**EEOB 563 – Final project outline**

**Introduction:**

Ebola virus is a highly pathogenic virus that can cause severe hemorrhagic fever in humans and other primates. The soluble glycoprotein (sGP) of Ebola virus is a protein that is secreted in large quantities during infection. It is distinct from the surface glycoprotein responsible for viral entry into host cells. sGP is thought to play a role in modulating the host immune response, possibly by competing with the surface glycoprotein for binding to host cell receptors.

Phylogenetic analysis of Ebola virus sGP can provide insights into the evolutionary

relationships and transmission patterns of the virus. By comparing the sequences of sGP from different isolates of the virus, we can construct a phylogenetic tree that reveals the relatedness of different strains and helps to identify the origin of outbreaks. This information can be used to track the spread of the virus during epidemics and inform public health responses to outbreaks. Overall, understanding the role of sGP in Ebola virus pathogenesis and using phylogenetic analysis to study the evolution of the virus can help to guide efforts to control and prevent the spread of this deadly disease.

**Main questions:**

How conserved is the EBOV sGP across the different strains of EBOV? There have been papers, which used whole genome or single gene (GP) to construct the phylogenetic tree of EBOV, how does a phylogenetic tree based on sGP or GP compares to the phylogenetic trees bases on genes/genomes?

**Methods:**

I’m going to perform Baysian analysis either using BEAST or MrBayes to construct the phylogenetic tree and perfom further analysis.

**Source of data:**

I’m going to use the amino acid sequence of EBOV sGP and/or GP of different strains of EBOV which are publicly available at NCBI (https://www.ncbi.nlm.nih.gov)