© 1999 Oxford University Press Nucleic Acids Research, 1999, Vol. 27, No. 1 **179–183**

Database on the structure of small subunit ribosomal RNA

Yves Van de Peer, Elmar Robbrecht[1], Sybren de Hoog[2], An Caers, Peter De Rijk and Rupert De Wachter*

Departement Biochemie, Universiteit Antwerpen (UIA), Universiteitsplein 1, B-2610 Antwerpen, Belgium, [1]Nationale Plantentuin, Domein van Bouchout, B-1860 Meise, Belgium and [2]Centraalbureau voor Schimmelcultures, PO Box 273, NL-3740 AG Baarn, The Netherlands

Received October 8, 1998; Accepted October 13, 1998

ABSTRACT

Over 11 500 complete or nearly complete sequences are now available from the Antwerp database on small subunit ribosomal 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 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 1998, the Antwerp small subunit (SSU) rRNA database contained 3166 eukaryotic, 7336 bacterial, 324 archaeal, 120 plastid and 601 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 a presumed close relative. All SSU rRNA sequences are stored in the form of an alignment and contain the postulated secondary structure pattern in encoded form (see the rRNA website at URL http://rrna.uia.ac.be/ for detailed information).

Table 1 lists the different eukaryotic taxa and the number of representatives in the database for which the SSU rRNA sequence has been determined. The taxonomic classification of the animals is according to Brusca and Brusca (1). For the plants and the Fungi, we have chosen to extend the taxonomic information supplied up to the level of orders, contrary to previous papers describing the contents of the SSU rRNA database. The classification of vascular plants is according to Mabberley (2), while the classification of Bryopsida is according to Crosby and Magill (3). Additional classificatory information for the terrestrial plants was taken from Sitte *et al.* (4) and Farr *et al.* (5). On the rRNA website, taxonomic information regarding the family to which the plant species are classified will also be available. The classification of the 'true' fungi or Eumycota is according to Hawksworth *et al.* (6), Kurtzman and Fell (7), and de Hoog and Guarro (8). The remaining eukaryotes, viz. the protoctists are classified according to Margulis *et al.* (9). Overall, species are included in the database under the binomial used for the publication of the sequence. We therefore refrained from doing any taxonomic name change, even when obviously needed.

Table 2 covers the prokaryotic SSU rRNA sequences. The classification of prokaryotes is, as before, based on the construction of evolutionary trees. New sequences retrieved from EMBL or GenBank, or from direct submissions, are aligned with their presumed closest relative. Evolutionary trees are then constructed by the neighbor-joining method (10), and according to the phylogenetic position observed, the species are assigned to one of the taxa previously described by Woese and coworkers (11,12) and our research group (13,14).

SECONDARY STRUCTURE AND NUCLEOTIDE VARIABILITY

Our SSU rRNA sequence alignment is based on two different secondary structure models. The first one is the prokaryotic model, which is applicable to Bacteria, Archaea, plastids and mitochondria, while the second one is the eukaryotic model applicable to all Eukaryotes. The two models are slightly different, each containing a number of structural elements specific for the group (see below). The prokaryotic model is essentially identical to those distributed by Gutell (15), but the model followed 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 multibranched 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, which are present in all SSU rRNAs from Archaea, Bacteria and plastids known to date. These 50 helices are also present in all known eukaryotic SSU rRNAs except in those of the microsporidians (such as *Vairimorpha*, *Nosema*, and relatives), where some of these helices are missing. Helix 11 is also missing in the

---

*To whom correspondence should be addressed. Tel: +32 3 820 2319; Fax: +32 3 820 2248; Email: dwachter@uia.ua.ac.be

Table 1. List of eukaryotic taxa represented in the database and number of their representatives (August 1998)

