Equine Veterinary Journal WILEY

ORIGINAL ARTICLE

Conservation of vaccine antigen sequences encoded by sequenced strains of *Streptococcus equi* subsp. *equi*

Sara Frosth[1] | Ellen Ruth A. Morris[2] | Hayley Wilson[3] | Lars Frykberg[1] | Karin Jacobsson[1] | Julian Parkhill[3] | Jan-Ingmar Flock[4,5] | Tim Wood[5] | Bengt Guss[1] | David M. Aanensen[6] | Ashley G. Boyle[7] | Miia Riihimäki[8] | Noah D. Cohen[2] | Andrew S. Waller[1,5]

[1]Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Uppsala, Sweden

[2]Department of Large Animal Clinical Sciences, College of Veterinary Medicine & Biomedical Sciences, Texas A&M University, Texas, USA

[3]University of Cambridge, Cambridge, UK

[4]Department of Microbiology, Tumour and Cell Biology, Karolinska Institutet, Stockholm, Sweden

[5]Intervacc AB, Stockholm, Sweden

[6]Big Data Institute, Li Ka Shing Centre for Health Information and Discovery, Nuffield Department of Medicine, University of Oxford, Oxford, UK

[7]Department of Clinical Studies New Bolton Center, School of Veterinary Medicine, University of Pennsylvania, Philadelphia, Pennsylvania, USA

[8]Department of Clinical Sciences, Swedish University of Agricultural Sciences, Uppsala, Sweden

Correspondence
Sara Frosth, Department of Biomedical Sciences and Veterinary Public Health, Swedish University of Agricultural Sciences, Uppsala, Sweden.
Email: sara.frosth@slu.se

Summary

Background: *Streptococcus equi* subspecies *equi* (*S equi*) is the cause of Strangles, one of the most prevalent diseases of horses worldwide. Variation within the immuno-dominant SeM protein has been documented, but a new eight-component fusion protein vaccine, Strangvac, does not contain live *S equi* or SeM and conservation of the antigens it contains have not been reported.

Objective: To define the diversity of the eight Strangvac antigens across a diverse *S equi* population.

Study design: Genomic description.

Methods: Antigen sequences from the genomes of 759 *S equi* isolates from 19 countries, recovered between 1955 and 2018, were analysed. Predicted amino acid sequences in the antigen fragments of SEQ0256(Eq5), SEQ0402(Eq8), SEQ0721(EAG), SEQ0855(SclF), SEQ0935(CNE), SEQ0999(IdeE), SEQ1817(SclI) and SEQ2101(SclC) in Strangvac and SeM were extracted from the 759 assembled genomes and compared.

Results: The predicted amino acid sequences of SclC, SclI and IdeE were identical across all 759 genomes. CNE was truncated in the genome of five (0.7%) isolates. SclF was absent from one genome and another encoded a single amino acid substitution. EAG was truncated in two genomes. Eq5 was truncated in four genomes and 123 genomes encoded a single amino acid substitution. Eq8 was truncated in three genomes, one genome encoded four amino acid substitutions and 398 genomes encoded a single amino acid substitution at the final amino acid of the Eq8 antigen fragment.

Therefore, at least 1579 (99.9%) of 1580 amino acids in Strangvac were identical in 743 (97.9%) genomes, and all genomes encoded identical amino acid sequences for at least six of the eight Strangvac antigens.

Main limitations: Three hundred and seven (40.4%) isolates in this study were recovered from horses in the UK.

Conclusions: The predicted amino acid sequences of antigens in Strangvac were highly conserved across this collection of *S equi*.

KEYWORDS
genetic conservation, horse, *S equi*, strangles, vaccine antigens

1 | INTRODUCTION

The host-restricted pathogen *Streptococcus equi* subspecies *equi* (*S equi*) causes the disease strangles, which is one of the most prevalent infectious diseases of horses worldwide.[1–5] *S equi* infects horses via the nose to nose contact with diseased animals, ingestion of contaminated food or water or contact with other fomites.[4] Once within the mouth or nose, *S equi* attaches to and invades the lingual and palatine tonsils *via* an array of cell surface receptors,[6–8] before transitioning to the lymph nodes of the head and neck within a few hours of infection.[9] Within the lymph nodes, *S equi* uses a multitude of immune-evasion strategies to neutralise the effects of the innate immune system and establish infection.[10–14] Active recruitment of neutrophils to infected lymph nodes and failure of the immune system to kill *S equi* results in enlargement of lymph nodes and formation of abscesses, which may be refractory to antibiotic treatment.[4] Treatment of cases with antibiotics may also impede the development of a humoral immune response[15] and select for resistant strains.[16–18]

Abscesses within the lymph nodes eventually burst, draining from the head of affected animals, releasing *S equi* into the local environment and providing an opportunity for transmission to naïve animals.[4] Most horses then recover from the disease. However, some recovered horses remain persistently infected "carriers" of *S equi*, providing long-term potential for transmission of *S equi* through contact with naïve animals.[4,19–21] Therefore, biosecurity, diagnostic testing and vaccination measures to prevent the establishment of infection are of vital importance to control the spread of this disease.[4,22–25]

The development of rapid methodologies for the generation and analysis of genome sequence information has shed unprecedented light on the evolution and transmission of *S equi*.[17–19,26,27] In a recent study, the genomes of a population of 670 isolates from 19 countries were found to cluster into six Bayesian analysis of population structure (BAPS) groups, based upon polymorphisms within a core genome comprising 1286 loci, with an average of 90.9 single nucleotide polymorphisms separating each group.[19] The calculated mean substitution rate per core genome site per year was 5.22×10^{-7}, suggesting that, within a core genome of 1.8 Mb,[26] approximately one base substitution accumulates per year. This dates the emergence of the contemporary strains of *S equi* to around the time of World War I.[17,28] However, genes encoding antigenic proteins may be subject to much greater selective pressure, leading to much more rapid genetic change. For example, genetic variation in the gene encoding the immunodominant SeM protein has been exploited in strain typing schemes, with 242 alleles described to date.[11,17,19,29–31] The SeM protein is used within several cell extract and live vaccines that target *S equi*.[32–36] However, the multicomponent fusion protein vaccine Strangvac does not contain SeM and instead uses eight different *S equi* proteins, the diversity of which has not been described previously.[37] Thus, in this study, we examined the diversity of Strangvac antigens by combining and analysing three published genome collections[18,19,27] that together comprised 759 *S equi* isolates from 19 countries, which were recovered from horses between 1955 and 2018.

