LBNL-59181

The Complete Sequence of the Mitochondrial Genome of the Chambered Nautilus (Mollusca: Cephalopoda)

Jeffrey L Boore

The Complete Sequence of the Mitochondrial Genome of the Chambered Nautilus (Mollusca: Cephalopoda)

Jeffrey L Boore

Evolutionary Genomics Program, DOE Joint Genome Institute and Lawrence Berkeley National Laboratory and

Department of Integrative Biology, University of California, Berkeley and

Genome Project Solution, Hercules, CA 94547

For correspondence: JLBoore@LBLgov, (925) 296-5691, FAX: (925) 296-5620

Corresponding author:

Jeffrey L Boore

DOE Joint Genome Institute and Lawrence Berkeley National Laboratory

2800 Mitchell Drive

Walnut Creek, CA 95498 USA

JLBoore@LBLgov

(925) 296-5691

FAX: (925) 296-5620

Abstract

Background

Mitochondria contain small genomes that are physically separate from those of nuclei. Their comparison serves as a model system for understanding the processes of genome evolution. Although complete mitochondrial genome sequences have been reported for more than 600 animals, the taxonomic sampling is highly biased toward vertebrates and arthropods, leaving much of the diversity yet uncharacterized.

Results

The mitochondrial genome of a cephalopod mollusk, the Chambered Nautilus, is 16,258 nts in length and 59.5% A+T, both values that are typical of animal mitochondrial genomes. It contains the 37 genes that are typical for animal mtDNAs, with 15 on one DNA strand and 22 on the other. The arrangement of these genes can be derived from that of the distantly related *Katharina tunicata* (Mollusca: Polyplacophora) by a switch in position of two large blocks of genes and transpositions of four tRNA genes. There is strong skew in the distribution of nucleotides between the two strands. There are an unusual number of non-coding regions and their function, if any, is not known; however, several of these demark abrupt shifts in nucleotide skew, suggesting that they may play roles in transcription and/or replication. One of the non-coding regions contains multiple repeats of a tRNA-like sequence. Some of the tRNA genes appear to overlap on the same strand, but this could be resolved if the polycistron were cleaved at the beginning of the downstream gene, followed by polyadenylation of the product of the upstream gene to form a fully paired structure.

Conclusions

Nautilus sp. mtDNA contains an expected gene content that has experienced few rearrangements since the evolutionary split between cephalopods and polyplacophorans. It contains an unusual number of non-coding regions, especially considering that these otherwise often are generated by the same processes that produce gene rearrangements. This appears to be yet another case where polyadenylation of mitochondrial tRNAs restores what would otherwise be an incomplete structure.

Background

Animal mitochondrial DNA (mtDNA) is nearly always a closed circular molecule and, with a few exceptions [e.g. 1-4], contains the same 37 genes, specifying 13 proteins, two ribosomal RNAs, and 22 tRNAs [5]. Sequences of these diminutive genomes have been broadly used to address phylogenetic questions ranging from the population [6,7] to the interphylum [8-11] levels and to model many processes of genome evolution [12,13]. Although there are exceptions, most mtDNAs contain no introns and are between 14 and 17 kb. Typically there are few intergenic nucleotides except for a single large non-coding region generally thought to contain elements that regulate replication and transcription [14]. Occasionally non-coding regions have been found that contain repeated elements [15] or contain pseudogenes [12,16] or that may be remnants of duplicated regions, perhaps those that mediate gene rearrangements [12,16,17]. Gene rearrangements tend to be uncommon and to occur in a saltatory manner [see 10]. The "universal" genetic code has been modified in many animal lineages, to include the use of alternative start codons and abbreviated stop codons [18,19]. In some mtDNAs there is pronounced skew in nucleotide composition, often with one strand being rich in G and T and the other in A and C [20]. Post-transcriptional modification of nucleotides has been observed for tRNAs [21,22].

Little study has been done to date on mollusk mtDNAs compared to those of vertebrates or arthropods [23], but it is already apparent that mollusks exhibit much variation in the features of their mitochondrial genomes, including losses and gains of genes [2], atypically large amounts of duplicated or non-coding nucleotides [15,24], highly rearranged genomes [2,25], and an unusual pattern of passage termed doubly uniparental inheritance [26,27]. We further this here by

reporting and comparing the features of the mitochondrial genome of the Chambered Nautilus (Mollusca: Cephalopoda).

Nautiloids were once abundant and diverse in the Paleozoic seas, but only a handful of species remain, all in the genus Nautilus. They are part of the molluscan class Cephalopoda, which otherwise contains octopi, squid, and cuttlefish. They are the earliest diverging lineage of this group and are often considered to be "living fossils" since living forms seem to have changed little from their ancient ancestors. Their common name, the chambered nautilus, comes from the chambers that develop in their spiral-shaped shells as they grow, beginning with about four in the young animal and progressing to 30 in the adult, and which are filled with gas to control buoyancy. The animal lives only in the largest, outermost chamber of the shell. They are carnivorous, using their many grooved tentacles to grasp prey and pass it to their mouth, where a beak-like jaw tears it and passes it to the shredding radula, and they move about by squirting jets of water. They live throughout the Southwest Pacific Ocean, at depths as great as 610 meters, and traverse a great range, as shallow as 90 meters, apparently in search of prey.

Results and Discussion

Gene content and organization

The *Nautilus* sp. (Mollusca: Cephalopoda) mitochondrial genome is 16,258 bp in length (GenBank accession number **NNNNNNN**) and contains the set of 37 genes most commonly found for animal mtDNAs [5]. Fifteen genes are located on one strand and 22 on the other (Fig. 1). There are several substantial non-coding regions (see below), the largest of which is 972 nts long and between *trnQ* and *trnT*. The mitochondrial gene arrangement of *Nautilus* sp. differs from that of the distantly related *Katharina tunicata* [28] (the only sampled representative of the

Polyplacophora, an early diverging class of the Mollusca) by only transpositions of four tRNA genes and the switch in position of two large blocks of genes (Fig. 2). By comparison with an outgroup, the phoronid *Phoronis architecta* [11] (and confirmed by others not shown), we can see that one of these tRNAs, *trnD*, remains in the ancestral condition in these two cephalopods shown in Fig. 2, with a transposition having occurred in the polyplacophoran, whereas all other changes are derived for *Nautilus* sp. The mtDNA of *Octopus vulgaris* [29] is nearly identical in arrangement to that of *K. tunicata*, differing otherwise only in the inversion of *trnP*, which it shares with *Nautilus* and many other mollusks.

The mtDNA sequence is available for two other cephalopod mollusks, the squids *Loligo bleekeri* [30] and *Todarodes pacificus* [29], each of which is more highly rearranged (Supplemental Table 1). There are only three gene blocks that *L. bleekeri* shares with *Nautilus* sp. mtDNA (minus symbol indicates opposite transcriptional orientation, and oriented here to correspond with Fig. 2): (1) *-trnS2*, *-cob*, *-nad6*, *-trnP*, *-nad1*; (2) *-rrnL*, *-trnV*, *-rrnS*; and (3) *trnS1*, *nad2*, *cox1*. *L bleekeri* mtDNA also shares the two blocks *-nad4*, *-nad4L*, *trnT* and *trnD*, *atp8*, *atp6* with *K. tunicata* and *O. vulgaris*. The mtDNA of *T. pacificus* also has several blocks of genes conserved with one or more mollusks: (1) *trnN*, *trnI*, *nad3*; (2) *cox1*, *cox2*, *trnD*, *atp8*, *atp6*, *-trnF*; (3) *-trnV*, *-rrnS*; (4) *cox3*, *trnK*; (5) *trnS1*, *nad2*, *cox1*, *cox2*, *trnD*, *atp8*, *atp6*; (6) *-nad5*, *-trnH*, *-nad4*, *-nad4L*, *trnT*, *-trnS2*, *-cob*, *-nad6*, *-trnP*, *-nad1*, *-trnL2*, *-trnL1*, *-rrnL*; (7) *-trnG*, *-trnE*. However, there are six genes that appear in duplicate, accounting for the listing of some genes more than once in differing arrangements here.

