1 Genetic characterization of *Streptococcus equi* subspecies *zooepidemicus* associated
2 with high swine mortality in United States

3 Xuhua Chen[§], Nubia Resende-De-Macedo[§], Panchan Sitthicharoenchai, Orhan Sahin, Eric
 4 Burrough, Maria Clavijo, Rachel Derscheid, Kent Schwartz, Kristina Lantz, Suelee
 5 Robbe-Austerman, Rodger Main, Ganwu Li*

6
 7 Department of Veterinary Diagnostic and Production Animal Medicine, College of
 8 Veterinary Medicine, Iowa State University, 1800 Christensen Drive, Ames, IA 50011,
 9 USA (X. Chen, N. Resende-De-Macedo, P. Sitthicharoenchai, O. Sahin, M. Clavijo, R.
 10 Derscheid, K. Schwartz, E. Burrough, R. Main, G. Li); National Veterinary Services
 11 Laboratories, 1920 Dayton Ave, Ames, IA 50010 (K. Lantz, S. Robbe-Austerman)

12
 13 [§]These authors contribute to this work equally

14
 15 *Address for corresponding: Ganwu Li, email: liganwu@iastate.edu

16 Department of Veterinary Diagnostic and Production Animal Medicine, College of
 17 Veterinary Medicine, Iowa State University, Ames, IA 50011, USA.

18
 19 Running title: *S. zooepidemicus* associated with high mortality in pigs

20

Abstract

High mortality events due to *Streptococcus equi* subspecies *zooepidemicus* (*S. zooepidemicus*) in swine have not previously been reported in the United States. In September and October 2019, outbreaks with swine mortality up to 50% due to *S. zooepidemicus* septicemia were reported in Ohio and Tennessee. Genomic epidemiological analysis revealed that the eight outbreak isolates were clustered together with ATCC 36246, a Chinese strain caused outbreaks with high mortality, also closely related to three isolates from human cases from Virginia, but significantly different from an outbreak-unrelated swine isolate from Arizona and most isolates from other animal species. Comparative genomic analysis on two outbreak isolates and another outbreak-unrelated isolate identified several genomic islands and virulence genes specifically in the outbreak isolates only, which are likely associated with the high mortality observed in the swine population. These findings have implications for understanding, tracking, and possibly preventing diseases caused by *S. zooepidemicus* in swine.

Introduction

Streptococcus equi subspecies *zooepidemicus* (*S. zooepidemicus*), a beta-hemolytic and Gram-positive bacterium, is most frequently isolated as an opportunistic pathogen of horses in the upper respiratory and lower genital tracts. It can also cause infections in a wide range of other animal species, including cats, ruminants, pigs, monkeys, dogs, and guinea pigs (1-6). *S. zooepidemicus* is zoonotic, with reported transmission from horses, dogs, and guinea pigs to humans (7) leading to either severe invasive disease (bacteremia, septic arthritis, pneumonia, and meningitis) or benign disease like pharyngitis, with potential to trigger acute post-streptococcal glomerulonephritis (APSGN) (8). The human patients usually acquire the bacteria through direct contact with infected animals or consumption of contaminated animal products such as milk or cheese.

The pathogenesis of *S. zooepidemicus* is not fully understood; however, virulence factors and the host immune status appear to have roles in the development of disease. The M protein, a member of M/M-like protein family, is a surface-associated protein and a classical virulence factor in group A *Streptococcus* (GAS; e.g., *S. pyogenes*) (9). The M-like protein SzP of *S. zooepidemicus* was reported to contribute to the virulence in animal model studies (10,11) and a second M-like protein SzM has been shown to bind fibrinogen, activate plasminogen, inhibit phagocytosis, and serve as a protective antigen for vaccination (12). Very recently, BifA, a Fic domain-containing protein was shown to disrupt the blood-brain barrier integrity by activating moesin in endothelial cells (13). In addition, capsular and other surface polysaccharides are important streptococcal virulence factors, several superantigen genes (*seeM*, *szeL*, and *szeM*) and several genes associated

with anti-phagocytic capability in *S. zooepidemicus* may be involved in the pathogenicity (14).

Although *S. zooepidemicus* was reported to cause epizootic outbreaks in swine resulting in significant economic losses in China and Indonesia (4, 15), previous isolation of *S. zooepidemicus* from clinically ill pigs have been rather limited in the US in past decades. The first high mortality event from *S. zooepidemicus* in North America was reported very recently in Canada in March 2019 (<https://www.biorxiv.org/content/10.1101/812636v2>). From late September to early October of 2019, three cases of cull sows and feeder pigs from Ohio and Tennessee were submitted to the Veterinary Diagnostic Laboratory at Iowa State University (ISU-VDL). High mortality ranging from 10-50% in groups of pigs was reported over the period of 8-10 days at the buying station in Ohio and similar high mortality (922 out of 2,222 sows in lairage) from an abattoir in Tennessee. The clinical signs included sudden death, weakness, lethargy, and high fever. Splenomegaly and hemorrhagic lymph nodes were the most consistent macroscopic findings. Microscopic lesions were consistent with acute bacterial septicemia. A laboratory diagnosis of *S. zooepidemicus* septicemia was given, which was corroborated by histopathology, PCR, and bacterial culture. To genetically characterize *S. zooepidemicus* strains associated with high mortality and gain insights into the epidemiology of these highly unusual and unexpected outbreaks, we performed whole genome sequencing on eight isolates from the Ohio and Tennessee outbreaks, another outbreak-unrelated swine isolate from Arizona, and 15 *S. zooepidemicus* isolates from other animal species. Three full-length complete genome sequences were further assembled and genomic epidemiological and comparative genomic analyses were conducted.

