Version of Record: https://www.sciencedirect.com/science/article/pii/S0378113519312726
Manuscript_de983cbaeb48d65a7caecdd1c483d5f1

Overview of spatio-temporal distribution inferred by multi-locus sequence typing of *Taylorella equigenitalis* isolated worldwide from 1977 to 2018 in equidae

Fabien Duquesne [a]*, Aurélie Merlin [a], Iratxe Pérez-Cobo [b], Kamil Sedlák [c], Falk Melzer [d], Gudrun Overesch [e], David Fretin [f], Wojciech Iwaniak [g], Marie-France Breuil [a], Ulrich Wernery [h], Jessica Hicks [i], Montserrat Agüero-García [b], Nieves Frías-Serrano [k], Elena San Miguel-Ibáñez [b], Eva Patrasová [c], Andreas S. Waldvogel [j], Krzysztof Szulowski [g], Marina Joseph [h], John Jeeba [h], Jose Shanty [h], Preethamol Varghese [h], Aymeric Hans [a], Sandrine Petry [a]

[a] ANSES, Laboratory for Animal Health, Physiopathology and Epidemiology of Equine Diseases Unit, Goustranville, France

[b] Laboratorio Central de Veterinaria-Sanidad Animal, Algete, Madrid, Spain

[c] State Veterinary Institute Prague, Prague, Czech Republic

[d] Friedrich-Loeffler-Institut, Institute of Bacterial Infections and Zoonoses, Jena, Germany

[e] Institute of Veterinary Bacteriology, University of Bern, Längassstrasse 122, Bern, Switzerland

[f] Veterinary and Agrochemical Research Center, Bacterial Zoonoses of livestock Unit, Operational Direction Bacterial Diseases, Brussels, Belgium

[g] National Veterinary Research Institute, Department of Microbiology, Pulawy, Poland

[h] Central Veterinary Research Laboratory, Dubai, United Arab Emirates

[i] Diagnostic Bacteriology and Pathology Laboratory, National Veterinary Services Laboratories, USDA, Ames, United States of America

[j] Alte Bernstrasse 5, Detligen, Switzerland

[k] TRAGSATEC. Tecnologías y servicios agrarios SA, Madrid, Spain

© 2020 published by Elsevier. This manuscript is made available under the CC BY NC user license
https://creativecommons.org/licenses/by-nc/4.0/

* Corresponding author. Tel.: +33 2 31 79 22 76. / fabien.duquesne@anses.fr (F. Duquesne).

Abstract

The accurate identification of *Taylorella equigenitalis* strains is essential to improve worldwide prevention and control strategies for contagious equine metritis (CEM). This study compared 367 worldwide equine strains using multilocus sequence typing according to the geographical origin, isolation year and equine breed. The strains were divided into 49 sequence types (STs), including 10 described for the first time. Three major and three minor clonal complexes (CCs), and 11 singletons, were identified. The genetic heterogeneity was low (0.13 STs/strain) despite the wide diversity of geographical origins (n=16), isolation years (1977 to 2018) and equine breeds (n=18). It was highest outside Europe and in the 1977-1997 period; current major STs and CCs already existed before 1998. Previous data associated the major CC1 with the first CEM outbreaks in 1977-1978 in the United Kingdom, Australia and the United States, and revealed its circulation in France. Our study confirms its circulation in France over a longer period of time (1992-2018) and its distribution in Spain and Germany but not throughout Europe. In addition to CC1, relationships between non-European and European countries were observed only through ST4, ST17 and ST30. Within Europe, several STs emerged with cross-border circulation, in particular ST16 and ST46 from the major complexes CC2 and CC8. These results constitute a baseline for monitoring the spread of CEM outbreaks. A retrospective analysis of a higher number of strains isolated worldwide between 1977 and the early 2000s would be helpful to obtain an exhaustive picture of the original CEM situation.

Keywords: infectious equine disease; contagious equine metritis; *Taylorella equigenitalis*; MLST.

Introduction

Contagious equine metritis (CEM) is a venereal disease of equids caused by *Taylorella equigenitalis*, a slow-growing Gram-negative, non-mobile capnophilic and pleiomorphic coccobacillus. The acute form of the disease is characterised by mucopurulent vaginal discharge and variable degrees of vaginitis, endometritis and cervicitis, leading to temporary infertility. All infected stallions and a variable proportion of infected mares are asymptomatic carriers (Sugimoto et al., 1983; Timoney, 2011). Because of its contagiousness and consequential economic losses, CEM was at the origin of the Horserace Betting Levy Board's Code of Practices to prevent and control specific diseases in all horse and pony breeds (Allen and Wilsher, 2018). CEM must be reported to the World Organisation for Animal Health (OIE) and is part of veterinary certification for international trade purposes. However, the CEM disease status of many countries remains unknown due to the absence of monitoring and import programmes.

CEM was first reported in 1977 in the United Kingdom and Ireland among Thoroughbred horses (Crowhurst, 1977; Timoney et al., 1977), but is currently a worldwide concern in various equine breeds (Jeoung et al., 2016; Schulman et al., 2013) with the hypothesis that the episodic "source of contagion" is often mainland Europe (Schulman et al., 2013). Pulsed-field gel electrophoresis (Aalsburg and Erdman, 2011; Sting et al., 2016) and several other molecular typing tools including field inversion gel electrophoresis (Bleumink-Pluym et al., 1990), chromosomal DNA fingerprinting (Thoresen et al., 1995), crossed-field gel electrophoresis (Miyazawa et al., 1995) and more recently repetitive extragenic palindromic PCR (Sting et al., 2016) have been used to genotype CEM isolates. However, these molecular epidemiological tools are not very portable and inter-laboratory results are difficult to compare (Maiden et al., 1998), making them ill-suited for global epidemiological studies of

CEM outbreaks. International equine trade stakeholders, however, need to identify and track the global spread of *T. equigenitalis*. With this in mind and starting by whole genome sequencing data for both *Taylorella* species (Hébert et al., 2011; Hébert et al., 2012), a *Taylorella* multilocus sequence typing (MLST) scheme was developed based on the nucleotide sequences of seven housekeeping genes (*glt*A, *gyr*B, *fh*, *shmt*, *tyr*B, *adk* and *txn*) (Duquesne et al., 2013). This scheme provides unambiguous results directly comparable between laboratories (Enright et al., 2000). The accumulation of nucleotide changes in housekeeping genes is a relatively slow process and their allelic profile is sufficiently stable over time to be used for global epidemiology purposes (Enright and Spratt, 1999).

