Erika Maldonado-Rosado

24 February 2019

Mauve Lab

BINF 6203

**Introduction:**

In this lab we are looking at 5 different genomes of the same family and/or class. One of the genomes is one that we assembled in lab 2. The other 4 were retrieved from NCBI. We are using a software called Mauve in order to look at gene difference and similarities between genomes. We are able to align them, look at homologies, and move them around. With this lab we are suppose to look at orthologs, look at which part of the nucleotide sequence of multiple genomes align, and understand how we decide what genes are and what they do to help interpret the meaning of genomic similarities and differences.

**Method:**

For this lab we needed to make sure we had the Java environment version 1.4 or higher. For the purpose of this lab I had to uninstall Java JRE and re-install it in order to get Mauve to work. Mauve software was provided to us by the Mave website. Then we retrieved four genomes from the NCBI website using the curl -s command. I used Ubuntu, Linux environment, to use the curl -s command and to get my files onto my desktop. All visualizations were seen on Mauve with the help of Align Progressive Mauve we were able to align our files, save the results, and view it.

**Results:**

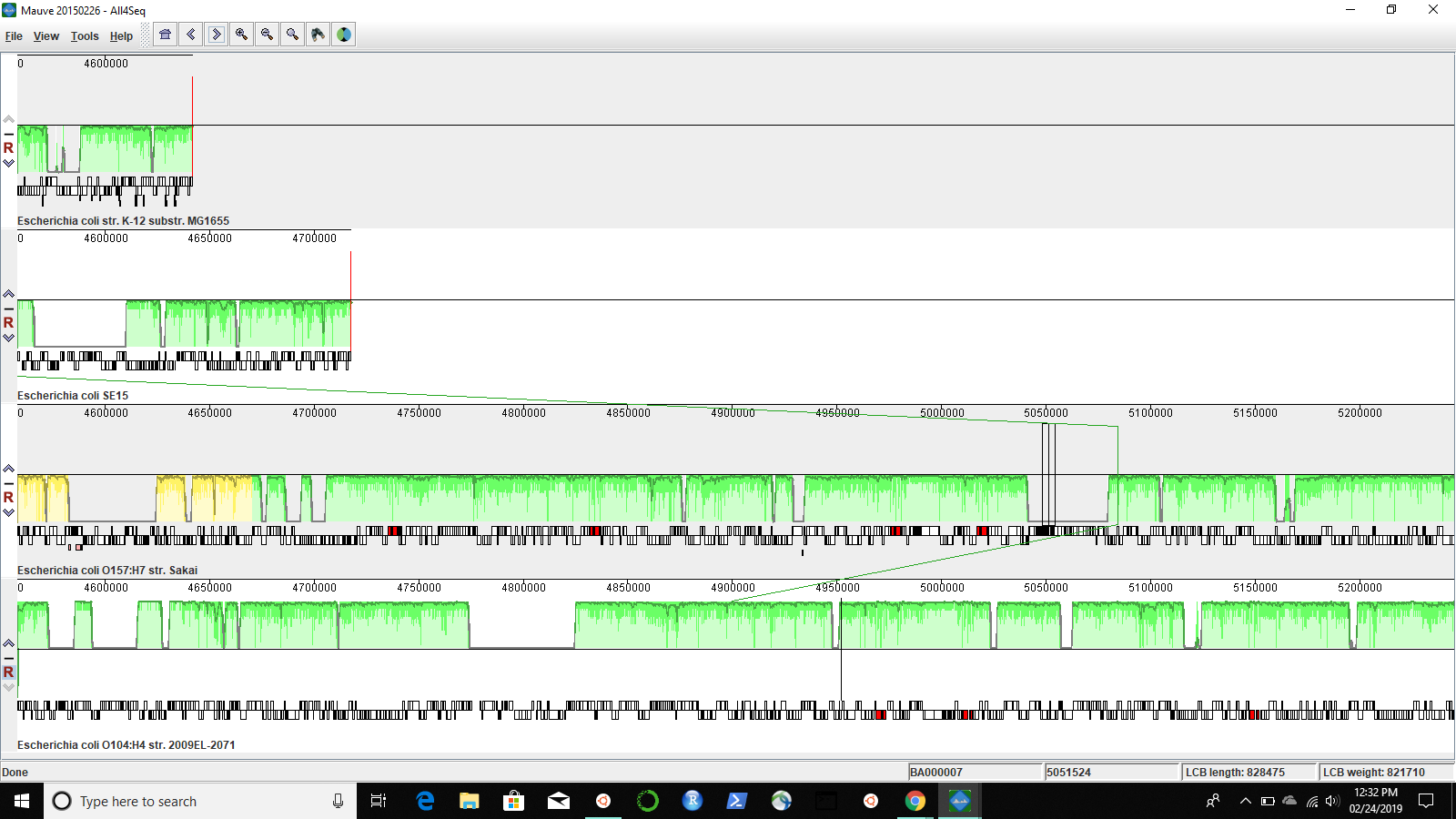
The first image is all four genome sequences together so you know I was able to pull it up. The next two images are of similar genes that I found yjhG KpLE2 gene and it exhibited the D-Xylonate dehydratase. I also decided to look at one other gene in the next two images. For two different genomes that had similar genes but different functions. They had ECSF\_1922 and 1921 for the same gene which was a putative acetyltransferase and putative glycosyltransferase. The similar gene from a different genomes was a Wbbk and it stood for putative lipopolysaccharide biosynthesis protein. It did not see any improvements with the tools, but I did find it easier to visualize my results with everything in the correct order. The move contig tool did give more optimal alignment because I could alter how the default results were.

