

Database on the structure of small ribosomal subunit RNA

Yves Van de Peer, An Caers, Peter De Rijk and Rupert De Wachter*

Departement Biochemie, Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium

Received October 6, 1997; Accepted October 8, 1997

ABSTRACT

About 8600 complete or nearly complete sequences are now available from the Antwerp database on small ribosomal subunit RNA. All these sequences are aligned with one another on the basis of the adopted secondary structure model, which is corroborated by the observation of compensating substitutions in the alignment. Literature references, accession numbers and detailed taxonomic information are also compiled. The database can be consulted via the World Wide Web at URL <http://rrna.uia.ac.be/ssu/>

CONTENTS OF THE DATABASE

In August 1997, the Antwerp database on small subunit (SSU) rRNA structure contained 2361 eukaryotic, 5487 bacterial, 238 archaeal, 98 plastid and 445 mitochondrial sequences. The database comprises complete or nearly-complete sequences while partial SSU rRNA sequences are included only if the combined length of the sequenced segments amounts to at least 70% of the estimated chain length of the molecule. The chain length of a partially determined sequence is estimated by comparing it to a complete sequence of the presumed closest relative. All SSU rRNA sequences are stored in the form of an alignment and contain the postulated secondary structure pattern in encoded form.

Table 1 lists the different eukaryotic taxa and the number of representatives in the database. The taxonomic classification of the species is according to Brusca and Brusca (1) for the Animalia, according to Cronquist (2) for the higher plants, according to Ainsworth *et al.* (3) for the zygomycetes and ascomycetes, according to Moore (4) for the basidiomycetes and ustomycetes, and according to Margulis *et al.* (5) for the remaining eukaryotes, viz. the Protoctista.

Table 2 covers the prokaryotic SSU rRNA sequences. The classification of prokaryotes is based on the construction of evolutionary trees. New sequences retrieved from EMBL (6), GenBank (7) or from direct submissions, are aligned with their presumed closest relative. Evolutionary trees are then constructed by the neighbor-joining method (8), and according to the phylogenetic position observed, the species are assigned to one of the taxa described by Woese and co-workers (9,10) and our research group (11,12).

SECONDARY STRUCTURE AND NUCLEOTIDE VARIABILITY

Two different secondary structure models form the basis of our SSU rRNA alignment. The first one is the prokaryotic model, which is applicable to Bacteria, Archaea, plastids and mitochondria, while the second one is the eukaryotic model which is applicable to all Eukaryotes. The two models are slightly different, each containing a number of structural elements specific for the group (see further). The prokaryotic model is essentially identical to that followed by Gutell (13), but the model for eukaryotic SSU rRNAs includes a secondary structure pattern in certain variable areas left undefined in the models of the latter author.

Helices in the SSU rRNA secondary structure model are given a different number if separated by a multibranching loop (e.g. helices 9 and 10), by a pseudoknot loop (e.g. helices 1 and 2), or by a single stranded area that does not form a loop (e.g. helices 2 and 32). A single number is given to 50 universal helices, defined as those present in all SSU rRNAs from Archaea, Bacteria and plastids known to date. The universal helices are also present in all known eukaryotic SSU rRNAs except in those of Microsporidia (Microspora) and in those of organisms classified as Parabasalia (*Trichomonas* and relatives), where some of these helices are missing. Helices specific to the eukaryotic model are numbered Ea_b, where a is the number of the preceding universal helix and b sequentially numbers all helices inserted between universal helices a and a+1. Helices specific to the prokaryotic model are similarly given composite numbers of the form Pa_b. It should be noted that mitochondrial sequences show extreme variability in length and in the number of helices present.

Figure 1 shows the secondary structure model of the SSU rRNA nucleotide sequence of *Drosophila melanogaster*. With respect to the previously published model (14), a different base pairing pattern has been adopted for the sequence separating the 5'-strands of helices E23₂ and E23₅. Following the proposal of Vogler *et al.* (15) for the local secondary structure in insect SSU rRNAs, two helices of the previous model have been combined into a single helix E23₃. It is not yet clear whether a similar change is applicable to SSU rRNAs of other taxa that contain helices between E23₂ and E23₅ (see 14), therefore the helix numbering has not yet been changed. Helices such as E23₃ and 43 are very variable in length and may actually contain additional branching points, so the proposed structures should be considered

* To whom correspondence should be addressed. Tel: +32 3 820 23 19; Fax: +32 3 820 22 48; Email: dwachter@uia.ua.ac.be