Gene initiation and termination

Mitochondrial genomes often use a variety of non-standard initiation codons [19], but *Nautilus* mtDNA has only one type of deviation; three genes (*nad3*, *nad4*, and *nad5*) start with GTG and all others use the standard ATG (Fig. 3). Seven genes have unambiguous termination codons, either TAG (*atp6*, *cox1*, *nad5*) or TAA (*atp8*, *cox3*, *nad1*, *nad2*). In four cases (*cox2*, *cob*, *nad3*, *nad4*) genes are probably abbreviated to a single T or to TA such that the excision of the adjacent, downstream tRNA from the polycistronic message leaves an mRNA that is polyadenylated to complete a TAA stop codon. However, in each of these cases, a complete stop codon is available if there is, alternatively, overlap of only one or two nucleotides with the downstream tRNA. Perhaps these act as a "backup" for cases where translation precedes message cleavage. The other two cases are more ambiguous. *nad4L* could have an abbreviated stop codon, but is inferred to overlap *nad4* by seven nucleotides to the first legitimate stop codon, since overlap of this pair has been commonly observed for other mtDNAs, where they are thought to be translated as a bicistron. *nad6* is inferred to overlap *cob* by eight nucleotides, perhaps suggesting that these are processed also as a bicistron, but could instead end on an abbreviated stop codon if there were some signal for message cleavage (i.e., other than a tRNA) that we do not recognize. Inferred in this way, all protein-encoding genes have lengths nearly identical to those of *K. tunicata* mtDNA (Supplemental Table 2).

Base composition and codon usage

The *Nautilus* sp. mtDNA is 59.5% A+T. The strand that includes *cox1*, which we will arbitrarily designate as the plus strand for the purpose of discussion, is 33.7% A, 25.8% T, 11.9% G, and 28.5% C. This strand is strongly skewed (as calculated in [20]) away from both T

(T-skew = – 013) and G (G-skew = – 041) in favor of A and C. As can be seen in Table 1, this is strongly reflected in the use of synonymous codons. For example, while TTT and TTC are used with approximately equal frequency to specify phenylalanine in plus strand genes, the bias is 158 to 3 for their usage in minus strand genes. The use of G vs. A in UUR (leucine) codons is in the ratio of 16 to 89 for plus strand genes but, even though the mtDNA is A+T-rich, it is 195 to 60 for minus strand genes. Presumably the biased use of synonymous codons is driven by strand-specific mutational propensity.

There are eight non-coding regions of this mtDNA that are larger than 20 nucleotides, several of which are associated with shifts in these skew values (Fig. 4, Table 2, Supplemental Table 3). **ADD MORE HERE.**

Transfer RNAs

Sequences were identified whose potential secondary structures indicate that they encode the 22 tRNAs typically found for animal mtDNA (Fig. 5). In general, these appear well paired with only a few mismatches.

There are three cases where tRNA genes appear to overlap, and these potential structures suggest how this is resolved. *trnL1* appears to overlap *trnL2* by only the former's discriminator nucleotide (A). *trnQ* appears to overlap *trnW* by two nucleotides. *trnK* appears to overlap *trnA* by four nucleotides, GGCT. These are well-paired in the potential structure of tRNA(A), but these four correspond to two G-T pairs, one mismatch, and the discriminator nucleotide of tRNA(K). It appears for each case that cleavage to form a complete downstream tRNA followed by (poly)adenylation of the upstream tRNA (as has been demonstrated for some mitochondrial tRNAs [Yokobori and Paabo 1995] would yield fully formed, well-paired structures for all. This

is illustrated in Fig. 5 by lower case, parenthetical letter "a" appended to the genome-encoded nucleotide to indicate likely nucleotides in the actual transcript.

Usually T is in the first anticodon position for tRNAs that recognize either four-fold degenerate codon families or to specifically recognize NNR codons; G is usually in this position only to specifically recognize NNY codons. (Due to the convention of always drawing RNAs from 5' to 3' in orientation, the first nucleotide listed for an anticodon pairs with the last nucleotide of a codon.) All but two of the *Nautilus* mitochondrial tRNAs follow this pattern. One exception is tRNA(M), which has the anticodon CAT (to recognize both ATA and ATA), as is almost universally the case for all animal mitochondrial systems. However, it is unusual that the tRNA(S-AGN) has a GCT anticodon, since this requires the G to pair with all four nucleotides in the wobble position of AGN codons. It is clear the AGA and AGG codons are being used and are not stop codons (as is the case in vertebrate mtDNAs), since they appear in the reading frames of protein encoding genes 117 times. It is possible that some anticodon nucleotides are post-transcriptionally modified to alter their base pairing properties.

Non-coding regions

The mtDNA of *Nautilus* sp. has 1,416 nucleotides that are not assigned to genes. This is not an unusually large number, but it is atypical that they are distributed among so many regions of the genome (Table 2 and Supplemental Table 3). It is particularly unusual to find this in a mitochondrial genome that has not undergone significant rearrangements, since intergenic non-coding regions appear in some cases to be vestiges of pseudogenes generated by the gene duplication-random loss process of rearrangement [Mueller and Boore 2005; Boore 2000; Macey et al 2004; Arendt and Smith 1996].

In the largest non-coding region, between *trnQ* and *trnT*, and beginning adjacent to a $(CA)_{13}$ run (see below), there are six repeats of a 62 nucleotide element followed by a partial repeat of 39 nucleotides. Within this are five overlapping regions that have potential for forming tRNA-like structures (Fig. 5). The anticodon portion of these structures is AGT, which would pair with codon ACT (or perhaps ACN) to specify threonine. However, having A in this anticodon position would be very unusual and there is little sequence similarity to *trnT* (or any other tRNA).

Tandem repeats of CA are common, with $(CA)_3$ in each of the intergenic regions of *trnA-trnR* and *trnG-atp6* and an especially noteworthy $(CA)_{13}$ in the region between *trnQ* and *trnT*. Homopolymer runs of T_{10}, nine C_9, and A_{20} are in the regions *trnQ-trnT*, *trnG-atp6*, and *trnE-cox3*, respectively. Non-coding, non-functional portions of mtDNA are generally eliminated rapidly (Ashley et al. 1989), presumably due to selection for small size at the point of entry into the primordial germ plasm during embryogenesis (Mignotte et al. 1987), but whether these or any particular motif plays any role in regulating replication of transcription awaits experimentation

Conclusions

ELABORATE ON THESE ASPECTS FROM THE ABSTRACT: *Nautilus* sp. mtDNA contains an expected gene content that has experienced few rearrangements since the evolutionary split between cephalopods and polyplacophorans. It contains an unusual number of non-coding regions, especially considering that these otherwise often are generated by the same processes that produce gene rearrangements. This appears to be yet

another case where polyadenylation of mitochondrial tRNAs restores what would otherwise be an incomplete structure.