Distinct sequence types (ST) reported by recent publications using the *Taylorella* MLST scheme show that new *T. equigenitalis* strains are constantly emerging (Hwang and Cho, 2018; Melzer et al., 2018). Some of these strains showed no genetic links with other identified STs according to the default clonal complex definition when single locus variants (SLV) were considered. This is the case, for example, of the inland South Korean strain KITE-1 isolated in 2016 (Hwang and Cho, 2018). The aim of the present study was to assess the distribution of specific *T. equigenitalis* genotypes from a large population of *T. equigenitalis* strains mainly originating from several European countries using MLST. Based on the resulting MLST data, the genetic diversity and phylogeny of 367 *T. equigenitalis* strains were analysed and compared on the basis of different factors such as the geographical origin, year of isolation and equine breed.

Materials and methods

Isolate collection

The study was conducted with the European Union Reference Laboratory for equine diseases other than African horse sickness (ANSES, Laboratory for Animal Health, France), the National Reference Laboratories for CEM from Belgium, Czech Republic, France, Germany, Poland, Spain and Switzerland and two laboratories outside Europe: the Central Veterinary Research Laboratory in Dubai, United Arab Emirates, and the National Veterinary Services Laboratories in Ames, United States.

A total of 367 *T. equigenitalis* strains isolated from horses, two ponies and two donkeys, were included in this study: 119 strains were obtained and characterised using MLST; the MLST sequence data for 242 strains were downloaded from the online *Taylorella* MLST database (https://pubmlst.org/taylorella/), and the MLST sequence data for six strains were obtained from Hicks et al. (Hicks et al., 2018) (n = 4) and Delerue et al. (Delerue et al., 2019) (n = 2).

The strains were isolated over a 42-year period between 1977 and 2018 (93.2%) or at an unknown date (6.8%). Most of them were isolated in Europe (81.8%), including Austria, Belgium, the Czech Republic, France, Germany, Poland, Spain, Switzerland, the Netherlands and the United Kingdom. The remaining strains were isolated in Australia, Japan, South Africa, South Korea, the United Arab Emirates and the United States (17.4%) or the geographical location was unknown (0.8%) (Table 1). At least 18 different equine breeds were represented (Table 2), which were grouped into 12 categories (Additional file 1) to perform phylogenetic and goeBURST analysis. The Draught horse category grouped Comtois Draught and Trait Breton breeds, while the Warmblood horse category grouped Belgian Warmblood/Warmblood, Hanoverian, Oldenburg, Westfalen and Zangersheide breeds. Detailed information on the geographical origin, year of isolation, clinical signs and equine

breed as well as the MLST sequence data relating to each strain are reported in Additional file 1.

Bacterial growth and DNA extraction

The 119 *T. equigenitalis* strains typed in this study were isolated and identified at the institutions of origin using the gold standard culture method according to OIE or national culturing instructions applicable at the time of isolation. Strains were subcultured onto chocolate agar at (37±2) °C with 5-10% (v/v) CO_2 in air prior to genomic DNA extraction and purification using commercially available DNA extraction kits according to the manufacturer's protocol.

MLST

The seven MLST loci (*glt*A, *gyr*B, *fh*, *shmt*, *tyr*B, *adk* and *txn*) were amplified and sequenced according to previously published experimental conditions (Duquesne et al., 2013). The *Taylorella* MLST database (https://pubmlst.org/taylorella/) was used to assign allele numbers and STs. Novel alleles and STs were allocated by the database curator. All the sequence data for the 119 strains typed in this study are available from the *Taylorella* MLST database.

Data analysis

The sequence data obtained from the online *Taylorella* MLST database and from the literature (Delerue et al., 2019; Hicks et al., 2018) were incorporated into the dataset. The sequences of every individual locus were trimmed to equivalent lengths, like in the *Taylorella* MLST database.

145 Sequences were aligned and a phylogenetic tree drawn up using MEGA software version 7.0
146 (Kumar et al., 2016). For each unique ST, sequences of the seven MLST loci were
147 concatenated, giving an in-frame sequence of 3,521 bp (not including nucleotide
148 insertions/deletions). The phylogenetic tree was based on the concatenated sequences using
149 the maximum likelihood method based on the Tamura-Nei model (Tamura and Nei, 1993).
150 Support for internal nodes was estimated using the nonparametric bootstrap method with
151 1,000 replications. The tree was rooted with the ST28-concatenated sequences of the
152 *Taylorella asinigenitalis* CIP 107673^{T} reference strain as the outgroup, extracted from the
153 *Taylorella* MLST database.
154 Relatedness between STs was analysed based on allelic profiles using eBURST software
155 version 3 (Feil et al., 2004; Spratt et al., 2004). The BURST algorithm, based on a set of
156 hierarchical rules relating to the number of single-locus variants (SLV) and double-locus
157 variants (DLV) involved, identifies mutually exclusive groups named clonal complexes (CCs)
158 of related STs in the population, and attempts to identify the primary and subgroup founding
159 ST for each CC. A CC was defined as major when a primary founding ST was predicted. All
160 others were considered minor.
161 Global optimal goeBURST (Francisco et al., 2009) diagrams were drawn up using PHYLOViZ
162 software version 2 (Francisco et al., 2012) to analyse relatedness between STs and
163 geographical origins or equine breeds.
164
165 **Results**
166 The MLST analysis of the 367 *T. equigenitalis* strains included in this study divided the strains
167 into 49 genotypes or STs, i.e. 0.13 STs/strain (Additional file 1). Thirty-one STs were
168 associated with a single strain per ST, and the other 18 STs were associated with between

two and 67 strains per ST. ST4 (n = 67), ST33 (n = 64), ST16 (n = 49) and ST1 (n = 41) were the most prevalent STs and concerned 60% of strains. MLST data for the 119 *T. equigenitalis* strains typed in this study allowed us to identify three new alleles (*fh*-25, *fh*-26 and *shmt*-15) and 10 new STs (ST32, ST36, ST53, ST54, ST59 and ST61 to ST65) never described before (Additional file 1).

