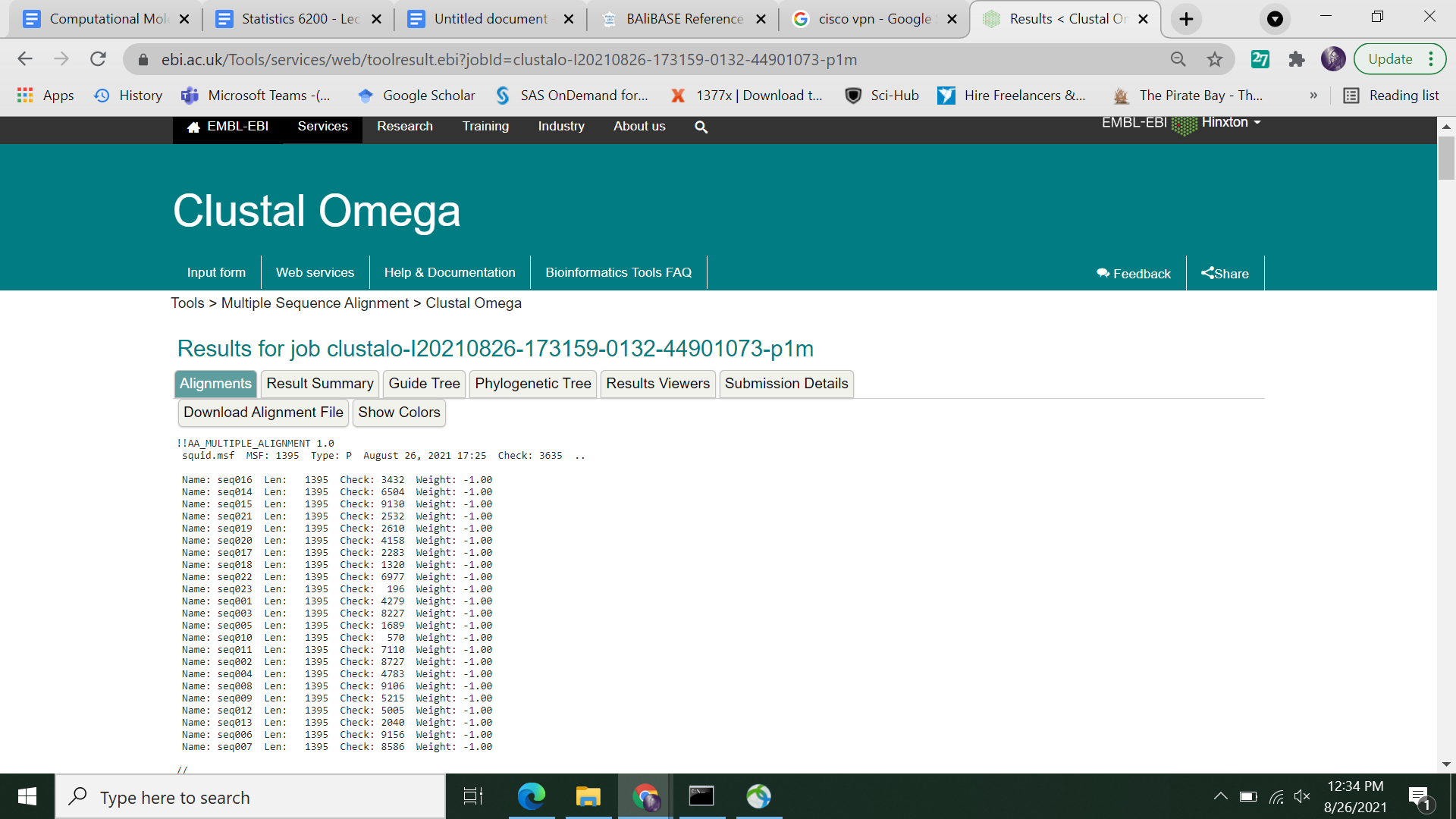
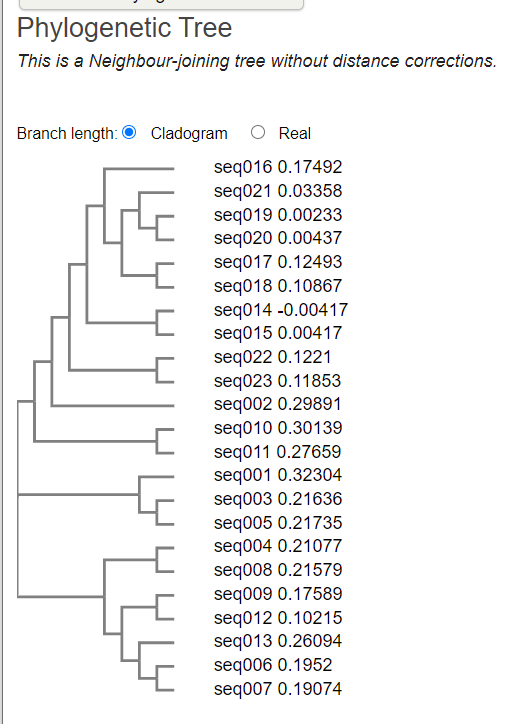
**CME: ASSIGNMENT 1**

