Positive Selection Drives Preferred Segment Combinations during Influenza Virus Reassortment

Konstantin B. Zeldovich,[1] Ping Liu,[2] Nicholas Renzette,[3] Matthieu Foll,[4,5] Serena T. Pham,[2] Sergey V. Venev,[1] Glen R. Gallagher,[2] Daniel N. Bolon,[6] Evelyn A. Kurt-Jones,[2] Jeffrey D. Jensen,[4,5] Daniel R. Caffrey,[2] Celia A. Schiffer,[6] Timothy F. Kowalik,[3] Jennifer P. Wang,*[,†,2] and Robert W. Finberg[†,2]

[1]Program in Bioinformatics and Integrative Biology, University of Massachusetts Medical School, Worcester

[2]Department of Medicine, University of Massachusetts Medical School, Worcester

[3]Department of Microbiology and Physiological Systems, University of Massachusetts Medical School, Worcester

[4]École Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland

[5]Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland

[6]Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester

[†]These authors contributed equally to this work.

***Corresponding author:** E-mail: jennifer.wang@umassmed.edu.

Associate editor: Joel Dudley

Abstract

Influenza A virus (IAV) has a segmented genome that allows for the exchange of genome segments between different strains. This reassortment accelerates evolution by breaking linkage, helping IAV cross species barriers to potentially create highly virulent strains. Challenges associated with monitoring the process of reassortment in molecular detail have limited our understanding of its evolutionary implications. We applied a novel deep sequencing approach with quantitative analysis to assess the in vitro temporal evolution of genomic reassortment in IAV. The combination of H1N1 and H3N2 strains reproducibly generated a new H1N2 strain with the hemagglutinin and nucleoprotein segments originating from H1N1 and the remaining six segments from H3N2. By deep sequencing the entire viral genome, we monitored the evolution of reassortment, quantifying the relative abundance of all IAV genome segments from the two parent strains over time and measuring the selection coefficients of the reassorting segments. Additionally, we observed several mutations coemerging with reassortment that were not found during passaging of pure parental IAV strains. Our results demonstrate how reassortment of the segmented genome can accelerate viral evolution in IAV, potentially enabled by the emergence of a small number of individual mutations.

Key words: **influenza, reassortment, evolution, deep sequencing, mutations.**

Article

Introduction

Influenza A virus (IAV) presents a major and persistent challenge to public health. A distinguishing feature of IAV is its segmented genome, comprised of eight RNA molecules (McGeoch et al. 1976) that encode at least 12 proteins (Jagger et al. 2012). The segmented genome allows IAV to undergo reassortment or segment mixing when the host is coinfected with two or more influenza strains. Reassortment can generate novel IAV strains with enhanced pathogenicity (Li et al. 2010), facilitating crossing of species barriers (Kawaoka et al. 1989) and thereby contributing to major influenza pandemics (Lindstrom et al. 2004). The molecular biology of influenza virus reassortment remains poorly understood but likely involves multiple mechanisms that select for functional and physical compatibility of highly divergent segments within a novel reassortant strain. The influenza viral genome contains packaging signals, specific sequences that are responsible for packaging the eight RNA segments together into infective virions (Luytjes et al. 1989; Fujii et al. 2003, 2005; Essere et al. 2013). Mutations in these signals can destroy the proper packaging of the virions (Gog et al. 2007) that is required for generating a viable reassortant after coinfection. In addition to packaging requirements, functional interactions between certain IAV proteins, especially hemagglutinin (HA) and neuraminidase (NA) (Wagner et al. 2002), may affect the viability of reassortants.

Various experimental and bioinformatic strategies have been employed to profile reassortment patterns. In traditional reassortment experiments, two influenza virus strains are used for coinfection either in cell culture (Lubeck et al. 1979) or in animal models (Angel et al. 2013). Alternatively, reverse genetics can be utilized in the generation of some (Chen et al. 2008) or all ($2^8 - 2$, or 254) (Li et al. 2010) of the possible reassortants between two strains which again can be monitored either in cell culture or in animal models. As sequence data from influenza viral genomes become increasingly available, reassortment can be studied by generating coalescent trees for each of the eight influenza viral segments and examining differing evolutionary histories between the segments (Lindstrom et al. 2004; Ghedin et al. 2005; Neverov et al. 2014).

Two important issues regarding influenza virus reassortment remain unresolved. First, the population dynamics of segments in the coinfecting viruses, including the temporal evolution of the viral population, is not well understood. Second, the effect of reassortment on the overall fitness landscape (Wright 1932), including the selection of mutations due to the reshuffling of the genetic background, has never been established. A clear understanding of associations between specific mutation patterns and reassortment would have broad implications on public health and potentially could be used for measuring the fitness advantages of genome recombination, and ultimately, sexual reproduction (Kilbourne 1981; Barton and Charlesworth 1998; Simon-Loriere and Holmes 2011).

To address these unresolved aspects, we employed a novel strategy to monitor reassortment of IAV in vitro through deep sequencing of the whole viral genome following coinfection. Madin-Darby canine kidney (MDCK) cells were coinfected with influenza A/Brisbane/59/2007 (H1N1) and A/Brisbane/10/2007 (H3N2) strains. Samples from each passage were deep sequenced using a universal influenza primer set, and the relative frequency of each viral segment from the two parent strains (16 total) was determined by bioinformatics analysis. By combining deep sequencing with bioinformatics methods, we were able to monitor the frequencies of influenza virus segments with high precision and dynamic range and quantify selection coefficients. Many of the shortcomings of traditional quantitative reverse transcription polymerase chain reaction (qRT-PCR) quantification of reassortment were avoided. Following coinfection, a novel H1N2 strain consistently emerged over serial passage, with reassortment-specific mutations in the viral genome distinct from those in the parent strain genetic background. Our detailed genetic analysis of the deep sequencing data permitted us to explicitly determine the selective advantage of reassortment during evolution of the influenza virus genome in vitro.

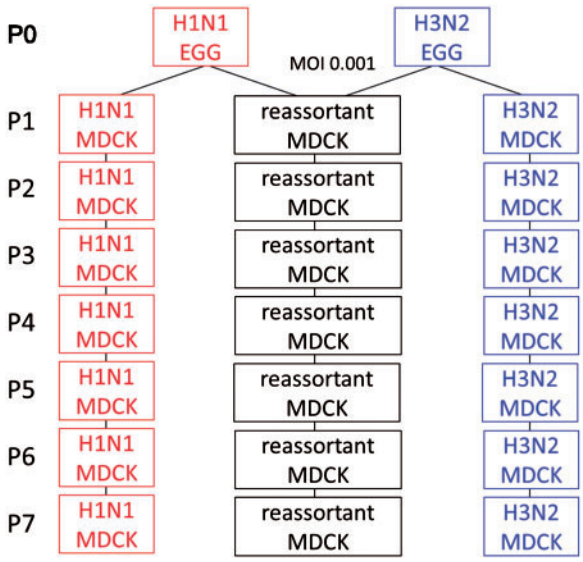

FIG. 1. Schematic of the serial passage experimental design. Pure strains of A/Brisbane/59/2007 (B59, H1N1) and A/Brisbane/10/2007 (B10, H3N2) were combined at a 1:1 ratio for a total MOI of 0.001 for the passage 1 (P1) reassortant in MDCK cells (black) and six additional passages were conducted in MDCK cells. In parallel, pure strains of B59 (red) and B10 (blue) were separately passaged following the same protocol.

Results

An In Vitro Platform to Examine Evolution during Influenza Reassortment

The overall design of the influenza evolution experiment is presented in figure 1. MDCK cells were coinfected with equal ratios of A/Brisbane/59/2007 (H1N1, henceforth denoted B59) and A/Brisbane/10/2007 (H3N2, henceforth denoted B10) influenza viruses, and the resultant viral populations were subjected to six additional serial passages. The experiment was performed in four complete, independent biological replicates, designated as Experiments 1–4. The initial infection was performed with a multiplicity of infection (MOI) of 5×10^{-4} for each parent strain (combined MOI of 10^{-3}). In Experiments 1–3, virus was continually passaged on cells for subsequent passages (P2–7) to avoid any freeze-thaw cycles. The amount of virus used to initiate a passage and the virus recovered at the end of each passage were subsequently empirically determined through plaque assays (see Materials and Methods). In one experiment (Experiment 4), a fixed MOI of 10^{-3} was used for every passage (P1–P7). The findings from all four experiments were highly similar, highlighting the robustness of the observations to variations in the passaging protocol. Pure B59 and B10 strains were similarly passaged in MDCK cells in parallel as nonreassortant controls.

Deep Sequencing-Based Quantitation of IAV Populations

We first established the unambiguity in the mapping of deep sequencing nucleotide reads to the genome of the two strains used in the experiments. The nucleotide sequence identity between the corresponding segments of the genomes of B59 and B10 strains varies between 57% and 91% (table 1). The 100-bp single-end reads were sufficiently long to map each read with high fidelity to either strain and to quantitate the

Table 1. Segment Lengths and Nucleotide Sequence Identity between A/Brisbane/59/2007 (B59, H1N1) and A/Brisbane/10/2007 (B10, H3N2) IAV Strains, Determined Using BLASTn.

Segment	Length in B59 (nt)	Length in B10 (nt)	Sequence Identity (%)
1, PB2	2,341	2,341	87
2, PB1	2,341	2,341	81
3, PA	2,233	2,233	89
4, HA	1,775	1,761	62
5, NP	1,565	1,565	89
6, NA	1,462	1,466	57
7, M1/M2	1,027	1,027	91
8, NS/NEP	890	890	89

relative abundance of viral genomic segments, as evident by the analysis of the control experiments below.