The eBURST analysis (Additional file 2) grouped the 49 STs into three major CCs (CC1, CC2 and CC8), three minor CCs (CC3, CC9 and CC10) and 11 singletons at the SLV threshold. CC1 was the largest CC, including 17 STs with the predicted primary founder ST1 and both predicted subgroup founders ST2 and ST33. CC2 and CC8 each included seven STs with the predicted primary founders ST17 and ST46 respectively. Minor complexes CC3, CC9 and CC10 included between two and three STs and no primary founders were predicted. The discovery of 10 new STs led to the emergence of the new CC10, and the identification of CC8 and CC2 as major CCs (last one following its merger with the former CC4 via ST66). At the DLV threshold, all the CCs and eight singletons previously described using the SLV threshold were linked; only ST9, ST31 and ST50 remained genetically independent (Additional file 2).

A phylogenetic tree (Fig. 1) drawn up using concatenated sequences of seven MLST loci did not reveal any major clade within the 49 unique STs. However, the distribution of STs within the six CCs reflects a correlation between the genetic distances and evolutionary relationships among the closely-related STs. The 11 singletons were distributed throughout the tree except on the upper part composed by CC1, and several of them were clustered along with CC2 and CC8. The 10 new STs were distributed throughout the tree, but 50% clustered in CC8.

Three hundred and twenty-nine *T. equigenitalis* strains out of the 367 included in this study were isolated over a 42-year period, but the strain distribution over time was heterogeneous

193 (Additional file 1): 32 strains were isolated from 1977 to 1997 versus 310 strains between
194 1998 and 2018. However, the ST diversity was greater during the 1977-1997 period (0.34
195 STs/strain) than during the 1998-2018 period (0.12 STs/strain). Twelve STs, grouped into the
196 current CC1, CC2, CC8 and CC10 complexes plus three singletons, had already been
197 identified before the 1998-2018 period; they included the four most currently prevalent, i.e.
198 ST1, ST4, ST16 and ST33.
199 The 64 non-European and 300 European *T. equigenitalis* strains out of the 367 strains
200 included in this study were grouped into 12 and 40 STs respectively (Table 1, Additional file
201 1). The goeBURST analysis (Fig. 2) allowed us to connect non-European and European
202 countries through ST1, ST4, ST17 and ST30. ST19 and ST20 were only associated with non-
203 European strains but were nonetheless related to CC1, which includes European strains.
204 Singletons ST3, ST9, ST31, ST42, ST55 and ST58 were associated only with non-European
205 strains whereas CC3, CC8, CC9, CC10 and singletons ST5 and ST50 were associated only with
206 European strains. Interestingly, (Fig. 2a) Japan (ST3 and ST9) and South Korea (ST55) had no
207 connections with any other countries, whereas Australia, the United States and Europe were
208 connected through CC1, the United Arab Emirates and Europe through CC2 and ST30, and
209 South Africa and Europe through ST4. Within Europe (Fig. 2b), the Czech Republic showed
210 the least connection with other European countries (only with Germany though CC10)
211 whereas Germany and France showed the most connections with other European countries.
212 While the French strains were the most represented in this study (30.8%), unlike the German
213 strains that only represented 5.4%, they are less genetically heterogeneous (0.21 STs/strain)
214 than German strains (0.60 STs/strain). A focus on the most prevalent STs (Fig. 2) suggested
215 that ST16 (CC2) may circulate within Europe, whereas ST1 (CC1) may circulate worldwide,
216 and ST33 (France and Spain) and ST4 (Austria, France and unknown) appear to be more

specific to a small number of countries. With fewer strains, ST46 (CC8) showed the same country distribution as ST16 without any Polish strains, and ST17 may be a European ST recently linked to Emirati strains.

Despite unknown equine breeds representing 35.5% of the 367 strains included in this study and the large number of equine breeds represented (Table 2, Additional file 1), we investigated the genetic relationships between the STs and equine breeds or category of equine breeds (Fig. 3). Singletons were mostly specific to an equine breed, except ST4 which was associated with both South African strains from Thoroughbred and Warmblood horses, and South African, Austrian and French strains from Lipizzaner horses. The largest complex, CC1, was associated with up to eight different equine breeds. Details on the seven prevalent equine breeds (Fig. 3) showed that except the Trotter from ST16, Draught and Trotter horses were clustered within CC1 and CC3, both closely located on the phylogenetic tree (Fig. 1). Thoroughbred, Warmblood and Lipizzaner horses were more dispersed through the goeBURST representation, but it may be noted that STs related to Lipizzaners were close to each other on the phylogenetic tree (Fig. 1). The Kladruber (ST56) and Pure Spanish (CC1) horses, on the other hand, were not only specific to a ST or CC, but were also country-specific (from the Czech Republic and Spain respectively).

Discussion

T. equigenitalis is the causative agent of CEM, a contagious equine disease leading to substantial economic losses. It is essential to accurately identify the strain for both epidemiological surveillance and to improve worldwide CEM prevention and control strategies. CEM was first reported in the United Kingdom and Ireland among Thoroughbred horses in 1977 (Crowhurst, 1977; Timoney et al., 1977). The disease spread rapidly from

1977 to 1978 in Europe (including Belgium, France and Germany), Australia and the United States (1978) (Matsuda and Moore, 2003). CEM was then reported in Italy and Japan in 1980 (Matsuda and Moore, 2003) and its presence was finally described in at least 30 countries worldwide in various equine breeds (Schulman et al., 2013), the most recent cases in the literature being in Croatia in 2014 or 2015 (Štritof et al., 2017) and South Korea in 2015 (Jeoung et al., 2016). Besides being the cradle of CEM emergence, the endemic status among the non-Thoroughbred horses of Europe supports the current hypothesis that the episodic "source of contagion" is often mainland Europe. MLST has already been implemented to examine the *T. equigenitalis* population structure. It revealed that the first CEM outbreaks in the United Kingdom, Australia and United States from 1977 to 1978 were associated with the founding complex, CC1 (Duquesne et al., 2013). In parallel, distinct genotypes emerged over time and in different countries (Japan, the United Arab Emirates, South Korea) suggesting the existence of an unidentified natural worldwide reservoir (Duquesne et al., 2013; Hwang and Cho, 2018). The present study focused on a large population of 367 *T. equigenitalis* strains to determine, through their MLST genotypes, their diversity and spatio-temporal distribution according to the geographical origin, year of isolation and equine breed. A hundred and nineteen strains were characterised using MLST, and the MLST sequence data of 248 additional strains were accessed from the online *Taylorella* MLST database and the literature (Delerue et al., 2019; Hicks et al., 2018). Despite the large selection of strains in terms of geographical origin, year of isolation and equine breed, one area of concern is the amount of emphasis put on the number of strains within the same ST. In this study, a *T. equigenitalis* strain was usually isolated from a CEM-positive animal reported in a single year. However, several strains with the same ST may be represented multiple times, subjected to repeated sampling (perhaps in relation to a treatment protocol)

or due to animals CEM-positive over a period of several years (Additional file 1); nevertheless, considering the known data on strains, this situation concerned only a limited number of strains and has not been considered a significant bias.

