Influenza Virus Evolution Within and Between Human Hosts

by

John T. McCrone IV

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Microbiology and Immunology)
in The University of Michigan
2018

Doctoral Committee:

Assistant Professor Adam Lauring, Chair
Assistant Professor Irina Grigorova
Professor Michael Imperiale
Professor Aaron King
Professor Pat Schloss

John Thomas McCrone IV

mccrone@umich.edu

ORCID iD: 0000-0002-9846-8917

To my family.

ACKNOWLEDGEMENTS

Thank you Adam Lauring for your patience, exciting ideas, and sound advice. Your mentorship and guidance has prepared me to think critically and deeply. I am a better scientist and a better person for having worked in your lab.

Thank you Will Fitzsimmons, M.S. for your friendship; Danny Lyons for sharing the excitement of virus evolution; and Kayla Peck for broadening my view of evolutionary models.

I am grateful to Matt Pauly for his example and discussions. Thank you to Kari Debbink for her contributions to Chapter III and her birthday cakes, and to Daniel Jorge for his kind ear.

Thank you to Bob Woods, who contributed so much of his time talking about modeling and evolution. Thank you for keeping your door open and the coffee pot full.

I would not be the scientist I am today without Pat Schloss's example. Aaron King's class changed how I code, and opened my mind to modeling stochastic processes. Thank you. Thank you also to Irina Grogoriva and Mike Imperiale, for their suggestions and critiques.

To all lab members past and present, to all those who have made the department such a full and friendly environment, and to everyone who has specifically invested in me and this research, thank you.

Finally, I am grateful for my wife Ellie and our children Peter and Anne. Without your encouragement and support this thesis would be worthless. Thank you for making my life so rich and joyful.

This research was supported by the NIH national Research Service Award T32GM00744 from the National Institute of General Medical Sciences, The University of Michigan Endowment for the Development of Graduate Education, and by grants to Adam Lauring from the University of Michigan and NIH.

TABLE OF CONTENTS

LIST OF FIGURES

Figure

LIST OF TABLES

Table

ABSTRACT

Rapid adaptive evolution significantly contributes to the size and severity of seasonal influenza epidemics. While influenza evolution has been well defined at the global scale, these dynamics ultimately derive from processes that take place within and between infected individuals. The dynamics of influenza evolution within and between individual hosts are poorly understood.

In my thesis, I have applied an empirically-validated, next-generation sequencing approach to over 300 patient-derived samples from two vaccinated cohorts to define influenza evolution within and between naturally infected individuals. I compared influenza diversity between vaccinated and unvaccinated individuals enrolled in the FLUVACS study, the last randomized, placebo-controlled trial of influenza vaccine efficacy. Phylogenetic analysis of consensus hemagglutinin and neuraminidase sequences showed no stratification by pre-season HAI and NAI titer, respectively. Additionally, within-host diversity did not significantly vary with day of sampling, vaccination status, or pre-season antibody titer. Contrary to what has been suggested in experimental systems, these data indicate that seasonal influenza vaccination has little impact on intrahost diversity in natural infections and that vaccine-induced immunity may be only a minor contributor to antigenic drift at local scales.

In the second study, I used quantitative models to define influenza virus dynamics in individuals enrolled in a prospective, community-based cohort. Sequence data from 35 serially sampled individuals suggested that within-host populations are dy-

namic and not shaped by antigenic selection. Classical population genetic models showed these dynamics were consistent with a within-host effective population size of 30-70 and an *in vivo* mutation rate of 4×10^{-5} per nucleotide. Additionally, I characterized the between-host effective transmission bottleneck in 43 epidemiologically linked and genetically-validated transmission pairs. Maximum likelihood optimization of multiple transmission models estimated an effective transmission bottleneck of 1-2 genomes.

These data suggest that contrary to the global dynamics, positive selection is inefficient at the level of the individual host. Genetic drift and other stochastic processes likely dominate the host-level evolution of influenza viruses.

CHAPTER I

Introduction

Note: The section entitled 'Genetic bottlenecks in intraspecies virus transmission' is a modified version of the published article:

McCrone JT, Lauring AS. 2017. ScienceDirect Genetic bottlenecks in intraspecies virus transmission. Current Opinion in Virology 28:2025.

Overview

Influenza virus causes significant morbidity and mortality worldwide through seasonal epidemics. The rapid evolution of the virus results in decreased drug and vaccine efficacy, as well as ineffective long-term immunity. Despite a well-developed, global, surveillance network, accurate forecasting and vaccine strain selection are difficult. Furthermore, vaccine efficacy remains low even when the vaccine strain and circulating strain are antigenically similar (Ohmit et al., 2008). The high attack rates and low vaccine efficacy that characterize influenza epidemics suggest there is a gap between our understanding of influenza evolution and the forces that influence infection at the host level. Combating the rapid evolution of influenza virus requires a full understanding of the evolutionary forces driving the process at all biological scales.

Influenza virus is an enveloped, negative-sense, RNA virus. Its genome consists of eight genomic segments that encode for twelve known proteins (Petrova and Russell, 2017). The three largest segments, basic polymerase 2 (PB2), basic polymerase 1 (PB1), and polymerase acidic (PA), encode the heterotrimeric RNA dependent RNA polymerase that is responsible for replicating the viral genome. PB2 and PA also encode two accessory proteins PB2-F2 and PA-X. The nucleoprotein (NP) which coats and organizes the genomic segments is encoded on the NP segment. The surface proteins hemagglutinin (HA) and neuraminidase (NA) are encoded on their own segments and are responsible for cell attachment and release. The M segment encodes two proteins M1 and M2 that act as the viral matrix protein and ion channel respectively. The nuclear export protein (NEP), which mediates the nuclear export of genomic complexes and NS2, which interferes with the innate immune response are encoded on the smallest segment, NS.

The study of influenza evolution has historically focused on hemagglutinin, which is the primary determinant of antigenicity (Caton et al., 1982; Wiley et al., 1981; Skehel et al., 1985). The accumulation of mutations in the antigenic epitopes of HA leads to episodic shifts in virus antigenicity, a process known as antigenic drift (Koelle et al., 2006). Antigenic mapping approaches have shown positive selection drives antigenic drift. Over time, circulating viruses have explored novel regions of antigenic space without returning to previously occupied areas (Smith et al., 2004).

Despite the undeniable role of positive selection at the global scale, phylogenetic analysis has shown that very little antigenic drift occurs within a localized epidemic. Instead, stochastic forces, such as migration of distinct lineages into the area and within-clade reassortment dominate local evolutionary dynamics (Nelson et al., 2006). The current model of influenza evolution accounts for the differences

between scales by partitioning the globe into regions, which play distinct roles in a source-sink model. Persistently-infected, tropical areas act as the source of antigenic evolution, while spillover from these regions seeds annual epidemics in more temperate zones (which constitute the sink) (Rambaut et al., 2008; Bedford et al., 2015). Although the global dynamics of influenza evolution have been well characterized, the underlying evolutionary forces driving these dynamics are not fully understood. Mutations that ultimately dominate the global population must first arise within and transmit between individual hosts. Influenza evolution within and between hosts is not well understood.

In my thesis, I have used next-generation sequencing of representative, patient-derived isolates to characterize influenza evolution within and between individuals.

Challenges for using deep sequencing in studies of virus evolution

Advances in next-generation sequencing have revolutionized the study of viral evolution and population genetics. It is now feasible to efficiently sequence patient isolates to a sufficient depth of coverage to identify rare mutations present in the population. Practically speaking, this technological advancement has removed the need to passage isolates in cell culture prior to sequencing. It is now possible to explore the dynamics of within-host virus populations directly from patient isolates.

While novel sequencing technologies have ushered in a new era of viral population genetics, these approaches are not error-free. In addition to sequencing errors, library preparation includes a number of error-prone reverse-transcription and PCR steps. These errors are often overlooked in studies of virus evolution and confound the accurate study of virus populations within human and animal hosts. As technological advances have rapidly opened the door to new and exciting avenues of research,

accounting for previously overlooked sources of error has become a rare but vital part of studying virus evolution in naturally occurring populations.

Error profiles in deep sequencing approaches

Each high-throughput sequencing method has its own short-comings and tendencies. Here, I focus on the Illumina platform as it is commonly used in identifying low-frequency mutations present in virus populations.

Briefly, Illumina sequencing begins by binding small fragments of cDNA to a flow cell and amplifying those segments to create small clusters of identical fragments. Each cluster's sequence is 'read' using a 'sequencing by synthesis' approach in which nucleotides containing reversible terminators and flurophores are added in a step-wise manner. At each step, the identity of the added nucleotide is determined photographically. The terminator is removed; the next base is added to the nascent strand, and the cycle repeats (Bentley et al., 2008). This process is highly efficient and can generate up to 300 million paired-end reads per sequencing lane.

Illumina sequencing has greatly enhanced our ability to sequence virus populations, but it is not perfect. Error rates can reach as high as 1% in some experimental set-ups (Schirmer et al., 2015). Additionally, the platform is plagued by biases that must be accounted for in studies of virus populations. As mentioned above, Illumina sequencers 'read' base calls from images of flow cells. The density and diversity of such clusters can vary from run to run resulting in variable error rates and batch effects. Furthermore, sequencing errors are not distributed uniformly across the sequencing read or the sample genome. Errors are disproportionately found near the end of reads (Kircher et al., 2009), in GC rich regions (Dohm et al., 2008), and after specific sequence epitopes and structures (Nakamura et al., 2011). Importantly, a significant proportion of these errors are base substitutions, which are indistinguishable

from the real single nucleotide variants (SNV) present in the population (Bentley et al., 2008)

Variant calling approaches

There are many methods to help identify and eliminate sequencing errors from studies of viral diversity. Bioinformatic and statistical approaches rely on estimates of sequence quality to identify true SNV from false positive sequencing errors (Flaherty et al., 2011; Gerstung et al., 2012; Isakov et al., 2015; Wilm et al., 2012; Macalalad et al., 2012; Koboldt et al., 2009; Watson et al., 2013; Koboldt et al., 2012; Yang et al., 2013; Gerstung et al., 2014). These approaches are based solely on sequence data and can be applied to a wide range of datasets regardless of experimental design. They have been shown to control for erroneous substitutions with a high level of accuracy when applied to *in silico* datasets or libraries prepared under favorable conditions.

Varscan and LoFreq are two commonly-used, statistical, variant callers (Koboldt et al., 2012, 2009; Wilm et al., 2012). Varscan takes a relatively simple approach, which requires that a putative SNV pass a number of quality thresholds, such as coverage, frequency, average base quality or Phred score, and position on a read (Koboldt et al., 2009). Lofreq, which was one of the first variant callers designed to identify SNV from high-coverage datasets, is slightly more complicated. It uses the Phred score of a base and a Poisson-binomial model to estimate the probability that sequencing error alone is responsible for a putative SNV (Wilm et al., 2012). Both of these methods were validated by comparing the overlap between SNV identified from two different sequencing runs of the same sample. Neither approach was tested on its ability to identify variants that were known *a priori*, nor were they benchmarked under conditions that mimic those of low-titer, patient-derived, isolates.

Another class of variant calling approaches aims to identify false positive SNV by 'tagging' template nucleic acid before the error-prone steps in library preparation. The presence of these identifiers provides a mechanism for eliminating errors that arise during sample preparation as well as sequencing. Three such approaches are widely used in the field of virus evolution.

Cirseq is a novel method in which template RNA is fragmented and circularized prior to reverse-transcription PCR and library preparation. In this protocol rolling-circle reverse transcription creates tandem cDNA repeats of the same template RNA fragment (Acevedo and Andino, 2014). Any true mutation will exist in all tandem repeats while RT-PCR, PCR, and sequencing errors are likely to occur only once. Cirseq has been reported to accurately detect SNV at frequencies as low as 10^{-5} in cell culture derived populations of poliovirus (Acevedo et al., 2014). However, its application to patient-derived isolates is limited by its poor efficiency and the high quantity of viral RNA required.

Primer ID is another well-known method that uses PCR primers with random bar-codes to tag nucleotide templates. The added bar-codes allow one to track the amplification of template strands during PCR. True SNV will be present in each sequencing read that contains a given bar-code, while errors are likely to be constrained to a small subset of the reads (Jabara et al., 2011). However, given the short read-length of the Illumina platform (125-250 bp) is it only possible to tag small regions of a genome at a time. Therefore, primer ID is an impractical approach for whole genome sequencing.

Sequence-independent single-primer amplification (SISPA) uses a Klenow reaction to shear and tag already amplified cDNA with random bar-codes. The approach does not account for errors that can arise in the first RT-PCR reaction that is often used to

enrich for viral genomes prior to sequencing, but it does control for errors that may be propagated during library preparation (Djikeng et al., 2008). SISPA is easily applied to whole genome sequencing of virus populations and is commonly used in studies of influenza diversity. Despite its low cost and ease of application, SISPA priming is biased in unpredictable ways and produces uneven coverage (Rosseel et al., 2013). The protocol has also been shown to have a much higher error rate than other library preparation methods (Kugelman et al., 2017b).

It is clear that deeply sequencing RNA virus populations from patient-derived samples is an error-prone process. Despite the widespread availability and application of computational and experimental approaches designed to mitigate these errors, the accuracy of sequencing patient-derived isolates was largely unknown when I began my thesis. In order to carefully and accurately study the within and between host dynamics of influenza, I needed to determine the accuracy and limitations of our sequencing approach. This work and its implications for studies of virus diversity are the focus of Chapter II.

Influenza Virus evolution within hosts

Like most RNA viruses, influenza replicates through an RNA dependent RNA polymerase that lacks proof-reading capabilities and has a very high mutation rate (Sanjuán et al., 2010; Pauly et al., 2017). This high error rate is a major contributing factor to the high rates of evolution often attributed to RNA viruses (Sanjuán, 2012). Because of the large number of mutations present in infected hosts, RNA virus populations exist as mutant swarms sometimes referred to as quasispecies (Lauring et al., 2013). The diversity of these populations, specifically the availablity of low-frequency mutations, is thought to be a virulence factor in some systems. (Vignuzzi

et al., 2006; Pfeiffer and Kirkegaard, 2005). In the case of influenza, the relationship among within-host diversity, the host immune response, and vaccine failure is largely unknown.

Since the vast majority of influenza infections are acute and only last about a week, acquiring representative sample isolates to study within-host dynamics has been challenging. Much of what is known regarding influenza evolution at the host level derives from animal models, human challenge experiments, and chronic infections.

Studies of experimental and naturally occurring infections in an equine model of influenza have shown that within-host populations are diverse, dynamic, and characterized by incomplete purifying selection. Clonal sequencing of longitudinal samples from experimentally infected horses failed to find any minority mutations that persisted between time points (Murcia et al., 2013). This rapid turnover of the population is consistent with experimental infections in swine, which showed the rapid fixation of synonymous mutations during infection (Murcia et al., 2012). However, such dynamics contradict findings from similar studies in equine models, which have found that intrahost diversity accumulates over time (Murcia et al., 2010).

Much has been made regarding the presence and transmission of putative antigenic variants in animal studies (Johnson et al., 2017); however, no robust differences have been observed between vaccinated and non-vaccinated subjects (Murcia et al., 2012, 2013). Furthermore, the average dN/dS ratios from these studies is around 0.8. The dN/dS ratio compares the number of nonsynonymous mutations per nonsynonymous sites with the number of synonymous mutations per synonymous sites. Under neutral conditions this ratio is expected to be one. A dN/dS below one suggest the population is under purifying selection. In one study, stop codons were shown to arise within and transmit between hosts (Murcia et al., 2012). Taken together, these data suggest

positive selection for novel antigenicity is rare and negative selection is inefficient at the within-host scale.

Contrary to the qualitative dynamics observed in swine and equine models of influenza, experimental infections in humans provide examples of selection shaping within-host populations. In infections with an egg-adapted virus, nonsynonymous mutations that had accumulated over the course of egg adaptation were rapidly purged (Sobel Leonard et al., 2016). This finding is consistent with the rapid reversions observed in ferrets infected with poorly adapted viruses (Lakdawala et al., 2015).

A quantitative study of the allele trajectories present in the human challenge study mentioned above identified four mutations under the influence of selection. Two of these resulted in nonsynonymous changes in HA and were hypothesized to be evidence of purifying selection acting to revert egg adaptation (Sobel Leonard et al., 2017a). Additionally, a similar, more complex, modeling of the longitudinal swine infections discussed previously (Murcia et al., 2012), estimated nine sites behaved according to non-neutral processes. Of these nine mutations, one was located in an antigenic epitope on HA and seemingly underwent time-dependent selection (Illingworth et al., 2014). These studies offer a proof of principle, suggesting that selective pressure can influence influenza evolution in the context of acute infections.

Studies of chronically infected and immunocompromised individuals offer more evidence of selection shaping within-host populations. Oseltamivir treatment, which is known to select for drug resistance in cell culture (Foll et al., 2014), has also been shown to select for resistant variants in chronically infected individuals (Ghedin et al., 2010). A longitudinal study of 4 chronically infected cancer patients found evidence of parallel evolution across biological scales (Xue et al., 2017). The same

nonsynonymous mutations arose independently of one another in distinct individuals. Interestingly, these mutations were also enriched in viruses that circulated at the global level in the years following sample collection. These findings have been interpreted to suggest that within-host populations may face similar selective pressures as global populations.

Despite the clear role of positive selection in global evolutionary dynamics, and its potential to affect within-host populations, antigenic evolution has not been characterized during representative, naturally-occurring, acute infections. Dinis et al. (2016) found putative antigenic mutations present in acutely infected individuals, including some present in the so-called 'antigenic ridge', which significantly influences antigenicity (Koel et al., 2013). However, these mutations were found in a minority of infections and were present at low frequencies. These data suggest that although putative antigenic variants may be present in acute infections, they do not reach fixation or even significantly higher frequencies than other mutations, even when expected to alter the antigenicity of the virus.

In Chapter III I use samples from a double-blind, placebo-controlled study to directly characterize the role of vaccination and immune selection in shaping the within-host population. In Chapter IV, I use longitudinal samples from acute infections to further explore the factors that influence frequency trajectories within individuals over time.

Genetic bottlenecks in intraspecies virus transmission

As previously discussed, many viral pathogens exist as diverse populations within infected hosts. The diversity present in this 'mutant swarm' provides the raw material on which selection can act. Although populations within a host may reach as high

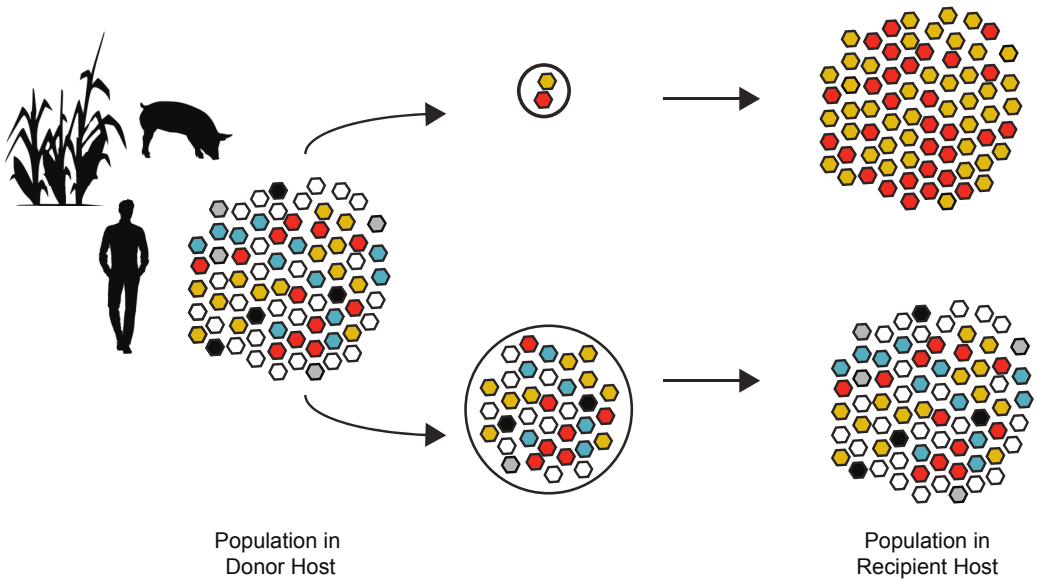


Figure 1.1: The effect of transmission bottlenecks on viral diversity. In a variety of hosts (e.g. humans, pigs, plant shown here), stringent bottlenecks (top) limit the size and diversity of a population and drastically alter their composition. The large populations that pass through loose bottlenecks (bottom) allow for transmission of rare variants. As a result the diversity of the population in the recipient approximates that of the donor.

as 10^{14} virions (Russell et al., 2012), viruses are frequently subject to bottleneck events as they spread within and between hosts (Pfeiffer and Kirkegaard, 2006). These bottlenecks drastically reduce the size of the population and, consequently, its genetic diversity.

Because the population that develops after a genetic bottleneck is derived from a small sample of the ancestral population, this process can dramatically alter the relative frequency of mutations in the population. The stringency of the transmission bottleneck plays an important role in linking within-host processes to a pathogen's larger evolutionary dynamics. Stringent, or tight, transmission bottlenecks limit the diversity of the founding population in the recipient and alter the mutational composition of the population in the recipient relative to that in the donor (Figure 1.1, top). By contrast, if the transmission bottleneck is loose, transmission does not significantly impact variant frequencies and the composition of the founding

population in the recipient more closely matches that present in the donor at the time of transmission (1.1, bottom).

Although transmission bottlenecks play an important role in viral evolution, relatively little is known about their size and determinants. In fact, only one study had estimated the influenza transmission bottleneck in humans prior to my work in Chapter IV. Here, I review what is know about transmission bottlenecks across all viral systems.

Many quantitative studies suggest that bottlenecks are tight (Zwart and Elena, 2015; Gutiérrez et al., 2012); however, there are exceptions and even conflicting reports for viruses with similar transmission pathways. Importantly, the factors that determine the stringency of the transmission bottleneck are poorly understood.

Measuring transmission bottlenecks

Transmission bottlenecks are measured by their effect on viral diversity. In experimental systems, within-host diversity can be approximated using a defined population of viruses that are tagged with genetic markers. If the markers are selectively neutral, the number of distinct markers that pass from donor to recipient reflects the sampling event of the bottleneck as opposed to selection within either host (Figure 1.2A). This technique has been used to qualitatively estimate a stringent bottleneck for aphid transmission of cucumber mosaic virus (an average of 3 of 12 markers were transmitted) (Ali et al., 2006) and aerosol transmission of influenza in ferrets and guinea pigs (2-5 of 100 sequence tags were transmitted) (Varble et al., 2014). In a particularly elegant experiment, Moury and colleagues artificially inoculated aphid vectors with mixtures of 2 Potato Y virus mutants prior to feeding the aphids on pepper plants (Moury et al., 2007). By modeling the number of plants exposed to

only one of the mutants, Moury *et al.* found that aphid transmission imposes a bottleneck of 0.5-3.2 virions on Potato Y virus.

Because natural systems do not offer the opportunity for a bar-coding approach, early studies characterized the transmission bottleneck qualitatively based on the degree of shared diversity found within transmission pairs (Figure 1.2B). Clonal sequencing of influenza virus isolates from swine and equine transmission chains found transmission pairs shared minority variants (Murcia et al., 2012, 2013; Hughes et al., 2012). Studies of aphid, mechanical, and vertical transmission of Zucchini Yellow Mosaic Virus found similar results (Simmons et al., 2012, 2013). These studies suggest that transmission bottlenecks are sometimes sufficiently loose to allow for the transmission of low-frequency mutations.

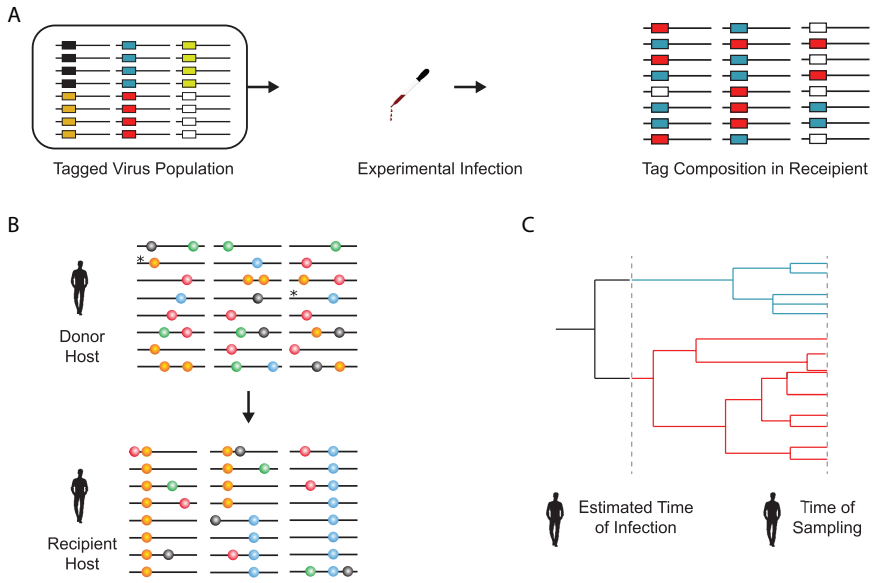


Figure 1.2: Measuring transmission bottlenecks. (A) The number of donor-derived, neutral markers detected in the recipient is an indication of the stringency of the transmission bottleneck. Here, 3 of the 6 markers were transmitted suggesting a stringent bottleneck. (B) Shared diversity data from natural systems can be used to estimate a bottleneck. In the example, only two donor genotypes, denoted with *, were transmitted to the recipient suggesting a stringent bottleneck. Other *de novo* mutations arise on these backgrounds after transmission. (C) Coalescent models allow one to work backward from the time of sampling and estimate the number of genotypes that could plausibly give rise to the observed diversity. In this case, the two lineages are traced back to two genetically distinct variants present at transmission.

More quantitative approaches can also be employed to estimate the transmission bottleneck from shared diversity data. In these models, the transmission process is assumed to be a random sampling of the donor population and individual variants are assumed to be transmitted independently of one another. The probability that a variant is transmitted is derived from a binomial distribution and is positively correlated with its frequency in the donor and the size of the bottleneck. More complexity can be incorporated into these models to tease apart the relative impact of within- and between-host processes (Sobel Leonard et al., 2017b). One such model has been used to estimate a loose bottleneck of roughly 200 genomes in a recent study of human transmission of influenza virus (Sobel Leonard et al., 2017b; Poon et al., 2016). This estimate is much larger than that provided by the bar-code experiments previously discussed. The large discrepancy in these studies highlights the need for a more complete understanding of the viral, host, and environmental factors that determine transmission bottleneck sizes.

When only one member of a transmission pair is available, the diversity present in the infected host can be used to estimate the number of genotypes in the founding population. Coalescent theory works backward in time, tracing the evolutionary history of the current population back to common ancestors (Kingman, 1982). Coalescent models based on the current diversity, the viral evolutionary rate, and the estimated time of infection can be used to determine how many genotypes were present in the founding population (Figure 1.2C). Phylogenetic analysis of HIV evolution suggests that most infections derive from small founding populations of only one genotype (Keele et al., 2008; Edwards et al., 2006). A similar approach has been used to estimate a stringent transmission bottleneck for HCV (Bull et al., 2011; Ho et al., 2017; D'Arienzo et al., 2013; Li et al., 2012).

Determinants of bottleneck size

Most transmission studies suggest tight bottlenecks and small founding populations (see tables in (Zwart and Elena, 2015) and (Gutiérrez et al., 2012)). However, as mentioned above, these estimates can vary significantly depending on the virus, host, route of transmission, and experimental design. Understanding the factors that determine the size of the transmission bottleneck is vital to interpreting the effect transmission has on viral evolution. Work in Tobacco etch virus (TEV) suggests that the size of the bottleneck is dose dependent, with higher exposure doses corresponding to larger founding populations (Zwart et al., 2011). Evidence from mixed infections of influenza virus in a guinea pig model is consistent with a dose dependence model (Tao et al., 2014). Further support comes from experimental infections with tagged influenza clones in ferret and guinea pig models, which indicate that the more limiting exposure dose of aerosol transmission imposes a significantly more stringent bottleneck than contact transmission (Varble et al., 2014). Additionally, coinfection by other pathogens, which can limit innate defenses and modulate the immune response, has been correlated with loose bottlenecks in HIV and HCV (Sagar et al., 2004; Haaland et al., 2009; Fauteux-Daniel et al., 2017). Taken together these data suggest that the transmission bottleneck is not constant, but rather a complex function of viral, host, and environmental factors.

Complicating matters is the observation that segregation of the viral population within the donor can also restrict the amount of diversity transmitted to the recipient. Work in animal models of influenza virus suggest that populations in the upper respiratory tract seed transmission and can be distinct from populations at other sites of infection (Varble et al., 2014; Lakdawala et al., 2015). The stringent bottleneck associated with aphid transmission of cucumber mosaic virus (Betancourt et al.,

2008) is likely the result of extreme viral segregation within the donor. Most plant cells are infected by only one genotype, and aphids are unlikely to feed on many donor cells prior to transmission (Takeshita et al., 2004). Other vector-transmitted viruses undergo an additional bottlenecking event within the vector. Smith et al. (2008) used fluorescent Venezuelan equine encephalitis virus (VEEV) replicons to show that an average of 28 midgut cells in the mosquito are initially infected by the virus. This small population size is consistent with observations of Dengue virus in infected mosquitoes (Lequime et al., 2016) and likely contributes to the stringent bottleneck observed during mosquito-mediated transmission of VEEV in a mouse model (Forrester et al., 2012).

Evolutionary consequences of transmission bottlenecks

Transmission bottlenecks determine the extent to which within-host diversity contributes to evolutionary trends at higher scales. While the relatively high mutation rates and large population sizes of many viruses may allow these pathogens to rapidly adapt to their host, the rate of adaptation is not unlimited. In particular, the rate depends on the effective population size (Rouzine et al., 2001). The effective population size can be roughly thought of as the number of viruses that replicate and contribute genomes to the next generation. It is usually smaller than the census population (Charlesworth, 2009). In large effective populations, selection is efficient, deleterious mutations are purged, and beneficial mutations increase in frequency over time (Willi et al., 2006). However, alleles in small populations are subject to sampling error known as random genetic drift. Drift introduces noise so that selection does not efficiently fix beneficial mutations or purge deleterious ones (Robertson, 1960). Stringent transmission bottlenecks reduce the effective population size of vi-

ral pathogens between hosts, increase genetic drift, and decrease the efficiency of selection.

Stringent transmission bottlenecks may therefore pose a significant barrier to adaptive evolution. Because most mutations are deleterious, repeated bottleneck events fix deleterious mutations and decrease viral fitness over time. This process, known as Muller's ratchet, opposes purifying selection and contributes to the deleterious load often observed at the tips of viral phylogenetic trees (Pybus et al., 2007; Koelle and Rasmussen, 2015). Although the fixation of deleterious mutations decreases fitness along a single transmission chain, it is unlikely to drastically decrease a virus' overall fitness at a global scale. Competition at the interhost level can serve to maintain viral fitness (Bergstrom et al., 1999; Elena et al., 2001). Notably, populations with low fitness are not as susceptible to Muller's ratchet as well-adapted populations with high fitness (Novella et al., 2008, 1995).

While transmission bottlenecks are expected to slow adaptive evolution, they may provide potential advantages to evolving pathogens. Stringent bottlenecks purge the population of defective interfering particles, which limit viral replication (Visser et al., 1999). Bottlenecks also increase genetic drift and provide a mechanism for virus populations to traverse potential fitness valleys and escape local fitness maxima (Rozen et al., 2008).

The available data suggest that transmission frequently imposes a stringent bottleneck that dramatically reduces the level of diversity in the founding population. In many cases, however, transmission bottlenecks appear to be sufficiently wide to transmit minority variants. A more complete understanding of viral transmission bottlenecks is necessary to link within-host population dynamics to larger evolutionary trends.

In Chapter IV I use samples from transmission pairs enrolled in a large, representative, household-based, prospective cohort to define the influenza transmission bottleneck.

CHAPTER II

Measurements of intrahost viral diversity are extremely sensitive to systematic errors in variant calling

Note: This chapter is a modified version of the published article:

McCrone JT, Lauring AS. 2016. Measurements of intrahost viral diversity are extremely sensitive to systematic errors in variant calling. J Virol 90:JVI.00667166895.

Introduction

Many viral pathogens are thought to exist as a cloud of closely related mutants within an infected individual (Lauring et al., 2013). Until recently, our understanding of intrahost viral dynamics and the impact of viral diversity on evolution and pathogenesis have been limited by low-throughput sequencing methods. However, with the advent of next generation sequencing (NGS), it is now feasible to sequence patient-derived samples at sufficient read depth to detect rare single nucleotide variants (SNV). There has been an explosion of studies that employ NGS to quantify viral diversity within and between hosts (e.g. (Andersen et al., 2015; Grubaugh et al., 2015; Rogers et al., 2015; Poon et al., 2016; Olp et al., 2015; Kugelman et al., 2015; Lakdawala et al., 2015; Van Slyke et al., 2015; Cuevas et al., 2015; Gire et al., 2014; Kundu et al., 2013; Lauck et al., 2012)). Although NGS produces the large quantities of sequence data needed to detect rare variants, the process is error prone (Lam

et al., 2011; Schirmer et al., 2015; Nakamura et al., 2011), and many bioinformatics tools do not explicitly address the challenges inherent in studies of patient-derived viral populations.

A number of sample preparation protocols have been developed to control for the errors in NGS-based studies of virus populations, but each approach has its own caveats that ultimately limit its application. Cirseq is an ingenious technique in which template RNA is sheared and circularized prior to reverse transcription (Acevedo et al., 2014; Acevedo and Andino, 2014). Subsequent rolling circle cDNA synthesis produces tandem reads, generating a consensus sequence for each RNA fragment. While this method is likely to be highly sensitive for rare variant detection and can control for reverse transcription, PCR, and sequencing errors, the requirement for a large and relatively pure population of viral RNA limits its applicability to patient-derived samples (Acevedo et al., 2014; Acevedo and Andino, 2014). "Primer ID" methods require less input and target sequencing to the viral genome. This approach relies on barcoded primers to construct consensus sequences for each cDNA template and can control for PCR and sequencing errors (Jabara et al., 2011; Zhou et al., 2015). Because Primer ID methods require that each bar code be physically attached to a PCR product, they are most easily applied to small, targeted regions of the genome. As such, they have limited application in whole genome sequencing.

Sequence independent single primer amplification (SISPA) is an alternative approach that allows for whole genome sequencing and controls for errors propagated during library preparation (Djikeng et al., 2008). In this method RT-PCR products are sheared and tagged with bar-coded random primers in a Klenow reaction, prior to library preparation. SISPA controls for any errors that may arise during library

amplification, including PCR biases. This method has been used in conjunction with statistical algorithms to control for accuracy in studies of intrahost influenza diversity (Poon et al., 2016; Nelson et al., 2014; Ghedin et al., 2010). However, the bar-coding reaction used in SISPA can be biased in unpredictable ways, resulting in uneven coverage and sensitivity across the genome (Rosseel et al., 2013).

Statistical algorithms have also been developed to distinguish true variants from sequencing errors (Flaherty et al., 2011; Gerstung et al., 2012; Isakov et al., 2015; Wilm et al., 2012; Macalalad et al., 2012; Koboldt et al., 2009; Watson et al., 2013; Koboldt et al., 2012; Yang et al., 2013; Gerstung et al., 2014). These methods rely solely on sequencing data, and are more easily applied to whole genome sequencing. In general, variant calling algorithms calculate base-specific error rates using various metrics including, but not limited to, mapping quality (MapQ), base quality (Phred), strand bias, and sequence context. True variants are identified as those with frequencies exceeding the expected error rate according to some predetermined statistical test. Despite being employed in many NGS-based studies of viral diversity (e.g. (Rogers et al., 2015; Combe et al., 2015; Beck et al., 2014; Grubaugh et al., 2015; Kundu et al., 2013) few of these algorithms have been benchmarked using defined viral populations. To our knowledge, none have been tested under conditions that mimic those found in patient-derived samples. The accuracy of such algorithms in the context of NGS studies of patient-derived viral populations is largely unknown.

Here, we use genetically defined populations of influenza A virus with variable input titers to determine the accuracy of rare variant detection in patient-derived samples. We highlight the challenges that accompany NGS-based studies of viral diversity and include a means for improved accuracy. This work exemplifies the

controls that should be run prior to any NGS-based study of viral populations and provides a comprehensive data set for benchmarking other pipelines.

Methods

Viruses and cells

Madin-Darby canine kidney (MDCK) cells were provided by Arnold S. Monto (University of Michigan School of Public Health) and were maintained in Dulbecco's Modified Eagle Medium (DMEM, Invitrogen) with 10% fetal bovine serum (Gibco and HyClone), 25mM HEPES (Invitogen), and 0.1875% bovine serum albumin (Life Technologies). Influenza A/WSN/33(H1N1) virus was rescued from transfected cells using the 8 plasmid reverse genetic system containing each genomic segment (pHW181-188), a kind gift from Robert Webster (St. Jude's Children's Research Hospital) (Hoffmann et al., 2002; Pauly and Lauring, 2015). A biological clone of influenza A/Puerto Rico/8/1934(H1N1) was obtained from ATCC (VR-1469) and the genomic segments were cloned into the pHW2000 reverse genetic system (Pauly and Lauring, 2015). The sequences of these clones were verified using Sanger sequencing.

Patient-derived samples of influenza A virus were collected as part of the Household Influenza Vaccine Effectiveness (HIVE) study (Ohmit et al., 2015; Monto et al., 2014) and kindly provided by Arnold S. Monto and colleagues at the University of Michigan School of Public Health. The HIVE study was approved by the Institutional Review Board at the University of Michigan, and all subjects provided informed consent.

Viral populations

We extracted viral RNA from infected supernatants using QIAamp Viral RNA kits (Qiagen) and generated cDNA using Superscript III one-step with HiFi platinum

Taq (Invitrogen). PCR products were purified using the GeneJet PCR Purification Kit (ThermoFisher Scientific) according to the manufacturer's instructions.

PR8-WSN33 population

For the experiment in Figure 2.2, WSN33 and PR8 viruses were plaque purified and passaged three times in MDCK cells. We then verified the sequences of these viruses by Sanger sequencing. Two microliters of RNA template were used to generate cDNA in eight segment-specific one-step RT-PCR reactions with 0.2 M of the following primers :

PB2-Forward-JT (5'-GCAGGTCAATTATATTCAATATGGAAA-3'),

PB2-Reverse-JT (5'-CAAGGTCGTTTTTAAACTATTCGACAC-3'),

PB1-Forward-JT (5'-GCAGGCAAACCATTTGAATGG-3'),

PB1-Reverse-JT (5'-CAAGGCATTTTTTCATGAAGGACAAG-3'),

PA-Forward-JT (GCAGGTACTGATTCAAAATGGAAG-3'),

PA-Reverse-JT (CAAGGTACTTTTTTGGACAGTATGG-3'),

NA-Forward-JT 5'-(GCAGGAGTTTAAATGAATCCAAACC-3'),

NA-Reverse-JT (5'-CAAGGAGTTTTTTGAACAAACTACTTG-3'),

HA-Forward-JT (5'-GCAGGGGAAAATAAAAACAACCAAAAT-3'),

HA-Reverse-JT (5'-CAAGGGTGTTTTTCCTTATATTTCTGAA-3'),

NP-Forward-JT (5'-GCAGGGTAGATAATCACTCACAG-3'),

NP-Reverse-JT (5'-CAAGGGTATTTTTCTTTAATTGTCGTACT-3'),

M-Forward-JT (5'-GCAGGTAGATATTGAAAGATGAGTC-3'),

M-Reverse-JT (5'-CAAGGTAGTTTTTTACTCCAGCTCT-3'),

NS-Forward-JT (5'-GCAGGGTGACAAAGACATAATG-3'),

NS-Reverse-JT (5'-CAAAGGGTGTTTTTTATTATTAAATAAGCTG-3').

