Diversity of *Streptococcus equi* subsp. *zooepidemicus* strains isolated from the Spanish sheep and goat population and the identification, function and prevalence of a novel arbutin utilisation system.

Karen F. Steward[a]#, Carl Robinson[a]#*, Matthew T. G. Holden[b], Simon R. Harris[b], Ana Fernández Ros[c], Gema Chacón Pérez[c], Rafael Baselga[c] and Andrew S. Waller[a].

Centre for Preventative Medicine, Animal Health Trust, Lanwades Park, Kentford, Newmarket, Suffolk, United Kingdom [a], Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom[b], Microbiology Department, Exopol, 50840 San Mateo de Gállego, Zaragoza, Spain[c].

#These people contributed equally to this work

*Corresponding author

Mailing Address: Animal Health Trust, Centre of Preventative Medicine, Lanwades Park, Newmarket, Suffolk, CB8 7UU, United Kingdom.

Phone: +44 (0) 1638 751000 ext. 1242.

Fax: +44 (0) 1638 555634.

E-Mail: carl.robinson@aht.org.uk

Abstract

The zoonotic bacterium *Streptococcus equi* subsp. *zooepidemicus* (*S. zooepidemicus*) is a diverse, opportunistic pathogen that can cause mastitis in dairy sheep and goats. We used multilocus sequence typing (MLST) to define the genetic diversity of 60 isolates of *S. zooepidemicus*, which were recovered

from sheep and goats in Spain between 2003 and 2010. We identify a novel clonal complex based on sequence type (ST), ST-236, which accounted for 39 of the 60 isolates. A representative ST-236 strain, *S. zooepidemicus* strain C7 (*Sz*C7), was sequenced and interrogated for the presence of novel nutritional uptake or utilisation systems, the acquisition of which have previously been shown to be important for environmental adaptation in other streptococcal pathogens. A novel phosphoenolpyruvate sugar phosphotransferase system (PTS), which enabled the utilisation of arbutin, was identified. Functionality of the PTS was confirmed following deletion of the PTS from *Sz*C7. Arbutin is found in multiple animal foodstuffs and we propose that the ability to utilise arbutin may have conferred a selective advantage to strains infecting animals, the diet of which contains this sugar.

Key words. *Streptococcus zooepidemicus*, mastitis, goat, sheep, arbutin utilisation, PTS System acquisition.

Introduction

Streptococcus equi subsp. *zooepidemicus* (*S. zooepidemicus*) is an opportunistic pathogen most commonly isolated from horses where its presence is significantly associated with inflammatory airway disease ($P < 0.0001$) [1, 2]. *S. zooepidemicus* also causes opportunistic infections in many other mammalian species including mastitis in cattle [3], sheep [4] and goats [5], pneumonia in ruminants [6], joint pain and respiratory complications in pigs [7] and fatal haemorrhagic pneumonia in dogs [8-10]. There is evidence for zoonotic transmission of *S. zooepidemicus* to humans, for example via consumption of unpasteurised milk or cheese, leading to nephritis [11, 12], meningitis [13-16] and septicaemia [17, 18].

Survival and dominance of emergent strains in new niches requires some genome flexibility, enabling them to function efficiently in new environments. *Streptococcus uberis* (*S. uberis*) is the primary cause of mastitis in cattle in the UK, responsible for 26.4 % of cases in 2014. (https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/458616/vida-cattle-07-14.pdf [last accessed 22/07/16]). Genome sequencing and comparison of this bovine pathogen to closely related streptococcal species showed that, whilst considered a pathogen of the mammary gland, flexibility within its genome and particularly within the nutritional pathways, has enabled it to flourish in a variety of nutritionally constrained habitats such as the bovine gut and in pasture [19]. In particular, phosphoenolpyruvate sugar phosphotransferase systems (PTSs) and glycoside hydrolase family I proteins that are sometimes associated with PTSs, were identified as major contributors to this capacity. Streptococci possess a variable number of PTSs involved in the utilisation of specific sugars, enabling them to survive in particular ecological niches. *S. uberis* strain 0140J contains 15 complete PTSs and five partial PTSs [19]. This is considerably more than the non-pathogenic dairy bacterium *Streptococcus thermophilis* which only has seven PTSs, four of which contain pseudogenes [20]. *Streptococcus pneumoniae* (TIGR4) has 21 PTSs, potentially reflecting its need for flexibility within its niche and PTSs are also known to be important for nutrient acquisition in group A streptococci [21]. The *Streptococcus mutans* (*S. mutans)* reference strain UA159 has 14 PTSs [22]; but lacks a high-affinity transporter for galactose, a sugar abundant in the oral cavity. Sequencing of *S. mutans* strain OMZ175 however, identified a galactose-specific PTS that was prevalent in the population, enhanced growth rates and the ability to compete with the commensal bacterium *Streptococcus gordonii* [23].

S. zooepidemicus is diverse both genetically, based on multilocus sequence type (MLST) data [24], and in the diseases it causes. *Streptococcus equi* subspecies *equi* (*S. equi*), is a host-restricted pathogen of horses that is believed to have descended from *S. zooepidemicus* by passing through a genetic bottleneck during its evolution [25]. Strains of *S. equi* can be easily differentiated from strains of *S.*

zooepidemicus as they have lost the ability to ferment lactose and sorbitol. Of the three publically available complete *S. zooepidemicus* strain genome sequences, H70 (*Sz*H70) [26], MGCS10565 (*Sz*MGCS10565) [15] and ATCC35246 (*Sz*ATCC35246) [27], and *S. equi* strain 4047 (*Se*4047) [26], all contain 16 homologous PTSs with the exception of *Sz*H70 which contains 17. Pseudogenes have been found in two PTSs of *Se*4047 and within one PTS of *Sz*MGCS10565 [25-28]. Additional PTS loci in *S. zooepidemicus* strains conferring abilities to utilise alternative sugar sources with the potential to confer a competitive advantage in certain growth conditions thus far have not been identified.

S. zooepidemicus is regularly isolated from cases of disease, including mastitis, wound infections, respiratory disease and uterine infections in the ruminant population of Spain. We utilised MLST, whole genome sequencing and sugar fermentation profiling to gain an understanding of the genetic structure in this population.

Materials and Methods

Strain collections

A sample of 60 epidemiologically independent Spanish caprine and ovine *S. zooepidemicus* isolates were used to examine the diversity of the population when differentiated by multilocus sequence typing (MLST). Isolate details are provided in Table S1. Strains were selected from disease-associated isolates, primarily mastitis, submitted to Exopol in Zaragoza, Spain, between 2003 and 2010 for diagnostic testing to maximise coverage of the country from the available samples, spanning 27 of the 50 provinces of Spain. A collection of 187 isolates of *S. zooepidemicus* from the UK, Ireland and Saudi Arabia in addition to Spain, collected between 1933 and 2010 were selected to represent a diverse strain set based upon their MLST profile (https://pubmlst.org/szooepidemicus/). These 187 strains were isolated from horses, humans, dogs, cows, goats, sheep and cats, which displayed a variety of disease

manifestations. This strain collection was utilised to determine the presence or absence of loci in a wider population of *S. zooepidemicus* by qPCR and sugar fermentation and are listed in Table S2.

Genomic DNA isolation

For isolation of genomic DNA (gDNA), strains were grown on COBA streptococcal selective agar (bioMérieux) overnight at 37 ^{o}C with 5 % CO_2. A single colony was then cultured in Todd-Hewitt broth (THB) (Oxoid) with hyaluronidase (30 µg/ml, Sigma) from which gDNA was extracted using a GenElute spin column kit as per the manufacturer's instructions (Sigma).