2 | MATERIALS AND METHODS

2.1 | Study collection

The origins of the genomes of the 759 isolates (three collections) of *S equi* analysed in this study are listed in Table S1. One of the collections, of 54 genomes of *S equi* isolates recovered from outbreaks in the USA, has been used previously to study the transmission of *S equi* in Texas and Kentucky (Collection A).[18] The largest collection of genomes used here, from 670 *S equi* isolates,[19] included the complete genome sequence of *Se*4047, which was used as the reference genome in this study as it was the first *S equi* genome sequenced to completion.[26] Of these, 224 genomes were from a study of the effects of persistent infection[17] and 445 genomes from a study of the international transmission of *S equi* (Collection B).[19] Finally, a collection of 35 genomes of isolates recovered from the USA (n = 21) or Sweden (n = 14) that have been used previously to examine the effects of persistent infection was included (Collection C).[27]

Together, the combined collection of 759 *S equi* genomes originated from isolates recovered between 1955 and 2018 from 19 countries that comprised: Argentina (n = 15), Australia (n = 26), Belgium (n = 14), Canada (n = 1), France (n = 14), Germany (n = 12), Ireland (n = 16), Israel (n = 14), Japan (n = 12), Kuwait (n = 1), the Netherlands (n = 17), New Zealand (n = 4), Poland (n = 11), Saudi Arabia (n = 5), Spain (n = 2), Sweden (n = 26), the United Arab Emirates UAE (n = 119), the UK (n = 307) and the USA (n = 143) (Table S1).

2.2 | Phylogenomic analysis

Genome assemblies for all 759 isolates were uploaded into the Pathogenwatch bioresource for *S equi* (https://cgps.gitbook.io/pathogenwatch/) and phylogenetic reconstruction of the combined populations was generated as described previously.[19] The collection in Pathogenwatch can be accessed at https://pathogen.watch/collection/j3qp5viupjjh-antigen-variation. A curated set of 1286 loci in the core genome of the *Se*4047 reference, excluding the mobile genetic elements (φSeq1, φSeq2, φSeq3, φSeq4, ICE*Se1* or ICE*Se2*), insertion sequences and sortase-processed proteins, was used for typing purposes.[17,26] Alleles of loci for which multiple copies were encoded within the *S equi* genome, including *hasC1* and *hasC2*, were also omitted.[26] BLAST matches of the 1286 loci across each genome relative to the core genome of the *Se*4047 reference were extracted and aligned using MAFFT,[38] and a database of the core genome segments with a per cent identity was constructed. Hits below 80% core gene length or identity were removed as fragments. Each specific combination of substitutions within the core genome loci relative to the *Se*4047 reference[26] was assigned an allele. Indels were excluded from further analysis, as they are often the result of assembly or sequencing error.[38] The variant sites between each pair of assemblies were then used to construct dendrograms using the APE package.[39] The resulting tree was midpoint-rooted using the phangorn package.[40] The phylogenetic reconstruction and associated metadata were visualised using Microreact[41] and can be viewed at https://microreact.org/project/8knxebFjP96CrKjv3uA9xY.

2.3 | Extraction of antigen sequences

The DNA sequence encoding the antigen fragments of SEQ0256(Eq5), SEQ0402(Eq8), SEQ0721(EAG), SEQ0855(SclF), SEQ0935(CNE), SEQ0999(IdeE), SEQ1817(SclI) and SEQ2101(SclC) in Strangvac were extracted in silico from the genome sequencing data using the Ridom SeqSphere+ version 7.0.5 software (Ridom GmbH, Münster, Germany) and the MLST and SeM-typing alleles using the BIGSdb software.[30,42–44] Differences in the predicted amino acid sequences of the antigen sequences and the N-terminus SeM were identified using MEGA X.[45]

3 | RESULTS

3.1 | The combined collections of *S equi* genomes clustered into six BAPS groups

Core genome multilocus sequence typing (cgMLST) analysis of genetic variation was in agreement with Mitchell et al,[19] differentiating the combined collection of 759 isolates of *S equi* into the same six BAPS groups (Figure 1). The genomes of isolates from the USA reported in Morris et al[18,27] clustered into BAPS groups 1 and 6, while the genomes of isolates in Sweden clustered into BAPS group

2. Overall, the combined collection contained 138 genomes in BAPS1, 339 genomes in BAPS2, 31 genomes in BAPS3, 56 genomes in BAPS4, 123 genomes in BAPS5 and 72 genomes in BAPS6. The BAPS4 group was split into two paraphyletic subgroups, containing 29 and 27 genomes. The increased diversity within BAPS4 could be an effect of recombination, creating similarities between lineages that are detected by the BAPS algorithm, but which are not represented in the consensus phylogenetic tree.

3.2 | The genomes of *S equi* isolates generally clustered with others from the same geographical regions

The genomes from Collections A and C clustered closest to those of isolates recovered from the same geographical region (the USA or Sweden) (https://microreact.org/project/8knxebFjP96CrKjv3uA9xY). However, isolates, ER14_125 and ER14_140, recovered from horses in Texas during 2014, which clustered most closely with a group of isolates from the UAE and Saudi Arabia recovered between 2013 and 2015 (Figure 2). These data provide further evidence of a link between outbreaks in the UAE, Saudi Arabia and the USA that may have been associated with the international transport of horses.

3.3 | The predicted amino acid sequences of the antigens targeted by Strangvac were conserved across the combined collections of *S equi*

Strangvac contains eight antigens based on the 1866 strain of *S equi*, which was recovered from a horse with strangles in Hälsingland, Sweden, in 2000 (Table S1). The core genome of strain 1866 clustered into BAPS2, which was the most prevalent type of *S equi* causing strangles in horses within Europe in Collections B and C. The core genome of strain 4047 (UK4047 in Table S1), which served as a reference genome for this study and was used as the challenge strain in Strangvac vaccine trials, clustered into BAPS5, the second most prevalent type of *S equi* recovered from European horses.

Sequence analysis confirmed that the amino acid sequences of SclC, SclI and IdeE in Strangvac were identical to the predicted amino acid sequences of the homologous proteins encoded by all 759 (100%) genomes in the combined collection (Figure 3, Table 1 and Table S1).

The predicted amino acid sequence of CNE was identical to that used in Strangvac in 754 (99.3%) of the 759 *S equi* genomes in the combined collection. The exceptions were five isolates recovered from horses in Glastonbury or Scone in Australia between 2014 and 2015, which clustered into BAPS3 and encoded truncated forms of CNE (Figure 3, Table 1 and Table S1). Therefore, the amino acid sequence of CNE in Strangvac was fully conserved in all 434 European genomes and all 144 North American genomes.

