

The RDP (Ribosomal Database Project)

Bonnie L. Maidak*, Gary J. Olsen, Niels Larsen¹, Ross Overbeek²,
Michael J. McCaughey and Carl R. Woese

Department of Microbiology, University of Illinois, B103 C&LSL, 601 South Goodwin Avenue, Urbana, IL 61801, USA, ¹Center for Microbial Ecology, Michigan State University, East Lansing, MI 48824, USA and ²Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA

Received October 9, 1996; Accepted October 9, 1996

ABSTRACT

The Ribosomal Database Project (RDP) is a curated database that offers ribosome-related data, analysis services and associated computer programs. The offerings include phylogenetically ordered alignments of ribosomal RNA (rRNA) sequences, derived phylogenetic trees, rRNA secondary structure diagrams, and various software for handling, analyzing and displaying alignments and trees. The data are available via anonymous FTP (rdp.life.uiuc.edu), electronic mail (server@rdp.life.uiuc.edu), gopher (rdpgopher.life.uiuc.edu) and WWW (<http://rdpwww.life.uiuc.edu/>). The electronic mail and WWW servers provide ribosomal probe checking, approximate phylogenetic placement of user-submitted sequences, screening for possible chimeric rRNA sequences, automated alignment, and a suggested placement of an unknown sequence on an existing phylogenetic tree.

DATA AND SERVICES

The Ribosomal Database Project (RDP) provides data, programs and services related to the ribosome. This paper describes changes that have been introduced since last year's description (1). Details about currently available analysis functions, data, and available programs can be found on the servers described below.

The ribosomal RNA sequences in the RDP alignments are drawn from major sequence repositories [GenBank (2) and EBI (3)] and direct submissions to the RDP. During the past year, RDP staff extended extensive effort to obtaining and examining all publications related to the sequencing of small subunit ribosomal RNA. Bibliographic citations from the sequence databases were updated from 'Unpublished' when possible and corrected when found to be in error. In each instance it was confirmed that (i) the published paper referred to the given public database sequence and (ii) the feature annotations were consistent with the published paper.

More than 800 small subunit (SSU) rRNA sequences have been added to the prokaryotic alignment, bringing the total to over 3600. The eukaryotic alignment has been extensively reworked, and several hundred new sequences added as well. New this year is an alignment of ~1000 small subunit mitochondrial sequences.

The full molecular length ranges from 611 bp for *Trypanosoma brucei* to 2023 bp for *Lupinus luteus* representing a tremendous amount of variation among organisms.

Phylogenetic trees are available for the sequences in the posted prokaryotic, eukaryotic and mitochondrial SSU rRNA alignments. They have been built by sequential addition of sequences to starting trees of full-length sequences, all using a maximum-likelihood criterion (4,5). The RDP also offers a collection of SSU and LSU rRNA secondary structure diagrams in PostScript format generated and supplied by R. Gutell and his collaborators.

To facilitate access to specific rRNA aligned and unaligned sequences, the RDP offers subdirectories containing GenBank-formatted files of each sequence (directory names: alignments/sequences/[A-Z] and unaligned/sequences/[A-Z]).

RDP ACCESS

All of the RDP data and analysis services can be accessed through the RDP World Wide Web server: <http://rdpwww.life.uiuc.edu/>. The Web pages have been extensively revised. Figure 1 depicts the current Home Page. Links to related sites have also been expanded.

The RDP automated electronic mail server also provides both data and analysis services. Its address is server@rdp.life.uiuc.edu. To obtain an overview of what data and services are currently available, send a mail message with the phrase 'help' as the body of the message. (Full command descriptions can be obtained by sending 'help complete'.) If your electronic mail address is not known to the e-mail server, you will also receive a registration form. After returning the completed registration form, you will be automatically notified when new data or services become available.

The RDP data can be accessed via anonymous FTP to rdp.life.uiuc.edu. Once you are logged in (using a user-id of 'anonymous' and your electronic mail address for password), examine the 00README files, which describe the organization of the data and programs.

The RDP gopher host name is rdpgopher.life.uiuc.edu. Gopher access to RDP data through the World Wide Web is also available (URL: gopher://rdpgopher.life.uiuc.edu/).

Electronic mail correspondence with RDP staff should be addressed to rdp@phylo.life.uiuc.edu. Those without access to

* To whom correspondence should be addressed. Tel: +1 217 333 5866; Fax: +1 217 244 6697; Email: b-maidak@uiuc.edu

