Understanding Changes in Influenza B Viral Genomes

Ewan P. Plant and Zhiping Ye
Division of Viral Products, Office of Vaccine Research and Review, Center for Biologics Evaluation and Research

Background

Genomes are malleable and change over time. However, it is not always clear why the changes occur or what impact they have. Understanding how viral genomes change is an important part of selecting which strains to use in vaccine development. Here we use published influenza B virus (IBV) genomes and Next Generation Sequencing (NGS) data from IBVs grown in the presence of antibodies to study how IBV genomes change.

It is apparent that some regions of viral genomes are less plastic than others. Strong selective pressures result in the maintenance of these regions which are fixed in the population. Other parts of viral genomes are variable, and this variability allows the viral population to cope with environmental challenges. Several respiratory viruses compete in the same host environment and it is hypothesized that each virus must constantly adapt, evolve and proliferate for the species to survive.

Viral genomes are generally small and tightly packed with information. This allows us to more easily discern the constant and variable regions. The proteins that dominate the outside of the influenza virion, hemagglutinin (HA) and neuraminidase (NA), have antigenic sites that evolve in response to immune pressure. Other regions of the genome encode proteins that are less exposed to antibodies but may be subject to other selective pressures. The polymerase proteins, for example, are critical for virus replication and must retain sequences that allow the proteins to interact and function.

Influenza viruses replicate quickly which facilitates experimental design and the study of evolutionary trends. The virally encoded RNA-depended RNA polymerases is more error prone than DNA polymerases. Influenza virus particles contain negative-sense genomes (gRNA). During infection the negative-sense gRNA is used as a template for generating both mRNAs for protein production and cRNAs which serve as templates for new genomes. Genomic mutations can occur during transcription of the cRNA or transcription of new gRNA. Influenza genomes are also segmented. This means that errors on different segments can be packaged together, or separately with other accurately transcribed segments, resulting in a more diverse viral populations.

Zoonotic events have led to the establishment of influenza A viruses (IAVs) with different HA and NA combinations in humans. These viruses adapt and evolve and can proliferate or become extinct. In contrast, IBVs are endemic in humans and are not widely found in other species. IBVs have evolved into two antigenically distinct clades that have co-circulated for several decades. Whole genome sequencing and NGS data provide swaths of information for analyzing genomic change.

Analyzing how IBVs have changed over time, and how IBV populations change under antigenic pressures, has provided insight into the what parts of the genome are important for viral viability and evolution. Viruses that have mutations resulting in diverse antigenic sites have the potential to escape immune pressure. Purifying selection results in a more homogenous viral population better adapted to the host environment. However, heterogenous viral genomes are still needed so that the virus can adapt, evolve and proliferate in new hosts. Ideally the quantity of new genomic variants should be sufficient so that they are available if needed but not so abundant as to risk extinction of the most fit genomes. We have previously shown synonymous changes occur in specific regions of the IBV genome. Here we analyze nucleotide composition of different types of influenza genomes and polymerase error frequency.

Materials and Methods

The dinucleotide frequency tables for the three influenza groups known to infect humans (influenza A, B and C), three mammalian genomes (human, pig and bat) and one avian genome (chicken) were obtained from codon usage tables (https://hive.biochemistry.gwo.edu/cuts/about). Sequence source information for viruses is provided in Table 1. The relative synonymous codon usage (RSCU) and nucleotide composition was calculated for each genome using CAIcal (http://genomes.urv.es/CAIcal). Results were exported to Microsoft Excel files to generate graphs. Polymerase error frequencies for IAVs are from Cheung PP, et al. (RNA 2015 PMID: 25404565). Error frequencies for IVB were calculated from deep sequence variant frequencies from Plant EP, et al. (PLoS ONE 2020 PMID: 32990724).

