Conserved Sequence Motifs, Alignment, and Secondary Structure for the Third Domain of Animal 12S rRNA

Robert E. Hickson,[1] *Chris Simon,*† *Alan Cooper,*‡[2] *Greg S. Spicer,*§ *Jack Sullivan,*† *and David Penny**

*School of Biological Sciences, Massey University; †Department of Ecology and Evolutionary Biology, University of Connecticut; ‡School of Biological Sciences, Victoria University of Wellington; §Institute of Molecular Medical Sciences

Secondary structure models are an important step for aligning sequences, understanding probabilities of nucleotide substitutions, and evaluating the reliability of phylogenetic reconstructions. A set of conserved sequence motifs is derived from comparative sequence analysis of 184 invertebrate and vertebrate taxa (including many taxa from the same genera, families, and orders) with reference to a secondary structure model for domain III of animal mitochondrial small subunit (12S) ribosomal RNA. A template is presented to assist with secondary structure drawing. Our model is similar to previous models but is more specific to mitochondrial DNA, fitting both invertebrate and vertebrate groups, including taxa with markedly different nucleotide compositions. The second half of the domain III sequence can be difficult to align precisely, even when secondary structure information is considered. This is especially true for comparsions of anciently diverged taxa, but well-conserved motifs assist in determining biologically meaningful alignments. Patterns of conservation and variability in both paired and unpaired regions make differential phylogenetic weighting in terms of "stems" and "loops" unsatisfactory. We emphasize looking carefully at the sequence data before and during analyses, and advocate the use of conserved motifs and other secondary structure information for assessing sequencing fidelity.

Introduction

Assessment of the reliability of phylogenetic reconstructions based on information contained within DNA sequences is now a major scientific endeavor. Much of the recent emphasis has been directed toward the reliability of phylogenetic analyses after sequences have been aligned (see, for example, Felsenstein 1988; Swofford and Olsen 1990; Bandelt and Dress 1992; Penny et al. 1993; Lento et al. 1995). The importance of alignment on the resulting phylogeny, however, has been recognized (see for example, Hein 1990; Lake 1991; Thorne, Kishino, and Felsenstein 1991; Gatesy, de Salle, and Wheeler 1993; Collins, Kraus, and Estabrook 1994; Kjer 1995). Structural information provides a framework for the alignment and analysis of protein-encoding genes (for example, Doolittle 1986; Irwin, Kocher, and Wilson 1991; Bell, Coggins, and Milner-White 1993), but alignments of ribosomal RNA (rRNA) genes have used such information less consistently. Sequences are frequently aligned by eye or with a multiple sequence alignment program without taking account of secondary structure. Analysis of genes in a structural context also provides a functional rationale for variation in rates of changes at different sites or regions of the molecule (Hickson 1993; van de Peer et al. 1993; Vawter and Brown 1993; Simon et al. 1994). Criteria are needed to derive satisfactory alignments and so improve homology assessments critical to phylogenetic comparisons.

Extensive experimental and comparative sequence analyses of prokaryotic and eukaryotic small and large subunit (SSU and LSU, respectively) rRNAs have greatly enhanced our understanding of rRNA secondary structures (see Noller et al. 1990; de Rijk et al. 1994; Gutell 1994; Maidak et al. 1994; van de Peer et al. 1994). In this paper we examine in more detail the secondary structure of one part of animal mitochondrial SSU (12S) rRNA. This is domain III of the molecule, encompassing helices (base paired regions) 32–48 (see fig. 1). Highly conserved flanking sequences (Kocher et al. 1989; Simon et al. 1994) make it the most commonly amplified region of the 12S rRNA (reviewed in Hillis and Dixon 1991; Simon et al. 1994). The present study is an extension and generalization of our use of 12S rRNA secondary structure information to investigate evolutionary questions among insects (Simon et al. 1990, 1994; Simon 1991; Spicer, unpublished), lizards (Hickson 1993), birds (Cooper 1994), and rodents (Sullivan, Holsinger, and Simon, in press). We found that alignment and analyses were greatly improved when secondary structure information was incorporated.

Mitochondrial sequences in SSU rRNA compilations are dominated by mammalian sequences (Gutell 1994; van de Peer et al. 1994), although models for

[1] Present address: Department of Genetics and Molecular Biology, University of Hawaii at Manoa.

[2] Present address: Molecular Genetics Laboratory, National Zoological Park.

Key words: 12S rRNA, SSU rRNA, secondary structure model, sequence alignment, motifs, molecular evolution.

Address for correspondence and reprints: R. E. Hickson, Department of Genetics and Molecular Biology, University of Hawaii at Manoa, Honolulu, Hawaii 96822. E-mail: hickson@hawaii.edu.

Mol. Biol. Evol. 13(1):150–169. 1996
© 1996 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1054489 by guest on 20 August 2022

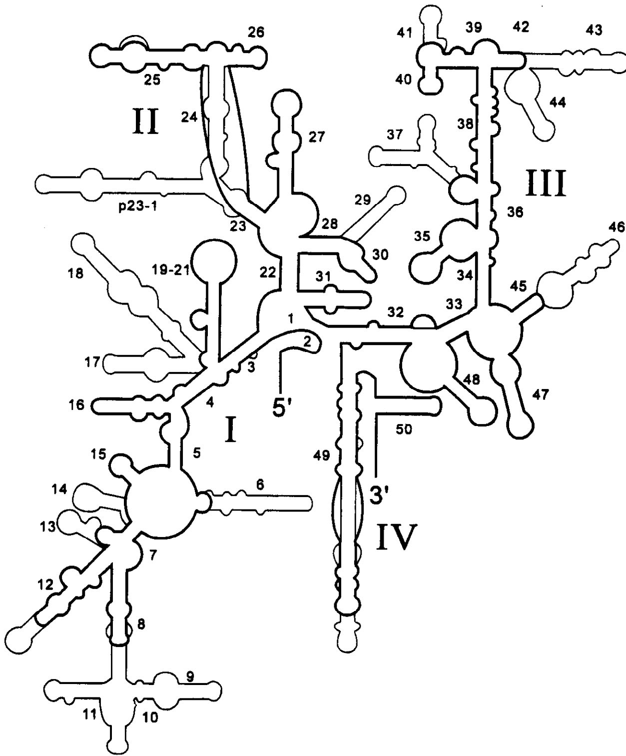


FIG. 1.—The secondary structure of the SSU (16S-like) rRNA of *E. coli.* All four domains are shown and helices are numbered (as in van de Peer et al. 1994). The parts of the molecule occurring in mitochondrial SSU rRNA are indicated by thick lines. Paired regions are shown as parallel lines while curved lines denote unpaired regions. These drawings were based on Gutell (1994) and so do not correspond precisely to the model we present here.

mitochondrial SSU secondary structures have been presented for a range of taxa (Bibb et al. 1981; Gutell et al. 1985; Clary and Wolstenholme 1985, 1987; Dunon-Bluteau and Brun 1986; Hixson and Brown 1986; Dams et al. 1988; Neefs et al. 1993; Gutell 1994; Pont-Kingdon et al. 1994; van de Peer et al. 1994). These structures differ from each other to varying degrees, as will be discussed below. A model originally constructed from comparison of domain III mitochondrial sequences from several avian orders (Cooper, unpublished) and vertebrate classes (Hickson 1993) proved to be very similar to the models of Gutell (1994) and van de Peer et al. (1994) but with minor modifications that may well be specific to mitochondria.

In the present paper we use a large number of sequences (184) from invertebrate and vertebrate taxa to examine the general SSU rRNA secondary structure models of Gutell (1994) and van de Peer et al. (1994) for domain III of animal SSU mtDNA. Using this approach we have identified well-conserved sequence motifs. These motifs are advantageous for both alignment and drawing of animal mitochondrial secondary structure diagrams. Comparison of sequences from taxa in the same genera, families, and orders provided a much clearer view of the mitochondrial secondary structure in animals because, as pointed out by Gutell, Larsen, and Woese (1994), the structural evidence from covariation of bases within phylogenetically restricted groups is more significant than evidence from covariation involving distantly related sequences.

An alignment and basic structure for domain III in animal mitochondrial SSU rRNA is presented indicating both conserved and variable sites. A 12S secondary structure template is included to assist determination of the secondary structure of domain III. The secondary structure-based alignment and the motifs also have value in identifying possible errors in sequencing and we discuss some published 12S rRNA sequence data that warrant further checking.

Materials and Methods

Secondary Structure Models for rRNA

Structurally, the SSU rRNA is divided into four domains (labeled I, II, III, and IV), each separated by a single-stranded region (fig. 1). Domains III and IV are the most highly conserved domains (de Rijk et al. 1992). The mitochondrial form of the SSU rRNA still has four domains but in comparison to the prokaryote structure, from which it is derived, lacks some helices (fig. 1). Several models for the prokaryote SSU rRNA were proposed based on experimental and comparative sequence analyses of the *Escherichia coli* 16S gene (Noller and Woese 1981; Stiegler et al. 1981; Zwieb, Glotz, and Brimacombe 1981). These have been steadily improved over the years using comparative sequence analysis by groups in the USA (see Gutell et al. 1985; Gutell 1994) and Belgium (see Huysmans and de Wachter 1985; van de Peer et al. 1994).

These two groups have produced compilations of rRNA sequences (both large and small subunits) for a wide range of organisms and the databases are regularly updated. Use of the most recent compilation for alignment and secondary structure analyses is critical since earlier versions, based on less information, can be inaccurate (Neefs et al. 1993; Kjer, Baldridge, and Fallon 1994). The compilations maintained by the de Wachter group (de Rijk et al. 1994; van de Peer et al. 1994) are available in electronic form from the anonymous FTP site uiam3.uia.ac.be, and compilations and structures from the Ribosomal Database Project (Maidak et al. 1994) can be retrieved from the FTP site rdp.life.uiuc.edu or pundit.colorado.edu (in subdirectory /pub/RNA/16S). The SSU rRNA compilation of van de Peer et al. (1994) and the model of Gutell (1994), formed the basis for our refinement of a secondary structure model for domain

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1055489 by guest on 20 August 2022

Table 1
Representative Vertebrate and Invertebrate Taxa Used in the Alignment of Domain III 12S rRNA

Phylum	Class	Order	Representative Taxon	Sequence No.	Source
Chordata	Mammalia:	Artiodactyla	Cow	1–15	J01394
		Carnivora	Sea lion	16–23	U12850
		Rodentia	White-footed mouse	24–29	Sullivan, unpublished
		Primates	Human	30–34	J01415
		Marsupialia	Opossum	35	Z29573
	Aves:	Dinornithiformes	Ratite bird	36–77	X67634
	Reptilia:	Squamata	Scincid lizard	78–97	Hickson (1993)
	Amphibia:	Anura	*Xenopus*	98–117	M10217
	Osteichthyes:	Lepidosireniformes	Lungfish	118–121	M87535
	Osteichthyes:	Cypriniformes	Cyprinid fish	122	M91245
	Osteichthyes:	Coelocanthiformes	Coelocanth	123	M87534
	Agnatha:	Petromyzontiformes	Lamprey	124	U11880
Echinodermata ..	Echinoidea	Echinoida	Sea urchin	125–130	J04815
Arthropoda	Insecta:	Diptera	*Drosophila*	131	X03240
	Insecta:	Hymenoptera	Honeybee	132	L06178
	Insecta:	Hemiptera	Cicada	133–138	Simon, unpublished
	Insecta:	Orthoptera	Locust	139	Flook, Rowell, and Gillessen (1995)
	Insecta:	Odonata	Damselfly	140–163	Spicer, unpublished
	Insecta:	Thysanura	Silverfish	164	L02381
	Chilopoda:	Scutigeromorpha	Centipede	165	L02376
	Arachnida:	Aranea	Spider	166–175	U00118
	Arachnida:	Scorpionida	Scorpion	176	L02397
	Crustacea:	Decapoda	Crab	177	L02396
	Crustacea:	Branchiopoda	Brine shrimp	178	X69067
	Onychophora:	Peripatopsidae	Onychophoran	179	L02380
Mollusca	Polyplacophora:	Ischnochitonida	Chiton	180	L02388
	Gastropoda:	Archaeogastropoda	Snail	181	L02389
Annelida	Oligochaeta:	Lumbricoidea	Earthworm	182	L02392
Nematoda	Nematoidea:	Ascaroidea	Nematode	183	X54252
Cnidaria	Actinaria:	Metridiidae	Sea anemone	184	Pont-Kingdon et al. (1994)

NOTE.—The number of species that each taxon summarizes is shown and a complete listing of all taxa used is given in the FTP file "TaxaList.12S" (see text). The GenBank accession numbers, if available, for the representative taxa are given. Note that several avian and amphibian orders were used.

