

1 **TITLE:**

2 Genomic dissection of an Icelandic epidemic of equine respiratory disease

3

4 **AUTHORS:**

5 Sigríður Björnsdóttir¹, Simon R. Harris², Vilhjálmur Svansson³, Eggert Gunnarsson³,

6 Ólöf G. Sigurðardóttir³, Kristina Gammeljord⁴, Karen F. Steward⁵, J. Richard

7 Newton⁵, Carl Robinson⁵, Amelia R. L. Charbonneau⁵, Julian Parkhill², Matthew T.G.

8 Holden^{2,6}, Andrew S. Waller⁵

9

10 **AFFILIATIONS:**

11 ¹ MAST Icelandic Food and Veterinary Authority, Austurvegur 64, IS-800 Selfoss,
12 Iceland.

13 ² Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton,
14 Cambridge, CB10 1SA, United Kingdom.

15 ³ Institute for Experimental Pathology, University of Iceland, Keldur, Vesturlandsveg,
16 IS-112 Reykjavík, Iceland.

17 ⁴ Department of Veterinary Disease Biology, University of Copenhagen, Denmark.

18 ⁵ Centre for Preventive Medicine, Animal Health Trust, Lanwades Park, Kentford,
19 Newmarket, Suffolk, CB8 7UU, United Kingdom.

20 ⁶ School of Medicine, University of St Andrews, St Andrews, Fife, KY16 9TF, United
21 Kingdom.

22

23 Correspondence should be addressed to A.S.W (andrew.waller@aht.org.uk; +44
24 (0)1638 751000).

25

26 **RUNNING TITLE:**

27 Genomic dissection of a respiratory epidemic

28

29 **KEYWORDS:**

30 Respiratory disease; Icelandic epidemic; Equine; Genome phylogeny; *Streptococcus*
31 *zooepidemicus*.

32

33 **ABSTRACT:**

34 The native horse population of Iceland has remained free of major infectious
35 diseases. Between May and July 2010 an epidemic of respiratory disease swept
36 through the population. Initial microbiological investigations ruled out known equine
37 viral agents as the cause of the infections, but identified the opportunistic pathogen
38 *Streptococcus zooepidemicus* as being frequently isolated from diseased animals.
39 This diverse bacterial species has a broad host range and is usually regarded as a
40 commensal of horses. By genome sequencing *S. zooepidemicus* recovered from
41 horses during the epidemic we show that although multiple clones of *S.*
42 *zooepidemicus* were present in the population, one particular clone, ST209, was
43 responsible for the epidemic. Concurrent with the epidemic, ST209 caused zoonotic
44 infections, highlighting the pathogenic potential of this clone. Phylogenetic analysis
45 suggests that the original ST209 strain entered Iceland in late 2008 or early 2009.
46 Epidemiological investigation revealed that the incursion of this strain into a training
47 yard that utilized a submerged treadmill between the 5th and 19th of February 2010
48 was a critical trigger for the ensuing epidemic of disease, provided a nidus for the
49 infection of multiple horses, and subsequent distribution of these animals to multiple
50 sites in Iceland.

51

52 **INTRODUCTION:**

53 The Icelandic horse population is geographically isolated and arose from animals
54 introduced by settlers in the 9th and 10th centuries, with virtually no import of horses
55 for the last 1,000 years. Icelandic horses are extremely hardy and are prized for their
56 maneuverability and movement; Icelandic horses possess five distinct gaits, instead

57 the three gaits common to most other horses. The isolation of the population has
58 meant that Icelandic horses have remained free from most common contagious
59 diseases of Equidae including equine influenza, equine rhinopneumonitis, equine
60 viral arteritis, *Rhodococcus equi* and strangles (*Streptococcus equi* subsp. *equi*)
61 (Robinson, Steward, Potts, Barker, Hammond, Pierce, Gunnarsson, Svansson,
62 Slater, Newton et al. 2013; Sanz, Oliveira, Loynachan, Page, Svansson, Giguere and
63 Horohov 2014; Torfason, Thorsteinsdottir, Torsteinsdottir and Svansson 2008). As
64 such the population is extremely vulnerable to these and other infectious agents and
65 strict biosecurity measures are in place to maintain a high health status.

66

67 Between April and November of 2010, an epidemic of respiratory disease that was
68 typified by a persistent dry cough and mucopurulent nasal discharge, swept across
69 Iceland affecting almost the entire population of an estimated 77,000 horses. In
70 addition to the paralysis of all equine activities it resulted in a self-imposed ban on the
71 export of horses; with important economic consequences for the Icelandic horse
72 industry.

73

74 Here we describe how the opportunistic bacterial pathogen *Streptococcus equi*
75 subsp. *zooepidemicus* (*S. zooepidemicus*) was identified as the causative agent of
76 the respiratory disease and how, by analyzing the genome sequences of *S.*
77 *zooepidemicus* isolates, a distinct rapidly expanding causal clone was identified.
78 Furthermore, we demonstrate that this clone was responsible for zoonotic infections
79 during the course of the epidemic, and that the epidemic strain has become endemic
80 within the Icelandic horse population.

81

82 **RESULTS:**

83 **Epidemiology of a respiratory disease epidemic**

84 The first cases of the epidemic were reported on the 7th of April at an equine center
85 in the north of Iceland (farm 16). At that time, 10 out of the 42 horses showed clinical
86 signs of respiratory tract infection, one of which had been coughing for 16 days.
87 Three other horses at farm 16 had a history of similar clinical signs two to six weeks
88 earlier. Information received from other parts of the country within that week,
89 revealed a widespread infection throughout the country and it became apparent that
90 an epidemic could not be avoided.

91

92 Dry coughing was usually the first clinical sign noted; most often coexisting with
93 mucopurulent nasal discharge and mild conjunctivitis, although rectal temperature
94 remained normal in most horses (Supplementary Table 1). Laryngoscopy revealed
95 laryngitis and inflammation of the upper trachea. The incubation period was between
96 two and four weeks and the duration of clinical signs varied from two to ten weeks.
97 The entire equine population in Iceland appeared to be susceptible to the disease
98 resulting in almost 100% morbidity, but very low mortality.

99

100 **Identification of *S. zooepidemicus* as the causative agent**

101 The initial assumption due to the rate of spread was that a virus was responsible.
102 Nasal swabs were tested by PCR for viruses known to cause respiratory disease in
103 horses and some common respiratory viruses of humans and animals
104 (Supplementary Table 2). Paired blood samples were used to determine if horses
105 developed an antibody response to equine respiratory viruses during the course of
106 their disease. Established and primary cell lines were also used in attempts to isolate
107 a novel causal virus. All of these tests proved negative with the exception of equine
108 gammaherpesviruses, which were isolated from small numbers of both healthy and
109 clinically affected horses and therefore these viruses were not regarded as being
110 related to the epidemic.

111

112 In the absence of a viral pathogen, it was noted that the Gram-positive bacterium *S.*
113 *zooepidemicus* was isolated from almost all of the nasal swabs taken from coughing
114 horses and from the diseased tissues of occasional fatal cases (Supplementary
115 Table 1). Initially this was not considered of relevance, as although *S. zooepidemicus*
116 is an opportunistic pathogen associated with a range of equine infections (Wood,
117 Newton, Chanter and Mumford 2005a; Wood, Newton, Chanter and Mumford 2005b)
118 and infections of other animals including humans (Abbott, Acke, Khan, Muldoon,
119 Markey, Pinilla, Leonard, Steward and Waller 2010; Balter, Benin, Pinto, Teixeira,
120 Alvim, Luna, Jackson, LaClaire, Elliott, Facklam et al. 2000), it is routinely isolated
121 from healthy horses and is widely considered to be a commensal. Evidence for the
122 role of *S. zooepidemicus* in the outbreak was provided when three healthy horses
123 (horses 16, 17 and 18, Supplementary Table 1) from the University of Iceland were
124 transferred into a barn (farm 5) on the 14th June. The seventeen resident horses had
125 been diagnosed with respiratory disease on the 31st May and *S. zooepidemicus* was
126 recovered from the nasal swabs of eleven horses, which were sampled on the 2nd
127 June. Following their introduction, the three healthy horses were monitored for
128 clinical signs of disease. Ten days post-arrival, nasal discharge was apparent, which
129 became mucopurulent 9 days later accompanied by the first observation of coughing.
130 *S. zooepidemicus* was recovered from these horses from 20 days post-introduction.
131 Post-mortem examination of these horses revealed signs of respiratory infection,
132 which included the presence of mucopurulent material in the nasal cavity, larynx and
133 trachea and enlarged cervical and mandibular lymph nodes. Histopathological
134 analysis identified sub-acute rhinitis, laryngitis and tracheitis with transmigration of
135 neutrophils through the mucosal epithelium. There was a mild hyperplasia and
136 metaplasia of the respiratory epithelium of the trachea in all three horses, with loss of
137 goblet and ciliated cells and focal erosions. Lympho-histiocytic and plasmacytic
138 inflammation was seen in the superficial submucosa, forming a broad inflammatory
139 band in the trachea of horse 17 and in the nasal mucosa of horse 16. *S.*

140 *zooepidemicus* was isolated in large numbers from the nasal cavity, larynx and
141 trachea from all three horses, and also isolated from the nasopharynx of horses 16
142 and 18, from a bronchus of horse 16, and from the guttural pouch of horse 18.