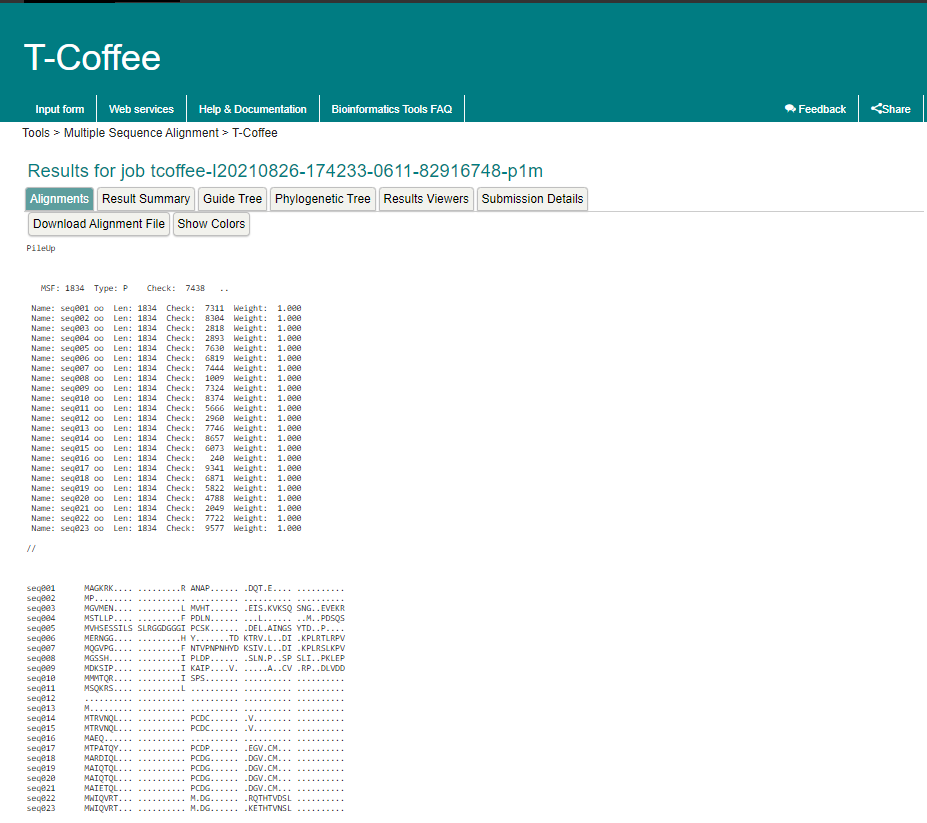
Name: Drashti Mehta; SID: 801262877

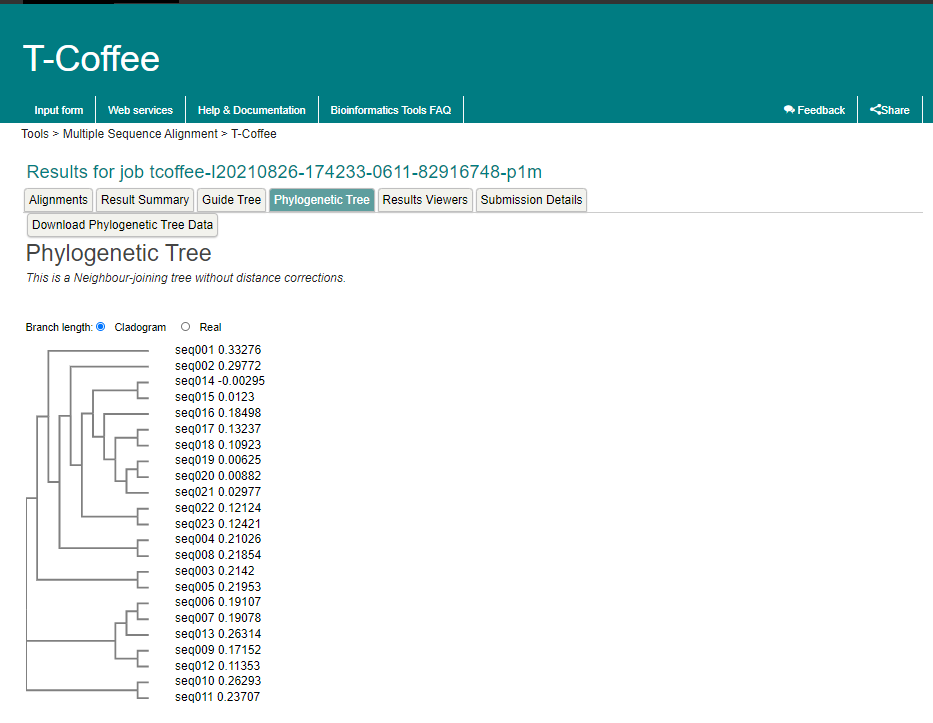
CLUSTAL OMEGA:



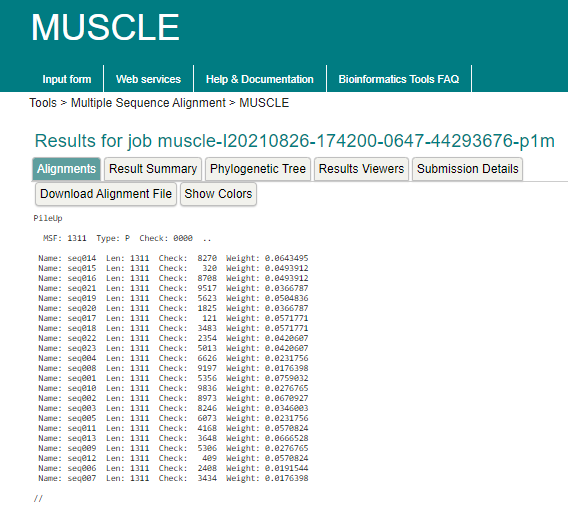


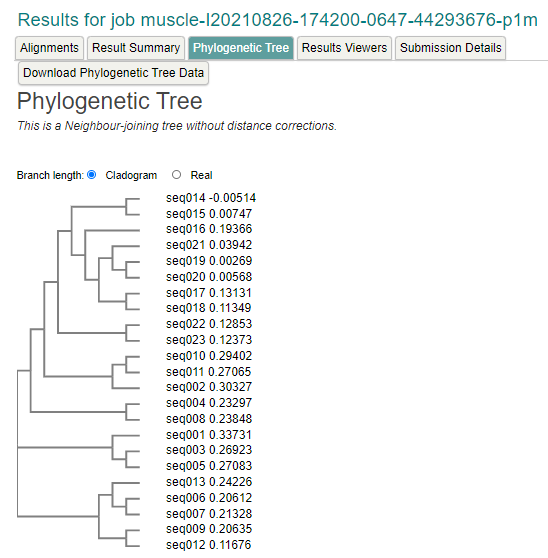
T-Coffee:





MUSCLE:





