Molecular diversity and antimicrobial susceptibility of *Streptococcus equi* ssp. *equi* isolates from equines

Inaugural Dissertation

submitted to the

Faculty of Veterinary Medicine

in partial fulfillment of the requirements

for the Ph.D.-Degree

of the Faculties of Veterinary Medicine and Medicine

of the Justus Liebig University Giessen

by

Rotinsulu, Dordia Anindita

of

Manado, Indonesia

Giessen, 2023

Printed and published with the support of

the German Academic Exchange Service

Molecular diversity and antimicrobial susceptibility of *Streptococcus equi* ssp. *equi* isolates from equines

Inaugural Dissertation

submitted to the

Faculty of Veterinary Medicine

in partial fulfillment of the requirements

for the Ph.D.-Degree

of the Faculties of Veterinary Medicine and Medicine

of the Justus Liebig University Giessen

by

Rotinsulu, Dordia Anindita

of

Manado, Indonesia

Giessen, 2023

From the Institute for Hygiene and Infectious Diseases of Animals

Director: Prof. Dr. Christa Ewers

of the Faculty of Veterinary Medicine of the Justus Liebig University Giessen

Supervisor, first reviewer, and committee member:	Prof. Dr. Rolf Bauerfeind,
Supervisor:	Prof. Dr. Christa Ewers
Co-supervisor and vice-chair of the committee:	Prof. Dr. Peter Kämpfer
Second (external) reviewer and committee member:	Prof. Dr. Peter Valentin-Weigand
Chair of the committee:	Prof. Dr. Ralph Schermuly
Date of doctoral defense:	April 30, 2024

This work is dedicated to

my beloved parents and family

Declaration

“I declare that I have completed this dissertation single-handedly without the unauthorized help of a second party and only with the assistance acknowledged therein. I have appropriately acknowledged and referenced all text passages that are derived literally from or are based on the content of published or unpublished work of others, and all information that relates to verbal communications. I have abided by the principles of good scientific conduct laid down in the charter of the Justus Liebig University of Giessen in carrying out the investigations described in the dissertation.”

_____

Dordia Anindita Rotinsulu

Giessen, December 28, 2023

List of publications

Parts of this thesis were presented or have been published previously:

Original article

Rotinsulu, D.A., C. Ewers, K. Kerner, A. Amrozi, R. D. Soejoedono, T. Semmler, R. Bauerfeind. 2023. Molecular features and antimicrobial susceptibilities of *Streptococcus equi* ssp. *equi* isolates from strangles cases in Indonesia. Veterinary Sciences 10 (1) : 49. https://doi.org/10.3390/vetsci10010049.

Oral presentations

Rotinsulu, D. A., K. Kerner, E. Prenger-Berninghoff, K. Fey, I. Stamm, A. Lübke-Becker, A. Amrozi, R. D. Soejoedono, T. Semmler, C. Ewers, R. Bauerfeind. 2020. Genetic diversity and antimicrobial susceptibility of *Streptococcus equi* ssp. *equi* isolates from horses. International Giessen Graduate Center for the Life Sciences, 13th Annual Conference. Online, September 29-30, 2020.

Rotinsulu, D. A., K. Kerner, E. Prenger-Berninghoff, K. Fey, I. Stamm, A. Lübke-Becker, A. Amrozi, R. D. Soejoedono, T. Semmler, C. Ewers, R. Bauerfeind. 2021. Analysis of whole-genome sequencing data of *Streptococcus equi* ssp. *equi* isolates from equines. International Giessen Graduate Center for the Life Sciences, 14th Annual Conference. Online, September 29-30, 2021.

Poster presentations

Rotinsulu, D. A., K. Kerner, E. Prenger-Berninghoff, A. Amrozi, R. D. Soejoedono, K. Fey, C. Ewers, R. Bauerfeind. 2019. Characterization of *Streptococcus equi* isolates from horses in Germany and Indonesia. International Giessen Graduate Center for the Life Sciences, 12th Annual Conference. Giessen, September 4-5, 2019.

Rotinsulu, D. A., K. Kerner, E. Prenger-Berninghoff, K. Fey, I. Stamm, A. Lübke-Becker, A. Amrozi, R. D. Soejoedono, T. Semmler, C. Ewers, R. Bauerfeind. 2020. Genetic diversity and antimicrobial resistance patterns of *Streptococcus equi* ssp. *equi* isolates from equines. 4th International Congress of Equine Medicine/DVG Vet Congress 2020. Online, October 16-17, 2020.

Rotinsulu, D. A., K. Kerner, E. Prenger-Berninghoff, K. Fey, I. Stamm, A. Lübke-Becker, A. Amrozi, R. D. Soejoedono, T. Semmler, C. Ewers, R. Bauerfeind. 2021. Molecular features and antimicrobial susceptibility patterns of *Streptococcus equi* ssp. *equi* isolates from equines. *Tagung der Deutschen Veterinärmedizinischen Gesellschaft* (DVG)/German Veterinary Medical Society Conference, Bacteriology and Mycology. Online, June 14-16, 2021.

Rotinsulu, D.A., C. Ewers, K. Kerner, A. Amrozi, R. D. Soejoedono, T. Semmler, R. Bauerfeind. 2023. Molecular features and antimicrobial susceptibilities of *Streptococcus equi* ssp. *equi* isolates from strangles cases in Indonesia. *Tagung der Deutschen Veterinärmedizinischen Gesellschaft* (DVG)/German Veterinary Medical Society Conference, Bacteriology and Mycology. Berlin, May 22-24, 2023.

Table of contents

List of abbreviations

Abbreviations	Meaning
APE	Analysis of phylogenetics and evolution
Aqua dest.	Aqua destilata
Aqua demin.	Aqua demineralisata
BAP	Sheep blood agar plate
BAPS	Bayesian analysis of population structure
BHI	Brain heart infusion
BLAST	Basic local alignment search tool
bp	Base pair
CC	Clonal complex
CDS	Coding sequence
cg-genotype	Core genome genotype
cgMLST	Core genome multilocus sequence typing
cgSNPs	Core genome single nucleotide polymorphisms
CLSI	Clinical and Laboratory Standards Institute
CO_2	Carbon dioxide
COBA	Columbia blood agar supplemented with colistin oxalinic acid
Da	Dalton (unit of molecular mass)
ddH_2O	Double-distilled water
DIVA	Differentiate infected from vaccinated animals
DLV	Double locus variant
DNA	Deoxyribonucleic acid
DSMZ	German Collection of Microorganisms and Cell Culture
dt	Direct transfer
edt	Extended direct transfer
EDTA	Ethylenediaminetetraacetic acid
EPS	Extracellular polymeric substances

FU Berlin	Free University of Berlin
hrs	Hours
ICE	Integrative conjugative element
IHIT	Institute for Hygiene and Infectious Diseases of Animals
IMT	Institute of Microbiology and Epizootic Diseases
IPB	*Institut Pertanian Bogor* (Bogor Agricultural University)
JLU	Justus Liebig University
kDa	Kilo Dalton
LB	Luria-Bertani
LGT	Lateral gene transfer
MALDI-TOF MS	Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry
MDR	Multidrug-resistant
MGE	Mobile genetic element
MIC	Minimum inhibitory concentration
min	Minutes
MLST	Multilocus sequence typing
MSCRAMMs	Microbial surface component recognizing adhesive matrix molecules
NCBI	National Centre for Biotechnology Information
nm	Nanometer
NRPS	Non-ribosomal peptide synthetase
OD	Optical density
PBMC	Peripheral blood mononucleated cells
PCR	Polymerase chain reaction
PFGE	Pulsed field gel electrophoresis
PMSF	Phenylmethyl sulfonyl flouride
PubMLST	Public databases for molecular typing and microbial genome diversity
RFLP	Restriction fragment length polymorphism
RKI	Robert Koch Institute

RNA	Ribonucleic acid
rpm	Rotation per minute
S.	*Streptococcus*
S. Braenderup	*Salmonella enterica* subspecies *enterica* serovar Braenderup
sAg	Superantigen
SBF	Specific biofilm formation
Scl	Streptococcal collagen-like
sec	Seconds
Se-mPCR	*S. equi*-multiplex polymerase chain reaction
See	*Streptococcus equi* subspecies *equi*
SeM	M-like protein of *S. equi* subspecies *equi*
Ser	*Streptococcus equi* subspecies *ruminatorum*
Sez	*Streptococcus equi* subspecies *zooepidemicus*
SVMBS	School of Veterinary Medicine and Biomedical Sciences
SLS	Streptolysin S
SLST	Single locus sequence typing
SLV	Single locus variant
SNP	Single nucleotide polymorphism
ssp.	Subspecies
ST	Sequence type
Staph.	*Staphylococcus*
Std	Standard deviation
THB	Tod-Hewitt broth
v/v	Volume/volume
VAG	Virulence-associated gene
Vt	Virotype
WGS	Whole genome sequence

1 Introduction

Strangles (dt. Druse) is one of the most important and frequently diagnosed infectious diseases of equines worldwide [192]. This disease is characterized by abscessation and rupture of lymph nodes of the head and neck [23]. The morbidity and case fatality rate of strangles may be up to 100 % [129,141] and 10 % [31], respectively. Although the disease has been known for centuries [155], it continues to cause a significant animal welfare burden and economic loss [192]. Carriers may shed the causative agent and transmit it to other equines [128].

Strangles is caused by *Streptococcus equi* subspecies *equi* (*See*), a beta haemolytic *Streptococcus* (*S.*) belonging to Lancefield Group C [23]. Phylogenetic analyses suggest that *See*, a pathogen restricted to equines, evolved from an ancestral strain of *S. equi* subspecies *zooepidemicus* (*Sez*) [73,196]. The animal suffering and costs associated with strangles necessitate effective treatment, control, and prevention, however, this can be challenging [23]. Antimicrobials, especially beta-lactam antibiotics, are used to treat strangles in equines depending on the stage and severity of the disease [23,25,174]. However, rates of resistance to trimethoprim/sulfamethoxazole, fluoroquinolones, and aminoglycosides have increased lately in *See* [48,79], indicating that *See* currently is under selective pressure and is adapting to these antimicrobials. This could further limit therapeutic options for this pathogen. Various *See* vaccines have been developed and used for prevention with varying degrees of success, for example, live attenuated vaccines [23,49] and cell-free protein extract vaccines [46]. The newest invention is a recombinant subunit vaccine named Strangvac® [148,150], which received marketing authorisation in the European Union in 2021 [50].

The pathogenicity of *See* is influenced by the presence and expression of specific virulence-associated genes (VAGs) [71,177,190]. The genome of *See* has been reported to mutate continuously, indicated by the loss and gain of VAGs, which may influence the virulence of this pathogen [30,71,73]. Furthermore, bacterial biofilm protects against antimicrobials and host immune responses *in vivo*, increasing their virulence and potential to cause a persistent infection in the host [17,58,68,106,193].

Strangles has been reported from equines worldwide [122], except in Iceland [18]. In Germany, strangles occurs sporadically, but epizootics involving 53 of 112 weanlings were also observed in 2014 on a farm in Neustadt-Glewe, Mecklenburg-Western Pomerania [122,185]. The molecular diversity of *See* strains causing strangles cases over the recent years in Germany remains unclear. Furthermore, the phylogenetic relationship between *See*

strains from Germany and other European countries as well as from other continents is elusive. Therefore, this study aimed to enlighten the molecular diversity of *See* by utilising modern whole genome sequence (WGS)-based analyses. WGS data were used to determine the phylogenetic relationship between *See* isolates recovered from equines in Germany and abroad, as well as the association between genotypes, VAGs, and epidemiological data of these *See* isolates. This study also determined the current spectrum and rates of antimicrobial susceptibilities of *See* to support effective treatment of strangles. In addition, biofilm formation by *See* isolates was assessed *in vitro* and correlated with genotypes in order to elucidate factors that may influence biofilm formation by *See in vivo*.

1.1 Strangles

Strangles is a highly infectious respiratory disease of equines caused by *Streptococcus equi* ssp. *equi* [23]. Clinical signs typical of strangles were reported for the first time by Jordanus Ruffus in 1251, an officer in the imperial court of Emperor Fredrick II [155]. Solleysel in 1664 Le Parfait Marechal mentioned that strangles is a disease that young horses must pass through. Interestingly, in 1811 Napoleon I, the French Emperor, requested horses that had recovered from strangles to be used on the battlefield [126], presumably to prevent horses from contracting this disease during military operations and therefore not being able to perform as required [126,198].

1.1.1 Clinical signs

Strangles primarily manifests as an inflammatory disease of the upper respiratory tract and its lymph nodes [23,174]. However, strangles and its sequelae can also affect many other organs, such as brain, liver, spleen, kidney, mammary gland, and limb due to metastatic abscessation [23,111]. Typical clinical findings of strangles are pyrexia, mucoid to mucopurulent nasal discharge, anorexia, depression, cough, local pain, and abscesses of lymph nodes in the head and neck, particularly of submandibular and retropharyngeal lymph nodes [23,111,174]. Pyrexia exceeding 42 °C and depression are the most clinical signs that develop in the first 3–14 days after infection [23]. Abscesses of the lymph nodes usually rupture between 7 days and 4 weeks after infection [23]. *See* can be detected three hours (hrs) after the infection in the head and neck lymph nodes [180], but does not colonize in the nasopharyngeal cavity; thus, it cannot be detected 24 hrs post-infection there [190,192]. The bacteria are shed per nasal after a latent period of 4–14 days and terminate after 3–6 weeks after the acute phase [177].

The morbidity rate of strangles may be 100 % in susceptible populations, and the case fatality rate may be 10 % [31,129]. The complication of strangles includes metastatic abscessation (termed as 'bastard strangles'), purpura hemorragica, and empyema of the guttural pouch [23,31,174]. Furthermore, concretions of inspissated pus can form 'chondroids', enabling *See* to persist for up to several years in the guttural pouch of its apparently healthy host [191,192]. Approximately 10 % of horses that recover from strangles become persistent carriers of *See* [84] by harbouring the bacteria in their guttural pouch for months or even years and thus remain a potential source of infection for other equines [129,191,192].

Strangles can occur in horses of any age with different severity depending on the immune status of the horse. While older horses appear to recover more quickly with less severity, young horses often experience more severe clinical signs, including formation and maturation of lymph node abscesses that may subsequently rupture [23]. If not treated with antimicrobials, around 75 % of horses develop immunity that persists for five years to reinfection after recovery from strangles [174].

It is extremely rare for *See* infections to occur in humans, however several cases have been reported recently [47,86,184,194]. Infection of *See* in human causes meningitis [184], meningoencephalitis [86], sepsis [184], and retropharyngeal abscess [194].

1.1.2 Epidemiology

1.1.2.1 Geographic distribution and frequency

Strangles is endemic worldwide [122,189], except in Iceland, where the horse population is geographically isolated, and a centuries-old self-imposed ban on horse imports has been applied [18]. Information about the prevalence of this disease is scarce in the literature. In the United Kingdom, around 700 strangles outbreaks were reported in a year, some involving more than 200 horses [77,138]. In Germany, a study reported that 53 % of weanlings of a horse breeding farm were infected by *See* during an outbreak in 2014 [185]. An outbreak of strangles involving 62 clinical cases occurred over a 15-month period on a large Standardbred breeding farm with a population of around 1,400 horses in the United States of America (USA) [31]. In Ontario, Canada, 10.3 % of samples (289 from 2799 samples) tested for *See* between 2008 and 2018 were positive [24].

It is assumed that current *See* strains, infecting horses around the world, are closely related to each other and are all descendants of the same strain that infected horses worldwide in 1909 [71]. As a result of the First World War, horses were transported, mixed, and killed on a mass scale throughout the world at the beginning of the 21st century [71]. On

the other side, horse breeding programs were established, providing a large number of naïve hosts for *See* [71]. The occurrence of all these events allowed different strains of *See* to mix together and cause problems for horses, resulting in many horses dying and being replaced by horses without strangles [71]. It is assumed that these events supported the survival of the fittest *See* strain that subsequently spread to equines all over the world [71]. International and national modern transportation of horses may drive the transmission of strangles in equine populations worldwide [71,122].

Phylogenomics analysis of 670 *See* genomes using a novel core genome multilocus sequence typing (cgMLST) revealed that *See* strains with the same type of Bayesian Analysis Population Structure (BAPS) were recovered from horses from different continents [122]. *See* isolates belonging to BAPS-1 were recovered from horses in North America, Asia, and Oceania; BAPS-2 in Europe and Asia; BAPS-3 in Oceania, Asia, and Europe; BAPS-4 in South America, Europe, and Asia; BAPS-5 in Europe and Asia; and BAPS-6 in North America, Europe, and Asia [122]. In particular, horses in the United Arab Emirates shared all six types of BAPS of *See* with countries worldwide [122].

1.1.2.2 Host range

Strangles cases have been reported from horses worldwide [122] and also from donkeys [43,207]. Based on a surveillance for important equine infectious respiratory diseases in the USA, *See* was the most common agent identified in horses 6–10 years old [144]. Donkey foals are more susceptible to strangles than older donkeys [43].

It is extremely rare for humans to become infected with *See*, but there have been reported infections in recent years [47,86,184,194]. Human cases of *See* infection have all been associated with previous horse-human contacts. [47,86,184,194]. *See* infections in humans were reported in a 69-years old adult [86] and children [47,184,194], including in an immunocompromised child [184].

1.1.2.3 Transmission

See can be transmitted directly through nose-to-nose contact with infected equids or indirectly by ingesting contaminated food or water, or by physical contact with contaminated equipment or even human skin [135,189]. *See* can survive approximately one to three days on the surface outside its host [197], while in a wet environment, it survives from four to 34 days, depending on the season [45]. Asymptomatic carrier equines are the most

common unrecognized source of *See* infections [23,129]. Strangles may be transmitted to other horses for at least 6 weeks following the resolution of their clinical signs [174].

1.1.3 Treatment

Equines with strangles should be treated according to the stage and severity of the disease. In most cases of strangles, no treatment is necessary except for rest in a comfortable warm stall and good quality palatable food [23]. If clinically indicated, antimicrobials are typically administered to treat equines with strangles [23,141,174]. Antimicrobials may be indicated during the acute stage with high fever with lethargy and respiratory distress, or cases of metastatic abscessation, as well as for local treatment of guttural pouch infections [23]. Administration of antimicrobials can have adverse effects on the course of *See* infection depending on the stage, manifestation, and severity of the disease [23,174]. In the acute phase of infection accompanied by pyrexia and depression, antimicrobials may prevent abscess formation [23,111]. However, if lymphadenopathy has already manifested itself, antimicrobials, for example trimethoprim/sulfadiazine, are ineffective until the abscess has ruptured by its own or has been drained [23,25]. Aside from antimicrobials, non-steroidal anti-inflammatory drugs are used to reduce pain and swelling at the site of the lymph node abscess [23,174].

Beta-haemolytic streptococci from equines are usually susceptible to beta-lactam antimicrobials (including penicillins, cephalothin, ticarcillin-clavulanate), erythromycin, bacitracin, and nitrofurantoin [48]. Penicillin is the drug of choice for *See* infection [23,25,174]. However, it must be considered that during the acute phase of strangles, penicillin administration can impair the persistence of humoral immunity to *See* [141]. Other beta-lactams, particularly ceftiofur, can be administered as alternatives to penicillin [23]. Furthermore, oxytetracycline [25], macrolides [23,174], and trimethoprim/sulfadiazine [23,25] can also be considered for treatment of equines with strangles. Although resistance against antimicrobials is still rare in *See*, significantly reduced susceptibility for trimethoprim/sulfamethoxazole, aminoglycosides, and fluoroquinolones has been reported for some clinical *See* isolates over the years [48,79].

1.1.4 Prevention

1.1.4.1 Biosecurity and outbreak management

Application of biosecurity and hygiene is essential to prevent strangles outbreaks since it is highly contagious. As a consequence of the known transmission routes, sharing of equipment should be avoided, and the principles of hygiene and sanitation should be implemented [135,189]. It is also important to practice biosecurity protocols to prevent the spread of the disease. New horses should be quarantined for at least three weeks and screened for possible clinical signs and serological indicators of infection [135,189]. It is also necessary to maintain a detailed record of equine movements and arrivals.

During a strangles outbreak, horses should be separated into three groups: red, amber, and green, based on their clinical signs and risk of infection [190]. The red group is infected horses with clinical signs [190]. The amber group is horses that have been in contact with infected horses and thus have been exposed to *See*, but are not showing any signs of disease yet [190]. Additionally, the green group is horses that did not have contact with horses of the red and amber groups, and do not show any suspicious clinical signs [190]. Appropriate screening and testing should be conducted at least four weeks after the clinical signs are resolved [135,190].

1.1.4.2 Vaccination

Several types of *See* vaccines have been developed using different technologies with various degrees of success, such as live attenuated vaccine [23,189], cell-free surface protein extract vaccine [46], and recombinant subunit vaccine [148,150]. A **live attenuated *See* vaccine** can be administered intranasally or into the lip, for example, Pinnacle I.N. (Zoetis), which is distributed in the USA, Canada, and New Zealand, and Equilis StrepE (MSD Animal Health), which is distributed in Europe [189]. Unfortunately, there is only limited capability to differentiate infected from vaccinated animals (DIVA) because animals vaccinated with the live attenuated vaccine become transiently culture-positive, polymerase chain reaction (PCR)-positive, and seropositive [40]. At least Equilis StrepE is based on a *See aroA* deletion mutant, which can be differentiated from wild type strains of *See* [78,85]. Safety concerns have been raised by reports that some cases of clinical diseases have occurred in Equilis StrepE vaccinated horses due to residual virulence and replication of the vaccine strain [85]. Furthermore, efficacy may be limited as well as strangles cases were reported in horses after vaccination with Pinnacle I.N. [40].

A **cell-free surface protein extract vaccine,** named Equivac S (Zoetis), distributed in the USA and Australia contains no live agent and can be administered intramuscularly [46]. It has some DIVA capability because the administered *See* antigen cannot be detected by culture methods. Depending on the primers and deoxyribonucleic acid (DNA) target used, the *See* antigen cannot be detected by PCR either. However, this strangles vaccine may interfere with serological results, thus inhibiting serological DIVA testing [23,46].

The latest development of strangles vaccine is Strangvac® (Intervacc AB), a **recombinant fusion protein** vaccine [148,150], which received marketing authorisation in the European Union in 2021 [50]. Strangvac® contains three recombinant fusion proteins, termed CCE, Eq85, and IdeE. CCE consists of of five cell anchored proteins, namely SEQ0935 (CNE), SEQ0855 (SclF), SEQ1817 (SclI) and SEQ2101 (SclC), SEQ0721 (EAG); Eq85 consists of two cell anchored protein, namely SEQ0402 (Eq8), SEQ0256 (Eq5), while SEQ0999 (IdeE) is a secreted protein [148,150]. Strangvac® is intended for horses at high risk of *See* infection in areas where the pathogen has been identified [50]. It can be administered via the intramuscular route to horses from the age of six months [150]. When tested by experimental infection, Strangvac® vaccination protected 95 % of ponies from developing strangles two weeks after the third vaccination [150]. Within two weeks and two months following the second vaccination, protection increased from 31 % to 58 %, respectively [150]. Subsequently, it further increased to 94% within three weeks after the third vaccination [150]. Strangvac® demonstrates DIVA capabilities by containing different antigens than those targeted by the diagnostic *See* PCR [4,37,195]. Since Strangvac® does not contain any live bacteria, the vaccine antigens are not detectable by any bacterial culture method. Furthermore, vaccinated horses remained seronegative in the commercial diagnostic indirect enzyme-linked immunosorbent assay (iELISA) for *See* [149,150].

1.2 *Streptococcus equi* ssp. *equi*

1.2.1 Taxonomy

The taxonomy of *Streptococcus equi* subspecies *equi* [42] is as follows:

Phylum:	Firmicutes
Class:	Bacilli
Order:	Lactobacillales
Family:	Streptococcaceae
Genus:	*Streptococcus*

Species: *Streptococcus equi*
Subspecies: *Streptococcus equi* subspecies *equi*

Genetically, *See* is closely related to *S. equi* subspecies *zooepidemicus* (*Sez*) [73,196] and *S. equi* subspecies *ruminatorum* (*Ser*) [53]. *See* belongs to Lancefield group C based on the classification of streptococci using serological identification of polysaccharide group antigen, known as Lancefield grouping [90]. According to another classification of streptococci introduced by Sherman in 1937 based on haemolytic reaction, group of carbohydrate antigens, and phenotypic tests, *See* is classified as a pyogenic *Streptococcus* subspecies [38,163].

1.2.2 Morphological and biochemical traits

Bacteria of *See* present as Gram-positive nonmotile cocci (0.6–1.0 µm) arranged in chains or pairs [38,167]. After incubation at 37 °C for 24 hrs on a sheep blood agar plate (BAP), colonies are small (approximately 1 mm in diameter), circular, translucent, and surrounded by a wide zone of beta-haemolysis. Some strains grow with mucoid colonies, while others are non-mucoid. *See* is a facultative anaerobic bacterium that grows well in 5–10 % CO_2 enriched atmosphere [38].

Based on carbohydrates fermentation tests, *See* is able to ferment D-glucose, maltose, sucrose, and salicin but can not ferment lactose, trehalose, L-arabinose, D-mannitol, trehalose, ribose, glycerol, inulin, raffinose, D-sorbitol [38,52,167]. The negative reactions of *See* in fermenting trehalose, lactose, and sorbitol [3,52] can be used to differentiate this bacteria from other beta-haemolytic Lancefield group C streptococci. Furthermore, *See* displays positive reactions for the phosphatase test and negative reactions for the catalase test, coagulase test, urease test, hippurate hydrolysis test, Voges-Proskauer test, and optochin test. [38,52,167].

1.2.3 Genome

Research on bacterial genomes has been revolutionized by the development of whole genome sequencing technologies. WGS-based analyses enable the study of the pan genome, core genome, and accessory genome. The pan genome encompasses the complete set of genes found in all the diverse strains of bacteria or other microorganisms within a specific clade, encompassing both core and accessory genes [176]. The core genome

consists of all the genes present in all the different strains of bacteria or other microorganisms within a particular clade [127,161]. These genes are essential for the fundamental biological functions and shared characteristics of the bacterial clade [161]. In contrast, the accessory genome refers to the genes that are present in some, but not all, strains or isolates within a bacterial clade [123,161]. Accessory genomes are considered to be strain-specific, contributing to genomic diversity within a population [161]. Genomic studies provide valuable resources for understanding the genetic diversity of a species, adaptation mechanisms, and uncovering the species' evolutionary history.

The early genome evolution of streptococci involved genome expansion through gene gain, followed by a period of streamlining through gene loss, leading to the diversification of major streptococci groups [147]. The current streptococci evolved through a period of genome expansion, with lateral gene transfer (LGT) playing a significant role [147]. The pyogenic streptococci group, including *See, Sez, Ser, S. pyogenes*, and *S. dysgalactiae*, exhibited high gene turnover through rapid adaptive radiation by LGT [147]. Pyogenic streptococci have enriched genes related primarily to transportation and metabolism, which may account for their diverse host and environment [147]. Genes involved in sugar uptake and carbohydrate metabolism were crucial in the streptococci pyogenic group's evolution, as were pathogenesis-related genes acquired through LGT after branching [147].

The genome of *See* shares around 80 % sequence identity with *S. pyogenes*, a prominent human pathogen classified under the Lancefield group A streptococci [73]. The 16S ribosomal ribonucleic acid (RNA) sequence of *See* has 98.8 % similarity with *Ser* [53]. Moreover, DNA-DNA pairing studies showed that *Ser* displayed more than 70 % relatedness to *See* and *Sez* [53].

The genome of *See* shares 98 % sequence identity with *Sez* [177]. A previous study compared the genome of *See* strain 4047 (GenBank accession number: FM204883.1) with *Sez* strain H70 (GenBank accession number: FM204884.1) [73]. *See* strain 4047 was isolated from a pony with strangles in the New Forest, England, in 1990, while *Sez* strain H70 was isolated from a healthy Thoroughbred racehorse in Newmarket, England, in 2000 [73,196]. The genome of *See* strain 4047 consists of 2,253,793 base pairs (bp) encoding 2,137 predicted coding sequences (CDSs), whereas the genome of *Sez* strain H70 consists of 2,149,866 bp, encoding 1,960 predicted CDSs [73]. Most CDSs of *See* strain 4047 (1,671 from 2,137, 78.2 %) have *Sez* strain H70 orthologs, and most non-orthologous *See* strain 4047 CDSs (422 from 466 CDSs) are found on mobile genetic elements (MGEs) [73]. The genome of *See* strain 4047 is larger than *Sez* strain H70 because of the integration of MGEs

that composed 16.4 % of the genome of *See* strain 4047, while only 7.5 % of the genome of *Sez* strain H70 is composed of MGEs [73].

MGEs influence bacterial fitness and virulence since they enhance gene acquisition and loss [139,186]. Certain MGEs can move between DNA molecules (insertion sequences/IS, transposons, and gene cassettes/integrons), while others can be transferred between bacterial cells (plasmid, prophage, integrative conjugative elements/ICE) [139]. *See* strain 4047 contains more IS elements (n = 73) compared to *Sez* strain H70 (n = 30) [73]. The IS elements in *See* strain 4047 are dominated by the member of ISSeq3. Additionally, *See* harbours more homologues of insertion IS861 compared to *Sez* [177]. The presence of IS elements in the genome of *See* interrupts the function of genes that are important for bacterial adaptation to environmental changes [73].

To date, two ICEs specific to *See* have been identified, termed ICE*Se*1 (35 kb) and ICE*Se*2 (63 kb) [73]. The ICE*Se*2 locus encodes equibactin, a siderophore that plays a role in iron acquisition [72]. Information about equibactin is explained in more detail in **chapter 1.2.4.**

Other important MGEs of *See* strain 4047 are its four prophages, namely φSeq1, φSeq2, φSeq3, φSeq4, that are 39 kb, 41 kb, 30 kb, and 40 kb in size, respectively [73]. The DNA sequences of these prophages are similar to the prophages of *S. pyogenes*, indicating a cross-species pathogen evolution [73]. No homology was identified between the genes encoded by φSeq1 (SEQ0133-SEQ0109) to known virulence factors [73]. Based on the accessory genome analyses of *See*, the φSeq1 prophage appears more dynamic than the other three prophages that are more conserved in the genome [71]. These prophages encode important virulence factors: φSeq2 (SEQ0787-SEQ0851) encodes phospholipase A2 toxin (SlaA), φSeq3 (SEQ1727-SEQ1765) encodes superantigens (sAgs) SeeL and SeeM, while φSeq4 encodes sAgs SeeH and SeeI [73].

Plasmids are extrachromosomal MGEs that have the capability of autonomous replication in the host and usually carry accessory genes that give advantages to their host [139]. Plasmids were identified in streptococci, including *S. suis* [203], *S. pneumoniae* [157], *S. pyogenes* [187], and *S. dysgalactiae* ssp. *equisimilis* [35]. Plasmids often carry genes that confer beneficial traits to the bacteria, such as antimicrobial resistance and VAGs [104,187]. There is no information in the literature regarding the plasmids of *See*. Plasmid can be detected using PCR-based replicon typing, pulsed-field gel electrophoresis (PFGE), optical mapping of plasmid [132], and sequencing-based plasmid detection [153].

Studies based on multilocus enzyme electrophoresis, 16S-23S RNA intergenic spacer sequence, and multilocus sequence typing (MLST) reported genetic homogeneity of various *See* isolates [29,81,196]. However, deep genome analysis revealed that the genome of *See* continues to change despite becoming equine-restricted, with evidence of gene acquisition and amplification, as well as decay and mutation [71]. Mutations and decay of the *See* genome cause a decrease in function, but this has been compensated by the acquisition of new traits such as prophage that carry a phospholipase A2 toxin and four superantigen exotoxins (SeeI, SeeH, SeeL, SeeM) [73] as well as an ICE that carries the equibactin iron acquisition system that is similar to the high pathogenicity island found in *Yersinia pestis* [72], indicating specialization for causing disease. Interestingly, mutations that lead to metabolic streamlining and loss of virulence determinants were reported as more common in persistent isolates than isolates from acute infection [71].

1.2.4 Virulence factors

Various virulence factors are used by *See* to infect the equine host and to cause strangles. Virulence factors of *See* have an important role in bacterial adherence [92,172,192], invasion, immune modulation [178,192], toxicity [3,8,136], and exoenzymatic host cell and tissue damage [93,177], and nutritional or metabolic activities, including iron acquisition [72,73]. The accepted and putative virulence factors of *See* are listed in **Table 1**. Capsule, M proteins, and equibactin as virulence factors are explained in more detail due to their role in this study.

Table 1. Accepted and putative virulence factors for *Streptococci equi* ssp. *equi*

Virulence Factors[1)]	Gene, operon, or locus tag	Gene Location	Function	Ref.
Adherence				
FNE	*fne*	Chromosome	Adhesin: binding to fibronectin and collagen	100
FNEB	*fneB*	Chromosome	Adhesin: binding to fibronectin and collagen	73, 92
FneC	*fneC* (SEQ1606)	Chromosome	(Putative) adhesin: binding to collagen	73
FneD	*fneD* (SEQ1607)	Chromosome	(Putative) adhesin: binding to collagen	73
FneE	*fneE* (SEQ0555)	Chromosome	(Putative) adhesin: binding to fibronectin and collagen	73
FneF	*fneF* (SEQ1649)	Chromosome	(Putative) adhesin: binding to collagen	73
SFS	*sfs* (SEQ0466)	Chromosome	Adhesin: binding to fibronectin	98

Virulence Factors[1)]	Gene, operon, or locus tag	Gene Location	Function	Ref.
Shr	*shr*	Chromosome	Adhesin: binding to fibronectin. Metabolism: putative protein associated with iron transport (binding to heme)	73
SclB, SclC, SclD, SclE, SclF, SclG, SclH, SclI	*sclB, sclC, sclD, sclE, sclF, sclG, sclH, sclI*	Chromosome	Adhesin: collagen-like surface anchored proteins	61, 73, 105, 148
Eq5	*eq5* (SEQ0256)	Chromosome	Adhesin: cell-surface anchored protein	148
Eq8	*eq8 (*SEQ0402)	Chromosome	Adhesin: cell-surface anchored protein	148
Pilus	FimI operon (*tetR, cne, SEQ0936, srtC1*)	Chromosome	Adhesin: binding to collagen (CNE)	148, 150, 172
Immune modulation				
Capsule	*has* operon (*hasA, hasB, hasC1, hasC2*)	Chromosome	Impedin: antiphagocytosis	173, 177, 188
SeM (M protein)	*seM*	Chromosome	Impedin: antiphagocytosis, binding to IgG	5, 84, 178
SzPSe (M protein)	*szpse*	Chromosome	Impedin: antiphagocytosis	76, 178
Se18.9	*se18.9*	Chromosome	Impedin: antiphagocytosis.	76, 182
EAG	*eag* (SEQ0721)	Chromosome	Impedin: binding to α2-macroglobulin, albumin and IgG	60, 61, 148
Exotoxins				
SeeH	*seeH*	Prophage φ-Seq4	Modulin: superantigenic exotoxin	4, 8, 73, 136
SeeI	*seeI*	Prophage φ-Seq4	Modulin: superantigenic exotoxin	4, 8, 73, 136
SeeL	*seeL*	Prophage φ-Seq3	Modulin: superantigenic exotoxin	3, 136
SeeM	*seeM*	Prophage φ-Seq3	Modulin: superantigenic exotoxin.	3, 136
SlaA	*slaA* (SEQ0849)	Prophage φ-Seq2	Aggressin: phospholipase A2 Toxin (cleaves phospholipid molecules)	73, 102, 190, 192
SlaB	*slaB* (SEQ2155)	Prophage φ-Seq2	Aggressin: phospholipase A2 Toxin (cleaves phospholipid molecules)	73, 102, 190, 192
Streptolysin S	*sagA* (SEQ0546, SEQ0547-SEQ0549)	Chromosome	Aggressin: haemolytic and cytolytic exotoxin, forms transmembrane pores in red blood cells resulting in osmotic lysis	73, 192
Exoenzymes				
Hyaluronidase/ hyaluronate lyase	*seq2045*	Prophage φ-Seq4	Aggressin: degrades hyaluronic acid and chondroitins	73, 101
Hyaluronidase/ hyaluronate lyase	*hylB* (SEQ1479)	Chromosome	Aggressin: degrades hyaluronic acid and chondroitins	73
Streptokinase	*skc* (SEQ2014)	Chromosome	Aggressin: plasminogen activation (aid in the dispersion of *S. equi* in the tissue via plasmin that hydrolyses fibrin)	73, 115, 177
EndoSe	*endoSe*	Chromosome	Aggressin: endoglycosidase, which hydrolyses glycosyl that has an antiphagocytic function by cleaving the IgG	59, 148
IdeE	*ideE* (SEQ0999, SEQ0938)	Chromosome	Aggressin: IgG-endopeptidase (cleaves horse IgG)	93, 148
IdeE2	*ideE2*	Chromosome	Aggressin: IgG-endopeptidase (cleaves horse IgG)	74
SodA	*sodA*	Chromosome	Aggressin: neutralise toxic levels of reactive oxygen species generated by the host	4, 140

Virulence Factors[1)]	Gene, operon, or locus tag	Gene Location	Function	Ref.
SortaseA	*srtA* (SEQ1171)	Chromosome	Adhesin: mediates covalent attachment of cell wall-anchored protein	[73]
SortaseC1	*srtC1* (SEQ0937)	Chromosome	Adhesin: mediates covalent attachment of cell wall-anchored protein	[73]
Prolipoprotein diacylglycerol transferase (lgt)	*lgt* (SEQ1537)	Chromosome	Synthesis of bacterial lipoprotein	[69]
Nutritional/Metabolic factors				
Equibactin	*Equibactin* operon (*eqbA-eqbN*)	ICE*Se*2	Siderophore: iron acquisition	[72]

[1)] Scl: Streptococcal collagen-like surface anchored proteins, SodA: superoxidase dismutase. Ref: Reference(s)

1.2.4.1 Hyaluronic acid capsule

The capsule of *See* is an extracellular layer superimposed on the bacterial cell wall from the outside. The capsule matrix is essentially formed by hyaluronic acid, a non-antigenic high molecular weight polymer consisting of alternating residues of N-acetylglucosamine and glucuronic acid [177]. *See* expresses capsule type III that is genetically similar to the genes for capsule production and regulation of other pathogenic streptococcal species like *Sez* and *S. dysgalactiae* ssp. *equisimilis*, having deletions in genes for autolysin and pneumolysin [177]. The synthesis of *See*-capsule is regulated by the *has* locus, which consist of *hasA* (hyaluronate synthase), *hasB* (UDP-glucose dehydrogenase), and *hasC* (UDP-glucose pyrophosphorylase) [173,177]. Amplification of the *has* locus results in an increase in the production of hyaluronic acid, while deletions of the locus result in a decrease in hyaluronic acid production [71]. In some *See* strains, deletions in *hasA* or *hasB* cause stable loss of capsule synthesis and lead to a non-mucoid phenotype of *See* [188]. Interestingly, a *See* strain 1691 displayed mixed colony phenotypes, although no mutation in the *has* locus was identified [173]. In this case, reduced capsule formation of *See* strain 1691 is correlated with a reduction in transcription of the *has* locus [173].

In *See*, capsule production contributes to its mucoid colony phenotype on BAP [6,173,177]. Highly encapsulated strains produce mucoid colonies, which are highly indicative of a virulent *See* isolate, whereas non-encapsulated strains are less virulent [6]. The hyaluronic acid capsule reduces the phagocytosis of *See* by neutrophils *in vitro* [6]. In the host tissue, hyaluronic acid is also prevalent, thus, its production by *See* helps mask the bacterium from the host immune system [173]. The capsule protects *See* from the phagocytic function of neutrophils [177]. Furthermore, an intact capsule supports the function of SeM and possibly other hydrophobic proteins exposed to the cell's surface [177].

1.2.4.2 M proteins

M proteins, also known as M-like proteins, are fibrillary cell-surface attached proteins composed of two polypeptide chains forming an alpha-helical coiled-coil arrangement [55,56,133]. M proteins have common structural features, including the LPXTG motif at their C-terminal region, which is a sorting signal required for the correct anchoring of the protein into the cell wall [133]. M proteins have variable N-terminal regions, conserved central repeats, and proline/glycine rich domain at the C-terminal regions [56,133]. M proteins have been widely identified in streptococci and play a role in adhesion and colonization [55,56,133,178].

Two M proteins have been discovered in *See*, namely SeM (also known as FgBP) and SzPSe, having a molecular weight of 58-kDa and 40-kDa, respectively [178]. There is no homology between SeM and SzPSe other than the signal sequence (39 % identity) and membrane anchor sequence (66 % identity) [178]. Neither of these proteins exhibits significant amino acid sequence similarity with the M proteins of *S. pyogenes*, which can facilitate intracellular survival in the neutrophils [171]. Still, both possess extensive regions of alpha-helical coiled coils [117] and bind strongly to equine fibrinogen [178].

SeM (534 amino acids) is encoded by the *seM* gene (1,605 bp). SeM has a signal sequence of 36 residues, two direct repeats located on residues 226 to 267, and an alpha helix region approximately on residues 120 to 480 [178]. The N-terminal region of the SeM is hypervariable, thus widely used for single locus sequence typing (SLST) of *See*, known as *seM* typing [5,84]. The *seM* typing is explained in more detail in **chapter 1.3.2**. SeM has antiphagocytic properties by its ability to bind fibrinogen and immunoglobulin G and to inhibit the deposition of C3b at the bacterial surface [84,178].

SzPSe of *See* is a homolog of SzP in *Sez* [178]. SzPSe (374 amino acids) is encoded by the *szpse* gene (1,125 bp) [178]. Secondary structure prediction of SzPSe indicates an extended alpha-helical structure with turns near the N-terminal region and a non-alphahelical hypervariable region [178]. SzPSe is a surface exposed immunoreactive protein [181]. It binds to equine fibrinogen, thus inhibiting phagocytosis [181]. The conservation of SzPSe and the lack of variation in its amino acid sequences among 25 See strains from North America, Europe, Australia, and Pakistan indicates that this virulence factor plays a vital role in the survival of *See* [76].

1.2.4.3 Equibactin

Equibactin, which is encoded by ICE*Se*2, mediates the iron acquisition in *See* [72,73]. ICE*Se2* was first identified in *See* strain 4047, while it was absent in *Sez* strain H70 [73]. The ICE*Se2* consists of 14 coding sequences, *eqbA* (SEQ1233) to *eqbN* (SEQ1246) [72]. The *eqbA* acts as a DtxR-like repressor and its deletion leads to a small-colony phenotype. This phenotype can be reversed by deleting *eqbE* as an irp2 homolog, deleting *eqbH-J* encoding a ferric-siderophore-like importer, or adding nitrilotriacetate as an iron chelator [72]. The *eqbB-G* and *eqbM*, *eqbN* genes encode non-ribosomal peptide biosynthetic proteins [72]. Due to its subspecies specificity, ICE*Se*2 serves as a target for PCR-based detection of *See* and to differentiate *See* from *Sez* [37,195].

1.2.5 Biofilm formation

Biofilm is an accumulation of microorganisms typically adhering to a solid-liquid matrix protected by a matrix of self-produced hydrated extracellular polymeric substances (EPS) [58]. EPS are a mixture of biopolymers such as polysaccharides, proteins, lipids, and nucleic acids. Within a biofilm, EPS facilitates adhesion of the biofilm layer to surfaces, aggregation of bacterial cells, cohesion of the biofilm matrix including cell-to-cell communication between bacteria, protection of bacteria against host immune responses and antimicrobials, absorption of organic and inorganic nutrients, nutrient source, and enzymatic activities [58]. As a result of the formation of biofilms, pathogens maintain a secure environment that makes sessile bacteria within a biofilm more resistant to antimicrobials than planktonic cells [193,202].

Several techniques have been developed to evaluate the biomass and viability of biofilms, encompassing microbiological and molecular techniques, as well as the physical and chemical properties of biofilms [10]. Microbiological and molecular techniques involve quantifying colony forming units on agar media, performing quantitative PCR, and employing microscopy to evaluate biofilm formation [10]. Physical methods include measuring biofilm volume or weight, employing microbial electrochemical systems through electrochemical impedance spectroscopy, and utilizing ultrasonic time-domain reflectometry to determine biofilm thickness [10]. Chemical methods utilize dyes and fluorochromes that can adhere to or adsorb onto biofilm, such as microtiter plate dye staining and phospholipid based biomass analysis [10].

Microtiter plate dye staining, also known as biofilm formation in microtiter plates, is an indirect method for measuring biofilm biomass by adsorption and desorption of a dye, mainly crystal violet [10]. The method is most commonly used for biofilm assay since it can be used with a wide variety of bacteria [10], including streptococci [20,64,193,201].

Various factors may influence biofilm formation, including environmental conditions, host interaction, protein expression, and bacterial communication through the quorum sensing system [193]. Only limited information is currently available in the literature regarding biofilm formation by *See*. It has been suggested that pili in Gram-positive bacteria play a role in biofilm formation [113]. Pili in *See* are encoded by the FimI locus [73,172]. In general, the deletion of genes in the FimI locus does not significantly reduce the *in vitro* biofilm formation by *See* in Tod Hewitt broth (THB) [172]. However, over transcription of the major pilin (SEQ0936) within the FimI locus significantly increased the biofilm formation by *See* in tryptone soya broth supplemented with 0.5 % glucose [172]. Furthermore, a comparative study of proteins expressed by *Sez* as sessile bacteria within a biofilm compared to planktonic bacteria revealed that of the 24 proteins tested, 13 were upregulated, while 11 were downregulated during biofilm formation [202]. Generally, proteins expressed during biofilm formation were associated with metabolism, adherence, and stress conditions [202].

1.3 Molecular typing methods for *Streptococcus equi* ssp. *equi*

1.3.1 Macrorestriction analysis

Macrorestriction analysis is commonly used for genotyping of bacteria, including *See* [91,97,99,175], and *Sez* [1,89,145]. The method is based on genome-wide restriction fragment length polymorphism (RFLP). Briefly, bacterial genomic DNA is made accessible for a rarely cutting restriction enzyme and cleaved infrequently into various large fragments. Subsequently, the fragments are separated by pulsed-field gel electrophoresis (PFGE) and then visualized [152]. Restriction enzymes used for successful macrorestriction analysis of *S. equi* have been *Sma*I [1,89,97,145], *Apa*I [89,97], and *Not*I [91,175]. The genotype of the investigated strain is defined by the pattern of DNA fragments called pulsotype. The degree of genetic relatedness between strains can be deduced from the similarity of their pulsotypes, mainly using bioinformatics software [145]. The Unweighted Pair Group Method using Arithmetic Averages (UPGMA) and Pearson's similarity coefficient expressed as a percentage are commonly applied to quantify genetic relatedness between bacterial strains by macrorestriction banding patterns of DNA [145].

Macrorestriction analysis can help to uncover genetic relationships within and between bacterial populations and, thus, help to investigate the spread of a certain pathogen in a host population [152]. Macrorestriction analysis has already been used to investigate strangles outbreaks in Sweden [97,99], Ireland [175], USA [91,175], and Japan [175]. A study reported that *See* isolates from Sweden displayed 11 different pulsotypes within 50 isolates [99]. In that study, the pulsotype of the *See* isolates recovered from horses of the same stable was identical [99]. Another study in Sweden also confirmed that *See* isolates from the same outbreak displayed the same pulsotype [97]. A study investigating *See* isolates from three continents revealed that all *See* isolates from independent outbreaks in Japan displayed the same pulsotype, while *See* isolates from USA and Ireland displayed various pulsotypes that differ from the isolates from Japan [175]. Macrorestriction analysis has also been used to distinguish *See* vaccine strain from wild type isolates [91].

Although useful in short-term epidemiological studies, this method, however, might be misleading in global epidemiology because the indexed variation evolves rapidly through unknown mechanisms [107,196]. Furthermore, this method is poorly portable because it is difficult to compare results between different laboratories [107,196]. Additionally, *See* isolates with different *seM* alleles may have the same pulsotype, indicating that macrorestriction analysis has a lower discriminatory power than *seM* typing [97].

