Microbiology Activity #11 - Analysis of 16S rRNA sequence data

In sexually reproducing organisms, species are defined by the ability to produce fertile offspring. In bacteria, species are defined by several factors including their 16S rRNA sequence. Each species of bacterium has an official strain called a "Type strain" that is used for reference. If the rRNA sequence of a bacterial isolate is less that 98.7% identical to the most similar "Type strain", the organism is a new species!

A. Identification of organisms with EzTaxon

1. Download **your** sequence data files and other files from this week's section on the class moodle site – I recommend that you create a folder called "rRNA" in your network space for these files.

2. Open the **NotePad** program (search in Apps – press windows key), then open the sequence data file ending in "**trimmed.seq**" (File menu, Open, then change "files of type box" to all files). On the format menu, check word wrap. Click & drag to select the entire sequence (**not** the header) press **CTRL+C** to copy your selection

3. Visit the web site: http://eztaxon-e.ezbiocloud.net/

   The EzTaxon.org server contains a manually curated database of type strains of prokaryotes and provides identification tools using similarity-based search. The current version has advanced tools and an up-to-date database for your identification jobs.

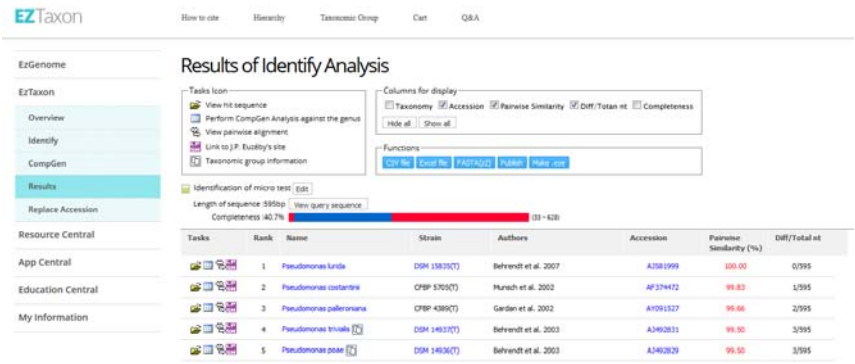
   When you use the EzTaxon server for your study, please cite the following article:

   > Kim, O.S., Cho, Y.J., Lee, K., Yoon, S.H., Kim, M., Na, H., Park, S.C., Jeon, Y.S., Lee, J.H., Yi, H., Won, S., Chun, J. (2012). Introducing EzTaxon-e: a prokaryotic 16S rRNA Gene sequence database with phylotypes that represent uncultured species. Int J Syst Evol Microbiol 62, 716–721 PubMed IJSEM

   This site can be used to search an exceptionally well-curated database of 16S rRNA sequences from bacterial "type strains". Type strains define a species and are used as a reference for comparison with new strains. These strains are generally accepted as "final" for the species and have been stored in major culture collections such as the American Type Culture Collection (ATCC).

4. If this is your first visit to the EzTaxon-e website, click the **sign-up** link, enter the requested information to register. After registering login to the site (if problems use newmanlab@gmail.com, pw 16srrna1).

5. Click on the Identify box on the left side of the window. Paste your sequence into the large sequence entry box, enter UK-XXX or K-XXX to the "Name of sequence" box and click "Identify"



6. On the results page, click on the link in the Query box. Scroll down to view the full list of high scores. Record the highest scoring organism below.

	Name	Accession	Pairwise similarity	Diff/Total nt
UK-XXX				
K-XXX				

7. Press CTRL+ several times to maximize the size of the view of the results page. Press the print screen key on the keyboard to capture a screen grab, open a Word document and paste the screengrab for inclusion in your lab report.

8. Open a new browser tab and go to this website: http://www.ncbi.nlm.nih.gov/blast/bl2seq/wblast2.cgi This is the **BLAST2** page that will compare the sequence from your unknown organism and the type strain(s) that had the highest similarity.

9. On the BLAST2 page, click the **browse button in the sequence 1** section and select the **sequence file from your rRNA folder** for uploading. Then type or paste the **accession number** from the top scoring organism identified by EzTaxon into the large search box in the **Sequence 2 section**. Click the **BLAST** button.

10. To save the alignment as a graphic, optimize the display of the alignment in your window and press the **"print screen" key**. Paste the captured image into a Word document and save with an appropriate name in your rRNA folder.

11. Return to the **BLAST2 result page and click on the link to the accession number**.

12. Record the phylogeny (Phylum, Class, Order, Family) of your organism (listed below the organism name) here

Known = _____

Unknown = _____

B. Literature research: Retrieve information about your organism.

1. Look for the link on the **Pubmed line** of the database entry. **Click it**. If the free full text is available for the original paper, **click and download the PDF file,** save it in your **rRNA folder using the organism name as the filename.**

2. If the paper for your organism does not have a free full text, **copy and paste the abstract and full citation** into a Word document and save the file with an appropriate name..

3. Click the organism name link in the GenBank record to go to the taxonomy browser. Retrieve any papers available in the comments and references section. Also, click the PubMed Central Link in the box in the upper right to identify relevant research articles. Alternatively, you may go back to the main NCBI website and click on PubMed. Type in the name of your organism in the search box and click Go.