Reaction conditions were 50°C (60 min), 94°C (2 min), followed by 30 cycles of 94°C (30 sec), 54°C (30 sec), and 68°C (3 min). Molar equivalents of each PCR product were pooled to generate reconstituted cDNA genomes of both WSN33 and PR8. The WSN33 cDNA pool was then serially diluted into the PR8 cDNA pool to yield WSN33-PR8 mixtures in which WSN33 made up 5, 2.5, 1.25, 0.63, and 0.16% of the population. Seven hundred and fifty nanograms of each mixture were sheared to an average size of 300-400 bp using a Covaris S220 focused ultrasonicator with the following settings: Intensity: 4, Duty cycle: 10%, Burst/second: 200, Duration: 80 seconds. Sequencing libraries were prepared from these fragmented products using the NEBNext Ultra DNA library prep kit (NEB), Agencourt AMPure XP beads (Beckman Coulter), and NEBNext multiplex oligonucleotides for Illumina (NEB). Pooled libraries were sequenced on an Illumina MiSeq machine with 2 x 250 paired end reads. A clonal plasmid control library was prepared from 8 plasmids containing PR8 genomic segments. These plasmids were mixed to equal molarity, and cDNA was generated using a multiplex one step RT-PCR with the primers Uni12/Inf1 (5'-GGGGGGAGCAAAAGCAGG-3'),Uni12/Inf3 (5'-GGGGGAGCGAAAGCAGG-3'), and Uni13/Inf1 (5'-5CGGGTTATTAGTAGAAA CAAGG-3') as in Hoffmann et al. (2001); Zhou et al. (2009). This library was prepared in identical fashion to the experimental populations and was sequenced in the same MiSeq lane.

Experimental intrahost population

Twenty point mutants were generated in the WSN33 background using the pHW-2000 reverse genetics system (Hoffmann et al., 2002; Visher et al., 2016). In short, we used overlap PCR mutagenesis to introduce the following mutations: HA T1583G, HA G1006T, HA G542T, M T861G, M A541C, NA G1168T, NA C454T, NP A454C, NP A1160T, NS G227T, NS A809G, PA T964G, PA T237A, PA A1358T, PB1 G599A, PB1 G1764T, PB1 T1288A, PB2 A1854G, PB2 A440T, PB2 A1167T. Viruses were rescued from transfected cells as in Pauly and Lauring (2015).

We passaged the 20 WSN33 point mutants and the WSN33 WT once in MDCK cells and verified the identity of the mutants by sequencing each on an Illumina MiSeq as above. We quantified the genome copy number of each supernatant using a SuperScript III Platinum One-Step RT-qPCR kit (Invitrogen) and universal influenza A/B primer and probe sets (Center of Disease Control, 2009). Equal genome equivalents of each infected supernatant were mixed and diluted to generate a population containing each of the 20 mutants present at 5% frequency and a total concentration of 10^5 genomes per microliter. We diluted this mixture into WT WSN33 supernatant to create populations in which each mutant was present at 2, 1, 0.5,and 0.2% frequency all with a total concentration of 10^5 genomes per microliter. These 5 populations were diluted serially into basal media to generate samples with total nucleic acid concentrations of 10^4 and 10^3 genomes per microliter. Viral RNA was extracted from these samples and cDNA was generated in a one-step multiplex RT-PCR as above. The WT WSN33 sample (10^5 genomes per microliter) was processed and sequenced in duplicate. We prepared libraries as before and used Quanti PicoGreen dsDNA quantification (ThermoFisher Scientific) to quantify the concentration of each indexed library. We pooled equal quantities (by nanogram) of each indexed library,

and removed adapter dimers by gel isolation with the GeneJET Gel Extraction Kit (ThermoFisher Scientific) prior to sequencing an on Illumina HiSeq 2500 with 2 x 125 paired end reads. A clonal control library was processed in an identical fashion starting from an equimolar mix of 8 plasmids containing the WSN33 genomic segments.

For the analysis in Figure 2.7, we isolated fresh RNA from the 5%, 2%, 1%, and 0.5% 10^4 genomes per microliter samples. These samples were processed and sequenced in duplicate as above.

Sequence analysis

Reads were aligned to either a PR8 or a WSN33 reference sequence using Bowtie2 (Langmead and Salzberg, 2012). Alignments were sorted and PCR duplicates removed using Picard (`http://broadinstitute.github.io/picard/`). Variants were called using either DeepSNV (Gerstung et al., 2012) or LoFreq (Wilm et al., 2012) and filtered using the Pysam module in python (`https://github.com/pysam-developers/pysam`) and custom R scripts available for download at `https://github.com/lauringlab/Benchmarking_paper`. Bases with a Phred < 30 were masked in the DeepSNV analysis. We connected all of these steps into an analytical pipeline using bpipe (Sadedin et al., 2012) which is available for download at `https://github.com/lauringlab/variant_pipeline`. To save memory during SNV processing only variants with $p < 0.9$ were included in our ROC curve analysis, as the vast majority of true negatives are trivial to identify and have a p=1. For ease of viewing, and to account for this analytical artifact, we extended the ROC curves horizontally from the last observed change in sensitivity. All raw fastq files were submitted to the Sequence Read Archive (SRA) under BioProject PRJNA317621, and all commands required to generate the figures in the manuscript are available for

anonymous download at `https://github.com/lauringlab/Benchmarking_paper`. An interactive Shiny app of our benchmarking work can be downloaded at `https://github.com/lauringlab/benchmarking_shiny`.

Diversity Metrics

The Shannon Entropy (H) of each genomic position was calculated as

$$H = -\sum_{i=1}^{n} x_i \ln(x_i)$$

where x_i represents the frequency of the ith allele, and n represents the number of alleles found at the given position. Since our data does not represent haplotypes, we report Shannon's Entropy as the mean across all genomic positions.

The L1-norm (L) between 2 populations was calculated as

$$L = \sum_{i=1}^{n} |p_i - q_i|$$

where n represents the union of variants between the two samples and p_i and q_i represent the frequencies of the ith variant in each sample respectively.

Results

The ability to reliably identify single nucleotide variants (SNV) is integral to accurate NGS-based studies of viral diversity. The accuracy of any SNV calling pipeline can be described in terms of its sensitivity and specificity. Sensitivity is the proportion of true variants that are properly identified, and specificity is the proportion of true negatives that are properly identified. In other words sensitivity measures an assay's ability to detect true variants present in a viral population; whereas, specificity is determined by how many false variants (errors of some kind) are erroneously identified. An assay with perfect accuracy, in which all the true

variants are found and only true variants are found, has a sensitivity and specificity of 1.

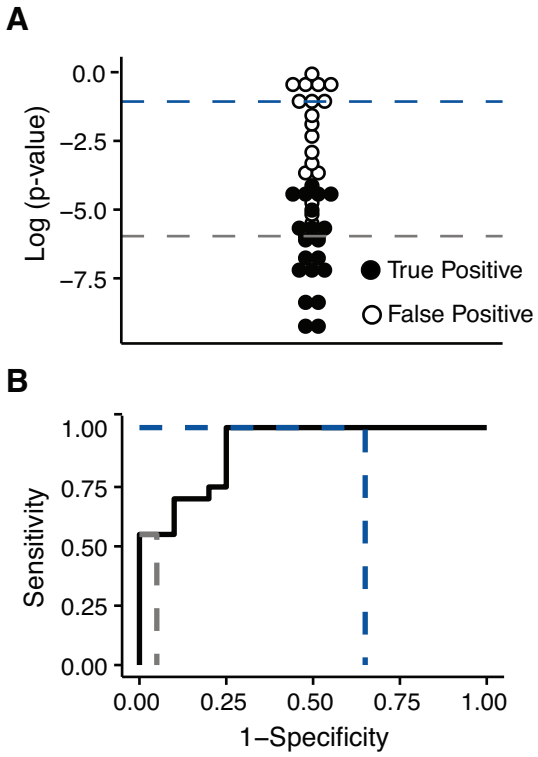


Figure 2.1: Example of an ROC curve. (A) Hypothetical variants are stratified by the log of the P value. P value thresholds are indicated as dashed colored lines. These data are intended to illustrate the concept and are not based on an actual experiment. (B) An ROC curve made from the hypothetical data shown in panel A. The dashed colored lines indicate the points on the curve corresponding to the thresholds in panel A.

There is an obvious trade-off between sensitivity and specificity. Improved sensitivity often requires less stringent criteria in variant calling, but reduces specificity. Conversely, increased stringency can improve specificity but often reduces sensitivity. This relationship can be visualized using a receiver operating characteristic (ROC) curve (Figure 2.1). An ROC curve plots the sensitivity of an assay along the y-axis and 1-specificity, or the false positive rate, along the x-axis. A variant calling pipeline must be tested against known data in order to construct an ROC curve. The outcomes can then be stratified according to a metric that quantifies the probability that

a given variant is real, often a p-value or quality score. In a controlled benchmarking experiment, all true variants are known, and the sensitivity and specificity can be calculated at different cut-offs (Figure 2.1A). These points are then used to construct the curve (Figure 2.1B). A perfect ROC curve in which all the true positives can be separated from all false positives is a right angle that follows the upper left perimeter of the plot.

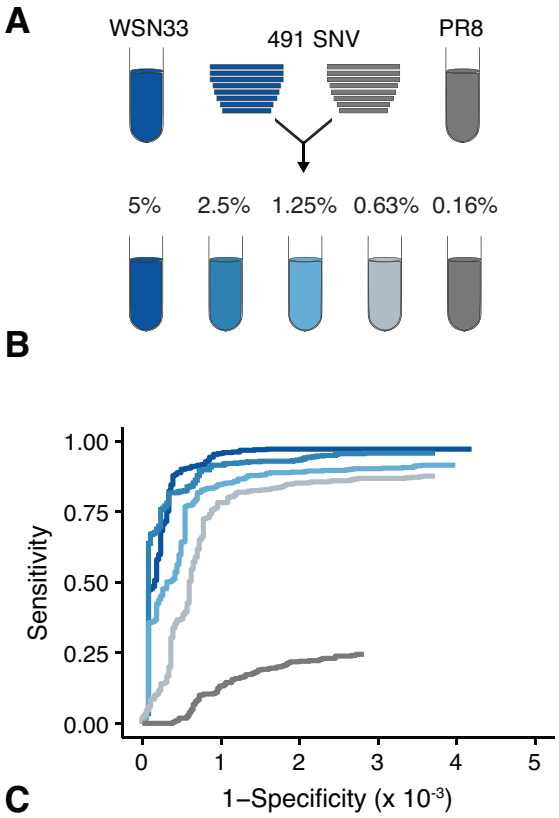
Initial accuracy

A comprehensive comparison of SNV calling approaches is beyond the scope of this work. Instead, we robustly benchmark one variant caller, DeepSNV, and highlight approaches for improving its accurate application to patient-derived populations. In doing so, we demonstrate the importance of validating any variant calling method under the experimental conditions to which it is applied. We chose DeepSNV as our starting point, because at the time, it was the only variant caller that had been benchmarked on a dataset of known viral variants (Gerstung et al., 2012).

DeepSNV is a variant calling algorithm that uses a clonal, plasmid-derived, control to estimate local error rates across the genome (Gerstung et al., 2012). Because it is clonal, the sequence of the control is known with a high degree of confidence, and any nonconsensus base is indicative of an error in library preparation or sequencing. Additionally, the control and experimental samples are processed together and are assumed to have identical noise characteristics, thereby minimizing issues of batch effect. DeepSNV then applies a hierarchical binomial model at each genomic position and identifies true variants as those with frequencies significantly above the noise found in the plasmid control. Like many variant calling algorithms, the accuracy of DeepSNV was initially determined using samples that required minimal PCR amplification. However, its accuracy has not been tested when applied to whole

genome sequencing of a viral population amplified by reverse transcription PCR (RT-PCR).

In our first benchmarking data set we created defined mixtures of two plaque purified and expanded influenza strains, WSN33 and PR8. Complementary DNA from both viruses was mixed serially such that WSN33 cDNA was present at frequencies of 5, 2.5, 1.25, 0.63 and 0.16% (Figure 2.2A). These viruses differ at 491 positions (primer sites used in RT-PCR were excluded from analysis), providing 491 true positives in each dilution. On plasmids subjected to limited PCR, DeepSNV identified known variants at 0.1% frequency with a sensitivity of 0.860 and specificity of 1.0 (Gerstung et al., 2012). Under our experimental conditions we found a reduction in sensitivity (0.851 for variants at 0.63% and 0.173 for variants at 0.16%), and specificity (0.9980 and 0.9987 for variants at 0.63 % and 0.16% respectively). We were able to more closely approach the perfect specificity previously reported for DeepSNV by applying a more stringent p value of 0.01. A minor decrease in sensitivity accompanied this slightly more stringent p value (Figure 2.2C). We used this p-value cut-off in all further experiments. The specificity was above 0.9980 in all dilutions. While this drop in specificity (from 1.0) appears small, it corresponds to 78 false positives when applied to the more than 39,000 potential variants in the 13,057 bp influenza genome.



Freq	Sens	TP	Spec	FP
5.0%	0.9715	477	0.9982	68
2.5%	0.9348	459	0.9980	78
1.25%	0.8900	437	0.9981	75
0.63%	0.8330	409	0.9983	66
0.16%	0.1426	70	0.9989	41

Figure 2.2: Accuracy of DeepSNV. (A) Reconstituted cDNA genomes of influenza virus strain WSN33 were diluted serially in reconstituted cDNA genomes of PR8, generating artificial populations with 491 single-nucleotide variants from WSN (relative to PR8) at the indicated frequencies. (B) ROC curve measuring the accuracy of DeepSNV in identifying WSN33 variants mixed with PR8 at the indicated (by colors matching those in panel A) frequencies. (C) Summary of the data in panel B at a P value threshold of 0.01. Freq, frequency; Sens, sensitivity; TP, true positives; FP, false positives.

An experimental intrahost population

Although the initial benchmarking experiment validated our ability to accurately detect rare variants in influenza populations, the experiment was run under relatively artificial conditions. Patient-derived populations are typically less divergent than WSN33 and PR8 (Ghedin et al., 2010; Rogers et al., 2015; Poon et al., 2016), and the number of viral genomes in patient-samples is much lower than that found in cell culture. To mimic patient specific conditions, we generated 20 viral clones, each with a single point mutation in the WSN33 background. We sequenced stocks of these mutants on an Illumina MiSeq instrument to account for any additional mutations that might have arisen between transfection and the passage 1 stock. Four additional mutations were found above 1% frequency (frequencies 1.2%-3.7%). We also determined the genome copy number of each stock using quantitative RT-PCR. We then mixed equal genome equivalents of these 20 viruses, to generate a sample population with 10^5 copies per microliter with each mutation present at 5% frequency. This population was serially diluted into a stock of wild type WSN33, generating samples with each of the 20 mutations present a 2, 1, 0.5, 0.2, and 0.1% frequency (Figure 2.3 A). We then serially diluted these populations into basal media to obtain mixtures with lower nucleic acid input. The range (10^3-10^5 copies per microliter) matches the inputs typically found in many patient-derived influenza virus samples (Table 2.2 and Figure 3.1). We sequenced these populations on the Illumina HiSeq platform, and called variants using DeepSNV. We also processed and sequenced the wild type WSN33 stock in duplicate to control for any mutations in the viral diluent.

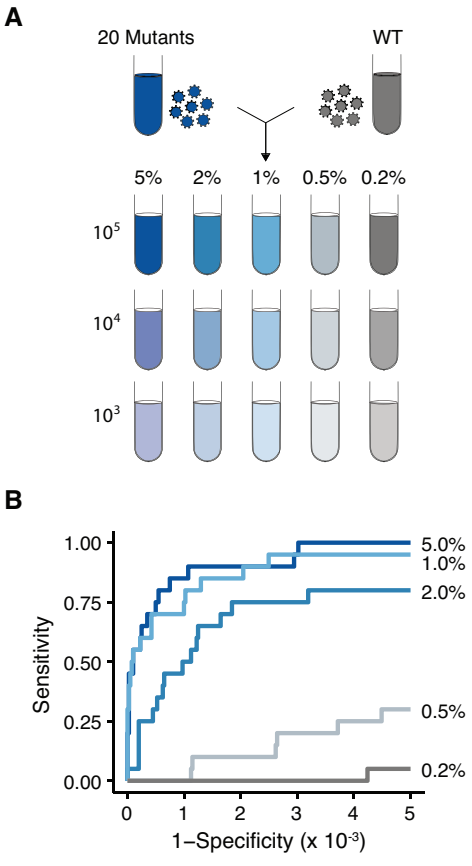
The 20 mutations present in our initial viral mixture were the only true positives considered in our analysis. Four SNV that were present at >1% frequency in either both duplicates of the wild type stock or any one of the viral clones were masked,

excluded from the analysis, and considered neither a true positive nor a true negative. By applying these thresholds we were able to validate our analysis using only variants identified a priori and avoid the circular logic of validating a variant pipeline using SNV identified by the same pipeline.

In these populations with lower diversity and input titer, we maintained greater than 0.85 sensitivity for SNV at 1% frequency or higher. Despite a high depth of coverage (>10,000 reads per bp), our sensitivity was considerably lower for variants at or below 0.5% frequency (Figure 2.3B and C). The drop in sensitivity, compared to the first data set (Figure 2.2), is most likely due to the 1,000-fold decrease in nucleic acid concentration, and the fact that library preparation requires a number of sampling steps that may limit detection.

In our initial analysis of these data using the same default DeepSNV settings as above, the specificity was significantly lower than what was observed in our PR8-WSN33 populations (mean of 0.9812 with a minimum of 0.9598). These lower input samples underwent more PCR cycles, which have been shown to skew the error distributions in the test libraries relative to the plasmid control (Gerstung and Beerenwinkel, 2015). We were able to partially account for this variation by using an alternative beta binomial model available in DeepSNV, which more appropriately fits these conditions. With these settings, the specificity was greater than 0.9900 in all the samples with 10^5 genomes per microliter (Figure 2.3C). As above, while 0.9900 specificity appears adequate, it results in over 200 false positive variants when applied to the over 39,000 potential variants in the influenza genome. The false positives outnumber the true positives by 10-fold in these populations with realistic diversity and input. We were able to increase our specificity by applying a more stringent p-value cutoff. However, as shown by the ROC curves in Figure 2.3, this

move towards the y-axis markedly reduces sensitivity. Our data demonstrate that with moderate concentrations of input nucleic acid, even statistically significant p-values from a robust variant caller are not sufficient to accurately separate true from false positive variants.



C

Freq	Sens	TP	Spec	FP
5.0%	1.00	20	0.9938	249
2.0%	0.85	17	0.9942	232
1.0%	0.95	19	0.9975	101
0.5%	0.35	7	0.9948	210
0.2%	0.05	1	0.9937	252

Figure 2.3: Accuracy of DeepSNV on populations approximating patient-derived samples. (A) Twenty viral supernatants, each with a single SNV, were diluted in a WSN33 viral supernatant to generate artificial viral populations with 20 mutations at the indicated frequencies. These populations were diluted further in basal medium to match the genome concentrations found in patient-derived samples (10^5 to 10^3 genomes/μl). (B) ROC curve measuring the accuracy of DeepSNV in identifying SNV at the indicated frequencies. (C) Summary of the data in panel B at a P value threshold of 0.01.

Additional filtering criteria

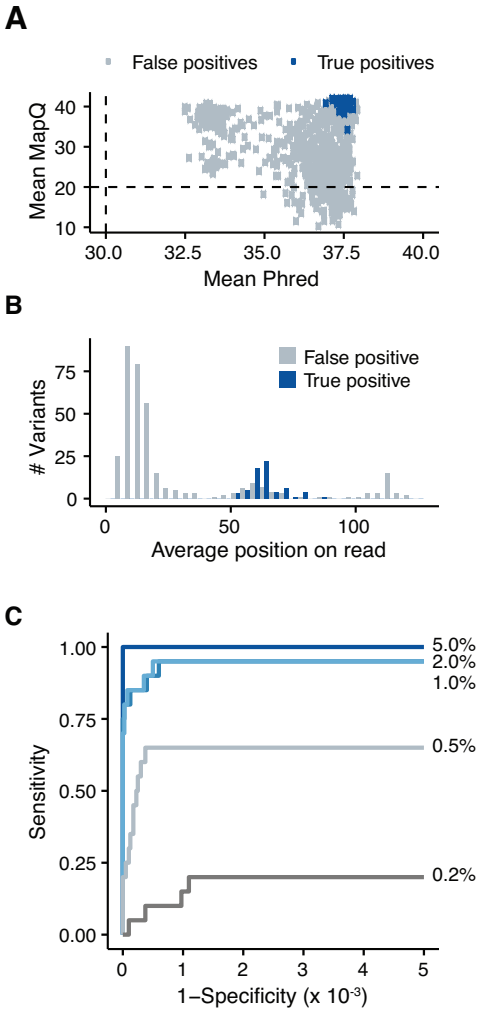
Many next generation sequencing studies utilize mapping quality (MapQ) and/or base quality (Phred) thresholds to ensure that only the highest caliber sequencing data is used to call variants. Mapping quality measures the probability that a given read is mapped to the correct position in the genome, while base quality estimates the likelihood that the base call by the sequencer is correct. In the above analysis we masked bases that had a Phred score less than 30 (0.001 probability of being incorrect) and did not apply any MapQ cut-offs. In our next analysis, we applied seemingly stringent cut-offs such as a MapQ of 20 and a Phred of 30 to our data (Koboldt et al., 2013, 2012; Dietz et al., 2013; Schmitt et al., 2012). These criteria were unable to distinguish true from false positives in our 10^5 samples (Figure 2.4A) and indicate that many false positives occur on well mapped reads with high quality base calls.

We further parsed our false variant calls by locating them within individual sequencing reads. It is well known that sequence quality drops near the end of a read (Schirmer et al., 2015; Wang et al., 2012), and we found that our false positives clustered at the termini of our paired end reads (Figure 2.4B). The average Phred score of these false positives was 37.1, further demonstrating that filtering on quality score alone is insufficient. In contrast, true positives were uniformly distributed across the reads resulting in an average read position near the middle of the read.

Based on these results, we applied a number of empirically determined cut-offs, which markedly improved our specificity to >0.9990 without sacrificing sensitivity (Figure 2.4C and 2.4D). For a variant to be considered in our analysis we required a mean mapping quality of ≥ 30, a mean Phred score of ≥ 35, and an average read position within the middle 50% of the read. Under these conditions we found 20 or fewer

false positives in all 5 of the samples. Given this success, we applied a number of other strategies to further increase our accuracy, including but not limited to Benjamini Hochberg p-value correction, more stringent p values (< 0.01) or frequency cut-offs ($> 0.2\%$), retention of duplicate PCR reads, trimming the ends of the influenza genome, and employing alternative statistical distributions to estimate the error rate in the control sample. None of these approaches significantly improved our accuracy over the above quality and read-position criteria. The impact of various filtering criteria on our data can be visualized in an interactive Shiny application available for download at `https://github.com/lauringlab/benchmarking_shiny.git`.

We also benchmarked the accuracy of our DeepSNV pipeline in estimating the frequency of the true positive variants (Figure 2.5). Although the medians of the measured frequencies match the expected values, we found substantial spread in each sample and the overall the fit was modest ($R^2 = 0.65$). The mean percent difference between the measured frequency and the expected was 41%. This error is likely due to amplification bias associated with RT-PCR (Zhou et al., 2015) and library preparation and should be kept in mind when employing downstream analyses that depend on frequency measurements (e.g. variant fitness, haplotype reconstruction, Shannon's Entropy, and other diversity metrics).



D

Freq	Sens	TP	Spec	FP
5.0%	1.00	20	>0.9999	1
2.0%	0.85	17	0.9999	6
1.0%	0.95	19	0.9995	20
0.5%	0.35	7	0.9999	5
0.2%	0.05	1	0.9997	11

Figure 2.4: Accuracy can be improved through more stringent quality thresholds. (A) All called variants from the five samples with 10^5 genomes/μl and P values of 0.01 stratified by the mean mapping quality of the reads containing the variant and the mean Phred scores of the variant bases. The dashed lines indicate common cutoffs of 20 and 30 for mapping quality and Phred, respectively. (B) Histogram of average positions on a paired-end read of the variants that passed our mean MapQ threshold of 30 and mean Phred threshold of 35. (C) ROC curve measuring the accuracy of our analysis after applying the following quality cutoffs: mean MapQ score, 30; mean Phred score, 35; average read position, between 32 and 94 (the middle 50% of the read). (D) Summary of the data in panel C at a P value threshold of 0.01.

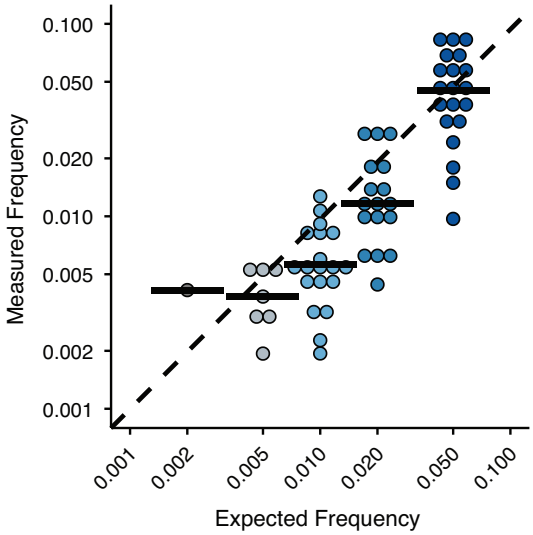
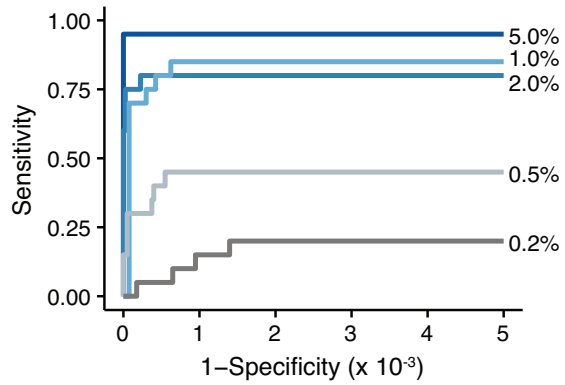


Figure 2.5: Accuracy of the frequency measurements for true-positive SNV in the samples with 10^5 genomes/μl. The black bars are the medians. The dashed line is where measured and expected frequencies are equal. Note that both axes are on a log scale.

Relatively low accuracy is not unique to DeepSNV

A



B

Freq	Sens	TP	Spec	FP
5.0%	0.95	19	0.9987	52
2.0%	0.80	16	0.9989	46
1.0%	0.85	17	0.9976	98
0.5%	0.45	9	0.9987	53
0.2%	0.20	4	0.9985	61

Figure 2.6: Accuracy of LoFreq on populations with 10^5 genomes/μl. (A) Accuracy of LoFreq using standard parameters. The specificity of LoFreq was scaled to account for the same number of tests as performed in DeepSNV. (B) Summary of the data in panel A at a P value threshold of 0.01.

DeepSNV is one of many variant callers that employ a combination of empiric and statistical approaches to model error rates. We asked whether the decreased accuracy observed in our dataset was due simply to peculiarities specific to DeepSNV. We analyzed our 10^5 input populations using LoFreq, another variant caller commonly used in next generation sequencing studies that has been reported to have perfect specificity (Wilm et al., 2012). Under our experimental conditions, LoFreq had marginally reduced sensitivity compared to DeepSNV when applied to variant frequencies $\geq 1.0\%$ but marginally increased sensitivity when applied to variant frequencies $< 1.0\%$ (Figure 2.6). The specificity of LoFreq was comparable to what we observed with DeepSNV in our high-input cell culture-derived populations (Figure

2.2), and better than DeepSNV in our initial analysis of the 20 mutant populations (Figure 2.3 prior to Phred, MapQ, and read position filtering). This increased specificity is most likely due to the fact that the LoFreq algorithm already takes MapQ and Phred scores into account when calling variants and has a stringent strand bias filter that removes many of the variants found only at one end of a paired-end read. Because it does not compare test samples to a plasmid control, it is also more robust to issues of PCR skewing than DeepSNV. However, even with these additional characteristics, the specificity of LoFreq was lower than our improved DeepSNV pipeline (compare Figure 2.6 and Figure 2.4), with over 40 false positives per sample. It appears that higher than expected false positive rates are not specific to DeepSNV and most likely plague many variant callers applied to patient-derived viral samples.

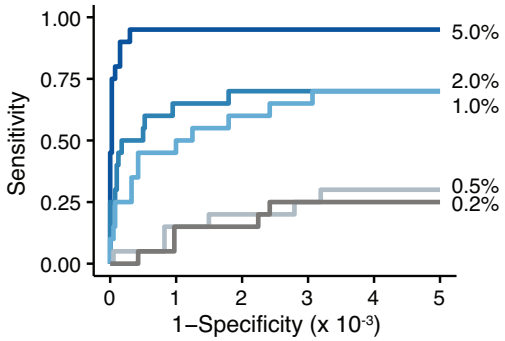
Accuracy at lower input levels

Host-derived viral populations vary in copy number and titer by several orders of magnitude (Lau et al., 2013; Teunis et al., 2015; Takeyama et al., 2016). This variability can be attributed to a variety of factors including, but not limited to: collection site, ease of nucleic acid isolation, the presence of host nucleic acid, efficiency of library preparation, and host and viral factors. To ensure accuracy across a range of input levels, we diluted our experimental populations serially into basal media (Figure 2.3A) and identified variants using our modified DeepSNV analysis pipeline (Figure 2.7). As expected, our sensitivity was lower in populations with fewer genomes. For example, a variant at 0.5% frequency in a 10^4 genomes per microliter sample is expected to be present on only 700 genomes in the initial RT-PCR. Many of these will be lost due to bottlenecks in the amplification and library preparation process. We also found reduced specificity in lower input samples. The increase in false positives is presumably due to a greater dependence on RT-PCR

amplification and consequent propagation of errors. These data highlight the importance of controlling for input levels when comparing diversity across experimental samples.

In most cases, RT-PCR errors should be sporadic and randomly distributed across the amplified region. If RT-PCR errors are responsible for the reduced specificity found at lower input levels, they should be easily identified as variants present in only one of two RT-PCR reactions performed on the same RNA (Robasky et al., 2014). To test this hypothesis, we sequenced duplicate RT-PCR reactions of the 5, 2, 1, and 0.5% variant frequency samples from our 10^4 genomes per microliter collection. The duplicates were processed separately, but sequenced on the same lane of an Illumina HiSeq. We applied the stringent quality cut-offs and required that a given variant be found in both duplicates. By analyzing samples in duplicate, we reduced the number of false positives in each sample to 10 or fewer resulting in and a specificity of >0.9998 (Figure 2.8). This increased specificity was not accompanied by a decreased sensitivity. In fact, we found a slight increase in sensitivity (compared to Figure 2.4), most likely due to variability in library preparation. Thus, accurate analysis of low input samples can be achieved through duplicate RT-PCR and careful benchmarking experiments.

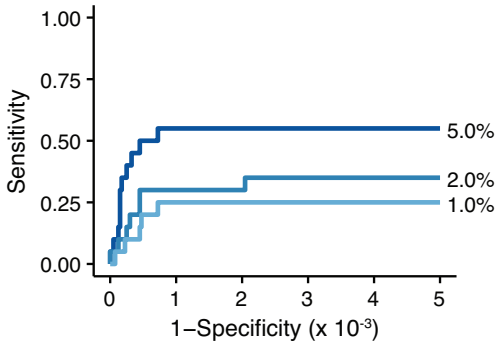
A



B

Freq	Sens	TP	Spec	FP
5.0%	0.95	19	0.9990	41
2.0%	0.65	13	0.9985	60
1.0%	0.55	11	0.9983	67
0.5%	0.20	4	0.9983	69
0.2%	0.15	3	0.9983	69

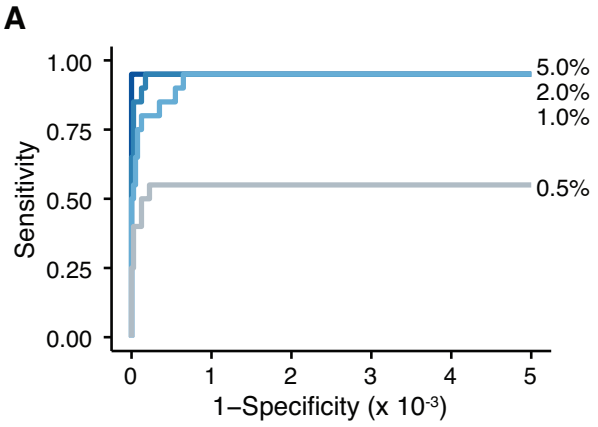
C



D

Freq	Sens	TP	Spec	FP
5.0%	0.55	11	0.9990	42
2.0%	0.30	6	0.9981	78
1.0%	0.25	5	0.9979	86

Figure 2.7: Accuracy of DeepSNV on populations with lower input nucleic acid levels. (A) ROC curve for the samples with 10^4 genomes/μl. (B) Summary of the data in panel A at a P value threshold of 0.01. (C) ROC curve for the samples with 10^3 genomes/μl. (D) Summary of the data in panel C at a P value threshold of 0.01.



B

Freq	Sens	TP	Spec	FP
5.0%	0.95	19	0.9999	4
2.0%	0.90	18	0.9999	6
1.0%	0.80	16	0.9998	10
0.5%	0.40	8	0.9999	3

Figure 2.8: At lower inputs, duplicate samples improve accuracy. (A) ROC curve of the samples with 10^4 genomes/μl processed in duplicate. Only SNV present in both samples were considered. (B) Summary of the data in panel A at a P value threshold of 0.01.

Sub-optimal SNV identification confounds diversity measurements

NGS of intrahost populations is commonly used to determine the impact of host or environmental factors on viral diversity. Because measurements of viral diversity rely entirely on SNV identified in NGS data, they are very sensitive to the accuracy of these variant calls. To illustrate this problem, we calculated the diversity of our 10^5 genomes per microliter samples at each step of our benchmarking process using three complementary metrics (Table 2.1). Richness is the count of nonconsensus variants present in a population (often referred to as intrahost SNV). Shannon's Entropy is a diversity metric that accounts for both the number of variants present (richness) and their frequencies (evenness). Because our data are unphased (i.e. no haplotypes), we have reported the average entropy per nucleotide position (Grubaugh et al., 2015). The last metric, L1-norm, is a distance measurement that describes how similar two populations are to one another based on the frequencies of variants present. Identical populations will have an L1-norm of 0. To mimic experimental conditions, we included all variants identified in each analysis regardless of whether or not subsequent benchmarking distinguished them as true or false positives. We found that the accuracy of the SNV calling method has a profound effect on measurements of diver-

Table 2.1: Diversity measurements in experimental populations.

Variant Frequency[a]	Diversity Metric	Expected	Lofreq	DeepSNV	DeepSNV modified
5%	Richness	20	71	269	21
	Entropy	2.97E-04	3.35E-04	1.60E-03	2.77E-04
	L1-norm	0	0.519	4.006	0.378
1%	Richness	20	115	120	39
	Entropy	8.37E-05	2.78E-04	3.14E-04	7.30E-05
	L1-norm	0	2.702	0.704	0.133
0.5%	Richness	20	62	217	12
	Entropy	4.71E-05	8.47E-05	1.12E-03	2.10E-05
	L1-norm	0	0.196	3.156	0.089

[a] Frequency of 20 true positive variants. Only the 10^5 genomes/μl input libraries were used

Table 2.2: Diversity measurements in patient-derived samples

Sample ID	DPS[a]	Strain	Season[b]	Log_{10} (genomes/μl)	DeepSNV		DeepSNV Modified	
					Richness	Entropy	Richness	Entropy
1376	1	A/H1N1	2013-2014	5.3	90	8.70E-04	5	6.04E-05
1401	2	A/H1N1	2013-2014	5.0	110	1.04E-03	22	1.16E-04
1405	3	A/H1N1	2013-2014	4.3	120	1.08E-03	30	1.55E-04
1374	4	A/H1N1	2013-2014	4.6	185	1.25E-03	13	1.05E-04
1227	1	A/H3N2	2012-2013	5.3	79	5.21E-04	8	4.43E-05
1321	1	A/H3N2	2012-2013	4.3	20	2.35E-04	3	1.70E-05
1229	2	A/H3N2	2012-2013	4.6	32	3.34E-04	8	7.07E-05
1245	3	A/H3N2	2012-2013	4.4	197	1.38E-03	3	1.07E-05

[a] Days post symptom onset.
[b] Influenza season was considered to run from September to May.

sity. It is clear from the richness measurements in Table 2.1 that the number of false SNV (i.e. the specificity) largely determines the accuracy of the downstream analyses. Thus, our adapted DeepSNV protocol, which was able to distinguish between true and false SNV with the highest accuracy, gave the most accurate measures of diversity, followed by LoFreq and the default version of DeepSNV.

To determine the impact of our improved variant calling pipeline on actual host-derived populations, we applied our approach to 8 patient-derived samples collected as part of a household-based cohort study of influenza (Ohmit et al., 2015; Monto et al., 2014) (Table 2.2). The samples were chosen from two influenza seasons and include H1N1 and H3N2 subtypes over a range of input titer and day of infection (measured as day post symptom onset). As in our benchmarking data set, the estimated diversity of each sample was greatly reduced when we applied our empirically determined quality thresholds. The number of intrahost SNV and the Shannon entropy were reduced by up to 10 fold, suggesting the presence of a large number of false positives in our unmodified pipeline. These data show that validation is necessary to avoid overestimating the diversity present in patient derived samples.

Discussion

Robust validation is essential in NGS-based studies of viral diversity. Differences in experimental design and sample preparation can lead to wide variability in the accuracy of SNV identification. We found that input nucleic acid concentration, which can vary greatly in patient-derived samples, had a large impact on both the sensitivity and specificity of rare variant detection. At moderate levels of nucleic acid input we could improve accuracy by filtering putative SNV based on quality metrics and read position. We further improved our accuracy at low input levels by processing these samples in duplicate. While our quality cut-offs may not be universally applicable to all samples and variant callers, our data suggest that experimental design is critical for accurate SNV detection. These findings are important as few, if any, variant callers have been benchmarked under patient-derived conditions. Finally, we showed that inaccuracies in SNV calling drastically impact downstream analysis and lead to overestimations of intrahost diversity in patient-derived samples.

We initially chose DeepSNV for our studies, because it is one of the few variant callers that has been validated on viral sequencing reads in which all true positive variants and their frequencies were known a priori and independent of NGS (Gerstung et al., 2012). A key strength of our study is that we applied DeepSNV to experimental populations that more closely mimic the diversity and levels of virus found in patient-derived samples of influenza virus. At a modest input of 10^5 genomes per microliter and default DeepSNV settings, false positives can outnumber true positives by a factor of 4. It should be noted that our specificity in all cases remained above 0.9900. When applied across an entire influenza genome, a specificity of > 0.9995 is required to reduce false positives to low levels. As above, the decreased accuracy

of DeepSNV under these conditions is due to small but important differences in our experimental design compared to what has been previously reported, namely input nucleic acid concentration and RT-PCR amplification.

Because DeepSNV is somewhat agnostic towards the mapping quality and the base quality of a given variant, we sought to improve our accuracy by identifying thresholds that more effectively distinguished true positive SNV. In our data sets, the distributions of average MapQ and Phred scores of putative SNV were bimodal with true SNV found in the higher of the two distributions. In the data presented, these cut-offs include $\geq$ 98% of the true positive variant calls. Our empirically determined thresholds were chosen to eliminate putative SNV found in the lower of the two distributions. These thresholds should be reproducible in our system, as we have observed consistent MapQ and Phred quality distributions over 300 influenza libraries and 5 HiSeq runs. We have also seen the same bimodal trend in libraries of poliovirus populations, but have applied a lower empirically determined MapQ cut-off to these data. The shift in MapQ is most likely rooted in differences in genomic structure between the two viruses. While our MapQ and Phred thresholds are robust in our system, they may need to be adjusted for use in others.