1.3.2 *SeM* typing

SeM typing is a single locus sequence typing (SLST) tool used for *See* and *Sez*. The method was introduced by two independent studies which documented a region in the N-terminal portion of the *seM* gene with nonsynonymous nucleotide base substitutions [5,84]. The *seM* typing method is essentially based on nucleotide sequence polymorphism of an internal fragment of 327 bp near the N-terminal of the *seM* gene [84].

The *seM* gene encodes for the cell wall-associated M protein (SeM) of *See* which is an important virulence factor and immunogen of *See* [178]. The molecular size of the SeM protein is approximately 58 kDa. The molecule has a mainly alpha-helical fibrillar structure which is 50-60 nm long [178]. SeM is exposed on the bacterial cell surface, has an antiphagocytic activity [177], and mediates bacterial adherence to equine fibrinogen with its N-terminal domain [117]. The central domain can bind to equine IgG [22] (**Figure 1**). Furthermore, SeM inhibits the deposition of complement C3b on the bacterial surface [118,119].

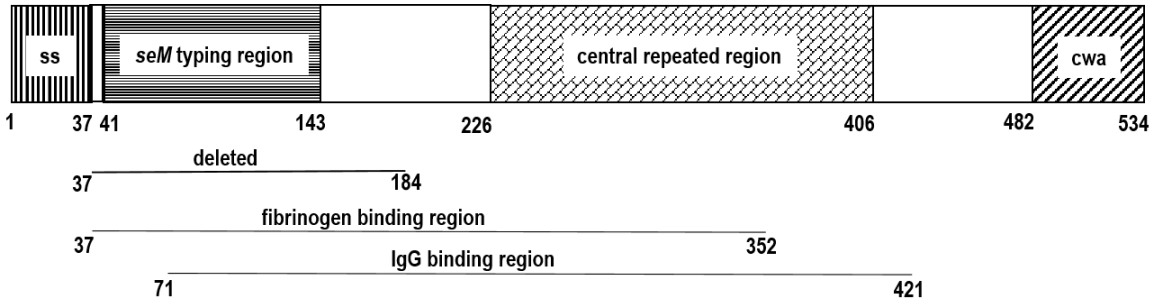


Figure 1. Schematic representation of the SeM protein of *S. equi* ssp. *equi.*

Numbers indicate amino acid (AA) positions. AA 1 to 37 (ss) comprise the M-protein signal sequence, AA 41 to 143 refer to the *seM* typing region [84], AA 226 to 406 (central repeat region) comprise the A and B repeat regions, AA 482 to 534 (cwa) comprise the wall-spanning region and the "LPSTG" cell wall anchor [178]. AA 37 to 352 are involved in fibrinogen binding [119], while AA 71 to 421 mediate binding to IgG [118]. AA 37 to 184 (deleted) refer to a *seM* gene fragment that was found deleted in some *See* strains [30]. The figure was adapted from Kelly *et al*. (2006) [84].

Application of *seM* typing by sequencing the target DNA region of the *seM* gene has been used as an epidemiological tool to investigate strangles outbreaks on various continents [5,27,77,82,84,122,138]. Currently, a database of *seM* alleles is maintained at the University of Oxford and is publicly available on the *Streptococcus zooepidemicus* pages of the Public databases for molecular typing and microbial genome diversity (PubMLST) website [80]. On September 30, 2022, the database contained records for 243 *seM* alleles encoding for 232 different SeM peptides (https://pubmlst.org/bigsdb?db=pubmlst_szooepidemicus_seqdef).

In spite of being a SLST method, *seM* typing has a greater discriminatory power for *See* than MLST [122,196]. Unfortunately, some *See* strains harbour a deletion in the *seM* gene, resultingn translation of a truncated SeM molecule missing several AA at positions 37 to 184 [30]. Thus, major parts or even the entire *seM* typing region is missing in those *See* strains and *seM* typing cannot be applied here. Since those *See* strains had been isolated from horses that were clinically healthy after recovery from strangles, it was hypothesized that truncated SeM might contribute to the persistence of *See* in the guttural pouches [30]. However, *See* strains with truncated SeM were more susceptible to phagocytosis by neutrophils *in vitro* than isolates with a complete SeM [30].

1.3.3 Multilocus sequence typing (MLST)

Multilocus sequence typing (MLST) of bacteria was first introduced in 1998 as a spin off development from multilocus enzyme electrophoresis (MLEE) [107]. MLST is essentially based on nucleotide sequence polymorphisms in defined internal fragments of several housekeeping genes [108,109]. The particular alleles of all analysed housekeeping genes represent the allelic profile or sequence type (ST) of the bacterial strain in question [108]. The method is highly portable, reproducible, and universal since the nucleotide sequences used for typing are publicly available from online platforms [109]. Currently, several online platforms maintain MLST scheme for various bacteria species, e.g. Public databases for molecular typing and microbial genome diversity (PubMLST) (https://pubmlst.org/), MLST 2.0 of Center for Genomic Epidemiology (CGE) (https://cge.food.dtu.dk/services/MLST/), and EnteroBase (https://enterobase.warwick.ac.uk/).

An MLST scheme for *Sez* and *See* was introduced by Webb *et al*. in 2008 [196] and is publicly available on the *Streptococcus equi* ssp*. zooepidemicus* pages of the PubMLST website [80] (https://pubmlst.org/organisms/streptococcus-zooepidemicus/). In this scheme, every sequence type (ST) is defined by the allelic profile of the internal fragments of seven housekeeping genes, namely *arcC* (carbamate kinase), *nrdE* (ribonucleoside-diphosphate reductase), *proS* (prolyl-tRNA synthetase), *spi* (signal peptidase I), *tdk* (thymidylate kinase), *tpi* (triosephosphate isomerase), and *yqiL* (acetyl-CoA acetyltransferase) [196].

With respect to MLST, the global population of *See* is almost clonal, dominated by ST-179 and its single locus variant (SLV) ST-151 [122,196]. Besides ST-179 and ST-151, only seven other STs have been identified so far: ST281, ST282, ST283, ST325, ST395, ST396, ST402 [122]. These other STs are also SLVs or Double Locus Variants (DLV) of ST-179 [122]. In contrast to *See*, the population of *Sez* is highly diverse, with currently 510 STs (PubMLST website, last accessed September 30, 2022).

1.3.4 Core genome multilocus sequence typing (cgMLST)

The development of whole genome sequencing technology has increased the use of core genome MLST (cgMLST) to identify, subtype, and epidemiologically tracking bacteria outbreaks [15,87,127]. Core genome MLST is a high-resolution typing technique for bacterial strain identification and epidemiological tracking by utilising WGS data [87]. In cgMLST, genome-wide gene-by-gene alleles collected from hundreds or thousands (usually 1,500-4,000) of the genes conserved across species are analysed, thereby providing a

much higher resolution compared to MLST [87]. The cgMLST uses a standardised allele numbering system to analyse single nucleotide polymorphism (SNP) diversity in the whole genome, making it highly portable and widely used [15].

A novel publicly available web-based cgMLST scheme for *See* was introduced in 2021 on the Pathogenwatch website maintained by the Wellcome Sanger Institute [122]. The cgMLST scheme for *See* uses 1,286 curated loci present in the reference genome of *See* strain 4047 [73]. Mobile genetic elements (MGEs), insertion sequences, *hasC1*, *hasC2,* and sortase-processed proteins are not used in this scheme [122]. The population structure of *See* is extrapolated using the population mixture analysis in Bayesian Analysis of Population Structure (BAPS) v. 6.0 [36]. The global population of *See* is classified into six BAPS clusters, dominated by *See* strains that belong to BAPS-2 , followed by BAPS-5, and BAPS-1 [62,122].

2 Objectives and research questions

In general, the objectives of this study were

(a) to enlighten the genetic diversity of *See,*

(b) to determine the spectrum and rates of antimicrobial susceptibility of *See,* and

(c) to analyse associations among epidemiological data, genetic diversity, virulence-associated genes (VAGs), antimicrobial susceptibility, and biofilm formation of *See*.

In order to achieve the objectives of this study, questions were addressed and categorized into four main topics:

1. Phylogeny and epidemiology
    - Do the *See* infection cases in this study refer to novel *seM* alleles, STs, and cgMLST BAPS clusters?
    - How is the genetic relationship between *See* isolates in this study and *See* isolates from other studies?
    - How is the association between the phylogeny and epidemiological data of the *See* isolates?
2. Virulence-associated genes and antigens
    - How is the genetic diversity of the *See* isolates according to their VAGs?
    - How is the diversity of the Strangvac® vaccine antigens in the *See* isolates of this study?
3. Antimicrobial susceptibility
    - What are the spectrum and rates of antimicrobial susceptibility of the *See* isolates?
    - Do *See* isolates harbour recognized antimicrobial resistance genes?
4. Biofilm formation
    - Do *See* isolates produce biofilm in different intensities?
    - How is the correlation between *in vitro* biofilm formation by *See* isolates and their genotypes?

3 Materials and methods

3.1 Materials

3.1.1 Culture media, biologicals, chemicals, and consumables

Culture media, biologicals, and chemicals used in this study are listed in **Supplementary Materials S1 Table 23** and **Supplementary Materials S2**. Consumables used in this study are listed in **Supplementary Materials S1 Table 24**.

3.1.2 Devices and software

Only devices other than basic laboratory equipment used in this study are listed in **Supplementary Material S1 Table 25.** The software and programmes used in this study are listed in **Table 2**.

Table 2. Software used in this study

Software	Manufacturer or supplier (source)
Geneious 8.1.9	Biomatters Ltd., Auckland, New Zealand
IBM SPSS Statistics, Version 27	IBM, Armonk, USA
Microreact	Wellcome Trust, London, United Kingdom (https://microreact.org/)
Micronaut Software MCN5	MERLIN Diagnostika GmbH, Bornheim-Hersel, Germany
MyDbFinder V.2.0	National Food Institute, Technical University of Denmark (https://cge.food.dtu.dk/services/MyDbFinder/)
Pathogenwatch	Wellcome Trust, London, United Kingdom (https://pathogen.watch/)
PubMLST	Wellcome Trust, London, United Kingdom (https://pubmlst.org/)
ResFinder V.4.1	National Food Institute, Technical University of Denmark (https://cge.food.dtu.dk/services/ResFinder/)
SAS 9.4	Statistical Analysis System Institute Inc., Cary, NC, USA
Skanlt Software 6.1.0.51 Research Edition	Thermo Fischer Scientific, Waltham, MA, USA

3.1.3 Reference strains

Reference strains used in this study are listed in **Table 3**.

Table 3. Reference strains used in this study

Species	Strain designation	IHIT No.	Origin or source*
Streptococcus equi ssp. *equi*	DSM 20561, ATCC 33398, NCDO 2493, NCTC 9682	IHIT38968	DSMZ Braunschweig, Germany
Streptococcus equi ssp. *zooepidemicus*	DSM 20727, ATCC 43079, NCDO 1358, NCTC 7023	IHIT38969	DSMZ Braunschweig, Germany
Escherichia coli	W3110	IHIT24619	IMT FU Berlin, Germany
Escherichia coli	C600	IHIT323	Helge Karch, Julius Maximilians University Würzburg, Germany
Salmonella enterica ssp. *enterica* serovar Braenderup	H9812	IHIT22759	Robert Koch Institute Wernigerode, Germany
Enterococcus faecalis	DSM 2570, ATCC 29212	IHIT9729	DSMZ Braunschweig, Germany
Staphylococcus aureus	DSM 2569, ATCC 29213	IHIT9730	DSMZ Braunschweig, Germany
Pseudomonas aeruginosa	DSM 1117, ATCC 27853	IHIT9728	DSMZ Braunschweig, Germany
Escherichia coli	DSM 1103, ATCC 25922	IHIT8068	DSMZ Braunschweig, Germany

* DSMZ: German Collection of Microorganisms and Cell Cultures GmbH, IMT FU: Institute of Microbiology and Epizootic Diseases Free University of Berlin.

3.1.4 Samples, bacterial isolates, and metadata

Clinical samples from equines investigated in this study were submitted to the Institute for Hygiene and Infectious Diseases of Animals (IHIT), JLU Giessen, from January 2019 to September 2020 for routine diagnostic purposes. Samples had been submitted by private veterinary practices as well as from the Clinic for Horses (Internal Medicine and Surgery) and from the Clinic for Obstetrics, Gynecology and Andrology, JLU Giessen. A total of 2,338 samples were investigated, 1,888 samples from private veterinary practices and 450 samples from the above mentioned veterinary clinics. With the exception of three samples from donkeys, all samples were from horses. During this time frame, 257 putative non-duplicate *S. equi* isolates (one isolate from the same equine at the same time of isolation) were collected.

Equine *See* isolates from bacterial culture collections at the following institutions were also investigated in this study: (a) IHIT, JLU Giessen (year 2002-2018, n = 45 isolates); (b) Institute of Microbiology and Epizootic Diseases (IMT), Free University (FU) Berlin, Germany (year 2001-2019, n = 29); (c) IDEXX GmbH Ludwigsburg, Germany (year 2019-2020, n = 171); and (d) School of Veterinary Medicine and Biomedical Sciences, IPB University (SVMBS IPB University), Indonesia, (year 2018, n = 7). Additionally, equine *Sez*

isolates from bacterial culture collection of IHIT JLU Giessen (year 2014-2018 n = 118) and an *Sez* isolate from SVMBS IPB University were investigated.

Metadata were collected for each sample or *S. equi* isolate, respectively, based on information submitted by the veterinarians. The metadata collected were date of isolation, sample type, information about the host (species, age, sex, breed, clinical signs), as well as the geographical location of the owner and the veterinarian.

3.2 Methods

3.2.1 Bacterial isolation, identification, and preservation

3.2.1.1 Sample collection and processing

Samples collected from equines were immediately shipped by the respective veterinarian to the diagnostic laboratories at ambient temperature. All samples were systematically examined for *S. equi* by standard bacterial culture methods. Briefly, samples were cultured on a 5 % (v/v) sheep blood agar plate (BAP) as well as on a 5 % (v/v) sheep blood Columbia agar supplemented with colistin sulphate and oxolinic acid (COBA), and incubated at 37 °C. BAP plates were incubated aerobically, whereas COBA plates were incubated in a 10 Vol % CO_2 atmosphere. The agar plates were inspected after 24 and 48 hrs of incubation. Single colonies of beta-haemolytic, Gram-positive cocci were then sub-cultured on BAP aerobically at 37 °C for 20 ± 2 hrs prior to further examination.

3.2.1.2 Processing of bacterial isolates

S. equi isolates from culture collections were cultured on BAP aerobically at 37 °C for 20 ± 2 hrs. Single beta haemolytic colonies were then sub-cultured on BAP and were regarded as putative beta-*Streptococcus* isolates to be submitted for further identification and re-assessment.

3.2.1.3 Microscopic assessment of bacterial isolates

Cell morphology, including Gram staining [38,125], was examined by bright field microscopy of fixed smears at a magnitude of 1,000 with a stereo microscope (SZX2).

3.2.1.4 MALDI-TOF mass spectrometry

All bacterial isolates were assessed for their species assignment by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS). MALDI-TOF MS was performed with a Microflex LT mass spectrometer V.3.3.1.0, Bruker Daltonics. The resulting spectra were compared against reference spectra (MBT 7854 MSP library, Bruker Daltonics) according to the manufacturer's instructions. Briefly, a bacterial colony was selected from the BAP and applied to a spot on the target plate (Micro Scout Plate 96 target polished steel BC) directly (**direct transfer, dt**) or after acetonitrile-formic acid extraction (**extended direct transfer, edt**) according to the manufacturer's instructions for Bruker Biotyper. For the dt method, 1 µl of matrix, a saturated solution of alfa-cyano-4-hydroxycinnamic acid in 50 % acetonitrile and 2.5 % trifluoroacetic acid, was applied onto the colony. For the edt method, 1 µl 70 % acetonitrile-formic acid were applied onto the colony and allowed to dried prior to the application of the matrix.

According to the MALDI-TOF MS scoring scheme provided by the manufacturer, resulting score values of a bacterial isolate were classified as follows:

2.30–3.00: highly probable species identification,
2.00–2.29: secure genus identification, probable species identification,
1.70–1.99: probable genus identification,
0.00–1.69: no reliable identification.

3.2.1.5 Preservation

Bacteria of the respective *S. equi* isolate were grown aerobically on BAP at 37 °C for 20 ± 2 hrs. Afterwards, bacteria were harvested by flushing the plate with 800 µl of Luria-Bertani (LB) broth containing 30 % (v/v) glycerine. The resulting bacterial suspension was transferred into a 1 ml Nunc Cryobank tube and homogenized by repeated up-and-down pipetting using a 1 ml tip. Finally, the tube was regarded as glycerine stock and stored at -70 °C until further use.

3.2.2 Antimicrobial susceptibility test

The susceptibility of *See* isolates to 16 antimicrobials was assessed by broth microdilution according to Clinical and Laboratory Standards (CLSI) standards. Minimum inhibitory concentrations (MICs) were determined with the Micronaut Veterinary System using Micronaut-S microtitration test plates for large animals 4 (E1-150-100) and the

Micronaut-H Medium according to the manufacturer's instructions. Bacteria cultured on BAP (aerobically, 37 °C, 20 ± 2 hrs) were used for the broth microdilution test. After inoculation with the tested bacteria, the Micronaut-S test plates were incubated in 10 Vol% enriched CO_2 atmosphere at 37 °C for 22 to 24 hrs. The optical density $(OD)_{405}$ of the bacterial growth was measured by photometric reading using a microplate photometer (Multiskan™ FC). MICs were interpreted using breakpoint tables by CLSI [34] and a published study [156], respectively (**Table 4**).

Table 4. Antimicrobials tested in this study and the MIC breakpoints used for interpretation

Class	Antimicrobial	Concentrations tested [µg/ml]	MICs [µg/ml]		
			S	I	R
Beta-lactams	Penicillin G*	0.0625–8	≤ 0.5	1	≥ 2
	Amoxicillin/ Clavulanic acid	2/1–16/8	-	-	-
	Ampicillin**	0.25–16	≤ 0.25	-	-
	Ceftiofur**	0.125–4	≤ 0.25	-	-
	Cephalothin	1–16	-	-	-
Fluoroquinolons	Enrofloxacin**	0.015625–1	≤ 0.12	0.25	≥ 0.5
Phenicols	Florfenicol	1–8	-	-	-
Aminoglycosides	Gentamicin	0.125–8	-	-	-
	Spectinomycin	4–64	-	-	-
Folate pathways inhibitors	Trimethoprim/ Sulfamethoxazole****	0.25/4.75–2/38	≤ 0.5/9.5	1/19–2/38	4/76
Tetracyclines	Tetracycline***	0.125–8	≤ 2	4	≥ 8
Pleuromutilins	Tiamulin	0.25–32			
Macrolides	Erythromycin***	0.125–4	≤ 0.25	0.5	≥ 1
	Tilmicosin	0.5–16	-	-	-
	Tulathromycin	1–64	-	-	-
Miscellaneous	Colistin	0.5–2	-	-	-

* horse-derived breakpoints for streptococci available [34], ** horse-derived breakpoints for *S. equi* ssp. *equi* and *S. equi* ssp. *zooepidemicus* available from CLSI [34] *** human-derived breakpoints for streptococci [34], **** horse-derived breakpoints for *S. equi* ssp. *equi* and *S. equi* ssp. *zooepidemicus* from Sadaka *et al*. (2017) [156]. (S) susceptible, (R) resistant, (I) intermediate. MICs: Minimum inhibitory concentrations.

3.2.3 Biofilm assay

The biofilm assay was performed by a microtitre plate assay as described by Yi *et al*. (2014) [202] with the following modifications. Biofilm formation by *See* was examined using three culture media, namely Brain Heart Infusion (BHI) broth, Todd-Hewitt broth (THB), and THB supplemented with 1 % human fibrinogen for 24 and 48 hrs, respectively. Bacteria from glycerin stock were cultured on BAP at aerobic atmosphere at 37 °C for 20 ± 2 hrs. Afterwards, the bacterial isolate from BAP was cultured in 3 ml BHI broth and THB, respectively (aerobically, 37°C, 20 ± 2 hrs). Then, the OD_{600} of the bacterial culture was measured using a spectrophotometer (Nanodrop 2000c) and adjusted to OD_{600} = 0.1 by dilution with the respective fresh medium. In those experiments where THB containing fibrinogen was used for bacterial culture, the bacterial overnight culture in THB was diluted with THB supplemented with 1 % human fibrinogen to OD_{600} = 0.1. Each well of a sterile flat-bottomed polystyrene microplate was filled with 200 µl of bacterial culture (OD_{600} = 0.1). Each bacterial culture was placed in triplicates wells for each condition. The plates were incubated aerobically at 37 °C for 24 and 48 hrs, respectively, without shaking. *Escherichia coli* (*E. coli*) strain W3110 (IHIT24619) was used as the positive control. *E. coli* C600 and the medium only were used as the negative control. The plate was sealed with a semipermeable sealing membrane (Breathe-Easy®).

After incubation, the bacterial growth was measured at OD_{595} using a microplate photometer (Multiskan™ FC). Supernatants of bacterial cultures with planktonic bacteria were discarded. The plates were washed three times with aqua demineralisata (aqua demin.) (300 µl/well). Biofilms were fixed with methanol (250 µl/well) for 15 minutes (min). Methanol was discarded, and the plate was air dried for 5 min. Biofilms were stained with 0.1 % (w/v) crystal violet solution (250 µl/well) for 30 min. The crystal violet solution was discarded, and the plate was washed three times with aqua demin. (300 µl/well). Afterwards, each well was loaded with 250 µl of an ethanol:aceton mix (80:20) before plates were shaken for 45 min on a horizontal shaker (Titramax 100) at 300 rpm to dissolve completely the crystal violet that bound to the biofilm. Finally, optical densities at OD_{570} were measured using the microplate photometer (Multiskan™ FC).

Assay data were interpreted using specific biofilm formation (SBF) index [114] as follows:

$$SBF = \frac{AB - CW}{G}$$

SBF = specific biofilm formation

AB = OD_{570} of the attached and stained bacteria

CW = OD_{570} of the stained control wells containing only broth medium to eliminate unspecific or abiotic OD values

G = OD_{595} of cell growth in broth medium after incubation and before staining.

3.2.4 Preparation of bacterial DNA

Genomic DNA was isolated from bacteria according to a published phenol-chloroform extraction method [110] with the following modifications. Bacteria were recovered from the glycerin stock by streaking the bacteria on BAP, followed by aerobic incubation at 37 °C for 20 ± 2 hrs. Bacteria were cultured in 15 ml BHI broth (aerobically, 37 °C, 20 ± 2 hrs), followed by centrifugation at 4 °C for 10 min at 4,500 *x g* (5804R). The supernatant was discarded carefully then the pellet was suspended in 200 µl of TES buffer and 15 µl of lysozyme solution (10 mg/ml) and incubated at 37 °C for 30 min, followed by the addition of 10 µl of 10 % sodium dodecyl sulfate (SDS), and 10 µl proteinase K solution (20 mg/ml). After incubation at 55 °C for 60 min, 12 µl of RNAse (500 µg/ml) was added prior to incubation at 37 °C for 30 min. The resulting solution was treated with phenol-chloroform and then with chloroform. Then, 20 µl natrium-acetate (3 M, pH 5.2) was added. DNA was precipitated from the resulting solution with ethanol, harvested by centrifugation (17,000 *x g*, ambient temperature, 15 min, Micro Star 17) and finally resuspended in elution buffer (10 mM Tris-HCl, pH 8.5).

For *seM* typing, the genomic DNA was isolated from bacteria using the GenElute™ Bacterial Genomic DNA Kit. Briefly, bacteria of *See* isolates were recovered from the glycerin stock by streaking the bacteria onto BAP (aerobically, 37 °C, 20 ± 2 hrs). Then, single bacterial colonies from BAP were cultured in 2 ml BHI broth (aerobically, 37 °C, 20 ± 2 hrs). Bacterial cells were harvested by centrifugation (Micro Star 17) at 16,000 *x g* for 2 min. To allow cell lysis, the pellet was resuspended in a 200 µl lysis solution containing 2 x 10^6 U/ml lysozyme solution and 250 U/ml mutanolysin solution and incubated at 37 °C for 30 min. The resulting solution was supplemented with 20 µl of RNAse solution and

incubated at ambient temperature for 2 min. This step was followed by adding 20 µl of Proteinase K and lysis solution C and incubation at 55 °C for 10 min. DNA was purified from the resulting solution using GenElute spin columns according to the manufacturer's instructions.

3.2.5 Polymerase chain reactions

3.2.5.1 *Streptococcus equi* multiplex PCR (Se-mPCR)

S. equi isolates were tested with a published *Streptococcus equi* multiplex PCR (Se-mPCR) for the presence of the ICE*Se2* locus of *See*, the ICE*Sz*1 locus of *Sez*, and the *sodA* gene (*S. equi*) [37]. Briefly, the bacteria in question were suspended in sterile aqua demin. (200 µl) and boiled for 10 min, followed by centrifugation at 17,000 *x g* for 5 min at ambient temperature using a centrifuge (Micro Star 17). The supernatant was used as a PCR template. The primer set was compiled as described by Cordoni *et al*. (2015) [37].

The PCR mix (30 µl/reaction) was composed as follows: 3 µl primer mix (each primer 5 µM: sodAGC1 F and R, ICESz1GC5 F and R, ICESe2GC2 F and R), 3 µl 10X DreamTaq Green Buffer, 1 µl 10 mM dNTPs, 0.2 µl DreamTaq DNA Polymerase solution, 19.8 µl sterile ddH_2O and 3 µl of the template. The established PCR protocol was performed on a thermal cycler (Biometra TRIO 48) using the following program: 94 °C for 5 min (initial denaturation), 35 cycles at 94 °C for 30 sec, 54 °C for 30 sec, and 72 °C for 1 min (amplification), and 72 °C for 5 min (final elongation). In each PCR run, reference strains of *See* (DSM 20521) and *Sez* (DSM 20727), as well as sterile ddH_2O, were used as positive and negative controls, respectively.

PCR products were separated by electrophoresis through 2 % (w/v) agarose gels containing ethidium bromide (f.c. 0.01 µg/ml). Electrophoresis was performed at 100 V for 1 hr and 15 min with 1 TAE as running buffer. Separated PCR products were visualized by UV illumination of the agarose gels on a transilluminator. The length of PCR products (in base pairs, bp) was calculated by comparison of their migration distances with those of reference DNA fragments contained in the GeneRuler 100 bp Plus DNA Ladder. The presence and absence of target amplicons were assessed by comparing existing versus expected lengths of PCR products as listed in **Table 5**. Images of stained gels were documented with an image documentation system (E.A.S.Y B-1393-347C).

Table 5. Oligonucleotide primers used in this study

Primer	Sequence 5′–3′	DNA target	Amplicon size [bp]	Reference
Se-mPCR				
sodAGC1F	TAAGCACCATGCCACCTATG	*sodA* of *S. equi*	380	[37]
sodAGC1R	TTGCCCTCTGAGATTGGTGT			
ICESE2GC2F	TTACCTCCATTACTTGACAATCCAT	ICE*Se*2 of *S. equi* ssp. *equi*	201	[37]
ICESE2GC2R	GATTTGCAACATGAAACATTTACAG			
ICESz1GC5F	CTTTTCTTCACCGCCCACT	ICE*Sz*1 of *S. equi* ssp. *zooepidemicus*	158	[37]
ICESz1GC5R	TGAGCTTTGGAGAAATGGAA			
***seM* PCR**				
ASW73	CAGAAAACTAAGTGCCGGTG	*seM*	541	[84]
ASW74	ATTCGGTAAGAGCTTGACGC			

3.2.5.2 *seM* PCR

The *seM* PCR was part the *seM* typing procedure and was essentially performed as described by Kelly *et al*. (2006) [84]. A 541 bp segment of the N-terminal region of the *seM* gene unique to *See* was amplified by PCR using published primers ASW73 and ASW74 [77,84,138]. The PCR mix (40 µl/reaction) was composed as follows: 1 µl forward primer ASW73 (10 µM), 1 µl reverse primer ASW74 (10 µM), 4 µl 10X DreamTaq Green Buffer, 1 µl 10 mM dNTPs, 0.3 µl DreamTaq DNA Polymerase (5 U/µl) and 27.7 µl of sterile ddH_2O, and finally, 5 µl of the respective bacterial DNA solution. Bacterial DNA of *See* strain DSM 20561 and sterile ddH_2O served as positive and negative controls, respectively.

The *seM* PCR was performed on a thermal cycler (Biometra TRIO 48) with 94 °C for 5 min (initial denaturation), 35 cycles at 94 °C for 30 sec, 58 °C for 30 sec, and 72 °C for 1 min (amplification), and 72 °C for 5 min (final elongation).

In order to confirm that the target DNA was amplified successfully, a sample of the PCR products was analysed by agarose gel electrophoresis as described for the Se-mPCR (see **chapter 3.2.5.1**).

Both strands of the amplified N-terminal region of the *seM* gene were sequenced as a commercial service provided by LGC Genomics GmbH (Berlin, Germany). The PCR products in question (10 µl) and solutions of primers ASW73 and ASW74 (4 µl, 5 µM) were submitted to LGC Genomics GmbH in sealed 1.5 ml tubes. Nucleotide sequences of the both DNA strands was determined using a DNA Analyzer (3730XL) by LGC Genomic GmbH. DNA sequence data were received as chromatograms in AB1 format.

3.2.6 Macrorestriction analysis

Selected *S. equi* isolates were compared by macrorestriction analysis including PFGE according to an in-house protocol adapted from Soedarmanto *et al*. (1996) [169] with modifications. Bacteria from the glycerin stock were cultured on BAP at aerobic atmosphere at 37 °C for 20 ± 2 hrs. Then, bacteria were cultured in BHI broth at 37 °C for 20 ± 2 hrs with shaking. Bacteria were harvested from the resulting suspension by centrifugation (4,700 *x g*, Micro Star 17) at ambient temperature for 10 min. Bacteria were washed with TE buffer and suspended in 250 µl TE buffer. This bacterial suspension was mixed with the same volume of 1.5 % melted agarose (Pulsed Field Agarose Biozym Gold in TE buffer) in a special plug mold and solidified at ambient temperature for 15 min. Bacteria were lysed by subsequent incubation of the resulting gel plug in 300 µl lysis buffer supplemented with 30 µl of lysozyme solution (10 mg/mL) at 37 °C for 24 hrs. This step was followed by proteinase K digestion (10 µl proteinase K, 20 mg/ml) at 56 °C overnight (20 ± 2 hrs). Subsequently, proteinase K was inactivated by serial washing steps: two times with 500 µl TE buffer for 30 min each, two times with 500 µl TE buffer supplemented with 5 µl phenylmethyl sulfonyl fluoride (PMSF) solution each at 56 °C (400 rpm, Thermomixer comfort) for 1 hr, and finally two times with TE buffer each at ambient temperature for 30 min.

Genomic bacterial DNA was digested by incubation of gel plug duplicates either with *Sma*I or *Apa*I [89] at 30 °C for 4 hrs. Each gel plug duplicate was digested with *Sma*I (20 U) in 1X Tango Buffer (2 µl *Sma*I 10 U/µl, 30 µl 10X Tango Buffer, 268 µl ddH_2O) or *Apa*I (30 U) in 1X Buffer B, respectively. Gel plug containing genomic DNA of *Salmonella* Braenderup [75] digested with 450 µl solution of *Xba*I (60 U) in 1X Tango Buffer was used as reference marker. The plug was finally washed with 500 µl of TE buffer. Each gel plug was loaded into the sample wells of a 25-well 1.2 % agarose gel (Pulsed Field Certified Agarose). Macrorestriction DNA fragments were separated in a CHEF Mapper XA PFGE system with 0.5X TBE buffer with thiourea at 14 °C with switch time ramping for 1 to 13 sec for 20 hrs at 6 volts/cm. Subsequently, the gel was incubated in ethidium bromide solution

(4 mg/l) for 30 min. Stained DNA fragments were visualized by UV illumination of the agarose gels on a transilluminator, and documented with the image documentation system (E.A.S.Y B-1393-347C).

3.2.7 Whole genome sequencing

For whole genome sequencing of selected *See* (n = 191) and *Sez* isolates (n = 4), a collaboration with Dr. Torsten Semmler, NG-1 Microbial Genomics, Robert Koch Institute (RKI), Berlin, Germany was initiated. The *See* isolates for whole genome sequencing were selected based on their geographical origin (country and federal states), year of isolation, and sample type. Two *See* isolates that had ICE*Se*2 negative results in the Se-mPCR were included. For *See* isolates from Germany, at least one isolate was chosen from every federal state where the isolates were collected. If more than 10 isolates originated from the same district, all isolates from these districts were submitted for whole genome sequencing. All isolates from foreign countries were submitted for whole genome sequencing as well. The *Sez* isolates were selected as representatives of ICESz1 negative isolates in the Se-mPCR.

Genomic DNA suspended in elution buffer (50 ng/µl) (see **chapter 3.2.4**) was sent to RKI. WGS libraries were generated using the Nextera XT DNA Library Preparation Kit (Illumina, USA) following the manufacturer's instructions at RKI. Whole genome sequencing was employed using a MiSeq sequencer (MiSeq Reagent Kit V.3; Illumina), resulting in 300 bp paired-end reads with an average coverage of 90x.

All Illumina paired-end reads were adapter-trimmed using Flexbar V.3.0.3 [151] and corrected by BayesHammer [130]. The *de novo* assembly of the genomes was employed using SPAdes V.3.11.1 [12]. The genomes were annotated with Prokka V.1.13 [160]. WGS data were received from RKI in FASTA and GenBanK (GBK) formats, respectively.

3.2.8 *In silico* analysis of DNA sequencing data

3.2.8.1 *seM* typing

The nucleotide sequence of a 327 bp internal fragment of the *seM* gene was used for *seM* typing of *See* isolates [77,84,138]. Target sequences were taken from the WGS data of selected *See* isolates (n = 191) and from *seM* gene sequencing data obtained for *See* isolates that were not submitted to whole genome sequencing (n = 74). For non-WGS data,

only high-quality chromatograms of the N-terminal *seM* gene were used for analysis. Chromatograms obtained for each DNA strand of the *seM* PCR amplicon were aligned to generate a consensus sequence of the 327 bp target sequence [84] using Geneious V.8.1.3 (Biomatters Ltd, Auckland, New Zealand).

DNA sequences in FASTA format were queried against the *seM* alleles using the publicly available software and database in PubMLST hosted by the University of Oxford, UK [80]. The *seM* alleles were assigned by comparison with the *seM* allele sequences in the PubMLST database of *S. equi* (http://pubmlst.org/szooepidemicus/) [80]. New alleles were submitted to the curator of PubMLST for allele number designation and for integration into the PubMLST database. In order to analyse the phylogenetic relationship among identified *seM* alleles, a neighbour-joining tree was constructed using Geneious V.8.1.3 (Biomatters Ltd).

3.2.8.2 Multilocus sequence typing (MLST)

WGS data of selected *See* isolates (n = 191) were used for multilocus sequence typing (MLST) according to the scheme of Webb *et al*. (2008) [196]. Briefly, nucleotide sequences of defined internal fragments of the following seven housekeeping genes were used for MLST: *arcC*, *nrdE*, *proS*, *spi*, *tdk*, *tpi*, and *yqiL* [196]. Assembled whole genome sequences in FASTA format were queried against all seven loci of interest of the MLST scheme of *See/Sez* using the software and database integrated in PubMLST [80]. For each locus, every sequence was assigned a specific allele number. Each sequence type (ST) was defined by a series of seven integers (the allelic profile) corresponding to the allele numbers at the seven gene loci sequenced. The allele numbers and STs were determined according to the MLST database for *S. equi* maintained at the University of Oxford (http://pubmlst.org/szooepidemicus/) [80,196].

3.2.8.3 Core genome MLST (cgMLST)

WGS data of selected *See* isolates (n = 191) were used for core genome MLST (cgMLST) analysis according to the scheme of Mitchell *et al*. (2021) [122]. This scheme comprised 1,286 core genes from the *See* reference isolate *See* strain 4047 by excluding mobile genetic elements (MGEs) (φSeq1, φSeq2, φSeq3, φSeq4, ICE*Se*1, and ICE*Se*2), insertion sequences, sortase-processed proteins, and repeats (*hasC1*, *hasC2*) [122]. Assembled *See* genomes in FASTA format were uploaded and analysed using the software packages in Pathogenwatch (https://pathogen.watch/). Briefly, BLAST matches of the

respective 1,286 loci in each genome relative to the core genome of *See* strain 4047 were extracted and aligned using MAFFT [83]. The tresholds used were both 80 % for sequence identity and length. Hits below this treshold were removed as fragments [122]. Alleles were assigned based on sequence differences in comparison to the *See* strain 4047 core genome (GenBank accession number: FM204883.1). Insertions and deletions in core genes of the isolate in question were not taken into account in the next step of the analysis. The mid-point rooted phylogenetic tree was constructed using Analysis of Phylogenetics and Evolution (APE) package [137] and phangorn package [158] as integrated in the Pathogenwatch [122]. The population structure was extrapolated using Bayesian Analysis of Population Structure (BAPS) v. 6.0 [36]. The cgMLST data of *See* isolates of this study (n = 191) were compared to those of 759 *See* isolates from other investigators that were retrieved from Pathogenwatch [62,71,122–124]. Phylogenetic reconstruction and metadata were visualized using Microreact [7].

3.2.8.4 Genome-wide search for virulence-associated genes

WGS data of selected *See* isolates (n = 191) were used *in silico* for a genome-wide search for virulence-associated genes (VAGs). A self-created database of VAGs was used for this screening. This database contained VAG sequences of *See* isolates from horses and was generated from published peer-reviewed papers including reviews as listed in **Appendices Table 22.** All VAG sequences were collected from the NCBI nucleotide database (https://www.ncbi.nlm.nih.gov/nuccore/). Assembled WGS data of *See* isolates in FASTA format were screened *in silico* for VAGs using MyDbFinder V.2.0 (https://cge.cbs.dtu.dk/services/MyDbFinder/). Sequence identity and coverage length thresholds were both 60 %. The VAGs were classified into adherence, immune modulation, and nutritional/metabolic factors as well as exoenzymes and exotoxins.

3.2.8.5 Genome-wide search assessment for antimicrobial resistance genes

WGS data of selected *See* isolates (n = 191) were used *in silico* for a genome-wide search for acquired resistance genes and chromosomal mutations mediating antimicrobial resistance using ResFinder V.4.1 (https://cge.food.dtu.dk/services/ResFinder/) [21,205,206], maintained by the Center for Genomic Epidemiology (CGE). Assembled WGS data in FASTA format were uploaded to the online bioinformatic tool ResFinder V.4.1 and screened against the acquired resistance genes. The ResFinder V.4.1 database comprises 2,690 antimicrobial resistance genes [21]. Since ResFinder 4.1 did not provide a database of

antimicrobial resistance due to chromosomal point mutations specific for streptococci, the respective database of *Staphylococcus* (*Staph.*) *aureus* was consulted. The selected thresholds were both 60 % for sequence identity and coverage length.

3.2.8.6 Genome-wide search for antigens used in Strangvac® vaccine

WGS data of selected *See* isolates (n = 191) were used *in silico* for a genome-wide search for sequences encoding the eight *See* protein fragments that were used as antigens in the commercial vaccine Strangvac® (Intervacc, Sweden). Nucleotide sequences encoding these antigens were retrieved from available genome data of *See* strain 1866 (NCBI Reference Sequence: NZ_CWFT01000000.1) [71], namely SEQ0935 (CNE), SEQ0855 (SclF), SEQ1817 (SclI) and SEQ2101 (SclC), SEQ0721 (EAG), SEQ0402 (Eq8), SEQ0256 (Eq5), and SEQ0999 (IdeE) [62,148]. Amino acid sequences of these antigens were kindly provided by Dr. Andrew Waller (Intervacc AB, Sweden). The *See* strain 1866 was isolated in Sweden in 2000 and was used as the Strangvac® vaccine sheet [62,65,148]. Information about the antigens used in Strangvac® is listed in **Table 6**. Geneious V.8.1.3 and MyDbFinder V.2.0 (https://cge.cbs.dtu.dk/services/MyDbFinder/) were used to detect target nucleotide sequences in WGS data. Sequence identity and coverage length thresholds were 80 % and 60 %, respectively. Geneious V.8.1.3 was also used for the translation of nucleotide sequences into amino acid (aa) sequences, as well as for alignments of nucleotide and aa sequences, respectively.

Table 6. Antigens contained in the recombinant protein vaccine Strangvac®

Name of the protein in the vaccine	Source of the antigen		Protein fragment used as antigen		
	Protein designation and locus tag *	**Size of the preprotein; size of its CDS**	**Size [aa]; aa positions**	**CDS length [nt]**	**CDS positions**
CCE (fusion protein)	**CNE** (SEQ0935)	657 aa; 1,974 nt	304; 28-331	912	82-993
	SclF (SEQ0855)	364 aa; 1,095 nt	50; 37-86	150	109-258
	SclI (SEQ1817)	491 aa; 1,476 nt	56; 37-92	168	109-276
	SclC (SEQ2101)	302 aa; 909 nt	58; 38-95	174	112-285
	EAG (SEQ0721)	429 aa; 1,290 nt	161; 35-195	483	103-585
Eq85 (fusion protein)	**Eq8** (SEQ402)	373 aa; 1,122 nt	196; 30-225	588	88-675
	Eq5 (SEQ0256)	608 aa; 1,827 nt	440; 39-478	1,320	115-1,434
IdeE (single protein)	**IdeE** (SEQ0999)	349 aa; 1,050 nt	315; 35-349	945	103-1,047

* Locus tag number in *See* strain 4047 (GenBank accession number: FM204883.1); aa: amino acids, nt: nucleotides. Data retrieved from Robinson *et al*. (2018) and Frosth *et al*. (2022) [62,148].

3.2.9 Statistical analysis

Statistical analyses were supported by the working group of Biomathematics and Data Processing, Faculty of Veterinary Medicine, JLU Giessen, Germany. Statistical analyses were performed using SAS 9.4 (SAS® Institute Inc, North Carolina, USA) and IBM SPSS Statistics V.27 (IBM Deutschland GmbH, Ehningen, Germany).

The Fisher's exact test was conducted to evaluate the frequency distribution of *seM* allele groups, STs, cgMLST BAPS clusters in relation to the data of the isolates (sample types, years of isolation, and geographical locations). Furthermore**,** Fisher's exact

test was used to determine whether proportions for one variable (antimicrobial susceptibilities) were different between phylogenetic groups (*seM* allele groups, STs, cgMLST BAPS clusters). In case of a significant correlation (p value ≤ 0.05) between the variables, the analyses were continued using multiple comparisons by considering Bonferroni correction.

For biofilm formation by *See*, the correlation between OD_{570} and SBF index was analysed using the Spearman's correlation test. The Wilcoxon signed-rank test was used to analyse the effect of various media and incubation times on biofilm formation. The association between the biofilm formation and phylogenetic groups was analysed by Wilcoxon-Mann-Whitney test.

The p values were considered as follows:

$p > 0.05$: not significant
$p \leq 0.05$: significant (*)
$p \leq 0.01$: moderate significant (**)
$p \leq 0.001$: highly significant (***).

4 Results

4.1 Subspecies-specific identification of putative *Streptococcus equi* isolates

4.1.1 Cell morphology and Gram staining

This study investigated 628 non-duplicate putative *S. equi* isolates. The term “non-duplicate” refers to the approach that only one isolate from the same equine at the same time of isolation was included in the study. Assessment of bacterial growth characteristics showed that isolates displayed single colonies that were circular, convex, and small (approximately 1–2 mm in diameter) on BAP after aerobic incubation at 37 °C for 24 hrs. Additionally, all single colonies were surrounded by a large translucent haemolytic area (approximately 3–5 mm in diameter). Most isolates grew with non mucoid single colonies (n = 440, 70.1 %), while the other isolates displayed a mucoid colony phenotype (n = 188, 29.9 %). Microscopically, all putative *S. equi* isolates presented themselves as Gram-positive cocci arranged in chains or pairs.

4.1.2 MALDI-TOF MS scores

All putative *S. equi* isolates (n = 628) were re-assessed by MALDI-TOF MS analysis for their reported or presumptive subspecies assignment. Most isolates (n = 556, 88.5 %) were successfully identified by MALDI-TOF MS and Bruker's MALDI-Biotyper database using the dt protocol, with score values greater than 2.0 (**Figure 2**). Isolates that achieved score values smaller than 2.00 (n = 72, 11.5 %) were re-tested using the edt protocol for preparation of bacteria for analysis. Using this protocol re-tested isolates scored values greater than 2.0 as well. Most of them (n = 48, 66.7 %) scored values of 2.30-3.00, while the others scored at least 2.00-2.29. Thus, MALDI-TOF MS analysis confirmed all presumptive *S. equi* isolates either as *See* (n = 265) or *Sez* (n = 363), respectively (**Figure 2**).

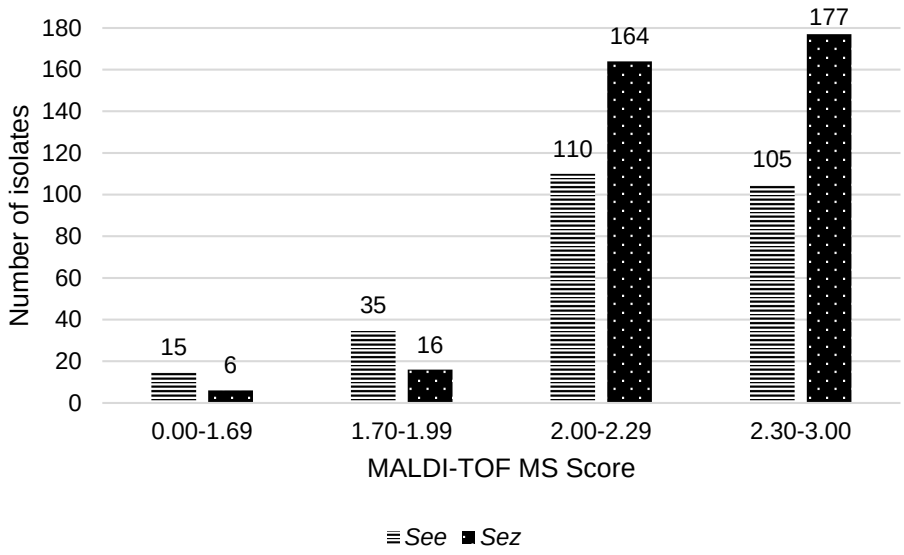


Figure 2. **The number of *S. equi* ssp. *equi* (*See*, n = 265) and *S. equi* ssp. *zooepidemicus* isolates (*Sez*, n = 363) by MALDI-TOF MS scores as they were prepared using the direct transfer (dt) protocol.**

4.1.3 Presence of *Streptococcus equi* signature loci as detected by multiplex PCR (Se-mPCR)

A published *S. equi* multiplex PCR (Se-mPCR) that detects the ICE*Se*2 signature locus of *See*, the ICE*Sz*1 signature locus of *Sez*, and the *sodA* gene (*S. equi*) [37] was used to determine the subspecies affiliation of isolates presumptively identified as *S. equi*.

The published Se-mPCR successfully detected the *sodA* gene (amplicon size: 380 bp) in all tested *S. equi* isolates. Furthermore, 263 of 265 isolates (99.2 %) that identified as *See* by MALDI-TOF MS also harboured the signature locus ICE*Se2* of that subspecies, and none of them harboured the ICE*Sz*1 locus of *Sez* (**Figure 3, Table 7**). The ICE*Se2* locus was only detected in isolates that were identified as *See* by MALDI-TOF MS, not in those identified as *Sez*. Whole genome sequencing revealed that the two *See* isolates that proved PCR-negative for ICE*Se*2, namely *See* isolates IHIT38875 and IHIT43005, actually harboured only parts of the ICE*Se*2 locus (**Figure 4**). The reference genome *See* strain 4047 had an ICE*Se*2 locus of 60,055 bp, while the ICESe2 locus of *See* IHIT38875 and *See* IHIT43005 comprised only 20,901 bp and 48,709 bp, respectively. In particular the ICE*Se*2 specific primer binding sites were missing in both of these isolates.