Multilocus sequencing typing

MLST was performed on the collection of 60 Spanish caprine and ovine *S. zooepidemicus* isolates (Table S1) as previously described by Webb *et al.* (2008). Sequence types (STs) were assigned using the MLST database (http://pubmlst.org/szooepidemicus) [29]. Clusters of related STs among the study set isolates and all other strains represented in the *S. zooepidemicus* MLST database were defined by the single- and double-locus variant grouping method implemented in eBURST (http://eburst.mlst.net) [30], to infer the genetic relationships of the isolates to each other and those already on the database. The evolutionary distances of 28 sequence variants, including all 18 STs found in the Spanish population, 3 closely related STs and the STs of the published *S. equi* and *S. zooepidemicus* genomes, were computed from concatenated MLST nucleotide sequences. The Maximum Composite Likelihood nucleotide substitution model was used in MEGA v4.0 and the tree inferred using the Neighbour-Joining method with bootstrap values after 5000 repetitions. The tree was rooted on concatenated sequence from the homologous regions in *S. uberis* strain 0140J. Chi squared or Fisher's exact tests were then used as appropriate to identify statistically significant changes in ST prevalence over time.

Sugar fermentation profiling

S. zooepidemicus strain C7 (*Sz*C7), which was isolated from a goat with mastitis in the Badajoz (BA) region of Spain on 29/02/2008, was selected as a representative ST-236 isolate. Single colonies of *Sz*C7 and the genome sequenced strain *Sz*H70 were inoculated into THB and grown overnight at 37 ^{o}C with 5 % CO_2. 167 µl aliquots of overnight cultures and 83 µl horse serum were then added to 1 ml purple broth supplemented with 0.5 % arbutin, glucose, dihydroxyacetone, glycerol, mannose, fructose, galactose, salicin or cellobiose and incubated at 37 ^{o}C in an atmosphere supplemented with 5 % CO_2 overnight. Cultures were also applied to the wells of analytical profile index (API) strips (bioMérieux), which demonstrate the ability of a strain to ferment 32 different substrates (Table 1), and processed according to the manufacturer's instructions. Colour changes on the API strips were compared to references and cultures were examined for the occurrence of a colour change from purple to yellow, signifying utilisation of the corresponding sugar.

Screening for arbutin fermentation

Single colonies of a collection of 187 *S. zooepidemicus* isolates selected on the basis of their diversity as determined by MLST (Table S2) were inoculated into THB and grown overnight at 37 ^{o}C with 5 % CO_2. 167 µl aliquots of overnight cultures and 83 µl horse serum were then added to 1 ml purple broth supplemented with 0.5 % arbutin and incubated at 37 ^{o}C with 5 % CO_2 overnight. Cultures were examined for the occurrence of a colour change from purple to yellow, signifying the utilisation of arbutin.

Genome sequencing and analysis

Genomic DNA was extracted from *Sz*C7. DNA was sequenced using Illumina technology at the Wellcome Trust Sanger Institute and data deposited in the European Nucleotide Archive under the accession number ERS134307. Sequence data was assembled using a *de novo* genome-assembly program, Velvet v0.7.03 [31], to generate a multi-contig draft genome. The *Sz*C7 genome was

compared to *Sz*H70 and orthologous proteins were identified as reciprocal best matches using FASTA with subsequent manual curation. PTS loci in the genomes of the published *S. zooepidemicus* strains *Sz*H70, *Sz*MGCS10565 and *Sz*ATCC35246 and the *S. equi* strain *Se*4047 were compared to those in the *Sz*C7 genome using Artemis Comparison Tool (ACT) [31]. The putative products of novel regions relative to the *Sz*H70 genome were identified by BLAST comparison with the NCBI NR database (http://www.ncbi.nlm.nih.gov/blast/Blast.cgi).

qPCR screen for PTS locus

Genomic DNA from a panel of 187 *S. zooepidemicus* strains (Table S2) was analysed by qPCR for the presence of a PTS similar to the PTS identified in *Sz*C7 using primers designed to the PTS β glucoside specific IIABC component gene, the second gene in the acquired locus in *Sz*C7. 20 µl reactions constituting 1 X Kapa SYBR fast qPCR mix (Anachem), 0.3 µM forward and reverse primers (PTSF 5' gcattgtatcacgcgattacg 3', PTSR: 5' caaaggcatcagccacatgag 3') and 10 ng DNA were thermocycled at 95 ^{o}C for 3 mins, 40 cycles of 95 ^{o}C for 3 secs, 60 ^{o}C for 10 secs, then 95 ^{o}C for 15 secs on an ABI StepOnePlus instrument. A ramp step from 60 ^{o}C to 95 ^{o}C with reads every 0.3 ^{o}C was performed to generate a dissociation curve of products. Data was analysed using StepOne Software v2.1 and presence/absence determined relative to *Sz*C7 and *Sz*H70 controls.

Deletion of the PTS locus

The beta-glucoside-specific PTS locus identified in the genome of *Sz*C7 was deleted from *Sz*C7 by allelic replacement mutagenesis using the method previously described for the deletion of *prtM* [32]. Regions flanking the region for deletion were amplified by PCR using primers C7PTS1new: 5' aaagaattcatgataagagtaaagaaggcac 3', C7PTS2new: 5' aaagatatcgtctcatttatcgttgattgctc 3' and C7PTS3: 5' aaagatatcgctagctcatagaataaggtac 3', C7PTS4: 5' aaagtcgacccctgattaagcatgggcag 3' and cloned into the *Eco*R1/*Sal*1 sites of the pGHost9 vector multiple cloning site [33]. The resultant vector construct

was transformed into *Sz*C7, integrated into the chromosome and then excised to give the strain *Sz*C7ΔPTS. Deletion of the PTS locus was confirmed by PCR and capillary sequencing using an ABI3100 DNA capillary sequencer.

PTS-linked sugar utilisation

Single colonies of *S. zooepidemicus* strains *Sz*C7 or *Sz*C7ΔPTS were inoculated into THB and grown overnight at 37 ^{o}C with 5 % CO_2. 3340 µl aliquots of overnight cultures and 1660 µl horse serum (Oxoid) were then added to 20 ml purple broth (BD) supplemented with 0.5 % sorbitol, salicin, arbutin or cellobiose (Sigma), incubated at 37 ^{o}C with 5 % CO_2 and the optical density (OD) of the cultures were measured at λ 590 nm and λ 430 nm hourly for eight hours. As sugars were utilised the pH reduced changing the broth cultures from purple to yellow, causing a corresponding reduction in OD at λ 590 nm and increase in OD at λ 430 nm. Experiments were repeated three times and the mean OD values plotted over time. ODs at each time point were compared between *Sz*C7 and *Sz*C7ΔPTS using the Student's t-test.

Results

Caprine and ovine strains of *S. zooepidemicus* are genetically related

STs were assigned to all 60 isolates of *S. zooepidemicus* with the generation of 17 new STs as a result of variation at one or more loci compared to those previously seen (Table S1). ST-30 was the only ST, found in one ovine and one caprine mastitis case here, to be seen in any other species or location, having been found previously in respiratory samples from two Thoroughbred horses stabled on the same yard in the UK during 2000.