4. Browse through the search results. Click on the links for articles that look interesting, read the abstracts, then if the paper seems useful, attempt to obtain the full text article in PDF format. If the full text article is not available online, print or save the abstract. If, after retrieving additional information, you find that articles not available online are still necessary, you can work with the librarians to obtain the article via document delivery.

5. You should also look up information about your organisms in your textbook and in the Bergey's Manuals; Determinative and **Systematic**. Compile published biochemical test information regarding your known and unknown organisms (file is available on the class moodle website). Do YOUR results match? If not, you MUST redo those tests next week to determine whether there was an error in your test results or if your strain really has different characteristics from the published strain.

6. Repeat the processes outlined above for the known organism

C. Download and analyze most closely related genome

1. Go to the NCBI Genome page at http://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome

2. Paste 16S sequence into the Query Sequence box

3. In the "Search Set" field, choose the "Whole genome shotgun contigs (wgs) database; select "limit by" organism, and enter the phylum of your organism; Be sure that the Program for "Optimized for" highly similar sequences (megablast) is selected.

4. Click on the Accession number on the right side of the results table, then click the download tab, and download the Genbank file to your network space. After it downloads, click run to unzip the compressed file.

5. Go to the Rapid Annotation with Subsystem Technology website (RAST) (Firefox preferred). Login with UserID: newmanlab pw:16srrna1

6. On the "Your Jobs" menu, choose "upload a new job", then upload the GenBank file to RAST, preserving NCBI Gene calls when asked.

7. Repeat the procedure above for your other sequence.

D. Construction of a Phylogenetic Tree

1. Click on the "rRNA seq for alignment.fasta" file on the class moodle site and save it to your rRNA folder.

2. Use Notepad to open one *.trimmed.seq file, edit it to get it into fasta format (use Genus species XXX for the name of your organism, where XXX corresponds to your initials or the previous strain designation). Add the name *novum* if your sequence is <98.5% identical to the closest type strain. CTRL+C to copy the sequence to the clipboard.

   >Genus species novum XXX
   AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

3. Open the other *.trimmed.seq file from the last lab meeting in Notepad, edit it to get it into fasta format (use Genus species YYY for the name of your organism, where YYY corresponds to your initials or the previous strain designation). CTRL+V to paste the previous sequence onto the end. CRTL+C to copy the combined sequences to the clipboard.

   >Genus_species_YYY
   AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA
   >Genus_species_novum XXX
   AAAAAAAAAAAAAAAAAAAAAAAAAAAAAA

4. Open the "rRNA seq for alignment.fas" file in your rRNA folder, CTRL+V to paste the sequences copied in the previous step, onto the top of the file. **Save** the file.

E. Perform multiple sequence alignment and create tree

1. Open the **MEGA6** program (under course programs, biology)

2. Click the **Align** Button and choose **Edit/Build Alignment**. *Choose* Retrieve Sequences from File, then select the **rRNA seq for alignment.fas** file from your network space.

3. On the **Alignment** menu, choose **Align by ClustalW** , then click OK to select all. And click OK to keep the default settings.

4. Save the current alignment session by going to the **Data** menu and choosing **Save Session**. This will allow the current alignment session to be restored for future editing.

5. On the **data** menu choose **export alignment** and **MEGA** format. Accept the default file name and click **save**. An input box will appear asking for a title for the data. Enter "rRNA Aligned by MEGA" as the title, and click the "OK" button. Another dialog box will appear asking you if the sequence data is protein coding. In this case, click "No."

6. On the **original MEGA** window, click the **Phylogeny** button and choose **Construct Test Neighbor-Joining tree**. Select the .meg file you saved in the previous step and click open.

7. In the Analysis preferences box, on the Phylogeny test entry, select **bootstrap method** and set to **1000** replications. This will provide information on the reproducibility of a particular branch. In the model/method section, choose Kimura 2-Parameter.

8. Click **Compute** to accept the defaults for the rest of the options and begin the computations. A progress indicator will appear briefly before the tree displays in the Tree Explorer.

9. On the **file** menu, choose **save** to save the current tree session. On the **image** menu, choose **copy to clipboard**. Paste the tree a Word document.

10. Return to the MEGA **Tree Explorer** page, click on **caption**, then copy and paste the caption and references onto Word document as well.

11. Click the distance button on the main Mega6 window, then compute pairwise distances. Click yes to use the currently active alignment.

12. Change Model/Method to P-distance, then click compute.

13. Move the rows into the same order that they are on the tree

14. Click the “XL” button, for the export type, choose matrix, then click print/save matrix.

15. Right click on column B, choose insert. Right click on row 1, choose insert.

16. Enter the number 1 into cell B1. Select that cell, then click and drag the lower right “handle” down to fill the series. Change the format to number with no decimals. Select the numbers and copy/transpose to cell C1.

17. Click and drag to select the data, on the home menu, choose conditional formatting, color scales, and red-green gradient to easily visualize differences.

18. Save all of your files as you close down the programs!