None of the isolates that identified as *Sez* by MALDI-TOF MS harboured the ICE*Se*2 locus. But surprisingly, only 44.6 % of these isolates harboured the respective PCR target ICE*Sz*1 (**Table 7**). WGS data of four PCR-ICE*Sz*1-negative *Sez* isolates (IHIT28143,

IHIT36273, IHIT36612, IHIT39397) revealed that the primer binding sites of ICE*Sz*1 were missing in these isolates. In *Sez* IHIT28143 and *Sez* IHIT36273 the entire ICE*Sz*1 locus was missing, while in *Sez* IHIT36612 and *Sez* IHIT39397 only 3,843 bp and 9,123 bp of 47,573 bp of the entire ICE*Sz*1 locus of *Sez* strain H70 (GenBank Accession number FM204884.1, locus tag SZO_12560–SZO_12981) were present, respectively (data not shown).

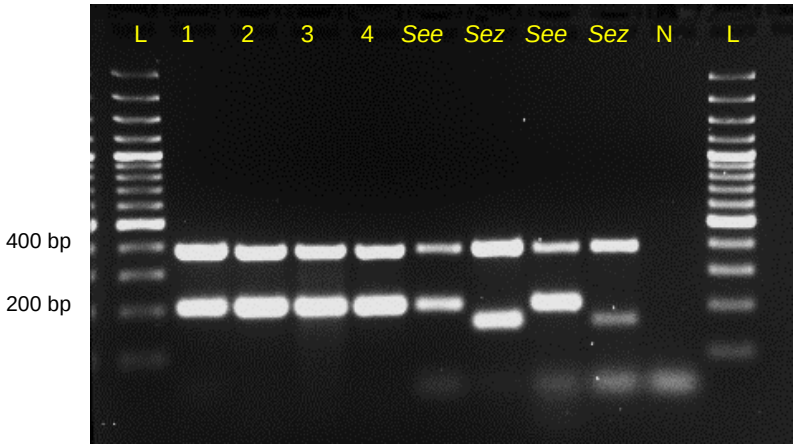


Figure 3. ***S. equi*** **multiplex PCR (Se-mPCR) results for identification of** ***S. equi*** **ssp.** ***equi*** **and** ***S. equi*** **ssp.** ***zooepidemicus***

L: DNA ladder, 1-4: *S. equi* ssp. *equi* (*See*) field isolates, *See*: *See* reference strain DSM 20561, *Sez: S. equi* ssp. *zooepidemicus* reference strain DSM 20727, N: Negative control. Amplicons sizes of the detected genes: 380 bp (*sodA*), 201 bp (ICE*Se*2), 158 bp (ICE*Sz*1).

Table 7. Comparison between the results of MALDI-TOF MS and the results of Se-mPCR for *S. equi* identification

MALDI-TOF MS results	**Number of isolates**			
	Tested	**Positive multiplex PCR result**		
		***sodA* (*S. equi*) (380 bp)**	**ICE*Se2* (*See*) (201 bp)**	**ICE*Sz1* (*Sez*) (158 bp)**
S. equi ssp. *equi (See)*	265	265 (100 %)	263 (99.2 %)	0 (0 %)
S. equi ssp. *zooepidemicus (Sez)*	363	363 (100 %)	0 (0 %)	162 (44.6 %)

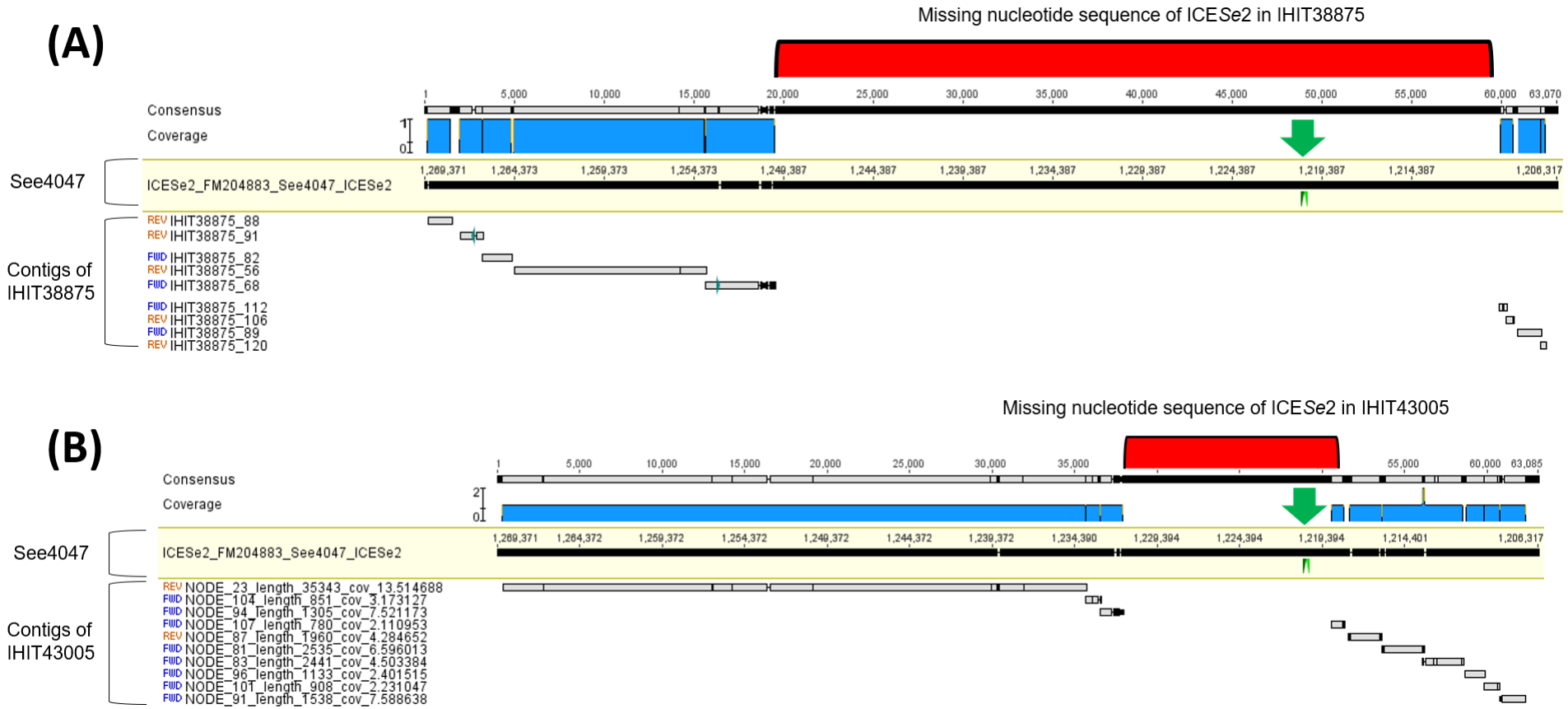


Figure 4. Nucleotide sequence alignment of the ICESe2 loci of *S. equi* ssp. *equi* (*See*) isolates IHIT38875 (A) and IHIT43005 (B) against ICE*Se*2 of the *See* reference strain 4047 (*See* 4047).

The missing nucleotide sequences of ICE*Se*2 in *See* isolates IHIT38875 and IHIT43005 are shown in red, PCR primer binding sites are shown by the green arrow.

4.1.4 Whole Genome Sequencing

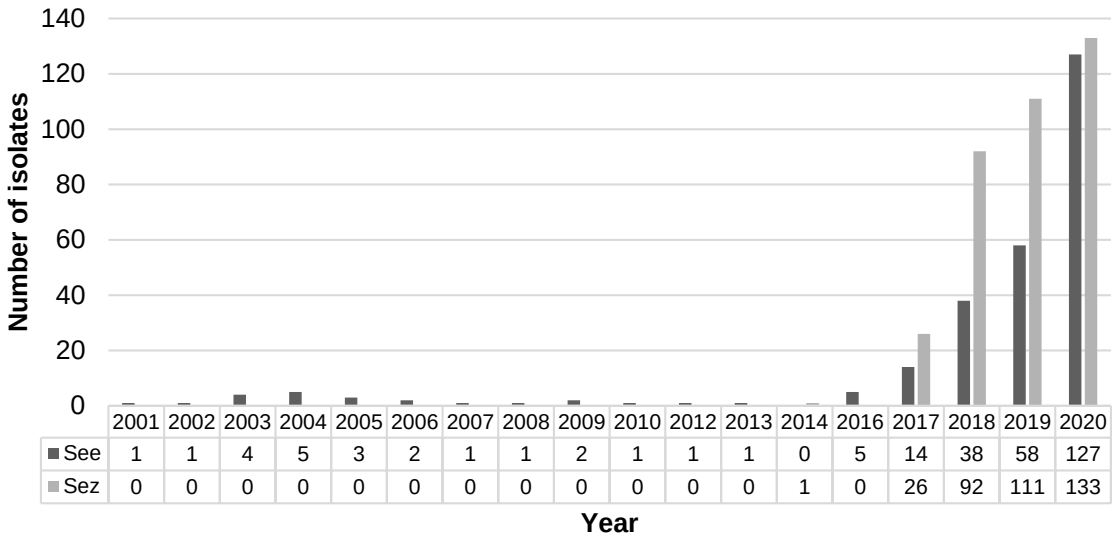
In total, 191 selected *See* isolates were submitted to whole genome sequencing. Subsequently, assembled genomes of these isolates were examined for their species/subspecies affiliation using Speciator, an integrated assignment tool in Pathogenwatch (https://cgps.gitbook.io/pathogenwatch/technical-descriptions/species-assignment/speciator).

All 191 isolates were confirmed as *See*. The arithmetic mean ± standard deviation of the genome sizes was 2,169,885 ± 34,598 bp (2,076,792–2,301,722 bp). The number of contigs was 236.5 ± 91.4 contigs (99–658 contigs), while the percentage of GC content was 41.27 ± 0.1 Mol% (41.1–41.7 Mol%). Genome size, number of contigs, and percentage of GC content for each genome analysed in this study can be accessed in Pathogenwatch at the following URL: https://pathogen.watch/collection/4j4q55j6u7s0-s-equi-equi-ihit.

4.2 Metadata of the *Streptococcus equi* isolates

This study investigated 628 non-duplicate putative *S. equi* isolates recovered from equines in Germany, Denmark, the Netherlands, Austria, Switzerland, and Indonesia (2001–2020). Examination by MALDI-TOF MS, Se-mPCR and WGS data analysis confirmed 265 and 363 isolates as *See* and *Sez,* respectively (**chapter 4.1**).

The *See* isolates had been recovered from field samples submitted by veterinary surgeons to four diagnostic laboratories between 2001 and 2020 (**Figure 5**). Some *See* isolates were isolated by IHIT (JLU Giessen, Germany) in 2002-2020 (n = 58), IMT (FU Berlin, Germany) in 2001-2019 (n = 29), IDEXX GmbH (Ludwigsburg, Germany) in 2019-2020 (n = 171), and SVMBS (IPB University, Indonesia) in 2018 (n = 7). The *Sez* isolates had been isolated from samples submitted by veterinarian surgeons to IHIT in 2014-2020 (n = 362) and to SVMBS in 2018. Most isolates had been isolated in 2017-2020 (n = 599, 95.4 %) (**Figure 5**), since this was the time frame where the collection of the *S. equi* isolates for this study was conducted intensively.



	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2012	2013	2014	2016	2017	2018	2019	2020
See	1	1	4	5	3	2	1	1	2	1	1	1	0	5	14	38	58	127
Sez	0	0	0	0	0	0	0	0	0	0	0	0	1	0	26	92	111	133

Figure 5. The number of *S. equi* ssp. *equi* (*See*, n = 265) and *S. equi* ssp. *zooepidemicus* (*Sez*, n = 363) isolates by year of isolation.

The geographical origin of the isolates was the location of the equine owner or the location of the submitting veterinarian practice. Whenever possible, the owner's location was used for isolates collected by IHIT, IMT, and SVMBS. Due to owner confidentiality, the geographical origin of isolates collected by IDEXX GmbH Ludwigsburg was assigned based on the location of the submitting veterinarian practice.

Most *See* isolates (n = 265) were recovered from equines in Germany (n = 251, 94.7 %) (**Table 8**). For comparison reasons, *See* isolates from Denmark (n = 3), Luxembourg (n = 2), the Netherlands (n = 1), Austria (n = 1), and Indonesia (n = 7) were included in this study. Furthermore, almost all *Sez* isolates were from equines in Germany (n = 359, 98.9 %), except three were from Switzerland and one was from Indonesia (**Table 8**).

The geographic origins of the *S. equi* isolates were unevenly distributed. Nevertheless, this study comprises isolates from 15 of the 16 federal states (*Bundesländer*) in Germany (**Table 8**). *See* isolates from Germany (n = 251) came from 14 federal states, mostly from Hesse (n = 57/251, 22.7 %), North Rhine-Westphalia (n = 52/251, 20.7 %), and Baden-Wuerttemberg (n = 41/251, 16.3 %) (**Table 8**). Seven *See* isolates (n = 7) had been isolated in 2018 from seven horses on four farms in three provinces of Indonesia. Most of the German *Sez* isolates (n = 359) originated in Hesse (n = 255/359, 71 %), followed by Rhineland-Palatinate (n = 37/359, 10.3 %), and North Rhine-Westphalia (n = 36/359, 10 %) (**Table 8**).

Table 8. The number of *S. equi* ssp. *equi* (n = 265) and *S. equi* ssp. *zooepidemicus* (n = 363) isolates by geographic origin

Country	Federal State	*S. equi* ssp. *equi*		*S. equi ssp. zooepidemicus*	
		Number of isolates	Percentage [%]	Number of isolates	Percentage [%]
Germany		251	94.7	359	98.9
	Baden-Wuerttemberg	41	15.5	9	2.5
	Bavaria	17	6.4	4	1.1
	Berlin	14	5.3	0	0
	Brandenburg	20	7.5	0	0
	Hamburg	0	0	1	0.3
	Hesse	57	21.5	255	70.2
	Mecklenburg-Western Pomerania	2	0.8	0	0
	Lower-Saxony	13	4.9	6	1.7
	North Rhine-Westphalia	52	19.6	36	9.9
	Rhineland-Palatinate	20	7.5	37	10.2
	Saarland	2	0.8	7	1.9
	Saxony	1	0.4	0	0
	Saxony-Anhalt	5	1.9	0	0
	Schleswig-Holstein	6	2.3	1	0.3
	Thuringia	1	0.4	3	0.8
Austria		1	0.4	0	0
Denmark		3	1.1	0	0
Luxembourg		2	0.8	0	0
The Netherlands		1	0.4	0	0
Switzerland		0	0	3	0.8
Indonesia		7	2.6	1	0.3
Total		**265**	**100 %**	**363**	**100 %**

The *S. equi* isolates had been cultured from various sample types (**Figure 6, Figure 7**). Nearly 75 % of the *See* isolates had been recovered from the respiratory tract, comprising sample types such as nasal swabs (44.2 %), guttural pouch lavage fluids (25.7 %), nasopharyngeal swabs (3.8 %), and chondroids (0.4 %) (**Figure 6**). Almost a quarter of the *See* isolates was from abscess material (sites not specified) (23.4 %). Most of the *Sez* isolates had been isolated from swabs or lavage fluids collected in the respiratory tract (43.5 %), but many came from swabs and discharges of reproductive organs (34.7 %), abscess material and wound swabs (sites not specified) (12.7 %) (**Figure 7**). In contrast to many *Sez* isolates none of the *See* isolates had been isolated from the reproductive organs.

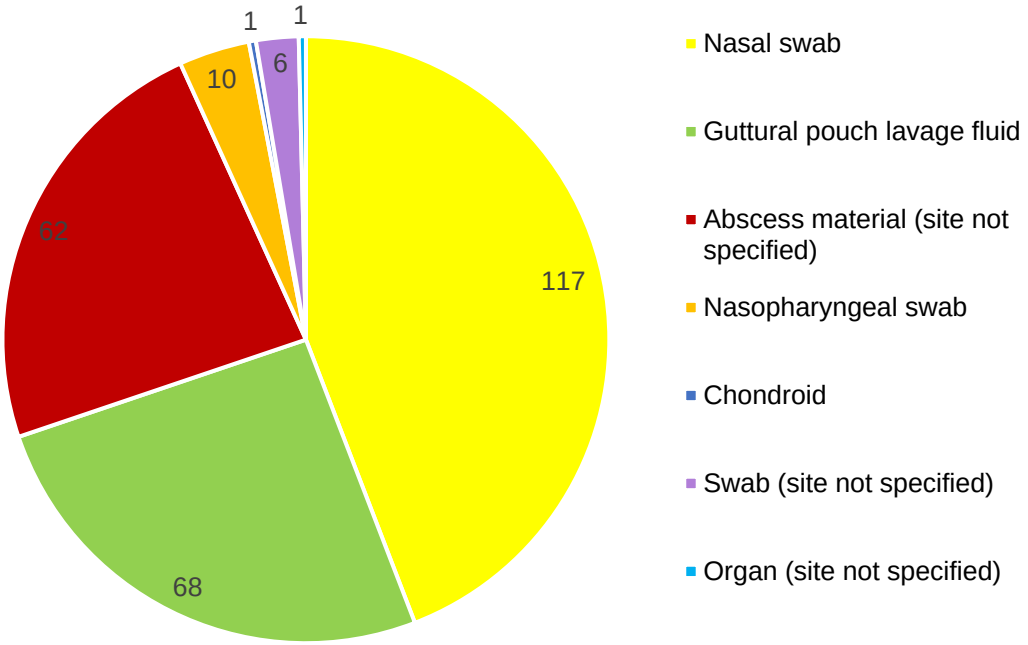


Figure 6. **The number of *S. equi* ssp. *equi* isolates (n = 265) by sample type.**

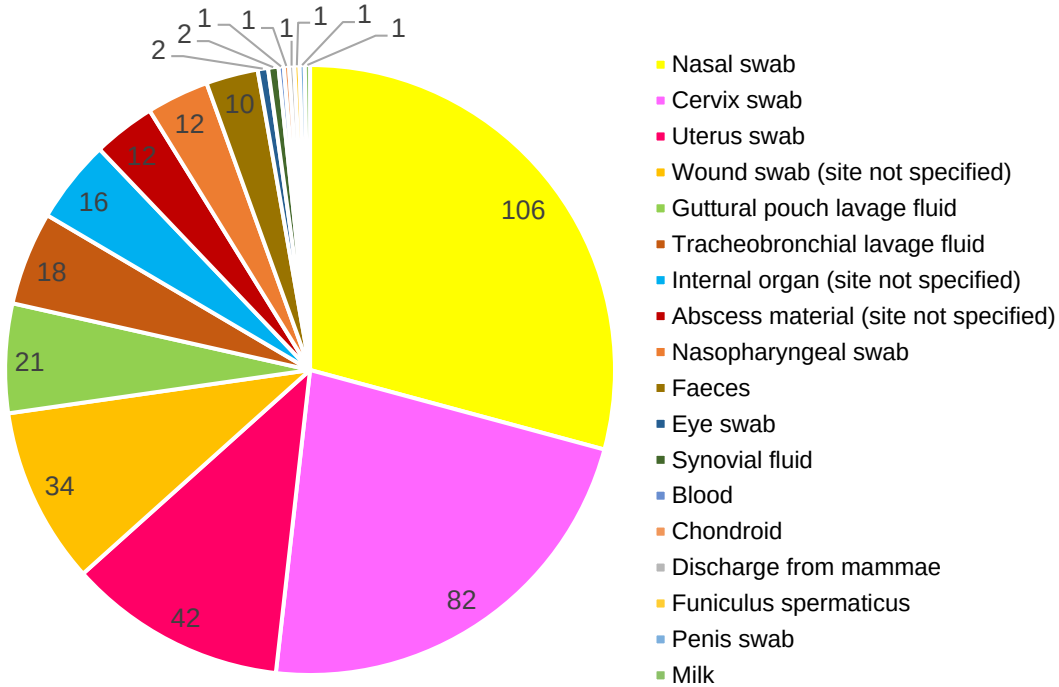


Figure 7. **The number of *S. equi* ssp. *zooepidemicus* isolates (n = 363) by sample type.**

4.3 Phylogenetic analysis of *Streptococcus equi* ssp. *equi* isolates

After the putative *S. equi* isolates were successfully identified as *See* and *Sez* (**chapter 4.1**), the further analyses of this study focused on the *See* isolates only.

4.3.1 Macrorestriction analysis of *Streptococcus equi* ssp. *equi* isolates from Indonesia

In a first step, the clonal relatedness between *See* isolates from Indonesia (n = 11) was investigated by macrorestriction analysis using restriction enzymes *Sma*I and *Apa*I. These 11 *See* isolates from Indonesia were from seven horses, including four isolates that had been isolated from the same horses at the same time ("duplicate isolates"). These four duplicate isolates had the same colony morphology as the isolates from the same horse and were confirmed as *See* by MALDI-TOF MS and Se-mPCR (data not shown). The Indonesian *See* isolates were compared with a *Sez* isolate from Indonesia, selected *See* isolates from Germany (n = 5) as well as *S. equi* reference strains DSM 20561 and 20727. The German isolates (n = 5) had been recovered from horses from various geographical origins and years of isolation.

Although macrorestriction with *Sma*I revealed easier to read banding patterns than macrorestriction with *Apa*I (**Figure 8, Figure 9**) it was obvious that all *See* isolates from Indonesia displayed the same pulsotype with each enzyme. These Indonesian *See* pulsotypes were clearly different from the 8 (*Sma*I) and 7 (*Apa*I) pulsotypes shown by all the other strains and isolates investigated here. Based on these results only one *See* isolate from each horse in Indonesia was selected for subsequent WGS analysis.

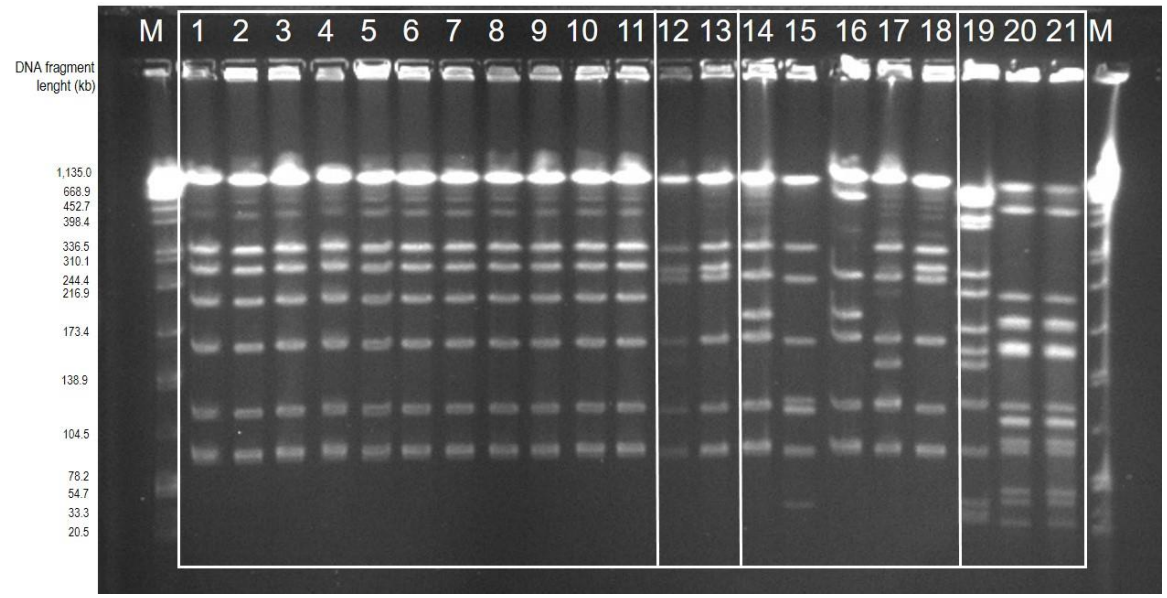


Figure 8. Macrorestriction analysis of *S. equi* isolates and reference strains using *Sma*I.

M: Marker *Salmonella enterica* Braenderup, lanes 1-11 *S. equi* ssp. *equi* (*See*) isolates from Indonesia, lanes 12-13: *See* reference strain DSM 20561, lanes 14-18: *See* isolates from Germany (14: IHIT38875, 15: IHIT12519, 16: IHIT36746, 17: IHIT37552, 18: IHIT38233), lane 19 *S: equi* ssp. *zooepidemicus* (*Sez*) IHIT39397 from Indonesia, lanes 20-21 *Sez* reference strain DSM 20727.

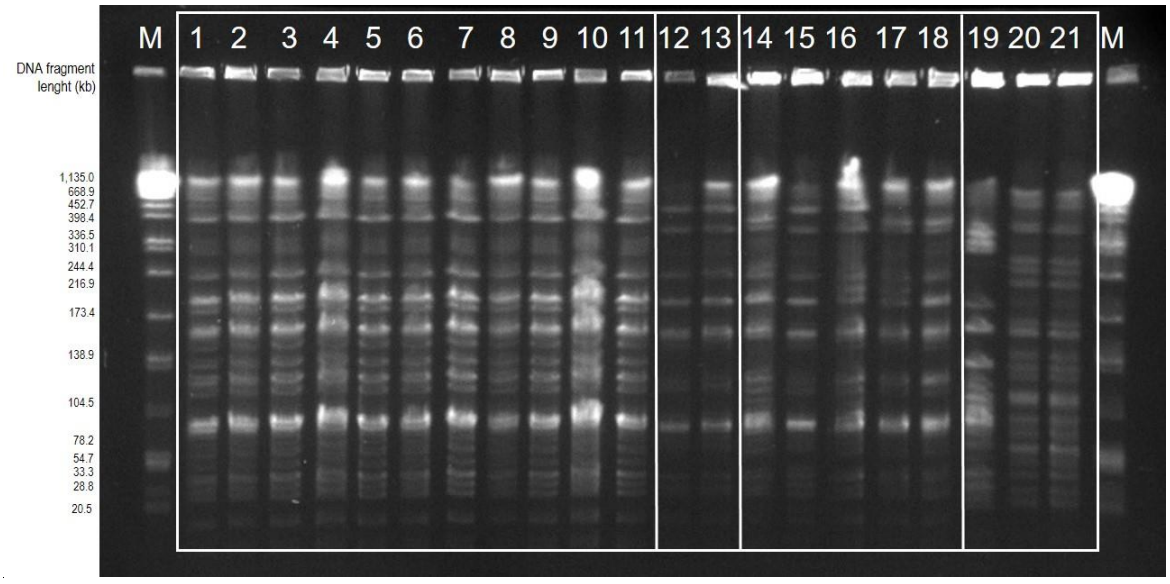


Figure 9. Macrorestriction analysis of *S. equi* isolates and reference strains using *Apa*I.

Legend as in Figure 8.

4.3.2 *Streptococcus equi ssp. equi* diversity according to *seM* typing

All non-duplicate *See* isolates (n = 265) were analysed for their *seM* allele according to the method of Kelly *et al*. (2006) [84]. WGS data or *seM* gene sequencing data were used for this typing method. Novel *seM* alleles were submitted to the PubMLST database (http://pubmlst.org/szooepidemicus/) for allele number designation.

The investigated *See* isolates encoded for a total of 47 different *seM* alleles, including 30 novel *seM* alleles. Identified *seM* alleles differed from each other by one to 13 nucleotides and were classified into five *seM* allele groups according to their phylogenetic relationships, namely *seM*-1, -2, -6, -8, and -9 (**Figure 10, Figure 11**). The most prevalent *seM* allele group was *seM*-9 which comprised 87.9 % of the isolates, followed by groups *seM*-6 (5.7 %), *seM*-8 (1.8 %), *seM*-1 (0.4 %), and *seM*-2 (0.4 %) (**Table 9**). The *seM* alleles of ten *See* isolates (3.8 %) could not be assigned due to deletions, insertions or nonsense mutations in the 327 bp-segment targeted for *seM* typing (**Table 10**). From here on, all of these alleles were referred to as “non-typeable *seM*”.

More than half of the isolates harboured the *seM*-9 allele (n = 133, 50.2 %), which was detected in *See* isolates from four of five European countries (**Table 9**). All *See* isolates from Indonesia (n = 7) carried *seM*-166 allele, which was also detected in seven isolates from Germany. *See* isolates from Germany (n = 251) displayed significant sequence variation in their *seM* gene, presenting 46 different *seM* alleles, including 29 novel *seM* alleles. Two *seM* alleles were detected in *See* isolates from Denmark, namely *seM*-9 and *seM*-77; these alleles differed by only one SNP from each other (T → C, at position 217 in the variable region targeted for *seM* typing. Two isolates from Luxembourg harboured alleles *seM*-9 and *seM*-242 that differed by a SNP at position 166 (G → T).

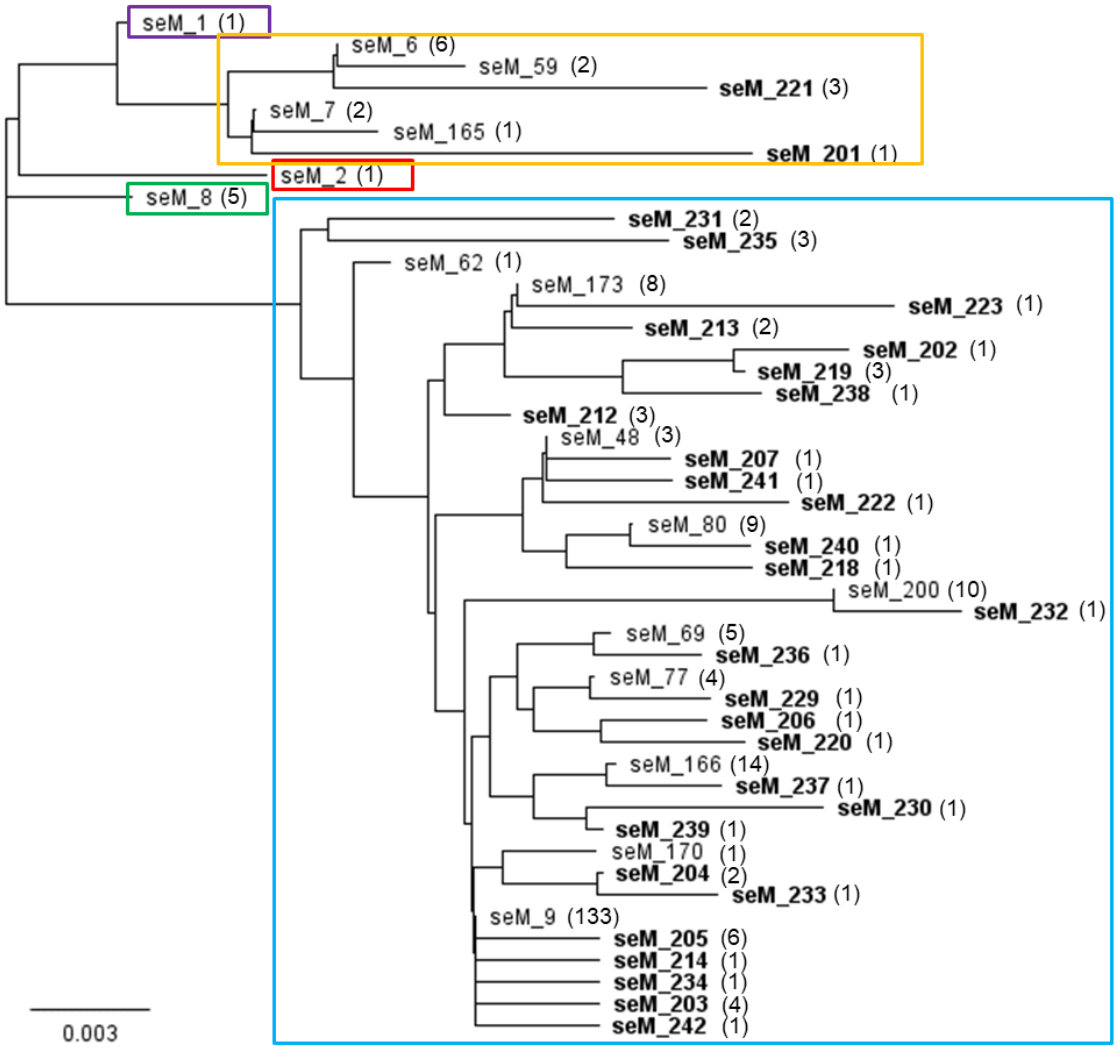


Figure 10. Neighbour-joining phylogenetic tree of *seM* alleles (n = 47) identified in the 265 *S. equi* ssp. *equi* isolates of this study.

Novel *seM* alleles are in bold. The number of isolates of each allele is in brackets. The scale bar represents the number of substitutions per site in the 327 bp-segment targeted for *seM* typing. Coloured frames are used to mark the five *seM*-groups detected (purple: *seM*-1 group; yellow: *seM*-6 group; red: *seM*-2 group; green: *seM*-8 group; blue: *seM*-9 group).

Table 9. Frequencies of *seM* alleles and SeM peptides in the investigated *S. equi* ssp. *equi* isolates (n = 265)

seM group	*seM* allele*	SeM peptide	Number of isolates in each country							Portion [%]
			Germany	Austria	Denmark	Luxembourg	The Netherlands	Indonesia	Total	
1	1	1	1	-	-	-	-	-	1	0.4
2	2	2	1	-	-	-	-	-	1	0.4
6	6	6	6	-	-	-	-	-	6	2.3
	7	7	2	-	-	-	-	-	2	0.8
	59	57	1	1	-	-	-	-	2	0.8
	165	158	1	-	-	-	-	-	1	0.4
	201	191	1	-	-	-	-	-	1	0.4
8	8	8	5	-	-	-	-	-	5	1.9
9	9	9	129	-	2	1	1	-	133	50.2
	48	43	3	-	-	-	-	-	3	1.1
	62	60	1	-	-	-	-	-	1	0.4
	69	66	5	-	-	-	-	-	5	1.9
	77	73	3	-	1	-	-	-	4	1.5
	80	76	9	-	-	-	-	-	9	3.4
	166	159	7	-	-	-	-	7	14	5.3
	170	163	1	-	-	-	-	-	1	0.4
	173	165	8	-	-	-	-	-	8	3.0
	200	190	10	-	-	-	-	-	10	3.8
	202	192	1	-	-	-	-	-	1	0.4
	203	193	4	-	-	-	-	-	4	1.5
	204	183	2	-	-	-	-	-	2	0.8
	205	194	6	-	-	-	-	-	6	2.3
	206	195	1	-	-	-	-	-	1	0.4
	207	196	1	-	-	-	-	-	1	0.4
9	**212**	201	3	-	-	-	-	-	3	1.1
	213	202	2	-	-	-	-	-	2	0.8
	214	203	1	-	-	-	-	-	1	0.4
	218	207	1	-	-	-	-	-	1	0.4
	219	208	3	-	-	-	-	-	3	1.1
	220	209	1	-	-	-	-	-	1	0.4
	221	210	3	-	-	-	-	-	3	1.1

seM group	*seM* allele*	SeM peptide	Number of isolates in each country							Portion [%]
			Germany	Austria	Denmark	Luxembourg	The Netherlands	Indonesia	Total	
	222	211	1	-	-	-	-	-	1	0.4
	223	212	1	-	-	-	-	-	1	0.4
	229	218	1	-	-	-	-	-	1	0.4
	230	219	1	-	-	-	-	-	1	0.4
	231	220	2	-	-	-	-	-	2	0.8
	232	221	1	-	-	-	-	-	1	0.4
	233	222	1	-	-	-	-	-	1	0.4
	234	223	1	-	-	-	-	-	1	0.4
	235	224	3	-	-	-	-	-	3	1.1
	236	225	1	-	-	-	-	-	1	0.4
	237	226	1	-	-	-	-	-	1	0.4
	238	227	1	-	-	-	-	-	1	0.4
	239	228	1	-	-	-	-	-	1	0.4
	240	229	1	-	-	-	-	-	1	0.4
	241	230	1	-	-	-	-	-	1	0.4
	242	231	-	-	-	1	-	-	1	0.4
Not applicable	Non-typeable	Truncated	10	-	-	-	-	-	10	3.8
Total			**251**	**1**	**3**	**2**	**1**	**7**	**265**	**100**

* Novel alleles are in bold and underlined.

Temporal analysis revealed that the frequency of *seM* groups of *See* changed over the years. When *See* isolates (n = 265) were grouped by five-year intervals based on their time of isolation, 14, 7, 2, and 242 *See* isolates were isolated in 2001-2005, 2006-2010, 2011-2015, and 2016-2020, respectively. Although much less *See* isolates were recovered in 2001-2005 than in 2006-2020 the former isolates were distributed among more groups (*seM*-groups 1, 2, 6, 8, and 9) than the latter (6 and 9) (**Figure 11**). *See* isolates of *seM*-6 group were detected in all periods of isolation. Isolates of *seM*-9 group dominated in years 2006-2010 (85.7 % of the *See* isolates of that period) and 2016-2020 (93 %). *See* isolates with non-typeable *seM* (n = 10) were only isolated in 2016-2020.

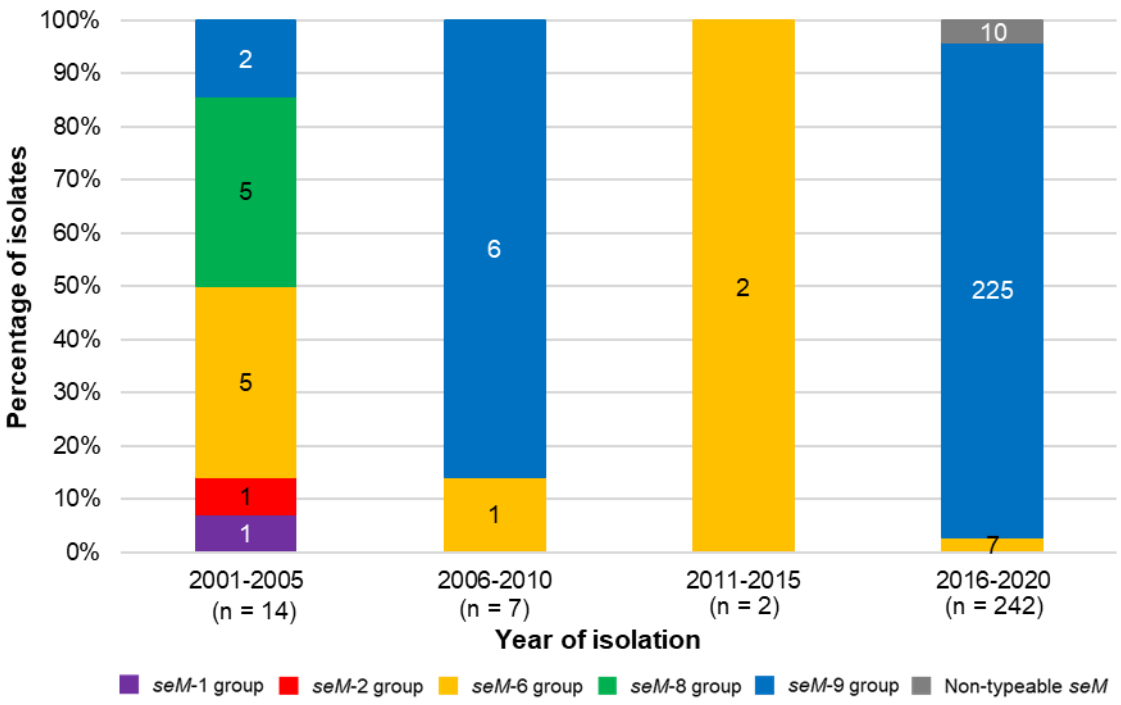


Figure 11. The relative frequency of *seM* groups of *S. equi* ssp. *equi* isolates (n = 265) in five-year intervals from 2001 to 2020.

The number of isolates of each *seM* group is indicated in the stacked bars.

Ten *See* isolates harboured a *seM* allele that was non-typeable due to a mutation in the 327 bp-target segment of the *seM* gene (**Table 10**). Although each isolate had its own mutation, two types of mutation occurred. Type 1 was a single nucleotide insertion or deletion or single nucleotide nonsense mutation which all created premature stop codons. In *See* isolate IHIT40799 the nucleotide guanine (G) compared to *See* strain 4047 at position 140 of *seM* was deleted while in *See* IHIT43619 an adenosinetriphosphate (A) was inserted at position 162 of *seM*. Both mutations led to frameshifts that would result in a truncated SeM protein comprising only the first 51 and 54 N-terminal amino acids, respectively, instead of the 534 amino acids of the entire SeM protein including the leader peptide. In *See* IHIT43010 a nonsense mutation at position 314 (T → G) in the *seM* gene would create a premature stop at codon 105. The type 2 mutations identified in this study were larger deletions within the *seM* gene. All these deletions preserved the reading frame (in-frame deletions). Therefore, peptides translated from these mutated *seM* genes would have the same N- and C-terminal domains as the protoype SeM but would lack an internal domain of 117 to 147 amino acids (AA) in size. AA residues 57–173 of the prototype SeM would be missing in all ten of the *seM* mutant *See* isolates.

The forward primer (ASW73) binding site of the *seM* PCR was present in all ten *seM* mutant *See* isolates. The reverse primer (ASW74) binding site was also present but altered in three isolates (IHIT40480, IHIT40795, IHIT40955) due to deletion or insertion.

Table 10. *Streptococcus equi* ssp. *equi* isolates (n = 10) with non-typeable *seM* alleles

Isolate ID	Sample type	Primer binding sites of the *seM* PCR		*seM* gene (1,605 bp)*		SeM Protein (534 AA)*	
		Forward primer (ASW73)	Reverse primer (ASW74)	Position of deleted/ mutated nucleotide(s)	Length of deletion [bp]	Number of deleted amino acid residues [AA]	Predicted length of the encoded SeM protein [AA]
IHIT 35378	Guttural pouch lavage fluid	Present	Present	Deletion: 170-520	351	117	417
IHIT 40246	Abscess material	Present	Present	Deletion: 115-525	411	137	397
IHIT 40480	Guttural pouch lavage fluid	Present	Present, deletion of 6 bp	Deletion: 123-563	441	147	387
IHIT 40795	Abscess material	Present	Present, deletion of 1 bp	Deletion: 128-550	423	141	393
IHIT 40799	Guttural pouch lavage fluid	Present	Present	Deletion, frameshift mutation: 140	1	483	51
IHIT 40955	Nasal swab	Present	Present, insertion of 6 bp	Deletion: 115-549	441	147	391
IHIT 41633	Nasal swab	Present	Present	Deletions: 111-120, 153-158, 180-555	390	130	404
IHIT 41784	Abscess material	Present	Present	Deletion: 108-548	441	147	387
IHIT 43010	Nasal swab	Present	Present	Nonsense mutation: 314 (T→G)	-	430	104
IHIT 43619	Guttural pouch lavage fluid	Present	Present	Insertion (A), frameshift mutation: 162	-	480	54

* The *seM* gene of *S. equi ssp. equi* strain 4047 (GenBank Accession Number: FM204883.1) was used as reference. AA: amino acid(s), bp: base pair(s)

4.3.3 *Streptococcus equi* ssp. *equi* diversity according to MLST

WGS data of 191 selected *See* isolates were submitted to MLST according to the method of Webb *et al*. (2008) [196]. The sequence types (STs) were determined according to the MLST database (http://pubmlst.org/szooepidemicus/) maintained at the University of Oxford, United Kingdom [80].

Only two and highly related STs were found among the 191 *See* isolates examined, namely ST-151 and ST-179. ST-179 is a single locus variant (SLV) of ST-151 that differed by one nucleotide in the *tdk* allele (**Table 11**). ST-151 isolates (n = 141, 73.5 %) were recovered from Germany (n = 135), Denmark (n = 3), Luxembourg (n = 2), and the Netherlands (n = 1) between 2016-2020 (**Table 11, Table 12**). ST-179 isolates (n = 50, 26.2 %) were recovered from Germany (n = 42), Indonesia (n = 7), and Austria (n = 1) for a more extended period (2002 to 2020). ST-151 isolates were more likely isolated from horses in 2011-2020 than in 2001-2010 (p = 0.0003; Fisher's exact test).

Isolates recovered from horses in Germany were dominated by ST-151 (135 from 177, 76.3 %). Furthermore, all isolates recovered from Denmark (n = 3), Luxembourg (n = 2), and the Netherlands (n = 1) also belonged to ST-151 (**Table 11**). In contrast, all seven *See* isolates recovered from three provinces in Indonesia, an isolate from Austria, and 42 isolates from Germany belonged to ST-179.

Table 11. The number of *S. equi* ssp. *equi* isolates (n = 191) belonging to each sequence type (ST) by country of origin

ST	Allele number of the respective gene							Country	Number of isolates	
	arcC	*nrdE*	*proS*	*spi*	*tdk*	*tpi*	*yqiL*		n	%
151	45	45	45	45	23	45	45	Germany	135	73.8
								Denmark	3	
								Luxembourg	2	
								The Netherlands	1	
179	45	45	45	45	45	45	45	Germany	42	26.2
								Indonesia	7	
								Austria	1	
Total									191	100

Table 12. The number of *S. equi* ssp. *equi* isolates (n = 191) belonging to each sequence type (ST) by year of isolation

ST	Number of isolates											Total
	2002	2003	2006	2007	2009	2012	2016	2017	2018	2019	2020	
151	0	0	0	0	0	0	4	5	13	39	80	141
179	1	1	2	1	1	1	0	6	13	8	16	50
All	1	1	2	1	1	1	4	11	26	47	96	191

4.3.4 *Streptococcus equi* ssp*. equi* diversity according to core genome MLST

WGS data of 191 selected *See* isolates were submitted to cgMLST using the published cgMLST scheme in Pathogenwatch [122]. The assembled genomes of this study and the corresponding cgMLST analysis can be accessed at Pathogenwatch at the URL address https://pathogen.watch/collection/4j4q55j6u7s0-s-equi-equi-ihit).

The cgMLST analysis revealed that the 191 *See* isolates displayed 159 core genome genotypes (cg-genotypes). All these cg-genotypes were affiliated to three of the six known BAPS clusters, namely BAPS-2, BAPS-5, and BAPS-6 [122] (**Figure 12, Figure 13**).

Based on the cgMLST analysis, all *See* isolates in this study were suited to the current global phylogenetic tree of *See* (**Figure 13**). The majority of the isolates (181/191, 94.8 %) belonged to BAPS-2, while 9 and 1 isolates belonged to BAPS-5 and BAPS-6, respectively. The isolates differed from one another by 0 to 110 pairwise core-genome SNPs (cgSNPs), while the isolates within each BAPS cluster differed from one another by 0 to 42 (BAPS-2), and 0 to 18 (BAPS-5) cgSNPs (**Figure 12, Table 13**). The single isolate in BAPS-6 differed from other isolates in the study collection by 87 to 110 cgSNPs.

See isolates from Germany (n = 177) were distributed in BAPS-2 (95 %), BAPS-5 (4.5 %), and BAPS-6 (0.5 %). All Indonesian isolates (n = 7) differed from each other by only 2 to 14 cgSNPs and built an exclusive subcluster in BAPS-2. Isolates from Denmark (n = 3) belonged to BAPS-2 and differed by 5 to 8 cgSNPs from each other, while two isolates from Luxembourg differed by 2 cgSNPs and belonged to BAPS-2. Moreover, the single isolate from the Netherlands belonged to BAPS-2 and was subclustered with isolates from Germany. The single isolate from Austria (*See* isolate IHIT41179) differed by 9 to 18 cgSNPs to the other eight isolates from Germany that belonged to BAPS-5.

