

## RESEARCH ARTICLE

# The effective rate of influenza reassortment is limited during human infection

Ashley Sobel Leonard<sup>1</sup>, Micah T. McClain<sup>2</sup>, Gavin J. D. Smith<sup>3</sup>, David E. Wentworth<sup>4a</sup>, Rebecca A. Halpin<sup>4</sup>, Xudong Lin<sup>4</sup>, Amy Ransier<sup>4</sup>, Timothy B. Stockwell<sup>4</sup>, Suman R. Das<sup>4</sup>, Anthony S. Gilbert<sup>5</sup>, Rob Lambkin-Williams<sup>5</sup>, Geoffrey S. Ginsburg<sup>2</sup>, Christopher W. Woods<sup>2</sup>, Katia Koelle<sup>1</sup>, Christopher J. R. Illingworth<sup>6,7\*</sup>

**1** Department of Biology, Duke University, Durham, North Carolina, United States of America, **2** Duke Center for Applied Genomics and Precision Medicine, Durham, North Carolina, United States of America, **3** Programme in Emerging Infectious Diseases, Duke-NUS Medical School, Singapore, **4** J. Craig Venter Institute, Rockville, Maryland, United States of America, **5** hVivo PLC, The QMB Innovation Centre, Queen Mary, University of London, London, United Kingdom, **6** Department of Genetics, University of Cambridge, Cambridge, United Kingdom, **7** Department of Applied Maths and Theoretical Physics, Centre for Mathematical Sciences, Wilberforce Road, University of Cambridge, Cambridge, United Kingdom

✉ Current Address: Centers for Disease Control and Prevention, MS-G16, Atlanta, Georgia, United States of America

\* [chris.illingworth@gen.cam.ac.uk](mailto:chris.illingworth@gen.cam.ac.uk)

 OPEN ACCESS

**Citation:** Sobel Leonard A, McClain MT, Smith GJD, Wentworth DE, Halpin RA, Lin X, et al. (2017) The effective rate of influenza reassortment is limited during human infection. *PLoS Pathog* 13(2): e1006203. doi:10.1371/journal.ppat.1006203

**Editor:** Adam S. Lauring, University of Michigan, UNITED STATES

**Received:** October 14, 2016

**Accepted:** January 26, 2017

**Published:** February 7, 2017

**Copyright:** © 2017 Sobel Leonard et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Data Availability Statement:** All sequence data is available from the SRA database (accession number SRP091397).

**Funding:** CJRI is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (Grant Number 101239/Z/13/Z) and received support from the National Science Foundation Research Coordination Network on Infectious Disease Evolution Across Scales. KK, ASL, CWW, and MTM were funded by NIGMS U54-GM111274, the MIDAS Center for

## Abstract

We characterise the evolutionary dynamics of influenza infection described by viral sequence data collected from two challenge studies conducted in human hosts. Viral sequence data were collected at regular intervals from infected hosts. Changes in the sequence data observed across time show that the within-host evolution of the virus was driven by the reversion of variants acquired during previous passaging of the virus. Treatment of some patients with oseltamivir on the first day of infection did not lead to the emergence of drug resistance variants in patients. Using an evolutionary model, we inferred the effective rate of reassortment between viral segments, measuring the extent to which randomly chosen viruses within the host exchange genetic material. We find strong evidence that the rate of effective reassortment is low, such that genetic associations between polymorphic loci in different segments are preserved during the course of an infection in a manner not compatible with epistasis. Combining our evidence with that of previous studies we suggest that spatial heterogeneity in the viral population may reduce the extent to which reassortment is observed. Our results do not contradict previous findings of high rates of viral reassortment *in vitro* and in small animal studies, but indicate that in human hosts the effective rate of reassortment may be substantially more limited.

## Author summary

The influenza virus is an important cause of disease in the human population. During the course of an infection the virus can evolve rapidly. An important mechanism of viral evolution is reassortment, whereby different segments of the influenza genome are shuffled with other segments, producing new viral combinations. Here we study natural selection

Inference and Dynamics of Infectious Disease. ASL acknowledges support from the MSTP training grant number T32 GM007171. GJDS was supported by the Duke-NUS Signature Research Programme funded by the Ministry of Health, Singapore and by contract HHSN272201400006C from the National Institute of Allergy and Infectious Disease, National Institutes of Health, Department of Health and Human Services, USA. DEW, RAH, XL, AR, TBS, SRD and also the influenza whole genome sequencing were supported with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract HHSN272200900007C. GSG was funded by the Defense Advanced Research Projects Agency under grant number DARPA-N66001-07-C-2024. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. Evolution Across Scales. KK, ASL, CWW, and MTM were funded by NIGMS U54- GM111274, the MIDAS Center for Inference and Dynamics of Infectious Disease. DEW, RAH, XL, AR, TBS, and SRD were supported with federal funds from the National Institute of Allergy and Infectious Diseases, National Institutes of Health, Department of Health and Human Services, under contract HHSN272200900007C. GSG was funded by the Defense Advanced Research Projects Agency under grant number DARPA-N66001-07-C-2024. This work was performed using the Darwin Supercomputer of the University of Cambridge High Performance Computing Service (<http://www.hpc.cam.ac.uk/>), provided by Dell Inc. using Strategic Research Infrastructure Funding from the Higher Education Funding Council for England and funding from the Science and Technology Facilities Council. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing interests:** ASG and RLW are employees of hVivo PLC.

and reassortment during the course of infections occurring in human hosts. Examining viral genome sequence data from these infections, we note that genetic variants that were acquired during the growth of viruses in culture are selected against in the human host. In addition, we find evidence that the effective rate of reassortment is low. We suggest that the spatial separation between viruses in different parts of the host airway may limit the extent to which genetically distinct segments reassort with one another. Within the global population of influenza viruses, reassortment remains an important factor. However, reassortment is not so rapid as to exclude the possibility of interactions between genome segments affecting the course of influenza evolution during a single infection.

## Introduction

Genome sequencing has provided multiple insights into the evolution of the influenza virus. At the global level, sequences collected from circulating strains of the virus have enabled the identification of codons in the virus evolving under positive selection [1, 2], and demonstrated the importance of interactions between selected alleles [3–5]. Collected sequence data have been used to understand the global migration dynamics of different seasonal strains [6–8] and to make predictions for the future evolution of the virus [9–11]. Considering infection at the level of a single host, sequencing technology has been used to explore the evolution of the influenza virus over the course of a single infection and during the process of viral transmission [12, 13]. Studies of this form have highlighted both the potential for rapid changes in sequence composition within a series of transmission events, as well as diversity in the antigenic properties of a viral population within a single host [14–16]. Experiments conducted in ferrets have considered the within-host evolution and transmission of novel viral strains [17–20].

While some phenotypic properties of the influenza virus can be understood in terms of genetic changes occurring within a single gene [21], interactions between genes have increasingly been identified as of importance for viral evolution. For example, evidence for epistasis in influenza has been found between variants both within single influenza genes [22–24] and across different gene segments [25–27], including the occurrence of permissive mutations which facilitate protein evolution and immune escape. A key multi-gene effect in influenza dynamics is the process of reassortment, whereby distinct influenza viruses within a single cell produce viruses with novel combinations of genome segments, potentially leading to the emergence of new, beneficial strains [28]; evidence for frequent reassortment has been identified in global viral populations [29]. While precise rates of reassortment are hard to quantify, both *in vitro* and *in vivo* experiments have suggested that reassortment between compatible strains of influenza occurs readily [30, 31]. Although the extent of reassortment in experimentally induced infections can depend upon the number of viruses received by a host [32], viral transmission gives a sufficient dose for within-host reassortment to be observed in small animal infection [33], suggesting that reassortment plays an important role in within-host viral dynamics.

Quantitative modelling of within-host influenza infection has a long history [34], but it is only recently that sequencing studies have provided the necessary information to fit evolutionary models to data. Standard techniques for detecting selection, such as dN/dS, are not appropriate for such populations, in which the time-scale for evolution is extremely short [35, 36]. Homologous recombination is rare [37–39], or potentially non-existent, such that selection acting upon one variant may, via linkage disequilibrium, cause changes in allele frequency across a single segment. Under the influence of selection, a population evolves according to a

‘fitness landscape’, which describes the expected within-host growth rate of each virus as a function of viral genotype [40]. Given time-resolved sequence data, approaches which account for the haplotype structure of the viral population have been applied in order to infer the core components of this landscape as it influences viral evolution [41, 42]. However, even in these inferences, the potential role of reassortment has not been considered. Given anything other than a rapid rate of reassortment, linkage disequilibrium between alleles in different segments may have an effect on viral dynamics.

Studies of within-host influenza growth and transmission have been conducted in a variety of animals, including birds, dogs, ferrets, pigs, horses and small mammals [13, 43–45]. Such studies are of great inherent importance, and have long been acknowledged as contributing to our understanding of human infections [46]. However, from the specific perspective of human health, while ferrets in particular provide a valuable model for understanding influenza infection [47, 48], studies conducted in human hosts remain the ultimate reference point.

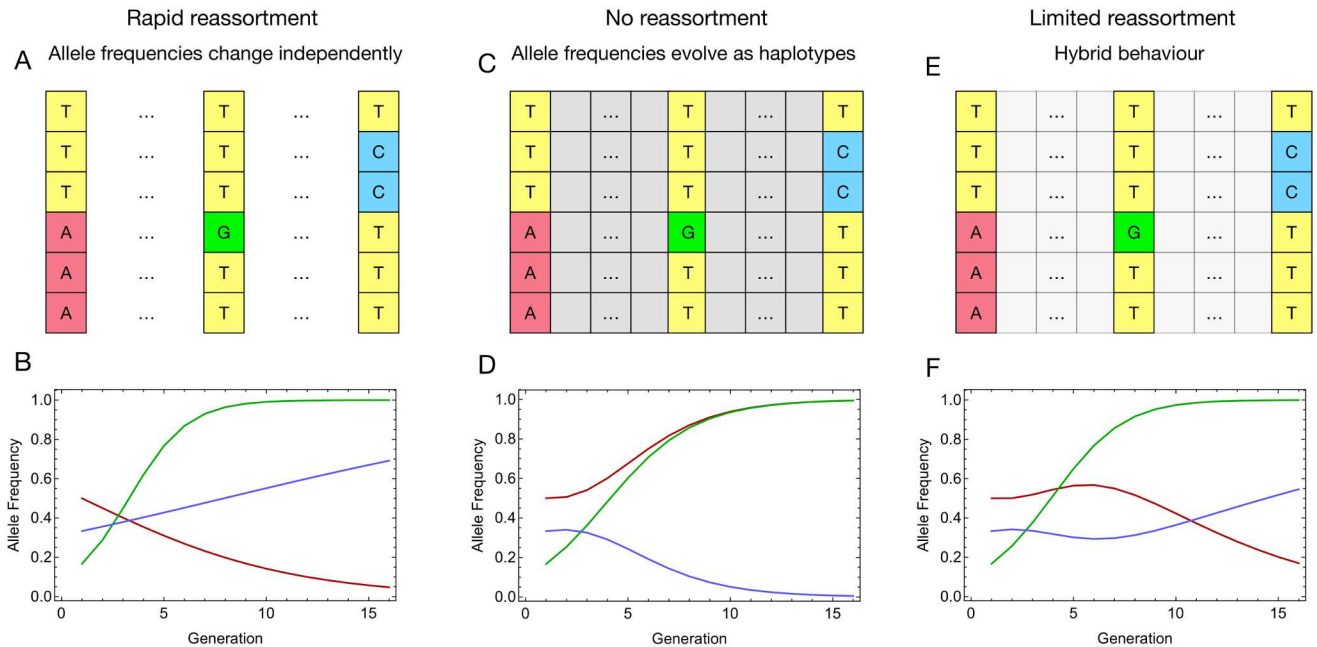
We here describe results from one of the first studies to examine influenza evolution in humans from the perspective of time-resolved viral sequence data, which describe in detail the process of viral evolution as it occurs. Using longitudinal viral sequence data collected from subjects in two related influenza challenge studies (described in previous publications [15, 49–54]) we evaluate how fitness effects shape within-host viral evolution. Using a maximum likelihood approach, we infer the location and magnitude of the selective forces underlying observed changes in the genotypic structure of the viral population. Capturing reassortment in real time on the scale of an infected individual has been noted as a difficult task [55]. Here, using an inference method which incorporates the effect of linkage disequilibrium between alleles on distinct viral segments, we estimate the effective rate of reassortment within the viral population in human subjects on the basis of changes in the genetic composition of the viral population observed over time.

On the basis of the collected data, we infer that selection acting upon variants in the HA, NP, and PA viral genes is responsible for driving within-host evolution in the subjects from which data were collected. Further, we find strong evidence to suggest that the effective rate of reassortment between variants found on different gene segments is limited, such that linkage disequilibrium between alleles on different segments is maintained throughout the course of an infection. The inferences derived via our model are consistent with previous findings that the extent of within-host recombination is linked to the dose of virus received by a patient. Our results are reconcilable with reports of high reassortment rates in *in vitro* and small animal studies given a scenario in which the absolute rate of reassortment is high, but the within-host influenza population is spatially distributed, forming a metapopulation within a patient. We conclude that interactions between variants on different viral segments may substantially affect within-host viral evolution.

