Problem-based Learning for Bioinformatics: Evolutionary genetics, phylogenetics, and molecular epidemiology

Instructor: Jeffrey Joy

February 05, 2020

Introduction to evolutionary genetics and molecular epidemiology:

The application of evolutionary genetic and phylogenetic techniques in epidemiology is a rapidly emerging field in medicine and biomedical research, it is crucially important in the analysis and understanding of the evolution of emerging human viral pathogens (especially rapidly evolving RNA viruses that have been sampled through time, for example the current outbreak of the novel, as yet unnamed, corona virus nCOV). This field is the future of many aspects of the study and response to infectious disease outbreaks. One of the core objectives of this field is to enable the transformation of large data sets into improvements in public health action, clinical practice, and preventative medicine. It is also inherently interesting from a pure evolutionary perspective. There are many important practical outcomes of work in this field for example the development of recommendations to optimize treatment and prevention of virus transmissions, and, in a public health context, the monitoring of viral transmission, and the prevalence of drug resistance in the population. Generally, molecular epidemiology involves the analysis of molecular sequence data from viral disease (or other pathogen) outbreaks to infer evolutionary and population dynamic processes in near to real-time.

Brief Introduction to Original Antigenic Sin

Original antigenic sin, also known as the Hoskins effect, refers to the propensity of the body's immune system to preferentially utilize immunological memory based on a previous infection when a second slightly different version of that foreign entity (e.g. a virus or bacterium) is encountered. This leaves the immune system "trapped" by the first response it has made to each antigen, and unable to mount potentially more effective responses during subsequent infections. The phenomenon of original antigenic sin has been described in relation to influenza virus, dengue fever, human immunodeficiency virus (HIV), and to several other viruses. This phenomenon was first described in 1960 by Thomas Francis, Jr. in the article "On the Doctrine of Original Antigenic Sin". It is named by analogy to the theological concept of original sin. (hint: the influenza literature is important to look at for this).

Evolutionary genetics and drug resistance

Sequences derived from patient infections can display mutations that confer resistance to the therapies taken to eradicate or suppress the pathogen (e.g. hepatitis C virus (HCV) or human immunodeficiency virus (HIV)). Virus sequences sampled from an infection may display resistance as a result of 2 main pathways "de novo" and "transmitted". A critical role of some clinical laboratories is to sequence pathogens and characterize the variation at known positions that confer resistance to therapy. Generally, amino acid sites in a sequence are reported in the format: M184V; where "M" is the most common "wild type" or susceptible amino acid, "184" is the amino acid position in the sequence (i.e. 184^{th} amino acid), and "V" is the resistant phenotype which confers resistance to in this case "nrti" therapies.

Learning Objectives:

1. Be able to describe the basic techniques in bioinformatics and how they can be applied to clinical research and practice
2. Demonstrate ability to use modern web bioinformatic and evolutionary genetic tools.
3. Describe how bioinformatics can be applied in a modern clinical laboratory setting to improve the care of HIV patients (and other infectious diseases such as HCV) and reduce transmission rates.
4. Determine how recommendations about antiretroviral treatment for HIV infected individuals can be optimized for individual patients in a personalized medicine framework.
5. Understand the difference between de novo drug resistance and transmitted drug resistance.
6. Understand the concept of "Original Antigenic Sin" and it's importance in infectious disease.
7. Understand how we use phylogenetics to optimize public health responses.

Problem:

As the resident public health scientist specializing in phylogenetic analysis you are faced with a rapidly emerging outbreak of an infectious virus. Luckily, genomes of the virus are being generated at a record pace and hundreds of genomes sampled through time from all over the globe are already publicly available with 15 new genomes added to the public databases on a daily basis. Also fortunately, you have unlimited computing resources. How do you leverage the accumulating data, along with your skills and resources to help combat the outbreak?

Guiding Questions:

1. One will often hear the term “phylodynamics” what does it mean and how is it different from “phylogenetics”?

2. How can molecular epidemiology play a role in public health policy and public health interventions?

3. How can evolutionary genetic and molecular evolutionary methods be utilized to optimize patient care?

4. Why might the study of RNA viral pathogens (and other pathogens) be inherently interesting from a pure evolutionary biology perspective?

5. HIV transmission remains criminalized in Canada what are the implications of this for: 1. The management of the disease in general; 2. The perspective of populations at risk of HIV transmission; 3. Public health practitioners involved in the response to the epidemic; and 4. What should you, the bioinformatician, take into consideration when performing analyses on HIV data? (Note that these same questions apply to other notifiable stigmatizing diseases e.g. HCV)

6. Given your discussion around question 5. how might you transmit the results of your phylogenetic and molecular epidemiological analyses to public health practitioners?

7. Following along from SARS there is a lot to be learned from the recently emergence of the novel corona virus (nCOV). What did we do better this time as nCOV emerged?

8. Following on from question 7, how can we improve upon the response that we have observed unfolding with nCOV for future oubreaks? i.e. the pending zombie apocalypse.

9. Original antigenic sin is most well known in the context of influenza, however what other scenarios/pathogens might this concept also be important and potentially overlooked?

10. Ebola is another repeated emerging disease of tremendous clinical and public health significance. Bioinformatic tools have been critical in understanding the previous Ebola epidemic outbreaks and will also be important for subsequent outbreaks. How might changing sequencing technology and phylogenetic tools prove even more valuable for emerging outbreaks such as Ebola in the future?