III of animal mitochondrial 12S rRNA. We have used the numbering system of Neefs et al. (1993) to identify individual helices (fig. 1).

Drawings of complete 12S rRNA secondary structures have been published for mouse (Bibb et al. 1981), cow (Gutell et al. 1985; Gutell 1994), rat (Gutell et al. 1985), *Drosophila* (Clary and Wolstenholme 1985, 1987), *Xenopus* (Dunon-Bluteau and Brun 1986), primates (Hixson and Brown 1986), human (Neefs et al. 1993; Gutell 1994), nematodes (Gutell 1994), a sea anemone (Pont-Kingdon et al. 1994), and carp (van de Peer et al. 1994). While generally similar, they often differ in the presence or structure of peripheral helices. Complete 12S rRNA sequences are currently available for a variety of animal taxa, but a much greater range of taxa have data available for domain III only. In this paper we have made use of sequence information from 184 taxa (table 1). While part of domain II can be amplified by the universal 12S rRNA primers we have have restricted our analyses to domain III (positions 1,174–1,477 in the human sequence of Anderson et al. [1981]) since many of the invertebrate taxa we used lacked the additional sequence information.

For the purposes of these analyses we consider four structural classes in rRNA (Varani and Pardi 1994):

1. Helices (or stems), which are regions of complementary base pairing.
2. Bulges, unpaired bases within a helix.
3. Loops, unpaired regions of nucleotides within helices (internal loops) or between the proximal and distal arms of a helix (hairpin loops).
4. Unpaired regions not within helices.

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1055489 by guest on 20 August 2022

As we discuss later however, it makes little sense to quantify paired versus unpaired rRNA bases in terms of variability because each of these classes has both highly variable and highly conserved members (Simon 1991). Furthermore, highly conserved helices and unpaired regions can have variable bases within them (van der Peer et al. 1993; Sullivan, Holsinger, and Simon, in press).

A Structurally Based Alignment

Comparative sequence analysis is now the most common method for refining secondary structure models, and the identification of both conserved motifs and covarying nucleotide substitutions is an important and powerful test of the model (Gutell, Larsen, and Woese 1994). Covarying, or compensatory, substitutions are those whose patterns of substitution correlate with nucleotide replacements at other positions and can be indicative of conventional, or non-conventional, base pairings (Gutell, Larsen, and Woese 1994). Conserved sequence motifs (fig. 2, Appendix 1) can serve as alignment foci and to localize helices when drawing secondary structures. Throughout this paper we assume that secondary structural features, such as helices, are often indicative of homology. This is justified because the SSU rRNA is a relatively slowly evolving molecule, the secondary structure of the *E. coli* SSU rRNA molecule has been examined in great detail (see Noller et al. 1990), and subsequent comparisons of a large number of SSU rRNAs have indicated common structures (Gutell 1994; van de Peer et al. 1994). Although, as we indicate later, there is variability in the size and position of some helices so not all bases in some helices can be considered homologous. Following Zuker (1989), hairpin loops had to consist of at least three, preferably four, nucleotides so that the helices were not distorted.

Identification of fixed compensatory mutations (changes preserving base-pairing in helical regions) was especially useful for finding helices whose position varied among groups. However, not all substitutions in a helical region will result in a fixed compensatory change in the pairing partner. A substitution on one side of a helix can have one of six results (fig. 3):

1. Noncanonical (non-Watson–Crick) pairing, but stable purine-pyrimidine bonds (fig. 3*a*).
2. Formation of a stable symmetrical bulge (fig. 3*b*).
3. Occurrence of a compensatory change that retains Watson–Crick base pairing (fig. 3*b*).
4. Return of the pair to its original state (fig. 3*b*).
5. The persistence of a nonsymmetrical bulge (fig. 3*c*; single base bulges, however, tend to be highly conserved in location and can be important in protein interactions).
6. Stem slippage (fig. 3*c*).

Base pairing between guanine and uracil was permitted, as such pairings appear to be structurally stable in RNAs (Topal and Fresco 1976). In some situations nonconventional base pairing (for instance between adenine and cytosine, adenine and guanine, or between two uracils) may also occur (Freier et al. 1986; SantaLucia, Kierzek, and Turner 1991; Gutell, Larsen, and Woese 1994; Varani and Pardi 1994). Support for these pairings can be inferred from covariation (Gutell, Larsen, and Woese 1994). For our mitochondrial data set adenine and cytosine pairings were included in some helices (A·C bonds were particularly common in the avian data set), while adeninie·guanine bonds were required less frequently (see Results).

Previous studies of rRNA structures have focused on distantly related taxa and taxa represented by only one or two species. We compared taxa from within the same genera, families, and orders to study substitutional changes in helices (see table 1). The avian and odonate (damselfly and dragonfly) data sets were especially informative in this regard (see below).

We also calculated minimal free energies for the folding of some potential helix structures using the MFOLD algorithm (Zuker 1989; Zuker, Jaeger, and Turner 1991) in the University of Wisconsin Genetics Computer Group (GCG) Package, version 7.2 (Genetics Computer Group 1991). The program takes account of the energetic costs and benefits of potential base pairings and the occurrence, locations, and interactions among unpaired nucleotides (Zuker 1989; Zuker, Jaeger, and Turner 1991). The MFOLD program can display both optimal and suboptimal folded structures. Some structures determined on the basis of comparative sequence analysis have been found to have a minimal free energy within 10% of the value of the optimal structure calculated by MFOLD (Zuker, Jaeger, and Turner 1991). Long-range interactions within the molecule and RNA–protein interactions may, however, result in foldings different from the optimal energetic solution (Zuker, Jaeger, and Turner 1991; and see Results), so energetic calculations by themselves can be poor predictors of secondary structure. Other folding programs are available (e.g., Abrahams et al. 1990) but were not used in this study.

Results

Figure 2 presents an alignment and secondary structure for domain III based on comparison of the 184 taxa (table 1). Space limitations prevent inclusion in figure 2 of all the groups we analyzed, but an alignment of more taxa is available from REH or CS, or via anonymous ftp from 130.123.1.3 (as file "12S.aln" in the subdirectory /pub/farside; available until at least January

Downloaded from https://academic.oup.com/mbe/article/13/1/150/…/by guest on 26 August 2022

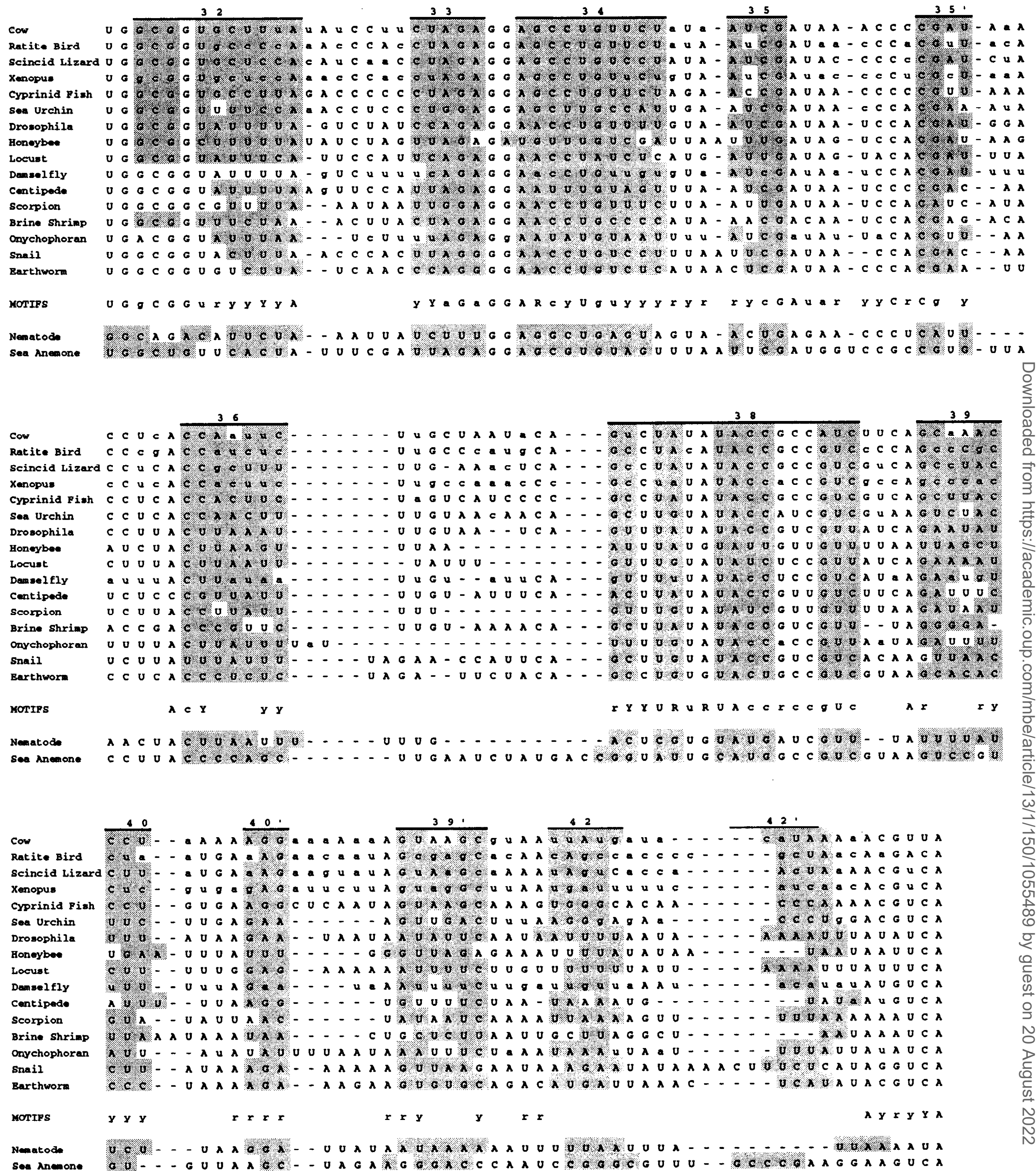


FIG. 2.—The secondary structure alignment of domain III animal 12S rRNA. The helices (stems) are listed above the alignment, while the bases involved in pairing are shaded for each taxon. The region shown corresponds to positions 1,174–1,477 in the human sequence of Anderson et al. (1981). Gaps are indicated by dashes, with some of the gaps reflecting our use of additional taxa that are not shown here. Where we have used sequences from more than one species (see table 1) variable positions relative to the representative taxon are indicated by lowercase letters, while the significance of lowercase characters for the motifs is discussed in the text. For some of the more conserved helices (e.g., helix 34) additional bases may pair, but we have only indicated these if several members of a group show this. The FTP file "TaxaList.12S" lists the species used in the determination of motifs.