## Results

### Model

Our approach to estimating reassortment rates is based upon some elementary principles of population genetics. Given the relatively short length of an influenza infection, and the large number of viruses created during that infection, changes in the viral population are driven by the influence of selection. Selection changes allele frequencies over time, favouring individuals with variants that grant beneficial traits to the virus. However, the manner in which allele frequencies changes under selection depends upon whether variants at different nucleotides are free to change in a manner independent from one another, or whether they are



**Fig 1. Patterns of allele frequency change are altered by the presence or absence of reassortment.** **A.** Representation of a three-locus system with large population size, and in which rapid reassortment wipes out linkage disequilibrium between alleles at different loci. The 'A' variant (red) is under mild negative selection, the 'G' variant (green) is under strong positive selection, while the 'C' variant is under weak positive selection. **B.** Given the rapid rate of reassortment, alleles at each locus evolve deterministically over time in accordance with the selection acting upon them. **C.** Representation of the same system, now with no reassortment. Alleles now evolve according to the changes in the frequencies of full sequences in the population; each horizontal row indicates a set of sequences with one of four distinct three-locus genotypes. **D.** While the initial state of the system, and magnitudes of selection are unchanged, the fates of the red and blue variants are reversed. The AGT genotype is the fittest of the four genotypes; in a process known as genetic hitchhiking, the deleterious red variant is carried by the green to fixation. In a process known as clonal interference, the beneficial blue variant is wiped out in the competition. **E.** Representation of the same system in which there is finite, but limited reassortment between the variants. **F.** Given limited reassortment, the association between the red and green variants is gradually eroded so that the red variant is lost over time. After an initial decline, formation of the maximally fit TGC haplotype leads to the blue variant increasing in frequency over time.

doi:10.1371/journal.ppat.1006203.g001

linked together, perhaps by physical linkage, through epistatic effects, or via a lack of reassortment. Examples of how different reassortment rates affect the evolution of a system are shown in Fig 1. Here we exploit this idea to make the converse step, using observed changes in the genetic composition of multiple viral populations to infer the rate of reassortment within human subjects. Our model exploits the fact that, within a single segment of influenza, short sequence reads may potentially describe alleles at more than one locus. Such reads grant direct information into the associations between alleles in a single segment. These multi-locus data can be exploited to gain a clearer view of the evolution of the population; [42].

Under our approach, we first identify sites in the influenza genome at which polymorphism exists, and at which the minor allele frequency changes significantly over time. The set of observed polymorphisms is used to construct a set of potential virus-wide haplotypes, from which variant alleles are observed. We then construct a model under which the frequencies of viral haplotypes change in frequency over time under mutation and selection, and under the effect of reassortment between viral segments. A likelihood optimisation method is used to fit the model to the data under a wide range of possible rates of reassortment, allowing an inference to be made of the effective rate of reassortment between segments. Full details of our model are given in the Methods section.

### Data summary

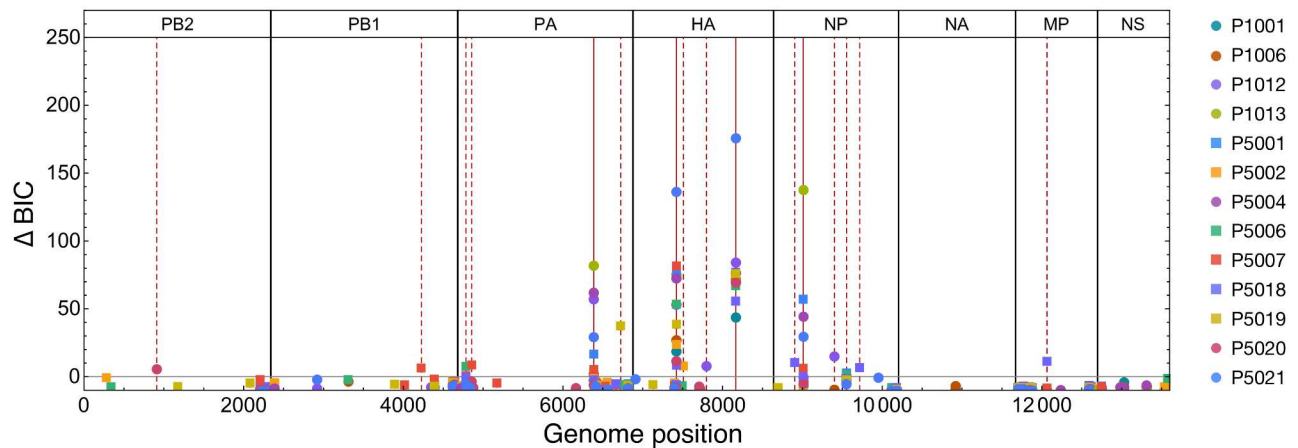
The data used for this study were next-generation sequencing samples generated from two human challenge studies with influenza A/Wisconsin/67/2005(H3N2) virus [56]. Briefly, individuals enrolled in these studies were inoculated intranasally with an influenza H3N2 viral inoculum and monitored throughout maximal symptom development, a total of 7 days. The two studies differed in the treatment given to individuals. In the first study, individuals received standard treatment, with oseltamivir being administered on the evening of the 5th day post-challenge. In the second study, individuals received either standard treatment or early treatment, defined as the administration of oseltamivir on the evening of the 1st day post-challenge. As such, subjects were divided into two cohorts, a ‘standard treatment’ cohort, and an ‘early treatment’ cohort (Table 1). The viral inoculum used for both studies consisted of a genetically heterogeneous population of virus generated from passaged viral reference strain A/Wisconsin/67/2005(H3N2). The virus was passaged 3X in chicken cells, 4X in eggs and 2X in vero cells. Subjects were challenged with doses of the viral inoculum ranging from 3.08–6.41 log<sub>10</sub>(TCID<sub>50</sub>/ml) [57]. Previous analysis of this experiment showed there was no association between the inoculum dose and whether the challenge subject became infected [58]. For those subjects who were infected, there was no association between the inoculum dose and the degree of disease symptoms or the magnitude/duration of viral shedding [58].

Nasal wash samples were collected daily and tested for the presence of virus using culture or quantitative PCR. Those samples containing virus (a total of 60 samples out of 266 samples obtained) and the viral inoculum, were sequenced at the JCVI, as described in [59, 60]. The next generation sequencing pipeline at the JCVI uses both the Illumina HiSeq 2000 and Ion Torrent platforms to compensate for platform specific errors [19]. The sequences from both platforms were used in this analysis. Sequences were obtained for 42 of the nasal wash samples from 17 subjects. As this analysis infers selection based upon on the trajectory of variant frequencies over the course of infection, we excluded samples from 4 individuals (Flu008, Flu010, Flu015 and Flu5017) who were only sequenced at one time-point. The remaining dataset consisted of 38 samples from 13 individuals (six of whom received early treatment and seven of whom received standard treatment), with between two and five samples per subject. Data showing the distribution of read lengths and the depth of sequencing are shown in S1 and S2 Figs. Sequence data are available from the SRA with the project accession number SRP091397.

**Table 1. Challenge study subjects.**

Subject	Experiment	Cohort	Inoculum (log <sub>10</sub> TCID <sub>50</sub> )ml <sup>-1</sup>	Symptomatic	Seroconversion
Flu001	1	Standard	6.41	Y	Y
Flu006	1	Standard	5.25	Y	Y
Flu012	1	Standard	4.41	Y	Y
Flu013	1	Standard	3.8	N	Y
Flu5001	2	Early	5.5	Y	Y
Flu5002	2	Early	5.5	Y	Y
Flu5004	2	Standard	5.5	Y	Y
Flu5006	2	Early	5.5	Y	N
Flu5007	2	Early	5.5	N	Y
Flu5018	2	Early	5.5	Y	Y
Flu5019	2	Early	5.5	Y	Y
Flu5020	2	Standard	5.5	Y	Y
Flu5021	2	Standard	5.5	N	Y

doi:10.1371/journal.ppat.1006203.t001



**Fig 2. Single-locus scan for alleles potentially under selection.** Bayesian Information Criterion (BIC) differences between the best selected model and the neutral model at each locus in the genome. Differences are reported for loci at which polymorphism was detected. A positive BIC difference indicates a weight of evidence in favour of selection, calculated using a single-locus model, applied to data from a single individual. Loci with a positive BIC difference are highlighted using vertical red dotted lines. Solid red vertical lines show loci at which selection was later identified using a multi-locus model with data from all individuals. Circles denote results from individuals receiving standard treatment; squares denote results from individuals receiving early treatment.

doi:10.1371/journal.ppat.1006203.g002

### Single locus model

Considering data from all time-points and all individuals, polymorphisms were identified at a total of 110 loci (nucleotide sites) across the influenza genome. Allele frequencies at these loci measured across time, which we refer to as trajectories, are shown in S3–S10 Figs. The inoculum sample was used as a proxy measurement of the population establishing infection within each host. Trajectories were assessed using a single-locus model of allele frequency change (c.f. [61]), looking for statistically significant deviation from a neutral model, under which the frequency of an allele remained constant over time. Applying this model a total of 16 loci, distributed across six of the eight influenza genome segments, showed potential evidence of non-neutral allele frequency change (Fig 2; equivalent data by segment are shown in S11 Fig). The absence of evidence for significant allele frequency change in either the NA or NS segments implies the absence of strong positive selection in each case. In NA the emergence of drug resistance mutations, notably the amino acid substitution H274T arising from the nucleotide variant C820T, has previously been reported following prolonged oseltamivir treatment in an immunocompromised host [62]. However, no such event was observed here, in either the standard or the early-treatment populations.

### Single-gene, multi-locus model

A single-segment, multi-locus model (SGML) [42] was applied to alleles at the potentially selected loci, accounting for the correlations between alleles that arise due to the lack of homologous recombination in influenza segments. Under this method, short-read data from each segment are used to identify potential haplotypes spanning the segment [63]. Considering potential haplotypes, we note that in a case where two alleles are observed at each of  $n$  different loci, a total of  $2^n$  potential haplotypes may exist. However, where reads span multiple loci, and these multi-locus reads indicate limited diversity across sites, the number of potential haplotypes required by the model may be reduced. Potential segment-wide haplotypes were used under a model of mutation and selection to construct the most parsimonious model of selection acting within each segment, inferring the initial frequency of each haplotype and noting

that selection at one locus may explain a change in allele frequency at another. Under the assumption of a consistent model of selection acting across all individuals, we identified potential evidence for selection at 9 loci, found in the HA, NP, PA, and PB2 segments (Table 2). A negative epistatic interaction between two variants in the HA segment was also inferred; negative epistasis implies that each variant incurs a fitness penalty in the presence of the other, in addition to its inherent effect upon viral fitness. This reduction in the number of potentially selected sites confirms the importance of linkage disequilibrium in influenza dynamics; changes in allele frequencies that under the single locus model appeared to result from the direct effect of selection were often better explained by the indirect effect of selection acting at another locus. For example, four different loci in the PA segment were identified as having significant changes in allele frequency using the single locus model, but when the segment was considered as a whole, selection at locus 1680 was sufficient to explain the changes observed across all four loci; adding selection at a second locus did not significantly improve the ability of the multi-locus model to fit the data.

### Multi-gene, multi-locus model

**Inferences of selection.** Data were further evaluated using a multi-segment, multi-locus (MGML) model, which allows for the modelling of linkage disequilibrium between loci on different segments. Segment-wide potential haplotypes constructed in the SGML model were joined across segments, constructing multi-segment potential haplotypes. In so far as no short reads spanning viral segments were available, all possible combinations of segment-wide haplotypes were incorporated in this model. Given these haplotypes, a model of mutation and selection was used to fit the data, this time accounting for the potential effects of reassortment in the model. As detailed in the Methods section, our model assumes that reassortment occurs as it would in a well-mixed viral population, such that any two viruses in the population have an equal probability of exchanging genetic material. As such, we refer throughout to an “effective reassortment rate”, describing the extent of reassortment between random viral segments in the host; we note that this value may differ from the absolute rate of reassortment within a single cell. Reassortment was assessed by calculating the overall model likelihood across a finite range of reassortment probabilities  $r = 0$ ,  $r = 1$ , and  $r = k \times 10^{-m}$  for  $k \in \{1, 2, \dots, 9\}$  and  $m \in \{1, 2, 3\}$ . All loci identified as potentially being under selection in the SGML model were included in this inference. With the NP segment, while the C372T variant was not identified as being under selection by the SGML model, polymorphisms at this locus were identified in multiple individuals; this locus was also included in the MGML inference.

Under an initial application of the MGML model, in which the effective rate of reassortment between segments was set to zero, selection was inferred to act at a total of four loci across the influenza genome (Table 3). Fitting a consistent fitness landscape to data from all 13 individuals, evidence for selection was inferred in the HA segment against the variants C516A and G1258A, in the NP segment, against the C327T variant, and in the PA segment against the T1680C variant. Evidence for epistatic effects between variants on different segments was found, with the best fitting model inferring epistasis between pairs of alleles on the HA, PA, and NP segments. However, the signal suggesting epistasis between variants within HA was lost. Thus, including interference effects across segments into the model led to a reduction in the number of loci at which selection was inferred.

Of the four variants for which evidence of selection was found, two, in the HA segment, are associated with non-synonymous substitutions. These substitutions have been documented in a previous analysis of these sequence data that focused exclusively on subjects from the first challenge study [15]. The variant at locus 516, representing the amino acid substitution

**Table 2. Inferred fitness coefficients for the single-segment, multi-locus model across all individuals.** Best fitting models are shown for each number of parameters; the best fitting model for data from each segment is highlighted in bold text. The parameter  $\chi$  denotes an epistatic interaction between alleles.