The MLST analysis performed in this study suggests little genetic heterogeneity, with only 49 STs characterised from the 367 *T. equigenitalis* strains (0.13 STs/strain), despite marked heterogeneity in terms of geographical origin (16 countries, including six outside Europe), year of isolation (1977 to 2018) and equine breed (n = 18) in the analysed strain collection. The genetic heterogeneity of *T. equigenitalis* was slightly higher outside Europe (0.19 STs/strain) than within Europe (0.13 STs/strain). More importantly, the genetic heterogeneity of *T. equigenitalis* was greater in the 1977-1997 period (0.34 STs/strain) than in the 1998-2018 period (0.12 STs/strain); it may be noted that the current major STs and CC had already been identified during the 1977-1997 period. The STs/strain value of 113 *T. equigenitalis* strains used to develop the *Taylorella* MLST was slightly higher (0.24 STs/strain) (Duquesne et al., 2013) than for the present study but probably corresponded to genotypic variability in the French strains, since 85% of the strains were from France. This is similar to our study, where the French data reported 0.21 STs/strains. A comparison of whole-genome sequencing reported by Hicks et al. (Hicks et al., 2018) showed diversity in genome size by up to 100,000 bp however, mainly due to repeat regions and four main variable regions. Moreover, the authors observed that *T. equigenitalis* was a more diverse species than the whole *Mycobacterium tuberculosis complex* by an SNP comparison (Hicks et al., 2018). The genomic diversity thus observed is not, however, noticeable here because of the MLST principle, being based on housekeeping genes.

Compared to previous MLST results, the present study confirms the circulation of the founding complex, CC1, in France over a long period of time (1992-2018) and confirms its

distribution in Spain (2015-2016) and Germany too, but not in the six other European countries investigated, i.e. Austria, Belgium, the Czech Republic, the Netherlands, Poland and Switzerland. In the 1970s-80s, and perhaps still today, CC1 was circulating worldwide yet it does not seem to have spread throughout Europe according to the dataset analysed. This observation should nevertheless be confirmed by a retrospective analysis of strains isolated from the first reported outbreaks in these European countries, if they are available.

Apart from the epidemiological relationships described through CC1, few epidemiological relationships between non-European and European countries were revealed in the present study. Thus, singleton ST4 was first found in Austria (1992-2007) and France (1993-2001) and then in South Africa (1996 to 2015), suggesting its spread from Europe to South Africa first within the Lipizzaner population then spreading to other breeds (Warmblood and Thoroughbred). Emirati strains were divided into four genotypes, three of which (ST17, ST30 and S42) had a DLV relationship and clustered on the phylogenetic tree, suggesting the presence of a common ancestral Emirati clone. The presence of this ancestral Emirati clone is, however, in contrast to the fact that ST17 (primary founder of CC2) was found in Europe before Emirati strains were isolated in 2012 (ST42), 2016-2017 (ST17) and 2018 (ST30).

Within Europe, ST16 included in CC2 has emerged as the dominant genotype with a cross-border circulation; it is composed of strains from Austria (2004), Belgium (2008-2013), France (2004-2012), Germany (1989-2009), Poland and Switzerland (2005-2008). ST46, the primary founder of CC8, seems to have the same profile as ST16 without any Polish strains, but it was found to be mainly composed of German strains plus the strain isolated in 2018 from the vaginal discharge of a French mare infected by cryopreserved stallion semen collected and processed in Germany in 2012 (Delerue et al., 2019). This Germanic origin could be extended to CC8, which is mainly composed of German strains first isolated in

1981. Interestingly, the genetic heterogeneity of German strains was much higher (0.60 STs/strain) than that of the entire dataset despite representing only 5.4% of the strain collection studied.

Conclusion

This study provides new insights into the molecular epidemiology of *T. equigenitalis* in Europe and constitutes a baseline for monitoring the spread of CEM outbreaks. However, it is difficult to give an exhaustive picture of the global CEM situation as there are very few *T. equigenitalis* strains genotyped by MLST from non-European countries and in the 10-15 years preceding the first CEM outbreaks reported in 1977. The genotyping of *T. equigenitalis* strains by MLST should be continued over an extended timeline, and a retrospective study carried out on a higher number of strains isolated from 1977 to the early 2000s. In the future, MLST data could be obtained using whole-genome sequencing, or a core genome MLST scheme could even be developed to the maximum resolution of the genealogy within the *T. equigenitalis* species.

Acknowledgements

We are very grateful to Isabelle Brodard from the Institute of Veterinary Bacteriology, University of Bern, Switzerland and to David Craig from the Emirates Racing Authority. We would also like to thank Delphine Libby-Claybrough, professional translator/editor and native English speaker, for her editorial assistance.

Financial support

This work was mostly supported by the European Commission (via the budget of the EURL for equine diseases other than African horse sickness).

Authors' contributions

Conception and design of the study and writing the article: FD and SP. Acquisition of data: FD, IPC, KS, FM, GO, DF, WI, MFB, UW, JK, MAG, NFS, ESMI, EP, AW, MJ, JJ, SJ and PV. Data analysis and interpretation: FD, AM and SP. Participation in the drafting and critical revision of the article: FD, AM, IPC, KS, WI, EP, AW, UW, DLC, AH and SP. All the authors read and approved the article.

Conflicts of interest

The authors declare that they have no conflicts of interest.

Abbreviations

CC, clonal complex; CEM, contagious equine metritis; DLV, double-locus variant; MLST, multilocus sequence typing; OIE, world organisation for animal health; SLV, single-locus variant; ST, sequence type.

355 **References**

356 Aalsburg, A.M., Erdman, M.M., 2011. Pulsed-field gel electrophoresis genotyping of Taylorella
357 equigenitalis isolates collected in the United States from 1978 to 2010. J Clin Microbiol 49,
358 829-833.

