

Evolution of the Influenza A Virus Genome during Development of Oseltamivir Resistance *In Vitro*

Nicholas Renzette,^a Daniel R. Caffrey,^b Konstantin B. Zeldovich,^c Ping Liu,^b Glen R. Gallagher,^b Daniel Aiello,^b Alyssa J. Porter,^b Evelyn A. Kurt-Jones,^b Daniel N. Bolon,^d Yu-Ping Poh,^{c,f} Jeffrey D. Jensen,^{e,f} Celia A. Schiffer,^d Timothy F. Kowalik,^a Robert W. Finberg,^b Jennifer P. Wang^b

Department of Microbiology and Physiological Systems,^a Department of Medicine,^b Program in Bioinformatics and Integrative Biology,^c and Department of Biochemistry and Molecular Pharmacology,^d University of Massachusetts Medical School, Worcester, Massachusetts, USA; Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne, Switzerland^e; Swiss Institute of Bioinformatics (SIB), Lausanne, Switzerland^f

Influenza A virus (IAV) is a major cause of morbidity and mortality throughout the world. Current antiviral therapies include oseltamivir, a neuraminidase inhibitor that prevents the release of nascent viral particles from infected cells. However, the IAV genome can evolve rapidly, and oseltamivir resistance mutations have been detected in numerous clinical samples. Using an *in vitro* evolution platform and whole-genome population sequencing, we investigated the population genomics of IAV during the development of oseltamivir resistance. Strain A/Brisbane/59/2007 (H1N1) was grown in Madin-Darby canine kidney cells with or without escalating concentrations of oseltamivir over serial passages. Following drug treatment, the H274Y resistance mutation fixed reproducibly within the population. The presence of the H274Y mutation in the viral population, at either a low or a high frequency, led to measurable changes in the neuraminidase inhibition assay. Surprisingly, fixation of the resistance mutation was not accompanied by alterations of viral population diversity or differentiation, and oseltamivir did not alter the selective environment. While the neighboring K248E mutation was also a target of positive selection prior to H274Y fixation, H274Y was the primary beneficial mutation in the population. In addition, once evolved, the H274Y mutation persisted after the withdrawal of the drug, even when not fixed in viral populations. We conclude that only selection of H274Y is required for oseltamivir resistance and that H274Y is not deleterious in the absence of the drug. These collective results could offer an explanation for the recent reproducible rise in oseltamivir resistance in seasonal H1N1 IAV strains in humans.

Influenza A virus (IAV) is a pathogen of primary importance for public health. The unique evolutionary mechanisms of IAV, including its high mutation rate, segment reassortment, and shifts between multiple host species, pose significant challenges for controlling disease and developing vaccination and treatment strategies. Therefore, a detailed understanding of influenza virus genome sequence evolution is imperative. The influenza virion consists of eight negative-strand RNA segments, which form protein-RNA complexes enveloped in a lipid membrane (1). The eight segments encode at least 10 proteins known to be essential for infectivity and replication. Within a given influenza virus strain, sequence evolution proceeds by mutation, selection, and genetic drift, all of which are affected by the environment, the host, and drug treatment. High mutation rates, together with the rapid development of influenza epidemics, make tracing the evolutionary history of the virus and discovering the principles governing IAV's evolution complex.

Neuraminidase (NA) inhibitors, especially the orally bioavailable drug oseltamivir, have been widely used for the treatment of IAV. These drugs bind to the influenza virus neuraminidase active site as competitive inhibitors. One great challenge in the treatment of influenza with NA inhibitors is the development of drug-resistant mutants. Several resistance mutations occur in response to oseltamivir treatment, including His274Tyr (H274Y), which is caused by a single nucleotide change and is specific to viruses with neuraminidase N1 (2, 3). (Nucleotide positions are given according to the N2 numbering system.) H274Y-associated oseltamivir resistance emerged rapidly and spread globally in H1N1 strains during the 2007–2008 and 2008–2009 influenza seasons (4, 5). The H274Y mutation is close to the active site of neuraminidase, and

crystal structures reveal the molecular basis of resistance (6). In addition, N2 viruses have specific neuraminidase mutations, including Arg292Lys (R292K) and Glu119Val (E119V), while Asn294Ser (N294S) mutants have been found among both N1 and N2 viruses (7, 8).

While IAV can develop drug resistance with a single mutation, how the development of resistance affects IAV population diversity and structure remains unclear. The recurrent selection of beneficial mutations is expected to eliminate variation in the region surrounding the selected loci (9) as well as to cause increased divergence between isolated populations (10). Thus, the development of drug resistance would be expected to alter the diversity of IAV populations and to increase divergence between drug-sensitive and drug-resistant populations. For segmented viruses, it has not been established whether such effects would be limited to segment 6 (neuraminidase), the target of selection during the development of oseltamivir resistance, or would be observed across the entire genome. Similarly, the presence of drug may alter the selective landscape by affecting the fitness of mutations in the drug or no-drug environment due to mechanisms such as altered structural constraint or epistatic interactions. Again, the genomewide

Received 8 July 2013 Accepted 15 October 2013

Published ahead of print 23 October 2013

Address correspondence to Jennifer P. Wang, jennifer.wang@umassmed.edu.

R.W.F. and J.P.W. are co-senior authors.

Copyright © 2014, American Society for Microbiology. All Rights Reserved.

doi:10.1128/JVI.01067-13

effects of this alteration are unknown. Better knowledge of these phenomena will yield insights into the evolutionary obstacles and costs associated with the development of drug resistance and may aid in future drug design efforts.

The availability of high-throughput sequencing technology enables observation beyond the consensus sequence of the virus and establishes the genetic structure and low-frequency single nucleotide polymorphisms (SNPs) of the viral population. Populations of RNA viruses such as IAV are likely not genetically uniform but instead form quasispecies, resulting in a degree of sequence diversity (11). The magnitude of sequence variance within a single IAV strain and the temporal patterns of evolution are key aspects of IAV biology and are paramount for predicting viral evolution, response to treatment, and drug resistance. Until now, the evolutionary trajectories leading to drug resistance in IAV have not been comprehensively elucidated. In this study, we combine serial passaging with high-throughput sequencing to explicitly trace *in vitro* evolutionary trajectories of the genome of influenza virus A/Brisbane/59/2007 (H1N1) populations in Madin-Darby canine kidney (MDCK) cells in the presence of escalating concentrations of oseltamivir. Population genetics analysis reveals the fine-scale changes associated with the presence of oseltamivir and suggests that few evolutionary obstacles prevent the development of oseltamivir resistance.

MATERIALS AND METHODS

Cells, virus stocks, and chemicals. MDCK cells were obtained from the American Type Culture Collection (Manassas, VA) and were propagated in Eagle's minimal essential medium (EMEM) with 10% fetal bovine serum (FBS; HyClone, Logan, UT) and 2 mM penicillin-streptomycin. Influenza virus A/Brisbane/59/2007 (H1N1), grown in chicken egg allantoic fluid, was obtained through the NIH Biodefense and Emerging Infections Research Resources Repository, NIAID, NIH (NR-12282; lot 58550257). Oseltamivir carboxylate (RO0640802-002; lot 91ST1126/1) was obtained from F. Hoffmann-La Roche Ltd., Basel, Switzerland.

Determination of viral titers by plaque assay. Viruses were quantified on MDCK cells to determine infectious titers (expressed in PFU per milliliter) as described previously (12). In brief, 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 unbound virus was washed off with phosphate-buffered saline (PBS), the cells were overlaid with agar (0.5%) in Dulbecco's modified Eagle medium (DMEM)-F-12 medium supplemented with penicillin-streptomycin, L-glutamine, bovine serum albumin, HEPES, sodium bicarbonate, and 20 µg/ml acetylated trypsin (Sigma, St. Louis, MO). After the agar solidified, the plates were incubated for ~48 h at 37°C. Cells were fixed and stained with an anti-H1 primary antibody (MAB8261; Millipore, Billerica, MA). Plaques were visualized with a horseradish peroxidase-conjugated anti-mouse secondary antibody (BD Biosciences, San Jose, CA) and were developed with a peroxidase substrate kit (Vector Laboratories, Burlingame, CA).