Model	Likelihood	BIC	Constant selection coefficients														
			PB2	PB1	PA		HA					NP			MP		
			G865T	G1862A	T1680C	G2018T	C516A	G603A	C891T	G1258A	$\chi_{516,1258}$	A219G	C327T	G868A	C1035T	A394C	
Neutral	-22.8	45.5															
<b>S1</b>	<b>-16.1</b>	<b>41.2</b>	<b>0.4</b>														
<b>Neutral</b>	<b>-63.9</b>	<b>127.8</b>															
S1	-60.6	130.4	0.2														
Neutral	-965.5	1931.0															
<b>S1</b>	<b>-932.8</b>	<b>1875.9</b>		<b>-0.1</b>													
S2	-931.9	1884.4		-0.1	0.1												
Neutral	-1682.02	3364.0															
S1	-732.8	1476.2								-1.0							
S2	-700.4	1421.7					-0.4			-0.8							
S2E	-675.0	1381.5					-0.4			-0.6							
S3E	-666.6	1375.0					-0.4		0.4	-0.6							
<b>S4E</b>	<b>-659.6</b>	<b>1371.6</b>					<b>-0.4</b>	<b>0.4</b>	<b>0.3</b>	<b>-0.7</b>							
Neutral	-1821.7	3643.5															
S1	-1739.0	3490.0												0.5			
S2	-1682.0	3388.1												1.0	1.1		
<b>S3</b>	<b>-1651.0</b>	<b>3338.3</b>												<b>0.9</b>	<b>1.2</b>		
S4	-1647.9	3344.2												0.2	-0.1	1.2	
<b>Neutral</b>	<b>-147.3</b>	<b>294.5</b>															
S1	-147.3	305.6														-0.0	

doi:10.1371/journal.ppat.1006203.t002



**Table 3. Inferred fitness coefficients for the multi-segment, multi-locus model across all individuals given a zero rate of reassortment.** Best fitting models are shown for each number of parameters; the best fitting model is highlighted in bold text. The parameter  $\chi$  denotes an epistatic interaction between alleles.

Model	Likelihood	BIC	Constant selection coefficients										
			HA		NP	PA	PB2						
			C516A	G1258A	C327T	T1680C	G865T	$\chi_{HA,1258}$	$\chi_{HA,NP}$	$\chi_{HA,PA}$	$\chi_{NP,PA}$		
Neutral	-2965.1	5930.1											
S1	-1798.0	3609.1		-1.1									
S2	-1693.7	3413.4	-0.6	-1.0									
S3	-1612.4	3282.0	-0.8	-1.1	-0.4								
S4	-1505.19	3062.6	-1.0	-1.0	-0.6	-0.7							
S4E	-1466.1	2997.4	-0.9	-1.3	-0.9	-0.6							1.5
S4E2	-1443.2	2964.7	-1.0	-0.9	-0.9	-1.1					-7.8		1.6
<b>S4E3</b>	<b>-1434.4</b>	<b>2960.2</b>	<b>-1.0</b>	<b>-0.9</b>	<b>-0.9</b>	<b>-1.2</b>				<b>-0.4</b>	<b>-19.0</b>		<b>1.6</b>
S4E4	-1433.7	2971.8	-1.0	-0.9	-0.9	-1.2			-0.2	-0.4	-19.0		1.6

doi:10.1371/journal.ppat.1006203.t003

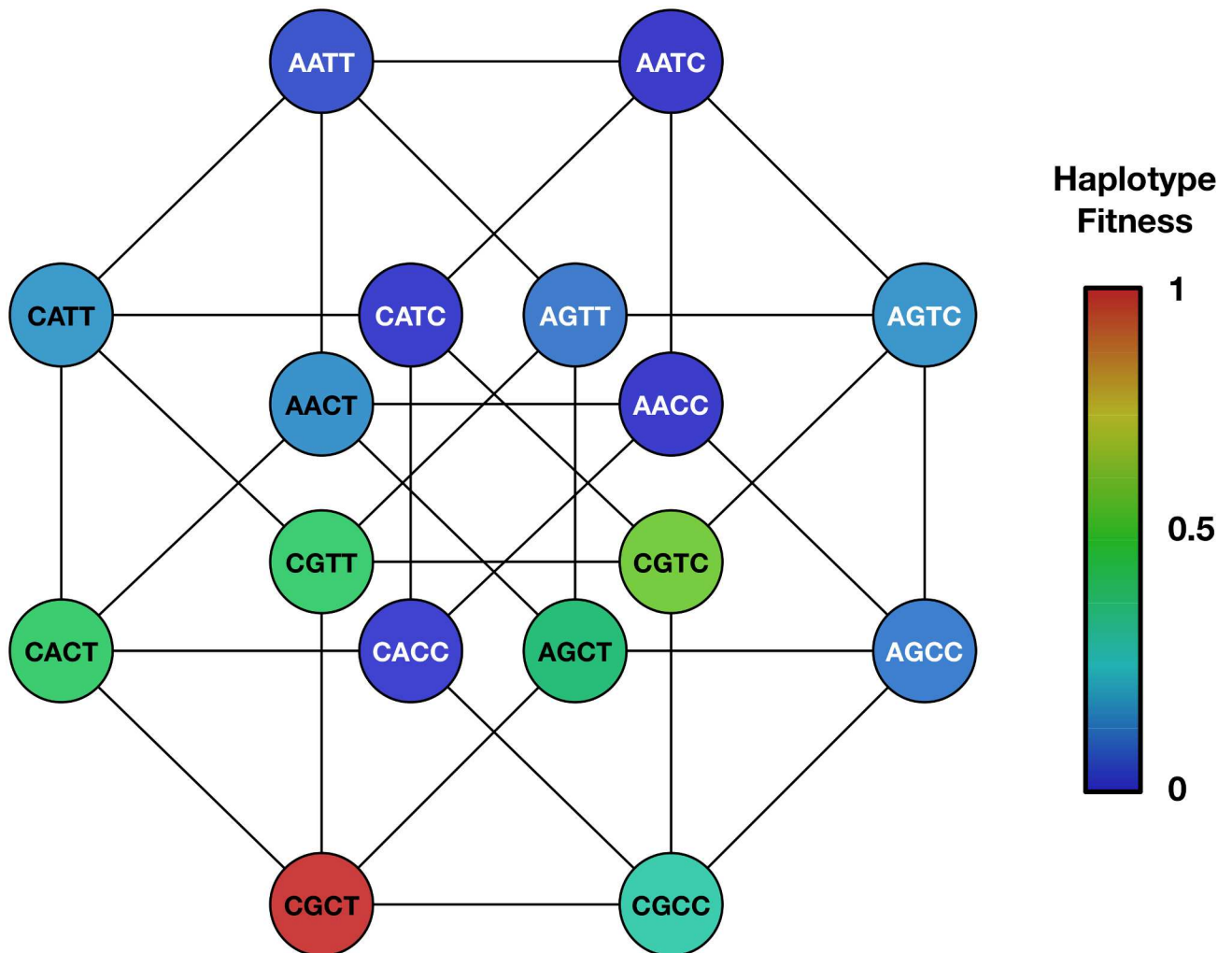
H156Q on the HA1 peptide, has previously been linked with adaptation to the *in ovo* environment [64, 65], while the variant at locus 1258, representing the amino acid substitution G404D (the 75<sup>th</sup> residue of the HA2 peptide), was identified as decreasing viral stability, such that the variant virus was less immunogenic in a ferret host [66]. As such, within-host evolution in HA appears to be driven by purifying selection for the reversion of passaging adaptation, and a gain in viral stability. Calculations based upon the inferred haplotype frequencies show the increase in the frequency of the reverted HA genotype, in addition to a general decrease in sequence diversity when measured in terms of information entropy; details are given in S12 Fig.

Dividing subjects by the treatment they received, inferences of selection were conducted across subjects receiving either the standard or early treatment regime. Similar results were obtained in each case, with selection being inferred on the four alleles mentioned above in each case. Within individuals under the standard treatment regimen, selection was also inferred to act in favour of the G865T variant in PB2; this variant was detected as being polymorphic in a single subject, in whom it increased in frequency over time. Differences in the inferred epistatic parameters were also evident for each model. However, a comparison of separately- and jointly-inferred models of fitness, calculated using the Bayesian Information Criterion, favoured a model of consistent selection acting across all individuals (Table 4); this model is represented in Fig 3. As may be observed from the figure, the inferred landscape has a single peak, representing the consensus sequence of the virus before passaging, with variants that increased in frequency during passaging incurring a fitness cost when the virus is returned

**Table 4. Inferred parameters and likelihoods for all subjects, and for subjects split by treatment, given a zero rate of reassortment.**

Dataset	Likelihood	BIC	Selection parameters											
			HA		NP	PA	PB2							
			C516A	G1258A	C327T	T1680C	G865T	$\chi_{HA,1258}$	$\chi_{HA,PA}$	$\chi_{HA,NP}$	$\chi_{NP,PA}$	$\chi_{NP,PA}$		
<b>All</b>	<b>-1434.4</b>	<b>2960.2</b>	-1.0	-0.9	-0.9	-1.2					-0.4		-19.0	1.6
Early	-520.4		-0.5	-0.4	-1.0	-1.2				-1.5		-21.6		2.1
Normal	-890.1		-0.8	-1.0	-0.5	-0.6	0.8	-1.2						-1.2
<b>Total (split models)</b>	<b>-1410.5</b>	<b>3003.8</b>												

doi:10.1371/journal.ppat.1006203.t004

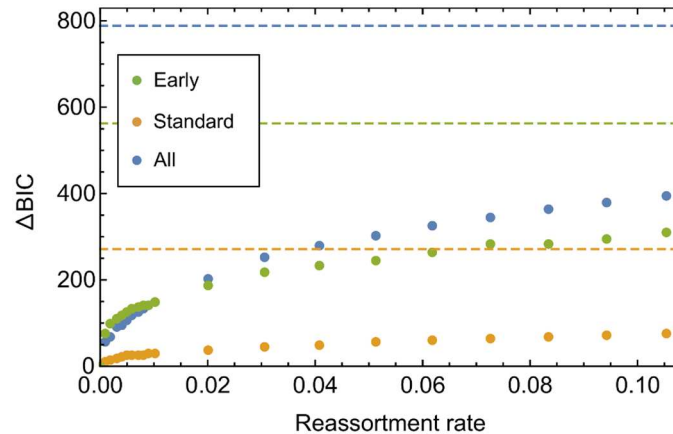


**Fig 3. Inferred haplotype fitnesses.** Reported haplotypes show the composition of the viral sequence at the nucleotide positions HA 516, HA 1258, NP 327 and PA 1680 respectively. Colour indicates inferred relative fitness from blue (0) to red (1). The haplotypes form a hypercube; lines indicate haplotypes differing by a single nucleotide.

doi:10.1371/journal.ppat.1006203.g003

to a human host. In general, as further mutations away from the consensus sequence are accumulated, viral fitness decreases, with the exception of the variants in NP and PA, between which positive epistasis was inferred; here positive epistasis implies that each variant imparts a fitness benefit to the virus in the presence of the other, offsetting the inherent cost of each variant.

**Inferences of effective reassortment rate.** Repeating the inference of selection at non-zero effective reassortment rates led to different rate-specific inferred fitness landscapes. However, after the optimisation of selection parameters, more rapid reassortment rates gave substantially worse fits to the data such that the best overall fit was obtained at a effective reassortment rate of zero (Fig 4) and with the fitness landscape described above. Considering the dataset as a whole, this model gave a improvement, measured via the Bayesian Information Criterion (BIC), of more than 50 units over a model with reassortment rate of  $1 \times 10^{-3} \text{ gen}^{-1}$ , indicating very strong support for a low rate of effective reassortment [67]. Our model implies



**Fig 4. Inferred reassortment rate across all individuals.** BIC values from the MGML model, relative to the optimal value, for the early-treatment, standard-treatment, and combined datasets. Horizontal dashed lines indicate the relative BIC value given rapid reassortment between segments. For each dataset, a reassortment rate of zero was inferred, a higher BIC indicating a worse fit to the data. For scale we note that a difference of 10 BIC units indicates strong support in favour of a model.

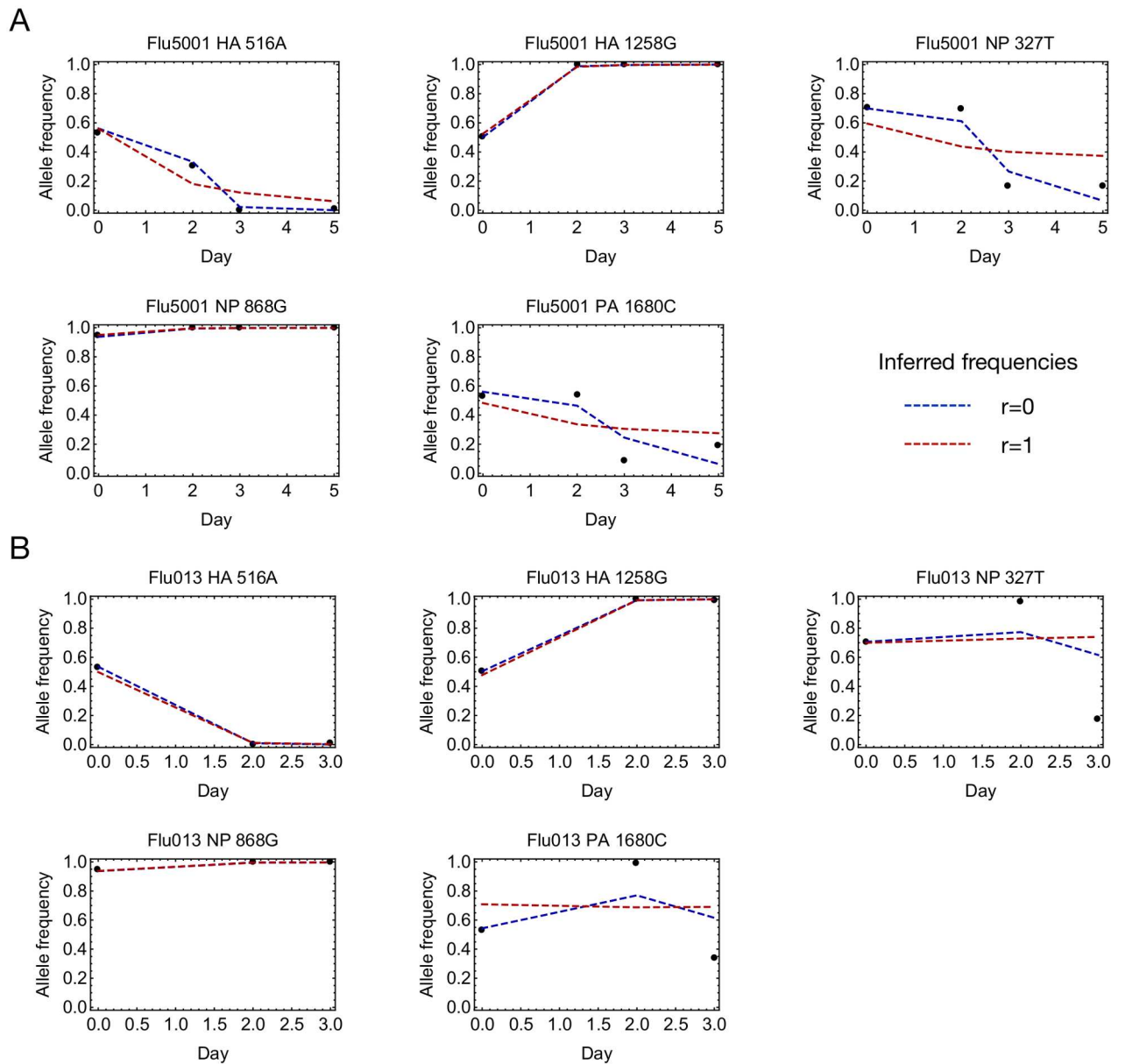
doi:10.1371/journal.ppat.1006203.g004

that linkage between sites on different genome segments had an important role in the observed dynamics of the system.