The 18 representative species used here are *Bos taurus* (cow), *Megalapteryx didinus* (ratite bird), *Leiolopisma nigriplantare polychroma* (scincid lizard), *Xenopus laevis* (*Xenopus*), *Crossostoma lacustre* (cyprinid fish), *Paracentrotus lividus* (sea urchin), *Drosophila yakuba* (*Dro-*

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1055489 by guest on 20 August 2022

```
               38'             36'          34'              45           45'
Cow            GGUCAAGGUGUAaCCuAU-GaaauGG-gAAGAAAUGGGCUA-CAUUcUCUacaccaAgAgaA-
Ratite Bird    GGUCAAGGUAUAGCauAu-gagauGG--aAGaAAUGGGCUA-CAUUuuCUAaca--UAGaacA
Scincid Lizard GGUCAAGGUGUAGCAcAU-aaaguGG--aAGAGAUGGGCUA-CACUCUCUcCC---AGAGa--
Xenopus        GGUCAAgGUGuAgcauAu-gaaguGg-gaaGAaAUgGGCUA-CAuUuucUaccuu-Agaaua-
Cyprinid Fish  GGUCGAGGUGUAGCGUAC-GAAGUGG-GAAGAGAUGGGCUA-CAUUUUCUACU---AGAAUA-
Sea Urchin     GAUCAAGGUGCAGCcUAU-AaGUUGG-GgAUAGCUGAGCUA-CAAUGUUUGAAc--AAAC---
Drosophila     GAUCAAGGUGUAGCUUAU-AUUUAAG--UAAUAAUGGGUUA-CAAUAAAUUUA---UUUA---
Honeybee       AAUCAAAAUGUAGU--AU-AUUUAAGAUAUAAGAUGAA-UAUCAAUAAAUUUAAU-UUUA---
Locust         GGUCAAGGUGCAGCUUAU-AGUUAAG-UGUAUGAUGGGUUA-CAAUAAUUUUUU-UUAU---
Damselfly      GGUCAAGGUGaAgcUuAU-guauAAG--uacuaAUGGgUUA-CAAUAauauUau--uUAu---
Centipede      GAUCAAGGUGUAGUUUAU-AAUAAGG-UUAC-UAUGAAUUA-CAAUUAAUGUA---UUAAU--
Scorpion       AAUCAAGGUGCAGCUU---AAUAUGG-AUGA-AAUGGGUUA-CAAUAUAAAUUA--UAUAU--
Brine Shrimp   GAUCAAGGUGUAGCUUAU-GUCCGGG--AGAGGGUGGGUUA-CAAUAAGAAC----UUA----
Onychophoran   GGUCAAGGUGCAAA--AU-AAAUAAG-UUAU-UAUGUAUUa-CAuUCAAaUAUa--UUG----
Snail          GAUCAAGGUGCAGCCAAU-AAAUAAA-GAAGAGAUGGGUUA-CAAUUAACAACUC-UUA----
Earthworm      GGUCAAAGUGCAGCCCAU-GGGAGGG -AGAUGAUGGGUUA-CACCCUAAACA---AAG----

MOTIFS         ggUCaAggUgyAgy  ry rr  Rg   rr raUgrgYUA CA U

Nematode       GAUAUAUACCCGGUUUAUGAUUUAAG-AAACAUUUGGCCUA-CAAUAUUUU-----AUAU---
Sea Anemone    GGUCUUAUGAUCCG-AAU-GUUGGGG--AUUUUAUGCGCUA-CAUUUUCUU*  *AAGUAAGA-
Anemone insert                    *AUAUAAUGAGAGACUUGAAAGGUAAGCCUCUA*
```

```
                          47                47'          33'
Cow            ucaagcACGAAAGuuauuA-----------UGAAAccaauaAC--cAAAGGAGGAUUUAG
Ratite Bird    ---cccACgaaAgagaaga----------UGAAAcccuccuc--agaAGGcGGAUUUAG
Scincid Lizard ----AcACgAAcaGCAuca---------AUGAAAcaCuGCu---caAAGGuGgAUUUAG
Xenopus        -----aaCgaaagaucu---------cuaUGAAAccagaucgagAaAaggcGGAUUUAG
Cyprinid Fish  ----AGACGAAUAGCAUCA----------UGAAAACUUAAUGCUUGAAGGAGGAUUUAG
Sea Urchin     ----CAGUGgAAggAGgg----------AUGAAAUaCccCUu--gGAAAUUGGAUUCAG
Drosophila     -----AACGGAUAAAAUU---------AUGAAAAAAUUUU---UGAAGGUGGAUUUGG
Honeybee       ---GAAUUAUUUAUUUU-AAUAUAAAUAUGAAAGAGAAUU---UAAAAGUAAA-UUAA
Locust         -----UAUGGAUUUAA----------UUUUGUAAUAUUUUAA---UGAAGGUGGAUUUGA
Damselfly      ------acGgAuuuuuuaau--uuauauuaaaau---------UaAAgguGGAUUugu
Centipede      ------AUGAAUAUUAUUU---------UGAAAAAGAUAAU---UGAAGGUGGAUUUAA
Scorpion       -----UACGGAUUUAAU-----------UGAAAAUUAAUU---UGAAGGUGGAUUCAA
Brine Shrimp   -----AACGGAACUUU--CAUUUUAUCAUGGAGA---------AGAAGGUGGAUUUAG
Onychophoran   -------uAgUGGAUAA--------uUAUUGAAAcAucUAUU--UGAAAUUGAAUCUac
Snail          -----CAUGAAAUUUUU----------AAAGAAAACACAAA--UGAAAUAGGACUUAA
Earthworm      ----AUACGGAAUAUAGU---------ACUAAAAGCUAUA----UAAAUUAUUACUUGG

MOTIFS               rYgrr                yrarr         yRaarr GrAuuuar

Nematode       -----UAUGGAUUUUAGUUUUAGUUAACUAAA------------UGAAAUUGUAAAAGA
Sea Anemone    ----GUUCGGAAGGUCGCU---------GUAAAAACGACU-----GGAAGUUGGAUUUAA
```

```
                   48                   48'                32'
Cow            cAGUAAaCUaagAAUAGAGuG------CUuAGuUGAAuuAGGCc-AUGAAGCACGC
Ratite Bird    CAGUAAAauaggAcaAgaacG------cccauUUUAAgcu-GGCc-CUggggcACGU
Scincid Lizard uAGUAAGAuaa-aCaAGAgA------CuuaucUuAAAcC-AGCc-CUGGAGCGCGC
Xenopus        caGUAAAgagaaacaagagag----uuccucuuUAaAac-ggcc-cuggagcgcGC
Cyprinid Fish  UAGUAAAAAGGAAAUAGAGUG------UCCUUUUGAACCCGGCU-CUGAGGCGCGU
Sea Urchin     CAGUAAGCCCCAcuaAGAc------AAUGGGaCUGAAAAGAGCU-CUGGAAUGCGU
Drosophila     UAGUAAAAUUAUAAAGAUUA-------AUAAUUUGAUUUUAGCU-CUAAAAUAUGU
Honeybee       -AGUAAUUUCUUU--------------AAUUGAAGAUAGUA-UUAAAAAGUGU
Locust         UAGUAAUUUAAAUUAUUUAA-------UUUAACUGAUAUUGGCU-CUGAAGUGUGU
Damselfly      aaGUAAagUaaauuaauu-------aauUuacuUGAuuuugGCU-C
Centipede      UUGUAAUUUCAUUUAUUUAU-------AUGAAAUUGAAUGAGUUC-CUAAAAU
Scorpion       AAGUAAUUUAAAAUUAAGAAG-------UUUAUUUGAAUAAUAAU-UUAAUG
Brine Shrimp   AAGUAAGGUCUAUCA-------------ACCUGAAGGAGGCU-CUGAGAAAUGU
Onychophoran   AAGUAAAAUuuUAauAAAG-------UaUUAUUUGAAuAaca----UuaaAU
Snail          AAGUAAUUAAUUAAAAACAAA------UAUUAAAUGAACAAGGA----AAAG
Earthworm      UUGUAACGUUUCUUCAA---------AACUAAAGUGAAUAUGAAU-CUAAG

MOTIFS            GUAA                        UgAr     RRyy  yurRrRyRcGY

Nematode       CAGUAAAAAAAUUCUUAAUGU---------AUUUUUGAAGAUUAU--CUAGAAGUGGU
Sea Anemone    UAGUAAAUCGGAAAGUAGGU-----CGUUCCGGUUGAAU-UGGUA-CAAGUGUUAGCACA
```

sophila), *Apis mellifera* (honeybee), *Locusta migratoria* (locust), *Ischnura cervula* (damselfly), *Cormocephalus amantiipes* (centipede), *Liocheles waigiensis* (scorpion), *Artemia franciscana* (brine shrimp), *Euperipatoides leuckartii* "Big Badja" (onychophoran), *Cellana tramoserica* (snail), *Aporrectodea rosea* (earthworm), *Caenorhabditis elegans* (nematode), and *Metridium senile* (sea anemone).

The nematode and sea anemone sequences are included but were not used for determination of motif sequences, although their sequences do conform to many of the motifs. The additional helix in the sea anemone is also indicated. An alignment of a larger set of sequences is available via anonymous FTP (see text), from EMBL, or from R.E.H.

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1055489 by guest on 20 August 2022

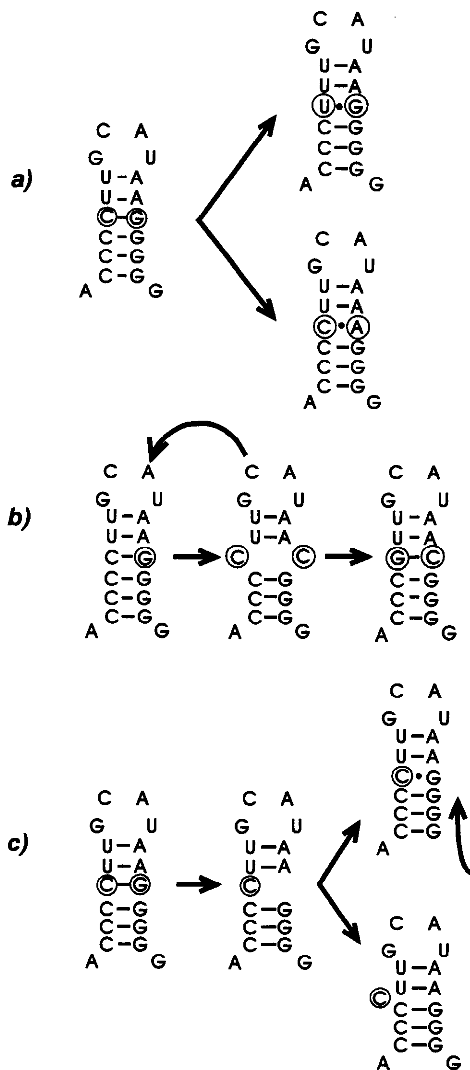


FIG. 3.—Potential consequences of a nucleotide substitution in a helix. Bases that could or do change are circled. (*a*) Nonconventional pair-bonds (indicated by "·") may form. (*b*) A bulge can be maintained or be subsequently replaced by a second substitution. (*c*) A deletion can result in a nonsymmetrical bulge, which may be preserved or be compensated for by slippage of one arm of the helix.

1998), and has also been submitted to EMBL. We focused on groups for which sequences from several relatively closely related taxa were available so that we could infer patterns of intragroup variation and compensatory substitutions. Sites of variability within specific groups are indicated by lowercase letters in fig. 2. The major features to note in the alignment are that:

1. It incorporates a wide range of vertebrate and invertebrate taxa including the phyla Chordata, Echinodermata, Arthropoda, Mollusca, Annelida, Nematoda, and Cnidaria. The largest phylum, Arthropoda, has representatives from four classes. The largest arthropod class, Insecta, includes representatives of basal wingless orders (Thysanura), early winged orders (Odonata, Orthoptera, and Hemiptera), and the more recent winged orders (Diptera and Hymenoptera). Chordate taxa include representatives from five vertebrate classes. Taxonomic designations are given in table 1 and a list of taxa is available from REH or as the file "TaxaList.12S" via anonymous FTP 130.123.1.3 from /pub/farside.

2. Base composition for domain III varies from 70 to 84% A+U in most of the invertebrates, whereas it is about 54% in the vertebrates. The sea anemone, earthworm, and sea urchin, however, have A+U values similar to the vertebrates. Despite their high A+U content (80%) the onychophora and honeybee sequences fit the secondary structure model well (the spider and nematode, with A+U contents of 84% and 76%, respectively, fit less well).

3. Relatively small insertions or deletions (indels), usually less than five nucleotides in length, occur in the first half of domain III (up to helix 38) and several well-conserved motifs in this region permit relatively unambiguous alignment of sequences (fig. 2). Larger and more numerous indels occur in the vicinity of helices 42, 45, 47, and 48, so alignment of parts of these regions is difficult when comparisons are made between distantly related taxa.

The conservation of the primary sequence and the secondary structure of domain III is remarkable given the range of taxa examined (which reflects more than 600 million years of separation). In only a few instances were the more variable helices hard to define; helix 42 in the brineshrimp, for example. However we have identified highly conserved motifs (fig. 2, and see below) that greatly facilitate alignment and identification of helices.