**Comparing with the ref seq:**

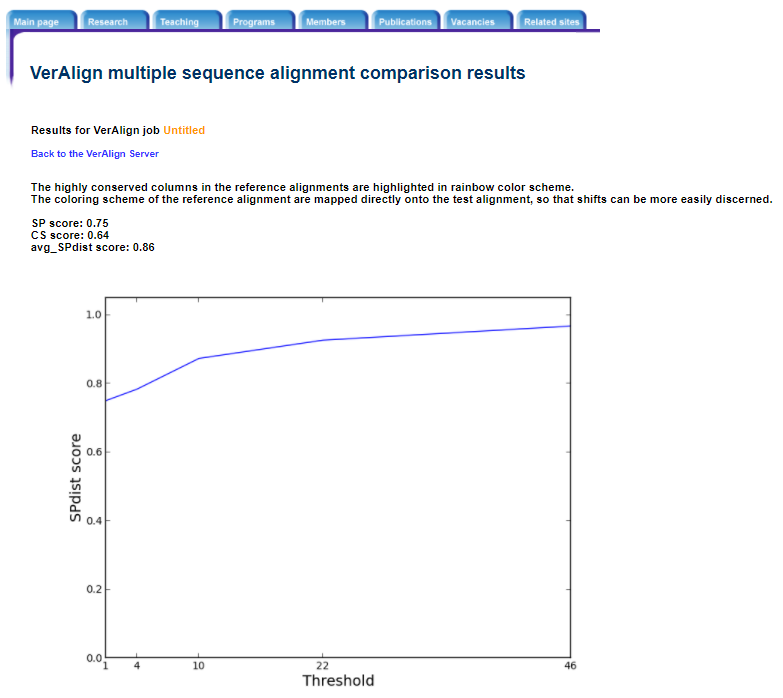


Fig 1a: Clustal W result .fa file alignment with test seq BBA0001.msf

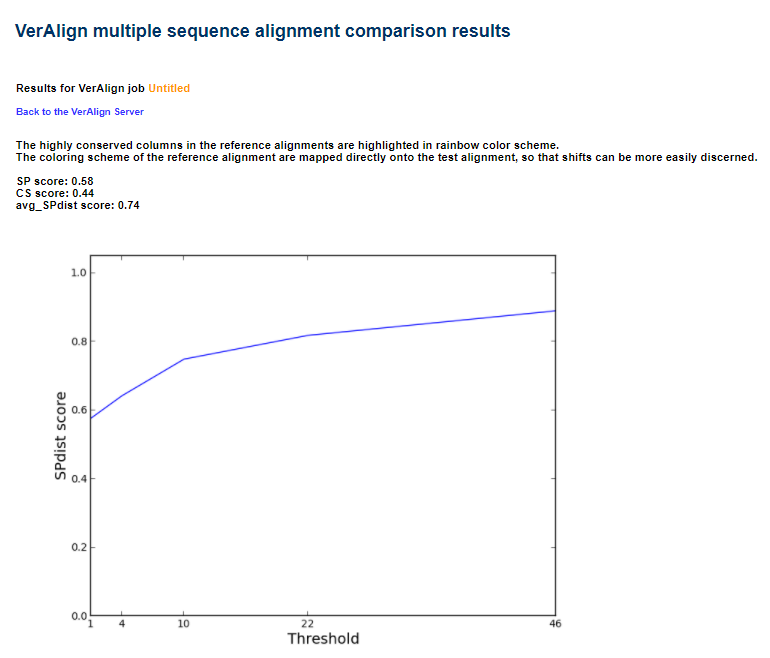


Fig 1b: MUSCLE result .fa file alignment with test seq BBA0001.msf

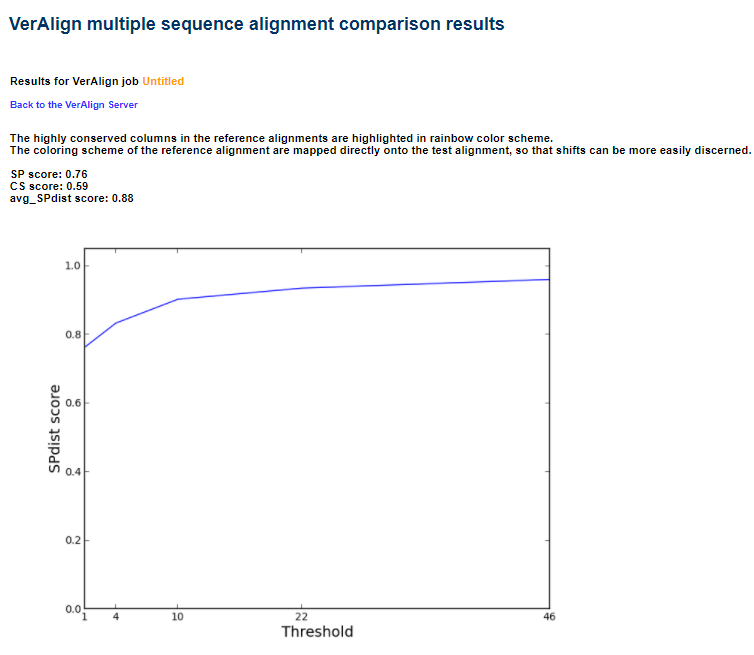


Fig 1c: T-Coffee result .fa file alignment with test seq BBA0001.msf

**ASSIGNMENT QUESTIONS:**

Answers:

