

The Ribosomal Database Project: Providing an Evolutionary Framework.

James R. Cole¹, Bonnie L. Maidak¹, Timothy G. Lilburn¹,
Charles T. Parker^{1,2}, Bing Li², George M. Garrity¹, Sakti Pramanik^{1,2},
Thomas M. Schmidt¹, and James M. Tiedje¹. Center for Microbial Ecology¹
and Dept. of Computer Science and Engineering²

The Ribosomal Database Project – II (RDP-II) provides rRNA related data and tools important for researchers from a number of fields. These RDP-II products are widely used in molecular phylogeny and evolutionary biology, microbial ecology, bacterial identification, characterizing microbial populations, and in understanding the diversity of life. As a value-added database, RDP-II offers aligned and annotated rRNA sequence data, analysis services, and phylogenetic inferences derived from these data to the research community. These services are available through the RDP-II website (<http://www.cme.msu.edu/RDP/>).

In December 1997, the RDP officially moved to The Center for Microbial Ecology at Michigan State University from its previous home at The University of Illinois. A new, greatly enhanced website and a major data update (version 7) were released on July 31, 1998. The new release contains 9835 aligned sequences, an increase of 66% over the previous release (June 1997). In addition, this is the first release to be generated from a new, custom dbms that provides the user with better, more consistent formatting of data within sequence records, and of shared data (eg. reference data) between records.

The new RDP-II web site offers a significant improvement over its progenitor. It exhibits a new, clean, easy to understand user interface, enhanced user data input, and improved, more informative output. Improved functions are provided for comparing a user submitted sequence to the RDP database (Sequence Match), aligning a user sequence against the nearest RDP sequence (Sequence Align), examining probe and primer specificity (Probe Match), testing for chimeric sequences (Chimera), and selecting a subset of an RDP alignment (Sequence Selection). In addition, we offer several new functions, including a similarity matrix generator (Distance Matrix), a T-RFLP analyzer (T-RFLP), and a java-based phylogenetic tree browser (Sub Trees). In the first full month of operation (August '98) the web site handled 23,032 requests from 1399 distinct hosts in 40 different countries.

RDP-II has interests in several areas of bioinformatics. We are currently involved in collaborations with researchers at Lawrence Berkeley National Laboratory to develop more practical methods for large scale phylogenetic reconstruction, and methods involving stochastic context-free grammars for incorporating secondary structure and other higher order information into multiple sequence alignments. In addition, we have interests in sequence comparison methods, sequence quality measures, and data visualization methods, particularly for large sequence alignments and phylogenetic trees.