Materials and Methods

S. zooepidemicus isolates

In total, twenty-four *S. zooepidemicus* isolates were whole genome sequenced and included in this bacterial genomic epidemiological study. Among them, eight isolates (OH-71905, TN-74097, NVSLTN-LUNG1, NVSLTN-LUNG2, NVSLTN-LUNG3, NVSLTN-LIVER4, NVSLTN-TB1 and NVSLTN-TC1) were from Ohio and Tennessee outbreaks, another swine isolate (AZ-45470) from a case not related to the outbreaks, and 15 isolates from different animal species (6 from equine, 3 from feline, 3 from guinea pig, 1 from canine, 1 from caprine, and 1 from chinchilla). Another 24 strains from different countries and years with complete or draft genomes that were publicly available from GenBank (<https://www.ncbi.nlm.nih.gov/genbank/>) were included in study. The detailed information of all 48 *S. zooepidemicus* strains are summarized in Appendix Table 1-2.

DNA extraction and library preparation

A single colony of each *S. zooepidemicus* strain was inoculated into Tryptic Soy Broth (TSB) with 10% bovine serum and incubated at 37°C with overnight shaking. ChargeSwitch gDNA mini bacteria kit (Life Technologies, Carlsbad, CA) was used to extract genomic DNA from *S. zooepidemicus* cells following the manufacturer's guidelines. DNA quality was determined by NanoDrop (Thermo Fisher Scientific) and accurate concentration was measured by Qubit fluorometer double-stranded DNA high sensitivity (dsDNA HS) kit (Life Technologies). At the National Veterinary Service Laboratories (NVSL), a loop of *S. zooepidemicus* from an overnight culture grown on blood agar was inoculated into 400 microliters of TE buffer containing 2.5 mg of lysozyme and incubated for 2 hours at 37°C. The entire volume was extracted using a

Promega Maxwell RSC Whole Blood DNA Kit on the Promega Maxwell RSC 48 (Promega, Madison, WI). Accurate DNA concentration was measured by Qubit fluorometer double-stranded DNA high sensitivity (dsDNA HS) kit (Life Technologies). Indexed genomic libraries were prepared at both laboratories by using Nextera XT DNA library prep kit (Life Technologies) for subsequent sequencing.

Genome sequencing and assembly

Bacterial genomes were sequenced using the Illumina MiSeq platform (Illumina) with 250x2 read length in the NGS Unit of the ISU-VDL or at the NVSL. Low-quality raw reads and adapters were filtered and trimmed by Seqtk and Trimmomatic-0.36 (16). The filtered reads were tested for quality by FastQC and were assembled utilizing SPAdes 3.13.1-Darwin (17). Assembly quality was evaluated using Quast (18) to determine N50, longest contigs, total length of contigs, GC content, and other parameters as appropriate. Contigs with low average depth (≤ 2) or small coverage (≤ 500) were removed from further genomic analysis. The genome sequencing and assembly data were deposited at NCBI under BioProject accessions PRJNA588803 (VDL) and PRJNA591128 (NVSL).

Phylogenetic analysis

Single nucleotide polymorphisms (SNPs) of 48 *S. zooepidemicus* isolates were identified by running kSNP3 with standard mode. The optimal k-mers size was calculated by Kchooser program and the whole-genome phylogeny was analyzed based on identified core genome SNPs (19). Sequence type (ST) based on 7 highly conserved housekeeping genes (*arc*, *nrdE*, *proS*, *spi*, *tdk*, *tpi* and *yqiL*) was assigned for each *S. zooepidemicus* genome according to the PubMLST *S. zooepidemicus* database

(<http://pubmlst.org/szooepidemicus>)(20). Tree Of Life (iTOL, <https://itol.embl.de/>) (21)

was used for display, manipulation and annotation on the base of core SNPs tree.

Genome gap closure and comparative genomic analysis

Three genomes of *S. zooepidemicus* including two outbreak strains of OH-71905 from Ohio and TN-74097 from Tennessee, and another outbreak-unrelated swine isolate AZ-45470 from Arizona were further sequenced using Oxford Nanopore Technologies (Oxford, United Kingdom) GridIONx5 in the DNA facility at ISU to generate longer reads for genome gap closure. The full-length circular genome sequences were obtained by hybrid assembly combining both Illumina short reads and Nanopore long reads using Unicycler v0.4.8 (22). Three complete genome sequences are available at NCBI under BioProject accession PRJNA588803. Comparative genomic studies were performed with BLAST Ring Image Generator (BRIG)(23) to generate a circular genomic map using three closed genome sequences in addition to two complete sequences of control strains ATCC 35246 and CY (Nanjing) strains (GenBank accession number CP002904.1 and CP006770.1). The circular graphical map was plotted to feature GC skew, GC content, and predicted genomic islands (GIs) along with genome comparisons.