The fidelity of mapping was assessed by analyzing experimental controls of pure, nonreassorted strains (fig. 2). Upon sequencing a pure B59 (H1N1) strain, its segment 1 mapped to the B59 genome (red) at a high sequencing depth ($\geq 10^5$) but to the B10 genome (blue) only minimally (fig. 2*A*). Similarly, segment 1 from pure B10 (H3N2) strongly mapped to the B10 genome (blue) rather than the B59 genome (red) (fig. 2*B*). A localized homology was observed toward the ends of the segment, which harbor highly conserved promoter sequences. Sequencing depths for all eight segments for each strain are shown in supplementary figure S1, Supplementary Material online. The degree of cross-mapping was low in all experiments regardless of the passage number of the virus being sequenced.

To quantitate the relationship between sequencing depth and RNA abundance, B10 RNA and B59 RNA from pure strains were directly mixed in ratios ranging from 1:1,000 to 1,000:1 for deep sequencing and bioinformatics analysis (see Materials and Methods). The observed ratios of sequencing depth and the proportion of B10 and B59 RNA in the mix for each of the eight segments of the IAV genome exhibited strong correlation ($R^2 > 0.98$), as shown in figure 3.

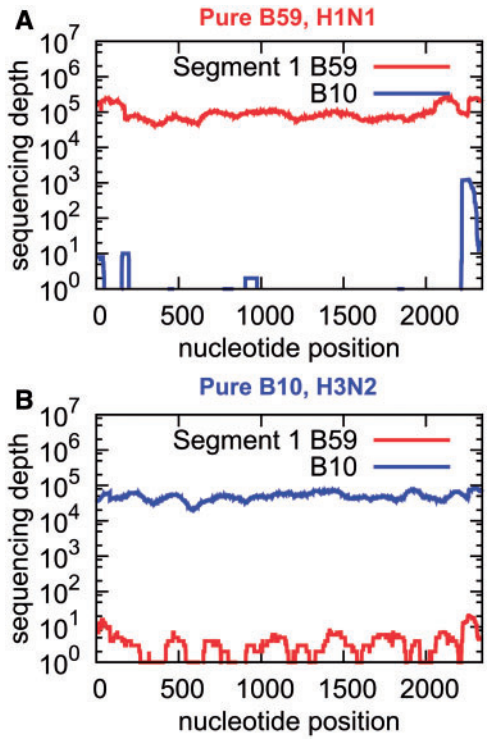

FIG. 2. Minimal cross-mapping is observed between pure strains of B10 (H3N2) and B59 (H1N1). (*A*) The sequencing depth for a B59 (H1N1) sample, segment 1, mapped to both the B59 (H1N1) and B10 (H3N2) genomes. Mapping to B59 produces a very high sequencing depth, whereas mapping to B10 is negligible. (*B*) The mapping of the B10 (H3N2) segment 1 to both the B59 (H1N1) and B10 (H3N2) genomes. The low degree of cross-mapping allows for accurate measurement of the relative abundance of each segment in the viral mixture. Data from passage 4 are shown.

To improve the discrimination between the two strains, the first and last 250 nt of each segment were excluded from the calculation of strain ratio, as these regions contain homologous sequences with lower strain specificity (fig. 2). Therefore, accurate measurement of the relative abundance of the corresponding segments from B59 and B10 in a viral mix over a wide range of mixing ratios can be performed with our deep sequencing and bioinformatics methodology. This same approach was then applied toward measuring segment fractions on serially passaged, pure (nonreassortant) viral strains (supplementary fig. S2, Supplementary Material online). When the B59 strain was passaged, the nonexistent B10 strain was spuriously detected at a median frequency of only 7×10^{-5}. Similarly, when B10 was passaged, the median fraction of B59 segments in the sample was estimated to be 5.9×10^{-5}. Both values are well below the sequencing error rate and are due to a small fraction of reads ambiguously mapping between the two strains' genomes. Thus, the deep sequencing and bioinformatics analyses described here can quantitatively discriminate between the two strains, with a dynamic range of segment ratio measurement exceeding 1:100.

Evolutionary Trajectory of Reassortment

Using this deep sequencing-based methodology, we tracked the frequencies of all 16 segments (eight from each parent virus) during reassortment following coinfection. Sequencing depths of the IAV segments from the first (P1) of seven serial passages of one reassortment experiment are shown in figure 4. The virus harvested at the first passage mainly consisted of the B10 (H3N2) strain, whereas the yield of B59 (H1N1) strain was relatively small. By the seventh passage

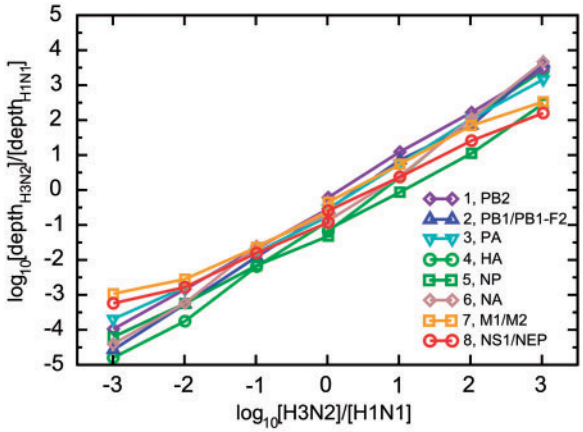

FIG. 3. Sequencing depth quantitatively reflects the relative abundance of viral RNA. A calibration curve was generated by extracting viral RNA from pure strains of B10 (H3N2) and B59 (H1N1) influenza virus and mixing at ratios of 1,000:1, 100:1, 10:1, 1:1, 1:10, 1:100, and 1:1,000, plotted on the *x*-axis as $\log_{10}$[H3N2]/[H1N1]. The ratio of the sequencing depths for the B10 (H3N2) and B59 (H1N1) viral RNAs is plotted on the *y*-axis on a $\log_{10}$ scale as $\log_{10}[\text{depth}_{H3N2}]/[\text{depth}_{H1N1}]$. A nearly perfect linear relationship allows extraction of the segment abundance ratio from the ratio of sequencing depths. PB2, polymerase basic 2; PB1/PB1-F2, polymerase basic 1; PA, polymerase acidic; M1/M2, matrix 1/matrix 2; NS1/NEP, nonstructural protein 1/nuclear export protein.

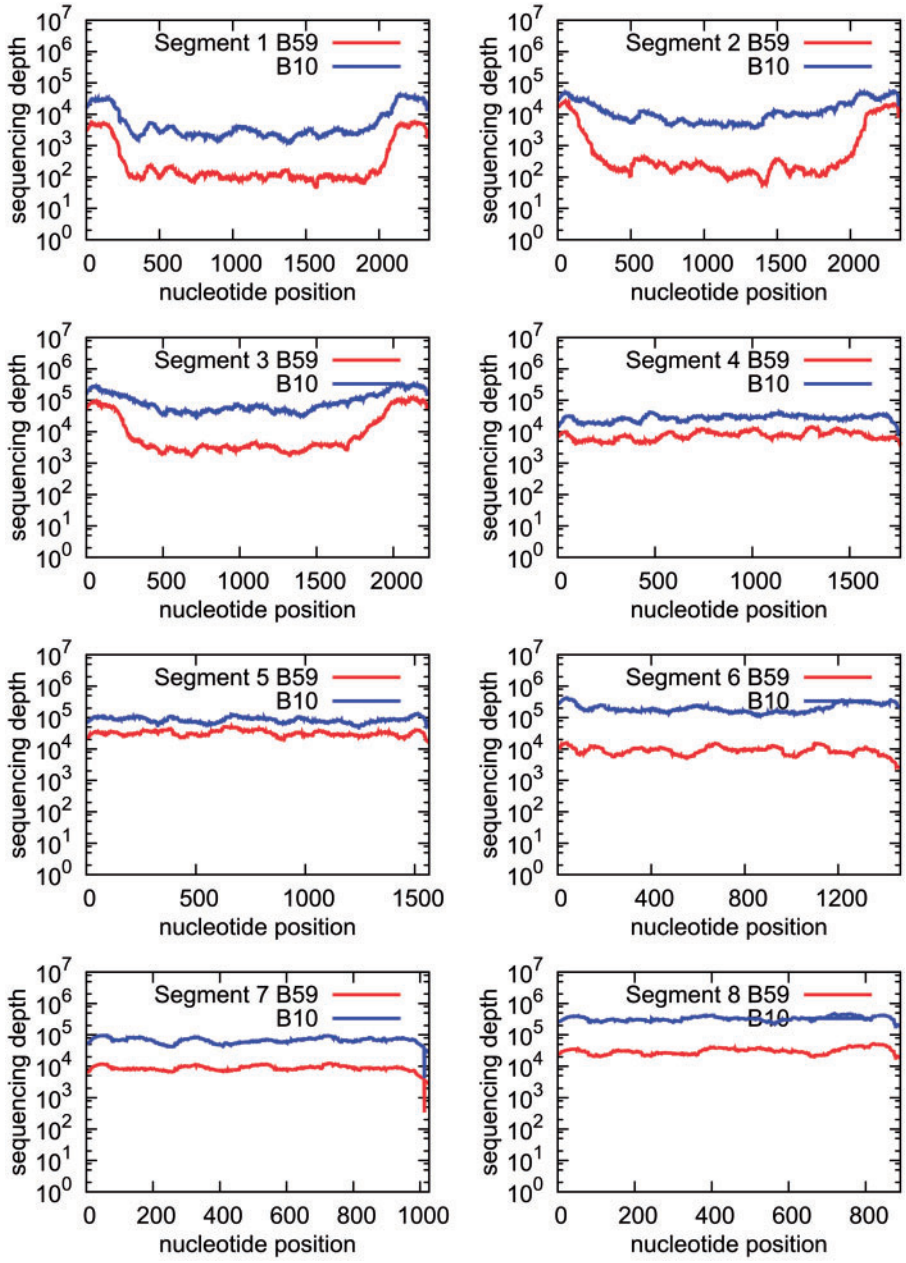

FIG. 4. Raw sequencing depth of all segments for a reassortant sample at passage 1 in MDCK cells. The sequenced viral populations mainly consist of the B10 (H3N2) strain (blue) in comparison to the B59 (H1N1) strain (red). The sequencing depths are shown on the *y*-axis, and each nucleotide position is represented on the *x*-axis. Data from Experiment 2 are shown.

