**How to Find Ribosomal RNA (rRNA) genes?**

By: Laura McCoy

Ribosomes are extremely important in both eukaryotic and prokaryotic cells because they translate messenger RNA to form proteins. The similarities in the process of translation among all organisms allow the genes that encode for rRNA, one of the main components of ribosomes, to be highly conserved from species to species. Therefore, the search for these genes begins by finding rRNA genes in closely related organisms.

**What genes are we looking for?**

Three types of rRNA are found in prokaryotes: 16S, 23S, 5S. Four types of rRNA are found in eukaryotes: 18S, 5S, 5.8S, and 28S. Each of these rRNAs is coded for by a different gene. Many copies—hundreds to thousands—are found in an organism’s genome, ensuring that a large amount of rRNA is produced by the cell.

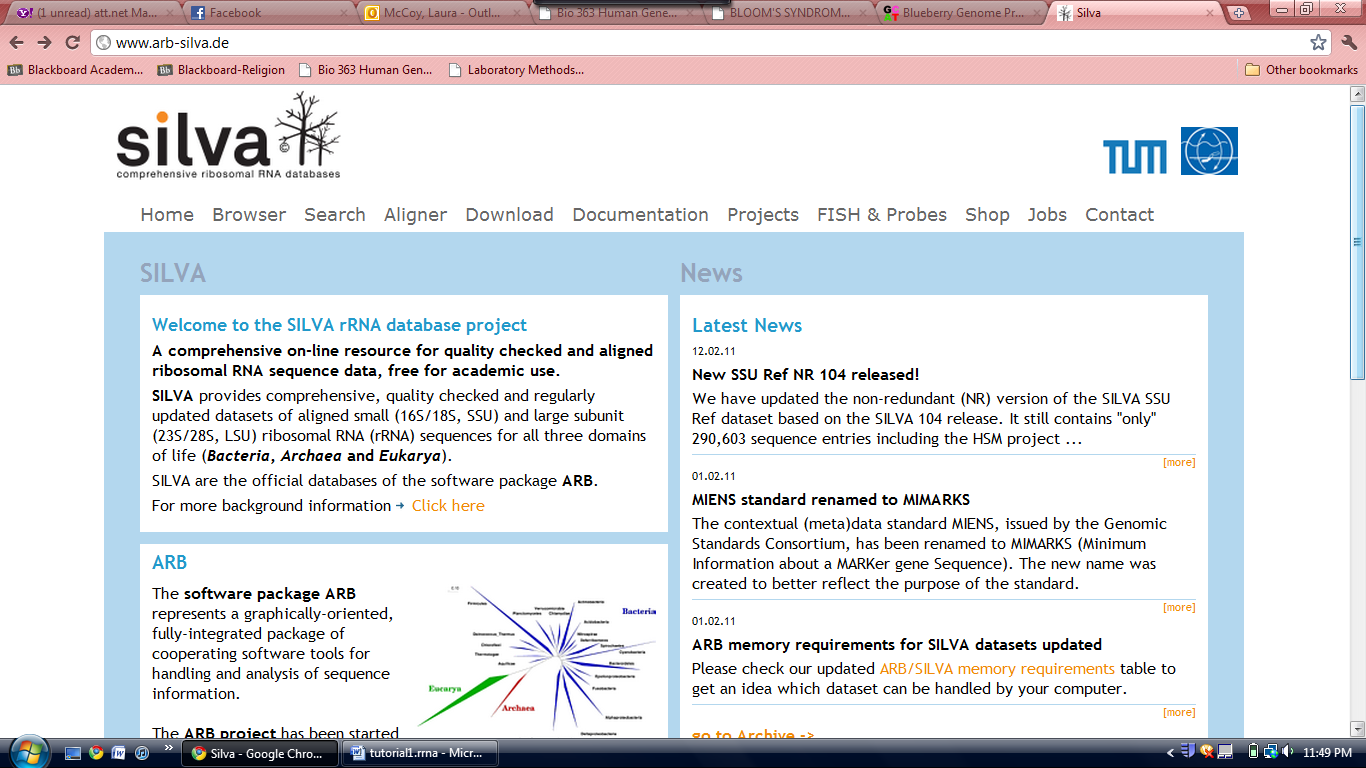
**Steps to finding rRNA genes:**

In this tutorial, rRNA genes in the blueberry (Vaccinium corymbosum) genome will be determined using orthologs from the grape (Vitis vinifera) and strawberry (fragaria vesca) genomes, but the technique is applicable to other organisms.