Kingdom Animalia[a]		Number of sequences[b]	
Phylum	Class	N	M
Placozoa		2	
Orthonectida		2	
Rhombozoa	Dicyemida	3	
Porifera	Calcarea	3	
	Demospongiae	3	
Cnidaria	Anthozoa	12	
	Cubozoa	1	
	Hydrozoa	3	
	Uncertain affiliation	7	
Ctenophora		2	
Platyhelminthes	Cestoda	2	
	Trematoda	15	
	Turbellaria	34	
	Uncertain affiliation	1	
Nemertea	Anopla	1	
	Enopla	1	
Rotifera		3	
Gastrotricha		2	
Kinorhyncha		1	
Nematoda	Adenophorea	27	
	Secernentea	60	
	Uncertain affiliation	10	
Nematomorpha	Gordioida	5	
Priapula		2	
Entoprocta		3	
Gnathostomulida		1	
Acanthocephala	Archiacanthocephala	3	
	Eoacanthocephala	2	
	Palaeacanthocephala	7	
Annelida	Hirudinida	6	
	Oligochaeta	8	1
	Polychaeta	16	
Sipuncula	Phascolosomida	1	
Echiura		1	
Pogonophora		1	
Vestimentifera	Basibranchia	1	
Arthropoda	Branchiopoda	5	
	Chelicerata	50	1
	Insecta	152	23
	Malacostraca	17	2
	Maxillopoda	12	
	Myriapoda	1	
Onychophora		1	
Tardigrada		5	
Pentastomida	Pentastomata	1	
Mollusca	Caudofoveata	1	
	Bivalvia	34	
	Gastropoda	43	
	Polyplacophora	3	
	Scaphopoda	1	
Phoronida		2	
Ectoprocta	Gymnolaemata	1	
	Phylactolaemata	2	
Brachiopoda	Articulata	8	
	Inarticulata	7	
Echinodermata	Asteroidea	4	1
	Crinoidea	2	1
	Echinoidea	24	4
	Holothuroidea	4	1
	Ophiuroidea	4	
Chaetognatha		3	
Hemichordata	Enteropneusta	2	1
Chordata	Agnatha	4	1
	Amphibia	18	62
	Aves	3	88
	Chondrichtyes	4	
	Mammalia	11	277
	Osteichtyes	16	32
	Reptilia	4	21
	Cephalochordata (Sub.)	1	
	Urochordata (Subphyl.)	4	
Total:		**706**	**516**

Kingdom Fungi[c]			Number of sequences[b]	
Phylum	Class	Order	N	M
Ascomycota				
	Archiascomycetes	Pneumocystidales	3	
		Protomycetales	4	
		Schizosaccharomycetales	5	2
		Taphrinales	18	
	Euascomyctes	Caliciales	14	
		Chaetothyriales	13	
		Cyttariales	1	
		Diaporthales	6	1
		Dothideales	14	
		Elaphomycetales	3	
		Erysiphales	1	
		Eurotiales	58	24
		Hypocreales	4	
		Lecanorales	9	
		Leotiales	11	
		Microascales	13	
		Neolectales	2	
		Onygenales	18	1
		Ophiostomatales	3	
		Ostropales	3	
		Peltigerales	2	
		Pezizales	68	
		Phyllachorales	1	1
		Pleosporales	23	
		Rhytismatales	1	
		Sordariales	8	1
		Xylariales	2	
		Uncertain affiliation	23	
	Hemiascomycetes	Saccharomycetales	120	7
		Uncertain affiliation	1	
	Uncertain affiliation		1	
Basidiomycota				
	Heterobasidiomycetes	Agaricostilbales	1	
		Atractiellales	1	
		Auriculariales	3	
		Filobasidiales	10	
		Heterogastridiales	1	
		Tremellales	30	
	Homobasidiomycetes	Agaricales	17	1
		Aphyllophorales	81	
		Boletales	1	
		Bondarzewiales	3	
		Cantharellales	9	
		Ceratobasidiales	18	
		Cortinariales	1	
		Dacrymycetales	4	
		Fistulinales	2	
		Gomphales	2	
		Hericiales	9	
		Lycoperdiales	3	
		Nidulariales	2	
		Phallales	1	
		Schizophyllales	1	
		Sclerodermatales	2	
		Tulostomatales	1	
	Teliomycetes	Uredinales	4	
	Ustomycetes	Graphiolales	2	
		Platygloeales	6	
		Sporidiales	31	
		Ustilaginales	10	
		Uncertain affiliation	1	
	Uncertain affiliation		13	
Zygomycota				
	Trichomycetes	Harpellales	5	
	Zygomycetes	Entomophthorales	3	
		Glomales	28	
		Kickxellales	9	
		Mucorales	5	
		Uncertain affiliation	1	
Uncertain affiliation			1	
Total:			**775**	**38**

trichomonads and relatives. 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. Mitochondrial sequences show extreme variability in length and in the number of helices present. Figure 1 shows the secondary structure model of the plastid SSU rRNA nucleotide sequence of *Zea mays*.