Conserved Sequence Motifs

The most important aspect of our alignment is the identification of sequence motifs that are conserved in all, or nearly all, of the taxa examined. Key motifs are indicated in figure 2 and a template showing consensus motifs is given in Appendix 1 (the nematode and sea anemone sequences appear to have some structural differences from the other taxa and so have not been included in the identifications of motifs, although they do match many of them). In the following discussion, lowercase letters identify variable positions. For example, lowercase "g" is used to imply generally (i.e., in at least 75% of the sequences) a guanine, but always a purine; "R" means that there is always a purine at the site, but that one specific base does not occur in more than 75% of the sequences; lowercase "r" identifies sites where a purine predominates but there may be pyrimidines (uracil or cytosine) at this position in some (< 25%) taxa. All the motifs are described reading in the 5′ to 3′ di-

Downloaded from https://academic.oup.com/mbe/article/13/7/1055489 by guest on 20 August 2022

rection. In some cases after comparison with the *E. coli* structure (Gutell 1994) if only one sequence deviated from the otherwise conserved motif we suspected this difference could have been a sequencing error and excluded the deviation from the motif assignment (for example, in the UACAnU motif between 34′ and 45; see below).

Motifs occur in both paired and unpaired regions of the molecule, with the more generally conserved motifs being found in bulges within helices (for example, the "CaA" motif in 38′), or in unpaired regions between helices ("helical junctions"; for example, the "GG" motif between helices 33 and 34; the invariant "A" preceding helix segments 36, 39, 38′, and 33′, and the "UACA" motif in the junction between helices 34′ and 45). Conserved motifs have been found to be common at helical junctions in other rRNA molecules (Huber et al. 1993; Varani and Pardi 1994).

The Model

As indicated in fig. 2, the central helices (numbers 32, 33, 34, 36, and 38) align very well, while the relative positions of some peripheral helices can vary. The secondary structure models for eight taxa are shown in fig. 4, with the more conserved sequence motifs identified with boxes or circles in the cow sequence. For the more well-conserved helices additional pair bonds could occur in some of the sequences, for instance at the end of helix 38 in the cow, but we have only indicated these if there are compensatory substitutions and/or the extension occurs within most members of a group (in most vertebrates, for example).

While the MFOLD algorithm often correctly identified peripheral helices (35, 45, 47, and 48, for example) it did not accurately fold the structural backbone of domain III (helices 33, 34, 36, and 38). Our results indicate that energetic calculations may best be suited for examination of short-range folding patterns rather than used to determine overall or long-range conformations. It is desirable to use folding algorithms in conjunction with other information (see also Zuker, Jaeger, and Turner 1991; Gutell, Larsen, and Woese 1994).

Several alternative models have been presented for the structure of domain III for animal mitochondrial SSU rRNAs and differ in their representations of helices 36–48 (Glotz and Brimacombe 1980; Clary and Wolstenholme 1985; Gutell et al. 1985; Hixson and Brown 1986; Dams et al. 1988; Simon et al. 1990; Hickson 1993; Gutell 1994; Pont-Kingdon et al. 1994; van de Peer et al. 1994). Examination of mitochondrial SSU rRNA sequences from a diverse range of taxa has, however, identified a common structure, and evidence supporting each of these helices is discussed below.

Helix 36

Helix 36 is usually 7 bp long and, like several of the helices, is preceded by a conserved adenine (figs. 2 and 4; Appendix 1). Compensatory changes in both halves of the helix occur in bovids, deer mice, great apes, birds, skinks, lungfish, damselflies, and dragonflies, though not all substitutions in this helix maintain purine–pyrimidine base pairing.

Compensatory substitutions are clearly illustrated by examination of changes in this helix among *Peromyscus* and *Onychomys* mice (fig. 5). *Neotoma* and *Sigmodon* species were used as paraphyletic outgroups to determine basal helix "I." In the *Onychomys* lineage, the C·A bond in helix "I" is converted to different Watson–Crick bonds in *O. leucogaster* (helix "III") versus *O. torridus* and *O. arenicola* (helix "II") by transitions at sites 566 and 675. In the *Peromyscus* lineage, the basal helix "I" changes to helix "IV" via two compensatory changes at sites 564/677 and 565/676, respectively. The "IV" condition is present in *P. eremicus, P. leucopus,* and *P. gossypinus.* The derived *Peromyscus* condition (helix "V"), present in *P. melanotis, P. polionotus,* and *P. keeni,* arises via compensatory substitutions at sites 564/677. This evolutionary picture is simplified by the absence of intermediate conditions present in unsampled taxa.

"Helix 37"

Earlier compilations (e.g., Neefs et al. 1990) included a helix between 36 and 38. It is possible to draw a 3-bp helix in this region for some avian taxa but few other taxa had bases able to pair in this region. Consequently, we have not included this helix in the model. The unpaired region between helices 36 and 38 is variable in length and sequence among taxa (figs. 2 and 4).

Helix 38

Nucleotides involved in this helix are some of the most conserved in domain III (fig. 2, Appendix 1). The helix can, however, be drawn in several ways (fig. 6). While there are few compensatory or covarying nucleotides to provide evidence (Gutell, Larsen, and Woese 1994) supporting one form of helix 38 over another, the mitochondrial sequence data suggest a common structure. Van de Peer et al. (1994) and Gutell, Larsen, and Woese (1994) present similar structures for helix 38, differing primarily in which of the three consecutive "uR" couplets in the proximal arm is unpaired (fig. 6*b,* 6*c*). In the mitochondrial data set (fig. 2), the second "uR" couplet is variable, while the third is a conserved "UA" and so we have chosen to pair the third "UA" couplet

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1025399 by guest on 20 August 2022

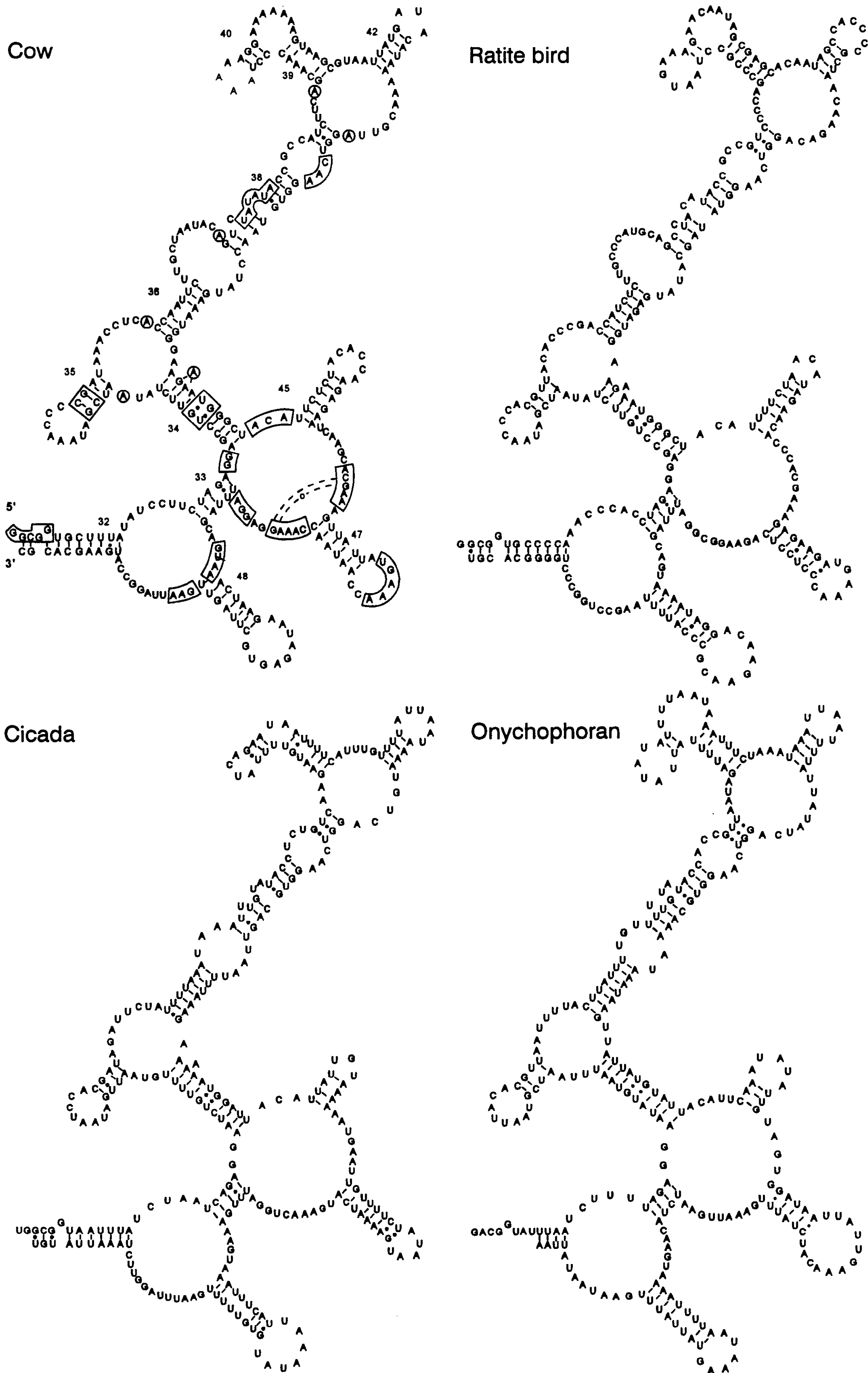
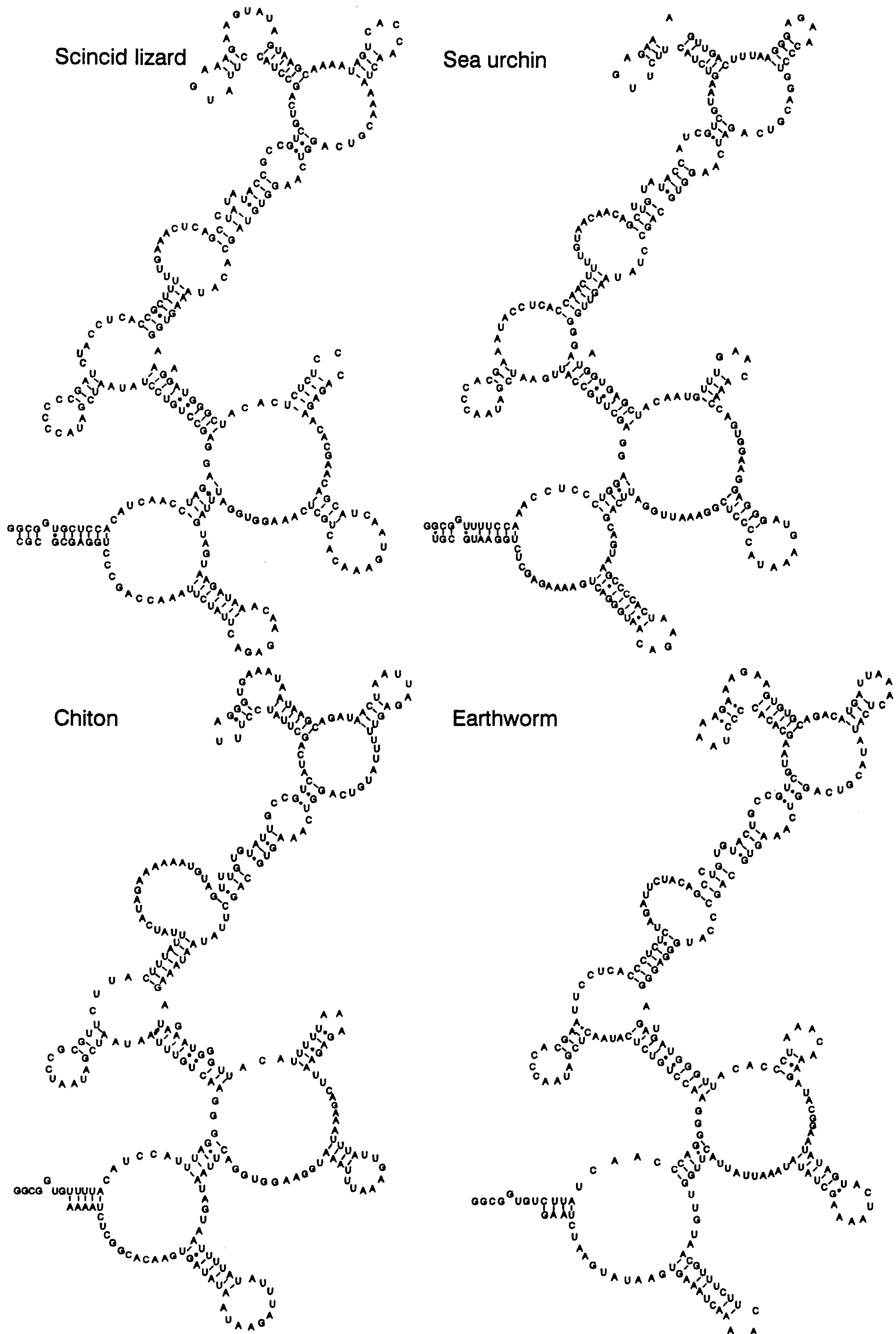


