

86. RDP: Data and Tools for Studying Structure and Function of Microbial Communities

Benli Chai* (chaibenl@msu.edu), Qiong Wang, Yanni Sun, C. Titus Brown, James M. Tiedje and James R. Cole

Michigan State University, East Lansing

<http://rdp.cme.msu.edu> <http://fungene.cme.msu.edu>

Project Goals: RDP offers aligned and annotated rRNA and important ecofunctional gene sequences with related analysis services to the research community. These services help researchers with the discovery and characterization of microbes important to bioenergy production, biogeochemical cycles, climate change, greenhouse gas production, and environmental bioremediation.

Over the past year, the RDP websites (Cole et al., 2014) were visited, on average, by 10,908 researchers (unique IP) in 21,842 analysis sessions each month. In addition to the tools and sequence data offered on our websites most RDP tools are available through our GitHub repository (<https://github.com/rdpstaff/>), with step-by-step instructions for each tool. These command-line RDP tools offer additional options not available on the web and are suitable for high-throughput analysis.

We enhanced our FunGene site (Fish et al., 2013) with new important functional genes. These included additional nitrogen cycle genes such as “clade II” nitrite reductase (atypical NosZ), important for understanding both nitrogen utilization in biofuel crops and greenhouse gas production. Use of the FunGene web increased to an average of 2117 researchers per month in 2878 analysis sessions. During the past year, we updated FunGene data releases five times from searches of the primary sequence databases.

In May 2014, two of our developers traveled to Argonne National Labs for an intensive two-day work session to gain hands-on experience with the new KBase deployment procedures and the new central KBase infrastructure components: Shock, a data management system for storing and sharing large data, and AWE, a workflow management system for job scheduling. While at Argonne, we were able to integrate a synchronous RDP Classifier developmental version into the new KBase infrastructure. This service accepts input files that were uploaded to the Shock server, executes the RDP Classifier command on the server host and returns results back to user.

The RDP Classifier has been extended with new features and training sets for new genes. The Classifier has been enhanced to use gene copy number data to adjust relative taxonomic assignment percentages based on copy number. The precompiled Classifier is trained with the 16S gene copy number data from rrnDB (<http://rrndb.umms.med.umich.edu>). Both the bacteria and archaea 16S rRNA and fungal LSU training sets have been updated in 2014 to incorporate new phyla and increase coverage of basal lineages. In addition to the previous two fungal LSU training sets, we now provide the Classifier trained on two new high-quality fungal ITS taxonomies: Warcup, a hand-curated set with mycological taxonomic names; and UNITE, grouped by “species hypothesis” accession codes. Both ITS training sets achieved high classification accuracy (85% and 88%) at species level with Warcup performing better at assigning novel sequences when the taxon is not present in the training set.

To demonstrate the utility of command-line RDP tools, we developed a detailed tutorial to walk researchers through the steps of data analysis using RDP tools and importing results to third-party tools. We included the R/Bioconductor package Phyloseq in our tutorial for its extensive suite of analysis and visualization functions. The tutorial provides both supervised and unsupervised methods with example

workflows and sample scripts.

RDP's mission includes user support; email rdpstaff@msu.edu or call +1(517) 432-4998.

References

1. Cole, J. R., Q. Wang, J. A. Fish, B. Chai, D. M. McGarrell, Y. Sun, C. T. Brown, A. Porras- Alfaro, C. R. Kuske, and J. M. Tiedje. 2014. Ribosomal Database Project: data and tools for high throughput rRNA analysis Nucl. Acids Res. 42(Database issue):D633-D642; doi: 10.1093/nar/gkt1244 [PMID: 24288368]
2. Fish, J. A., B. Chai, Q. Wang, Y. Sun, C. T. Brown, J. M. Tiedje, and J. R. Cole. 2013. FunGene: the Functional Gene Pipeline and Repository. Front. Microbiol. 4:291; doi: 10.3389/fmicb.2013.00291 [PMID: 24101916]

Funding Statement: RDP research was supported by the Office of Science (BER), U.S. Department of Energy, Grant No. DE-FG02-99ER62848; additional funding came from the DOE Biological Systems Research on the Role of Microbial Communities in Carbon Cycling Program (DE- SC0010715) and from the DOE Great Lakes Bioenergy Research Center via Office of Science (BER) U.S. Department of Energy (DE-FC02-07ER64494).