JVI Accepts, published online ahead of print on 5 May 2010
J. Virol. doi:10.1128/JVI.00112-10
Copyright © 2010, American Society for Microbiology and/or the Listed Authors/Institutions. All Rights Reserved.

JVI Accepts published online ahead of print

1 **Title page**

2 ***Title:*** The intra- and inter-host evolutionary dynamics of equine influenza virus

3 ***Running title:*** EIV evolutionary dynamics

4 ***Authors:*** Pablo R Murcia[a¶], Gregory J Baillie[a,b¶], Janet Daly[c†], Debra Elton[c], Carley
5 Jervis[c], Jennifer A. Mumford[a], Richard Newton[c], Colin R. Parrish[d], Karin Hoelzer[d],
6 Gordon Dougan[b], Julian Parkhill[b], Nicola Lennard[b], Doug Ormond[b], Sharon Moule[b],
7 Andrew Whitwham[b], John W. McCauley[e], Trevelyan J McKinley[a], Edward C. Holmes[f,g],
8 Bryan T Grenfell[f,g‡], and James L N Wood[a*].

9 ***Authors affiliations:***

10 [a] Cambridge Infectious Diseases Consortium, Department of Veterinary Medicine,
11 University of Cambridge, United Kingdom.

12 [b] Wellcome Trust Sanger Institute. Wellcome Trust Genome Campus, Hinxton,
13 Cambridge, United Kingdom.

14 [c] Animal Health Trust, Centre for Preventive Medicine, Lanwades Park, Newmarket,
15 United Kingdom.

16 [d] Baker Institute of Animal Health, Department of Microbiology and Immunology,
17 College of Veterinary Medicine, Cornell University, Ithaca, USA.

18 [e] Division of Virology, MRC National Institute for Medical Research, Mill Hill, London,
19 United Kingdom.

20 [f] Center for Infectious Disease Dynamics, Pennsylvania State University, University
21 Park, Pennsylvania, USA.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

22 [g] Fogarty International Center, National Institutes of Health, Bethesda, USA.

23 [†] Present address: School of Veterinary Medicine and Science, University of
24 Nottingham, Sutton Bonington, United Kingdom.

25 [‡] Present address: Department of Ecology and Evolutionary Biology, Princeton
26 University, Princeton, New Jersey, U.S.A.

27 [*]***Corresponding author:*** James L N Wood. Mailing address: Cambridge Infectious
28 Diseases Consortium, Department of Veterinary Medicine, University of Cambridge.
29 Madingley Road, CB3 0ES, Cambridge, England, United Kingdom. E-mail address:
30 jlnw2@cam.ac.uk. Tel: +44 (1223) 764666. Fax: +44 (1223) 764667.

31 ¶ These authors contributed equally to this work.

32 **Word count:** Abstract = 245 words; Main Text = 5376 words.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

33 **Abstract**

34 Determining the evolutionary basis of cross-species transmission and immune evasion
35 is key to understanding the mechanisms that control the emergence of either new
36 viruses or novel antigenic variants with pandemic potential. The hemagglutinin
37 glycoprotein of influenza A viruses is a critical host range determinant and a major
38 target of neutralizing antibodies. Equine influenza virus is a significant pathogen of the
39 horse that causes periodical outbreaks of disease even in populations with high
40 vaccination coverage. EIV has also jumped the species barrier and emerged as a novel
41 respiratory pathogen in dogs - canine influenza virus. We studied the dynamics of
42 equine influenza virus evolution in horses at the intra-host level, and how this
43 evolutionary process is affected by inter-host transmission in a natural setting. To this
44 end, we performed clonal sequencing of the hemagglutinin 1 gene derived from
45 individual animals at different times post-infection. Our results show that despite the
46 population consensus sequence remaining invariant, genetically distinct subpopulations
47 persist during the course of infection and are also transmitted, with some variants likely
48 to change antigenicity. We also detected a natural case of mixed infection in an animal
49 infected during an outbreak of equine influenza, raising the possibility of reassortment
50 between different strains of virus. In sum, our data suggest that transmission
51 bottlenecks may not be as narrow as originally perceived and that the genetic diversity
52 required to adapt to new host species may be partially present in the donor host and
53 potentially transmitted to the recipient host.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

54 **Introduction**

55 Influenza viruses inflict a persistent annual burden on human health, with the occasional
56 appearance of strains with pandemic potential. A major research focus has been to
57 determine the evolution of influenza A virus at the epidemic scale and across broad
58 geographical areas (15, 22, 28, 30). However, little is known about the original source
59 of viral genetic variation in natural animal hosts, or how this variation impacts on the
60 ability of the virus to adapt to new host species. RNA viruses are characterized by their
61 ability to rapidly generate genetic variation, a consequence of their replication with
62 highly error prone RNA polymerases (8). Such mutational power, coupled with
63 immense population sizes, is thought to be central to their classification as the most
64 common agents of emerging disease (44). While the processes of mutation, segment
65 reassortment and natural selection are expected to shape the evolutionary dynamics of
66 influenza viruses within an individual host, the nature, frequency and interaction of these
67 key evolutionary mechanisms have not been studied systematically *in vivo* in natural
68 hosts, nor integrated into our understanding of viral evolution at the scale of global
69 epidemics. Such data are also essential for addressing one of the most important
70 questions in the cross-species transmission of pathogens: whether the genetic variation
71 required for a virus to adapt to transmission in a new host species largely appears *de*
72 *novo* in the recipient host or is seeded from the donor host (20). Similarly, the ability of
73 natural selection to optimize such traits as host specificity is also in part dependent on
74 the proportion of total genetic variability that is passed between hosts at transmission:
75 the narrower the population bottleneck at transmission, the larger the stochastic
76 component to viral evolution.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

77 Influenza A viruses (IAVs) have their major reservoirs in wild aquatic birds (41), but
78 periodically transfer into humans and other mammals to cause epidemics or pandemics.
79 Contemporary human influenza A viruses appear to have originated from a transfer that
80 occurred shortly before the 1918 pandemic (39), and that virus subsequently exchanged
81 three and then two of its genome segments with avian viruses in 1957 and 1968 to
82 create new pandemic strains (31, 33). Phylogenetic analysis of recently emerged
83 swine-origin H1N1 in humans suggests that this new pandemic virus is a reassortant
84 between Eurasian and classic swine variants (11). While reassortment involves the
85 transfer of entire genome segments, the more gradual process of mutation
86 accumulation is required to facilitate both host adaptation and escape from immune
87 responses, and underlies the predictable seasonality of influenza in temperate regions
88 (28, 30).

89 To determine the patterns and consequences of genetic variation in mammalian
90 influenza viruses, we used equine influenza H3N8 virus (EIV) as an experimental model
91 system in horses. EIV of the H3N8 subtype was first isolated from horses in 1963 (40)
92 and is apparently the only subtype currently circulating in the horse population. EIV has
93 also recently jumped the species barrier and become established as a respiratory
94 pathogen of the dog (5), so we are able to compare the variation of the virus in its
95 established equine host, as well as in dogs which it initially infected at some point
96 before 2004 (14). We undertook transmission experiments in naïve horses and
97 examined the degree and composition of the within-host EIV genetic diversity of the
98 hemagglutinin 1 (HA1) at different times post-infection. HA1 encodes the major surface
99 viral antigen and the receptor binding domain (42). We focused on the proportion of

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

100 viral variability that persists during infection as well as that transmitted during natural
101 chains of transmission, with the latter informing on the magnitude of any transmission
102 bottleneck. To compare the variation observed within our experimental setting to that
103 observed in the field, we also examined the variation in HA1 sequences of viruses
104 recovered from natural cases of H3N8 influenza during an outbreak in the UK in 2003
105 (23). This allows us to link, for the first time, the process of viral evolution within and
106 among hosts.

