Streptococcus zooepidemicus: Commensal or Pathogen?

Andrew S. Waller, BSc, PhD*; and Hayley Wilson, BSc, PhD

1. Introduction

Streptococcus equi subspecies zooepidemicus is the bacterium most frequently recovered from the oropharynx of horses.1,2 The prevalence of S. zooepidemicus in samples recovered from healthy horses might suggest that this organism may not be a causal agent of disease. However, growing evidence supports an important role for S. zooepidemicus in a wide array of diseases of horses and other animals. This review examines the population structure of S. zooepidemicus, informed by the analysis of genome sequencing data, to shed new light on the role of this subspecies of bacteria as a causative agent of disease in animals, including humans.

2. The Many Lives of S. zooepidemicus

S. zooepidemicus is a β-hemolytic Gram-positive Lancefield group C Streptococcus that can use lactose and sorbitol but fails to ferment trehalose.3,4 Although S. zooepidemicus is frequently recovered from healthy horses,5 its presence is associated with respiratory disease in Thoroughbred racehorses,6,7 uterine infections in mares,8,9 and ulcerative keratitis.10 It is also associated with disease in a wide range of other animal hosts, including cattle,11 sheep,12–15 pigs,16–19 dogs,20,21 and humans.22–26 Streptococcus equi subspecies equi, a biovar of S. zooepidemicus that cannot use lactose or sorbitol,4 is the causative agent of strangles in horses.27 Strangles is one of the most frequently diagnosed disease of horses, which is characterized by fever and the abscessation of lymph nodes in the head and neck.28–31 The often obvious clinical signs and high prevalence of strangles led to this biovar lending its name to the species as a whole following its identification in 1888.27 However, S. equi actually clusters within the much broader group of S. zooepidemicus strains as just one of a wide range of variants.32 Therefore, these data32 and additional analyses using multilocus sequence typing (MLST)33 and genome sequencing data34 detailed below, provide evidence that S. equi evolved from an ancestral strain of S. zooepidemicus and is, in actual fact, a lineage of S. zooepidemicus. An enhanced understanding of the genetic diversity of S. zooepidemicus strains and the ability of particular lineages to cause specific diseases will shed new light on the evolution of this pathogenic group and direct the development of novel diagnostic tests and vaccines.

3. Identification of Different S. zooepidemicus Strain Types

Strains of S. zooepidemicus were originally differentiated using sugar fermentation. The S. equi biovar fails to ferment lactose and sorbitol, facilitating its
differentiation from the wider population of *S. zooepidemicus*. However, the resolution that biochemical methods provides is extremely limited and can be influenced by a variety of loss-of-function events that might suggest strains are related when in fact the same phenotypic properties arose independently. Therefore, although permitting the identification of *S. equi*, this methodology provided insufficient resolution to determine if certain strain types of *S. zooepidemicus* were more capable of causing specific types of disease in particular host animals. Over time, several other methods have been developed to improve the ability to differentiate strain types. MLST assigns specific allele numbers to each unique 400- to 500-base pair nucleotide sequence of internal fragments of seven housekeeping genes. Each combination of seven allele numbers is then assigned a specific sequence type (ST) number. Much more variation can be detected by MLST, and as the majority of nucleotide variation does not lead to changes in amino acid sequence, different alleles are usually not subjected to functional selective pressure. Furthermore, the nucleotide sequence data that are generated by MLST are fully portable and can easily be compared between different laboratories via electronic databases available on the Internet. These databases are therefore a powerful resource with which to conduct global epidemiological studies. The MLST scheme for *S. zooepidemicus* uses fragments of the seven housekeeping genes **areC**, **nrdE**, **proS**, **spi**, **tdk**, **tpi**, and **yqiL**. To date, 1,519 different isolates of *S. zooepidemicus* or *S. equi* from 21 different countries have been uploaded onto the online database (Fig. 1). A total of 865 different alleles have been assigned, generating 437 STs comprised of unique allele combinations (*Streptococcus zooepidemicus* [PubMLST, last accessed 27th January 2021]). Therefore, the MLST platform permits the identification of hundreds of different STs within populations of *S. zooepidemicus* in order to identify disease-causing strains.