Virulence gene and genomic islands identification

Putative virulence genes were retrieved from genome sequences according to previous publications (13, 24-27). The prediction of genomic island (GI) was based on IslandPath-DIMOB, SIGI-HMM, IslandPick and Islander using IslandViewer 4 (28). Coding sequences (CDS) in every GI were annotated by NCBI Prokaryotic Genome Annotation Pipeline (PGAP) (29). The distribution of putative virulence genes and

proportion of CDS of GIs in all of the *S. zooepidemicus* isolates were determined in R using pheatmap package. All gene sequences including putative virulence genes and CDSs in GIs were identified by local Blast+ (2.9.0 version) choosing BLASTn option.

Results

Phylogenetic characterization of *S. zooepidemicus*

Whole genome sequencing was performed on 24 *S. zooepidemicus* isolates. More than 90% of the paired-end reads processed had a Phred score of 35, indicating the high quality of sequencing data. The *de novo* assembly of raw reads generated contigs with total size ranging from 2.0 Mbp to 2.2 Mbp, and an overall GC content between 40% and 43%, which were similar to those of reference genome ATCC 35246 (2,167,164bp, 41.65% GC). Thus, the quality of all assembled contigs of 24 isolates were considered sufficient for whole genome analysis.

Whole genome phylogenetic analysis based on core genome SNPs was conducted with 48 *S. zooepidemicus* isolates including the 24 isolates sequenced in this study and 24 strains from different countries with publically available genome sequences (Figure 1). A total of 23,659 core genome SNPs were identified from all strains by kSNP3. The phylogenetic tree based on the core genome SNPs indicated a large genetic diversity and the 48 isolates could be clustered into 33 phylogenetic lineages (threshold = 0.01) with only 5 clusters consisting of more than one strain (isolate), while the other 28 clusters contained only one strain (isolate). No obvious geographic or historic distribution differences could be found in our phylogenetic analysis. The eight *S. zooepidemicus* isolates from Ohio and Tennessee outbreaks were indistinguishable based on whole genome phylogeny, suggesting that they shared a common source. Surprisingly, the eight

outbreak isolates were clustered together with the Chinese strain (ATCC 36246), which caused outbreaks of high mortality in 1975 (15). In addition, three isolates from human cases with guinea pig exposure (NVSLVA-S19, NVSLVA-S2, and NVSLVA-S22) (6) were also closely related to the outbreak isolates. In contrast, another swine isolate (AZ-45470), which did not cause a high mortality event, was distant to the outbreak isolates.

The MLST analysis on all 48 isolates revealed 24 previously described ST types, while 11 strains (22.91%) represented by 9 novel allelic profiles did not match to any STs available in the current PubMLST database of *S. zooepidemicus* as of December 2019 (Figure 2 and Appendix Table 3). Even though clusterings based on whole genome core SNPs and MLST allelic profiles were overall similar, strains with the same MLST types were distributed both within the same and adjacent whole genome core SNP clusters in the phylogenetic tree, indicating that the SNP-based genotyping was more discriminatory, as expected. The eight strains from Ohio and Tennessee outbreaks and two strains (CY and ATCC 36246) from China were grouped together in ST194, while the outbreak-unrelated Arizona swine isolate (AZ-45470) was clustered distantly in ST340. Three guinea pig isolates from human cases (NVSLVA-S19, NVSLVA-S2, and NVSLVA-S22) with the same allelic profile and an unassigned ST type, differing from ST194 at two allele sequences, were closely clustered with the eight outbreak strains from Ohio and Tennessee as well as the Chinese outbreak strain ATCC 36246 within the ST194 lineage.

