Genome-Scale Evolution and Phylodynamics of Equine H3N8 Influenza A Virus^{∇†}

Pablo R. Murcia,^{1*} James L. N. Wood,¹ and Edward C. Holmes^{2,3}

Cambridge Infectious Diseases Consortium, Department of Veterinary Medicine, University of Cambridge, Madingley Road, CB3 0ES, Cambridge, England, United Kingdom¹; Center for Infectious Disease Dynamics, Pennsylvania State University, University Park, Pennsylvania²; and Fogarty International Center, National Institutes of Health, Bethesda, Maryland³

Received 16 December 2010/Accepted 14 March 2011

Equine influenza viruses (EIVs) of the H3N8 and H7N7 subtypes are the causative agents of an important disease of horses. While EIV H7N7 apparently is extinct, H3N8 viruses have circulated for more than 50 years. Like human influenza viruses, EIV H3N8 caused a transcontinental pandemic followed by further outbreaks and epidemics, even in populations with high vaccination coverage. Recently, EIV H3N8 jumped the species barrier to infect dogs. Despite its importance as an agent of infectious disease, the mechanisms that underpin the evolutionary and epidemiological dynamics of EIV are poorly understood, particularly at a genomic scale. To determine the evolutionary history and phylodynamics of EIV H3N8, we conducted an extensive analysis of 82 complete viral genomes sampled during a 45-year span. We show that both intra- and intersubtype reassortment have played a major role in the evolution of EIV, and we suggest that intrasubtype reassortment resulted in enhanced virulence while heterosubtypic reassortment contributed to the extinction of EIV H7N7. We also show that EIV evolves at a slower rate than other influenza viruses, even though it seems to be subject to similar immune selection pressures. However, a relatively high rate of amino acid replacement is observed in the polymerase acidic (PA) segment, with some evidence for adaptive evolution. Most notably, an analysis of viral population dynamics provided evidence for a major population bottleneck of EIV H3N8 during the 1980s, which we suggest resulted from changes in herd immunity due to an increase in vaccination coverage.

Equine influenza is a significant disease of the horse, and it manifests as a rapidly transmitting respiratory disease associated with fever, coughing, and lethargy. Such a syndrome has been described in the historical literature for centuries (42), although the first serological demonstration of influenza A virus in horses came in 1955 in Sweden (21).

Equine influenza A (EIV) was first isolated from horses during an outbreak of disease in Czechoslovakia in 1956 and subsequently was demonstrated to be a subtype H7N7 virus (62). After that first isolation, infection with H7N7 EIV was associated with respiratory disease in European horse populations for approximately 20 years, with sporadic reports of virus isolation during this time period. EIV was first reported in North America in 1960 and in India in 1964.

A different subtype of EIV (H3N8) was first reported from horses in Florida during an outbreak of disease that started in January 1963 in animals recently imported from Argentina (59, 66). This virus was associated with a major transcontinental pandemic that reached Europe in early 1965 (50). The pandemic was distinguished in its early stages by its occurrence in horses of all ages rather than being restricted to younger horses, as had been the case in previous, presumably H7N7 influenza epidemics.

There was an international cocirculation of EIV H7N7 and H3N8 viruses during the 1960s and 1970s, although in the 1980s reports of equine influenza were increasingly largely restricted to H3N8. As EIV H7N7 has not been detected for more than two decades, it is now thought to be extinct (67, 68). In 1989 and 1990, two outbreaks of EIV with high morbidity (up to 80%) and mortality (up to 20%) affected large horse populations in northeast China. The virus that caused those outbreaks was of the H3N8 subtype, although it was phylogenetically related to avian rather than equine viruses (28).

The epidemiological patterns of clinical disease have tended to vary between countries, with many reports in the United Kingdom describing isolated outbreaks and occasional large epidemics, such as in 1979 and 1989 (9, 40). In other more isolated populations, such as the racehorse populations in Malaysia, South Africa, and Hong Kong or the competition horses of India and the Philippines and, more recently, in New South Wales and Queensland in Australia (<http://www.equineinfluenzainquiry.gov.au/>), there have been sporadic but violent disease outbreaks. These are associated with the movement of inadequately quarantined, often previously vaccinated, and subclinically infected horses into unvaccinated or inadequately vaccinated small populations with little or no herd immunity.

Importantly, equine influenza virus also has jumped the species barrier and emerged as a novel respiratory virus of dogs. Canine influenza virus (CIV) was established in the dog population in the early 2000s as a result of a complete transfer of EIV H3N8 (12). CIV was first detected in the United States (12) and subsequently in the United Kingdom and Australia (13, 37). Interestingly, the switch from horse to dog has clearly taken place more than once, as CIV sequences obtained from

* Corresponding author. Mailing address: Cambridge Infectious Diseases Consortium, Department of Veterinary Medicine, University of Cambridge, Madingley Road, CB3 0ES, Cambridge, England, United Kingdom. Phone: 44 (1223) 764963. Fax: 44 (1223) 764667. E-mail: prm42@cam.ac.uk.

† Supplemental material for this article may be found at <http://jvi.asm.org/>.

[∇] Published ahead of print on 23 March 2011.

infected dogs in Australia during a large outbreak in 2007 were identical to those of the source EIV strains circulating during the outbreak (37).

The first EIV vaccines were developed around 1965, although initial uptake was not high and the oil-adjuvanted products were associated with local adverse reactions. Vaccination policies vary among different countries, being mandatory in some types of horse in countries such as the United Kingdom. The focus of many field studies during the last 20 years has been the investigation of vaccine breakdown, which has been associated with both inadequately potent H3N8 vaccines (43, 46) and the antigenic drift of H3N8 viruses (14, 47). Indeed, most studies of EIV evolution have focused on the phylogenetic analysis of the hemagglutinin (HA) segment (7, 14, 15, 20, 34, 38, 39) and the resultant process of antigenic drift, which has been reported repeatedly for EIV H3N8 but less so for H7N7 subtype viruses. To date, only a few studies have considered the other EIV genomic segments, although occasionally they have been included in multispecies evolutionary studies of other influenza viruses (5, 25, 26, 57).