143

144 **Genomic investigation of the outbreak**

145 To determine if the epidemic was associated with the introduction and spread of a
146 specific *S. zooepidemicus* strain, we employed whole genome sequencing (WGS) to
147 interrogate the relationships of 251 *S. zooepidemicus* isolates recovered from the
148 nasal swabs of horses that were either actively showing signs of respiratory disease
149 or were stabled on affected premises during the course of the epidemic
150 (Supplementary Table 1). Included in this were multiple isolates recovered from the
151 same clinical sample or the same horse sampled over time, which optimized the
152 isolation of the epidemic strain in the face of concomitant colonization. During the
153 epidemic, cases of *S. zooepidemicus* disease were reported in companion animals
154 (two cats and one dog) and three people and these additional isolates were included
155 in our analysis. In order to provide a wider temporal and genetic context, we included
156 sequenced Icelandic isolates from seven horses, one dog and two sheep that pre-
157 dated the 2010 epidemic, and also included the genomes of 38 strains of *S.*
158 *zooepidemicus* from outside Iceland that were representative of the known species
159 diversity as a whole as defined by multilocus sequence typing (MLST) (Webb, Jolley,
160 Mitchell, Robinson, Newton, Maiden and Waller 2008) (Supplementary Table 1).

161

162 **Population structure of Icelandic *S. zooepidemicus* is dominated by a few** 163 **dominant clones**

164 The availability of MLST has permitted the identification of strains that were
165 associated with specific infections, including respiratory disease and abortion in
166 horses and acute fatal haemorrhagic pneumonia in dogs, suggesting that certain
167 strains of *S. zooepidemicus* may have greater potential to cause disease (Abbott,

168 Acke, Khan, Muldoon, Markey, Pinilla, Leonard, Steward and Waller 2010; Chalker,
169 Waller, Webb, Spearing, Crosse, Brownlie and Erles 2012; Webb, Jolley, Mitchell,
170 Robinson, Newton, Maiden and Waller 2008). However, the identification of specific
171 disease-causing strains can be confounded by the limited ability of MLST to
172 differentiate closely related strains. Phylogenetic analysis of the WGS data provided
173 a high-resolution view of the Icelandic *S. zooepidemicus* population structure. The
174 majority of *S. zooepidemicus* recovered during the epidemic (201 of 257 isolates
175 (78%) from 33 premises) fell into four distinct clades (Supplementary Table 1 and
176 Figure 1) that corresponded to four separate MLST sequence types (STs): Clade 4
177 (ST306) contained 37 isolates obtained from the Institute for Experimental Pathology
178 at Keldur, Reykjavik between September and November 2010 and differed by a
179 maximum of 44 SNPs; Clade 3 (ST248) contained 52 isolates obtained from 8
180 different farms, and a canine isolate, and differed by a maximum of 151 SNPs; Clade
181 2 (ST246) contained 29 isolates, obtained from 10 farms, and a human isolate; and
182 differed by a maximum of 153 SNPs; and Clade 1 (ST209) containing 83 isolates
183 from 22 farms, and a human and a feline isolate that differed by a maximum of 25
184 SNPs.

185

186 **Confirmation that the ST209 (clade 1) *S. zooepidemicus* population was the**
187 **cause of the epidemic**

188 The wide geographic dispersal of ST209 and its relative lack of diversity indicate that
189 this strain had been transmitted to horses throughout Iceland in a short period of
190 time, suggesting that this strain was responsible for the epidemic of respiratory
191 disease. Evidence to support this came from the investigation of the healthy horses
192 that were introduced to farm 5 and subsequently acquired disease. Seven of the 14
193 isolates recovered from these horses were ST209, whilst one was ST246. Five of the
194 six remaining isolates that were recovered from these horses clustered together as a
195 separate group within the Icelandic population of *S. zooepidemicus* and were not

196 recovered from horses that were already resident on farm 5 (Supplementary Table
197 1). One of the horses introduced into farm 5 (horse 17) did not provide a ST209
198 strain. Notably this horse was euthanized 23 days post-introduction in contrast to the
199 other two horses that were euthanized 28 days post-introduction. A possible
200 explanation for the absence of ST209 may be that these strains were not shed in any
201 great number until the later time point, and the heterogeneous population of *S.*
202 *zooepidemicus* may have confounded the identification of the epidemic strain during the
203 early stages of infection. In this regard it is worth noting that all of the isolates
204 recovered from nasal swabs taken from horses 16 and 18 from day 28 and at the
205 time of post-mortem examination on day 30 belonged to ST209.

206

207 Further evidence of the localized transmission of ST209 can be found in the
208 clustered regularly interspaced short palindromic repeats (CRISPRs), which provides
209 a snapshot of a strain's exposure to mobile genetic elements (Holden, Heather,
210 Paillot, Steward, Webb, Ainslie, Jourdan, Bason, Holroyd, Mungall et al. 2009; Waller
211 and Robinson 2013). In keeping with rapid transmission across the population, the
212 Icelandic isolates of ST209 predominantly shared the same complement of 41 spacer
213 sequences (Figure 2). However, ST209 isolates from horses on farm 5 were found to
214 contain a novel spacer 35 sequence. This unique spacer was present in isolates
215 recovered from four resident horses (3, 5, 7 and 14), but not from three other
216 residents (1, 2 and 9) on the 2nd June, or from any other affected horses from other
217 farms throughout Iceland. The ST209 isolates that were recovered from the horses,
218 which were introduced to farm 5 (16 and 18) contained the unique spacer 35
219 sequence directly linking the acquisition of this strain to their arrival and exposure to
220 the resident horses at farm 5.

221

222 **Epidemiological network analysis identifies an infection source**

223 Network analysis of affected farms identified a single common training yard (yard A)
224 as a primary center of transmission (Figure 3). Two horses leaving yard A on the 19th
225 of February transmitted the disease both to farm 16 and to a stable in the Reykjavik
226 area. All horses (n=20) that left yard A in March, transmitted the disease to new
227 premises (n=18). These new premises each had up to 50 horses stabled and
228 became secondary centers of transmission. However, two horses from yard A, which
229 had returned to farm 16 on the 4th of February, were not incubating the disease.
230 Therefore, it is likely that the transmission of ST209 strains to horses at yard A began
231 between the 5th and 19th of February 2010.

232

233 Yard A opened in January 2010 and offers rehabilitation for up to 35 horses
234 recovering from injury or poor condition as well as equipment training for competition
235 horses. Although the epidemic agent was generally transmitted through the coughing
236 of infected horses, coughing was not observed at yard A before April 2010, most
237 probably due to the relatively long incubation period permitting apparently healthy
238 horses to return to their original premises post-training. Instead, the most likely
239 source of transmission of the epidemic strain at yard A was a water treadmill, which
240 horses used on a daily basis. The water used in the treadmill contained no
241 disinfectant and was changed on a once- or twice-weekly basis, providing ideal
242 conditions for the transmission of *S. zooepidemicus* between the visiting horses.
243 Transmission via the water treadmill at yard A provides one possible explanation for
244 the efficiency of the spread of the epidemic throughout Iceland, with multiple
245 secondary sources of infection combining with the original point source to produce a
246 perfect storm of transmission.

247

248 **Estimation of the date of common ancestry for the Icelandic ST209 isolates**

249 Following the resolution of the epidemic we hypothesized that the ST209 strain had
250 become endemic within the Icelandic horse population, and that the genetic variation

251 that had accrued since the epidemic could be exploited to estimate the time of most
252 recent common ancestor (TMRCA) shared by the Icelandic ST209 population. To
253 determine if the ST209 strain persisted within the Icelandic horse population, we
254 sampled an additional 36 healthy horses three years after the epidemic. One
255 hundred and seventy-five *S. zooepidemicus* isolates (Supplementary Table 1) were
256 recovered, and of these 20 isolates were identified as ST209, obtained from four
257 healthy horses stabled at different farms throughout Iceland. WGS of the isolates
258 from each of the individual horses indicated that they were very closely related,
259 however, isolates from the different horses clustered separately, with longer branch
260 lengths relative to the isolates recovered from the epidemic in 2010. We analyzed a
261 subset of our dataset, for which the date of isolation was known, with the Bayesian
262 phylogenetics software, BEAST (Drummond and Rambaut 2007). Removal of strains
263 with identical sequences that were recovered from the same animal on the same
264 date produced a dataset of 59 ST209 isolates with 434 polymorphic core genome
265 positions. This dataset was comprised of 4 isolates from 2013, 48 isolates that were
266 recovered during the epidemic in 2010 and 7 non-Icelandic isolates from the
267 collection at the Animal Health Trust, which share an identical or closely related ST to
268 ST209. BEAST includes a number of relaxed molecular clock models that permit
269 modeling of variation in substitution rates on different branches of the tree, allowing
270 correction for the observed rate variation in our data. Utilizing these models, we
271 found that a strict clock with a skyline population model was the best fit of our data
272 and the mean substitution rate per core genome site per year was calculated as $2 \times$
273 10^{-6} . This substitution rate is similar to core genome rates reported for other
274 streptococci, including *Streptococcus pyogenes* (1.1×10^{-6}) (Davies, Holden,
275 Coupland, Chen, Venturini, Barnett, Zakour, Tse, Dougan, Yuen et al. 2014) and
276 *Streptococcus pneumoniae* (1.57×10^{-6}) (Croucher, Harris, Fraser, Quail, Burton, van
277 der Linden, McGee, von Gottberg, Song, Ko et al. 2011), and many other Gram-
278 positive bacteria, including *Staphylococcus aureus* (3.3×10^{-6}) (Harris, Feil, Holden,