(P7), the depths for segments 4 (HA) and 5 (nucleoprotein [NP]) became inverted; hence, these B59 (H1N1) segments outcompeted the B10 (H3N2) ones (fig. 5). Remarkably, B10 (H3N2) segment 4 nearly disappeared from the viral population. The other six segments, segment 1 (PB2), segment 2 (PB1), segment 3 (PA), segment 6 (NA), segment 7 (M), and segment 8 (NS), continued to be predominantly derived from B10 (H3N2).

The temporal change in the frequencies of all eight B59 (H1N1) segments in the viral population for every passage in four independent reassortment experiments is summarized in figure 6. The initial frequencies of B59 (H1N1) segments were consistently low in each case, but the frequencies of B59 segments 4 (HA) and 5 (NP) steadily increased over time in each experiment, achieving domination by P7. The rate of change of these frequencies differed slightly for each of the four experiments. Nevertheless, in all experiments, segments 4 and 5 from the B59 (H1N1) strain reproducibly outcompeted those from the B10 (H3N2) strain over time. Therefore, coinfection of MDCK cells by B10 (H3N2) and B59 (H1N1) influenza viruses produced a viable novel H1N2 reassortant viral strain comprising two segments encoding HA and NP from B59 and six segments from B10. Generation of this novel H1N2 strain required several serial passages, the temporal evolution of which would not be observed in

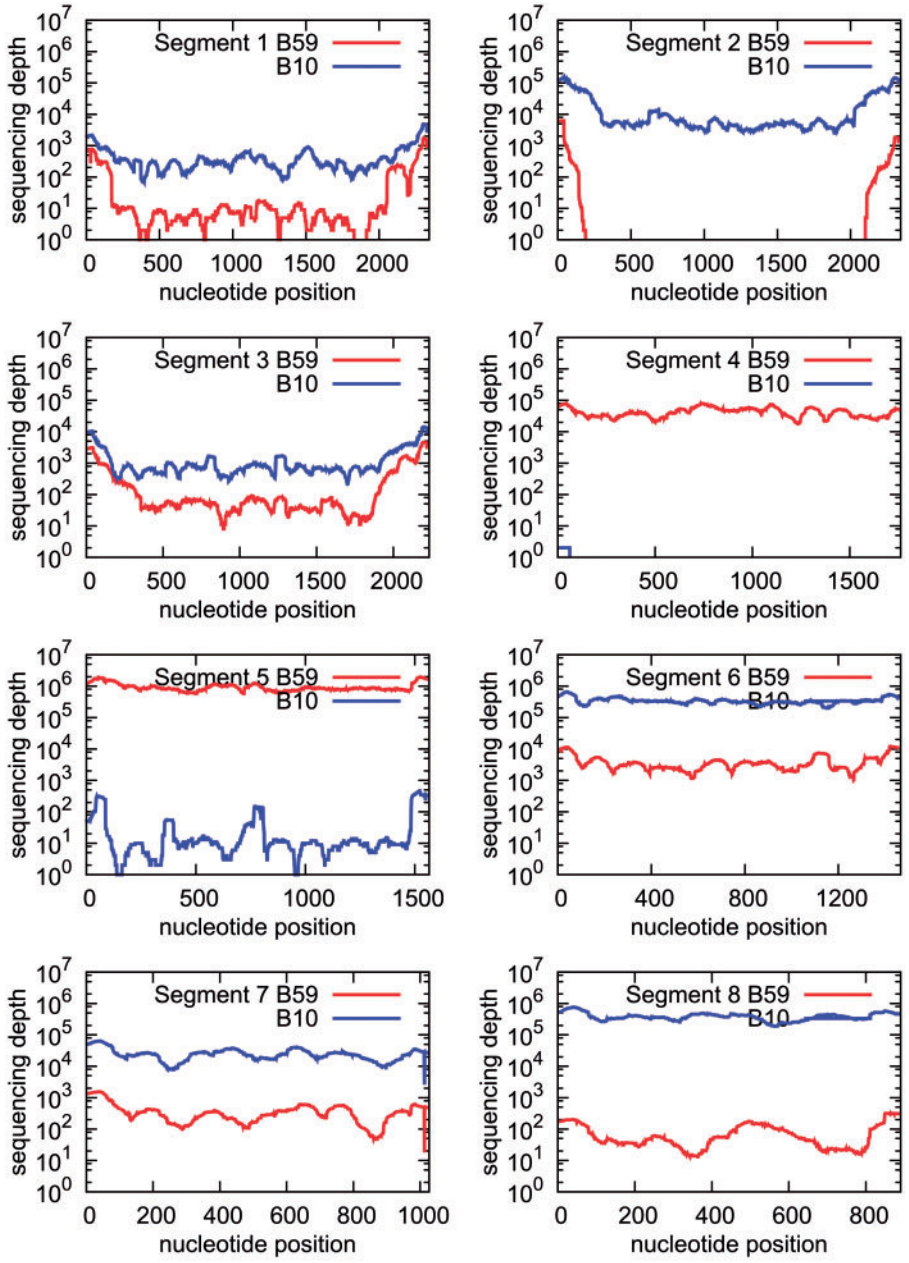

FIG. 5. Raw sequencing depth of all segments for a reassortant sample at passage 7 in MDCK cells. By passage 7, the depths for segments 4 and 5 become inverted compared with passage 1 and are predominately from the B59 (H1N1) strain (red). The other six segments continue to be largely derived from the B10 (H3N2) strain (blue), indicating the emergence of a well-defined H1N2 reassortant strain. Data from Experiment 2 are shown.

traditional single-passage coinfection experiments. Small fractions of the segments not included in the prevailing H1N2 reassortant could also be found in the evolved viral populations at P7 (supplementary table S1, Supplementary Material online)

Strength of Natural Selection of the Reassorting Segments

Our viral evolution experiment combined with bioinformatics analysis of deep sequencing data yielded an accurate determination of the change in frequencies of all viral segments over time, enabling the computation of effective population size and segment-specific selection coefficients. The selection coefficient s represents a measure of relative fitness, or contribution to the next generation, of one allele (in our case, B59 segments) compared with another (B10 segments). Positive values of s signify positive selection, $s = 0$ indicates neutrality, and $s = -1$ indicates lethal alleles. The fate of a beneficial segment is determined by both the effective population size and the selection coefficient. A beneficial segment has a greater chance to spread in large populations compared with small populations, where changes in frequency will be primarily governed by genetic drift. To quantify the strength of positive selection acting on B59 segment 4 (HA) and B59 segment 5 (NP) in the viral population, an approximate