Although the antigenic drift of HA is critical for survival, viral fitness is undoubtedly a polygenic trait that requires an optimal interaction between the proteins encoded by all of the genomic segments. Hence, to truly understand the evolutionary patterns and dynamics of EIV, which may assist in the future control of this virus, it is essential to undertake a genome-scale analysis. We therefore performed a comprehensive phylogeny-based analysis using 82 publicly available complete genome sequences of EIV and sampled during a time period of 45 years. Our aim was to infer fundamental aspects of EIV evolution, including the occurrence and approximate dates of reassortment (both within subtypes and between H7N7 and H3N8), the rates and time scale of evolutionary change, the sign and strength of natural selection, and the population dynamics of EIV H3N8.

MATERIALS AND METHODS

Sequence data. We collected 82 complete genomes of EIV H3N8 from the NCBI Influenza Virus Resource (4) (see Table S1 in the supplemental material). Isolate NewYork/VR-297/1983 (accession numbers CY028916 to CY028923) was excluded from the analysis, as the date of isolation could not be confirmed by the submitters (S. Krauss, personal communication). We added H7N7 full-length sequences to this data set (8 sequences of PB2, 7 of PB1, 7 of PA, 11 of NP, 10 of MP, and 15 of NS1 [see Table S2 in the supplemental material]) to determine reassortment events between the H3N8 and H7N7 subtypes. In every case, only complete gene segment sequences were analyzed.

Sequence analysis. We used Se-AL (<http://tree.bio.ed.ac.uk/software/seal/>) to manually align the major coding sequences of each genomic segment (PB2, 2,277 nucleotides [nt]; PB1, 2,271 nt; PA, 2,148 nt; HA, 1,701 nt; NP, 1,494 nt; NA, 1,410 nt; MP, 756 nt; NS1, 657 nt). Because of the presence of overlapping reading frames, the regions encoding M2 and NS2 were excluded from the analysis.

We inferred individual maximum-likelihood (ML) trees for each gene segment using PAUP* (63) under the best-fit model of nucleotide substitution provided by MODELTEST (55) (parameter settings are available from the authors on request). Tree bisection-reconnection (TBR) branch swapping was applied in each case. We assessed the robustness of individual nodes by applying a bootstrap resampling process (1,000 replicates) using the neighbor-joining (NJ) method that included the ML substitution model. We also used PAUP* to determine the amino acid changes along the main trunk branches of each phylogeny, employing a parsimony reconstruction. Additionally, we inferred trees using the Bayesian method that is available in the MrBayes package (56), again applying the best-fit substitution model determined by MODELTEST.

Bayesian MCMC analysis. To estimate the rates of nucleotide substitution and the time of the most recent common ancestor (TMRCA) for each gene

segment, we used a Bayesian Markov chain Monte Carlo (MCMC) approach that is available in the BEAST package (16) (version 1.5.3; <http://beast.bio.ed.ac.uk/>). We used both strict and relaxed (uncorrelated lognormal) molecular clocks. For all segments, values lower than the 95% highest probability density (HPD) of the coefficient of the variation of the relaxed molecular clock were >0 , so this clock model was used in preference to a strict molecular clock model. For PB2, PB1, MP, and NS, we used the HKY85 model of nucleotide substitution, whereas for PA, HA, NP, and NA, a general reversible substitution (GTR) model was used, and in all cases separate substitution rates and nucleotide frequencies were estimated for each codon position. We also employed a Bayesian skyline coalescent prior in all cases, and chains were run until convergence was achieved. To estimate the changes in EIV relative genetic diversity ($N_e t$; N_e represents the effective population size and t the generation time) through time, we used the Bayesian skyline plot (17), which also is available in the BEAST package. Finally, we also used BEAST to infer maximum clade credibility (MCC) trees for each segment based on the posterior distribution of trees determined as described above. In this case, support for each node is reflected in values of the Bayesian posterior probability.

Estimates of selection pressures. We estimated the relative numbers of non-synonymous (d_N) and synonymous (d_S) nucleotide substitutions per site (d_N/d_S ratio) for the protein-coding region of each gene segment. To this end, we used the single-likelihood ancestor-counting (SLAC) method and the two-rate fixed-effects likelihood method on both the entire tree (FEL) and internal branches only (IFEL), all available via the Datamonkey web interface (52) of the HyPhy package (54). Estimates of d_N/d_S ratios were based on neighbor-joining trees estimated under the GTR model. We also tested for variation in selection pressure along lineages using the genetic algorithm (GABRANCH) method available in HyPhy (54). This assigns four classes of d_N/d_S ratios to each lineage to determine the best-fit model of lineage-specific evolution, and it calculates the probability ($\geq 95\%$) of d_N/d_S ratios being >1 along a specific lineage from the averaged model probability of all models (53).

RESULTS

Phylogenetic analysis of complete H3N8 genomes from 1963 to 2008. We analyzed 82 EIV H3N8 complete genomes sampled during the period 1963 to 2008 from Europe, Asia, and the Americas (see Table S1 in the supplemental material). We inferred individual phylogenetic trees for each of the eight genomic segments using both maximum-likelihood and Bayesian coalescent methods (i.e., MCC trees from BEAST). In general, all of the trees displayed a similar phylogenetic pattern typical of other mammalian influenza viruses, in which viruses from consecutive seasons were linked by a main trunk lineage with short side branches stemming from it. Interestingly, the MCC phylogenies, which assume a relaxed molecular clock such that a root is automatically specified, were consistently rooted on viruses isolated from South America in 1963, in agreement with previous suggestions that EIV H3N8 originated in this subcontinent (59) (trees are available from the authors on request).