107 **Materials and Methods**

108 **Direct transmission experiments in naïve horses.** Two "transmitter" horses (7D36
109 and 0443) were nebulised with 20 ml of $\log_{10} 10^{6.3}$/ml EID_{50} of Eq/Newmarket/1/93 (21).
110 On confirmation of infection using a rapid diagnostic ELISA to detect influenza virus in
111 swab extracts (4), one transmitter (7D36) was housed with two naïve horses (7248 and
112 6005) in the same stable. When horses 7248 and 6005 became infected, they were
113 removed to clean separated stables and each was co-housed with two other naïve
114 horses (5447, 7C1C, 5257 and 282E) for another 72 hours, when the procedure was
115 repeated and horses 7C1C and 5257 were each co-housed as described above with
116 two further horses (2F50, 7A45, 780C, and 5D1A) in individual clean stables (Fig. 1).
117 Nasal swabs were collected for 2-6 days after infection or contact, immersed in viral
118 transport medium (5ml), and stored at -80°C. All horses included in this experiment
119 were considered naïve as they had no detectable antibodies against
120 Eq/Newmarket/1/93 (measured by single radial haemolysis (43)). The animal work was
121 done under Home Office license following full ethical approval.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

122 **RNA extraction, real-time PCR and PCR analysis.** Viral RNA from nasal swabs was
123 isolated from 280 µl-aliquots using the QIAamp viral RNA mini kit (Qiagen) according to
124 the manufacturer's instructions. Reverse transcription, PCR and real-time PCR (qPCR)
125 amplification were performed using a two-step reverse transcription-PCR protocol.
126 cDNAs of the viral genomic M and HA genes were generated using Superscript III
127 reverse transcriptase (Invitrogen) and primers Bm-M-1 and Bm-HA1 (14), respectively.
128 RT was performed at 55°C for 90 min, followed by incubation at 70°C for 10 min. Viral
129 copy numbers were estimated by a qPCR assay using the QuantiTect Probe PCR Kit
130 (Qiagen) with fluorogenic hydrolysis type probes following the manufacturer's
131 instructions. Primers and probe for qPCR were designed using Beacon designer
132 (Premier Biosoft, sequences available upon request). Standard curves were generated
133 using 10-fold dilutions of a plasmid containing the matrix segment (cloned from an egg-
134 grown Equine/Newmarket/1/1993 isolate), ranging from $1x10^2$ to $1x10^8$ copies $µl^{-1}$. For
135 each run, all samples, no template controls, plasmid standards, positive and negative
136 controls were run in triplicate and expressed as mean number of vRNA copies *per* µl of
137 cDNA. PCR amplification was performed using Platinum *Pfx* DNA polymerase
138 (Invitrogen) using primers Bm-HA1 and EHA1007rw
139 (5'TTGGGGCATTTTCCATATGT3'), spanning the region between nt -43 (upstream of
140 the HA1 start codon) and nt 1007. PCR amplification was performed for 40 cycles (94°C
141 for 30 sec, 55°C for 1 min and 68°C for 1 min), followed by a final extension at 68°C for
142 10 min. PCR products were gel-purified using the QIAquick Gel Extraction Kit (Qiagen)
143 and further cloned using the Zero-Blunt TOPO PCR Cloning kit for sequencing

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

144 (Invitrogen) following the manufacturer's instructions. Clones were sequenced at the
145 Influenza Sequencing Pipeline established at the Wellcome Trust Sanger Institute.

146 **Deep-amplicon sequencing of HA1 from nasal swabs and controls.** The
147 sequencing was performed using fluorescent sequencing chemistry and ABI 3730xl
148 capillary sequencers. Forward and reverse sequencing reads from each clone were
149 trimmed of vector sequence and poor quality regions, and assessed for quality; reads
150 with an average Phred quality score (9, 10) <20 were rejected. Overlapping forward
151 and reverse reads were merged into a single contig with quality scores. Only contigs
152 >900 bp in length were used for subsequent analyses. Contigs were aligned against
153 the HA1 consensus sequence of an egg grown isolate of A/Equine/Newmarket/1/1993
154 using Ssaha2 (24), and high-quality variants were identified using the SNP_analysis.pm
155 Perl module (http://sourceforge.net/projects/snpanalysis/), and our own Perl scripts.
156 Nucleotide variants were considered real if their Phred score was >25. Nucleotides with
157 a Phred score below that value were considered identical to the consensus nucleotide.
158 Amino acid variants were considered real if all of the nucleotides in the codon had
159 Phred scores >25. Sequences containing high-quality insertions or deletions (indels)
160 that altered the reading frame were counted and excluded from subsequent analyses.

161 **Analysis of assay-related mutations.** To measure errors associated with the
162 replication of plasmid DNA in bacteria and capillary sequencing, we sequenced 754
163 clones from a single plasmid clone. To assess for DNA polymerase errors during the
164 PCR step, we amplified the HA1 segment from a single plasmid and cloned it into
165 pCR®4Blunt-TOPO as described above. This PCR product was cloned and sequenced

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

166 as described above. To estimate mutations introduced during cDNA preparation, we
167 performed *in vitro* transcription from a single plasmid clone template, using the
168 Riboprobe® *in vitro* transcription systems (Promega) following the manufacturers'
169 instructions. *In vitro* transcribed RNA was subject to RT-PCR, cloned and sequenced as
170 described above.

171 **Evolutionary analysis.** A total of 2366 intra-host EIV sequences isolated from
172 experimentally infected animals (EMBL-Bank accession numbers FN398346 to
173 FN400711), 372 intra-host EIV sequences isolated from natural cases (EMBL-Bank
174 accession numbers FN422006 to FN422377), and 158 epidemiological-scale
175 sequences (Supporting Table 1, obtained from the Influenza Virus Resource
176 [http://www.ncbi.nlm.nih.gov/genomes/FLU/FLU.html]) extending from the start codon of
177 the HA open reading frame to nucleotide position 903 were collated. Sequence
178 alignments were generated from the Ssaha2 output for the intra-host EIV sequences,
179 and using Se-Al (http://tree.bio.ed.ac.uk/software/seal/) for the EIV population
180 sequences. Because of the very small genetic distances involved, the mean pairwise
181 genetic diversity within each sample was calculated from the uncorrected pairwise
182 distance matrix (*p*-distance) between taxa (available from the authors upon request).
183 Maximum likelihood (ML) trees were estimated using the PAUP* 4.0b10 package (38)
184 under the best-fit model of nucleotide substitution determined using Modeltest (27). For
185 those data sets representing individual horses expansive TBR branch-swapping was
186 used in all cases. However, for the extremely large data sets combining horses, simpler
187 NNI branch-swapping was utilized to provide computational tractability. Mean numbers
188 of nonsynonymous (d_N) and synonymous (d_S) substitutions per site (ratio d_N/d_S) were

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

189 estimated using the Single Likelihood Ancestor Counting (SLAC) algorithm available in
190 the Datamonkey web interface of the HyPhy software package (26). Minimum spanning
191 trees allowing hypothetical intermediate nodes were calculated from the sequence data
192 using Prim's algorithm in BIONUMERICS V4.5 (Applied Maths, Belgium) as described
193 previously (29).

JVI Accepts published online ahead of print

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

194 **Results**

195 **Transmission of equine influenza virus in naïve horses**

196 Our experimental system was designed to emulate natural transmission events of a
197 virus in its natural host and, with two separate chains of transmission initiated from the
198 first infected horse, internal comparison of any variation observed was also possible.
199 Two naïve horses were experimentally infected with equine influenza virus
200 Equine/Newmarket/1/1993 (Fig. 1A), and then one of those was housed with two further
201 naïve horses, which were then separated and used to initiate two transmission chains
202 (Fig. 1A). The aim of this first step was to re-adapt the egg-grown inoculums to the
203 horse. Of the ten exposed animals, nine became infected indicating that transmission
204 under these conditions was efficient. Viral shedding was determined from copies of
205 viral RNA in nasal swabs collected daily. The peak of excretion – in which millions of
206 viral particles were detected in a single nasal swab – was usually observed around 72
207 hours post-contact (hpc) (Fig. 1B). Thus, the sizes of the within-host viral populations
208 are extremely large, particularly since nasal swabs only capture a very small proportion
209 of the total population at a particular time-point. In one exposure only one of two
210 recipients (horse 5257) became infected despite the fact that the donor displayed high
211 levels of viral shedding (Fig. 1B). Horse 5257 subsequently exhibited lower levels of
212 viral shedding (Fig. 1B) but it transmitted the virus to two contacts which appeared to
213 shed less virus and showed a marked delay to reach the peak of viral shedding.

214 **Intra-host genetic variation**

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

215 To determine the extent and structure of viral genetic diversity within infected horses,
216 we examined the variation among multiple clones prepared from the RNA sequences of
217 the first 903 nucleotides of the HA1 gene. Because nasal swabs displayed variable
218 amounts of virus, we restricted our analysis to those days in which the viral populations
219 were large enough to allow PCR amplification and further sequencing. Between 44 and
220 154 clones were prepared and sequenced from the virus collected in nasal swabs from
221 each horse on each day examined, resulting in a total of 2,366 individual HA1
222 sequences. Table 1 summarizes the levels of sequence variability present in samples
223 examined, as well as the relative numbers of synonymous and nonsynonymous
224 substitutions per site as a measure of selection pressure.