279 Quail, Nickerson, Chantratita, Gardete, Tavares, Day, Lindsay et al. 2010a).
280 Interestingly, this rate is faster than the closely-related host-restricted pathogen
281 *Streptococcus equi* (5.22×10^{-7}) (Harris, Robinson, Steward, Webb, Paillot, Parkhill,
282 Holden and Waller 2015) with which *S. zooepidemicus* shares >97% DNA identity.
283 This difference most likely reflects the unusual lifestyle of *S. equi*, which can persist
284 within the guttural pouches of recovered horses (Harris, Robinson, Steward, Webb,
285 Paillot, Parkhill, Holden and Waller 2015). The analysis provided a median estimate
286 for the TMRCA of the Icelandic ST209 strains of July 2008 (95% HPD: August 2007
287 to May 2009). Our data suggest that the ST209 strains could have circulated in a
288 small number of Icelandic horses prior to the start of the epidemic in 2010. In
289 accordance with our network analysis, the movement of an infected animal to yard A,
290 which received horses from throughout Iceland, was critical for the wider
291 transmission of the ST209 strains. We have previously reported that the within-host
292 diversification of *S. equi* enables individual animals to shed multiple variants of the
293 original infecting strain (Harris, Robinson, Steward, Webb, Paillot, Parkhill, Holden
294 and Waller 2015). Therefore, it is also possible that more than one variant of ST209,
295 potentially originating from the same animal, was introduced to Iceland prior to the
296 start of the epidemic. In this latter scenario, the ST209 variants may have circulated
297 for a shorter period of time within the Icelandic horse population prior to February
298 2010. However, regardless of the date at which the epidemic strain was introduced to
299 Iceland, the transfer of an infected horse to yard A, identified through our network
300 analysis, remains essential to the rapid spread of the ST209 strains that was
301 characteristic of the epidemic.

302

303 **Origins of ST209 strains in Iceland**

304 The Icelandic epidemic ST209 strains share greatest genetic similarity to strain 435,
305 which was isolated from a coughing horse in Sweden during 2008. This strain was
306 selected for sequencing because it shared a common ST209 profile (Supplementary

307 Table 1). To date, with the exception of Iceland, ST209 strains have only been
308 isolated from Scandinavia and it is interesting to note the close relationships between
309 many training yards in Iceland and this region of Europe. The import of horses to
310 Iceland has been prohibited since 1882 for biosecurity reasons. Although the import
311 of used riding equipment is also prohibited, it is difficult to control. Therefore,
312 contaminated tack represents a possible horse-related route of introduction of the
313 ST209 strain.

314

315 During the epidemic a ST209 strain was recovered from an infected cat and a lady
316 with septicemia who had suffered a miscarriage that may have been linked to her
317 infection (Figure 4). Previously, a *S. zooepidemicus* strain of ST209 was associated
318 with a zoonotic infection in Finland (Pelkonen, Lindahl, Suomala, Karhukorpi,
319 Vuorinen, Koivula, Vaisanen, Pentikainen, Autio and Tuuminen 2013) and other
320 strains are known to infect companion animals (Webb, Jolley, Mitchell, Robinson,
321 Newton, Maiden and Waller 2008). The ability of ST209 strains to cross host
322 boundaries provides an alternative import mechanism whereby the infection of a
323 human may have facilitated onward anthroponotic transmission to horses resident in
324 Iceland.

325

326 **DISCUSSION:**

327 By analyzing WGS data of *S. zooepidemicus* isolates collected during the 2010
328 epidemic of equine respiratory disease, we obtained sufficient data to resolve recent
329 from distant transmission events, and so identify the causal strain. Our data fully
330 support the epidemiological analysis of this epidemic and points to the incursion of a
331 novel strain of *S. zooepidemicus* into the Icelandic horse population. It also reveals
332 the diverse pathogenomic properties of *S. zooepidemicus* and how a novel strain can
333 spread rapidly through a susceptible population devoid of sufficient cross-protective
334 immunity despite a background of concomitant colonization with endemic strains.

335 This study emphasizes the importance of national biosecurity as a barrier to protect
336 vulnerable populations such as Iceland's unique horse population.

337

338

339 **METHODS:**

340 **Study collection**

341 The origins and details of the isolates of *S. zooepidemicus* that were sequenced in
342 this study are listed in Supplementary Table 1. β -hemolytic colonies of *S.*
343 *zooepidemicus* strains were recovered from glycerol stocks following overnight
344 growth on COBA strep select plates (bioMérieux). Their identity was confirmed by
345 fermentation of ribose and sorbitol, but not trehalose in Purple broth (Becton
346 Dickinson). The published genome sequences for *S. zooepidemicus* strains
347 MGCS10565 (Beres, Sesso, Pinto, Hoe, Porcella, Deleo and Musser 2008), H70
348 (Holden, Heather, Paillot, Steward, Webb, Ainslie, Jourdan, Bason, Holroyd, Mungall
349 et al. 2009), BHS5 (Paillot, Darby, Robinson, Wright, Steward, Anderson, Webb,
350 Holden, Efstratiou, Broughton et al. 2010) and ATCC35246 (Ma, Geng, Zhang, Yu,
351 Yi, Lei, Lu, Fan and Hu 2011) were included in the analysis to capture the known
352 species diversity.

353

354 **Epidemiological investigation**

355 A questionnaire was sent to 200 premises, including all professional training yards
356 and breeding farms in the country in June 2010 with a follow up in September 2010
357 to determine whether and when their horses had shown signs of respiratory disease
358 within the preceding months. Further information pertaining to the movement of
359 horses into affected premises was collected by interviews, enabling the network of
360 connected farms and training centers to be investigated.

361

362 **Contact study**

363 Three clinically healthy Icelandic horses (horses 16, 17, 18) aged 18–21 years from
364 the Institute for Experimental Pathology at Keldur were transferred to farm 5 two
365 weeks after the first signs of respiratory disease were identified in the resident
366 horses. The experiment was conducted in accordance with the Icelandic animal care
367 guidelines for experimental animals (The Icelandic Animal welfare Act no. 15/1994),
368 the Icelandic Regulations on Animal Experimentation no. 279/2002, which is based
369 on and complies with the European Convention for the Protection of Vertebrate
370 Animals used for Experimental and other Scientific Purposes, European Treaty
371 Series no. 123, 18.III.1986, and after formal approval from the Icelandic ethical
372 committee on animal research, license number 0710-0401. The horses were
373 monitored for the onset of clinical signs by measuring rectal temperature,
374 auscultation of lung and trachea, palpation of submandibular lymph nodes and the
375 pharyngeal area, and assessment of nasal discharge and coughing. Nasal swabs
376 (Coban®) for bacteriological and virological examination and blood samples, serum
377 and EDTA (Vacurette) were collected every two days. Horse 17 was euthanized and a
378 post-mortem examination was conducted two days after the first signs of respiratory
379 disease (day 23 post-introduction), where muco-purulent nasal discharge and
380 coughing coexisted with positive cultivation of *S. zooepidemicus*. Horses 16 and 18
381 were euthanized seven days later and a post-mortem examination was conducted.

382

383 **Virus isolation and diagnostics**

384 Nasal swabs taken from horses with clinical signs of respiratory disease were
385 submitted for testing at the Institute of Experimental Pathology in Keldur in Iceland,
386 the National Veterinary Institute in Uppsala, Sweden and the Institute of Virology,
387 Justus-Liebig-Universität in Giessen, Germany. Cell free transport media from nasal
388 swabs and enriched peripheral blood leukocyte cells (PBLC)-plasma from EDTA
389 stabilized blood samples collected from the three contact study horses (horses 16, 17
390 and 18) post-introduction to farm 5 were used to inoculate retroviral vector

391 LXS116E6E7-transfected foetal lung and kidney cell lines (Thorsteinsdóttir *et al.*,
392 unpublished data). Further samples from swabs taken from several locations in the
393 respiratory tract and conjunctiva post-mortem were also used to inoculate these cells.
394 Inoculated cells were passaged and the conditioned cell culture media was also used
395 to inoculate new cell cultures. The sample was judged as being virus negative if no
396 cytopathic effect could be seen after the third passage. At each passage cells were
397 processed by cytospin and examined by indirect fluorescent antibody staining using
398 pooled convalescent serum samples from 6 horses as the primary antibody.
399 Additional virus isolation attempts were made in the same way with nasal swabs from
400 14 horses with clinical signs of disease using primary equine foetal lung and kidney
401 cells (Torfason, Thorsteinsdottir, Torsteinsdottir and Svansson 2008), RK-13 cells
402 (ATCC), MA-104 cells (Microbiological Associates) and Vero cells (ATCC).

403

404 A wide range of serological tests and PCR assays for equine herpesvirus types 1, 2,
405 4 and 5, equine infectious anemia virus, equine arteritis virus, equine rhinitis virus A
406 and B, equine influenza A type 1 (H7N7) and type 2 (H3N8), equine salemvirus,
407 human and equine adenovirus, mammalian reovirus serotypes 1, 2 and 3, carnivore
408 parvovirus, human rhinovirus, human enterovirus, canine pneumovirus, influenza A
409 and B, parainfluenza virus 3 and mammalian coronaviruses were used according to
410 the Manual of Diagnostic Tests and Vaccines for Terrestrial Animals produced by the
411 OIE and other published methods (Allard, Albinsson and Wadell 2001; Back, Penell,
412 Pringle, Isaksson, Roneus, Treiberg Berndtsson and Stahl 2015; Balasuriya,
413 Leutenegger, Topol, McCollum, Timoney and MacLachlan 2002; Coggins, Norcross
414 and Nusbaum 1972; Crabb, MacPherson, Reubel, Browning, Studdert and Drummer
415 1995; Dynon, Varrasso, Ficorilli, Holloway, Reubel, Li, Hartley, Studdert and
416 Drummer 2001; Escutenaire, Mohamed, Isaksson, Thoren, Klingeborn, Belak, Berg
417 and Blomberg 2007; Leary, Erker, Chalmers, Cruz, Wetzel, Desai, Mushahwar and
418 Dermody 2002; Peacey, Hall, Bocacao and Huang 2009; Puig, Jofre, Lucena, Allard,

419 Wadell and Girones 1994; Renshaw, Glaser, Van Campen, Weiland and Dubovi
420 2000; Renshaw, Zylich, Laverack, Glaser and Dubovi 2010; Schunck, Kraft and
421 Truyen 1995; Selvaraju and Selvarangan 2010; Svansson, Roelse, Olafsdottir,
422 Thorsteinsdottir, Torfason and Torsteinsdottir 2009; Thorsteinsdottir, Torfason,
423 Torsteinsdottir and Svansson 2010; Torfason, Thorsteinsdottir, Torsteinsdottir and
424 Svansson 2008) to identify a potential viral cause (Supplementary Table 2).