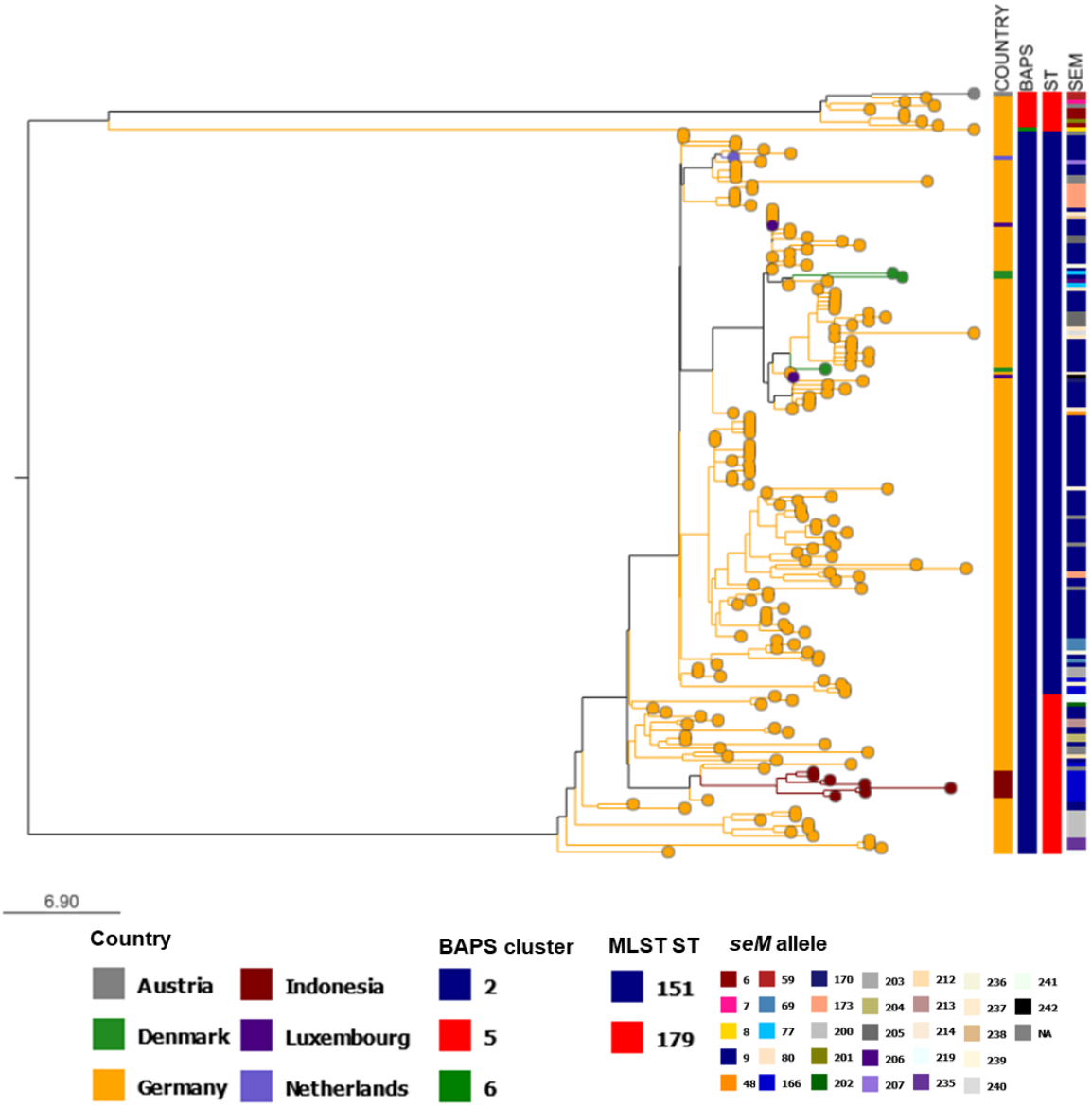


Figure 12. Phylogenetic tree of the *S. equi* ssp. *equi* isolates of this study (n = 191) based on pairwise cgMLST scores

The dendrogram was reconstructed from pairwise cgMLST scores [122] and visualized in Microreact [7]. The scale bar relates to horizontal branch lengths and indicates the number of cgSNPs proposed to have occurred on the horizontal branches. Coloured circles indicate the country of origin of the isolates. The cgMLST BAPS-cluster, MLST sequence type (ST) and *seM* allele of the respective isolate are indicated on the coloured metadata bars at the right edge of the figure. NA = non-typeable *seM* allele.

Table 13. Diversity of *S. equi* ssp. *equi* isolates (n = 191) within each BAPS cluster (cgMLST data)

BAPS cluster	Number of cg-genotypes	Pairwise difference between genotypes within the BAPS cluster [cgSNPs]	Number of *See* isolates	
			total	by country
BAPS-2	150	1–42	181	Germany (168) Denmark (3) Luxembourg (2) The Netherlands (1) Indonesia (7)
BAPS-5	8	1–18	9	Germany (8) Austria (1)
BAPS-6	1	-	1	Germany (1)
Total	159		191	

4.3.5 Concordance between molecular typing results

Molecular typing results for whole genome sequenced *See* isolates of this study (n = 191) were analysed for concordance, namely results obtained by *seM* typing, MLST, and cgMLST.

As summarized in **Table 14**, all ST-151 isolates (n = 141) belonged to BAPS-2, whereas ST-179 isolates (n = 50) were distributed in BAPS-2 (80 %), BAPS-5 (18 %), and BAPS-6 (2 %) (**Table 14**, **Figure 12**). *See* isolates that belonged to group *seM*-6 affiliated exclusively to BAPS-5, those of group *seM*-8 to BAPS-6, and group *seM*-9 to BAPS-2 (**Table 14**). Interestingly, *See* isolates habouring a non-typeable *seM* (n = 10) occured not only in BAPS-2 (n = 9) but also in BAPS-5 (n = 1) and in both STs. Furthermore, BAPS-2 isolates with a non-typeable *seM* were distributed in different subclusters (**Figure 12**), where they differed by 3 to 32 cgSNPs from each other. The single BAPS-5 isolate with a non-typeable *seM* (*See* isolate IHIT40246) differed by 88 to 104 cgSNPs from the BAPS-2 isolates with a non-typeable *seM* allele.

Table 14. Concordance between molecular typing methods used for the *S. equi* ssp. *equi* isolates (n = 191) in this study

cgMLST BAPS cluster	Sequence type (ST)	*seM* group	*seM* allele(s)	Number of *See* isolates
BAPS-2	ST-151	*seM*-9 group	9, 48, 69, 77, 80, 166, 170, 173, 203, 205, 206, 207, 214, 236, 237, 238, 239, 240, 242	135
		Not applicable	Non-typeable	6
	ST-179	*seM*-9 group	9, 166, 200, 202, 204, 212, 213, 219, 235, 241	37
		Not applicable	Non-typeable	3
BAPS-5	ST-179	*seM*-6 group	6, 7, 59, 201	8
		Not applicable	Non-typeable	1
BAPS-6	ST-179	*seM*-8 group	8	1
			Total	191

4.3.6 Association between phylogenetic groups of the *Streptococcus equi* ssp. *equi* and sample type for isolation

The association was analysed between the phylogenetic groups of the *See* isolates, based on *seM* typing, MLST, and cgMLST, and the sample type these isolates had been isolated from. *See* isolates belonging to *seM*-9 group (n = 233, 87.9 %) were recovered from all sample types of this study, while isolates belonging to *seM*-6 group (n = 15, 5.6 %) were recovered from all sample types where more than one isolate had been isolated from (**Table 15**). Isolates habouring a non-typeable *seM* (n = 10) were isolated from the guttural pouch lavage fluid (4), nasal swab (3), and abscess material (3). Since the number of isolates for some *seM* allele groups was too low, it was not justified to conduct a statistical analysis by the Fisher's exact test to determine significant associations.

See isolates (n = 191) of both STs (ST-151, ST-179) were recovered from all sample types where more than one isolate had been recovered from (**Table 15**). The portion of ST-179 was highest in abscess material (37.0 %) and lowest in guttural pouch lavage fluid (18.5 %). However, there was no significant association between MLST-ST and the sample type (p = 0.44; Fisher's exact test).

Phylogenetic grouping based on cgMLST showed that *See* isolates of the most prevalent BAPS-2 cluster were recovered from all sample types (**Table 15**). BAPS-5 isolates (n = 9) were recovered only from sample types where more than one isolate had been recovered from. The single BAPS-6 isolate was recovered from abscess material. There was no significant association between the cgMLST-BAPS and the sample type (p = 0.16; Fisher's exact test).

Table 15. The number of *S. equi* ssp. *equi* isolates by phylogenetic groups and sample type for their isolation

Typing method	Category	Number of isolates by sample type							Total
		Abscess material *	Chon-droid	Guttural pouch lavage fluid	Nasal swab	Naso-pharyngeal swab	Organ *	Swab *	
***seM* typing**	*seM*-2 group	-	-	-	-	1	-	-	1
	seM-3 group	1	-	-	-	-	-	-	1
	seM-6 group	6	-	2	3	1	-	3	15
	seM-8 group	2	-	-	1	-	-	2	5
	seM-9 group	50	1	62	110	8	1	1	233
	Non-typeable *seM*	3	-	4	3	-	-	-	10
	Total	**62**	**1**	**68**	**117**	**10**	**1**	**6**	**265**
MLST	ST-151	29	-	44	62	5	1	-	141
	ST-179	17	1	10	20	2	-	-	50
	Total	**46**	**1**	**54**	**82**	**7**	**1**	**-**	**191**
cgMLST	BAPS-2	41	1	53	79	6	1	-	181
	BAPS-5	4	-	1	3	1	-	-	9
	BAPS-6	1	-	-	-	-	-	-	1
	Total	**46**	**1**	**54**	**82**	**7**	**1**	**-**	**191**

* sites not specified

4.3.7 Genetic relationship between *Streptococcus equi* ssp. *equi* isolates in this study and other studies

A comparison of the *See* genomes in this study collection (n = 191, 2002–2020) with 759 publicly available *See* genomes collected between 1955 and 2018 [62,71,122–124] was conducted to visualize the phylogenetic relationship between *See* isolates from this study and *See* strains from other studies around the world.

The combined *See* genome collection (n = 950) can be accessed at Pathogen-watch at the following URL: https://pathogen.watch/collection/i9nku0mn2uay-streptococcus-equi-equi-ihit-and-other-studies. Furthermore, the phylogenetic reconstruction along with the metadata of the isolates was visualized in Microreact [7] and can be accessed at the following URL: https://microreact.org/project/nfobH83cQbKSksfYdQuEed-streptococcus-equi-equi-ihit-and-other-studies.

The combined *See* genome collection was recovered from 23 countries that comprised countries in Europe (n = 12, 603 genomes), Asia (n = 6, 158 genomes), North America (n = 2, 144 genomes), South America (n = 1, 15 genomes), and Australia/Oceania (n = 2, 30 genomes) **(Figure 13**). According to the results of the cgMLST analysis of the combined *See* genome collection, the global population structure of *See* appears to be consistent with prior studies where the *See* genomes were assorted into six BAPS clusters [62,122–124]. The See isolates of this study were distributed into the three of this six known BAPS clusters (**Figure 13**). The majority of the isolates in the combined collection were distributed in BAPS-2 (n = 520, 54.7 %), followed by BAPS-1 (n = 138, 14.5 %), BAPS-5 (n = 132, 13.9 %), BAPS-6 (n = 73, 7.7 %), BAPS-3 (n = 31, 3.3 %) (**Figure 13**). In the combined genome collection genomes differed from one another by 0 to 181 cgSNPs.

All *See* isolates of this study belonged to MLST sequence types ST-179 and ST-151 of the currently known nine STs. The vast majority of genomes in the combined *See* genome collection is assigned to these two STs: ST-179, n = 638, 67.2 %; ST-151, n = 295, 31.1 %; ST-395, n = 10, 1.1 %; ST-281, n = 3, 0.3 %; ST-396, n = 2, 0.2 %; ST-402, n = 2, 0.2 % (**Figure 13**). STs other than ST-179 are SLV or DLV of ST-179 and can be regarded as members of the same clonal complex (CC)-179 [122]. In the combined genome collection, all *See* isolates and strains of ST-151 belonged to BAPS-2, together with all strains of ST-281 (n = 3) and ST-402 (n = 2). Furthermore, all strains of ST-395 (n = 10) and ST-396 (n = 2) belonged to BAPS-1. Isolates and strains of ST-179 are distributed in all of the six BAPS clusters (**Figure 13**). BAPS-1 is dominated by *seM*-28 allele (35.5 %), BAPS-2 by *seM*-9 allele (62.3 %), BAPS-3 by *seM*-100 allele (83.9 %), BAPS-4 by *seM*-16 allele (19.6 %), BAPS-5 by *seM*-6 allele (25.8 %), and BAPS-6 by *seM*-39 allele (57.5 %).

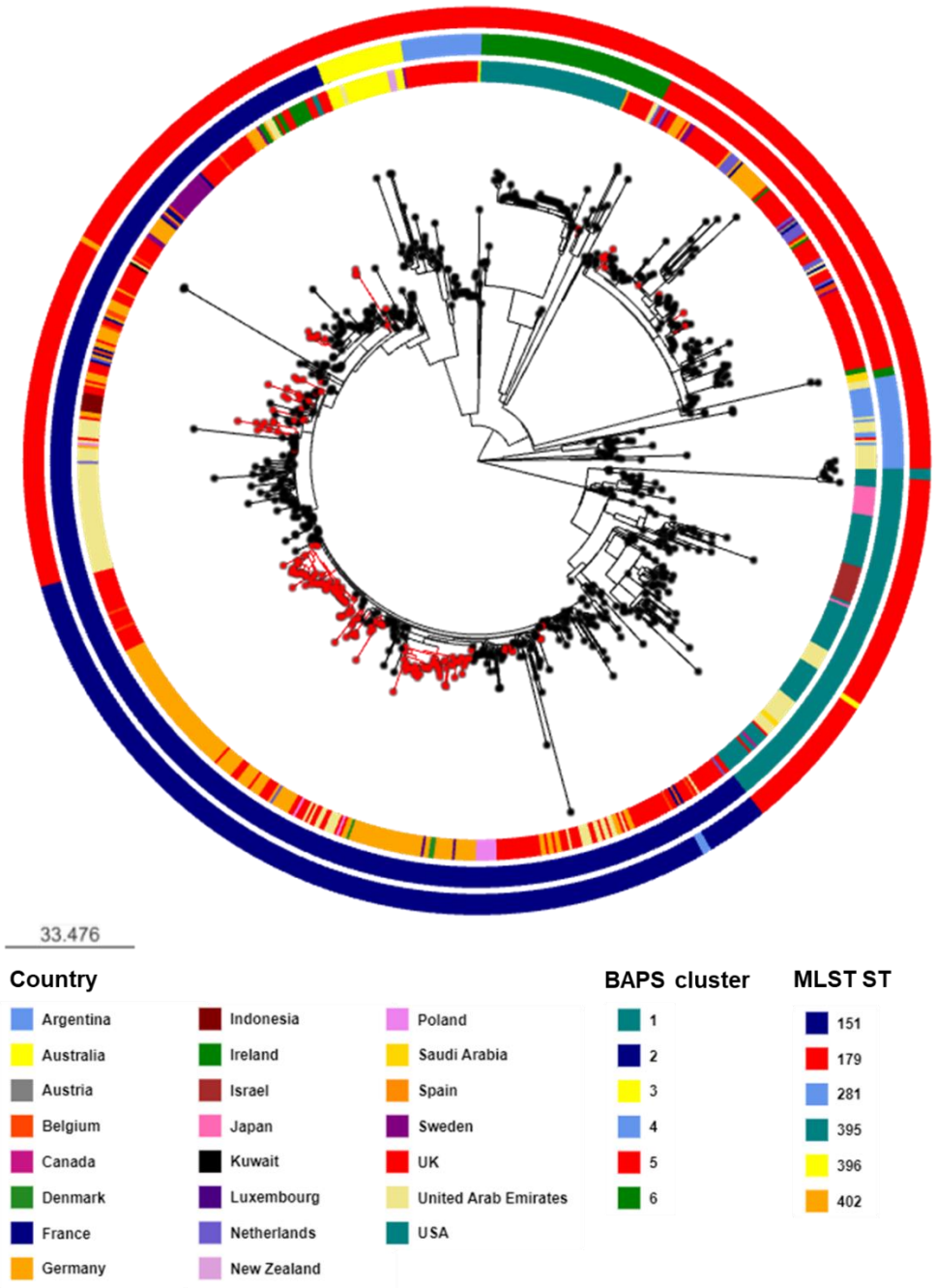


Figure 13. Phylogenetic reconstruction of the *S. equi* ssp. *equi* genomes (n = 950) in this study and other studies based on pairwise cgMLST scores.

The dendrogram was reconstructed from pairwise cgMLST scores [122] and visualized in Microreact [7]. The scale bars indicate the number of cgSNPs proposed to have occurred on branches. Coloured circles indicate the isolate collection (red: this study, black: other studies). The inner to outer circles surrounding the phylogenetic tree represent: (1) country of origin, (2) cgMLST BAPS cluster, (3) MLST sequence type (ST) of each isolate. Isolates from other studies were from published available genomes [71], [124], [122], [123].

4.4 Spatial-temporal distribution of *Streptococcus equi* ssp. *equi* in Germany

4.4.1 Genetic variation of *Streptococcus equi* ssp. *equi* in spatial-temporal clusters

Almost no information was provided with *See* isolates of this study that their equine hosts were stable mates or were associated with an infection chain or with a strangles epidemic. In attempt, to uncover epidemiological links and putative strangles outbreaks in the equine population retrospectively, those *See* isolates that had been isolated from horses in the same area within the same given time frame (spatial-temporal *See* clusters) were analysed for their phylogenetic relationship. The WGS data of selected *See* isolates from Germany (n = 177) were included in this *in silico* investigation. These isolates had been recovered in 2002-2020 from equines in 93 districts (*Landkreise*) in 14 federal states (*Bundesländer*) of Germany. The following criteria were used to define a spatial-temporal *See* cluster: (**a**) at least two *See* isolates, (**b**) all *See* isolates from different horses, (**c**) all *See* isolates originate in the same district, (**d**) all *See* isolates isolated apart from each other not longer than six months.

As summarized in **Table 16** this study identified 93 *See* isolates that formed 23 spatial-temporal *See* clusters. Most *See* infections occurred in small to medium-sized spatial-temporal clusters with 2–6 *See* isolates per cluster. Nine clusters were only formed by *See* isolates that were identical in their ST, BAPS cluster assignment, and *seM* allele (combined genotype; clusters 4, 5, 6, 9, 10, 11, 16, 18, 19). Eight of these clusters were formed by two *See* isolates only (**Table 16**). Six other clusters consisted purely of different *See* isolates (clusters 1, 13, 14, 15, 21, 22), while the rest (clusters 2, 3, 7, 8, 12, 17, 20, 23) contained both identical and different *See* isolates. The pairwise difference within each spatial-temporal cluster ranged from 0–97 cgSNPs (**Table 16**). The pairwise difference between *See* isolates of the same spatial-temporal cluster that were identical in their combined genotype (ST, BAPS cluster assignment, and *seM* allele) within the same spatial-temporal cluster was 0–14 cgSNPs. The pairwise difference between *See* isolates of the same cluster that were different in their combined genotype was 1–97 cgSNPs.

The number of identified cg-genotypes (same cg-allelic profile, pairwise difference = 0 cgSNPs) per spatial-temporal cluster correlated positively with the number of isolates per cluster. Five clusters contained isolates that displayed the same cg-genotype, as seen in clusters 3, 5, 6, 7, 9, 11, and 17 (**Table 16**). Furthermore, in several spatial-temporal clusters, isolates did not have identical but highly similar cg-genotypes and thus

belonged to the same subcluster within the phylogenetic BAPS tree (clusters 2, 3, 7, 8, 12, 14, 17, 19, and 22) (**Table 16**). This is shown by way of example for spatial-temporal clusters 12 and 17 in **Figure 14 A** and **B**.

The 10 *See* isolates in spatial-temporal cluster 12 displayed five different combined genotypes. Three isolates (ST-179, BAPS-5, *seM*-6 or *seM*-201; pairwise difference 1–7 cgSNPs) and two other isolates (ST-151, BAPS-2, *seM*-9; pairwise difference 2 cgSNPs), respectively, formed two subclusters that were localized separately from all other isolates in the BAPS tree (**Table 16, Figure 14 A**). The *seM*-6 and *seM*-201 alleles both belonged to *seM*-6 allele group and differed by 5 SNPs in the variable region targeted for *seM* typing.

Spatial-temporal cluster 17 consisted of 20 *See* isolates and was thus the largest of these clusters in the study. Based on pairwise cgMLST scores three BAPS subclusters were identified here. Subcluster 1 contained 3 isolates (ST-151, BAPS-2, *seM*-9; pairwise difference 0 cgSNPs), subcluster 2 comprised 4 isolates (ST-151, BAPS-2, *seM*-9; pairwise difference 1–6 cgSNPs), and subcluster 3 was composed of 5 isolates (ST-151, BAPS-2, *seM*-69 or *seM*-236; pairwise difference 1–7 cgSNPs) (**Table 16, Figure 14 B**). The *seM*-69 and *seM*-236 alleles both belonged to *seM*-9 allele group and differed by a SNP at position 217: T → C in the variable region targeted for *seM* typing.

Table 16. Spatial-temporal clusters (n = 23) of *S. equi* ssp. *equi* isolates in Germany

Spatial-temporal cluster	No of isolates	Dates of isolation [range]	Location(s)*	Federal state	Combined genotypes [ST_BAPS cluster_*seM* allele] (no of isolates)	No of pairwise differences within the cluster [cgSNPs]
1	2	09.03.2020 – 01.04.2020	1	BW	151_2_9 (1); 151_2_nt (1)	22
2	3	19.11.2019 – 28.05.2020	2, 3, 4	BW	151_2_9 (2); 151_2_nt (1)	5, 6, 9
3	12	01/08/2019 – 05.09.2020	5, 6	BW	151_2_9 (9); 151_2_77 (1); 179_2_235 (2)	0–34
4	2	19.09.2019 – 07.05.2020	7, 8	BW	151_2_9 (2)	5
5	2	05.02.2018 – 26.03.2018	9, 10	BY	151_2_173 (2)	0
6	2	19.07.2016 – 05.10.2016	11, 12	B	151_2_203 (2)	0

Spatial-temporal cluster	No of isolates	Dates of isolation [range]	Location(s)*	Federal state	Combined genotypes [ST_BAPS cluster_*seM* allele] (no of isolates)	No of pairwise differences within the cluster [cgSNPs]
7	6	23.03.2018 – 18.07.2019	13, 14, 15, 16, 17	H	151_2_9 (5); 179_2_204 (1)	0–21
8	3	30.03.2020 – 22.09.2020	18, 19	H	179_2_9 (2); 179_2_200 (1)	2, 26, 28
9	2	19.04.2018 – 16.08.2018	20, 21	H	179_2_200 (2)	0
10	2	27.12.2018 – 10.04.2019	22, 23	H	151_2_173 (2)	4
11	2	18.02.2020 – 11.03.2020	24, 25	H	151_2_9 (2)	0
12	10	17.09.2017 –24.05.2018	26, 27, 28, 29, 30, 31, 32, 33, 34	H	151_2_9 (5); 179_2_166 (1); 179_2_219 (1); 179_5_6 (2); 179_5_201 (1)	1–97
13	2	31.07.2019 –26.09.2019	35, 36	N	151_2_9 (1); 151_2_80 (1)	9
14	2	24.03.2020 – 18.09.2020	37, 38	NW	151_2_9 (1); 151_2_170 (1)	6
15	2	03.12.2019 – 14.03.2020	39	NW	151_2_9 (1); 151_2_207 (1)	9
16	3	13.05.2020 – 27.05.2020	40, 41, 42	NW	151_2_9 (3)	6, 8, 14
17	20	03.08.2019 – 23.07.2020	43	NW	151_2_9 (14); 151_2_69 (4); 151_2_236 (1); 179_2_9 (1)	0–26
18	2	19.09.2018 – 02.10.2018	44, 45	NW	151_2_9 (2)	3
19	2	11.09.2019 – 30.03.2020	46, 47	RP	151_2_9 (2)	2
20	3	27.11.2019 –15.09.2020	48, 49, 50	RP	151_2_9 (2); 179_2_241 (1)	16, 17, 23
21	2	05.04.2019 – 08.10.2019	51, 52	RP	151_2_9 (1); 179_5_59 (1)	95
22	2	07.04.2020 – 11.06.2020	53, 54	RP	151_2_205 (1); 151_2_9 (1)	5
23	5	20.08.2019 – 29.04.2020	55	SH	151_2_9 (2); 151_2_166 (1); 179_2_235 (1); 179_2_200 (1)	6–35

* location (encrypted) of the owner or veterinarian of the respective equine; BW: Baden Wuerttemberg, BY: Bavaria, B: Berlin, H: Hesse, N: Lower Saxony, NW: North Rhine-Westphalia, RP: Rhineland-Palatinate, SH: Schleswig-Holstein; nt: non-typeable.

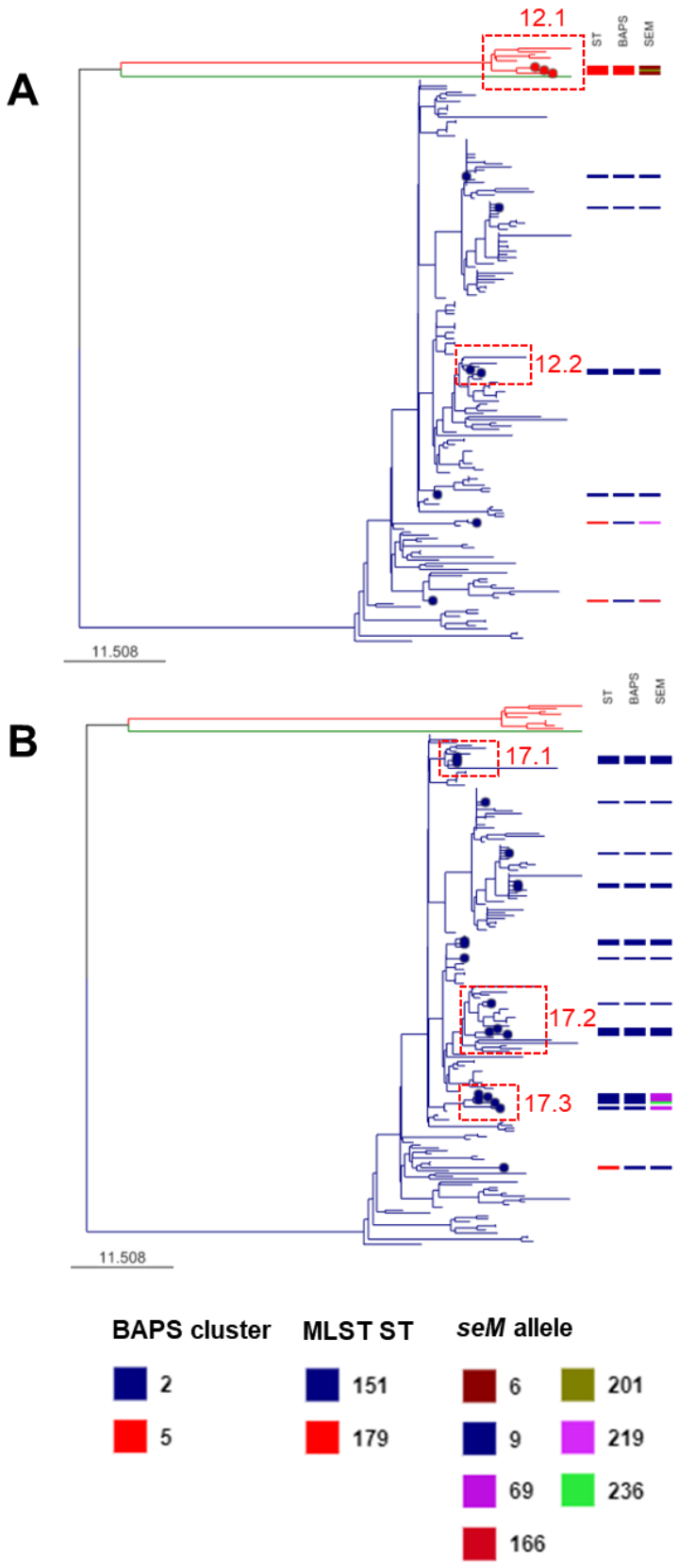


Figure 14. ***Streptococcus equi*** **ssp.** ***equi*** **isolates belonging to (A) spatial-temporal cluster 12 (10 isolates) and (B) spatial-temporal cluster 17 (20 isolates).**

Legend of Figure 14. The dendrogram was reconstructed from pairwise cgMLST scores [122] and visualized in Microreact [7]. The scale bar relates to horizontal branch lengths and indicates the number of cgSNPs proposed to have occurred on the horizontal branches. Coloured circles indicate the cgMLST BAPS cluster. Red rectangle boxes indicate isolates belonging to the same subcluster. The cgMLST BAPS-cluster, sequence type (ST) and *seM* allele of the respective *See* isolate are indicated on the coloured metadata bars at the right edge of the figure.

4.4.2 Spatially and/or temporally distant *Streptococcus equi* ssp. *equi* isolates with identical genotype

In a second approach, *See* isolates from Germany with identical genotypes (identical cg-genotype, ST, and *seM* allele) were analysed for their spatial-temporal distributions.

A total of 23 genotypes were detected in multiple *See* isolates (**Table 17**). The number of isolates with the same genotype was two to three isolates. Isolates with identical genotypes were identified not only in the same but also in different spatial-temporal clusters. Eight genotypes (17, 20, 52, 76, 89, 111, 117, 154) were identified in multiple isolates within the same spatial-temporal cluster (**Table 17**). The other genotypes were identified in those *See* isolates that did not came from the same spatial-temporal cluster. The time between isolations and the distance between geographical isolation sites of *See* isolates with the same genotype varied from 0–914 days and 0–690 kilometers, respectively. Isolates with the same genotype within the same spatial-temporal cluster mostly came from the same location (spatial distance 0 km) with a temporal distance of up to 143 days. On the other hand, *See* isolates with the same genotype and originating outside any spatial-temporal cluster were isolated from distant locations (spatial distance 3 to 690 km) and with a large temporal distance (up to 914 days). Interestingly, the two *See* isolates of cg-genotype 157 were isolated from very distant locations (690 km) but with a temporal distance of only 9 days. On the other hand, the three isolates of cg-genotype 21 were isolated from different locations with a distance of 60 to 223 km and large temporal distances of 397 to 914 days. Unfortunately, no information was available regarding movement of the horses before sampling.

Table 17. Features of spatially and/or temporally distant *Streptococcus equi* ssp. *equi* isolates with identical cg-genotype, ST, and *seM* allele

Genotype			Number of isolates			Time between isolations [days]	Distance between geographical isolation sites [range, km]
Core genome genotype	*seM* allele	MLST ST	Total	Within a spatial-temporal cluster	Outside the (largest) spatial-temporal cluster		
6	6	179	2	0	2	50	15
17	9	151	3	3	0	0–31	0
20	173	151	2	2	0	49	0
21	173	151	3	0	3	397–914	60–223
23	9	151	2	0	2	351	400
49	80	151	2	0	2	22	336
52	9	151	2	2	0	22	0
64	9	151	2	0	2	49	781
68	9	151	3	0	3	9–29	11–263
71	9	151	2	0	2	53	175
72	9	151	3	0	3	3–14	64–158
76	9	151	2	2	0	42	0
77	9	151	2	0	2	70	138
84	9	151	2	0	2	841	217
87	9	151	2	0	2	141	108
89	9	151	2	2	0	143	0
102	9	151	2	0	2	240	254
106	9	151	3	0	3	34–222	55–378
111	69	151	2	2	0	134	0
117	203	151	2	2	0	78	14
133	204	179	2	0	2	136	43
154	200	179	2	2	0	48	30
157	235	179	2	0	2	9	690

4.5 Occurrence of virulence-associated genes in the genetic lineages of *Streptococcus equi* ssp. *equi*

In order to investigate the genetic diversity of the *See* isolates according to their virulence-associated genes (VAGs), assembled WGS data of the 191 selected *See* isolates were examined *in silico* for recognized VAGs of *See.* This analysis used a self-created in-house database of VAGs with software packages MyDbFinder V.2.0 and Geneious V.8.1.3. In total, 67 VAGs specific for *See* were investigated, which were categorized into adherence

(n = 21), immune modulation (n = 10), exoenzymes and exoproteins (n = 11), exotoxins (n = 10), and nutritional/metabolic factors (n = 15) (**Appendices Table 22**). Thresholds for sequence identity and coverage length were both set at 60 %.

Among the examined VAGs, 38 (56.7 %) were detected in all isolates, while 29 VAGs (43.3 %) were detected in isolates at rates varying from 5 % to 99 %. The latter VAGs are listed in **Table 18**. All detected VAGs shared more than 90 % sequence identity with the reference gene. Overall, each VAG was detected in an average of 179.9 ± 26.4 isolates (arithmetic mean ± std). VAGs categorised to adherence were detected in 173.1 ± 40.9 isolates, VAGs associated with immune modulation in 185.3 ± 7.6 isolates, exoenzyme genes in 189.9 ± 3.4 isolates, exotoxin genes in 189.9 ± 3.4 isolates, and VAGs associated with nutrition/metabolism in 171.8 ± 21.0 isolates.

The presence of VAGs associated with adherence varied in *See* isolates. All isolates harboured genes encoding for CNE, SFS, FNE, FneB, FneC, FneD, FneE, FneF, SclC, SclH, Eq5, Eq8, Eq54 / FimI (SEQ0936), TetR-like (SEQ0934), and SEQ0944. However, genes encoding the following proteins were only present in some isolates: SclD (68 %), SclE (5 %), SclF (95 %), SclG (81 %), SclI (74 %), Scl3F (81 %) (**Table 18**). The *sclE* gene was only detected in 10 *See* isolates (5 %) which all belonged to BAPS-5 and BAPS-6. Interestingly, none of the BAPS-2, ST-151, or *seM*-9 group isolates harboured this gene (**Table 18**).

Concerning immune modulation, genes encoding for SeM, SezPSe, SzP, EAG, and Se18.9 were identified in all *See* isolates. The *has* locus, consisting of *hasA*, *hasB,* and *hasC1*, was identified in 90.6 % of the isolates. The *hasA, hasB,* and *hasC1* genes were lacking in 10, 10, and 15 isolates, respectively. Five *See* isolates in this study lacked two genes of the *has* locus, while five other isolates lacked all three genes of the *has* locus (**Table 18**).

All *See* isolates harboured genes for streptokinase and the recognized *See* exoenzymes hyaluronidase, EndoS, IdeE, IdeE2, SodA, sortase A, sortase C1, and Lgt. SEQ2045, the gene of a putative phage-encoded hyaluronidase, was present in 94 % of the isolates, including all BAPS-5 and BAPS-6 isolates and 93 % of the BAPS-2 isolates (**Table 18**).

Genes encoding for the exotoxins streptolysin S, SlaA, and SlaB were identified in all *See* isolates. Genes encoding for SeeH, SeeI, SeeL, and SeeM superantigenic exotoxins were missing in 2, 2, 3, and 3 isolates, respectively, which all belonged to BAPS-2 (**Table 18**). In two *See* isolates only one of these superantigen (sAg) genes was missing, while four other isolates did not have two of the sAg genes. Interestingly, three of these six isolates

were members of the exclusive BAPS-2 subcluster of *See* isolates from Indonesia. Genes *seeH* and *seeI* were not present in *See* isolate IHIT39393 while isolates IHIT40247, IHIT41177, and IHIT43010 lacked *seeL* and *seeM*.

Analysis of the 14 genes constructing the equibactin locus (*eqb*; genes *eqbA* to *eqbN*) revealed that all *See* isolates harboured at least one of these genes. The majority of *See* isolates (53.4 %) even carried a complete *eqb* locus while 46.6 % of the isolates did not harbour the one or the other equibactin gene. The number of missing *eqb* genes in an isolate varied from 1 to 13, where *eqbG* was the gene that was missing most often. The *eqbA* gene, which acts as the repressor in the equibactin locus, was missing in only two isolates, IHIT38875 and IHIT39393, which lacked most of the other genes of the equibactin locus as well. The *eqbE* and *eqbG* genes encoding NRPSs were detected in 73 % and 62 % of the isolates, respectively.

The analysis of WGS data for VAGs revealed that individual VAGs or various combinations of VAGs were missing each in at least one *See* isolate. This phenomenon made it possible to distinguish 107 virotypes (Vt) among the 191 *See* isolates investigated (**Figure 15)**. Most of the virotypes (79.4 %, n = 85), were only identified in one isolate, 14 virotypes were identified in two isolates, while each of 8 virotypes was identified in at least three isolates, respectively (**Table 19**). Vt-3 was the most frequent virotype in this study. This virotype was detected in 48 *See* isolates (25.1 %) and only lacked the *sclE* gene of the 67 VAGs that all isolates were searched for.

Table 18. Occurrence of virulence-associated genes in *S. equi* ssp. *equi* isolates (n = 191)

Virulence Factor	Virulence-associated gene(s)	Function	GenBank accession number (position); Locus tag	Percentage of isolates where the gene was identified [%]									
				Total (n=191)	cgMLST			MLST		*seM* allele groups			
					BAPS-2 (n=181)	BAPS-5 (n=9)	BAPS-6 (n=1)	ST-151 (n=141)	ST-179 (n=50)	*seM*-6 group (n=8)	*seM*-8 group (n=1)	*seM*-9 group (n=172)	Not applicable (n=10)
Adherence													
SclD	*sclD*	Collagen-like surface anchored protein	FM204883.1 (262550..263491); SEQ0280	68	69	67	0	63	82	75	0	67	80
SclE	*sclE*	Collagen-like surface anchored protein	FM204883.1 (616654..617592); SEQ0633	5	0	100	100	0	20	100	100	0	10
SclF	*sclF*	Putative collagen-like surface-anchored protein	CWFT01000003.1 (69953..71047); SEQ0855	95	94	100	100	94	98	100	100	94	100
SclG	*sclG*	Collagen-like surface anchored protein	FM204883.1 (95988..97181); SEQ0090	81	82	44	100	81	80	50	100	83	70
SclI	*sclI*	Collagen-like surface-anchored protein	CWFT01000015.1 (42440..43888)	74	72	100	100	71	82	100	100	72	90
Scl3F	*scl3F*	Putative collagen-like surface anchored protein Scl3F	AY326319.1	81	83	44	100	82	80	50	100	83	70
Immune modulation													
Hyaluronic acid capsule	*hasA*	Capsule synthesis	FM204883.1 (254370..255623); SEQ0269	95	94	100	100	94	96	100	100	94	100
	hasB	Capsule synthesis: UDP-glucose-6-dehydrogenase	FM204883.1 (255876..257081); SEQ0270	95	94	100	100	94	98	100	100	94	100
	hasC1/ gtaB1	Capsule synthesis	FM204883.1 (257131..2580399); SEQ0271	92	92	100	100	91	94	100	100	91	100
	hasC2/ gtaB2	Capsule synthesis	FM204883.1 (272684..273586); SEQ0289	88	89	78	100	89	88	75	100	90	70

Virulence Factor	Virulence-associated gene(s)	Function	GenBank accession number (position); Locus tag	Percentage of isolates where the gene was identified [%]									
				Total (n=191)	cgMLST			MLST		*seM* allele groups			
					BAPS-2 (n=181)	BAPS-5 (n=9)	BAPS-6 (n=1)	ST-151 (n=141)	ST-179 (n=50)	*seM*-6 group (n=8)	*seM*-8 group (n=1)	*seM*-9 group (n=172)	Not applicable (n=10)
Exoenzymes													
Hyaluroni-dase/ hyaluro-nate lyase	*seq2045*	Degrades connective tissue hyaluronan and chondroitins	FM204883.1 (2071724..2072842), SEQ2045	94	93	100	100	94	94	100	100	93	100
Exotoxins													
SeeH	*seeH*	Superantigenic exotoxin	AF186180.1	99	99	100	100	100	96	100	100	99	100
SeeI	*seeI*		FM204883.1 (2066635..2065856) SEQ2037	99	99	100	100	100	96	100	100	99	100
SeeL	*seeL*		AJ583527.1	98	98	100	100	99	96	100	100	99	90
SeeM	*seeM*		AJ583528.1	98	98	100	100	99	96	100	100	99	90
Nutritional/Metabolic factors													
Equibactin	*eqbA*	Iron acquisition, siderophore-like non-ribosomal peptide synthetase (NRPS)	FM204883.1 (1232968..1233348); SEQ1246	99	99	100	100	100	96	100	100	99	100
	eqbB		FM204883.1 (1231990..1232739); SEQ1245	99	99	100	100	100	98	100	100	99	100
	eqbC		FM204883.1 (1231320..1232003); SEQ1244	97	97	100	100	99	92	100	100	97	100
	eqbD		FM204883.1 (1229728..1231311); SEQ1243	99	99	100	100	99	98	100	100	99	100
	eqbE		FM204883.1 (1223642..1229713); SEQ1242	73	72	78	100	73	72	75	100	72	80
	eqbF		FM204883.1 (1221519..1223645); SEQ1241	87	87	100	100	89	84	100	100	86	100

Virulence Factor	Virulence-associated gene(s)	Function	GenBank accession number (position); Locus tag	Percentage of isolates where the gene was identified [%]									
				Total (n=191)	cgMLST			MLST		*seM* allele groups			
					BAPS-2 (n=181)	BAPS-5 (n=9)	BAPS-6 (n=1)	ST-151 (n=141)	ST-179 (n=50)	*seM*-6 group (n=8)	*seM*-8 group (n=1)	*seM*-9 group (n=172)	Not applicable (n=10)
	eqbG		FM204883.1 (1217717..1221535); SEQ1240	62	61	78	100	62	60	75	100	60	70
	eqbH		FM204883.1 (1217140..1217727); SEQ1239	99	99	100	100	100	96	100	100	99	100
	eqbI		FM204883.1 (1216467..1217138); SEQ1238	94	94	100	100	95	92	100	100	94	100
	eqbJ		FM204883.1 (1215077..1216453); SEQ1237	88	88	100	100	91	82	100	100	88	90
	eqbK		FM204883.1 (1213297..1215054); SEQ1236	90	90	100	100	91	86	100	100	90	90
	eqbL		FM204883.1 (1211547..1213292); SEQ1235	80	80	89	100	82	76	88	100	78	100
	eqbM		FM204883.1 (1210463..1211527); SEQ1234	83	83	78	100	87	70	75	100	83	80
	eqbN		FM204883.1 (1209536..1210441); SEQ1233	98	99	89	100	99	96	88	100	99	100

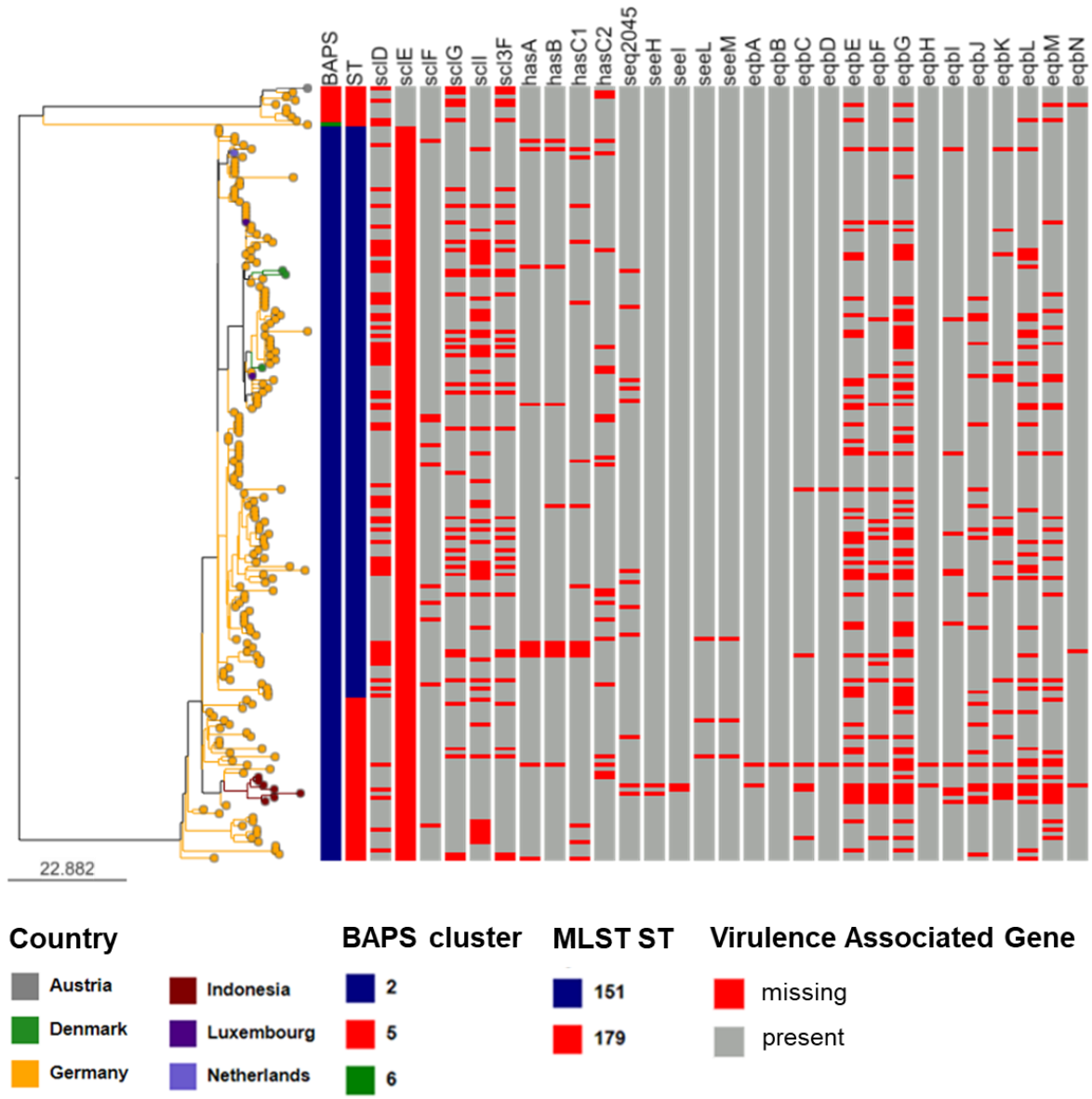


Figure 15. Presence of virulence-associated genes (n = 29) in *S. equi* ssp. *equi* isolates (n = 191) of this study.

See isolates (n = 191) were screend for 67 virulence-associated genes (VAGs) but only those 29 VAGs are indicated here that were missing in at least one of the isolates. The dendrogram was reconstructed from pairwise cgMLST scores [122] and visualized in Microreact [7]. The scale bar relates to horizontal branch lengths and indicates the number of cgSNPs proposed to have occurred on the horizontal branches. Coloured circles indicate the country of origin of the isolates (orange = Germany, grey = Austria, green = Denmark, purple = Luxembourg, blue = the Netherlands, dark red = Indonesia). The cgMLST BAPS-cluster, MLST ST, and the presence or absence of VAGs in the respective *See* isolate are indicated by colours on the vertical metadata bars.

Table 19. *Streptococcus equi* ssp. *equi* virotypes that were identified in more than two isolates

Virotype (Vt)	Missing virulence-associated gene(s)	No of isolates	Combined genotype [ST_BAPS cluster_*seM* allele] (no of isolates)
Vt-1	-	3	179_5_6 (2); 179_5_201 (1)
Vt-3	*sclE*	48	151_2_9 (17); 151_2_48 (1); 151_2_77 (1); 151_2_80 (2); 151_2_173 (5); 151_2_203 (2); 151_2_205 (3); 151_2_206 (1); 151_2_238 (1); 151_2_non-typeable (3); 179_2_9 (4); 179_2_166 (2); 179_2_200 (1); 179_2_202 (1); 179_2_204 (2); 179_2_213 (1); 179_2_235 (1)
Vt-5	*seq2045, sclE,*	4	151_2_9 (4)
Vt-6	*sclD, sclE*	5	151_2_9 (5)
Vt-11	*sclE, eqbG*	4	151_2_9 (2); 151_2_non-typeable (1) 179_2_166 (1)
Vt-12	*sclE, hasC2*	5	151_2_9 (3); 151_2_non-typeable (1) 179_2_non-typeable (1)
Vt-17	*sclD, sclE, sclI*	3	151_2_9 (2); 151_2_239 (1)
Vt-19	*sclE, sclF, hasC2*	6	151_2_9 (5); 151_2_166 (1)

4.6 Diversity of the Strangvac® vaccine antigens across the sequenced *S. equi* ssp. *equi* isolates of this study

Strangvac® is a recombinant subunit vaccine against strangles that contains two recombinant fusion proteins (CCE, Eq85) and a single recombinant protein (IdeE) constructed from parts of eight *See* proteins, namely CNE, SclF, SclI, SclC, EAG, Eq5, Eq8, and IdeE [148,150]. *See* isolates of this study were investigated to determine the extent to which they would match these antigens. With this objective, reference nucleotide sequences encoding for the eight Strangvac® antigens were retrieved from Robinson *et al*. 2018 [148], and the amino sequences were kindly provided by Dr. Andrew Waller (Intervacc AG). In a second step, reference nucleotide sequences and deduced amino sequences were compared with WGS data of the 191 selected *See* isolates of this study using the software package Geneious 8.1.9. All position details in this chapter refer to the numbering in the respective locus tag and the corresponding encoded preprotein, respectively.