225 Overall, we observed a total of 392 mutations, which results in a mutation frequency of
226 1.8×10^{-4} mutations per nucleotide site. Each sample (i.e. nasal swab) contained a
227 mixture of closely related genomes, with most clones being identical to each other, and
228 a proportion (14.2%) containing 1 or sometimes 2 mutations (Fig. 1C). Only a few
229 clones contained three, four, and six nucleotide mutations. Mutations were present at
230 246 nucleotide positions along the HA gene coding region examined. In most samples,
231 the number of nonsynonymous mutations was greater than the number of synonymous
232 mutations (Fig. 1D). Many clones shared a single mutation either in the same or in a
233 different sample, and only two clones shared two mutations (C476T, Ala144Val and
234 A884G, Glu280Gly), although these were derived from the same swab (animal 6005,
235 day 6 post-contact).

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

236 We observed a gradual increase in the complexity of genetic diversity over the course of
237 infection in individual animals, manifest as an increase in the number of synonymous
238 and nonsynonymous mutations, the number of clones possessing single and multiple
239 mutations, and the average pairwise divergence within samples (Table 1). Of note, the
240 average pairwise distance is always significantly higher at the end of the infection period
241 than at the beginning ($p<0.001$, after correcting for multiple tests using a Bonferroni
242 correction) in individual animals for which we had 3 or more samples. Some mutations
243 were observed at the same site on multiple days from the same horse and some of
244 these were also observed in other horses (Tables 2 and 3). Of the 392 detected
245 mutations, 144 (distributed in 43 nucleotide sites) were observed in more than one of
246 the 30 independent samples.

247 To estimate the selection pressures affecting EIV within its natural host, we calculated
248 the mean value of d_N/d_S for each sample (Table 1). Although individual values were
249 variable, mean d_N/d_S for the data as a whole was 0.80, close to the expected value of
250 ~1 expected under an entirely neutral evolutionary process. This value (and the
251 majority of the individual values) was higher than the 0.41 value estimated for
252 epidemiological populations of EIV H3N8, suggesting that different selection pressures
253 act at the individual than at the population level, with purifying selection dominating the
254 latter. However, since our intra-host data also contains some artifactual mutations that
255 are invariably introduced due to the experimental procedure undertaken (see
256 Supplementary Information), all conclusions on the overall extent and structure of
257 genetic diversity need to be drawn with caution. When we compared the sites and
258 mutations that arose in the experimental infections, most nonsynonymous mutations

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

259 were observed as single changes in one sequence (singletons) in sites that were not
260 polymorphic in the public database of equine H3N8 sequences (epidemiological scale
261 variation), suggesting that they may have been introduced either *in vivo*, and hence are
262 likely to represent transient deleterious mutations, or during the preparation of the
263 samples. This notion is also supported by the presence of sequences containing
264 mutations that either generated premature stop codons or frame-shifts (Table 1).

265 **Patterns of intra- and inter-host viral evolution in infected horses**

266 We investigated viral variation during infection of animals at different stages in the
267 transmission chain. In all cases the majority of sequences recovered were identical to
268 the consensus. In 21 of 30 data sets from individual swabs, predominantly single (or
269 occasionally double) nucleotide changes were seen from the consensus sequence.
270 Importantly, in nine of the sets of sequences, some of the variation was seen to give
271 rise to small phylogenetic clusters that were distinct from the consensus sequence.
272 Hence, during infection of individual hosts, the viral populations consist of very closely
273 related genomes, generated by an ongoing process of mutation, some of which then
274 acquire additional changes, producing phylogenetic structure. Such phylogenetic
275 structure is strongly suggestive of natural, rather than artifactual, variation.

276 To consider the dynamic properties of the within-host influenza populations, we
277 examined sequence variation from animals where more than one virus positive daily
278 nasal swab was available. Again, most sequences detected were identical to the
279 consensus, and variants generated through the course of infection mostly consisted of
280 singletons. For example, for horse 5D1A (for which we had three consecutive swabs

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

281 and 214 sequences), we detected 24 clones bearing a single mutation when compared
282 to the consensus sequence and three sequences displaying two mutations. Of this
283 entire mutational spectrum, only one mutation (A413G, Glu123Gly) was detected on two
284 different days, suggesting that it was either generated *de novo* twice, or more likely,
285 persisted through two days of the infection period (Fig. 2A). Maximum likelihood
286 phylogenetic trees for all the animals included in the transmission study are shown in
287 Supplementary Fig. 1.

288 We also repeatedly detected the same mutations in different animals. Of a total of 20
289 repeatedly observed nonsynonymous mutations, 17 were observed in animals linked
290 through transmission (Table 3 and Fig. 1A). In many cases such mutations were
291 located at internal nodes of a phylogenetic tree linking these sequences, suggesting
292 that they gave rise to subsequent variants. For example, nonsynonymous mutation
293 A61G (Ser6Gly) was detected in animal 6005 on days 3, 4, and 5 after contact, with one
294 sequence containing an additional nonsynonymous mutation (C395G, Thr117Arg) on
295 day 4 (Fig. 2B). Animal 5447 constitutes an interesting case, as on day 6 post-contact,
296 two lineages that differed from the consensus sequence were observed and the
297 difference between both branches was a nonsynonymous mutation at Asp145
298 (Supplementary Fig. 1C). One variant comprised six clones harboring Tyr145, while the
299 other branch exhibited eight clones with Asn145. Interestingly, position 145 lies within
300 antigenic site A of the HA (42), therefore mutations in that position could alter the
301 antigenic structure of the virus. More generally, the discovery that the mutations in that
302 antigenic site were detected six days after contact between the infected and uninfected
303 horses, together with their relatively high frequency, is compatible with the action of

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

304 immunological selection by the nascent adaptive immune response, as a significant rise
305 in serum antibodies is readily detected in immunologically naïve horses at day 7 post-
306 infection (13). Mutations at this site were not detected in the donor horse (7248), but
307 one clone bearing an Asp145Asn change was seen on day 6 post-contact in horse
308 7C1C, which was housed with horse 5447 and exposed at the same time, suggesting
309 that the variant was transmitted between these two horses. This indicates that novel
310 lineages bearing distinct antigenicity can evolve within a single individual despite the
311 short infection period of influenza.

312 The observation of intra-host viral populations comprising more than one lineage in
313 different horses, together with the detection of distinct variants harboring common
314 mutation/s in different animals (Table 3), prompted us to determine whether distinct
315 lineages were being transmitted between animals or if they were generated *de novo*
316 within individual horses. We therefore compiled all available data sets from each
317 branch of the transmission chain and determined their evolutionary relationships.
318 Phylogenetic analysis revealed that some variants from *different* animals shared a
319 common ancestor (Fig. 3A and B), indicating that multiple lineages were transmitted
320 between these animals (and a pattern that is highly unlikely to occur through
321 experimental error, see Supplementary Information). This result is of great importance
322 as it means that multiple viral lineages are transmitted between hosts, which in turn
323 means that transmission bottlenecks may not be especially narrow in the case of EIV.

324 **Detection of within-host variants of equine influenza virus associated with the**
325 **emergence of canine influenza**

JVI Accepts published online ahead of print

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

326 The HA mediates attachment and entry into target cells and is a significant determinant
327 of host range specificity (36). CIV emerged as a respiratory pathogen of dogs shortly
328 after 2000 from an apparent direct interspecies transfer of an H3N8 equine influenza
329 virus into dogs (5). Contemporary CIV and EIV share >96% sequence identity in all
330 eight segments (5, 25), with three signature amino acid substitutions located within the
331 stretch of 301 amino acid residues of HA that we examined: Asn54Lys, Asn83Ser and
332 Trp222Leu (5, 25). Strikingly, we found one variant bearing the Asn83Ser mutation in
333 one horse and three other variants harboring mutations in position 222, although none
334 of the latter exhibited leucine in that position (Trp222Gly, Trp222Arg, and Trp222stop).
335 No mutations in the codon for Asn54 were detected. Substitution Ser92Asn, which is
336 present in Canine/FL/04, was also detected, and a substitution of Gly7, albeit different
337 from the one present in Canine/FL/03, was also observed.