The predicted amino acid sequence of SclF was identical to that used in Strangvac in 757 (99.7%) of the 759 *S equi* genomes in the combined collection (Figure 3, Table 1 and Table S1). The BAPS1

BAPS5
(n = 123)
Collection A = 0
Collection B = 123
Collection C = 0

BAPS4
(n = 29)
Collection A = 0
Collection B = 29
Collection C = 0

BAPS3
(n = 31)
Collection A = 0
Collection B = 31
Collection C = 0

BAPS6
(n = 72)
Collection A = 34
Collection B = 29
Collection C = 9

BAPS4
(n = 27)
Collection A = 0
Collection B = 27
Collection C = 0

BAPS2
(n = 339)
Collection A = 0
Collection B = 325
Collection C = 14

BAPS1
(n = 138)
Collection A = 20
Collection B = 106
Collection C = 12

18.553
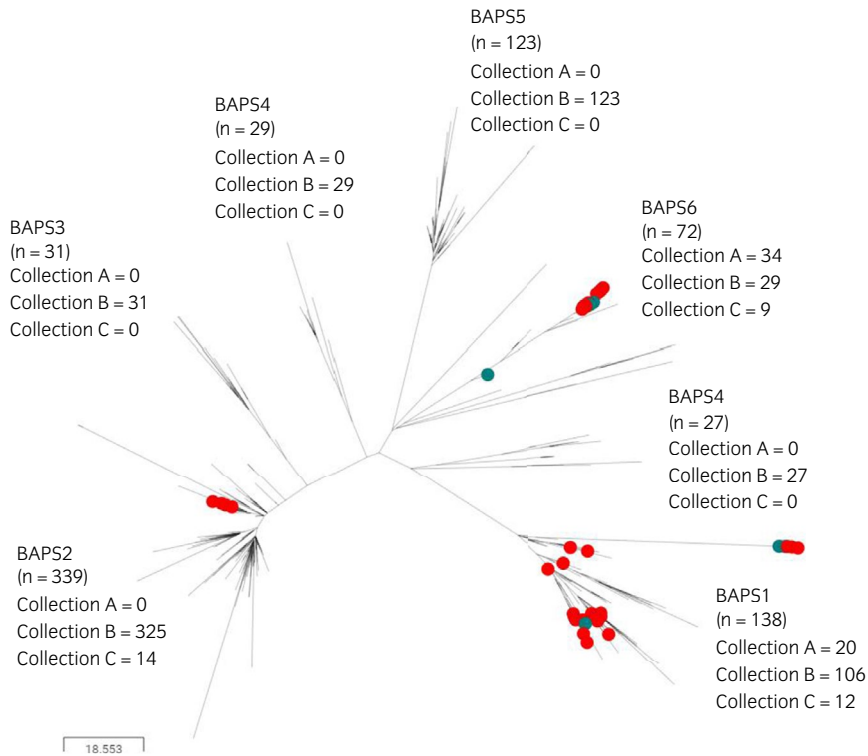
FIGURE 1 Distribution of isolates from Collections A, B and C into the six Bayesian analysis of population structure (BAPS) groups. Midpoint-rooted phylogenetic reconstruction of the *Streptococcus equi* population visualised in Microreact. The dendrogram was constructed from pairwise cgMLST scores using the APE package.[39] The resulting tree was midpoint-rooted using the phangorn package.[40] The scale bar relates to horizontal branch length and indicates the number of cgSNPs proposed to have occurred on the branches. Green and red circles indicate Collections A[18] and C,[27] respectively. Unlabelled branches indicate Collection B[19]

strain NZLU, which was recovered from a horse in New Zealand in 2011, contained a deletion of the gene encoding SclF, while the BAPS1 strain UAE3229, which was recovered from a horse in Dubai during 2009, contained a nonsynonymous amino acid change from proline to leucine at amino acid position 85 (the final amino acid position of the SclF fragment in Strangvac). Therefore, the amino acid sequence of SclF in Strangvac was fully conserved in all 434 European genomes and all 144 North American genomes.

The predicted amino acid sequence of EAG was identical to that used in Strangvac in 757 (99.7%) of the 759 *S equi* genomes in the combined collection (Figure 3, Table 1 and Table S1). The BAPS1 strain USA07_22, which was recovered from a horse in Indiana in 2009, and the BAPS2 strain UAE0015_3 (ST-179), which was recovered from a horse in Dubai in 2014, contained a truncation in the predicted EAG amino acid sequence. Therefore, the amino acid sequence of EAG in Strangvac was fully conserved in all 434 European genomes and 143 (99.3%) of 144 North American genomes.

The predicted amino acid sequence of Eq5 was identical to that used in Strangvac in 632 (83.3%) of the 759 *S equi* genomes in the combined collection (Figure 3, Table 1 and Table S1). The 123 strains in BAPS5 contained isoleucine to leucine substitution at amino acid position 201, which is conservative and not predicted to alter significantly the antigenicity of Eq5. In support of this, the *Se*4047 challenge strain, which encodes the same I^{201} to L amino acid variation, is clustered into BAPS5. Thus, the *Se*4047 strain provided a heterologous challenge and represented a diverse and relevant challenge strain to measure the protection conferred against the wider population of *S equi* circulating in horses. Eq5 was truncated in four BAPS4 strains recovered from horses in Argentina between 2013 and 2015. Therefore, the amino acid sequence of Eq5 in Strangvac was fully conserved in 313 (72.1%) of 434 European genomes and all 144 North American genomes.

The predicted amino acid sequence of Eq8 was identical to that used in Strangvac in 358 (47.2%) of the 759 *S equi* genomes in the combined collection. BAPS1 strain USA05_40, which was recovered from a horse in Iowa in 2003, contained four nonsynonymous amino acid changes (E^{212}–G, E^{214}–G, A^{218}–D and L^{223}–I) in the Eq8 antigen relative to Strangvac. Three strains, UAE0605 (BAPS2), Arg0107 (BAPS4) and UK501771 (BAPS6), each contained a different truncation of Eq8. Finally, all 338 remaining BAPS2 strains, all 31 BAPS3 strains and 29 (52%) of 56 BAPS4 strains contained a histidine to tyrosine amino acid substitution at position 225 (Figure 3, Table 1 and Table S1). This is the final amino acid of the Eq8 fragment in Strangvac and is not predicted to affect significantly the antigenicity of this protein relative to Strangvac. Therefore, the amino acid sequence of Eq8 in Strangvac was fully conserved in 161 (37.1%) of 434 European genomes and 141 (97.9%) of 144 North American genomes.