Examples of secondary structure models for eukaryotic and mitochondrial SSU rRNAs have been given in previous papers on our database (13,16,17). Color maps showing the distribution of

Table 1. *Continued*

Kingdom Plantae[c]				
Phylum	Order	Number of sequences[b]		
Class		N	M	P
Bryophyta				
Anthocerotopsida		7		
Bryopsida		31		1
Jungermanniopsida		12		
Marchantiopsida		11	1	1
Pteridophyta				
Equisetopsida		2		1
Filicopsida		19		8
Lycopsida		9		3
Psilotopsida		3		2
Pinophyta				
Cycadopsida		2		
Ginkgoopsida		1		
Gnetopsida		10		1
Pinopsida		43		2
Magnoliophyta				
Liliopsida	Alismatales	2		
	Arales	2		
	Arecales	2		
	Bromeliales	2		
	Commelinales	1		
	Cyperales	4	3	3
	Liliales	13		
	Orchidales	8		
	Poales		2	
	Typhales	1		
	Zingiberales	7		
Magnoliopsida	Apiales	10		1
	Aristolochiales	4		
	Asterales	2		
	Batidales	1		
	Campanulales	2		
	Capparidales	6	1	
	Caryophyllales	5		
	Casuarinales	1		
	Celastrales	1		
	Cornales	10		
	Daphniphyllales	1		
	Dilleniales	4		
	Dipsacales	3		
	Ebenales	4		
	Ericales	6		
	Eucommiales	1		
	Euphorbiales	7		3
	Fabales	4	2	4
	Fagales	2		1
	Gentianales	1		
	Geraniales	4		
	Haloragidales	3		
	Hamamelididales	6		
	Illiciales	2		
	Lamiales	3		
	Laurales	3		
	Linales	1		
	Magnoliales	9		
	Malvales	5		
	Myrtales	1	1	
	Nepenthales	3		
	Nymphaeales	6		
	Papaverales	3		
	Piperales	7		
	Plumbaginales	1		
	Polygalales	3		
	Polygonales	1		
	Proteales	4		
	Rafflesiales	5	3	5
	Ranunculales	21		
	Rhamnales	1		
	Rosales	44		
	Rubiales	1		
	Santalales	72	3	2
	Sapindales	4		
	Scophulariales	19		3
	Solanales	6	1	7
	Theales	2		
	Trochodendrales	2		
	Urticales	4		
	Violales	16		1
Total:		**529**	**17**	**49**

Kingdom Protoctista[c]				
Phylum	Class	Number of sequences[b]		
		N	M	P
Actinopoda	Heliozoa	1		
Apicomplexa	Coccidia	52		1
	Hematozoa	73		3
	Uncertain affiliation	2		
Bacillariophyta	Bacillariophyceae	10		3
	Coscinopiscophyceae	23		
Chlorarachnida		21	1	3
Chlorophyta	Charophyceae	30		3
	Chlorophyceae	167	6	19
	Prasinophyceae	7		1
	Ulvophyceae	39		
	Uncertain affiliation	4		
Chrysophyta	Chrysophyceae	26		2
	Dictyochophyceae	3		
	Uncertain affiliation	9		
Chytridiomycota		9	2	
Ciliophora	Colpodea	2		
	Karyorelictea	4		
	Litostomatea	16		
	Nassophorea	5	3	
	Oligohymenophorea	31	2	
	Phyllopharyngea	2		
	Prostomatea	1		
	Spirotrichea	13		
Conjugaphyta	Conjugatophyceae	9		1
Cryptophyta		30		4
Dictyostelida		1	1	
Dinoflagellata		50		
Euglenida		4		7
Eustigmatophyta	Eustigmatophyceae	5		
Glaucocystophyta	Glaucocystophyceae	1		5
Granuloreticulosa		1		
Haplosporidia	Haplosporea	7		
Hyphochytridiomycota		1		
Labyrinthulomycota		6		
Microspora		60		
Myxozoa	Myxosporea	20		
Oomycota		4		
Phaeophyta		11	1	1
Plasmodial Slime Molds	Myxomycota	1	2	
Plasmodiophoromycota		1		
Prymnesiophyta		17		5
Rhaphidophyta		4		
Rhizopoda	Filosea	1		
	Lobosea	72	2	
	Uncertain affiliation	1		
Rhodophyta		178	2	6
Xanthophyta		5		
Zoomastigina	Amebomastigota	6		
	Choanomastigotes	2		
	Diplomonadida	13		
	Kinetoplastida	56	7	
	Parabasalia	20		
	Proteromonadida	1		
	Zoomastigophora	2		
Stramenopiles	Uncertain affiliation	7		
Uncertain affiliation		9	1	7
Total:		**1156**	**30**	**71**

[a] The Metazoan taxa are listed in the same order as they appear in ref. 1.

[b] The 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.

[c] Except in the case of the plant phyla, taxa are ordered alphabetically.