194 The genetic relatedness of the 60 *S. zooepidemicus* isolates from goats and sheep were compared to the
195 MLST profiles available on https://pubmlst.org/szooepidemicus/. A neighbour-joining tree was
196 constructed from the concatenated nucleotide sequence of the MLST genes and rooted on its close
197 relative *S. uberis* strain 0140J (Figure 1A and 1B). The tree topology and branch lengths show that all
198 of the isolates of *S. zooepidemicus* from Spanish goats and sheep clustered together, with only eight
199 non-ruminant isolates (ST-62, ST-197, ST-211, ST-267, ST-334, ST-353, ST-356 and ST-358).
200 Interestingly, two of the non-ruminant STs: ST-197 and ST-267, were recovered from a human case of
201 septicaemia during 2007 in the UK and another undefined zoonotic case of infection from The
202 Netherlands in 2010, respectively.
203
204 eBURST identified three clonal complexes (Figure 2), none of which contained isolates from other
205 species or geographical locations. ST-254 was identified as the founder of the, predominantly caprine
206 (12/13 isolates, 9/10 STs), clonal complex 254 (CC-254) containing 10 of the 17 study set-specific STs.
207 The most dominant clonal complex (CC-236) contained the abundant ST-236, which accounted for 37
208 isolates of the 60 studied. Despite its dominance of the study population, ST-236 did not significantly
209 increase in prevalence from 2003 to 2010 (P = 0.685). Of the ovine isolates, 88 % (22/25) were CC-
210 236 compared to 49 % (17/35) of caprine isolates. None of the STs of the study set isolates formed
211 clonal complexes with STs from outside the study set.
212
213 **_Sz_C7 utilizes arbutin**
214 MLST of the 60 Spanish caprine and ovine isolates identified CC-236 as the most prevalent group in
215 the population (65 %, n = 39/60). *Sz*C7 was selected as a representative strain and its ability to utilize
216 41 different substrates was compared to the genome sequenced strain *Sz*H70 (Table S3). This showed
217 that *Sz*C7 and *Sz*H70 had the same profile for 38 of the 41 substrates (Table 1). With both 4-
218 nitrophenyl-βD-mannopyranoside and dihydroxyacetone, there was a very slight colour change with

219 *Sz*H70, indicating a low level of utilisation, which was absent in *Sz*C7. However, arbutin utilisation was
220 observed with *Sz*C7, but not *Sz*H70.
221
222 **Arbutin utilisation is ubiquitous within CC-236, but rare within the *S. zooepidemicus* population**
223 The panel of 187 *S. zooepidemicus* isolates were screened for the ability to utilise arbutin, to determine
224 the prevalence of this capability in the population. This revealed that all 37 ST-236 isolates tested
225 utilised arbutin as did the sole representatives of ST-266, ST-272, ST-55, ST-98, ST-131 and ST-192,
226 which fall into diverse areas around the neighbour joining tree. ST-236, ST-266 and ST-272 are all
227 Spanish caprine and ovine isolates that form a distinct clonal complex on eBURST analysis (Figure 2).
228 Only one other isolate, C25 (CC-254), that was able to utilise arbutin shared commonality with the
229 Spanish cluster, and this was isolated from a case of goat mastitis in Spain in 2008. The remaining
230 isolates that utilise arbutin appear to share no commonality with each other or the Spanish cluster in
231 terms of disease, host species, geography or timescale and the STs are not closely linked to each other
232 phylogenetically.
233
234 **The *Sz*C7 genome encodes a novel PTS**
235 The genome of ST-236 strain *Sz*C7 was sequenced to identify novel genomic regions which may have
236 conferred the ability to utilise arbutin. The draft genome was sequenced to an average depth of 80-fold.
237 The genome of *Sz*C7 has a predicted length of 2115377 bp encoding a predicted 2051 coding sequences
238 with an average GC content of 41.47 %. The *Sz*C7 genome was aligned to the *Sz*H70 genome to aid the
239 identification of novel regions that contain sequences involved in the utilisation of arbutin. Sequences
240 for all of the PTS genes labelled as such in the publically available complete *S. zooepidemicus* and *S.*
241 *equi* genomes were extracted and used to compare with the loci in the other published genomes of *S.*
242 *zooepidemicus* (summarised in Table S3). This information was used to confirm the presence of 16 of
243 these PTS in the *Sz*C7 strain. Through this process we identified a novel PTS in *Sz*C7 comprising a *licT*

transcriptional antiterminator (KC906571), a PTS system β glucoside specific IIABC component (KC906572), a β glucosidase (KC906573) and a phosphoglycerate mutase (KC906574) with greatest sequence identity to *Enterococcus faecalis* (Figures 3 and 4). The system was characterized as being putatively involved in sugar import and utilisation, but was not present in any other *S. zooepidemicus* strain sequenced to completion to date [25-27]. The mechanism by which this system was acquired is unclear, but there were no obvious mobile genetic elements associated with it or on the 52.16 kb contig in which it is assembled. At the same genomic location that *Sz*C7 has acquired the PTS locus described above, the *Sz*MGCS10565 and *Sz*ATCC 35246 genomes have acquired genes encoding a transcriptional regulator and an antibiotic efflux protein, and again there were no obvious mobile genetic elements associated with these acquisitions.

The presence of the novel *Sz*C7 PTS correlates with the ability to utilise arbutin

A PCR screen for the novel PTS β glucoside specific IIABC component gene of *Sz*C7 confirmed that all of the strains of *S. zooepidemicus* that were able to utilise arbutin generated a PCR product of the same size to that observed using *Sz*C7 (Table S2).

Deletion of the novel PTS locus in *Sz*C7 abolished utilisation of arbutin

To confirm that the novel PTS identified by genome sequencing conferred the ability to utilise arbutin to *Sz*C7, the locus was deleted and the fermentation of sorbitol, salicin, cellobiose and arbutin was measured. This showed that deletion of the PTS system from *Sz*C7 did not affect the utilisation of sorbitol (Figure 5A) or cellobiose (Figure 5D) relative to the wild-type strain. Utilisation of salicin (Figure 5B) was retarded in *Sz*C7ΔPTS compared to wild type *Sz*C7 with a significant difference in OD observed from three hours, and arbutin (Figure 5C) utilisation was abolished ($P < 0.0001$ at experimental end point).

Discussion

The widespread pathogen *S. zooepidemicus* is responsible for a diverse range of infections in many different host species with important welfare and economic consequences. Here we report the identity of *S. zooepidemicus* isolates infecting dairy sheep and goats across Spain from 2003 to 2010 and provide one possible explanation for the dominance of ST-236 in this population.

MLST of strains from sheep and goats across Spain identified 17 new STs and eBURST analysis identified three discrete clonal complexes consisting purely of isolates from this study. CC-236 alone accounted for 39 of the 60 isolates studied suggesting that there may have been gene gain or loss by this ST that has given it an advantage, enabling it to spread efficiently through the ruminant population of Spain. Our data could also be explained by a recent emergence of ST-236 although there does not appear to be a temporal increase in its prevalence during the seven-year study period (P = 0.685). Significantly more of the ovine isolates, 88 % (22/25), were CC-236 than any other CC compared to the caprine isolates where 49 % (17/35) were CC-236 (P = 0.0022). This could be explained by the relative ability of these strains to infect these two host species or the way in which the animals are housed, fed or mixed, permitting transmission.