For all gene segments, the main trunk of the trees first bifurcated when the so-called American and Eurasian lineages appeared in the late 1980s (14). These two lineages then co-circulated for a number of years, although after 1994 no other viruses from the European lineage were present in this data set. A second bifurcation event occurred in the early 2000s, when the so-called Florida clades 1 and 2 originated, and they continue to circulate today (48) (Fig. 1A to C; also see Fig. S1A to E in the supplemental material). The separation between these two clades is evident for all viruses isolated after 2005 in all phylogenies except MP. However, the boundaries of these two clades are less clear when looking at viruses from 2002 to 2003. For example, California/8560/02 and Kentucky/5/03 are in a postdivergence location for PB1, PA, NP, NA, and NS and

FIG. 1. (A) Maximum-likelihood (ML) phylogenetic tree for the HA segment of EIV H3N8. Colored boxes represent distinct clades of EIV that are numbered I to X in chronological fashion, apart from Florida clades 1 and 2 (FC1 and FC2). Bootstrap (>70%) values are shown in black. Numbers in black boxes along the main trunk branch indicate the number of amino acid changes. Individual isolates involved in reassortment events are indicated with colored asterisks (see the text). Internal branches with $d_N > d_S$ are shown in red. Horizontal branches are drawn to a scale of nucleotide substitutions per site, and the tree is rooted on the EIV Uruguay isolate from 1963. The same scale, color scheme, and rooting are used for each panel. (B) ML phylogeny of the NA gene of EIV H3N8. (C) ML phylogeny of the PA gene of EIV H3N8.

in a predivergence location for HA and PB2. Similarly, Newmarket/5/03 is in a postdivergence position for PB1, PA, HA, NP, NA, and NS but not for PB2. In turn, Wisconsin/1/03 is placed in a postdivergence position for PB2, PB1, HA, NP, and NA but not for PA and NS.

In general, viruses isolated from the same geographic region in the same year were closely related, although some exceptions were observed. For example, Kentucky/2/1980 and Kentucky/4/1980 were closely related for all gene segments with the exception of HA (Fig. 1A to C; also see Fig. S1A to E in the supplemental material), indicating that phylogenetically (and likely antigenically) distinct viruses were cocirculating within that state. Similarly, within Florida clade 2, all Chinese isolates group together in all phylogenies with the exception of PB2 and NA: for PB2, Xinjiang/3/2007 is separate from the rest of the Chinese isolates by a long branch, whereas for NA, Xinjiang/4/2007 is more closely related to viruses from America and the United Kingdom than to the other Chinese isolates,

again suggesting that distinct lineages with potentially different antigenicity were cocirculating within the same geographic location.

Besides the Florida clades 1 and 2, we identified 10 other distinct clades in six of the eight genomic segments (PB2, PB1, PA, HA, NP, and NA) (Fig. 1; also see Fig. S1 in the supplemental material). Each clade was comprised of a group of viruses that shared a common ancestor supported by high bootstrap and/or posterior probability values. Clades were numbered in a chronological fashion, with clade I being the oldest. Most clades contained contemporary viruses from a particular geographical region: clade III and X contained viruses from Europe, clade I and VIII were comprised of viruses from South America, clades IV, V, and IX contained viruses from Kentucky, and clades VI and VII displayed North American isolates from Georgia and California, respectively. The only exception to this geographical pattern was clade II, composed by viruses from Japan and Algiers (Sachiyama/1971 and

FIG. 1—Continued.

Algiers/1972), that were separated by long terminal branches but that shared a common ancestor.

Clades III and X contained European viruses isolated between 1978 and 1980 and between 1989 and 1994, respectively, and were consistent in all gene segments. In contrast, intrasubtype reassortment involving clades IV, V, and VI was reflected in minor topological differences between the phylogenies of each segment. For example, for PB2, clade V grouped with clade IV, while for PA it grouped with clade VI. With respect to specific viruses, Johannesburg/1/1986 also is likely to be a reassortant, as it grouped with Kentucky/1/1987 and Kentucky/2/1987 in the PB2 and PB1 phylogenies but not in the PA tree. Similarly, Italy/1062/1991 and Roma/5/1991 exhibited a clear phylogenetic divergence for the HA segment that contrasted with their close phylogenetic relationship in PB2, PB1, PA, NP, and NA (orange asterisks in Fig. 1; also see Fig. S1 in the supplemental material), again indicative of reassortment.

Amino acid differences along the main trunk branch of EIV H3N8 phylogenies. As an additional measure of evolutionary distance between the early and contemporary EIV H3N8 iso-

lates, we determined the number of amino acid changes along the main trunk of the phylogenies for each genomic segment. As expected, NA ($n = 50$) and HA ($n = 48$) exhibited the greatest number of amino acid changes, which is consistent with immune-driven antigenic drift (22, 24, 31, 45) (Table 1). Interestingly, we also observed a high number of amino acid changes in the main trunk of the PA phylogeny ($n = 45$), which is striking given that far fewer changes are seen in the other two segments that comprise the viral polymerase (PB1 and PB2). In some phylogenies we observed an unusually high number of amino acid substitutions along the main trunk branch connecting side branches. This was particularly evident between New York/1/1975 and Algiers/1/1972: the main trunk branch that separated them exhibited nine substitutions in PA and HA, seven in PB2, six in NA, and four in PB1, NP, and NS. Similarly, the main trunk branch that connected the Florida clade 2 (comprised of viruses isolated in China) with its more recent common ancestor exhibited seven substitutions in PA, six in PB1, and three in HA, NP, and MP. In addition, we detected eight amino acid changes within the main trunk branch of Florida clade 2 in the NA

FIG. 1—Continued.

phylogeny and two in PB2, which is consistent with *in situ* evolution in China.

As the mutations that occur on the main trunk branch are more likely to be selectively advantageous because they are at

TABLE 1. Number of amino acid substitutions along the main trunk branch for each gene segment of EIV H3N8

Segment	No. of amino acid changes	Site(s) associated with specific phenotypes ^a
PB2	20	<u>G65E</u> †, A588T†
PB1	18	M317I‡
PA	45	D55N†, V100A†, <u>R269K</u> †, <u>A337T</u> ‡, S409N†
HA	48	D172N*, A198E*, S137G*, V242I*, V196I*, T187S*, R140K*, P55S*, N172K*, T276I*, N189E*, E189K*, K189Q*
NP	20	<u>V312I</u> †, G16D†
NA	50	V213I‡, <u>N355T</u> ‡, L397P‡, <u>N396D</u> ‡
M	9	N/A
NS	21	S228P‡, <u>V194I</u> ‡, E227K†

^a Symbols: *, antigenic site; †, host marker site; ‡, virulence marker site. Amino acid positions located adjacent to sites of interest are underlined. N/A, not applicable.

a high frequency (even fixed) in the population, we next determined whether any of the trunk mutations were likely to be associated with immune evasion, host adaptation, and/or changes in virulence. Of the 48 amino acid changes that occurred across the HA phylogeny, 13 were located at putative (based on the human H3 HA structure) antigenic sites (of which two were positively selected; see below), which is consistent with antigenic drift (Table 1). With respect to the other genomic segments, we observed changes at or adjacent to sites associated with host specificity and virulence in all genomic segments but MP (Table 1) (2, 23, 41). Whereas PB2 and NP exhibited nonsynonymous mutations only at host marker sites, PB1 and NA displayed mutations only at sites associated with changes in virulence. In turn, PA and NS showed both types of mutations along the main trunk branch of their phylogenies (Table 1). Indeed, mutations D55N, V100A, and S409N in PA and G16D in NP not only occurred at host-differentiating sites but also resulted in the same amino acid changes suggested to confer avian-to-human adaptation (23). Similarly, mutation S228P in NS is associated with increased mammalian virulence (41).