425

426 **Post mortem examinations**

427 A full post mortem examination was performed on the three experimentally infected
428 horses. Samples were taken from all the major organs in addition to samples from
429 the nasal mucosa, larynx and trachea. Tissues were fixed in 10 % neutral buffered
430 formalin. The formalin-fixed material was processed by routine paraffin embedding,
431 and 4 µm thick sections were cut, mounted, and stained with haematoxylin and
432 eosin.

433

434 **DNA preparation**

435 A single colony of each *S. zooepidemicus* strain was grown overnight in 3 ml of Todd
436 Hewitt (TH) broth containing 30 µg/ml hyaluronidase (Sigma) at 37 °C in a 5 % CO₂
437 enriched atmosphere, centrifuged, the pellet re-suspended in 200 µl Gram +ve lysis
438 solution (GenElute, Sigma) containing 250 units/ml mutanolysin and 2 x 10⁶ units/ml
439 lysozyme and incubated for 1 hour at 37 °C to allow efficient cell lysis. DNA was then
440 purified using GenElute spin columns according to manufacturer's instructions (all
441 Sigma).

442

443 **MLST assignment**

444 MLST sequence types were identified from sequence data as previously described
445 (Webb, Jolley, Mitchell, Robinson, Newton, Maiden and Waller 2008).

446

447 **DNA Sequencing**

448 DNA libraries for isolates recovered pre-2011 were created using a method adapted
449 from the Illumina Indexing standard protocol. In brief, the steps taken (with clean-up
450 between each step) were: genomic DNA was fragmented by acoustic shearing to
451 enrich for 200 bp fragments using a Covaris E210, end-repaired and A-tailed.
452 Adapter ligation was followed by overlap extension PCR using the Illumina 3 primer
453 set to introduce specific tag sequences between the sequencing and flow-cell binding
454 sites of the Illumina adapter. After quantitation by qPCR followed by normalization
455 and pooling, pooled libraries were sequenced on Illumina GAII and HiSeq platforms
456 according to the manufacturer's protocols generating index tag-end sequences. For
457 isolates recovered post-2011, DNA libraries were prepared using the Illumina
458 Nextera XT DNA library prep kit, according to the standard protocol. The DNA
459 libraries were pooled and quantified using the KAPA Library Quantification Kit for
460 Illumina platforms prior to sequencing on Illumina MiSeq according to the
461 manufacturer's protocol.

462

463 **Variation detection**

464 Illumina reads were mapped onto the relevant reference sequences using SMALT
465 (<http://www.sanger.ac.uk/resources/software/smalt/>); H70 (accession number
466 FM204884) for the total population, and a de novo assembly of Clade 1. A minimum
467 of 30x depth of coverage for more than 92 % of the reference genomes was
468 achieved for both references (Supplementary Table 1). The default mapping
469 parameters recommended for reads were employed, but with the minimum score
470 required for mapping increased to 30 to make the mapping more conservative.
471 Candidate SNPs were identified using SAMtools mpileup (Li, Handsaker, Wysoker,
472 Fennell, Ruan, Homer, Marth, Abecasis and Durbin 2009), with SNPs filtered to
473 remove those at sites with a mapping depth less than 5 reads and a SNP score
474 below 60. SNPs at sites with heterogeneous mappings were filtered out if the SNP

475 was present in less than 75 % of reads at that site (Harris, Feil, Holden, Quail,
476 Nickerson, Chantratita, Gardete, Tavares, Day, Lindsay et al. 2010b). Identification of
477 the core genomes was performed as previously described (Harris, Feil, Holden,
478 Quail, Nickerson, Chantratita, Gardete, Tavares, Day, Lindsay et al. 2010b; Holden,
479 Hsu, Kurt, Weinert, Mather, Harris, Strommenger, Layer, Witte, de Lencastre et al.
480 2013). Recombination was detected in the genomes using Gubbins ([http://sanger-](http://sanger-pathogens.github.io/gubbins/)
481 [pathogens.github.io/gubbins/](http://sanger-pathogens.github.io/gubbins/)) (Croucher, Page, Connor, Delaney, Keane, Bentley,
482 Parkhill and Harris 2014). Phylogenetic trees for ST209 and ST22 were constructed
483 separately using RAxML v7.0.4 (Stamatakis 2006) for all sites in the core genomes
484 containing SNPs, using a GTR model with a gamma correction for among site rate
485 variation (Harris, Feil, Holden, Quail, Nickerson, Chantratita, Gardete, Tavares, Day,
486 Lindsay et al. 2010b). We used the Bayesian software package BEAST (v1.7.4)
487 (Drummond, Suchard, Xie and Rambaut 2012) to investigate the temporal, spatial
488 and demographic evolution of the clade 1 population. To estimate the substitution
489 rates and times for divergences of internal nodes on the tree, a GTR model with a
490 gamma correction for among-site rate variation was used. All combinations of strict,
491 relaxed lognormal, relaxed exponential and random clock models and constant,
492 exponential, expansion, logistic and skyline population models were evaluated. For
493 each, three independent chains were run for 100 million generations, sampling every
494 10 generations. On completion each model was checked for convergence, both by
495 checking ESS values were greater than 200 for key parameters, and by checking
496 independent runs had converged on similar results. Models including logistic
497 population models failed to converge so were discarded. Models were compared for
498 their fit to the data using Bayes Factors based on the harmonic mean estimator as
499 calculated by the program Tracer v1.4 from the BEAST package. The best-fit model
500 was found to be a strict skyline population model. A maximum clade credibility (MCC)
501 tree was created from the resulting combined trees using the treeAnnotator program,
502 also from the BEAST package.

503

504 **ST209-specific qPCR**

505 ST209 isolates (clade 1) were identified using a specific qPCR for the *nrdE* allele 31.
506 Forward (5'-ACCAAAAAGAAAATGCT-3') and reverse (5'-
507 TCAACACTATAAGGACTAAAGAGA-3') primers were used to amplify the *nrdE* allele
508 31 on a STEPONE plus machine (Applied Biosystems) with KAPA SYBR FAST ABI
509 PRISM reagents (Kapa Biosystems) and cycling conditions of 95 °C for 3 minutes
510 followed by 40 cycles of 95 °C for 30 seconds and 62 °C for 10 seconds. A melt
511 curve was performed between 60 °C and 95 °C reading SYBR every 0.3 °C. All
512 qPCR experiments were performed in triplicate.

513

514 **DATA ACCESS:**

515 The Illumina sequences generated and used in this study have been deposited in the
516 European Nucleotide Archive under the study accession number ERP000883. The
517 accession numbers for the sequences of each isolate are listed in Supplementary
518 Table 1.

519

520 **ACKNOWLEDGEMENTS:**

521 The authors thank Dr. Louise Treiberg Berndtsson and co-workers at the National
522 Veterinary Institute in Uppsala, Sweden and Dr. Christine Förster at the Institute of
523 Virology, Justus-Liebig-Universität in Giessen, Germany for their assistance with
524 virological testing and the personnel at the department of bacteriology at Keldur for
525 their technical assistance. The Icelandic human isolates were obtained from the
526 Department of Clinical Microbiology, Landspítali University Hospital, Reykjavík,
527 Iceland. The human isolates from the United Kingdom were provided by Professor
528 Androulla Efstratiou at Public Health England. Bovine isolates were provided by Dr.
529 Adrian Whatmore of the Animal and Plant Health Agency. The human isolate from
530 Finland was a kind gift from Professor Sinikka Pelkonen and Dr. Tamara Tuuminen

531 of the Finnish Food Safety Authority and University of Helsinki, respectively. We are
532 grateful to Dr. Alistair Darby of the University of Liverpool for providing the genome
533 sequence of *S. zooepidemicus* strain 1175. We also thank the core sequencing and
534 informatics teams at the Sanger Institute for their assistance and The Wellcome Trust
535 for its support of the Sanger Institute Pathogen Genomics and Biology groups. SRH,
536 JP and MTGH were supported by Wellcome Trust grant 098051.

537

538 **AUTHOR CONTRIBUTIONS:**

539 SB, VS, EG, OGS, JRN, MTGH and ASW designed the study. SB, SRH, VS, EG,
540 OGS, KG, KFS, CR, ARLC, MTGH and ASW carried out the research. SB, VS, EG,
541 OGS, JP and ASW supplied isolates, metadata and whole genome sequencing. SB,
542 SRH, VS, EG, KFS, JRN, CR, ARLC, MTGH and ASW analyzed the data. SB, SRH,
543 VS, EG, OGS, MTGH and ASW wrote the manuscript. All authors read and approved
544 the final manuscript.

