

# The RDP (Ribosomal Database Project) continues

Bonnie L. Maidak\*, James R. Cole, Timothy G. Lilburn, Charles T. Parker Jr, Paul R. Saxman, Jason M. Stredwick, George M. Garrity<sup>1</sup>, Bing Li, Gary J. Olsen<sup>2</sup>, Sakti Pramanik<sup>3</sup>, Thomas M. Schmidt<sup>1</sup> and James M. Tiedje<sup>1</sup>

Center for Microbial Ecology, 540 Plant and Soil Sciences Building, Michigan State University, East Lansing, MI 48824-1325, USA, <sup>1</sup>Department of Microbiology, Michigan State University, 294 Giltner Hall, East Lansing, MI 48824-1101, USA, <sup>2</sup>Department of Microbiology, University of Illinois, B-103 C&LSL Building, 601 South Goodwin Avenue, Urbana, IL 61801-3714, USA and <sup>3</sup>Department of Computer Science and Engineering, 2148 Engineering Building, Michigan State University, East Lansing, MI 48824-1226, USA

Received October 4, 1999; Accepted October 6, 1999

## ABSTRACT

**The Ribosomal Database Project (RDP-II), previously described by Maidak *et al.*, continued during the past year to add new rRNA sequences to the aligned data and to improve the analysis commands. Release 7.1 (September 17, 1999) included more than 10 700 small subunit rRNA sequences. More than 850 type strain sequences were identified and added to the prokaryotic alignment, bringing the total number of type sequences to 3324 representing 2460 different species. Availability of an RDP-II mirror site in Japan is also near completion. RDP-II provides aligned and annotated rRNA sequences, derived phylogenetic trees and taxonomic hierarchies, and analysis services through its WWW server (<http://rdp.cme.msu.edu/>). Analysis services include rRNA probe checking, approximate phylogenetic placement of user sequences, screening user sequences for possible chimeric rRNA sequences, automated alignment, production of similarity matrices and services to plan and analyze terminal restriction fragment length polymorphism (T-RFLP) experiments.**

## DESCRIPTION

The Ribosomal Database Project (RDP-II) provides data, programs and services related to the ribosome. This paper describes changes since the 1999 description (1). Details about specific analysis functions, data and available programs can be found at the WWW site <http://rdp.cme.msu.edu/>

## Data

The ribosomal RNA sequences in the RDP-II alignments are mainly drawn from the major sequence repositories [GenBank (2), EMBL Data Library (3) and DDBJ (4)].

Release 7.1, September 17, 1999, contained more than 10 700 small subunit (SSU) rRNA sequences in aligned form and more than 21 000 sequences in unaligned form. It also included 3324 type strain sequences from 2460 different bacterial

species. Phylogenetic trees continue to be available for the sequences in the rRNA alignments. They have been assembled by sequential addition of new sequences onto subtrees drawn from our previous trees, using a maximum-likelihood approach (5,6) for each addition. The RDP-II now offers a collection of SSU and LSU rRNA secondary structure diagrams in PDF format from the PostScript files generated and supplied by R. Gutell and his collaborators (7). To facilitate scientific research, RDP-II serves as a repository for published alignments and masks used in trees. Extracting these alignments and masks will support recalculation of published rRNA phylogenetic trees. These data are available for download from the RDP-II WWW (<http://rdp.cme.msu.edu/>) and anonymous ftp servers.

## Analysis services

Two new analysis commands have been added: Sequence Selection and TAP T-RFLP. In addition, the interfaces to several other commands have been updated. The documentation pages for all analysis services have also been extensively revised. A brief description of each analysis command available on the WWW server is listed in Table 1.

## RDP-II CITATION AND ACCESS

Research assisted by any RDP service should cite: the Ribosomal Database Project (RDP-II) at the Michigan State University in East Lansing, Michigan; the release number; and this article. Please state which data, programs and services were used and the method of access.

The RDP-II data and analysis services can be found at <http://rdp.cme.msu.edu/>. A mirror site is being developed in cooperation with the Laboratory for Molecular Classification in the Center for Information Biology at the National Institute of Genetics (NIG), Japan. This new mirror site should provide better access to RDP-II for researchers in that part of the world. The mirror site address and other news will be announced at the RDP-II WWW site.