338 In addition to the three HA signature substitutions that differentiate contemporary CIV
339 and EIV isolates, there are eight further amino acid changes between the consensus
340 sequence of the virus used in this study and other CIV viruses isolated in United States
341 since 2003: Val14Ala (in the signal peptide), Thr5Ile, Thr30Ser, Ile48Met, Val58Ile,
342 Val78Ala, Asn159Ser, Gln190Glu, and Glu193Lys. Of these, we detected substitutions
343 Thr5Ile, Ile48Met, and Val78Ala in samples of within-host EIV viruses. Six clones
344 harbored nonsynonymous mutations in the codon corresponding to Thr30, although
345 serine was not present in any of them. Other nucleotide substitutions present in
346 individual CIV isolates were also detected. Hence, the genetic diversity required to
347 successfully jump the species barrier and infect a new host may sometimes be partially
348 or completely present in viruses within individuals of the donor species.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

349 **Analysis of within-host variation in natural field cases**

350 To determine whether the variation observed within our experimental setting was
351 comparable to that observed in the field, we analyzed equine H3N8 influenza virus in
352 five nasal swabs from horses naturally infected during an outbreak that took place in
353 England in 2003 (23). Although the within-sample variation detected was similar to that
354 observed in the experimental infections (Table 4), important differences were observed.
355 First, we detected two distinct consensus viral populations that differed in one amino
356 acid position: while Arg62 was present at the consensus level in three samples, the
357 other two samples exhibited Lys62. Further, one of the natural cases examined
358 (OB151) exhibited an unusually high level of intra-host viral diversity, with nine clones
359 harbouring nine mutations, raising the possibility of a mixed infection. The mixed
360 infection in that case was confirmed by phylogenetic analysis of sequences of 158 EIV
361 epidemiological isolates, along with the 73 intra-host sequences derived from this
362 sample. Two phylogenetically distinct viral lineages were observed in this isolate and
363 those grouped with different clades of the Florida sub-lineage (Fig. 4). This represents
364 the first description of mixed infection in EIV, a process that also provides the raw
365 material for segment reassortment. In addition, since these two viruses exhibited four
366 nonsynonymous substitutions within HA1, it is possible that they also differ antigenically,
367 increasing the likelihood of emergence of novel antigenic variants with outbreak
368 potential.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

369 ## Discussion

370 Human cases of avian H5N1 and the recently isolated H1N1 of swine origin constitute
371 current examples of zoonotic infections with pandemic potential (2, 6). However, the
372 determining factors that allow only a small subset of emerging viruses to become
373 established in a new host species are still unclear. For example, although swine
374 influenza viruses periodically spill-over into humans, few have ever caused sustained
375 outbreaks (34). Although it is clear that ecological factors can play a key role in viral
376 emergence, it is likely that there is a genetic basis to cross-species transmission in most
377 cases. Accordingly, it is important to understand whether the mutations required for
378 successful host emergence can arise in the natural reservoir, or if they only arise to
379 significant levels within the new host subsequent to spill-over events, particularly in long
380 chains of transmission, should they exist (1).

381 Here, we applied a clonal sequencing approach to estimate the intra-host mutational
382 spectrum of EIV in its natural host and its implications for viral emergence. The novel
383 patterns of cross-scale viral variability revealed underlines the power of experimental
384 transmission and beyond-consensus viral sequencing to elucidate phylodynamic
385 patterns (12). Analysis of multiple samples taken from the same animal at different
386 times post-infection allowed us to study the evolutionary dynamics of EIV over very
387 short time periods. The HA, as the viral receptor binding protein, is a critical host range
388 determinant and a major target of neutralizing antibodies. Even allowing for the
389 presence of artifactual mutations introduced through error-prone reverse transcription,
390 our results show that the HA1 segment of EIV exhibits significant variation during the

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

391 course of infection, with up to 13% of the clones sequenced harboring single mutations
392 and ~1.4% of the total sequences exhibiting two mutations when compared with the
393 consensus population. The high frequency with which singletons fall into sites that are
394 invariant at the population level, together with the high d_N/d_S values estimated for the
395 intra-host data set and for the epidemiological scale viruses, strongly suggest that most
396 mutations are deleterious and will ultimately be removed by purifying selection. In
397 contrast, some other mutations were detected in multiple days and persisted throughout
398 infection, indicating that they were either neutral or advantageous. In particular, in one
399 animal we observed a high frequency of nonsynonymous mutations within an antigenic
400 site at day 6 post-contact, suggesting that immunological selection may take place in
401 the late stages of infection, when the evolutionary infectivity profile – the net
402 transmission rate of immunologically selected variants, which results from the
403 interaction between viral adaptation and immune history – will be the highest (12).
404 Transmission studies in vaccinated or previously infected animals will be key to
405 determine the role of immunity on within-host variation and its impact on emergence of
406 antigenic variants.

407 Clearly, some of the observed mutations were likely introduced during reverse
408 transcription and PCR amplification. However, it is highly unlikely that these
409 experimental artifacts would result in introduction of repeated changes in the same sites
410 in different samples, particularly in those that follow the transmission chain. On the
411 other hand, if artifactual mutations were commonly introduced as a result of RNA
412 secondary structure, one might expect a higher proportion of mutations at specific sites
413 than the distribution observed here. Indeed, the finding that mutations are passed on in

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

414 a manner that matches the transmission chain, together with a gradual increase in
415 diversity along the course of infection, acts as an independent verification of the validity
416 of this method of examining the viral variation. Finally, our results from control
417 experiments of artifactual mutations (see Supplementary Information) are also in
418 agreement with those of Descloux *et al*. who also evaluated *in vitro* introduced
419 mutations during reverse transcription and PCR in a study of intra-host genetic diversity
420 in dengue virus (7).

421 The frequency of individual mutations at any nucleotide site was only 1.8×10^{-4}, which is
422 similar to that observed by Iqbal *et al.* (17) where within-host variability of avian
423 influenza viruses was assessed *in vivo* using a variety of avian species. However,
424 because our sampling depth only allowed consistent detection of more common (>5-
425 10% of the population) variants, it is likely that greater variation exists at a lower
426 frequency than is reported here. Further studies including deeper sequencing of within-
427 host full genomes will therefore be essential to better understand the likelihood of
428 emergence, as all genomic segments are likely to play distinct roles in host-range
429 specificity (32, 35).

430 Based on these results, we conclude that influenza viral populations continually
431 generate the variants that would enable them to adapt to new host species, although
432 most are purged by purifying selection. Indeed, we detected mutations in HA1
433 associated with the emergence of canine influenza virus within the EIV mutational
434 spectra. As infection progresses, the mutational spectrum gradually gains complexity
435 and the frequency of certain variants in the viral population increases. Hence, individual

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

436 infections of EIV result in an on-going evolutionary process. The fact that within our
437 experimental setting we did not detect fixed substitutions contrasts with our
438 observations in natural cases collected from horses infected during the same outbreak,
439 where intra-host viral populations differed in one amino acid in different animals,
440 suggesting that replacement of the consensus population may require longer chains of
441 transmission.

442 By studying genetic variation both within and among hosts we are also able to examine
443 the process of evolution during inter-host transmission. Importantly, our observation
444 that clones bearing common mutations are found in animals that are directly linked by
445 transmission demonstrates that multiple viral lineages are passed among animals and
446 hence that transmission bottlenecks are not always narrow, such as down to the level of
447 a single virion as proposed for HIV (19). However, the size of the transmission
448 bottlenecks may vary depending on equine management; the bottleneck may be
449 narrower in horses in less direct contact, such as individually stabled animals, or
450 animals at pasture. The transmission of multiple influenza viral lineages will also assist
451 in the process of viral emergence, as the probability that new host species will be
452 exposed to mutations of adaptive value will be greater and natural selection will act
453 more efficiently. Along the same line, loose transmission bottlenecks will allow efficient
454 immune selection if individuals with different immunological histories are exposed to
455 antigenic mutations. The timing of transmission may therefore also play a role in the
456 extent of initial viral diversity, such that the later transmission takes place, the more
457 diverse the transmitted viral population is. In addition, the occurrence of mixed
458 infection, as we describe here for EIV in nature, provides the raw material for

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

459 reassortment (3). Finally, we suggest that understanding the dynamics of within-host
460 variation will assist in developing the theoretical tools required to predict and therefore
461 avoid or control the emergence of new viruses in humans and animals.

JVI Accepts published online ahead of print

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

462 **Acknowledgments**

463 This work was supported by a grant from DEFRA and HEFCE under the Veterinary
464 Training and Research Initiative to the Cambridge Infectious Disease Consortium
465 (CIDC) and also by a program grant from the Wellcome Trust. PM is supported by a
466 Wellcome Trust Veterinary Postdoctoral Fellowship. JLNW is supported by the Alborada
467 Trust. JD was supported by the Horserace Betting Levy Board equine influenza virus
468 surveillance programme. BG and JLNW were supported by the RAPIDD program of the
469 Science & Technology Directorate, Department of Homeland Security, and the Fogarty
470 International Center, National Institutes of Health. BG was also supported by grants
471 NSF0742373, NIH R01 GM083983-01. ECH and CRP were supported by grant R01
472 GM080533.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