Methods

Molecular techniques

Mitochondrial DNA was isolated from approximately 1 g of *Nautilus* sp. testis tissue (gift of Wesley Brown) that had been stored at -80º C by first grinding in liquid nitrogen using a mortar and pestle. This powder was dissolved in 14 ml of homogenization buffer (210 mM mannitol, 70 mM sucrose, 50 mM Tris HCl-pH 75, 3 mM $CaCl_2$) and processed using a Tissuemizer T-25 (Tekmar) with three strokes of five seconds each. Membranes were lysed by adding 1/10 volume of 20% SDS and incubating for 20 min at RT. A 1/6 volume of saturated CsCl in water was added and this mixture incubated on ice for 15 min. Debris was pelleted at 17,000 X G for 10 minutes at 4° C. Propidium iodide was added to the collected supernatant to a final concentration of 500 μg/ml and the CsCl concentration was adjusted to a density of 1.57 g/ml. Nuclear and mitochondrial DNA were separated by density gradient centrifugation in a VTi65 rotor at 55,000 X G for 15 hours at 21° C. Although no mitochondrial band was visible in the gradient, the region from about 2-10 mm below the nuclear band was collected using a needle. This was then extracted multiple times with water-saturated butanol to remove the propidium iodide and dialyzed against TE for 24 hours with three buffer changes to remove the CsCl, leaving the sample in a 100 μl volume.

This product was used in PCR as in [Boore and Brown 2000] to amplify first several short fragments of *cox1*, *rrnL*, and *cob* using primers found in [Folmer et al 1994, Palumbi 1996, Boore and Brown 2000]. The fragment of *cox1* was cloned into pBluescript (Stratagene) that

had been digested with *Eco*RV, T-tailed using Taq polymerase, and gel purified using GeneClean (QBiogene). A successful recombinant clone was selected and DNA prepared using standard techniques. The other fragments were purified by three serial passages through an Ultrafree (NMWL 30,000) spin column (Millipore) and sequenced directly. The sequences of these fragments were determined using an ABI377 automated DNA sequencer with BigDye chemistry (Applied Biosystems) according to supplier's instructions.

Primers were designed to known sequences for use in long PCR [Barnes 1994] with rTth-XL polymerase (Applied Biosystems) according to supplier's instructions, sometimes combined with primers to conserved mtDNA regions. Generously overlapping fragments were amplified from *cox1–nad1* (using conserved *nad1* primer CCTGATACTAATTCAGATTCTCCTTC), *nad1–cob*, *cob–rrnL*, and *rrnL—cox1* (using conserved primer 16SARL [Palumbi 1996]), jointly comprising the entire mtDNA. Because there was no information available for the gene arrangement, many combinations of primers were tried, but only these reactions gave bright, singular bands during electrophoretic analysis. Sequence was determined for each as above, then by primer walking through each fragment. To ensure accuracy, all sequence was determined from both strands. Sequencing reads were assembled manually and quality verified by eye using Sequence Navigator (Applied Biosystems).

Gene annotation and analysis

Genes encoding rRNAs and proteins were identified by matching nucleotide or inferred amino acid sequences to those of *Katharina tunicata* mtDNA [Boore and Brown 1994]. Since it is not possible to precisely determine the ends of rRNA genes by sequence data alone, they were assumed to extend to the boundaries of flanking genes. Each protein gene was inferred to begin

at an eligible initiation codon nearest to the beginning of its alignment with homologous genes that does not cause overlap with the preceding gene. In five cases, an abbreviated stop codon was inferred where cleavage of a downstream tRNA from the transcript would leave a partial codon of T or TA, such that subsequent mRNA polyadenylation could generate a TAA stop codon; however, in each of these cases, if the reading frame extended through the first legitimate stop codon there would be only a short overlap with the downstream gene. Genes for tRNAs were identified generically by their ability to fold into a cloverleaf structure and specifically by anticodon sequence. Subsequent sequence analyses were performed using MacVector (Accelrys).

Abbreviations

cox1, *cox2*, *cox3*, cytochrome oxidase subunit I, II, and III protein genes; *cob*, cytochrome b gene; *atp6*, *atp8*, ATP synthase subunit 6 and 8 genes; *nad1*, *nad2*, *nad3*, *nad4*, *nad4L*, *nad5*, *nad6*, NADH dehydrogenase subunit 1-6, 4L genes; *trnA*, *trnC*, *trnD*, *trnE*, *trnF*, *trnG*, *trnH*, *trnI*, *trnK*, *trnL1*, *trnL2*, *trnM*, *trnN*, *trnP*, *trnQ*, *trnR*, *trnS1*, *trnS2*, *trnT*, *trnV*, *trnW*, *trnY*, transfer RNA genes designated by the one-letter code for the specified amino acid, with numerals differentiating cases where there are two tRNAs for the same amino acid

Acknowledgments

I am grateful to Wesley Brown for tissue of *Nautilus* sp. This work was supported by DEB-9807100 from the National Science Foundation and by contract DE-AC02-05CH11231 between the US Department of Energy, Office of Biological and Environmental Research, and the University of California, Lawrence Berkeley National Laboratory.

References

Ashley MV, Laipis PJ, Hauswirth WW: **Rapid segregation of heteroplasmic bovine mitochondria.** *Nucleic Acids Res* 1989, 17(18): 7325-7331

16 Arndt A, Smith, MJ: **Mitochondrial gene rearrangement in the sea cucumber genus *Cucumaria*** *Mol Biol Evol* 1998, 15(8): 1009-1016

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

3 Beagley CT, Okimoto R, Wolstenholme DR: **The mitochondrial genome of the sea anemone *Metridium senile* (Cnidaria): Introns, a paucity of tRNA genes, and a near-standard genetic code** *Genetics* 1998, 148: 1091-1108

5 Boore JL: **Animal mitochondrial genomes** *Nucleic Acids Res* 1999, 27: 1767-1780

28 Boore JL, Brown WM: **Complete DNA sequence of the mitochondrial genome of the black chiton, *Katharina tunicata*** *Genetics* 1994, 138: 423-443

Boore JL, Brown WM: **Mitochondrial genomes of *Galathealinum*, *Helobdella*, and *Platynereis*: Sequence and gene arrangement comparisons indicate that Pogonophora is not a phylum and Annelida and Arthropoda are not sister taxa** *Mol Biol Evol* 2000, 17(1): 87-106

9 Boore JL, Collins TM, Stanton D, Daehler LL, Brown WM: **Deducing the pattern of arthropod phylogeny from mitochondrial DNA rearrangements** *Nature* 1995, 376: 163-165

25 Boore JL, Medina M, Rosenberg, LA: **Complete sequences of two highly rearranged molluscan mitochondrial genomes, those of the scaphopod *Graptacme eborea* and of the bivalve *Mytilus edulis*** *Mol Biol Evol* 2004, 21(8): 1492–1503

17 Boore JL: **The duplication/random loss model for gene rearrangement exemplified by mitochondrial genomes of deuterostome animals.** In: *Comparative Genomics*. Edited by Sankoff D, Nadeau J. Computational Biology Series vol 1, Kluwer Academic Publishers, Dordrecht, Netherlands 2000, 133-147.

