Nucleic Acids Research, 2001, Vol. 29, No. 1 173–174

The RDP-II (Ribosomal Database Project)

Bonnie L. Maidak, James R. Cole, Timothy G. Lilburn*, Charles T. Parker Jr, Paul R. Saxman, Ryan J. Farris, George M. Garrity, Gary J. Olsen[1], Thomas M. Schmidt[2] and James M. Tiedje[2]

Center for Microbial Ecology, 540 Plant and Soil Sciences Building, Michigan State University, East Lansing, MI 48824-1325, USA, [1]Department of Microbiology, University of Illinois, B-103 C&LSL Building, 601 South Goodwin Avenue, Urbana, IL 61801-3714, USA and [2]Department of Microbiology and Molecular Genetics, Michigan State University, 294 Giltner Hall, East Lansing, MI 48824-1101, USA

Received October 2, 2000; Accepted October 4, 2000

ABSTRACT

The Ribosomal Database Project (RDP-II), previously described by Maidak *et al.* [*Nucleic Acids Res.* (2000), 28, 173–174], continued during the past year to add new rRNA sequences to the aligned data and to improve the analysis commands. Release 8.0 (June 1, 2000) consisted of 16 277 aligned prokaryotic small subunit (SSU) rRNA sequences while the number of eukaryotic and mitochondrial SSU rRNA sequences in aligned form remained at 2055 and 1503, respectively. The number of prokaryotic SSU rRNA sequences more than doubled from the previous release 14 months earlier, and ~75% are longer than 899 bp. An RDP-II mirror site in Japan is now available (http://wdcm.nig.ac.jp/RDP/html/index.html). 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 polymorphism experiments. The RDP-II email address for questions and comments has been changed from curator@cme.msu.edu to rdpstaff@msu.edu.

DESCRIPTION

The Ribosomal Database Project (RDP-II) provides data, programs and services related to ribosomal RNA sequences. This paper describes changes since the 2000 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 8.0, June 1, 2000, contained 16 277 prokaryotic small subunit (SSU) rRNA sequences in aligned form with ~75% longer than 899 bp. Type strain status is marked for a sequence if it is determinable. The number of eukaryotic and mitochondrial SSU rRNA sequences in aligned form remains at 2055 and 1503. Besides the sequences from the aligned data, more than 10 000 additional sequences were added to create the unaligned data bringing the total number to more than 30 000. The unaligned data are available for downloading and for analyses that do not require alignment. The all-inclusive RDP phylogenetic tree has not been updated for Release 8.0 because its size precludes any utility and because it has become inaccurate. Instead, we have decided to build a hierarchical set of trees, with a single tree that encompasses the breadth of the prokaryotic sequence diversity at the top of the hierarchy (a so-called backbone tree) and subordinate trees that encompass less and less of the diversity as one moves down the hierarchy. The sequences represented in the subordinate trees are selected according to their position in the RDP Release 8.0 hierarchy. The backbone tree and 13 of these subordinate trees were calculated using the WEIGHBOR algorithm (5) for Release 8.0 and eventually all sequences in the RDP-II prokaryotic SSU rRNA alignment will be in one or more subordinate trees. A new backbone phylogenetic tree for 217 prokaryotic SSU rRNA sequences was calculated using the WEIGHBOR algorithm (5). Additional trees using this approach for 13 smaller groups were also prepared for Release 8.0. Eventually, all sequences in the RDP-II prokaryotic SSU rRNA alignment will be in one or more of these smaller grouped trees. To facilitate scientific research, RDP-II serves as a repository for alignments and masks used by authors in the preparation of phylogenetic trees. The availability of these alignments and masks supports the recalculation of published rRNA phylogenetic trees. These data are available for download from the RDP-II WWW (http://rdp.cme.msu.edu/) server.

Analysis services

A brief description of each analysis command available on the WWW server can be found in Table 1 from the Maidak *et al.* (1) description of the RDP-II or from the Documentation section of the RDP-II WWW server (http://www.cme.msu.edu/RDP/docs/documentation.html).

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

Visualization of large sets of sequence data

For some applications (e.g. the detection of sequencing or annotation errors, the definition of taxonomic boundaries and visualization of outliers) it is necessary to build models with a complete set of aligned sequences, rather than a small subset of sequences, drawn either at random or deliberately. However, current methods for constructing phylogenetic trees are inherently limited. Such methods are computationally too intensive and the output is too complex to permit accurate interpretation. To that end, in collaboration with the Bergey's Manual Trust, work on alternative means of visualizing extremely large sets of sequences using Principal Component Analysis (PCA) was initiated during 2000. Two-dimensional scatter plots using PCA are available in the Supplementary Material links.