473 ## References

474

475 1. **Antia, R., R. R. Regoes, J. C. Koella, and C. T. Bergstrom.** 2003. The role of evolution in
476 the emergence of infectious diseases. Nature **426:**658-61.
477 2. **Beigel, J. H., J. Farrar, A. M. Han, F. G. Hayden, R. Hyer, M. D. de Jong, S. Lochindarat,**
478 **T. K. Nguyen, T. H. Nguyen, T. H. Tran, A. Nicoll, S. Touch, and K. Y. Yuen.** 2005. Avian
479 influenza A (H5N1) infection in humans. N Engl J Med **353:**1374-85.
480 3. **Bryant, N. A., A. S. Rash, C. A. Russell, J. Ross, A. Cooke, S. Bowman, S. Macrae, N. S.**
481 **Lewis, R. Paillot, R. Zanoni, H. Meier, L. A. Griffiths, J. M. Daly, A. Tiwari, T. M.**
482 **Chambers, J. R. Newton, and D. M. Elton.** 2009. Antigenic and genetic variations in
483 European and North American equine influenza virus strains (H3N8) isolated from 2006
484 to 2007. Vet Microbiol **138:**41-52.
485 4. **Cook, R. F., R. Sinclair, and J. A. Mumford.** 1988. Detection of influenza nucleoprotein
486 antigen in nasal secretions from horses infected with A/equine influenza (H3N8) viruses.
487 J Virol Methods **20:**1-12.
488 5. **Crawford, P. C., E. J. Dubovi, W. L. Castleman, I. Stephenson, E. P. Gibbs, L. Chen, C.**
489 **Smith, R. C. Hill, P. Ferro, J. Pompey, R. A. Bright, M. J. Medina, C. M. Johnson, C. W.**
490 **Olsen, N. J. Cox, A. I. Klimov, J. M. Katz, and R. O. Donis.** 2005. Transmission of equine
491 influenza virus to dogs. Science **310:**482-5.
492 6. **Dawood, F. S., S. Jain, L. Finelli, M. W. Shaw, S. Lindstrom, R. J. Garten, L. V. Gubareva,**
493 **X. Xu, C. B. Bridges, and T. M. Uyeki.** 2009. Emergence of a novel swine-origin influenza
494 A (H1N1) virus in humans. N Engl J Med **360:**2605-15.
495 7. **Descloux, E., V. M. Cao-Lormeau, C. Roche, and X. De Lamballerie.** 2009. Dengue 1
496 diversity and microevolution, French Polynesia 2001-2006: connection with
497 epidemiology and clinics. PLoS Negl Trop Dis **3:**e493.
498 8. **Domingo, E., and J. J. Holland.** 1997. RNA virus mutations and fitness for survival. Annu
499 Rev Microbiol **51:**151-78.
500 9. **Ewing, B., and P. Green.** 1998. Base-calling of automated sequencer traces using phred.
501 II. Error probabilities. Genome Res **8:**186-94.
502 10. **Ewing, B., L. Hillier, M. C. Wendl, and P. Green.** 1998. Base-calling of automated
503 sequencer traces using phred. I. Accuracy assessment. Genome Res **8:**175-85.
504 11. **Garten, R. J., C. T. Davis, C. A. Russell, B. Shu, S. Lindstrom, A. Balish, W. M. Sessions,**
505 **X. Xu, E. Skepner, V. Deyde, M. Okomo-Adhiambo, L. Gubareva, J. Barnes, C. B. Smith,**
506 **S. L. Emery, M. J. Hillman, P. Rivailler, J. Smagala, M. de Graaf, D. F. Burke, R. A.**
507 **Fouchier, C. Pappas, C. M. Alpuche-Aranda, H. Lopez-Gatell, H. Olivera, I. Lopez, C. A.**
508 **Myers, D. Faix, P. J. Blair, C. Yu, K. M. Keene, P. D. Dotson, Jr., D. Boxrud, A. R. Sambol,**
509 **S. H. Abid, K. St George, T. Bannerman, A. L. Moore, D. J. Stringer, P. Blevins, G. J.**
510 **Demmler-Harrison, M. Ginsberg, P. Kriner, S. Waterman, S. Smole, H. F. Guevara, E. A.**
511 **Belongia, P. A. Clark, S. T. Beatrice, R. Donis, J. Katz, L. Finelli, C. B. Bridges, M. Shaw,**
512 **D. B. Jernigan, T. M. Uyeki, D. J. Smith, A. I. Klimov, and N. J. Cox.** 2009. Antigenic and

JVI Accepts published online ahead of print

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

513 Genetic Characteristics of Swine-Origin 2009 A(H1N1) Influenza Viruses Circulating in
514 Humans. Science.
515 12. **Grenfell, B. T., O. G. Pybus, J. R. Gog, J. L. Wood, J. M. Daly, J. A. Mumford, and E. C.**
516 **Holmes.** 2004. Unifying the epidemiological and evolutionary dynamics of pathogens.
517 Science **303:**327-32.
518 13. **Hannant, D., J. A. Mumford, and D. M. Jessett.** 1988. Duration of circulating antibody
519 and immunity following infection with equine influenza virus. Vet Rec **122:**125-8.
520 14. **Hoffmann, E., J. Stech, Y. Guan, R. G. Webster, and D. R. Perez.** 2001. Universal primer
521 set for the full-length amplification of all influenza A viruses. Arch Virol **146:**2275-89.
522 15. **Holmes, E. C., E. Ghedin, N. Miller, J. Taylor, Y. Bao, K. St George, B. T. Grenfell, S. L.**
523 **Salzberg, C. M. Fraser, D. J. Lipman, and J. K. Taubenberger.** 2005. Whole-genome
524 analysis of human influenza A virus reveals multiple persistent lineages and
525 reassortment among recent H3N2 viruses. PLoS Biol **3:**e300.
526 16. **Huang, J., L. G. Brieba, and R. Sousa.** 2000. Misincorporation by wild-type and mutant
527 T7 RNA polymerases: identification of interactions that reduce misincorporation rates by
528 stabilizing the catalytically incompetent open conformation. Biochemistry **39:**11571-80.
529 17. **Iqbal, M., H. Xiao, G. Baillie, A. Warry, S. C. Essen, B. Londt, S. M. Brookes, I. H. Brown,**
530 **and J. W. McCauley.** 2009. Within-host variation of avian influenza viruses. Philos Trans
531 R Soc Lond B Biol Sci **364:**2739-47.
532 18. **Ji, J. P., and L. A. Loeb.** 1992. Fidelity of HIV-1 reverse transcriptase copying RNA in vitro.
533 Biochemistry **31:**954-8.
534 19. **Keele, B. F., E. E. Giorgi, J. F. Salazar-Gonzalez, J. M. Decker, K. T. Pham, M. G. Salazar,**
535 **C. Sun, T. Grayson, S. Wang, H. Li, X. Wei, C. Jiang, J. L. Kirchherr, F. Gao, J. A.**
536 **Anderson, L. H. Ping, R. Swanstrom, G. D. Tomaras, W. A. Blattner, P. A. Goepfert, J.**
537 **M. Kilby, M. S. Saag, E. L. Delwart, M. P. Busch, M. S. Cohen, D. C. Montefiori, B. F.**
538 **Haynes, B. Gaschen, G. S. Athreya, H. Y. Lee, N. Wood, C. Seoighe, A. S. Perelson, T.**
539 **Bhattacharya, B. T. Korber, B. H. Hahn, and G. M. Shaw.** 2008. Identification and
540 characterization of transmitted and early founder virus envelopes in primary HIV-1
541 infection. Proc Natl Acad Sci U S A **105:**7552-7.
542 20. **Kuiken, T., E. C. Holmes, J. McCauley, G. F. Rimmelzwaan, C. S. Williams, and B. T.**
543 **Grenfell.** 2006. Host species barriers to influenza virus infections. Science **312:**394-7.
544 21. **Mumford, J. A., D. Hannant, and D. M. Jessett.** 1990. Experimental infection of ponies
545 with equine influenza (H3N8) viruses by intranasal inoculation or exposure to aerosols.
546 Equine Vet J **22:**93-8.
547 22. **Nelson, M. I., and E. C. Holmes.** 2007. The evolution of epidemic influenza. Nat Rev
548 Genet **8:**196-205.
549 23. **Newton, J. R., J. M. Daly, L. Spencer, and J. A. Mumford.** 2006. Description of the
550 outbreak of equine influenza (H3N8) in the United Kingdom in 2003, during which
551 recently vaccinated horses in Newmarket developed respiratory disease. Vet Rec
552 **158:**185-92.
553 24. **Ning, Z., A. J. Cox, and J. C. Mullikin.** 2001. SSAHA: a fast search method for large DNA
554 databases. Genome Res **11:**1725-9.