FIG. 4.—The secondary structure model of domain III for eight animal mitochondrial SSU rRNA sequences. Pairings between guanine and uracil, or adenine and cytosine, are indicated by a dot ("·"). Conserved motifs are shown boxed in the cow structure, and conserved adenine residues before helices 35, 36, 38, 39, and 38′ are circled. The bulged out adenine in helix 34 of cow is also circled to draw attention to the

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1055489 by guest on 20 August 2022



Downloaded from https://academic.oup.com/mbe/article/13/1/150/1055489 by guest on 20 August 2022

fact that a base (often adenine) is usually unpaired at this position. Appendix 1 presents a generalized structure that is very useful as a template for drawing mitochondrial SSU secondary structures.

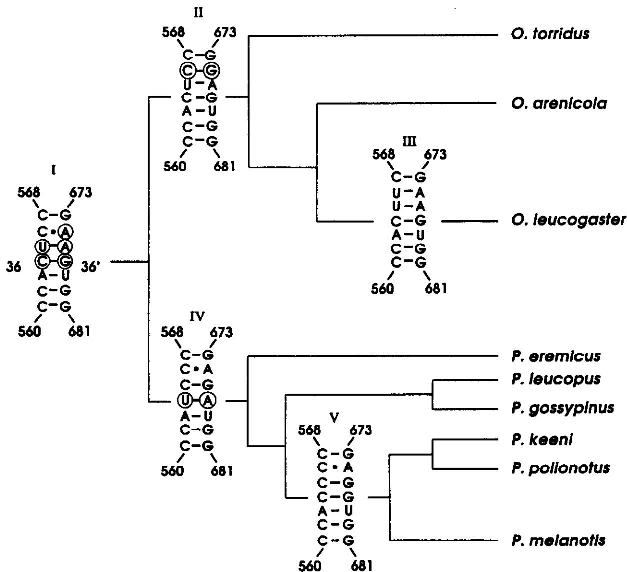


FIG. 5.—Evolution and compensatory mutations in helix 36 in *Peromyscus* and *Onychomys*. Circled bases change in the subsequent branch. Ancestral states were inferred using MacClade (Maddison and Maddison 1992) and accelerated transformations used to optimize changes on the well-corroborated relationships among these taxa using *Neotoma* and *Sigmodon* sequences as paraphyletic outgroups (helix I; Sullivan, Holsinger, and Simon, in press). The relationships are supported by phylogenetic analyses of mitochondrial cytochrome *b* sequences and the nodes are strongly supported, except for resolution of the *Onychomys* species (Sullivan et al., submitted). The compensatory changes creating helices II–V are discussed in the text.

with the distal "UG," leaving the "uR" unpaired, as in van de Peer et al.'s (1994) model (see figs. 2 and 6*d*, 6*f*, and 6*h*). Our representation for helix 38 differs from van de Peer et al. and Gutell's by postulating a symmetrical rather than an asymmetrical three-base bulge (compare fig. 6*c* and 6*d*). A uracil rather than a purine at the first position of the "rcc" motif (the ninth base pair in Gutell's structure; fig. 6*c*) in cicada and locust, as well as in other orthoptera (P. Flook, personal communication), all 24 odonate sequences, and in the frog *Rana catesbeiana,* results in a highly unusual four-base bulge if paired according to Gutell's model (see fig. 6*g*). Note that the optimal energetic folding of this helix (fig. 6*a*) has an asymmetric bulge but is otherwise different from the other structures. MFOLD produced different optimal energetic structures for different taxa.

Helix 39

Van de Peer et al. (1994) show helix 39 as a 2-bp structure, but it most commonly contains six nucleotide pairs in mitochondrial sequences (figs. 2 and 4, and Gutell [1994]). The helix is usually preceded by four unpaired bases, with an adenine always present immediately before the start of helix 39 (fig. 2, Appendix 1). Compensatory changes in the avian (table 2) and odonate (table 3) data sets provide support for the structure of this helix as we have drawn it. The third and/or fourth positions may or may not be be paired (fig. 2, tables 2 and 3), and among avian orders A·C bonds may be relatively frequent (table 2). The last pair bond of this helix is often well conserved (fig. 2, tables 2 and 3).

Helix 40

This helix is included in both Gutell's (1994) and van de Peer et al. (1994) compilations, and there are covarying substitutions strongly supporting its existence in mitochondrial SSU rRNA (tables 2 and 3, and McIntosh and Simon, unpublished). Helix 40 is only 2 bp long in ratite birds but there are three bonds in other avian orders (table 2). A·C bonds appear to be involved in helix 40 in amphibia and the earthworm (see figs. 2 and 4). The nematode *Caenorhabditis elegans* has 4 bp in this helix (fig. 2).

"Helix 41"

Helix 41, a 2-bp structure in the van de Peer et al. (1994) carp structure, is not supported by our analyses. Its suggested position is in a region with high nucleotide substitution events (including indels; fig. 2, and Hickson 1993), and the distal part of this helix falls within our elongated helix 39, making the formation of helix 41 energetically unfavorable.

Helix 42

A 2–4-bp helix can be drawn for all of the taxa in fig. 2, often preceded by an adenine residue and usually four bases from the end of helix 39. Pairings between adenine and cytosine are required to construct this helix in birds (table 2), as well as in *Mus* and *Rattus* (data not shown). In 12 of 24 odonate sequences examined helix 42 consists of 3 bp, while there are 4 bp in the other 12 taxa and various compensatory mutations occur (table 3). An energetically more favorable alternative helix 42 can be constructed closer to the 38′ helix for the primates (Hickson 1993) but not for other taxa, suggesting that it does not exist at the alternative location. The loop of helix 42 is the most variable region of domain III in birds, with indels of up to 10 bases (Cooper, unpublished).

Helix 45

Helix 45 follows the "UACAnU" motif (figs. 2 and 4), and is 4–7 bp in length (fig. 2, tables 2 and 3). The last "U" in the motif is a "C" in the earthworm, but this may be a sequencing error. The honeybee is

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1034469 by guest on 21 August 2022

a. Cow helix 38; optimal energetic solution
(-7.9 kcal/mol)

```
38-GUCUAUAUACC    GCCAUCUU
        ||•|||    •||
 38'-CaAUGUGGAACUGGAUUGCA
```

b. Cow helix 38; van de Peer et al's structure
(8.6 kcal/mol)

```
GUCUAUAUACC    GCC AUCUU
|| ||   •|||   •|| | |
Ca AU   GUGGAACUGGAUUGCA
```

c. Cow helix 38; Gutell's structure
(10.0 kcal/mol)

```
                 11 1 1 1
12 3456   78     901 2 3 4
GUCUAUAUACC      GCC AUCUU
|| ||•|   ||     •|| | | |
Ca AUGU   GGAACUGGAUUGCA
```

d. Cow helix 38; our structure
(5.3 kcal/mol)

```
12 34     78
GUCUAUAUACCGCCAUCUU
|| ||   •|||   |•|
Ca AU   GUGGAACUGGAUUGCA
```

e. Sea Urchin helix 38; following Gutell
(-3.2 kcal/mol)

```
                 11 1 1 1
12 3456   78     901 2 3 4
GCUUGUAUACC      AUC GUCGu
|| ||•|   ||     ||| | |||
CG ACGU   GGAACUAGACUGCA
```

f. Sea Urchin helix 38; our structure
(-0.6 kcal/mol)

```
12 34     78
GCUUGUAUACCAUCGUCGu
|| ||   •||| : •||
CG AC   GUGGAACUAGACUGCA
```

g. Periodical Cicada helix 38; following Gutell
(9.3 kcal/mol)

```
                 11 1 1
12 3 56   78     01 2 3
aUUUGUAUACC      UCU GUCAA
|• ||•|   ||     |• | ||
uG ACGU   GGAACUGGACUGUA
```

h. Periodical Cicada helix 38; our structure
(6.4 kcal/mol)

```
12 3      78
aUUUGUAUACCUCUGUCAA
|• ||   •|||:   ••|
uG AC   GUGGAACUGGACUGUA
```

FIG. 6.—Potential pairings for helix 38 in 12S rRNA domain III. Lowercase letters indicate variable sites in that taxonomic group. Nonconventional pair bonds are indicated by "·". (*a*) The structure in the cow favored on the basis of free energy calculations (determined using the MFOLD algorithm in the GCG package). Minimal free energy calculations are also shown for the other structures. (*b*) The cow structure following van de Peer et al's (1994) model. (*c*) The structure for the cow proposed by Gutell (1994), and its minimum free energy. Each base pair in the Gutell structure is numbered to facilitate comparison with the other structures. Van de Peer et al's and Gutell's structures are similar for this helix, differing primarily in which "UA" couplet is unpaired. (*d*) Our model for helix 38 in the cow based on comparative analysis of a large range of vertebrate and invertebrate mitochondrial sequences. Note the symmetrical "GCC/CAA" bulge. (*e*) The structure of helix 38 in the sea urchin, based on Gutell's (1994) model. (*f*) The sea urchin helix as represented in our model. Additional pair bonds that are not a general feature of our model are indicated by ":". (*g*). The structure of helix 38 in periodical cicada, based on Gutell's (1994) model. Note the four-base bulge. (*h*) The periodical cicada helix as represented in our model. Note that with our model, the symmetrical bulge associated with the "CAA" is a more general pattern (see fig. 2), and that it generally maintains the highly conserved ("dominant") G·U base pair adjacent to the "Ryy/CAA" bulge (see text).

highly unusual in having an extra "U" in the motif (fig. 2), which cannot be ascribed to a gel reading error (R. Crozier, personal communication).

The hairpin loop of helix 45 is usually six bases or fewer. Relatively large insertions can occur downstream of helix 45. The fin whale and the tinamou (a phylogenetic relative of ratite birds) have insertions of at least 10 bases in this region (Hickson 1993; Cooper unpublished). The sea anemone is unique in having an additional helix extended on from helix 45 (see fig. 2), making it much more similar to the *E. coli* structure (Pont-Kingdon et al. 1994; see fig. 1).

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1055489 by guest on 20 August 2022

and Appendix 1). The "Y" and "g" of this motif are suggested by Gutell, Larsen, and Woese (1994) to bond to the "r" and adjacent "A," respectively, of the "yRaarr" motif between helices 47 and 33. This is most commonly manifested as a G–C bond followed by an A·G pair (shown by dashed lines in the cow structure in fig. 4), but this cannot be drawn in some taxa, onychophora for instance.

Compensatory substitutions supporting the location of helix 47 occur in most of the groups for which we have several closely related taxa (see, for example, tables 2 and 3). Note that 13% of pairings for helix 47 may be A·C bonds in the bird sequences (table 2), while very few A·C pairings can be inferred among the odonates (table 3).

A "yRaarr" motif occurs between 47′ and 33′ in the mitochondrial sequences (fig. 2), and there can also be a similar "yrarr" motif upstream of this, either in the loop of 47 or associated with the distal arm of the helix (fig. 2).