Comparative genomic analysis of *S. zooepidemicus*

To further characterize the *S. zooepidemicus* isolates associated with the high mortality outbreaks in Ohio and Tennessee (suggesting a hypervirulence trait), the genome gaps of three isolates were closed to obtain full-length genome sequences and perform comparative genomic analysis. Considering the indistinguishability of seven porcine isolates from Tennessee, one representative isolate from Tennessee (TN-74097), the single swine outbreak isolate from Ohio (OH-71905), and the outbreak-unrelated swine isolate from Arizona (AZ-45470) were subjected to Oxford Nanopore Sequencing to generate long reads. Three full-length complete genome sequences (Accession numbers: CP046040, CP046042, and CP046041) were obtained using hybrid assembly of both Illumina short reads and Nanopore long reads. The general genomic features of all three genomes are summarized in Table 1. Both *S. zooepidemicus* strains OH-71905 and TN-74097 contained a circular chromosome of 2,189,155 bp with an average GC content of 41.65%. There was only one base pair nucleotide difference detected in these two outbreak isolates. Genome Annotation Pipeline (PGAP) detected 1,994 CDSs, 67 tRNA genes, 18 rRNA genes, and 4 ncRNA genes in both isolates OH-71905 and TN-74097. Compared to two outbreak isolates from Ohio and Tennessee, the genome size of *S. zooepidemicus* strain AZ-45470 from an outbreak-unrelated case was noticeably smaller with 2,074,453 bp of length and an average GC content of 41.54%. Accordingly, 1,869 CDSs were identified in the genome of AZ-45470, although it had the same numbers of tRNA, rRNA, and ncRNA genes as with OH-71905 and TN-74097 genomes. Comparative genomic analysis was performed with OH-71905, TN-74097, AZ-45470 genomes as well as the genome of ATCC 35246 and CY strains from China (Figure 3). Genome differences were visualized using BLAST Ring Image Generator (BRIG). As

previously mentioned, the genome sequences of OH-71905 and TN-74097 were nearly identical except one bp nucleotide (a single SNP) difference. Overall, OH-71905 and TN-74097 showed an average nucleotide identity of 99.96% with ATCC 35246 within the 99.36% aligned sequence length, indicating that they are highly similar. Compared with the ATCC 35246 genome, 18 insertions and 6 deletions were found in the genomes of OH-71905 and TN-74097, and the largest insertion was 12,494 bp of length from the position 786,295 bp to position 798,785 bp. In contrast, the genome sequence of AZ-45470 displayed an average nucleotide identity of 97.20% with those of OH-71905 and TN-74097 and covering 91.12% of aligned sequence length. Additionally, 189 deletions and 186 insertions were identified in the genome sequence of AZ-4570 compared with those of OH-71905 and TN-74097. The top six largest deletions ranged from 5,000 bp to 50,000 bp in length.

Identification and distribution of genomic islands

A genomic island (GI) is part of a bacterial genome that has evidence of horizontal transmission origins. The GIs in OH-71905 and TN-74097 were predicted using IslandViewer 4. Six GIs with significantly different GC contents compared with the core genome were identified in both strains. The sizes of identified GIs varied greatly from 6 kb to 89 kb. Several GIs encode putative virulence genes and possibly contribute to the virulence of *S. zooepidemicus* (Appendix Table 4-5). GI-2 encodes a putative holing-like toxin and a putative type VI secretion system protein; GI-3 encodes a virulence associated protein E; GI-4 encodes a putative toxin PezT, a nucleotidyl transferase which belong to the AbiEii/AbiGii toxin family protein, and a putative VirB4 of type IV secretion system.

The distribution of the identified six GIs in all of the 45 *S. zooepidemicus* strains (Figure 4, Appendix Table 6) was assessed by examining the percentage of CDS present in six GIs. All six GIs were present (CDSs >75%, with more than 75% of the CDS in a GI present in other genomes) in all eight outbreak isolates from Ohio and Tennessee. Five complete GIs could be detected in ATCC 35246 and CY strains of the Chinese outbreaks, and only partial CDS (90/106, 84.91%) of GI-2 were found in ATCC 35246. In contrast, only one complete GI (GI-1) was present in swine isolate AZ-45470, which originated from a swine case that was not associated with high swine mortality outbreaks. More than 82% of CDS in all GIs except for GI-2 were present in the three isolates from human cases with guinea pig exposure (NVSLVA-S19, NVSLVA-S2, and NVSLVA-S22). In addition, the occurrence of five GIs (GI-2, GI-3, GI-4, GI-5, and GI-6) in other *S. zooepidemicus* isolates from human and other animal species was extremely low, two of 35 (5.71%) isolates possessed GI-3, GI-4 and GI-6 (CDS > 81%), while none of 35 isolates contained GI-2 or GI-5 (CDS < 61%).

Detection and distribution of putative virulence genes

The presence of 15 putative virulence genes, which have been previously reported in *S. zooepidemicus* (13, 24-27), was examined in all 48 isolates included in this study (Figure 5, Appendix Table 7). The M-like protein gene *szP* was present in all *S. zooepidemicus* isolates in our study albeit with variable sequences. The second M-like protein *SzM* and the newly identified Fic domain-containing protein *BifA* were recently reported virulence factors of *S. zooepidemicus*. Both genes were encoded by all eight isolates from Ohio and Tennessee outbreaks, the ATCC 35246 strain from China associated with high swine mortality and another swine strain from China (CY), but absent from the

outbreak-unrelated swine isolate AZ-45470. The distribution of these two genes in other strains isolated from human or other animal species was only 18.42% (7/38). In addition, a fimbrial subunit protein coding gene *fszF* and a protective antigen-like protein coding gene *spaZ* were also present in all eight outbreak isolates from Ohio and Tennessee and the ATCC 35246 and CY strains from Chinese outbreaks, but negative in the outbreak-unrelated swine isolate AZ-45470. Several superantigen genes including *szcF*, *szcL*, *szcM*, *szcN* and *szcP* have been previously identified in *S. zooepidemicus* (14). These superantigen genes were all absent from all of the 11 swine isolates and infrequently present in other isolates from human and other animal species tested in our study (43.24%, 16/37).