In the bovine pathogen *S. uberis*, nutritional flexibility has been identified as a major contributor to the success of strains in a variety of nutritionally constrained environments [19]. Having identified ST-236 as a prevalent clone in the population we examined the ability of a representative strain, *Sz*C7, to utilise a panel of 41 substrates compared to the genome sequenced strain *Sz*H70. This identified arbutin as being utilized by *Sz*C7, but not *Sz*H70 and screening of a diverse panel of 187 *S. zooepidemicus* strains covering 129 STs revealed that all of CC-236 isolates tested (n = 38) but only one CC-254 isolate (n =

292 10) fermented arbutin. Arbutin fermentation was also infrequent in the wider population (isolates from
293 8/129 STs tested were able to ferment arbutin).
294
295 In order to identify genes or systems within the CC-236 isolates that confer the ability to utilise arbutin,
296 we sequenced the genome of *Sz*C7. This revealed the presence of a novel PTS, absent from the
297 publically available complete genome sequences. The presence of this locus as determined by qPCR in
298 the panel of 187 *S. zooepidemicus* strains correlated exactly with the ability to ferment arbutin. The
299 presence of the novel PTS in all CC-236 isolates, but only one CC-254 isolate, suggests that, despite
300 CC-236 and CC-254 sharing a more recent common ancestor than with the other strains of *S.*
301 *zooepidemicus* included in the phylogenetic analysis, the acquisition of the PTS was a more recent
302 genetic event. The CC-254 isolate (C25) that ferments arbutin and contains the arbutin PTS was
303 isolated in Spain from a mastitic goat in 2008, five years after the first identified member of the CC-
304 236 cluster, and it is interesting to hypothesise that it may have acquired the locus from one of the
305 concurrently circulating CC-236 isolates.
306
307 Deletion of the novel PTS from *Sz*C7 showed that it was essential for arbutin utilisation. Salicin
308 utilisation was also significantly retarded, but sorbitol and cellobiose utilisation were unaffected. The
309 structures of these sugars may be important to the function of the PTS as both arbutin and salicin
310 contain a benzene ring. It is therefore possible that the utilisation of other sugars, e.g. populin and
311 Benzoyl-*beta*-d-glucoside, with similar structures could also be affected.
312
313 We propose that the acquisition of the *Sz*C7 PTS was important for the adaptation of ST-236 to its
314 ability to infect ruminants. It is interesting that this PTS appears to be linked to arbutin utilisation as its
315 constituent genes are most similar to those from various *E. faecalis* strains, a gut pathogen. Arbutin,
316 used commercially in the cosmetics industry for skin whitening and by some herbalists to treat urinary

317 tract infections, is prepared from the bearberry plant (*Arctostaphylos uva-ursi)*. This plant is associated
318 with dry, sandy or rocky soil, and is widespread across Spain, Italy and the Balkans where it may be
319 grazed on by goats and sheep. Arbutin is also naturally found in wheat, potentially another important
320 component of the ruminant diet. A strain able to utilise arbutin may therefore have had a competitive
321 advantage over other strains unable to utilise this energy source.
322
323 Strain H042604571 (ST-192), which tested positive for arbutin utilisation and for the PTS found in
324 *Sz*C7, was isolated from a human septicaemia case. There have been a number of cases of *S.*
325 *zooepidemicus* infection in humans that have been traced to contaminated dairy products [14, 26].
326 Sampling and typing of human *S. zooepidemicus* isolates in Spain would conclude whether zoonotic
327 transmission of ST-236 has occurred. Sampling of ruminants outside of Spain would also be
328 enlightening as to the extent to which CC-236 has spread through the wider ruminant population and
329 may be informative as to other factors, such as housing, climate or diet, which may influence its
330 success.
331
332 In conclusion, CC-236 strains were found to be prevalent in the goat and sheep population in Spain.
333 Nutritional pathways/acquisition systems, shown to be important for the success of other streptococcal
334 species, were examined in a representative ST-236 strain, *Sz*C7, and a novel PTS identified. The PTS
335 was essential for arbutin utilisation, a sugar naturally occurring in potential foodstuffs of ruminants.
336 Although it remains to be determined if the acquisition of the PTS directly correlates with improved
337 colonisation, persistence or survival, this evidence highlights one potential route by which CC-236
338 isolates may have had a competitive advantage leading to their success.
339
340 **Conflict of interest statement.**
341 Authors have no conflict of interest to report.

Acknowledgments

This work was supported by the Animal Health Trust. The MLST website and database [http://pubmlst.org/szooepidemicus/] was developed and is managed by Dr Keith Jolley, Department of Zoology, University of Oxford, Oxford, United Kingdom. Adrian Whatmore, Animal Health & Veterinary Laboratories Agency, Addlestone, United Kingdom and Androulla Efstratiou, Public Health England, London, United Kingdom, kindly provided some of the strains used in the comparative analysis.

References

1. Wood, J.L., Burrell M.H., Roberts C.A., Chanter N. and Shaw Y. *Streptococci and Pasteurella spp. associated with disease of the equine lower respiratory tract.* Equine Vet J, 1993. **25**(4): p. 314-8.
2. Wood, J.L., Wood J.L, Newton J.R, Chanter N. and Mumford J.A. *Association between respiratory disease and bacterial and viral infections in British racehorses.* J Clin Microbiol, 2005. **43**(1): p. 120-6.
3. Sharp, M.W., M.J. Prince, and J. Gibbens, *S. zooepidemicus infection and bovine mastitis.* Vet Rec, 1995. **137**(5): p. 128.
4. Las Heras, A., Vela A.I., Fernández E., Legaz E., Domínguez L. and Fernández-Garayzábal J.F., *Unusual outbreak of clinical mastitis in dairy sheep caused by Streptococcus equi subsp. zooepidemicus.* J Clin Microbiol, 2002. **40**(3): p. 1106-8.
5. Pisoni, G., Zadoks R.N., Vimercati C., Locatelli C., Zanoni M.G. and Moroni P. *Epidemiological investigation of Streptococcus equi subspecies zooepidemicus involved in clinical mastitis in dairy goats.* J Dairy Sci, 2009. **92**(3): p. 943-51.
6. Stevenson, R.G., *Streptococcus zooepidemicus infection in sheep.* Can J Comp Med, 1974. **38**(3): p. 243-50.
7. Soedarmanto, I., Pasaribu F.H., Wibawan I.W. and Lämmler C. *Identification and molecular characterization of serological group C streptococci isolated from diseased pigs and monkeys in Indonesia.* J Clin Microbiol, 1996. **34**(9): p. 2201-4.
8. Chalker, V.J., H.W. Brooks, and J. Brownlie, *The association of Streptococcus equi subsp. zooepidemicus with canine infectious respiratory disease.* Vet Microbiol, 2003. **95**(1-2): p. 149-56.
9. Pesavento, P.A., Hurley K.F., Bannasch M.J., Artiushin S. and Timoney J.F. *A clonal outbreak of acute fatal hemorrhagic pneumonia in intensively housed (shelter) dogs caused by Streptococcus equi subsp. zooepidemicus.* Vet Pathol, 2008. **45**(1): p. 51-3.