Even in the face of stringent MapQ and Phred cut-offs, we found many high-quality false positive SNV that were identified only at the termini of paired end reads. We removed these by filtering putative SNV based on their average position in a paired end read. These false positives were found almost exclusively in regions of the genome that were enriched for read start sites. This enrichment may be a consequence of sequence context, the fragmentation process, or our size selection protocol. We suggest that there might also be a biological reason for this effect, as our PCR-amplified plasmid control samples did not exhibit this bias. For example,

defective interfering particles, which commonly arise during cell passage, contain truncated genomic segments and would only be present in infected supernatants and not the plasmid control. We hypothesize that the large deletions in these segments increase the number of reads that start at certain genomic positions. As the beginning of reads can also be error prone, this enrichment would result in false positive SNV. Our analysis was particularly vulnerable to this type of error, because we did not trim the ends of our reads, and DeepSNV, unlike other variant callers, does not directly test for strand bias or consider read position as a variable. While it is easy to diagnose these shortcomings in retrospect, such errors had not been previously reported for DeepSNV, and were only elucidated through our extensive validation.

While read filtering and trimming are common in NGS data sets (Del Fabbro et al., 2013; Nielsen et al., 2011; Bolger et al., 2014), we have taken a slightly different approach in our analysis. In the initial SNV identification step, we masked bases with Phred < 30 but made no additional restrictions on the raw data. We only imposed additional quality restrictions after putative SNV - those that exceeded the expected frequency given the plasmid control - were identified. While our approach treats variant nucleotides more stringently than consensus base calls, we do not think that this differential stringency introduces unnecessary bias. Because we identified specificity as the major problem in accurate SNV identification, stringent filtering of potential false positives seems appropriate. Furthermore, the vast majority of reads call a consensus base and our mean quality score thresholds would therefore not be expected to remove many consensus base calls from the analysis.

Frequency thresholds of 0.1 to 1% represent an additional quality filter that is applied to SNV after identification (Combe et al., 2015). We did not apply direct frequency thresholds in our analysis, as we found that arbitrary cut-offs limited

sensitivity without improving specificity. Read depth, or coverage, is another metric that can be used in conjunction with frequency to ensure accurate SNV identification. Although we did not apply a direct coverage cut-off, DeepSNV has been reported to require coverage of ten times the reciprocal of frequency for sufficient power to call SNV. For example, a coverage of 1,000x is needed to detect a variant at 1% frequency. In our analysis, the lowest coverage for a true positive was 1795 reads (4.8% frequency) while the lowest coverage for a false positive variant was 966 read (8.5% frequency). If a given data set has variability in read depth across the genome, SNV at identical frequencies may be detected with differing sensitivity. Under such conditions, the application of variant frequency or read depth thresholds would lead to severe ascertainment bias in subsequent analyses of diversity.

Few studies of intrahost diversity quantify or control for the number of genomes in a sample. This is important, because we found that input copy number is a key factor in variant detection. Despite high accuracy at 10^5 genomes per microliter we observed a decrease in sensitivity in our 10^3 genomes per microliter samples. More importantly, this drop in sensitivity was accompanied by reduced specificity. At lower nucleic acid concentrations, NGS pipelines rely more heavily on RT-PCR amplification, which tends to propagate errors that are otherwise indistinguishable from true positives in sequence data. We were able to limit these sporadic and random errors by processing low input samples in duplicate. Quantifying and controlling for RT-PCR errors in this way will allow us to accurately compare patient-derived samples that vary over a range of input.

Many variant callers are benchmarked on simulated data sets, plasmids, or PCR products, and may not have comparable sensitivity and specificity when applied to viral samples. Our goal was not to compare the strengths and weaknesses of a few

algorithms, but rather to highlight how accuracy can be experiment-specific. We recognize that some variant callers may perform better than DeepSNV, and that others may be better suited to other systems. However, our work with LoFreq suggests that all methods have inherent limitations and that understanding these limitations is essential. We have been able to greatly improve the accuracy of DeepSNV under our experimental conditions, and we are now equipped with an understanding of the limitations of our method.

Our study highlights previously under-recognized issues in variant calling and suggests factors that should be considered in future studies of viral diversity. The need for target amplification, the structure of the viral genome, and variation in input genome copy number may lead to errors specific to a given experiment. We have shown that these seemingly small differences in sensitivity and specificity (e.g. 0.9998 vs. 0.9900) can have profound effects on measurements of viral diversity in experimental and patient-derived populations. These differences are especially important in comparative studies of intrahost diversity. We realize that there are many solutions to the problem of NGS accuracy and have made our data sets and code available to the community. We hope that this will allow others to benchmark their own pipelines or to improve on our work. This process should improve the reliability of NGS in studies of virus evolution and molecular epidemiology.

Acknowledgements

We thank Pat Schloss, Mike Imperiale, and Robert Woods for helpful discussion and a critical reading of the manuscript. We thank Chris Gates and Peter Ulintz for help with bpipe and Arnold Monto, Suzanne Ohmit, Rachel Truscon, Emileigh Johnson, Ryan Malosh and Josh Petrie for providing the HIVE samples.

CHAPTER III

Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses

Note: This chapter is a modified version of the published article:

Debbink K*, McCrone JT*, Petrie JG, Truscon R, Johnson E, Mantlo EK, Monto AS, Lauring AS. 2017. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. PLoS Pathog 13:e1006194.

Introduction

Despite recommendations for universal influenza vaccination and the ample availability of vaccines in the United States, influenza continues to cause significant morbidity and mortality (World Health Organization, 2014). This is, in part, a result of the modest effectiveness of current vaccines, so that considerable numbers of vaccine failures occur each year. Within individuals, influenza populations exist as a collection of closely related, and at times antigenically distinct, variants that may exhibit diverse phenotypes (Ghedin et al., 2009, 2010, 2012; Poon et al., 2016; McCrone and Lauring, 2016). Intrahost single nucleotide variants (iSNV) can be transmitted as part of the infecting population (Poon et al., 2016; Murcia et al., 2010; Hughes et al., 2012; Wilker et al., 2013; Varble et al., 2014) or generated over the course of an infection due to the virus' low replication fidelity (Parvin et al., 1986; Sanjuán et al.,

2010). The evolutionary forces that shape the genetic structure of viral populations within hosts and ultimately give rise to novel antigenic variants at the host population level are poorly understood. A clear understanding of the intrahost diversity of influenza virus populations and its impact on influenza virus evolution is central to many questions of direct clinical and public health relevance (Holmes, 2009).

Influenza vaccines are considered for reformulation each year to counter the viral antigenic drift that enables escape from the previous year's vaccine (Center of Disease Control, 2016). Annual influenza vaccine effectiveness is 60% on average, and can be much lower during antigenically unmatched years (Osterholm et al., 2012; Skowronski et al., 2015). While antigenic drift is monitored annually on a global scale, the source of antigenic variation is ultimately at the level of the individual host. Phylogenetic studies of whole genome sequences from cities and smaller communities have demonstrated that multiple lineages circulate over the course of a single influenza season (Ghedin et al., 2009; Holmes et al., 2011), and individual hosts may harbor mixed infections that include antigenically novel variants (Ghedin et al., 2010, 2012). While human hosts could be preferentially infected with one lineage over another based on pre-infection immune status, the degree to which circulating escape variants contribute to vaccine failure is currently unknown.

Host immune selection is a major driver of influenza virus evolution on the global scale. Both phylogenetic analysis and antigenic cartography have demonstrated that antibodies exert positive selective pressure on the viral hemagglutinin (HA) and neuraminidase (NA) proteins (Smith et al., 2004; Nelson and Holmes, 2007). Vaccination and natural influenza infection often lead to partial, or non-sterilizing immunity, and post-vaccination antibody titers are only a moderate predictor of subsequent protection (Ohmit et al., 2011; Tsang et al., 2014). Previous work has demonstrated

that sub-neutralizing concentrations of immune sera can select for antigenic variants, and some have argued that vaccination can accelerate the process of antigenic drift (Archetti and Horsfall, 1950). A recent study in vaccinated people suggested that novel antigenic variants could be present at low frequencies (Dinis et al., 2016). Importantly, humans often differ in their prior exposure to influenza viruses and vaccines, and pre-existing immunity may confound such studies (Fonville et al., 2014). Therefore, the extent to which partial immunity selects for antigenically relevant variants during natural infection in humans is unclear.

By necessity, most of the available data on vaccination and intrahost evolution have come from analyses of HA sequences in large animal models of infection, including horses, pigs, and dogs. These studies have suggested that intrahost populations include a number of somewhat rare single nucleotide variants that increase and decrease in frequency over the course of an infection with sporadic fixation events occurring in some animals. The overall impact of vaccination on antigenic diversification was not clear (Murcia et al., 2010; Hoelzer et al., 2010; Murcia et al., 2012, 2013), and there were differences between experimentally and naturally infected animals (Murcia et al., 2010; Hughes et al., 2012; Murcia et al., 2012). These differences are also likely to be an issue in extrapolating results from human experimental challenge models (Killingley et al., 2011).

Some have suggested that high intrahost diversity reflects increased viral fitness, and mechanisms that alter intrahost diversity may impact evolutionary trajectories (Vignuzzi et al., 2006; Beck et al., 2014; Lauring et al., 2012). If the transmission bottleneck is sufficiently wide, low frequency variants that arise within a host can plausibly be transmitted and spread through host populations (Poon et al., 2016; Varble et al., 2014). Understanding how intrahost diversity is generated and main-

tained and the extent to which host immune status impacts this diversity may be important for defining influenza virus' larger evolutionary patterns (Dinis et al., 2016; Murcia et al., 2012, 2013).

Here we used next generation sequencing to define the impact of vaccine-induced immunity on the intrahost diversity of influenza virus during natural infection. We specifically asked: (i) whether influenza viruses in vaccinated individuals represent escape variants, (ii) whether novel antigenic variants are found in hosts with non-sterilizing immunity, and (iii) the degree to which vaccine-induced immunity impacts the overall diversity of intrahost populations. Because we analyzed influenza populations from individuals enrolled in a randomized, double-blind, placebo-controlled trial of influenza vaccine efficacy (Ohmit et al., 2006, 2008; Monto et al., 2009), we were uniquely positioned to define this aspect of immunity to natural infection.

Results

Study subjects and specimens

We utilized influenza A RT-PCR-positive throat swab samples from a randomized, double-blind, placebo-controlled study of vaccine efficacy that took place during the 2004-2008 influenza seasons at six study sites in Michigan (Ohmit et al., 2006, 2008; Monto et al., 2009). This trial measured vaccine efficacy of both the trivalent inactivated (IIV) and live attenuated influenza vaccine (LAIV) compared to placebo and each other. We sequenced patient-derived influenza populations without culturing from three seasons: 2004-2005, 2005-2006, and 2007-2008. The 2006-2007 influenza season did not have enough influenza-positive samples for our study (total n=16). Influenza A (H3N2) strains dominated the other three seasons, and the circulating 2004-2005 virus was considered at the time to be only a modest mismatch with the vaccine strain. The other seasons were antigenically matched. The num-

bers of subjects each year were as follows: 2004-2005 season, 522 IIV, 519 LAIV, and 206 placebo (Ohmit et al., 2006); 2005-2006 season, 867 IIV, 853 LAIV, and 338 placebo (Ohmit et al., 2008); 2007-2008 season, 813 IIV, 814 LAIV, and 325 placebo (Monto et al., 2009). Over these 5119 person-years of observation, 165 individuals had culture- or RT-PCR-confirmed influenza A (H3N2) infection and specimens available for analysis. Of these, 80 individuals had received LAIV, 42 had received IIV, and 43 had received placebo. For 2004-2005, flu-positive samples were available for 28 subjects: 7 IIV, 12 LAIV, and 9 placebo, and for 2005-2006, 32 samples were available to study: 13 IIV, 14 LAIV, and 5 placebo. During the 2007-2008 season, 105 flu-positive samples were available: 22 IIV, 54 LAIV, and 29 placebo. We were able to amplify and quantify genomes for 119 of the 165 influenza-positive samples (Table 3.1). The average age of this sequenced cohort from all years was 24.5, indicating that participants were generally young and likely shared similar influenza pre-exposure histories particularly after randomization. The age, sex, and race of the cohort were similar to that of the overall study cohort for each of the 3 seasons.

Despite subject randomization, differences in pre-existing immunity due to prior vaccination or influenza infection could provide a dominant immune background that is not significantly altered by vaccination against that season's strain. In the 2007-2008 season, 50.4% of individuals in the larger cohort and 44.3% in the sequenced cohort reported having ever received a prior influenza vaccination. To obtain a more reliable metric of strain-specific immunity, we measured pre-season antibody titers by hemagglutination inhibition (HAI) and neuraminidase inhibition (NAI) assays for all study participants against that season's vaccine strain (Figure 3.6). Vaccine-induced antibody titers and overall vaccine efficacy are generally stable for a single flu season (Petrie et al., 2016). Therefore, our pre-season HAI and NAI titers are likely to

Table 3.1: Samples analyzed over three FLU-VACS Seasons

	2004-2005	2005-2006	2007-2008
H3N2 Vaccine Strain	Wyoming/3/2003	California/7/2004	Wisconsin/67/2005
Circulating Strain(s)	Wyoming/3/2003 California/7/2004	California/7/2004 Wisconsin/67/2005	Wisconsin/67/2005 Brisbane/10/2007
Demographics			
Average Age (Years)	30.0	28.8	22.8
Sex (% Female)	71.4%	83.3%	70.1%
Race (% White)	92.9%	94.4%	86.2%
Sequenced Samples	14	18	87
IIV	4	9	16
LAIV	5	6	44
Placebo	5	3	27
Intrahost SNV Data	10	9	64
IIV	3	6	11
LAIV	4	1	30
Placebo	3	2	23

be similar to titers at the time of infection. Pre-season (post-vaccination) titers for individuals in the IIV group were above the geometric mean for the entire sequenced cohort, LAIV subjects had titers spanning the mean, and those in the placebo group were generally below the mean for all seasons. Since an HAI titer of 40 is typically considered to be associated with 50% protection given exposure (Ohmit et al., 2011; Tsang et al., 2014; Potter and Oxford, 1979), these data demonstrate that in the IIV group, and to a lesser degree, the LAIV group, individuals had strain-specific antibody levels that are sufficient to apply selective pressure against the infecting virus.

Viral load of sequenced samples

We have previously shown that viral load influences the sensitivity and specificity of iSNV detection (McCrone and Lauring, 2016). In order to determine whether viral load was different among the IIV, LAIV, and placebo samples that we sequenced,

we measured genome copy number by RT-qPCR for the 2004-2005, 2005-2006, and 2007-2008 seasons. For the 2007-2008 season, which had the most samples, there were no significant differences in copy number by vaccination group (Figure 3.1A). In agreement with the 2007-2008 data, we did not detect differences in copy number by vaccination group for the 2004-2005 and 2005-2006 seasons (Figure 3.7). Since copy number is dependent on time from illness onset (Lee et al., 2009; Tsang et al., 2015), we analyzed the data based on sample collection day (Figure 3.1B). Using days 2-4, for which there were at least 5 data points for each treatment group, we did not find any significant differences (p=0.24-0.57 for days 2-4, non-parametric one way ANOVA). We divided the larger group of 2007-2008 subjects into groups based on pre-season HAI and NAI titers ≥ 40 or <40 against that season's strain. This cutoff was identical to the HAI and NAI geometric mean titers for our sequenced cohort (61.9 and 34.5, respectively), given the dilutions used. We did not detect differences in copy number based on HAI or NAI titer (Figure 3.1C-D), even when accounting for day of symptom onset (p=0.25 for HAI, p=0.97 for NAI, Mann-Whitney U test). Because we only measured copy number in the subset of virus populations that were amplified and sequenced, these data should not be interpreted in the context of vaccination and overall shedding (Petrie et al., 2011).

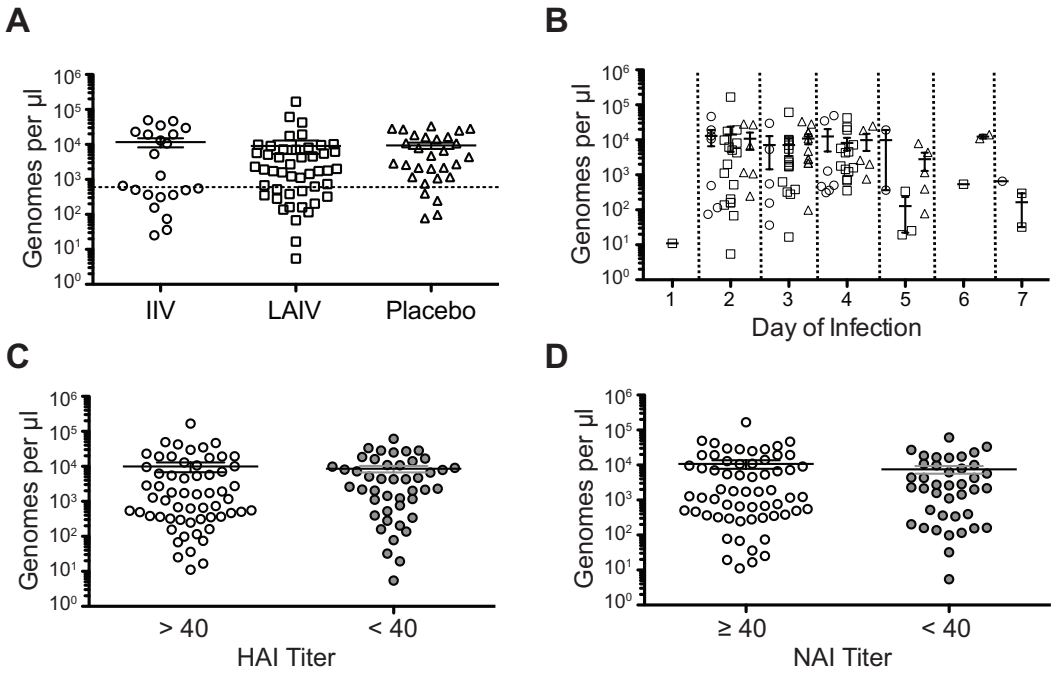


Figure 3.1: Viral shedding by vaccination status. Genome copy number per μl of transport media was determined by RT-qPCR all samples from the 2007-2008 season. (A) Copy number by vaccination status. IIV, inactivated influenza vaccine; LAIV, live attenuated influenza vaccine. (B) Copy number by day of infection (onset of symptoms is day 0) and vaccination status. Circle, IIV; Square, LAIV; Triangle, Placebo. (C) Copy number by HAI titer (D) Copy number by NAI titer. There were no differences among any of the groups by one-way ANOVA with Bonferroni correction.

Deep sequencing of intrahost influenza populations

We used the Illumina platform to determine the whole genome consensus sequence and to identify intrahost single nucleotide variants for each patient-derived sample. Importantly, we have developed and rigorously benchmarked a variant calling pipeline that maintains high sensitivity for rare iSNV detection while dramatically reducing false positive variant calls (McCrone and Lauring, 2016) (Supplementary-Table 3.3). We have found that the number of false positive iSNV calls is much higher in samples with genome copy numbers $< 10^3$ per μl of transport media; therefore, we only report iSNV from a high quality dataset that includes 64 samples from the 2007-2008 season. High quality iSNV data for the 2004-2005 and 2005-2006 influenza seasons are included in Supplementary Figures and Tables. Our libraries yielded an

average coverage above 20,000 reads per base with even coverage across the coding region of all segments for each season (Figure 3.8).

HA and NA sequences do not cluster by vaccination status or pre-season antibody titer

Given the frequent observation of community-level diversity in circulating influenza viruses (Ghedin et al., 2009; Holmes et al., 2011), we asked whether vaccinated individuals were infected with distinct strains relative to the placebo group. If vaccine failures were due to infection with an antigenically distinct variant, we would expect to see evidence of clustering by vaccination- or sero-status in HA and NA phylogenetic trees. We therefore analyzed HA and NA consensus sequences from all 87 individuals in the 2007-2008 season (Figure 3.2). There was very little diversity in either gene, and we found that sequences from individuals in each treatment group (e.g. IIV, LAIV, or placebo) were dispersed throughout the tree. More importantly, we found no evidence for clustering based on pre-season HAI and NAI titer (by colors in Figure 3.2, titers indicated at tips). We obtained similar results from the 2004-2005 and 2005-2006 seasons, albeit with fewer sequenced samples (Figure 3.9). These data suggest that within-season and within-host antigenic drift due to higher levels of vaccine-induced antibodies (HAI or NAI >40) were not major determinants of vaccine failure in these seasons.

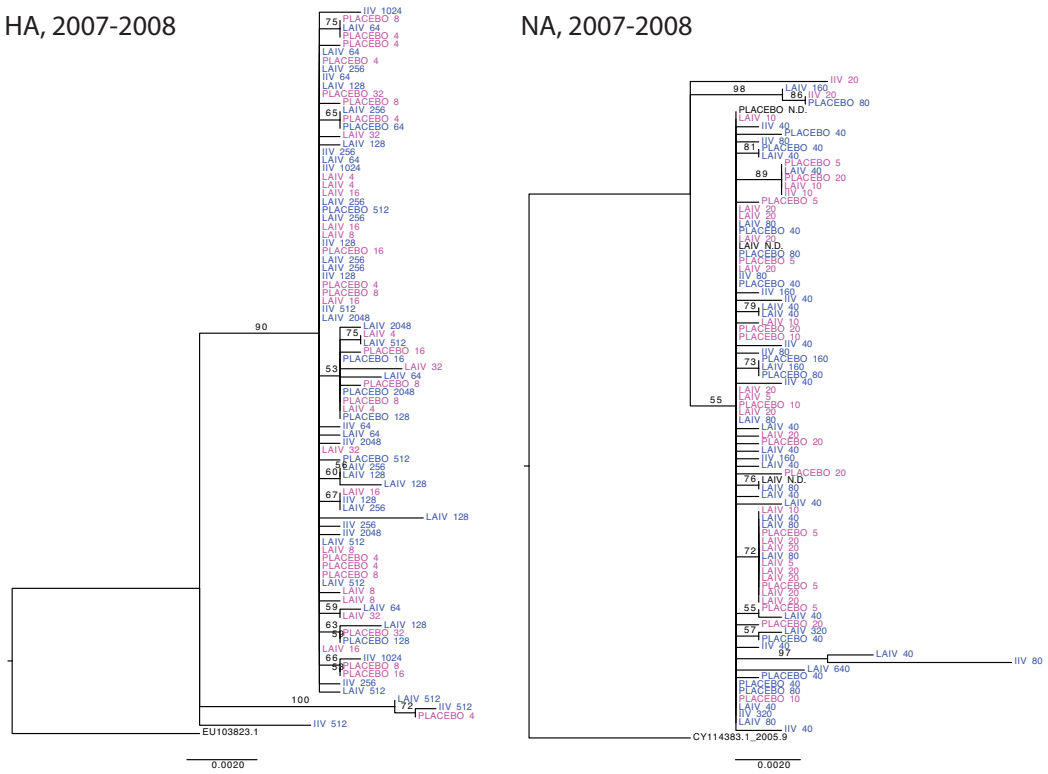


Figure 3.2: Phylogenetic trees of HA and NA consensus sequences from the 2007-2008 season. Maximum likelihood trees of HA (left) and NA (right) with tips coded by vaccine status and pre-season HAI (left; blue >40, magenta <40) or NAI (right; blue > 40, magenta <40) titer. HAI (left) and NAI (right) titers are shown on tips as well. Outgroups are HA (EU103823.1) and NA (CY114383.1) for the vaccine strain A/Wisconsin/67/2005. Bootstrap values (n = 1000 bootstraps) are shown and nodes with bootstrap values <50 are collapsed for easier visualization.

Intrahost diversity in vaccinated and unvaccinated individuals

We next analyzed the iSNVs present in 64 of our samples from the 2007-2008 season. We identified 360 minority variants across the entire genome, most of which were present at a frequency of <0.1 (Figure 3.3A). We did not observe many samples with a large number of higher frequency iSNV, which suggests that there were few, if any, mixed lineage infections in the samples from this season. The vast majority of iSNV were only found once (Figure 3.3B). We also evaluated whether partial immunity impacts viral diversity by comparing the number of iSNVs per sample in the HA and NA genes based on HAI and NAI titers, respectively. We did not

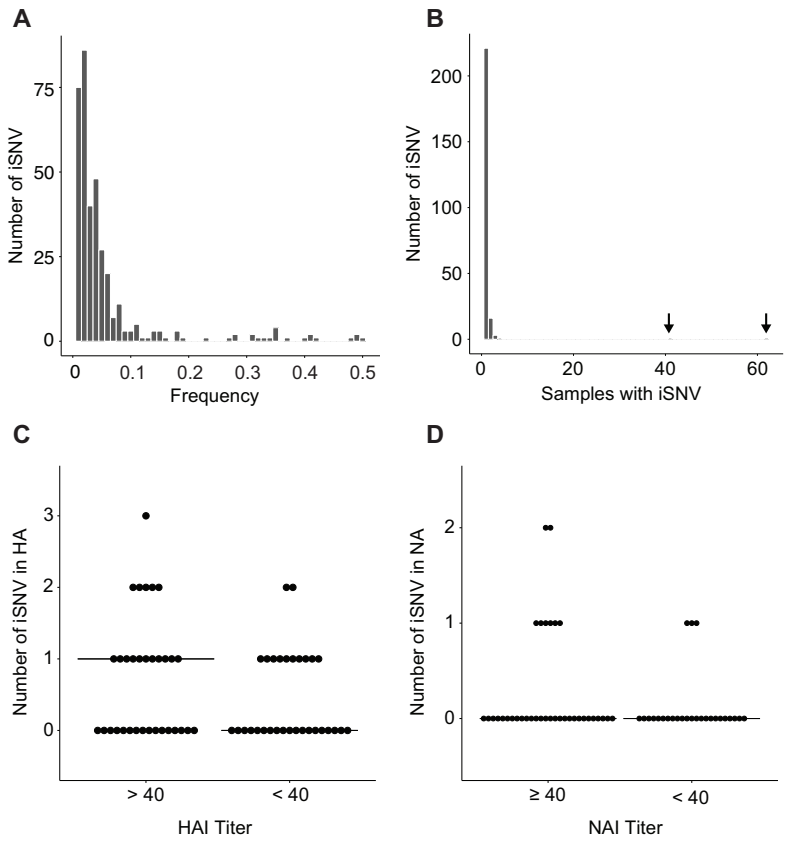


Figure 3.3: Intrahost diversity in samples from the 2007-2008 season. (A) Histogram of the number of iSNV at a given frequency. Bin width = 0.01. (B) Histogram of the number of samples in which each iSNV is found. Arrows indicate bars with one SNV, which are hard to discern in the histogram. These polymorphic SNV, at PB2 position 900 and PA position 515 respectively, were found at 4-6% frequency within hosts and in similar numbers of individuals across vaccination groups. (C) Number of HA iSNV per sample stratified by pre-season HAI titer. ≥ 40 =serologically immune, <40 = not serologically immune. (D) Number of NA iSNV per sample stratified by pre-season NAI titer. ≥ 40 = serologically immune, <40 = not serologically immune.

observe a difference in iSNV count based on HAI and NAI titers for either HA or NA (Figure 3.3C-D, p=0.20 for HA, p=0.26 for NA, Mann Whitney U test; see also Figure 3.10 for iSNV stratified by titer). The average number of iSNV per sample was similar across the genome regardless of host treatment group (Table 3.2, p=0.35, non-parametric one way ANOVA) or HAI and NAI titers (Figure 3.11, p=0.13 for HAI, p=0.22 for NAI, Mann Whitney U test).

Table 3.2: Number of iSNV (mean ± interquartile range) by segment and treatment group for samples from the 2007-2008 season.

Segment	IIV (n=11)	LAIV (n=30)	Placebo (n=23)
1 (PB2)	1.91 ± (2)	1.50 ± (1)	1.57 ± (1)
2 (PB1)	0.82 ± (1.5)	0.60 ± (1)	0.70 ± (1)
3 (PA)	1.64 ± (1)	1.57 ± (1)	1.65 ± (1)
4 (HA)	0.91 ± (1.5)	0.47 ± (1)	0.61 ± (1)
5 (NP)	0.64 ± (1)	0.27 ± (0)	0.30 ± (0)
6 (NA)	0.27 ± (0.5)	0.23 ± (0)	0.22 ± (0)
7 (M)	0.09 ± (0)	0.27 ± (0)	0.35 ± (1)
8 (NS)	0.36 ± (1)	0.23 ± (0)	0.39 ± (1)

Most iSNV from the 2004-2005 and 2005-2006 seasons were also found in only one sample each and at frequencies <0.1 (Figure 3.12). The number of iSNV did not differ based on treatment group or HAI and NAI titer (Supplementary Table 3.4). We did not identify any variants specifically associated with vaccination group or HAI and NAI titers for any of the influenza seasons analyzed. Unlike experimental challenge studies, we only had one sample per person and could not evaluate changes in diversity at the level of the individual host. However, we did not identify any significant differences in diversity by day of infection across the cohort by vaccination status (Figure 3.4, p=0.16-0.82 for days 2-4, non-parametric one way ANOVA) or antibody titer. We identified only a marginal difference in the number iSNV across the genome based on HAI titer on day 2 (uncorrected p-value 0.02 with >6 comparisons, Mann Whitney U test). These data suggest that our results are unlikely to be confounded by temporal sampling issues.

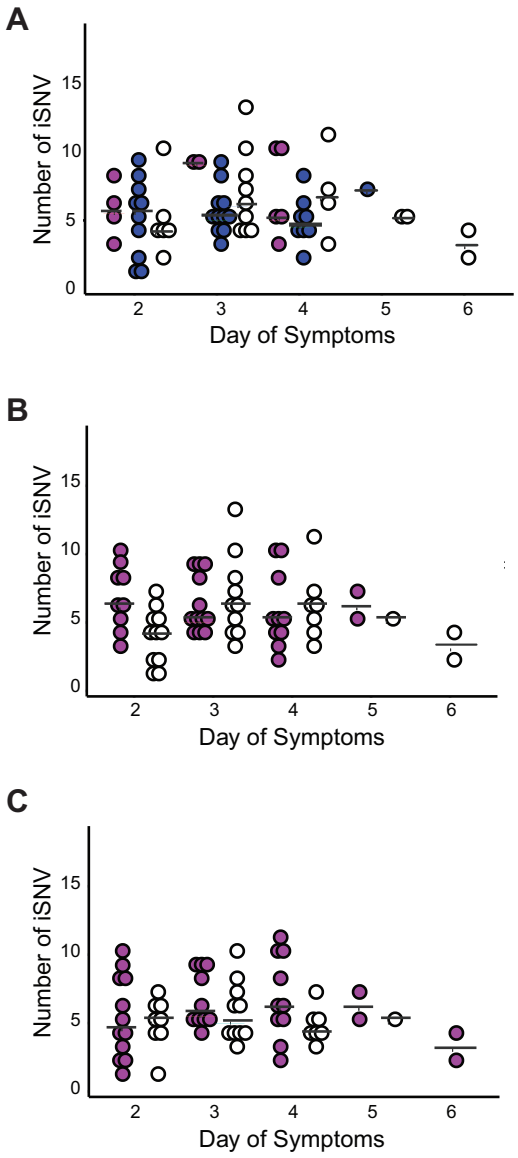


Figure 3.4: Temporal patterns of intrahost diversity. Number of genome-wide iSNV per sample (y-axis) by day of symptoms (x-axis) stratified by (A) recipients of IIV, magenta; LAIV, blue; placebo, white (B) HAI >40, magenta; HAI <40, white (C) NAI $\geq$ 40, magenta; NAI <40, white. Mean number of iSNV in each group is indicated (bar).

Evidence for low antigenic diversity within hosts

Some have suggested that vaccine-induced immunity will select for novel antigenic variants within hosts (Dinis et al., 2016). We therefore compared iSNV in the HA gene by vaccination group and serostatus. Of the 17 variants we identified that resulted in nonsynonymous changes within HA, 11 were in HA1 and 6 were in HA2 (Figure 3.5, Supplementary Table 3.5). Five antigenic sites, comprising 131 amino acid positions, have been described in HA1 for H3N2 viruses (Caton et al., 1982; Skehel et al., 1985; Lee and Chen, 2004). Six of the HA1 variants identified in our study were located in antigenic sites C and D (Figure 3.5B), one of which was found in two samples (Supplementary Table 3.5). Of these potential antigenic variants, three were found in vaccinated individuals (1 IIV, 2 LAIV) and four were found in those in the placebo group. When grouped by HAI titer, four were found in samples from individuals with titers ≥ 40, while three were found in individuals with titers <40. No variants were specific to vaccinated individuals, and antigenic diversity was similar across all groups. None of the identified mutations were observed in subsequent circulating H3N2 strains. There were no nonsynonymous iSNV in HA from the 2004-2005 samples and just 7 identified in individuals from the 2005-2006 season. Of these 7, N225D was observed in strains dominating prior, but not subsequent seasons.

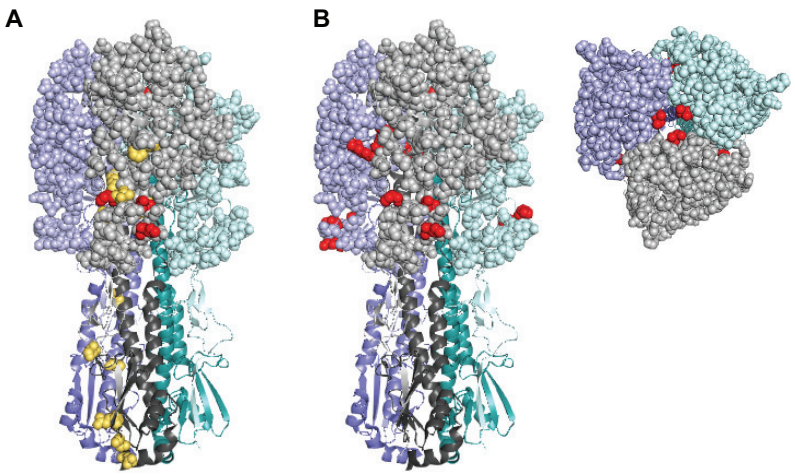


Figure 3.5: Structural mapping of HA variants. A homology model of the A/Brisbane/10/2007 (H3N2) HA trimer is shown, with each monomer represented by a different color (purple, grey, and teal) and HA1 and HA2 designated by lighter and darker shades of the same color, respectively. (A) A side view of HA. All identified non-synonymous mutations and known antigenic amino acid positions are shown as balls on the grey monomer. Those variants colored red are within known antigenic sites, while light orange mutations are not. (B) A side and top view of HA. All amino acid positions within known antigenic sites are displayed as balls, with the antigenic mutations identified here shown in red on all three monomers.

Discussion

We set out to define the relationship between vaccine-induced immunity and the intrahost diversity of influenza virus. We hypothesized that non-sterilizing immunity could potentially select for novel antigenic variants and contribute to larger scale patterns of influenza evolution. We were able to definitively address these questions using samples derived from a randomized, placebo-controlled vaccine trial in healthy, young adults who likely had similar prior exposure to influenza viruses (Ohmit et al., 2006; Monto et al., 2009). The availability of post-vaccination, preseason HAI and NAI titers allowed us to examine directly the impact of measured serologic antibody pressure rather than using vaccination alone as a surrogate marker. Because all individuals were infected naturally, our data provide a rare view of within-host influenza virus diversity in humans. We directly sequenced the samples without passage in

cell culture, eliminating the possibility of culture-adapted mutations and employed a well-benchmarked variant calling pipeline (McCrone and Lauring, 2016) that dramatically reduces the false positive iSNV calls that often plague next generation sequencing studies. In this exhaustive and well-controlled study, we found no differences in intrahost influenza diversity based on vaccination status or HAI and NAI titers.

Our findings in a natural infection system are concordant with an equine influenza virus evolution study in vaccinated horses. Intrahost variation was similar between nave and vaccinated horses, regardless of whether they were infected naturally or experimentally (Murcia et al., 2013). However, not all experimental infection studies mirror results seen during natural infection. A study investigating swine influenza virus found discrepancies in intrahost variation based on whether animals encountered natural infection or were experimentally infected (Murcia et al., 2012). Our data are in contrast with a study of experimentally-infected dogs that uncovered differences in intrahost diversity and evolution in antigenic sites based on vaccination status (Hoelzer et al., 2010). Two other studies of equine influenza virus found mixed infections of multiple influenza lineages during natural infection, which would not be seen in an experimental model but may be relevant to the transmission and spread of novel variants (Murcia et al., 2010; Hughes et al., 2012). Overall, it is challenging to compare our results to those obtained in various animal models, where the hosts are often immunologically naive and infected with a defined inoculum of a single genotype.

We did not detect phylogenetic clustering of HA and NA sequences based on vaccination status, type of administered vaccine, or pre-season HAI and NAI titers. These results are consistent with those of Dinis et al., who found no segregation of

HA sequences based on vaccination status in a case test-negative study of vaccine effectiveness (Dinis et al., 2016). Together, these data suggest that vaccinated and unvaccinated individuals are infected with similar strains and that within season antigenic drift is not a major contributor to reduced vaccine efficacy in the seasons analyzed. Because we also stratified our analysis by pre-season HAI and NAI titer, we can similarly exclude viral escape from a non-sterilizing antibody response. Our data further suggest that pre-existing antibody against circulating strains does not apply sufficiently strong selective pressure to drive the emergence of antigenically distinct strains within a given host.

High intrahost diversity may be an important factor in viral evolution, since it increases the number of novel variants on which natural selection can act. Some have proposed that the intrahost diversity of RNA viruses is linked to virulence (Vignuzzi et al., 2006; Beck et al., 2014), suggesting that processes that act to restrict or enhance intrahost diversity may alter disease phenotypes. We found that within-host diversity of influenza virus was quite low. Most iSNV were present at frequencies of less than 0.1, which means they would only plausibly be spread between hosts if the transmission bottleneck were reasonably large (Poon et al., 2016). The number of iSNV in a given host was similar between vaccine and placebo groups, both across the genome and on the segments coding for HA and NA. Furthermore, there were no significant differences in diversity when samples were grouped by HAI and NAI titer, and we did not find evidence of treatment-specific iSNV. Together our data suggest that vaccine-induced immunity does not significantly influence intrahost diversity and is a relatively weak selective pressure at the level of the individual host.

While the number of iSNV was similar in all groups, we considered it possible that partial immunity may drive the emergence of specific antigenic variants. Vaccination

and natural infection can induce a wide range of immune responses depending on host and viral factors. These can range from complete protection to leaky responses that allow infection, but influence disease severity or duration. Non-sterilizing immune responses have the potential to select for escape variants within each host. In the setting of current recommendations for universal vaccination, a highly vaccinated population could potentially select for antigenically evolved viruses more quickly than the spread of natural infection. For this reason, it is important to tease apart the role of vaccination on influenza evolution. We did find a number of variants in antigenic sites, but these were no more frequent than in other regions of the genome and did not vary with vaccination status or HAI and NAI titer. Importantly, our results are in conflict with earlier work in several model systems that demonstrate the rapid selection of antigenic variants in the presence of sub-neutralizing antibody or in experimentally infected animals (Murcia et al., 2012, 2013; GERBER et al., 1955; Gerber et al., 1956; Hamre et al., 1958). This discrepancy may be due to differences in infectious dose, host genetic background, history of prior influenza infection, immune correlates not captured by HAI and NAI titers, or the strains tested. Animal studies are often performed in immunologically naive or genetically identical animals, whereas humans have complex genetic backgrounds and immunological histories that could play a role in mediating population-wide immunity.