FIG. 2. Maximum clade credibility (MCC) trees showing the contrast in phylogenetic relationships between EIV H3N8 and H7N7 among gene segments. H3N8 and H7N7 are shown as orange and blue circles, respectively. Posterior probability values are shown for the key nodes relating to the position of EIV H7N7 reassortant sequences. Branches exhibiting reassortant viruses are highlighted. The trees are automatically rooted through the assumption of a relaxed molecular clock, and tip times represent times (years) of sampling.

Reassortment between EIV H3N8 and H7N7. H7N7 viruses cocirculated with H3N8 viruses for at least 15 years (68) and have been isolated from horses sharing the same yard (65). Previous studies based on the analysis of partial nucleotide sequences have provided evidence of reassortment between these two subtypes (1, 5, 25, 33). To determine the time and extent of reassortment between EIV H3N8 and H7N7, we added all available H7N7 full-length sequences to our original data set (see Table S2 in the supplemental material) and inferred phylogenetic trees for each genomic segment. This analysis reveals that reassortment between H7N7 and H3N8 oc-

curred between 1964 and 1973 (95% HPD of the molecular clock MCC trees; see below) and resulted in viruses harboring an H7N7 HA, NA, and MP and the PB1, PB2, PA, and NP segments from EIV H3N8 (Fig. 2). Indeed, reassortment is likely to have occurred more than once, as shown by the lack of monophyletic origin for the reassortant viruses. Intriguingly, we did not detect any reassortment event that included the HA, NA (i.e., H7N8 or H3N7), or H7N7 viruses carrying an H3N8-derived M segment (not shown).

Evolutionary dynamics of EIV. We used a Bayesian MCMC approach (16) to estimate the rates of nucleotide substitution

TABLE 2. Estimates of nucleotide substitution rates and time of the most recent common ancestor for each gene segment of EIV H3N8

Segment ^a	Mean substitution rate ^a (95% HPD)	Mean TMRCA ^a (95%)	TMRCA for reassortment ^b (95% HPD)
PB2 ^a	1.31×10^{-3} (1.03×10^{-3} – 1.58×10^{-3})	1955 (1946–1961)	1965–1972
PB1 ^a	1.28×10^{-3} (1.05×10^{-3} – 1.53×10^{-3})	1954 (1944–1961)	1964–1972
PA ^b	1.39×10^{-3} (1.13×10^{-3} – 1.65×10^{-3})	1955 (1946–1962)	1964–1972
HA ^b	1.74×10^{-3} (1.39×10^{-3} – 2.08×10^{-3})	1952 (1943–1960)	N/A
NP ^b	1.29×10^{-3} (1.08×10^{-3} – 1.54×10^{-3})	1958 (1953–1962)	1970–1973
NA ^c	1.70×10^{-3} (1.37×10^{-3} – 2.03×10^{-3})	1955 (1948–1961)	N/A
MP ^a	1.01×10^{-3} (7.73×10^{-4} – 1.26×10^{-3})	1955 (1947–1961)	N/A
NS ^a	1.07×10^{-3} (8.03×10^{-4} – 1.36×10^{-3})	1954 (1944–1962)	1968–1973

^a Data represent 82 EIV H3N8 genomes (see Table S1 in the supplemental material).

^b Data represent 82 EIV H3N8 genomes and all available EIV H7N7 complete segments (see Table S2 in the supplemental material). N/A, not applicable.

in each gene segment (Table 2). Although the substitution rates estimated were within the expected range for RNA viruses (18), they were consistently lower than those observed for other mammalian influenza viruses (19, 60). The only exception was observed in the NA segment, whose evolutionary rate was similar to that reported for European H1N1 swine viruses (19). The evolutionary rates for HA and NA not only were the highest in our data set (which is consistent with immune-mediated selection) but also were significantly different from those of MP and NS, which were the lowest (Table 2).

To date the origins of EIV H3N8, we estimated the TMRCA for each genomic segment. The sampled genetic diversity among the eight genes that make up the EIV H3N8 genome most likely arose in the 1950s (95% between 1943 and 1962; Table 2), which is consistent with previous hypotheses of H3N8 viruses circulating in South America (59) some years before the 1963 outbreak. We also estimated that the time of reassortment between H7N7 and H3N8 and the common ancestor of the H3N8-H7N7 reassortant viruses occurred between 1964 and 1973 (Table 2).

To estimate the changes in EIV relative genetic diversity ($N_e t$) through time, we used the Bayesian skyline plot (17). If evolution is neutral, $N_e t$ is also a measure of the effective population size of the virus. All segments analyzed revealed a similar demographic pattern that is characterized by a constant level of genetic diversity during the 1960s and 1970s, followed by an abrupt drop in the early 1980s, in turn followed by an increase and stabilization from about 1990 onwards (Fig. 3). That this phylodynamic pattern is present in all segments and all isolates indicates that it affected the EIV population on a global scale.