Table 1. List of eukaryotic taxa represented in the database and number of their representatives (August 1997)

Kingdom Animalia ^a				Kingdom Plantae ^a				
Phylum	Class	Number of sequences ^b		Phylum	Class	Number of sequences ^b		
		N	M			N	M	P
Placozoa		2		Bryophyta	Anthocerotopsida	11		
Rhombzoa		3		Bryopsida		17		
Orthonectida		2		Marchantiopsida		8	1	1
Rhombzoa		3		Equisetophyta		2		1
Porifera	Calcarea	3		Lycopodiophyta	Isotopsida	1		1
	Demospongiae	3			Lycopodiopsida	9		3
Cnidaria	Anthozoa	4		Magnoliophyta	Liliopsida	17	5	3
	Cubozoa	1			Magnoliopsida	308	4	20
	Hydrozoa	2		Pinophyta	Cycadopsida	2		
Ctenophora		2			Gnetopsida	10		1
Platyhelminthes	Cestoda	2			Ginkgoopsida	1		
	Trematoda	14			Pinopsida	24		2
	Turbellaria	25		Polypodiophyta		19		8
Nemertea	Anopla	2		Psilotophyta	Psilotopsida	3		2
Rotifera		3		Total:		432	10	42
Gastrotricha		1		Kingdom Protocista^a				
Kinorhyncha		1		Phylum	Class	Number of sequences ^b		
Nematoda	Secernentea	20	2			N	M	P
	Uncertain affiliation	3		Actinopoda	Heliozoa	1		
Nematomorpha		3		Apicomplexa	Coccidia	44		
Priapula		2			Hematozoa	62	1	1
Entoprocta		2			Uncertain affiliation	2		
Acanthocephala	Archiacanthocephala	1		Bacillariophyta	Bacillariophyceae	5		3
	Uncertain affiliation	2			Coccinopiscophyceae	22		
Annelida	Hirudinida	6		Chlorarachnida		18	1	3
	Oligochaeta	8	1	Chlorophyta	Charophyceae	29		3
	Polychaeta	16			Chlorophyceae	165	6	16
Sipuncula	Phascolosomida	1			Prasinophyceae	7		
Echiura		1			Ulvoophyceae	41		
Pogonophora		1			Uncertain affiliation	4		
Vestimentifera	Basibranchia	1		Chrysoophyta	Chrysoophyceae	24		2
Arthropoda	Branchiopoda	4	24		Dictyochophyceae	1		
	Chilopoda	1			Uncertain affiliation	12		
	Chelicerata	31		Chytridiomycota		9	2	
	Insecta	53	22	Ciliophora		62	5	
	Malacostraca	17	2	Conjugaphyta	Conjugatophyceae	9		1
	Maxillopoda	12		Cryptophyta		25		4
	Uncertain affiliation	1		Dictyostelida		1	1	
Onychophora		1		Dinoflagellata		30		
Tardigrada		5		Euglenida		1		7
Pentastomida	Pentastomata	1		Eustigmatophyta	Eustigmatophyceae	4		
Mollusca	Bivalvia	29	9	Glaucocystophyta	Glaucocystophyceae	1		3
	Gastropoda	15	4		Uncertain affiliation			1
	Polyplacophora	2	2	Granuloreticulosa		8		
Phoronida		2		Haplosporidia	Haplosporea	7		
Ectoprocta	Phylactolaemata	1		Hyphochytridiomycota		1		
Echinodermata	Asteroidea	2	1	Labyrinthulomycota		6		
	Crinoidea	1		Microspora		44		
	Echinoidea	24	4	Myxozoa	Myxosporea	13		
	Holothuroidea	1	1	Oomycota		4		
	Ophiuroidea	1		Phaeophyta		11	1	1
Chaetognatha		3		Plasmodial Slime Molds: Myxomycota		1	2	
Hemichordata	Enteropneusta	2		Plasmodiophoromycota		1		
Chordata	Agnatha	4	1	Prymnesiophyta		18		4
	Amphibia	18	4	Rhizopoda	Filosea	1		
	Aves	3	10		Lobosea	35	2	
	Chondrichthyes	4		Rhodophyta		159	2	5
	Mammalia	11	216	Xanthophyta		1		
	Osteichthyes	15	13	Zoomastigina	Amebomastigota	5		
	Reptilia	4	15		Choanomastigotes	2		
	Cephalochordata (Sub.)	1			Diplomonadida	10		
	Urochordata (Subphyl.)	4			Kinetoplastida	43	10	
					Parabasalia	11		
Total:		413	331		Zoomastigophora	2		
Kingdom Fungi^a				Stramenopiles	Uncertain affiliation	9		
Subphylum	Class	Number of sequences ^b		Uncertain affiliation		11		2
		N	M	Total:		982	33	56
Ascomycotina	Discomycetes	69						
	Hemiascomycetes	123	9					
	Loculoascomycetes	43						
	Plectomycetes	45	26					
	Pyrenomycetes	27	30					
	Uncertain affiliation	16	1					
Basidiomycotina	Heterobasidiomycetes	55						
	Hymenomycetes	62	2					
	Uncertain affiliation	3						
Ustomycotina	Ustomycetes	53						
Zygomycotina	Zygomycetes	38	3					
Total:		534	71					