Overall, all 759 genomes of *S equi* encoded at least six antigens identical to those in Strangvac and 1579 (99.9%) of the 1580 amino acids within the eight antigens in Strangvac were identical to those encoded by 744 (98.0%) of the 759 isolates of *S equi*.
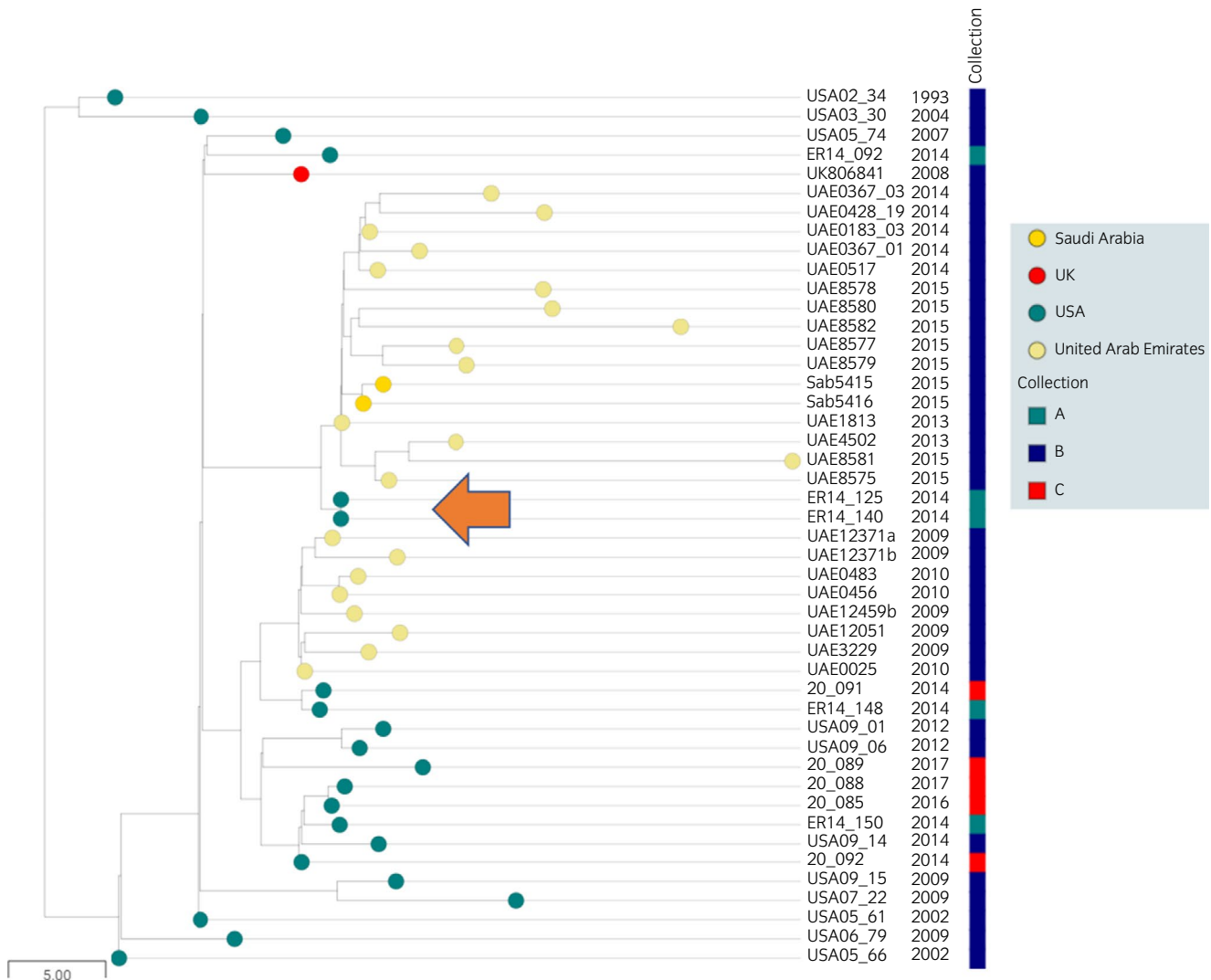
3.4 | The predicted amino acid sequences of SeM varied across the combined collections of *S equi*

A total of 111 different SeM alleles were identified across the combined collection of 759 *S equi* genomes (Table S1 and Table S2). Analysis of the predicted 109 amino acids encoded by the 5' variable region of the SeM gene of these 759 genomes revealed that 44 (40%)

FIGURE 2 Relationships of Bayesian analysis of population structure (BAPS) group 1 isolates from outbreaks in the USA, UAE and Saudi Arabia. Midpoint-rooted phylogenetic reconstruction of a BAPS1 subgroup of the *Streptococcus equi* population visualised in Microreact. The dendrogram was constructed from pairwise cgMLST scores using the APE package.[39] The resulting tree was midpoint-rooted using the phangorn package.[40] The scale bar relates to horizontal branch length and indicates the number of cgSNPs proposed to have occurred on the horizontal branches. Coloured circles indicate the country from which the isolates originated, as indicated in the key. Coloured bars indicate Collections A,[18] B[19] and C,[27] as indicated in the key. Date of isolation is shown. The arrow indicates the position of the genomes of isolates ER14_125 and ER14_140

of these amino acid positions had nonsynonymous variants generating 81 different amino acid changes across this population (Table S2).

Eleven (1.4%) of the 759 genomes in the combined collection, comprising one BAPS4 isolate and 10 BAPS5 isolates, encoded a SeM antigen identical to the consensus SeM1 allele. These 11 genomes included the isolate TW928 (NLTW928), which is used in the Equilis StrepE vaccine, the Arnica strain (NLArnica), which has been used in challenge studies to measure the efficacy of the Equilis StrepE vaccine, and two isolates, UK250223 and UK406006, recovered, respectively, from the lip and submandibular lymph node abscesses in horses post-vaccination with Equilis StrepE.

Forty-five (5.9%) genomes contained SeM alleles that differed from the consensus allele by one amino acid. Of the remaining genomes, 160 (21.1%), 429 (56.5%), 50 (6.6%), 16 (2.1%), 29 (3.8%), 3 (0.4%) and 1 (0.1%) encoded SeM alleles that differed from the

consensus allele by 2, 3, 4, 5, 6, 8 and 11 amino acids, respectively. Fifteen (2%) of the strains in the combined collection of *S equi* contained deletions or insertions in the 5' region of the SeM gene, leading to the production of a truncated product (Table S1). On average, the 744 *S equi* isolates which contained a full-length SeM gene encoded a product that contained 2.9 amino acid changes (2.7%) relative to the consensus 109 amino acids of the N-terminal region of SeM.