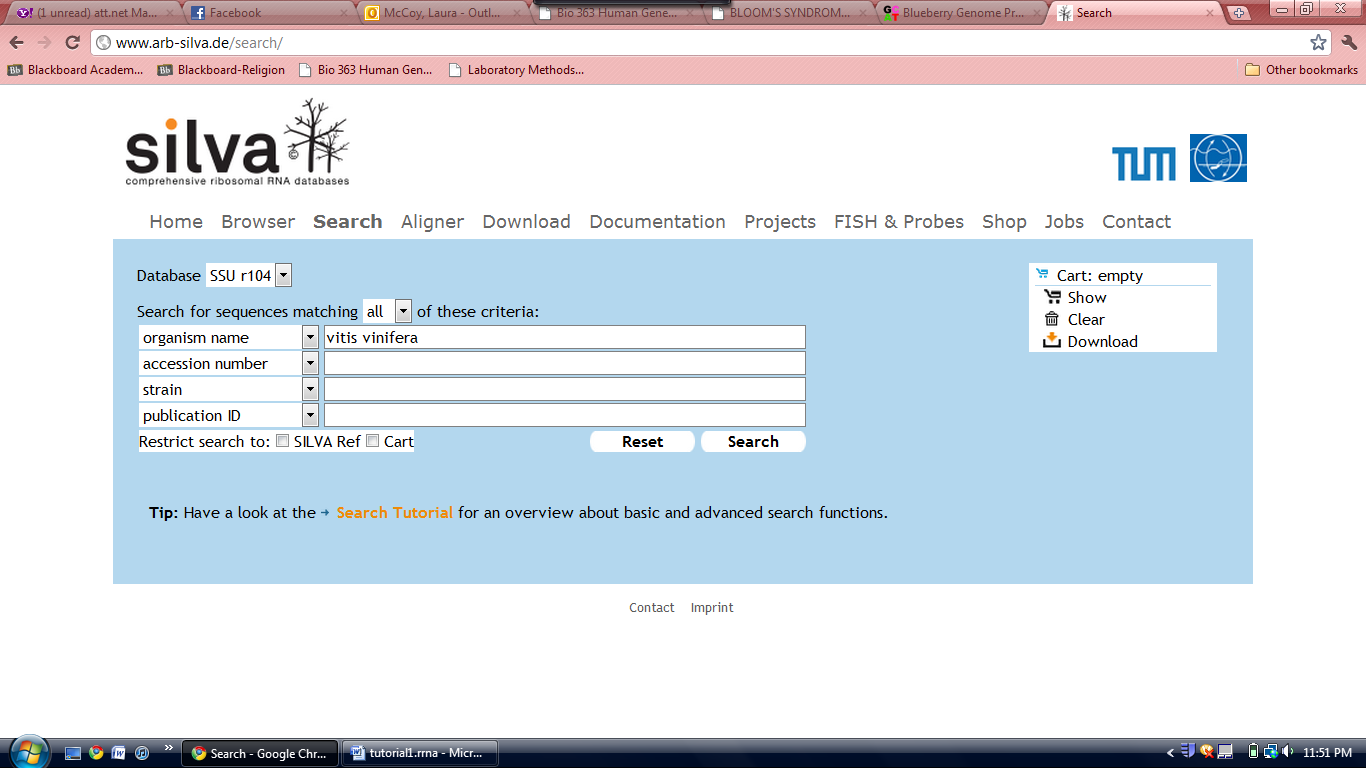
1. **Locate rRNA orthologs**

The best source of rRNA gene sequences is the Silva rRNA database (http://www.arb-silva.de/).