Examination of allele frequencies within different segments provided a qualitative understanding of the low inferred rate of reassortment (Fig 5); in some subjects allele frequencies within different segments changed in a non-monotonic and apparently correlated manner (allele frequencies for all trajectories included in the MGML model are shown in S13 Fig). A lack of reassortment would directly explain these observations. Firstly, alleles which are in linkage disequilibrium change frequency in a correlated manner. Secondly, while alleles evolving under constant selection are expected to change frequency in a consistent direction, linkage disequilibrium between alleles can produce non-monotonic change via clonal interference and hitchhiking effects.

Subject-specific likelihoods for the rate of reassortment were also recovered from the data. Under the assumption that a consistent fitness landscape applied to the virus across all subjects, non-zero rates of reassortment were sometimes inferred. For example, in the subject Flu5020, the maximum likelihood effective rate of reassortment was calculated to be  $0.005 \text{ gen}^{-1}$  (Fig 6A). In many cases the magnitude of reassortment could not be determined with high precision.

We note that, in general, high rates of reassortment may be difficult to infer directly from sequence data of the form considered here, in which each read describes alleles only from a single segment, with no multi-locus, multi-segment data. Specifically, a model of limited reassortment can produce a good fit to the behaviour of a system with very rapid reassortment, by fixing the linkage disequilibrium between cross-segment loci to be close to zero. As such, an inference of reassortment rate in a system with rapid reassortment, or in which there is no initial linkage disequilibrium, will produce a flat likelihood curve, with little potential to identify reassortment correctly. By contrast, in a system with significant linkage disequilibrium, and where reassortment is slow, there is far greater potential for a distinction to be made between proposed reassortment rates. An illustration of this effect, based upon simulated data, is shown in S14 Fig. In a system with rapid reassortment, inferences conducted using models with high and low reassortment rates give similar results. However, in a system without reassortment, there is a clear preference for the model in which the rate of reassortment is correct. In the real dataset, inferred frequencies for the data from subject Flu001, in whom the uncertainty in the

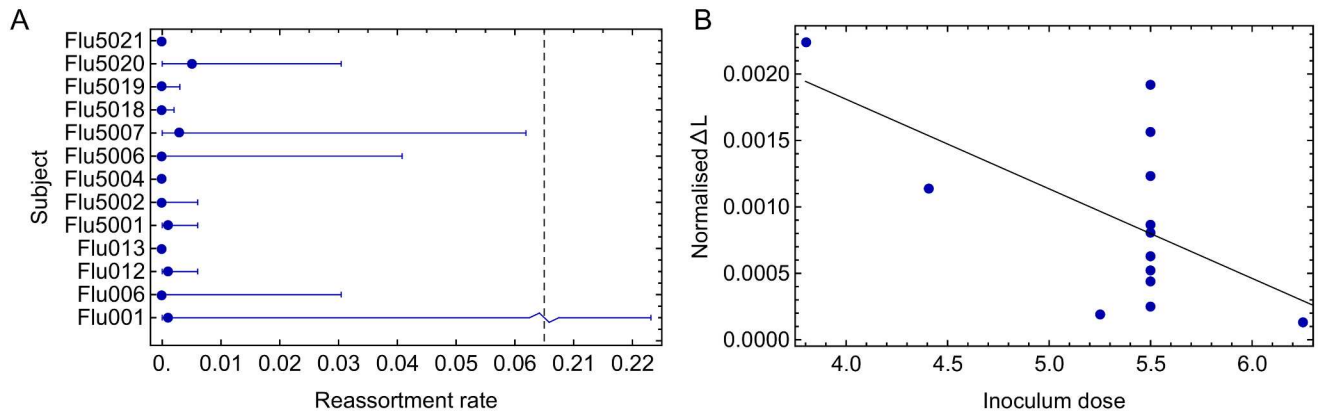


**Fig 5. Optimal inferred frequencies at high and low reassortment rates.** Viral allele frequencies are shown for the example subjects as black dots. The optimised fit to the data, based on the assumption of a consistent fitness landscape across all subjects, is shown as a red dotted line, for the case of rapid reassortment between segments, and as a blue dotted line for the case of no reassortment between segments. Limiting reassortment between segments allows for the possibility of linkage disequilibrium between alleles on different segments, fitting observed correlations in allele frequencies. Data are shown for all polymorphic alleles at loci included in the model for the subjects **A.** Flu5001 and **B.** Flu013.

doi:10.1371/journal.ppat.1006203.g005

inferred reassortment rate was greatest, show that both low and high reassortment rates produce similar model fits (S15 Fig).

Exploring the idea that the extent to which reassortment can be inferred is indicative of the reassortment rate itself, we obtain a result consistent with previous studies that have reported a link between the size of the inoculum dose used in experimental influenza infections, and the extent of reassortment between segments [32]. We consider as a proxy for the extent of recombination the magnitude of change in the subject likelihood function between the optimal value



**Fig 6. Individual-specific reassortment rates.** **A.** Optimal reassortment rates inferred for individual subjects. Error bars, shown as horizontal lines, correspond to confidence intervals obtained via a likelihood ratio test. **B.** Correlation between the extent of evidence for a limited reassortment rate, normalised by the extent of available data, and the size of the inoculum dose. The black line shows a linear regression model fitted to the data.

doi:10.1371/journal.ppat.1006203.g006

and that obtained for a model of rapid reassortment. While this ‘flatness’ of the likelihood function is determined both by the initial extent of linkage disequilibrium between alleles and the reassortment rate, all subjects were inoculated from the same viral stock, such that the former parameter is likely to be relatively consistent between infections. Given that doses were varied in only one of the two challenge studies, our ability to identify a link is somewhat limited. Nevertheless, as shown in Fig 6B, an increased inoculum dose was associated with a decreased likelihood difference between models ( $p = 0.028$ ), suggestive of a positive relationship between reassortment rate and inoculum dose.

**Validation against simulated data.** The ability of our approach to infer selection and reassortment parameters were validated against simulation. A low reassortment rate was successfully inferred, with variance between the inferred selection parameters being consistent with previous simulation studies [42]; in some cases it is possible for sets of selection parameters which contain compensatory errors to give a good fit to the correct relative haplotype fitness values. More details are given in S1 Text and S16–S18 Figs.

## Discussion

We have applied a novel multi-segment model of within-host evolution to data from infections in multiple individuals during two influenza challenge studies. Our approach for inferring selection is conservative with regard to the quantity of information contained within next-generation sequencing data, and in preferring simpler explanations of variant frequency change. Using our approach, we infer the key evolutionary effects influencing the adaptation of the within-host population in the human subjects under study.

Building upon previous approaches for inferring selection from influenza populations [41, 42, 61], our method infers selection not only at the level of events occurring at a single locus, or within a single segment, but across the entire influenza genome. As inferences were conducted across greater numbers of loci, in the SGML and MGML models, a decrease was observed in the number of loci at which selection was inferred. While the single locus test identifies potential non-neutrality at 16 loci over 6 gene segments, under the MGML model, selection was inferred to apply at only four of these positions within three gene segments. Two reasons likely underlie this reduction in the apparent complexity of the selective model. Firstly, due to linkage disequilibrium, selection acting at one locus can cause changes in the frequency

of alleles at multiple other loci; inter-segment associations between alleles may cause changes in the frequencies of alleles not directly under selection. Accounting for linkage disequilibrium between alleles reduces false positive inferences of selection caused in this manner. Secondly, as more data are brought into the model, the model penalty for adding another parameter for selection increases. Where a polymorphism is observed in only a single individual, evidence of genuine selection may not be sufficient to be detected by our model, leading to false negative calls of selection. As such, our inferences describe the core set of fitness effects shaping the evolution of the virus, rather than a comprehensive fitness landscape.

Across all subjects, we infer the key driver of evolution to be the reversion of mutations in HA acquired during passaging before inoculation into individual patients. However, selection was further inferred for variants in the NP and PA segments. The time of administration of oseltamivir to patients did not have a statistically significant effect upon the inferred fitness landscape. Although the emergence of oseltamivir resistance has been documented as having arisen during a single infection [68], in this case no evidence of drug resistance was observed; with regard to the genetic composition of the virus, drug therapy had no apparent impact. While early oseltamivir treatment led to reductions in the severity of symptoms and the length of infection [69], it did not have a distinct evolutionary effect on the viral genome.

Among the variants identified as being under selection, those occurring in HA could be attributed to the reversion of mutations acquired during passaging. The variants C327T in NP and T1680C in PA are both synonymous in nature, and identifying a biological rationale for selection acting upon them requires a little more speculation. Both variants were called as polymorphisms in all subjects, though the explanation for how these variants could affect viral fitness is less clear. One possible explanation would be that the variants interfere with a packaging signal within the vRNA. Packaging signals are a necessary component in the selective packaging model for the influenza virus, which asserts that there is some specific element within the primary sequence of the viral RNA ensuring that each of the eight gene segments is incorporated to new virions [70]. While the packaging signals so far identified have tended to be within the UTRs rather than the coding regions of the segments, a few exceptions to this rule exist [71, 72]. However, none of the packaging sequences identified within the coding regions for NP or PA coincide with the loci described here.

Another possible explanation for selection occurring on the synonymous mutations is that these variants interfere with the secondary structure of the viral RNA [73], structure being potentially related to viral fitness due to the need for the RNA to have a specific thermal stability [74]. For example, the synonymous variant C327T in NP identified by this analysis as under selection is located within a region of highly conserved secondary structure [75]. This region, spanning nucleotides 1031–1250 (numbers indicating the (+) RNA), has been characterized as having high stability. As such, this variant (at position 1193 of the (+) RNA), may be neutral or beneficial within the *in ovo* environment, but confer a fitness disadvantage in the human challenge subjects. While secondary structural elements have been identified within the PA segment [76], the regions described have not included the locus 1680, noted here.

In addition to inferring selection for specific variants, we also inferred the presence of epistatic effects acting between alleles in different segments of influenza. Following the discovery that reassortment viruses generated from distinct strains do not reassort in a random manner [77], a considerable body of literature has been developed studying the propensity for gene segments from distinct viruses to be separated, or found together, following a reassortment event. More formal statistical approaches for identifying associations between segments have been developed [78]. Biases in reassortment have been found in human influenza A and B viruses [79, 80], in cell culture [78, 81], and in swine influenza viruses

[82]. The focus of this study is somewhat different to these, being concerned with the rate at which effective reassortment takes place rather than what is produced given a reassortment event. Nevertheless, epistasis is an important factor in our study due to the potential for linkage disequilibrium to be maintained between alleles on separate segments via epistatic effects. By accounting for epistasis in our model, we separate the maintenance of linkage disequilibrium caused by epistasis from that arising from a lack of reassortment such that our result cannot be fully explained by non-random reassortment between segments. Our model infers the existence of epistatic interactions between the HA, NP, and PA segments. However, our study is not the ideal vehicle for which to consider such factors; the study of more diverse strains in an environment in which there was rapid effective reassortment would be desirable for making such inferences.

The restricted extent of effective reassortment between segments inferred during the course of infection is perhaps the most surprising result of our study, with a model of zero reassortment granting the best fit to the data. This result was supported by extremely strong model evidence, with a model of low reassortment providing a far better explanation of the data than a model with rapid reassortment. Our model is limited in so far as it requires the polymorphisms in a segment to be observed at significant allele frequencies; as such, our estimates of reassortment were made across a subset of the segments in the influenza genome. Our result should also be understood in proper context; a lower implied within-host reassortment rate does not contradict observations of reassortment events in global influenza populations.