359 Allen, W.R., Wilsher, S., 2018. Half a century of equine reproduction research and application: A
360 veterinary tour de force. Equine Vet J 50, 10-21.

361 Bleumink-Pluym, N., ter Laak, E.A., van der Zeijst, B.A., 1990. Epidemiologic study of Taylorella
362 equigenitalis strains by field inversion gel electrophoresis of genomic restriction
363 endonuclease fragments. J Clin Microbiol 28, 2012-2016.

364 Crowhurst, R.C., 1977. Genital infection in mares. Vet Rec 100, 476.

365 Delerue, M., Breuil, M.F., Duquesne, F., Bayon-Auboyer, M.H., Amenna-Bernard, N., Petry, S., 2019.
366 Acute Endometritis due to Taylorella equigenitalis Transmission by Insemination of
367 Cryopreserved Stallion Semen. J Equine Vet Sci 78, 10-13.

368 Duquesne, F., Hebert, L., Breuil, M.F., Matsuda, M., Laugier, C., Petry, S., 2013. Development of a
369 single multi-locus sequence typing scheme for Taylorella equigenitalis and Taylorella
370 asinigenitalis. Vet Microbiol 167, 609-618.

371 Enright, M.C., Day, N.P., Davies, C.E., Peacock, S.J., Spratt, B.G., 2000. Multilocus sequence typing for
372 characterization of methicillin-resistant and methicillin-susceptible clones of Staphylococcus
373 aureus. J Clin Microbiol 38, 1008-1015.

374 Enright, M.C., Spratt, B.G., 1999. Multilocus sequence typing. Trends Microbiol 7, 482-487.

375 Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., Spratt, B.G., 2004. eBURST: inferring patterns of
376 evolutionary descent among clusters of related bacterial genotypes from multilocus
377 sequence typing data. J Bacteriol 186, 1518-1530.

378 Francisco, A.P., Bugalho, M., Ramirez, M., Carrico, J.A., 2009. Global optimal eBURST analysis of
379 multilocus typing data using a graphic matroid approach. BMC Bioinformatics 10, 152.

Francisco, A.P., Vaz, C., Monteiro, P.T., Melo-Cristino, J., Ramirez, M., Carrico, J.A., 2012. PHYLOViZ: phylogenetic inference and data visualization for sequence based typing methods. BMC Bioinformatics 13, 87.

Hébert, L., Moumen, B., Duquesne, F., Breuil, M.F., Laugier, C., Batto, J.M., Renault, P., Petry, S., 2011. Genome sequence of Taylorella equigenitalis MCE9, the causative agent of contagious equine metritis. J Bacteriol 193, 1785.

Hébert, L., Moumen, B., Pons, N., Duquesne, F., Breuil, M.F., Goux, D., Batto, J.M., Laugier, C., Renault, P., Petry, S., 2012. Genomic characterization of the Taylorella genus. PLoS One 7, e29953.

Hicks, J., Stuber, T., Lantz, K., Erdman, M., Robbe-Austerman, S., Huang, X., 2018. Genomic diversity of Taylorella equigenitalis introduced into the United States from 1978 to 2012. PLoS One 13, e0194253.

Hwang, J.Y., Cho, G.J., 2018. First identification of Taylorella equigenitalis from genital tracts of thoroughbred horses from the inland area of South Korea by multilocus sequence typing. Journal of Equine Veterinary Science 60, 16-22.

Jeoung, H.Y., Lee, K.E., Yang, S.J., Park, T., Lee, S.K., Lee, J.H., Kim, S.H., Kim, B., Kim, Y.J., Park, J.Y., 2016. First Isolation of Taylorella equigenitalis From Thoroughbred Horses in South Korea. Journal of Equine Veterinary Science 47, 42-46.

Kumar, S., Stecher, G., Tamura, K., 2016. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Mol Biol Evol 33, 1870-1874.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russell, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc Natl Acad Sci U S A 95, 3140-3145.

404 Matsuda, M., Moore, J.E., 2003. Recent advances in molecular epidemiology and detection of
405 Taylorella equigenitalis associated with contagious equine metritis (CEM). Vet Microbiol 97,
406 111-122.
407 Melzer, F., Rassbach, A., Koenig-Mozes, A., Elschner, M.C., Tomaso, H., Busch, A., 2018. Draft
408 Genome Sequence of Taylorella equigenitalis Strain 210217RC10635, Isolated from a Pony
409 Stallion in Germany. Microbiol Resour Announc 7.
410 Miyazawa, T., Matsuda, M., Isayama, Y., Samata, T., Ishida, Y., Ogawa, S., Takei, K., Honda, M.,
411 Kamada, M., 1995. Genotyping of isolates of Taylorella equigenitalis from thoroughbred
412 brood mares in Japan. Vet Res Commun 19, 265-271.
413 Schulman, M.L., May, C.E., Keys, B., Guthrie, A.J., 2013. Contagious equine metritis: artificial
414 reproduction changes the epidemiologic paradigm. Vet Microbiol 167, 2-8.
415 Spratt, B.G., Hanage, W.P., Li, B., Aanensen, D.M., Feil, E.J., 2004. Displaying the relatedness among
416 isolates of bacterial species -- the eBURST approach. FEMS Microbiol Lett 241, 129-134.
417 Sting, R., Seeh, C., Mauder, N., Maurer, M., Loncaric, I., Stessl, B., Kopp, P., Banzhaf, K., Martin, B.,
418 Melzer, F., Rassbach, A., Spergser, J., 2016. Genotyping of German and Austrian Taylorella
419 equigenitalis isolates using repetitive extragenic palindromic (REP) PCR and pulsed-field gel
420 electrophoresis (PFGE). Res Vet Sci 109, 101-106.
421 Štritof, Z., Habuš, J., Mojčec Perko, V., Majhut, M., Brkljača Bottegaro, N., Perharić, M., Hađina, S.,
422 Milas, Z., Turk, N., 2017. Detection of Taylorella equigenitalis and Taylorella asinigenitalis in
423 horses in Croatia as a result of small scale survey. 87, 535-541.
424 Sugimoto, C., Isayama, Y., Sakazak, R., Kuramochi, S., 1983. Transfer ofHaemophilus equigenitalis
425 Taylor et al. 1978 to the genusTaylorella gen. nov. asTaylorella equigenitalis comb. nov.
426 Current Microbiology 9, 155-162.
427 Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region
428 of mitochondrial DNA in humans and chimpanzees. Mol Biol Evol 10, 512-526.