^aThe Metazoan taxa are listed in the same order as they appear in (1).

^bThe number of sequences listed in the database is larger than the number of species, because for certain species multiple SSU rRNA sequences have been determined, usually by different authors. The sequences are not necessarily identical because they may have been determined for different varieties or strains of a species, or for different genes of the same organism. The number is listed for sequences of nuclear (N), mitochondrial (M) and plastid (P) origin.

^cThe fungal, plant and protocist phyla and classes are ordered alphabetically.

Table 2. List of prokaryotic taxa represented in the database and number of their representatives (August 1997)

Bacteria		
Division	Number of sequences ^a	
Chlamydiae		86
Cyanobacteria		68
Fibrobacter		17
Flavobacteria and relatives		268
Fusobacterium and relatives		29
Gram Positives and relatives, Low G+C		1303
Gram Positives and relatives, High G+C		1083
Green Sulfur		29
Green non sulfur		10
Planctomyces and relatives		47
Proteobacteria α		671
Proteobacteria β		257
Proteobacteria γ		1029
Proteobacteria δ		132
Proteobacteria ϵ		134
Proteobacteria, uncertain affiliation		6
Radioresistant micrococci and relatives		34
Spirochetes		208
Thermophilic oxygen reducers		6
Thermotogales		13
Uncertain affiliation ^b		57
Total:		5487

Archaea		
Division	Subdivision	Number of sequences ^a
Crenarchaeota		89
Euryarchaeota	Archaeoglobales	3
	Halobacteria	48
	Methanobacteriales	24
	Methanococcales	17
	Methanomicrobium group	34
	Methanopyrales	1
	Thermococcales	20
	Thermoplasma	1
	Uncertain affiliation	1
Total:		238

^aThe number of sequences listed in the database is larger than the number of species (cf. Table 1).

^bIn some cases, it cannot be decided to which taxonomic group a species should be ascribed, since the clustering of its SSU rRNA sequence is unstable and depends on the tree construction method used and on the set of sequences included in the analysis.

as tentative. Small differences in lay-out with the previously proposed model (14) may also be noticed in the drawing of Figure 1, due to the fact that one of us (PDR) has developed a new sophisticated program to automatically draw secondary structures, starting from a sequence file in which the secondary structure information is encoded. This program, called RNAviz, will soon be made available to the scientific community. Examples of secondary structure models for prokaryotic and mitochondrial SSU rRNAs have been given in previous papers on our database (11,16–18).

Recently, we developed a new method for measuring the relative substitution rate of individual sites in extensive nucleotide sequence alignments (19). By dividing nucleotides into five variability subsets, and giving a different color to each of the subsets, color maps superimposed on the secondary structure of SSU rRNAs can be constructed (20). These color maps can be interpreted in terms of higher order structure, function and evolution of the molecules, and facilitate the selection of areas suitable for the design of PCR primers and hybridization probes

(20). Variability measurement is also important for the precise estimation of evolutionary distances and the inference of phylogenetic trees (e.g. 21). Color maps for bacterial 5S rRNA, SSU rRNA, LSU rRNA and eukaryotic SSU rRNA (22) can also be consulted via internet at URL <http://bioc-www.uia.ac.be/u/yvdp/>

AVAILABILITY OF THE DATA

In order to make access to the data as fast as possible, each SSU rRNA sequence is stored in a separate file. Each of these files contains primary and secondary structure information, as well as annotations such as accession number, literature reference and detailed taxonomic specifications. The SSU rRNA database is made available via the World Wide Web at URL <http://rrna.uia.ac.be/ssu/>. Through WWW, it is easy to select sequences either one by one, or by taxonomic group, or by a combination of both. Sophisticated query tools have been added to make retrieval of sequences as flexible and user-friendly as possible. When the search for sequences has been successful, sequences matching the query can be fetched and copied to the host computer in different formats. On-line information about the database is also available.