4. **Nonequine Transmission of *S. zooepidemicus* STs**

*S. zooepidemicus* is frequently recovered from horses both with and without clinical signs of disease, providing challenges in establishing a causal role in disease. However, *S. zooepidemicus* is less prevalent in other animal species, providing an opportunity to examine the potential of this agent to cause disease. The phylogenetic reconstruction shown in Fig. 2, which was generated using the concatenated DNA sequences of the seven MLST alleles, highlights the diversity of the population of *S. zooepidemicus* as measured by MLST. *S. zooepidemicus* is regularly isolated from cases of disease, including mastitis, wound infections, respiratory disease, and uterine infections in the ruminant population of Spain. MLST analysis identified that the ST-236 group of *S. zooepidemicus* was the dominant strain type recovered from goats and sheep in Spain and that none of the isolates within this ST had previously been recovered from horses. Thirty-two (82%) of the 39 isolates that are ST-236 or a single locus variant (ST-266 and ST-272) were recovered from cases of mastitis, suggesting that this subgroup of *S. zooepidemicus* may have evolved so as to specifically cause disease in sheep and goats rather than in horses. Genome sequencing of the C7 strain of ST-236 identified a novel phosphoenolpyruvate sugar phosphotransferase system, which enables this group of *S. zooepidemicus* to use arbutin. Arbutin is a common component of animal foodstuffs, and the ability to use arbutin may confer a selective advantage to strains infecting animals, the diet of which contains this sugar. Another subgroup of *S. zooepidemicus*, ST-72, was recovered from an outbreak of acute
nephritis in humans residing in Nova Serrana, Brazil, between 1997 and 1998. Two hundred and fifty-three people were affected, of which seven required dialysis and three patients died. The outbreak was linked to the consumption of unpasteurized dairy products. Interestingly, although two of the five ST-72 isolates listed in the MLST database were recovered from the respiratory tract of horses in the United Kingdom and the United Arab Emirates, another two isolates were recovered from humans. One of these isolates was recovered from the cerebrospinal fluid of a man in Barnsley, UK, during 2006 and the other from a case of nephritis in a dairy farmer in Northallerton, UK, during 1983, which was 14 years before the outbreak of nephritis in Brazil. Therefore, this subgroup of \textit{S. zooepidemicus} may have an increased ability to cross species boundaries and cause rare, but serious infection of humans. A double locus variant of ST-72, ST-10, was responsible for a large outbreak of acute fatal hemorrhagic pneumonia in kenned dogs residing in the United Kingdom between 2000 and 2002, providing further evidence of the ability of this subgroup to infect multiple mammalian hosts. ST-194 is a further subgroup of \textit{S. zooepidemicus} that has been responsible for severe disease in nonequine hosts. This ST was first recovered from pigs in Sichuan Province, China, during 1975, which suffered from painful swelling of the joints, respiratory disease, and diarrhea, leading to the death of over 300,000 pigs in a 2-week period. Cases of \textit{S. zooepidemicus} ST-194 infection in pigs in China has continued, with the genome of a further strain, CY, which was isolated from a pig in Nanjing Province during 1998 and sequenced in 2014. In 2019, thousands of cases of septicemia and death in pigs due to infection with a ST-194 strain of \textit{S. zooepidemicus} occurred in North America, providing evidence that the ST-194 strain had spread beyond...
Asia. Interestingly, the MLST database identifies a human case of septicemia due to an ST-194 strain of *S. zooepidemicus* that occurred in a woman during 2001 after her return to the United Kingdom from abroad. These findings highlight the significant threat already posed by the ST-194 strain of *S. zooepidemicus* to the pig industry in Asia and North America and the potential risk of zoonotic transmission to humans. Given the severity of disease in affected animals, the screening of pigs for *S. zooepidemicus* infection prior to their import into other parts of the world would be a prudent measure.