While global patterns of influenza transmission and evolution are complex, our study provides important insights regarding influenza evolution on intra- and inter-host scales. We were able to define aspects of intrahost evolution within a geographically-constrained and relatively young cohort with similar vaccination histories and previous influenza exposures across groups, limiting potential confounding factors. Still, we acknowledge that by reducing these confounders, we may be missing important

determinants of intrahost evolution. For example, intrahost diversity may be different in children, older adults, or those with high-risk conditions. Our study included 5119 person-years of observations to yield a dataset of 165 viruses, 119 of which were sequenced. Large, placebo-controlled, randomized influenza vaccine trials involving thousands of people are unlikely to be conducted in the future due to the recommendation for universal vaccination. Therefore, our sample set likely represents the best chance to directly assess the impact of vaccination on influenza evolution in the context of natural human infection.

Despite our well-controlled study, we found little evidence for vaccine-driven evolution in the context of the community that was sampled. None of the low frequency variants identified in the antigenic sites were found in subsequent seasons. While we cannot rule out the possibility that evolutionary patterns would be different in other geographic regions, different influenza seasons, or with other subtypes, we did not uncover differences between vaccinated and unvaccinated populations with respect to genome-wide or antigenic diversity in three H3N2 seasons. We were not able to evaluate the impact of vaccination on H1N1 or subtype B viruses. Furthermore, our sample set cannot discern whether there are differences in evolutionary pressure based on vaccine match/mismatch, as the vast majority of our samples were from antigenically matched seasons. A major limitation of our study was that only one sample was available for each participant, so we could not track changes in diversity or mutation accumulation within each individual over the course of their infection. However, we did not observe significant differences in the number of mutations over the first 6 days of infection, which is consistent with previous work in horses (Murcia et al., 2013) and a deep sequencing study of 7 humans in an experimental challenge model (Sobel Leonard et al., 2016). Together with these works, our data suggest

that within-host dynamics are dominated by purifying selection with the transient appearance of minority variants and little sustained fixation. Selection and transmission of antigenically and epidemiologically important variants is likely to be a rare event when studied at this scale. Given that our data are derived from 165 incident infections in 5119 season-years of observation, detection of such events in the course of a natural infection would require an unrealistic sample size.

We evaluated the potential for vaccine-induced immunity to drive intrahost evolution, as this is an important issue in light of the current recommendation for universal influenza vaccination. We did not find evidence for vaccine-induced pressure on the intrahost or consensus level, despite employing several methods used to study evolutionary processes. Our study is larger than other reports and involves extensive analysis of both placebo and vaccination groups. While randomization, placebo control, and reliance on a young healthy population allowed us to address this question in a rigorous manner, these factors may have lessened person-to-person variation. By better defining the intrahost evolutionary mechanisms and how they impact population-wide influenza evolution, we hope to address questions of import to clinicians and public health workers, improve vaccine design, and develop more efficient epidemiological control measures.

Materials and Methods

Ethics Statement

This study was approved by the Institutional Review Board of the University of Michigan Medical School, and all human subjects provided informed consent.

Subjects and specimens

We characterized host-derived influenza populations archived from a randomized, double-blind, placebo-controlled, clinical trial of influenza vaccine efficacy that ran from the 2004-2005 through the 2007-2008 influenza seasons (ClinicalTrials.gov number, NCT00133523, (Ohmit et al., 2006, 2008; Monto et al., 2009)). Each year, healthy adults, ages 18-49, were randomized to receive trivalent inactivated influenza vaccine (IIV), live attenuated influenza vaccine (LAIV), or placebo. Throat swab specimens were collected from individuals with influenza-like illness within 7 days of onset; residual specimen material was stored in veal infusion broth (VIB) at -80C. Viral RNA was extracted from 140μl of VIB using the QIAamp viral RNA mini kit (Qiagen 52906), eluted in 50μl buffer, and stored at -80C. Hemagglutination inhibition (HAI) and Neuraminidase agglutination inhibition (NAI) titers for subjects in this study were previously measured and reported in (Ohmit et al., 2011; Monto et al., 2015).

Determination of genome copy number

Quantitative reverse transcription polymerase chain reaction (RT-qPCR) was performed on 5μl RNA from each sample using CDC RT-PCR primers InfA Forward, InfA Reverse, and InfA probe, which bind to a portion of the influenza M gene (CDC protocol, 28 April 2009). Each reaction contained 5.4μl nuclease-free water, 0.5μl each primer/probe, 0.5μl SuperScript III RT/Platinum Taq mix (Invitrogen 111732) 12.5μl PCR Master Mix, 0.1μl ROX, 5μl RNA. The PCR master mix was thawed and stored at 4C, 24 hours before reaction set-up. A standard curve relating copy number to Ct values was generated based on 10-fold dilutions of a control plasmid run in duplicate.

Illumina library preparation and sequencing

We amplified cDNA corresponding to all 8 genomic segments from 3μl of the viral RNA using the SuperScript III One-Step RT-PCR Platinum Taq HiFi Kit (Invitrogen 12574). Reactions consisted of 0.5μl Superscript III Platinum Taq Mix, 12.5μl 2x reaction buffer, 8μl DEPC water, and 0.2μl of 10μM Uni12/Inf1, 0.3μl of 10μM Uni12/Inf3, and 0.5μl of 10μM Uni13/Inf1 universal influenza A primers (Zhou et al., 2009). The thermocycler protocol was: 42C for 60 min then 94C for 2 min then 5 cycles of 94C for 30 sec, 44C for 30 sec, 68C for 3 min, then 28 cycles of 94C for 30 sec, 57C for 30 sec, 68C for 3 min. Amplification of all 8 segments was confirmed by gel electrophoresis, and 750ng of each cDNA mixture were sheared to an average size of 300 to 400bp using a Covaris S220 focused ultrasonicator. Sequencing libraries were prepared using the NEBNext Ultra DNA library prep kit (NEB E7370L), Agencourt AMPure XP beads (Beckman Coulter A63881), and NEBNext multiplex oligonucleotides for Illumina (NEB E7600S). The final concentration of each barcoded library was determined by Quanti PicoGreen dsDNA quantification (ThermoFisher Scientific), and equal nanomolar concentrations were pooled. Residual primer dimers were removed by gel isolation of a 300-500bp band, which was purified using a GeneJet Gel Extraction Kit (ThermoFisher Scientific). Purified library pools were sequenced on an Illumina HiSeq 2500 with 2x125 nucleotide paired end reads. All raw sequence data have been deposited at the NCBI sequence read archive (BioProject submission ID: SUB1907046)

Variant detection

Sequencing reads that passed standard Illumina quality control filters were binned by index and aligned to the reference genome using Bowtie (Langmead et al., 2009).

Single nucleotide variants (SNV) were identified and analyzed using DeepSNV (Gerstung et al., 2012), which relies on a clonal control to estimate the local error rate within a given sequence context and to identify strand bias in base calling. The clonal control was a library prepared in an identical fashion from 8 plasmids containing the genome for the respective circulating reference strain and sequenced in the same flow cell to control for batch effects. True positive SNV were identified from the raw output tables by applying the following filtering criteria in R: (i) Bonferonni corrected p value <0.01, (ii) average MapQ score on variant reads >30, (iii) average phred score on variant positions >35, (iv) average position of variant call on a read >32 and <94, (v) variant frequency >0.01. We only considered SNV identified in a single RT-PCR reaction and sequencing library for samples with copy number $\geq 10^5$ genomes/μl transport media or in two separate RT-PCR reactions and sequencing libraries for samples with copy number 10^3-10^5 genomes per μl. For variants at a frequency of 50-99%, we called the minority base (i.e. 1-50%) as an iSNV. Our strategy for variant calling is described in (McCrone and Lauring, 2016) and all code can be found at `https://github.com/lauringlab/variant_pipeline`.

Phylogenetic analysis

Consensus nucleotide sequences for the HA and NA proteins were aligned using MUSCLE (Edgar, 2004). The best-fit models for nucleotide substitution was identified using jModelTest v2.1.10 (Darriba et al., 2012). Maximum likelihood phylogenetic trees were generated using RAxML v8 (Stamatakis, 2014) with a GTRGAMMA model, Genbank sequences for vaccine strains as outgroups, and 1000 bootstraps. Trees were visualized and annotated using FigTree (v1.4.2).

Data analysis and statistics

All statistical analyses were performed using Prism 6 and R. Description of the analysis and annotated code are available at `https://github.com/lauringlab/fluvacs_paper`. HA structural models were generated and visualized with PyMol.

Acknowledgements

We thank Suzanne Ohmit, Ryan Malosh, Emily Toth Martin, and Robert Woods for helpful discussion. I would like to also acknowledge Kari Debbink who significantly contributed to writing this chapter and is listed a co-first author in the published version (Debbink et al., 2017).

Supplemental Figures and Tables

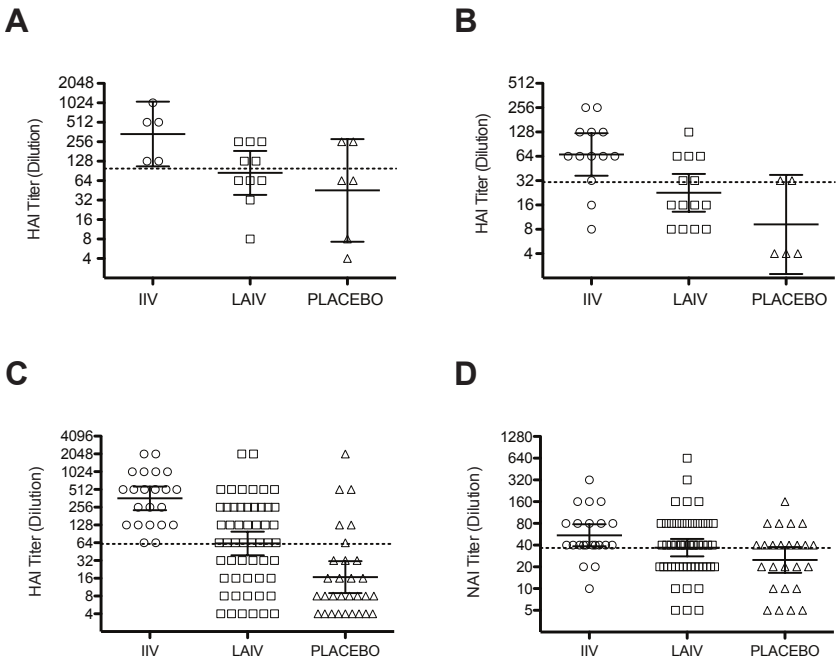


Figure 3.6: Pre-season hemagglutination inhibition (HAI, A-C) and neuraminidase inhibition (NAI, D) titers against that season's vaccine strain for all individuals in this study. (A) HAI titers for individuals from the 2004-2005 season. (B) HAI titers for individuals from the 2005-2006 season. (C) HAI titers for individuals from the 2007-2008 season. (D) NAI titers for individuals from the 2007-2008 season. IIV, inactivated influenza vaccine; LAIV, live attenuated influenza vaccine. Dotted line, geometric mean.

Table 3.3: Variant Detection Specificity and Sensitivity by Copy Number. Sensitivity and specificity for variant detection in experimental influenza populations with variants of known frequency and input titer. The benchmarking experiment and data are described in (McCrone and Lauring, 2016)

Copy Number[a]	Variant Frequency	Sensitivity	Specificity
$> 10^5$	0.050	1.00	>0.9999
	0.020	0.85	0.9999
	0.010	0.95	0.9995
	0.005	0.35	0.9999
$10^4 - 10^5$	0.050	0.95	0.9999
	0.020	0.90	0.9999
	0.010	0.80	0.9998
	0.005	0.40	0.9999
$10^3 - 10^4$	0.050	0.80	>0.9999
	0.020	0.45	0.9999
	0.010	0.20	0.9997
	0.005	0.10	0.9999

[a] Per μl transport media

Table 3.4: Number of iSNV (mean ± interquartile range) by segment and treatment group and HAI titer

	Segment	IIV	LAIV	Placebo	HAI > 1:40	HAI <1:40
2004-2005	1 (PB2)	1.67 ± (1.5)	2 ± (0.5)	3.33 ± (0.5)	2 ± (1.25)	4 ± (0)
	2 (PB1)	6 ± (8.5)	1 ± (0.5)	0.67 ± (0.5)	2.75 ± (1.25)	1 ± (0)
	3 (PA)	6.33 ± (9.5)	0.5 ± (1)	2 ± (1)	2.75 ± (1)	3 ± (0)
	4 (HA)	0 ± (0)	0 ± (0)	0 ± (0)	0 ± (0)	0 ± (0)
	5 (NP)	0.67 ± (0.5)	0.75 ± (1.25)	1 ± (1)	0.88 ± (1.25)	1 ± (0)
	6 (NA)	1.33 ± (0.5)	1 ± (0)	1.67 ± (0.5)	1.12 ± (0)	2 ± (0)
	7 (M)	2.67 ± (4)	0 ± (0)	0.33 ± (0.5)	1 ± (0)	1 ± (0)
	8 (NS)	0.33 ± (0.5)	0 ± (0)	0 ± (0)	0.12 ± (0)	0 ± (0)
	Segment	IIV	LAIV	Placebo	HAI > 1:40	HAI < 1:40
2005-2006	1 (PB2)	2.33 ± (0)	4 ± (0)	3 ± (0)	2.33 ± (0)	3.33 ± (0.5)
	2 (PB1)	0.83 ± (1)	1 ± (0)	0.5 ± (0.5)	0.83 ± (1)	0.67 ± (0.5)
	3 (PA)	0.33 ± (0.75)	5 ± (0)	0.5 ± (0.5)	0.33 ± (0.75)	2 ± (2.5)
	4 (HA)	1.67 ± (0.75)	5 ± (0)	0 ± (0)	1.67 ± (0.75)	1.67 ± (2.5)
	5 (NP)	0.5 ± (1)	1 ± (0)	2 ± (0)	0.5 ± (1)	1.67 ± (0.5)
	6 (NA)	1.5 ± (0)	1 ± (0)	1.5 ± (0.5)	1.5 ± (0)	1.33 ± (0.5)
	7 (M)	0.33 ± (0.75)	0 ± (0)	0.5 ± (0.5)	0.33 ± (0.75)	0.33 ± (0.5)
	8 (NS)	0.5 ± (0)	0 ± (0)	0.5 ± (0.5)	0.5 ± (0)	0.33 ± (0.5)

Table 3.5: Nonsynonymous variants in HA

HA Region	Amino Acid Substitution	Variant Frequency	Vaccination Status	Pre-season HAI Titer
HA1	G49S	0.030	Placebo	16
HA1	D53N*	0.012	Placebo	4
HA1	F79L	0.040	Placebo	8
HA1	K208R*	0.016	Placebo	8
HA1	K208R*	0.015	Placebo	4
HA1	I214V*	0.034	Placebo	512
HA1	K238N*	0.029	IIV	1024
HA1	R269K	0.079	LAIV	256
HA1	P273S*	0.014	LAIV	2048
HA1	I274M	0.010	LAIV	16
HA1	I278D*	0.030	LAIV	128
HA1	V323I	0.019	Placebo	512
HA2	A5S	0.017	LAIV	64
HA2	Q27L	0.021	Placebo	4
HA2	S29Y	0.057	IIV	512
HA2	Q65L	0.010	IIV	2048
HA2	A101T	0.028	LAIV	16
HA2	N169H	0.016	IIV	256

* Antigenic sites, as described in text and Figure 3.5

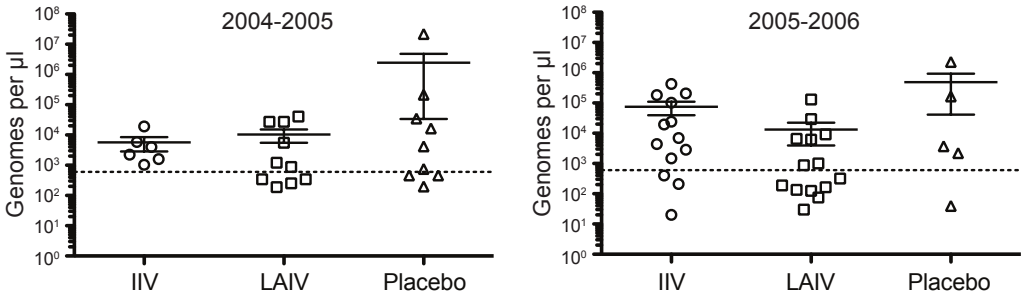


Figure 3.7: Genome copy number per μl viral transport media (y-axis) as determined by RT-qPCR for samples from 2004-2005 (left) and 2005-2006 (right) seasons by treatment group. IIV, inactivated influenza vaccine; LAIV, live attenuated influenza vaccine.

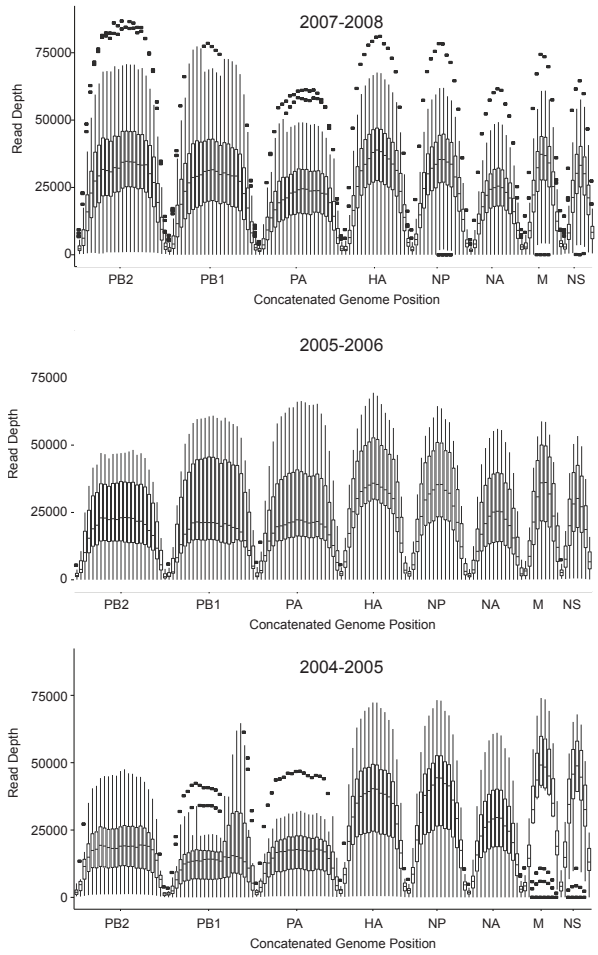


Figure 3.8: Sequence coverage for all samples in indicated seasons. The x-axis denotes each gene segment with bin width of 100 bases, while the y-axis shows average coverage per base within the bin. Boxplots are mean $\pm$ 1.5 interquartile range for samples within a season.

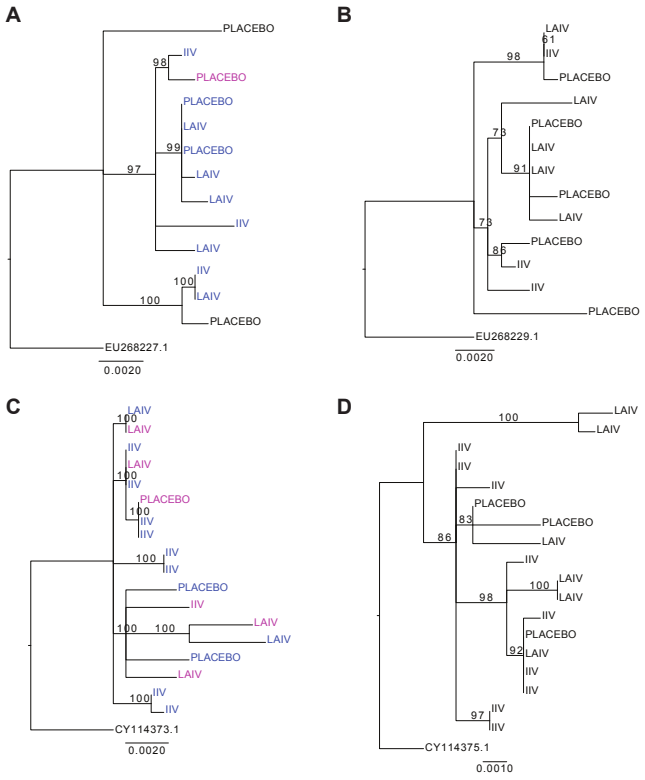


Figure 3.9: Maximum likelihood phylogenetic trees of consensus sequences for (A) HA, 2004-2005 (B) NA, 2004-2005 (C) HA 2005-2006 (D) NA 2005-2006 HA with tips coded by vaccine status and pre-season HAI (blue >40, magenta <40). Black, no data (note no NAI data). Outgroups are A/Wyoming/03/2003 (A,B) and A/California/7/2004 (C,D). Bootstrap values (n=1000 bootstraps) are shown and nodes with bootstrap values <50 are collapsed for easier visualization.

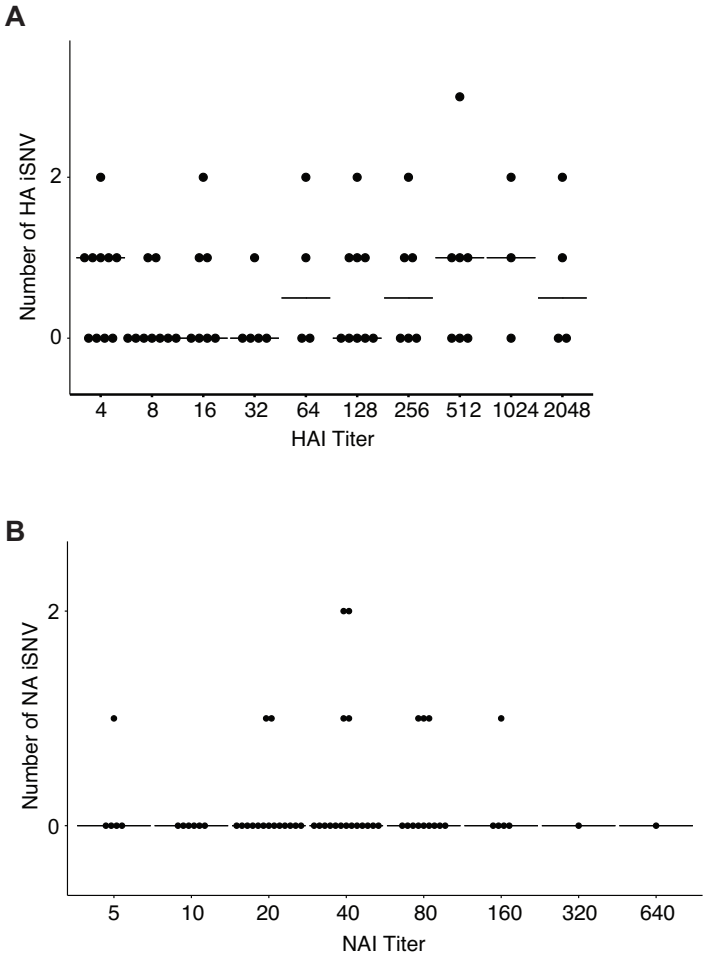


Figure 3.10: (A) Number of iSNV in HA per sample (y-axis) by HAI titer in that individual (x-axis). (B) Number of iSNV in NA per sample (y-axis) by NAI titer in that individual (x-axis).

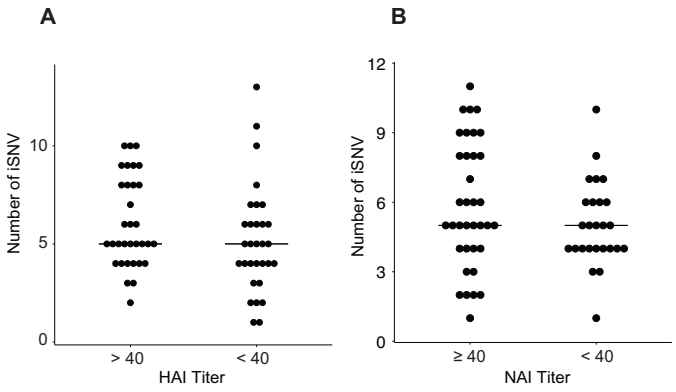


Figure 3.11: Number of genome-wide iSNV per sample by serological immune status for samples from the 2007-2008 season. (A) Whole genome iSNV count per sample based on (A) HAI titer or (B) NAI titer. ≥ 40 = serologically immune, < 40 = not serologically immune.

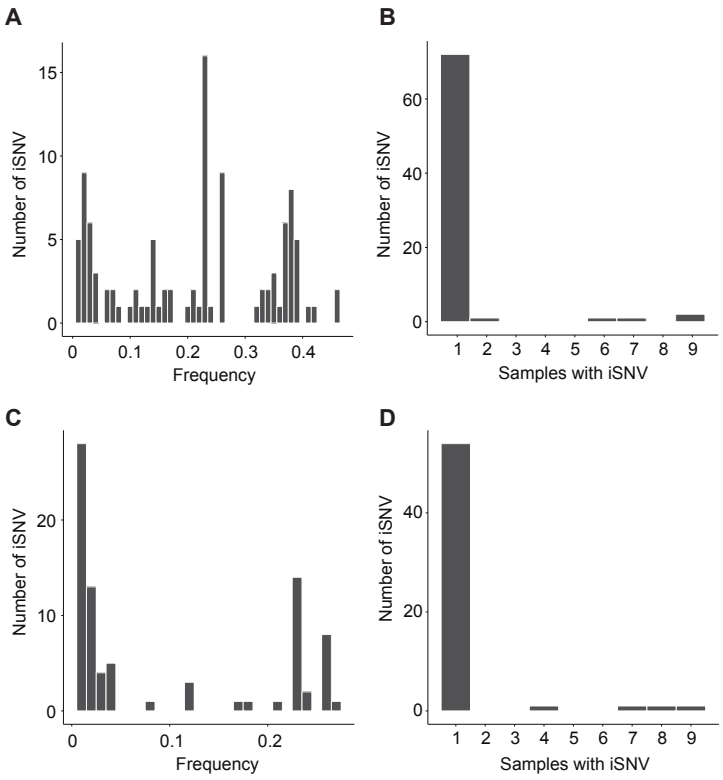


Figure 3.12: Histograms of iSNV frequency across entire cohort for the 2004-2005 (A) and 2005-2006 (C) seasons. Histograms of the number of specimens in which a given iSNV was found for the 2004-2005 (B) and 2005-2006 (D) seasons.

CHAPTER IV

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

Note: This chapter is a modified version of the preprint article:

McCrone JT, Woods RJ, Martin ET, Malosh RE, Monto AS, Lauring AS. 2017. The evolutionary dynamics of influenza A virus within and between human hosts. bioRxiv 153.

bioRxiv grants authors full rights to reuse their articles in dissertations. Adapted sections of this chapter have been submitted for future publication.

Introduction

The rapid evolution of influenza viruses has led to reduced vaccine efficacy, widespread drug resistance, and the continuing emergence of novel strains. Broadly speaking, evolution is the product of deterministic processes, such as selection, and stochastic processes, such as genetic drift (Kouyos et al., 2006). The relative contribution of each is greatly affected by the effective population size, or size of an idealized population whose dynamics are similar to that of the population in question (Rouzine et al., 2001). If the effective population size of a virus is large, as in quasispecies models, evolution is largely deterministic and the frequency of a mutation can be predicted based on its starting frequency and selection coefficient. In small populations, selection is inefficient, and changes in mutation frequency are strongly influenced by migration or genetic drift.

Viral dynamics may differ across spatial and temporal scales, and a complete understanding of influenza evolution requires studies at all levels (Nelson and Holmes, 2007; Holmes, 2009). The global evolution of influenza A virus (IAV) is dominated by the positive selection of novel antigenic variants that circulate in the tropics and subsequently seed annual epidemics in the Northern and Southern hemisphere (Rambaut et al., 2008). Whole genome sequencing has also demonstrated the importance of intrasubtype reassortment to the emergence of diverse strains that differ in their antigenicity. While continual positive selection of antigenically drifted variants drives global patterns, whole genome sequencing of viruses on more local scales suggests the importance of stochastic processes such as strain migration and within-clade reassortment (Nelson et al., 2006).

It is now feasible to efficiently sequence patient-derived isolates at sufficient depth of coverage to define the diversity and dynamics of virus evolution within individual hosts (Kao et al., 2014). Studies of IAV populations in animal and human systems suggest that most intrahost single nucleotide variants (iSNV) are rare and that intrahost populations are subject to strong purifying selection (Rogers et al., 2015; Murcia et al., 2010; Iqbal et al., 2009; Dinis et al., 2016; Debbink et al., 2017). While positive selection of adaptive variants is commonly observed in cell culture (Doud et al., 2017; Archetti and Horsfall, 1950; Foll et al., 2014), it has only been documented within human hosts in the extreme cases of drug resistance (Gubareva et al., 2001; Ghedin et al., 2010; Rogers et al., 2015), long-term infection of immunocompromised hosts (Xue et al., 2017), or experimental infections with attenuated viruses (Sobel Leonard et al., 2016). Indeed, we and others have been unable to identify evidence for positive selection in natural human infections (Debbink et al., 2017; Dinis et al., 2016), and its relevance to within host processes is unclear.

Despite limited evidence for positive selection, it is clear that novel mutations do arise within hosts. Their potential for subsequent spread through host populations is determined by the size of the transmission bottleneck (Alizon et al., 2011; Zwart and Elena, 2015). If the transmission bottleneck is sufficiently wide, low frequency variants can plausibly be transmitted and spread through host populations (Geoghegan et al., 2016). While experimental infections of ferrets suggest a very narrow transmission bottleneck (Varble et al., 2014; Wilker et al., 2013), studies of equine influenza support a bottleneck wide enough to allow transmission of rare iSNV (Hughes et al., 2012; Murcia et al., 2010). The only available genetic study of influenza virus transmission in humans estimated a large transmission bottleneck, allowing for transmission of 100-200 genomes (Poon et al., 2016; Sobel Leonard et al., 2017b).

Here, we use next generation sequencing of within-host influenza virus populations to elucidate the evolutionary dynamics of influenza A viruses (IAV) within and between human hosts. We apply a benchmarked analysis pipeline to identify iSNV and to characterize the genetic diversity of H3N2 and H1N1 populations collected over five post-pandemic seasons from individuals enrolled in a prospective household study of influenza. We use these data to estimate the *in vivo* mutation rate and the within and between-host effective population size. We find that intrahost populations are characterized by purifying selection, a small effective population size, and limited positive selection. Contrary to what has been previously reported for human influenza transmission (Poon et al., 2016), but consistent with what has been observed in other viruses (Zwart and Elena, 2015), we identify a very tight effective transmission bottleneck that limits the transmission of rare variants.

Results

We used next generation sequencing to characterize influenza virus populations collected from individuals enrolled in the Household Influenza Vaccine Effectiveness (HIVE) study (Monto et al., 2014; Ohmit et al., 2014, 2015, 2016; Petrie et al., 2013), a community-based cohort that enrolls 213-340 households of 3 or more individuals in Southeastern Michigan each year (Table 4.1). These households are followed prospectively from October to April, with symptom-triggered collection of nasal and throat swab specimens for identification of respiratory viruses by RT-PCR (see Methods). In contrast to case-ascertained studies, which identify households based on an index case who seeks medical care, the HIVE study identifies symptomatic individuals regardless of illness severity. In the first four seasons of the study (2010-2011 through 2013-2014), respiratory specimens were collected 0-7 days after illness onset. Beginning in the 2014-2015 season, each individual provided two samples, a self-collected specimen at the time of symptom onset and a clinic-collected specimen obtained 0-7 days later. Each year, 59-69% of individuals had self-reported or confirmed receipt of that season's vaccine prior to local circulation of influenza virus.

Over five seasons and nearly 6,290 person-seasons of observation, we identified 77 cases of influenza A/H1N1pdm09 infection and 313 cases of influenza A/H3N2 infection (Table 4.1). Approximately half of the cases (n=166) were identified in the 2014-2015 season, in which there was an antigenic mismatch between the vaccine and circulating strains (Flannery et al., 2016). All other seasons were antigenically matched. Individuals within a household were considered an epidemiologically linked transmission pair if they were both positive for the same subtype of influenza virus

Table 4.1: Influenza viruses over five seasons in a household cohort

	2010-2011	2011-2012	2012-2013	2013-2014	2014-2015
Households	328	213	321	232	340
Participants	1441	943	1426	1049	1431
Vaccinated, n (%)[a]	934 (65)	554 (59)	942 (66)	722 (69)	992 (69)
IAV Positive Individuals[b]	86	23	69	48	166
H1N1	26	1	3	47	0
H3N2	58	22	66	1	166
IAV Positive Households[c]					
Two individuals	13	5	-	2	2
Three individuals	-	9	3	1	7
Four individuals	3	2	23	11	4
High Quality NGS Pairs[d]	4	1	2	6	39

[a] Self reported or confirmed receipt of vaccine prior to the specified season.
[b] RT-PCR confirmed infection.
[c] Households in which two individuals were positive within 7 days of each other. In cases of trios and quartets, the putative chains could have no pair with onset >7 days apart.
[d] Samples with $> 10^3$ genome copies per μl of transport medium, adequate amplification of all 8 genomic segments, and average sequencing coverage $> 10^3$ per nucleotide.

within 7 days of each other. Several households had 3 or 4 symptomatic cases within this one-week window, suggestive of longer chains of transmission (Table 4.1).

Within-host populations have low genetic diversity

We processed all specimens for viral load quantification and next generation sequencing. Viral load measurements (genome copies per l) were used for quality control in variant calling, which we have shown is highly sensitive to input titer (McCrone and Lauring, 2016) (Figure 4.1A). Accordingly, we report data on 249 high quality specimens from 200 individuals, which had a viral load of $> 10^3$ copies per microliter of transport media, adequate RT-PCR amplification of all eight genomic segments, and an average read coverage of $> 10^3$ across the genome (Table 4.1, Supplementary Figure 4.5).

We identified intrahost single nucleotide variants (iSNV) using our empirically validated analysis pipeline (McCrone and Lauring, 2016). Our approach relies heavily on the variant caller DeepSNV, which uses a clonal plasmid control to distinguish between true iSNV and errors introduced during sample preparation and/or sequencing (Gerstung et al., 2012). There were a number of samples that differed from the plasmid control at the consensus level. DeepSNV is unable to estimate an error rate for the control or reference base at these positions. We therefore performed an additional benchmarking experiment to identify a threshold for majority iSNV at which we could correctly infer whether or not the corresponding minor allele was also present (see Methods). We found that we could correctly identify a minor allele at a frequency of $\geq 2\%$ at such sites. We therefore report data on iSNV present at frequencies between 2 and 98%. As expected, this threshold improved the specificity of our iSNV identification and decreased our sensitivity to detect variants below 5% compared to our initial validation experiment (McCrone and Lauring, 2016), which did not employ a frequency threshold (Supplemental Table 4.3).

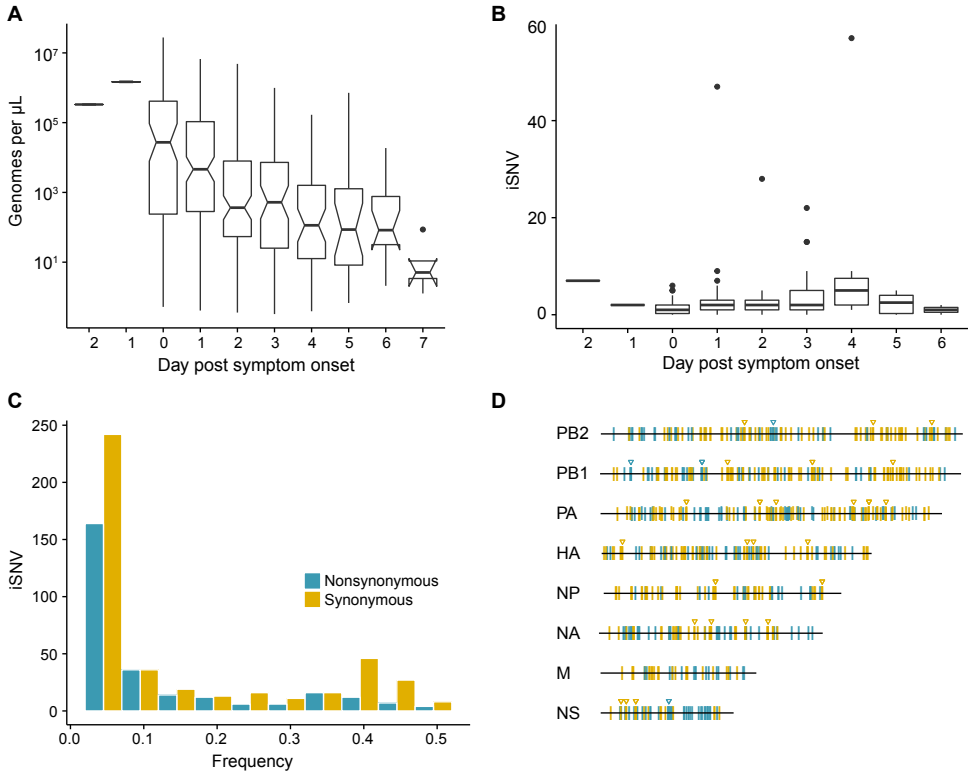


Figure 4.1: Within-host diversity of IAV populations. (A) Boxplots (median, 25th and 75th percentiles, whiskers extend to most extreme point within median 1.5 x IQR) of the number of viral genomes per microliter transport media stratified by day post symptom onset. Notches represent the approximate 95% 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) Histogram of within-host iSNV frequency in 249 high quality samples. Bin width is 0.05 beginning at 0.02. Mutations are colored nonsynonymous (blue) and synonymous (gold) (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. Triangles signify mutations that were found in more than one individual in a given season.

Consistent with our previous studies and those of others, we found that the within-host diversity of human influenza A virus (IAV) populations is low (Dinis et al., 2016; Debbink et al., 2017; Sobel Leonard et al., 2016; McCrone and Lauring, 2016). Two hundred forty-three out of the 249 samples had fewer than 10 minority(frequency below 50%) iSNV (median 2, IQR 1-3). There were 6 samples with greater than 10 minority iSNV. In 3 of these cases, the frequency of iSNVs were tightly distributed about a mean suggesting that the iSNV were linked and that the samples

represented mixed infections. Consistent with this hypothesis, putative genomic haplotypes based on these minority iSNV clustered with distinct isolates on phylogenetic trees (Supplementary Figures 4.6 and 4.7). While viral shedding was well correlated with days post symptom onset (Figure 4.1A) the number of minority iSNV identified was not affected by the day of infection, viral load, subtype, or vaccination status (Figure 4.1B and Supplementary Figure 4.8).

The vast majority of minority variants were rare (frequency 0.02-0.07), and iSNV were distributed evenly across the genome (Figure 4.1C and 4.1D). The ratio of nonsynonymous to synonymous variants was 0.64 and was never greater than 1 in any 5% bin, which suggests that within-host populations were under purifying selection. We also found that minority variants were rarely shared among multiple individuals. Ninety-five percent of minority iSNV were only found once, 4.7% were found in 2 individuals, and no minority iSNV were found in more than 3 individuals. The low level of shared diversity suggests that within-host populations were exploring distinct regions of sequence space with little evidence for parallel evolution. Of the 31 minority iSNV that were found in multiple individuals (triangles in Figure 4.1D), 4 were nonsynonymous.