Analysis revealed that all 191 *See* of this study had the coding sequences for the eight Strangvac® antigens in their genomes. Furthermore, in all these isolates the coding nucleotide sequences and deduced amino acid sequences of CNE, SclF, SclI, SclC, EAG, and IdeE were identical to those of the Strangvac® antigens (**Table 20**). Minor differences in comparison to the reference sequences were only detected for the Eq5- and the Eq8-derived antigen. In nine *See* isolates, which all were assigned to BAPS-5, ST-179, and *seM*-6 group, the coding sequence for the Eq5-derived antigen varied at nucleotide 601 (A → C) leading to an amino acid substitution at position 201 from isoleucine to leucine (**Table 20**).

Only 5.2 % isolates in this study had an identical Eq-8 derived antigen to that used in Strangvac® (**Table 20**). All 181 *See* isolates (94.8 %) belonging to BAPS-2 contained a nonsynonymous nucleotide difference at position 673 (C → T) of the *eq8* gene, which caused a single amino acid substitution at codon number 225 from histidine to tyrosine (**Table 20**). This codon refers to the C-terminal residue of the Eq8-derived antigen.

Table 20. In silico matching of *S. equi* ssp. *equi* isolates of this study (n = 191) with peptide antigens used in the Strangvac® vaccine

Source of the Strangvac® antigen (locus tag)	Length of coding sequence [bp]	Number of *S. equi* ssp. *equi* isolates …	
		… with identical coding sequence	… with identical coding sequence length
CNE (SEQ0935)	912	191 (100 %)	191 (100 %)
SclF (SEQ0855)	150	191 (100 %)	191 (100 %)
SclI (SEQ1817)	168	191 (100 %)	191 (100 %)
SclC (SEQ2101)	174	191 (100 %)	191 (100 %)
EAG (SEQ0721)	483	191 (100 %)	191 (100 %)
Eq5 (SEQ0256)	1,320	182 (95.3 %) *	191 (100 %)
Eq8 (SEQ0402)	588	10 (5.2 %) **	191 (100 %)
IdeE (SEQ0999)	945	191 (100 %)	191 (100 %)

* 9 *See* isolates (all BAPS-5 *See* isolates of this study) differed from the reference sequence

** 181 *See* isolates (all BAPS-2 *See* isolates of this study) differed from the reference sequence

4.7 Antimicrobial susceptibility rates and occurrence of antimicrobial resistance genes

4.7.1 Antimicrobial susceptibility of *S. equi* ssp. *equi* isolates

The spectrum and rates of antimicrobial susceptibilities of all non-duplicate *See* isolates (n = 265) in this study were determined using a broth microdilution test. MICs were interpreted according to CLSI [34] and a published study [156] and categorised into susceptible (S), intermediate (I), and resistant (R). MIC_{50} and MIC_{90} values were calculated for all tested antimicrobials [159] (**Table 21**). Furthermore, the association between phylogenetic groups (*seM* allele groups, STs, and cgMLST BAPS clusters) and antimicrobial susceptibilities (MICs and categories) was analysed statistically.

All 265 *See* isolates were susceptible to the tested three beta-lactam antimicrobials with available breakpoints in this study, namely penicillin G, ampicillin, and ceftiofur (**Table 21**). All *See* isolates proved also susceptible to tetracycline, and all but one (99.6 %) were susceptible to erythromycin. More than half of the *See* isolates showed intermediate susceptibility (55.1 %) or resistance (0.7 %) to the combination of trimethoprim/sulfamethoxazole. Furthermore, all *See* isolates displayed intermediate susceptibility (0.8 %) or resistance (99.2 %) to enrofloxacin.

There was no significant association between the phylogenetic groups and antimicrobial susceptibilities of *See* ($p > 0.05$, Fishers's exact test), except for trimethoprim/sulfamethoxazole and enrofloxacin. In the case of trimethoprim/sulfamethoxazole, there is a significant association between STs and MICs ($p = 0.0043$, Fisher's exact test). The pairwise comparison of MICs for trimethoprim/sulfamethoxazole between 0.5/9.5 µg/ml (susceptible) and 1/19 µg/ml (intermediate susceptible) was significant ($p = 0.0084$, Fisher's exact test, Bonferroni correction for multiple comparisons), with ST-151 isolates having higher MICs compared to ST-179 isolates. For enrofloxacin, STs and MICs were significantly associated ($p = 0.0099$, Fisher's exact test). The pairwise comparison of MICs for enrofloxacin between 0.5 µg/ml (resistant) and 1 µg/ml (resistant) was significant ($p = 0.0162$, Fisher's exact test, Bonferroni correction for multiple comparisons), with ST-179 isolates having higher MICs than ST-151 isolates.

4.7.2 Occurrence of antimicrobial resistance genes in *S. equi* ssp. *equi* isolates

WGS data of 191 *See* isolates were analysed for the presence of antimicrobial resistance genes using the ResFinder 4.1 database with 2,690 resistance genes entries [21].

In the *See* isolates examined, only the *dfrA1*-like trimethoprim resistance gene (folate pathway inhibitor) was identified in two isolates, namely in *See* isolate IHIT40242 and in *See* isolate IHIT40253). Both isolates were phenotypically susceptible to the combination of trimethoprim/sulfamethoxazole with a MIC value of 0.5/9.5 µg/ml. IHIT40424 had been recovered from the guttural pouch lavage fluid of a horse in North-Rhine Westphalia in 2019. This isolate harboured the *seM-9* allele (*seM*-9 group), belonged to ST-151, and BAPS-2. IHIT40253 had been recovered from the organ (site not specified) of a horse in Mecklenburg Western Pomerania in 2019. The later isolate harboured the *seM*-173 allele (*seM*-9 group), and also belonged to ST-151 and BAPS-2.

Unfortunately, the ResFinder 4.1 did not provide a database of antimicrobial resistance determinants due to chromosomal point mutations specific for streptococci. When the respective database of *Staph. aureus* was used for analysis of *See* isolates, a triple point mutation in the *fusA* gene (GTA → ATT) was detected in all 191 *See* isolates that would change the amino acid residue of FusA at position 90 from valine to isoleucine. This chromosomal point mutation may lead to resistance against fusidic acid, but phenotypic susceptibility of *See* isolates to this steroid antibacterial was not tested in the present study.

Table 21. Antimicrobial susceptibility of *S. equi* ssp. *equi* isolates (n = 265) from equines

Antimicrobial	Antimicrobial Class	Number of isolates with the respective MIC value [µg/ml]														Percentage [%]			MIC_{50} [µg/ml]	MIC_{90} [µg/ml]
		0.016	0.032	0.063	0.125	0.25	0.5	1	2	4	8	16	32	64	128	S	I	R		
Penicillin G*	Beta-lactam			263	2	0	0	0	0	0	0					100	0	0	0.063	0.063
Amoxicillin / Clavulanic acid	Beta-lactam								256	8	1	0				-	-	-	2/1	2/1
Ampicillin**	Beta-lactam					265	0	0	0	0	0	0				100	0	0	0.25	0.25
Ceftiofur*	Beta-lactam				260	5	0	0	0	0						100	0	0	0.125	0.125
Cephalothin	Beta-lactam							258	7	0	0	0				-	-	-	1	1
Enrofloxacin**	Fluoroquinolon	0	0	0	0	2	166	86	11							0	0.8	99.2	0.5	1
Florfenicol	Phenicol							241	24	0	0					-	-	-	1	1
Gentamicin	Aminoglycoside				0	0	0	4	28	218	7	8				-	-	-	4	4
Spectinomycin	Aminoglycoside									0	0	0	7	125	133	-	-	-	64	64
Trimethoprim / Sulfamethoxazol****	Folate pathways inhibitor					3	114	139	7	2						44.2	55.1	0.7	1/19	1/19
Tetracycline***	Tetracycline				241	12	2	9	1	0	0					100	0	0	0.125	0.125
Tiamulin	Pleuromutilin					3	4	47	201	4	5	1	0			-	-	-	2	2
Erythromycin***	Macrolide				261	3	1	0	0	0						99.6	0.4	-	0.125	0.125
Tilmicosin	Macrolide						23	232	7	3	0	0				-	-	-	1	1
Tulathromycin	Macrolide							3	5	121	129	5	1	1		-	-	-	8	8
Colistin	Miscellaneous						0	0	0	265						-	-	-	2	2

* horse-derived breakpoints for streptococci available [34], ** horse-derived breakpoints for *S. equi* ssp. *equi* (*See*), *S. equi* ssp. *zooepidemicus* (*Sez*) available [34], *** human-derived breakpoints for streptococci [34], **** horse-derived breakpoints for *See* and *Sez* available from Sadaka *et al*. (2017) [156]. S: susceptible, R: resistant, I: intermediate, MIC: minimum inhibitory concentration. White areas indicate the dilution ranges tested. Numbers shown above this range represent isolates with MICs greater than or equal to the concentration shown. Numbers at the lower end of tested dilution ranges represent isolates with MICs equal or lower than the lowest concentration tested. Where available, breakpoints are indicated by a single vertical line (susceptible – intermediate; I) and/or a double vertical line (susceptible or intermediate – resistant; R). Grey shaded areas indicate concentrations of antimicrobials that were not tested. Amoxicillin/clavulanic acid (2:1); trimethoprim/sulfamethoxazole (1:19). MICs were determined by broth microdilution susceptibility testing as recommended by CLSI [34].

4.8 Biofilm formation by *Streptococcus equi* ssp. *equi* isolates

WGS-submitted *See* isolates (n = 191) were examined for their capabilities to form a biofilm *in vitro* using the microtitre plate biofilm assay as described by Yi *et al*. (2014) [202] with modifications. Biofilm formation was assessed after bacterial growth at 37 °C in different media (BHI-broth, THB, and THB supplemented with 1 % fibrinogen) and over two different incubation times (24 and 48 hrs).

4.8.1 Effects of medium type and incubation time on biofilm formation by *Streptococcus equi* ssp. *equi* isolates

In the first step, all WGS-submitted *See* isolates (n = 191) were assessed for biofilm formation during growth at 37 °C in BHI-broth and in THB for 24 and 48 hrs, respectively. Attached and stained bacterial biofilms were quantified using photometric measurement of OD_{570} **(Figure 16 A**) and by calculation of the specific biofilm formation (SBF) index (**Figure 16 B**). Results are presented in box whisker plots generated with Microsoft Excel 2016 programme.

Analysis based on OD_{570} showed that *See* isolates had formed more biofilm in THB than in BHI-broth ($p < 0.0001$, Wilcoxon signed rank test), both after incubation for 24 and for 48 hrs (**Figure 16 A**). The incubation time had a significant effect on *See* biofilm formation in THB since isolates had formed significantly more biofilm at 48 hrs ($p < 0.0001$). The same results were obtained when the biofilm formation was quantified by the SBF index. *See* isolates significantly formed more biofilms in THB than in BHI-broth ($p < 0.0001$), both during incubation for 24 hrs and for 48 hrs (**Figure 16 B**). Regardless of the quantification method used, incubation time had no effect on biofilm production when *See* isolates were grown in BHI-broth ($p > 0.05$) (**Figure 16 A**).

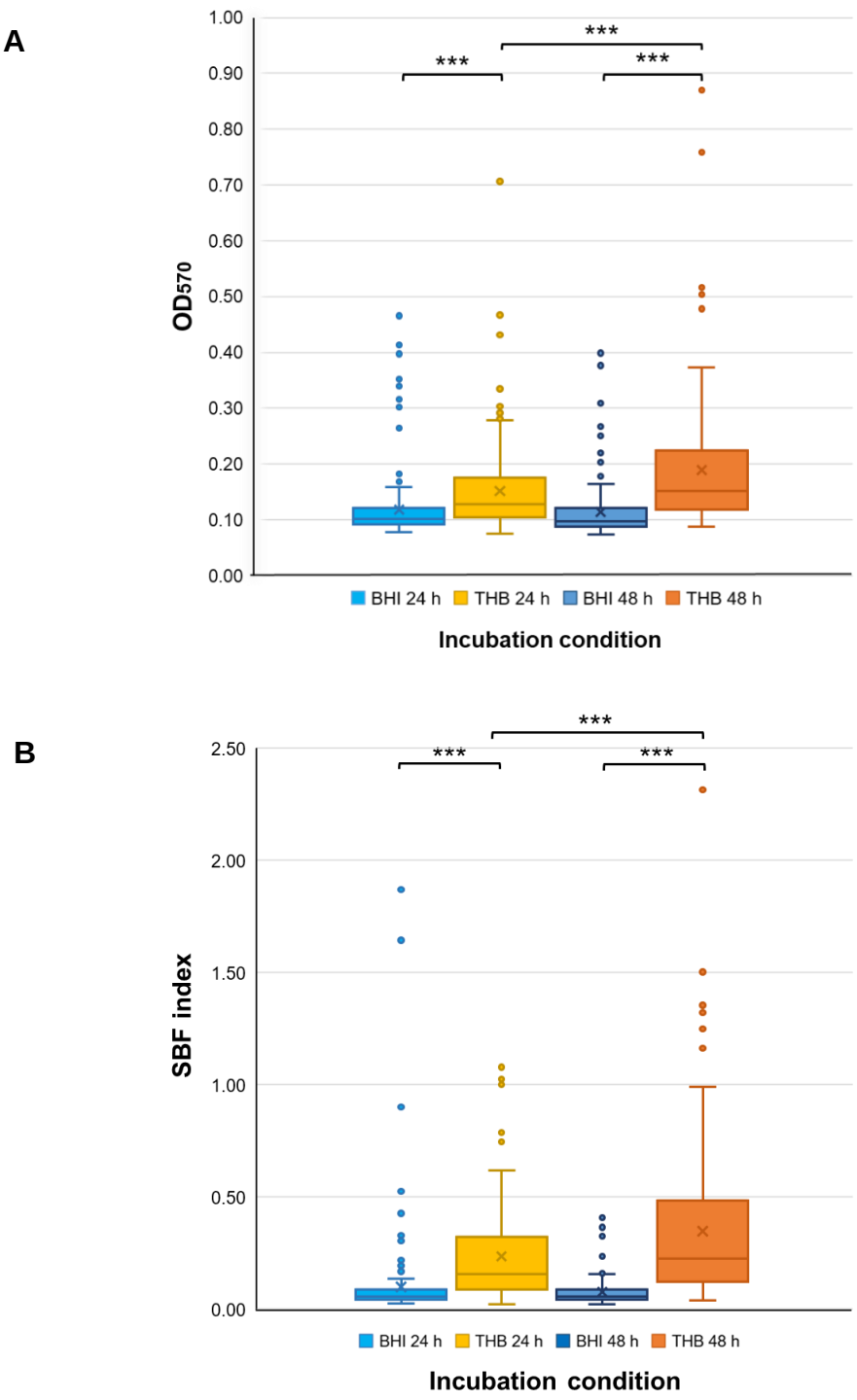


Figure 16. Biofilm formation by *S. equi* ssp. *equi* isolates (n = 191) after incubation at 37 °C in THB and BHI-broth for 24 and 48 hours.

Biofilm formation was quantified by assessing (**A**) the OD_{570} and **(B)** the specific biofilm formation (SBF) index. Box and whisker plots display minimum, maximum, quartile, arithmetic mean (x), median (horizontal lines), and outlier values at each incubation condition. THB: Tod-Hewitt broth, BHI: brain heart infusion broth, h: hours, ***: $p \leq 0.001$ (highly significant difference).

The correlation between OD_{570} and SBF index values is presented in **Figure 17**. There is a highly significant positive correlation between OD_{570} and SBF index ($r = 0.90$, $p < 0.001$, Spearman's correlation). The correlations between OD_{570} and SBF index for each incubation condition also proved positive and significant with the following Spearman's correlation coefficient (r) and p values: THB 24 hrs ($r = 0.80$, $p < 0.001$, Spearman's correlation test), BHI 24 hrs ($r = 0.88$, $p < 0.001$), THB 48 hrs ($r = 0.90$, $p < 0.001$), and BHI 48 hrs ($r = 0.93$, $p < 0.001$). The growth of bacteria after incubation was considered in calculating the SBF index. Since this serves as normalization and better represents the biofilm formation of the bacteria, the SBF index values were used for further analysis.

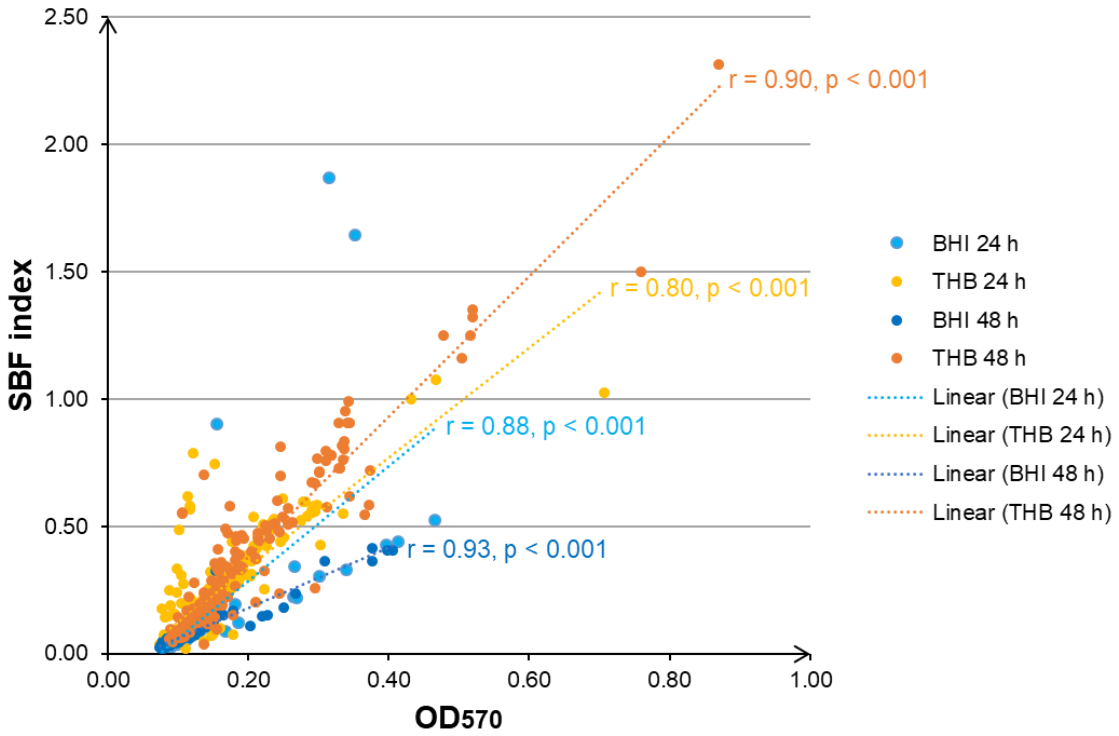


Figure 17. Correlation between OD_{570} and SBF index values for biofilm formation by *S. equi* ssp. *equi* isolates (n = 191) after incubation at 37 °C in THB and BHI-broth for 24 and 48 hours.

Spearman's correlation coefficient (r) and p values are indicated beside each line. SBF: specific biofilm formation, OD: optical density, THB: Tod-Hewitt broth, BHI: brain heart infusion broth, h: hours.

4.8.2 Effects of fibrinogen on biofilm formation by *Streptococcus equi* ssp. *equi* isolates

As demonstrated above, *See* isolates formed more biofilm during growth in THB. Therefore, only THB was used for further experiments to investigate the effect of fibrinogen supplementation on biofilm formation. Biofilm formation of selected *See* isolates (n = 58) was assessed after growth at 37 °C in THB supplemented with 1 % fibrinogen and in THB without fibrinogen for 24 and 48 hrs. Isolates were selected based on their genotype (all isolates with non-typeable *seM* alleles, representatives of every *seM*-allele group, ST, and BAPS-cluster), age of the host, and duration of the disease (acute or chronic).

After incubation for 24 hrs, *See* isolates formed significantly ($p < 0.0001$; Wilcoxon signed rank test) more biofilm in THB with fibrinogen (SBF index: 0.39 ± 0.3) than in THB without fibrinogen (0.16 ± 0.21) (**Figure 18**). This also applied to the incubation for 48 hrs (0.26 ± 0.2 vs. 0.13 ± 0.16). Interestingly, biofilm formation in THB with fibrinogen was significantly higher after 24 hrs than after 48 hrs of incubation ($p < 0.0001$).

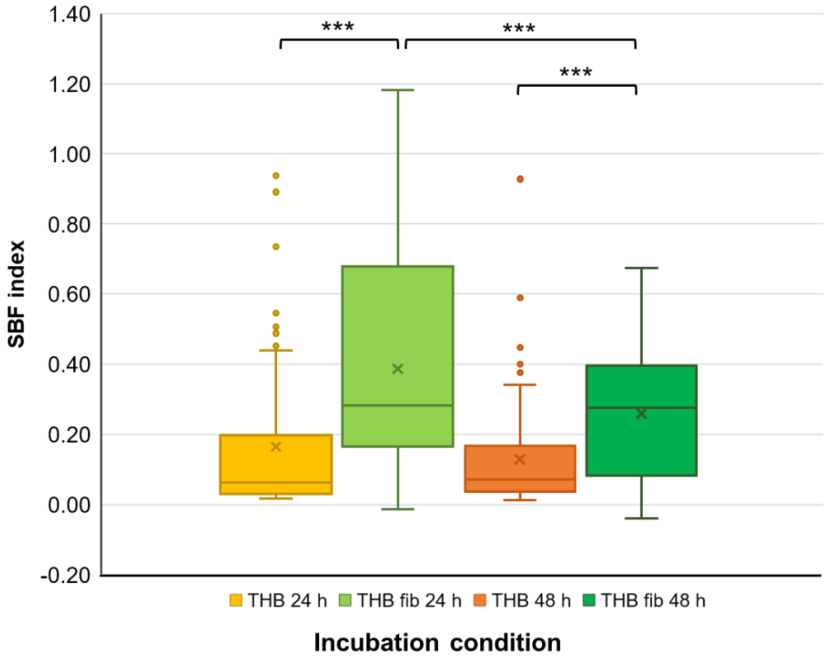


Figure 18. Biofilm formation by *S. equi* ssp. *equi* (n = 58) after incubation at 37 °C in THB and THB supplemented with 1% fibrinogen for 24 and 48 hours.

Box and whisker plots display minimum, maximum, quartile, arithmetic mean (x), median (horizontal lines), and outlier values at each incubation condition. SBF: Specific biofilm formation, THB: Tod-Hewitt broth, fib: 1 % fibrinogen, h: hours, ***: $p \leq 0.001$ (highly significant difference).

4.8.3 Comparative assessment of biofilm formation by different *Streptococcus equi* ssp. *equi* genotypes

The correlation between biofilm formation by *See* isolates and their genotypes (*seM* allele groups, STs, and BAPS clusters) was analysed. The analysis included

a) SBF index data of all WGS-submitted *See* isolates (n = 191) after incubation in BHI-broth and THB for 24 and 48 hrs, and
b) SBF index data of selected WGS-submitted *See* isolates (n = 58) after incubation in THB and THB supplemented with 1 % fibrinogen for 24 and 48 hrs.

4.8.3.1 Association between biofilm formation and *seM* allele group

SBF index values obtained from *See* isolates (n = 191) of groups *seM*-6 (8 isolates), *seM*-8 (1), and *seM*-9 (172) as well as from isolates with a non-typeable *seM* (10) were investigated. On average, biofilm formation by *See* isolates of *seM*-9 group was higher than by isolates of other *seM* allele groups (n = 19) for all incubation conditions, except the single isolate of *seM*-8 group in THB for 48 hrs. These differences were significant for incubation in THB for 24 and 48 hrs ($p < 0.001$, Wilcoxon-Mann-Whitney test), but not for incubation in BHI for 24 and 48 hrs ($p > 0.05$), respectively (**Figure 19 A**). *See* isolates harbouring a non-typeable *seM* allele produced significantly less biofilm compared to isolates of groups *seM*-9 in THB for 24 and 48 hrs, respectively ($p < 0.001$). These differences were not significant when isolates were incubated in BHI for 24 and 48 hrs ($p > 0.05$). *See* isolates harbouring a non-typeable *seM* allele also produced significantly less biofilm ($p = 0.002$) compared to isolates of all other *seM* groups (n = 181) in THB for 24 and 48 hrs, respectively.

In order to assess the effect of fibrinogen on biofilm formation in relation to *seM* allele groups, the SBF index values obtained from selected *See* isolates of groups *seM*-6 (4 isolates), *seM*-8 (1), *seM*-9 (43) and with a non-typeable *seM* (10 isolates) were investigated here. Most striking, *See* isolates encoding for a non-typeable *seM* allele produced very little biofilm even in THB supplemented with fibrinogen. In comparison to all other *See* isolates (n = 48) this biofilm formation was significantly less ($p < 0.001$; Wilcoxon-Mann-Whitney test) **(Figure 19 B).** On average and as seen before, isolates of groups *seM*-6, *seM*-8, *seM*-9 produced the greatest amount of biofilm when grown in THB with fibrinogen for 24 hrs. Differences between 24 hrs and 48 hrs of incubation were significant for isolates of group *seM*-9 ($p < 0.001$, Wilcoxon signed rank test) but not for isolates of group *seM*-6

(p > 0.05). In **Figure 19 B** significant differences between incubation conditions are not indicated within the same *seM* allele group.

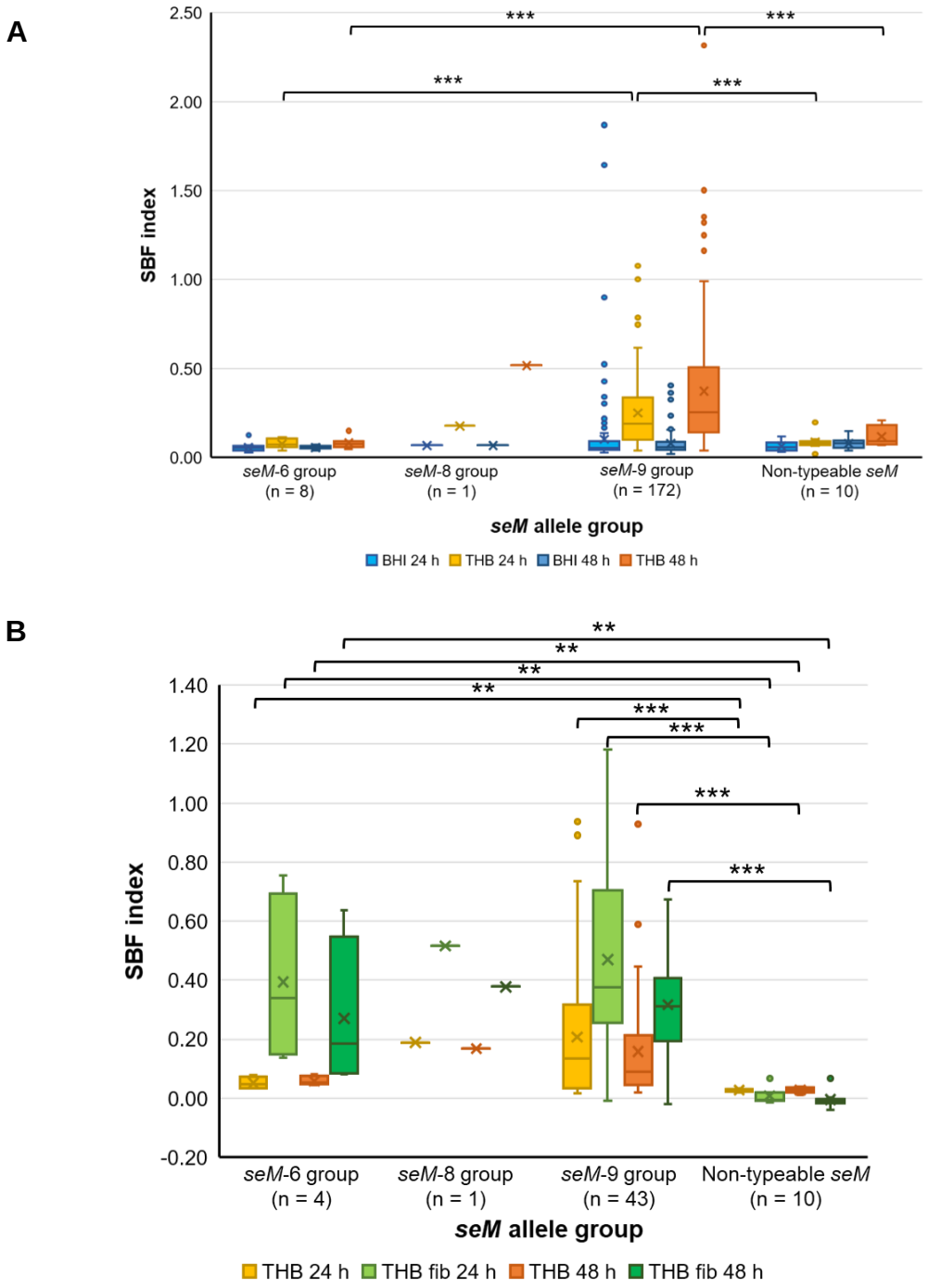


Figure 19. Biofilm formation by *S. equi* ssp. *equi* isolates of different *seM* allele groups under various incubation conditions.

Biofilm formation was quantified by assessing the SBF index values after bacterial growth at 37 °C for 24 or 48 hrs in **(A)** BHI-broth and THB (n = 191 isolates) and **(B)** THB and THB supplemented

with 1 % fibrinogen (n = 58 isolates). Box and whisker plots display minimum, maximum, quartile, arithmetic mean (x), median (horizontal lines) and outlier values at each incubation condition. SBF: Specific biofilm formation, THB: Tod-Hewitt broth, BHI: brain heart infusion broth, fib: fibrinogen, h: hours, ** p ≤ 0.01 (moderate significant), *** p ≤ 0.001 (highly significant).

4.8.3.2 Association between biofilm formation and MLST sequence type

As demonstrated above two highly related STs were identified among the 191 *See* isolates, namely ST-151 (n = 141 isolates) and ST-179 (n = 50). The comparison of the two STs in terms of biofilm production revealed that *See* isolates of ST-151 produced more biofilm than isolates of ST-179. This difference was highly significant under all tested incubation conditions ($p < 0.001$; Wilcoxon-Mann-Whitney test) with the exception of BHI 24 hrs ($p = 0.33$) (**Figure 20 A**).

Examination of the 58 selected *See* isolates showed that fibrinogen induced biofilm production both in ST-151 isolates (n = 36) and in ST-179 isolates (n = 22). On average induced SBF indices were higher in ST-151 than in ST-179 isolates but these differences were not significant ($p > 0.05$; Wilcoxon-Mann-Whitney test) (**Figure 20 B**).

4.8.3.3 Association between biofilm formation and cgMLST BAPS cluster

The *See* isolates of this study were distributed in three cgMLST BAPS clusters, namely BAPS-2 (181 isolates), BAPS-5 (9), and BAPS-6 (1). The comparison of the three BAPS clusters revealed that biofilm formation by BAPS-2 isolates was significantly higher than that of BAPS-5 isolates after incubation in THB for 24 and 48 hrs (**Figure 21 A**).

Examination of the 58 selected *See* isolates in terms of their biofilm formation in the presence of fibrinogen revealed that BAPS-2 isolates (n = 52) were more capable to form biofilm than BAPS-5 isolates (n = 5) both after 24 and 48 hrs of incubation. However, these difference were not statistically significant ($p > 0.05$, Wilcoxon-Mann-Whitney test) (**Figure 21 B**). Additionally, the SBF index of the single BAPS-6 isolate was higher than mean SBF index values of BAPS-2 and BAPS-5 isolates (**Figure 21 B**). The mean SBF index values determined after 24 hrs of incubation were significantly higher than those after 48 hrs in the BAPS-2 isolates ($p < 0.001$, Wilcoxon signed rank test), as well as in the BAPS-5 isolates ($p = 0.043$).

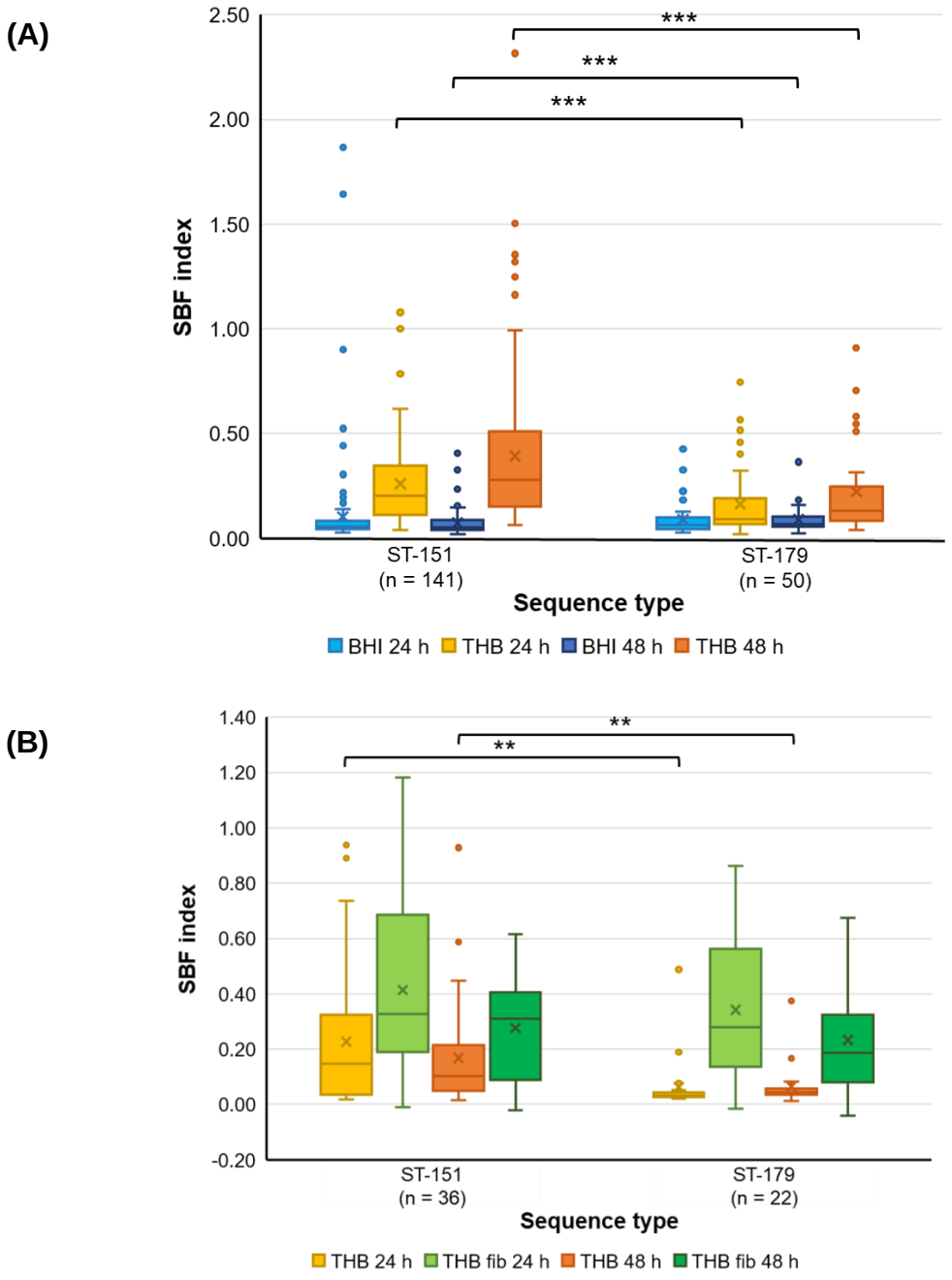


Figure 20. Biofilm formation by *S. equi* ssp. *equi* isolates of different MLST sequence types under various incubation conditions.

Biofilm formation was quantified by assesing the SBF index values at 37 °C for 24 or 48 hrs of incubation **(A)** in BHI-broth and THB (n = 191 isolates) and **(B)** in THB and THB supplemented with 1% fibrinogen (n = 58 isolates). Box and whisker plots display minimum, maximum, quartile, arithmetic mean (x), median (horizontal lines) and outlier values at each incubation condition. SBF: Specific biofilm formation, THB: Tod-Hewitt broth, BHI: brain heart infusion broth, fib: 1 % fibrinogen, h: hours, **: $p \leq 0.01$ (moderate significant), ***: $p \leq 0.001$ (highly significant).

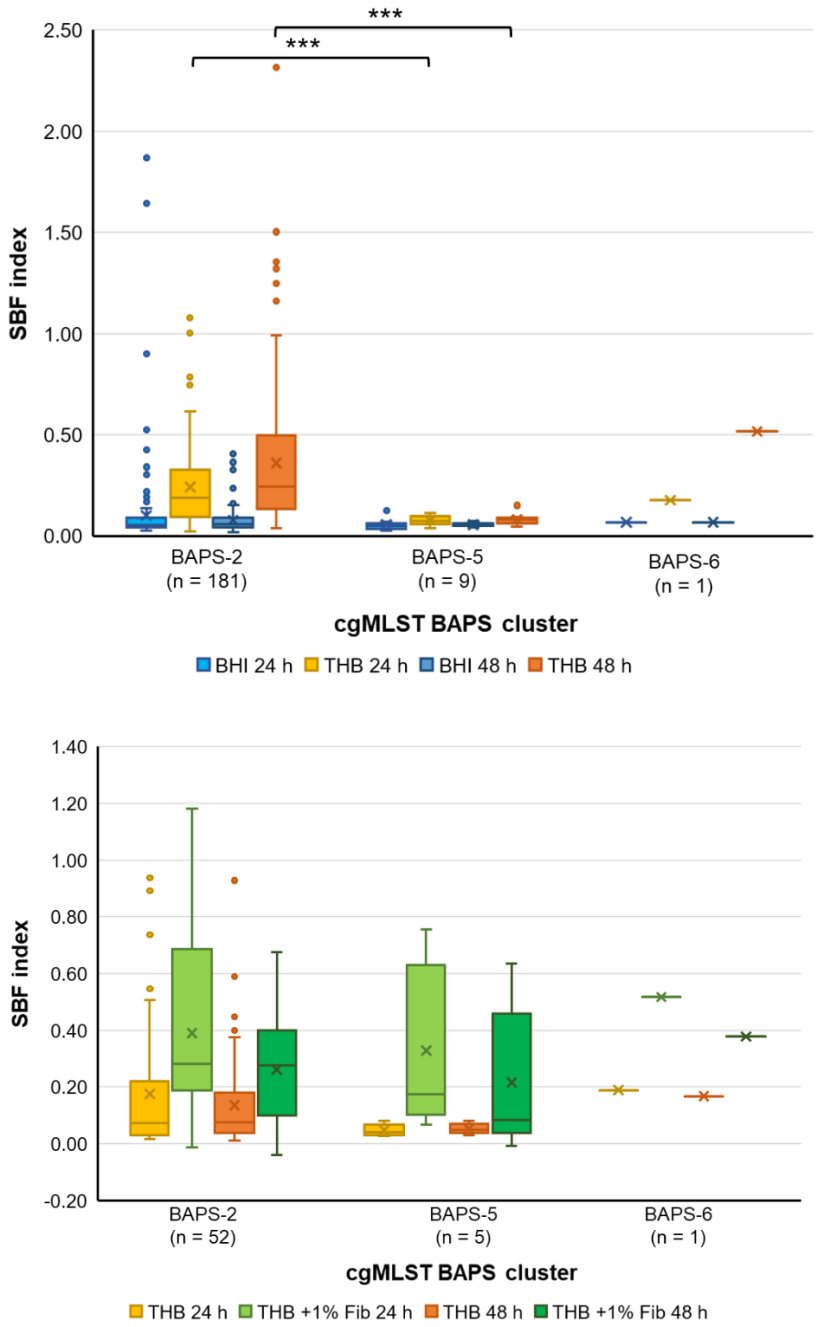


Figure 21. Biofilm formation by *S. equi* ssp. *equi* isolates of cgMLST BAPS clusters under various incubation conditions.

Biofilm formation was quantified by assesing the SBF index values at 37 °C for 24 or 48 hrs of incubation **(A)** in BHI-broth and THB (n = 191 isolates) and **(B)** in THB and THB supplemented with 1% fibrinogen (n = 58 isolates). Box and whisker plots display minimum, maximum, quartile, arithmetic mean (x), median (horizontal lines) and outlier values at each incubation condition. SBF: Specific biofilm formation, THB: Tod-Hewitt broth, BHI: brain heart infusion broth, fib: 1 % fibrinogen, h: hours, ***: $p \leq 0.001$ (highly significant).

5 Discussion

5.1 *Streptococcus equi* subspecies identification by MALDI-TOF MS and Se-mPCR

Results of the bacteriological examination supported the diagnoses of *See* and *Sez* infections in diseased equines. Both subspecies belong to Lancefield Group C and displayed a clear beta haemolytic hallo on BAP. At the genomic level, *See,* and *Sez* are closely related [73,177]. It is suggested that *See*, an equine-specific pathogen, has evolved from *Sez*, a mulitspecies-host pathogen [196]. Furthermore, around 85 % CDS of the *See* genome are orthologous to *Sez* [73]. The high phenotype and genotype similarity of these two subspecies make their discrimination difficult. Therefore, the identity of the putative *S. equi* was reassessed by MALDI-TOF and Se-mPCR.

The present study showed that MALDI-TOF MS could identify *See* correctly, since all isolates identified as *See* by MALDI-TOF MS were also identified as *See* by the Se-mPCR and WGS analysis. Spectra from 11 and 7 of *See* and *Sez* strains, respectively, were available in the MALDI-TOF MS database used in this study. Identification of microorganisms by MALDI-TOF MS relies on protein profiles of whole bacterial to the reference spectra [2,88]; thus, the number of reference spectra in the database influences the accuracy of the identification.

Furthermore, the method for the sample preparation influences the score value achieved by MALDI-TOF MS. Isolates that had a score value below 2.00 by using the direct transfer (dt) method reached a higher score value (≥ 2.00) as they were retested with the extended direct transfer (edt) method. Gram-positive bacteria have a thick peptidoglycan layer. In certain cases, the dt method fails to disrupt the peptidoglycan to prepare adequate proteins for bacterial identification by MALDI-TOF MS [2,168]. This phenomenon was mainly observed in mucoid colonies with higher hyaluronic acid capsules than in non-mucoid colonies. Generally, the spectra quality using the edt method is better compared to the dt method. Nevertheless, the edt method is more time and material-consuming.

The Se-mPCR detected the *sodA* gene encoding the manganese-dependent superoxide dismutase A protein (SodA) [37] in all *S. equi* isolates. The *sodA* gene was also used as a target to detect *S. equi* by PCR in other studies [4,14]. Furthermore, the Se-mPCR detected the ICE*Se*2 locus in all but two *See* isolates and none *Sez* isolates. Therefore, the Se-mPCR is valuable for identifying and distinguishing *See* from *Sez*. The ICE*Se*2 locus is

an integrative conjugative element that is specific for *See* and encodes for a non-ribosomal peptide synthetase system [72]. This system produces equibactin, a potential bacterial siderophore known to enhance iron acquisition and potentially contribute to the virulence of *See* [72,73]. Another study also targeted ICE*Se*2 to identify *See* using real-time PCR [195]. Analysis of the genome revealed that the two PCR-negative ICE*Se*2 isolates (*See* IHIT38875 and *See* IHIT43005) harboured the ICE*Se*2 locus only partially, and the Se-mPCR primer binding sites for the ICE*Se*2 locus were missing. *See* IHIT43005 was isolated from a nasopharyngeal swab, while *See* IHIT38875 was recovered from the guttural pouch lavage fluid of a persistently infected horse with chondroid. This finding correlates with a previous study by Harris *et al*. (2015), where *See* isolates with missing equibactin locus of the ICE*Se*2, were also recovered from persistently infected horses [71]. Studying the virulence of *See* isolates with missing ICESe2 compared to those with complete ICE*Se*2 will be an interesting endeavour.

The Se-mPCR was not able to detect the ICE*Sz*1 signature locus of *Sez* in more than half of the tested *Sez* isolates. Therefore, the identification of *Sez* using Se-mPCR needs to be improved by targeting a more stable and conserved *Sez* gene. The ICE*Sz*1 locus is a mobile genetic element (MGE) that encodes putative DNA-binding proteins, conserved hypothetical proteins, recombinase, relaxase, sensor histidine kinase, and ABC-transported proteins [73]. The target of the ICE*Sz*1 primers of the Se-mPCR is a putative hypothetical protein [73]. WGS analysis of four *Sez* isolates revealed that the ICE*Sz*1 locus was partially missing in two isolates and completely missing in two other isolates. However, none of the *See* isolates harboured the ICE*Sz*1 locus by PCR test (n = 265) and WGS analysis (n = 191), showing that ICE*Sz*1 is specific for *Sez*. *In silico* analysis using a publicly available NCBI database also confirmed the exclusivity of the ICE*Sz*1 locus to Sez.

5.2 Significance of *seM* typing for strangles outbreaks identification

The utilization of *seM* typing, which is based on nucleotide sequence polymorphism in the N-terminal, provides evidence that *See* strains with different *seM* alleles were responsible for strangles cases in Germany. Analysis by grouping the isolates into ten-year intervals revealed that the relative frequency of isolates belonging to the *seM*-9 allele group increased; on the contrary, isolates belonging to the *seM*-6 allele group decreased. A similar observation has been identified in strangles cases in the United

Kingdom [77,138]. Additionally, another study reported that the strangles cases in Sweden were dominated by isolates harbouring the *seM*-9 allele [97].

The results of the present study provide supporting evidence that *See* strains with identical *seM* alleles were identified within the same spatial-temporal cluster, indicating the same source of infection. In this case, *seM typing* is valuable in tracing the infection chain of strangles. Conversely, horses originating from distinct spatial-temporal clusters were infected by *See* harbouring the same *seM* allele, suggesting a widely distributed *seM*-allele, for example, the *seM*-9 allele. One plausible interpretation of this finding is that the role of contemporary horse transportation, such as for tournaments or breeding purposes, facilitates the transmission of *See*. Consequently, comprehensive anamneses of horses, particularly regarding contact and movement, prove essential for effectively identifying the outbreak source.

See isolates harbouring different *seM* alleles were identified in the same spatial-temporal cluster. This finding may be explained by the idea that mixed infections that are not epidemiologically related caused the strangles cases in this particular spatial-temporal cluster. Alternatively, the diversity of *seM* alleles within a spatial-temporal cluster could be attributed to mutations in the typing region of the *seM* gene, resulting in phylogenetically closely related but distinct *seM* alleles [84].

The 29 novel *seM* alleles identified in Germany indicate that the *seM* gene is under diversifying selection [84,96] and that its mutation is an ongoing process [138]. Nucleotide variation of the *seM* gene leads to synonymous and nonsynonymous amino acid changes [71,77,84]. *SeM* typing uses a region that binds fibrinogen and IgG [84,118,119], and may be subject to high immune pressure that causes mutation. Furthermore, a persistent infection might lead to mutation and deletion of the *seM* gene [30]. Identification of different *seM* alleles that are phylogenetically closely related within the same spatial-temporal cluster indicates that a particular *seM* allele might cause an outbreak, but subsequent immune pressure may lead to new alleles [84,96]. Since the mutation of *seM* is an ongoing process [84], it is suggested that interpretation of the *seM* allele based on its distinct phylogenetic group may provide more robust discriminatory evidence and is valuable for identifying the source of infection. Therefore, the identification of different but phylogenetically closely related *seM* alleles in a spatial-temporal cluster may not exclude their belonging to the same strangles epidemic [138].

