**Working title: Illustrating evolution of EBOV Glycoprotein**

Large Goals (Macro view of project):

We wish to create a visual web product that illustrates the geographic, chronological, genetic, and biochemical diversity of the Ebolavirus glycoprotein (EBOV GP) evolution over the past 40 years.

This will be accomplished by specific case studies of, but not limited to, Ebolavirus outbreaks in Yambuku, Zaire (1976); Kiwit, Zaire (1995); Sierra Leon (2014).

Specific stepping stones (Micro view of the project):

We wish to accomplish, at the minimum, the following roadmap.

(1) Perform multiple sequence alignment and discover a statistical significant and physically meaningful sequence divergence among EBOV genomes from Ebolavirus outbreaks since 1976 in West Africa. This includes, but is not limited to:

1976 - Yambuku, Zaire (318 infected, 280 killed - 88% fatality)

1995 - Kiwit, Zaire (315 infected, 254 killed - 81% fatality)

2014 - Sierra Leone (10,141 infected, 4,922 killed - 71% fatality)\*

(1.1) In fact, we specifically wish to increase the diversity of our sources)

(1.2) Sources include: Science\*, RCSB, NCBI & Genbank

(2) Reproduce finding from Science\*\* that it was observed that Ebolavirus mutating at rate of 8x10^(-4) mutations per site per year.

(2.1) Produce illustration of non-averaged Ebolavirus mutations over time.

(3) Investigate potential correlation between sequence divergence of Ebolavirus samples and fatality rates over time

(4) Produce choropleth map visualizing geographic and genetic diversity of Ebolavirus over time in sampled regions

(5) Investigate amino acid diversity of EBOV GP

(5.1) Are any amino acid changes EBOV GP significant to the overall biological function?

(5.2) Model amino acid structure of each Ebolavirus sample from each region using RCSB.

(6) Expand analysis to additional outbreak areas

(6.1) Congo

(6.2) Philippine Monkies

(6.3) Gabon

\*\*\* More structured list

(7) Web visualization

(7.1) Choropleth map visualizing studied geography

(7.2) Genome & Amino Acid viewer

(7.3) Protein structure viewer

\* Statistics taken from Wikipedia. 2014 values reflect the worldwide outbreak.

\*\* Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak Stephen K. Gire et al. Science 345, 1369 (2014); DOI: 10.1126/science.1259657

\*\*\* Structured list available in outbreak.gsheet