In the next image is a text file which is ortholog files of the U\* file and of my E. coli K-12 from lab 2. I found it interesting and thought I should share because it exhibited the similarities between the two genomes. Lastly, my last two images show blanks in my E. coli K-12 genome from lab two compared to the U\* file. I have some regions that are blank compared to the published annotated genome. However, I did notice that I saw just as many blanks in U\* file as in the E. coli K-12 from lab 2. I did move the two files to see if it would be easier to view my data and that is how i figured out that there were equal amounts of blanks. One of the blanks that I saw in my E. coli K-12 file was the dgcJ putative diguanylate cyclase DgcJ, however, there was enough of the genome to be able to tell that, that was it. I think we are seeing gaps because there were errors in the sequencing in both of them.

**Image 1**

****

**Image 2 & 3**

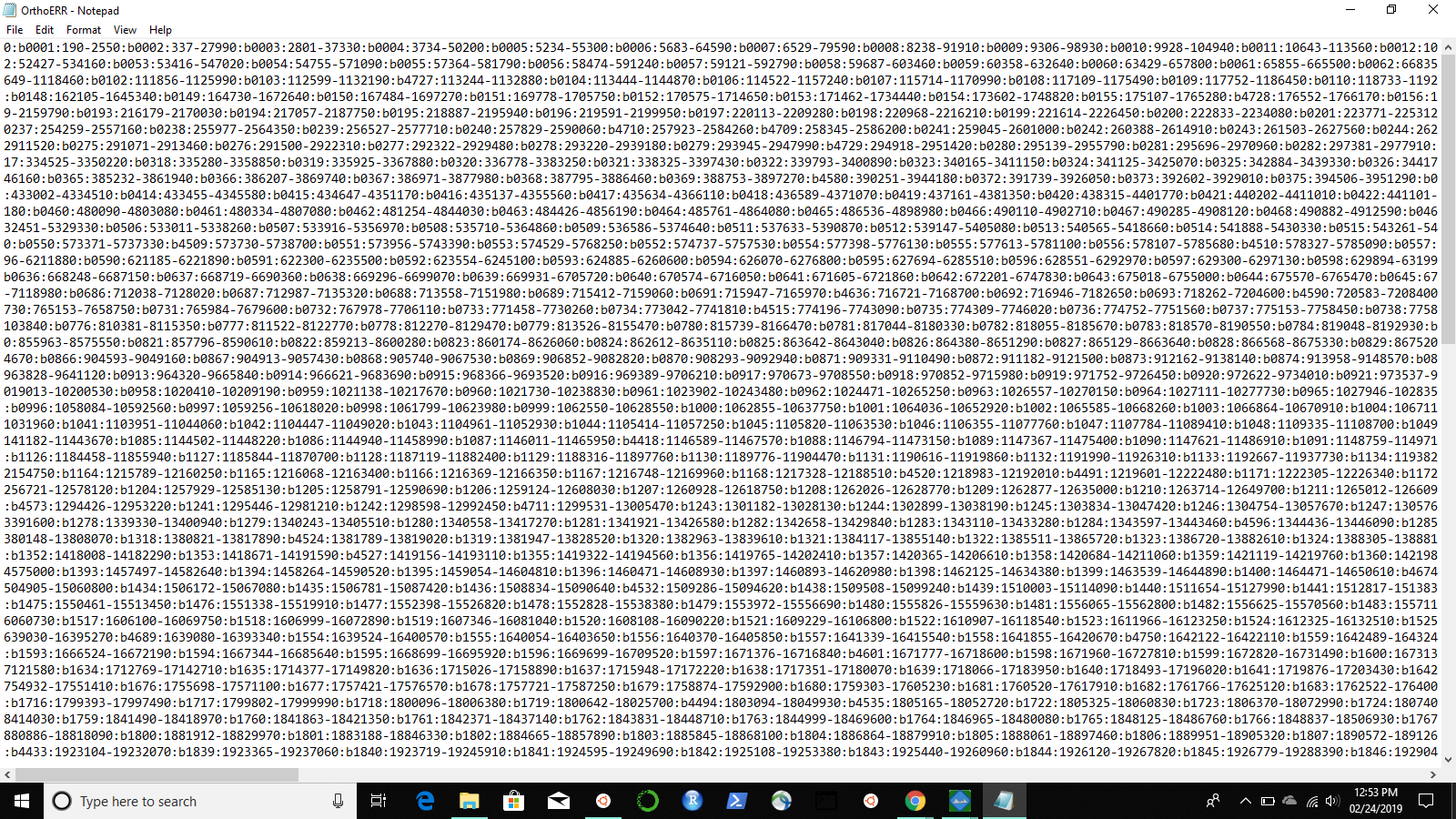
****

**Image 4 & 5**

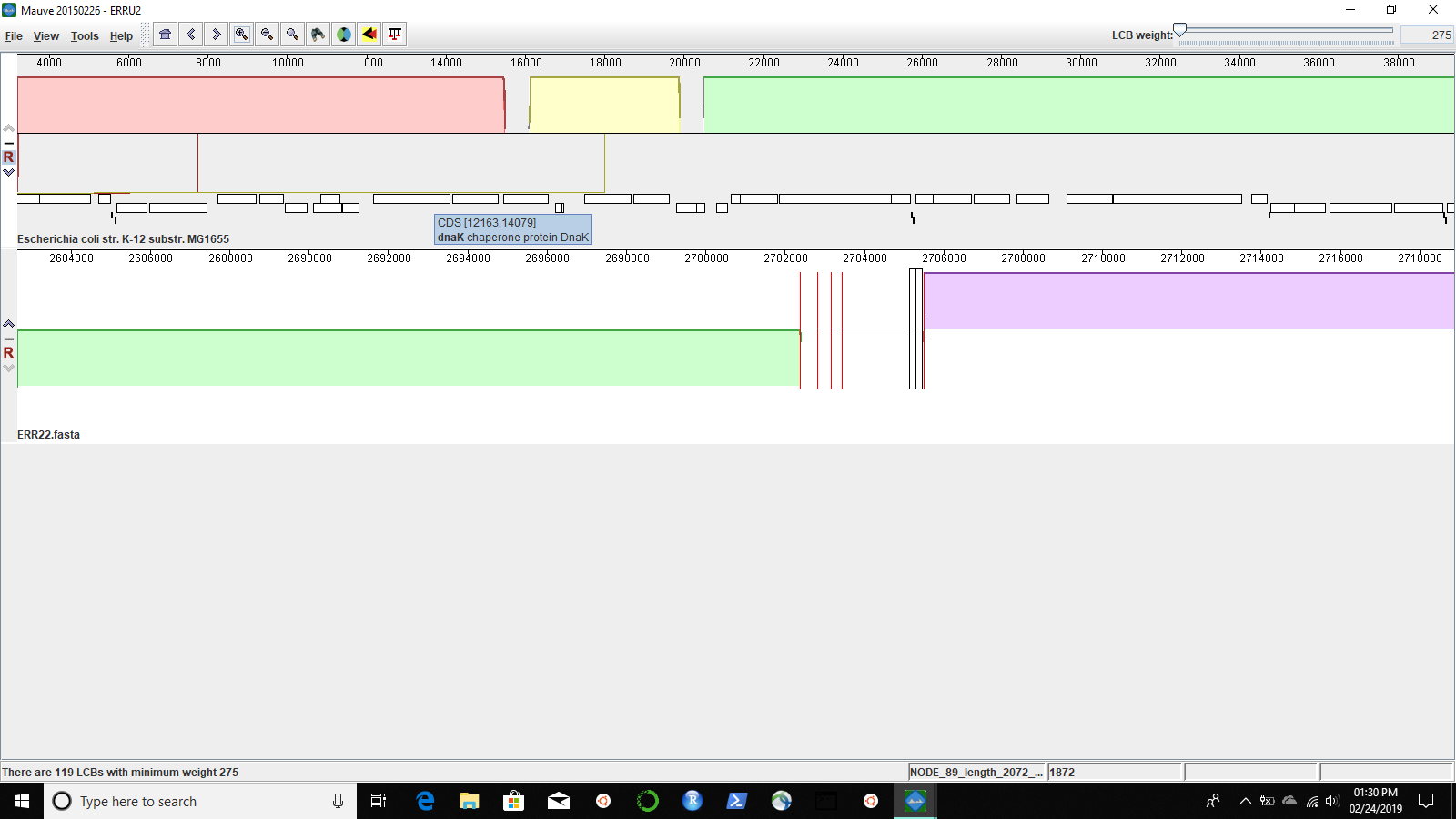
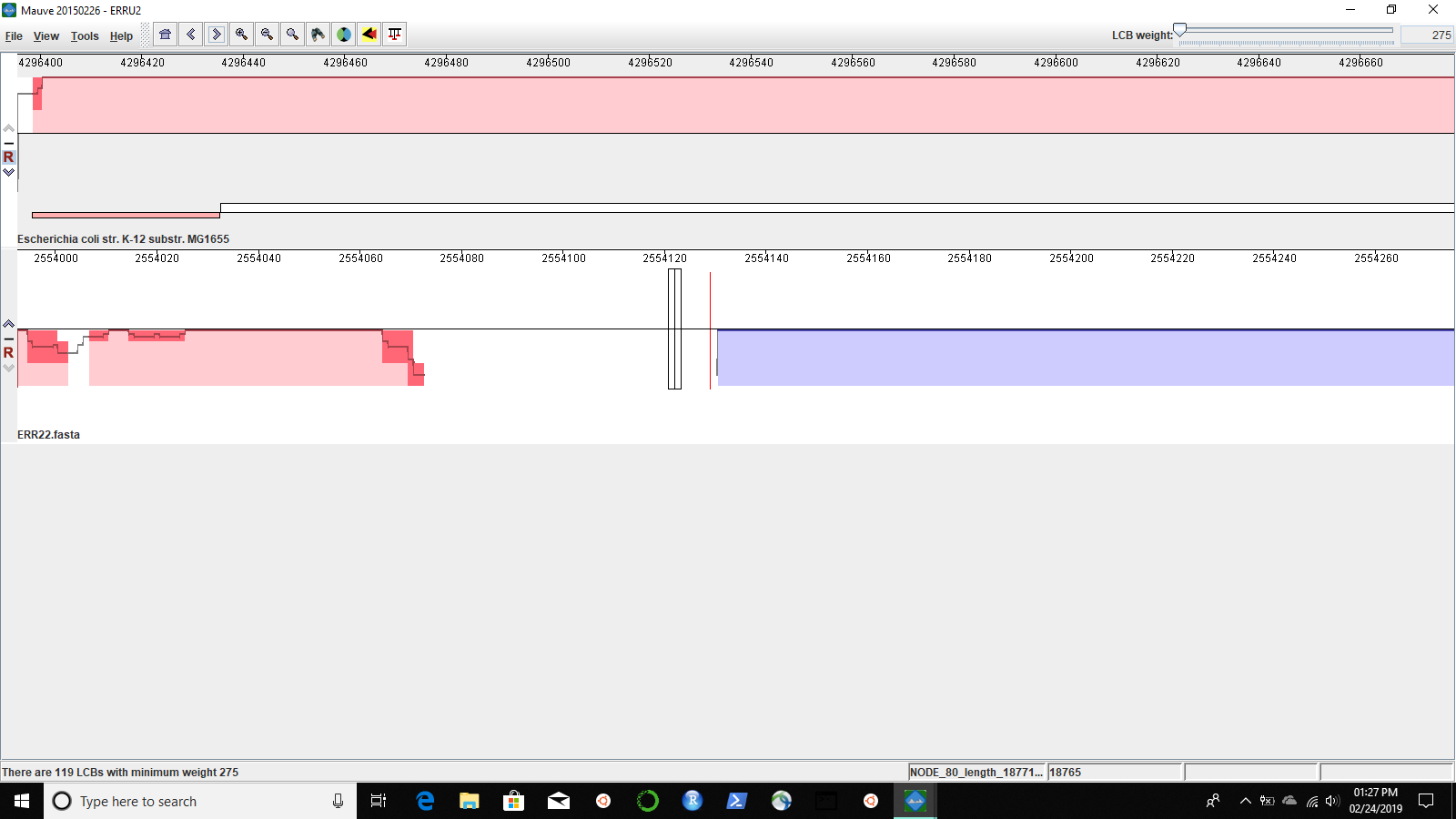
****