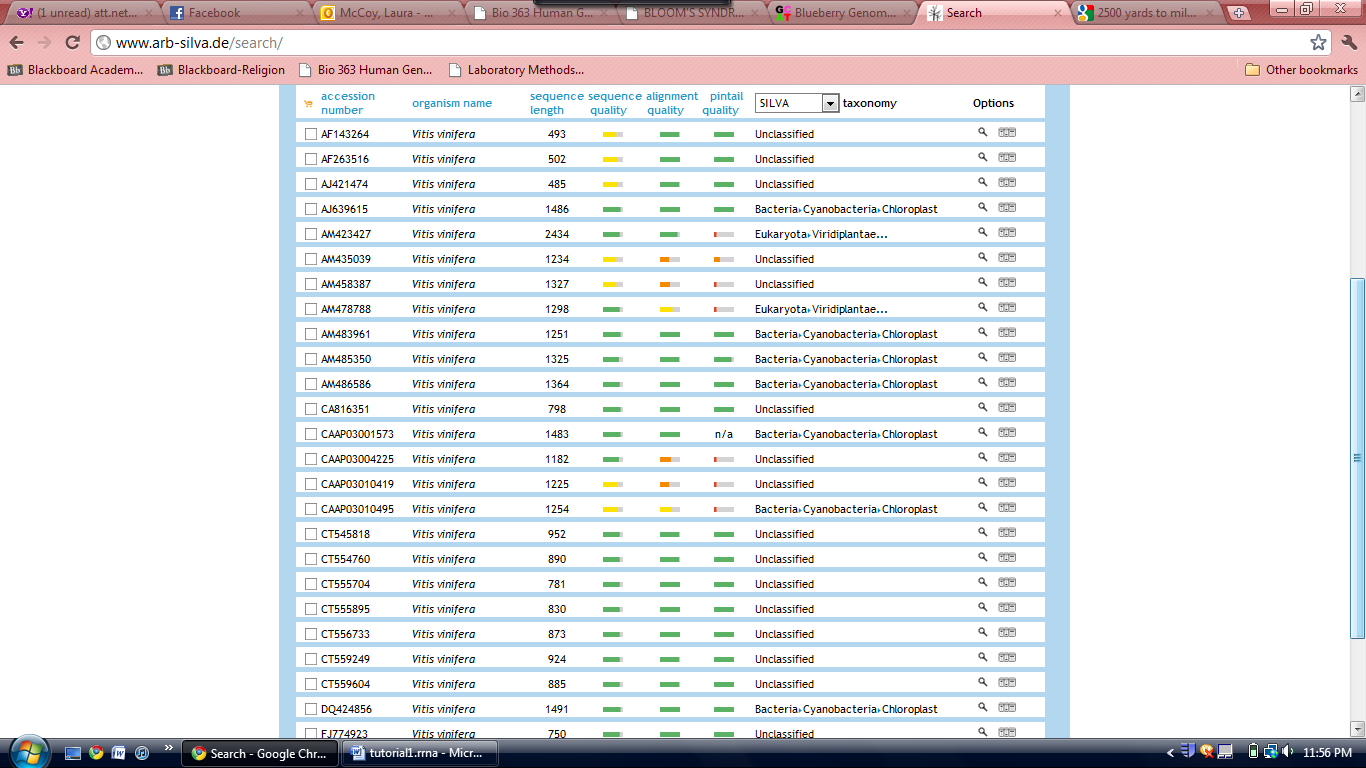
* Select “SEARCH” from the choices at the top of the page.