10. Abbott, Y., Acke E., Khan S., Muldoon E.G., Markey B.K., Pinilla M., Leonard F.C., Steward K. and Waller A., *Zoonotic transmission of Streptococcus equi subsp. zooepidemicus from a dog to a handler.* J Med Microbiol, 2010. **59**(Pt 1): p. 120-3.
11. Francis, A.J., Nimmo G.R., Efstratiou A., Galanis V., and Nuttall N. *Investigation of milk-borne Streptococcus zooepidemicus infection associated with glomerulonephritis in Australia.* J Infect, 1993. **27**(3): p. 317-23.
12. Balter, S., Benin A., Pinto S.W., Teixeira L.M., Alvim G.G., Luna E., Jackson D., LaClaire L., Elliott J., Facklam R. and Schuchat A.. *Epidemic nephritis in Nova Serrana, Brazil.* Lancet, 2000. **355**(9217): p. 1776-80.
13. Mohr, D.N., Feist D.J., Washington J.A. and Hermans P.E. *Meningitis due to group C streptococci in an adult.* Mayo Clin Proc, 1978. **53**(8): p. 529-32.
14. Edwards, A.T., M. Roulson, and M.J. Ironside, *A milk-borne outbreak of serious infection due to Streptococcus zooepidemicus (Lancefield Group C).* Epidemiol Infect, 1988. **101**(1): p. 43-51.
15. Bordes-Benitez, A., Sánchez-Oñoro M., Suárez-Bordón P., García-Rojas A.J., Saéz-Nieto J.A., González-García A., Alamo-Antúnez I., Sánchez-Maroto A. and Bolaños-Rivero M. *Outbreak of Streptococcus equi subsp. zooepidemicus infections on the island of Gran Canaria associated with the consumption of inadequately pasteurized cheese.* Eur J Clin Microbiol Infect Dis, 2006. **25**(4): p. 242-6.
16. Jovanovic, M., Stevanović G., Tosić T., Stosović B. and Zervos M.J. *Streptococcus equi subsp. zooepidemicus meningitis.* J Med Microbiol, 2008. **57**(Pt 3): p. 373-5.
17. Ghoneim, A.T. and E.M. Cooke, *Serious infection caused by group C streptococci.* J Clin Pathol, 1980. **33**(2): p. 188-90.
18. 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. and Huotari K. *An outbreak of Streptococcus equi subspecies zooepidemicus associated with consumption of fresh goat cheese.* BMC Infect Dis, 2006. **6**: p. 36.
19. Ward, P.N., Holden M.T., Leigh J.A., Lennard N., Bignell A., Barron A., Clark L., Quail M.A., Woodward J., Barrell B.G., Egan S.A., Field T.R., Maskell D., Kehoe M., Dowson C.G., Chanter N., Whatmore A.M., Bentley S.D. and Parkhill J. *Evidence for niche adaptation in the genome of the bovine pathogen Streptococcus uberis.* BMC Genomics, 2009. **10**(1): p. 54.
20. Bolotin, A., Quinquis B., Renault P., Sorokin A., Ehrlich S.D., Kulakauskas S., Lapidus A., Goltsman E., Mazur M., Pusch G.D., Fonstein M., Overbeek R., Kyprides N., Purnelle B., Prozzi D., Ngui K., Masuy D., Hancy F., Burteau S., Boutry M., Delcour J., Goffeau A. and Hols P. *Complete sequence and comparative genome analysis of the dairy bacterium Streptococcus thermophilus.* Nat Biotechnol, 2004. **22**(12): p. 1554-8.
21. Shelburne, S.A., 3rd, Keith D.B., Davenport M.T., Horstmann N., Brennan R.G. and Musser J.M. *Molecular characterization of group A Streptococcus maltodextrin catabolism and its role in pharyngitis.* Mol Microbiol, 2008. **69**(2): p. 436-52.
22. Ajdic, D., McShan W.M., McLaughlin R.E., Savić G., Chang J., Carson M.B., Primeaux C., Tian R., Kenton S., Jia H., Lin S., Qian Y., Li S., Zhu H., Najar F., Lai H., White J., Roe B.A. and Ferretti J.J. *Genome sequence of Streptococcus mutans UA159, a cariogenic dental pathogen.* Proc Natl Acad Sci U S A, 2002. **99**(22): p. 14434-9.
23. Zeng, L., Xue P., Stanhope M.J. and Burne R.A. *A galactose-specific sugar: phosphotransferase permease is prevalent in the non-core genome of Streptococcus mutans.* Mol Oral Microbiol, 2013. **28**(4): p. 292-301.
24. Webb, K., Jolley K.A., Mitchell Z., Robinson C., Newton J.R., Maiden M.C. and Waller A. *Development of an unambiguous and discriminatory multilocus sequence typing scheme for the Streptococcus zooepidemicus group.* Microbiology, 2008. **154**(Pt 10): p. 3016-24.

25. Holden, M.T., Heather Z., Paillot R., Steward K.F., Webb K., Ainslie F., Jourdan T., Bason N.C., Holroyd N.E., Mungall K., Quail M.A., Sanders M., Simmonds M., Willey D., Brooks K., Aanensen D.M., Spratt B.G., Jolley K.A., Maiden M.C., Kehoe M., Chanter N., Bentley S.D., Robinson C., Maskell D.J., Parkhill J. and Waller A.S. *Genomic evidence for the evolution of Streptococcus equi: host restriction, increased virulence, and genetic exchange with human pathogens.* PLoS Pathogens, 2009. **5**(3): p. e1000346.
26. Beres, S.B., Sesso R., Pinto S.W., Hoe N.P., Porcella S.F., Deleo F.R. and Musser J.M. *Genome sequence of a lancefield group C Streptococcus zooepidemicus strain causing epidemic nephritis: new information about an old disease.* PLoS ONE, 2008. **3**(8): p. e3026.
27. Ma, Z., Geng J., Zhang H., Yu H., Yi L., Lei M., Lu C.P., Fan H.J. and Hu S. *Complete genome sequence of Streptococcus equi subsp. zooepidemicus strain ATCC 35246.* J Bacteriol, 2011. **193**(19): p. 5583-4.
28. Paillot, R., Darby A.C., Robinson C., Wright N.L., Steward K.F., Anderson E., Webb K., Holden M.T., Efstratiou A., Broughton K., Jolley K.A., Priestnall S.L., Marotti Campi M.C., Hughes M.A., Radford A., Erles K. and Waller A.S. *Identification of three novel superantigen-encoding genes in Streptococcus zooepidemicus: szeF, szeN and szeP.* Infect Immun, 2010.
29. Jolley, K.A., M.S. Chan, and M.C. Maiden, *mlstdbNet - distributed multi-locus sequence typing (MLST) databases.* BMC Bioinformatics, 2004. **5**: p. 86.
30. Feil, E.J., Li B.C, Aanensen D.M., Hanage W.P. and Spratt B.G. *eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data.* J Bacteriol, 2004. **186**(5): p. 1518-30.
31. Zerbino, D.R. and E. Birney, *Velvet: algorithms for de novo short read assembly using de Bruijn graphs.* Genome Res, 2008. **18**(5): p. 821-9.
32. Hamilton, A., Robinson C., Sutcliffe I.C., Slater J., Maskell D.J., Davis-Poynter N., Smith K., Waller A. and Harrington D.J. *Mutation of the maturase lipoprotein attenuates the virulence of Streptococcus equi to a greater extent than does loss of general lipoprotein lipidation.* Infect Immun, 2006. **74**(12): p. 6907-19.
33. Maguin, E., Prévost H., Ehrlich S.D. and Gruss A. *Efficient insertional mutagenesis in lactococci and other gram-positive bacteria.* J Bacteriol, 1996. **178**(3): p. 931-5.

Figure and Table legends

Table 1 Substrate utilisation by *S. zooepidemicus* strains SzC7 and SzH70. Substrates utilised differentially are indicated by *.

Figure 1 Neighbour-joining trees of *S. zooepidemicus* isolates based on concatenated MLST nucleotide sequence.

A) Neighbour-joining tree of all concatenated MLST STs with complete profiles in the online database (http://pubmlst.org/szooepidemicus [last accessed 06/02/17). The branch containing Spanish isolates is indicated with a box. Tree rooted on *S. uberis* 0140J. Strains indicated with arrows have been fully

sequenced B) Expansion of the region of the neighbour-joining tree containing the Spanish caprine and ovine isolates and closely related STs (tree shown in linear format for ease of viewing). STs found only in the Spanish isolate collection in boxes, STs found within and outside the Spanish collection circled. Bootstrap values shown. The table indicates species associated with each ST.

