The ribosomal RNA Database project

Gary J.Olsen, Niels Larsen and Carl R.Woese*
Department of Micorobiology, University of Illinois, 131 Burrill Hall, 407 South Goodwin Avenue, Urbana, IL 61801, USA

The RDP (ribosomal RNA database project), currently at the University of Illinois, is a new start receiving support from the National Science Foundation, Division of Instrumentation and Resources. The RDP is becoming functional in stages, the initial one of which will be operational by the time of this publication.

The objectives of the RDP are to supply its user community with rRNA data, with various software packages for handling, analyzing and displaying it, and with certain useful services. (Eventually the RDP intends to collect, organize, cross-reference and distribute all information regarding the ribosome and translation, as well).

The individual sequences are drawn from various previously available rRNA database collections (1−3), from major sequence repositories [principally GenBank (4)] and from individuals who have kindly deposited their own laboratory's rRNA sequence collection with the RDP. They are aligned manually, invoking obvious primary structural homology and known secondary (and higher order) structural constraints (5). (This procedure will ultimately be automated, however). All ribosomal RNA sequence types (small subunit and large subunit, including 5S rRNA) from all categories of organisms, organelles, endosymbionts and direct natural population isolations, will be distributed in aligned and phylogenetically organized (or organizable) form. Three types of ribosomal RNA sequence releases will be available: (1) the 'complete' collection for any type of rRNA; (2) representative collections, comprising a small but phylogenetically representative selection of sequences; and (3) custom collections, selected from the complete listing by the user. Releases will be in any of several common formats [GenBank (with inserted alignment gaps), Olsen editor format, Macke editor format, ALMA editor format (6), PAUP (7), PHYLIP (8), or certain others]. They will be available in a variety of electronic media: tapes, diskettes of several types, ftp, and e-mail (when this is appropriate).

The initial RDP release will comprise approximately 450 complete (or nearly so) aligned prokaryotic small subunit rRNA sequences. The sequences included therein are summarized in the accompanying table. This release will be followed shortly by similar releases comprising eukaryotic, mitochondrial and plastid small subunit, large subunit, and 5S rRNA sequences. Non-aligned sequences can also be obtained if desired.

The RDP will later offer a variety of software and services. The former includes several sequence editors [one of which, the Olsen VAX/VMS-based editor is currently available; another of which, the Genetic Data Environment (GDE) X-window package (designed in collaboration with S. Smith) is under development],

phylogenetic analysis tools, a tree drawing program, and a 'bibliographic' editor for data accompanying rRNA sequences. The RDP will also distribute, by agreement, the various rRNA secondary structures created by R. Gutell, M. Gray and M. Schnare.

The RDP's services will include a 'sequence assessment' system, which aligns a given sequence against a reference alignment, and reports, among other things, salient characteristics (idiosyncrasies, group diagnostic features, possible sequencing errors, etc). Sequence alignment, phylogenetic analyses and secondary structural representation will be available for investigators who do not themselves have such capacities.

The RDP will in due course accept (properly documented) sequence submissions, which it will then format and deposit in the basal sequence databases, such as the GenBank/EMBL alliance and the NCBI backbone, and release for general distribution (at a date specified by the submitter).

The RDP is unique in its area in a number of respects: in offering the user considerable flexibility in formats and data selection; in resolving and documenting discrepancies among different versions of the same sequence; in integration of oligonucleotide catalog and sequence data; in the degree to which sequences are phylogenetically organized; and in terms of the software packages and services it will ultimately offer. At this time the RDP is attempting to identify its potential user community and define their needs. Those interested in utilizing the RDP offerings, or wishing to be informed of releases, services, software, etc., should contact the RDP at the e-mail address below, to be put on its mailing list. (Written contact is also possible for those not having access to e-mail.) To obtain sequences, software or services from the RDP, or information concerning these and future releases, the user should make initial contact via the RDP e-mail address, stating the nature of their request, the formats in which they wish to receive sequences, etc. Arrangements will then be made to transfer the information to the user. As the needs of the user community become defined, this user interface will become largely automated and the modes of access broadened and standardized. Since the RDP is committed to accumulating and distributing in machine readable form additional ribosome related data, we would appreciate suggestions as to data that might be included and formats in which they can be made most useful.

The RDP e-mail address is RDP@scotty.life.uiuc.edu. Telephone contact is through Terry Davis 217-333-1142.

---

* To whom correspondence should be addressed

REFERENCES

1. Neefs,J.-M., Van de Peer,Y., Hendriks,L. and De Wachter,R. (1990). *Nucleic Acids Res.*, **18**, 2237−2317.
2. Gutell,R.R., Schnare,M.N. and Gray,M.W. (1990) *Nucleic Acids Res.*, **18**, 2319−2330.
3. Specht,T., Wolters,J. and Erdmann,V.A. (1990) *Nucleic Acids Res.*, **18**, 2215−2235.
4. Bilofsky,H.S. and Burks,C. (1988) *Nucleic Acids Res.*, **16**, 1861−1864.
5. Woese,C.R., Gutell,R.R., Gupta,R. and Noller,H.F. (1983) *Microbiol. Rev.*, **47**, 621−669.
6. Thirup,S. and Larsen,N. (1990) *PROTEINS: Structure, Function and Genetics*, **7**, 291−295.
7. Felsenstein,J. (1990) *PHYLIP Manual Version 3.3.* University Herbarium, University of California, Berkeley, CA.
8. Swofford,D.L. (1990) *PAUP: Phylogenetic Analysis Using Parsimony, Version 3.0*, Illinois Natural History Survey, Champaign, IL.

Table 1. Prokaryotic small subunit ribosomal RNA sequences available in first RDP release.