5. Respiratory Infections of Horses Associated with *S. zooepidemicus*

Respiratory disease affects a large proportion of young horses around the world, with an incidence of approximately 5 cases per 100 horses per month. The mean duration of clinical signs for each episode is around 8 weeks, and the disease often reoccurs in individual animals, causing considerable disruption to the equine industry. Although several bacterial species, including *S. zooepidemicus*, *Streptococcus pneumoniae*, *Actinobacillus* spp., and *Mycoplasma equi* have been associated with respiratory disease in horses, only the prevalence and incidence of *S. zooepidemicus* and *S. pneumoniae* decreased in parallel with respiratory disease and age, consistent with the development of acquired immunity to infection with these pathogens. Recently, cases of respiratory disease in working horses residing in Ethiopia, characterized by coughing, nasal discharge, or altered respiration were significantly more likely to test positive for the presence of *S. zooepidemicus* (odds ratio: 12.4, 95% confidence interval: 3.6 to 42.4), with no evidence for the involvement of viral pathogens. Despite this evidence, the investigation of outbreaks of respiratory disease due to *S. zooepidemicus* has continued to be confounded by the diversity of this subspecies and the ability of different strains to establish persistent infection of the tonsils of recovered horses, which resemble the outbreak strain when using traditional typing methodologies. However, MLST analysis permits strains of *S. zooepidemicus* to be identified at much higher levels of resolution, making it possible to link particular STs with outbreaks of disease. *S. zooepidemicus* strain H70 (ST-1), which was recovered from the nasopharynx of a Thoroughbred racehorse in Newmarket, UK, during 2000, is typical of respiratory tract isolates being recovered from horses in the United Kingdom and New Zealand. The genome sequence of this strain was completed, highlighting an array of surface proteins and biochemical pathways that may assist this strain to infect horses. Fifty-five of 71 (77%) strains related to ST-1 were recovered from the respiratory tract of horses, suggesting that this group of strains may be better suited to this niche. In Sweden, a herd of 17 Icelandic horses developed respiratory disease during 2009. Clinical signs included an elevated body temperature, nasal discharge, coughing, and lethargy. MLST was used to identify that a ST-24 strain of *S. zooepidemicus* was recovered from all horses with signs of respiratory disease, and unrelated strains were isolated from healthy horses on the same farm. Interestingly, the ST-24 strain was recovered from one of the affected horses 8 months postresolution of its respiratory disease, providing evidence that this strain was able to persist, and opening up new opportunities for transmission to naive animals. ST-24 is related genetically to ST-79, ST-84, ST-161, ST-293, ST-339, and ST-418. Fifteen of the 20 (75%) isolates in the ST-24 group that are listed in the MLST database were recovered from the respiratory tract of horses, including from cases of respiratory disease in the United States during 1986 and 1988. Therefore, the ST-24 group appears to be adept at causing respiratory disease in horses. An outbreak of respiratory disease, characterized by purulent nasal discharge and coughing, was identified in New Caledonia-resident horses between October 1997 and July 1998. Attempts at the time to isolate a causal viral agent or to demonstrate seroconversion to equine influenza virus, equine herpesvirus-1, adenovirus, or rhinovirus were unsuccessful. However, *S. zooepidemicus* was recovered from 80% (25 of 31) of affected animals but only 4% (1 of 25) of healthy horses (*P < 0.0001, two-sided Fisher’s exact test*). An MLST analysis of isolates recovered from affected horses identified a ST-307 strain of *S. zooepidemicus* as the causal agent. ST-307 clusters with ST-3, ST-92, ST-167, ST-249, and ST-369. Seven of the 8 (88%) isolates within these related STs were recovered from cases of respiratory infection in horses, providing evidence that this subgroup of *S. zooepidemicus* is proficient at causing respiratory disease in horses. Although MLST has the ability to resolve different strains, it samples only a tiny portion (0.2%) of the core genome of *S. zooepidemicus*, and so the variation that is measured accrues slowly over many years. In some instances, it is important to be able to resolve differences within a ST in order to determine if the variation within that population of *S. zooepidemicus* arose over many years, indicative of an endemic strain, or whether there is little variation within an ST, which would be typical of an outbreak or epidemic strain. An epidemic of respiratory disease affected almost the entire native Icelandic horse population of 77,000 animals in 2010, resulting in a self-imposed ban on the export of horses and significant economic cost to associated industries. The disease was characterized by clinical signs of a dry cough coexisting with mucopurulent nasal discharge and mild conjunctivitis. Rectal temperature remained normal in most horses, and although morbidity rates approached 100%, mortality was infrequent. The incubation period was between 2 and 4 weeks and the duration of clinical signs varied from 2 to 10 weeks. The rate of spread suggested that a viral agent was responsible, but PCR of nasal swabs...