545

546 **REFERENCES:**

547 Abbott, Y., Acke, E., Khan, S., Muldoon, E.G., Markey, B.K., Pinilla, M., Leonard,
548 F.C., Steward, K., and Waller, A. 2010. Zoonotic transmission of
549 *Streptococcus equi* subsp. *zooepidemicus* from a dog to a handler. *J Med*
550 *Microbiol* **59**: 120-123.

551 Allard, A., Albinsson, B., and Wadell, G. 2001. Rapid typing of human adenoviruses
552 by a general PCR combined with restriction endonuclease analysis. *J Clin*
553 *Microbiol* **39**: 498-505.

554 Back, H., Penell, J., Pringle, J., Isaksson, M., Roneus, N., Treiberg Berndtsson, L.,
555 and Stahl, K. 2015. A longitudinal study of poor performance and subclinical
556 respiratory viral activity in Standardbred trotters. *Vet Rec Open* **2**: e000107.

- 557 Balasuriya, U.B., Leutenegger, C.M., Topol, J.B., McCollum, W.H., Timoney, P.J.,
558 and MacLachlan, N.J. 2002. Detection of equine arteritis virus by real-time
559 TaqMan reverse transcription-PCR assay. *J Virol Methods* **101**: 21-28.
- 560 Balter, S., Benin, A., Pinto, S.W., Teixeira, L.M., Alvim, G.G., Luna, E., Jackson, D.,
561 LaClaire, L., Elliott, J., Facklam, R. et al. 2000. Epidemic nephritis in Nova
562 Serrana, Brazil. *Lancet* **355**: 1776-1780.
- 563 Beres, S.B., Sesso, R., Pinto, S.W., Hoe, N.P., Porcella, S.F., Deleo, F.R., and
564 Musser, J.M. 2008. Genome sequence of a lancefield group C *Streptococcus*
565 *zooepidemicus* strain causing epidemic nephritis: new information about an
566 old disease. *PLoS ONE* **3**: e3026.
- 567 Chalker, V.J., Waller, A., Webb, K., Spearing, E., Crosse, P., Brownlie, J., and Erles,
568 K. 2012. Genetic diversity of *Streptococcus equi* subsp. *zooepidemicus* and
569 doxycycline resistance in kennelled dogs. *J Clin Microbiol* **50**: 2134-2136.
- 570 Coggins, L., Norcross, N.L., and Nusbaum, S.R. 1972. Diagnosis of equine infectious
571 anemia by immunodiffusion test. *Am J Vet Res* **33**: 11-18.
- 572 Crabb, B.S., MacPherson, C.M., Reubel, G.H., Browning, G.F., Studdert, M.J., and
573 Drummer, H.E. 1995. A type-specific serological test to distinguish antibodies
574 to equine herpesviruses 4 and 1. *Arch Virol* **140**: 245-258.
- 575 Croucher, N.J., Harris, S.R., Fraser, C., Quail, M.A., Burton, J., van der Linden, M.,
576 McGee, L., von Gottberg, A., Song, J.H., Ko, K.S. et al. 2011. Rapid
577 pneumococcal evolution in response to clinical interventions. *Science* **331**:
578 430-434.
- 579 Croucher, N.J., Page, A.J., Connor, T.R., Delaney, A.J., Keane, J.A., Bentley, S.D.,
580 Parkhill, J., and Harris, S.R. 2014. Rapid phylogenetic analysis of large
581 samples of recombinant bacterial whole genome sequences using Gubbins.
582 *Nucleic Acids Res.*
- 583 Davies, M.R., Holden, M.T., Coupland, P., Chen, J.H., Venturini, C., Barnett, T.C.,
584 Zakour, N.L., Tse, H., Dougan, G., Yuen, K.Y. et al. 2014. Emergence of

585 scarlet fever *Streptococcus pyogenes* emm12 clones in Hong Kong is
586 associated with toxin acquisition and multidrug resistance. *Nat Genet* **47**: 84-
587 87.

588 Drummond, A.J. and Rambaut, A. 2007. BEAST: Bayesian evolutionary analysis by
589 sampling trees. *BMC Evol Biol* **7**: 214.

590 Drummond, A.J., Suchard, M.A., Xie, D., and Rambaut, A. 2012. Bayesian
591 phylogenetics with BEAUti and the BEAST 1.7. *Molecular biology and*
592 *evolution* **29**: 1969-1973.

593 Dynon, K., Varrasso, A., Ficorilli, N., Holloway, S., Reubel, G., Li, F., Hartley, C.,
594 Studdert, M., and Drummer, H. 2001. Identification of equine herpesvirus 3
595 (equine coital exanthema virus), equine gammaherpesviruses 2 and 5, equine
596 adenoviruses 1 and 2, equine arteritis virus and equine rhinitis A virus by
597 polymerase chain reaction. *Aust Vet J* **79**: 695-702.

598 Escutenaire, S., Mohamed, N., Isaksson, M., Thoren, P., Klingeborn, B., Belak, S.,
599 Berg, M., and Blomberg, J. 2007. SYBR Green real-time reverse
600 transcription-polymerase chain reaction assay for the generic detection of
601 coronaviruses. *Arch Virol* **152**: 41-58.

602 Harris, S.R., Feil, E.J., Holden, M.T., Quail, M.A., Nickerson, E.K., Chantratita, N.,
603 Gardete, S., Tavares, A., Day, N., Lindsay, J.A. et al. 2010a. Evolution of
604 MRSA during hospital transmission and intercontinental spread. *Science* **327**:
605 469-474.

606 Harris, S.R., Feil, E.J., Holden, M.T.G., Quail, M.A., Nickerson, E.K., Chantratita, N.,
607 Gardete, S., Tavares, A., Day, N., Lindsay, J.A. et al. 2010b. Evolution of
608 MRSA during hospital transmission and intercontinental spread. *Science* **327**:
609 469-474.

610 Harris, S.R., Robinson, C., Steward, K.F., Webb, K.S., Paillot, R., Parkhill, J., Holden,
611 M.T., and Waller, A.S. 2015. Genome specialization and decay of the

- 612 strangles pathogen, *Streptococcus equi*, is driven by persistent infection.
613 *Genome Res* **25**: 1360-1371.
- 614 Holden, M.T., Heather, Z., Paillot, R., Steward, K.F., Webb, K., Ainslie, F., Jourdan,
615 T., Bason, N.C., Holroyd, N.E., Mungall, K. et al. 2009. Genomic evidence for
616 the evolution of *Streptococcus equi*: host restriction, increased virulence, and
617 genetic exchange with human pathogens. *PLoS Pathogens* **5**: e1000346.
- 618 Holden, M.T., Hsu, L.Y., Kurt, K., Weinert, L.A., Mather, A.E., Harris, S.R.,
619 Strommenger, B., Layer, F., Witte, W., de Lencastre, H. et al. 2013. A
620 genomic portrait of the emergence, evolution, and global spread of a
621 methicillin-resistant *Staphylococcus aureus* pandemic. *Genome Res* **23**: 653-
622 664.
- 623 Leary, T.P., Erker, J.C., Chalmers, M.L., Cruz, A.T., Wetzel, J.D., Desai, S.M.,
624 Mushahwar, I.K., and Dermody, T.S. 2002. Detection of mammalian reovirus
625 RNA by using reverse transcription-PCR: sequence diversity within the
626 lambda3-encoding L1 gene. *J Clin Microbiol* **40**: 1368-1375.
- 627 Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., Marth, G.,
628 Abecasis, G., and Durbin, R. 2009. The Sequence Alignment/Map format and
629 SAMtools. *Bioinformatics* **25**: 2078-2079.
- 630 Ma, Z., Geng, J., Zhang, H., Yu, H., Yi, L., Lei, M., Lu, C.P., Fan, H.J., and Hu, S.
631 2011. Complete genome sequence of *Streptococcus equi* subsp.
632 *zooepidemicus* strain ATCC 35246. *J Bacteriol* **193**: 5583-5584.
- 633 Paillot, R., Darby, A.C., Robinson, C., Wright, N.L., Steward, K.F., Anderson, E.,
634 Webb, K., Holden, M.T., Efstratiou, A., Broughton, K. et al. 2010. Identification
635 of three novel superantigen-encoding genes in *Streptococcus equi* subsp.
636 *zooepidemicus*, *szeF*, *szeN*, and *szeP*. *Infect Immun* **78**: 4817-4827.
- 637 Peacey, M., Hall, R.J., Bocacao, J., and Huang, Q.S. 2009. Diagnostic assay
638 recommended by the world health organization for swine origin influenza a

- 639 (h1n1) virus cross-reacts with H5N1 influenza virus. *J Clin Microbiol* **47**: 3789-
640 3790.
- 641 Pelkonen, S., Lindahl, S.B., Suomala, P., Karhukorpi, J., Vuorinen, S., Koivula, I.,
642 Vaisanen, T., Pentikainen, J., Autio, T., and Tuuminen, T. 2013. Transmission
643 of *Streptococcus equi* subspecies *zooepidemicus* infection from horses to
644 humans. *Emerg Infect Dis* **19**: 1041-1048.
- 645 Puig, M., Jofre, J., Lucena, F., Allard, A., Wadell, G., and Girones, R. 1994. Detection
646 of adenoviruses and enteroviruses in polluted waters by nested PCR
647 amplification. *Appl Environ Microbiol* **60**: 2963-2970.
- 648 Renshaw, R.W., Glaser, A.L., Van Campen, H., Weiland, F., and Dubovi, E.J. 2000.
649 Identification and phylogenetic comparison of Salem virus, a novel
650 paramyxovirus of horses. *Virology* **270**: 417-429.
- 651 Renshaw, R.W., Zyllich, N.C., Laverack, M.A., Glaser, A.L., and Dubovi, E.J. 2010.
652 Pneumovirus in dogs with acute respiratory disease. *Emerg Infect Dis* **16**:
653 993-995.
- 654 Robinson, C., Steward, K.F., Potts, N., Barker, C., Hammond, T.A., Pierce, K.,
655 Gunnarsson, E., Svansson, V., Slater, J., Newton, J.R. et al. 2013. Combining
656 two serological assays optimises sensitivity and specificity for the
657 identification of *Streptococcus equi* subsp. *equi* exposure. *Vet J* **197**: 188-
658 191.
- 659 Sanz, M.G., Oliveira, A.F., Loynachan, A., Page, A., Svansson, V., Giguere, S., and
660 Horohov, D.W. 2014. Validation and evaluation of VapA-specific IgG and IgG
661 subclass enzyme-linked immunosorbent assays (ELISAs) to identify foals with
662 *Rhodococcus equi* pneumonia. *Equine Vet J*.
- 663 Schunck, B., Kraft, W., and Truyen, U. 1995. A simple touch-down polymerase chain
664 reaction for the detection of canine parvovirus and feline panleukopenia virus
665 in feces. *J Virol Methods* **55**: 427-433.