If problems occur in connecting to the server or in retrieving data, the authors can be contacted by electronic mail to yvdp@uia.ua.ac.be or dwachter@uia.ua.ac.be. Users publishing results based on data retrieved from our database are requested to cite this paper.

ACKNOWLEDGEMENTS

Our research was supported by the BIOTECH program of the commission of European Communities (contract BIO2-CT94-3098), by a research project of the University of Antwerp (UA), by the Fund for Scientific Research Flanders, and by the Special Research Fund of the University (UIA). Yves Van de Peer and Peter De Rijk are Research Assistants of the Fund for Scientific Research Flanders.

REFERENCES

- 1 Brusca,R.C. and Brusca,G.J. (1990) *Invertebrates*. Sinauer Associates, Inc., Sunderland.
- 2 Cronquist,A. (1971) *Introductory Botany*. Harper & Row, New York.
- 3 Ainsworth,G.C., Sparrow,F.K. and Sussman,A.S. (1973) *The Fungi: and Advanced Treatise*. Academic Press, New York, Vol. 4A.
- 4 Moore,R.T. (1988) in Moriarty,C.H. (ed.), *Taxonomy Putting Plants and Animals in Their Place*. Royal Irish Academy, Dublin, pp. 61–88.
- 5 Margulis,L., Corliss,J.O., Melkonian,M. and Chapman,D.J. (eds) (1990) *Handbook of Protozoa*. Jones and Bartlett Publishers, Boston.
- 6 Stoesser,G., Sterk,P., Tuli,M.A., Stoehr,P.J. and Cameron,G.H. (1997) *Nucleic Acids Res.* **25**, 7–13 [see also this issue (1998) *Nucleic Acids Res.* **26**, 8–15].
- 7 Benson,D.A., Boguski,M., Lipman,D.J. and Ostell,J. (1997) *Nucleic Acids Res.* **25**, 1–6 [see also this issue (1998) *Nucleic Acids Res.* **26**, 1–7].
- 8 Saitou,N. and Nei,M. (1987) *Mol. Biol. Evol.* **4**, 406–425.
- 9 Woese,C.R. (1987) *Microbiol. Rev.* **51**, 221–271.
- 10 Olsen,G.J., Woese,C.R. and Overbeek,R. (1994) *J. Bacteriol.* **176**, 1–6.
- 11 Neefs,J.-M., Van de Peer,Y., De Rijk,P., Chapelle,S. and De Wachter,R. (1993) *Nucleic Acids Res.* **21**, 3025–3049.
- 12 Van de Peer,Y., Neefs,J.-M., De Vos,P. and De Wachter,R. (1994) *System. Appl. Microbiol.* **17**, 32–38.
- 13 Gutell,R.R. (1994) *Nucleic Acids Res.* **22**, 3502–3507.
- 14 De Rijk,P., Neefs,J.-M., Van de Peer,Y. and De Wachter,R. (1992) *Nucleic Acids Res.* **20**, 2075–2089.
- 15 Vogler,A.P., Welsh,A. and Hancock,J.M. (1997) *Mol. Biol. Evol.* **14**, 6–19.