Determination of oseltamivir ED₅₀. The 50% effective dose (ED₅₀) was defined as the concentration of drug reducing the number of plaques to 50% of that for the no-drug control. In brief, the ED₅₀ was determined by seeding 2.5×10^5 MDCK cells in each well of a 24-well plate and incubating overnight at 37°C under 5% CO₂. Virus was added to cells at a multiplicity of infection (MOI) of 0.01 in 100 µl of influenza virus growth medium (EMEM-10% FBS with 2 mM penicillin-streptomycin, 7.5% bovine serum albumin, and 1 µg/ml tosylsulfonyl phenylalanyl chloromethyl ketone [TPCK]-treated-trypsin [Sigma]) plus oseltamivir (0, 0.001, 0.01, 0.1, 1, 10, or 100 µM). After incubation at 37°C for 1 h, cells were washed once with PBS; 500 µl of influenza virus growth medium with the appropriate concentration of oseltamivir was added; and cells were again incubated at 37°C for several days. Supernatants were collected when >90% cytopathic effect (CPE) was achieved for at least one drug concen-

tration. Supernatants were centrifuged for 15 min at 300 relative centrifugal force (RCF) at 4°C and were stored at -80°C. The viral titer for each sample was determined by plaque assay. The ED₅₀ was the concentration of oseltamivir yielding a titer of 50% of that without drug.

Determination of oseltamivir IC₅₀. The 50% enzyme-inhibitory concentration (IC₅₀) of oseltamivir against IAV was determined by using a chemiluminescent neuraminidase activity inhibition (NI) assay, commercially available as the NA-Star Influenza Neuraminidase Inhibitor Resistance Detection kit (Applied Biosystems, Foster City, CA), according to the manufacturer's instructions. The values were determined from a dose-response curve.

Viral culture. Viruses were serially passaged in MDCK cells (2.5×10^5 cells per well). The MOI for the initial infection was 0.01. Multiple serial 10-fold dilutions of virus in supernatants were used to select virus for subsequent passages. Supernatants containing virus were harvested as the culture reached >50% CPE, and the cell-free virus was used for processing for high-throughput sequencing, quantification by plaque assay, and generation of the next passage. The mean MOI for additional passages was later determined to be 0.04 (range, 0.00001 to 0.1). Trajectories comprising at least 10 serial passages were prepared both in the presence and in the absence of escalating doses of oseltamivir. In the first passage with oseltamivir, the drug concentration was $1 \times$ the ED₅₀. For the next passage, the concentration was increased to $4 \times$ the ED₅₀, and it was then doubled for each subsequent passage as long as >50% CPE was present. If <50% CPE was present, the dose of oseltamivir was escalated at a lower rate. For select passages, oseltamivir was deliberately withdrawn.

High-throughput sequencing. We developed a high-throughput sample processing work flow, carried out in a 96-well format, including RNA purification, reverse transcription, and whole-genome PCR, followed by DNA bar coding and library preparation. Samples were purified using the RNeasy 96 kit (Qiagen, Gaithersburg, MD). Reverse transcription was performed with SuperScript III First-Strand Synthesis Supermix (Life Technologies, Grand Island, NY) by using the recommended protocol and IAV universal primers described previously (13). A universal, multiplex PCR was developed that produces full-length products of the entire IAV genome in a single reaction. The primers, described by Hoffmann et al. (14), were combined in a ratio that provided near-even sequence coverage across all segments. The amplification was performed using PfuUltra II HotStart high-fidelity polymerase (Agilent, Santa Clara, CA) with the following cycles: 95°C for 2 min; 32 cycles of 95°C for 15 s, 62°C for 15 s, and 72°C for 2.5 min; and finally 72°C for 10 min. Amplification products were prepared for Illumina sequencing using the MinElute 96 UF PCR purification kit (Qiagen) with the recommended protocol. For all subsequent steps, the same purification kit, with reagents from New England BioLabs (Ipswich, MA), was used between enzymatic reactions, except where otherwise noted. The amplified DNA was sheared using Fragmentase to produce fragments of 300 to 600 bp. The sheared DNA was repaired with the Blunt End Mix and was A tailed with Klenow fragment (Exo⁻). Bar codes containing either 3 or 6 nucleotide identifiers were ligated onto the fragments with T4 DNA ligase. The ligation products were size selected using AMPure XP beads (Agencourt, Beverly, MA) by adding $0.8 \times$ beads, collecting the supernatant, adding $1.6 \times$ beads, discarding the supernatant, and eluting in double-distilled water (ddH₂O). Size selection with beads replaced purification using the MinElute 96 UF system at this stage. The size-selected products were amplified using Illumina PE PCR primers and PfuUltra II HotStart high-fidelity polymerase with the following cycles: 95°C for 2 min; 12 cycles of 95°C for 20 s, 60°C for 20 s, and 72°C for 45 s; and finally 72°C for 10 min. The amplified products were purified, quantified on a NanoDrop 2000 spectrophotometer, combined into libraries, and sequenced on the Illumina HiSeq 2000 platform to generate 100 nucleotide reads. All libraries contained a control RNA product that was used for sequencing error analysis (described below).

Bioinformatics analysis. Sequence analysis of IAV populations could in theory be challenging with short-read sequencing. The high diversity of

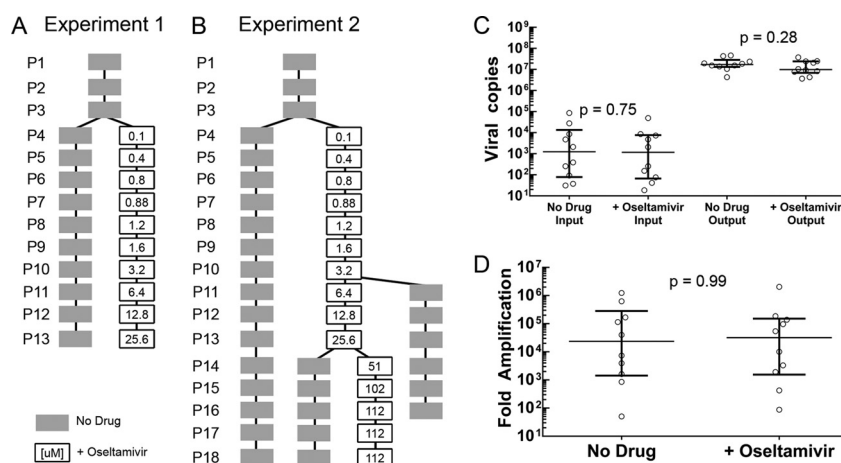


FIG 1 Experimental trajectories. (A) A/Brisbane/59/2007 (H1N1) was serially passaged in MDCK cells 13 times in the presence of escalating micromolar concentrations of oseltamivir (open boxes) or no drug (shaded boxes). (B) In a second experiment, A/Brisbane/59/2007 (H1N1) was subjected to similar growth conditions. Additionally, oseltamivir was withdrawn after passage 10 (P10) and P13, and higher concentrations of oseltamivir were used from P14 to P18. (C) Viral quantities were compared for no-drug versus oseltamivir-treated samples from experiment 1. Horizontal bars indicate the average values for samples. No significant differences are observed between input (P , 0.75 by the Mann-Whitney U test) or output (P , 0.28 by the Mann-Whitney U test) copies. (D) Viral amplification was compared for no-drug samples and oseltamivir-treated samples; again, no significant difference is observed (P , 0.99 by the Mann-Whitney U test).

the viral population could lead to incorrect or poor mapping of short reads with limited homology to a reference genome. Several publicly available alignment programs were evaluated for the ability to map reads from IAV populations. BLAST alignment was found to generate the highest mapping quality with the highest proportion of mapped reads from the data. The BLAST algorithm was integrated into an analysis pipeline to trim and bin the raw read data based on bar codes, align them to the appropriate reference genome, and obtain as output both the level of nucleotide variability and the level of amino acid variability within the viral population. To streamline the processing of large numbers of IAV samples, a structured query language (SQL) database with a Web interface has been developed, integrating sample growth conditions with DNA barcoding information. The database was directly accessed using the analysis pipeline, eliminating the potential for human error in correlating experimental conditions with large-scale IAV genomic data.