4 | DISCUSSION

The collection of *S equi* genomes used in this study is the most comprehensive to date. However, large proportions of the genomes were from isolates recovered from horses residing in the UK and the USA (40.4% and 18.8%, respectively). Continued
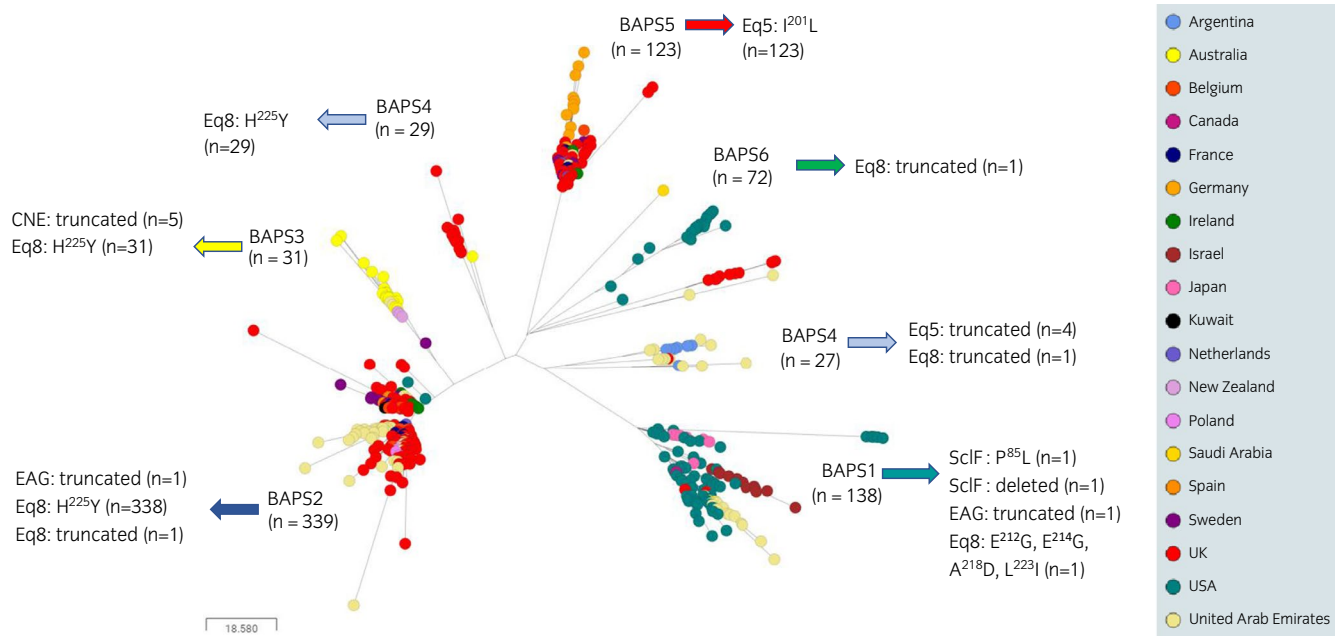
FIGURE 3 Variation in Strangvac antigens encoded by the six Bayesian analysis of population structure (BAPS) groups. Midpoint-rooted phylogenetic reconstruction of the *Streptococcus equi* population visualised in Microreact. The dendrogram was constructed from pairwise cgMLST scores using the APE package.[39] The resulting tree was midpoint-rooted using the phangorn package.[40] The scale bar relates to horizontal branch length and indicates the number of cgSNPs proposed to have occurred on the branches. Coloured circles indicate the country from which the isolates originated, as illustrated in the key. The variation in the Strangvac antigens encoded within each of the BAPS clusters is highlighted

sequencing of isolates with more diverse geographical origins would shed further light on the diversity of *S equi* genomes and the antigens they encode.

Despite the majority of the isolates originating from a small number of countries, the diversity identified within the encoded SeM protein was very high. The SeM protein is an important immunodominant component of *S equi* and the selective pressure exerted on this protein is evident in the variation in amino acid sequences encoded by different strains.[17,19,29–31,46] This selective pressure indicates the potential for SeM-containing vaccines to exert a protective effect.[47,48] However, sequencing of the SeM gene has shown that variants emerge over time within outbreaks and individual animals, suggesting that variation may permit evasion, at least in part, of an immune response directed only at this protein.[30,49–51] Eighty-six different amino acid variations, occurring at 44 (40.4%) of the 109 positions within the variable region, were identified across the population of *S equi* in the combined collection studied here. On average, each isolate contained 2.9 (2.7%) amino acid changes relative to the 109 amino acids within the SeM1 consensus sequence.

Much less variation in the 1580 amino acids encoded by the eight antigens included in Strangvac was identified, with all 759 isolates encoding at least six antigens that were identical to those in Strangvac. Seven different amino acid variations, across seven (0.4%) of the 1580 positions within the encoded antigens in Strangvac were identified in the population of *S equi* in the combined collection studied here. On average, each isolate contained 0.7 (0.04%) amino acid changes within

the 1580 amino acids encoded within the Strangvac antigens. Furthermore, 744 (98.0%) of the 759 genomes of *S equi* encoded eight antigens that had none or one amino acid variation, providing evidence that these antigens are well-conserved across this population of *S equi*. Further analysis of the gene sequences encoding these antigens is important to determine whether vaccination with Strangvac increases the selective pressure exerted by the immune response. However, the inclusion of eight antigens in Strangvac may reduce the selective pressure exerted on any one individual gene, which could drive the selection of variant strains.

Fourteen (1.8%) of the 759 *S equi* genomes in the combined collection contained a mutation that led to truncation or deletion in one of the genes encoding the eight antigens in Strangvac, likely preventing the production of the full-length encoded protein. One of these isolates, Arg0107, which was recovered from a horse in Buenos Aires in 2014, encoded CNE, SclC, SclF, SclI, EAG and IdeE that were identical to those in Strangvac, but the genes encoding Eq5 and Eq8 both encoded truncated products. Decay of the *S equi* genome of isolates recovered from persistently infected horses has been described previously.[17] Loss (and mutation) of SeM, equibactin and the *has* locus has been suggested to confer a fitness advantage to strains causing persistent infection in the guttural pouch, but such mutations are believed to be disadvantageous during acute infection.[17,49] In support of this, in a previous study on ponies and horses, an *eqbE* deletion mutant was shown to be significantly attenuated in ponies and none of the isolates recovered from horses with acute

TABLE 1 Variation in the eight Strangvac antigens in the combined population of 759 *Streptococcus equi* isolates