Helix 48

Helix 48 is also variable in size and position but occurs in all taxa between the two well-conserved helices 32 and 33, and motifs are associated with the beginning and end of this helix: "GUAA" ("yAGUAA" for most vertebrates, but "yuGUAA" in some frogs), and "UgAr," respectively (fig. 2). Compensatory substitutions in, for instance, birds (table 2) and odonates (table 3) also provide support for its location.

Domain III in Molluscs

The 12S rRNA sequence from the mussel *Mytilus edulis* (Hoffmann, Boore, and Brown 1992) was difficult to align to the other sequences because of large (up to 100 bases) insertions between 38 and 38′, and between 45′ and 47. Outside of the insertions, however, the conserved motifs were identifiable (data not shown) and the template (Appendix 1) was very useful for fitting *Mytilus* sequence onto the general structure. Other molluscan sequences, such as the snail *Cellana tramoserica* (fig. 2), the chiton (fig. 4), and scallop (GenBank accession number X67246; not shown) do not have such large insertions.

Discussion

Although rRNA genes can be difficult to align because they lack the triplet code pattern of protein genes (Mindell 1991), we find, along with Gutell (1994), van de Peer et al. (1994), and Kjer (in press), that secondary structure features of rRNA sequences can greatly facilitate alignment. Comparative analysis of rRNA sequences is the major method for refining the structural models (Gutell 1994; Gutell, Larsen, and Woese 1994; van de Peer et al. 1994), and identification of compensatory changes provides the strongest evidence for bases involved in pairings (Larsen 1992; Gutell, Larsen, and Woese 1994). Energetic calculations, while useful, can be less accurate than comparative sequence analysis for predicting the structure of some helices (see fig. 6, and also Gutell, Larsen, and Woese [1994]).

The conserved sequence motifs (Appendix 1), support from compensatory mutations (e.g., tables 2 and 3), and the ability to fit many diverse taxa onto the structure (fig. 4) suggest that our model has generality for animal mitochondrial SSU rRNA. In contrast to some previous models (Clary and Wolstenholme 1985, 1987; Hixson and Brown 1986; Dams et al. 1988; Simon et al. 1990), our analysis of a large number of mitochondrial sequences indicates a common structure for domain III. Utilization of a large number of taxa provides a refined definition of regions of intra- and intergroup sequence conservation and variation.

Differences between Models

Our 12S third domain model is very similar to those of Gutell (1994) and van de Peer et al. (1994). However, an important difference between our model and those of Gutell (1994) and van de Peer et al. (1994) is that in ours the "CaA" bulge in 38′ is symmetrically opposite three unpaired bases, "rcc" (fig. 4, Appendix 1). Gutell (1994) and van de Peer et al. (1994) have the "rcc" paired with a "ggU" motif adjacent to the "CaA" bulge (fig. 6*b* and 6*c*). Free energy calculations for these alternative structures do not favor one common structure (compare fig. 6*e* and 6*g*, and 6*f* and 6*h*). The structural models derived by Gutell (1994) and van de Peer et al. (1994) are based on comparative information from a very large and diverse group of prokaryote and eukaryote SSU rRNAs, while our alternative structure is based on comparison of just animal mitochondrial sequences. Examination of relatively closely related sequences is, however, an important aspect for refining secondary structure features and we have examined a much greater range of mitochondrial sequences than previous studies. In so doing we found that cicadas and the odonata cannot form the ninth base pair in the Gutell helix (see fig. 6*g*), and that cicadas, *Tetragnatha* spiders, and the brine shrimp would lack two bases from the end of helix 38 in the Gutell structure (see figs. 2 and 4).

Although this is not particularly strong evidence for a symmetrical bulge in helix 38, the inclusion of such a bulge results in a structure consistent across all of the mitochondrial sequences we examined, with the exception of the nematode and sea anemone (fig. 2). The symmetrical bulge maintains (except in mammals) the ad-

Downloaded from https://academic.oup.com/mbe/article/12/1/158/... by guest on ... August 2022

jacent G·U base pair that is highly conserved in SSU rRNAs (Appendix 1; Gutell, Larsen, and Woese 1994). Sea urchins (fig. 6*f*) and cicadas (fig. 6*h*), as well as dragonflies and *Tetragnatha* spiders, could have additional, nonconventional base pairings in this bulge, but adjacent pairs of C·A, C·C, or C·U bonds can destabilize a helix (SantaLucia, Kierzek, and Turner 1991) and so may not form. Similar comparative studies of other groups of taxa will help determine if a symmetrical bulge in helix 38 is a valid alternative structure for animal mitochondrial SSU rRNA. An asymmetrical bulge is favored in fungal and plant mitochondrial SSU rRNAs (see Gutell 1994). The secondary structure of helix 38 does not, however, affect the alignment of the sequences.

G·U Bonds

In domain III of SSU rRNA Gutell, Larsen, and Woese (1994) identified seven instances of G·U bonds that were "invariant" (occurred in 100% of SSU sequences examined) and five instances of "dominant" G·U bonds (in >55% of SSU sequences). In our mitochondrial secondary structural model, however, the same constraints are not observed, which may imply different selective forces operate on the mitochondrial rRNA. We can locate five of the seven invariant G·U pairs, although an examination of the 184 mitochondrial sequences reveals that all of these must be downgraded to "dominant." In addition, we have found two more dominant G·U sites near other G·U pairs; one in the middle of helix 34, and one near the distal end of helix 38. These seven G·U bonds are highlighted with boxes in Appendix 1. None of Gutell et al.'s "dominant" G·U pairs are dominant in our structure, but mitochondrial sequences are reduced in these areas compared to eukaryotic nuclear and prokaryotic sequences (see fig. 1).

Conserved Motifs

While the nematode *C. elegans* (as well as *Ascaris suum*), the sea anemone (*Metridium senile*), and the mussel (*Mytilus edulis)* differ in some aspects from our general model, they do match many of the motifs and structural features indicated in fig. 2 and Appendix 1. The mitochondrial SSU rRNA secondary structure of the alga *Chlamydomonas reinhardtii* is distinct from those of other plants (Gutell 1994), so some deviation from a general structure can be anticipated. Examination of sequences from near phylogenetic relatives of these taxa would be useful to determine how widespread such structural changes are. Preliminary sequence data for cestodes (P. Olson and C. Simon, unpublished) indicate that they have motif differences from molluscs, arthropods, and chordates.

Although there is still uncertainty in aligning some regions due to insertion or deletion events (around helices 42, 47, and 48 in particular), the conserved motifs (fig. 2 and Appendix 1) serve as a framework upon which alignment of 12S domain III sequences may be more reliably built. We have focused upon domain III because this region is the most frequently sequenced region of animal mitochondrial SSU rRNA. Alignment and refinement of the structure required a considerable amount of work, so aligning and refining the more variable domains I and II will be a much more difficult task. The secondary structure models of Gutell (1994) and van de Peer et al. (1994) are, however, valuable frameworks for aligning these regions. Similar approaches can be adopted for the mitochondrial large subunit (16S) rRNA, and the most recent general secondary structure models are given by Gutell, Gray, and Schnare (1993) and de Rijk et al. (1994). It is easiest to start with several relatively closely related sequences to first establish an alignment and provide evidence for compensatory changes. Increasingly divergent sequences can then be included to identify the more conserved motifs. Refinement of the alignment assists with refinement of a secondary structure model, and vice versa, so it is an interactive process.

Paired and Unpaired Regions

Helices and unpaired regions are relatively easy to draw in any piece of DNA sequence, especially when nucleotide bias is high and noncanonical base pairing is allowed, e.g., in A+U-rich insect mtDNA (Simon et al. 1990; Taylor et al. 1993). Simon et al. (1990) found that *Homo sapiens, Drosophila yakuba,* and *Magicicada tredecim* could roughly fit both the Glotz and Brimacombe (1980) and Dams et al. (1988) 12S rRNA models (but with so few sequences available, no motifs or compensatory mutations were identifiable). Rather than being a conflict in structure, Simon et al. suggested that there may be local switching between alternative structures, as has been documented for other regions of nuclear rRNAs. However, our analyses of a wider range of 12S rRNA sequences indicate that the Glotz and Brimacombe (1980) and Dams et al. (1988) structures are incorrect, reflecting instead the ease of drawing helices and loops. On the basis of comparative sequence analysis and the identification of compensatory changes, one structure common to all the vertebrates examined, as well as most of the invertebrates, is favored (figs. 2 and 4). As Simon et al. (1990) noted though, and the present analyses reinforce, experimental evidence for the structure of 12S rRNA would be useful (see also Gutell, Larsen, and Woese 1994).

Hixson and Brown (1986) suggested that "short-range" helices (where the two arms of the helix are

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1025889 by guest on 20 August 2022

close together in the primary structure of the DNA, e.g., helix 35) may be less well conserved than "long-range" helices (where the proximal and distal arms are separated by other helices). Simon (1991) pointed out that this is generally true but that tertiary structure and protein interactions can result in exceptions. Our analyses support Simon's view; the short-range helix 35, for example, is more conserved than expected, whereas the long-range helix 36 is more variable (see fig. 2). The nucleotide variability associated with helices 36, 45, 47, and 48, and the unpaired region between 40′ and 39′ (see fig. 2) may be of phylogenetic value for investigations of relatively closely related taxa if there are sufficient variable sites and multiple hits are not a problem (Sullivan, Holsinger, and Simon, in press; Simon et al., in preparation).

Vawter and Brown (1993) examined differences in substitution patterns between different rRNA structural classes but did not compare short-range versus long-range helices. We have shown that within helices and unpaired regions there can be extensive variability in the degree of nucleotide conservation (compare, for example, helices 36 and 38 in fig. 2), and that some regions of domain III appear to differ in their degree of variability between taxa. van de Peer et al. (1993) elegantly demonstrated this point and emphasized that closer analyses of sequence variation make labels such as "conserved" and "variable" sites too simplistic. Phylogenetic weighting of sequences using criteria such as "stem" and "loop" may be inappropriate (Hickson 1993; Simon et al. 1994). Nucleotides in unpaired regions can be sites for protein, tRNA or rRNA interactions (see, for example, Noller et al. 1990; von Ahsen and Noller 1995), and unpaired bases can be very well conserved (Huber et al. 1993; Gutell, Larsen, and Woese 1994; and this study).

Kraus et al. (1992) and Gatesy et al. (1994) noted that compensatory changes in rRNA may not occur for substantial periods of time and our analyses support this. Some bulges or nonconventional base-pairings may have a long evolutionary history (for example, the "A : A" bulge in helix 39 in mammals). The potential destabilizing effects of bulges may be lessened by protein binding or by nonconventional base pairing (for example the "CaA" bulge in helix 38′; Gutell et al. 1985; SantaLucia, Kierzek, and Turner 1991). Slippage of a helix arm rather than compensatory substitutions could also occur (fig. 3*c*) but appears to be rare in our data. Slippage is in essence a covarion shift and would be a serious problem for phylogenetic analysis because helix position would not correspond to nucleotide homology (Miyamoto and Fitch 1995).

Alignments and Phylogeny

Secondary structure information can be used to indicate the degree of reliability or confidence in the alignment of different parts of the sequence. The region between helices 32 and 38, for instance, can be aligned with confidence across a wide range of taxa, whereas alignment around helices 47 and 48 can be more difficult. Consequently, when highly divergent taxa are compared (as, for example, in the arthropod study by Ballard et al. [1992]) phylogenetic inferences based on alignment of domain III should be viewed with caution.

The close occurrence of similar motifs can generate misalignments when alignment programs are used. This is illustrated by the "yrarr" and "yRaar" motifs on either side of helix 47′, particularly since the first "UGAAA" is more conserved in vertebrates while the second "UGAAR" is more generally conserved among insects.

An important contribution to the evaluation of the reliability of phylogenetic analyses is the presentation of the sequence alignment, and inclusion of such data in submitted manuscripts should be encouraged (see also Kjer 1995). Kjer (1995) has identified examples in an amphibian rRNA data set where alignment of sequences based on secondary structure information results in phylogenies more congruent with nonmolecular data than phylogenies derived from rRNA sequences not aligned with respect to a structural model. In our view, deposition of the aligned sequences in a database or FTP site should be a requirement for manuscript submission. The availability of our alignment is given in the Results.