Selection pressures on EIV genes. The mean d_N/d_S values for each genomic segment were low in all cases (Table 3), which is consistent with an evolutionary process shaped mainly by purifying selection. The highest d_N/d_S ratios were observed in HA, NA, and NS, which is indicative of localized adaptive evolution. Notably, of the 16 positively selected sites detected here, six were located in PA (again suggesting that this segment is evolving in a qualitatively different manner from that of the other internal gene segments), five in HA and one in each of the other genomic segments with the exception of MP (Table 3). In addition, we used the genetic branch (GA) algorithm (53) to determine whether there was branch-to-branch variation in selection pressure along the EIV phylogenies. Only three genomic segments (PB2, PA, and HA) exhibited

branches with high support (>95%) for $d_N > d_S$ (Fig. 1A and C; also see Fig. S1A in the supplemental material). However, the majority of the branches with high support for $d_N > d_S$ were located at the tips of the phylogenetic trees, which could be explained by the presence of transient deleterious mutations rather than clearly beneficial mutations. In contrast, the HA phylogeny displayed 13 internal branches with high support for $d_N > d_S$, of which five were located in the main trunk branch (Fig. 1A). We also detected positive selection in one segment of the main trunk branch of the PA phylogeny (Fig. 1C), supporting our previous results that this gene is characterized by a surprisingly high number of mutations on the main trunk lineage, which may reflect the occurrence of adaptive evolution.

DISCUSSION

The traditional approach to study the evolution of EIV H3N8 has been based on the analysis of HA1 phylogenies as a means to understand the mechanisms of antigenic drift and its practical implications in the process of vaccine strain selection (8, 14, 38, 39). However, understanding the molecular evolution and phylodynamics of EIV evidently requires a more comprehensive, genome-scale analysis.

We show here that phylogenetically (and likely antigenically) distinct clades have cocirculated globally, and that intrasubtype reassortment is relatively frequent in EIV H3N8. This is consistent with our prior finding of a horse simultaneously infected with two phylogenetically distinct viral lineages (Florida clades 1 and 2) during an outbreak in 2003 (44). As such, our study further emphasizes the importance of intrasubtype reassortment in influenza evolution, as is well documented in both H1N1 and H3N2 human influenza viruses (29, 45). Moreover, it has been hypothesized that intrasubtype reassortment can play a role in the genesis of viruses with unusual epidemiological characteristics, such as sudden changes in antigenicity and/or virulence (45). Indeed, one of the viral isolates that we identified as a reassortant, Johannesburg/1/1986, caused an unusually severe outbreak of disease (36). Although it has not been directly determined whether Johannesburg/1/1986 was more virulent than contemporary circulating strains, it has been suggested that the severity of this outbreak was due to the lack of immunity to EIV in the South African horse population (36).

In addition to intrasubtype reassortment, our phylogenetic

FIG. 3. Population dynamics of EIV. Bayesian skyline plots showing the changing levels of relative genetic diversity through time ($N_e t$) for all gene segments of EIV H3N8. Mean values of $N_e t$ are given by the boldface line, while the gray lines denote the 95% HPD values. Note the genome-wide drop in genetic diversity during the 1980s. The approximate dates on which vaccination was made compulsory for competition horses in some countries are indicated with a vertical arrow (see the text).

analysis also provided strong evidence for heterosubtypic reassortment between EIV H3N8 and EIV H7N7, confirming previous reports (1, 5, 33) and suggesting that these reassortment events occurred between 1964 and 1973. Interestingly, this reassortment resulted in EIV H7N7 viruses carrying H3N8 internal genes (with the exception of MP) but not in H3N8

viruses carrying H7N7 internal genes or in H3N7 or H7N8 reassortants. Notably, all available EIV H7N7 sequences from 1973 onwards are reassortants until 1977, the year in which this virus was detected for the last time. Although it is not clear why EIV H7N7 disappeared, it is feasible that it was outcompeted by H3N8. EIV H3N8 had internal genes clearly compatible

TABLE 3. Summary of selection pressures in each EIV H3N8 gene segment^a

Segment	Mean d_N/d_S ratio (SLAC)	No. (position) of positively selected codons ^a
PB2	0.13	1 (377)
PB1	0.09	1 (584)
PA	0.16	6 (57, 277, 337, 348, 354, 400)
HA	0.26	5 (10, 23, <u>152</u> , <u>206</u> , 345)
NP	0.13	1 (452)
NA	0.25	1 (43)
MP	0.14	0
NS	0.3	1 (67)

^a Positive selection was done by at least one of the methods SLAC, FEL, and IFEL. $P < 0.1$. Underlined codons are located in putative antigenic sites.

with H7N7 HA, NA, and MP, whereas our analysis suggests that heterosubtypic reassortants bearing H7N7 internal genes were less fit than other viruses circulating at the time. It is also interesting that H7N7 was not isolated after vaccination became widespread around 1980; although it is not reported in the literature, it was recognized from an early stage that it was far easier to induce protective antibody levels against H7N7 viruses than for H3N8, and there are no reports of vaccine breakdown for this former virus. Hence, the persistence of H7N7 likely required the international transmission of the virus into susceptible equine populations, and it may have been that even the low levels of vaccination coverage initially attained, concentrated as they were in international competition horses, could have had some role in the reduction of H7N7 transmission.

Another possible explanation for H3N8 outcompeting H7N7 was that the latter virus exhibited lower fitness with respect to infection and transmission among horses due to the avian-like nature of its genome: it has been shown that the HA of EIV H7N7 can reassort with avian influenza viruses, resulting in viruses that are pathogenic in chickens (3). In contrast, EIV H3N8 appears to be very well adapted to mammals, as it can infect not only horses but also dogs and pigs (12, 13, 64). An intriguing aspect of the H7N7/H3N8 reassortment is that the only internal H7N7 gene that was maintained was MP. We speculate that this is due to structural constraints, as the MP is located beneath the envelope and likely interacts with the cytoplasmic tails of the HA and NA glycoproteins. In addition, it has been shown that MP plays a critical role in the generation of viable reassortants between human (H3N2) and avian (H5N1) viruses (10), although this has not been assessed experimentally for EIV.

The reassortment between H3N8 and H7N7, together with the apparent susceptibility of horses to be infected by avian-like viruses such as Jilin/1989 (28), which caused two outbreaks in China—one in 1989 (with extremely high morbidity and mortality rates of 81 and 20%, respectively, in some herds) and another the following year—suggests that reassortment in horses is commonplace and further highlights the importance of whole-genome sequencing in epidemiological surveillance. Most equine influenza surveillance still focuses on the HA gene alone, as reflected in the large number of HA sequences compared to the relatively limited number of complete genomes available. Although superior to surveillance programs based on the enzyme-linked immunosorbent assay (ELISA)

detection of virus NP, serology, and clinical symptoms, sequencing HA in isolation clearly cannot detect reassortment events involving internal genes that may have a major bearing on viral fitness. We therefore urge that a more widespread program of the complete genome sequencing of EIV be established.