The Illinois site remains available with an automated electronic mail server; its address is [server@rdp.life.uiuc.edu](mailto:server@rdp.life.uiuc.edu). To obtain an overview of what data and services are currently available,

\*To whom correspondence should be addressed. Tel: +1 517 432 4998; Fax: +1 517 353 8957; Email: [curator@cme.msu.edu](mailto:curator@cme.msu.edu)

**Table 1.** WWW analytical commands

Probe Match	Analyzes the occurrence of a specified 'probe' sequence in a data set.
Sequence Match	Determines the most similar RDP-II sequences to a user sequence. Uses the taxonomic hierarchy to organize matches for display.
Sequence Align	Aligns a user sequence to the most similar sequence in the RDP-II data sets.
Similarity Matrix	Calculates a similarity/dissimilarity matrix of RDP-II and/or user sequences.
Chimera Check	Detects possible chimeric user sequences.
Alignment Slices	Extracts columns from the full RDP-II alignment, with identical neighboring sequences condensed into one sequence. Highlights variations between neighboring sequences. Useful for examining probe candidates, covariation, or for simply browsing a region of an alignment.
Sequence Selection	Allows selection of sequences from a dynamically displayed taxonomic hierarchy. The selected sequences can be downloaded or included as input in other RDP-II analysis functions.
T-RFLP	Uses data in the ABI sequencer format to generate a similarity matrix of lane data from T-RFLP result data.
TAP T-RFLP	Performs 'in silico' T-RFLP experiments in the RDP data set, facilitating experiment design and analysis.
(Sub) Trees	A Java applet that displays and manipulates the phylogenetic trees, creates subtrees, or selects sequences for inclusion as input in other RDP-II analysis functions.

send an Email message with the phrase 'help' as the body of the message.

Electronic mail correspondence with RDP-II staff should be addressed to curator@cme.msu.edu. Those without access to Email may contact the RDP-II staff via telephone (+1 517 432 4998), fax (+1 517 353 8957) or regular mail.

## FUTURE CHANGES AND ADDITIONS

Release 8.0 will occur in 2000 and it represents an anticipated doubling of the number of rRNA sequences available in the SSU prokaryotic alignment. Changes in the methods used to produce phylogenetic trees are being investigated. Using major phenotypic groupings, based on the outline in the second edition of Bergey's Manual of Systematic Bacteriology (<http://www.cme.msu.edu/Bergeys/>), to move among tree nodes is being considered. Improved user query access to RDP data is also being developed.

## ACKNOWLEDGEMENTS

We thank several individuals for their past contributions: Robin Gutell (and his colleagues), Niels Larsen, Tom Macke,

Michael J. McCaughey, Ross Overbeek, Mitch Sogin and Carl R. Woese. The National Science Foundation's Science and Technology Center Program, the US Department of Energy Office of Science, and the State of Michigan currently support RDP-II.

## REFERENCES

1. Maidak, B.L., Cole, J.R., Parker, C.T., Jr, Garrity, G.M., Larsen, N., Li, B., Lilburn, T.G., McCaughey, M.J., Olsen, G.J., Overbeek, R., Pramanik, S., Schmidt, T.M., Tiedje, J.M. and Woese, C.R. (1999) *Nucleic Acids Res.*, **27**, 171–173.
2. Benson, D.A., Boguski, M.S., Lipman, D.J., Ostell, J., Ouellette, B.F., Rapp, B.A. and Wheeler, D.L. (1999) *Nucleic Acids Res.*, **27**, 12–17. Updated article in this issue: *Nucleic Acids Res.* (2000) **28**, 15–18.
3. Stoesser, G., Tuli, M.A., Lopez, R. and Sterk, P. (1999) *Nucleic Acids Res.*, **27**, 18–24. Updated article in this issue: *Nucleic Acids Res.* (2000) **28**, 19–23.
4. Sugawara, H., Miyazaki, S., Gojobori, T. and Tateno, Y. (1999) *Nucleic Acids Res.*, **27**, 25–28. Updated article in this issue: *Nucleic Acids Res.* (2000) **28**, 24–26.
5. Felsenstein, J. (1981) *J. Mol. Evol.*, **17**, 368–376.
6. Olsen, G.J., Matsuda, H., Hagstrom, R. and Overbeek, R. (1994) *Comput. Appl. Biosci.*, **10**, 41–48.
7. Gutell, R.R. (1994) *Nucleic Acids Res.*, **22**, 3502–3507.