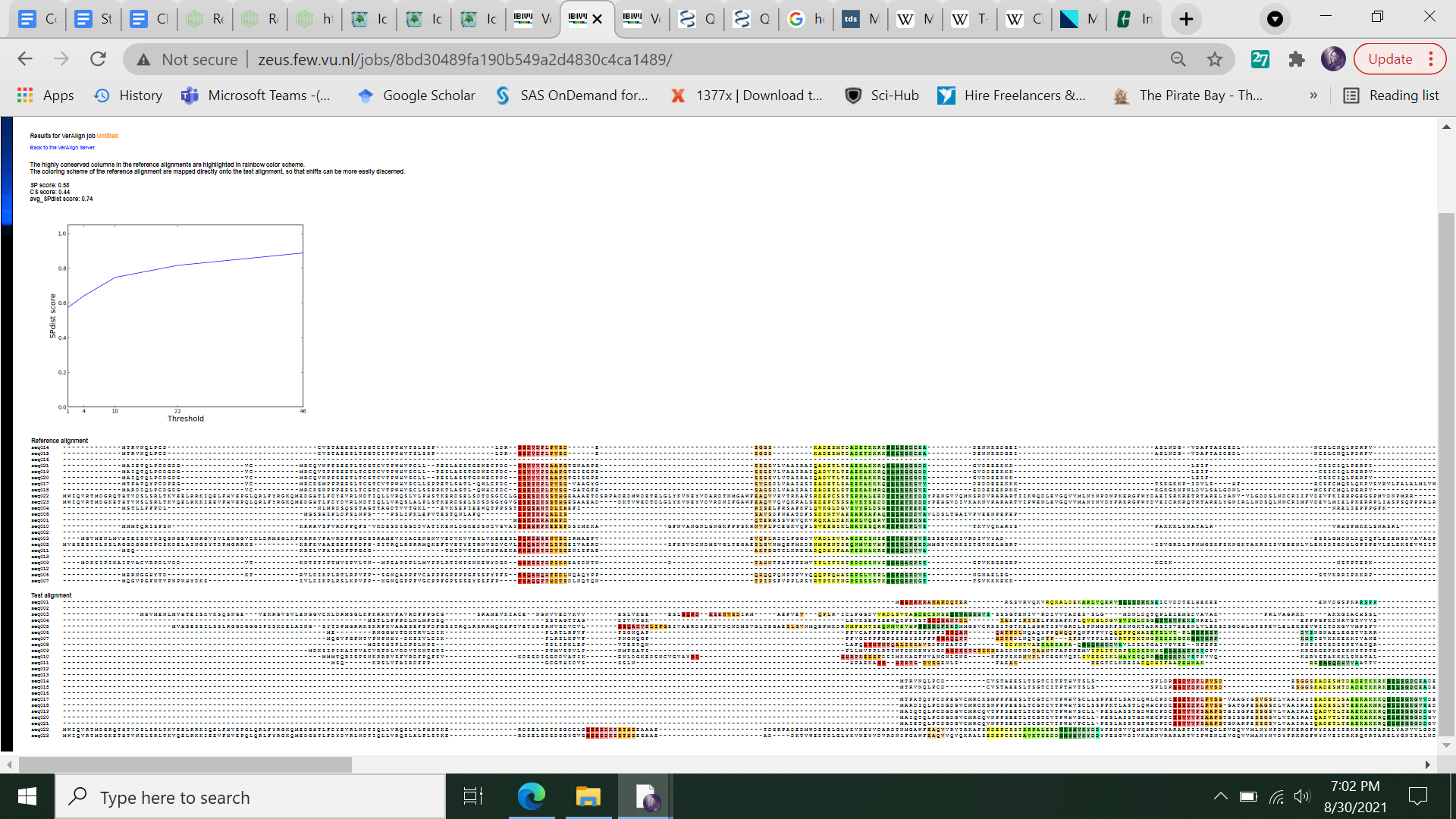
**1. How did the 3 MSA methods compare to each other? Did they all produce the same alignments, or not? Which of the methods performed the best? Which was the worst? Be sure to explain your answer. *5 pts***

A) In my opinion, all the methods have their own pros and cons. To exemplify, Clustal omega is the fastest method but in terms of accuracy T-Coffee is the most accurate but slower than the other methods.

All the 3 methods performed an alignment which was not exactly same to each other but was related to each other. The alignment output of the 3 MSA files was stored in the fasta format and then was compared with the BBA0001.msf file as the reference sequence via the web tool called VerAlign. This helped us to compare the 3 methods with the help of the graphs formed stating the sp\_score, cp\_score and the sp\_dist\_score for analysis.