JVI Accepts published online ahead of print

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

555 25. **Payungporn, S., P. C. Crawford, T. S. Kouo, L. M. Chen, J. Pompey, W. L. Castleman, E.**
556 **J. Dubovi, J. M. Katz, and R. O. Donis.** 2008. Influenza A virus (H3N8) in dogs with
557 respiratory disease, Florida. Emerg Infect Dis **14:**902-8.
558 26. **Pond, S. L., and S. D. Frost.** 2005. Datamonkey: rapid detection of selective pressure on
559 individual sites of codon alignments. Bioinformatics **21:**2531-3.
560 27. **Posada, D., and K. A. Crandall.** 1998. MODELTEST: testing the model of DNA
561 substitution. Bioinformatics **14:**817-8.
562 28. **Rambaut, A., O. G. Pybus, M. I. Nelson, C. Viboud, J. K. Taubenberger, and E. C.**
563 **Holmes.** 2008. The genomic and epidemiological dynamics of human influenza A virus.
564 Nature **453:**615-9.
565 29. **Roumagnac, P., F. X. Weill, C. Dolecek, S. Baker, S. Brisse, N. T. Chinh, T. A. Le, C. J.**
566 **Acosta, J. Farrar, G. Dougan, and M. Achtman.** 2006. Evolutionary history of Salmonella
567 typhi. Science **314:**1301-4.
568 30. **Russell, C. A., T. C. Jones, I. G. Barr, N. J. Cox, R. J. Garten, V. Gregory, I. D. Gust, A. W.**
569 **Hampson, A. J. Hay, A. C. Hurt, J. C. de Jong, A. Kelso, A. I. Klimov, T. Kageyama, N.**
570 **Komadina, A. S. Lapedes, Y. P. Lin, A. Mosterin, M. Obuchi, T. Odagiri, A. D. Osterhaus,**
571 **G. F. Rimmelzwaan, M. W. Shaw, E. Skepner, K. Stohr, M. Tashiro, R. A. Fouchier, and**
572 **D. J. Smith.** 2008. The global circulation of seasonal influenza A (H3N2) viruses. Science
573 **320:**340-6.
574 31. **Schafer, J. R., Y. Kawaoka, W. J. Bean, J. Suss, D. Senne, and R. G. Webster.** 1993.
575 Origin of the pandemic 1957 H2 influenza A virus and the persistence of its possible
576 progenitors in the avian reservoir. Virology **194:**781-8.
577 32. **Scholtissek, C., H. Burger, O. Kistner, and K. F. Shortridge.** 1985. The nucleoprotein as a
578 possible major factor in determining host specificity of influenza H3N2 viruses. Virology
579 **147:**287-94.
580 33. **Scholtissek, C., I. Koennecke, and R. Rott.** 1978. Host range recombinants of fowl
581 plague (influenza A) virus. Virology **91:**79-85.
582 34. **Shinde, V., C. B. Bridges, T. M. Uyeki, B. Shu, A. Balish, X. Xu, S. Lindstrom, L. V.**
583 **Gubareva, V. Deyde, R. J. Garten, M. Harris, S. Gerber, S. Vagasky, F. Smith, N. Pascoe,**
584 **K. Martin, D. Dufficy, K. Ritger, C. Conover, P. Quinlisk, A. Klimov, J. S. Bresee, and L.**
585 **Finelli.** 2009. Triple-Reassortant Swine Influenza A (H1) in Humans in the United States,
586 2005-2009. N Engl J Med.
587 35. **Subbarao, E. K., W. London, and B. R. Murphy.** 1993. A single amino acid in the PB2
588 gene of influenza A virus is a determinant of host range. J Virol **67:**1761-4.
589 36. **Suzuki, Y., T. Ito, T. Suzuki, R. E. Holland, Jr., T. M. Chambers, M. Kiso, H. Ishida, and Y.**
590 **Kawaoka.** 2000. Sialic acid species as a determinant of the host range of influenza A
591 viruses. J Virol **74:**11825-31.
592 37. **Svarovskaia, E. S., S. R. Cheslock, W. H. Zhang, W. S. Hu, and V. K. Pathak.** 2003.
593 Retroviral mutation rates and reverse transcriptase fidelity. Front Biosci **8:**d117-34.
594 38. **Swofford, D. L.** 2002. PAUP*. Phylogenetic Analysis Using Parsimony (*and Other
595 Methods). Version 4. Sunderland, Massachussetts. Sinauer Associates.
596 39. **Taubenberger, J. K., A. H. Reid, R. M. Lourens, R. Wang, G. Jin, and T. G. Fanning.** 2005.
597 Characterization of the 1918 influenza virus polymerase genes. Nature **437:**889-93.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

598 40. **Waddell, G. H., M. B. Teigland, and M. M. Sigel.** 1963. A New Influenza Virus Associated
599 with Equine Respiratory Disease. J Am Vet Med Assoc **143:**587-90.
600 41. **Webster, R. G., W. J. Bean, O. T. Gorman, T. M. Chambers, and Y. Kawaoka.** 1992.
601 Evolution and ecology of influenza A viruses. Microbiol Rev **56:**152-79.
602 42. **Wiley, D. C., and J. J. Skehel.** 1987. The structure and function of the hemagglutinin
603 membrane glycoprotein of influenza virus. Annu Rev Biochem **56:**365-94.
604 43. **Wood, J. M., G. C. Schild, C. Folkers, J. Mumford, and R. W. Newman.** 1983. The
605 standardization of inactivated equine influenza vaccines by single-radial
606 immunodiffusion. J Biol Stand **11:**133-6.
607 44. **Woolhouse, M. E., D. T. Haydon, and R. Antia.** 2005. Emerging pathogens: the
608 epidemiology and evolution of species jumps. Trends Ecol Evol **20:**238-44.
609
610
611

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

Horse	Chain	Day[a]	Number of sequences	Total number of nt mutations	Mean pairwise distance +/- SE	Global d_N/d_S	Stop codons	Indels[b]	Sites with > 1 nonsyn. mutation
7D36*	A/B-1	3	152	29	0.00042+-0.00001	0.71	4	1	0
		5	82	18	0.00048+-0.00002	1.04	0	0	1
7248	A-2	3	154	15	0.00022+-0.00000	0.72	0	0	0
		4	65	3	0.00010+-0.00001	0.20	0	0	0
		5	154	24	0.00034+-0.00001	0.37	0	0	1
		6	62	9	0.00032+-0.00001	0.37	0	0	1
5447	A-3	3	69	4	0.00013+-0.00001	1.81	0	0	0
		4	44	9	0.00045+-0.00003	1.25	0	0	0
		5	50	12	0.00053+-0.00002	0.58	1	0	0
		6	51	31	0.00123+-0.00003	1.11	2	0	2
7C1C	A-3	2	52	6	0.00026+-0.00001	0.56	0	0	0
		3	54	8	0.00033+-0.00002	1.13	0	0	0
		4	83	12	0.00032+-0.00001	0.90	0	0	2
		6	75	14	0.00041+-0.00001	1.10	1	0	2
2F50	A-4	3	49	4	0.00018+-0.00001	0.29	0	0	0
		4	54	9	0.00041+-0.00002	0.13	0	0	0
		5	80	13	0.00036+-0.00001	3.40	0	0	0
		7	46	7	0.00033+-0.00002	0.34	0	0	0
7A45	A-4	3	107	13	0.00027+-0.00001	0.58	0	0	0
6005	B-2	3	67	9	0.00033+-0.00001	N/A	1	0	2
		4	81	14	0.00038+-0.00001	3.49	1	1	1
		5	52	7	0.00029+-0.00002	N/A	1	0	1
		6	72	27	0.00073+-0.00002	1.98	0	0	3
5257	B-3	2	112	16	0.00034+-0.00001	0.30	1	0	1
		3	127	21	0.00038+-0.00001	0.40	3	0	0
780C	B-4	3	71	18	0.00056+-0.00002	0.85	0	0	0
5D1A	B-4	4	54	4	0.00016+-0.00001	N/A	0	0	0
		5	81	12	0.00033+-0.00001	0.87	2	0	1
		6	79	14	0.00039+-0.00001	1.60	0	0	3
0443*	0	2	87	10	0.00026+-0.00001	0.85	0	0	0

Table 1. Analysis of intra-host variation in EIV sequences from the transmission experiment.

* Experimentally inoculated animals.

[a] Day post-inoculation (for inoculated animals) or after initial contact with donor horse.

[b] Insertions/deletions.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

Horse	Mutation, amino acid change	Day[a]	Number of clones	Present at epidemiological scale?	Variable at epidemiological scale?	Other horses (days) with same mutation
7248	A228G[b]	4	1	-		5257 (D3), 7C1C (D5)
		5	1	-		
	A855G[b]	3	1	-		5447 (D5)
		5	1	-		
2F50	A231G[b]	4	2	-		
		7	2	-		
5447	T29C, Leu10Pro[c]	3	1	no	no (L)	7D36 (D5)
		5	1			
	G289A, Glu82Lys	4	1	yes	yes (E,G,K)	
		6	1			
	C407T, Thr121Ile	4	1	yes	yes (I,M,T)	5D1A (D5)
	A406C, Thr121Pro	6	1	no		
6005	C49A, Gln2Lys	4	1	no	no (Q)	
		6	11			
	A61G, Ser6Gly	3	2	yes	yes (D,G,N,S)	
		4	4			
		5	1			
	A884G, Glu280Gly	3	2	no	yes (E,K)	0443 (D2), 7D36 (D5), 5257 (D2)
		4	1			
		5	3			
		6	11			
5D1A	A413G, Glu123Gly	4	1	no	no (E)	
		5	1			

Table 2. Intra-host EIV mutations present on multiple days.