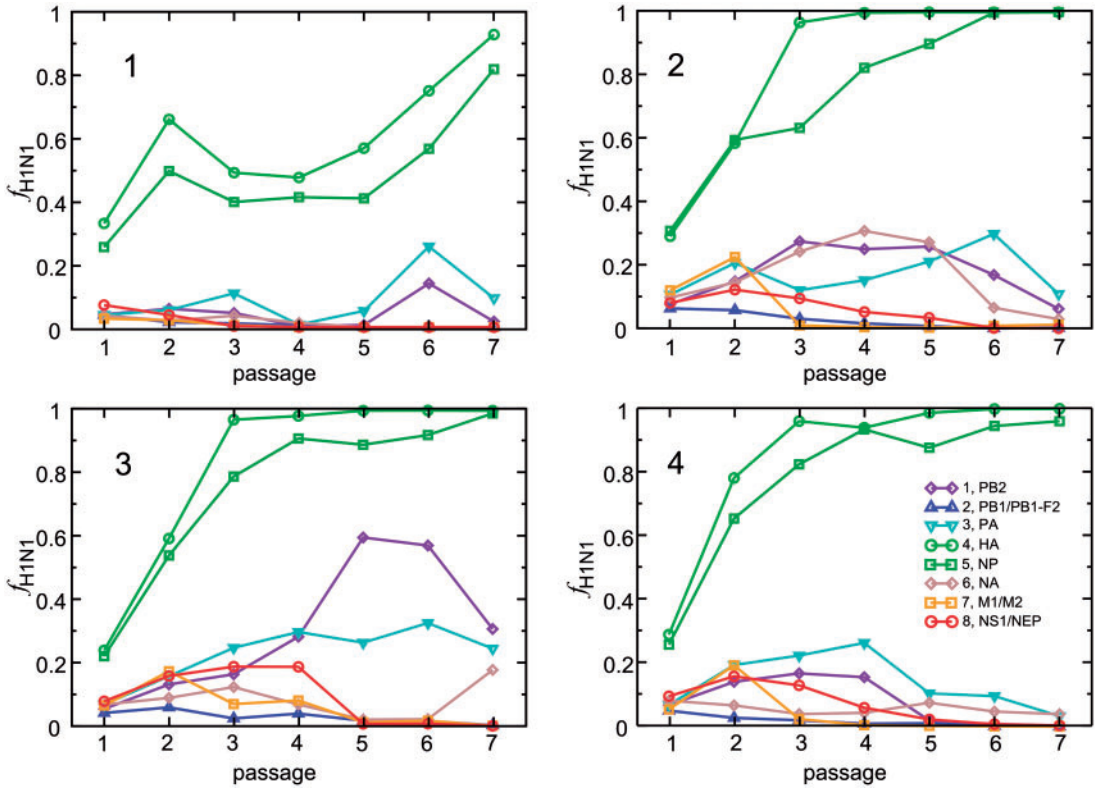

Fig. 6. Evolution of the relative abundances of the eight influenza virus genome segments over seven passages in four independent reassortment experiments. The frequency of sequences derived from B59 (H1N1) virus for each segment in the four experiments is plotted against passage number. The experiment number is indicated in the upper left hand corner of each panel. For all experiments, the frequency of B59 (H1N1) is low in early passages, so B10 (H3N2) dominates for all eight segments in the populations. Over the seven passages, the frequencies of six segments of B10 (H3N2) remain nearly constant whereas segment 4 (HA) and segment 5 (NP) of B10 are progressively replaced in the population by the segments from the B59 (H1N1) virus. PB2, polymerase basic 2; PB1/PB1-F2, polymerase basic 1; PA, polymerase acidic; M1/M2, matrix 1/matrix 2; NS1/NEP, nonstructural protein 1/nuclear export protein.

Bayesian computation (ABC) algorithm (Foll, Poh, et al. 2014; Foll, Shim, et al. 2015) was applied to infer the selection coefficients and effective population sizes from temporal data. The frequencies of B59 (H1N1) segments 4 (HA) and 5 (NP) in the population were considered as frequencies of segregating alleles. The s values and effective population sizes N_e for the B59 segment 4 (HA) and B59 segment 5 (NP) in the four experiments are presented in table 2. Positive selection was observed ($s > 0$) in all cases, with values reaching as high as 0.057. Importantly, evidence for positive selection was statistically significant (posterior probability $P > 0.95$) only for the B59 (H1N1) segments encoding HA and NP. Therefore, by analyzing quantitative temporal data on the reassortment process, we were able to measure the strong selective advantage of B59 segments 4 and 5 (HA, NP) in contributing to the viral fitness over their B10 counterparts.

Computational Identification of Prevalent Genotypes in the Reassortant Virus Populations

Deep sequencing of reassortant IAV populations precisely establishes the relative frequencies of the 16 viral genome segments in the pool, but cannot directly recover the frequencies of specific genotypes, or linkage between specific segments. Nevertheless, if segments are strongly segregated such as in the case of B59-HA, B59-NP in all four P7 pools, it becomes possible to mathematically determine the range of frequencies of all possible genotypes that provide overall segment frequencies observed by deep sequencing. For example, if the frequency of a segment from one strain in the population is 95%, no genotype having this segment can exceed 95% frequency. Conversely, the maximum population frequency of any genotype containing the same segment from the other strain is 5%. By considering all 256 possible genotypes, it becomes possible to recover their minimum and maximum frequencies compatible with eight overall segment frequencies observed in the population (see Materials and Methods). Figure 7*A* and *B* shows the ranges of the frequencies of the pure B10 and B59 strains (blue and red areas), the H1N2 reassortant genotype with B59-HA, B59-NP (green area), and the maximum population frequencies of all other 253 genotypes (black lines, most of which are very low in frequency and so invisible in the figure), in Experiments 2 and 4. In either case, the frequencies of the pure B10 and B59 genotypes were between 0 and 3.2×10^{-3}. Accordingly, pure strains would not be detectable at P7 using plaque purification (see Materials and Methods). The H1N2 reassortant genotype with B59-HA, B59-NP was the most prevalent genotype in both experiments, its frequency at P7 ranging between 78.6% and 89.1% in Experiment 2, and between 88% and 96% in Experiment 4. As follows from figures 6, 7, and

Table 2. Selection Coefficients (s) and Effective Population Size (N_e) for B59 Segments 4 (HA) and 5 (NP) in the Four Independent Reassortment Experiments[a].

	Segment 4 (HA)		Segment 5 (NP)		
Experiment	s	P ($s > 0$)	s	P ($s > 0$)	N_e
1	0.022 (−0.004, 0.051)	0.9543	0.024 (−0.004, 0.049)	0.9639	256 (183, 354)
2	0.051 (0.013, 0.127)	0.9983	0.046 (0.01, 0.144)	0.9976	219 (124, 414)
3	0.057 (0.02, 0.117)	> 0.9999	0.041 (0.014, 0.08)	0.9992	326 (160, 618)
4	0.056 (0.022, 0.113)	0.9999	0.041 (0.015, 0.079)	0.9993	382 (234, 685)

[a]The 95% confidence intervals (shown in parentheses) were determined by ABC analysis of each of the experiments. Positive selection for both HA and NP segments is statistically significant in all four experiments at significance level 0.05, after multiple hypothesis correction by Holm–Bonferroni method.

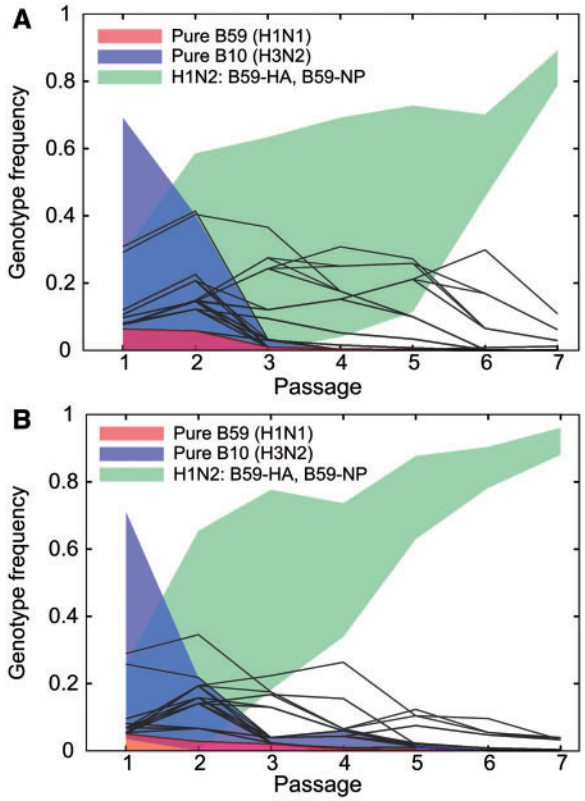

Fig. 7. Frequency ranges for the pure B59 genotype (red), pure B10 genotype (blue), and H1N2 genotype with B59-HA, B59-NP (green) for the seven passages of Experiments 2 (*A*) and 4 (*B*). The frequency ranges have been determined from population-level segment frequencies using a novel linear programming method (see Materials and Methods). The maximum frequencies of other 253 genotypes are shown as black lines. In these experiments, the H1N2 genotype with B59-HA, B59-NP is the most prevalent genotype at P7, with the average frequency above 78% and a narrow range of possible frequencies irrespective of segment linkage patterns.

table 2, positive selection for B59-HA and B59-NP results in the steady increase in frequency of the particular reassortant genotype, which at P7 constitutes the vast majority of the viral pool. Similar results were found for Experiments 1 and 3 (data not shown).

Individual Mutations Observed in the Reassortment Experiments

In addition to tracking the reassortment of gene segments, the complete coverage of the IAV genome at high depth in six independent evolutionary trajectories (four independently generated reassortants plus two pure strain controls) allowed for comparisons of the frequencies of single nucleotide mutations in all alleles across these experiments. These temporal changes in allele frequency reflect both genetic drift and selection (both positive and negative). Only a small number of mutations had significant changes in frequency during reassortment, similar to our observations of limited sequence diversity of IAV during evolution of the B59 (H1N1) strain in MDCK cells (Renzette et al. 2014).

We identified two single nucleotide substitutions that significantly increased in frequency in segments 2 and 3 of the pure strains (fig. 8*A* and *B*). Mutations B10:s2:T1960C (M646T in PB1) and B59:s3:C197G (synonymous in PA) both achieved high frequencies in the population by P7, with B59:s3:C197G nearly reaching fixation. The same two mutants remained at low levels throughout all four reassortment experiments.

Two other sites, in segments 3 and 8, B10:s3:T1379A (H452Q in PA) and B10:s8:G289A (synonymous in NS1), were polymorphic in the seed virus at frequencies of approximately 0.25 and 0.4, respectively. Their frequencies continuously decreased in the pure B10 strain, but remained significant in the reassortment experiments (fig. 8*C* and *D*). Both polymorphisms were present in the initial viral population at significant frequency, well above the noise level associated with sequence analysis artifacts such as residual cross-mapping (on the order of 10^{-2}). Additional investigation using reverse genetics will be necessary to evaluate the importance of these two polymorphisms in the reassortant, as neither of the sites has been functionally characterized yet.

In addition, several mutations rose to significant frequencies in one or more (but not all) reassortment experiments. In Experiment 4 only, the B59:s4:G1394A (D455N in HA) mutation rose to fixation (fig. 8*E*). Previously, we observed this mutation as positively selected during evolution of a pure B59 strain in MDCK cells (Foll, Poh, et al. 2014), and therefore it is unlikely to be directly related to reassortment. In segment 5, mutation B59:s5:A424G (D127G in NP) rose to approximately 45% frequency in Experiment 1 only (fig. 8*F*); it did not emerge in the pure B59 experiment or in our previous studies of pure B59 (Foll, Poh, et al. 2014).