**2. For the worst method, what did it seem to get wrong about the alignment? Can you describe any general trends when you look at the test alignment versus the reference alignment? For example, are there particular individual sequences that seem very off? Are the sizes and locations of the gaps really different in some way? Is there a single section of the sequence that seems more problematic than the rest? You do not need to list every possible difference-there will likely be way too many-just describe the patterns that stand out to you. *5 pts***

A) As per my observation, MUSCLE seems to be the least accurate method and consumes more time than Clustal - Omega method making itself the worst method when it comes to comparison with T-Coffee and Clustal-O, which is usually not the case. Generally, MUSCLE is used as a replacement to the Clustal - O for accuracy and speed. The CP score, SP score and SPdist score seem to be very less due to shifting in this case.



We can clearly see the shifting of the sequence as the conserved regions are highlighted by the rainbow colouring scheme and they don’t align exactly in the test alignment like that in the reference making the CP score to be lower and because of the shifting, the SPdist score is also lower.

**3. If you compare the resulting phylogenies from the best and the worst method, do they look different? What are some of the most noticeable changes? You should include pictures of each phylogeny with your answer (and label which phylogeny came from each method). *5 pts***

A) The arrangement of the sequences in the phylogenetic tree is different in all the methods. Although the grouping of the sequences is the same, the way of placing the sequences varies in all the methods.

