

# Phylogenetic Analysis

Hsiang-Yu Yuan

May 6, 2015

## Abstract

## 1 Materials and Methods

### 1.1 Immune escape model

The host immunity would be boosted during an infection and protects the host from being infected again if contact with the viruses. Once contacted with the same viruses, the probability of the virus to survive in the presence of the host immunity is defined as

$$f(V_i) = [1 - e^{-p(V_i+1)}]^\frac{1}{D} \quad (1)$$

where  $D$  is the antigenic distance between the host immunity and the same viruses. When the infection history  $H$  is considered, the probability of viruses survival becomes the survival of all cross-reactive immunity

$$f(H, V_i) = \prod_{j=1}^k [1 - e^{-p(V_i+1)}]^\frac{1}{D+\alpha_{ji}} \quad (2)$$

$$= [1 - e^{-p(V_i+1)}]^\sum_{j=1}^k \frac{1}{D+\alpha_{ji}} \quad (3)$$

where  $j$  indicates the virus strain caused previous infection and  $\alpha_{ji}$  is the distance of antigenic drift between the the previous strain  $j$  and the challenging strain  $i$ .

In order to evaluate the evolutionary dynamics of binding avidity, we made a simplified framework with certain assumptions. First, we consider individual viruses change on their viral receptor binding but have no individual differences on antigenicity. We assume a constant immune boosting rate. Each time when infection occurs, the immune boosting occurs which create the protection effect as antigenic distance  $D$ . We also assume the antigenic distance between the time of two consecutive infection is constant. Therefore, the distance of antigenic drift between strain  $j$  and  $i$  becomes

$$\alpha_{ji} = \delta \cdot (k + 1 - j) \quad (4)$$

The probability of within host virus infection per each virus becomes the product of virus survival and the cost of binding avidity (ref).

$$\phi(H_k, V_i) = [1 - e^{-p(V_i+1)}]^{\sum_{j=1}^k \frac{1}{D+\alpha_{ji}}} \cdot e^{-aV^b} \quad (5)$$

The probability of viral infection is

$$\rho = 1 - \left(\frac{1}{R_{in}}\right)^{-v} = 1 - \left(\frac{1}{n \cdot \phi(H_k, V_i)}\right)^{-v} \quad (6)$$

where  $R_{in}$  is within host reproductive number. We assume the constant number of offspring  $n$  during replication for each virus

[plot  $R_{in}$ ]

[plot  $\rho$ ]

## 2 Markers of netcharge

Provided that virus is isolated towards the end of a hosts infectious period, the binding avidity hypothesis for influenzas antigenic drift predicts a positive relationship between host immune status and the binding avidity of the virus isolated from the host. In our first analysis, we tested this prediction using influenza viral sequences isolated from New York State as part of an Influenza Genome Sequencing Project [20], along with metadata on host age. The sequence data, publicly available from the Influenza Virus Resource Database, consisted of a total of 686 full-length influenza A/H3N2 HA sequences and a total of 44 full-length influenza A/H1N1 HA sequences isolated across New York State between 1993 and 2006. For each of these viral isolates, the exact date of isolation and age of the individual from whom the virus was isolated were also downloaded.

[add]Since currently we still lack of binding avidity data for most of the virus strains. We are investigating whether netcharge is a good marker. Testing of the correlation of Binding avidity and Netcharge Chi square test to see whether binding avidity and netcharge are correlated. Binding Mutants source \$G:\MyPC2014Late\working\Projects.Duke\Projects\BindingAnal\validate\_netcharge\validate\_netcharge.xlsx

Check Adam J. Kucharski PLOS Biology paper. Estimating the Life Course of Influenza A (H3N2) Antibody Responses from CrossSectional Data  
If we use years 50 as a cutoff and test the regression from 3-50 and 51-100, what will show us?

[add]Studies show that the relationship between host age and immunity is not linear. Using serological data, the titres decrease from early children to mid-aged adults, then increase by age.

We used host age as a proxy for host immune status, while recognizing that it is an imperfect and only rough proxy. Our choice of host age for host immune status stems from several factors. First, it is well known that children are highly susceptible to infection with influenza because they lack prior exposure to the virus and because they have yet to develop a mature immune system exhibiting a full polyclonal response [2123]. As individuals grow older, each subsequent infection results in the proliferation of a number of B-cell clones and a polyclonal antibody response. It is therefore likely that prior exposure to previously circulating strains in older individuals would lead to cross-reactivity to newly circulating strains that are evolutionary descendants from these older strains. A general increase in seroprevalence by age against individual viral strains has also been observed [24,25], further supporting the use of host age as a proxy for immune status. Second, our choice of host age as a proxy for host immune status is a practical one: data on host immune status are not readily available, while data on host age are. Recognizing the limitations of using host age instead of host immune status directly, we will interpret a positive relationship between host age and binding avidity as evidence for immune-mediated selection on the binding avidity phenotype, while acknowledging that immune-mediated selection might still be at play if we do not find evidence for a positive relationship between host age and binding avidity.