Strangvac antigen	Number of genomes in each BAPS group with antigen sequences identical to those in Strangvac	Number of genomes in each BAPS group with variant antigen sequences
CNE	BAPS1 (n = 138, 100%) BAPS2 (n = 339, 100%) BAPS3 (n = 26, 84%) BAPS4 (n = 56, 100%) BAPS5 (n = 123, 100%) BAPS6 (n = 72, 100%) Total (n = 754, 99.3%)	BAPS3 (n = 5, 16%) Truncation (AusZola, AusPenn, AusJoy, AusGlaston, AusZadiym)
SclC	Total (n = 759, 100%)	
SclF	BAPS1 (n = 136, 98.6%) BAPS2 (n = 339, 100%) BAPS3 (n = 31, 100%) BAPS4 (n = 56, 100%) BAPS5 (n = 123, 100%) BAPS6 (n = 72, 100%) Total (n = 757, 99.7%)	BAPS1 (n = 1, 0.7%) P^{85} to L (UAE3229) BAPS1 (n = 1, 0.7%) Deletion (NZLU)
SclI	Total (n = 759, 100%)	
EAG	BAPS1 (n = 137, 99.3%) BAPS2 (n = 338, 99.7%) BAPS3 (n = 31, 100%) BAPS4 (n = 56, 100%) BAPS5 (n = 123, 100%) BAPS6 (n = 72, 100%) Total (n = 757, 99.7%)	BAPS1 (n = 1, 0.7%) Truncation (USA07_22) BAPS2, ST-179 (n = 1, 0.3%) Truncation (UAE0015_1)
Eq5	BAPS1 (n = 138, 100%) BAPS2 (n = 339, 100%) BAPS3 (n = 31, 100%) BAPS4 (n = 52, 92.9%) BAPS6 (n = 72, 100%) Total (n = 632, 83.3%)	BAPS4 (n = 4, 7.1%) Truncation (Arg0003, Arg0021, Arg0106, Arg0107) BAPS5 (n = 123, 100%) I^{201} to L
Eq8	BAPS1 (n = 137, 99.3%) BAPS4 (n = 26, 46%) BAPS5 (n = 123, 100%) BAPS6 (n = 71, 98.6%) Total (n = 358, 47.2%)	BAPS1, (n = 1, 0.7%) E^{212} to G, E^{214} to G, A^{218} to D and L^{223} to I (USA05_40) BAPS2, ST-179 (n = 1, 0.3%) Truncation (UAE0605) BAPS4 (n = 1, 2%) Truncation (Arg0107) BAPS6 (n = 1, 1.4%) Truncation (UK501771) BAPS2 (n = 338, 99.7%) H^{225} to Y BAPS3 (n = 31, 100%) H^{225} to Y BAPS4 (n = 29, 52%) H^{225} to Y
IdeE	Total (n = 759, 100%)	

Abbreviation: BAPS, Bayesian analysis of population structure groups 1-6.

disease contained deletions in the *has* locus, providing evidence that some decay events result in evolutionary dead ends.[17] Interestingly, an outbreak of unusually mild disease in yearling horses on a farm in Germany was linked to the presence of a 61 bp deletion that truncated SEQ0402(Eq8).[52] Therefore, mutations that lead to truncation in the genes encoding antigens targeted by Strangvac may reduce the virulence of *S equi*.

Immune responses towards the antigens within Strangvac conferred significant levels of protection to vaccinated ponies against infection with *S equi* following experimental challenge.[37] Therefore, the high level of conservation of the proteins used in Strangvac, relative to the variation identified in SeM was surprising. The variability of Strangvac antigens may be restricted by functional constraints that do not affect variation in SeM to the same degree.[46] The SeM protein of *S equi* is immunodominant[53] and it is evident that selective pressure on SeM leads to variation in this important protein.[46] However, vaccines based on SeM failed to confer protection to horses.[48] We speculate that the immunodominance of SeM during natural infection may divert selective pressure exerted by the equine immune response away from targeting other, less immunogenic, antigens the response to which may be more protective.

5 | CONCLUSION

The predicted amino acid sequences of antigens in Strangvac were highly conserved across this collection of *S equi*. At least 1579 (99.9%) of 1580 amino acids in Strangvac were identical in 743 (97.9%) genomes, and all genomes encoded identical amino acid sequences for

at least six of the eight Strangvac antigens. Our data suggest that the immune response generated following vaccination of horses with Strangvac will target the corresponding antigens produced by *S equi* regardless of geographical origin. Continued monitoring of the population of *S equi* for decay and variation of the genes encoding the antigens used in Strangvac will provide important insights into the continued evolution of *S equi* in the context of an adapted immune response post-vaccination.

CONFLICT OF INTERESTS

A.S. Waller, T. Wood and J-I. Flock are employed by Intervacc AB. B. Guss is a board member of Intervacc AB. S. Frosth, L. Frykberg and K. Jacobsson are funded by a research grant from Intervacc AB. None of the authors has any other financial or personal relationships that could inappropriately influence or bias the content of the paper.

AUTHOR CONTRIBUTIONS

All authors contributed to data analysis and interpretation and manuscript preparation and approved the final manuscript.

ETHICAL ANIMAL RESEARCH

This study analysed publicly available genome sequencing data from previously published studies. No experimental animals or clinical cases were sampled during this study.

INFORMED CONSENT

Not applicable.

PEER REVIEW

The peer review history for this article is available at https://publons.com/publon/10.1111/evj.13552.

DATA AVAILABILITY STATEMENT

The Illumina sequences used in this study were deposited previously at the National Center for Biotechnology Information (NCBI) under the accession numbers SUB6350545,[18] PRJEB38019[19] and PRJNA704656.[27] The Pathogenwatch web bioresource can be accessed at https://pathogen.watch/collection/j3qp5viupjjh-antigen-variation. Data can be visualised in Microreact at https://microreact.org/project/8knxebFjP96CrKjv3uA9xY.

ORCID

Sara Frosth https://orcid.org/0000-0002-4080-5183
Ellen Ruth A. Morris https://orcid.org/0000-0002-7957-4642
Hayley Wilson https://orcid.org/0000-0002-9914-2601
Julian Parkhill https://orcid.org/0000-0002-7069-5958
Bengt Guss https://orcid.org/0000-0002-7044-0701
David M. Aanensen https://orcid.org/0000-0001-6688-0854
Ashley G. Boyle https://orcid.org/0000-0003-4325-1005
Noah D. Cohen https://orcid.org/0000-0002-0384-2903
Andrew S. Waller https://orcid.org/0000-0002-7111-9549

REFERENCES

1. Laing G, Christley R, Stringer A, Aklilu N, Ashine T, Newton R, et al. Respiratory disease and sero-epidemiology of respiratory pathogens in the working horses of Ethiopia. Equine Vet J. 2018;50:793–9. https://doi.org/10.1111/evj.12834

2. Libardoni F, Machado G, Gressler LT, Kowalski AP, Diehl GN, dos Santos LC, et al. Prevalence of *Streptococcus equi* subsp. *equi* in horses and associated risk factors in the State of Rio Grande do Sul, Brazil. Res Vet Sci. 2016;104:53–7. https://doi.org/10.1016/j.rvsc.2015.11.009

3. Surveillance of equine strangles: a new initiative. Vet Rec. 2019;184:342–4. https://doi.org/10.1136/vr.l1188

4. Boyle AG, Timoney JF, Newton JR, Hines MT, Waller AS, Buchanan BR *Streptococcus equi* Infections in horses: guidelines for treatment, control, and prevention of strangles-revised consensus statement. J Vet Intern Med. 2018;32:633–47. https://doi.org/10.1111/jvim.15043