- 666 Selvaraju, S.B. and Selvarangan, R. 2010. Evaluation of three influenza A and B
667 real-time reverse transcription-PCR assays and a new 2009 H1N1 assay for
668 detection of influenza viruses. *J Clin Microbiol* **48**: 3870-3875.
- 669 Stamatakis, A. 2006. RAxML-VI-HPC: maximum likelihood-based phylogenetic
670 analyses with thousands of taxa and mixed models. *Bioinformatics* **22**: 2688-
671 2690.
- 672 Svansson, V., Roelse, M., Olafsdottir, G., Thorsteinsdottir, L., Torfason, E.G., and
673 Torsteinsdottir, S. 2009. Immune response against equine gammaherpesvirus
674 in Icelandic horses. *Vet Microbiol* **137**: 363-368.
- 675 Thorsteinsdottir, L., Torfason, E.G., Torsteinsdottir, S., and Svansson, V. 2010.
676 Isolation and partial sequencing of Equid herpesvirus 5 from a horse in
677 Iceland. *J Vet Diagn Invest* **22**: 420-423.
- 678 Torfason, E.G., Thorsteinsdottir, L., Torsteinsdottir, S., and Svansson, V. 2008. Study
679 of equid herpesviruses 2 and 5 in Iceland with a type-specific polymerase
680 chain reaction. *Res Vet Sci* **85**: 605-611.
- 681 Waller, A.S. and Robinson, C. 2013. *Streptococcus zooepidemicus* and
682 *Streptococcus equi* evolution: the role of CRISPRs. *Biochem Soc Trans* **41**:
683 1437-1443.
- 684 Webb, K., Jolley, K.A., Mitchell, Z., Robinson, C., Newton, J.R., Maiden, M.C., and
685 Waller, A. 2008. Development of an unambiguous and discriminatory
686 multilocus sequence typing scheme for the *Streptococcus zooepidemicus*
687 group. *Microbiology* **154**: 3016-3024.
- 688 Wood, J.L., Newton, J.R., Chanter, N., and Mumford, J.A. 2005a. Association
689 between respiratory disease and bacterial and viral infections in British
690 racehorses. *J Clin Microbiol* **43**: 120-126.
- 691 Wood, J.L., Newton, J.R., Chanter, N., and Mumford, J.A. 2005b. Inflammatory
692 airway disease, nasal discharge and respiratory infections in young British
693 racehorses. *Equine Vet J* **37**: 236-242.

694

695

696 **Figure 1.** Diversity of the Iceland *S. zooepidemicus* population. Phylogenetic reconstruction of the
697 Icelandic *S. zooepidemicus* population structure (center); Neighbor Joining phylogenetic tree was
698 built using core SNPs, with SNPs in regions of recombination removed. Included in the phylogeny
699 were *S. zooepidemicus* isolates from outside Iceland that were representative of the genetic
700 diversity of the species. The branches of the tree containing these isolates are colored in grey. The
701 four main clades in the Iceland population: clades 1, 2, 3, and 4, are colored magenta, green, red
702 and blue respectively. For each of the clades the distribution of the pairwise SNPs distance of
703 calculated for the core genome are displayed (top graph in panel), and also the geographic
704 distribution of the isolates' origins within Iceland (bottom image in panel).

705

706 **Figure 2.** Representation of shared spacer sequences within the CRISPR region of epidemic
707 ST209 isolates relative to the 435 strain that was recovered from Sweden in 2008. The left panel
708 represents the ML phylogeny of *S. zooepidemicus*. The presence of shared spacer sequences,
709 from 1 to 44 is indicated by colored boxes, with the date of isolation, farm and horse number
710 shown in columns to the right. The location of isolates recovered from farm 5 is highlighted by the
711 orange colored entries in the right panel. The isolate from farm 17 that contained an additional
712 three spacers in its genome sequence is highlighted in green in the right-hand panel.

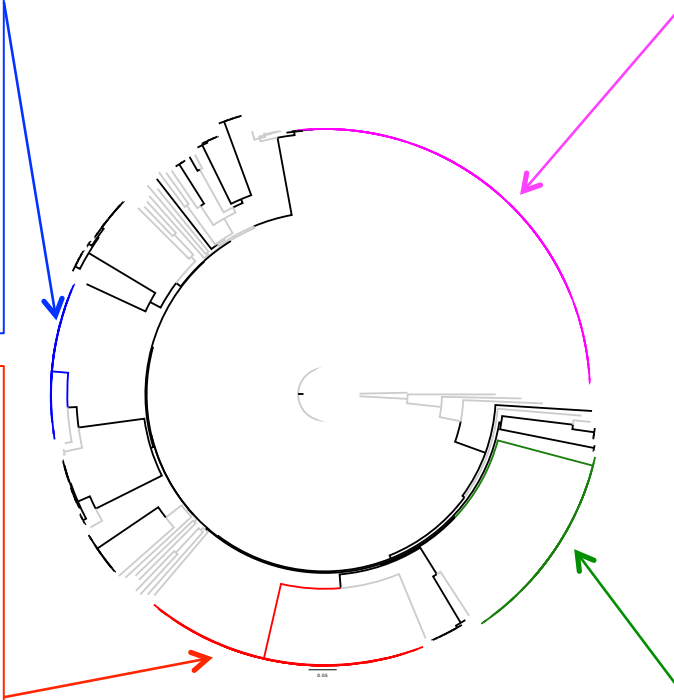
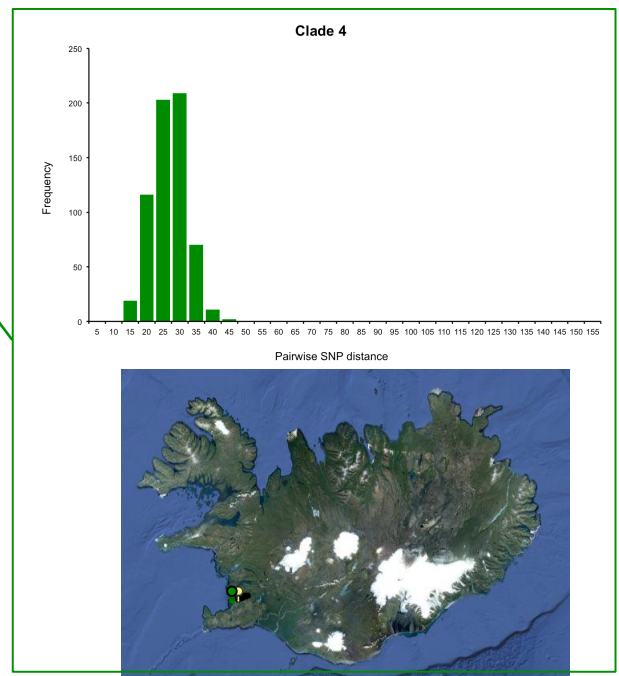
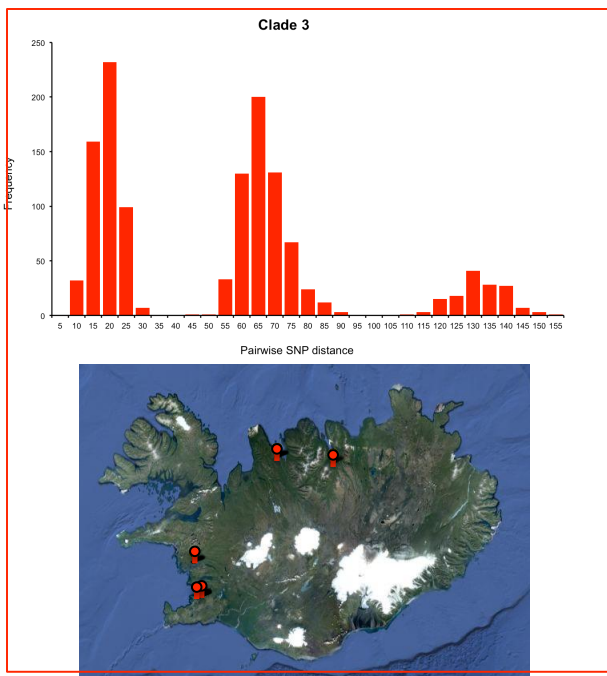
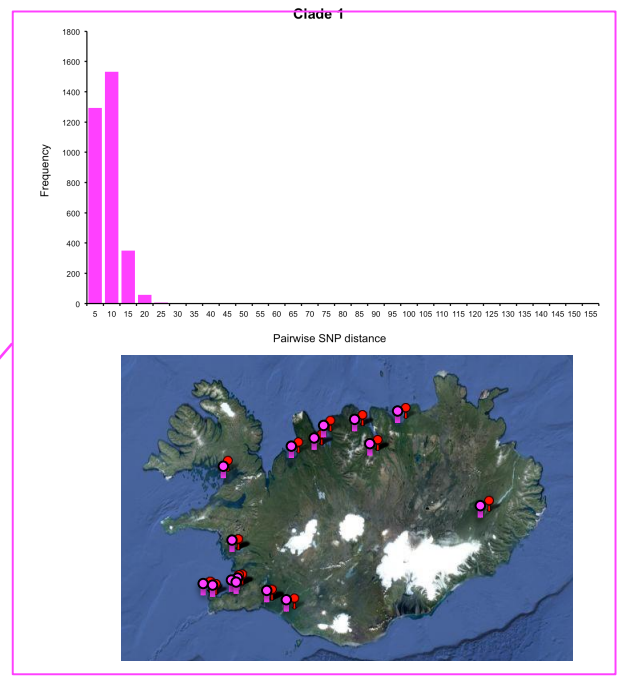
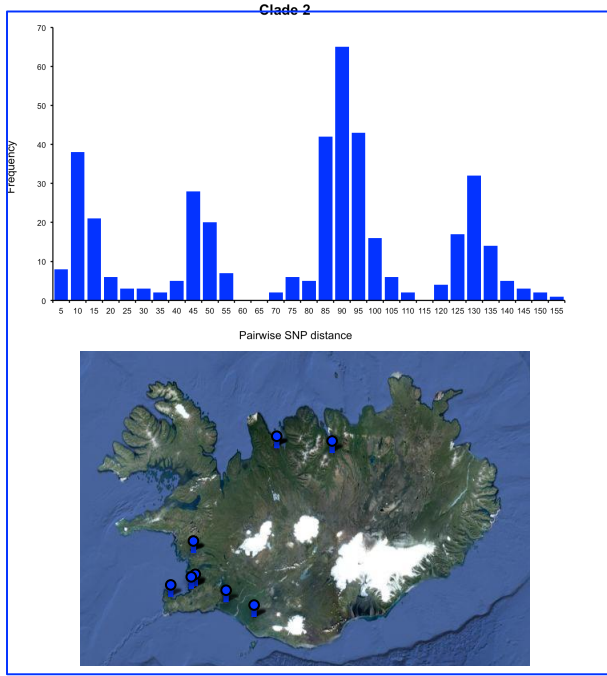
713