Although the present study supports that the *seM* typing may help to track the infection chain of strangles cases if strains with rare *seM* alleles are involved, the presence of a widely distributed *seM* allele decreases its discriminatory power. More than half of the isolates in this study harboured the *seM*-9 allele, which was distributed in many spatial-

temporal clusters and unrelated strangles cases. As the *seM*-9 allele is distributed globally (29.4%, 223 of 759 global isolates) [62], it is impossible to conclude recent transmission when *See* isolates with this allele are involved. It is tempting to speculate that host-pathogen interaction contributes to the mutation of the *seM* gene, leading to a dominant allele, as shown in this study by the *seM*-9 allele.

The ten isolates that had non-typeable *seM* were from different districts in Germany; thus, no spatial link could be identified. The deletion site of the *seM* gene varied in size and position from one isolate to another, indicating that they had evolved independently. *See* isolate IHIT40246, which has a deletion of 411 bp in its *seM* gene, was isolated from a three-month old foal in 2017. The foal came from the same geographical location of a strangles outbreak in 2014 as in another study [185], indicating a persistent infection of the herd. The non-typeable *seM* isolates were obtained from the respiratory tract (nasal and guttural pouch nasal fluid) and abscesses of naturally infected horses; however, the severity of the clinical signs was not recorded in the sample submission form.

Interestingly, an isolate with non-typeable *seM* and one with *seM*-9 allele were obtained from the same horse with a sampling interval of one month, indicating that the horse might have been infected with several *See* strains at the same time or that the *seM* gene might have mutated during infection. A previous study identified *See* with truncated SeM protein with different sizes and positions of nucleotide deletion from outwardly healthy horses with persistent infection [30]. The mechanism and reason for SeM truncation remain unclear. Neither the deletion sites nor the flanking DNA contained nucleotide sequences indicative of inverted repeats or transposons [30]. However, during infections and outbreaks, the M-proteins of *S. pyogenes* may change because of intragenic recombination between identical tandem repeats [57], insertion, and point mutations of the N-terminal [70]. It is unclear whether the mechanism of SeM truncation can occur in reverse. Possibly, *See* strains acquire a selective advantage in their host environment by altering their *seM* gene sequences.

The mutation of the *seM* gene has the potential to impact the virulence of *See*. Alterations in the N-terminal end of the *seM* gene lead to changes in the IgA epitope and influence the response of systemic T-cells [179]. However, these variations do not affect the binding of fibrinogen or antibody opsonization to the SeM protein [179]. Previous studies have indicated that the SeM proteins encoded by *seM*-1, -2, or -8 alleles were able to bind fibrinogen [30,178]. Isolates with truncated SeM, characterized by deletions in the fibrinogen and IgG binding regions of the *seM* gene, exhibit reduced resistance to phagocytosis [30]. Despite this vulnerability, these isolates can persist in their hosts under conditions of purulent inflammation [30]. It is noteworthy that the *See* isolates in the present study, including

those with non-typeable *seM* isolates, were obtained from naturally infected horses. This suggests that isolates with non-typeable *seM* may contribute to infection or that the deletion of *seM* occurred as a result of prolonged infection. Further investigations are necessary to explore the virulence of various *seM* alleles, including isolates with non-typeable *seM*. This includes an in-depth examination of the host immune response against different *seM* alleles and their capacity to bind fibrinogen and IgG.

The present study correlates with the previous research that the discriminatory power of *seM* typing is higher than MLST. So far, only nine STs but 244 *seM* alleles have been identified from *See* strains worldwide [122,196] (PubMLST database, last accessed September 30, 2022). The sequence typing of virulence or hypervariable genes has been utilized to increase the discriminatory power of the MLST scheme of *S. uberis* [204]. The variability of a virulence gene is especially appropriate for investigating local epidemiology, allowing the differentiation of strains without relying on variation in housekeeping genes [84].

In the present study, seven *See* isolates obtained from horses in Indonesia exhibited the same *seM*-166 allele and pulsotype in macrorestriction analysis using *Sma*I and *Apa*I, indicating a potential infection by a clonal *See* lineage. An epidemiological study of *See* strains in Sweden found that *seM* typing has higher discriminatory power than macrorestriction analysis [97]. Notably, *seM* typing proves to be more efficient in terms of material and time consumption compared to both MLST and macrorestriction analysis.

SeM typing has been used to identify outbreaks in the United Kingdom [77,138], Sweden [97], United States of America [124], Brazil [96], Thailand [183], Indonesia [154], imported horses in Japan [82], and donkeys in China [43]. *SeM* typing could help to identify infection clusters and develop risk assessments regarding outbreak identification, policy for importation of horses, preparation of equine events, and vaccination. However, WGS-based typing schemes are required to resolve the *See* population structure appropriately.

5.3 Population structure of *Streptococcus equi* ssp. *equi*

The results of this study confirmed the low phylogenetic diversity of *See* with respect to MLST [71,122,196], since only two highly related STs (ST-151 and ST-179) were identified in the 191 investigated *See* isolates. *See* isolates recovered from horses in Germany were dominated by ST-151. This finding differed from the previous studies, where ST-179 dominated the global population structure of *See*, and *See* isolates from Germany all belonged to ST-179 [122,196]. ST-151 isolates were more likely to have been isolated from horses in 2011-2020 compared to 2001-2020 ($p = 0.0003$, Fisher's exact test), indicating

this ST has been fitter in the last decade. Interestingly, another study reported that ST-151 isolates were recovered more often in horses vaccinated with Equilis StrepE and Pinnacle IN vaccine [71]. Equilis StrepE vaccine has been used to vaccinate horses against strangles in Europe since 2004 [49].

So far, only nine STs of *See* have been identified globally from 1955 until 2020, namely ST151, ST179, ST281, ST282, ST283, ST325, ST395, ST396, and ST402 (PubMLST database, last accessed September 30, 2022) [122]. All STs of *See* are either single or double locus variants of ST-179, consequently belonging to the CC-179. In contrast, using the same MLST scheme, 510 STs have been identified globally for *Sez* (PubMLST database, last accessed September 30, 2022). MLST schemes identify variations in housekeeping genes that accumulate very slowly over time [196]. Genomic analyses of *See* worldwide suggest that international transportation, mixing of horses, and establishment of horse breeding centres during the global conflicts in the nineteenth and twentieth centuries had led to the emergence of a fitter strain of *See*, from which all contemporary strains descended [71]. As a consequence of this evolution and based on the result of the MLST analysis, there is insufficient genetic diversity in the seven housekeeping genes to adequately distinguish the population structure of *See* [122]. Therefore, typing methods with higher discriminatory power than MLST, for example, using WGS-based analyses, are needed to resolve the population structure of *See* in more detail.

See strains in this study (n = 191) fitted into the current global phylogenetic tree of *See* based on the cgMLST analysis. This result strongly implies that all isolates in the present study can be regarded as descendants of the abovementioned *See* strain that has globally expanded since the end of the 19th or beginning of the 20th century. Most of the German isolates belonged to BAPS-2, which is consistent with a previous study that BAPS-2 is the predominant genetic lineage in Europe [122]. The single German isolate that belonged to BAPS-6 (*See* isolate IHIT9732) was isolated in 2003. This isolate is phylogenetically closely related to *See* strain UK231350 (pairwise difference = 4 cgSNPs), isolated during the same year in the United Kingdom, indicating an international transmission.

In the present study, only eight isolates from Germany (4.5 %) belonged to BAPS-5. The low number of BAPS-5 isolates identified in this study aligns with prior research suggesting a declining prevalence of BAPS-5 isolates in Europe [122]. Mitchell *et al*. (2021) documented 12 strangles cases in Lewitz, Germany, during 2014–2015, all attributed to BAPS-5 isolates [122]. Notably, in this study, *See* isolate IHIT40246, which belonged to BAPS-5, was recovered from a foal at the same aforementioned location in 2017. *See* IHIT40246 differed from one another by 8 to 34 pairwise cgSNPs with the isolates recovered in Lewitz in 2014-2015, forming a distinct yet closely related subcluster. A noteworthy observation is

that *See* IHIT40246 harboured a non-typeable *seM*, indicating that this strain probably has been circulating in the farm and mutated, having a decayed genome. Strangles outbreaks have been reported in weanlings in Lewitz with decreased severity [185]. It seems that the descendants of this strain are still circulating in the herd. It would be beneficial to expand the current findings by examining more isolates causing strangles cases in Lewitz for a more extended period to study the microevolution of *See*.

The results of this study provide supporting evidence that there is a concordance between SLST, MLST, and cgMLST for *See*. All ST-151 isolates belonged to BAPS-2, while ST-179 isolates were distributed in several BAPS clusters, showing a dependency between MLST and cgMLST ($p < 0.0001$, Fisher's exact test). This result pattern is consistent with another literature [122]. The analysis of the global *See* population revealed that there is a dominant *seM* allele in each of the BAPS clusters, as reported by another study [122]. In addition, all isolates belonging to a particular *seM* allele group clustered to the same BAPS cluster, indicating a correlation between *seM* typing and cgMLST analysis. However, this did not apply for isolates with non-typeable *seM*, since they were identified in BAPS-2 and BAPS-5 of this study and in all but not in BAPS-3 of the global *See* population [62,122].

The spatial-temporal analysis of *See* isolates from horses in Germany aimed to uncover epidemiological links and potential strangles outbreaks within the equine population retrospectively. Spatial-temporal analysis of isolates that were recovered from horses in Germany revealed that some isolates within a spatial-temporal cluster were from the same cg-genotype (pairwise difference = 0 cgSNPs). In several clusters, isolates exhibited highly similar cg-genotypes, forming subclusters within the phylogenetic cgMLTS tree. Ten spatial-temporal clusters exclusively comprised *See* isolates with identical combined genotypes (ST, BAPS cluster assignment, and seM allele), suggesting a localized presence of specific *See* strains. This finding underscores the utility of molecular typing using cgMLST in tracing the epidemiological links of strangles cases in Germany. Furthermore, *See* isolates recovered from strangles cases in three provinces in Indonesia during 2018 clustered together as an exclusive subcluster in BAPS-2 with pairwise differences of 2–14 cgSNPs, indicating that these horses were infected by closely related strains [154]. Contrastingly, spatial-temporal analysis revealed that *See* infections in the same spatial-temporal cluster must not necessarily all be epidemiologically related. This observation could be clarified by considering the possibility that unrelated mixed infections contributed to the occurrence of strangles cases in this specific spatial-temporal cluster.

The spatial-temporal analysis of *See* isolates with identical genotypes (identical cg-genotype, ST, and *seM* allele) in Germany reveals a complex pattern of distribution, suggesting both localized persistence and broader dissemination of specific *See* strains.

Genetically identical or highly similar *See* strains were identified in different spatial-temporal clusters, indicating the potential existence of unrecognized epidemiological links. These findings can be explained by the modern transportation of horses by bringing horses from different distant locations to a specified location at the time of a tournament or breeding. Transportation of asymptomatic carrier horses is a severe factor in the transmission of strangles [23,128].

Although the present results support the utility of the *See* cgMLST analysis publicly available in Pathogenwatch [122] for tracing epidemiological links between strangles cases, it is appropriate to recognize several potential limitations of the spatial-temporal analysis performed in this study. First, the geographical location used for the spatial-temporal analysis was a combination of the location of the owner and the submitting veterinarian practice due to data protection and owner confidentiality. It cannot be ruled out that the geographical locations used here did not correctly represent the location of the horses. Second, the data interpretation relied on the information submitted by the corresponding veterinarian, which in many cases did not include the anamneses and clinical histories of the horses. Despite these limitations, the present study has enhanced the understanding of the molecular epidemiology of strangles, especially in Germany.

In terms of future research, it is crucial to define a threshold of cgSNPs to differentiate isolates that are variants of a *See* outbreak strain and epidemically non-related *See* isolates. However, it is challenging to determine this threshold since the infection status (acute or persistent) influences the genomic diversity of *See* [71]. Additionally, this threshold may differ from that applicable to other streptococci, considering that the mutation rate of *See* is comparatively slower [71]. It has been reported that *See* has an average substitution rate per core genome site of 5.22×10^{-7} per year [71], equivalent to approximately one substitution yearly [122]. This rate is notably slower than the core genome substitution rates reported for other streptococci, like *S. pyogenes* (1.1×10^{-6}) [41] and *S. pneumoniae* (1.57×10^{-6}) [39].

5.4 Pathotyping *S. equi* ssp. *equi* using virulence-associated genes

See utilizes various virulence factors to infect the host and cause strangles. In this study, the VAGs of *See* were classified into adherence, immune modulation, nutritional metabolic factors, exoenzyme, and exotoxins. The genetic diversity of *See* isolates with respect to their VAGs were investigated *in silico*.

The results of this study underscore the diversity in the VAGs composition among the *See* isolates, resulting in 107 virotypes within the 191 investigated *See* isolates. More than half (n = 38, 56.7 %) of the examined VAGs were detected in all investigated isolates. Certain VAGs or various combinations of VAGs were missing in at least one *See* isolate. Most of the virotypes (n = 85, 79.4 %) were only found in one isolate, which suggests a high degree of variability at the level of individual strains according to their VAGs. Meanwhile, the presence of 14 virotypes in two isolates and eight virotypes in at least three isolates indicate shared patterns of VAGs composition among subsets of isolates. These findings hint at a potential commonality in the virulence profiles of *See* strains having the same virotype. The diverse virotypes imply that the virulence potential of *See* strains may vary, possibly contributing to differences in clinical manifestations and persistence of *See* infections. Mutations and gene losses can cause changes in the structure, function, and expression of the VAGs, which can lead to differences in the pathogenicity of the bacteria [73]. The genome of *See* has been observed to undergo continuous mutation, manifested by the acquisition and loss of VAGs, potentially impacting the virulence of this pathogen [71,73].

The results indicate that VAGs for adherence vary in *See* isolates. *See* produces various adhesins that can bind to the fibronectin or collagen of the host. Although most of the VAGs for adherence were present in all *See* isolates, genes encoding the following putative streptococcal collagen-like (Scl) surface anchored proteins termed as SclD (SEQ0280), SclE (SEQ0633), SclF (SEQ0855), SclG (SEQ0090), SclI (SEQ1817), and Scl3F were not identified in all isolates. Notably, none of the BAPS-2 isolates harboured the gene encoding SclE, suggesting that another Scl protein probably replaced its function in BAPS-2 isolates. Scl proteins are involved in adhesion to host tissues, evasion of host immune responses, and pathogenesis by mediating interactions with other microorganisms and host components [105]. Scl proteins contribute to the virulence of *S. pyogenes*, *S. agalactiae*, *S. pneumoniae*, *See*, and *Sez* and are essential for establishing infections [105]. However, the precise role of the above-mentioned Scl proteins produced by *See* remains unclear. Based on cluster analysis of sequences, SclC, D, E, G, H, I of *See* and *Sez* belong to the Scl cluster 3, whereas SclF belongs to an independent cluster 4 [105]. The amino acid sequences of Scl from *See* were similar to those from *Sez*, where SEQ0280, SEQ0633, SEQ0855, SEQ0090, and SEQ1817 of *See* strain 4047 are homologs to SZO17540 (76.5 % amino acid similarity), SZO13850 (80.1 %), SZO12230 (83.3 %), SZO00840 (66.7 %), and SZO03720 (79.3 %) of *Sez* strain H70 [73].

Most isolates in this study harboured all genes of the *has* locus. The *has* locus, composed of *hasA* (hyaluronate synthase), *hasB* (UDP-glucose dehydrogenase), and *hasC1* (UDP-glucose pyrophosphorylase), is responsible for hyaluronic acid capsule

synthesis in *See* [73]. Genes of the *has* locus were identified in most of isolates (90.6%). The sample type of the *See* isolates where the genes of the *has* locus were missing were mainly from the guttural pouch lavage fluid. Infection of the guttural pouch may indicate that the host suffered from a persistent infection. Accordingly, a previous study reported gene deletion in the *has* locus was identified in isolates recovered from persistently infected horses [71].

The hyaluronic acid capsule production of *See* correlates with increased resistance to phagocytosis by masking the bacterium from the host immune system since host tissue also has hyaluronic acid [180]. Deletion of the genes in the *has* locus reduces the capsule synthesis [177,188], influences its phenotypic morphology on BAP [71,173], and reduces the virulence of *See* [177,188]. In a study using reverse transcription qPCR technique to measure the amount of *hasA* gene expression, a *See* strain with deleted *hasA* gene had significantly less hyaluronic acid capsule than the wild-type strain [71]. Accordingly, a *See* strain with an amplification of the *has* locus had significantly higher transcription and hyaluronic acid levels on its surface than a wild-type strain [71]. The production of the hyaluronic acid capsule by *See* gives a phenotypic mucoid colony appearance on BAP or COBA [6,173,180]. Interestingly, phenotypic switching of *See* from the same isolate is caused by the transcription level of the *has* locus rather than genome mutations [173]. A study showed that the transcription level of genes in the *has* locus was significantly lower in less mucoid colonies compared to highly mucoid colonies of the same *See* strain, which showed an *in vitro* switching phenotype on COBA plates [173]. The ability of bacteria to switch their phenotype gives advantages to maximize their survival and infectivity in a changing environment.

The data suggest that *See* isolates are less diverse in term of genes coding for exoenzymes, as all isolates harboured the VAGs for streptokinase, endoglycosidase, and IgG-endopeptidase. Variability was observed in VAGs encoding hyaluronate lyase, namely SEQ1479 and SEQ2045. SEQ1479 was identified in all isolates, while SEQ2045 was not identified in 6 % of the isolates belonging to BAPS-2. Streptococci utilize hyaluronate lyases to invade the host by degrading connective tissue hyaluronic acid and chondroitins [11]. SEQ2045 is encoded by prophage φ-Seq4 and shares 75 % amino sequence identity with HylP1 from *S. pyogenes*, which is also encoded by a prophage [28,73,101]. Phage hyaluronate lyase promotes bacterial lysogenization by facilitating the penetration of streptococcal hyaluronan capsules [101]. While a study reported that SEQ2045 could not degrade chondroitin [101], another study reported that SEQ2045 could completely degrade hyaluronan and other glycosaminoglycans [166]. Additionally, infected horses produced antibodies against SEQ2045, indicating that SEQ2045 has an essential role in pathogenesis [101].

The VAGs encoding superantigenic toxins SeeH, SeeI, SeeL, and SeeM were widely distributed in *See* isolates of this study. The results of this study support that superantigens (sAgs) play a crucial role in the pathogenesis of strangles, since only 3 % of the isolates had deleted sAgs, and there was no isolate where all of the sAgs were deleted. It is known that sAgs disrupt the immune system by activating non-specific T-cells and producing excessive cytokines as a pro-inflammatory host response [142,170]. Every sAg can bind to the MHC class II to facilitate T cell recognition [142]. Superantigens are upregulated as the bacteria infect the host [142]. Considering the role of sAgs, deletion of one or more of these genes may result in a less virulent strain. However, although one or two of their sAgs were deleted, these *See* isolates in this study could still cause infection with clinical signs of strangles and were isolated from nasal swabs and abscesses. All four sAgs were detected in an isolate recovered from chondroid (IHIT42729), indicating a persistently infected horse. This finding supports that isolates from persistent infection can produce sAgs [116]. A study reported that mitogenic activity of sAgs was detected during both acute and persistent infection of *See* [116]. Additionally, mitogenic activity by sAgs was detected in abscess material and chondroid, but not in the serum of diseased horses [116].

The VAGs encoding for sAgs were missing in only six BAPS-2 isolates, with three of them originating from Indonesia. Notably, the results indicate that the absence of two sAgs in an isolate consistently corresponds to the specific sAgs carried on the same prophage. This finding aligns with existing literature, affirming that the sAgs of *See* are phage-encoded [73], mirroring th situation in *S. pyogenes* [8]. Genes encoding for SeeL and SeeM are carried by the prophage φ-Seq3, while genes encoding for SeeH and SeeI are carried by the prophage φ-Seq4 [73]. Amino acid sequence analyses revealed that the sAgs of *See* are highly similar to their homologues in *S. pyogenes* [8,143], indicating a potential horizontal gene transfer between streptococci. However, no homologues of SeeI and SeeH have been found in *Sez* [73], which implies that the prosecution of *seeH* and *seeI* has led to the evolution of a more virulent strain of *See* compared with its *Sez* ancestor. It has been documented that certain strains of *Sez* harbour *seeL* and *seeM*, but this occurrence is rare [73].

Iron acquisition serves as a crucial virulence factor for bacteria, and gram-positive bacteria utilize various pathways for this purpose. To acquire iron, *See* encodes siderophores using a non-ribosomal peptide synthetase (NRPS) system, termed equibactin [72,73]. *In silico* analyses conducted in this study revealed missing genes of the equibactin locus in nearly half of the See isolates. The missing genes of the equibactin locus varied within the isolates, from 1-13 genes of the 14 genes constructing the equibactin locus. Equibactin, encoded by the equibactin locus (*eqbA-N)*, is carried by the streptococcal

integrative conjugative element ICE*Se*2 [72]. Equibactin is specific to *See* and was not identified in *Sez* [73]. Notably, similarities were observed between equibactin and the yersiniabactin system found in the high-pathogenicity island of *Yersinia* sp. [72,120]

In this study, *See* isolates with missing genes of the equibactin locus were identified in acute as well as persistent infections of *See*. This contradicts the finding of past researchers that deletion of the equibactin locus was only detected in isolates from persistently infected horses [71]. It is tempting to speculate that *See* has another system to acquire iron or that equibactin can still function, although specific genes of the locus are missing. Another study hypothesized that a partial or entire gene deletion in the equibactin locus might reduce the virulence of *See* to cause acute infection [73]. In the present study, the *eqbA* gene was missing in only two isolates, while *eqbB* was lacking in only one isolate. Deletion of *eqbA*, which functioned as the repressor, was reported to increase iron importation, causing intoxicity, which was shown phenotypically by the small colony appearance of Δ*eqbA* mutants [72]. The promotor of the equibactin locus is located in the *eqbB* gene [72]. The *eqbE* gene, which product is responsible for aminoacylation, was missing in 27 % of the isolates in this study. Another study reported that Δ*eqbE* mutants showed no difference in bacteria growth rate in Todd-Hewitt agar compared to wild-type strains [72]. Additionally, Harris *et al*. (2015) reported that although ponies that were infected with Δ*eqbE* showed less pyrexia and had less post-mortem pathological changes compared to those infected with wild type strain, infection of Δ*eqbE* mutants was still able to cause abscessation [71]. Furthermore, this study identified *See* isolates in which the ABC transporter genes (*eqb*H-L) of the equibactin locus were missing. The lack of *eqbH-J*, *eqbK,* and *eqbL* genes may inhibit iron transportation since they encode ABC transporters essential for iron transport [72].

The current study underscores the significance of VAGs in bacterial pathogenesis. However, the limitation of this VAGs analysis lies in its reliance solely on *in silico* genomic examination. A valuable avenue for future investigation would involve scrutinizing the transcription levels of VAGs across diverse phylogenetic groups of *See*. Moreover, further research examining the implication of deleted VAGs mutants in horses may shed light on the role of VAGs. Despite the limitation of this study, the findings represent a progressive step in the identification of antigens for vaccine development or the selection of isolates for clinical trials involving *See*, distinguishing between potentially pathogenic and less pathogenic strains based on the presence of VAGs.

5.5 The diversity of Strangvac® vaccine antigen sequences in the *S. equi* ssp. *equi* genomes

This study assessed the diversity of Strangvac® vaccine antigens among 191 whole genome sequenced *See* isolates. Strangvac® comprises two recombinant fusion proteins (CCE, Eq85) and a single recombinant protein (IdeE), constructed from eight *See* antigen fragments with a total of 1,580 amino acids [62].

Analysis revealed that all isolates had coding sequences for the eight Strangvac® antigens, with identical nucleotide and deduced amino acid sequences for CNE, SclF, SclI, SclC, EAG, and IdeE, indicating that the six abovementioned antigen fragments are conserved in the *See* population. Minor differences were observed in Eq5- and Eq8-derived antigens. The Eq5- and Eq8-derived antigens were identical in 95.3 % and 5.2 % of the studied isolates, respectively.

For the Eq5-derived antigen, nine *See* isolates (4.7 %) exhibited a variation in the coding sequence at nucleotide 601 (A → C), resulting in a nonsynonymous amino acid substitution at position 201 from isoleucine to leucine. Interestingly, all of these nine isolates belonged to BAPS-5, ST-179, and *seM*-6 group. The same substitution was also identified in *See* strain 4047 [73], which also belonged to BAPS-5 and was used as a challenge strain of ponies vaccinated with Strangvac® [62,148]. The effectiveness of fusion proteins as vaccine antigens depends on the ability of the immune system to recognize all antigens without interference from neighbouring epitopes [148]. The findings of another study indicate that Strangvac® is likely to provide protection against *See* strain 4047 [148], even though it has a nonsynonymous substitution in Eq5. Therefore, Frosth *et al*. (2022) assumed that the substitution does not significantly alter the antigenicity of Eq5 [62].

Only 5.2 % of *See* isolates in this study possessed an identical Eq8-derived antigen when compared to Strangvac®. All BAPS-2 isolates (n = 181) encoded for a protein that carried a tyrosine instead of histidine at position 225 of the Eq8 protein. This substitution appears prevalent in *See* strains of BAPS-2, since the same substitution was also detected in 99.7 % of other BAPS-2 isolates reported by Frosth *et al*. (2022) [62]. Amino acid 225 represents the C-terminal amino acid of the Eq8-derived antigen fragment used in Strangvac® [62,148]. It was suspected that the substitution of terminal amino acid would not significantly affect the antigenicity of this protein relative to Strangvac® [62]. Therefore, it is assumed that immunization horses with Strangvac might also be protective despite the nonsynonymous amino acid substitution in Eq8.

Although most of the *See* genomes (92.7 %) in this study were recovered from horses in Germany, similar results were identified for *See* genomes recovered from horses in four other European countries and Indonesia. The *in silico* analysis supports that vaccination with Strangvac® might trigger protective immunity [150] against *See* circulating in Germany. Previous *in silico* genomic study using global *See* strains showed that the predicted amino acid sequences of Strangvac® antigens are highly conserved [62]. Strangvac® can be used to protect high-risk horses in areas where *See* was identified [50]. Horses can receive the vaccine intramuscularly as early as six months of age [50,150]. A prior study documented that protection increased from 31% to 58% within two weeks and over two months after the second vaccination [150]. Then, the protection elevated to 94% within three weeks after the third vaccination [150]. Strangvac® has DIVA capability since it uses different antigens targeted for *See* PCR for diagnostic properties [4,37,195] and does not contain live bacteria. It is crucial to continuously monitor the diversity of genes encoding antigens used in Strangvac® to gain insight into the evolution of *See* with respect to the adaptation of the immune response post-vaccination.

5.6 Frequency of phenotypic and genotypic antimicrobial resistance

The appropriate use of antibiotics in the treatment of horses with strangles remains a subject of ongoing scientific debate [174]. Many strangles cases resolve without the application of antimicrobial drugs [185], and the administration of penicillin during acute strangles may interfere with the persistence of humoral immunity to *See* [141]. Nonetheless, abscesses can develop slower or recur if antimicrobial treatment is discontinued [23]. Antimicrobial treatment can be indicated in certain strangles cases, such as acute infection with high fever and depression prior to the formation of abscesses, respiratory distress, profound lymphadenopathy, or metastatic abscessation and guttural pouch infections treated locally and systemically to eliminate the carrier state [23]. While antibiotics have been shown to reduce the severity of symptoms and the duration of the infection in some cases, there are concerns that their overuse can lead to antimicrobial resistance. Consequently, continuous monitoring of bacterial pathogens for resistance to antimicrobials is an essential component of antimicrobial stewardship.

In the present study, all *See* isolates were susceptible to beta-lactam antibiotics such as penicillins (penicillin G and ampicillin) and cephalosporins (ceftiofur). Previous studies have reported no penicillin resistance among *See* isolates [48,79,95]. Accordingly, our

study confirms that penicillin G can be considered the first-line antimicrobial for strangles in which antimicrobial chemotherapy is clinically indicated [23,25]. Following previous studies reporting minimal *Sez* resistance to ampicillin [103,112], all isolates in this study were susceptible to ampicillin. All isolates were susceptible to ceftiofur, with an MIC_{90} of 0.125 µg/ml, while the CLSI-approved susceptible breakpoint is at ≤ 0.25 µg/ml [34]. In a large outbreak of strangles in the USA, ceftiofur showed greater efficacy than penicillin G and doxycyclin concerning the duration of symptoms [31]. Nevertheless, ceftiofur, like other third-generation cephalosporins, has been classified as a critically important antimicrobial drug in human medicine [199]. In horses, it should be used with restraint and be reserved for cases where compliance with the treatment scheme is challenging [23]. Although no *See* isolates were resistant to ceftiofur in this study, the use of ceftiofur without proper consideration may result in increased resistance, as reported in the United Kingdom [79]. Furthermore, cephalosporin administration in horses might promote multidrug resistance in commensal faecal *E. coli* [44]; thus should not be to treat *See* infections.

The use of macrolides such as erythromycin should also be restricted and carefully considered [67,199], even though all *See* isolates in this study were susceptible to erythromycin. Tetracyclines, such as oxytetracycline and doxycycline, and trimethoprim/sulfonamide combinations are regarded as valuable antimicrobials to treat bacterial infectious diseases in horses [23,25]. In fact, all isolates in this study were susceptible to tetracycline, which is in agreement with a previous finding that resistance to this agent is rare in *See* [48,79].

The finding of intermediate susceptibility or resistance to enrofloxacin in all *See* isolates in this study is also in accordance with previous studies [32,48,79]. Resistance to enrofloxacin has been increased significantly for *See* [79] and *Sez* [79,103,131]. Furthermore, the French surveillance network for antimicrobial resistance in bacteria from diseased animals (RESAPATH) reported that the highest prevalence of streptococci resistance was observed for enrofloxacin [66]. Streptococci exhibits intrinsic low-level resistance to fluoroquinolones, which may explain that finding [66]. The resistance to fluoroquinolones among beta-hemolytic streptococci results from nucleotide substitutions in the quinolone resistance determinant regions (QRDR) of gyrase and topoisomerase IV genes, particularly *gyrA* and *parC* [13,94].

One of the most concerning findings of the present study was the substantial levels of resistance to trimethoprim/sulfamethoxazole, which is commonly used in equine practice, including to treat strangles [25]. An increased resistance level to trimethoprim/ sulfamethoxazole was also reported previously [9,32,48,79]. Pus inactivates trimethoprim/sulfamethoxazole, making this drug unsuitable for treating purulent bacterial infection [67]. The treatment of strangles with trimethoprim/sulfadiazine has been reported to be unsuccessful [23]. Intermediate susceptibility or resistance to trimethoprim/

sulfamethoxazole in more than half of *See* isolates in this study also argues against its use in horses with strangles.

The *dfrA1-like* resistance gene encoding resistance to trimethoprim (a folate pathway inhibitor) was detected in two isolates, *See* IHIT40242 and IHIT40253. *See* IHIT40242 was isolated from the guttural pouch lavage fluid of a horse in North Rhine-Westphalia, Gemany in 2019, while *See* IHIT40253 was isolated from an organ of a two-months-old foal in Mecklenburg-Western Pomerania in 2019. No epidemiological link was recorded between both isolates. Interestingly, both were phenotypically sensitive to trimethoprim/sulfamethoxazole. Among streptococci, the *drfA* gene was detected in *S. pyogenes* [16], and *S. agalactiae* [26]. A previous study showed that the *dfrA1* gene was detected in 24.2 % of 99 *Escherichia* spp. isolated in healthy food animals across Europe [51]. In *E. coli*, trimethoprim resistance genes were frequently located in integrons. The *dfrA1* gene was reported as an often identified class I integron gene cassette found in *E. coli* consisting of *dfrA1-aadA1-qacEdelta1-sul1* [51].

In this study, the interpretation of antimicrobial susceptibility for *See* against gentamicin and spectinomycin was hampered due to the lack of specific breakpoints in the CLSI [34]; therefore, the results should be interpreted cautiously. Streptococci have been reported to have low-level intrinsic resistance against aminoglycosides [94,112], including gentamycin and spectinomycin. Aminoglycoside should preferably not be used to treat *See* infections due to its low-level intrinsic resistance.

The suitability of the other antimicrobials in this study remains unclear since veterinary breakpoints for MIC interpretation, particularly for *Streptococcus* spp. from horses, are not available [34,156]. It should be noted that CLSI's interpretation categories and MIC have changed over time, which could result in a misclassification bias in the resistance trend and thus should be interpreted with caution. For instance, new enrofloxacin MIC breakpoints were published for *See* and *Sez* isolated from horses in 2018 [33].

In silico analysis revealed that all isolates had a point mutation of the *fusA* gene (GTA $\rightarrow$ ATT), which changed the amino acid from valine to isoleucine (V $\rightarrow$ I). This alteration may confer resistance to fusidic acid. It is noteworthy that streptococci inherently exhibit resistance to fusidic acid [94]. This study did not assess the phenotypic resistance against fusidic acid, as it was not included in the antimicrobial susceptibility test panel using microbroth dilution. Furthermore, fusidic acid is not a preferable antimicrobial for strangles treatment.

Surveillance programs for antimicrobial resistance of bacteria from animal origin exist globally. However, most programs focus on zoonotic bacteria only, while streptococci

of animal origin are seldom included in the national and international antimicrobial resistance surveillance programs [66]. So far, in surveillance systems, genotypic techniques to identify antibiotic resistance are only optionally implemented as a second-line characterization [66]. However, the decrease cost of next-generation sequencing methodologies may support that genotypic characterization of antimicrobial resistance may be implemented in monitoring programs for long-term surveillance purposes [21,66,205].

5.7 Biofilm formation by *Streptococcus equi* ssp. *equi* genotypes

The ability of *See* to form biofilm is poorly understood; therefore, this study aimed to better understand biofilm formation by *See* in different media across various phylogenetic groups *in vitro* by using a microtitre plate assay. The present study showed that *See* isolates were able to form biofilm *in vitro* on a polystyrene surface after incubation in BHI broth, THB, and THB supplemented with 1 % human fibrinogen. The results underscored the impact of the incubation medium on *See*'s biofilm formation, with THB proving as a more favourable medium for biofilm formation by *See* than BHI broth, and supplementation with fibrinogen enhanced this activity. These findings provide supporting evidence that biofilm formation is influenced by environmental parameters, including the availability of nutrients [20,54]. BHI broth and THB were recommended for cultivating fastidious pathogenic microorganisms, including streptococci [134,164]. Both media contain beef heart infusion, sodium chloride, and disodium phosphate [134,164]. Additionally, BHI broth contains calf brain infusion, while THB contains a peptic digest of animal tissue. The composition of THB seems to be more favourable for *See* to form biofilm. Another study about the biofilm formation of *See* used THB and tryptone soya broth (TSB) supplemented with 0.5 % glucose, in which *See* strain 4047 produced more biofilm in THB than in TSB [172]. Growth medium was also reported to influence the *in vitro* biofilm formation of other streptococci, such as *S. agalactiae* [20,121], *S. suis* [19,64], and *S. pyogenes* [54].

In general, the results demonstrated that fibrinogen enhanced biofilm formation by *See*, except for *See* with non-typeable *seM*. The present result is consistent with other publications where biofilm formation by *S. suis* increased after supplementation of human fibrinogen [19]. Studies on biofilm formation by *Sez* also used THB supplemented with 1 % fibrinogen to enhance biofilm formation [201,202]. Interestingly, biofilm formation by *See* isolates with non-typeable *seM* allele was significantly impaired, even in THB supplemented with fibrinogen. The SeM protein possesses a fibrinogen binding domain [84,119], suggesting that

this domain may play a crucial role in the biofilm formation mechanism. Insertion, deletion, and non-sense mutation of the *seM* gene, particularly at the fibrinogen binding domain, may inhibit bacterial attachment to fibrinogen, thereby potentially inhibiting the biofilm formation mechanism. Fibrinogen from humans, bovine, and porcine enhanced biofilm formation by *S. suis*, suggesting fibrinogen-mediated cross-bridging supports bacteria in attaching to each other, thus stimulating biofilm formation [19]. Proteomics analysis of *Staph. aureus* biofilm formation revealed that biofilm cells expressed higher levels of fibrinogen-binding proteins than planktonic cells [146]. Another study of biofilm formation by *Staph. aureus* showed that heparin enhanced biofilm formation, suggesting that this compound stimulates the formation of adhesion molecules that assist bacteria in adhering to each other through a heparin-mediated cross-bridge [162]. In *S. pneumonia*, hyaluronic acid was reported to support biofilm formation [200].

To the best of the author's knowledge, the role of the SeM in *See* biofilm formation has not been reported. A study of biofilm formation by *S. canis* reported that isolates harbouring SCM-10 allele were strong biofilm producers, while isolates harbouring SCM-1 allele were non-biofilm producers [63], suggesting that M-protein influences the biofilm formation of streptococci. M-like protein was also meant to influence the biofilm formation by *S. pyogenes* [54]. Other virulence factors, for example, bacteria pili were reported to be responsible for biofilm formation by *S. pyogenes* [113]. However, Steward *et al*. (2017) revealed that Fim1 locus encoding the pili in *See* was not required for biofilm formation by *See* strain 4047. Proteomics analysis of biofilm formation by *Sez* demonstrated that proteins upregulated during biofilm formation mainly were associated with adhesion, metabolism, and stress conditions [202].

This study also underscores significant differences in biofilm formation between STs of *See*, since on average ST-151 isolates formed a greater amount of biofilm mass than ST-179 isolates. The majority of isolates in this study belonged to ST-151 (73.8%), and this ST was significantly more frequently isolated in the last ten years than ST-179. It is tempting to speculate that the better ability of ST-151 isolates to form biofilm contributes to their fitness advantage over ST-179. This observation aligns with findings in a study on biofilm formation by *S. agalactiae,* where ST-19 isolates demonstrated a higher increase in biofilm formation than ST-17 isolates in the presence of human plasma [165]. Additionally, adherent-invasive *Escherichia coli* (AIEC) strains produced more biofilm than non-AIEC strains [114]. Future research should explore biofilm formation in a more extensive isolate collection from various phylogenetic groups.

Biofilm formation by bacteria is an important virulence factor that supports their ability to persist in the environment, evade the host immune system, increase their ability of

antimicrobial resistance, and cause persistent infections [17,54,58,68,165]. Therefore, developing a better understanding of the molecular mechanism of biofilms, which may function as a survival strategy for bacteria, could give new insights into the prevention and therapeutic measures of *See* infections.

.

Summary

Streptococcus equi subspecies *equi* (*See*) is the causative agent of strangles, a highly infectious disease of equines worldwide. This study utilized various molecular typing methods, including analysis of whole genome sequence (WGS) data, to enlighten the phylogenetic relationships of *See* isolates obtained from equines in Germany and other countries, as well as to examine the associations between genotype, virulence-associated genes (VAGs), and epidemiological data. Additionally, antimicrobial susceptibilities, biofilm formation *in vitro*, and the relationship between genotype and biofilm formation were investigated.

This study examined 628 non-duplicate putative *S. equi* isolates obtained from equines between 2001 and 2020. Most isolates originated in Germany (97.1 %). Matrix-assisted laser desorption/ionization time-of-flight mass spectrometry (MALDI-TOF MS) analysis confirmed the identity of all isolates as *See* (n = 265) and *Streptococcus equi* subspecies *zooepidemicus* (*Sez*) (n = 363), respectively. The published Se-mPCR successfully detected the *sodA* gene in all tested isolates, thus also confirming their assignment to the species *S. equi*. Additionally, among the 265 isolates identified as *See* by MALDI-TOF MS, 99.2 % also harboured the *See* signature locus ICE*Se*2, while none of them carried the ICE*Sz*1 locus of *Sez*. Surprisingly, only 44.6 % of the isolates identified as *Sez* by MALDI-TOF MS harboured the respective PCR target ICE*Sz*1. While Se-mPCR appears valuable for identifying and distinguishing *See* from *Sez*, it is necessary to enhance PCR test sensitivity for *Sez* by targeting a more stable and conserved gene of *Sez*.

Among the 265 non-duplicate *See* isolates, analysis of their *seM* gene revealed a total of 47 different alleles, including 30 novel ones. Pairwise differences between isolates ranged from one to 13 nucleotides in the 327 bp DNA section targeted for *seM* typing. Based on their phylogenetic relationship, these *seM* alleles were categorized into five groups (*seM*-1, -2, -6, -8, -9) with *seM*-9 group as the most prevalent one (87.9 %). The *seM* alleles of ten isolates could not be assigned due to deletions, insertions, or nonsense mutations in the *seM* target section for typing (non-typeable *seM*). Results indicated that various *See* strains caused strangles cases in Germany. The high number of novel *seM* alleles suggests continued diversification in the *seM* gene. Therefore, *seM* typing can be a valuable tool in tracing strangles outbreaks and *See* infection chains since it outperforms MLST in discrimination. However, the domination of *seM*-9 allele reduces its discriminatory power. It seems appropriate to recommend WGS data analysis to enhance understanding of the molecular epidemiology of *See*.

Multilocus sequence typing (MLST) of 191 *See* isolates selected for whole genome sequencing confirmed a low phylogenetic diversity of *See* since only two highly related sequence types (STs) were detected: ST-151 (73.8 %) and ST-179 (26.2 %). However, utilizing the core genome MLST (cgMLST) scheme on the global genome platform Pathogenwatch (https://pathogen.watch/) isolates exhibited 159 core genome (cg)-genotypes affiliated with three globally recognized BAPS clusters of *See*: BAPS-2 (94.8 %), BAPS-5 (4.7 %), and BAPS-6 (0.5 %). Within these clusters, isolates differed by zero to 110 pairwise core genome single nucleotide polymorphisms (cgSNPs). ST-151 isolates exclusively belonged to BAPS-2, while ST-179 isolates were distributed across BAPS-2 (80 %), BAPS-5 (18 %), and BAPS-6 (2 %). *See* isolates with *seM* alleles of groups *seM*-6, *seM*-8, and *seM*-9 were associated with BAPS-5, BAPS-6, and BAPS-2, respectively. Comparison of the cgMLST results with those publicly available for 759 other *See* isolates from around the world revealed that all genotypes of this study fit perfectly into the general phylogenetic tree of *See*, supporting the hypothesis of a common *See* ancestor strain and global expansion of its descendants since the late 19th or early 20th century.

Isolates of *See* obtained from horses in the same geographical area in Germany during a specified time period were considered as a spatial-temporal *See* cluster and were examined subsequently for their phylogenetic relationships. The WGS data of *See* isolates recovered from horses in Germany between 2002 and 2020 (n = 177) were included in this analysis. Twenty-three spatial-temporal clusters were identified, each comprising 2 to 20 *See* isolates (total n = 93). Nine of these clusters were exclusively composed of *See* isolates with identical STs, BAPS cluster assignments, and *seM* alleles. Six other clusters consisted solely of different *See* isolates, while the remaining eigth clusters included both identical and different *See* isolates. Additionally, 23 genotypes (identical cg-genotype, ST, and *seM* allele) were each observed in two or three *See* isolates. Notably, isolates with an identical genotype were found not only within the same spatial-temporal cluster but also in different clusters or outside from any cluster. These results indicate that *See* infections in the same spatial-temporal cluster must not necessarily all be epidemiologically related. Furthermore, the identification of genetically identical or closely related *See* strains at distant sites might be explained by the modern transportation of equids over long distances in a short time.

In silico screening of *See* genome data for known virulence-associated genes (VAGs) revealed 38 VAGs consistently present in all 191 tested isolates, with an additional 29 VAGs occurring only in 5 % to 99 % of the isolates. Differences in VAG combinations made it possible to distinguish 107 virotypes. Most virotypes (79.4 %) were found in only one isolate, while 14 virotypes in two isolates, and 8 virotypes in at least three isolates. The

differences between *See* strains in armament with VAGs may imply differences in their virulence and capacity to survive in the equine host.

The coding nucleotide and deduced amino acid sequences of six out of eight antigens used in the commercial Strangvac® vaccine were identical with those of the 191 investigated *See* isolates of this study. For antigens Eq5 and Eq8, only a single amino acid difference was detected. These results indicate that these antigens are highly-conserved in the investigated isolates and suggest that immunization with Strangvac® may also be protective against these strains.

All 265 *See* isolates in this study displayed susceptibility to the tested beta-lactam antimicrobials (penicillin G, ampicillin, and ceftiofur) and tetracycline, and almost all were susceptible to erythromycin. Over half of the isolates exhibited intermediate susceptibility or resistance to trimethoprim/sulfamethoxazole, and all isolates displayed intermediate susceptibility or resistance to enrofloxacin. According to these findings, penicillin G can be recommended as the first-line antimicrobial for strangles cases where antimicrobial therapy is indicated clinically.

In vitro biofilm formation by *See* isolates (n = 191) was examined by using a published microtitre plate assay. Results indicated that Tod Hewitt Broth (THB) is a more favourable medium for biofilm formation by *See* compared to Brain Hearth Infusion (BHI) broth, and supplementation with fibrinogen enhanced this activity. However, biofilm formation was significantly impaired in *See* isolates encoding for a non-typeable *seM* allele, even in THB supplemented with fibrinogen. This finding suggests that the fibrinogen binding domain of the SeM protein may play a role in the mechanism of biofilm assembly. The study also highlights significant differences between STs of *See*, with on average ST-151 isolates producing more biofilm mass than ST-179 isolates. Increased biofilm formation may contribute to a better fitness of certain *See* strains in the environment whether inside or outside the host organism.

In conclusion, this study sheds light on the genetic and phenotypic diversity of *See*, including antimicrobial susceptibility and virulence-associated factors, and provides valuable impetus for strangles surveillance, treatment, and prophylaxis.

Zusammenfassung

Streptococcus equi subspecies *equi* (*See*), ist der Krankheitserreger der Druse, einer weltweit auftretenden, hochansteckenden Infektionskrankheit der Einhufer. In der hier vorgelegten Arbeit wurden verschiedene molekulare Typisierungsmethoden einschließlich der *in silico*-Analyse von Vollgenomsequenzen (WGS) eingesetzt, um die phylogenetischen Beziehungen zwischen *See*-Isolaten aus Deutschland und anderen Ländern aufzudecken sowie die Zusammenhänge zwischen Genotyp, virulenzassoziierten Genen (VAGs) und epidemiologischen Daten zu prüfen. Darüber hinaus wurden die antimikrobielle Empfindlichkeit, die *in vitro*-Biofilmbildung und die Beziehung zwischen Genotyp und Grad der Biofilmbildung bestimmt.

In dieser Arbeit wurden 628 putative *S. equi*-Isolate untersucht, die zwischen 2001 und 2020 von Equiden gewonnen worden waren. Die meisten Isolate stammten aus Deutschland (97,1 %). Die Spezies- und Subspeziesidentifizierung erfolgte mittels *matrix-assisted laser desorption/ionization time-of-flight*-Massenspektrometrie (MALDI-TOF-MS). Mit ihr wurden 265 Isolate als *See* und 363 Isolate als *Streptococcus equi* subspecies *zooepidemicus* (*Sez*) identifiziert. Mit der von anderen Autoren publizierten Se-mPCR ließ sich das *sodA*-Gen in allen getesteten Isolaten erfolgreich nachweisen und somit deren Zugehörigkeit zur Spezies *S. equi* gleichfalls bestätigen. Darüber hinaus waren 99,2 % der 265 Isolate, die mit der MALDI-TOF-MS als *See* identifiziert wurden, auch Träger des Signatur-Locus ICE*Se*2 von *See*, während keines von ihnen den für *Sez* typischen ICE*Sz*1-Locus aufwies. Umgekehrt wurde aber nur bei 44,6 % der durch MALDI-TOF-MS als *Sez* identifizierten Isolate das entsprechende PCR-Target ICE*Sz*1 nachgewiesen. Obwohl die Se-mPCR für die Identifizierung von *See* und die Differenzierung gegenüber *Sez* hilfreich ist, erscheint es notwendig, die Sensitivität des PCR-Tests für *Sez* zu erhöhen, indem ein stabileres und stärker konserviertes Gen des *Sez*-Genoms als PCR-Target genutzt wird.