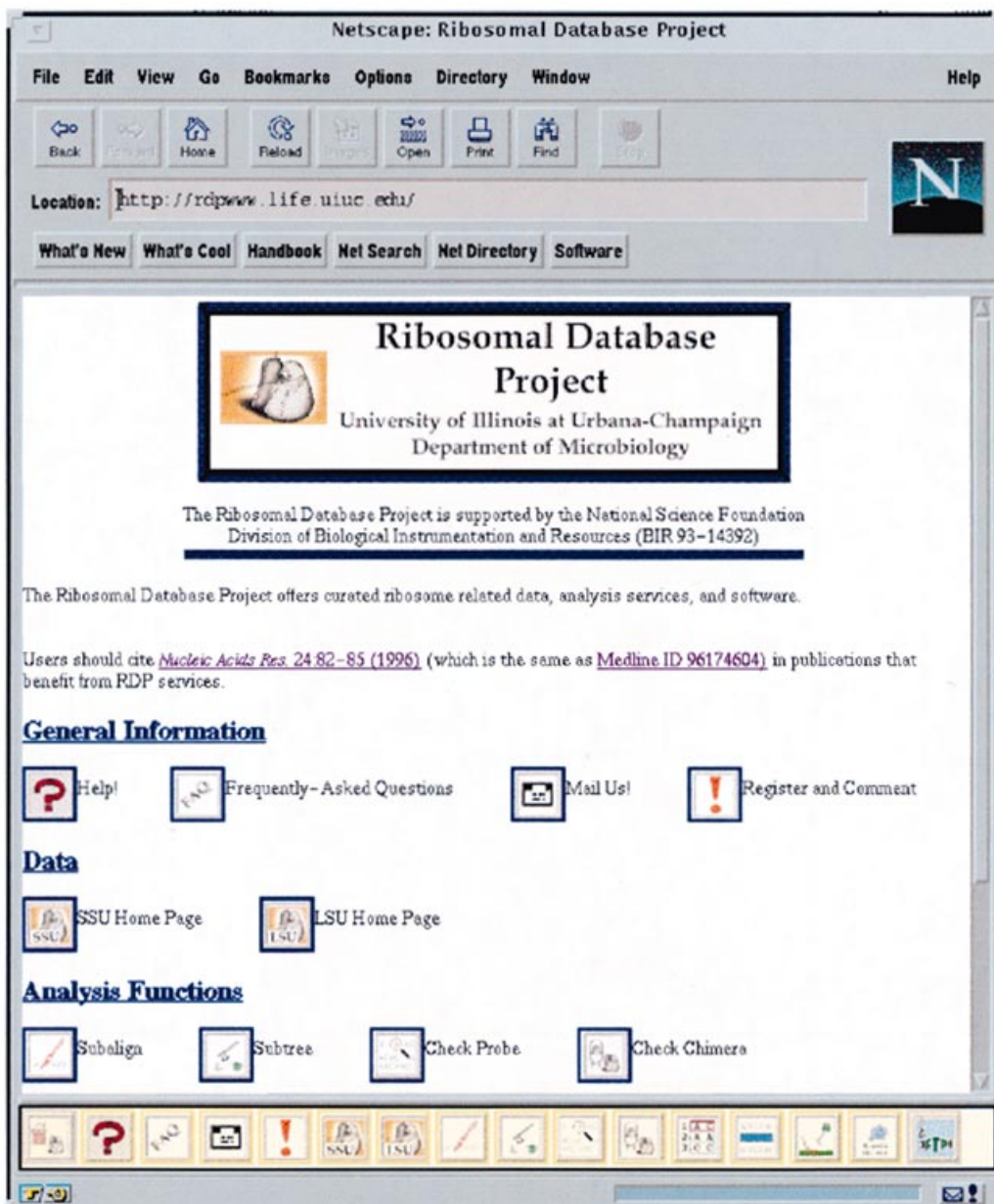


Figure 1. The RDP Home Page.

electronic mail may contact the RDP curator (B.L.M.) via telephone (+ 1 217 333 5866), fax (+1 217 244 6697) or regular mail.

RDP CITATION

Research assisted by any RDP service should cite: the Ribosomal Database Project (RDP) at the University of Illinois in Urbana, Illinois; the release number; and this article (i.e., Maidak *et al.*, 1997). Please state which data, programs and services were used and the method of access.

ACKNOWLEDGEMENTS

We thank R. Gutell (and his colleagues) and M. L. Sogin for

providing their data collections. The RDP is largely and currently supported by the National Science Foundation, Biological Instrumentation and Resources Division.

REFERENCES

- 1 Maidak, B.L., Olsen, G.J., Larsen, N., Overbeek, R., McCaughey, M.J. and Woese, C.R. (1996) *Nucleic Acids Res.*, **24**, 82–85.
- 2 Benson, D.A., Boguski, M., Lipman, D.J. and Ostell, J. (1996) *Nucleic Acids Res.*, **24**, 1–5.
- 3 Rodriguez-Tome, P., Stoehr, P.J., Cameron, G.N. and Flores, T.P. (1996) *Nucleic Acids Res.*, **24**, 6–12.
- 4 Felsenstein, J. (1981) *J. Mol. Evol.*, **17**, 368–376.
- 5 Olsen, G.J., Matsuda, H., Hagstrom, R. and Overbeek, R. (1994) *Comput. Appl. Biosci.*, **10**, 41–48.