**Sequence Similarity Inquiry Exercise – William Tapprich**

**Ebola Bioinformatics Exercise (Exercise 4)**

Ebola virus is one of the most devastating human pathogens. It first emerged in 1976 in Zaire (Democratic Republic of the Congo, DRC) and Sudan. It has been responsible for sporadic outbreaks since that time. In most cases, the outbreaks are self-limiting and small in scope. The usual pattern for an outbreak is infection of a small number of individuals followed by infection of family members and health care workers. After the initial outbreak, high mortality and isolated geography leads to a diminishing number of infections until the outbreak ends. Mortality varies between 40% and 95%, making Ebola one of the most deadly infectious diseases in humans. In addition to high mortality, the disease course is devastating. Ebola Virus Disease (EVD) begins with fever, headache, muscle pain, diarrhea and vomiting then progresses to hemorrhagic fever. During hemorrhagic fever, there is major damage to liver, kidney, and gastrointestinal tract in addition to severe bleeding. Fatalities usually occur as a result of multiple organ failure or shock due to lack of fluids.

The West Africa Ebola outbreak of 2014-2015 taught us much about Ebola. We now know that Ebola has unique features that make it a source of worldwide concern. Health agencies around the world, as well as the World Health Organization, responded massively and effectively to the outbreak. This response is credited with reducing the mortality considerably, but still, 40% of those infected died from the disease.

A current Ebola outbreak in DRC has become a major public health concern. The Ebola outbreak that began in August 2018 in North Kivu Province is still ongoing. As of January 2019, with over 600 cases and 360 deaths, the DRC outbreak is now the second largest in history, but pales in comparison to the 28,000 infected and 11,000 dead in the 14-15 West Africa outbreak. An experimental vaccine has been developed and is being deployed, as completely as the difficulties related to the current conflict in the region will allow, in the current DRC outbreak. To develop the vaccine, it is imperative to know the strain of Ebola causing the current outbreak.

As part of our investigation of Ebola virus, we will use bioinformatics analysis to explore the Ebola strain(s) responsible for the most recent outbreak in Democratic Republic of Congo (DRC).

You have completed three exercises that introduce bioinformatics principles and approaches that enable you to address questions about the viral strain(s) that are present in the DRC outbreak. Table 1 provides information about the Ebola virus Glycoprotein (GP) amino acid sequence for reference strains of Ebola virus and Marburg virus. Table 2 provides information about the GP amino acid sequence from 4 different viruses isolated from individuals infected with Ebola in the current North Kivu outbreak.

Table 1. Accession Numbers for Ebola Reference Strains

|  |  |  |
| --- | --- | --- |
| Virus | Genome Accession | GP Accession |
| Zaire ebolavirus isolate Ebola virus/H.sapiens-tc/COD/1976 | NC\_002549.1 | NP\_066246 |
| Reston ebolavirus isolate Reston virus/M.fascicularis-tc/USA/1989/ | NC\_004161.1 | NP\_690583 |
| Sudan ebolavirus isolate Sudan virus/H.sapiens-tc/UGA/2000/ | NC\_006432.1 | YP\_138523.1 |
| Bundibugyo ebolavirus | NC\_014373.1 | YP\_003815435 |
| Bombali ebolavirus isolate Bombali ebolavirus/Mops condylurus/SLE/2016/ | NC\_039345.1 | YP\_009513277.1 |
| Tai Forest ebolavirus isolate Tai Forest virus/H.sapiens-tc/CIV/1994/ | NC\_014372.1 | YP\_003815426.1 |
| Marburg marburgvirus isolate Ravn virus/H.sapiens-tc/KEN/1987/Kitum Cave | NC\_024781.1 | YP\_009055225.1 |
| Marburg marburgvirus isolate Marburg virus/H.sapiens-tc/KEN/1980/Mt. Elgon-Musoke, | NC\_001608.3 | YP\_001531156.1 |

Table 2. Accession Numbers for North Kivu Ebola Strains

|  |  |  |
| --- | --- | --- |
| Virus | Genome Accession | GP Accession |
| Ebola virus 1 outbreak in North Kivu and Ituri Provinces | MK007341.1 | AYN74157.1 |
| Ebola virus 2 outbreak in North Kivu and Ituri Provinces | MK007340.1 | AYN74148.1 |
| Ebola virus 3 outbreak in North Kivu and Ituri Provinces | MK007342.1 | AYN74166.1 |
| Ebola virus 4 outbreak in North Kivu and Ituri Provinces | MK007343.1 | AYN74175.1 |

Follow the approaches of Bioinformatics exercise 3 to create a multiple sequence alignment and a phylogenetic tree that shows the evolutionary relationships between the GP sequences of the reference strains and at least one of the GP sequences derived from a North Kivu virus.

**Assignment:**

1. Turn in a screen shot of your multiple sequence alignment showing at least the first 119 amino acids and a screen shot of your phylogenetic tree. Please use the Virus Pathogen Database and Analysis Resource (VIPR): <https://www.viprbrc.org/brc/tree.spg?method=ShowCleanInputPage&decorator=flavi_zika>
2. Input a name for your phylogenetic tree
3. Choose Quick Tree followed by the sequence type being analyzed.
4. Choose “Paste sequence in FASTA” and cut and paste your sequences into the text box and select “unaligned FASTA” prior to clicking “Build Tree”.
5. On the following screen choose “Archaeoptryx-js as the tree viewer mode and click “View Tree”.
6. Based on your analysis, describe what you have learned about the North Kivu virus you investigated. For example:
   1. Are you confident that the North Kivu virus is Ebola and not Marburg? What evidence are you using to make that decision?
   2. Is the virus a new strain or is it closely related to a virus from a previous outbreak?
7. Having worked through the bioinformatics analysis, what new questions about Ebola should be investigated?