[a] Day after initial contact with donor horse.

[b] Synonymous mutations.

[c] Signal peptide (absent in mature HA1).

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

Mutation, amino acid change	Horses (days)	Direct contact?	Horses from same chain?	Motif
T26C,Leu9Pro	7248(5) 5D1A(5)	No	No	
T29C,Leu10Pro	7D36(5) 5447(3) 5447(5)	No	Yes	
T257C,Leu71Pro	5257(3) 5D1A(6)	Yes	Yes	
T337C,Tyr98His	7D36(5) 5D1A(6)	No	Yes	
T358C,Tyr105His	5447(3) 7C1C(3)	Yes	Yes	
C370T,Arg109Trp	7C1C(3) 780C(3)	No	No	
C395G,Tyr117Arg	6005(4) 7A45(3)	No	No	
C407T,Thr121Ile	5447(4) 5D1A(5)	No	No	Ag[b] A
G409A,Ala122Thr	6005(5) 5447(3) 5D1A(5)	No	Yes (6005 and 5D1A)	Ag A
T460C,Cys139Arg	6005(5) 5D1A(6)	No	Yes	
G478A,Asp145Asn	7C1C(6) 5447(6)	Yes	Yes	Ag A
T497G,Leu151Arg	5257(2) 7C1C(3)	No	No	
C545T,Thr167Ile	7D36(3) 7248(3)	Yes	Yes	Gly[c]
A635G,Gln197Arg	7248(6) 780C(3)	No	No	Ag B
T709A,Trp222Arg	7D36(5) 5447(6)	No	Yes	
A730G,Arg229Gly	7D36(3) 7248(5) 2F50(5)	Yes (7D36 and 7248)	Yes (all)	receptor-binding-associated
A835G,Lys264Glu	0443(2) 7D36(5) 7248(5) 2F50(5)	Yes (7D36 and 0443, 7D36 and 7248)	Yes (all)	
C863T,Pro273Leu	780C(3) 2F50(7)	No	No	Ag C
G868A,Asp275Asn	0443(2) 7248(3)	No	Yes?	Ag C
A884G,Glu280Gly	0443(2) 7D36(5) 6005(3) 6005(4) 6005(5) 6005(6) 5257(2)	Yes (0443 and 7D36, 7D36 and 6005, 6005 and 5257)	Yes (all)	

Table 3. Intra-host EIV non-synonymous mutations present in multiple horses.

[a] Signal peptide (absent in mature HA1).

[b] Ag: Antigenic site.

[c] Glycosylation site.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

Horse	Number of sequences	Total number of nt mutations[a]	Mean pairwise distance	Global d_N/d_S	Stop codons	Indels[b]	Sites with > 1 nonsyn. mutation
OB150	68	6	0.0001	0.43	1	0	0
OB151	73	145	0.002	0.35	0	0	5
OB152	82	8	0.0002	0.58	0	0	0
OB153	75	92	0.0005	0.49	0	0	2
OB172	74	6	0.0001	0.42	0	0	1

Table 4. Analysis of patterns of intra-host variation in EIV sequences from natural field cases.

[a] Mutations relative to the reference sequence A/Equine/Newmarket/5/2003 (Accession number: FJ375213).

[b] Insertions/deletions.

JVI Accepts published online ahead of print

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

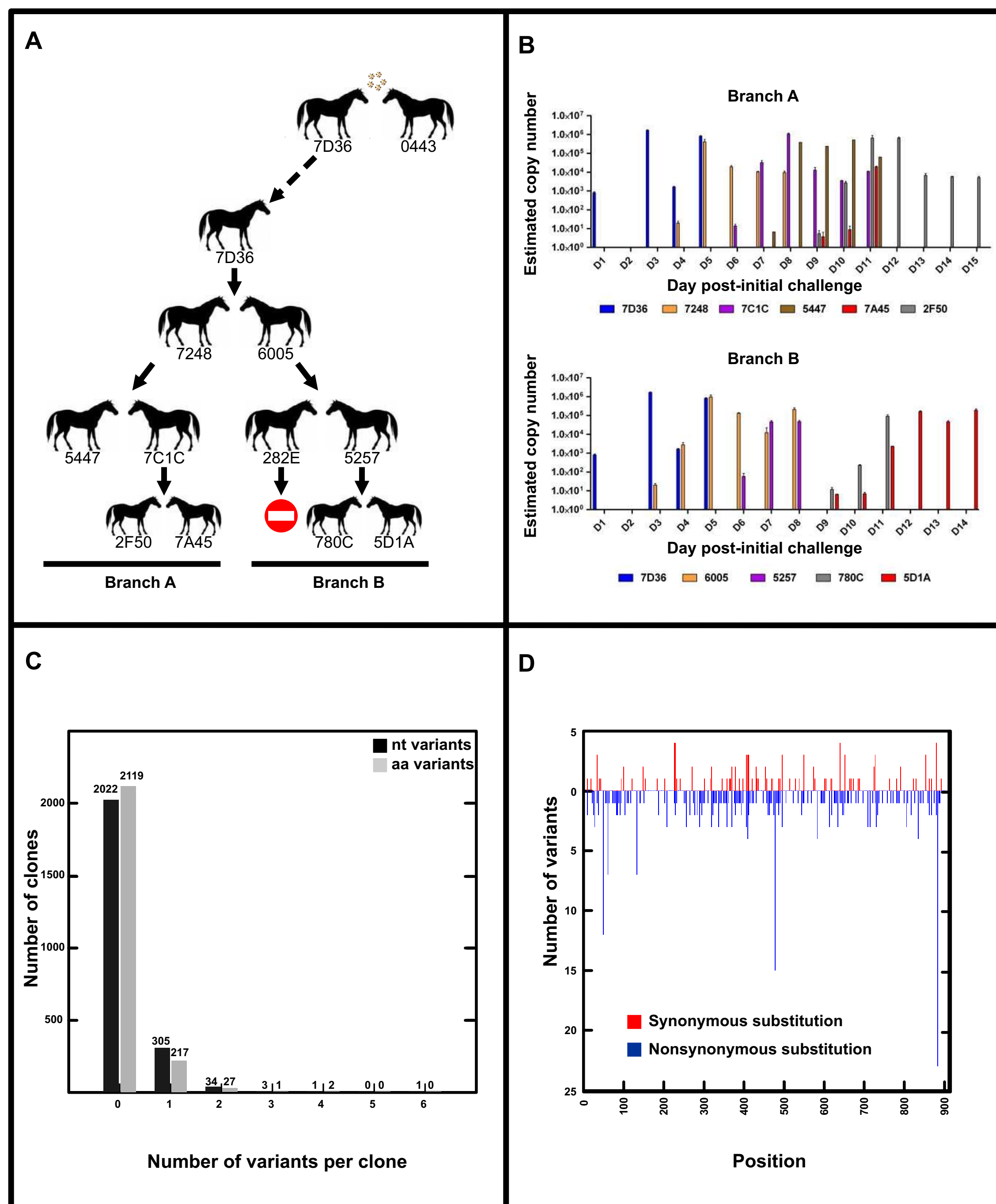

Figure 1. Intra-host genetic variation of EIV in horses. (A) Layout of the transmission chain. Of the two horses experimentally inoculated with EIV, only one (7D36) was used to start the transmission chain (dotted arrow). No virus was detected in nasal swabs from horse 282E (indicated with a stop sign). (B) Estimation of viral copy numbers from individual swabs by real-time PCR. (C) Number of clones with nucleotide (black columns) and amino acid (grey columns) mutations. (D) Position and frequency of synonymous (red) and nonsynonymous (blue) mutations relative to the reference sequence (nucleotide level).