Several mutations were occasionally observed in B10 (H3N2) NA during reassortment, but not in the pure B10 strain (fig. 8*G* and *H*). Mutation B10:s6:G1404A (D463N in NA, N2 numbering) was observed in two out of four

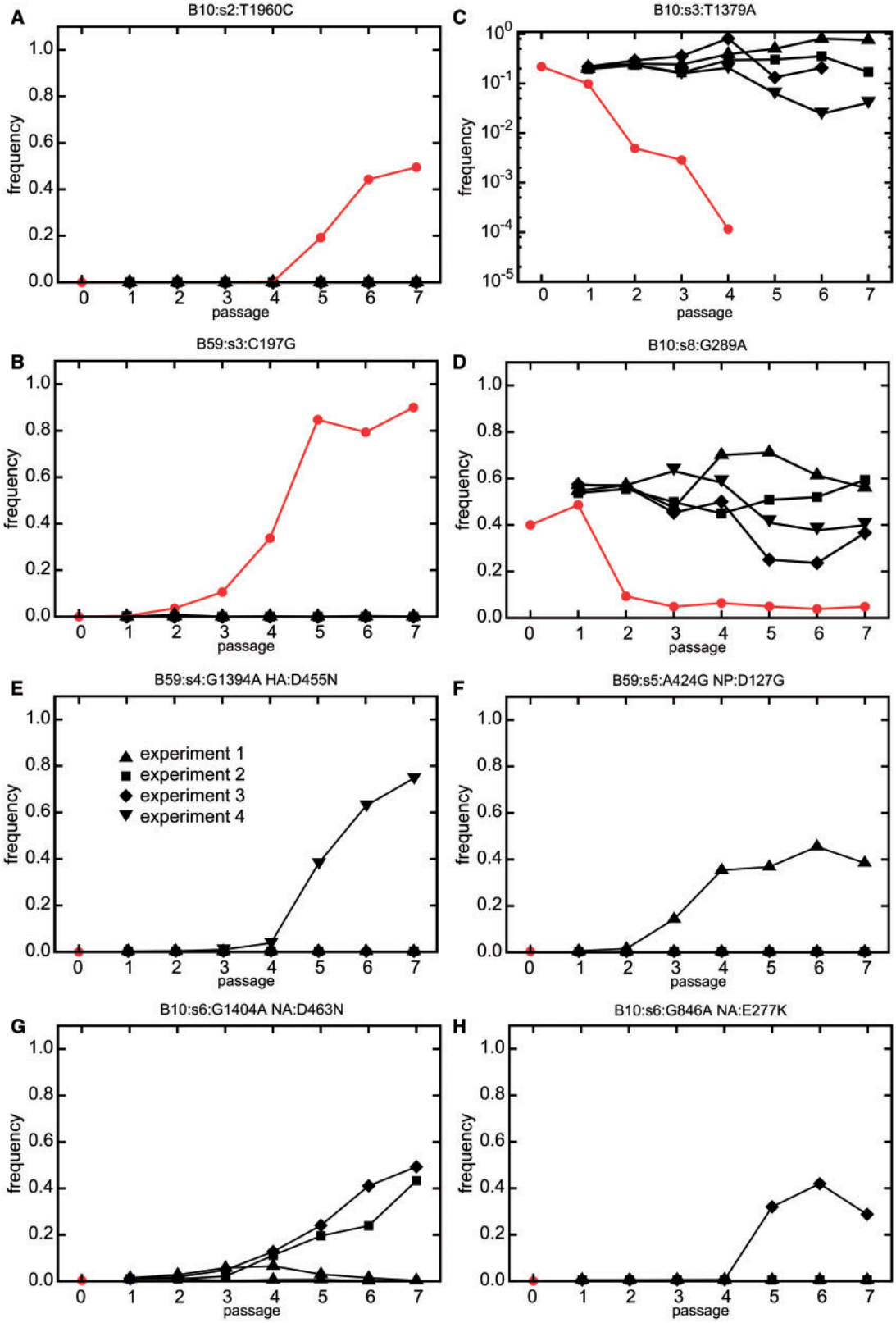

FIG. 8. Differing mutational trajectories between passaging of reassortants and pure strains. (*A*, *B*) Frequencies of B10:s2:T1960C and B59:s3:C197G mutations during passaging of the pure strain (red) and the four reassortment experiments (black) are plotted against the passage number. The mutations become fixed in the pure strain but fail to reach significant frequencies during reassortment. (*C*, *D*) Frequencies of the B10:s3:T1379A and B10:s8:G289A polymorphisms demonstrate the opposite behavior, being present at very low frequencies in the pure strains, and at high frequencies in all four reassortment experiments. Passage 0 denotes the original virus grown in egg. Passages 1–7 were performed in MDCK cells. Missing or unconnected points on the logarithmic scale plot (*C*) correspond to zero frequency of the mutations. (E–H) Frequencies of select mutations observed in one or more reassortment experiments, but not during passaging of the corresponding pure strain. The frequencies of these mutations in the pure strains were either zero or negligible.

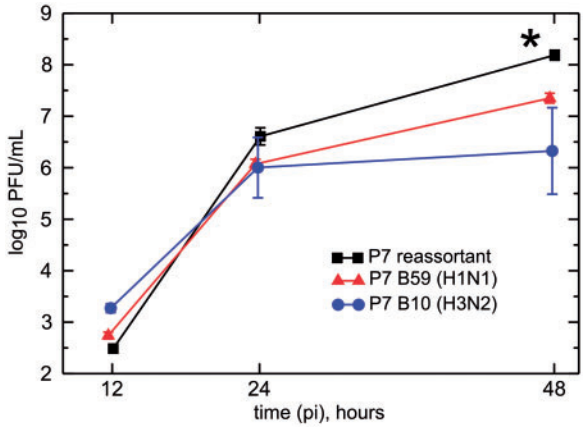

FIG. 9. Replication kinetics of pure and reassortant viruses. Multistep growth curves for the passage 7 (P7) pure B59 (H1N1) and B10 (H3N2) viruses, as well as the P7 reassortant pool from Experiment 2, in MDCK cells are shown. Confluent cells were infected with viruses at an MOI of 0.0005 PFU/cell. The virus yield was titered in MDCK cells at 12, 24, and 48 h post-infection (pi). Each data point represents viral yield (PFU/ml ± S.E.M.). Data are combined from two independent experiments using P7 reassortant generated from Experiment 2. The reassortant virus titer is an order of magnitude higher (*, $P < 0.05$) compared to the value for each parent virus.

reassortment experiments. Previously, emergence of D463N was observed during passaging of A/Victoria/3/75 (H3N2) in MDCK cells (Pérez-Cidoncha et al. 2014), and also in several influenza B strains such as B/England/45/2008. Another mutation, B10:s6:G846A (E277K in NA) rose to approximately 40% frequency in reassortment Experiment 2 only. This mutation lies close to the NA active site and has been previously found in a clinical study of zanamivir resistance (Yates et al. 2014). Together, these observations support the hypothesis that certain polymorphisms may be selected for only after reassortment, owing to changes in fitness landscape and epistatic interactions relative to the pure strains (Neverov et al. 2014).

Fitness Advantage of the Evolved Reassortants

To assess the fitness advantage of the reassortant virions relative to pure strains, we compared the replication kinetics of virus from the P7 reassortant generated from Experiment 2 with each of the P7 pure strains in MDCK cells (fig. 9). MDCK cells were inoculated with an equal amount of each virus. Although at 12 h postinfection (hpi) the yields of infectious virus in the supernatants for each strain were similar, the reassortant virus titer was at least a log higher than that for either pure strain by 48 hpi ($P < 0.05$, Student's t-test). To quantify this fitness advantage, we determined the selection coefficient from the slope of the ratio of the reassortant to B10 titers between 0 and 48 h, $s = \frac{d}{dt}\log_2 \frac{R}{B10}$, where R and B10 are the titers of P7 reassortant and P7 B10 virus, respectively. The value obtained, $s = 0.27$, indicates strong positive selection and is consistent with results obtained from population genetic based analysis (table 2). As shown in figure 7A, the viral population used for this kinetics experiment contained between 78.6% and 89.1% of the prevalent reassortant H1N2 genotype, and negligible amounts of the pure strains (below 1.5×10^{-4} in Experiment 2). Given the low MOI of the kinetics experiments (0.0005, resulting in a population bottleneck of ~125 infective particles), we conclude that the fitness advantage of the evolved P7 pool was due to reassortment. Predominance of the H1N2 reassortant genotype with B59-HA, B59-NP in the P7 pool, as well as its steady increase in frequency during passaging, strongly suggests that the increased growth of the P7 viral pool is due to the specific H1N2 genotype containing B59-HA and B59-NP segments.

Discussion

Reassortment provides unique evolutionary opportunities to IAV compared with other pathogens. Although the outcome of reassortment has been shown to be highly nonrandom in vitro (Lubeck et al. 1979), temporal evolution of reassortment could not be monitored in the previously described single-passage experiments. Our extensive deep sequencing data spanning multiple passages enabled us to observe the evolution of the viral genome during reassortment with an excellent detection limit ($<10^{-4}$), discovering the temporal changes in segment frequencies of IAV following in vitro co-infection, and tracking selection of individual mutations during reassortment.