Our study contains a result matching earlier findings in suggesting that the rate of reassortment within a host is linked to the dose received by a single patient. However, our study contrasts with previous work in the extent of reassortment that was observed; while we identify low reassortment rates, an earlier study noted 86% of viruses reassorting during the course of a single infection in a case where a guinea pig was infected with  $10^6$  PFU of virus [31]. We suggest that these results may be reconciled under the assumption of a metapopulation model of infection (c.f. [83, 84]) in which interactions between viruses are limited by spatial diversity within the host. Such a model is well-supported by existing knowledge of influenza virus; viruses in cell culture form distinct plaques [85], while genetically distinct influenza populations have been observed at different locations within a single ferret [19], implying that viruses are not uniformly distributed throughout a host. Under such a model, reassortment, as a local process occurring in single cells, will tend to occur between viruses that are genetically more self-similar than viruses drawn randomly from the population; the effective rate of reassortment, as we define it, will be lower than the absolute rate at which segments are swapped between viruses. Assuming a constant absolute rate of reassortment throughout the host, the rate of effective reassortment will be a function of the local genetic diversity of the population, which itself will depend upon the number of distinct viruses founding an infection in any given region of the host respiratory system. As the total number of viruses founding infection in a host increases, more, and therefore statistically more diverse viruses, found each local infection, such that the observed extent of reassortment also increases. This understanding suggests that in a larger host, where the respiratory system occupies a larger physical space, more viruses would be required to create the same local genetic diversity, and hence the same observed rate of reassortment; given the same infectious dose, the observed rate of reassortment will be lower.

Our study does not oppose the vital role played by animal experiments in understanding influenza infection. However with specific regard to reassortment, experiments conducted *in vitro* or in small animal experiments may over-estimate the effective reassortment rate that might be expected in a human infection. Multi-segment effects may have a substantial and previously under-appreciated role in the evolutionary dynamics of within-host influenza.

## Methods

### Sequence processing

The SAMFIRE package [63] was used to call time-resolved multi-locus sequence variants from the data. High quality short reads were identified using the criteria of a median PHRED quality score of at least 30, trimming sequences where necessary to obtain this. Single nucleotide polymorphisms (SNPs) were then identified from reported nucleotides with PHRED score at least 30, for which at least 10 reads, and at least 2% of the reads in a sample, reported the variant. To account for possible PCR bias, variant alleles were only considered if they were identified in at least two samples from the dataset. SNP trajectories, describing the frequency of each SNP over time, were calculated.

In order to quantify the influence of selection, a conservative estimate was made of the extent of noise in the reported sequence data. Under the assumption that changes in allele frequency over time resulted purely from noise, the extent of noise was characterized within the framework of a Dirichlet multinomial model. Where in a given individual  $i$ ,  $n_a^i(t)$  is the number of copies of a variant allele observed at the polymorphic locus  $a$  at time  $t$ , and  $N_a^i(t)$  is the total number of alleles observed in this individual at this locus at time  $t$ , we calculated the mean allele frequency across all reads

$$\hat{q}_a^i = \frac{\sum_t n_a^i(t)}{\sum_t N_a^i(t)} \quad (1)$$

for each allele. Assuming that observed changes in allele frequency occur only through noise, we then optimised the parameter  $C$  to maximise the likelihood

$$\mathcal{L} = \sum_{a,i,t} \mathcal{L}^D(C, \{n_a^i(t)\}, \{\hat{q}_a^i\}) \quad (2)$$

where  $\mathcal{L}^D$  is the Dirichlet multinomial distribution with parameters  $\alpha_i = Cq_a^i$ . This approach produces a conservative estimate of the extent of noise in the system, characterised by  $C$ .

### Single-locus (SL) model

Given an estimate of the extent of noise in the sequenced data, a likelihood model was used to identify potentially non-neutral alleles across the influenza genome using a single-locus model of selection similar to those applied in earlier studies of viral evolution [61]. Here, for data from allele  $a$  within individual  $i$ , we generate a set of model allele frequencies  $q_a^i(t)$ , calculating the model likelihood as

$$\mathcal{L}^m = \sum_t \mathcal{L}^D(C, \{n_a^i(t)\}, \{q_a^i(t)\}) \quad (3)$$

and comparing models using the Bayesian Information Criterion (BIC) for model selection [86]; under this comparison, a difference of 10 in BIC values indicates strong evidence for the more complex model [67]. In fitting a model to data, our model assumes a “generation time”, representing a round of intracellular reproduction, as occurring in 12 hours [87]. Fitted models describe the deterministic evolution of an allele frequency under either a neutral model, or a model of selection for the allele in question. In order to account for potential hitchhiking effects, we also examined a time-dependent model of selection, in which a different magnitude of selection was used to model selection between each pair of samples within an individual.



Neutral model:

$$q_a(t) = q \quad \forall t \quad (4)$$

Constant selection:

$$q_a(t) = \frac{q_a(0) \exp(\sigma t)}{1 - q_a(0) + q_a(0) \exp(\sigma t)} \quad (5)$$

Time-dependent selection:

$$q_a(t) = \frac{q_a(t-1) \exp(\sigma_t)}{1 - q_a(t-1) + q_a(t-1) \exp(\sigma_t)} \quad (6)$$

At the single-locus level, time-dependent selection was considered so as to identify significant non-monotonic changes in allele frequency, as might occur via linkage disequilibrium between alleles; in further models where linkage disequilibrium between alleles was explicit to the model, only models of constant selection were considered. As in the models described below, fitness differences between viruses are understood to arise from differences in the respective within-host growth rates of genetically distinct viral strains.

### Single-segment multi-locus (SGML) model

Where more than one variant occurs within a single segment, selection at a single locus may result in changes in the genetic composition of the virus across multiple loci [88]; under such conditions, single-locus measures of selection can produce misleading results [89]. A multi-locus model was used to account for such effects. Data within each segment were compiled to call multi-locus variants across potentially non-neutral loci. Multi-locus variants were then used to construct a set of potential viral haplotypes for each segment. Our approach to haplotype reconstruction, implemented within the SAMFIRE package, is two-fold by design. Given only single-locus reads, the number of potential haplotypes in a segment is an exponential function of the number of loci at which variation was observed. However, given multi-locus reads, limitations in the diversity of the population may be directly observed. Firstly, therefore, we infer a set of potential segment-wide haplotypes sufficient (though perhaps not necessary) to explain the observed multi-locus data, with no concern for the frequencies at which these haplotypes may or may not exist in the population. Secondly, we fit an evolutionary model describing mutation and selection to the variant data, under the constraint that the population is composed of the identified segment-wide haplotypes in some proportion at each recorded point in time, the frequencies of these haplotypes changing according to the evolutionary models. Mutation was modelled as occurring at constant rate between haplotypes:

$$M : q_a(t+1) = q_a(t) + \mu \sum_b (q_b(t) - q_a(t)) \quad (7)$$

where  $q_a(t)$  is the frequency of the haplotype  $a$  in the population at time  $t$ , and the sum is conducted over haplotypes differing from  $a$  by a single nucleotide. The mutation rate was assumed to be  $\mu = 10^{-5}$  per site per replication cycle [90]. Selection was modelled to occur deterministically, changing the frequencies of different viral haplotypes according to their fitness.

$$S : q_a(t+1) = \frac{q_a(t) \exp(\sigma_a)}{\sum_b q_b(t) \exp(\sigma_b)} \quad (8)$$

where the sum is conducted over all haplotypes  $b$ , and the fitness of haplotype  $a$  is determined

by  $\exp(\sigma_a)$ , where

$$\sigma_a = \sum_i s_i + \sum_{i,j} \chi_{i,j} + \sum_{i,j,k} \chi_{i,j,k} + \dots \quad (9)$$

is a sum of single- and multi-locus (i.e. epistatic) terms, over variant nucleotides  $i, j, k, \dots$ , contained within the haplotype  $a$ . Hierarchical models of fitness were considered, gradually increasing the number of component terms in the fitness model, and using BIC to select the model giving the best explanation of the observed changes in viral genetic composition [41, 42].

### Multi-segment multi-locus (MGML) model

While reassortment between influenza segments is potentially a rapid process [31], associations have been observed to exist between alleles on different segments [91]. Further to previous modelling approaches, an inference was made of the effect of reassortment upon the viral dynamics, via the use of a multi-segment, multi-locus model. Sites in each segment at which potential non-neutrality was inferred in the multi-locus model were combined, forming potential viral haplotypes spanning all segments at which selection had been inferred. As above, selection and mutation were evaluated across haplotypes, the fitness of any given viral genotype being calculated at the multi-segment haplotype level. However, in addition, reassortment was evaluated between segments, being modelled to occur between haplotypes with probability  $r$ :

$$R : q_a(t + 1) = (1 - r)q_a(t) + r \prod_g q_{a,g}(t) \quad (10)$$

where  $q_a$  now denotes the frequency of viruses with a specific multi-segment haplotype, and  $q_{a,g}(t)$  is the total frequency at time  $t$  of viruses having the same alleles as  $a$  across all loci in the segment  $g$ . The precise mechanism via which influenza segments are packaged into virions is a subject of ongoing research [92]. Following our assumptions of a large, well-mixed influenza population, we here assume a multi-parental model of reassortment, whereby reassortment occurs freely between a proportion of viruses in each generation. Recalculation of statistics under a strict bi-parental model of reassortment reproduced our key inference of a low effective reassortment rate; see [S1 Text](#) and [S19 Fig](#). Given a Poisson model, we note that an equivalent reassortment rate  $\rho$  can be calculated as:

$$\rho = -\log(1 - r) \quad (11)$$

In the above model, reassortment occurs between random genotypes in the population, as would occur if the population was well-mixed. In the same way that the effective population size is used in population genetics to seek to quantify the extent of genetic drift in terms of an idealised population [93], the statistic we measure in this way describes an effective reassortment rate in the viral population. We note that this effective rate may be significantly lower than the absolute rate of reassortment between viruses, as in a case where reassortment primarily occurs between viruses with similar genotypes.

In order to reduce the computational cost of the calculation, the multi-segment model was evaluated only across sites within each individual at which polymorphism was identified; as such, selection for a genetic variant was only considered to affect the evolution within a host if the variant under selection was observed in that host. Polymorphic loci in an individual form a subset of the total set of loci; individual-specific multi-segment haplotypes were constructed as a projection of the complete set of multi-segment haplotypes onto these loci, evaluating the effect of a consistent model of selection on the dynamics of each system.

Calculations involving reassortment rate were conducted across a range of values of  $r$ . Subject-specific estimates of reassortment were derived from the component of the likelihood function corresponding to the data from that subject. A further measure quantifying the extent of evidence for limited reassortment in an individual subject was calculated as the normalised difference between the maximum likelihood reassortment model at any value of  $r$ , and under a model of rapid reassortment.

$$\frac{\max \{ \mathcal{L}^r \}_r - \mathcal{L}^{(r=1)}}{N} \quad (12)$$

where  $N$  was the total number of datapoints collected from that individual.

## Ethics statement

The details of this human challenge study have been previously described in [15, 49–54]. The procedures performed under this study were in accordance with the Declaration of Helsinki. The protocols followed for this study were approved by the institutional review boards (IRBs) of Duke University Medical Center (Durham, NC), the Space and Naval Warfare Systems Center San Diego (SSD-SD) of the US Department of Defense (Washington, DC), the East London and City Research Ethics Committee 1 (London, UK), and the Independent Western Institutional Review Board (Olympia, WA). All participants enrolled in this challenge study provided written consent, as in accordance with standard IRB protocol.

## Code

Code for the MGML model is available from <http://websvc.gen.cam.ac.uk/~cjri2/MGML>. The SAMFIRE code contains the single-locus model and is available at <https://github.com/cjri/samfire>. The SGML code is available as part of a previous publication [42].

## Supporting information

**S1 Text. Statistical properties of the dataset.** Properties of the sequence data. Inferences conducted using simulated data. Model validation.  
(PDF)

**S1 Fig. Length of short reads.** Lengths of short reads are reported following processing via the SAMFIRE software package. During processing, reads of shorter than 30 nucleotides were removed from the dataset. Where paired-end data were available, paired-end reads were joined into single continuous reads. The length of a short read is then defined as the total distance in the genome between the first and last reported nucleotides.  
(TIF)

**S2 Fig. Genome-wide read depths.** Read depths are reported following filtering via the SAMFIRE software package. During processing, nucleotide calls with PHRED score less than 30 were removed; only high-quality data are here reported. Across the 39 samples collected, including the sampling of the innoculum, data are shown for the minimum (blue), median (yellow) and maximum (green) read depths for each genome position.  
(TIF)

**S3 Fig. Allele frequency trajectories of polymorphisms in the PB2 gene segment.** Observed allele frequency values are colour-coded by individual.  
(TIF)

**S4 Fig. Allele frequency trajectories of polymorphisms in the PB1 gene segment.** Observed allele frequency values are colour-coded by individual.

(TIF)

**S5 Fig. Allele frequency trajectories of polymorphisms in the PA gene segment.** Observed allele frequency values are colour-coded by individual.

(TIF)

**S6 Fig. Allele frequency trajectories of polymorphisms in the HA gene segment.** Observed allele frequency values are colour-coded by individual.

(TIF)

**S7 Fig. Allele frequency trajectories of polymorphisms in the NP gene segment.** Observed allele frequency values are colour-coded by individual.

(TIF)

**S8 Fig. Allele frequency trajectories of polymorphisms in the NA gene segment.** Observed allele frequency values are colour-coded by individual.

(TIF)

**S9 Fig. Allele frequency trajectories of polymorphisms in the MP gene segment.** Observed allele frequency values are colour-coded by individual.

(TIF)

**S10 Fig. Allele frequency trajectories of polymorphisms in the NS gene segment.** Observed allele frequency values are colour-coded by individual.