Acetomaculum ruminis	Clathrochloris sulfurica
Acholeplasma (3 spp.)	Clostridium (40 spp.) *
Acinetobacter calcoaceticus	Corynebacterium (3 spp.)
Acyrthosyphon pisum symbiont (2 spp.)	Coxiella burnetii
Aerococcus viridans	Cytophaga (11 spp.) *
Aeromicrobium erythreus	Deinococcus radiodurans *
Aeromonas hydrophila	Dermatophilus congolensis
Agrobacterium tumefaciens	Desulfobacter postgatei
Alcaligenes (4 spp.)	Desulfomonile tiedjei
Anaeroplasma (3 spp.)	Desulfosarcina variabilis
Ancylobacter aquaticus	Desulfovibrio desulfuricans *
Archaeoglobus fulgidus *	Desulfurococcus mobilis
Arhodomonas oleiferhydrans	Desulfuromonas acetoxidans
Arthrobacter (2 spp.) *	Ectothiorhodospira (3 spp.)
Asteroleplasma anaerobium	Ehrlichia risticii
Azospirillum lipoferum	Eikenella (4 spp.)
Bacillus subtilis *	Enterococcus faecalis
Bacteroides (5 spp.) *	Erwinia (2 spp.)
Bdellovibrio (2 spp.)	Erysipelothrix rhusiopathiae
Beijerinckia indica	Erythrobacter (2 spp.)
Bifidobacterium (16 spp.) *	Escherichia coli *
Borrelia burgdorferi	Eubacterium (3 spp.)
Brochothrix thermosphacta	Faenia rectivirgula
Brucella abortus	Fervidobacterium (2 spp.) *
Campylobacter (2 spp.)	Flavobacterium (16 spp.) *
Cardiobacterium hominis	Flectobacillus (3 spp.)
Carnobacterium (2 spp.)	Flexibacter (13 spp.) *
Chlamydia psittaci *	Flexistipes sinusarabici
Chlorobium (3 spp.) *	Frankia sp.
Chloroflexus aurantiacus *	Fusobacterium (8 spp.) *
Chromatium (2 spp.) *	Gardnerella vaginalis
Chromobacterium (2 spp.)	Gemella haemolysans
Citrobacter freundii	Haemophilus influenzae

Hafnia alvei

Haliscomenobacter hydrossis

Haloanaerobium praevalens

Halobacterium (2 spp.)

Halococcus morrhuae

Haloferax volcanii *

Heliobacterium chlorum

Herpetosiphon aurantiacus

Hyphomicrobium vulgare

Kingella (2 spp.)

Kurthia zopfii

Lactobacillus (40 spp.) *

Lactococcus (6 spp.)

Legionella pneumophila

Leptonema illini

Leptospira sp.

Leptotrichia buccalis

Leuconostoc (5 spp.)

Listeria monocytogenes

Megasphaera elsdenii

Methanobacterium (3 spp.) *

Methanobrevibacter arboriphilicus

Methanococcus (5 spp.) *

Methanocorpusculum parvum

Methanogenium (4 spp.)

Methanohalophilus (3 spp.)

Methanolobus tindarius

Methanomicrobium mobile

Methanoplanus limicola

Methanosarcina (5 spp.) *

Methanosphaera stadtmanii

Methanospirillum hungatei *

Methanothermus fervidus *

Methanothrix (2 spp.)

Methylcoccoides methylutens

Methylobacterium (2 spp.)

Methylococcus capsulatus

Methylocystis parvus

Methylomonas methanica

Methylophilus methylotrophus

Methylosinus trichosporium

Methylosporovibrio methanica

Micrococcus luteus

Microscilla (5 spp.)

MLO infecting Oenothera

Mycobacterium (22 spp.)

Mycoplasma (28 spp.)

Myxococcus xanthus *

Neisseria (3 spp.)

Nitrosolobus multiformis

Nitrosomonas europae

Nocardia otitidis-caviarum

Oceanospirillum linum

Pediococcus (2 spp.)

Planctomyces staleyi *

Plesiomonas shigelloides

Propionibacterium (5 spp.)

Proteus vulgaris

Pseudomonas (5 spp.)

Pseudonocardia thermophila

Pyrodictium occultum

Rhodobacter capsulatus

Rhodococcus equi

Rhodocyclus (2 spp.) *

Rhodomicrobium vannielii

Rhodopila globiformis

Rhodopseudomonas (3 spp.) *

Rhodospirillum (7 spp.) *

Rickettsia (3 spp.)

Rochalimaea quintana

Rothia dentocariosa

Ruminobacter amylophilus

Runella slithyformis

Saccharopolyspora hirsuta

Saprospira grandis *

Serpula hyodysenteriae *

Serratia marcescens

Simonsiella muelleri

Sphingobacter mizutae

Spirillum volutans *

Spirochaeta (8 spp.) *

Spiroplasma (10 spp.)

Spirosoma linguale

Sporohalobacter lortetii

Sporolactobacillus inulinis

Sporomusa paucivorans

Streptococcus (4 spp.)

Streptomyces (3 spp.)

Sulfolobus solfataricus *

Synechococcus sp. *

Syntrophomonas wolfei

Thermococcus celer *

Thermomicrobium roseum *

Thermoplasma acidophilum

Thermoproteus tenax *

Thermosipho africanus

Thermotoga maritima *

Thermus thermophilus *

Treponema (4 spp.) *

Tsukamurella paurometabolu

Ureaplasma urealyticum

Vagococcus fluvialis

Vibrio parahaemolyticus

Vitreoscilla stercoraria

Wolbachia persica

Wolinella succinogenes *

Xanthomonas maltophilia

Xylella fastidiosa

Yersinia enterocolitica

* Species or genera included in the representative collection.