At the final passage, the evolved viral population consisted mainly of the reassortant comprised of two segments, encoding HA and NP, from the B59 (H1N1) strain, and six segments from the B10 (H3N2) strain, resulting in a novel H1N2 strain. This pattern of reassortment is in part consistent with the original observations (Lubeck et al. 1979) of reassortment between human H1N1 and H3N2, in which different strains (A/PR8/34 [H1N1] and A/HK/8/68 [H3N2]) were used and significant linkage between segments 1, 2, and 3 forming the polymerase complex was observed. Similarly, in a study of reassortment of a seasonal H1N1 (A/Hong Kong/226654/97) and 2009 pdm (A/California/4/09), the strongest association, quantified as mutual information, was observed between segments 2 (PB1) and 3 (PA) (Greenbaum et al. 2012). In our case, segments 1, 2, and 3 were consistently provided by the B10 (H3N2) strain. The strong positive selection for the B59 segment 4 (HA) agrees with earlier observations, where HA was dominantly provided by one of the reassorting strains (Li et al. 2008; Greenbaum et al. 2012), but differs from the Lubeck et al. study (1979), where HA was taken from either strain at random after the single passage. Consistent incorporation of only one version of HA in our experiments is likely due to its selective advantage as the virus binds to MDCK cells during infection. To our knowledge, reproducible reassortment involving segment 5 (NP) has not been previously reported; our population genetics analysis suggests that B59 segment 5 is positively selected, for a reason still unknown. The possibility exists that B59 NP does not confer a selective advantage to the virus, but hitchhikes on the positively selected B59 HA. This hypothesis, however, must be reconciled with the reproducible reassortment pattern in the four experiments (fig. 6), and its statistical significance (table 2). For reproducible hitchhiking (assuming

no epistasis between HA and NP), association of B59 NP and B59 HA would likely require a physical RNA–RNA interaction between the HA and NP segments, for example, during virion packaging. Such an interaction has been detected in vitro in an avian H5N2 strain (Gavazzi et al. 2013), but not in a human H3N2 strain (Fournier et al. 2012). Further studies are required to elucidate the molecular mechanisms leading to the apparent selective advantage of B59 NP over B10 NP, epistatic interactions between B59 HA and B59 NP, and/or their physical association.

Our in vitro reassortment experiment is relevant to the evolution of influenza viruses in natural populations as well. Reassortments between H1N1 and H3N2 serotypes to generate H1N2 viruses have been previously recovered from humans (Nishikawa and Sugiyama 1983). H1N2 reassortant viruses were characterized at the genomic level and were reported to have seven segments originating from H3N2 and only one, encoding HA, from H1N1 (Gregory et al. 2002). Recently, emergence of an H1N2 reassortant virus has been demonstrated in a ferret model after coinfection with recombinant viruses of seasonal H3N2 and pandemic H1N1 strains (Angel et al. 2013), paralleling what is observed in swine in nature (Pascua et al. 2008). In this case, the emerging H1N2 viruses carried six segments from the H1N1 strain, and only two segments (encoding NA and PB1) from the H3N2 strain. Hence, the original viral sequences and the relative fitness of the possible reassortants likely determine the resulting specific reassortant H1N2 strain.

The segregating B59 segments 4 (HA) and 5 (NP) may be incorporated into the B10 strain background either together or individually. Although the population-averaged segment frequencies cannot precisely delineate between these two mechanisms, the marked divergence of the evolutionary trajectories of these two segments, particularly at passages 3 and 4 of Experiment 2 (fig. 6), suggests individual incorporation into the predominantly B10 virions at the early stages of reassortment. Direct evidence for individual versus simultaneous incorporation can be obtained either from plaque experiments, or higher throughput single virion studies, such as single molecule fluorescence of IAV (Chou et al. 2012) or single virion sequencing (Allen et al. 2011), which has not been yet demonstrated for IAV. Nevertheless, the strong, reproducible selection for B59-HA and B59-NP segments in the viral populations severely limited the range of possible reassortant genotypes at P7. The evolved H1N2 genotype was the most frequent one at P7, and consistently increased in frequency over time, suggestive of positive selection for the specific genotype, rather than for individual segments.

The reproducibility of the results in four independent replicates demonstrates that the evolved reassortant is unlikely to be the result of random fluctuations of segment frequency (i.e., genetic drift) after coinfection, but rather that evolved segment combinations are driven by positive selection among the 16 varieties of IAV segments originally introduced to the host after coinfection. The patterns of change in segment frequencies during serial passaging and the concomitant increase of viral fitness together reflect selective pressures acting on the population of two coevolving strains of influenza virus. Accordingly, the temporal change in the frequencies of the segregating B59 segments 4 (HA) and segment 5 (NP) yielded significantly positive selection coefficients. The selective advantage of reassortment inferred through population genetics analysis had a clear phenotypic manifestation, as the reassortant viral population achieved titers an order of magnitude higher than either of the original strains.

The high sensitivity of our assay to very minor fractions of IAV segments suggests revisiting the operational definition of a monotypic viral population in the context of a reassortment experiment. For detection of rare segments, deep sequencing dramatically outperforms the common approach of genotyping of multiple (between 40 and 242) plaques (Lubeck et al. 1979; Greenbaum et al. 2012; Marshall et al. 2013). As described in Materials and Methods, not observing a specific segment in 100 genotyped plaques only means that its frequency in the population is below 4.6% with $P = 0.99$. In contrast, the deep sequencing approach allows us to measure segment frequencies in the population on the order of 10^{-4} (supplementary fig. S2, Supplementary Material online), theoretically limited by sequence homology and practically also by sequencing errors. Detection of a rare segment at 1% frequency in the viral pool with $P > 0.99$ would require extensive genotyping of 460 individual plaques, using 16 different primers for each. Our detection of B59 segments 1–3, 6–8, and B10 segments 4–5 at very low frequencies in the passage 7 viral pool reveals the fine structure of IAV reassortant populations not accessible by earlier methods. At the same time, although our method provides a precise statistical description of large viral populations, it does not allow us to observe linkage of the segments in individual virions, readily extracted from plaque analysis.

Our observation of different trajectories of mutation frequencies during reassortment compared with pure strains may be indicative of epistatic interactions between IAV segments, potentially allowing previously deleterious mutants to become beneficial. This mechanism is complementary to the elevated mutation rates characteristic of RNA viruses. A similar observation was recently made during infection of guinea pigs by mouse-adapted A/PR8/34 influenza virus (Ince et al. 2013). Sequencing of multiple viral clones demonstrated rapid reassortment of the viruses carrying beneficial Single nucleotide polymorphisms (SNPs) on NP and M1 segments. On a larger scale, a recent bioinformatics study of human H3N2 strains found an acceleration of amino acid replacements following reassortment events, unraveling coevolution between segmental composition of IAV strains and mutations in individual segment sequences (Neverov et al. 2014). Our observations of reassortment-specific mutations can be seen as direct experimental support of those findings.

Hence, the segmented nature of the influenza viral genome facilitates the reassortment of segments and the selection of reassortment-specific mutations potentially enabling rapid adaptation. The possibility of epistatic interactions leading to reassortant-specific mutants during influenza virus evolution supports the seminal works of Kilbourne (1981) who proposed that influenza virus reassortment be considered an asymmetric form of sexual

reproduction. Indeed, this system represents a potential avenue for quantifying the fundamental importance of epistasis in driving the evolution of sex and recombination (Kondrashov 1988; Barton and Charlesworth 1998). Further experimental studies of influenza virus reassortment, using a combination of reverse genetics with deep sequencing-based quantification of the reassorting genomes, will shed light on the molecular mechanisms in place for influenza virus genome assembly and reassortment.

Materials and Methods

Cells, Virus Stocks, and Chemicals

MDCK cells were obtained from American Type Culture Collection (Manassas, VA) and propagated in Eagle's minimal essential medium (MEM) with 10% fetal bovine serum (Hyclone, Thermo Fisher Scientific Inc., Waltham, MA) and 2 mM penicillin/streptomycin. Influenza virus A/Brisbane/59/2007 (H1N1) and A/Brisbane/10/2007 (H3N2), originally grown in the chicken egg allantoic fluid, were obtained through the NIH Biodefense and Emerging Infections Research Resources Repository, NIAID, NIH, Washington, DC (Brisbane/59/2007, NR-12282, lot 58550257; Brisbane/10/2007, NR-12283, lot 58550258).

Viral Titer Determination by Plaque Assay

Viruses were quantified on MDCK cells to determine infectious titer (plaque forming units [PFU] per ml, or PFU/ml). Six 10-fold serial dilutions were performed on the viral samples followed by 1 h of binding at 37°C on confluent MDCK cells in 12-well plates. After washing off unbound virus with phosphate buffered saline, the cells were overlaid with agar (0.5%) in Dulbecco's Modified Eagle Medium-F12 supplemented with penicillin/streptomycin, L-glutamine, bovine serum albumin, HEPES, sodium bicarbonate, and 20 μg/ml acetylated trypsin (Sigma-Aldrich, St. Louis, MO). After the agar solidified, the plates were incubated for approximately 48 h at 37°C. Cells were fixed and stained with anti-NP antibody ab20343 (Abcam, Cambridge, MA). Plaques were visualized with antimouse horseradish peroxidase-conjugated secondary antibody (BD Biosciences, San Jose, CA) and developed with peroxidase substrate kit (Vector Laboratories, Burlingame, CA). Viral plaques in the MDCK monolayer were quantified by visual inspection.

