Figures and figure supplements

Stochastic processes constrain the within and between host evolution of influenza virus

John T McCrone et al
**Figure 1.** Within-host diversity of IAV populations. (A) Boxplots (median, 25th and 75th percentiles, whiskers extend to most extreme point within median ±1.5 x IQR) of the number of viral genomes per microliter transport media stratified by day post symptom onset. Notches represent the approximate 95% confidence interval of the median. (B) Boxplots (median, 25th and 75th percentiles, whiskers extend to most extreme point within median ±1.5 x IQR) of the number of iSNV in 249 high quality samples stratified by day post symptom onset. (C) The number of iSNV in each isolate stratified by vaccination status. The red lines indicate the median. (D) Location of all identified iSNV in the influenza A genome. Mutations are colored nonsynonymous (blue) and synonymous (gold) relative to that sample’s consensus sequence. Mutations are considered nonsynonymous if they are nonsynonymous in any known influenza open reading frame. Triangles signify mutations that were found in more than one individual in a given season.

DOI: https://doi.org/10.7554/eLife.35962.003
Figure 1—figure supplement 1. Sequence coverage for all samples. For each sample, the sliding window mean coverage was calculated using a window size of 200 and a step of 100. The distributions of these means are plotted as box plots (median, 25th and 75th percentiles, whiskers extend to most extreme point within median ±1.5 x IQR) where the y-axis represents the read depth and the x-axis indicates the position of the window in a concatenated IAV genome.

DOI: https://doi.org/10.7554/eLife.35962.004
Figure 1—figure supplement 2. Approximate maximum likelihood trees of the concatenated coding sequences for high quality H1N1 samples. The branches are colored by season; the tip identifiers are colored by household. Arrows with numbers indicate consensus and putative minor haplotypes for samples with greater than 10 iSNV. Trees were made using FastTree (Price et al., 2010). ‘Caledonia’ refers to A/New Caledonia/20/1999(H1N1). ‘H1N1’ refers to sequence from a plasmid clone derived from a clinical sample corresponding to A/California/07/2009.

DOI: https://doi.org/10.7554/eLife.35962.005
Figure 1—figure supplement 3. Approximate maximum likelihood trees of the concatenated coding sequences for high quality H3N2 samples. The branches are colored by season; the tip identifiers are colored by household. Arrows with numbers indicate consensus and putative minor haplotypes for samples with greater than 10 iSNV. Trees were made using FastTree (Price et al., 2010). "Wisconsin_2005" refers to A/Wisconsin/67/2005(H3N2). "Perth", "Victoria", and "HK" refer to sequences from plasmid clones derived from clinical samples corresponding to A/Perth/16/2009, A/Victoria/361/2011, A/Hong Kong/4801/2014, respectively.

DOI: https://doi.org/10.7554/eLife.35962.006
Figure 1—figure supplement 4. The effect of titer on the number of iSNV identified. (A) The number of iSNV identified in an isolate (y-axis) plotted against the titer (x-axis, genomes/μl transport media).
DOI: https://doi.org/10.7554/eLife.35962.007
Figure 2. Within-host dynamics of IAV. (A) Histogram of within-host iSNV frequency in 249 high quality samples. Bin width is 0.05 beginning at 0.02. As in Figure 1, mutations were classified as nonsynonymous (blue) if they were nonsynonymous in any known influenza reading frame. Synonymous mutations are gold. (B) The within-host frequency of nonsynonymous mutations in HA stratified by whether or not they are in known antigenic sites (p=0.46 Wilcoxon rank sum). (C) The global frequency of putative antigenic minority iSNV identified in our cohort that have circulated at frequencies above 5% globally since their time of collection. Each variant is labeled according the H3 numbering scheme. The dashed line indicates when samples were collected. Frequency traces are faded prior to the collection date. (D) Timing of sample collection for 43 paired longitudinal samples relative to day of symptom onset. Of the 49 total, 43 pairs had minority iSNV present in either sample. (E) The change in frequency over time for minority iSNV identified for the paired samples in (A). Nonsynonymous and synonymous iSNV are plotted separately. Mutations are colored according to whether they were detected in both isolates (blue), detected only the first isolate (red), or detected only in the second isolate (yellow). The threshold of detection was 2%. The arrows indicate mutations in known antigenic sites.

DOI: https://doi.org/10.7554/eLife.35962.012
Figure 2—figure supplement 1. (A) Reproducibility of iSNV identification for paired samples acquired on the same day. The x-axis represents iSNV frequencies found in the home-acquired nasal swab. The y-axis represents iSNV frequencies found the clinic-acquired combined throat and nasal swab. Dashed line is a one to one expectation. (B) Frequency of all minority iSNV as determined from replicate RT-PCR and sequencing libraries (see Materials and methods). Data are stratified by viral load of original samples. Note that samples with viral loads > 10^5 were not run in duplicate and those with viral loads < 10^3 were not used in the study. Dashed line is a one to one expectation.

DOI: https://doi.org/10.7554/eLife.35962.013
Figure 3. Between host dynamics of IAV. (A) The distribution of pairwise L1-norm distances for household (blue) and randomly-assigned community (gold) pairs. The bar heights are normalized to the height of the highest bar for each given subset (47 for household, 1592 for community). The red line represents the 5th percentile of the community distribution. (B) Timing of symptom onset for 52 epidemiologically linked transmission pairs. Days of symptom onset for both donor and recipient individuals are indicated by black dots. Dashed lines represent pairs that were removed due to abnormally high genetic distance between isolates, see (A). (C) The frequency of donor iSNV in both donor and recipient samples. Frequencies below 2% and above 98% were set to 0% and 100% respectively. (D) The presence-absence model fit compared with the observed data. The x-axis represents the frequency of donor iSNV with transmitted iSNV plotted along the top and nontransmitted iSNV plotted along the bottom. The line represents the predicted probability of transmission given by the presence-absence model with a mean bottleneck of 1.68. The shaded regions represent the 95% confidence interval. Black points on the plot represent the probability of transmission estimated as the proportion of iSNV transmitted within a sliding window of width 5% and a step of 1%. The error bars represent the 95% confidence interval and were derived from a binomial distribution as in (Sobel Leonard et al., 2017). Only those windows with more than 5 iSNV are plotted. Blue curve shows the probability of transmission at a given frequency given a bottleneck size of 10 in the presence-absence model. (E) The beta-binomial model fit. Similar to (D), except the predicted outcomes are based on a beta-binomial model using a mean bottleneck of 1.75. Blue curve shows the probability of transmission at a given frequency given a bottleneck size of 10 in the beta-binomial model.

DOI: https://doi.org/10.7554/eLife.35962.019
**Figure 3—figure supplement 1.** Estimate of effective bottleneck size with relaxed variant calling criteria. (A) The frequency of iSNV in both recipient and donor isolates. iSNV were identified using the original variant calling pipeline as in (McCrone and Lauring, 2016). (B) The presence-absence model fit as in Figure 3D with relaxed variant calling.

DOI: https://doi.org/10.7554/eLife.35962.020