Thoresen, S.I., Jenkins, A., Ask, E., 1995. Genetic homogeneity of Taylorella equigenitalis from Norwegian trotting horses revealed by chromosomal DNA fingerprinting. J Clin Microbiol 33, 233-234.

Timoney, P.J., 2011. Horse species symposium: contagious equine metritis: an insidious threat to the horse breeding industry in the United States. J Anim Sci 89, 1552-1560.

Timoney, P.J., Ward, J., Kelly, P., 1977. A contagious genital infection of mares. Vet Rec 101, 103.

436 **Table 1.** Countries of origin for the 367 *T. equigenitalis* strains included in this study.

Country of origin	Strain typed in this study	Strain extracted from the MLST database	Strain extracted from the literature *	Total number of strains
Australia	-	6	-	6
Austria	-	24	1	25
Belgium	1	16	-	17
Czech Republic	27	2	-	29
France	13	99	1	113
Germany	15	3	2	20
Japan	-	7	-	7
Netherlands	1	-	1	2
Poland	-	17	-	17
South Africa	-	34	-	34
South Korea	-	1	-	1
Spain	57	-	-	57
Switzerland	-	19	-	19
United Arab Emirates	5	5	-	10
United Kingdom	-	1	-	1
United States	-	5	1	6
Unknown	-	3	-	3
Total	119	242	6	367

437 *Delerue et al., 2019; Hicks et al., 2018

438 **Table 2.** Equine breeds for the 367 *T. equigenitalis* strains included in this study.

Equine breed	Strain typed in this study	Strain extracted from the MLST database	Strain extracted from the literature *	Total number of strains
Anglo-Arabian	2	-	-	2
Arabian	1	17	-	18
Baudet du Poitou	-	2	-	2
Belgian Warmblood / Warmblood	-	15	-	15
Comtois Draught Horse	-	7	-	7
French Saddlebred Horse	-	11	-	11
French Trotter / Trotter	-	44	-	44
Hanoverian	-	-	1	1
Kladruber horse	27	-	-	27
Lipizzaner	5	54	1	60
Oldenburg	-	-	1	1
Pony	-	2	-	2
Pure Spanish horse	34	-	-	34
Quarter Horse	-	1	-	1
Thoroughbred	7	7	1	15
Trait Breton	-	2	-	2
Westfalen riding horse	-	1	-	1
Zangersheide horse	-	1	-	1
Unknown	43	78	2	123
Total	119	242	6	368