(TIF)

**S11 Fig. Single-locus scan for alleles potentially under selection.** Bayesian Information Criterion (BIC) differences between the best selected model and the neutral model at each locus in the genome. Differences are reported for loci at which polymorphism was detected. A positive BIC difference indicates a weight of evidence in favour of selection, calculated using a single-locus model, applied to data from a single individual. Loci with a positive BIC difference are highlighted with vertical red dotted lines. Solid red vertical lines show loci at which selection was later identified using a multi-locus model with data from all individuals. Circles denote results from individuals receiving standard treatment; squares denote results from individuals receiving early treatment.

(TIF)

**S12 Fig. Inference of purifying selection within the genome. A.** Inferred frequency of individuals within each population having the wild-type consensus alleles at loci 516 and 1258 within the HA segment. Under purifying selection, a return to the consensus over time is inferred. **B.** Inferred sequence entropy calculated across the haplotypes used to model viral evolution within each individual. A general decrease in entropy over time is consistent with the action of purifying selection upon the viral population.

(TIF)

**S13 Fig. Allele frequency trajectories of trajectories used in the inference process.** Observed allele frequency values are colour-coded by gene segment.

(TIF)

**S14 Fig. Inferrability of high and low reassortment rates.** Simulated data (dots) and inferences (lines) are shown describing two-locus haplotype frequencies  $q_{ij}^{11}$  (blue),  $q_{ij}^{10}$  (green),  $q_{ij}^{01}$  (yellow),  $q_{ij}^{00}$  (orange), and single-locus allele frequencies for systems with reassortment rates

$R^S$  inferred using models with reassortment rates  $R^M$ . Inferences were conducted by fitting single-locus allele frequency data. The maximum log likelihood for the system,  $L$ , is reported. (a, b) Where the reassortment rate of the system is high, models with either high or low reassortment rate provide a good fit to the data. (c, d) Where the reassortment rate of the system is low, a model with high reassortment rate may be unable to correctly reproduce the behaviour of the system. Where the model reassortment rate is high, the inferred model may nevertheless have instantaneous linkage disequilibrium at the first time point.

(TIF)

**S15 Fig. Optimal inferred frequencies at high and low reassortment rates.** Viral allele frequencies are shown for subject Flu001 as black dots. The optimised fit to the data, based on the assumption of a consistent fitness landscape across all subjects, is shown as a red dotted line, for the case of rapid reassortment between genes, and as a blue dotted line for the case of no reassortment between genes.

(TIF)

**S16 Fig. BIC values inferred from a simulated population at different model reassortment rates.** Inferences were performed at different model reassortment rates for a simulated population based upon parameters inferred from the real population. A model which correctly reproduces the low reassortment rate of the input population was clearly favoured by the inference.

(TIF)

**S17 Fig. Allele frequencies inferred from a simulated population.** Dots show samples of allele frequencies collected from each population. Equivalent lines show model fits to the same data.

(TIF)

**S18 Fig. Fitness landscape inferred from a simulated population.** Reported haplotypes show the composition of the viral sequence at the nucleotide positions HA 516, HA 1258, NP 327 and PA 1680 respectively. Colour indicates inferred relative fitness from blue (0) to red (1). Lines indicate haplotypes accessible via a single mutation.

(TIF)

**S19 Fig. Inferred reassortment rate across all individuals under a pairwise reassortment model.** BIC values from the MGML model, relative to the optimal value, for the combined dataset. The data give a close qualitative fit to the values obtained under a multi-parental model.

(TIF)

## Author Contributions

**Conceptualization:** ASL MTM GSG ASG RLW CWW KK CJRI.

**Data curation:** ASL GJDS RAH XL TBS CJRI.

**Formal analysis:** ASL CJRI.

**Funding acquisition:** SRD GSG KK CJRI.

**Investigation:** MTM GJDS RAH ASG AR CWW CJRI.

**Methodology:** ASG RLW CJRI.

**Project administration:** DEW SRD ASG GSG KK CJRI.

**Resources:** GJDS RLW ASG GSG.

**Software:** CJRI.

**Supervision:** DEW SRD ASG RLW GSG KK CJRI.

**Validation:** RLW CJRI.

**Visualization:** CJRI.

**Writing – original draft:** ASL KK CJRI.

**Writing – review & editing:** ASL MTM GJDS DEW RAH AR TBS SRD ASG RLW GSG CWW KK CJRI.

## References

1. Fitch WM, Bush RM, Bender CA, Cox NJ. Long Term Trends in the Evolution of H(3) HA1 Human Influenza Type A. *Proceedings of the National Academy of Sciences of the United States of America*. 1997; 94(15):7712–7718. doi: [10.1073/pnas.94.15.7712](https://doi.org/10.1073/pnas.94.15.7712) PMID: [9223253](https://pubmed.ncbi.nlm.nih.gov/9223253/)
2. Wolf YI, Viboud C, Holmes EC, Koonin EV, Lipman DJ. Long intervals of stasis punctuated by bursts of positive selection in the seasonal evolution of influenza A virus. *Biology direct*. 2006; 1:34. doi: [10.1186/1745-6150-1-34](https://doi.org/10.1186/1745-6150-1-34) PMID: [17067369](https://pubmed.ncbi.nlm.nih.gov/17067369/)
3. Strelkova N, Lässig M. Clonal interference in the evolution of influenza. *Genetics*. 2012; 192(2):671–682. doi: [10.1534/genetics.112.143396](https://doi.org/10.1534/genetics.112.143396) PMID: [22851649](https://pubmed.ncbi.nlm.nih.gov/22851649/)
4. Illingworth CJR, Mustonen V. Components of Selection in the Evolution of the Influenza Virus: Linkage Effects Beat Inherent Selection. *PLoS Pathogens*. 2012; 8(12):e1003091. doi: [10.1371/journal.ppat.1003091](https://doi.org/10.1371/journal.ppat.1003091) PMID: [23300444](https://pubmed.ncbi.nlm.nih.gov/23300444/)
5. Koelle K, Rasmussen DA. The effects of a deleterious mutation load on patterns of influenza A/H3N2's antigenic evolution in humans. *eLife*. 2015; 4:e07361. doi: [10.7554/eLife.07361](https://doi.org/10.7554/eLife.07361) PMID: [26371556](https://pubmed.ncbi.nlm.nih.gov/26371556/)
6. Nelson MI, Simonsen L, Viboud C, Miller MA, Holmes EC. Phylogenetic Analysis Reveals the Global Migration of Seasonal Influenza A Viruses. *PLoS Pathogens*. 2007; 3(9):e131. doi: [10.1371/journal.ppat.0030131](https://doi.org/10.1371/journal.ppat.0030131)
7. Bedford T, Cobey S, Beerli P, Pascual M. Global Migration Dynamics Underlie Evolution and Persistence of Human Influenza A (H3N2). *PLoS Pathogens*. 2010; 6(5):e1000918. doi: [10.1371/journal.ppat.1000918](https://doi.org/10.1371/journal.ppat.1000918) PMID: [20523898](https://pubmed.ncbi.nlm.nih.gov/20523898/)
8. Bedford T, Riley S, Barr IG, Broor S, Chadha M, Cox NJ, et al. Global circulation patterns of seasonal influenza viruses vary with antigenic drift. *Nature*. 2015; 523(7559):217–220. doi: [10.1038/nature14460](https://doi.org/10.1038/nature14460) PMID: [26053121](https://pubmed.ncbi.nlm.nih.gov/26053121/)
9. Bush RM, Bender CA, Subbarao K, Cox NJ, Fitch WM. Predicting the evolution of human influenza A. *Science*. 1999; 286(5446):1921–1925. doi: [10.1126/science.286.5446.1921](https://doi.org/10.1126/science.286.5446.1921) PMID: [10583948](https://pubmed.ncbi.nlm.nih.gov/10583948/)
10. Łuksza M, Lässig M. A predictive fitness model for influenza. *Nature*. 2014; 507(7490):57–61. doi: [10.1038/nature13087](https://doi.org/10.1038/nature13087) PMID: [24572367](https://pubmed.ncbi.nlm.nih.gov/24572367/)
11. Neher RA, Russell CA, Shraiman BI. Predicting evolution from the shape of genealogical trees. *eLife*. 2014; 3:e03568. doi: [10.7554/eLife.03568](https://doi.org/10.7554/eLife.03568)
12. Murcia PR, Baillie GJ, Daly J, Elton D, Jervis C, Mumford JA, et al. Intra- and Interhost Evolutionary Dynamics of Equine Influenza Virus. *Journal of Virology*. 2010; 84(14):6943–6954. doi: [10.1128/JVI.00112-10](https://doi.org/10.1128/JVI.00112-10) PMID: [20444896](https://pubmed.ncbi.nlm.nih.gov/20444896/)
13. Varble A, Albrecht RA, Backes S, Crumiller M, Bouvier NM, Sachs D, et al. Influenza A Virus Transmission Bottlenecks Are Defined by Infection Route and Recipient Host. *Cell Host and Microbe*. 2014; 16(5):691–700. doi: [10.1016/j.chom.2014.09.020](https://doi.org/10.1016/j.chom.2014.09.020) PMID: [25456074](https://pubmed.ncbi.nlm.nih.gov/25456074/)
14. Murcia PR, Hughes J, Battista P, Lloyd L, Baillie GJ, Ramirez-Gonzalez RH, et al. Evolution of an Eurasian Avian-like Influenza Virus in Naïve and Vaccinated Pigs. *PLoS Pathogens*. 2012; 8(5):e1002730. doi: [10.1371/journal.ppat.1002730](https://doi.org/10.1371/journal.ppat.1002730) PMID: [22693449](https://pubmed.ncbi.nlm.nih.gov/22693449/)
15. Sobel Leonard A, McClain M, Smith G, Wentworth D, Halpin R, Lin X, et al. Deep Sequencing of Influenza A Virus from a Human Challenge Study Reveals a Selective Bottleneck and Only Limited Intra-host Genetic Diversification. *J Virol*, in press. 2016;. doi: [10.1128/JVI.01657-16](https://doi.org/10.1128/JVI.01657-16) PMID: [27707932](https://pubmed.ncbi.nlm.nih.gov/27707932/)
16. Dinis JM, Florek NW, Fatola OO, Moncla LH, Mutschler JP, Charlier OK, et al. Deep Sequencing Reveals Potential Antigenic Variants at Low Frequencies in Influenza A Virus-Infected Humans. *Journal of Virology*. 2016; 90(7):3355–3365. doi: [10.1128/JVI.03248-15](https://doi.org/10.1128/JVI.03248-15) PMID: [26739054](https://pubmed.ncbi.nlm.nih.gov/26739054/)