****

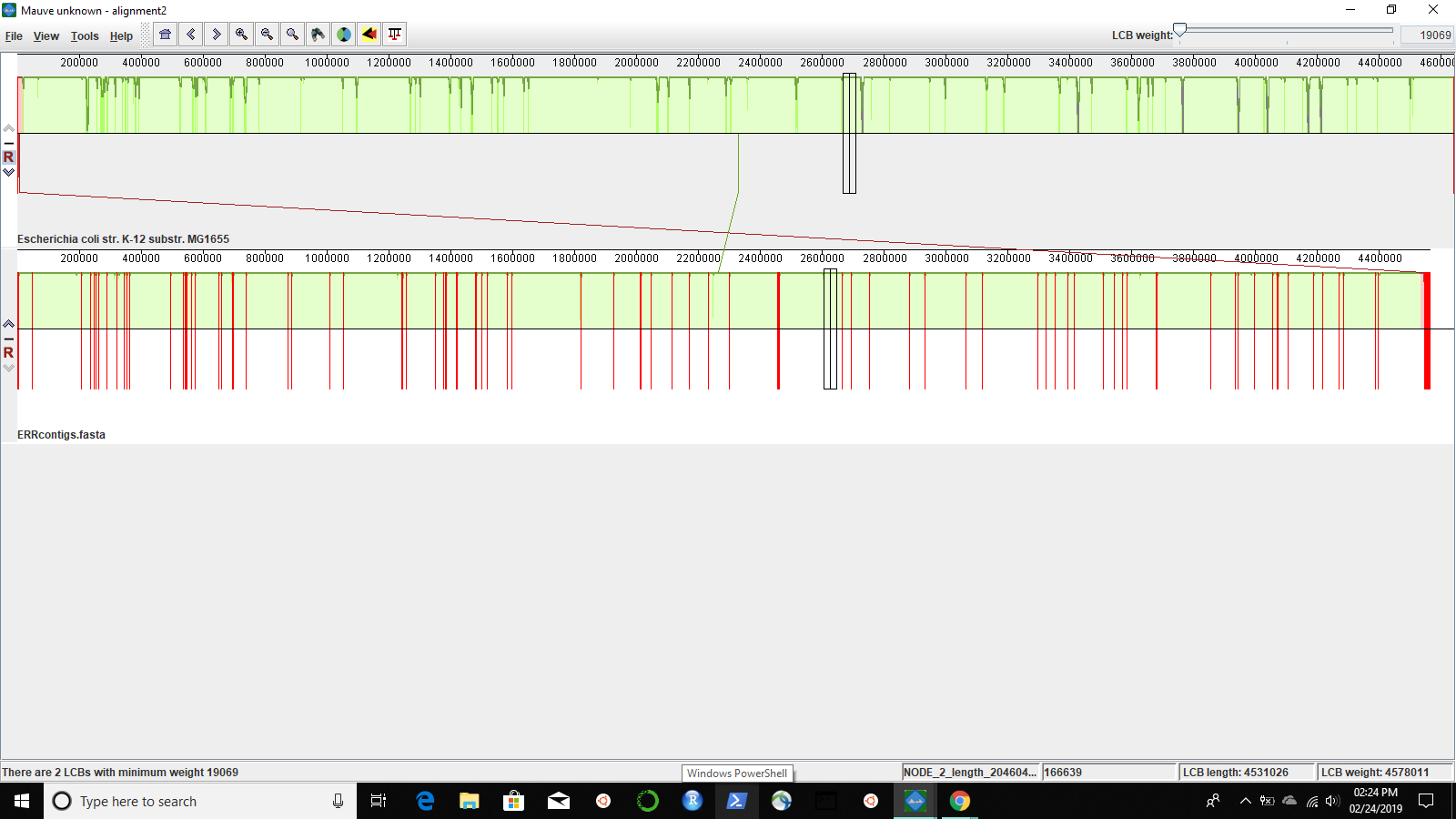
**Image 6**

****

**Image 7 & 8**

****

**Image 9 - (image 7&8 but moved)**

****