Discussion

Although *S. zooepidemicus* is considered as an opportunistic pathogen in a large variety of host species including cats, rodents, mink, monkeys, and seals, the majority of cases were reported from domestic animals such as horses, dogs, ruminants, and pigs (30). *S. zooepidemicus* is a commensal organism in horses, but may act as an opportunistic pathogen causing abscesses, neonatal septicemia, and endometritis (31). Several studies revealed that *S. zooepidemicus* was the major causative agent of purulent infections in horses and foals leading to severe respiratory diseases (32, 33). *S. zooepidemicus* has also been reported as a causative agent of mastitis in ruminants inducing severe and deep tissue infections (2). In addition, emerging infections of *S. zooepidemicus* in dogs frequently occur as outbreaks causing severe and life-threatening diseases such as hemorrhagic pneumonia and septicemia (34, 35). In the swine industry, *S. zooepidemicus* has been a significant pathogen in some Asian countries with outbreaks reported in China

(1975) where more than 300,000 pigs died (15) and in Indonesia nearly twenty years later (1994) (4). However, high mortality events due to *S. zooepidemicus* in swine have not previously been reported in the United States. Our genomic epidemiological analysis revealed that all eight *S. zooepidemicus* isolates from Ohio and Tennessee outbreaks were indistinguishable, yet clearly different from those isolated from horses, dogs, ruminants, and most other host species regardless of European or North American origin. These eight outbreak isolates were also significantly divergent from another outbreak-unrelated swine strain AZ-45470 from Arizona. What is especially concerning is that these eight outbreak isolates were clustered together with the ATCC 35246, the strain that caused the high mortality outbreak in China (15), and showed high similarity in their contents of genomic islands and virulence genes.

S. zooepidemicus is a known zoonotic pathogen and nearly 30 human cases of meningitis, septicemia, pneumonia, and glomerulonephritis have been documented (36-41). These human cases are usually associated with ingestion of animal products including unpasteurized milk and cheese or contact with companion animals such as horses, dogs, and guinea pigs (30). Transmission from pigs to human has, thus far, never been reported. Our results showed that the eight outbreak isolates of *S. zooepidemicus* in this study were classified as ST194, an ST type recorded in the *S. zooepidemicus* database from two human blood isolates recovered during 2001. Based on the whole genome phylogeny, the eight isolates were also closely clustered with three isolates from human patients who had severe clinical illness with guinea pig exposure (6), and moreover, their genomic islands and virulence genes were very similar. These results highlight significant public health concerns with these recent United States outbreak isolates. Swine producers,

veterinarians, and other personnel who may directly or indirectly have contact with pigs should be aware of the potential of this organism to cause serious disease and death and *S. zooepidemicus* infection should be considered if they have purulent wounds or systemic symptoms of infection.

Our results also suggests the contributions of two previously identified virulence genes *szM* and *bifA* to the pathogenicity of *S. zooepidemicus*. Both genes were present in all eight *S. zooepidemicus* isolates and the ATCC 3546 strain, all of which have been involved in high mortalities. Conversely, both genes were absent from another outbreak-unrelated swine isolate AZ-45470, and were also extremely rare in strains from other animal species. It is possible that these two virulence genes are specifically associated with the hypervirulence of *S. zooepidemicus* to pigs, and further study of these specific genes is warranted. Superantigens are potent toxins, which may disrupt both innate and adaptive immune response and trigger non-specific T-cell proliferation and overzealous inflammatory production in the host (42). Several superantigen genes were shown to be significantly associated with non-strangles lymph node abscessation in the horse and probably contribute to virulence (24). However, in our study, all 11 swine isolates were negative for any superantigen gene, suggesting that they might not be necessary for the virulence of *S. zooepidemicus* in pigs. In addition, many deletions were identified from the genome of the outbreak-unrelated swine isolate AZ-45470 compared with those of eight *S. zooepidemicus* outbreak isolates and the ATCC 3546 strain. Among them, six large deletions were predicted genomic islands that were absent from this strain and several of these islands encoded putative virulence genes including putative toxin genes and type IV and VI secretion system proteins. It will be very interesting to

determine if these genes and islands would contribute to the virulence of *S. zooepidemicus* in pigs.

In summary, we performed genomic epidemiological and comparative genomic analyses with *S. zooepidemicus* isolates associated with high swine mortality in the United States. Our findings provide significant and timely insights for a better understanding of the epidemiology and virulence of *S. zooepidemicus* isolates associated with highly unexpected and severe outbreaks that occurred very recently in the US swine population. In addition, identification of specific virulence genes and genomic islands may lead to the development of novel molecular diagnostic tools, and provide the basis for future investigation of virulence mechanisms and control measures.