Although the full range of the H3 antigenic sites have not been functionally defined, it is estimated that 131 of the 329 amino acids in HA1 lie in or near these sites (Lee and Chen, 2004). We identified 17 minority nonsynonymous iSNV in these regions (Supplemental Table 4.4). Six of these were in positions that differ among antigenically drifted viruses (Smith et al., 2004; Wiley et al., 1981), and two (193S and 189N) lie in the 'antigenic ridge', a region that has been identified as a major determinant of antigenicity (Koel et al., 2013). Three of these have been detected at the global level as consensus variants since the time of isolation (128A, 193S and

262N) with two (193S and 262N) seemingly increasing in global frequency (Neher and Bedford, 2015) (Supplementary Figure 4.9). Additionally, we identified 1 putative H1N1 antigenic variant (208K in Ca) (Caton et al., 1982; Xu et al., 2010). In total, putative antigenic variants account for 1.0-2.5% of minority iSNV identified and were found in 3.5-8.0% of infections. None of these iSNV were shared among multiple individuals.

Estimation of effective population size

Given the above observations, we hypothesized that within-host populations of IAV are under purifying selection and that variants that rise to detectable levels do so by a neutral process as opposed to positive selection. Consistent with this hypothesis, we found that nonsynonymous and synonymous iSNV exhibited similar changes in frequency over time in the 35 individuals who provided serial specimens that contained iSNV (Figure 4.2A and 4.2B). We used a maximum likelihood approach to estimate the within-host effective population size (Ne) of IAV by fitting a diffusion approximation of the Wright-Fisher model (Kimura, 1955). This model assumes that changes in iSNV frequency are due solely to random genetic drift and not selection, that iSNV are independent of one another, and that the effective population is sufficiently large to justify a continuous approximation to changes in allele frequency. The diffusion approximation of the Wright-Fisher model assigns probabilities to frequency changes given an Ne and the number of generations between sample times. In our model we fixed the within host generation time as either 6 or 12 hours (Geoghegan et al. 2016) and report the findings for the 6 hour generation time below. We then asked what population size makes the observed changes in frequency most likely (Figure 4.2B). We restricted this analysis to samples taken at least 1 day apart (n = 29), as there was very little change in iSNV frequency in populations

Table 4.2: Within-host effective population size of IAV

Model	SNV Used	Generation Time (h)	Effective Population Size (95% CI)
Diffusion approximation	All	6	35 (26-46)
	All	12	17 (13-23)
Discrete model	All	6	32 (28-41)
	Nonsynonymous	6	30 (21-40)
	Synonymous	6	37 (27-54)
	All	12	23 (23-29)
	Nonsynonymous	12	19 (19-21)
	Synonymous	12	27 (22-33)

sampled twice on the same day ($R^2 = 0.986$, Figure 4.2B and Supplementary Figure 4.10). The concordance of same day samples suggests that our sampling procedure is reproducible. Maximum likelihood optimization of this diffusion model revealed a within-host effective population size of 35 (95% CI 26-46, Table 4.2).

The diffusion approximation makes several simplifying assumptions, which if violated could influence our findings. To ensure our results were robust to the assumption of a large population, we employed a discrete interpretation of the Wright-Fisher model, which does not assume the population is large enough to justify continuous allele trajectories (Williamson and Slatkin, 1999). In this case we found an effective population size of 32 (95% CI 28-41), very close to our original estimate (Table 4.2). Both models assume complete independence of iSNV. To ensure this assumption did not affect our results, we fit the discrete model 1000 times, each time randomly subsetting our data such that only one iSNV per individual was included. This simulates a situation in which all modeled iSNV are independent and our assumption is met. Under these conditions we found a median effective population size of 33 (IQR 32-40), demonstrating negligible bias in the initial analysis due to correlation between iSNV.

As above, most iSNV in the longitudinal samples were rare ($< 10\%$) and many became extinct between samplings. To ensure that our models were capable of accurately estimating the effective population size from such data, we simulated 1000 Wright-Fisher populations with iSNV present at approximately the same starting frequencies as in our data set and an Ne of 30, 50, or 100. In these simulations, we found mean Ne of 34, 56 and 117 (Figure 4.2C). These simulations suggest that although this method may slightly overestimate the Ne, our estimates are not simply artifacts of the data structure.

To this point, we have assumed that neutral processes are responsible for the observed changes in iSNV frequency within hosts. Although this assumption seems justified at least in part by the analysis above, we tested the robustness of our models by fitting the nonsynonymous (n = 27) and synonymous iSNV (n = 36) separately. Here, we estimated an effective population size of 30 using the nonsynonymous iSNV and an effective population size of 37 using the synonymous iSNV (Table 4.2). These estimates are very close to those derived from the whole dataset and suggest that nonsynonymous and synonymous mutations are influenced by similar within-host processes. To further ensure that our results were not driven by a few outliers subject to strong selection, we ranked iSNV by their change in frequency over time and consecutively removed iSNV with the most extreme changes. We estimated the effective population size at each iteration and found that removing the top 50% most extreme iSNV increased the effective population size to 161 (Figure 4.2D). Therefore, our estimates are robust to a reasonable number of non-neutral sites. Finally, we also applied a separate Approximate Bayesian Computational (ABC) method, which uses a non-biased moment estimator in conjunction with ABC to estimate the effective population size of a population as well as selection coefficients for the iSNV present

(Foll et al., 2014). This distinct approach relaxes the previous assumption regarding neutrality. We applied this analysis to the 16 longitudinal pairs that were sampled 1 day apart and estimated an effective population of 69. We were unable to reject neutrality for just 7 of the 35 iSNV in this data set (Figure 4.2E). These seven mutations consisted of 3 nonsynonymous and 4 synonymous mutations and were split between two individuals. None were putative antigenic variants.

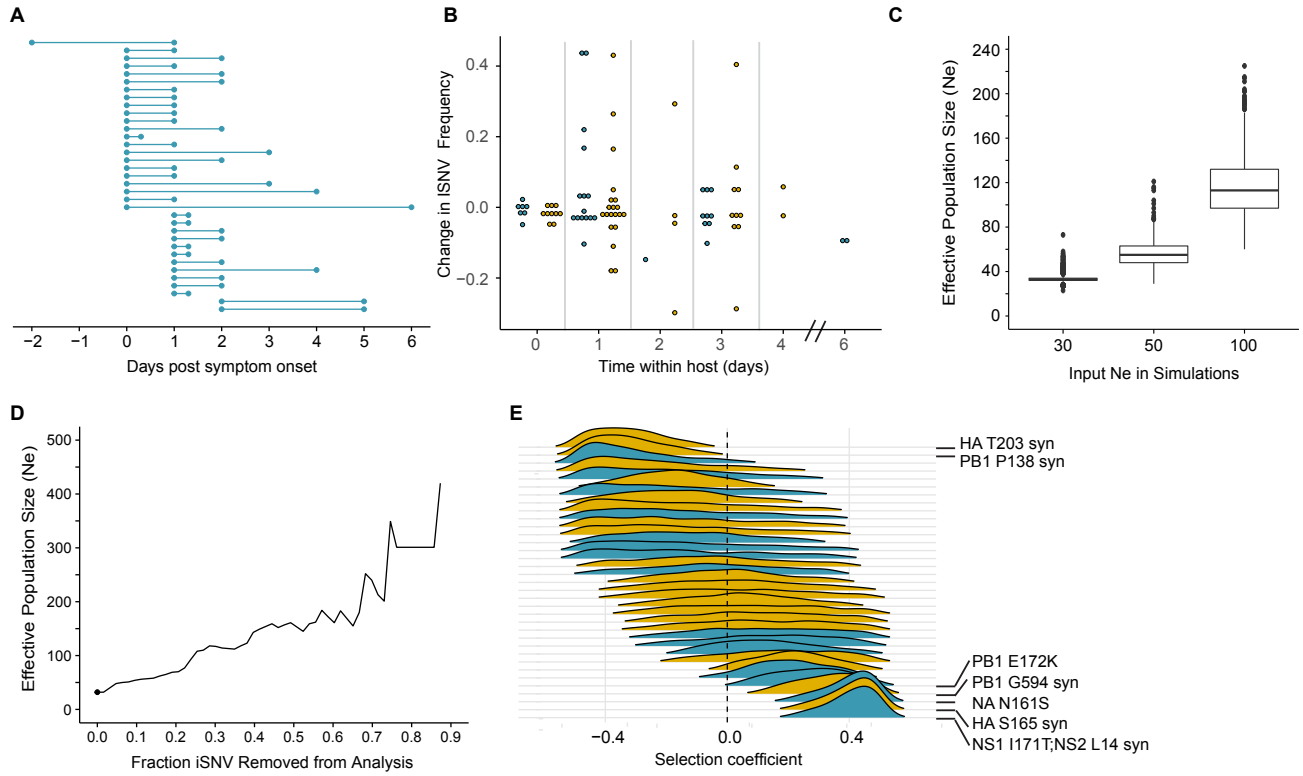


Figure 4.2: Within-host dynamics of IAV. (A) Timing of sample collection for 35 paired longitudinal samples relative to day of symptom onset. Of the 49 total, 35 pairs had minor iSNV present in the first sample. (B) The change in frequency over time for minority nonsynonymous (blue) and synonymous (gold) iSNV identified for the paired samples in (A). (C) The distribution of effective population sizes estimated from 1,000 simulated populations. Simulations were run on populations with characteristics similar to the actual patient-derived populations and with the specified effective population size (x-axis). (D) The effect of iteratively removing iSNV with the most extreme change in frequency (fraction of iSNV removed, x-axis) on the estimated effective population size. The point represents the estimate when all iSNV are included. (E) The posterior distributions of selection coefficients estimated for the 35 iSNV present in isolates sampled one day apart. Distributions are colored according to class relative to the sample consensus sequence, nonsynonymous (blue) synonymous (gold). Variants for which the 95% highest posterior density intervals exclude 0.0 are noted in the margin.

Identification of forty-three transmission pairs

We analyzed virus populations from 85 households with concurrent infections to quantify the level of shared viral diversity and to estimate the size of the IAV transmission bottleneck (Table 4.1). Because epidemiological linkage does not guarantee that concurrent cases constitute a transmission pair (Petrie et al., 2017), we used a stringent rubric to eliminate individuals in a household with co-incident community acquisition of distinct viruses. We considered all individuals in a household with symptom onset within a 7-day window to be epidemiologically linked. The donor in each putative pair was defined as the individual with the earlier onset of symptoms. We discarded a transmission event if there were multiple possible donors with the same day of symptom onset. Donor and recipients were not allowed to have symptom onset on the same day, unless the individuals were both index cases for the household. In these 6 instances, we analyzed the data for both possible donor-recipient directionalities. Based on these criteria, our cohort had 124 putative household transmission events over 5 seasons (Table 4.1). Of these, 52 pairs had samples of sufficient quality for reliable identification of iSNV from both individuals.

We next used sequence data to determine which of these 52 epidemiologically linked pairs represented true household transmission events as opposed to coincident community-acquired infections. We measured the genetic distance between influenza populations from each household pair by L1-norm and compared these distances to those of randomly assigned community pairs within each season (Figure 4.3A, see also trees in Supplementary Figures 2 and 3). While the L1-norm of a pair captures differences between the populations at all levels, in our cohort, it was largely driven by differences at the consensus level. We only considered individuals to be a true transmission pair if they had a genetic distance below the 5th percentile of

the community distribution of randomly assigned pairs (Figure 4.3A). Forty-seven household transmission events met this criterion (Figure 4.3B). Among these 47 sequence-validated transmission pairs, 3 had no iSNV in the donor and 1 additional donor appeared to have a mixed infection. These four transmission events were removed from our bottleneck analysis as donors without iSNV are uninformative and mixed infections violate model assumptions of site independence (see Methods). We estimated the transmission bottleneck in the remaining 43 high-quality pairs (37 H3N2, 6 H1N1, Figure 4.3B).

A transmission bottleneck restricts the amount of genetic diversity that is shared by both members of a pair. We found that few minority iSNV were polymorphic in both the donor and recipient populations (Figure 4.3C). Minority iSNV in the donor were either absent or fixed in the recipient (top and bottom of plot). The lack of shared polymorphic sites (which would lie in the middle of the plot in Figure 4.3C) suggests a stringent effective bottleneck in which only one allele is passed from donor to recipient.

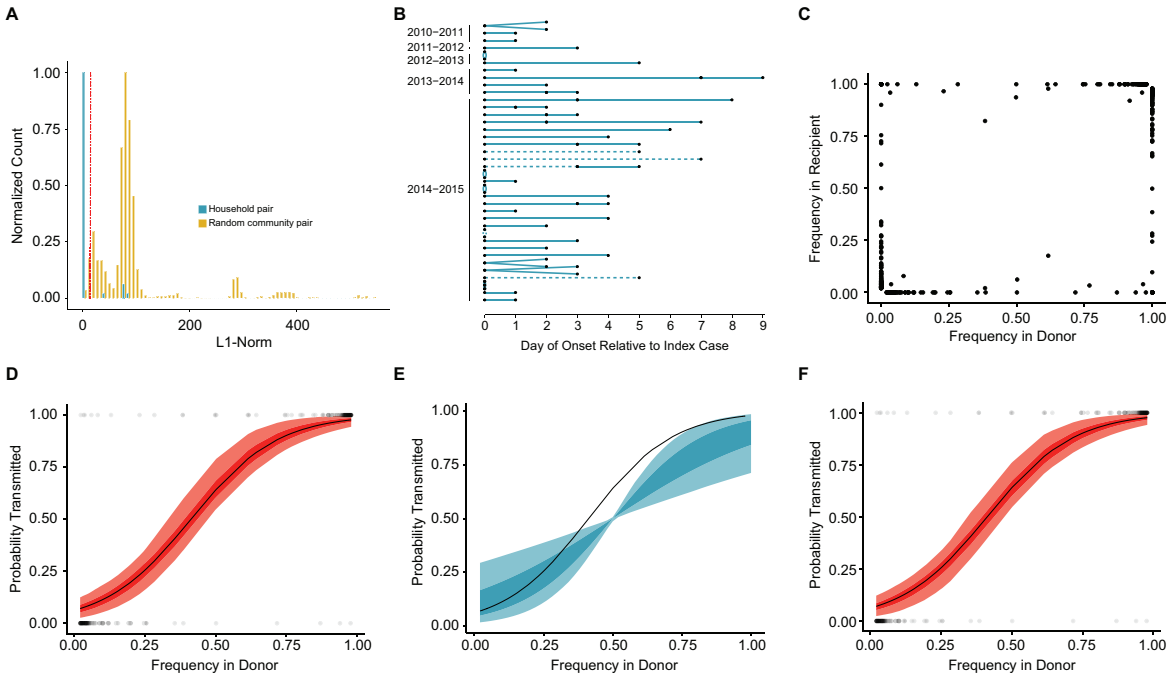


Figure 4.3: Between-host dynamics of IAV. (A) The distribution of pairwise L1-norm distances for household (blue) and randomly-assigned community (gold) pairs. The bar heights are normalized to the height of the highest bar for each given subset (47 for household, 1,592 for community). The red line represents the 5th percentile of the community distribution. (B) Timing of symptom onset for 52 epidemiologically linked transmission pairs. Day of symptom onset for both donor and recipient individuals is 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 black line indicates the probability of transmission for a given iSNV frequency as determined by logistic regression. Similar fits were calculated for 1,000 simulations with a mean bottleneck size of 1.66. Fifty percent of simulated outcomes lie in the darkly shaded region and 95% lie in the lightly shaded regions. (E) The outcome from 1,000 simulated 'transmission' events with randomly assigned pairings. The black line represents the observed data, as in (D) the shaded regions represent the middle 50% and 95% of simulated outcomes. The results from the simulated logit models were smoothed by plotting the predicted probability of transmission at 0.02 intervals. (F) The beta-binomial model fit. Similar to (D) except the simulated outcomes are the based on a beta-binomial model using a mean bottleneck of 1.73.

Estimation of the transmission bottleneck

We applied a simple presence-absence model to quantify the effective transmission bottleneck in our cohort. The presence-absence model considers only whether or not a donor allele is present or absent in the recipient sample. Under this model, transmission is a neutral, random sampling process, and the probability of transmission is simply the probability that the iSNV will be included at least once in the sample given its frequency in the donor and the sample size, or bottleneck. We estimated a distinct bottleneck for each transmission pair and assumed these bottlenecks followed a zero-truncated Poisson distribution. This model also assumes that the sensitivity for detection of transmitted iSNVs is perfect and that each genomic site is independent of all others. We then used maximum likelihood optimization to determine the distribution of bottleneck sizes that best fit the data. We found a zero-truncated Poisson distribution with a mean of 1.66 (lambda = 1.12; 0.51-1.99, 95% CI) best described the data. This distribution indicates that the majority of bottlenecks are 1, and that very few are greater than 5 (probability 0.2%). There were no apparent differences between H3N2 and H1N1 pairs. The model fit was evaluated by simulating each transmission event 1,000 times. The presence or absence of each iSNV in the recipient was noted and the probability of transmission given donor frequency determined. The range of simulated outcomes matched the data well, which suggests that transmission is a selectively neutral event characterized by a stringent bottleneck (Figure 4.3D).

The majority of transmitted iSNV were fixed in the recipients. Although this trend matches the expectation given a small bottleneck, these data could also be consistent with a model in which the probability of transmission is determined by the frequency at which iSNV are found at the community level. To ensure our bot-

tleneck estimates were an outcome of neutral transmission and not an artifact of the larger community population structure or selection for the community consensus, we created a null model by randomly assigning community 'recipient-donor' pairings. Each community 'recipient' was drawn from the pool of individuals that were infected after the 'donor' but in the same season and with the same subtype as the donor. We then identified whether or not each donor iSNV was found in the community recipient and determined the relationship between 'donor' frequency and probability of 'transmission' for 1,000 such simulations. Given the low level of diversity in our cohort, we predicted that rare iSNV would be unlikely to be found in a random sample, while the major alleles should be fixed in most random pairs. This trend is clearly demonstrated in Figure 4.3E. It is also clear that this null model fit the data much more poorly than the presence/absence model, suggesting that the observed data in our bona fide transmission pairs were not a product of community metapopulation structure, but rather an outcome of neutral sampling events.

Because our bottleneck estimates were much lower than what has previously been reported for human influenza (Poon et al., 2016), we investigated the impact that our simplifying assumptions could have on our results. In particular, the presence-absence model assumes perfect detection of variants in donor and recipient, and it can therefore underestimate the size of a bottleneck in the setting of donor-derived variants that are transmitted but not detected in the recipient. These 'false negative' variants can occur when the frequency of an iSNV drifts below the level of detection (e.g. 2% frequency) or when the sensitivity of sequencing is less than perfect for variants at that threshold (e.g. 15% sensitivity for variants at a frequency 2-5%). To determine the impact of sequencing sensitivity and specificity on our bottleneck estimates, we re-called variants using our original pipeline without the 2% frequency

cut-off. As shown in Supplemental Table 4.3, this increases the sensitivity of iSNV detection in the 1-5% frequency range, and also the number of false positive variant calls (McCrone and Lauring, 2016). This analysis only slightly increased the average transmission bottleneck to 2.10 (lambda = 1.67; 0.91-2.71, 95% CI), and indicates that our results are not biased by the added stringency used in the initial analysis (Supplementary Figures 7A and 7B).

To further investigate the impact of sequencing accuracy on our estimates, we inferred minor variants in our current pipeline (see above and methods) without a frequency cutoff. Ultimately, this reduced variant calling to a count method at a number of positions and greatly increased the number of shared minority iSNV in our samples (Supplementary Figure 4.11C). Many of these presumed false positive variant calls were at similar frequencies (0.1-2%) in donor and recipient. As such, the 'apparent transmission' of rare variants drives an inflated estimate of the transmission bottleneck (118, see Supplementary Figure 4.11D). Simulation showed that this inflated bottleneck no longer fit the trend in the data, likely because the model is now forced to accommodate shared iSNV that are biased toward sequencing error as opposed to the actual transmission process.

We also estimated bottleneck size using a beta binomial model, which Leonard et al. have used to account for the stochastic loss of transmitted variants. This model allows for a limited amount of time-independent genetic drift within the recipient (Sobel Leonard et al., 2017b), and we modified it to also account for our benchmarked sensitivity for rare variants (Supplemental Table 4.3, Current Pipeline). For all donor-derived iSNV that were absent in the recipient, we estimated the likelihood that these variants were transmitted but either drifted below our level of detection or drifted below 10% and were missed by our variant identification. Despite the relaxed

assumptions provided by this modified beta binomial model, maximum likelihood estimation only marginally increased the average bottleneck size (mean 1.73: lambda 1.22; 0.57-2.17, 95%CI) relative to the simpler presence-absence model. We simulated transmission and subsequent random drift using the beta binomial model and the estimated bottleneck distribution as above (Figure 4.3F). Although the model matched the data well, the fit was not better than that of the presence-absence model (AIC 83.0 for beta-binomial compared to 76.7 for the presence-absence model).

The mutation rate of influenza A virus within human hosts

The stringent influenza transmission bottleneck suggests that most infections are founded by one lineage and develop under essentially clonal processes. The diffusion approximation to the Wright-Fisher model (see above and Figure 4.2) can be used to predict the rate at which homogenous populations diversify from a clonal ancestor as a function of mutation rate and effective population size (Rouzine et al., 2001). Maximum likelihood optimization of this model suggested an in vivo neutral mutation rate of $4\text{x}10^{-6}$ mutations per nucleotide per replication cycle and a within-host effective population size of 36 (given a generation time of 6 hours, Figure 4.4). These estimates are consistent with those above (Table 4.2). As we have recently estimated that 13% of mutations in influenza A virus are neutral (Visher et al., 2016), we estimated that the true in vivo mutation rate would be approximately 8 fold higher than our neutral rate on the order of $3 - 4 \times 10^{-5}$. This in vivo mutation rate is close to our recently published estimate of influenza A mutation rates in epithelial cells by fluctuation test (Pauly et al., 2017) and within the range of other estimates for IAV (Sanjuán et al., 2010).

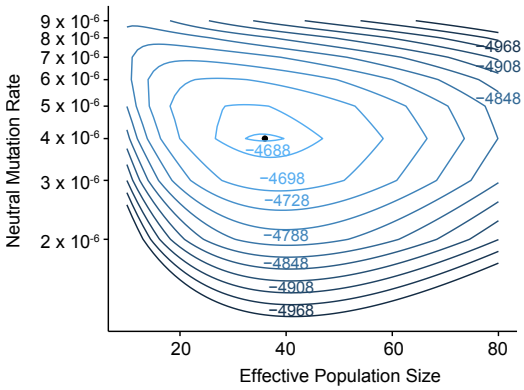


Figure 4.4: Combined estimates of within-host mutation rate and effective population size. Contour plot shows the log likelihood surface for estimates of the effective population size and neutral mutation rate. The point represents the peak ($\mu = 4 \times 10^{-6}$, $N_e = 36$, log likelihood = -4,687). Log likelihoods for each contour are indicated.

Discussion

We find that seasonal influenza A viruses replicate within and spread among human hosts with very small effective population sizes. Because we used viruses collected over five influenza seasons from individuals enrolled in a prospective household cohort, these dynamics are likely to be broadly representative of many seasonal influenza infections in their natural transmission context. Our results are further strengthened by the use of a validated sequence analysis pipeline and models that are robust to the underlying assumptions. The small effective size of intrahost populations and the tight effective transmission bottleneck suggest that stochastic processes, such as genetic drift, dominate influenza virus evolution at the level of individual hosts. This stands in contrast to prominent role of positive selection in the global evolution of seasonal influenza.

While influenza virus populations are subject to continuous natural selection, selection is an inefficient driver of evolution in small populations (Rouzine et al., 2001). Despite a large viral copy number, our findings demonstrate that intrahost populations of influenza behave like much smaller populations. We therefore expect

stochastic fluctuations to be the major force driving the fixation of novel variants within human hosts. This finding contradicts previous studies, which have found signatures of adaptive evolution in infected hosts (Gubareva et al., 2001; Rogers et al., 2015; Ghedin et al., 2010; Sobel Leonard et al., 2016). However, these studies rely on data from infections in which selective pressures are likely to be particularly strong (e.g. due to drug treatment or infection with a poorly adapted virus), or in which the virus has been allowed to propagate for extended periods of time (Xue et al., 2017). Under these conditions, one can identify the action of positive selection on within-host populations. We suggest that these are important and informative exceptions to the drift regime defined in our cohort.

We used both a simple presence-absence model and a more complex beta binomial model to estimate an extremely tight transmission bottleneck. The estimation of a small bottleneck size is driven by low within-host diversity and very few minority iSNV shared among individuals in a transmission pair. While our methods for variant calling may be more conservative than those used in similar studies, we found that relaxing our variant calling criteria led to the inclusion of false positive variants that inflated our estimates. Furthermore, the beta binomial model accounts for false negative iSNV (i.e. variants that are transmitted but not detected in the donor), which can lead to underestimated transmission bottlenecks (Sobel Leonard et al., 2017b). Our formulation of this model incorporates empirically determined sensitivity and specificity metrics to account for both false negative iSNV and false positive iSNV (McCrone and Lauring, 2016). Finally, if rare, undetected, iSNV were shared between linked individuals, we would expect to see transmission of more common iSNV (frequency 5-10%), which we can detect with high sensitivity. In our data, the

transmission probability iSNVs > 5% frequency in the donor were also well predicted by small bottleneck size (Figure 4.3D).

Although the size of our transmission bottleneck is consistent with estimates obtained for other viruses and in experimental animal models of influenza (Zwart and Elena, 2015; Varble et al., 2014), it differs substantially from the only other study of bottlenecks in natural human infection (Poon et al., 2016; Sobel Leonard et al., 2017b). While there are significant differences in the design and demographics of the cohorts, the influenza seasons under study, and sequencing methodology (Kugelman et al., 2017a), the bottleneck size estimates are fundamentally driven by the amount of viral diversity shared among individuals in a household. Importantly, we used both epidemiologic linkage and the genetic relatedness of viruses in households to define transmission pairs and to exclude confounding from the observed background diversity in the community. This stringency is not expected to bias our bottleneck estimate toward small founding populations. We removed cases in which the genetic distance between the donor and host is large, and expect this would only bias our analysis towards loose transmission bottlenecks.

In our analysis, we find that household transmission pairs and randomly assigned community pairs had distinct patterns of shared consensus and minority variant diversity. The comparison to random community pairs is important, as an unexplained aspect of the work of Poon et al. is that rare iSNV were frequently shared by randomly selected individuals, and more common ones were not (Poon et al., 2016). Reasons for these shared minority variants could include sequencing artifacts, parallel within-host evolution, and a stable and diverse population that is maintained during transmission. The within-host data described in my study suggests that

these shared variants could sequencing artifacts; however, a more mechanistic study influenza transmission is needed to definitely rule out other possibilities.

Our estimates of IAV population dynamics are consistent across three separate models and partitions of the data. We jointly estimated the *in vivo* mutation rate and effective population size based on the frequency distribution of minor alleles observed in the entire cohort. This model assumed a small transmission bottleneck, produced a mutation rate that is consistent with previous estimates, and independently reproduced the within-host population size estimate. Given the concordance among these distinct approaches, it is unlikely that our findings are biased by hidden assumptions or model limitations.

Accurately modeling and predicting influenza virus evolution requires a thorough understanding of the virus' population structure. Some models have assumed a large intrahost population and a relatively loose transmission bottleneck (Geoghegan et al., 2016; Russell et al., 2012; Peck et al., 2015). Here, adaptive iSNV can rapidly rise in frequency and low frequency variants can have a high probability of transmission. In such a model, it would be possible for the highly pathogenic H5N1 virus to develop the requisite 4-5 mutations to become transmissible through aerosols during a single acute infection of a human host (Herfst et al., 2012; Russell et al., 2012). Although the dynamics of emergent avian influenza and human adapted seasonal viruses likely differ (Petrova and Russell, 2017), our work suggests that fixation of multiple mutations over the course of a single acute infection is unlikely.

While it may seem counterintuitive that influenza evolution is dominated by drift on local scales and positive selection on global scales, these models are certainly not in conflict. Within individuals we have shown that the effective population is quite small, which suggests that selection is inefficient. Indeed, we have deeply se-

quenced 332 intrahost populations from 283 individuals collected over more than 11,000 person-seasons of observation and only identified a handful of minority antigenic variants with little evidence for positive selection (this work and (Debbink et al., 2017)). However, with several million infected individuals each year, even inefficient processes and rare events are likely to happen at a reasonable frequency on a global scale.

Methods

Description of the cohort

The HIVE cohort (Monto et al., 2014; Ohmit et al., 2014, 2015, 2016; Petrie et al., 2013), established at the UM School of Public Health in 2010, enrolled and followed households of at least 3 individuals with at least two children <18 years of age; households were then followed prospectively throughout the year for ascertainment of acute respiratory illnesses. Study participants were queried weekly about the onset of illnesses meeting our standard case definition (two or more of: cough, fever/feverishness, nasal congestion, sore throat, body aches, chills, headache if >3 yrs old; cough, fever/feverishness, nasal congestion/runny nose, trouble breathing, fussiness/irritability, decreased appetite, fatigue in <3 yrs old), and the symptomatic participants then attended a study visit at the research clinic on site at UM School of Public Health for sample collection. For the 2010-2011 through 2013-2014 seasons, a combined nasal and throat swab (or nasal swab only in children < 3 years of age) was collected at the onsite research clinic by the study team. Beginning with the 2014-2015 seasons, respiratory samples were collected at two time points in each participant meeting the case definition; the first collection was a self- or parent-collected nasal swab collected at illness onset. Subsequently, a combined nasal and throat swab (or nasal swab only in children < 3 years of age) was collected at the onsite

research clinic by the study team. Families with very young children (< 3 years of age) were followed using home visits by a trained medical assistant.

Active illness surveillance and sample collection for cases were conducted October through May and fully captured the influenza season in Southeast Michigan in each of the study years. Data on participant, family and household characteristics, and on high-risk conditions were additionally collected by annual interview and review of each participant's electronic medical record. In the current cohort, serum specimens were also collected twice yearly during fall (November-December) and spring (May-June) for serologic testing for antibodies against influenza.

This study was approved by the Institutional Review Board of the University of Michigan Medical School, and all human subjects provided informed consent.

Identification of influenza virus

Respiratory specimens were processed daily to determine laboratory-confirmed influenza infection. Viral RNA was extracted (Qiagen QIAamp Viral RNA Mini Kit) and tested by RT-PCR for universal detection of influenza A and B. Samples with positive results by the universal assay were then subtyped to determine A(H3N2), A(H1N1), A(pH1N1) subtypes and B(Yamagata) and B(Victoria) lineages. We used primers, probes and amplification parameters developed by the Centers for Disease Control and Prevention Influenza Division for use on the ABI 7500 Fast Real-Time PCR System platform. An RNAseP detection step was run for each specimen to confirm specimen quality and successful RNA extraction.

Quantification of viral load

Quantitative reverse transcription polymerase chain reaction (RT-qPCR) was performed on 5μl RNA from each sample using CDC RT-PCR primers InfA Forward,

InfA Reverse, and InfA probe, which bind to a portion of the influenza M gene (CDC protocol, 28 April 2009). Each reaction contained 5.4μl nuclease-free water, 0.5μl each primer/probe,0.5μl SuperScript III RT/Platinum Taq mix (Invitrogen 111732) 12.5μl PCR Master Mix, 0.1μl ROX, 5μl RNA. The PCR master mix was thawed and stored at 4C, 24 hours before reaction set-up. A standard curve relating copy number to Ct value was generated based on 10-fold dilutions of a control plasmid run in duplicate.

Illumina library preparation and sequencing

We amplified cDNA corresponding to all 8 genomic segments from 5μl of viral RNA using the SuperScript III One-Step RT-PCR Platinum Taq HiFi Kit (Invitrogen 12574). Reactions consisted of 0.5μl Superscript III Platinum Taq Mix, 12.5μl 2x reaction buffer, 6μl DEPC water, and 0.2μl of 10μM Uni12/Inf1, 0.3μl of 10μM Uni12/Inf3, and 0.5μl of 10μM Uni13/Inf1 universal influenza A primers (Zhou et al., 2009). The thermocycler protocol was: 42°C for 60 min then 94°C for 2 min then 5 cycles of 94°C for 30 sec, 44°C for 30 sec, 68°C for 3 min, then 28 cycles of 94°C for 30 sec, 57°C for 30 sec, 68°C for 3 min. Amplification of all 8 segments was confirmed by gel electrophoresis, and 750ng of each cDNA mixture were sheared to an average size of 300 to 400bp using a Covaris S220 focused ultrasonicator. Sequencing libraries were prepared using the NEBNext Ultra DNA library prep kit (NEB E7370L), Agencourt AMPure XP beads (Beckman Coulter A63881), and NEBNext multiplex oligonucleotides for Illumina (NEB E7600S). The final concentration of each barcoded library was determined by Quanti PicoGreen dsDNA quantification (ThermoFisher Scientific), and equal nanomolar concentrations were pooled. Residual primer dimers were removed by gel isolation of a 300-500bp band, which was purified using a GeneJet Gel Extraction Kit (ThermoFisher Scientific). Purified li-

brary pools were sequenced on an Illumina HiSeq 2500 with 2x125 nucleotide paired end reads. All raw sequence data have been deposited at the NCBI sequence read archive (BioProject submission ID: SUB2951236). PCR amplicons derived from an equimolar mixture of eight clonal plasmids, each containing a genomic segment of the circulating strain were processed in similar fashion and sequenced on the same HiSeq flow cell as the appropriate patient derived samples. These clonally derived samples served as internal controls to improve the accuracy of variant identification and control for batch effects that confound sequencing experiments.

Identification of iSNV

Intrahost single nucleotide variants were identified in samples that had greater than 10^3 genomes/μl and an average coverage >1000x across the genome. Variants were identified using DeepSNV and scripts available at `https://github.com/lauringlab/variant_pipeline` as described previously i(McCrone and Lauring, 2016) with a few minor and necessary modifications. Briefly, reads were aligned to the reference sequence (H3N2 2010-2011 & 2011-2012 : GenBank CY121496-503, H3N2 2012-2013:GenBank KJ942680-8, H3N2 2014-2015 : Genbank CY207731-8, H1N1 GenBank : CY121680-8) using Bowtie2 (35). Duplicate reads were then marked and removed using Picard (`http://broadinstitute.github.io/picard/`). We identified putative iSNV using DeepSNV. Bases with phred <30 were masked. Minority iSNV (frequency <50%) were then filtered for quality using our empirically determined quality thresholds (p-value <0.01 DeepSNV, average mapping quality >30, average Phred >35, average read position between 31 and 94). To control for PCR errors in samples with lower input titers, all isolates with titers between 10^3 and 10^5 genomes/μl were processed and sequenced in duplicate. Only iSNV that were found in both replicates were included in down stream analysis. The frequency of the vari-

ant in the replicate with higher coverage at the iSNV location was assigned as the frequency of the iSNV. Finally, any SNV with a frequency below 2% was discarded.

Given the diversity of the circulating strain in a given season, there were a number of cases in which isolates contained mutations that were essentially fixed (>95%) relative to the plasmid control. Often in these cases, the minor allele in the sample matched the major allele in the plasmid control. We were, therefore, unable to use DeepSNV in estimating the base specific error rate at this site for these minor alleles and required an alternative means of eliminating true and false minority iSNV. To this end we applied stringent quality thresholds to these putative iSNV and implemented a 2% frequency threshold. In order to ensure we were not introducing a large number of false positive iSNV into our analysis, we performed the following experiment. Perth (H3N2) samples were sequenced on the same flow cell as both the Perth and Victoria (H3N2) plasmid controls. Minority iSNV were identified using both plasmid controls. This allowed us to identify rare iSNV at positions in which the plasmid controls differed both with and without the error rates provided by DeepSNV. We found that at a frequency threshold of 2% the methods were nearly identical (NPV of 1, and PPV of 0.94 compared to DeepSNV).

dN/dS calculation

We calculated the genome-wide dN/dS as in (Miyata and Yasunaga, 1980) without correction for iSNV frequency in the population. This analysis was run only on canonical open reading frames (excluding PB1-F2 and PA-X).

Overview of models for effective population size

We estimated the effective population size using two separate interpretations of a Wright-Fisher population (Ewens, 2004). At its base, the Wright-Fisher model

describes the expected changes in allele frequency of an ideal population, which is characterized by non-overlapping generations, no migration, no novel mutation, and no population structure. We then asked what size effective population would make the changes in frequency observed in our dataset most likely. We calculated these values using two applications of the Wright-Fisher model (i) a diffusion approximation (Kimura, 1955) and (ii) a maximum likelihood approach based on the discrete interpretation (Williamson and Slatkin, 1999).

For these estimates we restricted our analysis to longitudinal samples from a single individual that were separated by at least 1 day and only used sites that were polymorphic in the initial sample (29 of the 49 total serial sample pairs). We modeled only the iSNV that were the minor allele at the first time point, and we assumed a within-host generation time of either 6 or 12 hours as proposed by Geoghegan et al. (Geoghegan et al., 2016).

Diffusion approximation

The diffusion approximation was first solved by Kimura in 1955 (Kimura 1955). This approximation to the discrete Wright-Fisher model has enjoyed widespread use in population genetics as it allows one to treat the random time dependent probability distribution of final allele frequencies as a continuous function (e.g.(Zanini et al., 2017; Kimura and Ohta, 1969; Kimura, 1971; Myers et al., 2008)). Here, we also included the limitations in our sensitivity to detect rare iSNV by integrating over regions of this probability density that were either below our limit of detection or within ranges where we expect less than perfect sensitivity as follows.

Let $P(p_0, p_t, t|N_e)$ be the time dependent probability of a variant drifting from an initial frequency of p_0 at time 0 p_t at time t generations given an effective population size of N_e where $0 < p_t < 1$.

The time dependent derivative of this probability has been defined using the Kolmogorov forward equation (Kimura, 1955) and for haploid populations is:

$$(4.1)\quad P(p_0, p_t, t|N_e) = \sum_{i=1}^{\infty} p_0 q_0 i(i+1)(2i+1)F(1-i, i+2, 2, p)\times F(1-i, i+2, 2, x)e^{-[i(i+1)/2N_e]t}$$

Where $q = 1 - p$ and F is the hypergeometric function. We approximated the infinite sum by summing over the first 50 terms. When we added an additional 50 terms (100 in total) we found no appreciable change in the final log likelihoods.

We denote the frequency of allele that is not observed at the second time point as $p_t \approx 0$ and the probability of such an event as $P(p_0, p_t \approx 0, t|N_e)$. This probability is given in equation 2 as the sum of the probability that the variant is truly lost by generation t (i.e. the other allele is fixed $P(q_0, 1, t|N_e)$)), the probability that it is present but below the limit of detection (i.e. $P(p, p_t \approx 0, t|0 < p_t < 0.02, N_e)$) and the probability the variant is not detected due to low sensitivity for rare variant detection (i.e. $P(p_0, p_t \approx 0, t|0.02 < p_t < 0.1, N_e)$). The probability of not observing an allele at the second time is then

$$(4.2)\quad P(p_0, p_t \approx 0, t|N_e) = P(q_0, 1, t|N_e) + P(p, p_t \approx 0, t|0 < p_t < 0.02, N_e) + P(p_0, p_t \approx 0, t|0.02 < p_t < 0.1, N_e)$$

The first term in equation 4.2 is given by Kimura 1955 as

$$(4.3)\quad P(q_0, 1, t|N_e) = q_0 + \sum_{i=1}^{\infty}(2i+1)p_0 q_0(-1)^i F(1-i, i+2, 2, q_0)e^{-[i(i+1)/2N]t}$$

Where q is defined as above. (Note : this is simply the probability of fixation for a variant at initial frequency q). As in equation 4.1 the infinite was approximated with a partial sum of 50 terms.