Another notable observation from this study was that the whole genome of EIV H3N8 evolves more slowly than that of other mammalian influenza viruses, which is consistent with other studies in which the evolutionary rates of individual gene segments were studied (6, 20, 25, 57). At present, the explanation for this reduced rate of nucleotide substitution is unclear. Although theoretically possible, it seems unlikely that the mutation rate of the EIV viral polymerase is either inherently lower than those of other mammalian influenza viruses or affected by host-associated factors. Similarly, that influenza viruses cause a distinct clinical syndrome in horses and are able to replicate to high titers (44) further suggests that the generation time of EIV is unlikely to be significantly lower than that of other mammalian influenza viruses. Consequently, the most likely explanation for the reduced rate of nucleotide substitution in EIV is that it is subject to weaker immune selection. However, it is striking that the mean d_N/d_S ratios for each EIV genomic segment are not significantly different from those estimated for swine-origin H1N1/09 and human H3N2 influenza A viruses (11, 60). As such, the reasons for the reduced rate of evolutionary change in EIV clearly merit further investigation.

One marked difference in the selective environment in EIV compared to that seen in other mammalian influenza viruses is that the PA segment exhibited an unusually high number of amino acid substitutions along the main trunk branch of its phylogeny, as well as the greatest number of positively selected codons. For example, the internal branch linking Newmarket/5/2003 with Florida clade 2, which spanned a similar time period, exhibited seven amino acid changes in the PA phylogeny (95% HPD between 1.4 and 3.4 years) and three in HA (95% HPD between 1.6 and 3.2 years). Although it has been shown that PA displays proteolytic activity (58) and that it contributes to the adaptation of low-pathogenic avian influenza viruses to mice (61), the exact function of this protein has not yet been elucidated (51). Of the 45 amino acid changes that we detected in the main trunk branch of the PA phylogeny, four were located at host marker sites (Table 1), suggesting that this gene has an important role in the adaptation of EIV H3N8 to the horse. Indeed, it is striking that substitutions D55N and V100A, which occurred early in the evolutionary history of EIV H3N8 (along the branch that links Sachiya/1/1971 and Algiers/1/1972 with New York/1/1975), have been identified as mutations that meet all standards of host differentiation and host-dependent conservation between avian and human influenza viruses (23). S409N is another mutation associated with host adaptation (23), but in this case it appears only in Florida clade 1.

The ultimate origins of EIVs are unknown. However, previous studies of the evolution of individual virus segments from multiple host species revealed no closely related lineages, and hence obvious progenitors, for the PB1, PA, HA, NP, and NA of EIV H3N8 (6, 25, 49, 57). In contrast, it has been suggested that the PB2, MP, and NS segments of EIV were more closely

related to those of avian influenza viruses (26, 32, 35). Although determining the origin of EIV H3N8 was beyond the scope of this work, we provide evidence that the most recent common ancestor of the first EIV H3N8 isolate (Miami/1/1963) likely arose in the 1950s (95% HPD between 1943 and 1962). This information could aid future sequencing of archived material.

A final, and highly intriguing, observation from our analysis was the marked drop in the relative genetic diversity of each genomic segment that coincides with a dramatic increase in vaccination rates after major epidemics took place in 1979 and 1981 in Europe and North America, respectively. For example, vaccination against EIV was made compulsory for racehorses in 1981 in the United Kingdom, France, and Ireland, and this intervention measure was extended to competition horses soon thereafter. We speculate that the increase in herd immunity, in particular in horses traveling internationally, caused by this vaccination program (coincident with a general increase in awareness of infectious diseases in the horse industry), and the concomitant reduction in the size of the susceptible host population resulted in a major population bottleneck. In addition, vaccine-induced selection pressure might have favored the fixation of advantageous mutations at key antigenic sites. For example, mutations S137G, V242I, V196I, and T187S were at antigenic sites and were fixed along the main trunk branch that links New York/1/1975 and clade VI (constituted by viruses isolated from Georgia in 1981; Table 1). The increase in relative genetic diversity observed after 1990 coincided with a major series of international epidemics associated with significant antigenic drift. Vaccine strains were not updated until the late 1990s with strains more recent than Fontainebleau/79 and Kentucky/1981.

In sum, we suggest that EIV H3N8 played an important role in the extinction of H7N7 in horses, and that specific intervention measures, such as widespread vaccination, have had a major effect on the global population genetic structure of EIV. Clearly, most of the available EIV sequences are likely to be derived from competition horses and from countries with active veterinary surveillance. However, as this population of horses exhibits a relatively long life span, with frequent global travel and mixing with horses from various geographical locations, we believe that equine influenza constitutes a powerful model system in the emerging field of viral phylogenetics (27, 30).

ACKNOWLEDGMENTS

We thank Mariana Varela and Simon Frost for helpful discussions. P.R.M. is supported by a Wellcome Trust Veterinary Postdoctoral Fellowship. J.L.N.W. is supported by the Alborada Trust and the RAPIDD program of the Science and Technology Directorate, Department of Homeland Security. E.C.H. is supported in part by grant R01 GM080533 from the National Institutes of Health.