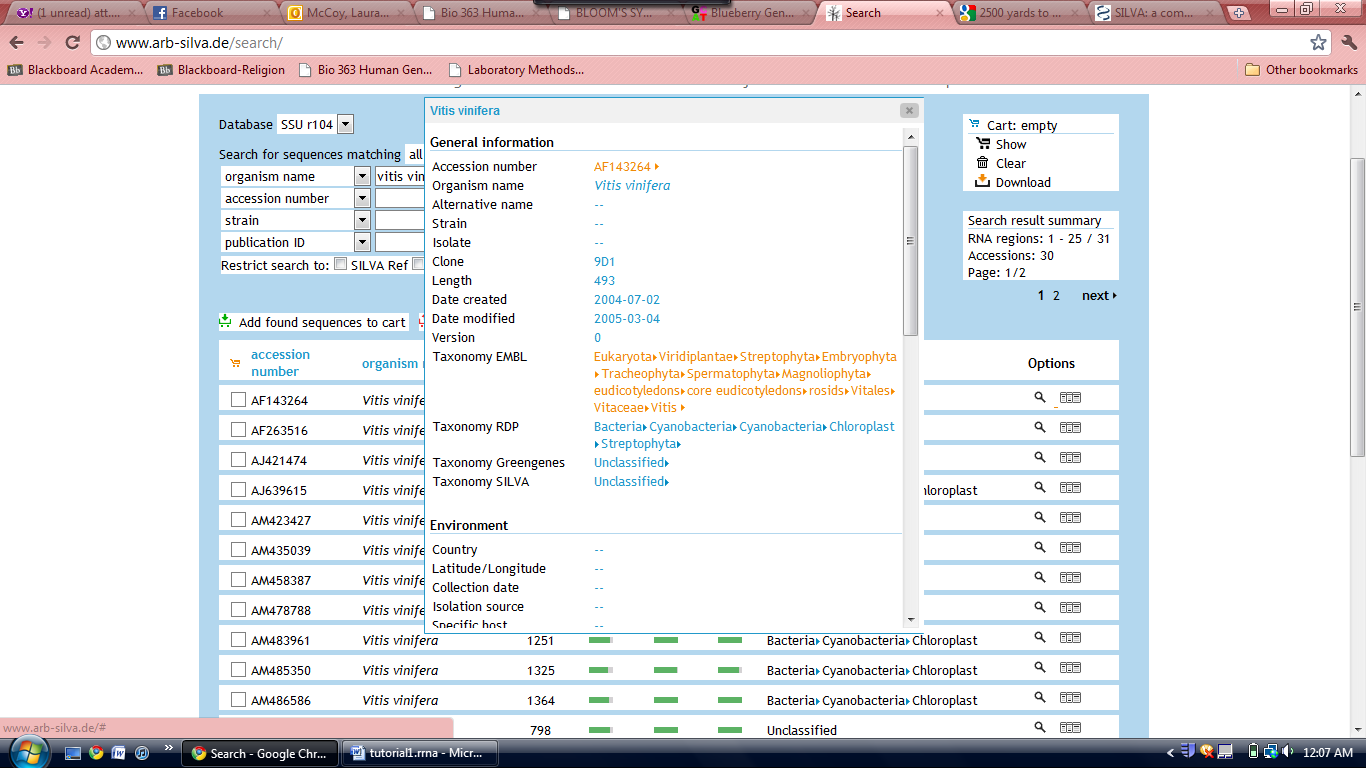
* In the search bar to the right of “organism name”, type in the scientific name of the organism you are searching (“vitis vinifera” in the picture below). The accession number, publication ID, and/or strain can also be entered to narrow your search. Click “Search.”