Comparison of 12S domain III alignments derived from a range of automatic alignment programs with the alignment based on the domain III motifs and secondary structure model will be presented elsewhere (Hickson and Simon, in preparation).

Accuracy of Sequences

Sequencing errors are a recognized problem in the databases (see for example Clark and Whittam 1992; States 1992; G. Olsen, personal communication). Errors can be detected by simple consistency checks, such as the secondary structure alignment illustrated here. There are several cases where one or a few taxa differ from the rest at positions that are otherwise constant. These may be real differences or the artifacts of cloning, polymerase chain reaction (PCR), sequencing, or data entry. The motifs (fig. 2 and Appendix 1) are informative in this respect and deviations from some of the motifs provide a valuable feedback to check sequencing precision.

As an illustration, the two complete sea urchin 12S rRNA sequences in GenBank (*Strongylocentrotus purpuratus*; accession number X12631, Jacobs et al. [1988]; *Paracentrotus lividus*; J04815; Cantatore et al. [1989])

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1025589 by guest on 20 August 2022

both lack one of the highly conserved guanines in the bulge between helices 33 and 34. Four other urchin species sequences (generated by the PCR rather than cloning) had this double guanine bulge, as did another individual of *S. purpuartus* (Thomas, Maa, and Wilson 1989). Given the conservation of this unpaired "GG" doublet in animal 12S rRNA and other SSU rRNAs (see van de Peer et al. 1994), and its presence in the Thomas, Maa, and Wilson (1989) urchin sequences, we included it in the *P. lividus* sequence shown in figure 2. Thomas, Maa, and Wilson also noted that their *S. purpuratus* sequence differed at several positions from the Jacobs et al. sequence, so reconfirmation of the cloned sea urchin sequences is recommended.

The honeybee mitochondrial genome is very A+U rich and appears to evolve rapidly in relation to other insect taxa (Crozier and Crozier 1993). Although its 12S sequence fits the general model very well, there are several other differences between the honeybee sequence and the motifs we identified. For example, the bee (as well as earthworm) do not match the "GGA" motif before helix 33′ (fig. 2). The first "G" in this motif is very well conserved over a very broad range of taxonomic groups and is involved with tRNA binding in *E. coli* (von Ahsen and Noller 1995) and so may have a similar role for other SSU rRNAs. Rechecking of the original sequencing gels did not indicate inaccuracy in gel reading or data entry (R. H. Crozier, personal communication), but (as with the sea urchins) cloning artifacts cannot be excluded. Sequence data from additional individuals and related taxa is required to determine whether these ambiguities are a feature common to other hymenoptera.

Deviations from the motifs (fig. 2) may, however, be real in some cases and sequences from several individuals or from relatively closely related taxa are useful for resolving this. An example is the "CAA" motif in helix 38′. While the majority of invertebrate and vertebrate taxa have "CAA" (including articulate brachiopods [S. Stark and B. L. Cohen, personal communication] and the amphioxus [*Branchiostoma*; L. Daehler, personal communication]), it is "CGA" in fish (including nine genera of sharks [A. Martin, personal communication] and lamprey).

Spiders (10 species), nematodes, and sea anemone are unusual in not having a cytosine at the first position of this bulge. Fungal and plant mitochondrial SSU rRNA sequences, as well as many other prokaryotic and eukaryotic SSU rRNAs, have cytosine at this position (see Gutell 1994).

Nuclear copies of fragments of the mitochondrial genome, including 12S rRNA sequences, are now being found (for example, Lopez et al. 1994) so greater care must be taken to authenticate mitochondrial sequences. Deviations from secondary structural features or well-conserved sequence motifs can have utility for identifying potential nuclear copies of mitochondrial genes. A relatively long period of time may, however, be necessary for such changes to occur; the cat nuclear 12S fragment (Lopez et al. 1994) cannot be distinguished from the mitochondrial sequence on the basis of the motifs presented here (R.E.H., unpublished observations).

Concluding Remarks

In this paper we have advocated the use of secondary structure information as a framework for aligning and analyzing rRNA sequences with increased reliability. Visual inspection has always been an important part of data analysis. The increasingly rapid generation of DNA sequence data and the development of more sophisticated software for DNA sequence analysis should not detract from looking at the data before and during analyses (see also Lento et al. 1995).

Acknowledgments

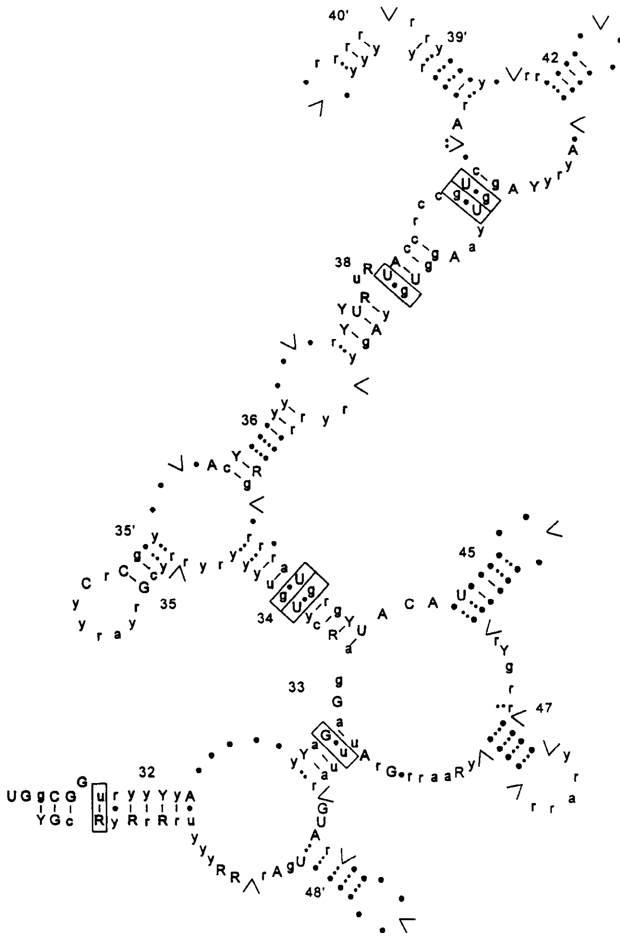
We are indebted to Mary Jane Spring for producing the secondary structure illustrations. Gina Lento generously allowed us to use her pinniped sequence data prior to publication. Some of the 12S sequence data for the parrots was contributed by Adrian Baker and James Hellewell. Ross Crozier indulged us by rechecking the honeybee sequence data. Bernard Cohen, John Gatesy, Robin Gutell, Des Higgins, Karl Kjer, Andrew Martin, Marcie McClure, Jean-Marc Neefs, Gary Olsen, Yves van de Peer, Graham Wallis, and Michael Zuker provided useful discussions, advice, and/or unpublished manuscripts. Funding for this study was provided in part by Massey University and the Alfred P. Sloan Foundation (to R.E.H.), by NSF (to C.S.), and in part by Victoria University of Wellington, the Museum of New Zealand Te Papa Tongarewa, and the Smithsonian Institution (to A.C.). R.E.H. thanks Charles Daugherty for access to the *Leiolopisma* specimens. C.S. thanks Carl McIntosh and Adrian Franke for laboratory assistance. A.C. thanks the many museum curators and field workers who generously provided avian samples. G.S.S. thanks Marina Vainer and Carol Spicer for laboratory assistance.

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1054999 by guest on 26 August 2022

APPENDIX 1

Secondary structure template for domain III of animal mitochondrial SSU rRNA showing conserved residues and motifs. Nonconserved bases are represented as dots (●), helix positions that are not paired in every taxon are indicated by double dots (:), and areas of insertions shown by V. Bases in capitals signify positions conserved in all the taxa examined (excluding nematode

and sea anemone), while lowercase letters identify nucleotides conserved in at least 75% of the taxa, as discussed in the text. Well-conserved ("dominant" in Gutell, Larsen, and Woese's [1994] terminology) guanine–uracil pair bonds are shown boxed. Note that the sequence and length of helices 42, 45, 47, and 48 are variable, but their locations can be identified by reference to conserved motifs as shown in the figure. To easily draw a structure, locate the conserved motifs in the sequence to be analyzed and use these as guides when you overlay the sequence on the template.



LITERATURE CITED

ABRAHAMS, J. P., M. VAN DEN BERG, E. VAN BATENBURG, and C. PLEIJ.. 1990. Prediction of RNA secondary structure, including pseudoknotting, by computer simulation. Nucleic Acids Res. **18**:3035–3044.

ANDERSON, S., A. T. BANKIER, B. G. BARRELL et al. 1981. Sequence and organization of the human mitochondrial genome. Nature **290**:457–465.

BALLARD, J. W. O., G. J. OLSEN, D. P. FAITH, W. A. ODGERS, D. M. ROWELL, and P. W. ATKINSON. 1992. Evidence from 12S ribosomal RNA sequences that onychophorans are modified arthropods. Science **258**:1345–1348.

BANDELT, H. J., and A. W. M. DRESS. 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Mol. Phylogenet. Evol. **1**:242–252.

BELL, L. H., J. R. COGGINS, and E. J. MILNER-WHITE. 1993. Mix 'n' match: an improved multiple sequence alignment procedure for distantly related proteins using secondary structure predictions, designed to be independent of the choice of gap penalty and scoring matrix. Protein Engineering **6**:683–690.

BIBB, M. J., R. A. VAN ETTEN, C. T. WRIGHT, M. W. WALBERG, and D. A. CLAYTON. 1981. Sequence and gene organization of mouse mitochondrial DNA. Cell **26**:167–180.

CANTATORE, P., M. ROBERTI, G. RAINALDI, M. N. GADALETA, and C. SACCONE. 1989. The complete nucleotide sequence, gene organization, and genetic code of the mitochondrial genome of *Paracentrotus lividus*. J. Biol. Chem. **264**: 10965–10975.

CLARK, A. G., and T. S. WHITTAM. 1992. Sequencing errors and molecular evolutionary analysis. Mol. Biol. Evol. **9**: 744–752.

CLARY, D. O., and D. R. WOLSTENHOLME. 1985. The ribosomal RNA genes of *Drosophila* mitochondrial DNA. Nucleic Acids Res. **13**:4029–4044.

———. 1987. *Drosophila* mitochondrial DNA:conserved sequences in the A+T-rich region and supporting evidence for a secondary structure model of the small ribosomal RNA. J. Mol. Evol. **25**:116–125.

COLLINS, T. M., F. KRAUS, and G. ESTABROOK. 1994. Compositional effects and weighting of nucleotide sequences for phylogenetic analysis. Syst. Biol. **43**:449–459.

COOPER, A. 1994. Molecular evolutionary studies of New Zealand birds. Unpublished Ph.D. Thesis, Victoria University of Wellington, New Zealand.

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.

DAMS, E., L. HENDRIKS, Y. VAN DE PEER, J.-M. NEEFS, G. SMITS, I. VANDENBEMPT, and R. DE WACHTER. 1988. Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res. **16**:Suppl., r87–r173.

DE RIJK, P., J.-M. NEEFS, Y. VAN DE PEER, and R. DE WACHTER. 1992. Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res. **20**:Suppl., 2075–2089.

DE RIJK, P., Y. VAN DE PEER, S. CHAPELLE, and R. DE WACHTER. 1994. Database on the structure of large ribosomal subunit RNA. Nucleic Acids Res. **22**:3495–3501.

DOOLITTLE, R. F. 1986. Of ORFs and URFs. University Science Books, Mill Valley, Calif.

DUNON-BLUTEAU, D., and G. BRUN. 1986. The secondary structures of the *Xenopus laevis* and human mitochondrial small ribosomal subunit RNA are similar. FEBS Lett. **198**: 333–337.

FELSENSTEIN, J. 1988. Phylogenies from molecular sequences: inference and reliability. Annu. Rev. Genet. **22**:521–565.

FLOOK, P. K., C. H. F. ROWELL, and G. GILLISSEN. 1995. The sequence, organization and evolution of the *Locusta migratoria* mitochondrial genome. J. Mol. Evol. (in press).

