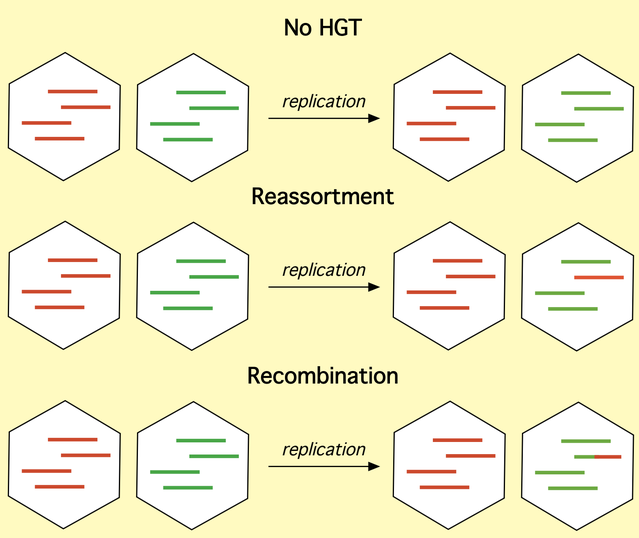
Recombination and Reassortment in the Evolutionary History of Influenza A

Paul-Michael Agapow (paul-michael.agapow@hpa.org.uk)
Bioinformatics, Centre for Infections, Health Protection Agency

Introduction

Much of the impact of influenza is due to the viruses adaption via the exchange of gene segments between different strains. This horizontal gene transfer (HGT) allows the combination of advantageous traits from different viruses, allowing large adaptive leaps and comes in two forms. The first is *reassortment*: the transfer of entire gene segments without alteration. This has been complicit in virtually all major pandemics and the host shifts that can suddenly create virulent strains (e.g. H1N1pdm). At the same time, its evolutionary parameters are poorly delineated, with little understanding of which genes reassort at what rate under which circumstances (e.g. host and strain specific factors).

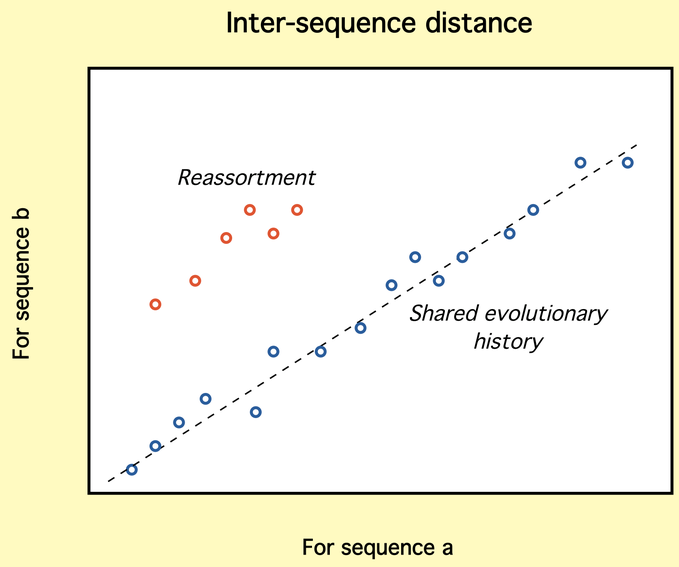


The second - *recombination*, the merging of homologous gene segments - is more widely understood outside of influenza than within. Recombination has been widely observed in pathogen evolution and intuitively should function similar to reassortment, in the transfer of whole functional adaptive traits. However, its place in influenza evolution is unclear with some studies finding it ubiquitous [Niman] but others "essentially non-existent" [Boni]

This study is the start of a comprehensive study into the parameters surrounding horizontal gene transfer in influenza, so as to better understand and model the evolution of influenza and the generation of virulent pandemic strains. It uses three divergent whole genome sequence datasets from the NCBI Influenza Virus Resource – 67 swine-lineage H1N1, 167 avian-lineage H5N1 and 525 human lineage H3N2 – and focuses on how reassortment and recombination can be detected in influenza and differs across genes and hosts.

Comparing sequence distances

Reassortment was identified by the method of Rabadan [2009] via Khiabanian [2009]. Essentially for each coding segment of the influenza genome, a distance is calculated between each strain. If two genes share the same evolutionary history (i.e. no HGT involving them has taken place), their distances should vary proportional to each other and when plotted should cluster on a diagonal running through zero. However, if two genes do not share the same evolutionary history, the time that separates two strains will differ across genes and when compared in a plot, points will diverge from the diagonal:



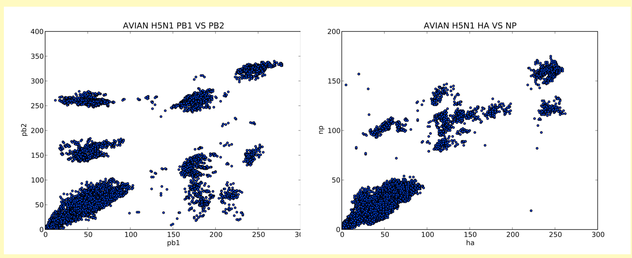
To calculate intra-strain distances, within each dataset, each influenza coding region was aligned. Sequences were then reduced to third codon positions only, so as to avoid any effect of positive or negative selection. For each unique pair of strains (e.g. A/chicken/Egypt/2253/2006 vs A/duck/Vietnam/201/2005), a distance was calculated as a simple number of differences between the sequences for each gene. Finally, the distances for each unique pair of genes (e.g. HA vs NA, NA vs PB1) were plotted.

