A new version of the RDP (Ribosomal Database Project)

Bonnie L. Maidak*, James R. Cole, Charles T. Parker, Jr, George M. Garrity, Niels Larsen[1], Bing Li, Timothy G. Lilburn, Michael J. McCaughey[2], Gary J. Olsen[2], Ross Overbeek[3], Sakti Pramanik[4], Thomas M. Schmidt[1], James M. Tiedje[1] and Carl R. Woese[2]

Department of Microbiology, Michigan State University, 294 Giltner Hall, East Lansing, MI 48824-1101, USA, [1]Center for Microbial Ecology, 540 Plant and Soil Sciences Building, Michigan State University, East Lansing, MI 48824-1325, USA, [2]Department of Microbiology, University of Illinois, B-103 C&LSL Building, 601 South Goodwin Avenue, Urbana, IL 61801-3714, USA, [3]Mathematics and Computer Science Division, Argonne National Laboratory, Argonne, IL 60439, USA and [4]Department of Computer Science and Engineering, 2148 Engineering Building, Michigan State University, East Lansing, MI 48824-1226, USA

ABSTRACT

The Ribosomal Database Project (RDP-II), previously described by Maidak *et al*. [*Nucleic Acids Res.* (1997), 25, 109–111], is now hosted by the Center for Microbial Ecology at Michigan State University. RDP-II is a curated database that offers ribosomal RNA (rRNA) nucleotide sequence data in aligned and unaligned forms, analysis services, and associated computer programs. During the past two years, data alignments have been updated and now include >9700 small subunit rRNA sequences. The recent development of an ObjectStore database will provide more rapid updating of data, better data accuracy and increased user access. RDP-II includes phylogenetically ordered alignments of rRNA sequences, derived phylogenetic trees, rRNA secondary structure diagrams, and various software programs for handling, analyzing and displaying alignments and trees. The data are available via anonymous ftp (ftp.cme.msu.edu) and WWW (http://www.cme.msu.edu/RDP). The WWW server provides 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. Additional utilities also exist at RDP-II, including distance matrix, T-RFLP, and a Java-based viewer of the phylogenetic trees that can be used to create subtrees.

DESCRIPTION

The Ribosomal Database Project (RDP-II) provides data, programs and services related to the ribosome. This paper describes changes since the 1997 description (1). Details about specific analysis functions, data, and available programs can be found at the WWW site (http://www.cme.msu.edu/RDP).
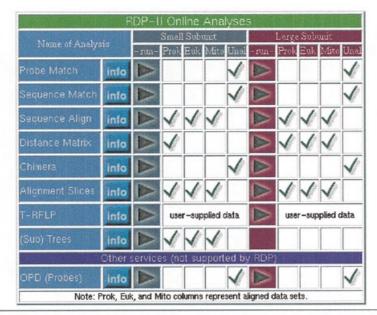
In December 1997, the previous RDP site hosted by the University of Illinois, officially moved to the Center for Microbial Ecology (CME) (http://www.cme.msu.edu), located at Michigan State University (MSU). Release 7.0, available on July 31, 1998, represented the first data release of RDP-II at CME. With the location change, all WWW pages have been redesigned, and the steps to use many of the previous online analysis tools have changed. Some functionality previously available at the University of Illinois web site currently exists only at that site. The URL for the original RDP WWW site is now http://rdpwww.life.uiuc.edu/index2.html , and it will remain in service until all functionality at Illinois is supported at MSU.

DBMS

A new database containing the RDP-II data using ObjectStore, a commercially available object-oriented database product, available from Object Design (http://www.odi.com), now exists. Development of an Annotation Editor, using web-based forms, also occurred during the past year. These web-based forms are being used to edit and curate the data in the database. Tools to be used in aligning the ribosomal RNA sequences are being re-evaluated for their compatibility with the new DBMS. Better curation and improved data reliability also exist with the new DBMS and Annotation Editor.

Data

The ribosomal RNA sequences in the RDP-II alignments are drawn from major sequence repositories [GenBank (2) and EBI (3)] and from direct submissions to the RDP.

---

*To whom correspondence should be addressed. Tel: +1 517 432 4998; Fax: +1 517 353 8957; Email: rdp@vitro.cme.msu.edu

Figure 1. Online analysis commands available on the RDP-II WWW server.

