)(n-2)\}$, so that the expected number of shared arrangements with three genes is $E(S_3) = n/\{(n-1)(n-2)\}$. In general the expected number of shared arrangements with k genes is given by

$$E(S_k) = n(n-k)!/(n-1)!. \qquad \text{(A1)}$$

In case II the direction of transcription is random, so that we have

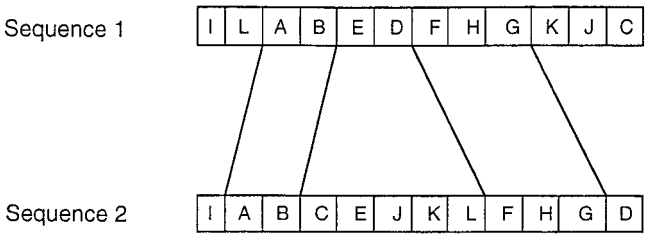

Figure 4.—Schematic example to explain the definition of shared gene arrangement. In this example case I is assumed. There are two shared gene arrangements: A-B and F-H-G.

$$E(S_k) = n(n-k)!/\{2^{k-1}(n-1)!\}. \quad \text{(A2)}$$

It is difficult to obtain the distribution of S_k when n is large. It might be expected, however, that the distribution of S_k follows the Poisson distribution with mean $E(S_k)$. To know whether the Poisson distribution is a good approximation or not, we conducted a computer simulation. In this simulation, using pseudorandom numbers, we generated a pair of random sequences 10,000,000 times and observed S_2. The results are shown in Table 6. From this table we can see that the distribution of S_2 approximately follows the Poisson distribution with mean $E(S_2)$ when $n \geq 10$. This means that S_2 can be used to test for random gene arrangement. For example, when $n = 37$ and $S_2 = 4$, the probability of having $S_2 \geq 4$ is 0.021 so that we can reject the hypothesis of random gene arrangement at the 5% level in case I.

Next, we consider the number of shared arrangements that have exactly k genes (L_k). In the example shown in Figure 4, we have $L_2 = 1$ and $L_3 = 1$. The expected number of arrangements that have exactly k genes, $E(L_k)$, can be obtained as follows. We note $S_n = nL_n$, which can occur when the gene arrangement is exactly the same between two sequences. Since the gene arrangement with exactly $k + i$ genes contributes $i + 1$ times to S_k and since the completely identical gene arrangements between two sequences contribute n times to S_k, we have $S_k = L_k + 2L_{k+1} + 3L_{k+2} + \ldots + (n-k)L_{n-1} + nL_n$. Therefore, when $k \leq n-2$, we have $L_k = S_k - 2S_{k+1} + S_{k+2}$. From these results, $E(L_k)$ can be given by

$$E(L_k) = E(S_k) - 2E(S_{k+1}) + E(S_{k+2}) \quad \text{for } k \leq n-2,$$
$$E(L_{n-1}) = E(S_{n-1}) - E(S_n) \quad \text{and} \quad E(L_n) = E(S)/n. \quad \text{(A3)}$$

Now, let us consider the longest shared arrangement. First, we note that when $k \geq n/2$ in case I and when $k > n/2$ in case II, the probability that the longest shared arrangement has k genes is the same as $E(L_k)$. This is because the longest shared arrangement exists only once. When $k < n/2$ in case I and when $k \leq n/2$ in case II, more than one longest shared arrangement can occur. Such events, however, might be very rare when $k \geq 3$, since $E(L_3)$ is quite small. Thus, when $k \geq 3$, the probability that the longest shared arrangement has k genes is approximately given by

$$\text{Prob}\{\text{Lmax} = k\} = E(L_k). \quad \text{(A4)}$$

To know the accuracy of Equation A4, we conducted a computer simulation. The method is the same as the previous one, and the results are shown in Table 7. We can see from this table that Equation A4 is quite accurate and can be used to test for random gene arrangement. For example, when $n = 37$ and Lmax = 3, we have $\text{Prob}\{\text{Lmax} \geq 3\} = 0.0072$ so that we reject the hypothesis of random gene arrangement at the 1% level in case II.

Table 8 shows the critical values for S_k and Lmax, which can be used for the hypothesis testing.

TABLE 7

Probability distribution of Lmax

	Lmax			
	3	4	5	6
	Case I			
$n = 10$				
$E(L_k)$	0.10251	0.01389	0.00215	0.00039
Simulation	0.09736	0.01372	0.00216	0.00039
$n = 15$				
$E(L_k)$	0.06931	0.00568	0.00051	0.00005
Simulation	0.06684	0.00565	0.00051	0.00005
$n = 20$				
$E(L_k)$	0.05181	0.00302	0.00019	0.00001
Simulation	0.05046	0.00304	0.00019	0.00001
$n = 37$				
$E(L_k)$	0.02766	0.00081	0.00002	0.00000
Simulation	0.02733	0.00081	0.00003	0.00000
	Case II			
$n = 10$				
$E(L_k)$	0.02997	0.00209	0.00017	0.00002
Simulation	0.02951	0.00210	0.00017	0.00001
$n = 15$				
$E(L_k)$	0.01893	0.00078	0.00004	0.00000
Simulation	0.01878	0.00078	0.00004	0.00000
$n = 20$				
$E(L_k)$	0.01377	0.00040	0.00001	0.00000
Simulation	0.01373	0.00040	0.00001	0.00000
$n = 37$				
$E(L_k)$	0.00713	0.00010	0.00000	0.00000
Simulation	0.00713	0.00011	0.00000	0.00000

TABLE 8

Critical values for S_2 and Lmax

	S_2						Lmax					
	$\alpha = 0.05$		$\alpha = 0.01$		$\alpha = 0.001$		$\alpha = 0.05$		$\alpha = 0.01$		$\alpha = 0.001$	
n	I	II	I	II	I	II	I	II	I	II	I	II
10	4	3	5	4	7	5	4	3	5	4	6	5
11–12	4	3	5	4	6	5	4	3	5	4	6	5
13	4	3	5	4	6	5	4	3	4	4	5	5
14–21	4	3	5	4	6	5	4	3	4	4	5	4
22–27	4	3	5	4	6	5	3	3	4	4	5	4
28–34	4	3	5	4	6	5	3	3	4	3	5	4
35–40	4	3	5	4	6	5	3	3	4	3	4	4

α and n are the confidence level and the number of genes, respectively. I and II correspond to cases I and II.