Bei der Analyse des *seM*-Gens wurden in den 265 *See*-Isolaten 47 verschiedene Allele gefunden, einschließlich 30 bisher unbekannte Allele. Im paarweisen Vergleich unterschieden sich die Isolate voneinander in einem bis 13 Nukleotiden in dem 327 bp langen DNA-Abschnitt, der die Grundlage für die s*eM*-Typisierung ist. Basierend auf ihrer phylogenetischen Verwandtschaft wurden die *seM*-Allele in fünf Gruppen eingeteilt (*seM*-1, -2, -6, -8, -9-Gruppe), wobei die *seM*-9-Gruppe am häufigsten vorkam (87,9 %). Die *seM*-Allele von zehn Isolaten konnten aufgrund von Deletionen, Insertionen oder *Nonsense*-Mutationen in dem oben genannten 327 bp-Abschnitt des *seM*-Gens keiner Gruppe zugeordnet werden (nicht typisierbares *seM*). Die Ergebnisse zeigen auf, dass die in Deutschland

nachgewiesenen Fälle von Druse vielfach durch unterschiedliche *See*-Stämme verursacht wurden. Die hohe Anzahl neuer *seM*-Allele lässt darauf schließen, dass eine fortlaufende Diversifizierung des *seM*-Gens stattfindet. Daher kann die *seM*-Typisierung ein wichtiges Instrument bei der Verfolgung von Druse-Ausbrüchen und der Erkennung von Infektketten sein, da sie das *multilocus sequence typing* (MLST) in der Diskriminationskraft übertrifft. Die Diskriminationskraft wird jedoch von der Dominanz des Allels *seM*-9 gemindert. Es erscheint deshalb angebracht, die Analyse von WGS-Daten zu empfehlen, um die molekulare Epidemiologie von *See* besser zu verstehen.

Die MLST-Analyse von 191 *See*-Isolaten, die zur Gesamtgenomsequenzierung ausgewählt wurden, bestätigte eine geringe phylogenetische Diversität von *See*, da nur zwei eng miteinander verwandte Sequenztypen (STs) gefunden wurden: ST-151 (73,8 %) und ST-179 (26,2 %). Wurde das *core genome* MLST (cgMLST)-Schema der globalen Genomplattform Pathogenwatch (https://pathogen.watch/) verwendet, waren dagegen 159 *core genome* (cg)-Genotypen zu unterscheiden. Diese ließen sich alle den drei weltweit anerkannten BAPS-Clusters BAPS-2 (94,8 %), BAPS-5 (4,7 %) und BAPS-6 (0,5 %) zuordnen. Isolate, die zu demselben BAPS-Cluster gehörten, unterschieden sich beim paarweisen Vergleich in null bis 110 *core genome single nucleotide polymorphisms* (cgSNPs) voneinander. ST-151-Isolate gehörten ausschließlich zu BAPS-2, während die ST-179-Isolate auf BAPS-2 (80 %), BAPS-5 (18 %) und BAPS-6 (2 %) verteilt waren. *See*-Isolate mit *seM*-Allelen der Gruppen *seM*-6, *seM*-8 und *seM*-9 waren mit BAPS-5 bzw. BAPS-6 bzw. BAPS-2 assoziiert. Der Vergleich der cgMLST-Ergebnisse mit den öffentlich zugänglichen Genomdaten von 759 anderen *See*-Isolaten aus der ganzen Welt ergab, dass die untersuchten *See*-Stämme der vorliegenden Studie perfekt in den aktuellen globalen phylogenetischen Stammbaum von *See* passten. Dies unterstützt die Hypothese, dass heutige Stämme von einem gemeinsamen *See*-Vorfahren abstammen, dessen Abkömmlinge sich seit Ende des 19. oder Anfang des 20. Jahrhunderts weltweit verbreiteten.

See-Isolate, die während eines bestimmten Zeitraums von Pferden im gleichen geografischen Gebiet in Deutschland isoliert worden waren, wurden als räumlich-zeitlicher *See*-Cluster betrachtet und auf ihre phylogenetische Verwandtschaft untersucht. In dieser Analyse wurden WGS-Daten von *See*-Isolaten (n = 177), die zwischen 2002 und 2020 von Pferden in Deutschland gewonnen worden waren, betrachtet. Insgesamt wurden 23 räumlich-zeitliche Clusters identifiziert, die jeweils aus 2 bis 20 *See*-Isolaten bestanden und zusammen 93 Isolate umfassten. Zehn Clusters bestanden ausschließlich aus *See*-Isolaten mit identischem ST, BAPS-Cluster und *seM*-Allel. In sechs weiteren Clusters waren jeweils alle *See*-Isolate verschieden, während in den übrigen sieben Clusters sowohl identische als

auch unterschiedliche *See*-Isolate vorkamen. Darüber hinaus wurden 23 Genotypen (identischer cg-Genotyp, ST und *seM*-Allel) identifiziert, die jeweils bei zwei oder drei *See*-Isolaten vorkamen. Bemerkenswert war, dass Isolate mit identischem Genotyp nicht nur innerhalb desselben räumlich-zeitlichen Clusters, sondern auch in verschiedenen Clusters oder außerhalb jeglicher Clusters isoliert worden waren. Diese Ergebnisse deuten darauf hin, dass *See*-Infektionen in demselben räumlich-zeitlichen Cluster nicht unbedingt alle epidemiologisch zusammenhängen müssen. Die Identifizierung genetisch identischer oder eng verwandter *See*-Stämme an weit entfernten Standorten kann wiederum durch den modernen Tiertransport erklärt werden, bei dem auch Einhufer in kurzer Zeit über große Entfernungen verbracht werden.

Die *in silico*-Suche nach bekannten virulenzassoziierten Genen (VAGs) in den *See*-Genomdaten ergab, dass 38 VAGs in allen 191 getesteten Isolaten vorhanden waren und weitere 29 VAGs jeweils nur in 5 % bis 99 % der Isolate. Anhand von Unterschieden zwischen den VAG-Kombinationen konnten 107 Virotypen differenziert werden. Die meisten Virotypen (79,4 %) wurden jeweils in nur einem Isolat gefunden, während 14 Virotypen in zwei Isolaten und 8 Virotypen in mindestens drei Isolaten festgestellt wurden. Die Unterschiede zwischen den *See*-Stämmen in Bezug auf ihre VAGs könnten auf eine unterschiedliche Virulenz und Überlebensfähigkeit im equinen Wirtsorganismus hindeuten.

Die Nukleotidsequenzen und die davon abgeleiteten Aminosäuresequenzen von sechs der acht im kommerziellen Strangvac®-Impfstoff verwendeten Antigene waren mit denen der 191 untersuchten *See*-Isolaten identisch. Bei den Antigenen Eq5 und Eq8 wurde jeweils nur ein einziger Aminosäureunterschied festgestellt. Diese Ergebnisse deuten darauf hin, dass die betreffenden Antigene in den untersuchten Isolaten hochgradig konserviert sind, sodass die Impfung mit Strangvac® wahrscheinlich auch vor diesen Stämmen schützt.

Alle 265 untersuchten *See*-Isolate waren empfindlich gegenüber den getesteten Beta-Lactam-Antibiotika (Penicillin G, Ampicillin und Ceftiofur) und Tetracyclin, und fast alle waren auch Erythromycin-sensibel. Dagegen wurde mehr als die Hälfte der Isolate als intermediär oder resistent gegen Trimethoprim/Sulfamethoxazol eingestuft, und alle Isolate waren entweder intermediär oder resistent gegen Enrofloxacin. Aufgrund dieser Ergebnisse kann Penicillin G als Antibiotikum der ersten Wahl (*first line*-Antibiotikum) bei Druse empfohlen werden, falls aus klinischer Sicht eine antimikrobielle Therapie indiziert ist.

Die *in vitro*-Biofilmbildung von *See*-Isolaten (n = 191) wurde mit Hilfe eines veröffentlichten Mikrotiterplatten-Assays untersucht. Die Ergebnisse belegten, dass die *Todd-Hewitt-Broth* (THB) im Vergleich zu der *Brain-Hearth-Infusion* (BHI)-Bouillon für die

Biofilmbildung von *See* besser geeignet ist, und dass die Supplementierung mit Fibrinogen die Bildung von Biofilm fördert. Allerdings war die Biofilmbildung bei *See*-Isolaten, die für ein nicht typisierbares *seM*-Allel kodierten, signifikant beeinträchtigt, auch in mit Fibrinogen supplementiertem THB. Dieser Befund deutet darauf hin, dass die bekannte Fibrinogen-Bindungsdomäne im SeM-Protein eine Rolle bei der Assemblierung des *See*-Biofilms spielen könnte. Auch konnten signifikante Unterschiede zwischen den Sequenztypen nachgewiesen werden, wobei ST-151-Isolate im Durchschnitt mehr Biofilmmasse bildeten als ST-179-Isolate. Die stärkere Biofilmbildung trägt möglicherweise zu einer besseren Fitness bestimmter *See*-Stämme in der Umwelt bei, sei es innerhalb oder außerhalb des Wirtsorganismus.

Zusammenfassend bietet die vorgelegte Arbeit zahlreiche, neue Erkenntnisse über die genetische und phänotypische Vielfalt des Krankheitserregers *See*, einschließlich seiner antimikrobiellen Empfindlichkeit und seiner virulenzassoziierten Faktoren. Sie liefert damit wertvolle Impulse für die Überwachung, Behandlung und Prophylaxe von Druse.

References

1. Akineden O, Alber J, Lämmler C, Weiss R, Siebert U, Foster G, Tougaard S, Brasseur SMJM, Reijnders PJH. Relatedness of *Streptococcus equi* subsp. *zooepidemicus* strains isolated from harbour seals (*Phoca vitulina*) and grey seals (*Halichoerus grypus*) of various origins of the North Sea during 1988-2005. *Veterinary microbiology* 2007;121(1-2):158–62. https://doi.org/10.1016/j.vetmic.2006.11.015.

2. Alatoom AA, Cunningham SA, Ihde SM, Mandrekar J, Patel R. Comparison of direct colony method versus extraction method for identification of gram-positive cocci by use of Bruker Biotyper matrix-assisted laser desorption ionization-time of flight mass spectrometry. *Journal of clinical microbiology* 2011;49(8):2868–73. https://doi.org/10.1128/JCM.00506-11.

3. Alber J, El-Sayed A, Estoepangestie S, Lämmler C, Zschöck M. Dissemination of the superantigen encoding genes *seeL, seeM, szeL* and *szeM* in *Streptococcus equi* subsp. *equi* and *Streptococcus equi* subsp. *zooepidemicus*. *Veterinary microbiology* 2005;109(1-2):135–41. https://doi.org/10.1016/j.vetmic.2005.05.001.

4. Alber J, El-Sayed A, Lämmler C, Hassan AA, Weiss R, Zschöck M. Multiplex polymerase chain reaction for identification and differentiation of *Streptococcus equi* subsp. *zooepidemicus* and *Streptococcus equi* subsp. *equi*. *Journal of veterinary medicine. B, Infectious diseases and veterinary public health* 2004;51(10):455–8. https://doi.org/10.1111/j.1439-0450.2004.00799.x.

5. Anzai T, Kuwamoto Y, Wada R, Sugita S, Kakuda T, Takai S, Higuchi T, Timoney JF. 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. *American journal of veterinary research* 2005;66(12):2167–71. https://doi.org/10.2460/ajvr.2005.66.2167.

6. Anzai T, Timoney JF, Kuwamoto Y, Fujita Y, Wada R, Inoue T. In vivo pathogenicity and resistance to phagocytosis of *Streptococcus equi* strains with different levels of capsule expression. *Veterinary microbiology* 1999;67(4):277–86. https://doi.org/10.1016/s0378-1135(99)00051-6.

7. Argimón S, Abudahab K, Goater RJE, Fedosejev A, Bhai J, Glasner C, Feil EJ, Holden MTG, Yeats CA, Grundmann H, Spratt BG, Aanensen DM. Microreact: visualizing and sharing data for genomic epidemiology and phylogeography. *Microbial genomics* 2016;2(11):e000093. https://doi.org/10.1099/mgen.0.000093.

8. Artiushin SC, Timoney JF, Sheoran AS, Muthupalani SK. Characterization and immunogenicity of pyrogenic mitogens SePE-H and SePE-I of *Streptococcus equi*. *Microbial pathogenesis* 2002;32(2):71–85. https://doi.org/10.1006/mpat.2001.0482.

9. Awosile BB, Heider LC, Saab ME, McClure JT. Antimicrobial resistance in bacteria isolated from horses from the Atlantic Provinces, Canada (1994 to 2013). *The Canadian veterinary journal* 2018;59(9):951–7.

10. Azeredo J, Azevedo NF, Briandet R, Cerca N, Coenye T, Costa AR, Desvaux M, Di Bonaventura G, Hébraud M, Jaglic Z, Kačániová M, Knøchel S, Lourenço A, Mergulhão F, Meyer RL, Nychas G, Simões M, Tresse O, Sternberg C. Critical review on biofilm methods. *Critical reviews in microbiology* 2017;43(3):313–51. https://doi.org/10.1080/1040841X.2016.1208146.

11. Baker JR, Pritchard DG. Action pattern and substrate specificity of the hyaluronan lyase from group B streptococci. *Biochemical journal* 2000;348 Pt 2(Pt 2):465–71.

12. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV, Sirotkin AV, Vyahhi N, Tesler G, Alekseyev MA, Pevzner PA. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* 2012;19(5):455–77. https://doi.org/10.1089/cmb.2012.0021.

13. Barros RR. Antimicrobial Resistance among Beta-Hemolytic *Streptococcus* in Brazil: An Overview. *Antibiotics* 2021;10(8). https://doi.org/10.3390/antibiotics10080973.

14. Båverud V, Johansson SK, Aspan A. Real-time PCR for detection and differentiation of *Streptococcus equi* subsp. *equi* and *Streptococcus equi* subsp. *zooepidemicus*. *Veterinary microbiology* 2007;124(3-4):219–29. https://doi.org/10.1016/j.vetmic.2007.04.020.

15. Been M de, Pinholt M, Top J, Bletz S, Mellmann A, van Schaik W, Brouwer E, Rogers M, Kraat Y, Bonten M, Corander J, Westh H, Harmsen D, Willems RJL. Core genome multilocus sequence typing dcheme for high- Resolution typing of *Enterococcus faecium*. *Journal of clinical microbiology* 2015;53(12):3788–97. https://doi.org/10.1128/JCM.01946-15.

16. Bergmann R, van der Linden M, Chhatwal GS, Nitsche-Schmitz DP. Factors that cause trimethoprim resistance in *Streptococcus pyogenes*. *Antimicrobial agents and chemotherapy* 2014;58(4):2281–8. https://doi.org/10.1128/AAC.02282-13.

17. Bjarnsholt T. The role of bacterial biofilms in chronic infections. *APMIS. Supplementum* 2013(136):1–51. https://doi.org/10.1111/apm.12099.

18. Björnsdóttir S, Harris SR, Svansson V, Gunnarsson E, Sigurðardóttir ÓG, Gammeljord K, Steward KF, Newton JR, Robinson C, Charbonneau ARL, Parkhill J, Holden MTG, Waller AS. Genomic dissection of an Icelandic epidemic of respiratory disease in horses and associated zoonotic cases. *mBio* 2017;8(4). https://doi.org/10.1128/mBio.00826-17.

19. Bonifait L, Grignon L, Grenier D. Fibrinogen induces biofilm formation by *Streptococcus suis* and enhances its antibiotic resistance. *Applied and environmental microbiology* 2008;74(15):4969–72. https://doi.org/10.1128/AEM.00558-08.

20. Bonsaglia ECR, Latosinski GS, Rossi RS, Rossi BF, Possebon FS, Pantoja JCF, Fernandes Júnior A, Rall VLM. Biofilm production under different atmospheres and growth media by *Streptococcus agalactiae* isolated from milk of cows with subclinical mastitis. *Archives of microbiology* 2020;202(1):209–12. https://doi.org/10.1007/s00203-019-01727-8.

21. Bortolaia V, Kaas RS, Ruppe E, Roberts MC, Schwarz S, Cattoir V, Philippon A, Allesoe RL, Rebelo AR, Florensa AF, Fagelhauer L, Chakraborty T, Neumann B, Werner G, Bender JK, Stingl K, Nguyen M, Coppens J, Xavier BB, Malhotra-Kumar S, Westh H, Pinholt M, Anjum MF, Duggett NA, Kempf I, Nykäsenoja S, Olkkola S, Wieczorek K, Amaro A, Clemente L, Mossong J, Losch S, Ragimbeau C, Lund O, Aarestrup FM. ResFinder 4.0 for predictions of phenotypes from genotypes. *The Journal of antimicrobial chemotherapy* 2020;75(12):3491–500. https://doi.org/10.1093/jac/dkaa345.

22. Boschwitz JS, Timoney JF. Characterization of the antiphagocytic activity of equine fibrinogen for *Streptococcus equi* subsp. *equi*. *Microbial pathogenesis* 1994;17(2):121–9. https://doi.org/10.1006/mpat.1994.1058.

23. 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. *Journal of veterinary internal medicine* 2018;32(2):633–47. https://doi.org/10.1111/jvim.15043.

24. Brankston G, Rossi TM, O'Sullivan TL, Greer AL. Diagnostic testing patterns for *Streptococcus equi* subsp. *equi* in Ontario horses during the years 2008 to 2018. *The Canadian veterinary journal* 2021;62(6):629–36.

25. British Equine Veterinary Association (BEVA). Protect ME Practice policy: Dose and routes of administration of common antimicrobials 2018.

26. Brochet M, Couvé E, Zouine M, Poyart C, Glaser P. A naturally occurring gene amplification leading to sulfonamide and trimethoprim resistance in *Streptococcus agalactiae*. *Journal of Bacteriology* 2008;190(2):672–80. https://doi.org/10.1128/JB.01357-07.

27. Bustos CP, Muñoz AJ, Guida N, Waller A, Mesplet M. Novel *seM*-types of *Streptococcus equi* subsp. *equi* identified in isolates circulating in Argentina. *Equine veterinary journal* 2022;54(1):132–8. https://doi.org/10.1111/evj.13420.

28. Canchaya C, Desiere F, McShan WM, Ferretti JJ, Parkhill J, Brüssow H. Genome analysis of an inducible prophage and prophage remnants integrated in the *Streptococcus pyogenes* strain SF370. *Virology* 2002;302(2):245–58. https://doi.org/10.1006/viro.2002.1570.

29. Chanter N, Collin N, Holmes N, Binns M, Mumford J. Characterization of the Lancefield group C *streptococcus* 16S-23S RNA gene intergenic spacer and its potential for identification and sub-specific typing. *Epidemiology and infection* 1997;118(2):125–35. https://doi.org/10.1017/s0950268896007285.

30. Chanter N, Talbot N, Newton JR, Hewson D, Verheyen K. *Streptococcus equi* with truncated M-proteins isolated from outwardly healthy horses. *Microbiology* 2000;146:1361–9.

31. Christmann U, Pink C. Lessons learned from a strangles outbreak on a large standardbred farm. *Equine veterinary journal* 2017;29(3):138–43. https://doi.org/10.1111/eve.12451.

32. Clark C, Greenwood S, Boison JO, Chirino-Trejo M, Dowling PM. Bacterial isolates from equine infections in western Canada (1998-2003). *The Canadian veterinary journal* 2008;49(2):153–60.

33. Clinical and Laboratory Standards Institute (CLSI). *Performance standards for antimicrobial disk and dilution susceptibility tests for bacteria isolated from animals, 4th edition: CLSI*; 2018.

34. Clinical and Laboratory Standards Institute (CLSI). *Performance standards for antimicrobial disk and dilution susceptibility test for bacteria isolated from animals, 5th edition: CLSI VET01S ED5*; 2020.

35. Coccitto SN, Massacci FR, Cinthi M, Albini E, Cucco L, D'Achille G, Morroni G, Mingoia M, Orsini M, Giovanetti E, Brenciani A, Magistrali CF. Co-location of the oxazolidinone resistance *poxtA2* and *cfr(D)* genes on a multiresistance plasmid from a porcine *Streptococcus dysgalactiae* subsp. *equisimilis*, Italy. *The journal of antimicrobial chemotherapy* 2023. https://doi.org/10.1093/jac/dkad169.

36. Corander J, Marttinen P, Sirén J, Tang J. Enhanced Bayesian modelling in BAPS software for learning genetic structures of populations. *BMC bioinformatics* 2008;9:539. https://doi.org/10.1186/1471-2105-9-539.

37. Cordoni G, Williams A, Durham A, Florio D, Zanoni RG, La Ragione RM. Rapid diagnosis of strangles (*Streptococcus equi* subspecies *equi*) using PCR. *Research in veterinary science* 2015;102:162–6. https://doi.org/10.1016/j.rvsc.2015.08.008.

38. *Cowan and Steel's Manual for the Identification of Medical Bacteria 3rd Edition*. Cambridge, United Kingdom: Cambridge University Press; 2004.

39. Croucher NJ, Harris SR, Fraser C, Quail MA, Burton J, van der Linden M, McGee L, Gottberg A von, Song JH, Ko KS, Pichon B, Baker S, Parry CM, Lambertsen LM, Shahinas D, Pillai DR, Mitchell TJ, Dougan G, Tomasz A, Klugman KP, Parkhill J, Hanage WP, Bentley SD. Rapid pneumococcal evolution in response to clinical interventions. *Science* 2011;331(6016):430–4. https://doi.org/10.1126/science.1198545.

40. Cursons R, Patty O, Steward KF, Waller AS. Strangles in horses can be caused by vaccination with Pinnacle I. N. *Vaccine* 2015;33(30):3440–3. https://doi.org/10.1016/j.vaccine.2015.05.009.

41. Davies MR, Holden MT, Coupland P, Chen JHK, Venturini C, Barnett TC, Zakour NLB, Tse H, Dougan G, Yuen K-Y, Walker MJ. Emergence of scarlet fever *Streptococcus pyogenes* emm12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance. *Nature genetics* 2015;47(1):84–7. https://doi.org/10.1038/ng.3147.

42. De Vos P, Garrity GM, Jones D, Krieg NR, Ludwig W, Rainey FA, Schleifer K-H, Whitman WB. *Bergey's manual of systematic bacteriology second edition: volume three the Firmicutes.* 3rd ed. New York: Springer; 2009.

43. Dong J, Gao N, Waller AS, Cook FR, Fan S, Yuan D, Du Y, Li F, Norimine J, Zhu W. An outbreak of strangles associated with a novel genotype of *Streptococcus equi* subspecies *equi* in donkeys in China during 2018. *Equine veterinary journal* 2019;51(6):743–8. https://doi.org/10.1111/evj.13114.

44. Dunowska M, Morley PS, Traub-Dargatz JL, Hyatt DR, Dargatz DA. Impact of hospitalization and antimicrobial drug administration on antimicrobial susceptibility patterns of commensal *Escherichia coli* isolated from the feces of horses. *Journal of the American Veterinary Medical Association* 2006;228(12):1909–17. https://doi.org/10.2460/javma.228.12.1909.

45. Durham AE, Hall YS, Kulp L, Underwood C. A study of the environmental survival of *Streptococcus equi* subspecies *equi*. *Equine veterinary journal* 2018;50(6):861–4. https://doi.org/10.1111/evj.12840.

46. El-Hage CM, Bannai H, Wiethoelter AK, Firestone SM, Heislers CM, Allen JL, Waller AS, Gilkerson JR. Serological responses of Australian horses using a commercial duplex indirect ELISA following vaccination against strangles. *Australian veterinary journal* 2019;97(7):220–4. https://doi.org/10.1111/avj.12825.

47. Elsayed S, Hammerberg O, Massey V, Hussain Z. *Streptococcus equi* subspecies *equi* (Lancefield group C) meningitis in a child. *Clinical microbiology and infection* 2003;9(8):869–72. https://doi.org/10.1046/j.1469-0691.2003.00663.x.

48. Erol E, Locke SJ, Donahoe JK, Mackin MA, Carter CN. Beta-hemolytic *Streptococcus* spp. from horses: a retrospective study (2000-2010). *Journal of veterinary diagnostic investigation* 2012;24(1):142–7. https://doi.org/10.1177/1040638711434138.

49. European Medicines Agency (EMA). *Equilis StrepE (live deletion mutant Streptococcus equi strain TW928)*. Available at: https://www.ema.europa.eu/en/documents/overview/equilis-strepe-epar-medicine-overview_de.pdf; 2020 [accessed February 14, 2023].

50. European Medicines Agency (EMA). CVMP assessment report for Strangvac (EMEA/V/C/005309/0000) Vaccine common name: *Streptococcus equi* vaccine (recombinant proteins) 2021;2021:1–39.

51. Ewers C, Jong A de, Prenger-Berninghoff E, El Garch F, Leidner U, Tiwari SK, Semmler T. Genomic diversity and virulence potential of ESBL- and AmpC-β-Lactamase-Producing *Escherichia coli* strains from healthy food animals across

Europe. *Frontiers in microbiology* 2021;12:626774. https://doi.org/10.3389/fmicb.2021.626774.

52. Facklam R. What happened to the streptococci: overview of taxonomic and nomenclature changes. *Clinical microbiology reviews* 2002;15(4):613–30. https://doi.org/10.1128/CMR.15.4.613-630.2002.

53. Fernández E, Blume V, Garrido P, Collins MD, Mateos A, Domínguez L, Fernández-Garayzábal JF. *Streptococcus equi* subsp. *ruminatorum* subsp. nov., isolated from mastitis in small ruminants. *International journal of systematic and evolutionary microbiology* 2004;54(Pt 6):2291–6. https://doi.org/10.1099/ijs.0.63145-0.

54. Fiedler T, Köller T, Kreikemeyer B. *Streptococcus pyogenes* biofilms-formation, biology, and clinical relevance. *Frontiers in cellular and infection microbiology* 2015;5:15. https://doi.org/10.3389/fcimb.2015.00015.

55. Fischetti VA. Streptococcal M protein: molecular design and biological behavior. *Clinical microbiology reviews* 1989;2(3):285–314. https://doi.org/10.1128/CMR.2.3.285.

56. Fischetti VA. Surface Proteins on Gram-Positive Bacteria. *Microbiology spectrum* 2019;7(4). https://doi.org/10.1128/microbiolspec.GPP3-0012-2018.

57. Fischetti VA, Jarymowycz M, Jones KF, Scott JR. Streptococcal M protein size mutants occur at high frequency within a single strain. *The journal of experimental medicine* 1986;164(4):971–80. https://doi.org/10.1084/jem.164.4.971.

58. Flemming H-C, Wingender J. The biofilm matrix. *Nature reviews microbiology* 2010;8(9):623–33. https://doi.org/10.1038/nrmicro2415.

59. Flock M, Frykberg L, Sköld M, Guss B, Flock J-I. Antiphagocytic function of an IgG glycosyl hydrolase from *Streptococcus equi* subsp. *equi* and its use as a vaccine component. *Infection and immunity* 2012;80(8):2914–9. https://doi.org/10.1128/IAI.06083-11.

60. Flock M, Jacobsson K, Frykberg L, Hirst TR, Franklin A, Guss B, Flock J-I. Recombinant *Streptococcus equi* proteins protect mice in challenge experiments and induce immune response in horses. *Infection and immunity* 2004;72(6):3228–36. https://doi.org/10.1128/IAI.72.6.3228-3236.2004.

61. Flock M, Karlström A, Lannergård J, Guss B, Flock J-I. Protective effect of vaccination with recombinant proteins from *Streptococcus equi* subspecies *equi* in a strangles model in the mouse. *Vaccine* 2006;24(19):4144–51. https://doi.org/10.1016/j.vaccine.2006.02.016.

62. Frosth S, Morris ERA, Wilson H, Frykberg L, Jacobsson K, Parkhill J, Flock J-I, Wood T, Guss B, Aanensen DM, Boyle AG, Riihimäki M, Cohen ND, Waller AS. Conservation of vaccine antigen sequences encoded by sequenced strains of *Streptococcus equi* subsp. *equi*. *Equine veterinary journal* 2022. https://doi.org/10.1111/evj.13552.

63. Fukushima Y, Tsuyuki Y, Goto M, Yoshida H, Takahashi T. Biofilm production ability and other microbiological features of *Streptococcus canis*. *Japanese journal of infectious diseases* 2022;75(1):63–9. https://doi.org/10.7883/yoken.JJID.2020.1086.

64. Grenier D, Grignon L, Gottschalk M. Characterisation of biofilm formation by a *Streptococcus suis* meningitis isolate. *Veterinary journal* 2009;179(2):292–5. https://doi.org/10.1016/j.tvjl.2007.09.005.

65. Guss B, Flock M, Frykberg L, Waller AS, Robinson C, Smith KC, Flock J-I. Getting to grips with strangles: an effective multi-component recombinant vaccine for the protection of horses from *Streptococcus equi* infection. *PLoS pathogens* 2009;5(9):e1000584. https://doi.org/10.1371/journal.ppat.1000584.

66. Haenni M, Lupo A, Madec J-Y. Antimicrobial resistance in *Streptococcus* spp. *Microbiology spectrum* 2018;6(2). https://doi.org/10.1128/microbiolspec.ARBA-0008-2017.

67. Haggett EF, Wilson WD. Overview of the use of antimicrobials for the treatment of bacterial infections in horses. *Equine veterinary education* 2008;20(8):433–48. https://doi.org/10.2746/095777308X338893.

68. Hall CW, Mah T-F. Molecular mechanisms of biofilm-based antibiotic resistance and tolerance in pathogenic bacteria. *FEMS microbiology reviews* 2017;41(3):276–301. https://doi.org/10.1093/femsre/fux010.

69. Hamilton A, Robinson C, Sutcliffe IC, Slater J, Maskell DJ, Davis-Poynter N, Smith K, Waller A, Harrington DJ. Mutation of the maturase lipoprotein attenuates the virulence of *Streptococcus equi* to a greater extent than does loss of general lipoprotein lipidation. *Infection and immunity* 2006;74(12):6907–19. https://doi.org/10.1128/IAI.01116-06.

70. Harbaugh MP, Podbielski A, Hügl S, Cleary PP. Nucleotide substitutions and small-scale insertion produce size and antigenic variation in group A streptococcal M1 protein. *Molecular microbiology* 1993;8(5):981–91. https://doi.org/10.1111/j.1365-2958.1993.tb01642.x.

71. Harris SR, Robinson C, Steward KF, Webb KS, Paillot R, Parkhill J, Holden MTG, Waller AS. Genome specialization and decay of the strangles pathogen, *Streptococcus equi*, is driven by persistent infection. *Genome research* 2015;25(9):1360–71. https://doi.org/10.1101/gr.189803.115.

72. Heather Z, Holden MTG, Steward KF, Parkhill J, Song L, Challis GL, Robinson C, Davis-Poynter N, Waller AS. A novel streptococcal integrative conjugative element involved in iron acquisition. *Molecular microbiology* 2008;70(5):1274–92. https://doi.org/10.1111/j.1365-2958.2008.06481.x.

73. Holden MTG, Heather Z, Paillot R, Steward KF, Webb K, Ainslie F, Jourdan T, Bason NC, Holroyd NE, Mungall K, Quail MA, Sanders M, Simmonds M, Willey D, Brooks K, Aanensen DM, Spratt BG, Jolley KA, Maiden MCJ, Kehoe M, Chanter N, Bentley SD, Robinson C, Maskell DJ, Parkhill J, Waller AS. Genomic evidence for the evolution of *Streptococcus equi*: host restriction, increased virulence, and genetic exchange with human pathogens. *PLoS pathogens* 2009;5(3):e1000346. https://doi.org/10.1371/journal.ppat.1000346.

74. Hulting G, Flock M, Frykberg L, Lannergård J, Flock J-I, Guss B. Two novel IgG endopeptidases of *Streptococcus equi*. *FEMS microbiology letters* 2009;298(1):44–50. https://doi.org/10.1111/j.1574-6968.2009.01698.x.

75. Hunter SB, Vauterin P, Lambert-Fair MA, van Duyne MS, Kubota K, Graves L, Wrigley D, Barrett T, Ribot E. Establishment of a universal size standard strain for use with the PulseNet standardized pulsed-field gel electrophoresis protocols: converting the national databases to the new size standard. *Journal of clinical microbiology* 2005;43(3):1045–50. https://doi.org/10.1128/JCM.43.3.1045-1050.2005.

76. Ijaz M, Velineni S, Timoney JF. Selective pressure for allelic diversity in SeM of *Streptococcus equi* does not affect immunoreactive proteins SzPSe or Se18.9. *Infection, genetics and evolution* 2011;11(5):1159–63. https://doi.org/10.1016/j.meegid.2011.01.011.

77. Ivens PAS, Matthews D, Webb K, Newton JR, Steward K, Waller AS, Robinson C, Slater JD. Molecular characterisation of 'strangles' outbreaks in the UK: the use of M-protein typing of *Streptococcus equi* ssp. *equi*. *Equine veterinary journal* 2011;43(3):359–64. https://doi.org/10.1111/j.2042-3306.2010.00177.x.

78. 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*. *The veterinary record* 2000;147(20):563–7. https://doi.org/10.1136/vr.147.20.563.

79. Johns IC, Adams E-L. Trends in antimicrobial resistance in equine bacterial isolates: 1999-2012. *The Veterinary record* 2015;176(13):334. https://doi.org/10.1136/vr.102708.

80. Jolley KA, Bray JE, Maiden MCJ. Open-access bacterial population genomics: BIGSdb software, the PubMLST.org website and their applications. *Wellcome open research* 2018;3:124. https://doi.org/10.12688/wellcomeopenres.14826.1.

81. Jorm LR, Love DN, Bailey GD, McKay GM, Briscoe DA. Genetic structure of populations of beta-haemolytic Lancefield group C streptococci from horses and their association with disease. *Research in veterinary science* 1994;57(3):292–9. https://doi.org/10.1016/0034-5288(94)90120-1.

82. Kasuya K, Tanaka N, Oshima F, Fujisawa N, Saito M, Tagami K, Niwa H, Sasai K. Genetic analysis of *Streptococcus equi* subsp. *equi* isolated from horses imported into Japan. *The journal of veterinary medical science* 2019;81(6):924–7. https://doi.org/10.1292/jvms.18-0656.

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

84. Kelly C, Bugg M, Robinson C, Mitchell Z, Davis-Poynter N, Newton JR, Jolley KA, Maiden MCJ, Waller AS. Sequence variation of the SeM gene of *Streptococcus equi* allows discrimination of the source of strangles outbreaks. *Journal of clinical microbiology* 2006;44(2):480–6. https://doi.org/10.1128/JCM.44.2.480-486.2006.

85. Kemp-Symonds J, Kemble T, Waller A. Modified live *Streptococcus equi* ('strangles') vaccination followed by clinically adverse reactions associated with bacterial replication. *Equine veterinary journal* 2007;39(3):284–6. https://doi.org/10.2746/042516407x195961.

86. Kerstens J, Durmus B, Lambrecht S, Baar I, Ieven MM, van der Zijden T, Parizel PM, Menovsky T, Lammens MMY, Jorens PG. Meningoencephalitis with *Streptococcus equi s*ubspecies *equi* leading to a dural arteriovenous fistula. *Case reports in neurological medicine* 2021;2021:1–6. https://doi.org/10.1155/2021/9898364.

87. Kimura B. Will the emergence of core genome MLST end the role of in silico MLST? *Food microbiology* 2018;75:28–36. https://doi.org/10.1016/j.fm.2017.09.003.

88. Kudirkiene E, Welker M, Knudsen NR, Bojesen AM. Rapid and accurate identification of *Streptococcus equi* subspecies by MALDI-TOF MS. *Systematic and applied microbiology* 2015;38(5):315–22. https://doi.org/10.1016/j.syapm.2015.02.010.

89. Kuusi M, Lahti E, Virolainen A, Hatakka M, Vuento R, Rantala L, Vuopio-Varkila J, Seuna E, Karppelin M, Hakkinen M, Takkinen J, Gindonis V, Siponen K, Huotari K. An outbreak of *Streptococcus equi* subspecies *zooepidemicus* associated with consumption of fresh goat cheese. *BMC infectious diseases* 2006;6:36. https://doi.org/10.1186/1471-2334-6-36.

90. Lancefield RC. A serological differentiation of human and other groups of hemolytic streptococci. *The journal of experimental medicine* 1933;57(4):571–95. https://doi.org/10.1084/jem.57.4.571.

91. Lanka S, Borst LB, Patterson SK, Maddox CW. A multiphasic typing approach to subtype *Streptococcus equi s*ubspecies *equi*. *Journal of veterinary diagnostic investigation* 2010;22(6):928–36. https://doi.org/10.1177/104063871002200612.

92. Lannergård J, Flock M, Johansson S, Flock J-I, Guss B. Studies of fibronectin-binding proteins of *Streptococcus equi*. *Infection and immunity* 2005;73(11):7243–51. https://doi.org/10.1128/IAI.73.11.7243-7251.2005.

93. Lannergård J, Guss B. IdeE, an IgG-endopeptidase of *Streptococcus equi* ssp. *equi*. *FEMS microbiology letters* 2006;262(2):230–5. https://doi.org/10.1111/j.1574-6968.2006.00404.x.

94. Leclercq R, Cantón R, Brown DFJ, Giske CG, Heisig P, MacGowan AP, Mouton JW, Nordmann P, Rodloff AC, Rossolini GM, Soussy C-J, Steinbakk M, Winstanley TG, Kahlmeter G. EUCAST expert rules in antimicrobial susceptibility testing. *Clinical microbiology and infection* 2013;19(2):141–60. https://doi.org/10.1111/j.1469-0691.2011.03703.x.

95. Léon A, Castagnet S, Maillard K, Paillot R, Giard J-C. Evolution of in vitro antimicrobial susceptibility of equine clinical isolates in France between 2016 and 2019. *Animals* 2020;10(5). https://doi.org/10.3390/ani10050812.

96. Libardoni F, Vielmo A, Farias L, Matter LB, Pötter L, Spilki FR, Vargas AC de. Diversity of seM in *Streptococcus equi* subsp. *equi* isolated from strangles outbreaks. *Veterinary microbiology* 2013;162(2-4):663–9. https://doi.org/10.1016/j.vetmic.2012.09.010.

97. 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. *Veterinary microbiology* 2011;153(1-2):144–9. https://doi.org/10.1016/j.vetmic.2011.03.027.

98. Lindmark H, Guus B. SFS, a novel fibronectin-binding protein from *Streptococcus equi,* inhibits the binding between fibronectin and collagen. *Infection and immunity* 1999;May:2383–8.

99. Lindmark H, Jonsson P, Engvall E, Guss B. Pulsed-field gel electrophoresis and distribution of the genes *zag* and *fnz* in isolates of *Streptococcus equi*. *Research in veterinary science* 1999;66(2):93–9. https://doi.org/10.1053/rvsc.1998.0250.

100. Lindmark H, Nilsson M, Guss B. Comparison of the fibronectin-binding protein FNE from *Streptococcus equi* subspecies *equi* with FNZ from *S. equi* subspecies *zooepidemicus* reveals a major and conserved difference. *Infection and immunity* 2001;69(5):3159–63. https://doi.org/10.1128/IAI.69.5.3159-3163.2001.

101. Lindsay A-M, Zhang M, Mitchell Z, Holden MTG, Waller AS, Sutcliffe IC, Black GW. The *Streptococcus equi* prophage-encoded protein SEQ2045 is a hyaluronan-specific hyaluronate lyase that is produced during equine infection. *Microbiology* 2009;155(Pt 2):443–9. https://doi.org/10.1099/mic.0.020826-0.

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

103. Lord J, Carter C, Smith J, Locke S, Phillips E, Odoi A. Antimicrobial resistance among *Streptococcus equi* subspecies *zooepidemicus* and *Rhodococcus equi* isolated from equine specimens submitted to a diagnostic laboratory in Kentucky, USA. *PeerJ* 2022;10:e13682. https://doi.org/10.7717/peerj.13682.

104. Lu M, Gong T, Zhang A, Tang B, Chen J, Zhang Z, Li Y, Zhou X. Mobile genetic elements in streptococci. *Current issues in molecular biology* 2019;32:123–66. https://doi.org/10.21775/cimb.032.123.

105. Lukomski S, Bachert BA, Squeglia F, Berisio R. Collagen-like proteins of pathogenic streptococci. *Molecular microbiology* 2017;103(6):919–30. https://doi.org/10.1111/mmi.13604.

106. Macià MD, Rojo-Molinero E, Oliver A. Antimicrobial susceptibility testing in biofilm-growing bacteria. *Clinical microbiology and infection* 2014;20(10):981–90. https://doi.org/10.1111/1469-0691.12651.

107. Maiden MC, Bygraves JA, Feil E, Morelli G, Russell JE, Urwin R, Zhang Q, Zhou J, Zurth K, Caugant DA, Feavers IM, Achtman M, Spratt BG. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proceedings of the National Academy of Sciences of the United States of America* 1998;95(6):3140–5. https://doi.org/10.1073/pnas.95.6.3140.

108. Maiden MCJ. Multilocus sequence typing of bacteria. *Annual review of microbiology* 2006;60:561–88. https://doi.org/10.1146/annurev.micro.59.030804.121325.

109. Maiden MCJ, van Jansen Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, McCarthy ND. MLST revisited: the gene-by-gene approach to bacterial genomics. *Nature reviews microbiology* 2013;11(10):728–36. https://doi.org/10.1038/nrmicro3093.

110. Makino S, Okada Y, Maruyama T, Ishikawa K, Takahashi T, Nakamura M, Ezaki T, Morita H. Direct and rapid detection of *Erysipelothrix rhusiopathiae* DNA in animals by PCR. *Journal of clinical microbiology* 1994;32(6):1526–31.

111. Mallicote M. Update on *Streptococcus equi* subsp *equi* infections. *The veterinary clinics of North America equine practice* 2015;31(1):27–41. https://doi.org/10.1016/j.cveq.2014.11.003.

112. Malo A, Cluzel C, Labrecque O, Beauchamp G, Lavoie J-P, Leclere M. Evolution of in vitro antimicrobial resistance in an equine hospital over 3 decades. *The Canadian veterinary journal* 2016;57(7):747–51.

113. Manetti AGO, Zingaretti C, Falugi F, Capo S, Bombaci M, Bagnoli F, Gambellini G, Bensi G, Mora M, Edwards AM, Musser JM, Graviss EA, Telford JL, Grandi G, Margarit I. *Streptococcus pyogenes* pili promote pharyngeal cell adhesion and biofilm formation. *Molecular microbiology* 2007;64(4):968–83. https://doi.org/10.1111/j.1365-2958.2007.05704.x.

114. Martinez-Medina M, Naves P, Blanco J, Aldeguer X, Blanco JE, Blanco M, Ponte C, Soriano F, Darfeuille-Michaud A, Garcia-Gil LJ. Biofilm formation as a novel phenotypic feature of adherent-invasive *Escherichia coli* (AIEC). *BMC microbiology* 2009;9:202. https://doi.org/10.1186/1471-2180-9-202.

115. McCoy HE, Broder CC, Lottenberg R. Streptokinases produced by pathogenic group C streptococci demonstrate species-specific plasminogen activation. *The journal of infectious diseases* 1991;164(3):515–21. https://doi.org/10.1093/infdis/164.3.515.

116. McLean R, Rash NL, Robinson C, Waller AS, Paillot R. Localised mitogenic activity in horses following infection with *Streptococcus equi*. *Research in veterinary science* 2015;100:100–4. https://doi.org/10.1016/j.rvsc.2015.03.015.

117. Meehan M, Kelly SM, Price NC, Owen P. The C-terminal portion of the fibrinogen-binding protein of *Streptococcus equi* subsp. *equi* contains extensive alpha-helical coiled-coil structure and contributes to thermal stability. *FEMS microbiology letters* 2002;206(1):81–6. https://doi.org/10.1111/j.1574-6968.2002.tb10990.x.

118. Meehan M, Lynagh Y, Woods C, Owen P. The fibrinogen-binding protein (FgBP) of *Streptococcus equi* subsp. *equi* additionally binds IgG and contributes to virulence in a mouse model. *Microbiology* 2001;147(Pt 12):3311–22. https://doi.org/10.1099/00221287-147-12-3311.

119. Meehan M, Muldowney DA, Watkins NJ, Owen P. Localization and characterization of the ligand-binding domain of the fibrinogen-binding protein (FgBP) of *Streptococcus equi* subsp. *equi*. *Microbiology* 2000;146 (Pt 5):1187–94. https://doi.org/10.1099/00221287-146-5-1187.

120. Miller DA, Luo L, Hillson N, Keating TA, Walsh CT. *Yersiniabactin* synthetase: a four-protein assembly line producing the nonribosomal peptide/polyketide hybrid siderophore of *Yersinia pestis*. *Chemistry and biology* 2002;9(3):333–44. https://doi.org/10.1016/s1074-5521(02)00115-1.

121. Miranda PSD, Lannes-Costa PS, Pimentel BAS, Silva LG, Ferreira-Carvalho BT, Menezes GC, Mattos-Guaraldi AL, Hirata R, Mota RA, Nagao PE. Biofilm formation on different pH conditions by *Streptococcus agalactiae* isolated from bovine mastitic milk. *Letters in applied microbiology* 2018;67(3):235–43. https://doi.org/10.1111/lam.13015.

122. Mitchell C, Steward KF, Charbonneau ARL, Walsh S, Wilson H, Timoney JF, Wernery U, Joseph M, Craig D, van Maanen K, Hoogkamer-van Gennep A, Leon A, Witkowski L, Rzewuska M, Stefańska I, Żychska M, van Loon G, Cursons R, Patty O, Acke E, Gilkerson JR, El-Hage C, Allen J, Bannai H, Kinoshita Y, Niwa H, Becú T, Pringle J, Guss B, Böse R, Abbott Y, Katz L, Leggett B, Buckley TC, Blum SE, Cruz López F, Fernández Ros A, Marotti Campi MC, Preziuso S, Robinson C, Newton JR, Schofield E, Brooke B, Boursnell M, Brauwere N de, Kirton R, Barton CK, Abudahab K, Taylor B, Yeats CA, Goater R, Aanensen DM, Harris SR, Parkhill J, Holden MTG, Waller AS. Globetrotting strangles: the unbridled national and international transmission of *Streptococcus equi* between horses. *Microbial genomics* 2021;7(3). https://doi.org/10.1099/mgen.0.000528.

123. Morris ERA, Boyle AG, Riihimäki M, Aspán A, Anis E, Hillhouse AE, Ivanov I, Bordin AI, Pringle J, Cohen ND. 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(6):e0252804. https://doi.org/10.1371/journal.pone.0252804.

124. Morris ERA, Hillhouse AE, Konganti K, Wu J, Lawhon SD, Bordin AI, Cohen ND. 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. *Veterinary microbiology* 2020;243:108638. https://doi.org/10.1016/j.vetmic.2020.108638.

125. Moyes RB, Reynolds J, Breakwell DP. Differential staining of bacteria: gram stain. *Current protocols in microbiology* 2009; https://doi.org/10.1002/9780471729259.mca03cs15.

126. Napoleon I. *Correspondence de Napoléon I: publiée par ordre de l'empereur Napoléon III. H. Plon, J. Dumaine*. Available at: https://books.google.co.uk/books?id=EPQtAAAAYAAJ&pg=PA100&lpg=PA100&dq=gourme+du+cheval+napoléon&source=bl&ots=HnR1O8FGbX&sig=EMuErNAbM-xPYY-uv1Gb5Qz-DZM&hl=en&sa=X&ved=0ahUKEwiyuaOR987UAhWDZVAKHRQoAZUQ6AEIQTAE#v=onepage&q=gourme du cheval napoléon&f=false; 1811 [accessed October 22, 2021].

127. Neumann B, Prior K, Bender JK, Harmsen D, Klare I, Fuchs S, Bethe A, Zühlke D, Göhler A, Schwarz S, Schaffer K, Riedel K, Wieler LH, Werner G. A core genome

multilocus sequence typing scheme for *Enterococcus faecalis*. *Journal of clinical microbiology* 2019;57(3). https://doi.org/10.1128/JCM.01686-18.

128. Newton JR, Verheyen K, Talbot NC, Timoney JF, Wood JLN, Lakhani KH, Chanter N. Control of strangles outbreaks by isolation of guttural pouch carriers identified using PCR and culture of *Streptococcus equi*. *Equine veterinary journal* 2000;32(6):515–26.

129. Newton JR, Wood JLN, Dunn KA, DeBrauwere MN, Chanter N. Naturally occurring persistent and asymptomatic infection of the guttural pouches of horses with *Streptococcus equi*. *The veterinary record* 1997;1997(140):84–90.

130. Nikolenko SI, Korobeynikov AI, Alekseyev MA. BayesHammer: Bayesian clustering for error correction in single-cell sequencing. *BMC genomics* 2013;14(1):S7. https://doi.org/10.1186/1471-2164-14-S1-S7.