10 Boore, JL, Brown, WM: **Big trees from little genomes: Mitochondrial gene order as a phylogenetic tool** *Curr Opin Genet Dev* 1998, 8(6): 668-674

Folmer O, Black M, Hoeh W, Lutz R, Vrijenhoek R: **DNA primers for amplification of mitochondrial cytochrome *c* oxidase subunit I from diverse metazoan invertebrates** *Mol Mar Biol Biotech* 1994, 3: 294-299

24 Fuller KM, Zouros E: **Dispersed length polymorphism of mitochondrial DNA in the scallop *Placopecten magellanicus* (Gmelin)** *Curr Genet* 1993, 23: 365-369

11 Helfenbein KG, Boore JL: **The mitochondrial genome of *Phoronis architecta*— Comparisons demonstrate that phoronids are lophotrochozoan protostomes** *Mol Biol Evol* 2004, **21(1):** 153-157

13 Helfenbein KG, Brown WM, Boore JL: **The complete mitochondrial genome of a lophophorate, the brachiopod *Terebratalia transversa*** *Mol Biol Evol* 2001, 18(9): 1734-1744

4 Helfenbein KG, Fourcade HM, Vanjani RG, Boore, JL: **The mitochondrial genome of *Paraspadella gotoi* is highly reduced and reveals that chaetognaths are a sister-group to protostomes** *Proc Natl Acad Sci USA* 2004, 101 (29): 10639-10643

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

7 Ingman M, Kaessmann H, Pääbo S, Gyllensten U: **Mitochondrial genome variation and the origin of modern humans** *Nature* 2001, 408: 708-713

21 Lavrov D, Brown WM, Boore JL: **A novel type of RNA editing occurs in the mitochondrial tRNAs of the centipede *Lithobius forficatus*** *Proc Natl Acad Sci USA* 2000, 97(25): 13738-13742

Mignotte F, Tourte M, Mounolou J-C: **Segregation of mitochondria in the cytoplasm of *Xenopus* vitellogenic oocytes** *Biol. of the Cell* 1987, 60: 97-102

12 Mueller RL, Boore JL: **Molecular mechanisms of extensive mitochondrial gene rearrangement in plethodontid salamanders** *Mol Biol Evol* 2005, 22: 2104-2112

6 Nyakaana S, Arctander P, Siegismund H: **Population structure of the African savannah elephant inferred from mitochondrial control region sequences and nuclear microsatellite loci** *Heredity* 2002, 89(2): 90-98

18 Ojala D, Montoya J, Attardi G: **tRNA punctuation model of RNA processing in human mitochondria** *Nature* 1981, 290: 470-474

1 Okimoto R, Macfarlane JL, Clary DO, Wolstenholme DR: **The mitochondrial genomes of two nematodes, *Caenorhabditis elegans* and *Ascaris suum*** *Genetics* 1992, 130: 471-498

Palumbi SR: **Nucleic acids II: The polymerase chain reaction.** In: *Molecular Systematics.* Edited by Hillis DM, Moritz C, Mable BK. Sinauer Associates, Sunderland, Massachusetts, USA 1996, 205-247.

27 Passamonti M, Boore, JL, Scali, V: **Molecular evolution and recombination in gender-associated mitochondrial DNAs of the Manila clam *Tapes philippinarum*** *Genetics* 2003, **164:** 603–611

20 Perna NT, Kocher TD: **Patterns of nucleotide composition at fourfold degenerate sites of animal mitochondrial genomes** *J Mol Evol* 1995, 41: 353-358

15 Rigaa A, Monnerot M, Sellos D: **Molecular cloning and complete nucleotide sequence of the repeated unit and flanking gene of the scallop *Pecten maximus* mitochondrial DNA: Putative replication origin features** *J Mol Evol* 1995, 41: 189-195

14 Shadel GS, Clayton DA: **Mitochondrial DNA maintenance in vertebrates** *Annu Rev Biochem* 1997, 66: 409-435

8 Smith MJ, Arndt A, Gorski S, Fajber E: **The phylogeny of echinoderm classes based on mitochondrial gene arrangements** *J Mol Evol* 1993, 36: 545-554

26 Stewart DT, Saavedra C, Stanwood RR, Ball AO, Zouros E: **Male and female mitochondrial DNA lineages in the Blue Mussel (*Mytilus edulis*) species group** *Mol Biol Evol* 1995, 12: 735-747

30 Tomita K, Yokobori S, Oshima T, Ueda T, Watanabe K: **The cephalopod *Loligo bleekeri* mitochondrial genome: Multiplied noncoding regions and transposition of tRNA genes** *J Mol Evol* 2002, 54 (4): 486-500

23 Vallès Y, Boore, JL: **Lophotrochozoan mitochondrial genomes** *Integrative Comp Biol* 2006, in press

19 Wolstenholme DR: **Animal mitochondrial DNA: structure and evolution** *Intl Rev. Cytology* 1992, 141: 173-216

29 Yokobori S, Fukuda N, Nakamura M, Aoyama T, Oshima T: **Long-term conservation of six duplicated structural genes in cephalopod mitochondrial genomes** *Mol Biol Evol* 2004, 21(11): 2034-2046

22 Yokobori S, Pääbo S: **Transfer RNA editing in land snail mitochondria** *Proc Natl Acad Sci* 1995, 92(22): 10432-10435

Figures

Figure 1 – Mitochondrial gene map of the cephalopod mollusk *Nautilus* sp.

Genes for proteins and rRNAs are shown with standard abbreviations with an arrow indicating the direction of transcription. Genes for tRNAs are designated by a single letter for the corresponding amino acid, with the two leucine and two serine tRNAs differentiated by numeral (S1, S2, L1, and L2 recognizing codons AGN, UCN, CUN, and UUR, respectively). tRNA genes shown outside the circle are transcribed clockwise and those inside the circle are transcribed counter-clockwise. The largest non-coding region is designated "nc".

Figure 2 – Reconstruction of mitochondrial genome rearrangements for *Nautilus* sp.

At the top is the nearly complete gene arrangement for *Phoronis architecta* [Helfenbein and Boore 2004], a presumed outgroup to the mollusks, shown to polarize two of the cephalopod rearrangements: Having *trnP* in opposite orientation to *nad6* and *nad1* is the ancestral condition, as is having *trnD* between *cox2* and *atp8*. The only two differences between the chiton *Katharina tunicata* [Boore and Brown 1994] and the octopus is the inversion of *trnP* in the octopus and the transposition of *trnD* in the chiton. (No attempt is being made here to reconstruct all of the rearrangements between the phoronid and the chiton.) The arrangement found in the Chambered Nautilus, then, can be reconstructed by the additional switch in order of two large blocks of genes plus transpositions of *trnF* and *trnT*; whether these changes were sequential or simultaneous is undetermined. Genes are not drawn to scale and are abbreviated as in Fig. 1 except that underlining signifies right-to-left transcriptional orientation. All genomes are circular and only graphically linearized at an arbitrarily chosen point.

Figure 3 – Greatly abbreviated sequence of *Nautilus* sp. mtDNA

To save space the middle portions of many genes are replaced by a numeral indicating the number of omitted nucleotides. Gene orientation is specified by a dart (>). Stop codons are shown by asterisks whether complete or abbreviated, with a plus symbol indicating an alternative that overlaps the downstream gene. Down-facing arrows mark repeats found in the largest non-coding region. When not conforming to the genetic code, the presumed initiator methionine (M) is in parentheses.

Figure 4 – Plot of A+C and G+T composition along mtDNAs of *Nautilus* sp. and *Katharina tunicata* using a sliding window of 100 nucleotides

Numbering of nucleotides begins as the arbitrarily chosen *cox1* (as in Fig. 3 for *Nautilus* sp.). The scaled gene maps are also presented. tRNA genes are pictured but not labeled. Underlining indicates right-to-left transcriptional orientation. Numerals label each non-coding region larger than 20 nts, which are then projected onto the plot by gray highlighting. Several of these correspond to positions where there is a shift in nucleotide bias. Asterisks beside two of the numerals for *K. tunicata* indicate some ambiguity where these may instead be supernumerary tRNA genes (Boore and Brown 1994). Red bars show the major transposition between the two genomes (see Fig. 2).

Figure 5 – *Nautilus* sp. mitochondrial tRNA gene sequences folded into typical cloverleaf structures

Lower case "a" in parentheses indicates likely replacements by (poly)adenylation after transcript cleavage at the downstream tRNA (see text for explanation). Structural features are shown on

tRNA(V). Also shown is the secondary structure possible for the repeats in the large non-coding region that appear to be pseudogenes.