5. Parkinson NJ, Robin C, Newton JR, Slater J, Waller AS. Molecular epidemiology of strangles outbreaks in the UK during 2010. Vet Rec. 2011;168:666. https://doi.org/10.1136/vr.d1485

6. Lannergård J, Flock M, Johansson S, Flock JI, Guss B. Studies of fibronectin-binding proteins of *Streptococcus equi*. Infect Immun. 2005;73:7243–51. https://doi.org/10.1128/IAI.73.11.7243-7251.2005

7. Lannergård J, Frykberg L, Guss B. CNE, a collagen-binding protein of *Streptococcus equi*. FEMS Microbiol Lett. 2003;222:69–74. https://doi.org/10.1016/S0378-1097(03)00222-2

8. Harrington DJ, Sutcliffe IC, Chanter N. The molecular basis of *Streptococcus equi* infection and disease. Microbes Infect. 2002;4:501–10. https://doi.org/10.1016/s1286-4579(02)01565-4

9. Timoney JF, Kumar P. Early pathogenesis of equine *Streptococcus equi* infection (strangles). Equine Vet J. 2008;40:637–42. https://doi.org/10.2746/042516408x322120

10. Lannergård J, Guss B. IdeE, an IgG-endopeptidase of *Streptococcus equi* ssp. *equi*. FEMS Microbiol Lett. 2006;262:230–5. https://doi.org/10.1111/j.1574-6968.2006.00404.x

11. Boschwitz JS, Timoney JF. Inhibition of C3 deposition on *Streptococcus equi* subsp. *equi* by M protein: a mechanism for survival in equine blood. Infect Immun. 1994;62:3515–20. https://doi.org/10.1128/IAI.62.8.3515-3520.1994

12. Artiushin SC, Timoney JF, Sheoran AS, Muthupalani SK. Characterization and immunogenicity of pyrogenic mitogens SePE-H and SePE-I of *Streptococcus equi*. Microb Pathog. 2002;32:71–85. https://doi.org/10.1006/mpat.2001.0482

13. Paillot R, Robinson C, Steward K, et al. Contribution of each of four Superantigens to *Streptococcus equi*-induced mitogenicity, gamma interferon synthesis, and immunity. Infect Immun. 2010;78:1728–39. https://doi.org/10.1128/IAI.01079-09

14. Lopez-Alvarez MR, Salze M, Cenier A, Robinson C, Paillot R, Waller AS. Immunogenicity of phospholipase A2 toxins and their role in *Streptococcus equi* pathogenicity. Vet Microbiol. 2017;204:15–9. https://doi.org/10.1016/j.vetmic.2017.04.002

15. Pringle J, Storm E, Waller A, Riihimäki M. Influence of penicillin treatment of horses with strangles on seropositivity to *Streptococcus equi* ssp. *equi*-specific antibodies. J Vet Intern Med. 2020;34:294–9. https://doi.org/10.1111/jvim.15668

16. Fonseca JD, Mavrides DE, Morgan AL, Na JG, Graham PA, McHugh TD. Antibiotic resistance in bacteria associated with equine respiratory disease in the United Kingdom. Vet Rec. 2020;187:189. https://doi.org/10.1136/vr.105842

17. Harris SR, Robinson C, Steward KF, et al. Genome specialization and decay of the strangles pathogen, *Streptococcus equi*, is driven by persistent infection. Genome Res. 2015;25:1360–71. https://doi.org/10.1101/gr.189803.115

18. Morris ERA, Hillhouse AE, Konganti K, Wu J, Lawhon SD, Bordin AI, et al. Comparison of whole genome sequences of *Streptococcus equi* subsp. *equi* from an outbreak in Texas with isolates from within the region, Kentucky, USA, and other countries. Vet Microbiol. 2020;243:108638. https://doi.org/10.1016/j.vetmic.2020.108638

19. Mitchell C, Steward KF, Charbonneau ARL, Walsh S, Wilson H, Timoney JF, et al. Globetrotting strangles: the unbridled national and international transmission of *Streptococcus equi* between horses. Microb Genom. 2021;7(3):mgen000528. https://doi.org/10.1099/mgen.0.000528

20. Newton JR, Verheyen K, Talbot NC, Timoney JF, Wood JLN, Lakhani KH, et al. Control of strangles outbreaks by isolation of guttural pouch carriers identified using PCR and culture of *Streptococcus equi*. Equine Vet J. 2000;32:515–26. https://doi.org/10.2746/042516400777584721

21. Verheyen K, Newton JR, Talbot NC, de Brauwere MN, Chanter N. Elimination of guttural pouch infection and inflammation in asymptomatic carriers of *Streptococcus equi*. Equine Vet J. 2000;32:527–32. https://doi.org/10.2746/042516400777584703

22. Robinson C, Steward KF, Potts N, Barker C, Hammond T-A, Pierce K, et al. Combining two serological assays optimises sensitivity and specificity for the identification of *Streptococcus equi* subsp. *equi* exposure. Vet J. 2013;197:188–91. https://doi.org/10.1016/j.tvjl.2013.01.033

23. Webb K, Barker C, Harrison T, Heather Z, Steward KF, Robinson C, et al. Detection of *Streptococcus equi* subspecies *equi* using a triplex qPCR assay. Vet J. 2013;195:300–4. https://doi.org/10.1016/j.tvjl.2012.07.007

24. Båverud V, Johansson SK, Aspán A. Real-time PCR for detection and differentiation of *Streptococcus equi* subsp. *equi* and *Streptococcus equi* subsp. *zooepidemicus*. Vet Microbiol. 2007;124:219–29. https://doi.org/10.1016/j.vetmic.2007.04.020

25. Waller AS. New perspectives for the diagnosis, control, treatment, and prevention of strangles in horses. Vet Clin North Am Equine Pract. 2014;30:591–607. https://doi.org/10.1016/j.cveq.2014.08.007

26. Holden MTG, Heather Z, Paillot R, Steward KF, Webb K, Ainslie F, et al. Genomic evidence for the evolution of *Streptococcus equi*: host restriction, increased virulence, and genetic exchange with human pathogens. PLoS Pathog. 2009;5:e1000346. https://doi.org/10.1371/journal.ppat.1000346

27. Morris ERA, Boyle AG, Riihimäki M, Aspán A, Anis E, Hillhouse AE, et al. Differences in the genome, methylome, and transcriptome do not differentiate isolates of *Streptococcus equi* subsp. *equi* from horses with acute clinical signs from isolates of inapparent carriers. PLoS One. 2021;16:e0252804. https://doi.org/10.1371/journal.pone.0252804