New auxiliary WWW sites

The Center for Microbial Ecology WWW server now supports two additional WWW sites that contain data related to the RDP-II. The Biodegradative Strain Database (http://bsd.cme.msu.edu) provides corresponding microbiological data to complement and integrate the phylogenetic data of the RDP-II with the chemical and metabolic data of the University of Minnesota Biocatalysis/Biodegradation Database (http://www.labmed.umn.edu/umbbd/index.html) (6). The second auxiliary WWW site is rrndb (http://rrndb.cme.msu.edu), which provides information pertaining to the number of rRNA operons contained on prokaryotic genomes. (7).

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.

The RDP-II data and analysis services can be found at URL: http://rdp.cme.msu.edu/. A mirror site is available at the Laboratory for Molecular Classification in the Center for Information Biology at the National Institute of Genetics (NIG), Japan (http://wdcm.nig.ac.jp/RDP/html/index.html). This new mirror site should provide better access to RDP-II for researchers in that part of the world.

The address for email correspondence with RDP-II staff is now rdpstaff@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

Several upgrades to the WWW analysis programs are planned for release in the near future. An improved sequence selection tool will allow searching and provide a graphical display of sequence completeness. A new analysis program will allow users to create phylogenetic trees incorporating RDP sequences along with their own data. In addition, Version 2.0 of the terminal restriction fragment polymorphism (T-RFLP) program (8) is under development. To keep abreast of the increasing volume of rRNA sequence data, we are evaluating changes in workflow, additional automation of annotation and more robust automated alignment procedures. These back-end changes should enable the RDP to provide timely release of rRNA data.

SUPPLEMENTARY MATERIAL

Additional material related to the RDP-II and described in the Supplementary Data section of this article at NAR Online consists of the following:

(i) a PDF file of a poster from the American Society for Microbiology (ASM) May 2000 meeting describing the RDP-II and some historical aspects of the RDP and RDP-II rRNA sequence data;

(ii) a PDF file of the new backbone phylogenetic tree of 217 SSU rRNA prokaryotic sequences;

(iii) a PDF file detailing the diversity found in RDP releases;

(iv) a PDF file of PCA two-dimensional scatter plots for prokaryotic SSU rRNA sequences (figure 5 of the ASM May 2000 poster, above)

ACKNOWLEDGEMENTS

We thank several individuals for their past contributions: Robin Gutell (and his colleagues), Niels Larsen, Tom Macke, Michael J. McCaughey, Ross Overbeek, Sakti Pramanik, Mitch L. 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., Lilburn,T.G., Parker,C.T.,Jr, Saxman,P.R., Stredwick,J.M., Garrity,G.M., Li,B., Olsen,G.J., Pramanik,S., Schmidt,T.M. and Tiedje,J.M. (2000) The RDP (Ribosomal Database Project) continues. *Nucleic Acids Res.*, **28**, 173–174.
2. Benson,D.A., Karsch-Mizrachi,I., Lipman,D.J., Ostell,J., Rapp,B.A. and Wheeler,D.L. (2000) GenBank. *Nucleic Acids Res.*, **28**, 15–18.
3. Baker,W., van den Broek,A., Camon,E., Hingamp,P., Sterk,P., Stoesser,G. and Tuli,M.A. (2000) The EMBL Nucleotide Sequence Database. *Nucleic Acids Res.*, **28**, 19–23.
4. Tateno,Y., Miyazaki,S., Ota,M., Sugawara,H. and Gojobori,T. (2000) DNA Data Bank of Japan (DDBJ) in collaboration with mass sequencing teams. *Nucleic Acids Res.*, **28**, 24–26.
5. Bruno,W.J., Socci,N.D. and Halpern,A.L. (2000) Weighted Neighbor Joining: a likelihood-based approach to distance-based phylogeny reconstruction. *Mol. Biol. Evol.*, **17**, 189–197.
6. Ellis,L.B.M., Hershberger,C.D. and Wackett,L.P. (2000) The University of Minnesota Biocatalysis/Biodegradation Database: microorganisms, genomics and prediction. *Nucleic Acids Res.*, **28**, 377–379.
7. Klappenbach,J.A., Saxman,P.R., Cole,J.R. and Schmidt,T.A. (2001) rrndb: the ribosomal RNA operon copy number database. *Nucleic Acids Res.*, **29**, 181–184.
8. Marsh,T.L., Saxman,P., Cole,J. and Tiedje,J. (2000) Terminal restriction fragment length polymorphism analysis program, a web-based research tool for microbial community analysis. *Appl. Environ. Microbiol.*, **66**, 3616–3620.