Eukaryotes in the Silva database

Laura Wegener Parfrey
Knight lab, CU Boulder

Pelin Yilmaz – MPI Bremen
Silva in general

- SSU and LSU rDNA only
- Curated based on alignment, sequence quality metrics and length (reference sets greater than 1200 bp)
Silva 108

• Qiime formatted files currently available (from Qiime.org resources)
• Tony Walters implemented 18S tutorial

Processing 18S data

This tutorial explains how to use the QIIME pipeline to analyze 18S, or mixed 18S/16S, data. While most of the steps are identical to the standard 16S pipeline described in the overview tutorial, reference data that include eukaryotic sequences may be required for OTU picking, taxonomic assignments, and template-based alignment building.

To this end, a QIIME-compatible version of the Silva 104 release, which contains data for all three domains of life, is available here: http://www.arb-silva.de/download/archive/qiime1. The Qlime_files_r104.zip archive contains several files that will be utilized in this tutorial, and can be used for analyzing 18S datasets. Notes about how these datasets and tree were constructed are listed at the bottom of this tutorial.

A sample fasta file containing sequences from archaea, bacteria, and eukaryotes can be found in the /qiimeTutorial/18S_tutorial_files/ which will be used along with the Silva reference set in this tutorial.

Each of the following steps are described in this tutorial (the remaining analyses are identical to those described in the general 16S tutorial):
Problems with Silva 108

• About 1/5 listed as uncultured eukaryote
• Taxonomy based on NCBI
• Not standardized for computational analyses
  – Variable numbers of levels/ranks
  – Tony generated RDP formatted taxa maps – but just grab first categories
Eukaryotic Taxonomy Working Group

• Collaboration between Silva ribosomal database (Pelin Yilmaz), ISOP systematics committee and others with computational or taxonomic expertise.

• Goals for the revised classification
  – reflect phylogeny
  – Interface with computational tools
Eukaryotic Taxonomy Working Group

• Revised classification based on ISOP taxonomy (Adl et al 2005 and updates)
  – Implemented in the Silva 111 release

• Other databases (Greengenes and RDP) also committed to incorporating eukaryotes in next release.
Collaboration between Silva ribosomal database (Pelin Yilmaz), ISOP systems committee and others with computational or taxonomic expertise.

Goals for the revised classification – reflect phylogeny – Interface with computational tools

Eukaryotic Taxonomy Working Group

Members of the Working Group
- Pelin Yilmaz, MPI-Bremen (Chair)
- Frank Oliver Glöckner, MPI-Bremen
- Laura Wegener Parfrey, Department of Chemistry and Biochemistry University of Colorado
- Rob Knight, Department of Chemistry and Biochemistry University of Colorado
- James R. Cole, Center for Microbial Ecology, Michigan State University
- Linda Amaral Zettler, Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole
- Susan M. Huse, Josephine Bay Paul Center, Marine Biological Laboratory, Woods Hole
- Alastair Simpson, Department of Biology Dalhousie University
- Colomban de Vargas, Station biologique de Roscoff, CNRS, France
- Alexandra Z. Worden, Monterey Bay Aquarium Research Institute
- Brian Leander, Departments of Zoology and Botany The University of British Columbia
- Chris Lane, Department of Biological Sciences, University of Rhode Island
- Daniel McDonald, Department of Chemistry and Biochemistry University of Colorado
- George M. Garrity, Department of Microbiology and Molecular Genetics Michigan State University
- Holly Bik, UC Davis Genome Center University of California, Davis
- Jason Stajich, Dept of Plant Pathology & Microbiology University of California
- Naiara Rodríguez-Ezpeleta, AZTI Tecnalia
- Noah Fierer, Department of Ecology & Evolutionary Biology University of Colorado
- Philip Hugenholtz, Australian Centre for Ecogenomics
- Scott Bates, Department of Ecology & Evolutionary Biology University of Colorado
- Sina Adl, College of Agriculture and Bioresources University of Saskatchewan
- Virginia Edgcomb, Woods Hole Oceanographic Institution

silva ©
comprehensive ribosomal RNA databases
Silva 111

• 71787 eukaryotic sequences
• 13582 fungal sequences
• Taxonomy assigned to uncultured sequences according to tree (e.g. uncultured ciliate rather than uncultured eukaryote)
Silva 111 – taxonomy sources

• Deep eukaryotes and protists – ISOP (International Society of Protistologists) classification
The Revised Classification of Eukaryotes

SINA M. ADL, ALASTAIR G. B. SIMPSON, CHRISTOPHER E. LANE, JULIUS LUKEŠ, DAVID BASS, SAMUEL S. BOWSER, MATTHEW W. BROWN, FABIEN BURKI, MICAH DUNTHORN, VLADIMIR HAMPL, AARON HEISS, MONA HOPPENRATH, ENRIQUE LARA, LINE LE GALL, DENIS H. LYNN, HILARY MCMANUS, EDWARD A. D. MITCHELL, SHARON E. MOZLEY-STANBRIDGE, LAURA W. PARFREY, JAN PAWLOWSKI, SONJA RUECKERT, LAURA SHADWICK, CONRAD L. SCHOCH, ALEXEY SMIRNOV, and FREDERICK W. SPIEGEL.