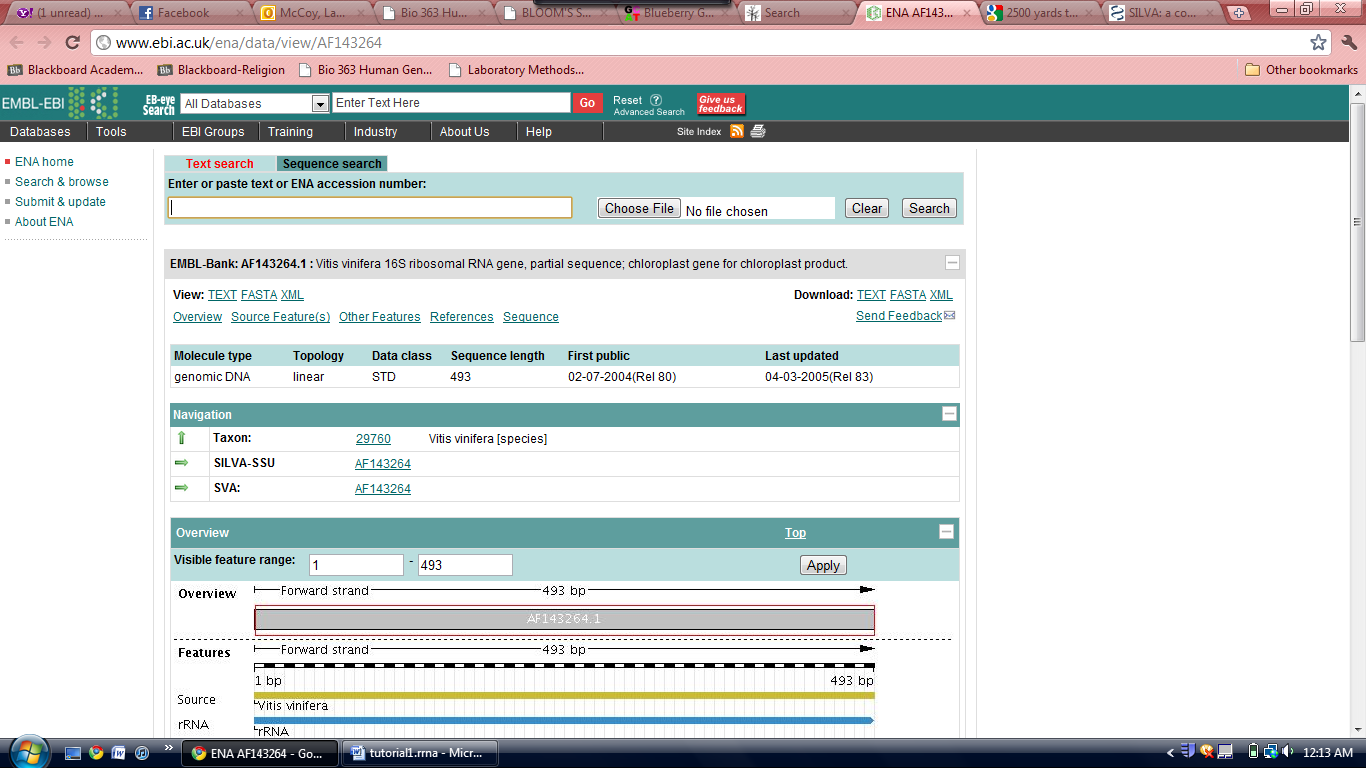
* Examine results. Information about each rRNA gene that has been found in that species is given—accession number, sequence length, sequence quality, alignment quality, and pintail quality. (note: pintail quality is determined by “pintail software” that detects anomalies between sequences.)



* More in depth information can be found for each result by clicking the magnifying glass icon (shown in red above). After clicking this, a window with information for that individual accession number will appear.