Viral Culture

Viruses were serially passaged in MDCK cells in influenza virus growth media, comprised of DMEM supplemented with penicillin/streptomycin, 0.2% bovine serum albumin fraction V solution (Invitrogen, Carlsbad, CA), and 1 μg/ml tosylsulfonyl phenylalanyl chloromethyl ketone-treated trypsin (Sigma-Aldrich). For the initial passage, each virus was added at an MOI of 0.0005. Virus was continually passaged on MDCK cells to minimize any freeze–thaw cycles. Multiple serial 10-fold dilutions of virus supernatants were inoculated on MDCK cells so that greater than 50% cytopathic effect would be achieved in samples within 72 h. Cell-free virus was collected and used for deep sequencing, quantification by plaque assay (to determine the MOI), and for generation of the next passage. The MOI ranged between 1.5×10^{-5} and 0.63 for each passage. Of note, one reassortment passaging experiment (Experiment 4) was performed with constant MOIs of 10^{-3} (quantified by plaque assay).

Replication Kinetics

To determine multistep growth curves, MDCK cells were infected with either pure or reassortant viruses at an MOI of 0.0005. After incubation, the cells were washed and overlaid with influenza virus growth medium. Supernatants were collected at 12, 24, and 48 hpi and stored at −80°C for titration. *P* values were determined using Student's *t*-test (GraphPad Prism 6).

Deep Sequencing

A high-throughput sample processing workflow was carried out in 96-well format, including RNA purification, reverse transcription, whole genome PCR, followed by DNA barcoding and library preparation, as described previously (Renzette et al. 2014). Sequencing was performed on Illumina HiSeq2000 platform using 100-bp reads. All libraries contained an error control RNA product that was produced from a plasmid clone of full-length influenza A/Brisbane/59/2007 (H1N1) NA segment that was processed in parallel with the experimental samples.

Bioinformatics Analysis

To assess the feasibility of discrimination between B10 and B59 strains by deep sequencing, we simulated the deep sequencing experiment by randomly cutting the strains' genomes into 80-bp long reads and mapping them to the concatenated reference genome. Intrinsic sequence diversity of the viral population was modeled by considering each strain as a set of 200 quasispecies with 1% sequence divergence from the master genome, and the allele frequency spectrum taken from Renzette et al. (2014). Random sequencing errors, including indels, were introduced in each simulated read with the probability of 10^{-2} per nucleotide. Mapping of the reads using BLAST and Bowtie2 showed that the sequence divergence between B10 and B59 strains was sufficient to correctly map a mix of reads to the proper strains with the probability exceeding 99% for representative sequencing error rates and several models of quasispecies structure of the viral population.

When processing actual deep sequencing data, short reads from the Illumina platform were filtered for quality scores greater than 20 throughout the read and aligned to the strains' reference genomes (accessions CY030232, CY031391, CY058484–CY058486, CY058488–CY058489, CY058491 for A/Brisbane/59/2007 and CY035022, CY031812, EU199420, CY035025–CY035029 for A/Brisbane/10/2007) using BLAST (Altschul et al. 1990), with *E* value cutoff of 10^{-5}, and Bowtie2 (Langmead and Salzberg 2012) with the very-sensitive-local option. The differences in nucleotide frequencies detected by either aligner were insignificant. Analysis of the alignment revealed that these reference

genomes lack the 3′-RNA untranslated region (UTR) sequences. We reassembled the missing UTR sequence and formed the complete reference genomes for the pure A/Brisbane/59/2007 and A/Brisbane/10/2007 strains used in our experiments. We confirmed the newly derived reference genome sequences by Sanger sequencing of the viruses. Data analysis was performed using the new, complete reference genomes. Reads producing significant mapping to the same segment in both strains were assigned to the strain producing the better alignment score. Only alignments longer than 40 nt were retained. Over 95% of the filtered reads produced significant alignments. From the alignments, the sequencing depth and frequency of all four nucleotides at every position were calculated. The median sequencing depth was 35,000. To quantitate the relative abundance of a segment in the viral populations, the median sequencing depth of each segment was calculated, and the ratio of the sequencing depths was used as a measure of relative abundance of the segments. To exclude ambiguous mapping due to local sequence homology and position-specific amplification biases during sample processing, the first and last 250 nt of each segment were excluded from depth ratio calculations. The ratio of the corresponding segments in the viral population was calculated as the ratio of the median sequencing depths (excluding the first and last 250 nt) after calibration correction (see below). Nucleotide mutations are reported with respect to the DNA sequence of the reference genome, with first nucleotide of each segment numbered zero. Protein mutations are reported in the numbering system of the corresponding open reading frame, with first amino acid numbered one. Assembled reference genomes and all sequencing data sets are available at http://bib.umassmed.edu/influenza, last accessed February 23, 2015.

Sequencing Error Analysis

Sequence errors can be introduced either at sample processing stage (e.g., PCR) or during amplification or sequencing-by-synthesis. Thus, a plasmid encoding the full-length NA segment from A/Brisbane/59/2007 was created. RNA was generated from these plasmids using T7 RNA polymerase and these products were assumed to accurately represent the sequence of the cDNA. The RNA pools were then introduced into the sample processing workflow and processed and sequenced in an identical manner as viral samples. All sequencing runs included this error control construct to account for any run-to-run variation in error rates. From these data, the combined level of error introduced from sample processing and sequencing was estimated. Mapping the reads from the sequencing control to the reference sequence established the total error rate of amplification and sequencing of 3.2×10^{-3} per nucleotide (95% upper confidence limit). SNPs exceeding this frequency were considered significant.

Calibration

To transform the ratio of sequencing depths to the ratio of segment concentrations in the viral mix and account for the possible biases during sample processing, sequencing, and analysis, we extracted RNA from strains A/Brisbane/59/2007 and A/Brisbane/10/2007, quantified the RNA by qRT-PCR, and generated mixtures of RNA from each strain in the ratios of 1,000:1, 100:1, 10:1, 1:1, 1:10, 1:100, and 1:1,000. These mixtures were processed exactly as the viral samples and deep sequenced. The ratio of sequencing depths was plotted as a function of known concentration ratio. A near-perfect linear dependence is observed for all segments ($R^2 > 0.98$, fig. 3). This calibration was used to infer the ratio of segment abundances from the ratio of sequencing depths.

Determination of Selection Coefficients

Selection coefficients (s) and effective population sizes (N_e) were determined from time-course allele (segment) frequency data using an ABC method (Foll, Poh, et al. 2014; Foll, Shim, et al. 2015). The method provides posterior distributions for the selection coefficients and effective population sizes. We considered segments as significantly positively selected when the posterior probability $P(s > 0)$ was higher than 0.95. Posterior distributions of the selection coefficients are presented in supplementary figure S3, Supplementary Material online.

Determination of the Possible Frequencies of Reassortant Genotypes from Population Level Segment Frequencies

Assume that 256 possible genotypes of strains A and B are in the viral population, with frequencies x_i, $1 \le i < 256$. The population frequencies f_j of each of the eight segments of type A are then given by $f_j = \sum_{k=1}^{256} x_k a_{kj}$, where $a_{kj} = 1$ if the genotype k contains segment j from strain A, and 0 otherwise. These eight linear equations, together with the condition $\sum_{k=1}^{256} x_k = 1,\ 0 \le x_k \le 1$, impose mathematical limits on the possible minimum and maximum values of each of the genotype frequencies x_i compatible with the observed values of f_j. Similar optimization problems often occur in economics in the context of determining the maximum profit or lowest production costs subject to linear constraints, for example, cost and availability of materials and other resources (Dantzig 1997). To determine these limits, we used a linear programming approach, with the objective function seeking the minimum or maximum value of each of the variables x_i individually. The linear programming problem was solved using the simplex method (Kantorovich 1940) and the frequency ranges of all possible genotype are presented in figure 7. Importantly, the method makes no assumptions about segment linkage patterns, and establishes rigorous bounds on genotype frequencies compatible with population frequencies of each segment. Detailed description of the method will be presented elsewhere (Venev SV, Zeldovich KB, in preparation).

Estimation of Reassortment Detection Limits in Plaque Assays

Consider an assay where N plaques are genotyped, and suppose that the viral population is nearly monotypic but

contains a small fraction λ of a particular reassortant. In N plaques, this reassortant would be observed k times, and k follows the Poisson distribution with parameter $N\lambda$. The average value of k is $N\lambda$, and the probability P that the reassortant would be observed one or more times is $P(k > 0) = 1 - P(0) = 1 - \exp(-N\lambda)$. Therefore, reliable detection of the reassortant with probability $P > 0.99$ (significance level 0.01) is only possible if $1 - \exp(-N\lambda) > P$, or $\lambda > -\ln(1-P)/N \approx 4.60/N$. For a typical plaque experiment, $N = 100$, and $\lambda > 0.046 = 4.6\%$. Therefore, in an experiment with 100 plaques one can reliably detect only reassortants present at 4.6% frequency or above. Conversely, if a set of 100 plaques has been found to be monotypic (the reassortant was not observed), it only means that the frequency of the reassortant was below 4.6%, at the confidence level of 0.99. Those detection limits are vastly inferior to the deep sequencing approach or qRT-PCR of bulk viral populations. Incidentally, no less than $4.60 \times 256 = 1{,}178$ plaques must be genotyped to capture all 256 possible reassortants with $P > 0.99$, even under the unrealistic assumption that reassortment is purely random and no rare segment constellations are present. Setting a lower confidence level, for example, $P = 0.95$, has limited effect, improving the detection limit to 3% and the requisite number of plaques to 768.

Supplementary Material

Supplementary figures S1–S3 and table S1 are available at *Molecular Biology and Evolution* online (http://www.mbe.oxfordjournals.org/).