Short reads from the Illumina platform were filtered for quality scores of >20 throughout the read and were aligned to the strain's reference genome using BLAST. More than 95% of the selected reads could be mapped to the IAV reference genome obtained from GenBank (accession numbers [CY030232](#), [CY031391](#), [CY058484](#) to [CY058486](#), [CY058488](#), [CY058489](#), and [CY058491](#)). Only alignments longer than 80 nucleotides were retained. The median sequencing depth was 34,650. Amino acid frequencies were calculated after alignment of translated reads to the corresponding positions in the reference proteins. We confirmed that nucleotide and amino acid frequencies were identical between passages. Unfolded SNP frequencies were generated using the IAV reference genome and were used for the population genetics analyses; the amino acid frequencies were used for the structural analysis. The sequencing data sets generated in this study are available at <http://bib.umassmed.edu/influenza>.

Sequencing error analysis. Sequence errors can be introduced either at the sample-processing (PCR) stage, during amplification, or during sequencing by synthesis. Plasmids encoding the full-length segment 6 (NA) from A/Brisbane/59/2007 or a region of human cytomegalovirus (HCMV) with a length and GC content similar to those of the NA segment were created. RNA was generated from these plasmids using T7 RNA polymerase. The RNA pools were then introduced into the sample-processing pipeline and were processed and sequenced as controls in a manner identical to that for viral samples. All sequencing runs included the

error constructs to account for any run-to-run variation in error rates. From these data, the level of error introduced from both sample processing and sequencing was estimated. Briefly, we calculated the percentage of nonreference nucleotides in all reads that were mapped to the control sequence. The cumulative probability of these nonreference nucleotides indicated that the 95% upper confidence limit of the error rate was 0 to 0.17%. Therefore, in order to confidently identify low-frequency variants, we discarded all variants that occurred at a frequency of 0.17% or less (see Fig. 2A).

Population genetic analysis. Under a standard single-locus population genetic model, the expected change in gene frequency per generation owing to selection was described by Wright (15). Assuming a constant population size, it is thus conceivable to directly estimate the selection coefficients (s) by solving the equation $x_{n+1} = x_n \times (1 + s)^{\Delta t}$, where x_n and x_{n+1} are the allele frequencies at passages n and $n + 1$, respectively, and Δt is the number of generations between passages n and $n + 1$. In this way, s values per mutation were inferred from the derived allele frequency for passages 4 to 9 (P4 to P9) in experiment 1 (and P4 to P12 in experiment 2) in both drug-treated and untreated samples.

Structural analysis. The selection coefficients (described above) were averaged across replicates (experiments 1 and 2) for each nucleotide. Using PFAAT (16), the maximum values for the three positions within each codon were loaded as residue annotations and were mapped to a related NA structure (Protein Data Bank [PDB] code 3CL0). The selection coefficients were converted to a color spectrum, mapped to the structure, and visualized in PyMOL.

RESULTS

To delineate the genetic mechanisms that contribute to drug resistance, IAV strain A/Brisbane/59/2007 (H1N1) was subjected to progressively increasing concentrations of oseltamivir, and the whole-genome sequence was analyzed over 13 serial passages. Two evolutionary trajectories for A/Brisbane/59/2007 in MDCK cells, one in the presence and one in the absence of oseltamivir, are summarized in Fig. 1. IAV from chicken eggs was propagated in MDCK cells for 3 passages (passages 1 to 3 [P1 to P3]) to establish a viral stock for parallel trajectories in the presence or absence of drug. For the first experiment (Fig. 1A), the virus was passaged 10 additional times in the presence of escalating concentrations of

oseltamivir, starting with $1 \times$ the ED_{50} ($0.1 \mu\text{M}$) at P4. By P13, the evolved virus was able to tolerate a concentration of oseltamivir 256-fold higher than the ED_{50} . In the second experiment (Fig. 1B), conditions were expanded to include further escalation of the oseltamivir concentration, as well as drug withdrawal beyond P10 and P13. The MOI for each passage was variable. Input virus and output virus were subsequently quantified for all samples. The average amounts of input virus in the presence and absence of oseltamivir were similar (Fig. 1C). No significant differences were observed in the amounts of output virus collected (Fig. 1D). Hence, oseltamivir had no overall impact on viral amplification.

Influenza A virus genome sequence diversity with or without oseltamivir. To initially characterize the populations, the distributions of SNP frequencies were analyzed. Several hundred sites out of 13,500 total nucleotide positions were polymorphic at frequencies above the error frequency cutoff (defined as $\geq 0.17\%$) (Fig. 2A and Table 1). Yet the majority of SNPs were present at extremely low frequencies. Figure 2B shows the SNP minor allele frequency spectra for A/Brisbane/59/2007 at P4 and P13 without oseltamivir in experiment 1: only a small number of minor alleles had an increased frequency in the population. Figure 2C shows the minor allele frequency spectra for A/Brisbane/59/2007 at P13 with oseltamivir in the same experiment. The striking similarity of the frequency spectra suggests that the selective pressure induced by oseltamivir does not cause genomewide alterations in levels of variation. Table 1 shows the numbers of sites in the IAV genome where the minor alleles were present at frequencies exceeding various thresholds, ranging from 0.17% (sequencing noise level) to 90% (nearly complete fixation of a mutation), for P4 and P13 with and without oseltamivir in the two experiments. Fewer than 100 sites, or 0.7% of the genome, showed appreciable variation (frequency of minor allele, $>2\%$ in the population) during the experiment, demonstrating that the A/Brisbane/59/2007 influenza virus genome sequence is strongly constrained. For example, the numbers of sites with variation of $>5\%$ at P4 and P13 were only 35 and 42, respectively, in experiment 1 without the drug. These numbers reflect a strong degree of conservation of the influenza virus genome, suggesting that only a small number of sites experience appreciable variation during the serial passages.

To further characterize the influence of oseltamivir on IAV genetic variation, population-level nucleotide diversity was calculated either for the whole genome (Fig. 3A) or for segment 6 (NA) alone (Fig. 3B). Nucleotide diversity is an estimate of the average pairwise percentage of difference between any sequences in a population. Nucleotide diversity increased slowly with serial passaging in the A/Brisbane/59/2007 populations. Because strong selection is expected to reduce variation, attention was given to the diversity of the oseltamivir-treated populations after the introduction of drug at P4. The diversity of segment 6 (NA) alone accounted for approximately half of the average genomewide diversity. However, the addition of oseltamivir at P4 seemed to have little effect on IAV population diversity, whether analyzed as a genomewide average (Fig. 3A) or in segment 6 (NA) alone (Fig. 3B). In contrast, the degree of viral amplification appeared to have a considerable impact on nucleotide diversity. We observed that samples with smaller amounts of input virus and a greater degree of viral amplification had higher nucleotide diversity (Fig. 3C), showing that the level of replication in each passage is a stronger predictor of population diversity than the presence or absence of drug. Strong selection acting on one population is also expected to