for viral agents of horses and some common respiratory viruses of humans and other animals were negative. Paired blood samples showed a lack of seroconversion to viral agents. Only equine herpesvirus-4 was inconsistently isolated from small numbers of both healthy and clinically affected horses. In the absence of a viral pathogen, it was noted that \( S. \text{ zooepidemicus} \) was isolated from almost all of the nasal swabs taken from coughing horses and from the diseased tissues of occasional fatal cases. Initial MLST analysis identified that four STs accounted for 198 of 257 (77%) isolates recovered from the Icelandic horse population during the epidemic. ST-306 contained 37 isolates obtained from eight horses at the Icelandic Veterinary Institute at Keldur, Reykjavik. Fifty-two ST-248 strains were recovered from 14 horses and 1 dog residing at 8 different farms. Twenty-eight isolates of ST-246 were obtained from 14 horses residing at 10 farms and 1 human isolate. Finally, 81 isolates of ST-209 were recovered from 45 horses residing at 21 premises across Iceland, 1 human, and 1 cat. Mixed populations of 2 different isolates of \( S. \text{ zooepidemicus} \) were recovered from 13 horses, and 3 different isolates were recovered from 2 horses, highlighting the challenges in differentiating endemic from epidemic strains. Based on these MLST data, the epidemic could have been caused by ST-248, ST-246, or ST-209 strains of \( S. \text{ zooepidemicus} \), and so a whole-genome sequencing approach was used to resolve these STs further. The analysis of genome sequencing data identified that the isolates of ST-248, ST-246, and ST-209 differed from other isolates of their same ST by a maximum of 151, 153, and 25 single nucleotide polymorphisms (SNPs), respectively. Therefore, ST-248 and ST-246 strains had accrued relatively high levels of variation, indicating that these strains were endemic within the Icelandic horse population. The 81 isolates of ST-209 had little variation within their core genome, indicating that they had been transmitted through the Icelandic horse population over a short period of time and were responsible for the epidemic of respiratory disease. The Icelandic-resident horse population is geographically isolated, arising from animals introduced by settlers in the 17th and 18th centuries, with virtually no import of horses for the last 1,000 years. The most likely source of the ST-209 strain was believed to be through the import of contaminated tack. Network analysis identified that the epidemic traced back to yard A, which used a submerged treadmill for the training and rehabilitation of horses from farms located across Iceland. The water used in the treadmill contained no disinfectant and was changed on a once- or twice-weekly basis, providing ideal conditions for the transmission of \( S. \text{ zooepidemicus} \) between the visiting horses. Therefore, it is likely that following entry into Iceland, the rapid epidemic spread of ST-209 was facilitated by exposure of horses from across Iceland to contaminated water in the submerged treadmill. The relatively long incubation period permitted infected horses to return to their original farms before exhibiting clinical signs of disease. On identification of the likely transmission route, the addition of chlorine coupled with regular cleaning and disinfection of water treadmills has been introduced to minimize or eliminate the transmission of \( S. \text{ zooepidemicus} \) or other infectious agents via this route. The epidemic ST-209 was also recovered from a cat and the blood of an Icelandic woman who had suffered a miscarriage. A closely related strain of \( S. \text{ zooepidemicus} \) ST-209, Hum3, was recovered from a psoas abscess in a Finnish man during 2011, indicating that ST-209 strains of \( S. \text{ zooepidemicus} \) have the potential to cause zoonotic disease. Bjornsdottir et al. also identified other cases of human infection caused by ST-246 and ST-2 strains of \( S. \text{ zooepidemicus} \), suggesting that the prevalence of zoonotic transmission of \( S. \text{ zooepidemicus} \) may be underreported.