Virus Type	Strain Name	Sequence Source
B ancestral	B/Lee/40	Influenza Research Database
B ancestral	B/Russia/69	Influenza Research Database
B Yamagata	B/Phuket/3073/2013	GISAID EPI_ISL_166957
B Victoria	B/Colorado/06/2017	NCBI (txid1987257)
H1N1	A/Brevig Mission/1/1918	NCBI (txid88776)
H1N1	A/New Jersey/11/1976	Influenza Research Database
H1N1	A/New Caledonia/20/1999	GISAID EPI_ISL_22626
pdmH1N1	A/California/7/2009	GISAID EPI_ISL_391380
H3N2	A/Aichi/2/1968	Influenza Research Database
H3N2	A/Brisbane/10/2007	Influenza Research Database
C	C/Ann Arbor/1/50	Influenza Research Database
C	C/Victoria/2/2012	Influenza Research Database

Table 1. Influenza strain name and source location for sequences used for nucleotide and dinucleotide composition and RSCU analyses.

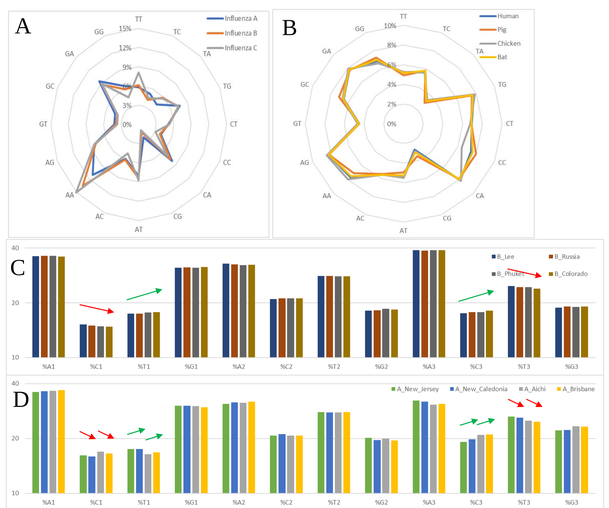

Figure 1. Genome nucleotide composition and temporal trends. The dinucleotide composition from HIVE-CUT reference sequences for influenza viruses (A), and mammalian and chicken genomes (B). The percentage of nucleotides at each codon position for four temporally distinct influenza B viruses (C) and two H1N1 and H3N2 viruses (D).

Results and Discussion

The nucleotide composition of influenza viruses does not closely match the composition of the host species (Figure 1A and 1B). The genomic nucleotide composition has changed for influenza B and influenza A (H1N1 and H3N2) genomes over time (Figure 1C and 1D). A reduction of cytosines at the first position and an increase at the third position over time is observed for all three viruses but these trends are not due to similar RSCU preferences.

For example, the phenylalanine codons (TTT and TTC) for B Phuket and B Colorado diverge in different directions from the ancestral strains (Figure 2A). Where the trends for influenza B viruses do remain consistent over time, they may differ from the influenza A and C trends (Figure 2B). For example, compare the proline codons CCX. This suggests similar genome evolution among influenza types is not tightly constrained by RSCU.