Figure 2 eBURST of *S. zooepidemicus* isolates based on concatenated MLST nucleotide sequence.

Associated clonal complexes as determined by eBURST, purple lines indicate single-locus variants, light blue lines indicate double locus-variants, blue dots show predicted founder STs of clonal complexes and dot size increases with the number of isolates in the group (CC-236 n = 39, CC-254 n = 13, CC-265 n = 4 – all Spanish caprine and ovine isolates). Study-specific STs are encircled and those with multiple STs are shown expanded and below.

Figure 3 Genomic region encoding the novel *Sz*C7 PTS. Artemis comparison tool alignment between *Sz*C7 (top), *Sz*H70 (middle) and *Sz*MGCS10565 (bottom) of the region encoding the novel PTS acquired by *Sz*C7. Blue and red blocks represent conserved regions within the sequences in the forward and reverse orientations respectively.

Figure 4 Gene map of the PTS system acquired by *Sz*C7. Percentage amino acid identities, species and predicted gene function to closest matches are indicated. Acquired PTS indicated by black arrows, flanking genes indicated by grey arrows. Accession numbers: *licT* transcriptional regulator (AGS46817), PTS system beta glucoside specific IIABC component (AGS46818), beta glucosidase (AGS46819) and phosphoglycerate mutase (AGS46820).

Figure 5 Utilisation of sorbitol, salicin, arbutin and cellobiose by *Sz*C7 and *Sz*C7ΔPTS. Optical densities of *Sz*C7 and *Sz*C7ΔPTS cultures as a result of utilisation of A) sorbitol, B) salicin, C) arbutin and D) cellobiose over time, 95 % confidence intervals shown. As sugars were utilised the pH reduced changing the broth cultures from purple to yellow, causing a corresponding reduction in OD at λ 590 nm and increase in OD at λ 430 nm. Chemical structures of sugars are shown. ODs significantly different to wild type are shown, * indicates $0.001 < P < 0.05$ and Δ indicates $P < 0.001$.

Table S1 Spanish *S. zooepidemicus* isolates analysed by MLST, data available on the MLST database (http://pubmlst.org/szooepidemicus).

Table S2 *S. zooepidemicus* isolates tested by qPCR for the presence of the PTS like that found in *Sz*C7 and the ability to ferment arbutin. ND indicates not determined. Isolates from 8/129 STs tested were able to ferment arbutin. Study associated strains are shaded grey.

Table S3 Summary of PTSs identified in genome sequenced *S. equi* and *S. zooepidemicus* strains.

516 Table 1

Substrate	*Sz*H70	*Sz*C7
Arbutin *	-	+
2-naphthyl-βD-galactopyranoside	-	-
4-nitrophenyl-αD-galactopyranoside	-	-
4-nitrophenyl-βD-galactopyranoside-2-CHA *	±	-
4-nitrophenyl-βD-mannopyranoside	-	-
6-bromo-2-naphthyl-N-acetyl-βD-glucosaminide	-	-
Cellobiose	+	+
D-arabitol	-	-
Dihydroxyacetone *	±	-
D-lactose	+	+
D-maltose	+	+
D-mannitol	-	-
D-melezitose	-	-
D-melibiose	-	-
D-raffinose	-	-
D-ribose	-	-
D-saccharose	+	+
D-sorbitol	+	+
D-tagatose	-	-
D-trehalose	-	-
Fructose	+	+
Galactose	+	+
Glucose	+	+
Glycerol	-	-
Glycogen	+	+
L-alanyl-L-phenylalanyl-L-proline-β-naphthylamide	+	+
L-arabinose	-	-
L-arginine	+	+
L-glycyl-L-tryptophan-β-naphthylamide	-	-
Mannose	+	+
methyl-βD-glucopyranoside	-	-
Pullulan	+	+
pyroglutamic acid-β-naphthylamide	-	-
resorufin-βD-galactopyranoside	-	-
resorufin-βD-glucopyranoside	-	-
resorufin-βD-glucuronideside	+	+
Salicin	+	+
sodium hippurate	+	+
sodium pyruvate	-	-
urea	-	-
αcyclodextrin	+	+

517

Figure 1

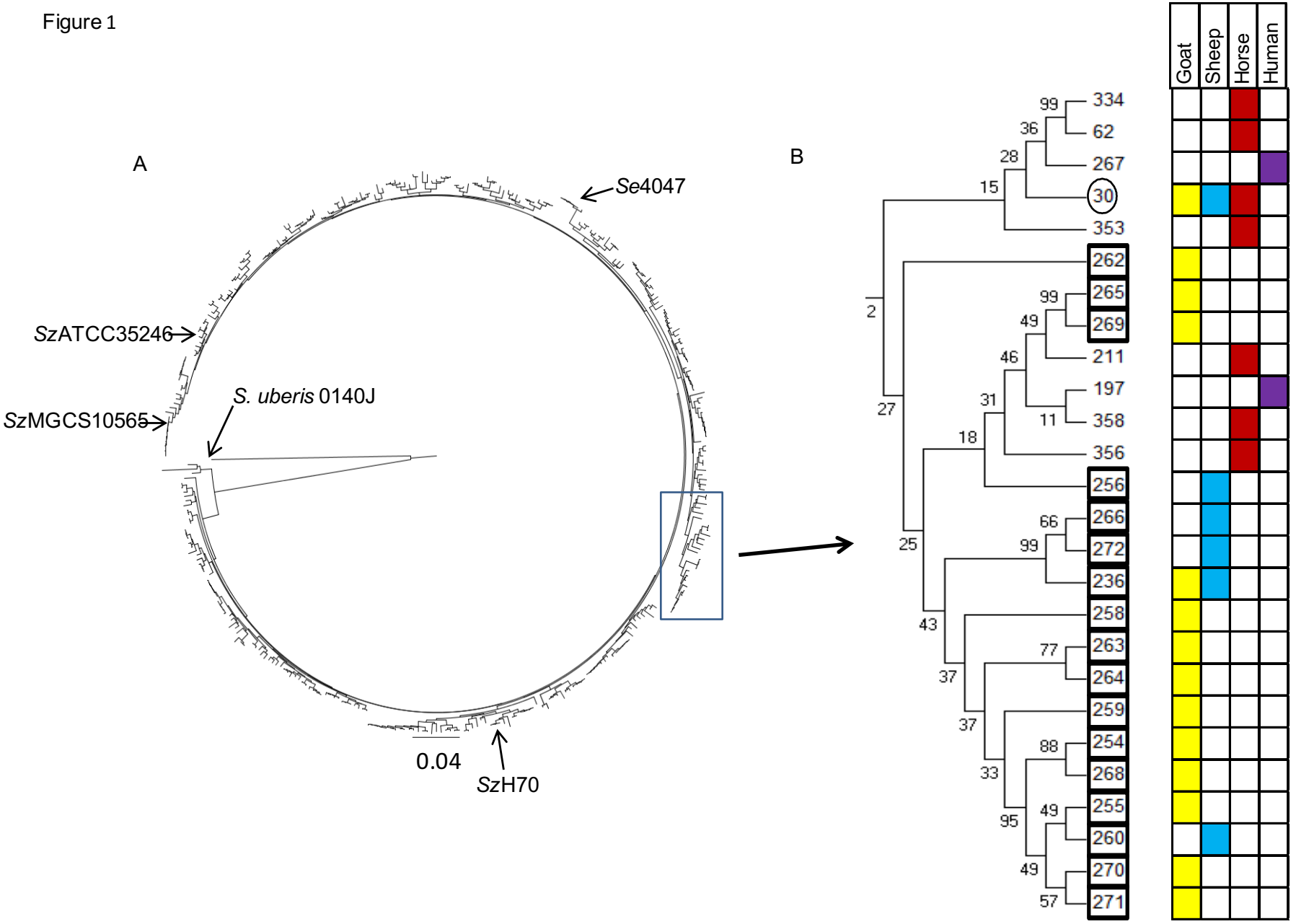

A
Se4047
SzATCC35246
S. uberis 0140J
SzMGCS10565
0.04
SzH70
B
Goat
Sheep
Horse
Human
334
62
267
30
353
262
265
269
211
197
358
356
256
266
272
236
258
263
264
259
254
268
255
260
270
271
99
36
28
15
2
27
99
49
46
31
11
18
25
66
99
43
77
37
37
88
33
95
49
49
57