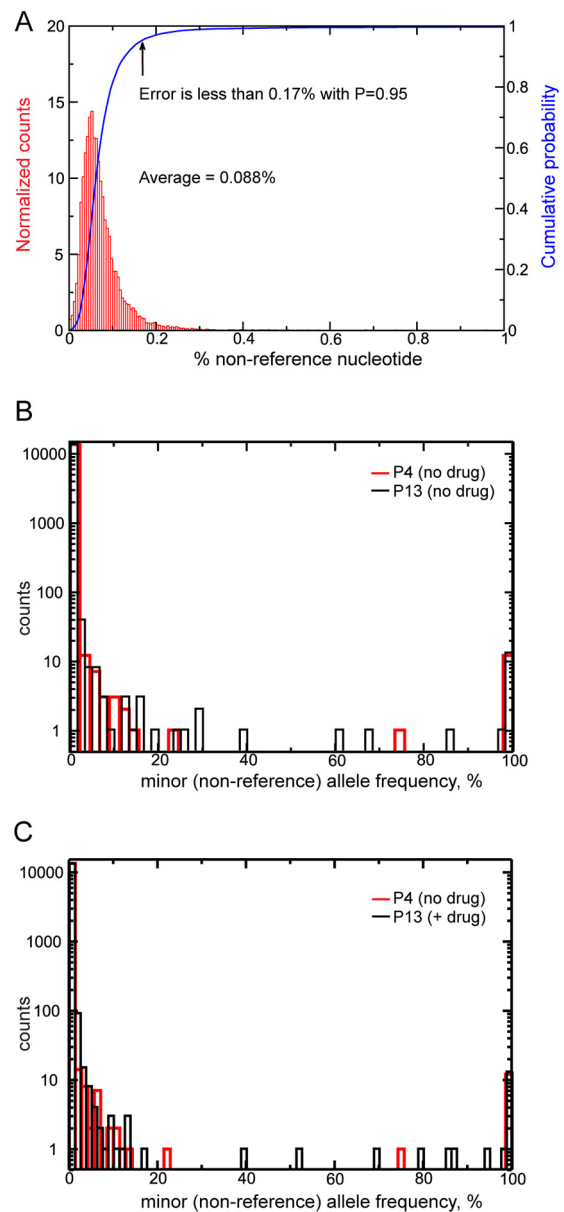


FIG 2 (A) Analysis of error frequency distribution. RNA error controls were processed and sequenced in a manner identical to that for viral population samples. These data were used to apply a filtering threshold to the experimental sequence data. The distribution of mutations in the error control data was highly skewed to very low frequencies. By filtering data to remove all variation of $\leq 0.17\%$, $\geq 95\%$ of errors could be removed (arrow), allowing for SNPs to be called in the population with high confidence. (B) Allele frequency spectrum of A/Brisbane/59/2007. Only a very small number of sites in the 13.5-kb IAV genome exhibited polymorphisms or changes during serial passages in the absence of oseltamivir. P4 (no drug) is designated by red columns and P13 (no drug) by black columns. (C) Serial passaging with increasing doses of oseltamivir did not increase the sequence diversity of influenza virus. Bars outlined in red, P4 (no drug); bars outlined in black, P13 (with oseltamivir). Data for experiment 1 are shown; data for experiment 2 are similar.

increase the rate of differentiation between two isolated populations. F_{ST} , a metric of population differentiation (17, 18), was estimated between all samples of this study for each passage. The population differentiation between drug-treated and untreated controls was similar to that between two untreated populations

TABLE 1 Numbers of sites where SNP frequencies exceed various thresholds^a

Threshold (%)	No. of sites where the SNP ^b frequency exceeds the threshold							
	Expt 1				Expt 2			
	No drug		Oseltamivir		No drug		Oseltamivir	
	P4	P13	P4	P13	P4	P13	P4	P13
>0.17	950	1,665	637	1,825	4,134	1,867	3,528	1,698
>0.5	358	353	157	557	728	769	700	543
>1	172	157	72	206	129	327	180	227
>2	76	75	39	83	52	98	114	84
>5	35	42	28	37	38	32	38	50
>10	22	30	23	27	28	25	27	38
>25	14	21	13	21	20	18	21	25
>50	13	17	13	20	14	16	14	19
>90	12	14	13	15	12	15	13	17

^a From 0.17% (sequencing error limit) to 90% (fixation).
^b SNP refers to sequence variation with respect to the GenBank reference sequence.

from experiments 1 and 2 (Fig. 3D). These data suggest that oseltamivir treatment does not cause a genomewide differentiation of IAV populations.

Population dynamics of oseltamivir resistance in A/Brisbane/59/2007. Changes in the frequency of individual SNPs were analyzed as a function of passage number. In all experiments and passages, 50% of SNPs were present at low frequencies ($\leq 0.25\%$) (Fig. 2 and 4). For each experiment, a greater number of SNPs fixed in the trajectories with oseltamivir than in the trajectories with no drug. The most striking effect of the drug was observed in experiment 1, in which sharp rises in the frequencies of four SNPs, located in segment 4 (hemagglutinin [HA]) and segment 6 (NA), were observed beginning at P6 (Fig. 4). Yet in experiment 2, the overall patterns in the presence and absence of the drug were similar. In total, the frequency changes for the majority of SNPs in both experiments did not appear to be altered by the presence of oseltamivir.

Analysis of the changes in the frequencies of individual SNPs showed that the known resistance mutation H274Y (2) was the only SNP that fixed in the presence of oseltamivir in both experi-

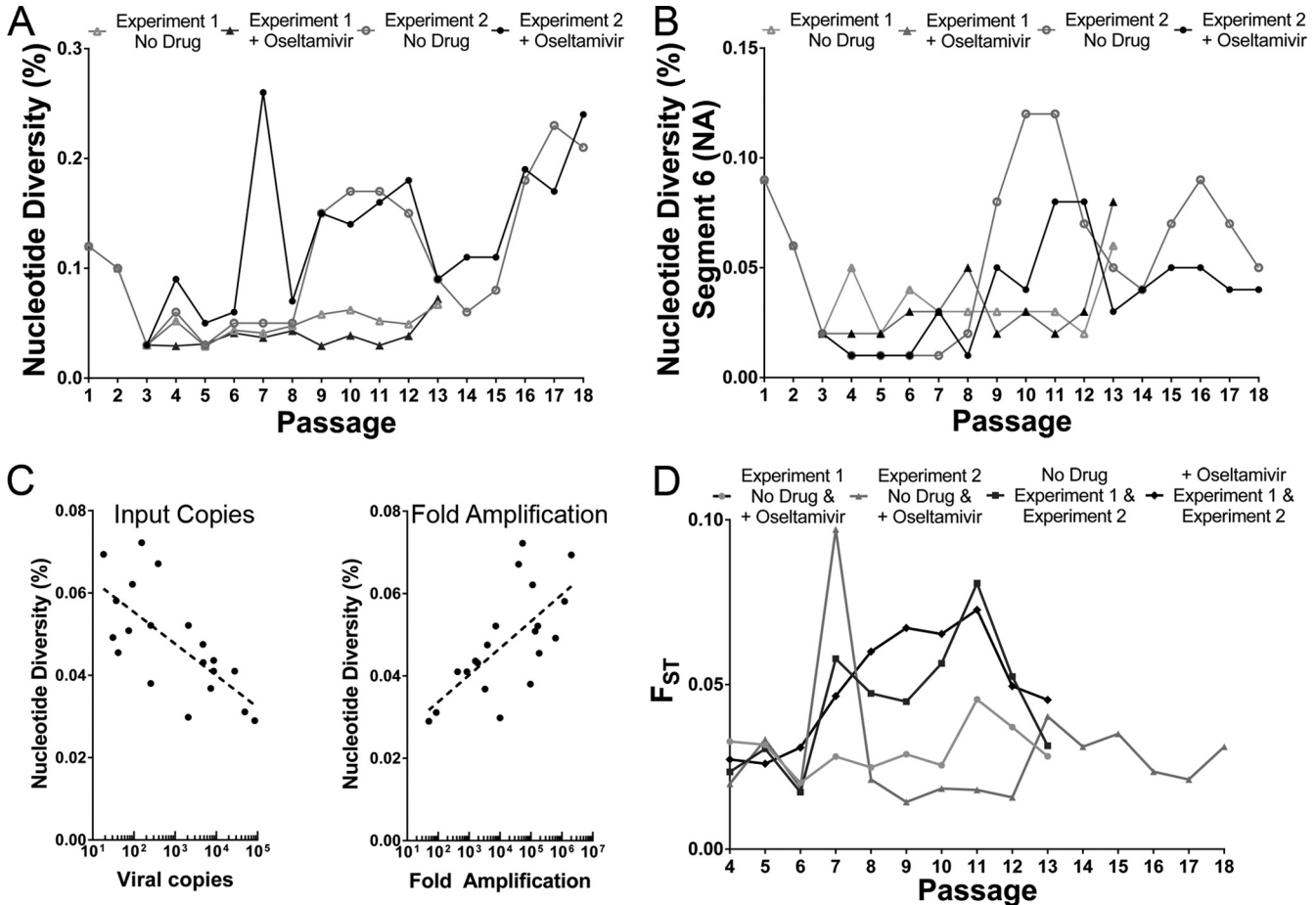


FIG 3 IAV population diversity and divergence in the presence and absence of oseltamivir. The nucleotide diversity of IAV populations was measured for each passage of *in vitro* growth. Population divergence was measured by comparing two populations at the same passage using the F_{ST} statistic, a metric of population differentiation. (A) The presence of oseltamivir did not cause a significant change in the amount of variation in the total population. (B) Oseltamivir did not cause changes in variation in segment 6 (neuraminidase [NA]). (C) Nucleotide diversity was influenced by the amount of input virus and by fold amplification. Trends are shown by the dashed lines. R^2 (Pearson correlation), 0.49 for input virus versus nucleotide diversity and 0.46 for fold amplification versus nucleotide diversity. Data from experiment 1 are shown. (D) Two populations grown with and without oseltamivir were no more divergent than two populations grown without drug from the two experiments.