REFERENCES

- Adeyefa, C. A., K. Quayle, and J. W. McCauley. 1994. A rapid method for the analysis of influenza virus genes: application to the reassortment of equine influenza virus genes. *Virus Res.* **32**:391–399.
- Allen, J. E., S. N. Gardner, E. A. Vitalis, and T. R. Slezak. 2009. Conserved amino acid markers from past influenza pandemic strains. *BMC Microbiol.* **9**:77.
- Banbura, M. W., Y. Kawaoka, T. L. Thomas, and R. G. Webster. 1991. Reassortants with equine 1 (H7N7) influenza virus hemagglutinin in an avian influenza virus genetic background are pathogenic in chickens. *Virology* **184**:469–471.
- Bao, Y., et al. 2008. The influenza virus resource at the National Center for Biotechnology Information. *J. Virol.* **82**:596–601.
- Bean, W. J. 1984. Correlation of influenza A virus nucleoprotein genes with host species. *Virology* **133**:438–442.
- Bean, W. J., et al. 1992. Evolution of the H3 influenza virus hemagglutinin from human and nonhuman hosts. *J. Virol.* **66**:1129–1138.
- Bryant, N. A., et al. 2009. Antigenic and genetic variations in European and North American equine influenza virus strains (H3N8) isolated from 2006 to 2007. *Vet. Microbiol.* **138**:41–52.
- Bryant, N. A., et al. 2011. Isolation and characterisation of equine influenza viruses (H3N8) from Europe and North America from 2008 to 2009. *Vet. Microbiol.* **147**:19–27.
- Burrows, R., D. Goodridge, M. Denyer, G. Hutchings, and C. J. Frank. 1982. Equine influenza infections in Great Britain, 1979. *Vet. Rec.* **110**:494–497.
- Chen, L. M., C. T. Davis, H. Zhou, N. J. Cox, and R. O. Donis. 2008. Genetic compatibility and virulence of reassortants derived from contemporary avian H5N1 and human H3N2 influenza A viruses. *PLoS Pathog.* **4**:e1000072.
- Chen, R., and E. C. Holmes. 2006. Avian influenza virus exhibits rapid evolutionary dynamics. *Mol. Biol. Evol.* **23**:2336–2341.
- Crawford, P. C., et al. 2005. Transmission of equine influenza virus to dogs. *Science* **310**:482–485.
- Daly, J. M., et al. 2008. Transmission of equine influenza virus to English foxhounds. *Emerg. Infect. Dis.* **14**:461–464.
- Daly, J. M., et al. 1996. Antigenic and genetic evolution of equine H3N8 influenza A viruses. *J. Gen. Virol.* **77**:661–671.
- Damiani, A. M., et al. 2008. Genetic characterization of equine influenza viruses isolated in Italy between 1999 and 2005. *Virus Res.* **131**:100–105.
- Drummond, A. J., and A. Rambaut. 2007. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol. Biol.* **7**:214.
- Drummond, A. J., A. Rambaut, B. Shapiro, and O. G. Pybus. 2005. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol. Biol. Evol.* **22**:1185–1192.
- Duffy, S., L. A. Shackelton, and E. C. Holmes. 2008. Rates of evolutionary change in viruses: patterns and determinants. *Nat. Rev. Genet.* **9**:267–276.
- Dunham, E. J., et al. 2009. Different evolutionary trajectories of European avian-like and classical swine H1N1 influenza A viruses. *J. Virol.* **83**:5485–5494.
- Endo, A., R. Pecoraro, S. Sugita, and K. Nerome. 1992. Evolutionary pattern of the H3 haemagglutinin of equine influenza viruses: multiple evolutionary lineages and frozen replication. *Arch. Virol.* **123**:73–87.
- Espmark, A., L. Heller, and P. Viriden. 1956. Immunological relationship between infectious cough in horses and human influenza A. *Arch. Gesamte Virusforsch.* **7**:120–124.
- Ferguson, N. M., A. P. Galvani, and R. M. Bush. 2003. Ecological and immunological determinants of influenza evolution. *Nature* **422**:428–433.
- Finkelstein, D. B., et al. 2007. Persistent host markers in pandemic and H5N1 influenza viruses. *J. Virol.* **81**:10292–10299.
- Fitch, W. M., J. M. Leiter, X. Q. Li, and P. Palese. 1991. Positive Darwinian evolution in human influenza A viruses. *Proc. Natl. Acad. Sci. U.S.A.* **88**:4270–4274.
- Gorman, O. T., W. J. Bean, Y. Kawaoka, and R. G. Webster. 1990. Evolution of the nucleoprotein gene of influenza A virus. *J. Virol.* **64**:1487–1497.
- Gorman, O. T., R. O. Donis, Y. Kawaoka, and R. G. Webster. 1990. Evolution of influenza A virus PB2 genes: implications for evolution of the ribonucleoprotein complex and origin of human influenza A virus. *J. Virol.* **64**:4893–4902.
- Grenfell, B. T., et al. 2004. Unifying the epidemiological and evolutionary dynamics of pathogens. *Science* **303**:327–332.
- Guo, Y., et al. 1992. Characterization of a new avian-like influenza A virus from horses in China. *Virology* **188**:245–255.
- Holmes, E. C., et al. 2005. Whole-genome analysis of human influenza A virus reveals multiple persistent lineages and reassortment among recent H3N2 viruses. *PLoS Biol.* **3**:e300.
- Holmes, E. C., and B. T. Grenfell. 2009. Discovering the phylodynamics of RNA viruses. *PLoS Comput. Biol.* **5**:e1000505.
- Ina, Y., and T. Gojobori. 1994. Statistical analysis of nucleotide sequences of the hemagglutinin gene of human influenza A viruses. *Proc. Natl. Acad. Sci. U.S.A.* **91**:8388–8392.
- Ito, T., O. T. Gorman, Y. Kawaoka, W. J. Bean, and R. G. Webster. 1991. Evolutionary analysis of the influenza A virus M gene with comparison of the M1 and M2 proteins. *J. Virol.* **65**:5491–5498.
- Ito, T., et al. 1999. Replacement of internal protein genes, with the exception of the matrix, in equine 1 viruses by equine 2 influenza virus genes during evolution in nature. *J. Vet. Med. Sci.* **61**:987–989.
- Kawaoka, Y., W. J. Bean, and R. G. Webster. 1989. Evolution of the hemagglutinin of equine H3 influenza viruses. *Virology* **169**:283–292.
- Kawaoka, Y., et al. 1998. Influence of host species on the evolution of the nonstructural (NS) gene of influenza A viruses. *Virus Res.* **55**:143–156.
- Kawaoka, Y., and R. G. Webster. 1989. Origin of the hemagglutinin on A/Equine/Johannesburg/86 (H3N8): the first known equine influenza outbreak in South Africa. *Arch. Virol.* **106**:159–164.
- Kirkland, P. D., D. S. Finlaison, E. Crispe, and A. C. Hurt. 2010. Influenza