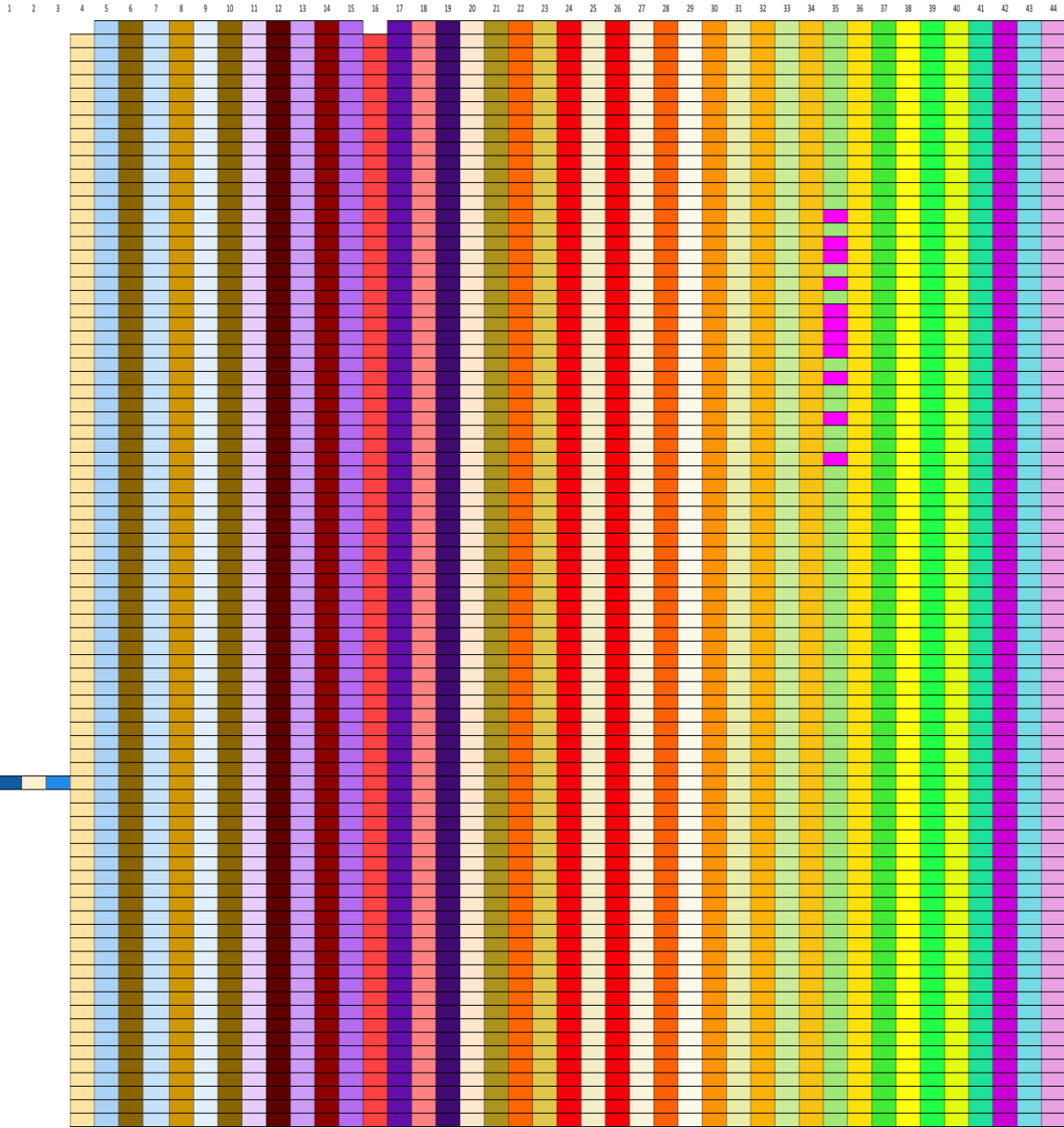
714 **Figure 3.** Depiction of the movement of infected horses from the primary center of transmission
715 (red) and the secondary centers of transmission (turkey) to affected farms (blue) during February
716 and March of 2010.

717

718 **Figure 4.** Bayesian phylogenetic visualization of isolate metadata produced with BEAST. Branch
719 lengths represent time with dates shown beneath the tree. Country, farm and host of origin are
720 indicated in columns adjacent to the tree.