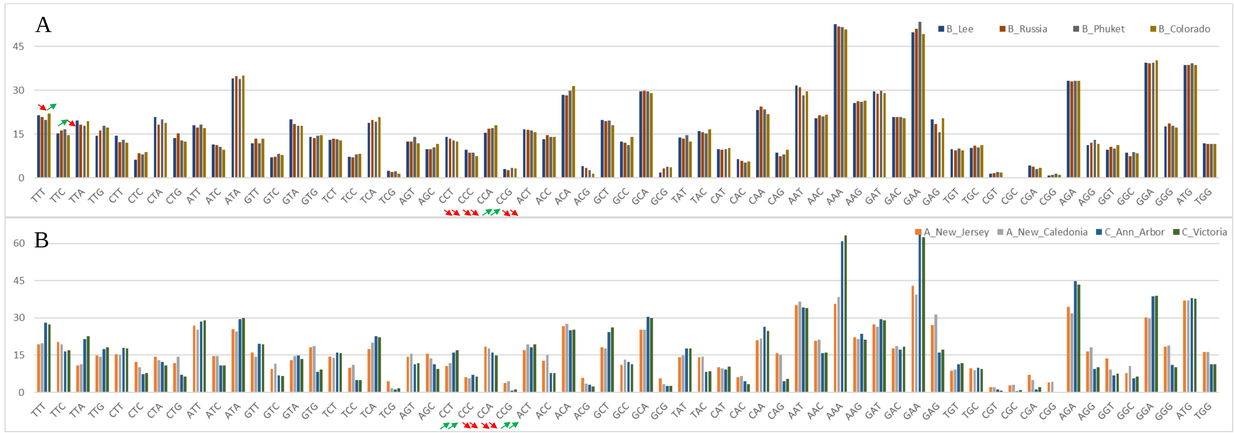

Figure 2. RSCU for four temporally distinct influenza B viruses (A) and two genomes from the influenza A and C virus strains (B)

The frequency of polymerase error in IBVs differs from that in IAVs (Table 2). For example, G-to-A and C-to-U transitions are more frequent in IBVs. Error frequencies calculated from NGS data differ from those calculated by tabulating changes from an ancestral strain (compare 'NGS' and 'vs RS' in Table 2). The difference between B/Phuket and B/Colorado frequencies is greater using *in vitro* NGS data than *in silico* sequence data indicating different polymerase error rates.

Nucleotide Change	H3 mean (Cheung)	H5 mean (Cheung)	B/PK (NGS)	B/CO (NGS)	B/PK (vs RS)	B/CO (vs RS)
G to A	0.1325	0.1175	0.224	0.226	0.244	0.251
G to U	0.042	0.016	0.014	0.009	0.017	0.021
G to C	0	0.0125	0.004	0.005	0.003	0.004
A to G	0.3915	0.44	0.248	0.223	0.253	0.266
A to U	0.003	0.0095	0.022	0.037	0.026	0.015
A to C	0.014	0.013	0.037	0.040	0.031	0.023
U to G	0.014	0.022	0.011	0.018	0.016	0.004
U to A	0.011	0	0.027	0.024	0.016	0.021
U to C	0.3355	0.2875	0.203	0.176	0.176	0.185
C to G	0.0085	0.009	0.005	0.008	0.008	0.006
C to A	0.003	0	0.027	0.037	0.039	0.037
C to U	0.0845	0.097	0.178	0.196	0.172	0.169

Table 2. Average polymerase error frequencies for H3 and H5 IAVs (from Cheung et al., 2014) are shown. Error frequencies for B/Phuket/3073/2013 (B/PK) and B/Colorado/06/2017 (B/CO) calculated from NGS data and from phylogenetic differences from B/Russia/69 (vs RS) are shown.

Conclusion

Influenza viruses evolve and adapt within an ecological niche. IAVs have jumped the species barrier, proliferated, and in some instances gone extinct. During the same timeframe IBVs have evolved, dividing into new antigenic lineages that have remained endemic in humans. Clues to the robustness of IBVs are evident in genomic information.

All influenza viruses have a similar dinucleotide composition that differs from the host composition. Changes in nucleotide composition over time are also similar among influenza viruses (Figure 1) but the changes in codon usage differ (Figure 2). This indicates trends in polymerase error are not limited by codon usage.

There are differences in polymerase errors between influenza types and lineages. Differences in IAV and IBV host cofactors may account for some differences between influenza types. IBV polymerase fidelity values from *in vitro* data have a wider range than values extrapolated from phylogenetic comparisons. This likely reflects both limited susceptible loci and the overrepresentation of variants subject to selective pressures in databases.

Our analysis of NGS data from cultured IBVs shows that the potential for IBV adaptation and evolution lies in the quasispecies population. We have previously shown that synonymous change frequently occurs at specific loci in different IBV lineages (Plant et al., 2020). Identifying IAV loci that are more susceptible to mutation would inform our understanding of why some types of influenza that compete for the same host environment are more robust (and remain endemic) and why others become extinct.