The probability of the allele drifting below our limit of detection can be found by integrating equation 4.1 between 0 and our limit of detection, 0.02. This was done numerically using the python package scipy.

$$P(p, p_t \approx 0, t | 0 < p_t < 0.02, N_e) = \int_0^{0.02} P(p_0, p_t, t, N) dp_t \tag{4.4}$$

Finally, the probability of an iSNV being present at the second time point, but escaping detection, is given by the integral of equation 4.1 between our benchmarked frequencies (0.02,0.05) times the false negative rate for that range. Here, we assumed the entire range had the same sensitivity as the benchmarked frequency at the lower bound and rounded recipient titers down to the nearest $\log_{10}$ titer (e.g.10^3,10^4, 10^5). We also assumed perfect sensitivity above 10%.

$$P(p_0, p_t \approx 0, t | 0.02 < p_t < 0.1, N_e)) = \sum_{f_i}^{[0.02,0.05,0.10)} \left(\text{FNR}|\text{Titer}_r, f_i\right) \int_{f_i}^{f_{i+1}} P(p_0, p_t, t | N_e) dp_t \tag{4.5}$$

Where $(\text{FNR}|\text{Titer}_r, f_i)$ is the false negative rate given the frequency and the sample titer (See Supplemental Table 4.3) and $P(p_0, p_t, t|N_e)$ is defined in equation 4.1 .

The log likelihood of an effective population size is the sum of the log of $P(p_0, p_t, t|N_e)$ for each minor allele in the data set, where either the position is polymorphic at time t (i.e. equation 4.1) or the allele is not observed at time t (i.e. equation 4.2).

Discrete Wright-Fisher estimation of N_e

The diffusion approximation treats changes in frequency as a continuous process because it assumes sufficiently large N_e. That assumption can be relaxed, and the effective population size can be determined, by applying a maximum likelihood method developed by Williamsom and Slatkin 1999 (Williamson and Slatkin, 1999). In this model, the true allele frequencies move between discrete states (i.e. the frequency must be of the form i/N_e where i is a whole number in the range $[0, N_e]$. In the original application, allele counts were used, and sampling error was added to the model as a binomial distribution with n determined by the sample size. Here, we use the frequencies available from next generation sequencing and estimate sampling error as a normal distribution with mean equal to the observed frequency and a standard deviation equal to that observed in our benchmarking study for the 10^4 genomes/μl samples (σ=0.014) (McCrone and Lauring, 2016).

In this model, the probability of observing an allele frequency shift from $\hat{p}_0$ to $\hat{p}_t$ in t generations provided an effective population of N_e is the probability of observing $\hat{p}_0$ given some initial state p_0 and the probability of the population having that state, times the probability of observing $\hat{p}_t$ given some final state p_t and the probability of moving from the initial to the final state summed across all possible states.

$$P(p_0, p_t|N_e) = \sum_{p_0,p_t} P(\hat{p}_0|p_0)P(p_0|N_e)P(\hat{p}_t|p_t)P(p_t|p_0|N_e) \tag{4.6}$$

Where $\hat{p}_x$ are the observed probabilities and p_x are the real ones (of the form i/N_e discussed above). The likelihood of observing a given frequency $\hat{p}_x$ given a defined state p_x is given by the likelihood of drawing $\hat{p_x}$ from a normal distribution with mean p_x and standard deviation 0.014.

$$P(\hat{p}_x|p_x) = N(p_x, 0.014) \tag{4.7}$$

As in Williamson and Slatkin 1999, we assume a uniform prior on the initial state. Because we know that our specificity is near perfect (above 2%, Supplemental Table 4.3) and we restrict our analysis to only polymorphic sites, the probability of any initial state is given by

$$P(p_0|N_e) = \frac{1}{N_e - 1} \tag{4.8}$$

and finally the probability of moving from one state to another in t generations is given by

$$P(p_t, p_0|N_e) = v_0 M^t v_t \tag{4.9}$$

Where M is a square transmission matrix with $N_e + 1$ rows and columns. Where i$m_{i,j}$ is the probability of going from the ith configuration to the jth or the probability of drawing $j - 1$ out of binomial distribution with mean $(i - 1)/N_e$ and a sample size N_e. v_0 is a row vector of initial frequencies p_0 with 100% chance of initial state p_0, and v_t is column vector of the frequencies at time point t with 100% chance of the final state. In other words v_0 is a row vector of $N_e + 1$ states with 0 everywhere except in the ith position where $(i-1)/N_e = p_0$, and v_t is a column vector of $N_e + 1$ states with 0 everywhere except the jth position where $(j - 1)/N_e = p_t$.

Using the scalar and cumulative properties of matrix multiplication equation 6 reduces to

$$(4.10)\quad P(\hat{p_0}, \hat{p_t}|N_e) =$$

$$[0, P(\hat{p_0}|p_{0_2})P(p_{0_2}|N_e), ..., P(\hat{p_0}|p_{0_{N_e-1}})P(p_{0_{N_e-1}}|N_e), 0]M^t \begin{bmatrix} P(\hat{p_t}|p_{t_1}) \\ \vdots \\ P(\hat{p_t}|p_{t_{N_e}}) \end{bmatrix}$$

The first and last entries in v_0 are 0 because we assume all measured sites represent polymorphisms at the first time of sampling.

As above, the log likelihood of a given population size is then simply the sum of the log of $P(\hat{p_0}, \hat{p_t}, t|N_e)$ for each minor allele in the data set.

Simulations

To simulate within-host evolution we set N_e in equation 4.10 to either 30, 50 or 100. For each minor allele we used the closest available non-zero state given the effective population size as the starting state. We then calculated the probability of moving to any other state and selected a final state from this distribution. We then drew a final measured frequency from the normal distribution accounting for measurement errors.

ABC model

We estimated both the effective population size and selection coefficients using the approximate Bayesian computation (ABC) described in (Foll et al., 2014) with the WFACB_v1.1 software provided in (Foll et al., 2015). In its current implementation, this analysis requires the same time points for each sample, and we restricted this analysis to longitudinal samples taken 1 day apart. This subset constitutes 16 of the 29 modeled longitudinal samples. Briefly, we subsampled polymorphic sites to 1,000x coverage to estimate allele counts from frequency data as in (Foll et al. 2014).

We then estimated the prior distribution of the effective population size using 10,000 bootstrap replicates. We selected a uniform distribution on the range [-0.5,0.5] as the prior distribution for the selection coefficients. The posterior distributions were determined from accepting the top 0.01% of 100,000 simulations.

Overview of models used for estimating the transmission bottleneck

We model transmission as a simple binomial sampling process (Sobel Leonard et al., 2017b). In our first model, we assume any transmitted iSNV, no matter the frequency, will be detected in the recipient. In the second, we relax this assumption and account for false negative iSNV in the recipient. To include the variance in the transmission bottlenecks between pairs we use maximum likelihood optimization to fit the average bottleneck size assuming the distribution follows a zero-truncated Poisson distribution.

Presence/Absence model

The presence/absence model makes several assumptions. We assume perfect detection of all transmitted iSNV in the recipient. For each donor iSNV, we measure only whether or not the variant is present in the recipient. Any iSNV that is not found in the recipient is assumed to have not been transmitted. We also assume the probability of transmission is determined only by the frequency of the iSNV in the donor at the time of sampling (regardless of how much time passes between sampling and transmission). The probability of transmission is simply the probability that the iSNV is included at least once in a sample size equal to the bottleneck. Finally, we assume all genomic sites are independent of one another. For this reason, we discarded the one case where the donor was likely infected by two strains, as the iSNV were certainly linked.

Because the presence/absence model is unaware of the frequency of alleles in the recipient we must track both alleles at each donor polymorphic site. Let A_1 and A_2 be alleles in donor j at genomic site i. Let $P(A_1)$ be the probability that A_1 is the only transmitted allele. There are three possible outcomes for each site. Either onlyA_1 is transmitted, only A_2 is transmitted, or both A_1 and A_2 are transmitted. The probability of only A_1 being transmitted given a bottleneck size of N_b is

$$P_{i,j}(A_1|N_b) = p_1^{N_b} \tag{4.11}$$

where p_1 is the frequency of A_1 in the donor. In other words, this is simply the probability of only drawing A_1 in N_b draws. The probability that only A_2 is transmitted is similarly defined.

The probability of both alleles being transmitted is given by

$$P_{i,j}(A_1, A_2|N_b) = 1 - \left(p_1^{N_b} + p_2^{N_b}\right) \tag{4.12}$$

where p_1 and p_2 are the frequencies of the alleles respectively. This is simply the probability of not picking only A_1 or only A_2 in N_b draws.

This system could easily be extended to cases where there are more than 2 alleles present at a site; however, that never occurs in our data set.

For ease we will denote the likelihood of observing the data at a polymorphic site i in each donor j given the bottleneck size N_b as $P_{i,j}(N_b)$ where $P_{i,j}(N_b)$ is defined by equation 4.11 if only one allele is transmitted and equation 4.12 if two alleles are transmitted.

The log likelihood of a bottleneck of size N_b is given by

$$LL(N_b) = \sum_j \sum_i \mathrm{Ln}(P_{i,j}) \tag{4.13}$$

where i, j refers to the ith polymorphic site in the jth donor. This is the log of the probability of observing the data summed over all polymorphic sites across all donors.

Because the bottleneck size is likely to vary across transmission events, we used maximum likelihood to fit the bottleneck distribution as oppose to fitting a single bottleneck value. Under this model we assumed the bottlenecks were distributed according to a zero-truncated Poisson distribution parameterized by λ. The likelihood of observing the data given a polymorphic site i in donor j and λ is

$$P_{i,j}(\lambda) = \sum_{N_b=1}^{\infty} P_{i,j}(N_b)P(N_b|\lambda) \tag{4.14}$$

where $P_{i,j}(N_b)$ is defined as above, $P(N_b|\lambda)$ is the probability of drawing a bottleneck of size N_b from a zero-truncated Poisson distribution with a mean of $\frac{\lambda}{1-e^{-\lambda}}$. The sum is across all possible N_b defined on $[1, \infty)$. Although for practical purposes we only investigated bottleneck sizes up to 100 as λ is quite small and the probability of drawing a bottleneck size of 100 from a zero-truncated Poisson distribution with $\lambda = 10$ is negligible. We follow this convention whenever this sum appears.

The log likelihood of λ for the data set is given by

$$LL(\lambda) = \sum_{j} \sum_{i} \mathrm{Ln}\Big(\sum_{N_b=1}^{\infty} P_{i,j}(N_b)P(N_b|\lambda)\Big) \tag{4.15}$$

Beta Binomial model

The Beta binomial model is explained in detail in Leonard *et al.* (Sobel Leonard et al., 2017b). It is similar to the presence/absence model in that transmission is modeled as a simple sampling process; however, it relaxes the following assumptions.

In this model, the frequencies of transmitted variants are allowed to change between transmission and sampling according a beta distribution. The distribution is not dependent on the amount of time that passes between transmission and sampling, but rather depends on the size of the founding population (here assumed to equal to N_b) and the number of variant genomes present in founding population k. Note the frequency in the donor is assumed to be the same between sampling and transmission.

The equations below are very similar to those presented by Leonard *et al.* with one exception. Because we know the sensitivity of our method to detect rare variants based on the expected frequency and the titer, we can include the possibility that iSNV are transmitted but are missed due to poor sensitivity. Because the beta binomial model is aware of the frequency of the iSNV in the recipient, no information is added by tracking both alleles at a genomic site i. Let p_{i,j_d}represent the frequency of the minor allele at position i in the donor of some transmission pair$j$. Similarly, let p_{i,j_d} be the frequency of that same allele in the recipient of the jth transmission pair. Then, as in Leonard *et al.*, the likelihood of some bottleneck N_b for the data at site i in pair$j$ where the minor allele is transmitted is given by

$$L(N_b)_{i,j} = \sum_{k=1}^{N_b} \text{p\_beta}\Big(p_{i,j_r}|k, N_b - k\Big)\text{p\_bin}\Big(k|N_b, p_{i,j_d}\Big) \tag{4.16}$$

Where p_beta is the probability density function for the beta distribution and p_bin is the probability mass function for the binomial distribution.

This is the probability density that the transmitted allele is found in the recipient at a frequency of p_{i,j_r} given that the variant was in k genomes in a founding population of size N_b times the probability k variant genomes would be drawn in a sample size of N_b from the donor where the variant frequency was p_{i,j_d}. This is then summed for all possible k where $1 \geq k \leq N_b$.

As in equation 4.14 the likelihood of a zero truncated Poisson with a mean of $\frac{\lambda}{1-e^{-\lambda}}$ given this transmitted variants is then given by

$$L(\lambda)_{i,j}^{\text{transmitted}} = \sum_{N_b=1}^{\infty} L(N_b)_{i,j} P(N_b|\lambda) \tag{4.17}$$

This is simply the likelihood of each N_b weighted by the probability of drawing a bottleneck size of N_b from the bottleneck distribution.

In this model, there are three possible mechanisms for a donor iSNV to not be detected in the recipient. (i) The variant was not transmitted. (ii) The variant was transmitted but is present below our level of detection (2%). (iii) The variant was transmitted and present above our level of detection but represents a false negative in iSNV identification.

As in Leonard *et al.* the likelihood of senerios (i) and (ii) for a given N_b are expressed as

$$L(N_b)_{i,j}^{\text{lost}} = \sum_{k=0}^{N_b} \text{p\_beta\_cdf}\Big(p_{i,j_r} < 0.02 | k, N_b - k\Big) \text{p\_bin}\Big(k | N_b, p_{i,j_d}\Big) \tag{4.18}$$

Where p_beta_cdf is the cumulative distribution function for the beta distribution. Note that if $k = 0$ (i.e. the iSNV was not transmitted) then the term reduces to the probability of not drawing the variant in N_b draws.

The likelihood of the variant being transmitted but not detected in the recipient given a bottleneck of N_b is described by

$$L(N_b)_{i,j}^{\text{missed}} = \sum_{k=0}^{N_b} \sum_{f_e}^{[0.02,0.05,0.1)} \text{p\_beta\_cdf}\big(f_e < p_{i,j_r} < f_{e+1} \big| k, N_b - k\big) \times \text{p\_bin}\Big(k | N_b, p_{i,j_d}\Big) \big(\text{FNR} | \text{Titer}_r, f\big) \tag{4.19}$$

This is the likelihood of the variant existing in the ranges [0.02,0.05] or [0.05,0.1] given an initial frequency of k/N_b and a bottleneck size of N_b multiplied by the expected False Negative Rate (FNR) given the titer of the recipient and the lower frequency tested. As in our diffusion model, we assumed perfect sensitivity for detection of iSNV present above 10%,rounded recipient titers down to the nearest $\log_{10}$titer (e.g. 10^3,10^4, 10^5) and assumed the entire range $[f_e, f_{e+1}]$ has the same sensitivity as the lower bound.

The likelihood of λ for iSNV that are not observed in the recipient is then given by summing equations 4.18 and 4.19 across all possible N_b.

$$L(\lambda)_{i,j}^{\text{nontransmitted}} = \sum_{N_b=1}^{\infty} \left(\left(L(N_b)_{i,j}^{\text{lost}} + L(N_b)_{i,j}^{\text{missed}} \right) P(N_b|\lambda) \right) \tag{4.20}$$

The log likelihood of the total dataset is then determined by summing log of equations 4.17 and 4.20 (as applicable) across all polymorphic sites in each donor. (As before here we sum of N_b within the range $[1, 100]$.)

Simulation

In order evaluate the fits of the two transmission models, we simulated whether or not each donor iSNV was transmitted or not. This involved converting each model to a presence absence model. In each simulation, we assigned a bottleneck from the bottleneck distribution for each transmission pair. We then determined the probability of only transmitting one allele (A_x where $x \in [1, 2]$ as in the presence/absence model above) and the probability of transmitted both alleles (A_1, A_2 above) for each polymorphic site.

For the presence/absence model the probabilities for each possible outcome are given by equations 4.11 and 4.12

For the beta binomial model the probability of only observing A_x at site i is given by

$$P(A_x|N_b) = L(N_b)_{i,j}^{\text{lost}} + L(N_b)_{i,j}^{\text{missed}} \tag{4.21}$$

where $L(N_b)_{i,j}^{\text{lost}}$ and $L(N_b)_{i,j}^{\text{missed}}$ are defined as in equations 4.18 and 4.19 respectively, but with p_{i,j_d} replaced by $1 - p_{i,j_d}$. This is simply the probability of not observing the other allele in the recipient.

Again, the probability of observing both alleles is

$$P(A_1, A_2|N_b) = 1 - \Big(P(A_1) + P(A_2)\Big) \tag{4.22}$$

where $P(A_1)$ and $P(A_2)$ are defined as in equation 4.21.

Fitting mutation rate and N_e

The diffusion approximation to the Wright - Fischer model allows us to make predictions on the allele frequency spectrum of a population given a mutation rate and an effective population size. The probability of observing a mutation at frequency p_t given an initial frequency of 0 can be approximated as in (Rouzine et al., 2001).

$$P(0, p_t, t, |\mu, N_e) = \frac{2\mu N_e}{p_t} e^{-\frac{2N_e p_t}{t}} \tag{4.23}$$

Where μ is the mutation rate. In this model mutation increases an allele's frequency from 0 but after that initial jump, drift is responsible for allowing the mutation to reach its observed frequency. Because the limit of equation 23 approaches infinity as p_t approaches 0 and for ease in numerical integration, we assumed that any variant present at less than 0.1% was essentially at 0%.

We then assumed each infection began as a clonal infection matching the consensus sequence observed at the time of sampling. The likelihood of observing minor alleles at the observed frequency is the given by equation 4.23.

As in the other within-host models, we can account for nonpolymorphic sites by adding the likelihood that no mutation is present $P(0, p_t \approx 0, t|p_t < 0.001, \mu, N_e)$, that a mutation is present but below our level of detection $P(0, p_t \approx 0, t|p_t < 0.02, \mu, N_e)$, and that a mutation is present but missed due to low sensitivity at low frequencies $P(0, p_t \approx 0, t|0.02 < p_t < 0.1, \mu, N_e)$. In this model we assumed 13133 mutagenic targets in each sample (the number of coding sites present in the reference strain from 2014-2015).

The probability of not observing a mutation is given by

$$P(0,0,t,|\mu,N_e) = P(0,p_t \approx 0,t|p_t < 0.001,\mu,N_e) + P(0,p_t \approx 0,t|p_t < 0.02,\mu,N_e) + P(0,p_t \approx 0,t|0.02 < p_t < 0.1,\mu,N_e) \tag{4.24}$$

Where

$$P(0,p_t \approx 0,t|p_t < 0.001,\mu,N_e) = 1 - \int_{0.001}^{1} P(0,p_t,t,|\mu,N_e)dp_t \tag{4.25}$$

and

$$P(0,p_t \approx 0,t|p_t < 0.02,\mu,N_e) = \int_{0.001}^{0.02} P(0,p_t,t,|\mu,N_e)dp_t \tag{4.26}$$

and

$$(4.27)\quad P(0, p_t \approx 0, t | 0.02 < p_t < 0.1, \mu, N_e) = \sum_{f_i}^{[0.02,0.05,0.10)} \left(\text{FNR}|\text{Titer}_r, f_i\right) \int_f^{f_{i+1}} P(p_0, p_t, t|\mu, N_e) dp_t$$

Where we follow the same convention as in equation 4.5. The log likelihood of a given μ and N_e pair is then the sum of the log of equations 4.23 and 4.24 for all possible sites in the data set.

Supplemental Figures and Tables

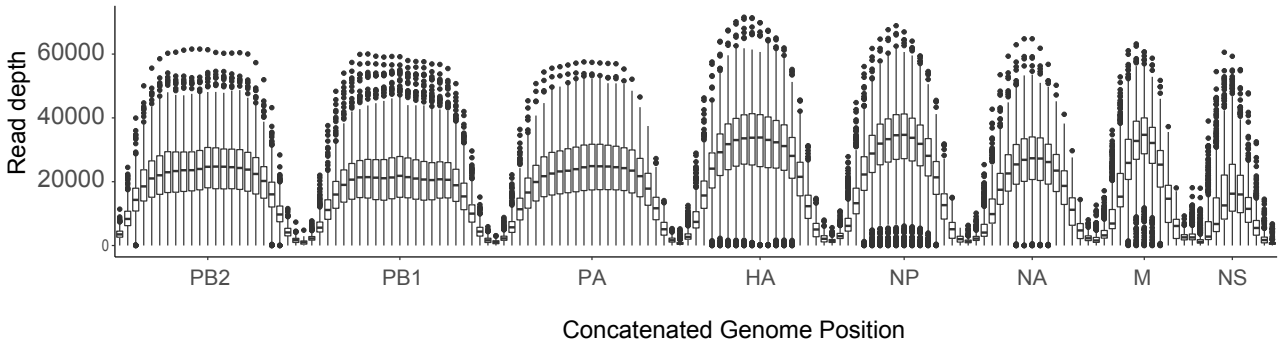


Figure 4.5: Sequence coverage for all samples. For each sample, the sliding window mean coverage was calculated using a window size of 200 and a step of 100. The distributions of these means are plotted as box plots (median, 25th and 75th percentiles, whiskers extend to most extreme point within median 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.

Table 4.3: Sensitivity and specificity of variant detection

Copy Number[a]	Variant Frequency	Original Pipeline[b] Sensitivity	Original Pipeline[b] Specificity	Current Pipeline[c] Sensitivity	Current Pipeline[c] Specificity
$> 10^5$	0.05	1	>0.9999	0.85	1.000
	0.02	0.85	0.9999	0.15	1.000
	0.01	0.95	0.9995	-	-
	0.005	0.35	0.9999	-	-
$10^4 - 10^5$	0.05	0.95	0.9999	0.85	1.000
	0.02	0.9	0.9999	0.15	1.000
	0.01	0.8	0.9998	-	-
	0.005	0.4	0.9999	-	-
$10^3 - 10^4$	0.05	0.8	>0.9999	0.70	1.000
	0.02	0.45	0.9999	0.15	1.000
	0.01	0.2	0.9997	-	-
	0.005	0.1	0.9999	-	-

[a] Per μl transport media
[b] As described in (McCrone and Lauring, 2016)
[c] As described in Methods, benchmarked for frequencies 0.02-0.98 only

Table 4.4: Nonsynonymous substitutions in HA antigenic sites

House ID	Enrollment ID	Symptom Onset	Subtype	Frequency	Amino Acid Change	Antigenic Site	Vaccinated	Day Day of Symptoms
1111	300481	3-30-2011	H3N2	0.071	E62G	E*	No	0
2166	320661	2-13-2012	H3N2	0.071	V297A	C	Yes	1
1302	301355	3-20-2011	H3N2	0.088	L86I	E	Yes	1
3075	331045	12-10-2012	H3N2	0.066	I214T	D	Yes	1
5219	50935	12-5-2014	H3N2	0.175	F193S	B*[a]	No	3
5263	51106	12-6-2014	H3N2	0.111	T128A	B	Yes	3
5290	51225	12-15-2014	H3N2	0.405	I260V	E*	Yes	1
5302	51273	12-13-2014	H3N2	0.030	S262N	E*	Yes	0
5098	50419	12-22-2014	H3N2	0.364	G208R	D	Yes	4
5033	50141	12-3-2014	H3N2	0.032	A163T	B	Yes	2
5034	50143	1-11-2015	H3N2	0.119	I307R	C	Yes	1
5289	51220	12-13-2014	H3N2	0.038	K189N	B*[a]	Yes	-1
5033	50141	12-3-2014	H3N2	0.025	D53E	C*	Yes	1
5033	50141	12-3-2014	H3N2	0.023	S312G	C	Yes	1
5269	51132	12-6-2014	H3N2	0.028	I242T	D	Yes	2
5147	50630	11-18-2014	H3N2	0.164	I242L	D	Yes	1
5034	50143	1-11-2015	H3N2	0.161	I307R	C	Yes	2
4185	UM40738	12-14-2013	H1N1	0.021	R208K	Ca	No	2

* Sites observed to vary between antigenically distinct strains in (Wiley et al., 1981) and (Smith et al., 2004).
[a] Sites located in the 'antigenic ridge' identified in (Koel et al., 2013).

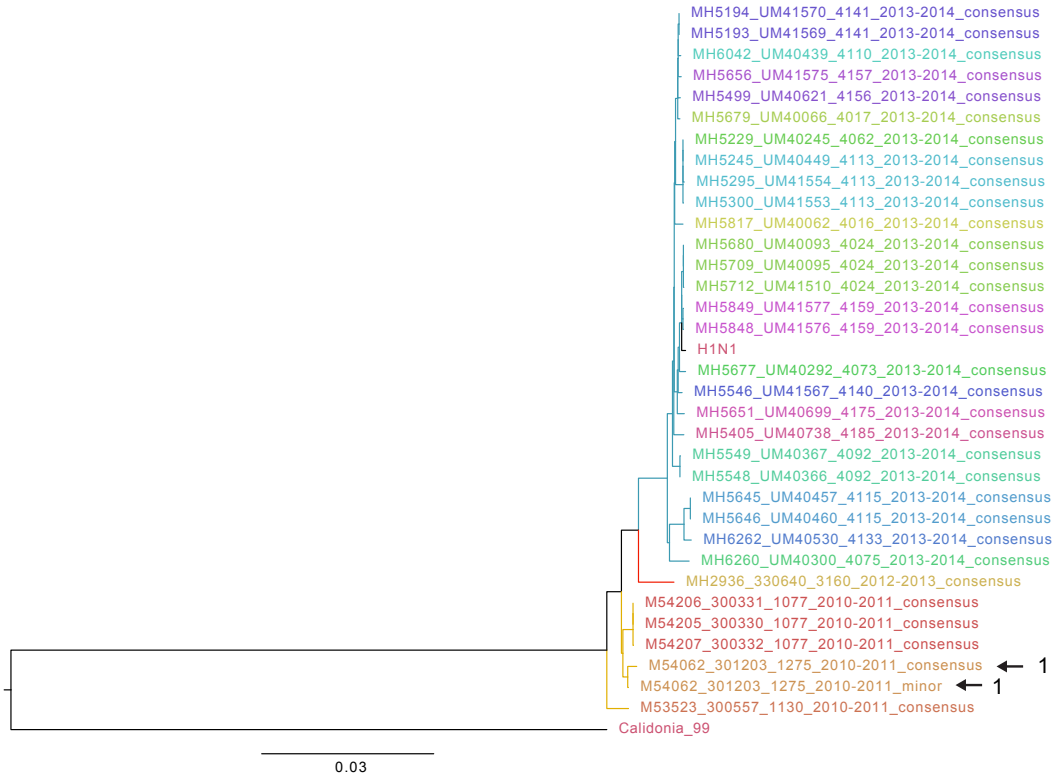


Figure 4.6: Approximate maximum likelihood trees of the concatenated coding sequences for high quality H1N1 samples. The branches are colored by season; the tip identifiers are colored by household. Arrows with numbers indicate consensus and putative minor haplotypes for samples with greater than 10 iSNV. Trees were made using FastTree (Price et al., 2010).

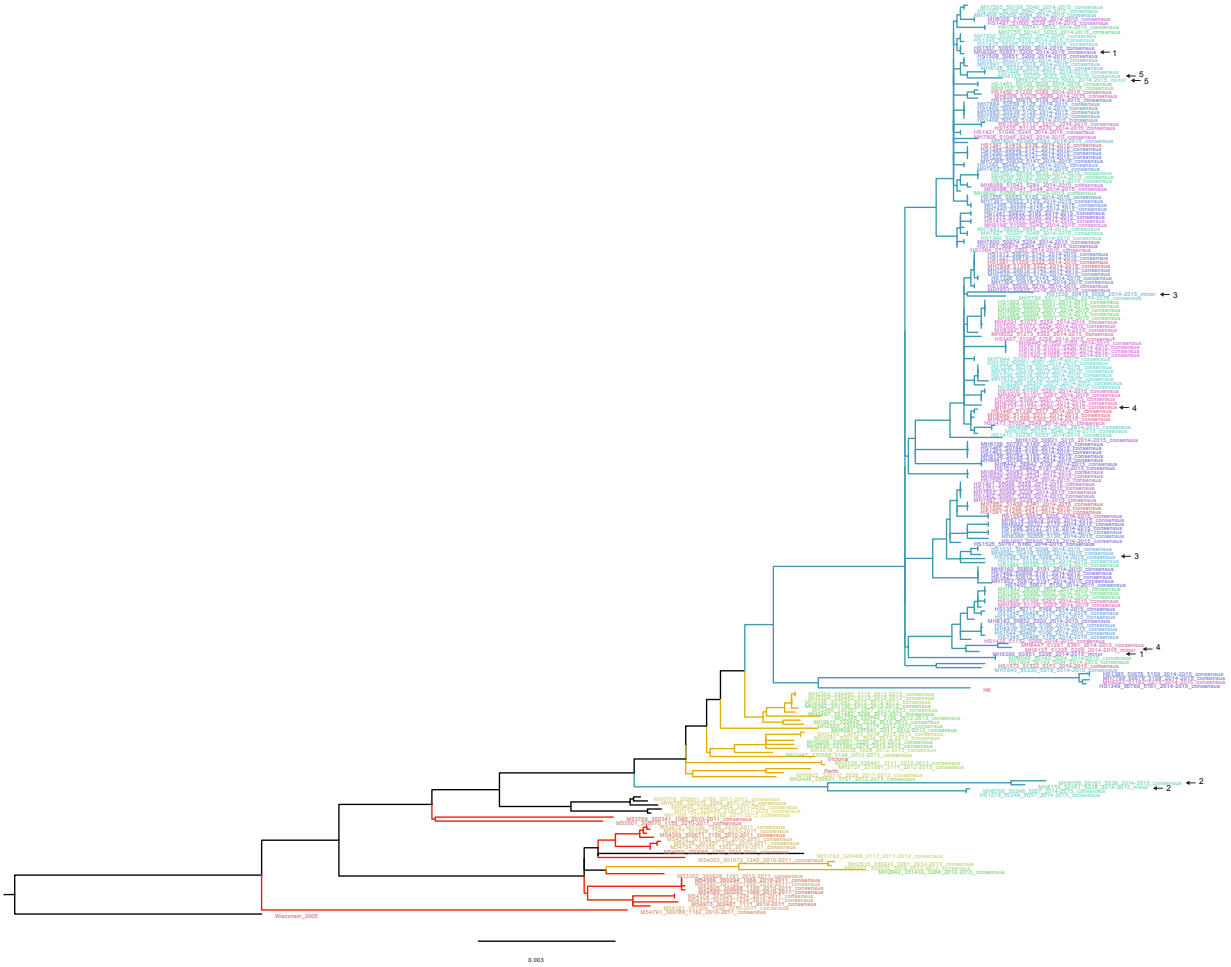


Figure 4.7: Approximate maximum likelihood trees of the concatenated coding sequences for high quality H3N2 samples. The branches are colored by season; the tip identifiers are colored by household. Arrows with numbers indicate consensus and putative minor haplotypes for samples with greater than 10 iSNV. Trees were made using FastTree (Price et al., 2010).

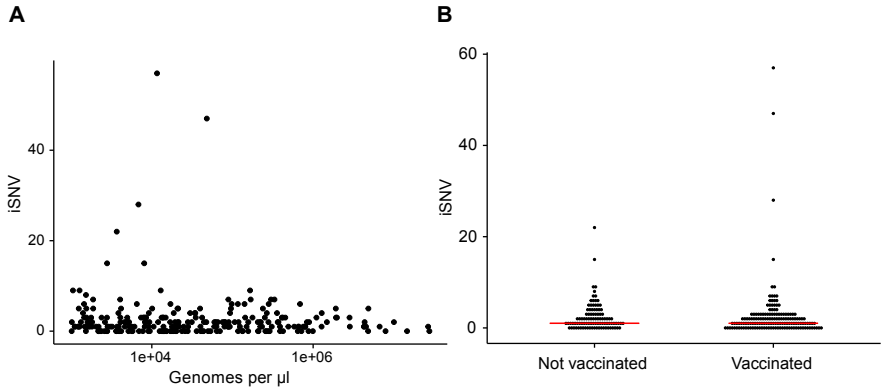


Figure 4.8: The effect of titer and vaccination on the number of iSNV identified. (A) The number of iSNV identified in an isolate (y-axis) plotted against the titer (x-axis, genomes/μl transport media). (B) The number of iSNV identified in each isolate stratified by whether that individual was vaccinated or not. Red bars indicate the median of each distribution.

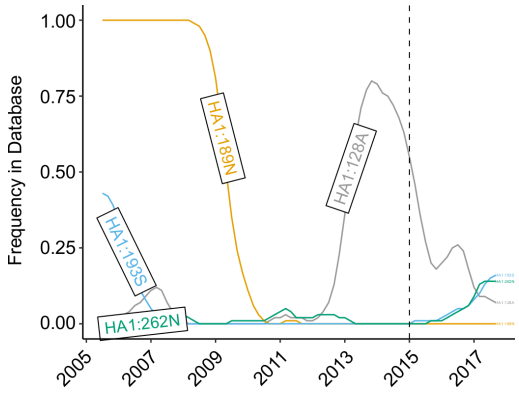


Figure 4.9: Minority nonsynonymous iSNV in global circulation. The global frequencies of the amino acids that were found as minority variants in sample isolates (x-axis) plotted overtime (y-axis). Each amino acid trace is labeled according to the H3 number scheme. All samples were isolated in December of 2014 (gray line).

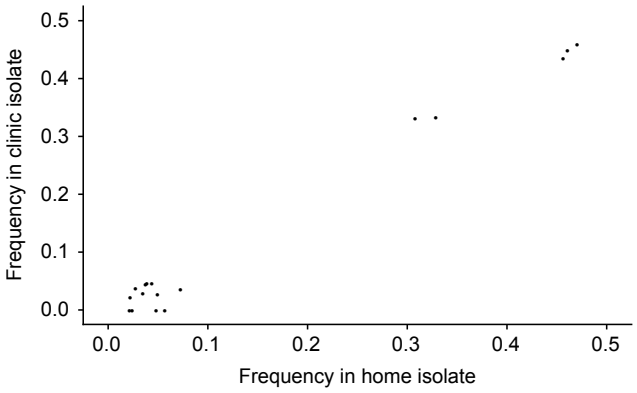


Figure 4.10: Reproducibility of iSNV identification for paired samples acquired on the same day. The x-axis represents iSNV frequencies found in the home-acquired nasal swab. The y-axis represents iSNV frequencies found the clinic-acquired combined throat and nasal swab.

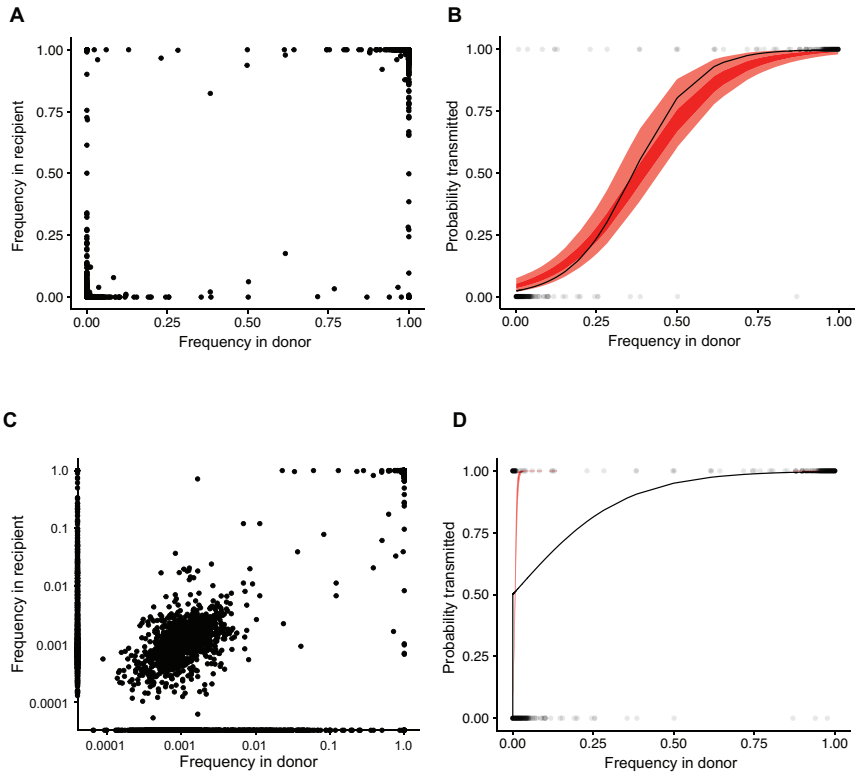


Figure 4.11: Estimate of effective bottleneck size with relaxed variant calling criteria. (A) The frequency of iSNV in both recipient and donor isolates. iSNV were identified using the original variant calling pipeline. (B) The presence-absence model fit compared to the observed data for iSNV identified using the original variant calling pipeline. The x-axis represents the frequency of donor iSNV with transmitted iSNV plotted along the top and nontransmitted iSNV plotted along the bottom. The black line indicates the probability of transmission for a given iSNV frequency as determined by logistic regression. Similar fits were calculated for 1,000 simulations with a mean bottleneck size of 2.10. Fifty percent of simulated outcomes lie in the darkly shaded region and 95% lie in the lightly shaded regions. (C) Similar to (A) but with minority iSNV identified using the current analytical framework without a frequency threshold. (D) Similar to B but with minority iSNV identified using the current analytical framework without a frequency threshold.

CHAPTER V

Discussion

Rapid antigenic evolution allows influenza virus to infect a significant proportion of the population each year. While the dynamics of influenza evolution are observed on the global scale, selective forces act at the level of individual hosts. In my thesis, I have used a validated sequencing approach to explore the evolutionary dynamics of influenza within human hosts during naturally-occurring, acute infections, and between susceptible hosts during transmission. I have found that within-host populations are dynamic, and yet putative antigenic mutations are rare and undistinguishable from other single nucleotide variants (SNV). Additionally, the evolutionary dynamics during transmission are consistent with a random sampling process and a tight transmission bottleneck. These findings challenge the assumption that antigenic selection shapes influenza diversity at all levels. In over 10,000 person-seasons of observations, I found that selection for antigenic variants was not a significant factor in determining within-host diversity. My results suggest that global evolutionary dynamics likely result from selection for traits that increase infectivity (i.e., host susceptibility) and allow a given lineage to infect more hosts than other circulating strains. These findings help reshape our view of influenza evolution and

will inspire a more nuanced study of the forces that drive the evolutionary trajectory of the virus.

Sequencing methodology and evolutionary questions

Before studying patient-derived populations, I validated our sequencing approach on known samples (Chapter II). This experiment provided estimates of our sensitivity and specificity to detect low-frequency SNV. I found that previously validated variant callers were unable to accurately discriminate true and false positive SNV present in low-titer RNA virus populations. For samples with titers below 10^{-5} genome copies/μl, these inaccuracies persisted even after applying stringent quality thresholds. Only by sequencing low-titer isolates in duplicate, was I able to control for RT and PCR errors.

I identified true SNV from patient-derived samples by first including all possible SNV and then removing those with sub-par mapping quality, Phred score, or average position on the read. Although this method accurately identified low-frequency SNV, it created discord between the raw data and the variant frequency measurements. Under the current approach, sequencing artifacts and RT-PCR errors confounded the frequency of SNV in the raw alignment data. These erroneous bases were removed during the filtering stage, and the frequencies of true SNV were corrected before secondary analysis. A more robust method would be to filter and trim sequencing reads before variant identification. I have tested this approach in a pilot experiment done in conjunction with the bioinformatic core on campus and found it identified variants with similar accuracy as the method in Chapter II. This new approach would also allow for additional analyses, like haplotype reconstruction, which rely on alignment files and not variant calls alone. The dataset provided by my original

benchmarking experiment offers a valuable resource for future validation of this and other approaches to sequencing patient-derived isolates.

Throughout my thesis, I required SNV meet stringent criteria to be included in the secondary analysis. Because these SNV represented real mutations, I was able to interpret their dynamics as outcomes of biological processes and not methodological artifacts. However, this approach limited my power to detect parallel evolution within individuals. At the global level, antigenic mutations arise on similar, but not identical backgrounds (Bedford et al., 2015), which suggests there is at least a small degree of parallel evolution. If the same mutations were to arise repeatedly within individuals, their presence could influence the evolutionary trajectories observed at higher scales.