721

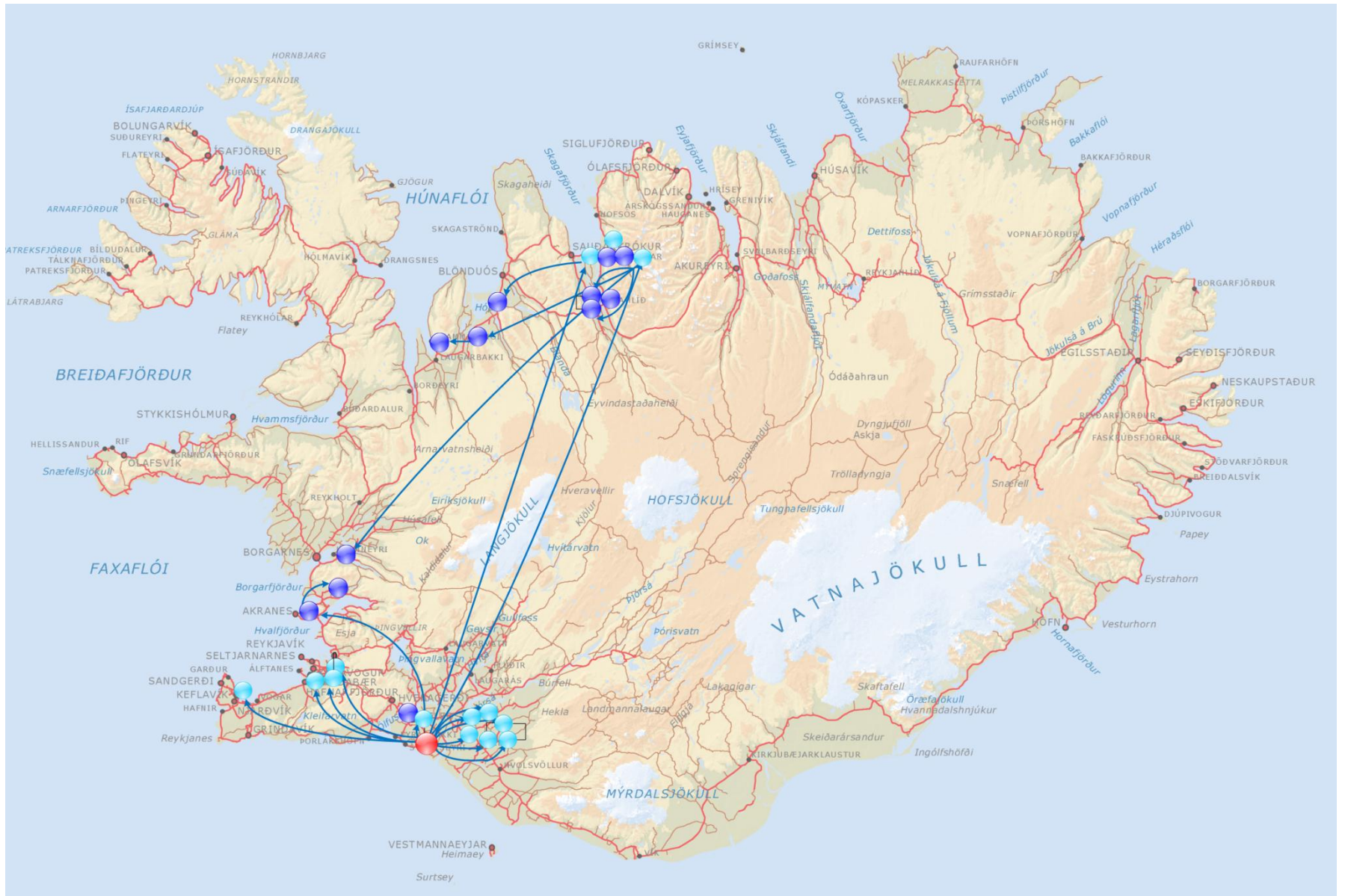


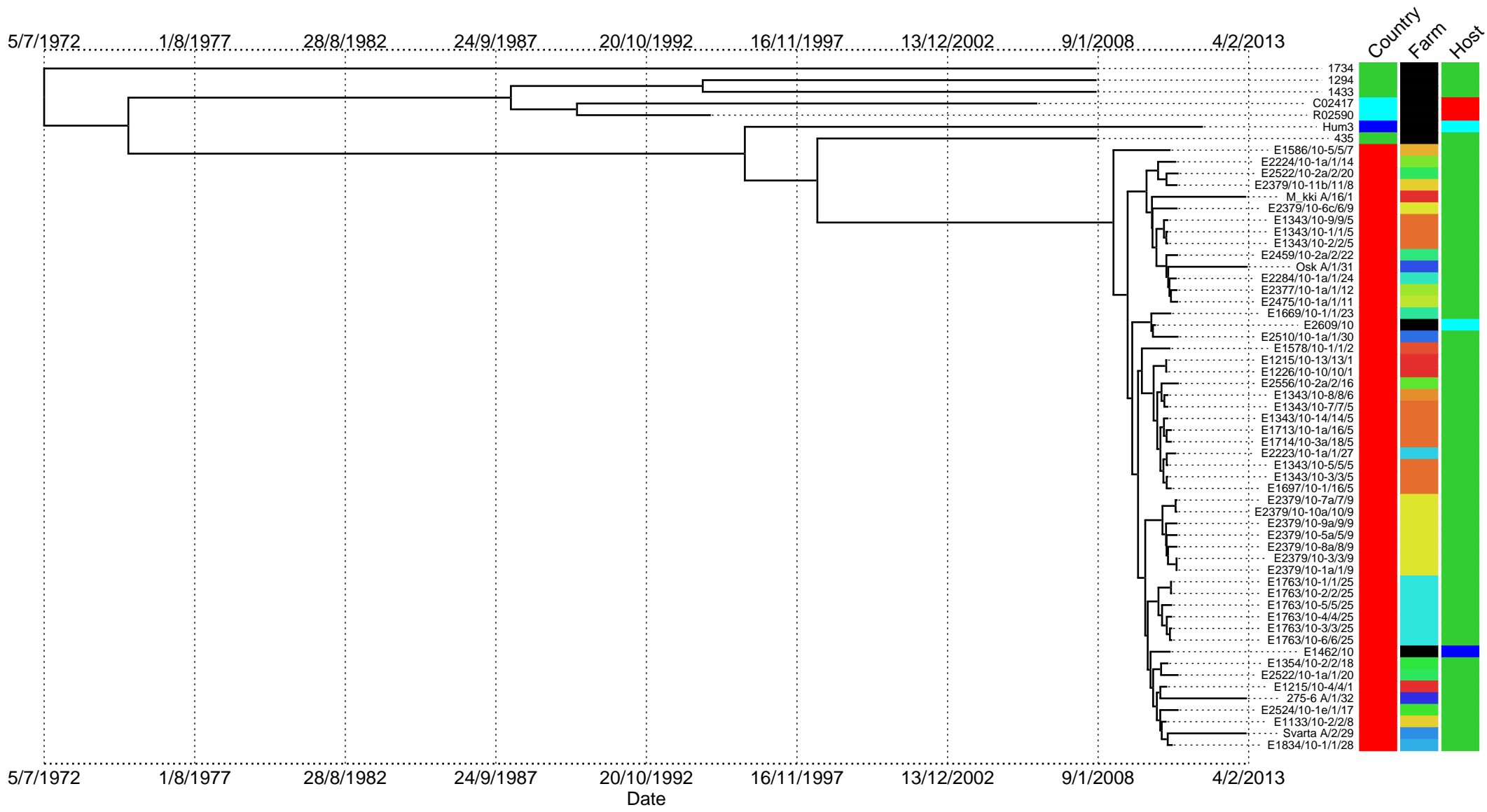


Farm	horse #	Date of sample
Sweden		28/01/2008
27	1	07/09/2010
27	1	07/09/2010
7	5	29/06/2010
30	1	05/10/2010
28	1	26/07/2010
20	1	06/10/2010
20	1	06/10/2010
20	1	06/10/2010
18	2	02/06/2010
2	1	28/06/2010
	Human	2010
1	10	20/05/2010
12	1	22/09/2010
5	16	14/07/2010
27	1	07/09/2010
5	5	02/06/2010
5	16	12/07/2010
16	2	08/10/2010
5	16	14/07/2010
12	1	22/09/2010
5	16	12/07/2010
5	18	14/07/2010
5	18	14/07/2010
5	18	14/07/2010
16	2	08/10/2010
5	14	02/06/2010
12	1	22/09/2010
6	8	02/06/2010
5	7	02/06/2010
1	13	19/05/2010
12	1	22/09/2010
5	3	02/06/2010
12	1	22/09/2010
24	1	13/09/2010
24	1	13/09/2010
11	1	30/09/2010
9	6	22/09/2010
24	1	13/09/2010
14	1	07/09/2010
20	2	06/10/2010
20	2	06/10/2010
20	2	06/10/2010
8	11	22/09/2010
8	11	22/09/2010
11	1	30/09/2010
11	1	30/09/2010
5	9	02/06/2010
5	2	02/06/2010
25	3	19/07/2010
25	5	19/07/2010
25	1	19/07/2010
25	2	19/07/2010
25	4	19/07/2010
25	6	19/07/2010
1	4	19/05/2010
17	1	06/10/2010
23	1	08/07/2010
30	1	05/10/2010
5	1	02/06/2010
22	2	29/09/2010
22	2	29/09/2010
22	2	29/09/2010
8	2	10/05/2010
	Feline	2010
9	10	22/09/2010
9	10	22/09/2010
9	7	22/09/2010
9	8	22/09/2010
9	5	22/09/2010
9	10	22/09/2010
9	1	22/09/2010
9	3	22/09/2010
9	1	22/09/2010
9	1	22/09/2010
9	8	22/09/2010
9	8	22/09/2010
9	9	22/09/2010
9	7	22/09/2010
9	7	22/09/2010
9	5	22/09/2010
9	5	22/09/2010

435

E2223/10-1b
 E2223/10-1c
 E1586/10-5
 E2510/10-1a
 E1834/10-1
 E2522/10-1
 E2522/10-1b
 E2522/10-1c
 E1354/10-2
 E1578/10-1
 E2609/10
 E1226/10-10
 E2377/10-1b
 E1713/10-1b
 E2223/10-1a
 E1343/10-5
 E1697/10-1
 E2556/10-2b
 E1713/10-1a
 E2377/10-1d
 E1700/10-1
 E1714/10-3a
 E1714/10-3c
 E1714/10-3b
 E2556/10-2a
 E1343/10-14
 E2377/10-1e
 E1343/10-8
 E1343/10-7
 E1215/10-13
 E2377/10-1c
 E1343/10-3
 E2377/10-1a
 E2284/10-1b
 E2284/10-1a
 E2475/10-1b
 E2379/10-6c
 E2284/10-1c
 E2224/10-1a
 E2522/10-2c
 E2522/10-2b
 E2522/10-2a
 E2379/10-11b
 E2379/10-11c
 E2475/10-1a
 E2475/10-1c
 E1343/10-9
 E1343/10-2
 E1763/10-3
 E1763/10-5
 E1763/10-1
 E1763/10-2
 E1763/10-4
 E1763/10-6
 E1215/10-4
 E2524/10-1e
 E1669/10-1
 E2510/10-1b
 E1343/10-1
 E2459/10-2b
 E2459/10-2c
 E2459/10-2a
 E1133/10-2
 E1462/10
 E2379/10-10a
 E2379/10-10c
 E2379/10-7c
 E2379/10-8b
 E2379/10-5c
 E2379/10-10b
 E2379/10-1b
 E2379/10-3
 E2379/10-1c
 E2379/10-1a
 E2379/10-8a
 E2379/10-8c
 E2379/10-9a
 E2379/10-7a
 E2379/10-7b
 E2379/10-5a
 E2379/10-5b





Key (Country): Finland Iceland Sweden UK

Key (Farm): 1 2 5 6 7 8 9 11 12 14 16 17 18 20 22 23 24 25 27 28 29 30 31 32

Key (Host): cat cow horse human