28. Waller AS. Strangles: a pathogenic legacy of the war horse. Vet Rec. 2016;178:91–2. https://doi.org/10.1136/vr.i123

29. Anzai T, Kuwamoto Y, Wada R, et al. Variation in the N-terminal region of an M-like protein of *Streptococcus equi* and evaluation of its potential as a tool in epidemiologic studies. Am J Vet Res. 2005;66:2167–71. https://doi.org/10.2460/ajvr.2005.66.2167

30. Kelly C, Bugg M, Robinson C, et al. Sequence variation of the SeM gene of *Streptococcus equi* allows discrimination of the source of strangles outbreaks. J Clin Microbiol. 2006;44:480–6. https://doi.org/10.1128/JCM.44.2.480-486.2006

31. Tartor YH, El-Naenaeey EY, Gharieb NM, Ali WS, Ammar AM. Novel *Streptococcus equi* strains causing strangles outbreaks in Arabian horses in Egypt. Transbound Emerg Dis. 2020;67:2455–66. https://doi.org/10.1111/tbed.13584

32. Boyle AG, Mitchell C, Stefanovski D, Waller AS. Horses vaccinated with live attenuated intranasal strangles vaccine seroconvert to SEQ2190 and SeM. Equine Vet J. 2021;1–7. Epub ahead of print. https://doi.org/10.1111/evj.13443

33. El-Hage CM, Bannai H, Wiethoelter AK, Firestone SM, Heislers CM, Allen JL, et al. Serological responses of Australian horses using a commercial duplex indirect ELISA following vaccination against strangles. Aust Vet J. 2019;97:220–4. https://doi.org/10.1111/avj.12825

34. Boyle AG, Sweeney CR, Kristula M, Boston R, Smith G. Factors associated with likelihood of horses having a high serum *Streptococcus equi* SeM-specific antibody titer. J Am Vet Med Assoc. 2009;235:973–7. https://doi.org/10.2460/javma.235.8.973

35. Jacobs AA, Goovaerts D, Nuijten PJ, Theelen RP, Hartford OM, Foster TJ. Investigations towards an efficacious and safe strangles vaccine: submucosal vaccination with a live attenuated *Streptococcus equi*. Vet Rec. 2000;147:563–7.

36. Hoffman AM, Staempfli HR, Prescott JF, Viel L. Field evaluation of a commercial M-protein vaccine against *Streptococcus equi* infection in foals. Am J Vet Res. 1991;52:589–92.

37. Robinson C, Waller AS, Frykberg L, Flock M, Zachrisson O, Guss B, et al. Intramuscular vaccination with Strangvac is safe and induces protection against equine strangles caused by *Streptococcus equi*. Vaccine. 2020;38:4861–8. https://doi.org/10.1016/j.vaccine.2020.05.046

38. Katoh K, Asimenos G, Toh H. Multiple alignment of DNA sequences with MAFFT. Methods Mol Biol. 2009;537:39–64. https://doi.org/10.1007/978-1-59745-251-9_3

39. Paradis E, Claude J, Strimmer K. APE: analyses of phylogenetics and evolution in R language. Bioinformatics. 2004;20:289–90. https://doi.org/10.1093/bioinformatics/btg412

40. Schliep KP. phangorn: phylogenetic analysis in R. Bioinformatics. 2011;27:592–3. https://doi.org/10.1093/bioinformatics/btq706

41. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, et al. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. Microb Genom. 2016;2:e000093. https://doi.org/10.1099/mgen.0.000093

42. Webb K, Jolley KA, Mitchell Z, Robinson C, Newton JR, Maiden MCJ, et al. Development of an unambiguous and discriminatory multilocus sequence typing scheme for the *Streptococcus zooepidemicus* group. Microbiol (Reading). 2008;154:3016–24. https://doi.org/10.1099/mic.0.2008/018911-0

43. Jolley KA, Maiden MC. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595. https://doi.org/10.1186/1471-2105-11-595

44. Jünemann S, Sedlazeck FJ, Prior K, Albersmeier A, John U, Kalinowski J, et al. Updating benchtop sequencing performance comparison. Nat Biotechnol. 2013;31:294–6. https://doi.org/10.1038/nbt.2522

45. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018;35:1547–9. https://doi.org/10.1093/molbev/msy096

46. Timoney JF, DeNegri R, Sheoran A, Forster N. Affects of N-terminal variation in the SeM protein of *Streptococcus equi* on antibody and fibrinogen binding. Vaccine. 2010;28:1522–7. https://doi.org/10.1016/j.vaccine.2009.11.064

47. Meehan M, Nowlan P, Owen P. Affinity purification and characterization of a fibrinogen-binding protein complex which protects mice against lethal challenge with *Streptococcus equi* subsp. *equi*. Microbiol (Reading). 1998;144:993–1003. https://doi.org/10.1099/00221287-144-4-993

48. Sheoran AS, Artiushin S, Timoney JF. Nasal mucosal immunogenicity for the horse of a SeM peptide of *Streptococcus equi* genetically coupled to cholera toxin. Vaccine. 2002;20:1653–9. https://doi.org/10.1016/s0264-410x(01)00488-1

49. Chanter N, Talbot NC, Newton JR, Hewson D, Verheyen K *Streptococcus equi* with truncated M-proteins isolated from outwardly healthy horses. Microbiol (Reading). 2000;146:1361–9. https://doi.org/10.1099/00221287-146-6-1361

50. Riihimäki M, Aspán A, Ljung H, Pringle J. Long term dynamics of a *Streptococcus equi* ssp *equi* outbreak, assessed by qPCR and culture

and seM sequencing in silent carriers of strangles. Vet Microbiol. 2018;223:107–12. https://doi.org/10.1016/j.vetmic.2018.07.016

51. Lindahl S, Söderlund R, Frosth S, Pringle J, Båverud V, Aspán A. Tracing outbreaks of *Streptococcus equi* infection (strangles) in horses using sequence variation in the seM gene and pulsed-field gel electrophoresis. Vet Microbiol. 2011;153:144–9. https://doi.org/10.1016/j.vetmic.2011.03.027

52. Tscheschlok L, Venner M, Steward K, Bose R, Riihimäki M, Pringle J. Decreased clinical severity of strangles in weanlings associated with restricted seroconversion to optimized *Streptococcus equi* ssp *equi* assays. J Vet Intern Med. 2018;32:459–64. https://doi.org/10.1111/jvim.15037

53. Galan JE, Timoney JF. Mucosal nasopharyngeal immune responses of horses to protein antigens of *Streptococcus equi*. Infect Immun. 1985;47:623–8. https://doi.org/10.1128/IAI.47.3.623-628.1985

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.