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

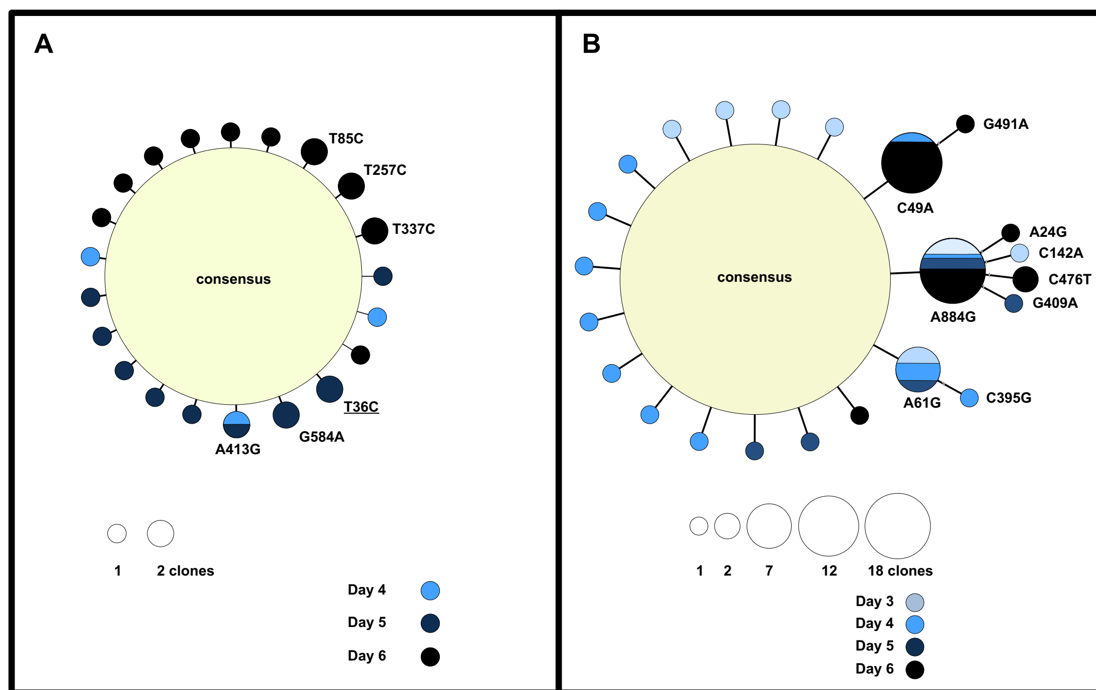

Figure 2. Intra-host EIV populations are constituted by a mixture of closely-related genomes. Minimum spanning tree of animals 5D1A (a) and 6005 (b). Each tree was inferred by compiling sequences from multiple days. With the exception of singletons, mutations are indicated on respective branches. Synonymous mutations are underlined. Circle sizes are proportional to the number of sequences that exhibit each variant.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

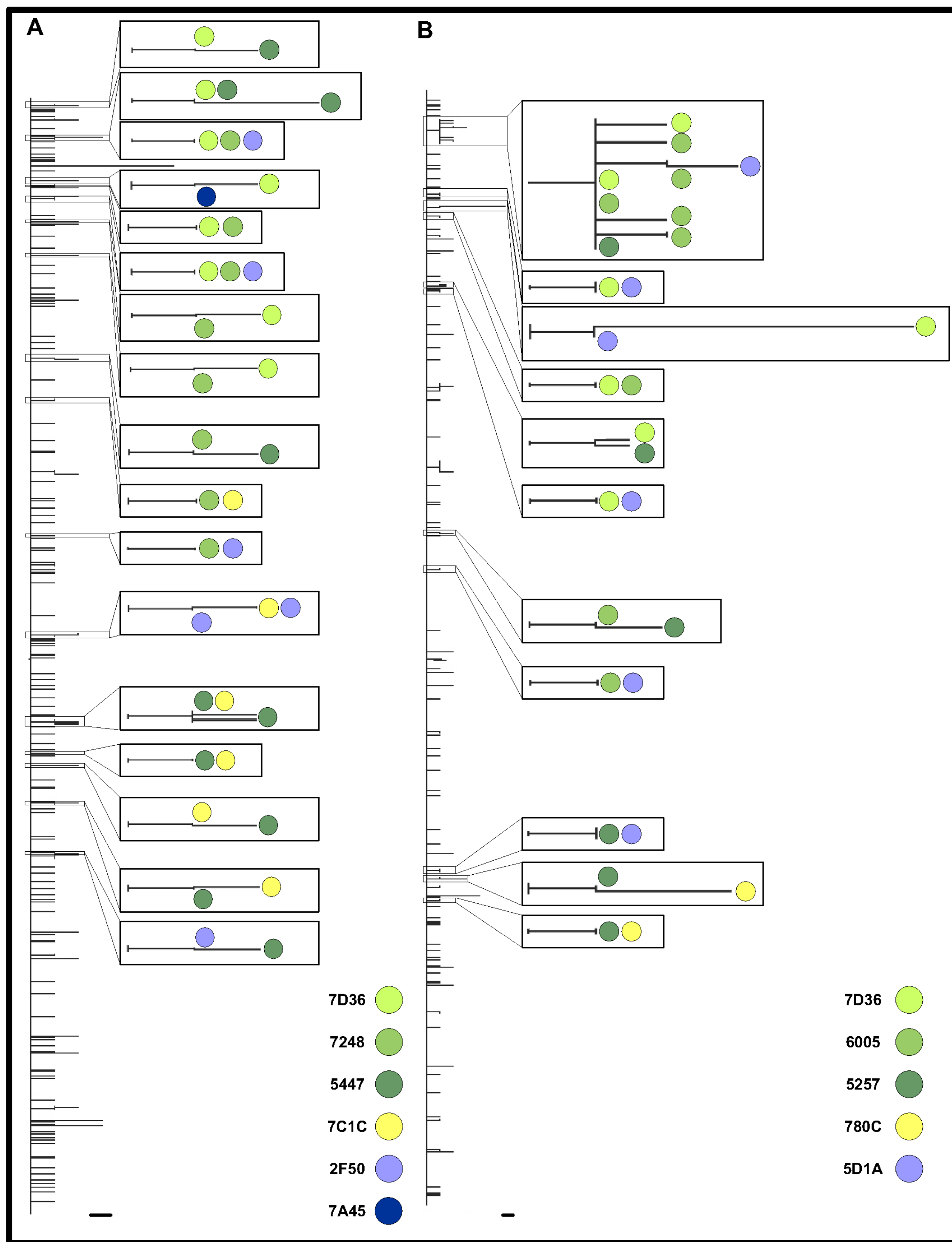

Figure 3. Unrooted maximum likelihood phylogenetic trees for HA1 segment clones from samples obtained from chain 1 (A) and chain 2 (B). Branch lengths are drawn to scale, with those displaying variants detected in more than one animal magnified. Colored circles represent different animals. Horizontal branch lengths are drawn to a scale of nucleotide substitutions per site (scale bar = 0.001 subs/site).

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010

JVI Accepts published online ahead of print

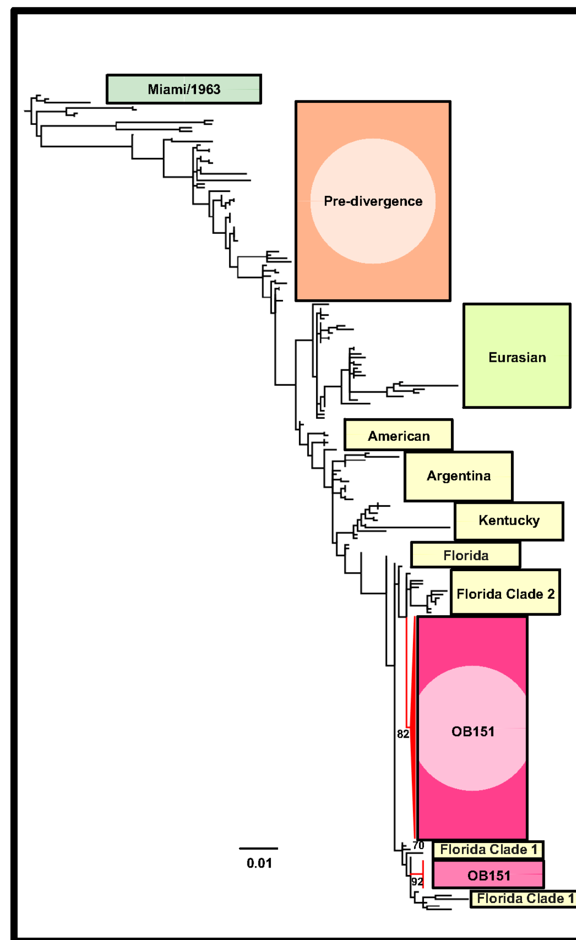

Figure 4. Maximum likelihood phylogenetic tree for HA1 segment clones from a natural case of EIV (sample OB151) compared with the global phylogeny of EIV H3N8. Colored boxes represent distinct EIV phylogenetic groups according to the OIE nomenclature of H3N8 strains(43). Clones from sample OB151 are represented in red boxes. Bootstrap values (>70%) are shown for key nodes relating to the phylogenetic position of those sequences sampled from animal OB151. Horizontal branches are drawn to a scale of nucleotide substitutions per site and the tree is rooted on the EIV Miami isolate from 1963.

Downloaded from jvi.asm.org at NAT INST OF HEALTH LIB on May 19, 2010