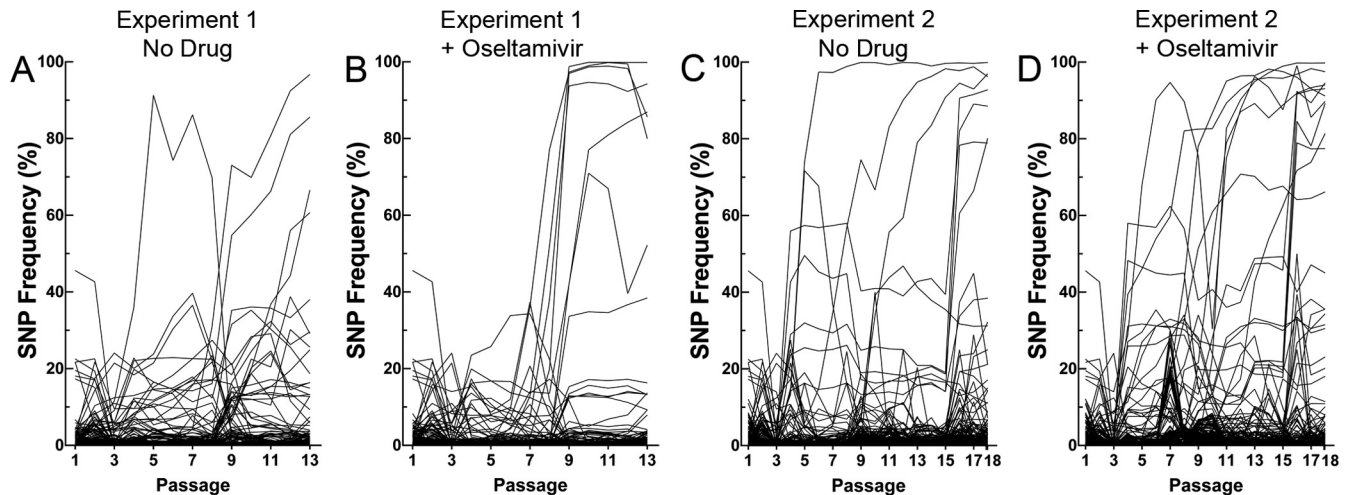


FIG 4 SNP frequency changes during *in vitro* growth with or without oseltamivir. The frequencies of all SNPs across the IAV genome were tracked for 13 (A and B) or 18 (C and D) passages. The presence of oseltamivir caused an increase in the number of SNPs that fixed during passaging. The effect of oseltamivir was especially pronounced in experiment 1, where a sudden increase in the frequency of a cluster of SNPs can be observed beginning at P6. However, the majority of SNP frequency changes were not affected by the presence of oseltamivir.

ments 1 and 2. Remarkably, before the introduction of the drug, the maximum frequency of the H274Y mutation was 0.085%, below the accurate detection limit of 0.17% (Fig. 5). Therefore, the H274Y mutation either occurred at extremely low frequencies or was a *de novo* mutation that arose during the experimental trajectory. The H274Y mutation occurred via a single nucleotide substitution, C822T.

These data suggest that positive selection causes a sharp rise in the frequency of the H274Y mutation and possibly others across the genome. The presence of oseltamivir may have caused a change in the selective environment, altering the fitness of SNPs

across the genome. To test this hypothesis, selection coefficients were calculated using time-sampled changes in SNP frequencies (15). Selection coefficients quantify the fitness advantage or disadvantage of a SNP. Values can range from negative for deleterious SNPs to zero for neutral SNPs and positive for beneficial SNPs. Oseltamivir did not cause a significant skewing in the genomewide distributions of selection coefficients in either experiment (Fig. 6). These results suggest that the selective environment at the genomic level is not altered by oseltamivir. However, a fine-scale analysis of selection coefficients for all SNPs of segment 6 (NA) showed that the strength and targets of positive selection were altered in the presence of oseltamivir. The H274Y mutation was shown to be the most strongly beneficial (Fig. 7), although there were additionally several low-frequency mutations (G193R, E227D, K248E, and P326T) near the oseltamivir binding site with positive selection coefficients. E227D had positive selection coefficients both in the presence and in the absence of drug and thus was not specific for passage with oseltamivir. Notably, the frequency of the K248E mutation, which is adjacent to H274Y, increased to 19% in the second experiment (P12). As the more beneficial H274Y mutation approached 100% fixation (P13 onward), the frequency of the K248E mutation was reduced accordingly to less than 1% in the population. The frequency of the G193R mutation was less than 2%, and the other two mutations (E227D and P326T) occurred at low frequencies in both the presence and the absence of drug. Taken together, these results demonstrate that the H274Y mutation is the primary target of selection in NA and that the fixation of H274Y in the presence of oseltamivir supersedes other potential targets of selection (e.g., K248E).

The H274Y mutation is neutral in the absence of oseltamivir.

To assess the fitness of the H274Y mutation in the absence of drug, oseltamivir was withdrawn at key stages in experiment 2. Following P13, where the frequency of H274Y was 95%, two parallel trajectories were continued either in the presence or in the absence of drug (Fig. 1B). Oseltamivir was either continued at escalating concentrations or withdrawn for five additional passages (P14 to P18). The frequency of H274Y fixed to 99% by

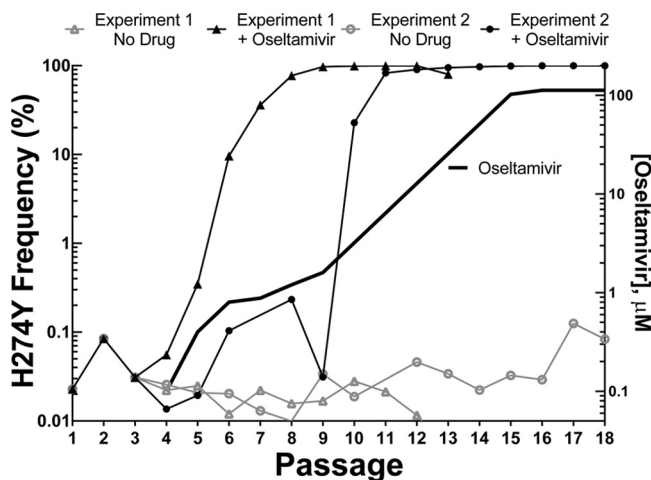


FIG 5 Frequency of the H274Y mutation versus the passage number in A/Brisbane/59/2007 grown in MDCK cells in the presence or absence of oseltamivir. At oseltamivir concentrations of $>1 \mu\text{M}$, the known resistance mutation H274Y emerged and fixed in the population in both experiment 1 (filled triangles) and experiment 2 (filled circles). Prior to drug treatment, the frequency of H274Y in the population was below 0.1%. The frequency of H274Y was below 0.1% in the absence of oseltamivir in both experiments (gray lines). Oseltamivir was added at P4 and beyond; the concentration (along the right y axis) is indicated by the heavy black line.