- virus transmission from horses to dogs, Australia. *Emerg. Infect. Dis.* **16**:699–702.
38. **Lai, A. C., et al.** 2001. Diverged evolution of recent equine-2 influenza (H3N8) viruses in the Western Hemisphere. *Arch. Virol.* **146**:1063–1074.
 39. **Lai, A. C., K. M. Rogers, A. Glaser, L. Tudor, and T. Chambers.** 2004. Alternate circulation of recent equine-2 influenza viruses (H3N8) from two distinct lineages in the United States. *Virus Res.* **100**:159–164.
 40. **Livesay, G. J., T. O'Neill, D. Hannant, M. P. Yadav, and J. A. Mumford.** 1993. The outbreak of equine influenza (H3N8) in the United Kingdom in 1989: diagnostic use of an antigen capture ELISA. *Vet. Rec.* **133**:515–519.
 41. **Lycett, S. J., et al.** 2009. Detection of mammalian virulence determinants in highly pathogenic avian influenza H5N1 viruses: multivariate analysis of published data. *J. Virol.* **83**:9901–9910.
 42. **Morens, D. M., and J. K. Taubenberger.** 2010. Historical thoughts on influenza viral ecosystems, or behold a pale horse, dead dogs, failing fowl, and sick swine. *In* A. W. Hampson (ed.), *Influenza and other respiratory viruses*. Wiley, Hoboken, NJ.
 43. **Mumford, J. A.** 1999. The equine influenza surveillance program. *Adv. Vet. Med.* **41**:379–387.
 44. **Murcia, P. R., et al.** 2010. Intra- and interhost evolutionary dynamics of equine influenza virus. *J. Virol.* **84**:6943–6954.
 45. **Nelson, M. I., et al.** 2008. Multiple reassortment events in the evolutionary history of H1N1 influenza A virus since 1918. *PLoS Pathog.* **4**:e1000012.
 46. **Newton, J. R., et al.** 2000. Immunity to equine influenza: relationship of vaccine-induced antibody in young Thoroughbred racehorses to protection against field infection with influenza A/equine-2 viruses (H3N8). *Equine Vet. J.* **32**:65–74.
 47. **Newton, J. R., K. Verheyen, J. L. Wood, P. J. Yates, and J. A. Mumford.** 1999. Equine influenza in the United Kingdom in 1998. *Vet. Rec.* **145**:449–452.
 48. **Office International des Epizooties.** 2008. Expert surveillance panel on equine influenza vaccines. Conclusions and recommendations. Office International des Epizooties, Paris, France.
 49. **Okazaki, K., Y. Kawaoka, and R. G. Webster.** 1989. Evolutionary pathways of the PA genes of influenza A viruses. *Virology* **172**:601–608.
 50. **Paccaud, M. F., M. Couard, F. Bürki, H. Gerbber, and J. Löhrer.** 1966. Outbreak of influenza A/equi-2 in Switzerland. *Nature* **211**:101–102.
 51. **Palese, P., and M. L. Shaw.** 2007. Orthomyxoviridae: The viruses and their replication, p. 1647–1690. *In* D. M. Knipe and P. M. Howley (ed.), *Fields virology*, 5th ed., vol. 1. Lippincott Williams & Wilkins, Philadelphia, PA.
 52. **Pond, S. L., and S. D. Frost.** 2005. Datamonkey: rapid detection of selective pressure on individual sites of codon alignments. *Bioinformatics* **21**:2531–2533.
 53. **Pond, S. L., and S. D. Frost.** 2005. A genetic algorithm approach to detecting lineage-specific variation in selection pressure. *Mol. Biol. Evol.* **22**:478–485.
 54. **Pond, S. L., S. D. Frost, and S. V. Muse.** 2005. HyPhy: hypothesis testing using phylogenies. *Bioinformatics* **21**:676–679.
 55. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA substitution. *Bioinformatics* **14**:817–818.
 56. **Ronquist, F., and J. P. Huelsenbeck.** 2003. MrBayes 3: Bayesian phylogenetic inference under mixed models. *Bioinformatics* **19**:1572–1574.
 57. **Saito, T., Y. Kawaoka, and R. G. Webster.** 1993. Phylogenetic analysis of the N8 neuraminidase gene of influenza A viruses. *Virology* **193**:868–876.
 58. **Sanz-Ezquerro, J. J., T. Zurcher, S. de la Luna, J. Ortin, and A. Nieto.** 1996. The amino-terminal one-third of the influenza virus PA protein is responsible for the induction of proteolysis. *J. Virol.* **70**:1905–1911.
 59. **Scholtens, R. G., J. H. Steele, W. R. Dowdle, W. B. Yarbrough, and R. Q. Robinson.** 1964. U.S. epizootic of equine influenza, 1963. *Public Health Rep.* **79**:393–402.
 60. **Smith, G. J., et al.** 2009. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature* **459**:1122–1125.
 61. **Song, M. S., et al.** 2009. The polymerase acidic protein gene of influenza A virus contributes to pathogenicity in a mouse model. *J. Virol.* **83**:12325–12335.
 62. **Sovinova, O., B. Tumova, F. Pouska, and J. Nemec.** 1958. Isolation of a virus causing respiratory disease in horses. *Acta Virol.* **2**:52–61.
 63. **Swofford, D. L.** 2002. PAUP*. Phylogenetic analysis using parsimony (*and other methods). Version 4. Sinauer Associates, Sunderland, MA.
 64. **Tu, J., et al.** 2009. Isolation and molecular characterization of equine H3N8 influenza viruses from pigs in China. *Arch. Virol.* **154**:887–890.
 65. **Tumova, B., B. C. Easterday, and A. Stumpa.** 1972. Simultaneous occurrence of A-equi-1 and A-equi-2 influenza viruses in a small group of horses. *Am. J. Epidemiol.* **95**:80–87.
 66. **Waddell, G. H., M. B. Teigland, and M. M. Sigel.** 1963. A new influenza virus associated with equine respiratory disease. *J. Am. Vet. Med. Assoc.* **143**:587–590.
 67. **Webster, R. G.** 1993. Are equine 1 influenza viruses still present in horses? *Equine Vet. J.* **25**:537–538.
 68. **Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka.** 1992. Evolution and ecology of influenza A viruses. *Microbiol. Rev.* **56**:152–179.