Figure 2

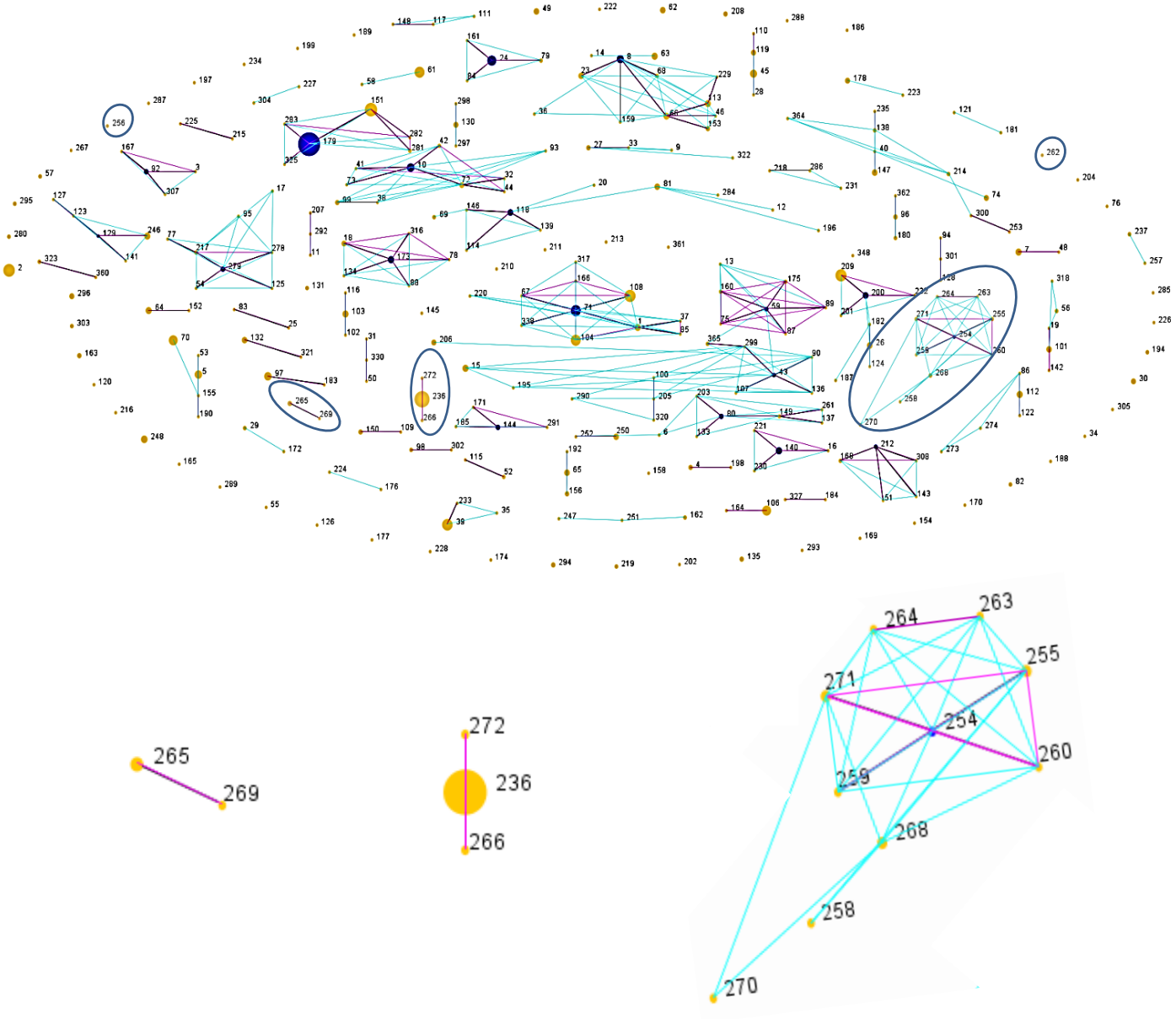


Figure 3

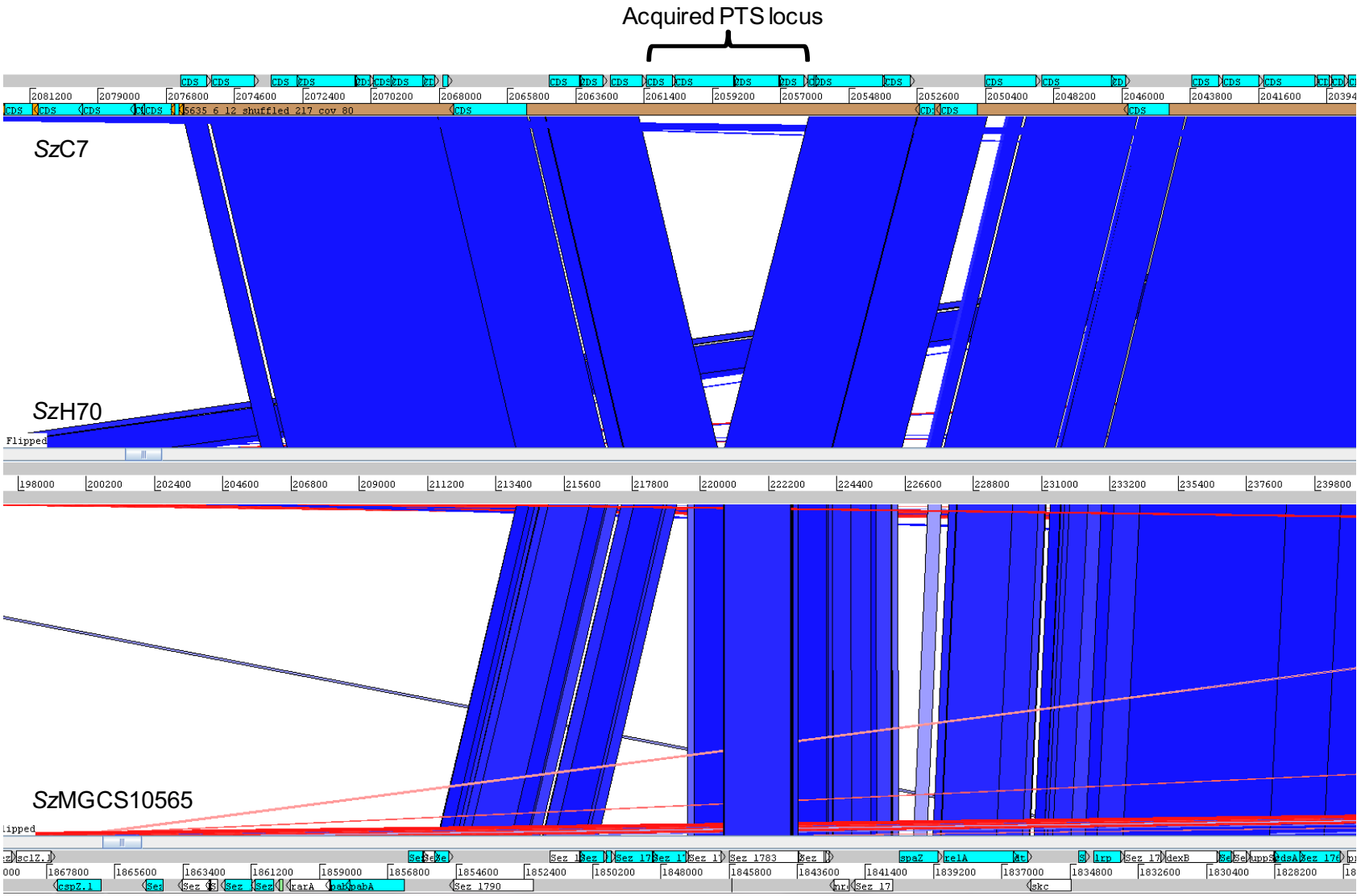


Figure 4

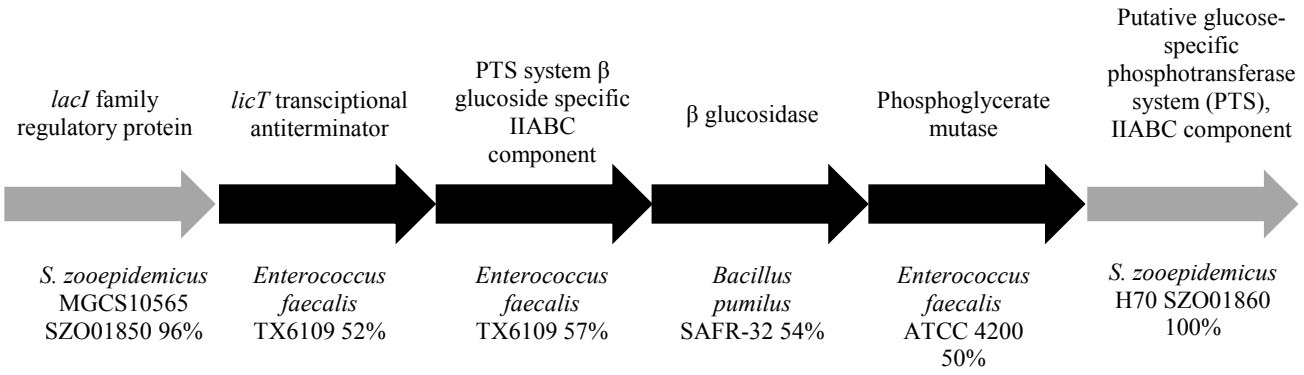
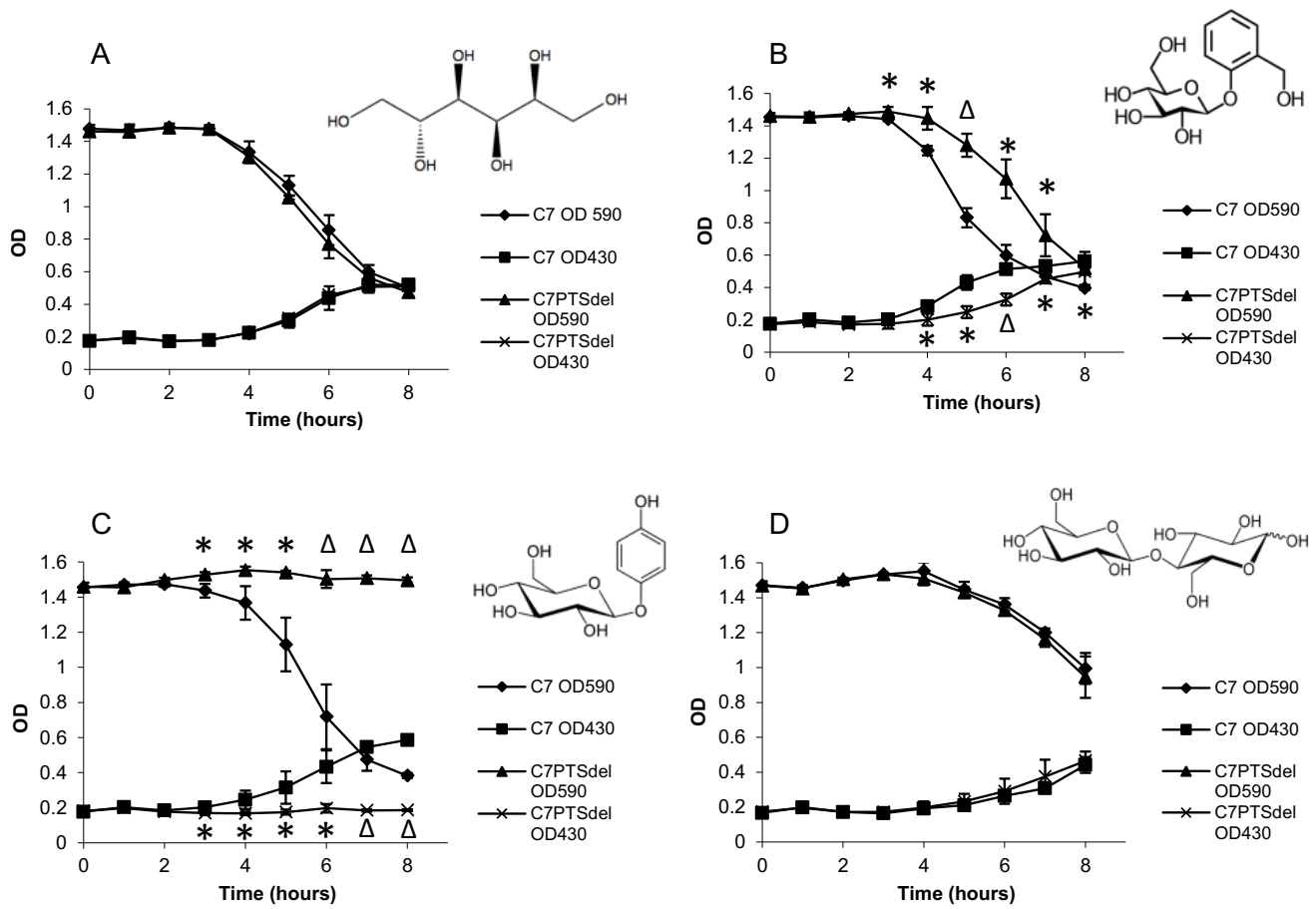

lacI family regulatory protein
licT transciptional antiterminator
PTS system β glucoside specific IIABC component
β glucosidase
Phosphoglycerate mutase
Putative glucose-specific phosphotransferase system (PTS), IIABC component
S. zooepidemicus MGCS10565 SZO01850 96%
Enterococcus faecalis TX6109 52%
Enterococcus faecalis TX6109 57%
Bacillus pumilus SAFR-32 54%
Enterococcus faecalis ATCC 4200 50%
S. zooepidemicus H70 SZO01860 100%

Figure 5


A
B
C
D
OD
Time (hours)
C7 OD 590
C7 OD590
C7 OD430
C7PTSdel OD590
C7PTSdel OD430