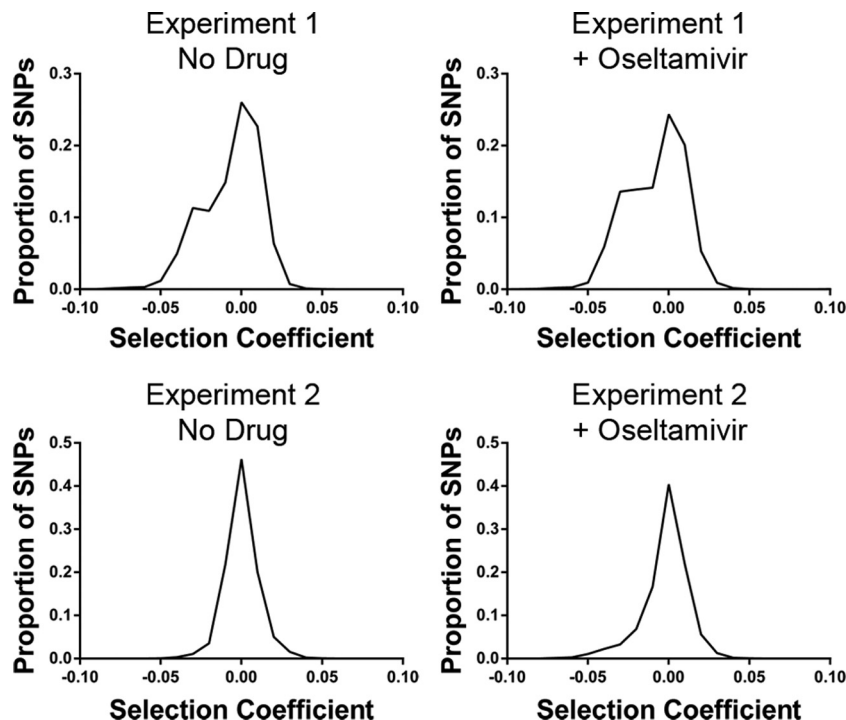


FIG 6 Genomewide distribution of selection coefficients. Selection coefficients for all SNPs across the genome were calculated using time-sampled changes in frequency. The distributions were centered on an s of zero, suggesting that most SNPs are neutral and are subject to genetic drift. Further, growth in oseltamivir had little effect on the distribution, suggesting that the genomewide selective environment is not altered by the presence of the drug.

P17 regardless of the presence (Fig. 8, filled triangles) or absence (Fig. 8, open circles) of oseltamivir. Similarly, oseltamivir was withdrawn following P10, where the frequency of H274Y was 23%. Despite the withdrawal of oseltamivir for six additional passages (P11 to P16), the frequency of the mutation was

maintained at comparable levels (30 to 40%) and appeared to be governed by genetic drift (Fig. 8, open triangles). Collectively, these results demonstrate that the H274Y mutation is likely neutral in the absence of drug during serial passaging of A/Brisbane/59/2007 in MDCK cells.

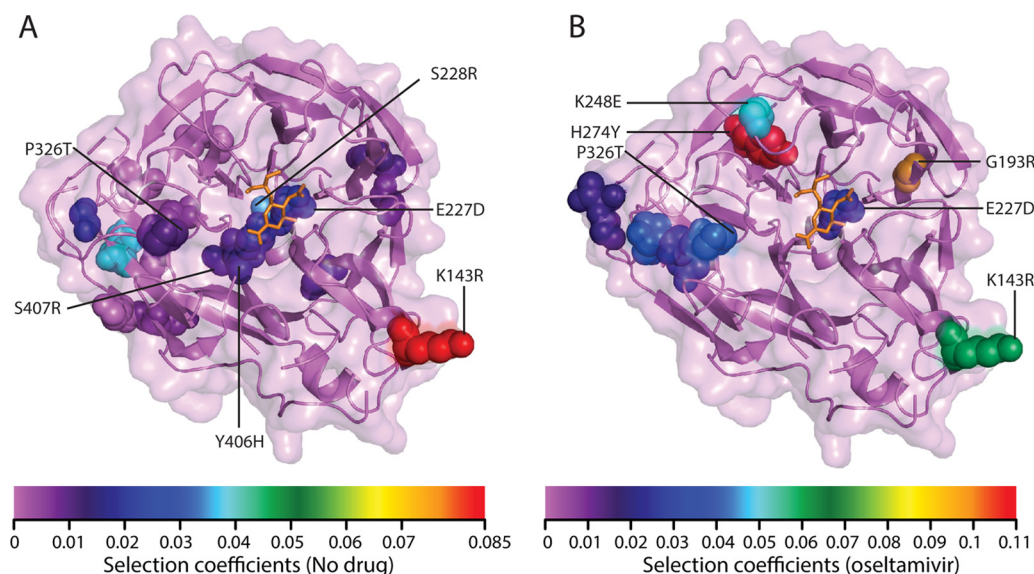


FIG 7 Selection coefficients for neuraminidase in A/Brisbane/59/2007. Selection coefficients were calculated for the periods corresponding to the development of drug resistance (P4 onward) with no drug (A) and with increasing concentrations of oseltamivir (B). Data were compiled from experiments 1 and 2. The selection coefficients were mapped to the crystal structure of neuraminidase in complex with oseltamivir (PDB 3CL0). The H274Y resistance mutation is the most beneficial mutation in the population. The K248E mutation is adjacent to H274Y and is a target of positive selection with oseltamivir in one of two experiments. The spectrum of colors ranges from the lowest-ranked (violet) to the highest-ranked (red) selection coefficients.

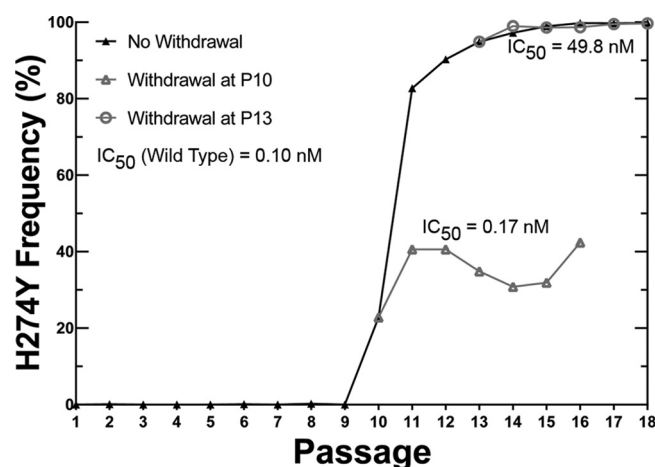


FIG 8 Impact of oseltamivir withdrawal on H274Y frequency in A/Brisbane/59/2007 and correlation with IC_{50} s for neuraminidase. Shown is the frequency of the H274Y mutation for select samples in the second experimental trajectory with A/Brisbane/59/2007. After P13, the removal of the drug has no effect on H274Y frequency (open circles). The removal of the drug at P10, before H274Y reached fixation, did not lead to a significant change in H274Y frequency (open triangles). Mean oseltamivir IC_{50} s for influenza virus samples beyond P13 and P10 versus those for control virus (wild type) are given. Values were determined using a chemiluminescent neuraminidase activity inhibition assay. For samples in which H274Y remains fixed in the population (open circles), the IC_{50} is >100 -fold that for the wild-type control virus (means, 49.8 ± 6.8 nM versus 0.10 ± 0.01 nM; $P < 0.001$ by Student's t test). For samples in which the frequency of H274Y is $\sim 40\%$ in the population (open triangles), the IC_{50} is slightly but significantly greater than that for the wild-type control virus (means, 0.17 ± 0.02 nM versus 0.10 ± 0.01 nM; $P < 0.001$ by Student's t test).