17. Imai M, Watanabe T, Hatta M, Das SC, Ozawa M, Shinya K, et al. Experimental adaptation of an influenza H5 HA confers respiratory droplet transmission to a reassortant H5 HA/H1N1 virus in ferrets. *Nature*. 2012; 486(7403):420–428. doi: [10.1038/nature10831](https://doi.org/10.1038/nature10831) PMID: [22722205](https://pubmed.ncbi.nlm.nih.gov/22722205/)
18. Herfst S, Schrauwen EJA, Linster M, Chutinimitkul S, de Wit E, Munster VJ, et al. Airborne Transmission of Influenza A/H5N1 Virus Between Ferrets. *Science*. 2012; 336(6088):1534–1541. doi: [10.1126/science.1213362](https://doi.org/10.1126/science.1213362) PMID: [22723413](https://pubmed.ncbi.nlm.nih.gov/22723413/)
19. Lakdawala SS, Jayaraman A, Halpin RA, Lamirande EW, Shih AR, Stockwell TB, et al. The soft palate is an important site of adaptation for transmissible influenza viruses. *Nature*. 2015; 526(7571):122–125. doi: [10.1038/nature15379](https://doi.org/10.1038/nature15379) PMID: [26416728](https://pubmed.ncbi.nlm.nih.gov/26416728/)
20. Moncla LH, Zhong G, Nelson CW, Dinis JM, Mutschler J, Hughes AL, et al. Selective Bottlenecks Shape Evolutionary Pathways Taken during Mammalian Adaptation of a 1918-like Avian Influenza Virus. *Cell Host and Microbe*. 2016; 19(2):169–180. doi: [10.1016/j.chom.2016.01.011](https://doi.org/10.1016/j.chom.2016.01.011) PMID: [26867176](https://pubmed.ncbi.nlm.nih.gov/26867176/)
21. Koel BF, van der Vliet S, Burke DF, Bestebroer TM, Bharoto EE, Yasa IWW, et al. Antigenic variation of clade 2.1 H5N1 virus is determined by a few amino acid substitutions immediately adjacent to the receptor binding site. *mBio*. 2014; 5(3):e01070–14. doi: [10.1128/mBio.01070-14](https://doi.org/10.1128/mBio.01070-14) PMID: [24917596](https://pubmed.ncbi.nlm.nih.gov/24917596/)
22. Gong LI, Bloom JD. Epistatically Interacting Substitutions Are Enriched during Adaptive Protein Evolution. *PLoS Genetics*. 2014; 10(5):e1004328–10. doi: [10.1371/journal.pgen.1004328](https://doi.org/10.1371/journal.pgen.1004328) PMID: [24811236](https://pubmed.ncbi.nlm.nih.gov/24811236/)
23. Gong LI, Suchard MA, Bloom JD. Stability-mediated epistasis constrains the evolution of an influenza protein. *eLife*. 2013; 2:e00631. doi: [10.7554/eLife.00631](https://doi.org/10.7554/eLife.00631) PMID: [23682315](https://pubmed.ncbi.nlm.nih.gov/23682315/)
24. Kryazhimskiy S, Dushoff J, Bazykin GA, Plotkin JB. Prevalence of Epistasis in the Evolution of Influenza A Surface Proteins. *PLoS Genetics*. 2011; 7(2):e1001301. doi: [10.1371/journal.pgen.1001301](https://doi.org/10.1371/journal.pgen.1001301) PMID: [21390205](https://pubmed.ncbi.nlm.nih.gov/21390205/)
25. Kaverin NV, Gambaryan AS, Bovin NV, Rudneva IA, Shilov AA, Khodova OM, et al. Postreassortment changes in influenza A virus hemagglutinin restoring HA-NA functional match. *Virology*. 1998; 244(2):315–321. doi: [10.1006/viro.1998.9119](https://doi.org/10.1006/viro.1998.9119) PMID: [9601502](https://pubmed.ncbi.nlm.nih.gov/9601502/)
26. Mitnaul LJ, Matrosovich MN, Castrucci MR, Tuzikov AB, Bovin NV, Kobasa D, et al. Balanced hemagglutinin and neuraminidase activities are critical for efficient replication of influenza A virus. *Journal of Virology*. 2000; 74(13):6015–6020. doi: [10.1128/JVI.74.13.6015-6020.2000](https://doi.org/10.1128/JVI.74.13.6015-6020.2000) PMID: [10846083](https://pubmed.ncbi.nlm.nih.gov/10846083/)
27. Xu R, Zhu X, McBride R, Nycholat CM, Yu W, Paulson JC, et al. Functional balance of the hemagglutinin and neuraminidase activities accompanies the emergence of the 2009 H1N1 influenza pandemic. *Journal of Virology*. 2012; 86(17):9221–9232. doi: [10.1128/JVI.00697-12](https://doi.org/10.1128/JVI.00697-12) PMID: [22718832](https://pubmed.ncbi.nlm.nih.gov/22718832/)
28. Zeldovich KB, Liu P, Renzette N, Foll M, Pham ST, Venev SV, et al. Positive Selection Drives Preferred Segment Combinations during Influenza Virus Reassortment. *Molecular Biology and Evolution*. 2015; p. 1–14.
29. Holmes EC, Ghedin E, Miller N, Taylor J, Bao Y, St George K, et al. Whole-Genome Analysis of Human Influenza A Virus Reveals Multiple Persistent Lineages and Reassortment among Recent H3N2 Viruses. *PLoS Biology*. 2005; 3(9):e300. doi: [10.1371/journal.pbio.0030300](https://doi.org/10.1371/journal.pbio.0030300) PMID: [16026181](https://pubmed.ncbi.nlm.nih.gov/16026181/)
30. Ince WL, Gueye-Mbaye A, Bennink JR, Yewdell JW. Reassortment complements spontaneous mutation in influenza A virus NP and M1 genes to accelerate adaptation to a new host. *Journal of Virology*. 2013; 87(8):4330–4338. doi: [10.1128/JVI.02749-12](https://doi.org/10.1128/JVI.02749-12) PMID: [23365443](https://pubmed.ncbi.nlm.nih.gov/23365443/)
31. Marshall N, Priyamvada L, Ende Z, Steel J, Lowen AC. Influenza Virus Reassortment Occurs with High Frequency in the Absence of Segment Mismatch. *PLoS Pathogens*. 2013; 9(6):e1003421. doi: [10.1371/journal.ppat.1003421](https://doi.org/10.1371/journal.ppat.1003421) PMID: [23785286](https://pubmed.ncbi.nlm.nih.gov/23785286/)
32. Tao H, Steel J, Lowen AC. Intrahost Dynamics of Influenza Virus Reassortment. *Journal of Virology*. 2014; 88(13):7485–7492. doi: [10.1128/JVI.00715-14](https://doi.org/10.1128/JVI.00715-14) PMID: [24741099](https://pubmed.ncbi.nlm.nih.gov/24741099/)
33. Tao H, Li L, White MC, Steel J, Lowen AC. Influenza A Virus Coinfection through Transmission Can Support High Levels of Reassortment. *Journal of Virology*. 2015; 89(16):8453–8461. doi: [10.1128/JVI.01162-15](https://doi.org/10.1128/JVI.01162-15) PMID: [26041285](https://pubmed.ncbi.nlm.nih.gov/26041285/)
34. Murillo LN, Murillo MS, Perelson AS. Towards multiscale modeling of influenza infection. *Journal of Theoretical Biology*. 2013; 332(C):267–290. doi: [10.1016/j.jtbi.2013.03.024](https://doi.org/10.1016/j.jtbi.2013.03.024) PMID: [23608630](https://pubmed.ncbi.nlm.nih.gov/23608630/)
35. Kryazhimskiy S, Plotkin JB. The Population Genetics of dN/dS. *PLoS Genetics*. 2008; 4(12):e1000304. doi: [10.1371/journal.pgen.1000304](https://doi.org/10.1371/journal.pgen.1000304) PMID: [19081788](https://pubmed.ncbi.nlm.nih.gov/19081788/)
36. Mugal CF, Wolf JBW, Kaj I. Why Time Matters: Codon Evolution and the Temporal Dynamics of dN/dS. *Molecular Biology and Evolution*. 2013; 31(1):212–231. doi: [10.1093/molbev/mst192](https://doi.org/10.1093/molbev/mst192) PMID: [24129904](https://pubmed.ncbi.nlm.nih.gov/24129904/)
37. Boni MF, Zhou Y, Taubenberger JK, Holmes EC. Homologous recombination is very rare or absent in human influenza A virus. *Journal of Virology*. 2008; 82(10):4807–4811. doi: [10.1128/JVI.02683-07](https://doi.org/10.1128/JVI.02683-07) PMID: [18353939](https://pubmed.ncbi.nlm.nih.gov/18353939/)
38. Han GZ, Worobey M. Homologous recombination in negative sense RNA viruses. *Viruses*. 2011; 3(8):1358–1373. doi: [10.3390/v3081358](https://doi.org/10.3390/v3081358) PMID: [21994784](https://pubmed.ncbi.nlm.nih.gov/21994784/)

39. Lam TTY, Chong YL, Shi M, Hon CC, Li J, Martin DP, et al. Systematic phylogenetic analysis of influenza A virus reveals many novel mosaic genome segments. *Infection, genetics and evolution: journal of molecular epidemiology and evolutionary genetics in infectious diseases*. 2013; 18:367–378. doi: [10.1016/j.meegid.2013.03.015](https://doi.org/10.1016/j.meegid.2013.03.015) PMID: [23548803](https://pubmed.ncbi.nlm.nih.gov/23548803/)
40. de Visser JAGM, Krug J. Empirical fitness landscapes and the predictability of evolution. *Nature Reviews Genetics*. 2014; 15(7):480–490. doi: [10.1038/nrg3744](https://doi.org/10.1038/nrg3744) PMID: [24913663](https://pubmed.ncbi.nlm.nih.gov/24913663/)
41. Illingworth CJR, Fischer A, Mustonen V. Identifying selection in the within-host evolution of influenza using viral sequence data. *PLoS Computational Biology*. 2014; 10(7):e1003755. doi: [10.1371/journal.pcbi.1003755](https://doi.org/10.1371/journal.pcbi.1003755) PMID: [25080215](https://pubmed.ncbi.nlm.nih.gov/25080215/)
42. Illingworth CJR. Fitness Inference from Short-Read Data: Within-Host Evolution of a Reassortant H5N1 Influenza Virus. *Molecular Biology and Evolution*. 2015; 32(11):3012–3026. doi: [10.1093/molbev/msv171](https://doi.org/10.1093/molbev/msv171) PMID: [26243288](https://pubmed.ncbi.nlm.nih.gov/26243288/)
43. Iqbal M, Essen SC, Xiao H, Brookes SM, Brown IH, McCauley JW. Selection of variant viruses during replication and transmission of H7N1 viruses in chickens and turkeys. *Virology*. 2012; 433(2):282–295. doi: [10.1016/j.virol.2012.08.001](https://doi.org/10.1016/j.virol.2012.08.001) PMID: [22944111](https://pubmed.ncbi.nlm.nih.gov/22944111/)
44. Hoelzer K, Murcia PR, Baillie GJ, Wood JLN, Metzger SM, Osterrieder N, et al. Intrahost evolutionary dynamics of canine influenza virus in naive and partially immune dogs. *Journal of Virology*. 2010; 84(10):5329–5335. doi: [10.1128/JVI.02469-09](https://doi.org/10.1128/JVI.02469-09) PMID: [20219907](https://pubmed.ncbi.nlm.nih.gov/20219907/)
45. Murcia PR, Baillie GJ, Stack JC, Jervis C, Elton D, Mumford JA, et al. Evolution of Equine Influenza Virus in Vaccinated Horses. *Journal of Virology*. 2013; 87(8):4768–4771. doi: [10.1128/JVI.03379-12](https://doi.org/10.1128/JVI.03379-12) PMID: [23388708](https://pubmed.ncbi.nlm.nih.gov/23388708/)
46. Easterday BC. Animals in the Influenza World. *Philosophical Transactions of the Royal Society B: Biological Sciences*. 1980; 288(1029):433–437. doi: [10.1098/rstb.1980.0020](https://doi.org/10.1098/rstb.1980.0020)
47. Belser JA, Katz JM, Tumpey TM. The ferret as a model organism to study influenza A virus infection. *Disease models & mechanisms*. 2011; 4(5):575–579. doi: [10.1242/dmm.007823](https://doi.org/10.1242/dmm.007823)
48. Buhnerkempe MG, Gostic K, Park M, Ahsan P. Mapping influenza transmission in the ferret model to transmission in humans. *eLife*. 2015; 4:e07969. doi: [10.7554/eLife.07969](https://doi.org/10.7554/eLife.07969)
49. Zaas AK, Chen M, Varkey J, Veldman T, Hero AO, Lucas J, et al. Gene expression signatures diagnose influenza and other symptomatic respiratory viral infections in humans. *Cell Host and Microbe*. 2009; 6(3):207–217. doi: [10.1016/j.chom.2009.07.006](https://doi.org/10.1016/j.chom.2009.07.006) PMID: [19664979](https://pubmed.ncbi.nlm.nih.gov/19664979/)
50. Huang Y, Zaas AK, Rao A, Dobigeon N, Woolf PJ, Veldman T, et al. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. *PLoS Genetics*. 2011; 7(8):e1002234. doi: [10.1371/journal.pgen.1002234](https://doi.org/10.1371/journal.pgen.1002234) PMID: [21901105](https://pubmed.ncbi.nlm.nih.gov/21901105/)
51. Moody MA, Zhang R, Walter EB, Woods CW, Ginsburg GS, McClain MT, et al. H3N2 influenza infection elicits more cross-reactive and less clonally expanded anti-hemagglutinin antibodies than influenza vaccination. *PLoS ONE*. 2011; 6(10):e25797. doi: [10.1371/journal.pone.0025797](https://doi.org/10.1371/journal.pone.0025797) PMID: [22039424](https://pubmed.ncbi.nlm.nih.gov/22039424/)
52. Wilkinson TM, Li CKF, Chui CSC, Huang AKY, Perkins M, Liebner JC, et al. Preexisting influenza-specific CD4+ T cells correlate with disease protection against influenza challenge in humans. *Nature Medicine*. 2012; 18(2):274–280. doi: [10.1038/nm.2612](https://doi.org/10.1038/nm.2612) PMID: [22286307](https://pubmed.ncbi.nlm.nih.gov/22286307/)
53. Zaas AK, Burke T, Chen M, McClain M, Nicholson B, Veldman T, et al. A host-based RT-PCR gene expression signature to identify acute respiratory viral infection. *Science Translational Medicine*. 2013; 5(203):203ra126. doi: [10.1126/scitranslmed.3006280](https://doi.org/10.1126/scitranslmed.3006280) PMID: [24048524](https://pubmed.ncbi.nlm.nih.gov/24048524/)
54. Woods CW, McClain MT, Chen M, Zaas AK, Nicholson BP, Varkey J, et al. A host transcriptional signature for presymptomatic detection of infection in humans exposed to influenza H1N1 or H3N2. *PLoS ONE*. 2013; 8(1):e52198. doi: [10.1371/journal.pone.0052198](https://doi.org/10.1371/journal.pone.0052198) PMID: [23326326](https://pubmed.ncbi.nlm.nih.gov/23326326/)
55. Greenbaum BD, Ghedin E. ScienceDirectViral evolution: beyond drift and shift. *Current Opinion in Microbiology*. 2015; 26:109–115. doi: [10.1016/j.mib.2015.06.015](https://doi.org/10.1016/j.mib.2015.06.015) PMID: [26189048](https://pubmed.ncbi.nlm.nih.gov/26189048/)
56. Zaas AK, Chen M, Varkey J, Veldman T, Hero AO III, Lucas J, et al. Gene Expression Signatures Diagnose Influenza and Other Symptomatic Respiratory Viral Infections in Humans. *Cell Host and Microbe*. 2009; 6(3):207–217. doi: [10.1016/j.chom.2009.07.006](https://doi.org/10.1016/j.chom.2009.07.006) PMID: [19664979](https://pubmed.ncbi.nlm.nih.gov/19664979/)
57. Moody MA, Zhang R, Walter EB, Woods CW, Ginsburg GS, McClain MT, et al. H3N2 influenza infection elicits more cross-reactive and less clonally expanded anti-hemagglutinin antibodies than influenza vaccination. *PLoS ONE*. 2011; 6(10):e25797. doi: [10.1371/journal.pone.0025797](https://doi.org/10.1371/journal.pone.0025797) PMID: [22039424](https://pubmed.ncbi.nlm.nih.gov/22039424/)
58. Huang Y, Zaas AK, Rao A, Dobigeon N, Woolf PJ, Veldman T, et al. Temporal dynamics of host molecular responses differentiate symptomatic and asymptomatic influenza a infection. *PLoS Genetics*. 2011; 7(8):e1002234. doi: [10.1371/journal.pgen.1002234](https://doi.org/10.1371/journal.pgen.1002234) PMID: [21901105](https://pubmed.ncbi.nlm.nih.gov/21901105/)
59. Zhou B, Donnelly ME, Scholes DT, St George K, Hatta M, Kawaoka Y, et al. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and swine origin human