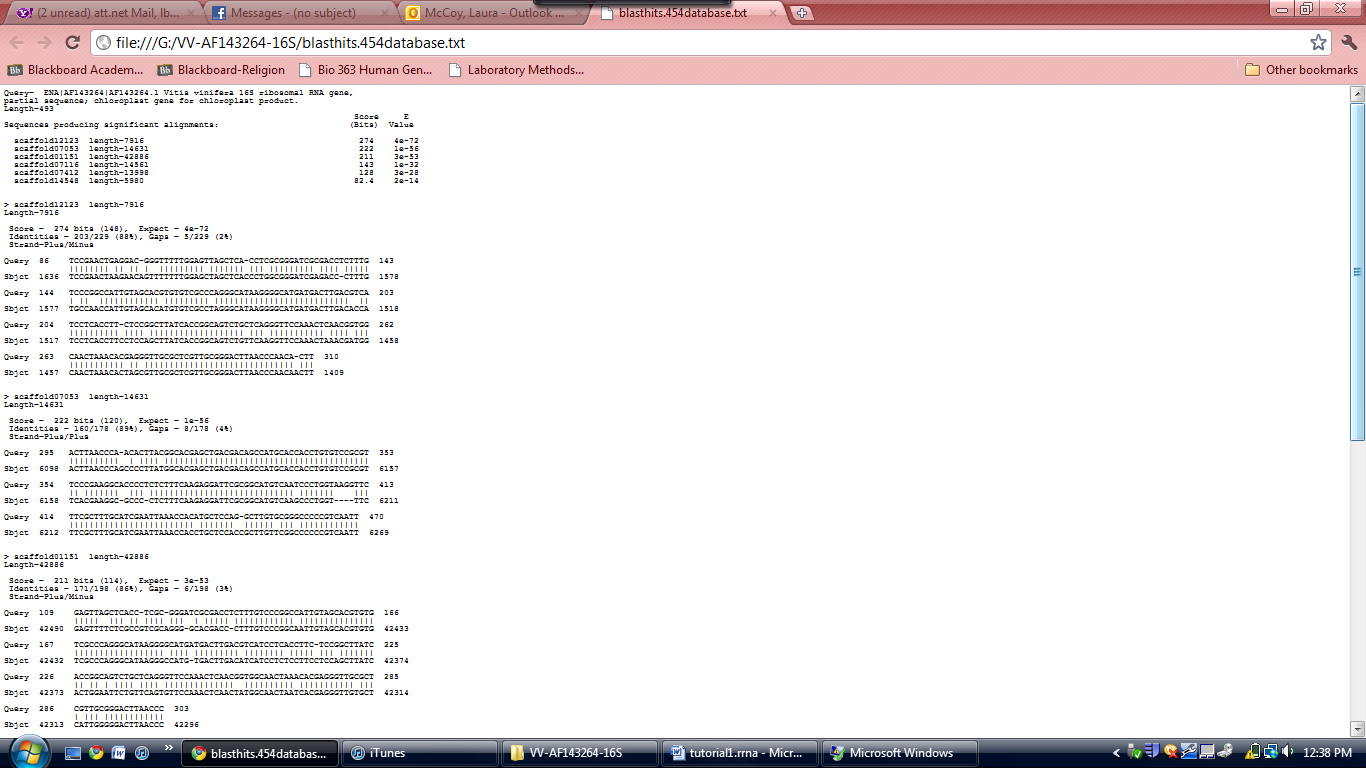
* Click the link to the right of accession number (shown in red above). This directs you to a different website that offers very specific information for that accession number. From this page, the actual sequence (in FASTA format) and the type of rRNA gene, as well as other information, are given.



* Use this method for the remainder of the accession numbers from the search page to find the sequences of different rRNA genes. Be sure to record which type of rRNA gene (18S, 5S, 5.8S, 28S, chloroplast, or mitochondria) each sequence is.
* Repeat this entire process to find the strawberry rRNA genes.