IC_{50} measurements correlate with the frequency of the oseltamivir resistance mutation. To assess if oseltamivir resistance correlated with the relative frequency of the H274Y mutation in strain A/Brisbane/59/2007, enzyme inhibition studies were performed on select samples. The 50% enzyme-inhibitory concentrations (IC_{50}) of oseltamivir were assessed on samples without oseltamivir in the supernatant to provide for accurate measurements. Successful replication of resistant viruses in the presence of increasing concentrations of oseltamivir was accompanied by the increases in the enzymatic IC_{50} of oseltamivir for neuraminidase (Fig. 8). At P13 and beyond, the H274Y mutation was present at a frequency of $>99\%$. The mean IC_{50} for these samples was 49.8 ± 6.8 nM ($n = 3$). This was significantly higher than the IC_{50} for neuraminidase in the oseltamivir-sensitive control samples (0.10 ± 0.01 nM [$n = 6$]; $P < 0.0001$ by Student's unpaired t test), in which the H274Y mutation was below the detection limit, and indicates oseltamivir resistance at P13. Following oseltamivir withdrawal at P10, the frequency of H274Y drifted between 30% and 40%. The corresponding mean IC_{50} was 0.17 ± 0.02 nM ($n = 3$), also significantly higher than that for oseltamivir-sensitive control samples (0.10 ± 0.01 nM) ($P < 0.01$ by Student's unpaired t test). Therefore, the presence of the H274Y mutation in the viral population, either at a low or at a high frequency, led to measurable changes in the neuraminidase inhibition assay results.

DISCUSSION

Accurately unraveling the evolutionary changes of a viral population in response to a selective pressure was previously elusive. In the past, information about influenza virus sequence diversity was

typically inferred from large sets of consensus sequences determined from clinical isolates (19), and it was difficult to resolve the genetic structure of well-defined individual IAV populations. In the current study, we have used high-throughput sequencing to closely monitor the population genetics of IAV strain A/Brisbane/59/2007 in response to oseltamivir. Surprisingly, the genomewide effects of oseltamivir were minimal, although high-resolution analysis revealed the subtleties associated with drug resistance.

IAV can rapidly evolve and adapt to drugs or changes in the environment through rapid mutation and replication, strong natural selection, and genome segment reassortment. By subjecting A/Brisbane/59/2007 to oseltamivir under controlled conditions, we were able to track the frequency of resistant mutants over time. Initially, mutants occurred at exceptionally low frequencies ($\leq 0.085\%$), and they rose to fixation within only a few serial passages in each case. Since the mutation rate of influenza virus *in vitro* is between 10^{-6} and 10^{-5} mutation per site per infectious cycle (20), the resistance mutation may have developed *de novo*. Strong population bottlenecks during serial passaging at a low MOI would make it unlikely for the preexisting resistance mutation to remain at such a low frequency ($\sim 10^{-4}$) over multiple passages prior to the introduction of the drug. We also expect the evolutionary trajectories to be similar in individual infected humans, since the H274Y resistance mutation was observed at low frequencies ($\leq 3.7\%$) in influenza virus samples isolated from a single patient prior to oseltamivir therapy (21, 22). Our results show that the development of oseltamivir resistance in A/Brisbane/59/2007 is not associated with large changes in population sequence diversity or divergence. Further, the genomewide distribution of selection coefficients is not significantly altered by the presence of the drug, suggesting that the overall selective environments for A/Brisbane/59/2007 in MDCK cells are similar in the presence and absence of drug. If the development of resistance were associated with genomewide changes in constraint or increased strength of epistatic interactions, one would expect genomewide alteration of the distribution of selection coefficients due to alterations of the fitness landscape (23). Given these results and the fact that the H274Y mutation arises from a single nucleotide change, not much may hinder the conversion of a drug-sensitive population to drug resistance. These data support the observed rise in the occurrence of oseltamivir resistance in seasonal H1N1 isolates from 2007 to 2009 (5). Others have found epistatic interactions between the H274Y mutation and other mutations in NA: the fitness of the resistant virus is increased by the presence of the V234M and R222Q mutations (24). Because these mutations are encoded by strain A/Brisbane/59/2007 as well as by many recent H1N1 isolates (24), our data show how current H1N1 strains are "primed" for the development of resistance with little impediment at the genomewide or population level. The absence of a fitness penalty for the H274Y mutation in quasispecies populations may contribute to its prevalence in humans. An interesting question to address using our system is whether other resistance mutations in the neuraminidase gene are stable or would disappear from the population in the absence of permissive secondary mutations. Future drug design should be influenced not only by structural considerations, as was the case for oseltamivir (25, 26), but also by evolutionary information, to reduce the likelihood of resistance.

Surprisingly, variation in the influenza virus genome sequence was constrained. Out of the entire 13.5-kb IAV genome, signifi-

cant variation developed in only about 100 sites ($\sim 0.7\%$) for the A/Brisbane/59/2007 experiments over 13 passages (Table 1). Further, the nucleotide diversity of the genome was significantly less than that of other RNA viruses, such as hepatitis C virus (27), although the latter was sampled from *in vivo* populations. Such a strong degree of conservation in a rapidly mutating RNA virus can be due to a combination of factors, including constraints on protein function, correct folding and packaging of the RNA segments, and maintenance of appropriate codon bias. In an initial effort to better understand these events, we investigated how influenza virus genome packaging can be impacted by factors such as segment self-repulsion (28), highlighting the importance of packaging signals for viral replication. Further understanding of the fitness consequences of SNPs across the genome will help elucidate the mechanism underlying the conservation.

Several neuraminidase inhibitor resistance mutations have been reported in the literature (29). However, when A/Brisbane/59/2007 was treated with oseltamivir in two separate experiments, the H274Y mutation fixed rapidly in both populations. Although the H274Y mutation was the primary target of selection in A/Brisbane/59/2007, the neighboring K248E mutation was observed at relatively high frequencies in one of the experiments (19% in P12). The K248E mutation was also identified as a target of positive selection during this period. However, as the more beneficial H274Y mutation approached fixation (P13 onward), there was a corresponding reduction in the frequency of the K248E mutation in the population (below 1%), highlighting the importance of competition during viral evolution.

Our enzyme inhibition studies demonstrated a relationship between oseltamivir resistance and the relative frequency of the H274Y mutation in the population. In the presence of oseltamivir, the mutation provides an unequivocal selective advantage that leads to its rapid fixation, indicated by the reproducibly positive selection coefficients for this mutation. However, as shown in the enzymatic assays, the full functional advantage of the mutation is not apparent until it is present at very high frequencies. When this mutation was present in approximately one-third of the population, the enzymatic activity was slightly, though significantly, higher than that of a control population without the mutation. These data provide a functional basis for the rapid fixation of the mutation in the presence of oseltamivir during *in vitro* growth in this study, or during *in vivo* replication, as observed by others (21).

Collectively, temporal high-throughput sequencing studies for oseltamivir provide a framework for analysis of the evolution of IAV under additional selective pressures, such as novel antiviral agents and antiviral immunity. Such analyses will be crucial for understanding the evolution of drug resistance and for improving vaccine design. Future temporal high-throughput sequencing studies of influenza virus will help address many of the challenges posed by this pathogen and will bring us closer to relevant, robust *in vitro* modeling of influenza virus evolution.

ACKNOWLEDGMENTS

We acknowledge the support of DARPA (Prophecy Program, Defense Advanced Research Projects Agency, Defense Sciences Office [DSO], contract HR0011-11-C-0095).

We also acknowledge the contributions of all the members of the ALiVE (Algorithms to Limit Viral Epidemics) working group.