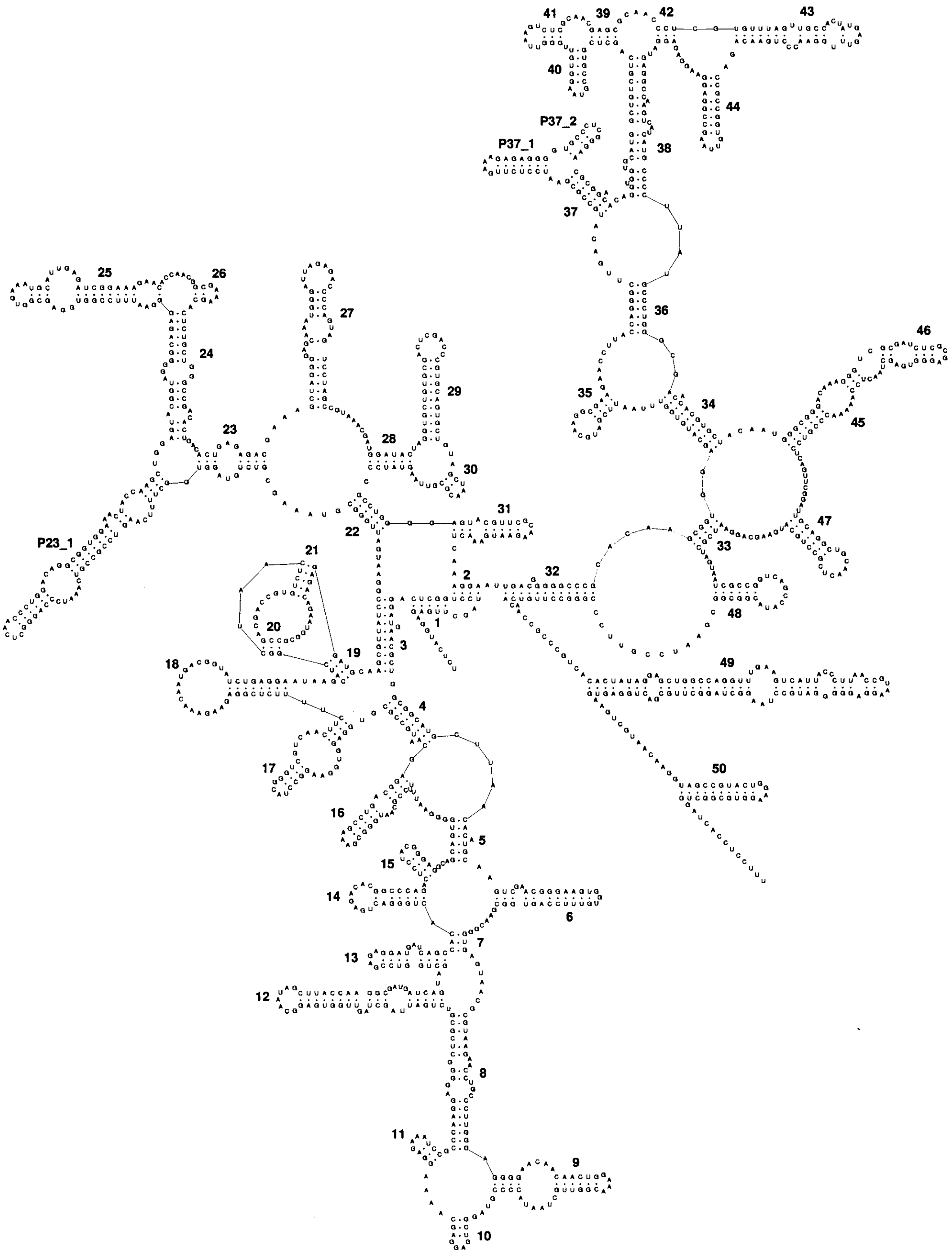

Figure 1. Secondary structure model for the plastid SSU rRNA of *Zea mays*. The sequence is written clockwise from 5′ to 3′ terminus.

conserved and variable sites in bacterial and eukaryotic SSU rRNAs (18,19) can be consulted via the Internet at URL http://bioc-www.uia.ac.be/u/yvdp/

AVAILABILITY OF THE DATA

Each SSU rRNA sequence is stored in a separate file in order to simplify access to the data. 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 the WWW, it is very easy to select sequences either one by one, or by taxonomic group, or by a combination of both. Sequences can be retrieved in different formats. On-line information about the database is also available.