131. Nocera FP, D'Eletto E, Ambrosio M, Fiorito F, Pagnini U, Martino L de. Occurrence and antimicrobial susceptibility profiles of *Streptococcus equi* subsp. *zooepidemicus* strains isolated from mares with fertility problems. *Antibiotics* 2021;11(1). https://doi.org/10.3390/antibiotics11010025.

132. Nyberg LK, Quaderi S, Emilsson G, Karami N, Lagerstedt E, Müller V, Noble C, Hammarberg S, Nilsson AN, Sjöberg F, Fritzsche J, Kristiansson E, Sandegren L, Ambjörnsson T, Westerlund F. Rapid identification of intact bacterial resistance plasmids via optical mapping of single DNA molecules. *Scientific reports* 2016;6:30410. https://doi.org/10.1038/srep30410.

133. Oehmcke S, Shannon O, Mörgelin M, Herwald H. Streptococcal M proteins and their role as virulence determinants. *Clinica chimica acta* 2010;411(17-18):1172–80. https://doi.org/10.1016/j.cca.2010.04.032.

134. Oxoid. *Brain Heart Infusion broth: dehydrated culture media*. Available at: http://www.oxoid.com/UK/blue/prod_detail/prod_detail.asp?pr=CM1135&c#:~:text=Oxoid%20Brain%20Heart%20Infusion%20Broth,cultivation%20of%20dental%20pathogens2; 2023 [accessed February 10, 2023].

135. Paillot R, Lopez-Alvarez MR, Newton JR, Waller AS. Strangles: A modern clinical view from the 17th century. *Equine veterinary journal* 2017;49(2):141–5. https://doi.org/10.1111/evj.12659.

136. Paillot R, Robinson C, Steward K, Wright N, Jourdan T, Butcher N, Heather Z, Waller AS. Contribution of each of four Superantigens to *Streptococcus equi*-induced mitogenicity, gamma interferon synthesis, and immunity. *Infection and immunity* 2010;78(4):1728–39. https://doi.org/10.1128/IAI.01079-09.

137. Paradis E, Claude J, Strimmer K. APE: Analyses of Phylogenetics and Evolution in R language. *Bioinformatics* 2004;20(2):289–90. https://doi.org/10.1093/bioinformatics/btg412.

138. Parkinson NJ, Robin C, Newton JR, Slater J, Waller AS. Molecular epidemiology of strangles outbreaks in the UK during 2010. *The veterinary record* 2011;168(25):666. https://doi.org/10.1136/vr.d1485.

139. Partridge SR, Kwong SM, Firth N, Jensen SO. Mobile genetic elements associated with antimicrobial resistance. *Clinical microbiology reviews* 2018;31(4). https://doi.org/10.1128/CMR.00088-17.

140. Poyart C, Quesne G, Coulon S, Berche P, Trieu-Cuot P. Identification of streptococci to species level by sequencing the gene encoding the manganese-dependent superoxide dismutase. *Journal of clinical microbiology* 1998;36(1):41–7. https://doi.org/10.1128/JCM.36.1.41-47.1998.

141. 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. *Journal of veterinary internal medicine* 2019;34(1):294–9. https://doi.org/10.1111/jvim.15668.

142. Proft T, Fraser JD. Streptococcal superantigens. *Chemical immunology and allergy* 2007;93:1–23. https://doi.org/10.1159/000100851.

143. Proft T, Webb PD, Handley V, Fraser JD. Two novel superantigens found in both group A and group C *Streptococcus*. *Infection and immunity* 2003;71(3):1361–9. https://doi.org/10.1128/iai.71.3.1361-1369.2003.

144. Pusterla N, Kass PH, Mapes S, Johnson C, Barnett DC, Vaala W, Gutierrez C, McDaniel R, Whitehead B, Manning J. Surveillance programme for important equine infectious respiratory pathogens in the USA. *The veterinary record* 2011;169(1):12. https://doi.org/10.1136/vr.d2157.

145. Rasmussen CD, Haugaard MM, Petersen MR, Nielsen JM, Pedersen HG, Bojesen AM. *Streptococcus equi* subsp. *zooepidemicus* isolates from equine infectious endometritis belong to a distinct genetic group. *Veterinary research* 2013;44(1):26. https://doi.org/10.1186/1297-9716-44-26.

146. Resch A, Leicht S, Saric M, Pásztor L, Jakob A, Götz F, Nordheim A. Comparative proteome analysis of *Staphylococcus aureus* biofilm and planktonic cells and correlation with transcriptome profiling. *Proteomics* 2006;6(7):2327. https://doi.org/10.1002/pmic.200690035.

147. Richards VP, Palmer SR, Pavinski Bitar PD, Qin X, Weinstock GM, Highlander SK, Town CD, Burne RA, Stanhope MJ. Phylogenomics and the dynamic genome evolution of the *genus Streptococcus*. *Genome biology and evolution* 2014;6(4):741–53. https://doi.org/10.1093/gbe/evu048.

148. Robinson C, Frykberg L, Flock M, Guss B, Waller AS, Flock J-I. Strangvac: A recombinant fusion protein vaccine that protects against strangles, caused by *Streptococcus equi*. *Vaccine* 2018;36(11):1484–90. https://doi.org/10.1016/j.vaccine.2018.01.030.

149. Robinson C, Steward KF, Potts N, Barker C, Hammond T, Pierce K, Gunnarsson E, Svansson V, Slater J, Newton JR, Waller AS. Combining two serological assays optimises sensitivity and specificity for the identification of *Streptococcus equi* subsp. *equi* exposure. *Veterinary journal* 2013;197(2):188–91. https://doi.org/10.1016/j.tvjl.2013.01.033.

150. Robinson C, Waller AS, Frykberg L, Flock M, Zachrisson O, Guss B, Flock J-I. Intramuscular vaccination with Strangvac is safe and induces protection against equine strangles caused by *Streptococcus equi*. *Vaccine* 2020;38(31):4861–8. https://doi.org/10.1016/j.vaccine.2020.05.046.

151. Roehr JT, Dieterich C, Reinert K. Flexbar 3.0 - SIMD and multicore parallelization. *Bioinformatics* 2017;33(18):2941–2. https://doi.org/10.1093/bioinformatics/btx330.

152. Römling U, Fislage R, Tümmler B. Theory and practice of macrorestriction analysis for the clonal analysis of pathogens. *Immunitat und Infektion* 1995;23(1):4–8.

153. Roosaare M, Puustusmaa M, Möls M, Vaher M, Remm M. PlasmidSeeker: identification of known plasmids from bacterial whole genome sequencing reads. *PeerJ* 2018;6:e4588. https://doi.org/10.7717/peerj.4588.

154. Rotinsulu DA, Ewers C, Kerner K, Amrozi A, Soejoedono RD, Semmler T, Bauerfeind R. Molecular features and antimicrobial susceptibilities of *Streptococcus equi* ssp. *equi*

isolates from strangles cases in Indonesia. *Veterinary sciences* 2023;10(1). https://doi.org/10.3390/vetsci10010049.

155. Ruffo G. *De medicina equorum*. Available at: http://wellcomelibrary.org/player/b19689755; 1251.

156. Sadaka C, Kanellos T, Guardabassi L, Boucher J, Watts JL. Evaluation of veterinary-specific interpretive criteria for susceptibility testing of *Streptococcus equi s*ubspecies with trimethoprim-sulfamethoxazole and trimethoprim-sulfadiazine. *Journal of clinical microbiology* 2017;55(1):326–30. https://doi.org/10.1128/JCM.01610-16.

157. Santoro F, Iannelli F, Pozzi G. Genomics and genetics of *Streptococcus pneumoniae*. *Microbiology spectrum* 2019;7(3). https://doi.org/10.1128/microbiolspec.GPP3-0025-2018.

158. Schliep KP. phangorn: phylogenetic analysis in R. *Bioinformatics* 2011;27(4):592–3. https://doi.org/10.1093/bioinformatics/btq706.

159. Schwarz S, Silley P, Simjee S, Woodford N, van Duijkeren E, Johnson AP, Gaastra W. Editorial: assessing the antimicrobial susceptibility of bacteria obtained from animals. *The Journal of antimicrobial chemotherapy* 2010;65(4):601–4. https://doi.org/10.1093/jac/dkq037.

160. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 2014;30(14):2068–9. https://doi.org/10.1093/bioinformatics/btu153.

161. Segerman B. The genetic integrity of bacterial species: the core genome and the accessory genome, two different stories. *Frontiers in cellular and infection microbiology* 2012;2:116. https://doi.org/10.3389/fcimb.2012.00116.

162. Shanks RMQ, Donegan NP, Graber ML, Buckingham SE, Zegans ME, Cheung AL, O'Toole GA. Heparin stimulates *Staphylococcus aureus* biofilm formation. *Infection and immunity* 2005;73(8):4596–606. https://doi.org/10.1128/IAI.73.8.4596-4606.2005.

163. Sherman JM. The Streptococci. *Bacteriological reviews* 1937;1(1):3–97. https://doi.org/10.1128/br.1.1.3-97.1937.

164. Sigma Aldrich. *T1438 Todd Hewitt Broth*. Available at: https://www.sigmaaldrich.com/deepweb/assets/sigmaaldrich/product/documents/119/730/T1438dat.pdf; 2018 [accessed February 10, 2023].

165. Silvestre I, Borrego MJ, Jordão L. Biofilm formation by ST17 and ST19 strains of *Streptococcus agalactiae*. *Research in microbiology* 2020;171(8):311–8. https://doi.org/10.1016/j.resmic.2020.08.001.

166. Singh SK, Bharati AP, Singh N, Pandey P, Joshi P, Singh K, Mitra K, Gayen JR, Sarkar J, Akhtar MS. The prophage-encoded hyaluronate lyase has broad substrate specificity and is regulated by the N-terminal domain. *The journal of biological chemistry* 2014;289(51):35225–36. https://doi.org/10.1074/jbc.M113.507673.

167. SLU. *Veterinary bacteriology: information about important bacteria: Streptococcus equi subsp. equi*. Available at: http://www.vetbact.org/species/14; 2021 [accessed October 27, 2021].

168. Smole SC, King LA, Leopold PE, Arbeit RD. Sample preparation of Gram-positive bacteria for identification by matrix assisted laser desorption/ionization time-of-flight. *Journal of microbiological methods* 2002;48(2-3):107–15. https://doi.org/10.1016/s0167-7012(01)00315-3.

169. Soedarmanto I, Pasaribu FH, Wibawan IW, Lämmler C. Identification and molecular characterization of serological group C streptococci isolated from diseased pigs and

monkeys in Indonesia. *Journal of clinical microbiology* 1996;34(9):2201–4. https://doi.org/10.1128/jcm.34.9.2201-2204.1996.

170. Sriskandan S, Faulkner L, Hopkins P. *Streptococcus pyogenes*: Insight into the function of the streptococcal superantigens. *The international journal of biochemistry & cell biology* 2007;39(1):12–9. https://doi.org/10.1016/j.biocel.2006.08.009.

171. Staali L, Mörgelin M, Björck L, Tapper H. *Streptococcus pyogenes* expressing M and M-like surface proteins are phagocytosed but survive inside human neutrophils. *Cellular microbiology* 2003;5(4):253–65. https://doi.org/10.1046/j.1462-5822.2003.00272.x.

172. Steward KF, Robinson C, Maskell DJ, Nenci C, Waller AS. Investigation of the Fim1 putative pilus locus of *Streptococcus equi* subspecies *equi*. *Microbiology* 2017;163(8):1217–28. https://doi.org/10.1099/mic.0.000506.

173. Steward KF, Robinson C, Waller AS. Transcriptional changes are involved in phenotype switching in *Streptococcus equi* subspecies *equi*. *Molecular bioSystems* 2016;12(4):1194–200. https://doi.org/10.1039/c5mb00780a.

174. Sweeney CR, Timoney JF, Newton JR, Hines MT. *Streptococcus equi* infections in horses: guidelines for treatment, control, and prevention of strangles. *Journal of veterinary internal medicine* 2005;2005(19):123–34.

175. Takai S, Anzai T, Yashiro H, Ishii C, Tsubaki S, Wada R, Timoney JF. Detection of DNA restriction fragment polymorphisms in *Streptococcus equi*. *The veterinary record* 2000;146(6):159–61. https://doi.org/10.1136/vr.146.6.159.

176. Tettelin H, Medini D. *The pangenome: diversity, dynamics and evolution of genomes*. Cham (CH); 2020.

177. Timoney JF. The pathogenic equine streptococci. *Veterinary research* 2004;35(4):397–409. https://doi.org/10.1051/vetres:2004025.

178. Timoney JF, Artiushin SC, Boschwitz JS. Comparison of the sequences and functions of *Streptoccus equi* M-Like Proteins SeM and SzPSe. *Infection and immunity* 1997;65(9):3600–5.

179. 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(6):1522–7. https://doi.org/10.1016/j.vaccine.2009.11.064.

180. Timoney JF, Kumar P. Early pathogenesis of equine *Streptococcus equi* infection (strangles). *Equine veterinary journal* 2008;40(7):637–42. https://doi.org/10.2746/042516408x322120.

181. Timoney JF, Qin A, Muthupalani S, Artiushin S. Vaccine potential of novel surface exposed and secreted proteins of *Streptococcus equi*. *Vaccine* 2007;25(30):5583–90. https://doi.org/10.1016/j.vaccine.2007.02.040.

182. Tiwari R, Qin A, Artiushin S, Timoney JF. Se18.9, an anti-phagocytic factor H binding protein of *Streptococcus equi*. *Veterinary microbiology* 2007;121(1-2):105–15. https://doi.org/10.1016/j.vetmic.2006.11.023.

183. Tonpitak W, Sornklien C, Wutthiwithayaphong S. Characterization of a *Streptococcus equi* ssp. *equi* isolate from a strangles outbreak in Thailand. *Journal of equine veterinary science* 2016;38:30–2. https://doi.org/10.1016/j.jevs.2016.01.001.

184. Torpiano P, Nestorova N, Vella C. *Streptococcus equi* subsp. *equi* meningitis, septicemia and subdural empyema in a child. *IDCases* 2020;21:e00808. https://doi.org/10.1016/j.idcr.2020.e00808.

185. Tscheschlok L, Venner M, Steward K, Böse R, Riihimäki M, Pringle J. Decreased clinical severity of strangles in weanlings associated with restricted seroconversion to optimized *Streptococcus equi* ssp *equi a*ssays. *Journal of veterinary internal medicine* 2018;32(1):459–64. https://doi.org/10.1111/jvim.15037.

186. Vale FF, Lehours P, Yamaoka Y. Editorial: The role of mobile genetic elements in bacterial evolution and their adaptability. *Frontiers in microbiology* 2022;13:849667. https://doi.org/10.3389/fmicb.2022.849667.

187. Varaldo PE, Montanari MP, Giovanetti E. Genetic elements responsible for erythromycin resistance in streptococci. *Antimicrobial agents and chemotherapy* 2009;53(2):343–53. https://doi.org/10.1128/AAC.00781-08.

188. Walker JA, Timoney JF. Construction of a stable non-mucoid deletion mutant of the *Streptococcus equi* Pinnacle vaccine strain. *Veterinary microbiology* 2002;89(4):311–21. https://doi.org/10.1016/s0378-1135(02)00205-5.

189. Waller A. *Streptococcus equi*: breaking its strangles-hold. *The Veterinary record* 2018;182(11):316–8. https://doi.org/10.1136/vr.k1231.

190. Waller AS. Strangles: taking steps towards eradication. *Veterinary microbiology* 2013;167(1-2):50–60. https://doi.org/10.1016/j.vetmic.2013.03.033.

191. Waller AS. New perspectives for the diagnosis, control, treatment, and prevention of strangles in horses. *The Veterinary clinics of North America: Equine practice* 2014;30(3):591–607. https://doi.org/10.1016/j.cveq.2014.08.007.

192. Waller AS, Paillot R, Timoney JF. *Streptococcus equi*: a pathogen restricted to one host. *Journal of medical microbiology* 2011;60(Pt 9):1231–40. https://doi.org/10.1099/jmm.0.028233-0.

193. Wang Y, Wang Y, Sun L, Grenier D, Yi L. *Streptococcus suis* biofilm: regulation, drug-resistance mechanisms, and disinfection strategies. *Applied microbiology and biotechnology* 2018;102(21):9121–9. https://doi.org/10.1007/s00253-018-9356-z.

194. Waśniewska-Włodarczyk A, Pepaś R, Janowicz R, Konopka W. *Streptococcus equi* subsp. *equi* in retropharyngeal abscess: case report and review of literature. *Microorganisms* 2022;10(10). https://doi.org/10.3390/microorganisms10102032.

195. Webb K, Barker C, Harrison T, Heather Z, Steward KF, Robinson C, Newton JR, Waller AS. Detection of *Streptococcus equi* subspecies *equi* using a triplex qPCR assay. *Veterinary journal* 2013;195(3):300–4. https://doi.org/10.1016/j.tvjl.2012.07.007.

196. Webb K, Jolley KA, Mitchell Z, Robinson C, Newton JR, Maiden MCJ, Waller A. Development of an unambiguous and discriminatory multilocus sequence typing scheme for the *Streptococcus zooepidemicus* group. *Microbiology* 2008;154(Pt 10):3016–24. https://doi.org/10.1099/mic.0.2008/018911-0.

197. Weese JS, Jarlot C, Morley PS. Survival of *Streptococcus equi* on surfaces in an outdoor environment. *Canadian veterinary journal* 2009;2009(50):968–70.

198. Whitlock F, Newton R, Waller A. 'Subtle strangles' – the more elusive signs of the disease. *Equine health* 2019.

199. World Health Organization (WHO). *Critically important antimicrobials for human medicine: ranking of medically important antimicrobials for risk management of antimicrobial resistance due to non human use; 6th revision*. Geneva, Switzerland; 2019.

200. Yadav MK, Chae S-W, Park K, Song J-J, van der Mei HC. Hyaluronic acid derived from other streptococci supports *Streptococcus pneumoniae in vitro* biofilm formation.

BioMed research international 2013;2013:690217. https://doi.org/10.1155/2013/690217.

201. Yi L, Wang Y, Ma Z, Lin H-X, Xu B, Grenier D, Fan H, Lu C. Identification and characterization of a *Streptococcus equi* ssp. *zooepidemicus* immunogenic GroEL protein involved in biofilm formation. *Veterinary research* 2016;47:50. https://doi.org/10.1186/s13567-016-0334-0.

202. Yi L, Wang Y, Ma Z, Zhang H, Li Y, Zheng J, Yang Y, Fan H, Lu C. Biofilm formation of *Streptococcus equi* ssp. *zooepidemicus* and comparative proteomic analysis of biofilm and planktonic cells. *Current microbiology* 2014;69(3):227–33. https://doi.org/10.1007/s00284-014-0574-z.

203. Yu R, Xu Y, Schwarz S, Shang Y, Yuan X, Zhang Y, Li D, Du X-D. *erm*(T)-Mediated Macrolide-Lincosamide Resistance in *Streptococcus suis*. *Microbiology spectrum* 2022;10(1):e0165721. https://doi.org/10.1128/spectrum.01657-21.

204. Zadoks RN, Schukken YH, Wiedmann M. Multilocus sequence typing of *Streptococcus uberis* provides sensitive and epidemiologically relevant subtype information and reveals positive selection in the virulence gene pauA. *Journal of clinical microbiology* 2005;43(5):2407–17. https://doi.org/10.1128/JCM.43.5.2407-2417.2005.

205. Zankari E, Allesøe R, Joensen KG, Cavaco LM, Lund O, Aarestrup FM. PointFinder: a novel web tool for WGS-based detection of antimicrobial resistance associated with chromosomal point mutations in bacterial pathogens. *The journal of antimicrobial chemotherapy* 2017;72(10):2764–8. https://doi.org/10.1093/jac/dkx217.

206. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. Identification of acquired antimicrobial resistance genes. *The Journal of antimicrobial chemotherapy* 2012;67(11):2640–4. https://doi.org/10.1093/jac/dks261.

207. Zhu Y, Chen S, Yi Z, Holyoak R, Wang T, Ding Z, Li J. Nasopharyngeal microbiomes in donkeys shedding *Streptococcus equi* subspecies *equi* in comparison to healthy donkeys. *Frontiers in veterinary science* 2021;8.

Acknowledgements

This research and Ph.D. thesis would not have been possible without the help and support of many people. Therefore, I would like to thank all the institutions and individuals involved.

I would like to express my deepest gratitude to Prof. Dr. Rolf Bauerfeind, my supervisor and promotor from the Institute for Hygiene and Infectious Diseases of Animals (IHIT), Justus Liebig University (JLU) Giessen, for his excellent supervision and guidance during my Ph.D. programme and research. My sincerest gratitude extends to him for his support of my scholarship applications, for providing the research facilities, and for his valuable advice during the entire research and writing process. His critical questions, mentorship, and passion for transferring knowledge extensively developed my scientific thinking capability, which will indeed contribute to my future career. He always has an open door for discussion and a goodwill to find the best solution during challenging situations.

I am very grateful to Prof. Dr. Christa Ewers, my supervisor and the managing director of IHIT, JLU Giessen, for her profound support, valuable advice, and for the research facilities. I am very thankful to her for facilitating the network to obtain *S. equi* isolates from the Free University of Berlin and IDEXX Ludwigsburg as well as the cooperation with Robert Koch Institute, Germany. Her expertise in genome analysis and bacteriology enhanced this research in many ways.

I would like to sincerely thank Prof. Dr. Peter Kämpfer, my co-supervisor from the Institute for Applied Microbiology, JLU Giessen, for his valuable input in this research. I am very grateful for the scientific advice of Prof. Dr. Peter Valentin-Weigand from the Institute of Microbiology, University of Veterinary Medicine Hannover, as the reviewer of my Ph.D. thesis, and Prof. Dr. Ralph Schermuly from the Department of Internal Medicine, Faculty of Medicine, Justus Liebig University Giessen, as the chair of the examination committee. Their insightful questions during the oral thesis defense enhanced the quality of this study.

I am very grateful to the rector of IPB University, Indonesia, and the dean of the School of Veterinary Medicine and Biomedical Science, IPB University, Indonesia, who permitted me to temporarily leave my home university to pursue a Ph.D. degree abroad. I am thankful to Prof. Dr. I Wayan T. Wibawan for his support and for attending my oral thesis defense.

I am very much grateful to *Deutscher Akademischer Austauschdienst* (DAAD) (German Academic Exchange Service), which financially supported my Ph.D. study in

Germany through the funding programme Research Grants - Doctoral Programmes in Germany. It is an honour for me to be a DAAD alumni, and I definitely experienced the motto of DAAD “to change by exchange” during my stay in Germany. I would also like to thank JLU Giessen for awarding me the dissertation completion grant on the basis of the JLU’s Gender Equality Concept. I am thankful to Giessen Graduate Centre for the Life Sciences (GGL), which enabled me to participate in excellent soft skill programmes, conferences, and seminars.

I would like to express my sincere gratitude to our cooperation partners Dr. Ivonne Stamm and Dr. Antina Lübke-Becker, for providing *S. equi* isolates from IDEXX Laboratories Ludwigsburg and the Institute of Microbiology and Epizootic Diseases, the Free University of Berlin, respectively. Many thanks to Prof. Dr. Kerstin Fey from the Clinic for Horses Internal Medicine, JLU Giessen, for her expert input regarding the clinical side of strangles. Many thanks also to Dr. Amrozi, and Prof. Dr. Retno D. Soejoedono for sending us valuable *S. equi* isolates from IPB University, Indonesia. I sincerely thank Dr. Torsten Semmler and his working group at NG1-Microbial Genomics, Robert Koch Institute, Germany, for their cooperation in the whole genome sequencing of the *S. equi* isolates in this study. I am thankful to Dr. Katrin Büttner from the working group of Biomathematics and Data Analysis, JLU Giessen, for her expert advice on statistical analyses. Many thanks to Prof. Tobias Eisenberg and Dr. Karin Schlez, for the opportunity to perform an internship at the Department of Veterinary Medicine, Hessian State Laboratory (Landesbetrieb Hessisches Landeslabor), Germany.

I am very grateful to Dr. Andrew Waller, formerly at the Animal Health Trust, United Kingdom, and now at Intervacc AB, Sweden, for the fascinating and inspiring discussions about *S. equi*, explaining the functions of Pathogenwatch and Microreact during cgMLST analysis as well as for providing the amino acid sequences data of Strangvac® antigens.

I am thankful to Dr. Katharina Kerner for her support during the early phase of this research. I am very thankful to Dr. Ellen Prenger-Berninghoff, Sophie Aurich, Dr. Christiane Hoffmann, Dr. Carsten Heydel, and the diagnostic team at IHIT (especially Peppi, Sonja, Julita, and Joerg) for their excellent bacteriological diagnostic support. I am very grateful to Anja Schwanitz and Ursula Leidner for explaining various laboratory methods, for their excellent technical assistance, and the nice talks during coffee breaks. I am very thankful to all the staffs at IHIT who administratively and technically prepared all materials required for this research. My thanks also go to Ismet for his technical assistance at IPB University.

I am very thankful to my friends and the doctorates at IHIT: Steffen Hermanns, Sophie Aurich, Carina Leitner, Johanna Rühl-Teichner, Aung Zaw Moe, Lisa Göpel, Dr.

Franziska Dambon, Dr. Lisa Jacobmeyer, and Tammy Schmidt for their friendship, sharing, and support, both scientifically and privately. You made my life in Giessen more colorful! Thank you very much Mahran Sasmaja, Dr. Riyanti, Dr. Benedicta, Dr. Zahady, Dominic, May Linn, family Obadja, and Giessen Stronghold community for our friendship.

I am very grateful and forever in debt to Prof. Dr. Dr. Bernd Hoffmann, Mrs. Irene Hoffmann, Dr. Ernst Vockert, Dr. Hannelore Vockert (†), Mr. Ulrich Zeidler, and Mrs. Marliese Zeidler for their never-ending kindness and care for my family and me. I am very blessed to find a piece of family and home through them.

To my beloved mother and father, Dr. Sri Adiani (†) and Dr. Fredrik D. Rotinsulu, thank you very much for your love, sacrifices, and everything. You inspired me in many ways, including becoming a veterinarian and a good parent. You always remind me to have faith that there are blessings in disguise, and to keep striving to reach my dreams. Mama, may you rest in eternal peace.

To my beloved children, Diobert and Deandra Hutapea, thank you very much for your unconditional love, patience, and support, which always succeeded in recharging my energy and motivation. To my husband, Dungdang Hutapea, thank you so much for your love, trust, patience, and sacrifices for our family. Your character and actions strengthened us to go through ups and downs in reaching our family vision. To the families Sadu-Soekarno, Hutapea-Simanjuntak, Rotinsulu-Wakari, and Hutahaean-Sianipar, thank you very much for your support. I am very blessed to be surrounded by wonderful people.

Appendices

Table 22. Virulence-associated genes of *Streptococcus equi* ssp. *equi* investigated in this study

Virulence Factor	Virulence-associated gene(s)	Accession number (position); locus tag
Adherence		
CNE	*cne*	CWFT01000004.1 (58425..60398); SEQ0935
FNE	*fne*	FM204883.1 8354323..356212); SEQ0375
FneB	*fneB*	FM204883.1 (2019108..2020535); SEQ1999
FneC	*fneC*	FM204883.1 (1611888..1613774); SEQ1606
FneD	*fneD*	FM204883.1 (1614087..1615119); SEQ1607
FneE	*fneE*	FM204883.1 (534129..535364); SEQ0555
FneF	*fneF*	FM204883.1 (1656030..1657580); SEQ1649
SFS	*sfs*	FM204883.1 (452510..453625); SEQ0466
SclC	*sclC*	CWFT01000042.1 (15864..16772); SEQ2101
SclD	*sclD*	FM204883.1 (262550..263491); SEQ0280
SclE	*sclE*	FM204883.1 (616654..6175929); SEQ0633
SclF	*sclF*	CWFT01000003.1 (69953..71047); SEQ0855
SclG	*sclG*	FM204883.1 (95988..97181); SEQ0090
SclH	*sclH*	FM204883.1 (245894..247138); SEQ0260
SclI	*sclI*	CWFT01000015.1 (42440..43888)
Scl3F	*scl3F*	AY326319.1
Eq5	*eq5*	CWFT01000007.1 (18069..19895); SEQ0256
Eq8	*eq8*	CWFT01000026.1 (4738..5787); SEQ0402
Eq54/FimI	*eq54/fimI*	FM204883.1 (916183..917625); SEQ0936
TetR-like	*tetR*	FM204883.1 (913524..914051); SEQ0934
SEQ0944	*Seq0944*	FM204883.1 (923891..927607); SEQ0944
Immune modulation		
Hyaluronic acid capsule	*hasA*	FM204883.1 (254370..255623); SEQ0269
	hasB	FM204883.1 (255876..257081); SEQ0270
	hasB	FM204883.1 (280432..281793); SEQ0296
	hasC1/gtaB1	FM204883.1 (257131..2580399); SEQ0271
	hasC2 gtaB2	FM204883.1 (272684..273586); SEQ0289
SeM	*seM/fbp*	FM204883.1 (2043361..2044965); SEQ2017
SzPSe	*szPSe*	FM204883.1 (912238..913362); SEQ0933
SzP	*szP*	FM204883.1 (549915..550936); SEQ0566
EAG	*eag*	CWFT01000025.1 (13552..14841); SEQ0721
Se18.9	*se18.9*	FM204883.1 (216947..217438); SEQ0235

Exoenzymes and exoproteins		
Hyaluronidase/ hyaluronate lyase	*seq2045*	FM204883.1 (2071724..2072842), SEQ2045
	hylB (seq1479)	FM204883.1 (1486225..1489412); SEQ1479
Streptokinase	*Skc*	FM204883.1 (2039042..2040334); SEQ2014
EndoS	*endoS*	CWFT01000047.1 (7071..10127)
IdeE (SEQ0999)	*ideE*	CWFT01000043.1 (7849..8898)
IdeE (SEQ0938)	*ideE*	FM204883.1 (918830..919984); SEQ0938
IdeE2	*ideE2*	FJ792821.1
SodA	*sodA*	Z95902.1
SortaseA	*srtA*	FM204883.1 (1156008..1156721); SEQ1171
SortaseC1	*srtC1*	FM204883.1 (917706..918515); SEQ0937
Prolipoprotein diacylglycerol transferase (Lgt)	*lgt*	FM204883.1 (1537563..1538342); SEQ1537
Exotoxins		
Streptolysin S	*sagA (streptolysin S precursor)*	FM204883.1 (525696..525860); SEQ0546
	seq0547	FM204883.1 (526088..527038); SEQ0547
	seq0548	FM204883.1 (527035..528099); SEQ0548
	seq0549	FM204883.1 (528112..529470); SEQ0549
SeeH	*seeH*	AF186180.1
SeeI	*seeI*	KU215885.1
SeeL	*seeL*	AJ583527.1
SeeM	*seeM*	AJ583528.1
SlaA	*slaA*	FM204883.1 (824084..824659); SEQ0849
SlaB	*slaB*	FM204883.1 (2164542..2165114); SEQ2155
Nutritional/metabolic factors		
Equibactin	*eqbA*	FM204883.1 (1232968..1233348); SEQ1246
	eqbB	FM204883.1 (1231990..1232739); SEQ1245
	eqbC	FM204883.1 (1231320..1232003); SEQ1244
	eqbD	FM204883.1 (1229728..1231311); SEQ1243
	eqbE	FM204883.1 (1223642..1229713); SEQ1242
	eqbF	FM204883.1 (1221519..1223645); SEQ1241
	eqbG	FM204883.1 (1217717..1221535); SEQ1240
	eqbH	FM204883.1 (1217140..1217727); SEQ1239
	eqbI	FM204883.1 (1216467..1217138); SEQ1238
	eqbJ	FM204883.1 (1215077..1216453); SEQ1237
	eqbK	FM204883.1 (1213297..1215054); SEQ1236
	eqbL	FM204883.1 (1211547..1213292); SEQ1235
	eqbM	FM204883.1 (1210463..1211527); SEQ1234
	eqbN	FM204883.1 (1209536..1210441); SEQ1233
Shr	*shr*	FM204883.1 (424196..427947); SEQ0443

Supplementary materials

S1. Culture media, biologicals, chemicals, consumables, and devices used in this study

Table 23. Culture media, biologicals, and chemicals used in this study

Chemicals	Supplier	Article Number
Acetic acid	Carl Roth GmbH & Co KG, Karlsruhe, Germany	3738.2
Aceton	Carl Roth GmbH & Co KG, Karlsruhe, Germany	CP40.6
Agar-agar BioScience	Carl Roth GmbH & Co KG, Karlsruhe, Germany	6494.4
Agarose	Anprotec, Bruckberg, Germany	AC-GN 00009
Alfa-cyano-4-hydroxycinnamic acid	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	476870
Bacterial test standard for Mass Spectrometry	Bruker Daltonik GmbH, Bremen, Germany	8255343
Bacto-trypton	Becton, Dickinson and Company, Le Pont de Claix, France	211705
Blood agar base	Oxoid GmbH, Wesel, Germany	CM0055
Boric acid	Carl Roth GmbH & Co KG, Karlsruhe, Germany	6943
Brain Heart Infusion	Oxoid GmbH, Wesel, Germany	CM1135
Brij® 35	Carl Roth GmbH & Co KG, Karlsruhe, Germany	CN21.1
Buffer B (10X)	Thermo Fisher Scientific, Waltham, United States	BB5
Chloroform	VWR International GmbH, Darmstadt, Germany	22711260
Columbia agar base	Carl Roth GmbH & Co KG, Karlsruhe, Germany	1507.1
Crystal violet	Merck KGaA, Darmstadt, Germany	1014080025
Desoxycholat	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	D6750
DreamTaq Green Buffer (10X) (includes 20 mM$MgCl_2$)	Thermo Fisher Scientific Baltics, Vilnius, Lithuania	B71
DreamTaq Green DNA Polymerase	Thermo Fisher Scientific Baltics, Vilnius, Lithuania	EP0712
Ethylenediamine tetraacetic acid (EDTA)	Serva Electrophoresis GmbH, Heidelberg, Germany	11280
Ethanol absolute for analysis	Merck KGaA, Darmstadt, Germany	1.00983.1011
Ethidium bromide 1 %	Serva Electrophoresis GmbH, Heidelberg	21251
Fibrinogen from human plasma	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	F3879-1G
Formic acid	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	F0507
Gassner agar	Sifin Diagnostic GmbH, Berlin, Germany	TN1194

Chemicals	Supplier	Article Number
GenElute ™ Bacterial Genomics DNA Kit	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	NA2110-1KT, NA2100
GeneRuler 100 bp plus DNA Ladder	Thermo Fisher Scientific Baltics, Vilnius, Lithuania	SM0322
Glycerol	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	8.18709.1000
Gram's crystal violet solution	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	94448-2.5L-F
Hydrochloric acid (HCl)	Merck KGaA, Darmstadt, Germany	1.00319.2511
Isopropanol	Merck KGaA, Darmstadt, Germany	1.09634.2511
Lysozyme from chicken egg white	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	L-6876
Lugol's solution (diluted iodine-potassium iodine solution)	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	1.09261.1000
Methanol	Merck KGaA, Darmstadt, Germany	1060096025
Micronaut H-Medium	Sifin Diagnostic GmbH, Berlin, Germany	M/E2-311-100
Mutanolysin	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	M9901
Natrium acetate	Carl Roth GmbH & Co KG, Karlsruhe, Germany	8779
Natrium chloride (NaCl)	Merck KGaA, Darmstadt, Germany	1.06400.5000
Natrium hydroxide (NaOH)	Carl Roth GmbH & Co KG, Karlsruhe, Germany	6771.1
Nucleotides (dNTP)	Rapidozym, Berlin, Germany	GEN-009-250
Phenol	Carl Roth GmbH & Co KG, Karlsruhe, Germany	38.2
Phenylmethyl sulfonyl flouride (PMSF)	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	P7626
Proteinase K	Sigma-Aldrich Chemie B.V., Zwijndrecht, NL	P6556
Pulse Field Certified Agarose	BIO-RAD Laboratories GmbH, München	1620137
Pulsed Field Agarose Biozym Gold Agarose	Biozym Scientific GmbH, Hessisch Oldendorf, Germany	850152
Restriction enzyme: *Apa*I	Thermo Fisher Scientific, Waltham, United States	ER1411
Restriction enzyme: *Sma*I	Thermo Fisher Scientific, Waltham, United States	ER0661
Restriction enzyme: *Xba*I	Thermo Fisher Scientific, Waltham, United States	ER0681
RNAse, DNAse-free from bovine pancreas (500 µg, 1 ml)	Roche Diagnostics GmbH, Mannheim, Germany	11119915001
Sheep blood defibrinated	Thermo Scientific, Landsmeer, the Netherlands	SR0051E
Sodium dodecyl sulfate (SDS)	Carl Roth GmbH & Co KG, Karlsruhe, Germany	2326.2
Standard solvent (OS) for MALDI-TOF	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	19182
Streptococcus selective supplement	Oxoid LTD, Hampshire, England	SR0126E
Tango Buffer (10X)	Thermo Fisher Scientific, Waltham, United States	BY5

Chemicals	Supplier	Article Number
Thiourea	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	88810
Todd Hewitt Broth	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	T1438
Tris base	Carl Roth GmbH & Co KG, Karlsruhe, Germany	4855.2
Yeast extract	Merck KGaA, Darmstadt, Germany	1.03753.0500
Watter, DNAse free	Sigma-Aldrich Chemie GmbH, Steinheim, Germany	W4502
Ziehl-Neelsen carbol fuchsin solution	Merck KGaA, Darmstadt, Germany	1.09215.2500

Table 24. Consumables used in this study

Article	Supplier	Article number
8-strip optical clear flat caps	Sarstedt AG & Co., Nümbrecht, Gemany	651998400
96 PCR plates, half skirt flat	Sarstedt AG & Co., Nümbrecht, Gemany	721979102
Cellstar tubes (50 ml, PP, conical bottom, sterile, single use)	Greiner Bio-One GmbH, Frickenhausen, Germany	227261
Cotton buds (sterile)	Applimed SA, Châtel-Saint-Denis, via MAGV GmbH, Rabenau-Londorf	1102257
Cuvettes, 1.5 ml semimicro (single use)	MAGV, Rabenau-Londorf, Germany	759015
Erlenmeyers, 250 ml, 500 ml, 1,000 ml, 2,000 ml	Duran GmbH, Duisburg-Hochemmerich, Germany	Various
Gloves nitrile	VWR International GmbH, Darmstadt, Germany	112-4195
Inoculation loops, 1 µl, 200 mm, (single use)	Greiner Bio-One GmbH, Frickenhausen, Germany	731101
Inoculation loops,10 µl (single use)	Greiner Bio-One GmbH, Frickenhausen, Germany	731170
Lids ultra low profile, clear, 127/85 mm,	Greiner Bio-One GmbH, Frickenhausen, Germany	691101
Microtitration test plates, Micronaut-S for large animals 4	Merlin Diagnostika GmbH, Bornheim-Hersel, Germany	E1-150-100
Microtitration plates, Rotilabo®, F-profile	Carl Roth GmbH & Co KG, Karlsruhe, Germany	9293.1
Nunc cryotubes, 1 ml	Thermo Fisher Scientific, Waltham, United States	374081
Object glass, 26x76 mm	MAGV, Rabenau-Londorf, Germany	190000001
Parafilm	MAGV, Rabenau-Londorf, Germany	31011
Petri dishes	Greiner Bio-One GmbH, Frickenhausen, Germany	633180
Pipette tips, 10 µl, 200 µl, 1,000 µl	Sarstedt AG & Co., Nümbrecht, Gemany	701116, 70.3050.020
Pipette tips, 200 µl with filter	Sarstedt AG & Co., Nümbrecht, Gemany	70.762.211
Pipette tips, 0-300 µl	Nerbe plus GmbH & Co.KG, Winsen/Luhe, Germany	06-367-2018

Article	Supplier	Article number
Plating spatula (single use)	Sarstedt AG & Co., Nümbrecht, Gemany	861569005
Plug molds, CHEF (disposable)	BIO-RAD Laboratories GmbH, München, Germany	170-3713
Semipermeable sealing membrane, Breathe-Easy®	Diversified Biotech via Sigma Aldrich	Z380059-1PAK
Toothpicks	Carl Roth GmbH & Co KG, Karlsruhe, Germany	EC48.1
Tubes, 1.5 ml (safe lock, single use)	Eppendorf AG, Hamburg, via MAGV GmbH, Rabenau-Londorf, Germany	30120086
Tubes, 2.0 ml (safe lock, single use)	Eppendorf AG, Hamburg, via MAGV GmbH, Rabenau-Londorf, Germany	296920094
Tubes, 1.5 ml (single use)	Sarstedt SG & Co., Nümbrecht, Gemany	72690550
Tubes 2.0 ml (single use)	Sarstedt AG & Co., Nümbrecht, Gemany	72691
Wipes, Kimtech Science delicate task wipes	Kimberly-Clark via MAGV, Rabenau-Londorf, Germany	7558
Wipes, Kimtech Science precision wipes	Kimberly-Clark via MAGV, Rabenau-Londorf, Germany	7552

Table 25. Devices used in this study

Device, Model	Supplier
Air Luminar Chamber, Heraeus LaminAir HB 2448 S	Heraeus Instruments, Hanau, Germany
Centrifuge, 5804R	Eppendorf, Hamburg, Germany
Centrifuge, Micro Star 17	VWR International GmbH, Darmstadt, Germany
DNA Analyzer, 3730XL	Applied Biosystem, Waltham, USA
Electrophoresis chamber	HYBAID GmbH, Heidelberg, Germany
Electrophoresis electric source, Pharmacia LKB GPS 200/400	Pharmacia, New Jersey, USA
Flame hood	Köttermann, Uetze-Hänigsen, Germany
Gel pouring stand and comb for PCR	HYBAID GmbH, Heidelberg, Germany
Gel pouring stand and comb for PFGE	BIO-RAD Laboratories GmbH, München, Germany
Horizontal shaker, Titramax 100	Heidolph Instruments GmbH & Co. KG, Schwabach, Germany
Incubator (aerobe)	Heraeus Instruments, Hanau, Germany
Incubator (CO_2)	BINDER, Tuttingen, Germany
MBT 7854 MSP library	Bruker Daltonics, Bremen, Germany
Micro Scout Plate 96 target polished steel BC	Bruker Daltonics, Bremen, Germany
Microflex LT mass spectrometer system V3.3.1.0	Bruker Daltonics, Bremen, Germany
Multichannel pipette electric, Biohit, 10-300 µl	Biohit Deutschland GmbH, Rosbach, Germany
Photometer, Multiskan™ FC Microplate Photometer	Thermo Fisher ScientificTM, Waltham, Massachusetts, USA
Pipette 8 channel, Picus electronic, 10-300 µl	Sartorius, Helsinki, Finland
Pipetts, Eppendorf Research, 0.5-10 µl, 2-20 µl, 10-100 µl, 100-1000 µl	Eppendorf AG, Hamburg, Germany

Device, Model	Supplier
Plate shaker, Heidolph Tirtamax 100	Heidolph Instruments GmbH & CO.KG, Schwabach, Germany
Pulsed Field Electrophoresis System, Bio Rad CHEF Mapper XA	Bio-RadBio-Rad Laboratories GmbH, Feldkirchen, Germany
Spectrophotometer, Nanodrop 2000c	Thermo Fisher ScientificTM, Waltham, Massachusetts, USA
Stereo microscope, SZX2	Olympus, Hamburg, Germany
Thermocycler, Biometra Trio 48	Analytik Jena, Jena, Germany
Thermomixer, comfort 5355	Eppendorf, Hamburg, Germany
Transilluminator, Image documentation system E.A.S.Y B-1393-347C	Herolab Laborgeräte GmbH, Wiesloch, Germany
Vortex, Genie 2TM	Bender and Hobein GmbH, Zurich, Switzerland
Vortex, Minnishaker MS1	IKA-Werke GmbH & Co. KG, Staufen, Germany

S2. Overview of the Culture media, biologicals, and chemicals

Culture Media

5 % sheep blood agar (BAP)

Blood agar base	40 g
Defibrinated sheep blood	50 ml
Aqua demin.	Ad 1000 ml

Brain Heart Infusion (BHI) broth

Brain Heart Infusion (Oxoid)	37 g
Aqua demin.	Ad 1000 ml

Colistin oxalinic acid blood agar (COBA)

Columbia agar base granulated (Roth)	22 g
Defibrinated sheep blood	25 ml
Aqua demin.	Ad 500 ml
Streptococcus selective supplement (added with 1 ml ethanol, 1 ml sterile aqua dest.)	1 vial

Gassner agar

Gassner agar base (Sifin)	77 g
Aqua demin.	Ad 1000 ml

Luria-Bertani (LB) broth

NaCl	5 g
Yeast extract	5 g
Bacto-trypton	10 g
5 M NaOH	4 ml
Aqua demin.	Ad 1000 ml

LB-broth with 30 % glycerol

NaCl	5 g
Yeast extract	5 g
Bacto-trypton	10 g
5 M NaOH	4 ml
Glycerol	300 ml
Aqua demin.	Ad 1000 ml

Todd Hewitt broth (THB)

Todd Hewitt Broth	77 g
Aqua demin.	Ad 1000 ml

THB with 1 % fibrinogen

Todd Hewitt Broth	1.54 g
Aqua demin.	Ad 15 ml
Suspense with:	
Fibrinogen from human plasma	200 mg
0.9 % NaCl	5 ml
Total end volume	20 ml

Buffers and Solutions

0.1 % crystal violet

Crystal violet	0.5 g
Aqua demin.	Ad 500 ml

0.9 % NaCl

NaCl	0.9 g
Aqua demin.	Ad 100 ml

Electrophoresis working solution (1X TAE)

Electrophoresis stock solution	200 ml
Aqua demin.	10 l

Electrophoresis stock solution (10X TAE)

Tris base	242 g
Acetic acid	57.1 ml
0.5 M EDTA pH 8.0	100 ml
Aqua demin.	Ad 1,000 ml

Elution buffer (10 mM Tris-Cl, pH 8.5)

Tris base	0.12 g
DNAse free water	Ad 100 ml
Set pH to 8.5	

Ethylenediaminetetraacetic acid (EDTA) (0.5 M, pH 8.0)

EDTA	186.15 g
Aqua demin.	Ad 1,000 ml
Set pH to 8.00	

Lysis buffer (for macrorestriction analysis)

5 M NaOH	20 ml
1 M Tris-HCl pH 8.0	0.6 ml

0.5 EDTA pH 8.0	20 ml
Brij 35	0.5 g
N-lauroylsarcosin	0.5 g
Na-desoxycholat	0.2 g
Aqua demin.	Ad 100 ml

Phenylmethyl sulfonyl fluoride (PMSF) solution

Phenylmethyl sulfonyl fluoride	0.17419 g
Isopropanol	Ad 10 ml

Sodium dodecyl sulfate (SDS) (10 %)

SDS	100 g
Aqua demin.	Ad 1,000 ml

TBE buffer (10X)

Tris-base	107.815 g
Boric acid	55.03 g
0.5 M EDTA pH 8.0	40 ml
Aqua demin.	Ad 1000 ml

TBE buffer (0.5X) with thiourea

TBE buffer (10X)	125 ml
Thiourea	15 mg
Aqua demin.	Ad 2,500 ml

TE buffer

1 M Tris HCl pH 8.0	10 ml
0.5 M EDTA, pH 8.0	2 ml
Aqua demin.	Ad 1,000 ml
Set pH to 8.00	

TES buffer

1 M Tris-HCl pH 8.0	5 ml
0.5 M EDTA pH 8.0	1 ml
5 M NaCl	1 ml
Aqua demin.	Ad 100 ml

Tris-HCl (1 M, pH 8.0)

Tris base	121.1 g
Aqua demin.	1,000 ml
Set pH to 8.00	

Na-acetate (3 M, pH 5.2)

Na-acetate	24.61 g
Aqua demin.	100 ml
Set pH to 5.2	

NaCl (5 M)

NaCl	292.2 g
Aqua demin.	Ad 1,000 ml

NaOH (5 M)

NaOH	200 g
Aqua demin.	Ad 1,000 ml