- influenza A viruses. *Journal of Virology*. 2009; 83(19):10309–10313. doi: [10.1128/JVI.01109-09](https://doi.org/10.1128/JVI.01109-09) PMID: [19605485](https://pubmed.ncbi.nlm.nih.gov/19605485/)
60. Zhou B, Wentworth DE. Influenza A virus molecular virology techniques. *Methods in molecular biology* (Clifton, NJ). 2012; 865:175–192. doi: [10.1007/978-1-61779-621-0\\_11](https://doi.org/10.1007/978-1-61779-621-0_11)
  61. Foll M, Poh YP, Renzette N, Ferrer-Admetlla A, Bank C, Shim H, et al. Influenza Virus Drug Resistance: A Time-Sampled Population Genetics Perspective. *PLoS Genetics*. 2014; 10(2):e1004185. doi: [10.1371/journal.pgen.1004185](https://doi.org/10.1371/journal.pgen.1004185) PMID: [24586206](https://pubmed.ncbi.nlm.nih.gov/24586206/)
  62. Ghedin E, Holmes EC, DePasse JV, Pinilla LT, Fitch A, Hamelin ME, et al. Presence of oseltamivir-resistant pandemic A/H1N1 minor variants before drug therapy with subsequent selection and transmission. *The Journal of infectious diseases*. 2012; 206(10):1504–1511. doi: [10.1093/infdis/jis571](https://doi.org/10.1093/infdis/jis571) PMID: [22966122](https://pubmed.ncbi.nlm.nih.gov/22966122/)
  63. Illingworth CJR. SAMFIRE: multi-locus variant calling for time-resolved sequence data. *Bioinformatics*. 2016; p. btw205.
  64. Stevens J, Chen LM, Carney PJ, Garten R, Foust A, Le J, et al. Receptor specificity of influenza A H3N2 viruses isolated in mammalian cells and embryonated chicken eggs. *Journal of Virology*. 2010; 84(16):8287–8299. doi: [10.1128/JVI.00058-10](https://doi.org/10.1128/JVI.00058-10) PMID: [20519409](https://pubmed.ncbi.nlm.nih.gov/20519409/)
  65. Skowronski DM, Janjua NZ, De Serres G, Sabaiduc S, Eshaghi A, Dickinson JA, et al. Low 2012–13 Influenza Vaccine Effectiveness Associated with Mutation in the Egg-Adapted H3N2 Vaccine Strain Not Antigenic Drift in Circulating Viruses. *PLoS ONE*. 2014; 9(3):e92153–15. doi: [10.1371/journal.pone.0092153](https://doi.org/10.1371/journal.pone.0092153) PMID: [24667168](https://pubmed.ncbi.nlm.nih.gov/24667168/)
  66. Nakowitsch S, Wolschek M, Morokutti A, Ruthsatz T, Krenn BM, Ferko B, et al. Mutations affecting the stability of the haemagglutinin molecule impair the immunogenicity of live attenuated H3N2 intranasal influenza vaccine candidates lacking NS1. *Vaccine*. 2011; 29(19):3517–3524. doi: [10.1016/j.vaccine.2011.02.100](https://doi.org/10.1016/j.vaccine.2011.02.100) PMID: [21406268](https://pubmed.ncbi.nlm.nih.gov/21406268/)
  67. Kass RE, Raftery AE. Bayes factors. *Journal of the American Statistical Association*. 1995; 90(430):773–795. doi: [10.1080/01621459.1995.10476572](https://doi.org/10.1080/01621459.1995.10476572)
  68. Kiso M, Mitamura K, Sakai-Tagawa Y, Shiraishi K, Kawakami C, Kimura K, et al. Resistant influenza A viruses in children treated with oseltamivir: descriptive study. *The Lancet*. 2004; 364(9436):759–765. doi: [10.1016/S0140-6736\(04\)16934-1](https://doi.org/10.1016/S0140-6736(04)16934-1)
  69. McClain MT, Nicholson BP, Park LP, Liu TY, Hero AO, Tsalik EL, et al. A Genomic Signature of Influenza Infection Shows Potential for Presymptomatic Detection, Guiding Early Therapy, and Monitoring Clinical Responses. *Open forum infectious diseases*. 2016; 3(1):ofw007. doi: [10.1093/ofid/ofw007](https://doi.org/10.1093/ofid/ofw007) PMID: [26933666](https://pubmed.ncbi.nlm.nih.gov/26933666/)
  70. Fujii Y, Goto H, Watanabe T, Yoshida T, Kawaoka Y. Selective incorporation of influenza virus RNA segments into virions. *Proceedings of the National Academy of Sciences of the United States of America*. 2003; 100(4):2002–2007. doi: [10.1073/pnas.0437772100](https://doi.org/10.1073/pnas.0437772100) PMID: [12574509](https://pubmed.ncbi.nlm.nih.gov/12574509/)
  71. Marsh GA, Rabadan R, Levine AJ, Palese P. Highly conserved regions of influenza A virus polymerase gene segments are critical for efficient viral RNA packaging. *Journal of Virology*. 2008; 82(5):2295–2304. doi: [10.1128/JVI.02267-07](https://doi.org/10.1128/JVI.02267-07) PMID: [18094182](https://pubmed.ncbi.nlm.nih.gov/18094182/)
  72. Hutchinson EC, von Kirchbach JC, Gog JR, Digard P. Genome packaging in influenza A virus. *Journal of General Virology*. 2010; 91(Pt 2):313–328. doi: [10.1099/vir.0.017608-0](https://doi.org/10.1099/vir.0.017608-0) PMID: [19955561](https://pubmed.ncbi.nlm.nih.gov/19955561/)
  73. Gulyaev AP, Foucher RAM, Olsthoorn RCL. Influenza virus RNA Structure: Unique and Common Features. *International Reviews of Immunology*. 2010; 29(6):533–556. doi: [10.3109/08830185.2010.507828](https://doi.org/10.3109/08830185.2010.507828) PMID: [20923332](https://pubmed.ncbi.nlm.nih.gov/20923332/)
  74. Brower-Sinning R, Carter DM, Crevar CJ, Ghedin E, Ross TM, Benos PV. The role of RNA folding free energy in the evolution of the polymerase genes of the influenza A virus. *Genome Biology*. 2009; 10(2):R18. doi: [10.1186/gb-2009-10-2-r18](https://doi.org/10.1186/gb-2009-10-2-r18) PMID: [19216739](https://pubmed.ncbi.nlm.nih.gov/19216739/)
  75. Soszynska-Jozwiak M, Michalak P, Moss WN, Kierzek R, Kierzek E. A Conserved Secondary Structural Element in the Coding Region of the Influenza A Virus Nucleoprotein (NP) mRNA Is Important for the Regulation of Viral Proliferation. *PLoS ONE*. 2015; 10(10):e0141132–16. doi: [10.1371/journal.pone.0141132](https://doi.org/10.1371/journal.pone.0141132) PMID: [26488402](https://pubmed.ncbi.nlm.nih.gov/26488402/)
  76. Moss WN, Priore SF, Turner DH. Identification of potential conserved RNA secondary structure throughout influenza A coding regions. *RNA*. 2011; 17(6):991–1011. doi: [10.1261/rna.2619511](https://doi.org/10.1261/rna.2619511) PMID: [21536710](https://pubmed.ncbi.nlm.nih.gov/21536710/)
  77. Lubeck MD, Palese P, Schulman JL. Nonrandom association of parental genes in influenza A virus recombinants. *Virology*. 1979; 95(1):269–274. doi: [10.1016/0042-6822\(79\)90430-6](https://doi.org/10.1016/0042-6822(79)90430-6) PMID: [442543](https://pubmed.ncbi.nlm.nih.gov/442543/)
  78. Greenbaum BD, Li OTW, Poon LLM, Levine AJ, Rabadan R. Viral reassortment as an information exchange between viral segments. *Proceedings of the National Academy of Sciences of the United States of America*. 2012; 109(9):3341–3346. doi: [10.1073/pnas.1113300109](https://doi.org/10.1073/pnas.1113300109) PMID: [22331898](https://pubmed.ncbi.nlm.nih.gov/22331898/)

79. Rabadan R, Levine AJ, Krasnitz M. Non-random reassortment in human influenza A viruses. *Influenza and Other Respiratory Viruses*. 2008; 2(1):9–22. doi: [10.1111/j.1750-2659.2007.00030.x](https://doi.org/10.1111/j.1750-2659.2007.00030.x) PMID: [19453489](https://pubmed.ncbi.nlm.nih.gov/19453489/)
80. Dudas G, Bedford T, Lycett S, Rambaut A. Reassortment between Influenza B Lineages and the Emergence of a Coadapted PB1-PB2-HA Gene Complex. *Molecular Biology and Evolution*. 2014; 32(1):162–172. doi: [10.1093/molbev/msu287](https://doi.org/10.1093/molbev/msu287) PMID: [25323575](https://pubmed.ncbi.nlm.nih.gov/25323575/)
81. Zeldovich KB, Liu P, Renzette N, Foll M, Pham ST, Venev SV, et al. Positive Selection Drives Preferred Segment Combinations during Influenza Virus Reassortment. *Molecular Biology and Evolution*. 2015; 32(6):1519–1532. doi: [10.1093/molbev/msv044](https://doi.org/10.1093/molbev/msv044) PMID: [25713211](https://pubmed.ncbi.nlm.nih.gov/25713211/)
82. Khiabanian H, Trifonov V, Rabadan R. Reassortment Patterns in Swine Influenza Viruses. *PLoS ONE*. 2009; 4(10):e7366–7. doi: [10.1371/journal.pone.0007366](https://doi.org/10.1371/journal.pone.0007366) PMID: [19809504](https://pubmed.ncbi.nlm.nih.gov/19809504/)
83. Frost SD, Dumaurier MJ, Wain-Hobson S, Brown AJ. Genetic drift and within-host metapopulation dynamics of HIV-1 infection. *Proceedings of the National Academy of Sciences of the United States of America*. 2001; 98(12):6975–6980. doi: [10.1073/pnas.131056998](https://doi.org/10.1073/pnas.131056998) PMID: [11381143](https://pubmed.ncbi.nlm.nih.gov/11381143/)
84. Lythgoe KA, Blanquart F, Pellis L, Fraser C. Large Variations in HIV-1 Viral Load Explained by Shifting-Mosaic Metapopulation Dynamics. *PLoS Biology*. 2016; 14(10):e1002567–25. doi: [10.1371/journal.pbio.1002567](https://doi.org/10.1371/journal.pbio.1002567) PMID: [27706164](https://pubmed.ncbi.nlm.nih.gov/27706164/)
85. Baer A, Kehn-Hall K. Viral concentration determination through plaque assays: using traditional and novel overlay systems. *Journal of visualized experiments: JoVE*. 2014;(93):e52065. doi: [10.3791/52065](https://doi.org/10.3791/52065) PMID: [25407402](https://pubmed.ncbi.nlm.nih.gov/25407402/)
86. Schwarz G. Estimating the Dimension of a Model. *The Annals of Statistics*. 1978; 6(2):461–464. doi: [10.1214/aos/1176344136](https://doi.org/10.1214/aos/1176344136)
87. Baccam P, Beauchemin C, Macken CA, Hayden FG, Perelson AS. Kinetics of Influenza A Virus Infection in Humans. *Journal of Virology*. 2006; 80(15):7590–7599. doi: [10.1128/JVI.01623-05](https://doi.org/10.1128/JVI.01623-05) PMID: [16840338](https://pubmed.ncbi.nlm.nih.gov/16840338/)
88. Hill WG, Robertson A. The effect of linkage on limits to artificial selection. *Genetics Research*. 1966; 8(3):269–294. doi: [10.1017/S0016672300010156](https://doi.org/10.1017/S0016672300010156)
89. Illingworth CJR, Mustonen V. Distinguishing Driver and Passenger Mutations in an Evolutionary History Categorized by Interference. *Genetics*. 2011; 189(3):989–1000. doi: [10.1534/genetics.111.133975](https://doi.org/10.1534/genetics.111.133975) PMID: [21900272](https://pubmed.ncbi.nlm.nih.gov/21900272/)
90. Sanjuan R, Nebot MR, Chirico N, Mansky LM, Belshaw R. Viral Mutation Rates. *Journal of Virology*. 2010; 84(19):9733–9748. doi: [10.1128/JVI.00694-10](https://doi.org/10.1128/JVI.00694-10) PMID: [20660197](https://pubmed.ncbi.nlm.nih.gov/20660197/)
91. Rogers MB, Song T, Sebra R, Greenbaum BD, Hamelin ME, Fitch A, et al. Intra-host Dynamics of Antiviral Resistance in Influenza A Virus Reflect Complex Patterns of Segment Linkage, Reassortment, and Natural Selection. *mBio*. 2015; 6(2):e02464–14–8. doi: [10.1128/mBio.02464-14](https://doi.org/10.1128/mBio.02464-14) PMID: [25852163](https://pubmed.ncbi.nlm.nih.gov/25852163/)
92. McDonald SM, Nelson MI, Turner PE, Patton JT. Reassortment in segmented RNA viruses: mechanisms and outcomes. *Nature Publishing Group*. 2016; 14(7):448–460.
93. Wright S. Evolution in Mendelian Populations. *Genetics*. 1931; 16(2):97–159. PMID: [17246615](https://pubmed.ncbi.nlm.nih.gov/17246615/)