Release 7.0, July 31, 1998, contained >9700 small subunit (SSU) rRNA sequences in aligned form, and >22 000 sequences in unaligned form. During the past two years, the RDP completely reworked the eukaryotic alignment file, resulting in a >400% increase in the number of SSU rRNA sequences (from 436 to 2055). Representative lists of organisms for all three SSU data sets (prokaryotic, eukaryotic and mitochondrial) now exist, and additional compilation statistics for previous RDP releases are also available from the WWW and ftp servers. Alphabetical and taxonomic listings have been split into smaller files for easier web browser loading. Statistical tallies are also provided. The large subunit (LSU) rRNA alignment file has not changed from that previously described (1).

Phylogenetic trees continue to be available for the sequences in the posted SSU rRNA alignments. They have been assembled from appropriately overlapping subtrees, each of which has been inferred using maximum-likelihood analysis (4,5). Additional trees for the representative and taxonomic nodes for each aligned SSU rRNA data set are also now available. The RDP-II offers a collection of SSU and LSU rRNA secondary structure diagrams in PostScript format generated and supplied by R.Gutell and collaborators (6).

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

Data servers

The RDP-II World Wide Web server pages have been extensively revised. Figure 1 illustrates the current Online Analyses WWW page (minus the header information), with a brief description of each command available on the WWW server in Table 1.

RDP-II CITATION AND ACCESS

Research assisted by any RDP-II 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 URL: http://www.cme.msu.edu/RDP

The RDP-II data can be accessed via anonymous ftp to ftp.cme.msu.edu. Once you are logged in (using a user-id of 'anonymous' and your Email address for password), cd 'pub/RDP', and examine the 00README files, which describe the organization of the data and programs.

Currently, only the Illinois site supports an automated electronic mail server; its address is server@rdp.life.uiuc.edu. To obtain an overview of what data and services are currently available, send an Email message with the phrase 'help' as the body of the message. Plans to create a new Email server are underway at MSU.

Email correspondence with RDP-II staff should be addressed to rdp@vitro.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

Scientists at Michigan State are currently maintaining and working on further development of RDP-II as a cooperative effort. RDP-II will expend considerable effort over the next several months to incorporate new data into the ObjectStore DBMS and make these data available in aligned form to database users. Additional changes in the tools to produce phylogenetic

Table 1. WWW analytical commands

Probe Match	Analyzes the occurrence of a specified "probe" sequence in a data set.
Sequence Match	Uses the unaligned data set to determine the most similar sequence to the submitted one. Uses the taxonomic hierarchy to organize matches for display.
Sequence Align	The submitted sequence is aligned to the most similar sequence in the RDP-II aligned data sets using the Sequence Match program.
Distance Matrix	Calculates a fractional similarity/dissimilarity matrix of aligned submitted sequences.
Chimera	Detects possible chimeric sequences
Alignment Slices	Extracts regions of the full 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.
T-RFLP	Uses data in the ABI sequencing machine format to generate a similarity value and a matrix of lane data.
(Sub) Trees	Uses a Java applet to display and manipulate the phylogenetic trees.

trees are being investigated. Connections between the RDP-II and Bergey's Manual of Systematic Bacteriology (http://www.cme.msu.edu/Bergeys) will be established for data that are common to the two projects. To facilitate scientific research, RDP-II plans to serve as a repository for published alignments and masks used in trees. Extracting these alignments and masks will support recalculation of published rRNA phylogenetic trees.

REFERENCES

1 Maidak,B.L., Olsen,G.J., Larsen,N., Overbeek,R., McCaughey,M.J. and Woese,C.R. (1997) *Nucleic Acids Res.*, **25**, 109–111.
2 Benson,D.A., Boguski,M.S., Lipman,D.J., Ostell,J. and Ouellette,B.F. (1998) *Nucleic Acids Res.*, **26**, 1–7.
3 Stoesser,G., Moseley,M.A., Sleep,J., McGowran,M., Garcia-Pastor,M. and Sterk,P. (1998) *Nucleic Acids Res.*, **26**, 8–15.
4 Felsenstein,J. (1981) *J. Mol. Evol.*, **17**, 368–376.
5 Olsen,G.J., Matsuda,H., Hagstrom,R. and Overbeek,R. (1994) *Comput. Applic. Biosci.*, **10**, 41–48.
6 Gutell,R.R. (1994) *Nucleic Acids Res.*, **22**, 3502–3507.