Acknowledgments

This work was supported by the Defense Advanced Research Projects Agency (DARPA) Prophecy Program, Defense Sciences Office (DSO), Contract No. HR0011-11-C-0095, and D13AP00041. The authors acknowledge the contributions of all the members of the ALiVE (Algorithms to Limit Viral Epidemics) working group. The authors thank Nese Yilmaz, Ashwini Sunkavalli, and Melanie Trombly for their helpful edits.

References

Allen LZ, Ishoey T, Novotny MA, McLean JS, Lasken RS, Williamson SJ. 2011. Single virus genomics: a new tool for virus discovery. *PLoS One* 6:e17722.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ. 1990. Basic local alignment search tool. *J Mol Biol.* 215:403–410.

Angel M, Kimble JB, Pena L, Wan H, Perez DR. 2013. In vivo selection of H1N2 influenza virus reassortants in the ferret model. *J Virol.* 87: 3277–3283.

Barton NH, Charlesworth B. 1998. Why sex and recombination? *Science* 281:1986–1990.

Chen L-M, Davis CT, Zhou H, Cox NJ, Donis RO. 2008. Genetic compatibility and virulence of reassortants derived from contemporary Avian H5N1 and human H3N2 influenza A viruses. *PLoS Pathog.* 4: e1000072.

Chou YY, Vafabakhsh R, Doganay S, Gao Q, Ha T, Palese P. 2012. One influenza virus particle packages eight unique viral RNAs as shown by FISH analysis. *Proc Natl Acad Sci U S A.* 109:9101–9106.

Dantzig GB, Thapa MN. 1997. Linear programming 1: introduction. New York: Springe-Verlag.

Essere B, Yver M, Gavazzi C, Terrier O, Isel C, Fournier E, Giroux F, Textoris J, Julien T, Socratous C, et al. 2013. Critical role of segment-specific packaging signals in genetic reassortment of influenza A viruses. *Proc Natl Acad Sci U S A.* 110:E3840–E3848.

Foll M, Poh YP, Renzette N, Ferrer-Admetlla A, Bank C, Shim H, Malaspinas AS, Ewing G, Liu P, Wegmann D, et al. 2014. Influenza virus drug resistance: a time-sampled population genetics perspective. *PLoS Genet.* 10:e1004185.

Foll M, Shim H, Jensen JD. 2015. WFABC: a Wright-Fisher ABC-based approach for inferring effective population sizes and selection coefficients from time-sampled data. *Mol Ecol Resour.* 15(1):87–98.

Fournier E, Moules V, Essere B, Paillart J-CC, Sirbat J-DD, Isel C, Cavalier A, Rolland J-PP, Thomas D, Lina B, et al. 2012. A supramolecular assembly formed by influenza A virus genomic RNA segments. *Nucleic Acids Res.* 40:2197–2209.

Fujii K, Fujii Y, Noda T, Muramoto Y, Watanabe T, Takada A, Goto H, Horimoto T, Kawaoka Y. 2005. Importance of both the coding and the segment-specific noncoding regions of the influenza A virus NS segment for its efficient incorporation into virions. *J Virol.* 79: 3766–3774.

Fujii Y, Goto H, Watanabe T, Yoshida T, Kawaoka Y. 2003. Selective incorporation of influenza virus RNA segments into virions. *Proc Natl Acad Sci U S A.* 100:2002–2007.

Gavazzi C, Isel C, Fournier E, Moules V, Cavalier A, Thomas D, Lina B, Marquet R. 2013. An in vitro network of intermolecular interactions between viral RNA segments of an avian H5N2 influenza A virus: comparison with a human H3N2 virus. *Nucleic Acids Res.* 41: 1241–1254.

Ghedin E, Sengamalay NA, Shumway M, Zaborsky J, Feldblyum T, Subbu V, Spiro DJ, Sitz J, Koo H, Bolotov P, et al. 2005. Large-scale sequencing of human influenza reveals the dynamic nature of viral genome evolution. *Nature* 437:1162–1166.

Gog JR, Afonso Edos S, Dalton RM, Leclercq I, Tiley L, Elton D, von Kirchbach JC, Naffakh N, Escriou N, Digard P. 2007. Codon conservation in the influenza A virus genome defines RNA packaging signals. *Nucleic Acids Res.* 35:1897–1907.

Greenbaum BD, Li OTW, Poon LLM, Levine AJ, Rabadan R. 2012. Viral reassortment as an information exchange between viral segments. *Proc Natl Acad Sci U S A.* 109(9):3341–3346.

Gregory V, Bennett M, Orkhan MH, Al Hajjar S, Varsano N, Mendelson E, Zambon M, Ellis J, Hay A, Lin YP. 2002. Emergence of influenza A H1N2 reassortant viruses in the human population during 2001. *Virology* 300:1–7.

Ince WL, Gueye-Mbaye A, Bennink JR, Yewdell JW. 2013. Reassortment complements spontaneous mutation in influenza A virus NP and M1 genes to accelerate adaptation to a new host. *J Virol.* 87: 4330–4338.

Jagger BW, Wise HM, Kash JC, Walters K-A, Wills NM, Xiao Y-L, Dunfee RL, Schwartzman LM, Ozinsky A, Bell GL, et al. 2012. An overlapping protein-coding region in influenza A virus segment 3 modulates the host response. *Science* 337:199–204.

Kantorovich LV. 1940. A new method of solving some classes of extremal problems. *Dokl Akad Sci USSR.* 28:211–214.

Kawaoka Y, Krauss S, Webster RG. 1989. Avian-to-human transmission of the PB1 gene of influenza A viruses in the 1957 and 1968 pandemics. *J Virol.* 63:4603–4608.

Kilbourne ED. 1981. Segmented genome viruses and the evolutionary potential of asymmetrical sex. *Perspect Biol Med.* 25:66–77.

Kondrashov AS. 1988. Deleterious mutations and the evolution of sexual reproduction. *Nature* 336:435–440.

Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods.* 9:357–359.

Li C, Hatta M, Nidom CA, Muramoto Y, Watanabe S, Neumann G, Kawaoka Y. 2010. Reassortment between avian H5N1 and human H3N2 influenza viruses creates hybrid viruses with substantial virulence. *Proc Natl Acad Sci U S A.* 107(10):4687–4692.

Li C, Hatta M, Watanabe S, Neumann G, Kawaoka Y. 2008. Compatibility among polymerase subunit proteins is a restricting

factor in reassortment between equine H7N7 and human H3N2 influenza viruses. *J Virol.* 82:11880–11888.

Lindstrom SE, Cox NJ, Klimov A. 2004. Genetic analysis of human H2N2 and early H3N2 influenza viruses, 1957–1972: evidence for genetic divergence and multiple reassortment events. *Virology* 328: 101–119.

Lubeck MD, Palese P, Schulman JL. 1979. Nonrandom association of parental genes in influenza A virus recombinants. *Virology* 95: 269–274.

Luytjes W, Krystal M, Enami M, Parvin JD, Palese P. 1989. Amplification, expression, and packaging of foreign gene by influenza virus. *Cell* 59: 1107–1113.

Marshall N, Priyamvada L, Ende Z, Steel J, Lowen AC. 2013. Influenza virus reassortment occurs with high frequency in the absence of segment mismatch. *PLoS Pathog.* 9:e1003421.

McGeoch D, Fellner P, Newton C. 1976. Influenza virus genome consists of eight distinct RNA species. *Proc Natl Acad Sci U S A.* 73: 3045–3049.

Neverov AD, Lezhnina KV, Kondrashov AS, Bazykin GA. 2014. Intrasubtype reassortments cause adaptive amino acid replacements in H3N2 influenza genes. *PLoS Genet.* 10:e1004037.

Nishikawa F, Sugiyama T. 1983. Direct isolation of H1N2 recombinant virus from a throat swab of a patient simultaneously infected with H1N1 and H3N2 influenza A viruses. *J Clin Microbiol.* 18:425–427.

Pascua PN, Song MS, Lee JH, Choi HW, Han JH, Kim JH, Yoo GJ, Kim CJ, Choi YK. 2008. Seroprevalence and genetic evolutions of swine influenza viruses under vaccination pressure in Korean swine herds. *Virus Res.* 138:43–49.

Pérez-Cidoncha M, Killip MJ, Oliveros JC, Asensio VJ, Fernández Y, Bengoechea JA, Randall RE, Ortín J. 2014. An unbiased genetic screen reveals the polygenic nature of the influenza virus anti-interferon response. *J Virol.*

Renzette N, Caffrey DR, Zeldovich KB, Liu P, Gallagher GR, Aiello D, Porter AJ, Kurt-Jones EA, Bolon DN, Poh YP, et al. 2014. Evolution of the influenza A virus genome during development of oseltamivir resistance in vitro. *J Virol.* 88:272–281.

Simon-Loriere E, Holmes EC. 2011. Why do RNA viruses recombine? *Nat Rev Microbiol.* 9:617–626.

Wagner R, Matrosovich M, Klenk HD. 2002. Functional balance between haemagglutinin and neuraminidase in influenza virus infections. *Rev Med Virol.* 12:159–166.

Wright S. 1932. The roles of mutation, inbreeding, crossbreeding and selection in evolution. In Jones DF, editor. Proceedings of the Sixth International Congress on Genetics; Ithaca, NY: Brooklyn Botanical Gardens. p. 356–366.

Yates PJ, Mehta N, Hasan S, Choy M, Peppercorn A. 2014. Identification of resistance mutations as minority species in clinical specimens from hospitalised adults with influenza and treated with intravenous Zanamivir. Influenza and other respiratory virus infections: Advances in Clinical Management, Third isirv-Antiviral Group Conference; June 4–6, 2014; Tokyo, Japan. Presentation P19.