Published by the International Society of Protistologists

DOI: 10.1111/j.1551-7004.2012.00644.x
Summary

451 taxa,
72 lineages
16 genes

Parfrey et al. 2010
Silva 111 – taxonomy sources

- Deep eukaryotes and protists - ISOP classification
- Fungi
  - originally Pelin using Index Fungorum
  - Currently looking into Mycobank
  - Ideal: adopt same taxonomy as ITS database
- Long term: aim to sync taxonomy with other efforts, e.g. Open tree of life project
Silva 111 – taxonomy

- Also released flat file of taxonomy
- Encourage users to use taxonomy of choice
- Customize ranks (for RDP classifier) based on study
Topiary Explorer

Visualizing the high-throughput sequencing data within a phylogenetic context

Developer: Meg Pirrung in Knight lab

http://topiaryexplorer.sourceforge.net/
Environmental sequences colored by clade

Fungi important in most environments.
<table>
<thead>
<tr>
<th>Path</th>
<th>node_name</th>
<th>rank_name</th>
<th>level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Eukaryota</td>
<td>Eukaryota</td>
<td>domain</td>
<td>Level 1</td>
</tr>
<tr>
<td>Eukaryota;Amb-18S-6341</td>
<td>Amb-18S-6341</td>
<td>phylum</td>
<td>Level 2</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa</td>
<td>Amoebozoaa</td>
<td>kingdom</td>
<td>Level 2</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Archamoebae</td>
<td>Archamoebae</td>
<td>phylum</td>
<td>Level 3</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Archamoebae;Entamoebida</td>
<td>Entamoebida</td>
<td>order</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Archamoebae;Entamoebida;Entamoeba</td>
<td>Entamoeba</td>
<td>genus</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Archamoebae;Mastigamoebaeaa</td>
<td>Mastigamoebaeaa</td>
<td>order</td>
<td>Level 5</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Archamoebae;Mastigamoebaeaa;Mastigamoeba</td>
<td>Mastigamoeba</td>
<td>genus</td>
<td>Level 5</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Cavosteliida</td>
<td>Cavosteliida</td>
<td>class</td>
<td>Level 3</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Cavosteliida;Cavostelium</td>
<td>Cavostelium</td>
<td>genus</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Cavosteliida;Cavostelium;MPE1-14</td>
<td>MPE1-14</td>
<td>order</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Cavosteliida;Schizoplasmodiopsis</td>
<td>Schizoplasmodiopsis</td>
<td>genus</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Cavosteliida;Tychosporium</td>
<td>Tychosporium</td>
<td>genus</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Dictyostelia</td>
<td>Dictyostelia</td>
<td>class</td>
<td>Level 3</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Dictyostelia;Acytostelium</td>
<td>Acytostelium</td>
<td>genus</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Dictyostelia;Dictyostelium</td>
<td>Dictyostelium</td>
<td>genus</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Dictyostelia;Dictyostelium;Polysphondylium</td>
<td>Polysphondylium</td>
<td>genus</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Discosea</td>
<td>Discosea</td>
<td>phylum</td>
<td>Level 3</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Discosea;Flabellinia</td>
<td>Flabellinia</td>
<td>class</td>
<td>Level 4</td>
</tr>
<tr>
<td>Eukaryota;Amoebozoaa;Discosea;Flabellinia;Portulacodium</td>
<td>Portulacodium</td>
<td>genus</td>
<td>Level 5</td>
</tr>
</tbody>
</table>
Release of Qiime formatted files

• Later this fall
• 99% and 97% representative sequences
  – Aligned and unaligned
  – Genbank accession is the identifier
• Taxa map
  – Full
  – RDP formatted
• Tree for eukaryotes only
How to integrate ITS and 18S based classification / databases?
Acknowledgements

Collaborators:
Rob Knight (CU)
Pelin Yilmaz (Silva)
Frank Oliver Glockner (Silva)

Knight Lab:
Tony Walters
Matt Gebert
Chris Lauber
Daniel McDonald
Jose C Clemente
Scott Bates
Jessica Metcalf