Tables

Table 1 – Codon usage for the 13 mitochondrial proteins of *Nautilus* sp.

The total number of codons is 3711. Stop codons were not included in this count. Here the plus strand refers arbitrarily to the one that contains *cox1*.

AA	Codon	All protein encoding genes		Plus strand only	Minus strand only	AA	Codon	All protein encoding genes		Plus strand only	Minus strand only
F	TTT	61 %	225	67	158	I	ATT	38 %	141	83	58
	TTC	18 %	65	62	3		ATC	15 %	57	50	7
L	TTA	40 %	149	89	60	M	ATA	19 %	71	49	22
	TTG	57 %	211	16	195		ATG	25 %	92	21	71
S	TCT	22 %	83	35	48	T	ACT	11 %	41	24	17
	TCC	13 %	50	49	1		ACC	12 %	43	39	4
	TCA	15 %	57	49	8		ACA	12 %	44	40	4
	TCG	04 %	16	3	13		ACG	04 %	14	4	10
Y	TAT	25 %	92	14	78	N	AAT	14 %	53	19	34
	TAC	13 %	49	41	8		AAC	12 %	45	40	5
*	TAA	00 %	0	0	0	K	AAA	10 %	36	24	12
*	TAG	00 %	0	0	0		AAG	11 %	42	10	32
C	TGT	18 %	66	6	60	S	AGT	11 %	39	8	31
	TGC	01 %	5	3	2		AGC	05 %	17	12	5
W	TGA	14 %	52	41	11		AGA	12 %	44	24	20

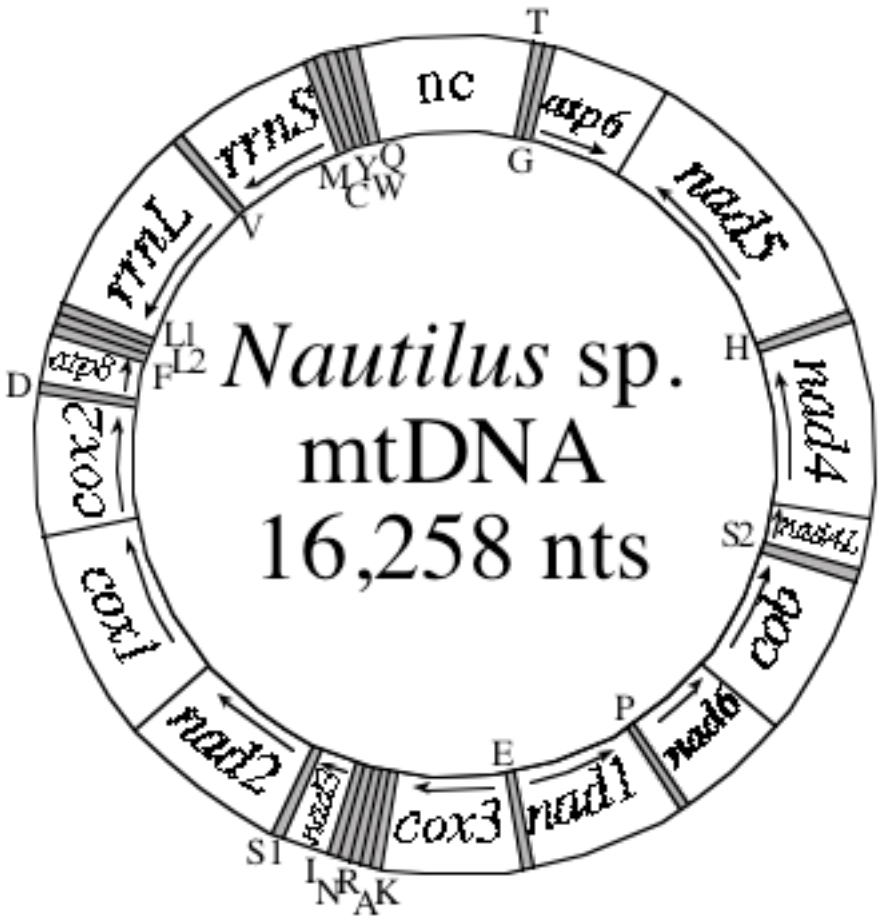
	TGG	17 %	64	16	48		AGG	20 %	73	4	69
L	CTT	22 %	80	53	27	V	GTT	45 %	168	37	131
	CTC	13 %	48	47	1		GTC	08 %	28	26	2
	CTA	21 %	77	72	5		GTA	16 %	61	36	25
	CTG	08 %	28	13	15		GTG	39 %	145	19	126
P	CCT	14 %	53	22	31	A	GCT	24 %	90	31	59
	CCC	08 %	31	26	5		GCC	09 %	35	29	6
	CCA	09 %	33	31	2		GCA	12 %	45	35	10
	CCG	05 %	19	5	14		GCG	06 %	23	0	23
H	CAT	12 %	44	24	20	D	GAT	12 %	45	18	27
	CAC	10 %	36	29	7		GAC	07 %	25	21	4
Q	CAA	12 %	43	34	9	E	GAA	09 %	35	27	8
	CAG	05 %	20	5	15		GAG	15 %	56	14	42
R	CGT	05 %	17	5	12	G	GGT	17 %	62	19	43
	CGC	02 %	9	8	1		GGC	08 %	29	19	10
	CGA	04 %	16	14	2		GGA	28 %	104	54	50
	CGG	05 %	17	4	13		GGG	41 %	153	19	134

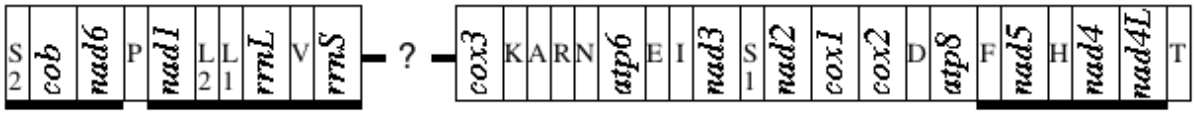
Table 2 – Number of nucleotides at gene boundaries

Negative numbers refer to overlapping nucleotides.

Boundary	nts	Boundary	nts
cox1 – cox2	2	*trnH – nad4*	0
cox2 – trnD	0	*nad4 – nad4L*	– 7
trnD – atp8	1	*nad4L – trnS2*	14
atp8 – trnF	20	*trnS2 – cob*	0
trnF – trnL2	0	*cob – nad6*	0
trnL2 – trnL1	– 1	*nad6 – trnP*	2
trnL1 – rrnL	0	*trnP – nad1*	5
rrnL – trnV	0	*nad1 – trnE*	26
trnV – rrnS	0	*trnE – cox3*	102
rrnS – trnM	0	*cox3 – trnK*	0
trnM – trnC	3	*trnK – trnA*	– 4
trnC – trnY	0	*trnA – trnR*	14
trnY – trnW	0	*trnR – trnN*	39
trnW – trnQ	– 2	*trnN – trnI*	4
trnQ – trnT	972	*trnI – nad3*	0
trnT – trnG	90	*nad3 – trnS1*	0
trnG – atp6	97	*trnS1 – nad2*	0
atp6 – nad5	23	*nad2 – cox1*	2
nad5 – trnH	0		