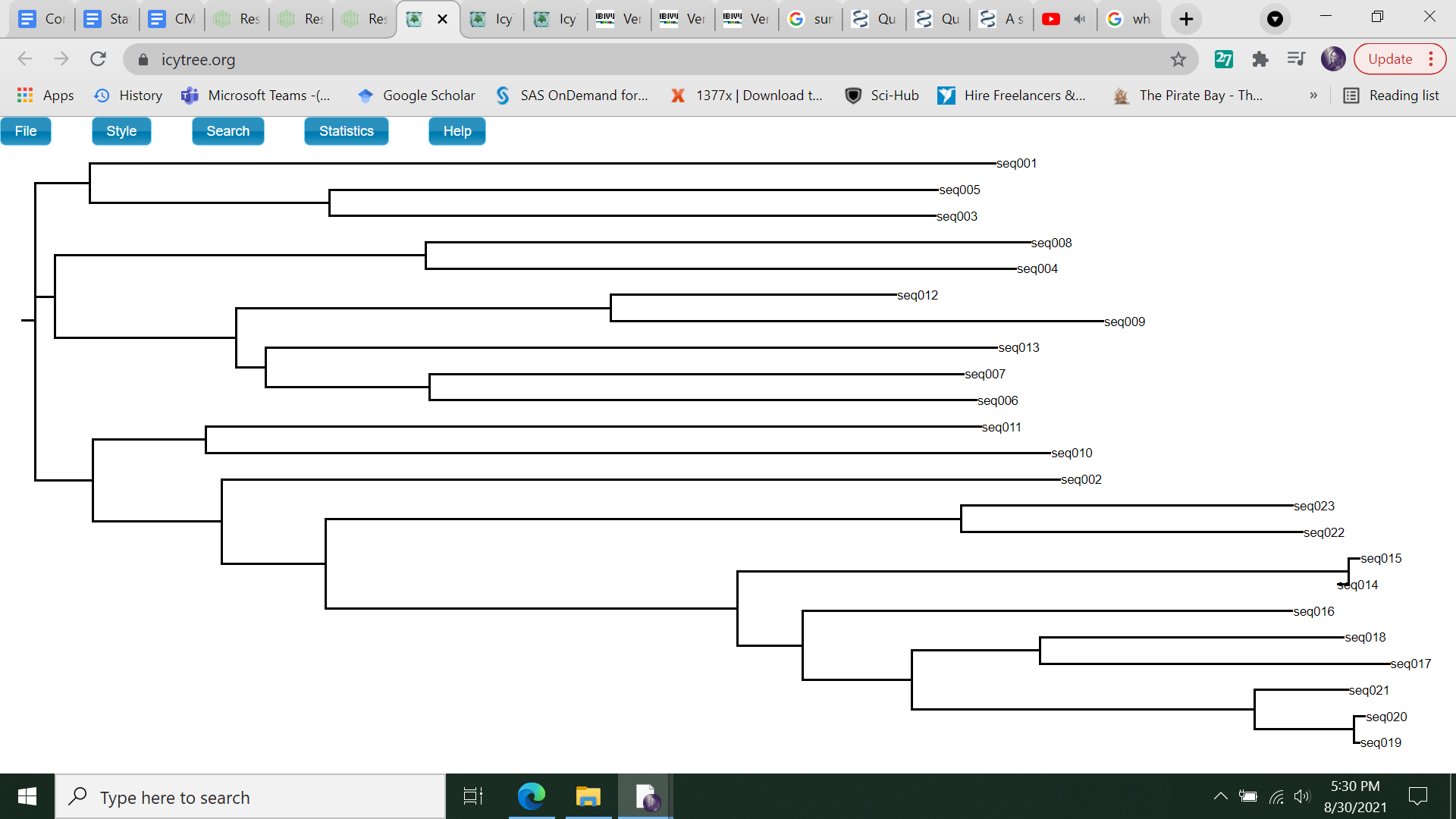


Fig 2a: Clustal omega phylogenetic tree

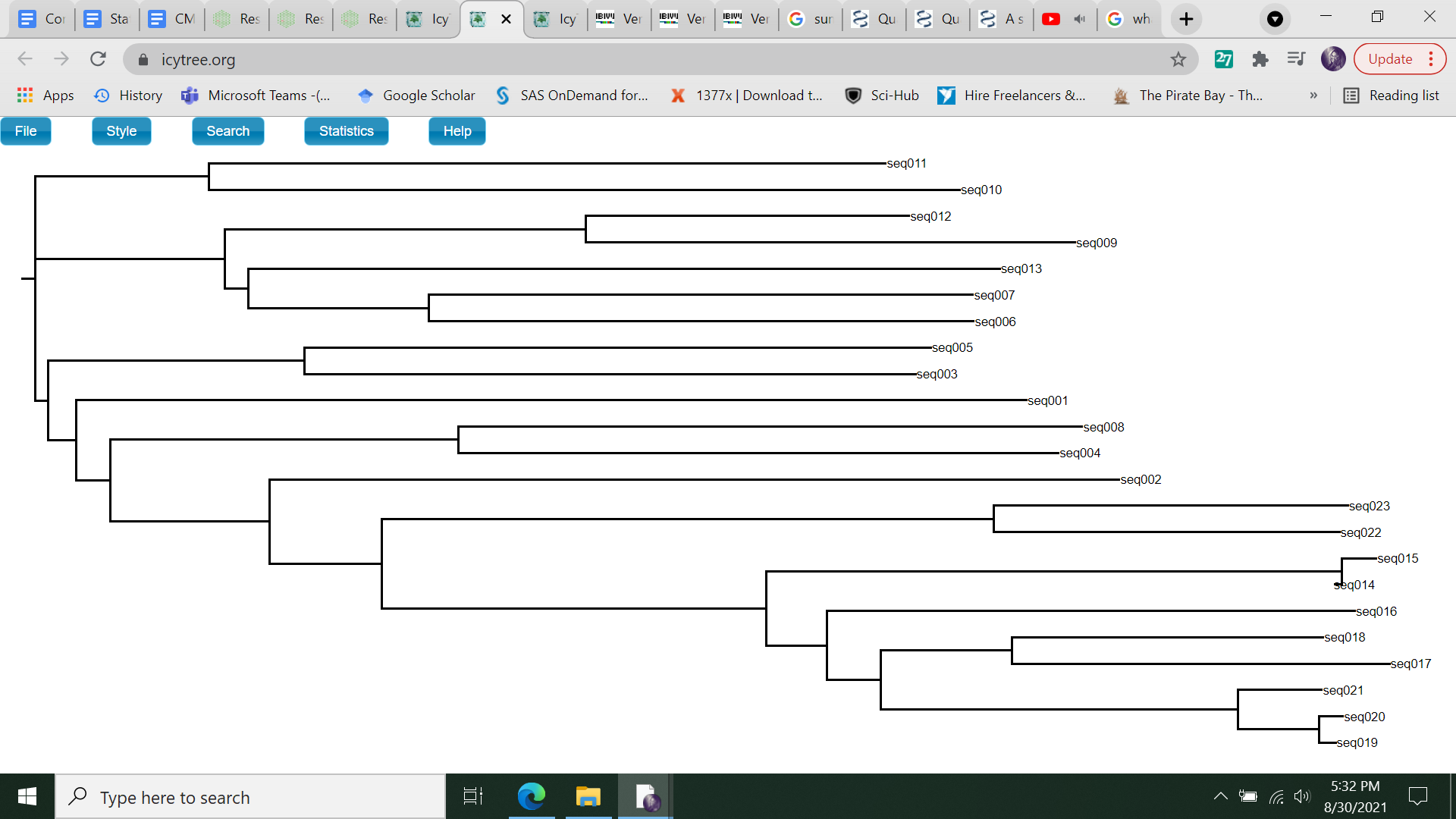


Fig 2b:T-Coffee phylogenetic tree

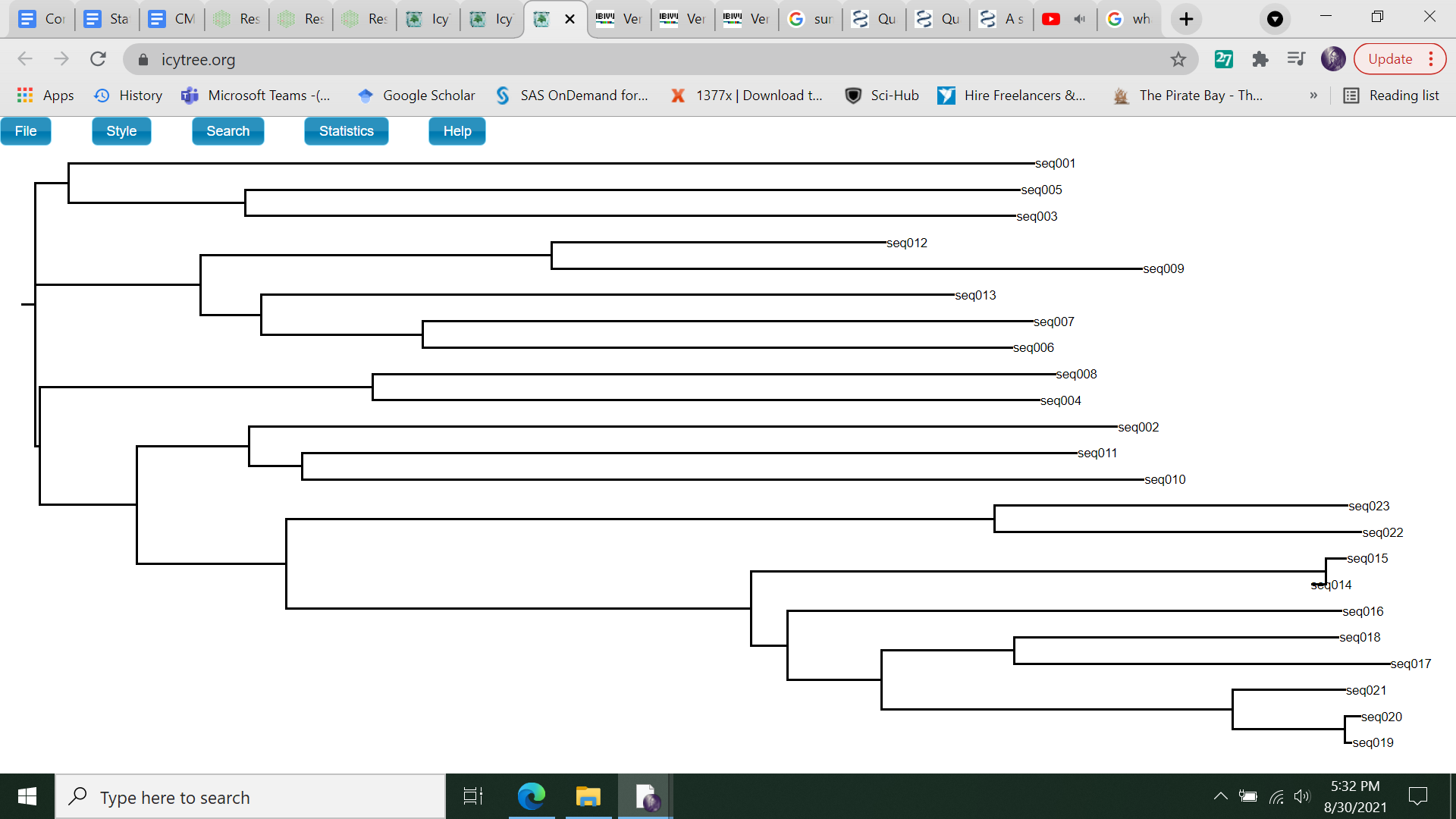


Fig 2c:MUSCLE phylogenetic tree

For instance, as shown in Fig 2a, Clustal omega tree, the sequences 3 and 5 are closely related and parent clad share a relationship with seq 1, they are viewed at the top, but in fig 2b, seq 10 and 11 are shown on the top. Because of the difference in the arrangement of the clads, the length of the branches, parent age, child age etc are also different.