**Towards a whole-virus view of Influenza B virus evolution**

Influenza pandemics originate when antigenically new influenza viruses enter and spread in the human population. In fact, four of the past five influenza pandemics in the past 100 years arose from the introduction of avian and swine virus genes into the human population. The viruses that led to these pandemics represent only a fraction of the total diversity of influenza viruses present in nature, making it challenging to identify the viruses likely to result in the next human pandemic. [1]. Seasonal influenza viruses evolve rapidly to escape the collective immunity of the human population. Other zoonotic strains, such as A/H5N1, may or may not be able to evolve transmissibility between humans. If they can attain transmissibility, they can lead to new influenza pandemics. [2].

Recent research on the mutations that allow for receptor binding in engineering ferret-transmissible H5N1 strains, illustrates that the predictability of disease phenotype from viral sequence is complex. When these mutations were introduced in the context of the genetic background of more recent avian isolates of H5N1, affinity to human receptors was lost; the phenotype caused by a mutation is likely mediated by its interactions with the genetic background, making epistasis an important factor in the attainment of transmissibility[3]. Recent studies of molecular sequence data have suggested that epistasis is highly important for sequence evolution. A study examining large protein sequence alignments noted that about 90 percent of all amino-acid substitutions are neutral or beneficial only in their own genetic background, and deleterious in the genetic backgrounds characteristic of other species [4]. The majority of possible amino acids are not observed at any given position in an alignment perhaps due to the effect of epistasis upon the fitness of substitutions. The goal of this project is to obtain a better understanding of the extent to which epistatic interactions are constrained by the need to form appropriate interactions within the virus and also with the host.

**Specific Aim 1. Determine the extent to which structural epistatic interactions are constrained by the virus.** The outcome of viral pathogenesis requires appropriate interactions between the viral proteins. A multiple-sequence alignment of influenza B virus, which exclusively infects humans, will be conducted to identify correlated mutations using a method by Jacob et al [5]. We will explore the effect of mutated positions on the tertiary structure using EVmutation, a de novo structure prediction method proposed by Hopf et al [6]. We will determine how well correlated amino acids can be explained by protein structure, by identifying parameters such as connect distance between correlated amino acids and residue accessibility.

**Specific Aim 2. Determine the extent to which structural epistatic interactions are determined by host-virus interactions.** The different outcomes of viral pathogenesis depend on the interactions between the virus and the host cellular proteins. These interactions are vital to the virus hijacking the cell’s molecular machinery to promote the viral life cycle or trigger the host immune system against the virus. A total of 136 host-virus protein interactions have been found for the Influenza B Virus in 3 studies [7]. We will identify the proteins that are vital for the viral cycle. Using similar analyses as in specific aim 1, will consider how epistatic interactions are limited by the need to drive the viral cycle by appropriately interacting with the host proteins.

**References**

[1] Russell et al, Science forum: Improving pandemic influenza risk assessment, eLife, 2014

[2] Gong et al, eLife, 2013

[3] Lipstich and Galvani, Ethical Alternatives to Experiments with Novel Potential Pandemic Pathogens, PLoS Medicine, 2014

[4] Breen et al, Epistasis as the primary factor in molecular evolution, Nature, 2012

[5] <https://www-nature-com.proxy.uchicago.edu/articles/nmeth.3634>

[6] <https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5383098/> [7] [https://www.ncbi.nlm.nih.gov/pmc/articles/PMC557’ ‘](https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5578063/)

1) What is the theme of the proposed research? What are the major questions to be addressed?

2) Do the methods and preliminary analyses used integrate concepts that lie at the interface of molecular evolution and biochemistry? Do they address the questions being asked?

3) What major improvements would you suggest?