Table 2. List of prokaryotic taxa represented in the database and number of their representatives (August 1998)

Bacteria

Division	Number of sequences[a]
Chlamydiae	86
Cyanobacteria	107
Fibrobacter	18
Flavobacteria and relatives	442
Fusobacterium and relatives	32
Gram Positives and relatives, Low G+C	1645
Gram Positives and relatives, High G+C	1308
Green Sulfur	29
Green non sulfur	25
Planctomyces and relatives	60
Proteobacteria α	877
Proteobacteria β	385
Proteobacteria γ	1417
Proteobacteria δ	181
Proteobacteria ε	167
Proteobacteria, uncertain affiliation	9
Radioresistant micrococci and relatives	51
Spirochetes	279
Thermophilic oxygen reducers	8
Thermotogales	18
Uncertain affiliation[b]	192
Total:	**7336**

Archaea

Division	Subdivision	Number of sequences[a]
Crenarchaeota		93
Euryarchaeota	Archaeoglobales	4
	Halobacteria	70
	Methanobacteriales	34
	Methanococcales	19
	Methanomicrobium group	63
	Methanopyrales	1
	Thermococcales	29
	Thermoplasma	2
	Uncertain affiliation	9
Total:		**324**

[a] The number of sequences listed in the database is larger than the number of species (cf. Table 1).

[b] In 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.

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 is supported by the Fund for Scientific Research–Flanders and by the Special Research Fund of the University of Antwerp (Belgium). Y.V.deP. and P.DeR. are Research Fellows of the Fund for Scientific Research—Flanders.

REFERENCES

1 Brusca,R.C. and Brusca,G.J. (1990) *Invertebrates*. Sinauer Associates Inc., Sunderland.

2 Mabberley,D.J. (1987; reprint 1996) *The Plant Book. A Portable Dictionary of the Higher Plants*. Cambridge University Press, Cambridge, UK.

3 Crosby,M.R. and Magill,R.E. (1978) *A Dictionary of Mosses* (second printing). Monogr. Syst. Bot. Missouri Bot. Garden **3**.

4 Sitte,P., Ziegler,H., Ehrendorfer,F. and Bresinsky,A. (1991) *Lehrbuch der Botanik für Hochschulen*. Übersicht des Pflanzenreiches: 530–828. Stuttgart, Fischer.

5 Farr,E.R., Leussink,J.A. and Stafleu,F., (1979). *Index Nominum Genericorum (Plantarum)*. Regn. Veget. 100–102.

6 Hawksworth,D.L., Kirk,P.M., Sutton,B.C. and Pegler,D.N. (1995) *Ainsworth & Bisby's Dictionary of the Fungi*, 8th ed. CAB International, Egham.

7 Kurtzman,C.P. and Fell,J.W. (1998) *The Yeasts, a Taxonomic Study*, 4th ed. Elsevier, Amsterdam.

8 de Hoog,G.S. and Guarro,J. (1995) *Atlas of Clinical Fungi*. Centraalbureau voor Schimmelcultures, Baarn; Universitat Rovira i Virgili, Reus.

9 Margulis,L., Corliss,J.O., Melkonian,M. and Chapman,D.J. (eds) (1990) *Handbook of Protoctista*, Jones and Bartlett Publishers, Boston.

10 Saitou,N. and Nei,M. (1987) *Mol. Biol. Evol*., **4**, 406–425.

11 Woese,C.R. (1987) *Microbiol. Rev*., **51**, 221–271.

12 Olsen,G.J., Woese,C.R. and Overbeek,R. (1994) *J. Bacteriol*., **176**, 1–6.

13 Neefs,J.-M., Van de Peer,Y., De Rijk,P., Chapelle,S. and De Wachter,R. (1993) *Nucleic Acids Res*., **20**, 3025–3049.

14 Van de Peer,Y., Neefs,J.-M., De Rijk,P., De Vos,P. and De Wachter,R. (1994) *System. Appl. Microbiol*., **17**, 32–38.

15 Gutell,R.R. (1994) *Nucleic Acids Res*., **22**, 3502–3507.

16 Van de Peer,Y., Van den Broeck,I., De Rijk,P. and De Wachter,R. (1994) *Nucleic Acids Res*., **22**, 3488–3494.

17 Van de Peer,Y., Nicolaï,S., De Rijk,P. and De Wachter,R. (1996) *Nucleic Acids Res*., **24**, 86–91.

18 Van de Peer,Y., Chapelle,S. and De Wachter,R. (1996) *Nucleic Acids Res*., **24**, 3381–3391.

19 Van de Peer,Y., Jansen,J., De Rijk,P. and De Wachter,R. (1997) *Nucleic Acids Res*., **25**, 111–116.