6. The Future Application of Whole-Genome Sequencing to Track \( S. \text{ zooepidemicus} \) Infection

The analysis reported by Bjornsdtotir et al. was the first to report the application of whole-genome sequencing to elucidate the cause of an outbreak on a national scale by the differentiation of an epidemic strain from strains that were endemic in the population. \( S. \text{ equi} \) isolates are closely related to one another and can currently be differentiated into nine different STs by MLST (\( Strepotococcus \) \( zooepide-\) micus | PubMLST, last accessed 27th January 2021), which cluster together as a subgroup of \( S. \text{ zooepidemi-}\) micus (Fig 2). Therefore, MLST captures insufficient genetic diversity to adequately differentiate the population structure of \( S. \text{ equi} \) and to resolve transmission events. An alternative single locus typing scheme for \( S. \text{ equi} \) measures variation within the 5’ variable region of the SeM gene. This sortase-processed cell surface protein binds to fibrinogen and immunoglobulin to impede the phagocytosis of \( S. \text{ equi} \) by immune cells. However, the utility of SeM typing as an epidemiological tool is limited by homoplasies and the high rate of mutation of SeM. Recently, Mitchell et al. described the application of a novel web-based core genome MLST (cgMLST) bioresource, implemented in Pathogenwatch (https://pathogen.watch/collection/8c022xfgu8v-globetrotting; last accessed 28th January 2021) to determine the population structure of 670 isolates of \( S. \text{ equi} \), which were recovered from horses residing in 19 countries (Fig 3). A total of 2,962 variant sites were identified across the 1,286 loci within the core genome, and each pair of \( S. \text{ equi} \) isolates differed from one another by an average of 69 SNPs across the core genome (range, 0 to 181). The analysis identified numerous examples of closely related strains of \( S. \text{ equi} \) in geographically distant nations, highlighting that the lack of pre-export testing, used routinely for many animal diseases, facilitated the international transmission.
of *S. equi* (Mitchell et al., in press). The Pathogenwatch resource that was developed for *S. equi* is also available for the analysis of *S. zooepidemicus* genomes. This resource enabled the rapid differentiation of 923 isolates of *S. zooepidemicus* based on 145,408 variable sites across 1,286 loci within the core genome (Fig. 4). The *S. zooepidemicus* population differed by an average of 32,954 SNPs across the core genome (range, 0 to 52,715), providing unprecedented capacity to resolve differences between isolates and track the transmission of pathogenic strains as they are transmitted at national and international levels through multiple animal hosts (https://pathogen.watch/).

![Fig. 3. Midpoint rooted phylogenetic reconstruction of the *S. equi* population generated within Pathogenwatch and visualized in Microreact. The dendrogram was constructed from pairwise cgMLST scores using the APE package. The resulting tree was midpoint rooted using the phangorn package. The scale bars relate to branch lengths and indicate the number of core genome SNPs (cgSNPs) that are proposed to have occurred on the branches. Colored circles indicate the country from which the isolates originated as indicated in the key.](image1)

![Fig. 4. Midpoint rooted phylogenetic reconstruction of the *S. zooepidemicus* population generated within Pathogenwatch. The dendrogram was constructed from pairwise cgMLST scores using the APE package. The resulting tree was midpoint rooted using the phangorn package. The scale bars relate to branch lengths and indicate the number of cgSNPs that are proposed to have occurred on the branches. Colored circles indicate the location of STs described in the text.](image2)
7. Summary

The diversity of *S. zooepidemicus* cannot be captured by traditional phenotypic strain typing methods, which were only capable of differentiating *S. equi* subspecies *zooepidemicus* from *S. equi* subspecies *equi*. The failure of these early typing schemes resulted in an assumption that all strains of *S. zooepidemicus* were commensal organisms despite the fact that certain strains can be highly virulent, leading to severe disease and the death of some animals, including humans. The availability of the MLST and Pathogenwatch resources for the identification and tracking of pathogenic strains herald a new era for the study of this neglected pathogen. With these tools, it is now possible to identify and track outbreak and epidemic strains as they are transmitted through and between populations of animals around the world. The finding that some strains such as ST-194, ST-236, and ST-72 are adapted to cause disease in pigs, ruminants, and humans, respectively, and that other strains including ST-1, ST-24, ST-209, and ST-307 may be more capable of causing respiratory disease in horses provides an opportunity to identify genes that influence host and tissue specificity and that are required for *S. zooepidemicus* to cause disease. This knowledge will have enormous potential for the rational development of novel vaccines and therapeutics with which to prevent and treat disease.

Acknowledgments

Dr. Hayley Wilson is funded by a grant from the Petplan Charitable Trust (S19-741-780).

Declaration of Ethics

The Authors have adhered to the Principles of Veterinary Medical Ethics of the AVMA.

Conflict of Interest

Dr. Andrew Waller is employed as Chief Scientific Officer for Intervacc AB, Stockholm, Sweden. Dr. Hayley Wilson has no conflict of interest.

References