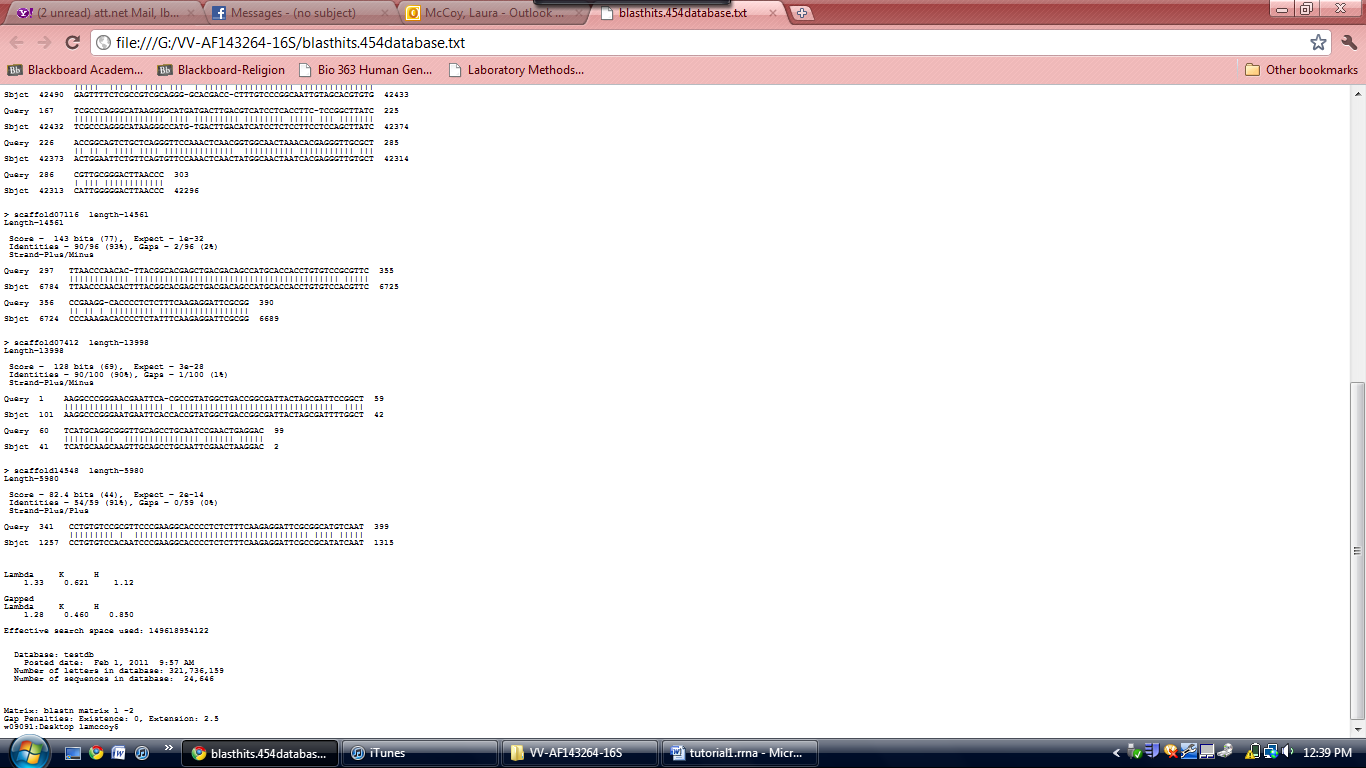
1. **BLAST against blueberry scaffolds**

Once you have the FASTA for a particular rRNA gene, BLAST it against the blueberry genome. The BLASTn version of BLAST can be used locally on your computer. After running the BLAST, the results will tell you if there were any hits. If the ortholog sequence matches up to a sequence in the blueberry genome, then you will get a hit. Running BLAST so that alignments of the ortholog sequence and blueberry sequence appear along with the hits is the easiest way to find out where the hit is in the blueberry genome and what the sequence of that hit is.



Sequence alignments

6 hits found



Repeat this process with each of the orthologs found from the Silva rRNA database. If an ortholog sequence does not give you any hits or if you cannot find an ortholog for a particular gene in the Silva database, then look for rRNA gene sequences from other sources. Many organisms that have been sequenced have their own genome browser, so this is another source of rRNA gene sequences.

1. **Analyze results**

After completing these steps, look at the results of the BLAST as a whole. Count up the number of hits for each type of rRNA gene. Since there are multiple copies of each rRNA gene within a genome, it is not unusual to get many hits for each ortholog. Be sure to look at the E value for each hit and only count the hit if the E value is very close to zero.

**Possible Problems:**

The largest problem that you may have is finding sequences for the orthologs. Not every organism’s genome has been fully sequenced and annotated; therefore, some of the sequences on Silva are either partial sequences or “working draft” sequences. In the above example, the *vitis vinifera* accession number AF143264 is a partial sequence of the 18S rRNA gene. The fact that it is a partial sequence explains why there were only 6 hits. Some sequences from Silva are working draft sequences and do not give the specific rRNA gene for which the sequence corresponds because it is not yet known. An example of this is shown below.