Downloaded from https://academic.oup.com/sysbio/article/... by guest on 20 August 2022

Freier, S. M., R. Kierzek, J. A. Jaeger, N. Sugimoto, M. H. Caruthers, T. Neilson, and D. H. Turner. 1986. Improved free-energy parameters for predictions of RNA duplex stability. Proc. Natl. Acad. Sci. USA **83**:9373–9377.

Gatesy, J., R. de Salle, and W. Wheeler. 1993. Alignment—ambiguous nucleotide sites and the exclusion of systematic data. Mol. Phylogenet. Evol. **2**:152–157.

Gatesy, J., C. Hayashi, R. de Salle, and E. Vrba. 1994. Rate limits for mispairing and compensatory change: the mitochondrial ribosomal DNA of antelopes. Evolution **48**: 188–196.

Glotz, C., and R. Brimacombe. 1980. An experimentally-derived model for the secondary structure of the 16S ribosomal RNA from *Escherichia coli.* Nucleic Acids Res. **8**: 2377–2395.

Gutell, R. R. 1994. Collection of small subunit (16S- and 16S-like) ribosomal RNA structures: 1994. Nucleic Acids Res. **22**:3502–3507.

Gutell, R. R., M. W. Gray, and M. N. Schnare. 1993. Compilation of large subunit (23S- & 23S-like) ribosomal RNA structures: 1993. Nucleic Acids Res. **21**:3055–3074.

Gutell, R. R., N. Larsen, and C. R. Woese. 1994. Lessons from an evolving rRNA:16S and 23S rRNA structures from a comparative perspective. Microbiol. Rev. **58**:10–26.

Gutell, R. R., B. Weiser, C. R. Woese, and H. F. Noller. 1985. Comparative anatomy of 16-S-like ribosomal RNA. Prog. Nucleic Acids Res. **32**:155–216.

Hein, J. 1990. A unified approach to alignment and phylogenies. Pp. 625–645 *in* R. F. Doolittle, ed. Methods in enzymology, vol. 183. Academic Press, San Diego.

Hickson, R. E. 1993. Evolutionary tails from the South Pacific. Being a wondrous and exciting account of investigations into the histories of New Zealand skinks, and in which it is shown that things are not what they seem. Unpublished PhD Thesis, Massey University, New Zealand.

Hillis, D. M., and M. T. Dixon. 1991. Ribosomal DNA: molecular evolution and phylogenetic inference. Q. Rev. Biol. **66**:411–453.

Hixson, J. E., and W. M. Brown. 1986. A comparison of the small ribosomal RNA genes from the mitochondrial DNA of the great apes and humans: sequence, structure, evolution, and phylogenetic implications. Mol. Biol. Evol. **3**:1–18.

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.

Huber, K. C., T. S. Haider, M. W. Müller, B. A. Huber, R. J. Schweyen, and F. G. Barth. 1993. DNA sequence data indicates the polyphyly of the family ctenidae (Araneae). J. Arachnol. **21**:194–201.

Huysmans, E., and R. de Wachter. 1985. Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res. **14**(Suppl.):r73–r118.

Irwin, D. M., T. D. Kocher, and A. C. Wilson. 1991. Evolution of the cytochrome b gene of mammals. J. Mol. Evol. **32**:128–144.

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

Kjer, K. M. 1995. Use of rRNA secondary structure in phylogenetic studies to identify homologous positions: an example of alignment and data presentation from the frogs. Mol. Phylogenet. Evol. **4**:314–330.

Kjer, K. M., G. D. Baldridge, and A. M. Fallon. 1994. Mosquito large subunit ribosomal RNA: simultaneous alignment of primary and secondary structure. Biochim. Biophys. Acta **1217**:147–155.

Kocher, T. D., W. K. Thomas, A. Meyer, S. V. Edwards, S. Pääbo, F. X. Villablanca, and A. C. Wilson. 1989. Dynamics of mitochondrial DNA evolution in animals: amplification and sequencing with conserved primers. Proc. Natl. Acad. Sci. USA **86**:6196–6200.

Kraus, F., L. Jarecki, M. M. Miyamoto, S. M. Tanhauser, and P. J. Laipis. 1992. Mispairing and compensational changes during the evolution of mitochondrial ribosomal RNA. Mol. Biol. Evol. **9**:770–774.

Lake, J. A. 1991. The order of sequence alignment can bias the selection of tree topology. Mol. Biol. Evol. **8**:378–385.

Larsen, N. 1992. Higher order interactions in 23S rRNA. Proc. Natl. Acad. Sci. USA **89**:5044–5048.

Lento, G. M., R. E. Hickson, G. K. Chambers, and D. Penny. 1995. Use of spectral analysis to test hypotheses on the origin of pinnipeds. Mol. Biol. Evol. **12**:28–52.

Lopez, J. V., N. Yuhki, R. Masuda, W. Modi, and S. J. O'Brien. 1994. Numt, a recent transfer and tandem amplification of mitochondrial DNA to the nuclear genome of the domestic cat. J. Mol. Evol. **39**:174–190.

Maddison, W. P., and D. R. Maddison. 1992. MacClade: Analysis of phylogeny and character evolution, version 3. Sinauer Associates, Inc. Sunderland, Mass.

Maidak, B. L., N. Larsen, M. J. McCaughley, R. Overbeek, G. J. Olsen, K. Fogel, J. Blandy, and C. R. Woese. 1994. The ribosomal database project. Nucleic Acids Res. **22**:3485–3487.

Mindell, D. P. 1991. Aligning DNA sequences: homology and phylogenetic weighting. Pp. 73–89 *in* M. M. Miyamoto and J. Cracraft, eds. Phylogenetic analysis of DNA sequences. Oxford University Press.

Miyamoto, M. M., and W. Fitch. 1995. Testing the covarion hypothesis of molecular evolution. Mol. Biol. Evol. **12**: 503–513.

Neefs, J.-M., Y. van de Peer, P. de Rijk, S. Chapelle, and R. de Wachter. 1993. Compilation of small ribosomal subunit RNA structures. Nucleic Acids Res. **21**:3025–3049.

Neefs, J.-M., Y. van de Peer, L. Hendriks, and R. de Wachter. 1990. Compilation of small ribosomal subunit RNA sequences. Nucleic Acids Res. **18**:Suppl., 2237–2317.

Noller, H. F., D. Moazed, S. Stern, T. Powers, P. N. Allen, J. M. Robertson, B. Weiser, and K. Triman. 1990. Structure of rRNA and its functional interactions in translation. Pp. 73–92 *in* W. E. Hill, A. Dahlberg, R. A. Garrett, P. B. Moore, D. Schlessinger, and J. R. Warner, eds. The ribosome: structure, function, and evolution. American Society for Microbiology, Washington, D.C.

Noller, H. F., and C. R. Woese. 1981. Secondary structure of 16S ribosomal RNA. Science **212**:403–411.

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1055489 by guest on 20 August 2024

PEATTIE, D. A., S. DOUTHWAITE, R. A. GARRETT, and H. F. NOLLER. 1981. A "bulged" double helix in a RNA–protein contact site. Proc. Natl. Acad. Sci. USA **78**:7331–7335.

PENNY, D., E. E. WATSON, R. E. HICKSON, and P. J. LOCKHART. 1993. Some recent progress with methods for evolutionary trees. N.Z. J. Bot. **31**:275–288.

PONT-KINGDON, G. A., C. T. BEAGLEY, R. OKIMOTO, and D. R. WOLSTENHOLME. 1994. Mitochondrial DNA of the sea anemone, *Metridium senile* (Cnidaria): prokaryote-like genes for $tRNA^{f\text{-}Met}$ and small-subunit ribosomal RNA, and standard genetic code specificities for AGR and ATA codons. J. Mol. Evol. **39**:387–399.

SANTALUCIA, J., R. KIERZEK, and D. H. TURNER. 1991. Stabilities of consecutive A·C, C·C, G·G, U·C, and U·U mismatches in RNA internal loops: evidence for stable hydrogen-bonded U·U and C·C+ pairs. Biochemistry **30**:8242–8251.

SIMON, C. 1991. Molecular systematics at the species boundary: exploiting conserved and variable regions of the mitochondrial genome of animals via direct sequencing from amplified DNA. Pp. 33–71 *in* G. M. HEWITT, A. W. B. JOHNSTON, and J. P. W. YOUNG, eds. Molecular techniques in taxonomy. NATO ASI Series, Vol. H57.

SIMON, C., S. PÄÄBO, T. D. KOCHER, and A. C. WILSON. 1990. Evolution of mitochondrial ribosomal RNA in insects as shown by the polymerase chain reaction. Pp. 235–244 *in* M. T. CLEGG, and S. J. O'BRIEN, eds. UCLA Symposia on Molecular and Cellular Biology, New Series, Vol. 122.

SIMON, C., F. FRATI, A. BECKENBACH, B. CRESPI, H. LIU, and P. FLOOK. 1994. Evolution, weighting, and phylogenetic utility of mitochondrial gene sequences and a compilation of conserved polymerase chain reaction primers. Ann. Entomol. Soc. Am. **87**:651–701.

STATES, D. J. 1992. Molecular sequence accuracy: analysing imperfect data. Trends Genet. **8**:52–55.

STIEGLER, P., P. CARBON, J. P. EBEL, and C. EHRESMANN. 1981. A general secondary-structure model for procaryotic and eucaryotic RNAs of the small ribosomal subunits. Eur. J. Biochem. **120**:487–495.

SULLIVAN, J., K. E. HOLSINGER, and C. SIMON. In press. Among-site rate variation and phylogenetic analysis of 12S rRNA in sigmodontine rodents. Mol. Biol. Evol.

SWOFFORD, D. L., and G. J. OLSEN. 1990. Phylogeny reconstructions. Pp. 411–501 *in* D. M. HILLIS and C. MORITZ, eds. Molecular systematics. Sinauer, Sunderland, Mass.

TAYLOR, M. F. J., S. W. MCKECHNIE, N. PIERCE, and M. KREITMAN. 1993. The lepidopteran mitochondrial control region: structure and evolution. Mol. Biol. Evol. **10**:1259–1272.

THOMAS, W. K., J. MAA, and A. C. WILSON. 1989. Shifting constraints on tRNA genes during mitochondrial DNA evolution in animals. New Biol. **1**:93–100.

THORNE, J. L., H. KISHINO, and J. FELSENSTEIN. 1991. An evolutionary model for maximum likelihood alignment of DNA sequences. J. Mol. Evol. **33**:114–124.

TOPAL, M. D., and J. R. FRESCO. 1976. Complementary base-pairing and the origin of substitution mutations. Nature **263**:285–289.

VAN DE PEER, Y., J.-M. NEEFS, P. DE RIJK, and R. DE WACHTER. 1993. Reconstructing evolution from eukaryotic small-ribosomal-subunit RNA sequences: calibration of the molecular clock. J. Mol. Evol. **37**:221–232.

VAN DE PEER, Y., I. VAN DEN BROECK, P. DE RIJK, and R. DE WACHTER. 1994. Database on the structure of small ribosomal subunit RNA. Nucleic Acids Res. **22**:3488–3494.

VARANI, G., and A. PARDI. 1994. Structure of RNA. Pp. 1–24 *in* K. NAGAI and I. MATTAJ, eds. RNA–protein interactions. Oxford University Press, Oxford.

VAWTER, L., and W. M. BROWN. 1993. Rates and patterns of base change in the SSU ribosomal RNA gene. Genetics **134**:597–608.

VON AHSEN, U., and H. F. NOLLER. 1995. Identification of bases in 16S rRNA essential for tRNA binding at the 30S ribosomal P site. Science **267**: 234–237.

ZUKER, M. 1989. On finding all suboptimal foldings of an RNA molecule. Science **244**:48–52.

ZUKER, M., J. A. JAEGER, and D. H. TURNER. 1991. A comparison of optimal and suboptimal RNA secondary structures predicted by free energy minimization with structures determined by phylogenetic comparison. Nucleic Acids Res. **19**:2707–2714.

ZWIEB, C., C. GLOTZ, and R. BRIMACOMBE. 1981. Secondary structure comparisons between SSU ribosomal RNA molecules from six different species. Nucleic Acids Res. **9**:3621–3640.

SIMON EASTEAL, reviewing editor

Accepted June 14, 1995

Downloaded from https://academic.oup.com/mbe/article/13/1/150/1055489 by guest on 20 August 2022