BIO 331

Homework 2

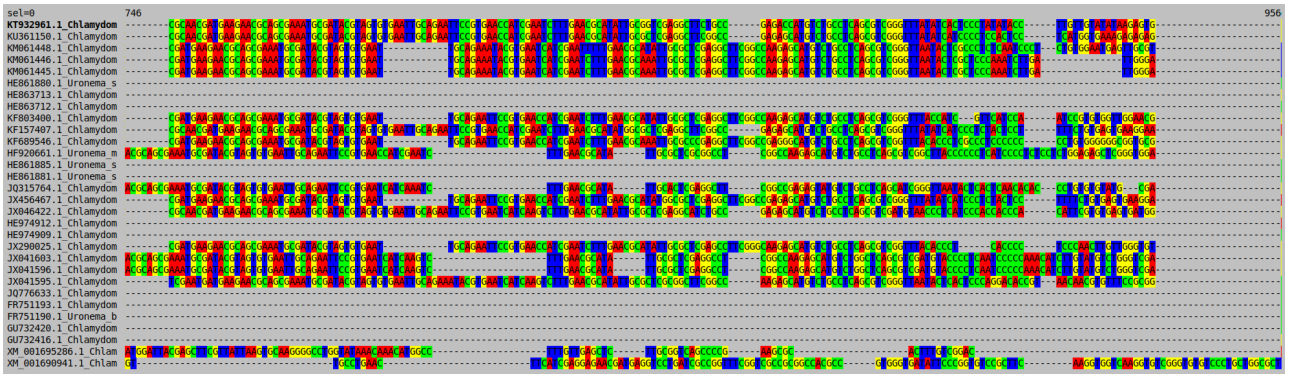
- Section 1
- Section 2

Section 3 - answer all

Questions 1-6 will require that you download the fastA formatted text file "28S uronema Chlamydomonas file1.txt", opening it in SeaView.

The sequences here were obtained from NCBI based on a search using the query term: "uronema OR chlamydomonas AND 28S AND 700:1000[SLEN]". In short, we are trying to study the 28S rRNA gene.

1. Align the sequences in SeaView. Take a screenshot of 2 regions where there is a strong, quality alignment for a good % of the sequences (hint: it won't be all of them in any region). Next, take a screenshot where there seems to be sequence data for most of the represented taxa, but where the alignment seems poor for several sequences. ***6 pts***



1. Do you see any sequences in this file that don't align well at all (i.e. across any part of their length)? A good rule of thumb is that if you are finding only about 25% nucleotide identity between 2 sequences across a decent window length, they are probably not homologous. *This would be considered a bad alignment.* Identify any sequences that seem to never align well. Then search their accession #s in NCBI (i.e. go into their GenBank flatfile pages). Is there information in the flat files that tells you why these sequences did not align well? Explain. ***6 pts***
2. Do a search in NCBI for the *Uronema* sequence under HF920661.1. Extract only the 28S rRNA gene in FastA format (i.e. you don't want any intergenic DNA or DNA from other genes). Paste it in here in FastA format. ***6 pts***
3. Suppose you wanted to extract only the 28S rRNA gene for the following Accession #s: JX041603.1, JX041596.1, KF803400.1, KM061448.1, and KU361150.1. Do these sequences contain only 28S? If so, can you easily extract it (go to the flat files to find out)? Explain. ***6 pts***
4. BLAST is a solution that can help you overcome the problem mentioned in question #2. Use the extracted 28S rRNA sequence from HF920661.1 as the query to find homologous 28S rRNA regions from *Chlamydomonas* using BLAST. Explain how you did this (i.e. which BLAST tool, why?) and how

you limited your findings to *Chlamydomonas*. Describe the total score, query coverage, E-values, and accession #s for the first 5 hits. ***6 pts***