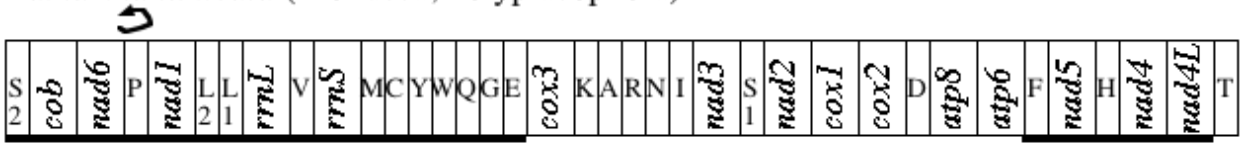
Keywords: Chambered nautilus, mollusk, mitochondria, mtDNA, evolution, genome


Nautilus sp.
mtDNA
16,258 nts
nc
atp6
nad5
nad4
nad4L
cob
nad6
nad1
cox3
nad3
nad2
cox1
cox2
atp8
rrnL
rrnS
T
G
H
S2
P
E
S1
I
N
R
A
K
L1
L2
F
D
V
M
Y
Q
C
W

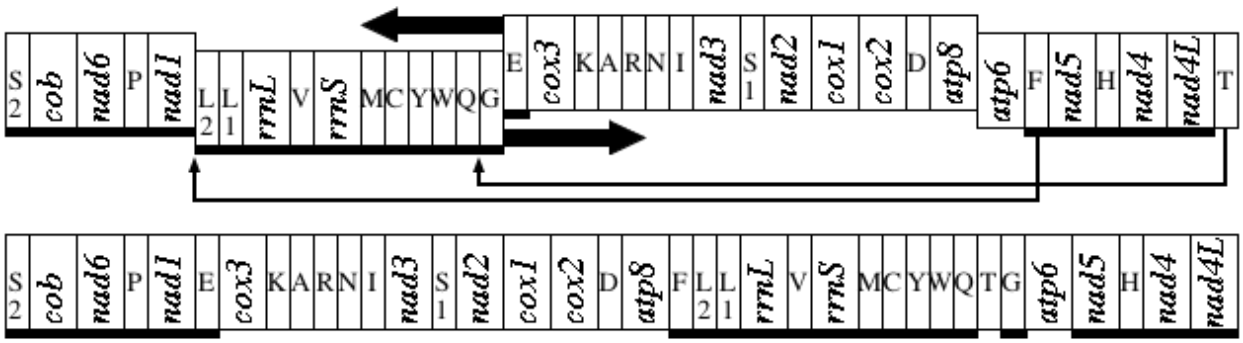


Phoronis architecta (Phoronida)

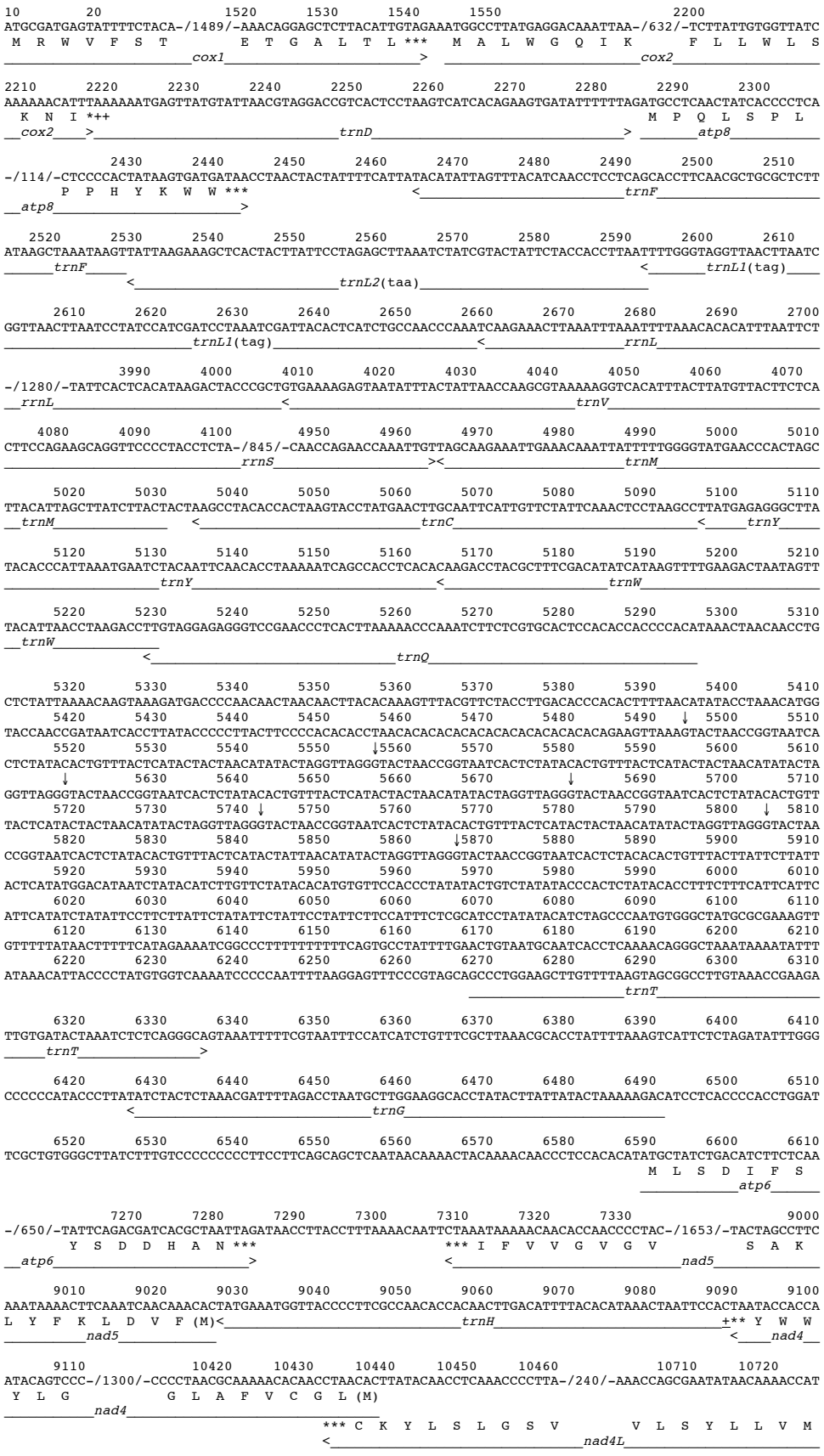
Katharina tunicata (Mollusca, Polyplacophora)