A recent study of within-host populations of HIV took an alternative approach to variant calling, which could provide more power to capture parallel evolution in within-host populations. Instead of limiting the number of putative SNV to only those likely to be real, Zanini et al. (2017) included all putative variants. The authors then partitioned loci into various groups and looked for differences in SNV prevalence and frequency between groups. If the groups are sufficiently large, then methodological errors should affect all groups equally so that any differences between groups indicate underlying biological processes.

The forces that shape within-host populations

Many studies, including those in Chapters III and IV, have found evidence of antigenic SNV present in within-host isolates from both human and animal models (Debbink et al., 2017; Dinis et al., 2016; Murcia et al., 2013; Hughes et al., 2012; Murcia et al., 2012, 2010). However, in all cases, these mutations were present at low

frequencies and only found in a minority of infections. The presence of these mutations alone is not evidence of selection. There are; however, a few instances where within-host selection has been reported. These studies have often relied on poorly adapted viruses or chronically infected patients. Under these conditions,beneficial mutations are easily accessible, and selection is sufficiently strong to shape the population. These situations do not capture the characteristics present in acute infections. The two studies that have detected selection during acute infections have relied on a recently developed algorithm which constructs haplotypes from short-read data and then estimates selection coefficients consistent with the frequency changes present in the data (Sobel Leonard et al., 2017a; Illingworth et al., 2014). This approach assumes that any change in frequency, outside the variance of frequency measurements, is evidence of selection. Such a model ignores neutral processes such as drift, migration, and hitchhiking, which affect allele frequencies independent of selection.

If it exists, positive selection for growth and survival within a host is likely a rare event. Influenza infections are acute, lasting only a week. By definition, infected individuals do not have a strong, targeted immune response against the infecting strain. Also, the adaptive immune system takes a week to mount a naive response, by which point most infections are over (Petrova and Russell, 2017). Under these circumstances, it would be remarkable for selection to efficiently drive antigenic alleles to detectable frequencies over the course of an infection.

The findings in Chapter III are consistent with high barriers to efficient within-host selection. I found that the strength of an individual's immune response did not predict the presence of putative antigenic mutations at least above a frequency of 1%. There are several reasons why this cross-sectional study could have failed to detected evidence of antigenic selection. It could be that serum samples provide

little indication of the immunity faced by the virus replicating in the mucosal regions of the upper respiratory track (Tumpey et al., 2001). It is also possible that the HAI and NAI assays used to measure antibody titer do not perfectly capture all the possible protective, interactions between antibodies and virus. In each of these cases, my analysis would be confounded by an inaccurate proxy for immune pressure. However, the sparsity of SNV on HA and NA, suggests that antigenic selection over the course of an infection is unlikely.

The longitudinal samples used in Chapter IV allowed for a more detailed exploration of within-host processes. Consistent with the findings in Chapter III, I found that antigenic mutations follow similar within-host trajectories as other nonsynonymous mutations and do not reach higher frequencies (Figure 5.1). Mutations outside of antigenic regions are expected to decrease HA stability and viral fitness (Łuksza and Lässig, 2014). The similar frequencies observed between these groups of mutations, which are expected to lie on opposite ends of the fitness spectrum suggest the populations behave according to neutral processes. This observation is captured in the small N_e size estimated in Chapter IV and is supported by the similar N_e sizes found for nonsynonymous and synonymous mutations.

While the effective population size provides an interesting starting point, it does not provide a mechanistic understanding of influenza evolutionary dynamics during acute infections. The effective population size reported in Chapter IV is orders of magnitude smaller than the census population. Understanding the mechanisms behind the small N_e is essential for developing a more detailed understanding of influenza evolution at the host level.

Of particular interest is the role of viral growth and decay over the course of an infection. Influenza is known to undergo two to four days of exponential growth

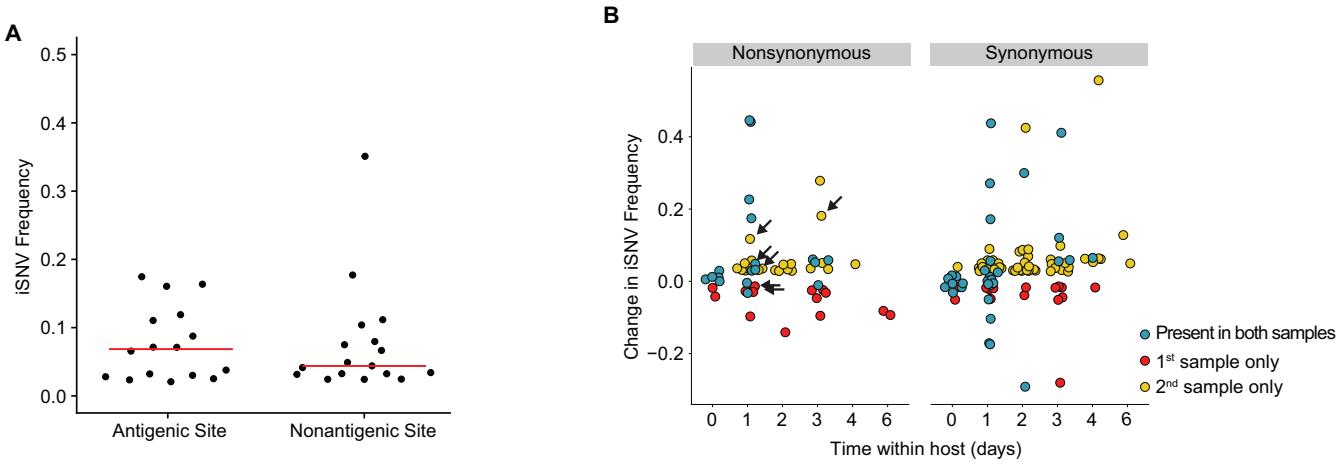


Figure 5.1: Within-host dynamics of antigenic iSNV. (A) The within-host frequency of nonsynonymous mutations in HA stratified by whether or not they are in known antigenic sites (p=0.46 Wilcoxon rank sum). (B) The change in frequency over time for minority iSNV identified for longitudinal samples. Nonsynonymous 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.

followed by a rapid decay (Figure 4.1 A and (Xue et al., 2018)). During exponential growth, minor fitness differences can allow slightly beneficial mutations to reach meaningful frequencies. A mutation with a growth advantage of 5% made during the first replication cycle can reach a frequency of 16% in only ten generations. If the same mutation arises in a similar population that is not expanding, a selective advantage of 5% will only increase its frequency to 8% in ten generations. This simple example highlights the importance of accounting for infection dynamics in models of influenza evolution. Because the mutation rate is 4×10^{-5} mutations/nucleotide/replication cycle and the transmission bottleneck is stringent, we would have to sample tens of thousands of individuals to capture such an event. However, given the large number of infections that take place each year, these rare events could impact global trends.

Population structure, and migration can also decrease the N_e and has been reported in a ferret model of influenza (Lakdawala et al., 2015). Mutations may be under strong positive selection at various locations within the respiratory tract, and yet there would be little to no evidence of selection shaping the sampled population. Understanding the factors that shape the nose and throat population is important as this region likely contributes the most to transmission.

My work has not definitively shown that within-host populations of influenza evolve neutrally. However, the work in Chapters III and IV suggest neutral evolution is a plausible explanation and provides a framework from which to address this question in the future. As more longitudinal samples from acutely infected individuals become available, it will be possible to include selection coefficients in the models discussed in Chapter IV. Better defined allele trajectories will provide more convincing evidence of which sites (if any) deviate from neutral processes. More longitudinal sampling would also allow us to move beyond the assumption of site independence. It would be possible to apply haplotype reconstruction algorithms, which take into account overlapping reads and similarities in frequency trajectories to explore which SNV are likely linked. This method would provide a fuller depiction of the within-host population, and make the data more amenable to classical population genetic approaches, in which allele frequencies represent genotypes.

Linking within- and between-host evolution

The transmission bottleneck links within-host processes to dynamics at higher levels. My data suggest that even if positive selection affects the frequency of alleles below the level of detection, the transmission of rare variants is extremely unlikely (with probability roughly equal to the within-host frequency of the variant). How-

ever, my estimate is confounded by the within-host dynamics that are likely to have occurred in both the donor and recipient between the time of transmission and the time of sampling. The reported bottleneck represents the effective or observed transmission bottleneck. Rare variants may have been transmitted but then lost before sampling. Stochastic forces early in infection have been shown to affect the progression of the infection (Duneau et al., 2017). Understanding the diversity of the founding population would enhance our understanding of how early events shape both influenza infection and evolution.

Rapid exponential growth early in infection could explain the stringent bottleneck observed in Chatper IV. If selection is strong, growth advantages can have extreme affects on allele frequencies. During exponential growth, a variant with a growth advantage of 30% can increase in frequency from 0.1% to 98% within ten generations. This dramatic change in frequency is a possible mechanism for the stringent transmission bottleneck. During exponential growth, less fit variants would be purged from the population rapidly. However, strong selection alone can not explain the diversity observed between transmission pairs. There are a few transmission pairs in which recipient hosts contain multiple donor SNV. These events as well as the mixed infections reported in Chapter IV would not be possible under a very intense selective regime.

The independent action hypothesis offers a complementary explanation for the tight bottleneck (Zwart and Elena, 2015). Under this model, each virion in the inoculating dose has a small probability of successfully founding the infection. The expected number of founders is the product of the inoculating dose and the probability of infection. Our data suggest the likelihood of any virion successfully initiating infection is very low. Looser transmission bottlenecks are possible if the probability

of infection increases, either due to host or viral factors. Increased host susceptibility may in part explain the discrepancy between my results and the much larger transmission bottleneck reported by Poon et al. (2016) during a pandemic. However, to explain the entire discrepancy, host susceptibility in the Poon et al. (2016) study would have to be 100 fold higher than in our cohort. This is unlikely even during a pandemic.

The within-host models developed in Chapter IV provide a solution for dissecting the within and between-host effects on the effective transmission bottleneck. I have shown that it is possible to measure evolutionary parameters such as the within-host effective population size and the within-host mutation rate from patient-derived samples. Combining my within-host and transmission models would allow us to account for the time that passes in both the donor and recipient between sampling and transmission. The aggregated model would estimate the largest transmission bottleneck consistent with the data while providing a framework to account for SNV that present in the recipient but not the donor.

A combined within and between host model could also be employed in epidemiological studies to identify transmission pairs. Understanding the difference between household and community-acquired infections is vital in determining accurate estimates of vaccine efficacy (Petrie et al., 2017). In Chapter IV I used the genetic distance between isolates to differentiate between household and community-acquired infection. However, this approach largely ignores the information provided by minority SNV. Current methods that account for SNV are based on partitioning the genome into regions and then constructing a phylogeny for each region (Wymant et al., 2017). However, the connection between the underlying biology and observed trees is unclear. The proposed model directly relates to the evolutionary processes

at play. Additionally, many households contain more than two concurrent infections, which can be explained by multiple possible transmission chains. Combining the within and between-host models would provide a framework for testing which of the possible chains most likely occurred.

Evolutionary dynamics at larger scales

The next step in understanding influenza evolution is to connect host-level dynamics to processes at the global scale. It is possible to use a combined within- and between-host model to simulate the evolution of influenza within the HIVE cohort. This model would have to account for the number of migration events into the HIVE cohort as well as unobserved infections from the outside community. This approach would allow us to determine whether or not the rate of evolution in the cohort matches the dynamics expected under neutral, host-level processes. The combined model would also help estimate to what degree (if any) selection contributes to local dynamics.

I suspect that selection for infectivity (i.e., increased host susceptibility) is the principal mechanism behind antigenic evolution. This selection is expected to act early in infection during the rapid exponential expansion of the virus. Depending on the host's specific immune response, antigenic mutations that are present early in infection could have a sufficient fitness advantage to reach fixation quickly. In the future, members of HIVE will provide self-samples as soon as a family member becomes symptomatic. These additional time points will help characterize the importance of early infection dynamics in influenza evolution.

Selection for increased infectivity also has the potential to drive influenza evolution at higher scales. The transmission study in Chapter IV did not include cases where

transmission did not occur. Characterizing the virus and host factors that restrict transmission, would provide insight into the conditions required for antigenic selection at the host level. An SIR model, which incorporates selection as competition for susceptible hosts, was able to account for several aspects of the global dynamics of influenza (Zinder et al., 2013). The HIVE cohort provides an ideal setting for exploring what viral and host characteristics contribute to antigenic selection in naturally occurring infections. Do antigenically drifted samples have a higher probability of transmission or transmit to more individuals? Are there differences in host immunity not captured by only measuring the response against the vaccine strain? If so, do these differences contribute to vaccine failure and infection?

The majority of samples included in my study came from the 2014-2015 season in which there was an antigenic mismatch between the circulating strain and the vaccine strain. However, in the intervening years, herd immunity in the HIVE cohort has likely increased against this circulating strain. A careful analysis of household transmission events in coordination with antibody titers and antigenic measurements of circulating viruses may find evidence of higher transmission rates when there is a mismatch between virus antigenicity and host immunity. These processes could be modeled to determine how much of the global dynamics can be explained by neutral dynamics within-hosts and herd immunity.

The global evolutionary dynamics of influenza virus emerge from a combination of factors acting across biological scales. In my thesis, I have shown that antigenic selection, the dominant global phenomena, is not a major contributing factor over the course of an individual infection. Importantly, I have also estimated several evolutionary parameters including the within-host effective population size, the within-host mutation rate, and the effective transmission bottleneck. These findings,

parameters, and models provide the first steps in quantitatively linking host-level processes to evolutionary dynamics at larger scales.

Bibliography

Ashley Acevedo and Raul Andino. Library preparation for highly accurate population sequencing of RNA viruses. *Nature protocols*, 9(7):1760–1769, July 2014.

Ashley Acevedo, Leonid Brodsky, and Raul Andino. Mutational and fitness landscapes of an RNA virus revealed through population sequencing. *Nature*, 505(7485):686–690, January 2014.

Akhtar Ali, Hongye Li, William L Schneider, Diana J Sherman, Stewart Gray, Dawn Smith, and Marilyn J Roossinck. Analysis of genetic bottlenecks during horizontal transmission of Cucumber mosaic virus. *Journal of virology*, 80(17):8345–8350, September 2006.

Samuel Alizon, Fabio Luciani, and Roland R Regoes. Epidemiological and clinical consequences of within-host evolution. *Trends in Microbiology*, 19(1):24–32, January 2011.

Kristian G Andersen, B Jesse Shapiro, Christian B Matranga, Rachel Sealfon, Aaron E Lin, Lina M Moses, Onikepe A Folarin, Augustine Goba, Ikponmwonsa Odia, Philomena E Ehiane, Mambu Momoh, Eleina M England, Sarah Winnicki, Luis M Branco, Stephen K Gire, Eric Phelan, Ridhi Tariyal, Ryan Tewhey, Omowunmi Omoniwa, Mohammed Fullah, Richard Fonnie, Mbalu Fonnie, Lansana Kanneh, Simbirie Jalloh, Michael Gbakie, Sidiki Saffa, Kandeh Karbo, Adrianne D Gladden, James Qu, Matthew Stremlau, Mahan Nekoui, Hilary K Finucane, Shervin Tabrizi, Joseph J Vitti, Bruce Birren, Michael Fitzgerald, Caryn McCowan, Andrea Ireland, Aaron M Berlin, James Bochicchio, Barbara Tazon-Vega, Niall J Lennon, Elizabeth M Ryan, Zach Bjornson, Danny A Milner Jr, Amanda K Lukens, Nisha Broodie, Megan Rowland, Megan Heinrich, Marjan Akdag, John S Schieffelin, Danielle Levy, Henry Akpan, Daniel G Bausch, Kathleen Rubins, Joseph B McCormick, Eric S Lander, Stephan Günther, Lisa Hensley, Sylvanus Okogbenin, Stephen F Schaffner, Peter O Okokhere, S Humarr Khan, Donald S Grant, George O Akpede, Danny A Asogun, Andreas Gnirke, Joshua Z Levin, Christian T Happi, Robert F Garry, Pardis C Sabeti, and Viral Hemorrhagic Fever Consortium20. Clinical Sequencing Uncovers Origins and Evolution of Lassa Virus. *Cell*, 162(4):738–750, August 2015.

Italo Archetti and Frank L Horsfall. Persistent Antigenic Variation of Influenza a Viruses After Incomplete Neutralization in Ovo With Heterologous Immune Serum. *Journal of Experimental Medicine*, 92(5):441–462, November 1950.

A Beck, R B Tesh, T G Wood, S G Widen, K D Ryman, and A D T Barrett. Comparison of the Live Attenuated Yellow Fever Vaccine 17D-204 Strain to Its Virulent Parental Strain Asibi by Deep Sequencing. *Journal of Infectious Diseases*, 209(3):334–344, January 2014.

Trevor Bedford, Steven Riley, Ian G Barr, Shobha Broor, Mandeep Chadha, Nancy J Cox, Rodney S Daniels, C Palani Gunasekaran, Aeron C Hurt, Anne Kelso, Alexander Klimov, Nicola S Lewis, Xiyan Li, John W McCauley, Takato Odagiri, Varsha Potdar, Andrew Rambaut, Yuelong Shu, Eugene Skepner, Derek J Smith, Marc A Suchard, Masato Tashiro, Dayan Wang, Xiyan Xu, Philippe Lemey, and Colin A Russell. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*, 523(7559):217–220, June 2015.

David R Bentley, Shankar Balasubramanian, Harold P Swerdlow, Geoffrey P Smith, John Milton, Clive G Brown, Kevin P Hall, Dirk J Evers, Colin L Barnes, Helen R Bignell, Jonathan M

Boutell, Jason Bryant, Richard J Carter, R Keira Cheetham, Anthony J Cox, Darren J Ellis, Michael R Flatbush, Niall A Gormley, Sean J Humphray, Leslie J Irving, Mirian S Karbelashvili, Scott M Kirk, Heng Li, Xiaohai Liu, Klaus S Maisinger, Lisa J Murray, Bojan Obradovic, Tobias Ost, Michael L Parkinson, Mark R Pratt, Isabelle M J Rasolonjatovo, Mark T Reed, Roberto Rigatti, Chiara Rodighiero, Mark T Ross, Andrea Sabot, Subramanian V Sankar, Aylwyn Scally, Gary P Schroth, Mark E Smith, Vincent P Smith, Anastassia Spiridou, Peta E Torrance, Svilen S Tzonev, Eric H Vermaas, Klaudia Walter, Xiaolin Wu, Lu Zhang, Mohammed D Alam, Carole Anastasi, Ify C Aniebo, David M D Bailey, Iain R Bancarz, Saibal Banerjee, Selena G Barbour, Primo A Baybayan, Vincent A Benoit, Kevin F Benson, Claire Bevis, Phillip J Black, Asha Boodhun, Joe S Brennan, John A Bridgham, Rob C Brown, Andrew A Brown, Dale H Buermann, Abass A Bundu, James C Burrows, Nigel P Carter, Nestor Castillo, Maria Chiara E Catenazzi, Simon Chang, R Neil Cooley, Natasha R Crake, Olubunmi O Dada, Konstantinos D Diakoumakos, Belen Dominguez-Fernandez, David J Earnshaw, Ugonna C Egbujor, David W Elmore, Sergey S Etchin, Mark R Ewan, Milan Fedurco, Louise J Fraser, Karin V Fuentes Fajardo, W Scott Furey, David George, Kimberley J Gietzen, Colin P Goddard, George S Golda, Philip A Granieri, David E Green, David L Gustafson, Nancy F Hansen, Kevin Harnish, Christian D Haudenschild, Narinder I Heyer, Matthew M Hims, Johnny T Ho, Adrian M Horgan, Katya Hoschler, Steve Hurwitz, Denis V Ivanov, Maria Q Johnson, Terena James, T A Huw Jones, Gyoung-Dong Kang, Tzvetana H Kerelska, Alan D Kersey, Irina Khrebtukova, Alex P Kindwall, Zoya Kingsbury, Paula I Kokko-Gonzales, Anil Kumar, Marc A Laurent, Cynthia T Lawley, Sarah E Lee, Xavier Lee, Arnold K Liao, Jennifer A Loch, Mitch Lok, Shujun Luo, Radhika M Mammen, John W Martin, Patrick G McCauley, Paul McNitt, Parul Mehta, Keith W Moon, Joe W Mullens, Taksina Newington, Zemin Ning, Bee Ling Ng, Sonia M Novo, Michael J O'Neill, Mark A Osborne, Andrew Osnowski, Omead Ostadan, Lambros L Paraschos, Lea Pickering, Andrew C Pike, Alger C Pike, D Chris Pinkard, Daniel P Pliskin, Joe Podhasky, Victor J Quijano, Come Raczy, Vicki H Rae, Stephen R Rawlings, Ana Chiva Rodriguez, Phyllida M Roe, John Rogers, Maria C Rogert Bacigalupo, Nikolai Romanov, Anthony Romieu, Rithy K Roth, Natalie J Rourke, Silke T Ruediger, Eli Rusman, Raquel M Sanches-Kuiper, Martin R Schenker, Josefina M Seoane, Richard J Shaw, Mitch K Shiver, Steven W Short, Ning L Sizto, Johannes P Sluis, Melanie A Smith, Jean Ernest Sohna Sohna, Eric J Spence, Kim Stevens, Neil Sutton, Lukasz Szajkowski, Carolyn L Tregidgo, Gerardo Turcatti, Stephanie vandeVondele, Yuli Verhovsky, Selene M Virk, Suzanne Wakelin, Gregory C Walcott, Jingwen Wang, Graham J Worsley, Juying Yan, Ling Yau, Mike Zuerlein, Jane Rogers, James C Mullikin, Matthew E Hurles, Nick J McCooke, John S West, Frank L Oaks, Peter L Lundberg, David Klenerman, Richard Durbin, and Anthony J Smith. Accurate whole human genome sequencing using reversible terminator chemistry. *Nature Publishing Group*, 456(7218):53–59, November 2008.

C T Bergstrom, P McElhany, and L A Real. Transmission bottlenecks as determinants of virulence in rapidly evolving pathogens. *Proceedings of the National Academy of Sciences*, 96(9):5095–5100, April 1999.

Mónica Betancourt, Alberto Fereres, Aurora Fraile, and Fernando García-Arenal. Estimation of the effective number of founders that initiate an infection after aphid transmission of a multipartite plant virus. *Journal of virology*, 82(24):12416–12421, December 2008.

Anthony M Bolger, Marc Lohse, and Bjoern Usadel. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics (Oxford, England)*, 30(15):2114–2120, August 2014.

Rowena A Bull, Fabio Luciani, Kerensa McElroy, Silvana Gaudieri, Son T Pham, Abha Chopra, Barbara Cameron, Lisa Maher, Gregory J Dore, Peter A White, and Andrew R Lloyd. Sequential bottlenecks drive viral evolution in early acute hepatitis C virus infection. *PLOS Pathogens*, 7 (9):e1002243, September 2011.

A J Caton, G G Brownlee, J W Yewdell, and W Gerhard. The antigenic structure of the influenza virus A/PR/8/34 hemagglutinin (H1 subtype). *Cell*, 31(2 Pt 1):417–427, December 1982.

Center of Disease Control. CDC protocol of realtime RTPCR for influenza A(H1N1), 2009.

Center of Disease Control. **Selecting viruses for the seasonal influenza vaccine**, 2016. URL `http://www.cdc.gov/flu/about/season/vaccine-selection.htm`.

Brian Charlesworth. Fundamental concepts in genetics: effective population size and patterns of molecular evolution and variation. *Nature Reviews Genetics*, 10(3):195–205, March 2009.

Marine Combe, Raquel Garijo, Ron Geller, José M Cuevas, and Rafael Sanjuán. Single-Cell Analysis of RNA Virus Infection Identifies Multiple Genetically Diverse Viral Genomes within Single Infectious Units. *Cell host & microbe*, 18(4):424–432, October 2015.

J M Cuevas, A Willemsen, J Hillung, M P Zwart, and S F Elena. Temporal Dynamics of Intrahost Molecular Evolution for a Plant RNA Virus. *Molecular Biology and Evolution*, 32(5):1132–1147, April 2015.

Valentina D'Arienzo, Alain Moreau, Louis D'Alteroche, Valérie Gissot, Emmanuelle Blanchard, Catherine Gaudy-Graffin, Emmanuelle Roch, Frédéric Dubois, Bruno Giraudeau, Jean-Christophe Plantier, Alain Goudeau, Philippe Roingeard, and Denys Brand. Sequence and functional analysis of the envelope glycoproteins of hepatitis C virus variants selectively transmitted to a new host. *Journal of virology*, 87(24):13609–13618, December 2013.

Diego Darriba, Guillermo L Taboada, Ramón Doallo, and David Posada. jModelTest 2: more models, new heuristics and parallel computing. *Nature methods*, 9(8):772–772, August 2012.

Kari Debbink, John T McCrone, Joshua G Petrie, Rachel Truscon, Emileigh Johnson, Emily K Mantlo, Arnold S Monto, and Adam S Lauring. Vaccination has minimal impact on the intrahost diversity of H3N2 influenza viruses. *PLOS Pathogens*, 13(1):e1006194, January 2017.

Cristian Del Fabbro, Simone Scalabrin, Michele Morgante, and Federico M Giorgi. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLoS One*, 8(12):e85024–13, December 2013.

Julia Dietz, Sven-Eric Schelhorn, Daniel Fitting, Ulrike Mihm, Simone Susser, Martin-Walter Welker, Caterina Füller, Martin Däumer, Gerlinde Teuber, Heiner Wedemeyer, Thomas Berg, Thomas Lengauer, Stefan Zeuzem, Eva Herrmann, and Christoph Sarrazin. Deep sequencing reveals mutagenic effects of ribavirin during monotherapy of hepatitis C virus genotype 1-infected patients. *Journal of virology*, 87(11):6172–6181, June 2013.

Jorge M Dinis, Nicholas W Florek, Omolayo O Fatola, Louise H Moncla, James P Mutschler, Olivia K Charlier, Jennifer K Meece, Edward A Belongia, and Thomas C Friedrich. Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. *Journal of virology*, 90(7):3355–3365, 2016.

Appolinaire Djikeng, Rebecca Halpin, Ryan Kuzmickas, Jay DePasse, Jeremy Feldblyum, Naomi Sengamalay, Claudio Afonso, Xinsheng Zhang, Norman G Anderson, Elodie Ghedin, and David J Spiro. Viral genome sequencing by random priming methods. *BMC Genomics*, 9(1):5–9, 2008.

Juliane C Dohm, Claudio Lottaz, Tatiana Borodina, and Heinz Himmelbauer. Substantial biases in ultra-short read data sets from high-throughput DNA sequencing. *Nucleic acids research*, 36 (16):e105, September 2008.

Michael B Doud, Scott E Hensley, and Jesse D Bloom. Complete mapping of viral escape from neutralizing antibodies. *PLOS Pathogens*, 13(3):e1006271, March 2017.

David Duneau, Jean-Baptiste Ferdy, Jonathan Revah, Hannah Kondolf, Gerardo A Ortiz, Brian P Lazzaro, and Nicolas Buchon. Stochastic variation in the initial phase of bacterial infection predicts the probability of survival inD. melanogaster. *eLife*, 6, October 2017.

Robert C Edgar. MUSCLE: multiple sequence alignment with high accuracy and high throughput. *Nucleic acids research*, 32(5):1792–1797, 2004.

Charles T T Edwards, Edward C Holmes, Daniel J Wilson, Raphael P Viscidi, Elaine J Abrams, Rodney E Phillips, and Alexei J Drummond. Population genetic estimation of the loss of genetic diversity during horizontal transmission of HIV-1. *BMC Evolutionary Biology*, 6(1):28, March 2006.

S F Elena, R Sanjuan, A V Bordería, and P E Turner. Transmission bottlenecks and the evolution of fitness in rapidly evolving RNA viruses. *INFECTION, GENETICS AND EVOLUTION*, 1(1): 41–48, July 2001.

Warren J Ewens. *Mathematical Population Genetics*, volume 27 of *Interdisciplinary Applied Mathematics*. Springer New York, New York, NY, 2004.

Sébastien Fauteux-Daniel, Ariane Larouche, Virginie Calderon, Jonathan Boulais, Chanel Béland, Doris G Ransy, Marc Boucher, Valérie Lamarre, Normand Lapointe, Isabelle Boucoiran, Armelle Le Campion, and Hugo Soudeyns. Vertical transmission of hepatitis C virus: variable transmission bottleneck and evidence of mid-gestation in utero infection. *Journal of virology*, pages JVI.01372–17–49, September 2017.

P Flaherty, G Natsoulis, O Muralidharan, M Winters, J Buenrostro, J Bell, S Brown, M Holodniy, N Zhang, and H P Ji. Ultrasensitive detection of rare mutations using next-generation targeted resequencing. *Nucleic acids research*, 40(1):e2–e2, December 2011.

Brendan Flannery, Richard K Zimmerman, Larisa V Gubareva, Rebecca J Garten, Jessie R Chung, Mary Patricia Nowalk, Michael L Jackson, Lisa A Jackson, Arnold S Monto, Suzanne E Ohmit, Edward A Belongia, Huong Q McLean, Manjusha Gaglani, Pedro A Piedra, Vasiliy P Mishin, Anton P Chesnokov, Sarah Spencer, Swathi N Thaker, John R Barnes, Angie Foust, Wendy Sessions, Xiyan Xu, Jacqueline Katz, and Alicia M Fry. Enhanced Genetic Characterization of Influenza A(H3N2) Viruses and Vaccine Effectiveness by Genetic Group, 2014-2015. *The Journal of infectious diseases*, 214(7):1010–1019, October 2016.

Matthieu Foll, Yu-Ping Poh, Nicholas Renzette, Anna Ferrer-Admetlla, Claudia Bank, Hyunjin Shim, Anna-Sapfo Malaspinas, Gregory Ewing, Ping Liu, Daniel Wegmann, Daniel R Caffrey, Konstantin B Zeldovich, Daniel N Bolon, Jennifer P Wang, Timothy F Kowalik, Celia A Schiffer, Robert W Finberg, and Jeffrey D Jensen. Influenza Virus Drug Resistance: A Time-Sampled Population Genetics Perspective. *PLOS Genetics*, 10(2):e1004185–17, February 2014.

Matthieu Foll, Hyunjin Shim, and Jeffrey D Jensen. WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Molecular ecology resources*, 15(1):87–98, January 2015.

J M Fonville, S H Wilks, S L James, A Fox, M Ventresca, M Aban, L Xue, T C Jones, N M H Le, Q T Pham, N D Tran, Y Wong, A Mosterin, L C Katzelnick, D Labonte, T T Le, G van der Net, E Skepner, C A Russell, T D Kaplan, G F Rimmelzwaan, N Masurel, J C de Jong, A Palache, W E P Beyer, Q M Le, T H Nguyen, H F L Wertheim, A C Hurt, A D M E Osterhaus, I G Barr, R A M Fouchier, P W Horby, and D J Smith. Antibody landscapes after influenza virus infection or vaccination. *Science*, 346(6212):996–1000, November 2014.

Naomi L Forrester, Mathilde Guerbois, Robert L Seymour, Heidi Spratt, and Scott C Weaver. Vector-borne transmission imposes a severe bottleneck on an RNA virus population. *PLOS Pathogens*, 8(9):e1002897, September 2012.

Jemma L Geoghegan, Alistair M Senior, and Edward C Holmes. Pathogen population bottlenecks and adaptive landscapes: overcoming the barriers to disease emergence. *Proceedings. Biological sciences*, 283(1837):20160727–9, August 2016.

P GERBER, C G LOOSLI, and D HAMBRE. Antigenic variants of influenza A virus, PR8 strain. I. Their development during serial passage in the lungs of partially immune mice. *Journal of Experimental Medicine*, 101(6):627–638, June 1955.

P Gerber, D Hamre, and C G Loosli. Antigenic variants of influenza A virus (PR8 strain). II. Serological and immunological characteristics of variants derived from variants. *Journal of Experimental Medicine*, 103(4):413–424, April 1956.

Moritz Gerstung and Niko Beerenwinkel. *Calling subclonal mutations with deepSNV*, October 2015.

Moritz Gerstung, Christian Beisel, Markus Rechsteiner, Peter Wild, Peter Schraml, Holger Moch, and Niko Beerenwinkel. Reliable detection of subclonal single-nucleotide variants in tumour cell populations. *Nature Communications*, 3:811, 2012.

Moritz Gerstung, Elli Papaemmanuil, and Peter J Campbell. Subclonal variant calling with multiple samples and prior knowledge. *Bioinformatics (Oxford, England)*, 30(9):1198–1204, May 2014.

E Ghedin, A Fitch, A Boyne, S Griesemer, J DePasse, J Bera, X Zhang, R A Halpin, M Smit, L Jennings, K St George, E C Holmes, and D J Spiro. Mixed Infection and the Genesis of Influenza Virus Diversity. *Journal of virology*, 83(17):8832–8841, August 2009.

E Ghedin, J Laplante, J DePasse, D E Wentworth, R P Santos, M L Lepow, J Porter, K Stellrecht, X Lin, D Operario, S Griesemer, A Fitch, R A Halpin, T B Stockwell, D J Spiro, E C Holmes, and K S George. Deep Sequencing Reveals Mixed Infection with 2009 Pandemic Influenza A (H1N1) Virus Strains and the Emergence of Oseltamivir Resistance. *Journal of Infectious Diseases*, 203 (2):168–174, December 2010.

Elodie Ghedin, Edward C Holmes, Jay V DePasse, Lady Tatiana Pinilla, Adam Fitch, Marie-Eve Hamelin, Jesse Papenburg, and Guy Boivin. Presence of oseltamivir-resistant pandemic A/H1N1 minor variants before drug therapy with subsequent selection and transmission. *The Journal of infectious diseases*, 206(10):1504–1511, November 2012.

Stephen K Gire, Augustine Goba, Kristian G Andersen, Rachel S G Sealfon, Daniel J Park, Lansana Kanneh, Simbirie Jalloh, Mambu Momoh, Mohamed Fullah, Gytis Dudas, Shirlee Wohl, Lina M Moses, Nathan L Yozwiak, Sarah Winnicki, Christian B Matranga, Christine M Malboeuf, James Qu, Adrianne D Gladden, Stephen F Schaffner, Xiao Yang, Pan-Pan Jiang, Mahan Nekoui, Andres Colubri, Moinya Ruth Coomber, Mbalu Fonnie, Alex Moigboi, Michael Gbakie, Fatima K Kamara, Veronica Tucker, Edwin Konuwa, Sidiki Saffa, Josephine Sellu, Abdul Azziz Jalloh, Alice Kovoma, James Koninga, Ibrahim Mustapha, Kandeh Kargbo, Momoh Foday, Mohamed Yillah, Franklyn Kanneh, Willie Robert, James L B Massally, Sinéad B Chapman, James Bochicchio, Cheryl Murphy, Chad Nusbaum, Sarah Young, Bruce W Birren, Donald S Grant, John S Scheiffelin, Eric S Lander, Christian Happi, Sahr M Gevao, Andreas Gnirke, Andrew Rambaut, Robert F Garry, S Humarr Khan, and Pardis C Sabeti. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*, 345(6202):1369–1372, September 2014.

Nathan D Grubaugh, Darci R Smith, Doug E Brackney, Angela M Bosco-Lauth, Joseph R Fauver, Corey L Campbell, Todd A Felix, Hannah Romo, Nisha K Duggal, Elizabeth A Dietrich, Tyler Eike, Jennifer E Beane, Richard A Bowen, William C Black, Aaron C Brault, and Gregory D Ebel. Experimental evolution of an RNA virus in wild birds: evidence for host-dependent impacts on population structure and competitive fitness. *PLOS Pathogens*, 11(5):e1004874, May 2015.

Larisa V Gubareva, Laurent Kaiser, Mikhail N Matrosovich, Yee Soo-Hoo, and Frederick G Hayden. Selection of Influenza Virus Mutants in Experimentally Infected Volunteers Treated with Oseltamivir. *Journal of Infectious Diseases*, 183(4):523–531, February 2001.

Serafin Gutiérrez, Yannis Michalakis, and Stéphane Blanc. Virus population bottlenecks during within-host progression and host-to-host transmission. *Current Opinion in Virology*, 2(5):546–555, October 2012.

Richard E Haaland, Paulina A Hawkins, Jesus Salazar-Gonzalez, Amber Johnson, Amanda Tichacek, Etienne Karita, Olivier Manigart, Joseph Mulenga, Brandon F Keele, George M Shaw, Beatrice H Hahn, Susan A Allen, Cynthia A Derdeyn, and Eric Hunter. Inflammatory Genital Infections Mitigate a Severe Genetic Bottleneck in Heterosexual Transmission of Subtype A and C HIV-1. *PLOS Pathogens*, 5(1):e1000274, January 2009.

D Hamre, C G Loosli, and P Gerber. Antigenic variants of influenza A virus (PR8 strain) III. Serological relationships of a line of variants derived in sequence in mice given homologous vaccine. *Journal of Experimental Medicine*, 107(6):829–844, June 1958.

S Herfst, E J A Schrauwen, M Linster, S Chutinimitkul, E de Wit, V J Munster, E M Sorrell, T M Bestebroer, D F Burke, D J Smith, G F Rimmelzwaan, A D M E Osterhaus, and R A M Fouchier. Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets. *Science*, 336 (6088):1534–1541, June 2012.

Cynthia K Y Ho, Jayna Raghwani, Sylvie Koekkoek, Richard H Liang, Jan T M Van der Meer, Marc Van Der Valk, Menno De Jong, Oliver G Pybus, Janke Schinkel, and Richard Molenkamp. Characterization of Hepatitis C Virus (HCV) Envelope Diversification from Acute to Chronic Infection within a Sexually Transmitted HCV Cluster by Using Single-Molecule, Real-Time Sequencing. *Journal of virology*, 91(6):e02262–16, March 2017.

Karin Hoelzer, Pablo R Murcia, Gregory J Baillie, James L N Wood, Stephan M Metzger, Nikolaus Osterrieder, Edward J Dubovi, Edward C Holmes, and Colin R Parrish. Intrahost evolutionary dynamics of canine influenza virus in naive and partially immune dogs. *Journal of virology*, 84 (10):5329–5335, May 2010.

E Hoffmann, J Stech, Y Guan, R G Webster, and D R Perez. Universal primer set for the full-length amplification of all influenza A viruses. *Archives of Virology*, 146(12):2275–2289, December 2001.

Erich Hoffmann, Kutubuddin Mahmood, Chin-Fen Yang, Robert G Webster, Harry B Greenberg, and George Kemble. Rescue of influenza B virus from eight plasmids. *Proceedings of the National Academy of Sciences*, 99(17):11411–11416, August 2002.

E C Holmes, E Ghedin, R A Halpin, T B Stockwell, X Q Zhang, R Fleming, R Davey, C A Benson, S Mehta, R Taplitz, Y T Liu, K C Brouwer, D E Wentworth, X Lin, the INSIGHT FLU002 Study Group, and R T Schooley. Extensive Geographical Mixing of 2009 Human H1N1 Influenza A Virus in a Single University Community. *Journal of virology*, 85(14):6923–6929, June 2011.

Edward C Holmes. RNA virus genomics: a world of possibilities. *Journal of Clinical Investigation*, 119(9):2488–2495, September 2009.

Joseph Hughes, Richard C Allen, Marc Baguelin, Katie Hampson, Gregory J Baillie, Debra Elton, J Richard Newton, Paul Kellam, James L N Wood, Edward C Holmes, and Pablo R Murcia. Transmission of Equine Influenza Virus during an Outbreak Is Characterized by Frequent Mixed Infections and Loose Transmission Bottlenecks. *PLOS Pathogens*, 8(12):e1003081, December 2012.