Ms. Chen is a graduate student in the College of Veterinary Medicine at Iowa State University. Her Primary research interests include molecular epidemiology and bacterial pathogenesis.

Acknowledgement

This study was partially supported by the Swine Health Information Center (SHIC#19-236). We thank Ying Zheng and Dr. Huigang Shen for excellent technical assistance.

Reference

- 353 1. Blum S, Elad D, Zukin N, Lysnyansky I, Weisblith L, Perl S, et al. Outbreak of
354 *Streptococcus equi* subsp. *zooepidemicus* infections in cats. *Vet Microbiol.* 2010 Jul
355 29;144(1-2):236-9.
- 356 2. Sharp MW, Prince MJ, Gibbens J. S *zooepidemicus* infection and bovine mastitis.
357 *Vet Rec.* 1995 Jul 29;137(5):128.
- 358 3. Fan H, Wang Y, Tang F, Lu C. Determination of the mimic epitope of the M-like
359 protein adhesin in swine *Streptococcus equi* subsp. *zooepidemicus*. *BMC Microbiol.*
360 2008 Oct 7;8:170.
- 361 4. Soedarmanto I, Pasaribu FH, Wibawan IW, Lammler C. Identification and molecular
362 characterization of serological group C streptococci isolated from diseased pigs and
363 monkeys in Indonesia. *J Clin Microbiol.* 1996 Sep;34(9):2201-4.
- 364 5. Byun JW, Yoon SS, Woo GH, Jung BY, Joo YS. An outbreak of fatal hemorrhagic
365 pneumonia caused by *Streptococcus equi* subsp. *zooepidemicus* in shelter dogs. *J Vet Sci.*
366 2009 Sep;10(3):269-71.
- 367 6. Gruszynski K, Young A, Levine SJ, Garvin JP, Brown S, Turner L, et al.
368 *Streptococcus equi* subsp. *zooepidemicus* infections associated with guinea pigs. *Emerg*
369 *Infect Dis.* 2015 Jan;21(1):156-8.
- 370 7. Pelkonen S, Lindahl SB, Suomala P, Karhukorpi J, Vuorinen S, Koivula I, et al.
371 Transmission of *Streptococcus equi* subspecies *zooepidemicus* infection from horses to
372 humans. *Emerg Infect Dis.* 2013 Jul;19(7):1041-8.
- 373 8. Torres R, Santos TZ, Bernardes AFL, Soares PA, Soares ACC, Dias RS. Outbreak of
374 Glomerulonephritis Caused by *Streptococcus zooepidemicus* SzPHV5 Type in Monte
375 Santo de Minas, Minas Gerais, Brazil. *J Clin Microbiol.* 2018 Oct;56(10).

- 376 9. Metzgar D, Zampolli A. The M protein of group A Streptococcus is a key virulence
377 factor and a clinically relevant strain identification marker. *Virulence*. 2011
378 Sep-Oct;2(5):402-12.
- 379 10. Walker JA, Timoney JF. Molecular basis of variation in protective SzP proteins of
380 Streptococcus zooepidemicus. *Am J Vet Res*. 1998 Sep;59(9):1129-33.
- 381 11. Ma Z, Zhang H, Zheng J, Li Y, Yi L, Fan H, et al. Interaction between M-like
382 protein and macrophage thioredoxin facilitates antiphagocytosis for Streptococcus equi
383 ssp. zooepidemicus. *PLoS One*. 2012;7(2):e32099.
- 384 12. Velineni S, Timoney JF. Characterization and protective immunogenicity of the SzM
385 protein of Streptococcus zooepidemicus NC78 from a clonal outbreak of equine
386 respiratory disease. *Clin Vaccine Immunol*. 2013 Aug;20(8):1181-8.
- 387 13. Ma Z, Peng J, Yu D, Park JS, Lin H, Xu B, et al. A streptococcal Fic
388 domain-containing protein disrupts blood-brain barrier integrity by activating moesin in
389 endothelial cells. *PLoS Pathog*. 2019 May;15(5):e1007737.
- 390 14. Xu B, Zhang P, Zhou H, Sun Y, Tang J, Fan H. Identification of novel genes
391 associated with anti-phagocytic functions in Streptococcus equi subsp. zooepidemicus.
392 *Vet Microbiol*. 2019 Jun;233:28-38.
- 393 15. Ma Z, Geng J, Zhang H, Yu H, Yi L, Lei M, et al. Complete genome sequence of
394 Streptococcus equi subsp. zooepidemicus strain ATCC 35246. *J Bacteriol*. 2011
395 Oct;193(19):5583-4.
- 396 16. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina
397 sequence data. *Bioinformatics*. 2014 Aug 1;30(15):2114-20.

398 17. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, et al.
399 SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing.
400 J Comput Biol. 2012 May;19(5):455-77.

401 18. Mikheenko A, Prjibelski A, Saveliev V, Antipov D, Gurevich A. Versatile genome
402 assembly evaluation with QUAST-LG. Bioinformatics. 2018 Jul 1;34(13):i142-i50.