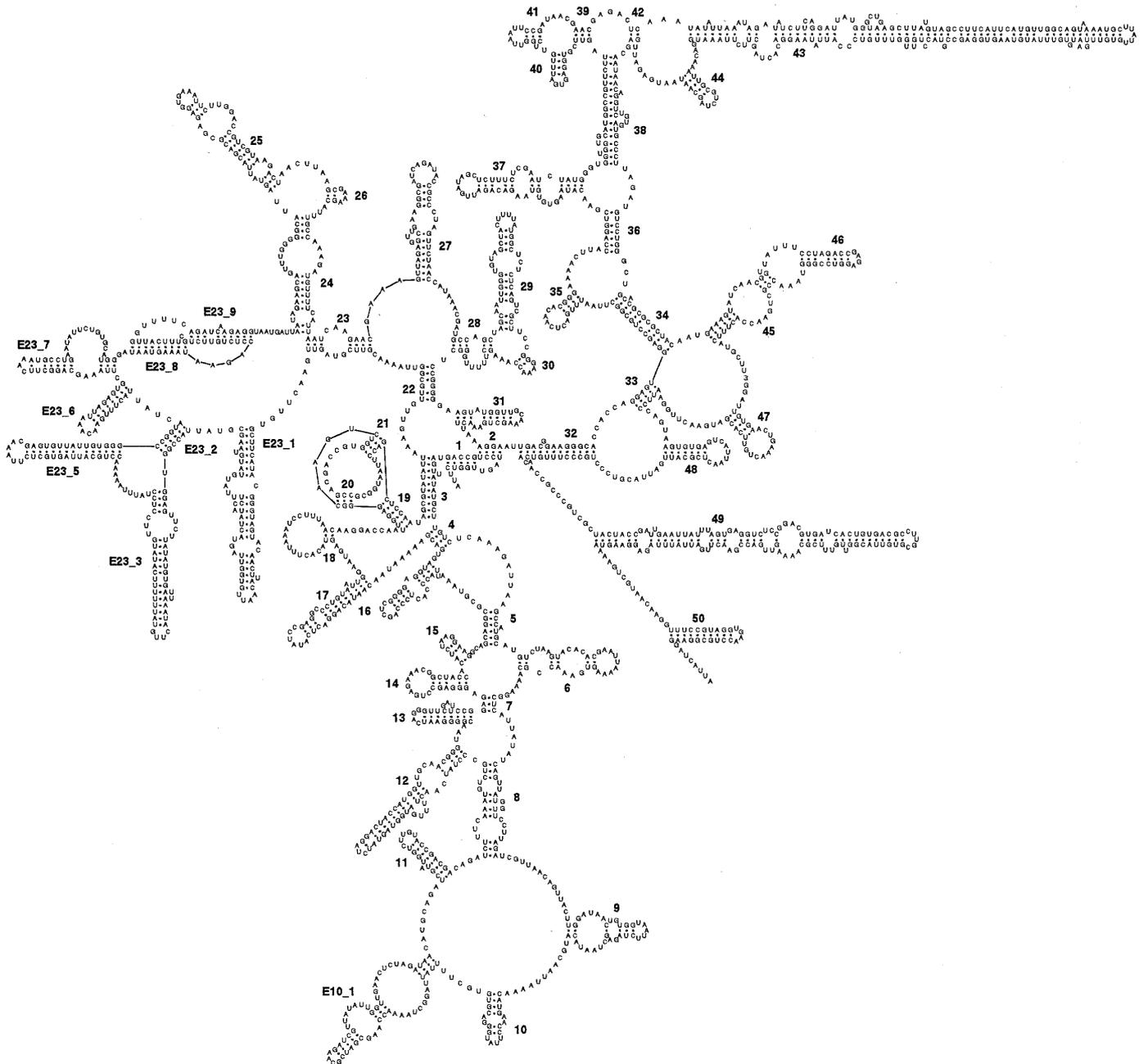


Figure 1. Secondary structure model for the nuclear SSU rRNA of the insect *D.melanogaster*. The sequence is written clockwise from 5' to 3' terminus.

<p>16 Neefs,J.-M., Van de Peer,Y., De Rijk,P., Goris,A. and De Wachter,R. (1991) <i>Nucleic Acids Res.</i> 19, 1987–2015.</p> <p>17 Van de Peer,Y., Van den Broeck,I., De Rijk,P. and De Wachter,R. (1994) <i>Nucleic Acids Res.</i> 22, 3488–3494.</p> <p>18 Van de Peer,Y., Nicolai,S., De Rijk,P. and De Wachter,R. (1996) <i>Nucleic Acids Res.</i> 24, 86–91.</p> <p>19 Van de Peer,Y., Van der Auwera,G. and De Wachter,R. (1996) <i>J. Mol. Evol.</i> 42, 201–210.</p>	<p>20 Van de Peer,Y., Chapelle,S. and De Wachter,R. (1996) <i>Nucleic Acids Res.</i> 24, 3381–3391.</p> <p>21 Van de Peer,Y., Rensing,S., Maier,U.-G. and De Wachter,R. (1996) <i>Proc. Natl. Acad. Sci. USA</i> 93, 7732–7736.</p> <p>22 Van de Peer,Y., Jansen,J., De Rijk,P. and De Wachter,R. (1997) <i>Nucleic Acids Res.</i> 25, 111–116.</p>
-----------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------	--------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------------