Christopher J R Illingworth, Andrej Fischer, and Ville Mustonen. Identifying Selection in the Within-Host Evolution of Influenza Using Viral Sequence Data. *PLoS Computational Biology*, 10(7):e1003755, July 2014.

M Iqbal, H Xiao, G Baillie, A Warry, S C Essen, B Londt, S M Brookes, I H Brown, and J W McCauley. Within-host variation of avian influenza viruses. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 364(1530):2739–2747, August 2009.

Ofer Isakov, Antonio V Bordería, David Golan, Amir Hamenahem, Gershon Celniker, Liron Yoffe, Hervé Blanc, Marco Vignuzzi, and Noam Shomron. Deep sequencing analysis of viral infection and evolution allows rapid and detailed characterization of viral mutant spectrum. *Bioinformatics (Oxford, England)*, 31(13):2141–2150, July 2015.

Cassandra B Jabara, Corbin D Jones, Jeffrey Roach, Jeffrey A Anderson, and Ronald Swanstrom. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proceedings of the National Academy of Sciences of the United States of America*, 108(50):20166–20171, December 2011.

Katherine E E Johnson, Timothy Song, Benjamin Greenbaum, and Elodie Ghedin. Getting the flu: 5 key facts about influenza virus evolution. *PLOS Pathogens*, 13(8):e1006450, August 2017.

Rowland R Kao, Daniel T Haydon, Samantha J Lycett, and Pablo R Murcia. Supersize me: how whole-genome sequencing and big data are transforming epidemiology. *Trends in Microbiology*, 22(5):282–291, May 2014.

Brandon F Keele, Elena E Giorgi, Jesus F Salazar-Gonzalez, Julie M Decker, Kimmy T Pham, Maria G Salazar, Chuanxi Sun, Truman Grayson, Shuyi Wang, Hui Li, Xiping Wei, Chunlai Jiang, Jennifer L Kirchherr, Feng Gao, Jeffery A Anderson, Li-Hua Ping, Ronald Swanstrom, Georgia D Tomaras, William A Blattner, Paul A Goepfert, J Michael Kilby, Michael S Saag, Eric L Delwart, Michael P Busch, Myron S Cohen, David C Montefiori, Barton F Haynes, Brian Gaschen, Gayathri S Athreya, Ha Y Lee, Natasha Wood, Cathal Seoighe, Alan S Perelson, Tanmoy Bhattacharya, Bette T Korber, Beatrice H Hahn, and George M Shaw. Identification and characterization of transmitted and early founder virus envelopes in primary HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America*, 105(21):7552–7557, May 2008.

Ben Killingley, Joanne Enstone, Robert Booy, Andrew Hayward, John Oxford, Neil Ferguson, and Jonathan Nguyen Van-Tam. Potential role of human challenge studies for investigation of influenza transmission. *The Lancet Infectious Diseases*, 11(11):879–886, November 2011.

M Kimura. SOLUTION OF A PROCESS OF RANDOM GENETIC DRIFT WITH A CONTINUOUS MODEL. *Proceedings of the National Academy of Sciences*, 41(3):144–150, March 1955.

M Kimura and T Ohta. The Average Number of Generations until Fixation of a Mutant Gene in a Finite Population. *Genetics*, 61(3):763–771, March 1969.

Motoo Kimura. Theoretical foundation of population genetics at the molecular level. *Theoretical Population Biology*, 2(2):174–208, June 1971.

J F C Kingman. On the Genealogy of Large Populations. *Journal of Applied Probability*, 19:27, 1982.

Martin Kircher, Udo Stenzel, and Janet Kelso. Improved base calling for the Illumina Genome Analyzer using machine learning strategies. *Genome biology*, 10(8):R83, 2009.

Daniel C Koboldt, Ken Chen, Todd Wylie, David E Larson, Michael D McLellan, Elaine R Mardis, George M Weinstock, Richard K Wilson, and Li Ding. VarScan: variant detection in massively parallel sequencing of individual and pooled samples. *Bioinformatics (Oxford, England)*, 25(17): 2283–2285, September 2009.

Daniel C Koboldt, Qunyuan Zhang, David E Larson, Dong Shen, Michael D McLellan, Ling Lin, Christopher A Miller, Elaine R Mardis, Li Ding, and Richard K Wilson. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome research*, 22(3):568–576, March 2012.

Daniel C Koboldt, David E Larson, and Richard K Wilson. Using VarScan 2 for Germline Variant Calling and Somatic Mutation Detection. *Current protocols in bioinformatics*, 44:15.4.1–15.4.17, December 2013.

Björn F Koel, David F Burke, Theo M Bestebroer, Stefan van der Vliet, Gerben C M Zondag, Gaby Vervaet, Eugene Skepner, Nicola S Lewis, Monique I J Spronken, Colin A Russell, Mikhail Y Eropkin, Aeron C Hurt, Ian G Barr, Jan C de Jong, Guus F Rimmelzwaan, Albert D M E Osterhaus, Ron A M Fouchier, and Derek J Smith. Substitutions near the receptor binding site determine major antigenic change during influenza virus evolution. *Science*, 342(6161):976–979, November 2013.

Katia Koelle and David A Rasmussen. The effects of a deleterious mutation load on patterns of influenza A/H3N2's antigenic evolution in humans. *eLife*, 4:e07361, September 2015.

Katia Koelle, Sarah Cobey, Bryan Grenfell, and Mercedes Pascual. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science*, 314(5807):1898–1903, December 2006.

Roger D Kouyos, Christian L Althaus, and Sebastian Bonhoeffer. Stochastic or deterministic: what is the effective population size of HIV-1? *Trends in Microbiology*, 14(12):507–511, December 2006.

Jeffrey R Kugelman, Johanny Kugelman-Tonos, Jason T Ladner, James Pettit, Carolyn M Keeton, Elyse R Nagle, Karla Y Garcia, Jeffrey W Froude, Ana I Kuehne, Jens H Kuhn, Sina Bavari, Larry Zeitlin, John M Dye, Gene G Olinger, Mariano Sanchez-Lockhart, and Gustavo F Palacios. Emergence of Ebola Virus Escape Variants in Infected Nonhuman Primates Treated with the MB-003 Antibody Cocktail. *CellReports*, 12(12):2111–2120, September 2015.

Jeffrey R Kugelman, Michael R Wiley, Elyse R Nagle, Daniel Reyes, Brad P Pfeffer, Jens H Kuhn, Mariano Sanchez-Lockhart, and Gustavo F Palacios. Error baseline rates of five sample preparation methods used to characterize RNA virus populations. *PLoS One*, 12(2):e0171333–13, February 2017a.

Jeffrey R Kugelman, Michael R Wiley, Elyse R Nagle, Daniel Reyes, Brad P Pfeffer, Jens H Kuhn, Mariano Sanchez-Lockhart, and Gustavo F Palacios. Error baseline rates of five sample preparation methods used to characterize RNA virus populations. *PLoS One*, 12(2):e0171333, February 2017b.

S Kundu, J Lockwood, D P Depledge, Y Chaudhry, A Aston, K Rao, J C Hartley, I Goodfellow, and J Breuer. Next-Generation Whole Genome Sequencing Identifies the Direction of Norovirus Transmission in Linked Patients. *Clinical Infectious Diseases*, 57(3):407–414, July 2013.

Seema S Lakdawala, Akila Jayaraman, Rebecca A Halpin, Elaine W Lamirande, Angela R Shih, Timothy B Stockwell, Xudong Lin, Ari Simenauer, Christopher T Hanson, Leatrice Vogel, Myeisha Paskel, Mahnaz Minai, Ian Moore, Marlene Orandle, Suman R Das, David E Wentworth, Ram Sasisekharan, and Kanta Subbarao. The soft palate is an important site of adaptation for transmissible influenza viruses. *Nature*, 526(7571):122–125, October 2015.

Hugo Y K Lam, Michael J Clark, Rui Chen, Rong Chen, Georges Natsoulis, Maeve O'Huallachain, Frederick E Dewey, Lukas Habegger, Euan A Ashley, Mark B Gerstein, Atul J Butte, Hanlee P Ji, and Michael Snyder. Performance comparison of whole-genome sequencing platforms. *Nature Biotechnology*, 30(1):78–82, December 2011.

Ben Langmead and Steven L Salzberg. Fast gapped-read alignment with Bowtie 2. *Nature methods*, 9(4):357–359, March 2012.

Ben Langmead, Cole Trapnell, Mihai Pop, and Steven L Salzberg. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome biology*, 10(3):R25, 2009.

L L H Lau, D K M Ip, H Nishiura, V J Fang, K H Chan, J S M Peiris, G M Leung, and B J Cowling. Heterogeneity in Viral Shedding Among Individuals With Medically Attended Influenza A Virus Infection. *Journal of Infectious Diseases*, 207(8):1281–1285, March 2013.

Michael Lauck, Mónica V Alvarado-Mora, Ericka A Becker, Dipankar Bhattacharya, Rob Striker, Austin L Hughes, Flair J Carrilho, David H O'Connor, and João R Rebello Pinho. Analysis of hepatitis C virus intrahost diversity across the coding region by ultradeep pyrosequencing. *Journal of virology*, 86(7):3952–3960, April 2012.

Adam S Lauring, Ashley Acevedo, Samantha B Cooper, and Raul Andino. Codon usage determines the mutational robustness, evolutionary capacity, and virulence of an RNA virus. *Cell host & microbe*, 12(5):623–632, November 2012.

Adam S Lauring, Judith Frydman, and Raul Andino. The role of mutational robustness in RNA virus evolution. *Nature Reviews Microbiology*, 11(5):327–336, May 2013.

Min-Shi Lee and Jack Si-En Chen. Predicting Antigenic Variants of Influenza A/H3N2 Viruses. *Emerging Infectious Diseases*, 10(8):1385–1390, August 2004.

Nelson Lee, Paul K S Chan, David S C Hui, Timothy H Rainer, Eric Wong, Kin-Wing Choi, Grace C Y Lui, Bonnie C K Wong, Rita Y K Wong, Wai-Yip Lam, Ida M T Chu, Raymond W M Lai, Clive S Cockram, and Joseph J Y Sung. Viral Loads and Duration of Viral Shedding in Adult Patients Hospitalized with Influenza. *Journal of Infectious Diseases*, 200(4):492–500, August 2009.

Sebastian Lequime, Albin Fontaine, Meriadeg Ar Gouilh, Isabelle Moltini-Conclois, and Louis Lambrechts. Genetic Drift, Purifying Selection and Vector Genotype Shape Dengue Virus Intra-host Genetic Diversity in Mosquitoes. *PLOS Genetics*, 12(6):e1006111, June 2016.

Hui Li, Mark B Stoddard, Shuyi Wang, Lily M Blair, Elena E Giorgi, Erica H Parrish, Gerald H Learn, Peter Hraber, Paul A Goepfert, Michael S Saag, Thomas N Denny, Barton F Haynes, Beatrice H Hahn, Ruy M Ribeiro, Alan S Perelson, Bette T Korber, Tanmoy Bhattacharya, and George M Shaw. Elucidation of hepatitis C virus transmission and early diversification by single genome sequencing. *PLOS Pathogens*, 8(8):e1002880, 2012.

Marta Łuksza and Michael Lässig. A predictive fitness model for influenza. *Nature Publishing Group*, 507(7490):57, March 2014.

Alexander R Macalalad, Michael C Zody, Patrick Charlebois, Niall J Lennon, Ruchi M Newman, Christine M Malboeuf, Elizabeth M Ryan, Christian L Boutwell, Karen A Power, Doug E Brackney, Kendra N Pesko, Joshua Z Levin, Gregory D Ebel, Todd M Allen, Bruce W Birren, and Matthew R Henn. Highly sensitive and specific detection of rare variants in mixed viral populations from massively parallel sequence data. *PLoS Computational Biology*, 8(3):e1002417, 2012.

John T McCrone and Adam S Lauring. Measurements of intrahost viral diversity are extremely sensitive to systematic errors in variant calling. *Journal of virology*, 90(15):JVI.00667–16–6895, May 2016.

T Miyata and T Yasunaga. Molecular evolution of mRNA: a method for estimating evolutionary rates of synonymous and amino acid substitutions from homologous nucleotide sequences and its application. *Journal of molecular evolution*, 16(1):23–36, September 1980.

Arnold S Monto, Suzanne E Ohmit, Joshua G Petrie, Emileigh Johnson, Rachel Truscon, Esther Teich, Judy Rotthoff, Matthew Boulton, and John C Victor. Comparative efficacy of inactivated and live attenuated influenza vaccines. *New England Journal of Medicine*, 361(13):1260–1267, September 2009.

Arnold S Monto, Ryan E Malosh, Joshua G Petrie, Mark G Thompson, and Suzanne E Ohmit. Frequency of acute respiratory illnesses and circulation of respiratory viruses in households with children over 3 surveillance seasons. *The Journal of infectious diseases*, 210(11):1792–1799, December 2014.

Arnold S Monto, Joshua G Petrie, Rachel T Cross, Emileigh Johnson, Merry Liu, Weimin Zhong, Min Levine, Jacqueline M Katz, and Suzanne E Ohmit. Antibody to Influenza Virus Neuraminidase: An Independent Correlate of Protection. *The Journal of infectious diseases*, 212(8): 1191–1199, October 2015.

Benoît Moury, Frédéric Fabre, and Rachid Senoussi. Estimation of the number of virus particles transmitted by an insect vector. *Proceedings of the National Academy of Sciences*, 104(45): 17891–17896, November 2007.

Pablo R Murcia, Gregory J Baillie, Janet Daly, Debra Elton, Carley Jervis, Jennifer A Mumford, Richard Newton, Colin R Parrish, Karin Hoelzer, Gordon Dougan, Julian Parkhill, Nicola Lennard, Doug Ormond, Sharon Moule, Andrew Whitwham, John W McCauley, Trevelyan J McKinley, Edward C Holmes, Bryan T Grenfell, and James L N Wood. Intra- and interhost evolutionary dynamics of equine influenza virus. *Journal of virology*, 84(14):6943–6954, July 2010.

Pablo R Murcia, Joseph Hughes, Patrizia Battista, Lucy Lloyd, Gregory J Baillie, Ricardo H Ramirez-Gonzalez, Doug Ormond, Karen Oliver, Debra Elton, Jennifer A Mumford, Mario Caccamo, Paul Kellam, Bryan T Grenfell, Edward C Holmes, and James L N Wood. Evolution of an Eurasian Avian-like Influenza Virus in Naïve and Vaccinated Pigs. *PLOS Pathogens*, 8(5): e1002730, May 2012.

Pablo R Murcia, Gregory J Baillie, J Conrad Stack, Carley Jervis, Debra Elton, Jennifer A Mumford, Paul Kellam, Bryan T Grenfell, Edward C Holmes, and James L N Wood. Evolution of equine influenza virus in vaccinated horses. *Journal of virology*, 87(8):4768–4771, April 2013.

Simon Myers, Charles Fefferman, and Nick Patterson. Can one learn history from the allelic spectrum? *Theoretical Population Biology*, 73(3):342–348, May 2008.

Kensuke Nakamura, Taku Oshima, Takuya Morimoto, Shun Ikeda, Hirofumi Yoshikawa, Yuh Shiwa, Shu Ishikawa, Margaret C Linak, Aki Hirai, Hiroki Takahashi, Md Altaf-Ul-Amin, Naotake Ogasawara, and Shigehiko Kanaya. Sequence-specific error profile of Illumina sequencers. *Nucleic acids research*, 39(13):e90–e90, July 2011.

Richard A Neher and Trevor Bedford. nextflu: real-time tracking of seasonal influenza virus evolution in humans. *Bioinformatics (Oxford, England)*, 31(21):3546–3548, October 2015.

Martha I Nelson and Edward C Holmes. The evolution of epidemic influenza. *Nature Reviews Genetics*, 8(3):196–205, January 2007.

Martha I Nelson, Lone Simonsen, Cécile Viboud, Mark A Miller, Jill Taylor, Kirsten St George, Sara B Griesemer, Elodie Ghedin, Naomi A Sengamalay, David J Spiro, Igor Volkov, Bryan T Grenfell, David J Lipman, Jeffery K Taubenberger, and Edward C Holmes. Stochastic Processes Are Key Determinants of Short-Term Evolution in Influenza A Virus. *PLOS Pathogens*, 2(12): e125, December 2006.

Martha I Nelson, Angel Balmaseda, Guillermina Kuan, Saira Saborio, Xudong Lin, Rebecca A Halpin, Timothy B Stockwell, David E Wentworth, Eva Harris, and Aubree Gordon. The evolutionary dynamics of influenza A and B viruses in the tropical city of Managua, Nicaragua. *Virology*, 462-463:81–90, August 2014.

Rasmus Nielsen, Joshua S Paul, Anders Albrechtsen, and Yun S Song. Genotype and SNP calling from next-generation sequencing data. *Nature Publishing Group*, 12(6):443–451, June 2011.

I S Novella, S F Elena, A Moya, E Domingo, and J J Holland. Size of genetic bottlenecks leading to virus fitness loss is determined by mean initial population fitness. *Journal of virology*, 69(5): 2869–2872, May 1995.

Isabel S Novella, Ranendra N Dutta, and Claus O Wilke. A linear relationship between fitness and the logarithm of the critical bottleneck size in vesicular stomatitis virus populations. *Journal of virology*, 82(24):12589–12590, December 2008.

S E Ohmit, J G Petrie, R E Malosh, A M Fry, M G Thompson, and A S Monto. Influenza Vaccine Effectiveness in Households With Children During the 2012-2013 Season: Assessments of Prior Vaccination and Serologic Susceptibility. *Journal of Infectious Diseases*, 211(10):1519–1528, April 2015.

Suzanne E Ohmit, John C Victor, Judy R Rotthoff, Esther R Teich, Rachel K Truscon, Laura L Baum, Bhavya Rangarajan, Duane W Newton, Matthew L Boulton, and Arnold S Monto. Prevention of antigenically drifted influenza by inactivated and live attenuated vaccines. *New England Journal of Medicine*, 355(24):2513–2522, December 2006.

Suzanne E Ohmit, John C Victor, Esther R Teich, Rachel K Truscon, Judy R Rotthoff, Duane W Newton, Sarah A Campbell, Matthew L Boulton, and Arnold S Monto. Prevention of symptomatic seasonal influenza in 2005-2006 by inactivated and live attenuated vaccines. *Journal of Infectious Diseases*, 198(3):312–317, August 2008.

Suzanne E Ohmit, Joshua G Petrie, Rachel T Cross, Emileigh Johnson, and Arnold S Monto. Influenza hemagglutination-inhibition antibody titer as a correlate of vaccine-induced protection. *The Journal of infectious diseases*, 204(12):1879–1885, December 2011.

Suzanne E Ohmit, Mark G Thompson, Joshua G Petrie, Swathi N Thaker, Michael L Jackson, Edward A Belongia, Richard K Zimmerman, Manjusha Gaglani, Lois Lamerato, Sarah M Spencer, Lisa Jackson, Jennifer K Meece, Mary Patricia Nowalk, Juhee Song, Marcus Zervos, Po-Yung Cheng, Charles R Rinaldo, Lydia Clipper, David K Shay, Pedro Piedra, and Arnold S Monto. Influenza vaccine effectiveness in the 2011-2012 season: protection against each circulating virus and the effect of prior vaccination on estimates. *Clinical infectious diseases : an official publication of the Infectious Diseases Society of America*, 58(3):319–327, February 2014.

Suzanne E Ohmit, Joshua G Petrie, Ryan E Malosh, Emileigh Johnson, Rachel Truscon, Barbara Aaron, Casey Martens, Caroline Cheng, Alicia M Fry, and Arnold S Monto. Substantial Influenza Vaccine Effectiveness in Households With Children During the 2013–2014 Influenza Season, When 2009 Pandemic Influenza A(H1N1) Virus Predominated. *Journal of Infectious Diseases*, 213(8): 1229–1236, March 2016.

Landon N Olp, Adrien Jeanniard, Clemence Marimo, John T West, and Charles Wood. Whole-genome sequencing of KSHV from Zambian Kaposi's sarcoma biopsies reveals unique viral diversity. *Journal of virology*, pages JVI.01712–15–10, September 2015.

Michael T Osterholm, Nicholas S Kelley, Alfred Sommer, and Edward A Belongia. Efficacy and effectiveness of influenza vaccines: a systematic review and meta-analysis. *The Lancet Infectious Diseases*, 12(1):36–44, January 2012.

J D Parvin, A Moscona, W T Pan, J M Leider, and P Palese. Measurement of the mutation rates of animal viruses: influenza A virus and poliovirus type 1. *Journal of virology*, 59(2):377–383, August 1986.

Matthew D Pauly and Adam S Lauring. Effective Lethal Mutagenesis of Influenza Virus by Three Nucleoside Analogs. *Journal of virology*, 89(7):3584–3597, March 2015.

Matthew D Pauly, Megan C Procario, and Adam S Lauring. A novel twelve class fluctuation test reveals higher than expected mutation rates for influenza A viruses. *eLife*, 6:686, June 2017.

Kayla M Peck, Carmen H S Chan, and Mark M Tanaka. Connecting within-host dynamics to the rate of viral molecular evolution. *Virus Evolution*, 1(1), March 2015.

Joshua G Petrie, Suzanne E Ohmit, Emileigh Johnson, Rachel T Cross, and Arnold S Monto. Efficacy studies of influenza vaccines: effect of end points used and characteristics of vaccine failures. *The Journal of infectious diseases*, 203(9):1309–1315, May 2011.

Joshua G Petrie, Suzanne E Ohmit, Benjamin J Cowling, Emileigh Johnson, Rachel T Cross, Ryan E Malosh, Mark G Thompson, and Arnold S Monto. Influenza transmission in a cohort of households with children: 2010-2011. *PLoS One*, 8(9):e75339, 2013.

Joshua G Petrie, Suzanne E Ohmit, Rachel Truscon, Emileigh Johnson, Thomas M Braun, Min Z Levine, Maryna C Eichelberger, and Arnold S Monto. Modest Waning of Influenza Vaccine Efficacy and Antibody Titers During the 2007-2008 Influenza Season. *The Journal of infectious diseases*, 214(8):1142–1149, October 2016.

Joshua G Petrie, Marisa C Eisenberg, Sophia Ng, Ryan E Malosh, Kyu Han Lee, Suzanne E Ohmit, and Arnold S Monto. Application of an Individual-Based Transmission Hazard Model for Estimation of Influenza Vaccine Effectiveness in a Household Cohort. *American Journal of Epidemiology*, pages 1–9, June 2017.

Velislava N Petrova and Colin A Russell. The evolution of seasonal influenza viruses. *Nature Reviews Microbiology*, pages 1–14, October 2017.

Julie K Pfeiffer and Karla Kirkegaard. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLOS Pathogens*, 1(2):e11, October 2005.

Julie K Pfeiffer and Karla Kirkegaard. Bottleneck-mediated quasispecies restriction during spread of an RNA virus from inoculation site to brain. *Proceedings of the National Academy of Sciences*, 103(14):5520–5525, April 2006.

Leo L M Poon, Timothy Song, Roni Rosenfeld, Xudong Lin, Matthew B Rogers, Bin Zhou, Robert Sebra, Rebecca A Halpin, Yi Guan, Alan Twaddle, Jay V DePasse, Timothy B Stockwell, David E Wentworth, Edward C Holmes, Benjamin Greenbaum, Joseph S M Peiris, Benjamin J Cowling, and Elodie Ghedin. Quantifying influenza virus diversity and transmission in humans. *Nature genetics*, 48(2):195–200, February 2016.

C W Potter and J S Oxford. Determinants of immunity to influenza infection in man. *British medical bulletin*, 35(1):69–75, January 1979.

Morgan N Price, Paramvir S Dehal, and Adam P Arkin. FastTree 2–approximately maximum-likelihood trees for large alignments. *PloS one*, 5(3):e9490, March 2010.

Oliver G Pybus, Andrew Rambaut, Robert Belshaw, Robert P Freckleton, Alexei J Drummond, and Edward C Holmes. Phylogenetic Evidence for Deleterious Mutation Load in RNA Viruses and Its Contribution to Viral Evolution. *Molecular Biology and Evolution*, 24(3):845–852, January 2007.

Andrew Rambaut, Oliver G Pybus, Martha I Nelson, Cécile Viboud, Jeffery K Taubenberger, and Edward C Holmes. The genomic and epidemiological dynamics of human influenza A virus. *Nature*, 453(7195):615–619, May 2008.

Kimberly Robasky, Nathan E Lewis, and George M Church. The role of replicates for error mitigation in next-generation sequencing. *Nature Reviews Genetics*, 15(1):56–62, January 2014.

A Robertson. A Theory of Limits in Artificial Selection. *Proceedings of the Royal Society B: Biological Sciences*, 153(951):234–249, November 1960.

Matthew B Rogers, Timothy Song, Robert Sebra, Benjamin D Greenbaum, Marie-Eve Hamelin, Adam Fitch, Alan Twaddle, Lijia Cui, Edward C Holmes, Guy Boivin, and Elodie Ghedin. Intrahost dynamics of antiviral resistance in influenza a virus reflect complex patterns of segment linkage, reassortment, and natural selection. *mBio*, 6(2):e02464–14, 2015.

Toon Rosseel, Steven Van Borm, Frank Vandenbussche, Bernd Hoffmann, Thierry van den Berg, Martin Beer, and Dirk Höper. The Origin of Biased Sequence Depth in Sequence-Independent Nucleic Acid Amplification and Optimization for Efficient Massive Parallel Sequencing. *PLoS One*, 8(9):e76144–9, September 2013.

I M Rouzine, A Rodrigo, and J M Coffin. Transition between stochastic evolution and deterministic evolution in the presence of selection: general theory and application to virology. *Microbiology and Molecular Biology Reviews*, 65(1):151–185, March 2001.

Daniel E Rozen, Michelle G J L Habets, Andreas Handel, and J Arjan G M de Visser. Heterogeneous adaptive trajectories of small populations on complex fitness landscapes. *PLoS One*, 3(3):e1715, March 2008.

Colin A Russell, Judith M Fonville, André E X Brown, David F Burke, David L Smith, Sarah L James, Sander Herfst, Sander van Boheemen, Martin Linster, Eefje J Schrauwen, Leah Katzelnick, Ana Mosterín, Thijs Kuiken, Eileen Maher, Gabriele Neumann, Albert D M E Osterhaus, Yoshihiro Kawaoka, Ron A M Fouchier, and Derek J Smith. The potential for respiratory droplet-transmissible A/H5N1 influenza virus to evolve in a mammalian host. *Science*, 336(6088): 1541–1547, June 2012.

Simon P Sadedin, Bernard Pope, and Alicia Oshlack. Bpipe: a tool for running and managing bioinformatics pipelines. *Bioinformatics (Oxford, England)*, 28(11):1525–1526, June 2012.

Manish Sagar, Ludo Lavreys, Jared M Baeten, Barbra A Richardson, Kishorchandra Mandaliya, Jeckoniah O Ndinya-Achola, Joan K Kreiss, and Julie Overbaugh. Identification of modifiable factors that affect the genetic diversity of the transmitted HIV-1 population. *AIDS (London, England)*, 18(4):615, March 2004.

Rafael Sanjuán. From Molecular Genetics to Phylodynamics: Evolutionary Relevance of Mutation Rates Across Viruses. *PLOS Pathogens*, 8(5):e1002685, May 2012.

Rafael Sanjuán, Miguel R Nebot, Nicola Chirico, Louis M Mansky, and Robert Belshaw. Viral mutation rates. *Journal of virology*, 84(19):9733–9748, October 2010.

M Schirmer, U Z Ijaz, R D'Amore, N Hall, W T Sloan, and C Quince. Insight into biases and sequencing errors for amplicon sequencing with the Illumina MiSeq platform. *Nucleic acids research*, 43(6):e37–e37, March 2015.

Michael W Schmitt, Scott R Kennedy, Jesse J Salk, Edward J Fox, Joseph B Hiatt, and Lawrence A Loeb. Detection of ultra-rare mutations by next-generation sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, 109(36):14508–14513, September 2012.

H E Simmons, J P Dunham, J C Stack, B J A Dickins, I Pagán, E C Holmes, and A G Stephenson. Deep sequencing reveals persistence of intra- and inter-host genetic diversity in natural and greenhouse populations of zucchini yellow mosaic virus. *The Journal of general virology*, 93(Pt 8):1831–1840, August 2012.

H E Simmons, J P Dunham, K E Zinn, G P Munkvold, E C Holmes, and A G Stephenson. Zucchini yellow mosaic virus (ZYMV, Potyvirus): vertical transmission, seed infection and cryptic infections. *Virus Research*, 176(1-2):259–264, September 2013.

J J Skehel, E Brown, R S Daniels, A R Douglas, M KNOSSOW, I A Wilson, N G WRIGLEY, and D C Wiley. Studies with monoclonal antibodies prepared against X-31 influenza virus haemagglutinin. *Biochemical Society Transactions*, 13(1):12–14, February 1985.

Danuta M Skowronski, Catharine Chambers, Suzana Sabaiduc, Gaston De Serres, Anne-Luise Winter, James A Dickinson, Jonathan Gubbay, Kevin Fonseca, Hugues Charest, Mel Krajden, Martin Petric, Salaheddin M Mahmud, Paul Van Caeseele, Nathalie Bastien, Alireza Eshaghi, and Yan Li. Integrated Sentinel Surveillance Linking Genetic, Antigenic, and Epidemiologic Monitoring of Influenza Vaccine-Virus Relatedness and Effectiveness During the 2013–2014 Influenza Season. *Journal of Infectious Diseases*, 212(5):726–739, August 2015.

Darci R Smith, A Paige Adams, Joan L Kenney, Eryu Wang, and Scott C Weaver. Venezuelan equine encephalitis virus in the mosquito vector Aedes taeniorhynchus: infection initiated by a small number of susceptible epithelial cells and a population bottleneck. *Virology*, 372(1):176–186, March 2008.

Derek J Smith, Alan S Lapedes, Jan C de Jong, Theo M Bestebroer, Guus F Rimmelzwaan, Albert D M E Osterhaus, and Ron A M Fouchier. Mapping the antigenic and genetic evolution of influenza virus. *Science*, 305(5682):371–376, July 2004.

Ashley Sobel Leonard, Micah T McClain, Gavin J D Smith, David E Wentworth, Rebecca A Halpin, Xudong Lin, Amy Ransier, Timothy B Stockwell, Suman R Das, Anthony S Gilbert, Robert Lambkin-Williams, Geoffrey S Ginsburg, Christopher W Woods, and Katia Koelle. Deep Sequencing of Influenza A Virus from a Human Challenge Study Reveals a Selective Bottleneck and Only Limited Intrahost Genetic Diversification. *Journal of virology*, 90(24):11247–11258, December 2016.

Ashley Sobel Leonard, Micah T McClain, Gavin J D Smith, David E Wentworth, Rebecca A Halpin, Xudong Lin, Amy Ransier, Timothy B Stockwell, Suman R Das, Anthony S Gilbert, Rob Lambkin-Williams, Geoffrey S Ginsburg, Christopher W Woods, Katia Koelle, and Christopher J R Illingworth. The effective rate of influenza reassortment is limited during human infection. *PLOS Pathogens*, 13(2):e1006203, February 2017a.

Ashley Sobel Leonard, Daniel B Weissman, Benjamin Greenbaum, Elodie Ghedin, and Katia Koelle. Transmission Bottleneck Size Estimation from Pathogen Deep-Sequencing Data, with an Application to Human Influenza A Virus. *Journal of virology*, 91(14):e00171–17, July 2017b.

Alexandros Stamatakis. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics (Oxford, England)*, 30(9):1312–1313, May 2014.

Minoru Takeshita, Naofumi Shigemune, Kenji Kikuhara, Naruto Furuya, and Yoichi Takanami. Spatial analysis for exclusive interactions between subgroups I and II of Cucumber mosaic virus in cowpea. *Virology*, 328(1):45–51, October 2004.

Aya Takeyama, Koichi Hashimoto, Masatoki Sato, Ryoko Kawashima, Yukihiko Kawasaki, and Mitsuaki Hosoya. Respiratory syncytial virus shedding by children hospitalized with lower respiratory tract infection. *Journal of medical virology*, 88(6):938–946, June 2016.

H Tao, J Steel, and A C Lowen. Intrahost Dynamics of Influenza Virus Reassortment. *Journal of virology*, 88(13):7485–7492, June 2014.

P F M Teunis, F H A Sukhrie, H Vennema, J Bogerman, M F C Beersma, and M P G Koopmans. Shedding of norovirus in symptomatic and asymptomatic infections. *Epidemiology and Infection*, 143(8):1710–1717, June 2015.

Tim K Tsang, Simon Cauchemez, Ranawaka A P M Perera, Guy Freeman, Vicky J Fang, Dennis K M Ip, Gabriel M Leung, Joseph Sriyal Malik Peiris, and Benjamin J Cowling. Association between antibody titers and protection against influenza virus infection within households. *The Journal of infectious diseases*, 210(5):684–692, September 2014.

Tim K Tsang, Benjamin J Cowling, Vicky J Fang, Kwok-Hung Chan, Dennis K M Ip, Gabriel M Leung, J S Malik Peiris, and Simon Cauchemez. Influenza A Virus Shedding and Infectivity in Households. *The Journal of infectious diseases*, 212(9):1420–1428, November 2015.

T M Tumpey, M Renshaw, J D Clements, and J M Katz. Mucosal delivery of inactivated influenza vaccine induces B-cell-dependent heterosubtypic cross-protection against lethal influenza A H5N1 virus infection. *Journal of virology*, 75(11):5141–5150, June 2001.

Greta A Van Slyke, Jamie J Arnold, Alex J Lugo, Sara B Griesemer, Ibrahim M Moustafa, Laura D Kramer, Craig E Cameron, and Alexander T Ciota. Sequence-Specific Fidelity Alterations Associated with West Nile Virus Attenuation in Mosquitoes. *PLOS Pathogens*, 11(6):e1005009–21, June 2015.

Andrew Varble, Randy A Albrecht, Simone Backes, Marshall Crumiller, Nicole M Bouvier, David Sachs, Adolfo García-Sastre, and Benjamin R tenOever. Influenza A virus transmission bottlenecks are defined by infection route and recipient host. *Cell host & microbe*, 16(5):691–700, November 2014.

Marco Vignuzzi, Jeffrey K Stone, Jamie J Arnold, Craig E Cameron, and Raul Andino. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature*, 439(7074):344–348, January 2006.

Elisa Visher, Shawn E Whitefield, John T McCrone, William Fitzsimmons, and Adam S Lauring. The Mutational Robustness of Influenza A Virus. *PLOS Pathogens*, 12(8):e1005856, August 2016.

P B Visser, D J Brown, F T Brederode, and J F Bol. Nematode transmission of tobacco rattle virus serves as a bottleneck to clear the virus population from defective interfering RNAs. *Virology*, 263(1):155–165, October 1999.

Xin Victoria Wang, Natalie Blades, Jie Ding, Razvan Sultana, and Giovanni Parmigiani. Estimation of sequencing error rates in short reads. *BMC bioinformatics*, 13(1):185, 2012.

S J Watson, M R A Welkers, D P Depledge, E Coulter, J M Breuer, M D de Jong, and P Kellam. Viral population analysis and minority-variant detection using short read next-generation sequencing. *Philosophical Transactions of the Royal Society B: Biological Sciences*, 368(1614): 20120205–20120205, February 2013.

D C Wiley, I A Wilson, and J J Skehel. Structural identification of the antibody-binding sites of Hong Kong influenza haemagglutinin and their involvement in antigenic variation. *Nature*, 1981.

Peter R Wilker, Jorge M Dinis, Gabriel Starrett, Masaki Imai, Masato Hatta, Chase W Nelson, David H O'Connor, Austin L Hughes, Gabriele Neumann, Yoshihiro Kawaoka, and Thomas C Friedrich. Selection on haemagglutinin imposes a bottleneck during mammalian transmission of reassortant H5N1 influenza viruses. *Nature Communications*, 4:2636, 2013.

Yvonne Willi, Josh Van Buskirk, and Ary A Hoffmann. Limits to the Adaptive Potential of Small Populations. *Annual Review of Ecology, Evolution, and Systematics*, 37(1):433–458, December 2006.

E G Williamson and M Slatkin. Using maximum likelihood to estimate population size from temporal changes in allele frequencies. *Genetics*, 152(2):755–761, June 1999.

Andreas Wilm, Pauline Poh Kim Aw, Denis Bertrand, Grace Hui Ting Yeo, Swee Hoe Ong, Chang Hua Wong, Chiea Chuen Khor, Rosemary Petric, Martin Lloyd Hibberd, and Niranjan Nagarajan. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. *Nucleic acids research*, 40(22):11189–11201, December 2012.

World Health Organization. WHO — World Health Organization, 2014. URL `http://www.who.int/mediacentre/factsheets/fs211/en/`.

Chris Wymant, Matthew Hall, Oliver Ratmann, David Bonsall, Tanya Golubchik, Mariateresa de Cesare, Astrid Gall, Marion Cornelissen, Christophe Fraser, and STOP-HCV Consortium, The Maela Pneumococcal Collaboration, and The BEEHIVE Collaboration. PHYLOSCANNER: Inferring Transmission from Within- and Between-Host Pathogen Genetic Diversity. *Molecular Biology and Evolution*, 35(3):719–733, November 2017.

Rui Xu, Damian C Ekiert, Jens C Krause, Rong Hai, James E Crowe, and Ian A Wilson. Structural basis of preexisting immunity to the 2009 H1N1 pandemic influenza virus. *Science*, 328(5976): 357–360, April 2010.

Katherine S Xue, Terry Stevens-Ayers, Angela P Campbell, Janet A Englund, Steven A Pergam, Michael Boeckh, and Jesse D Bloom. Parallel evolution of influenza across multiple spatiotemporal scales. *eLife*, 6:46, June 2017.

Katherine S Xue, Louise H Moncla, Trevor Bedford, and Jesse D Bloom. Within-Host Evolution of Human Influenza Virus. *Trends in Microbiology*, pages 1–13, March 2018.

Xiao Yang, Patrick Charlebois, Alex Macalalad, Matthew R Henn, and Michael C Zody. V-Phaser 2: variant inference for viral populations. *BMC Genomics*, 14(1):674, 2013.

Fabio Zanini, Vadim Puller, Johanna Brodin, Jan Albert, and Richard A Neher. In vivo mutation rates and the landscape of fitness costs of HIV-1. *Virus Evolution*, 3(1):1–12, March 2017.

Bin Zhou, Matthew E Donnelly, Derek T Scholes, Kirsten St George, Masato Hatta, Yoshihiro Kawaoka, and David E Wentworth. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and Swine origin human influenza a viruses. *Journal of virology*, 83(19):10309–10313, October 2009.

Shuntai Zhou, Corbin Jones, Piotr Mieczkowski, and Ronald Swanstrom. Primer ID Validates Template Sampling Depth and Greatly Reduces the Error Rate of Next-Generation Sequencing of HIV-1 Genomic RNA Populations. *Journal of virology*, 89(16):8540–8555, August 2015.

Daniel Zinder, Trevor Bedford, Sunetra Gupta, and Mercedes Pascual. The Roles of Competition and Mutation in Shaping Antigenic and Genetic Diversity in Influenza. *PLOS Pathogens*, 9(1): e1003104–13, January 2013.

Mark P Zwart and Santiago F Elena. Matters of Size: Genetic Bottlenecks in Virus Infection and Their Potential Impact on Evolution. *Annual Review of Virology*, 2(1):161–179, November 2015.

Mark P Zwart, José-Antonio Daròs, and Santiago F Elena. One is enough: in vivo effective population size is dose-dependent for a plant RNA virus. *PLOS Pathogens*, 7(7):e1002122, July 2011.