403 19. Gardner SN, Slezak T, Hall BG. kSNP3.0: SNP detection and phylogenetic analysis
404 of genomes without genome alignment or reference genome. Bioinformatics. 2015 Sep
405 1;31(17):2877-8.

406 20. Jolley KA, Chan MS, Maiden MC. mlstdbNet - distributed multi-locus sequence
407 typing (MLST) databases. BMC Bioinformatics. 2004 Jul 1;5:86.

408 21. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v4: recent updates and new
409 developments. Nucleic Acids Res. 2019 Jul 2;47(W1):W256-W9.

410 22. Wick RR, Judd LM, Gorrie CL, Holt KE. Unicycler: Resolving bacterial genome
411 assemblies from short and long sequencing reads. PLoS Comput Biol. 2017
412 Jun;13(6):e1005595.

413 23. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. BLAST Ring Image Generator
414 (BRIG): simple prokaryote genome comparisons. BMC Genomics. 2011 Aug 8;12:402.

415 24. Rash NL, Robinson C, DeSouza N, Nair S, Hodgson H, Steward K, et al. Prevalence
416 and disease associations of superantigens *szf*, *szn* and *szp* in the *S. zooepidemicus*
417 population and possible functional redundancy of *szf*. Res Vet Sci. 2014
418 Dec;97(3):481-7.

- 419 25. Kittang BR, Pettersen VK, Oppegaard O, Skutlaberg DH, Dale H, Wiker HG, et al.
420 Zoonotic necrotizing myositis caused by *Streptococcus equi* subsp. *zooepidemicus* in a
421 farmer. *BMC Infect Dis.* 2017 Feb 15;17(1):147.
- 422 26. Alber J, El-Sayed A, Estoepangestie S, Lammler C, Zschock M. Dissemination of
423 the superantigen encoding genes *seeL*, *seeM*, *szeL* and *szeM* in *Streptococcus equi* subsp.
424 *equi* and *Streptococcus equi* subsp. *zooepidemicus*. *Vet Microbiol.* 2005 Aug
425 10;109(1-2):135-41.
- 426 27. Bergmann R, Jentsch MC, Uhlig A, Muller U, van der Linden M, Rasmussen M, et
427 al. Prominent binding of human and equine fibrinogen to *Streptococcus equi* subsp.
428 *zooepidemicus* is mediated by specific *SzM*-types and is a distinct phenotype of zoonotic
429 isolates. *Infect Immun.* 2019 Oct 21.
- 430 28. Bertelli C, Laird MR, Williams KP, Simon Fraser University Research Computing G,
431 Lau BY, Hoad G, et al. IslandViewer 4: expanded prediction of genomic islands for
432 larger-scale datasets. *Nucleic Acids Res.* 2017 Jul 3;45(W1):W30-W5.
- 433 29. Tatusova T, DiCuccio M, Badretdin A, Chetvernin V, Nawrocki EP, Zaslavsky L, et
434 al. NCBI prokaryotic genome annotation pipeline. *Nucleic Acids Res.* 2016 Aug
435 19;44(14):6614-24.
- 436 30. Fulde M, Valentin-Weigand P. Epidemiology and pathogenicity of zoonotic
437 streptococci. *Curr Top Microbiol Immunol.* 2013;368:49-81.
- 438 31. Timoney JF. The pathogenic equine streptococci. *Vet Res.* 2004
439 Jul-Aug;35(4):397-409.
- 440 32. Laus F, Preziuso S, Spaterna A, Beribe F, Tesei B, Cuteri V. Clinical and
441 epidemiological investigation of chronic upper respiratory diseases caused by

442 beta-haemolytic Streptococci in horses. *Comp Immunol Microbiol Infect Dis.* 2007
 443 Jul;30(4):247-60.

444 33. Newton JR, Laxton R, Wood JL, Chanter N. Molecular epidemiology of
 445 *Streptococcus zooepidemicus* infection in naturally occurring equine respiratory disease.
 446 *Vet J.* 2008 Mar;175(3):338-45.

447 34. Pesavento PA, Hurley KF, Bannasch MJ, Artiushin S, Timoney JF. A clonal
 448 outbreak of acute fatal hemorrhagic pneumonia in intensively housed (shelter) dogs
 449 caused by *Streptococcus equi* subsp. *zooepidemicus*. *Vet Pathol.* 2008 Jan;45(1):51-3.

450 35. Priestnall S, Erles K. *Streptococcus zooepidemicus*: an emerging canine pathogen.
 451 *Vet J.* 2011 May;188(2):142-8.

452 36. Abbott Y, Acke E, Khan S, Muldoon EG, Markey BK, Pinilla M, et al. Zoonotic
 453 transmission of *Streptococcus equi* subsp. *zooepidemicus* from a dog to a handler. *J Med*
 454 *Microbiol.* 2010 Jan;59(Pt 1):120-3.

455 37. Minces LR, Brown PJ, Veldkamp PJ. Human meningitis from *Streptococcus equi*
 456 subsp. *zooepidemicus* acquired as zoonoses. *Epidemiol Infect.* 2011 Mar;139(3):406-10.