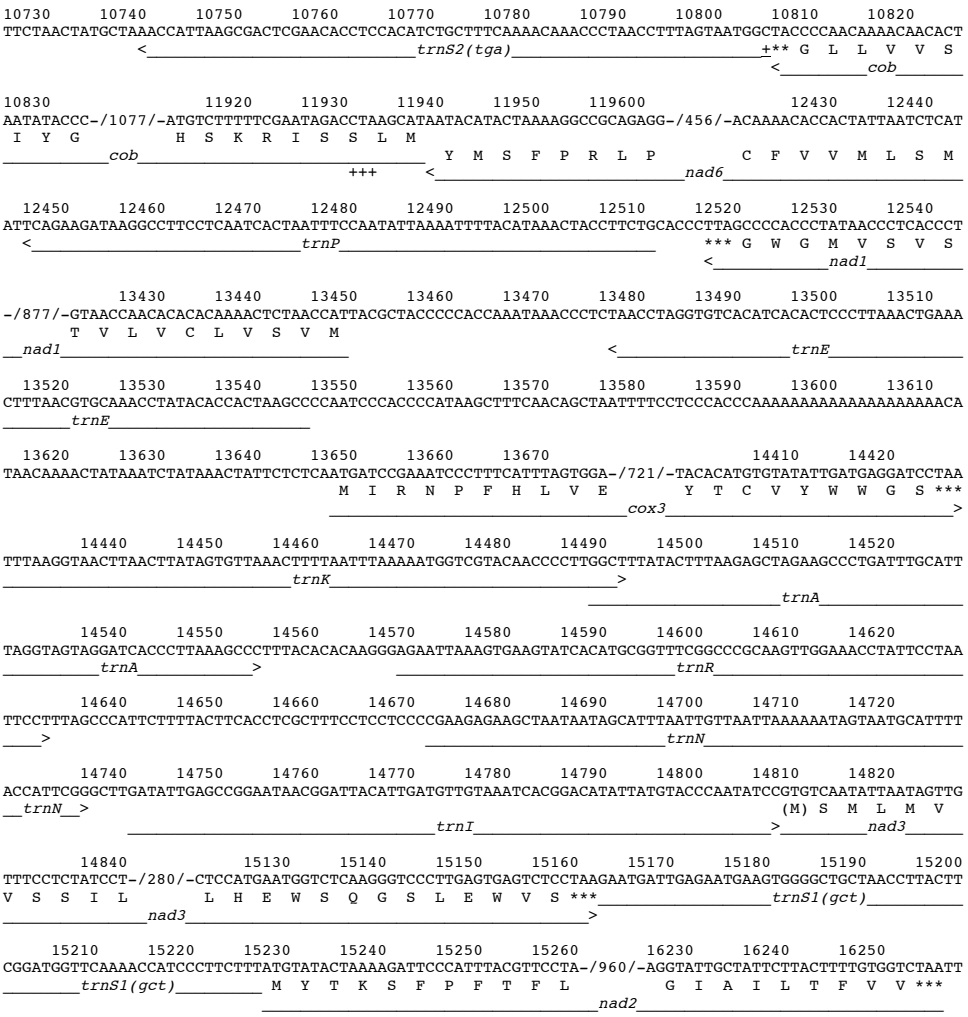


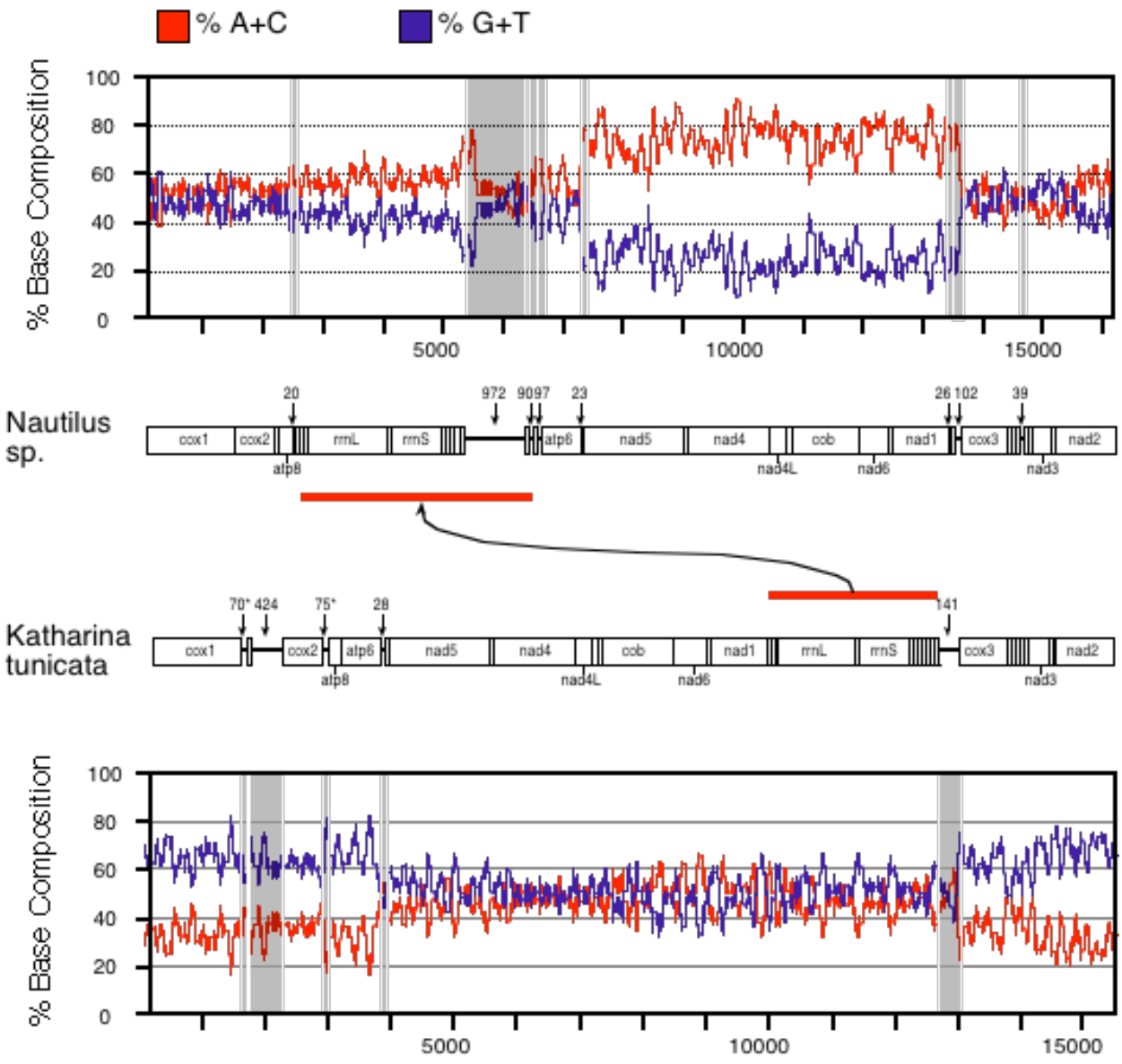
Octopus vulgaris (Mollusca, Cephalopoda) (and common ancestor with *Nautilus*)