To examine recombination, each coding segment was broken into two halves and these analysed and compared as above, so as to see if the 3′ and 5′ "genes" had different evolutionary histories.

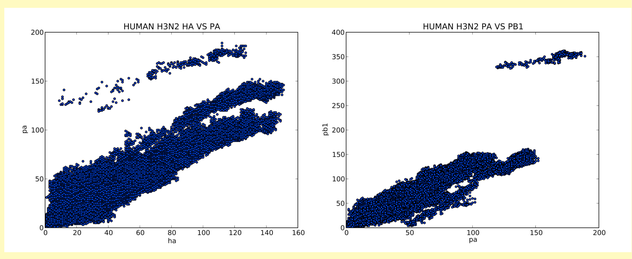
A large amount of data and graphs were generated. Space restricts this poster to a qualitative summary of the highlights and precludes more complex statistical analysis.

Results

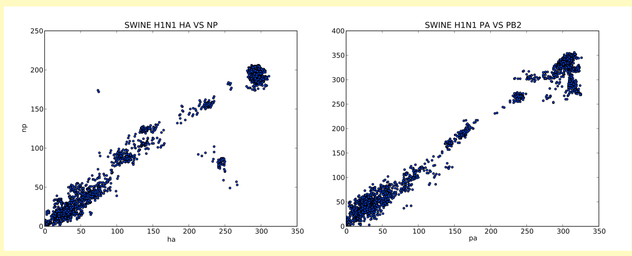
The avian H5N1 dataset contained complex and rich signatures of reassortment for every pair of genes save - interestingly - HA and NA. The presence of several distinct clades (diagonals) is consistent with multiple independent reassortment events even in nuclear / non-antigenic genes such as the polymerase genes:



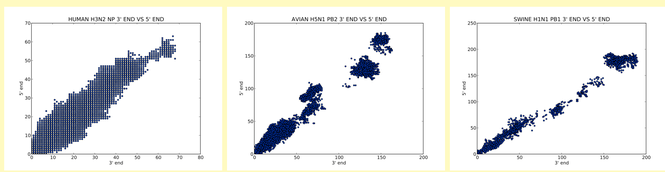
The human dataset showed distinct reassortment signatures primarily for antigenic and polymerase genes. Curiously, there were fewer clear multiple events than the avian data and - despite the larger dataset - signatures and clades were much more clearly defined.



Reassortment was less distinct in swine H1N1, primarily concentrated in antigenic genes and possibly in NA (nucleoprotein).



No clear sign of recombination was found. A faint signal may be present in the avian H5N1, representing a clade with a recent recombination within PB2. This awaits further analysis. It is interesting to note that the human H3N2 data was more widely dispersed that other hosts. This may reflect a more complex adaptive environment for influenza in humans or simply be a side effect of sampling and dataset size.



Conclusions

With the caveats expressed above, some broad conclusions can be drawn. The existence reassortment in influenza is a given, but requirement of co-infection has lead some to suggest it is a rare event [Hamilton]. In fact, each dataset shows at least several reassortment events suggesting it is infrequent but far from extraordinary. Considering the evidence for co-circulation of distinct strains, co-infection and opportunities for HGT may be more common than suspected.

There are also signs of how reassortment rate varies. While the antigenic genes are most commonly. This may simply reflect the powerful selection that could be placed on such reassortants, but non-antigenic genes (e.g. polymerase) are also exchanged. There are also varying rates across strains and genes, e.g. the distinct multiple events in avian H1N1 versus the few in H3N2. Further analysis will be required to identify the responsible host and/or strain factors.

The absence of recombination is interesting. Possibly reassortment is so powerful that the benefit of recombination is minimal. It should be noted that an analysis of 2009 pandemic H1N1 showed no signal of recombination. However, any HGT might be lost or undetectable if it occurs between close relatives.

Thanks to Monica Galiano, Xu-Sheng Zhang, Jonathan Green and the staff and facilities of the HPA Centre for Infections.

- Niman (2007) Swine Influenza A evolution via recombination - Genetic drift reservoir, *Nature Precedings*
- Khiabanian et al. (2009) Reassortment patterns in Swine Influenza Viruses, *PLoS One* 4(10):1-7
- Rabadan et al. (2009) Comparison of avian and human influenza A viruses reveals a mutational bias on the viral genomes, *J. Virol.* 80(23):11887-91
- Boni et al. (2008) Homologous recombination is very rare or absent in human influenza A virus, *J. Virol.* 82: 4807Ð4811
- Hamilton (2002) *Narrow Roads of Gene Land, Volume 2: Evolution of Sex*