439 *Delerue et al., 2019; Hicks et al., 2018

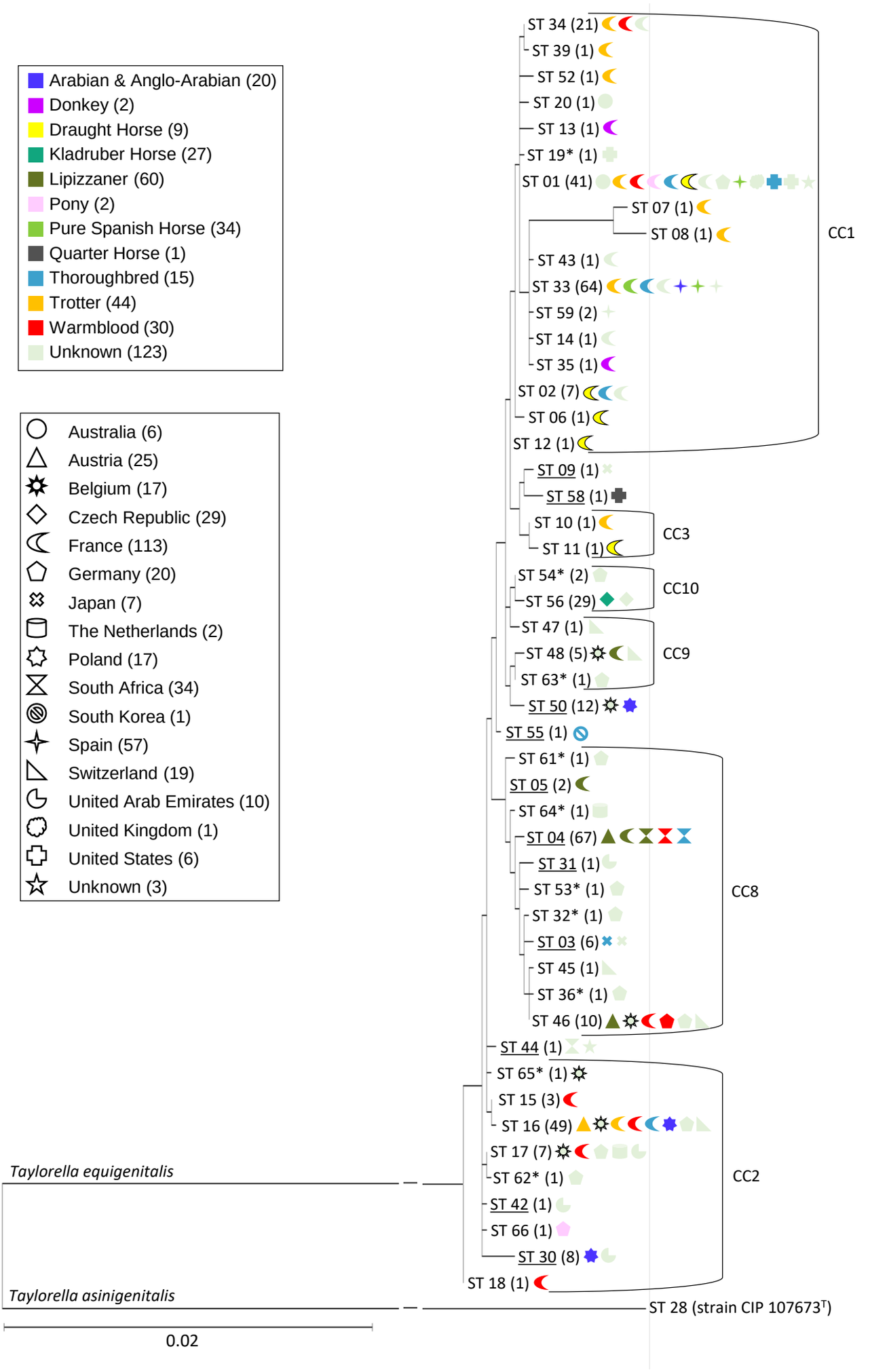
440

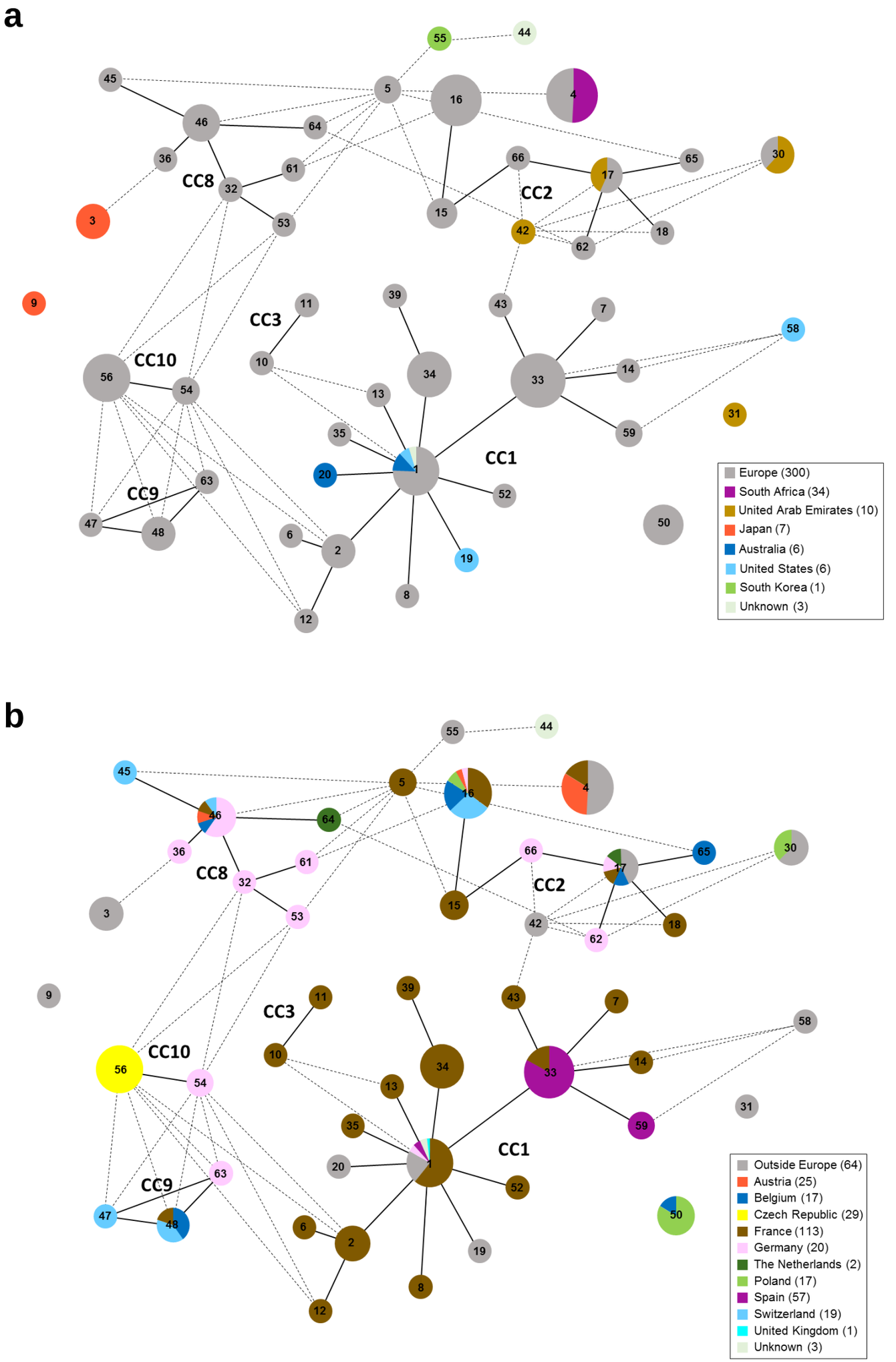
Figure 1. Phylogenetic tree based on 3,521-bp concatenated sequences of seven MLST housekeeping genes. The tree was constructed from the MLST data of 367 *T. equigenitalis* strains (49 STs) using the maximum likelihood method, and rooted with the *T. asinigenitalis* CIP 107673^T reference strain as the outgroup. The bootstrap values were calculated from 1,000 replications. The total number of strains associated with each ST is indicated in brackets. The equine breeds are indicated next to the STs using geometrical shapes, and the geographical origins correspond to the colours of those geometrical shapes (see the legend). The grouping of STs into CCs defined by eBURST analysis at the SLV threshold is indicated by braces, while singletons are underlined. STs identified for the first time in this study are identified by an asterisk. The numbers in brackets in the legend correspond to the number of strains.