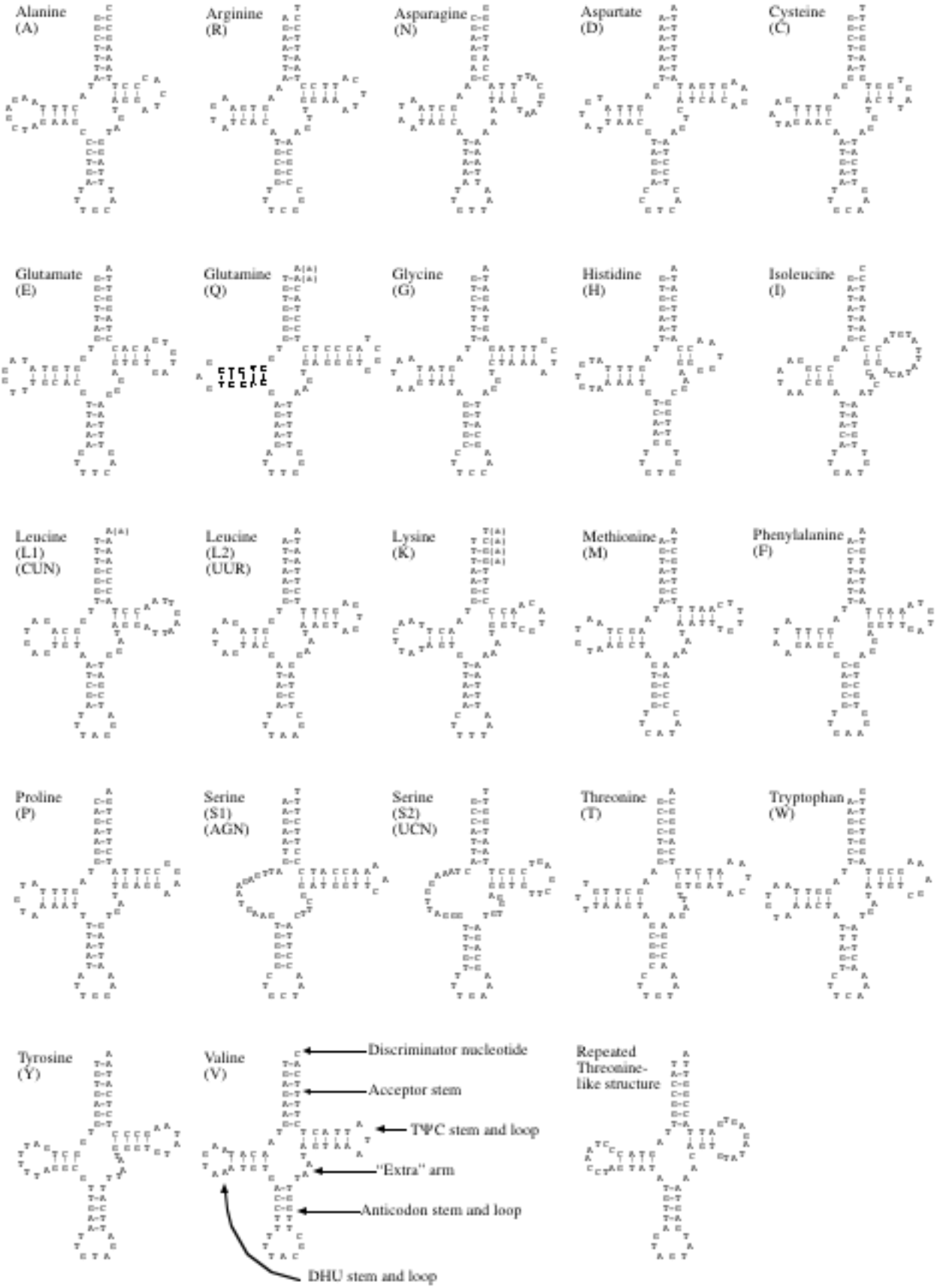


Nautilus sp. (Mollusca, Cephalopoda)


cox1
cox2
trnD
atp8
trnF
trnL2(taa)
trnL1(tag)
rrnL
trnV
rrnS
trnM
trnC
trnY
trnW
trnQ
trnT
trnG
atp6
nad5
trnH
nad4
nad4L


10730 10740 10750 10760 10770 10780 10790 10800 10810 10820
TTCTAACTATGCTAAACCATTAAGCGACTCGAACACCTCCACATCTGCTTTCAAAACAAACCCTAACCTTTAGTAATGGCTACCCCAACAAAACAACACT
trnS2(tga)
** G L L V V S
cob
10830 11920 11930 11940 11950 119600 12430 12440
AATATACCC-/1077/-ATGTCTTTTTCGAATAGACCTAAGCATAATACATACTAAAAGGCCGCAGAGG-/456/-ACAAAACACCACTATTAATCTCAT
I Y G H S K R I S S L M
cob
Y M S F P R L P C F V V M L S M
nad6
12450 12460 12470 12480 12490 12500 12510 12520 12530 12540
ATTCAGAAGATAAGGCCTTCCTCAATCACTAATTTCCAATATTAAAATTTTACATAAACTACCTTCTGCACCCTTAGCCCCACCCTATAACCCTCACCCT
trnP
*** G W G M V S V S
nad1
13430 13440 13450 13460 13470 13480 13490 13500 13510
-/877/-GTAACCAACACACACAAAACTCTAACCATTACGCTACCCCCACCAAATAAACCCTCTAACCTAGGTGTCACATCACACTCCCTTAAACTGAAA
T V L V C L V S V M
nad1
trnE
13520 13530 13540 13550 13560 13570 13580 13590 13600 13610
CTTTAACGTGCAAACCTATACACCACTAAGCCCCAATCCCACCCCATAAGCTTTCAACAGCTAATTTTCCTCCCACCCAAAAAAAAAAAAAAAAAAAAAA
trnE
13620 13630 13640 13650 13660 13670 14410 14420
TAACAAAACTATAAATCTATAAACTATTCTCTCAATGATCCGAAATCCCTTTCATTTAGTGGA-/721/-TACACATGTGTATATTGATGAGGATCCTAA
M I R N P F H L V E Y T C V Y W W G S ***
cox3
14440 14450 14460 14470 14480 14490 14500 14510 14520
TTTAAGGTAACTTAACTTATAGTGTTAAACTTTTAATTTAAAAATGGTCGTACAACCCCTTGGCTTTATACTTTAAGAGCTAGAAGCCCTGATTTGCATT
trnK
trnA
14540 14550 14560 14570 14580 14590 14600 14610 14620
TAGGTAGTAGGATCACCCTTAAAGCCCTTTACACACAAGGGAGAATTAAAGTGAAGTATCACATGCGGTTTCGGCCCGCAAGTTGGAAACCTATTCCTAA
trnA
trnR
14640 14650 14660 14670 14680 14690 14700 14710 14720
TTCCTTTAGCCCATTCTTTTACTTCACCTCGCTTTCCTCCTCCCCGAAGAGAAGCTAATAATAGCATTTAATTGTTAATTAAAAAATAGTAATGCATTTT
trnN
14740 14750 14760 14770 14780 14790 14800 14810 14820
ACCATTCGGGCTTGATATTGAGCCGGAATAACGGATTACATTGATGTTGTAAATCACGGACATATTATGTACCCAATATCCGTGTCAATATTAATAGTTG
trnN
(M) S M L M V
trnI
nad3
14840 15130 15140 15150 15160 15170 15180 15190 15200
TTTCCTCTATCCT-/280/-CTCCATGAATGGTCTCAAGGGTCCCTTGAGTGAGTCTCCTAAGAATGATTGAGAATGAAGTGGGGCTGCTAACCTTACTT
V S S I L L H E W S Q G S L E W V S ***
trnS1(gct)
nad3
15210 15220 15230 15240 15250 15260 16230 16240 16250
CGGATGGTTCAAAACCATCCCTTCTTTATGTATACTAAAAGATTCCCATTTACGTTCCTA-/960/-AGGTATTGCTATTCTTACTTTTGTGGTCTAATT
trnS1(gct)
M Y T K S F P F T F L G I A I L T F V V ***
nad2


% A+C
% G+T
% Base Composition
100
80
60
40
20
0
5000
10000
15000
Nautilus sp.
20
972
9097
23
26
102
39
cox1
cox2
rrnL
rrnS
atp6
nad5
nad4
cob
nad1
cox3
nad2
atp8
nad4L
nad6
nad3
Katharina tunicata
70*
424
75*
28
141
cox1
cox2
atp6
nad5
nad4
cob
nad1
rrnL
rrnS
cox3
nad2
atp8
nad4L
nad6
nad3
% Base Composition
100
80
60
40
20
0
5000
10000
15000


Alanine
(A)
Arginine
(R)
Asparagine
(N)
Aspartate
(D)
Cysteine
(C)
Glutamate
(E)
Glutamine
(Q)
Glycine
(G)
Histidine
(H)
Isoleucine
(I)
Leucine
(L1)
(CUN)
Leucine
(L2)
(UUR)
Lysine
(K)
Methionine
(M)
Phenylalanine
(F)
Proline
(P)
Serine
(S1)
(AGN)
Serine
(S2)
(UCN)
Threonine
(T)
Tryptophan
(W)
Tyrosine
(Y)
Valine
(V)
Discriminator nucleotide
Acceptor stem
TΨC stem and loop
"Extra" arm
Anticodon stem and loop
DHU stem and loop
Repeated
Threonine-
like structure