REFERENCES

1. Palese P, Shaw M. 2007. *Orthomyxoviridae*: the viruses and their replication, p 1647–1689. In Knipe DM, Howley PM, Griffin DE, Lamb RA, Martin MA, Roizman B, Straus SE (ed), *Fields virology*, 5th ed. Lippincott Williams & Wilkins, Philadelphia, PA.
2. Gubareva LV, Kaiser L, Matrosovich MN, Soo-Hoo Y, Hayden FG. 2001. Selection of influenza virus mutants in experimentally infected volunteers treated with oseltamivir. *J. Infect. Dis.* 183:523–531. <http://dx.doi.org/10.1086/318537>.
3. Lackenby A, Hungnes O, Dudman SG, Meijer A, Paget WJ, Hay AJ, Zambon MC. 2008. Emergence of resistance to oseltamivir among influenza A (H1N1) viruses in Europe. *Euro Surveill.* 13(5):8026. <http://www.eurosurveillance.org/ViewArticle.aspx?ArticleId=8026>.
4. Dharan NJ, Gubareva LV, Meyer JJ, Okomo-Adhiambo M, McClinton RC, Marshall SA, St George K, Epperson S, Brammer L, Klimov AI, Bresee JS, Fry AM. 2009. Infections with oseltamivir-resistant influenza A (H1N1) virus in the United States. *JAMA* 301:1034–1041. <http://dx.doi.org/10.1001/jama.2009.294>.
5. Moscona A. 2009. Global transmission of oseltamivir-resistant influenza. *N. Engl. J. Med.* 360:953–956. <http://dx.doi.org/10.1056/NEJMp0900648>.
6. Collins PJ, Haire LF, Lin YP, Liu J, Russell RJ, Walker PA, Skehel JJ, Martin SR, Hay AJ, Gamblin SJ. 2008. Crystal structures of oseltamivir-resistant influenza virus neuraminidase mutants. *Nature* 453:1258–1261. <http://dx.doi.org/10.1038/nature06956>.
7. Kiso M, Mitamura K, Sakai-Tagawa Y, Shiraishi K, Kawakami C, Kimura K, Hayden FG, Sugaya N, Kawaoka Y. 2004. Resistant influenza A viruses in children treated with oseltamivir: descriptive study. *Lancet* 364:759–765. [http://dx.doi.org/10.1016/S0140-6736\(04\)16934-1](http://dx.doi.org/10.1016/S0140-6736(04)16934-1).
8. Le QM, Kiso M, Someya K, Sakai YT, Nguyen TH, Nguyen KH, Pham ND, Nguyen HH, Yamada S, Muramoto Y, Horimoto T, Takada A, Goto H, Suzuki T, Suzuki Y, Kawaoka Y. 2005. Avian flu: isolation of drug-resistant H5N1 virus. *Nature* 437:1108. <http://dx.doi.org/10.1038/4371108a>.
9. Nielsen R. 2005. Molecular signatures of natural selection. *Annu. Rev. Genet.* 39:197–218. <http://dx.doi.org/10.1146/annurev.genet.39.073003.112420>.
10. Barreiro LB, Laval G, Quach H, Patin E, Quintana-Murci L. 2008. Natural selection has driven population differentiation in modern humans. *Nat. Genet.* 40:340–345. <http://dx.doi.org/10.1038/ng.78>.
11. Llaure AS, Andino R. 2010. Quasispecies theory and the behavior of RNA viruses. *PLoS Pathog.* 6:e1001005. <http://dx.doi.org/10.1371/journal.ppat.1001005>.
12. Hendricks GL, Weirich KL, Viswanathan K, Li J, Shriver ZH, Ashour J, Ploegh HL, Kurt-Jones EA, Fygenon DK, Finberg RW, Comolli JC, Wang JP. 2013. Sialylneolacto-*N*-tetraose c (LSTc)-bearing liposomal decoys capture influenza A virus. *J. Biol. Chem.* 288:8061–8073. <http://dx.doi.org/10.1074/jbc.M112.437202>.
13. Zhou B, Donnelly ME, Scholes DT, St George K, Hatta M, Kawaoka Y, Wentworth DE. 2009. Single-reaction genomic amplification accelerates sequencing and vaccine production for classical and swine origin human influenza A viruses. *J. Virol.* 83:10309–10313. <http://dx.doi.org/10.1128/JVI.01109-09>.
14. Hoffmann E, Stech J, Guan Y, Webster RG, Perez DR. 2001. Universal primer set for the full-length amplification of all influenza A viruses. *Arch. Virol.* 146:2275–2289. <http://dx.doi.org/10.1007/s007050170002>.
15. Wright S. 1931. Evolution in Mendelian populations. *Genetics* 16:97–159.
16. Caffrey DR, Dana PH, Mathur V, Ocano M, Hong EJ, Wang YE, Somaroo S, Caffrey BE, Potluri S, Huang ES. 2007. PFAAT version 2.0: a tool for editing, annotating, and analyzing multiple sequence alignments. *BMC Bioinformatics* 8:381. <http://dx.doi.org/10.1186/1471-2105-8-381>.
17. Holsinger KE, Weir BS. 2009. Genetics in geographically structured populations: defining, estimating and interpreting FST. *Nat. Rev. Genet.* 10: 639–650. <http://dx.doi.org/10.1038/nrg2611>.
18. Weir BS, Cockerham CC. 1984. Estimating F-statistics for the analysis of population structure. *Evolution* 38:1358–1370. <http://dx.doi.org/10.2307/2408641>.
19. Holmes EC, Ghedin E, Miller N, Taylor J, Bao Y, St George K, Grenfell BT, Salzberg SL, Fraser CM, Lipman DJ, Taubenberger JK. 2005. Whole-genome analysis of human influenza A virus reveals multiple per-

- sistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* 3:e300. <http://dx.doi.org/10.1371/journal.pbio.0030300>.
20. Nobusawa E, Sato K. 2006. Comparison of the mutation rates of human influenza A and B viruses. *J. Virol.* 80:3675–3678. <http://dx.doi.org/10.1128/JVI.80.7.3675-3678.2006>.
 21. Ghedin E, Holmes EC, DePasse JV, Pinilla LT, Fitch A, Hamelin ME, Papenburg J, Boivin G. 2012. Presence of oseltamivir-resistant pandemic A/H1N1 minor variants before drug therapy with subsequent selection and transmission. *J. Infect. Dis.* 206:1504–1511. <http://dx.doi.org/10.1093/infdis/jis571>.
 22. Ghedin E, Laplante J, DePasse J, Wentworth DE, Santos RP, Lepow ML, Porter J, Stellrecht K, Lin X, Operario D, Griesemer S, Fitch A, Halpin RA, Stockwell TB, Spiro DJ, Holmes EC, St George K. 2011. Deep sequencing reveals mixed infection with 2009 pandemic influenza A (H1N1) virus strains and the emergence of oseltamivir resistance. *J. Infect. Dis.* 203:168–174. <http://dx.doi.org/10.1093/infdis/jiq040>.
 23. Wright S. 1932. The roles of mutation, inbreeding, crossbreeding, and selection in evolution, p 355–366. *In* Jones DF (ed), *Proceedings of the Sixth International Congress on Genetics*, vol 1. Brooklyn Botanic Garden, Menasha, WI.
 24. Bloom JD, Gong LI, Baltimore D. 2010. Permissive secondary mutations enable the evolution of influenza oseltamivir resistance. *Science* 328:1272–1275. <http://dx.doi.org/10.1126/science.1187816>.
 25. Kim CU, Lew W, Williams MA, Liu H, Zhang L, Swaminathan S, Bischofberger N, Chen MS, Mendel DB, Tai CY, Laver WG, Stevens RC. 1997. Influenza neuraminidase inhibitors possessing a novel hydrophobic interaction in the enzyme active site: design, synthesis, and structural analysis of carbocyclic sialic acid analogues with potent anti-influenza activity. *J. Am. Chem. Soc.* 119:681–690. <http://dx.doi.org/10.1021/ja963036t>.
 26. Moscona A. 2005. Oseltamivir resistance—disabling our influenza defenses. *N. Engl. J. Med.* 353:2633–2636. <http://dx.doi.org/10.1056/NEJMp058291>.
 27. Sakai A, Kaneko S, Honda M, Matsushita E, Kobayashi K. 1999. Quasispecies of hepatitis C virus in serum and in three different parts of the liver of patients with chronic hepatitis. *Hepatology* 30:556–561. <http://dx.doi.org/10.1002/hep.510300234>.
 28. Venev SV, Zeldovich KB. 2013. Segment self-repulsion is the major driving force of influenza genome packaging. *Phys. Rev. Lett.* 110:098104. <http://dx.doi.org/10.1103/PhysRevLett.110.098104>.
 29. Aoki FY, Boivin G, Roberts N. 2007. Influenza virus susceptibility and resistance to oseltamivir. *Antivir. Ther.* 12:603–616. <http://www.intmedpress.com/journals/avt/abstract.cfm?id=446&pid=88>.