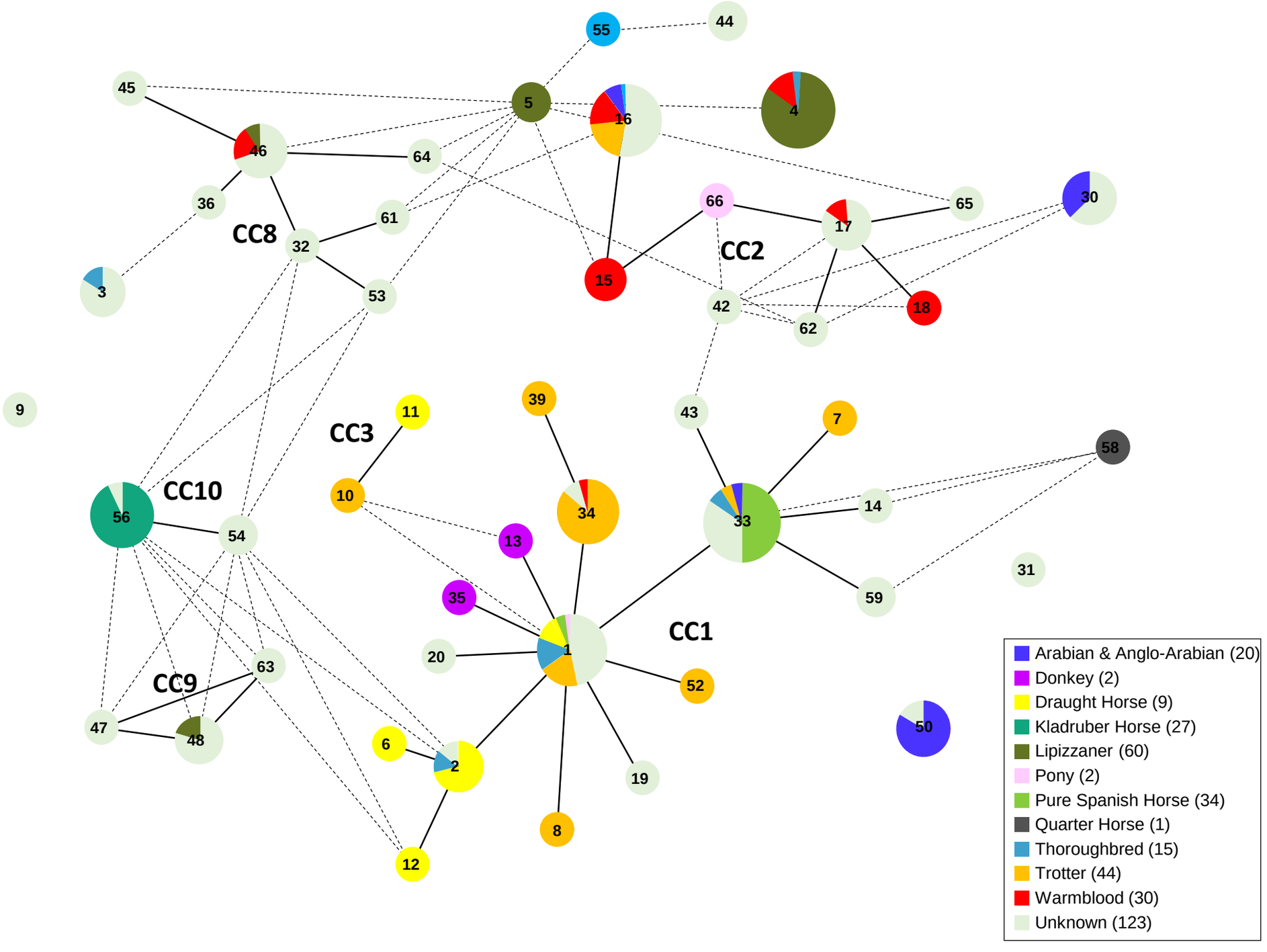
Figure 2. This population snapshot of *T. equigenitalis* displays genetic relationships between the STs and countries of isolation, with a focus on non-European (**a**) and European countries (**b**). It is based on a goeBURST (PHYLOViZ) analysis of country distribution among 367 *T. equigenitalis* strains characterised by MLST. STs with an SLV relationship were linked together by solid lines to form six CCs, and STs with a DLV relationship were linked together by dashed lines (the length of lines are not representative of the distances between STs). Dot colour fractions refer to countries of isolation associated with STs (see legend). The dot size is proportional to the prevalence of STs in the analysed collection. The numbers in brackets in the legend correspond to the number of strains.

Figure 3. This population snapshot of *T. equigenitalis* displays genetic relationships between the STs and equine breeds. It is based on a goeBURST (PHYLOViZ) analysis of equine breed

distribution among 367 *T. equigenitalis* strains characterised by MLST. STs with an SLV relationship were linked together by solid lines to form six CCs, and STs with a DLV relationship were linked together by dashed lines (the length of lines are not representative of the distances between STs). Dot colour fractions refer to equine breeds associated with STs (see legend). The dot size is proportional to the prevalence of STs in the analysed collection. The numbers in brackets in the legend correspond to the number of strains.


Arabian & Anglo-Arabian (20)
Donkey (2)
Draught Horse (9)
Kladruber Horse (27)
Lipizzaner (60)
Pony (2)
Pure Spanish Horse (34)
Quarter Horse (1)
Thoroughbred (15)
Trotter (44)
Warmblood (30)
Unknown (123)
Australia (6)
Austria (25)
Belgium (17)
Czech Republic (29)
France (113)
Germany (20)
Japan (7)
The Netherlands (2)
Poland (17)
South Africa (34)
South Korea (1)
Spain (57)
Switzerland (19)
United Arab Emirates (10)
United Kingdom (1)
United States (6)
Unknown (3)
ST 34 (21)
ST 39 (1)
ST 52 (1)
ST 20 (1)
ST 13 (1)
ST 19* (1)
ST 01 (41)
ST 07 (1)
ST 08 (1)
CC1
ST 43 (1)
ST 33 (64)
ST 59 (2)
ST 14 (1)
ST 35 (1)
ST 02 (7)
ST 06 (1)
ST 12 (1)
ST 09 (1)
ST 58 (1)
ST 10 (1)
ST 11 (1)
CC3
ST 54* (2)
ST 56 (29)
CC10
ST 47 (1)
ST 48 (5)
ST 63* (1)
CC9
ST 50 (12)
ST 55 (1)
ST 61* (1)
ST 05 (2)
ST 64* (1)
ST 04 (67)
ST 31 (1)
ST 53* (1)
ST 32* (1)
ST 03 (6)
ST 45 (1)
ST 36* (1)
ST 46 (10)
CC8
ST 44 (1)
ST 65* (1)
ST 15 (3)
ST 16 (49)
ST 17 (7)
ST 62* (1)
ST 42 (1)
ST 66 (1)
ST 30 (8)
ST 18 (1)
CC2
Taylorella equigenitalis
Taylorella asinigenitalis
ST 28 (strain CIP 107673T)
0.02


a
CC1
CC2
CC3
CC8
CC9
CC10
Europe (300)
South Africa (34)
United Arab Emirates (10)
Japan (7)
Australia (6)
United States (6)
South Korea (1)
Unknown (3)
b
CC1
CC2
CC3
CC8
CC9
CC10
Outside Europe (64)
Austria (25)
Belgium (17)
Czech Republic (29)
France (113)
Germany (20)
The Netherlands (2)
Poland (17)
Spain (57)
Switzerland (19)
United Kingdom (1)
Unknown (3)


CC1
CC2
CC3
CC8
CC9
CC10
Arabian & Anglo-Arabian (20)
Donkey (2)
Draught Horse (9)
Kladruber Horse (27)
Lipizzaner (60)
Pony (2)
Pure Spanish Horse (34)
Quarter Horse (1)
Thoroughbred (15)
Trotter (44)
Warmblood (30)
Unknown (123)