457 38. Bordes-Benitez A, Sanchez-Onoro M, Suarez-Bordon P, Garcia-Rojas AJ,
 458 Saez-Nieto JA, Gonzalez-Garcia A, et al. Outbreak of *Streptococcus equi* subsp.
 459 *zooepidemicus* infections on the island of Gran Canaria associated with the consumption
 460 of inadequately pasteurized cheese. *Eur J Clin Microbiol Infect Dis.* 2006
 461 Apr;25(4):242-6.

462 39. Kuusi M, Lahti E, Virolainen A, Hatakka M, Vuento R, Rantala L, et al. An outbreak
 463 of *Streptococcus equi* subspecies *zooepidemicus* associated with consumption of fresh
 464 goat cheese. *BMC Infect Dis.* 2006 Feb 27;6:36.

- 465 40. Sesso R, Pinto SW. Five-year follow-up of patients with epidemic
466 glomerulonephritis due to Streptococcus zooepidemicus. Nephrol Dial Transplant. 2005
467 Sep;20(9):1808-12.
- 468 41. Beres SB, Sesso R, Pinto SW, Hoe NP, Porcella SF, Deleo FR, et al. Genome
469 sequence of a Lancefield group C Streptococcus zooepidemicus strain causing epidemic
470 nephritis: new information about an old disease. PLoS One. 2008 Aug 21;3(8):e3026.
- 471 42. Fraser J, Arcus V, Kong P, Baker E, Proft T. Superantigens - powerful modifiers of
472 the immune system. Mol Med Today. 2000 Mar;6(3):125-32.
- 473

474 **Table 1. General genomic features of three swine *S. zooepidemicus* isolates**
 475 **OH-71905, TN-714097 and AZ-45470.**

Feature	OH-71905	TN-714097	AZ-45470
Length, bp	2,189,155	2,189,155	2,074,453
GC Content	41.65%	41.65%	41.54%
Genes (total)	2,083	2,083	1,958
CDSs (total)	1,994	1,994	1,869
rRNAs	6, 6, 6 (5S, 16S, 23S)	6, 6, 6 (5S, 16S, 23S)	6, 6, 6 (5S, 16S, 23S)
tRNAs	67	67	67
ncRNAs	4	4	4
Pseudo Genes (total)	75	75	47

476

477 *CDS, coding sequence; tRNA, transfer RNA; rRNA, ribosomal RNA; ncRNA,
 478 non-coding RNA.

479

Figure legends

Figure 1. Whole-genome sequence-based phylogenetic analysis was conducted using SNPs located in all tested *S. zooepidemicus* genome to generate a core SNP parsimony tree. The branches of the tree are proportional to the distance between the isolates. 11 color strips demonstrate different hosts of total 48 isolates.

Figure 2. MLST analysis of 48 *S. zooepidemicus* isolates. The outer color ring represents different hosts. ST type of every isolate is listed at the inner circle. The neighbor-joining tree was based on a concatenated alignment of 7 housekeeping genes. Color shades indicate different isolates with identical ST type.

Figure 3. Comparative genome analysis between *S. zooepidemicus* isolates from pigs. Rings from outside to inside: 1) predicted genomic islands of OH-71905 (brown); 2) AZ-45470 from Arizona (blue); 3) CY from China (orange); 4) ATCC 35246 from China (Pink); 5) TN-714097 from Tennessee (yellow); 6) GC skew of OH-71905; 8) GC content of OH-71905; 9) OH-71905 genome. The blank spaces in the rings represented matches with less than 70% identity to the reference genome.

Figure 4. Two-way clustering of GIs prevalence among *S. zooepidemicus* isolates. A red-white-blue heat map was constructed based upon the percentage of the CDSs in each predicted OH-71905 GIs presented in all *S. zooepidemicus* isolates. Clustering was performed to illustrate the similarities between the prevalence of the GIs examined and between the *S. zooepidemicus* isolates with regard to CDS proportion. Red shade

indicates porcine isolates associated with high mortality, while orange shade indicates another porcine isolate AZ-45470 without high mortality.

Figure 5. Distribution of putative virulence genes among 48 *S. zooepidemicus* isolates. Right side of the vertical line shows *S. zooepidemicus* isolates clustered by presence or absence of the putative virulence genes. Colored boxes indicate the presence of 3 classified groups: 1, M-like protein coding genes, yellow; 2, superantigen coding genes, green; 3, other putative virulence genes, red; white boxes indicate the absence of virulence genes. Red shade indicates porcine isolates associated with high mortality, while orange shade indicates another porcine isolate AZ-45470 without high mortality.

host

GC Content

GC Skew

GC Skew(-)

GC Skew(+)

TN-714097

100% identity

90% identity

70% identity

ATCC 35246

100% identity

90% identity

70% identity

CY

100% identity

90% identity

70% identity

AZ-45470

100% identity

90% identity

70% identity

Genomic Islands