If there would have been another take home exam, it would have looked like this.

1. True/False: All Orthologs and paralogs can be determined using unrooted trees. (It is synapomorphies and clades that cannot).

2. Which of the following is a PROPER technique to find transferred genes and not just putatively transferred genes?
a. Gene presence absence data for closely related genomes (*if a gene is absent in many deeper branching relatives, then a gene showing up in a single lineage as been created in that lineage, likely through HGT. If you compare to many closely related genomes, the presence absence analysis results is somewhat like a phylogenetic conflict*)
b. Phylogenetic conflict
c. Composition based analyses
d. Taxplot at NCBI
e. All of the above
f. a+b

3. When performing a Taxplot of one bacterial genome as query and one other bacterial genome and one archaeal genome for comparison, what do the hits that are closer to the archaeal genome represent?
A. Horizontal gene transfers
B. Long branch attractions
C. Putative HGTs that MUST be verified, for example by making phylogenetic trees, or a gene presence absence analysis using a larger taxon sampling.
D. Lineage specific gene sorting
E. Errors in gene calling

4. In a gene plot, what do the two axes represent?
a) Number of introns
b) E-values
c) The two genome sequences
d) The number of substitutions
e) None of the above

5. Many papers are published using very few sub-Saharan African sequences or genomes (the Neanderthal paper had 2, the Denisovian paper 3). What does the

saga of Y-chromosome Adam tell us about the advisability of having a larger sample size?
A. It's all cool, as long as you sample hundreds of Europeans and Asians
B. No amount of Europeans and Asians can make up for having a reasonable number of sub-Saharan African samples
C. Analyses with a small sample size may be problematic (depends a little on what one studies)
D. A sample size of less than 40 is plenty

6. It was once thought that Y-Chromosome Adam lived ~40,000 years ago. What happened to this story?
A. Careful study of the Bible revealed…something.
B. Further studied proved that there was a massive migration of men from China into sub-Saharan Africa and this story totally makes sense to someone, despite the lack of fossils showing such a migration.
C. An African American was found whose genome had all of the ancestral states, which finally made people think that it might be a good idea to sequence a few more sub-Saharan Africans. Y-Chromosome Adam now dates back to well over 200,000 years.
D. The genome of the Neanderthals revealed that it really goes back to only 30,000 years.

7. Sub-Saharan Africans account for ____ percentage of human diversity and should therefore account for the same percentage of any sample size when studying the human species?
A. 1%
B. 3%
C. 25%
D. 70%
E. 99%

8. Which part of a circular bacterial genome often is least conserved?
A. The origin
B. The terminus
C. Some other part

9. True/False Supermatrixes are super fun, because if horizontally transferred genes are included together with vertically inherited genes, the resulting tree represents neither the history of the organism not the history of any single gene, but it may do so with high support.

10. True/False When super trees fail, instead of producing bastardized trees, they tend to produce polytomies with low support, which aren't nearly as fun, explaining why supermatrixes are so popular.

11. True/False Super trees are better able to handle HGT, but are slightly more prone to lack of signal.

12. If you want really high support values and don't really care how meaningful they are, which of the following approaches should you use?
A. Mr. Bayes
B. Supermatrix
C. Maximum Likelihood (using approximate Likelihood Ratio Test - aLRT) support values.
D. Super tree analyses
E. Bootstrapping

13. When sequenceing a genome and a ridiculous amount appears to be HGT, what could have happened?
A. All those genes could really be HGTs
B. Someone used a BLAST based approach and didn't bother to verify the HGTs
C. The sample wasn't properly cleaned; the HGTs are contamination
D. The sequence is of the Hologenome, and including many bacterial genomes.
E. All of the above

14. A lesson learned from the genome of the tardigrade is:
A) sequence data is only as good as the specimen
B) secrecy in science detracts from identifying problems early on
C) 1/6 of the tardigrade genome is from HGT
D) A&B

15. You are looking for genes that might have been transferred from Gammaproteobacteria to Prochlorococcus (after they split from Synecchococcus). Which genomes could you use for the query genome, and which for the two reference genomes?
Any of the Prochlorococcus as query and any of the Synecchococcus as the first, and a Gammaproteobacterium (e.g. E. coli) as the second:

Protein homologs in Complete Microbial / Eukaryotic genomes

To compare the similarity of the query genome proteins to different species choose two organisms by Taxonomy id or select them from the menu

Select your query genome

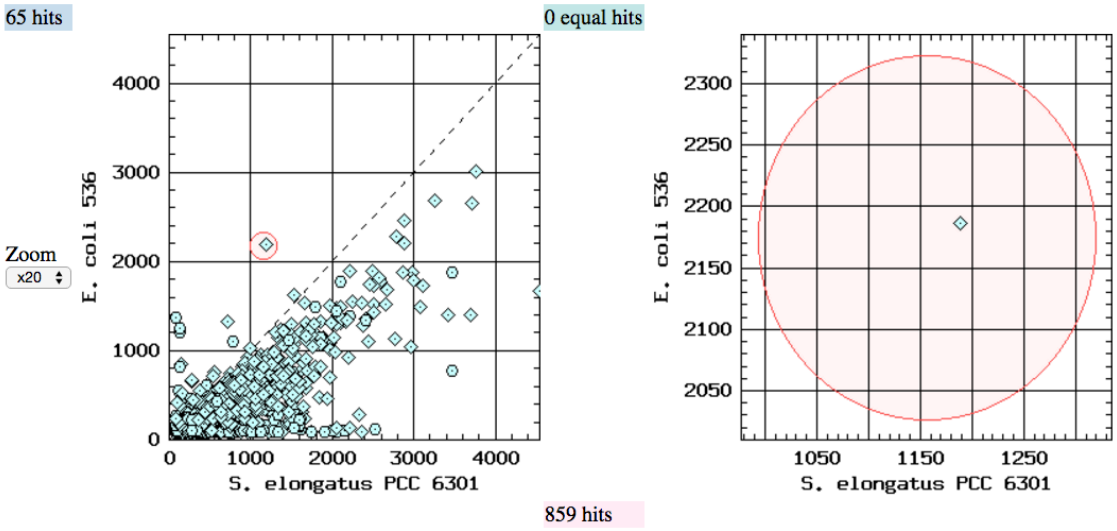
93059 P. marinus str. MIT 9211

Choose two species for comparison

362663 E. coli 536

269084 S. elongatus PCC 6301

Distribution of *P. marinus str. MIT 9211* homologs



1854 query proteins produced **924 hits**, from which 1 is selected .
Each circle represents a single query genome protein, plotted by its BLAST scores to the highest scoring protein from each of the selected organis the homologs in two chosen organisms.

Cutoff: 10 Linear scale compare

Query: compare clear

Blast2Seq	Accession	Blink	Definition
	YP_001550	=>	**threonyl-tRNA synthetase [Prochlorococcus marinus str. MIT 9211]**
2186	YP_669570	=>	threonyl-tRNA synthetase [Escherichia coli 536]
1189	YP_172214	=>	threonyl-tRNA synthetase [Synechococcus elongatus PCC 6301]

16. EMBOSS is a

A.) program package similar to phylip that can be used in phylogenetic reconstruction.

B) program package that contains a diverse collection of programs that can be run from the commandline. This includes many programs useful in molecular biology and bioinformatics.

C) a collection of commandline version of all the programs available at the European Bioinformatics Institute (ebi).