PinR mediates the generation of reversible population diversity in *Streptococcus zooepidemicus*

Karen F. Steward[1], Tihana Harrison[1], Carl Robinson[1], Josh Slater[2], Duncan J. Maskell[3] Simon R. Harris[4], Matthew T.G. Holden[4] and Andrew S. Waller[1*]

[1]*Animal Health Trust, Kentford, Newmarket, Suffolk, CB8 7UU, UK*

[2]*Royal Veterinary College, Hawkshead Lane, Hatfield, Hertfordshire, AL9 7TA, UK*

[3]*Department of Veterinary Medicine, University of Cambridge, Madingley Road, Cambridge, CB3 OES, United Kingdom*

[4]*Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridgeshire, CB10 1SA, UK*

* Corresponding author

E-mail address: andrew.waller@aht.org.uk

Contents category: Cell and Molecular Biology of Microbes

Word count of Abstract: 246

Word count of main body of text: 3,960

Key words: Streptococcus zooepidemicus, PinR, recombinase, surface proteins

Accession number for the *Streptococcus zooepidemicus* strain 2329 genome: JTJH00000000

Abstract:

Opportunistic pathogens must adapt to and survive in a wide range of complex ecosystems. *Streptococcus zooepidemicus* is an opportunistic pathogen of horses and many other animals, including man. The assembly of different surface architecture phenotypes from one genotype is likely to be crucial to the successful exploitation of such an opportunistic lifestyle. Construction of a series of mutants revealed that a serine recombinase, PinR, inverts 114 bp of the promoter of SZO_08560, which is bordered by GTAGACTTTA and TAAAGTCTAC inverted repeats. Inversion acts as a switch, controlling the transcription of this sortase-processed protein, which may enhance the attachment of *S. zooepidemicus* to equine trachea. The genome of a recently sequenced strain of *S. zooepidemicus*, strain 2329 (*Sz*2329), was found to contain a disruptive internal inversion of 7 kb of the FimIV pilus locus, which is bordered by TAGAAA and TTTCTA inverted repeats. This strain lacks *pinR* and we hypothesized that this inversion may have become irreversible following the loss of this recombinase. Active inversion of FimIV was detected in three strains of *S. zooepidemicus*: 1770 (*Sz*1770), B260863 (*Sz*B260863) and H050840501 (*Sz*H050840501), all of which encoded *pinR*. A deletion mutant of *Sz*1770 that lacked *pinR* was no longer capable of inverting its internal region of FimIV. Our data highlight redundancy in the PinR sequence recognition motif around a short TAGA consensus and suggest that PinR can reversibly influence the wider surface architecture of *S. zooepidemicus*, providing this organism with a bet-hedging solution to survival in fluctuating environments.

Abbreviations:

ACT, Artemis Comparison Tool

AFHP, acute fatal haemorrhagic pneumonia

DMEM, Dulbecco's modified Eagle's medium

ST, sequence type

S. equi, Streptococcus equi subspecies *equi*

S. zooepidemicus, Streptococcus equi subspecies *zooepidemicus*

THA, Todd Hewitt Agar

THAE, Todd Hewitt Agar containing at 0.5 μg ml^{-1}

THB, Todd Hewitt Broth

THBE, Todd Hewitt Broth containing at 0.5 μg ml^{-1}

Introduction:

The Gram-positive organism *Streptococcus equi* subspecies *zooepidemicus* (*S. zooepidemicus*) is the most frequently isolated opportunistic pathogen of horses, associated with respiratory disease in young horses (Lindahl *et al.*, 2013; Velineni *et al.*, 2014; Wood *et al.*, 1993; Wood *et al.*, 2005) and uterine infections in mares (Hong *et al.*, 1993; Rasmussen *et al.*, 2013; Smith *et al.*, 2003). The bacterium is also associated with disease in a wide range of other animal hosts including dogs (Abbott *et al.*, 2010; Chalker *et al.*, 2003; Pesavento *et al.*, 2008) and humans (Abbott *et al.*, 2010; Balter *et al.*, 2000). The *S. zooepidemicus* group contains a wide variety of strain types, reflecting the diverse array of hosts and tissues that this species of bacteria can infect, and there are 324 distinct sequence types (ST) currently listed on the multilocus sequence typing (MLST) online database http://pubmlst.org/szooepidemicus/ [last accessed 24th November 2014], (Webb *et al.*, 2008). However, *S. zooepidemicus* strains of the same ST are frequently isolated from several host species, highlighting that at least some strains are equipped to exploit new pathogenic niches as and when the opportunity arises.

Within the *S. zooepidemicus* group, *Streptococcus equi* subspecies *equi* (*S. equi*) is the causative agent of strangles, which is the most frequently diagnosed infectious disease of horses worldwide. *S. equi* is host-restricted and only causes strangles, which is characterized by abscessation of the lymph nodes of the head and neck. Comparison of the genomes of *S. zooepidemicus* strain H70 (*Sz*H70) and *S. equi* strain 4047 (*Se*4047) provided evidence of functional loss in the genome of *Se*4047 due to mutation and deletion, coupled with pathogenic specialization through the acquisition of mobile genetic elements (Heather *et al.*, 2008; Holden *et al.*, 2009). The majority of *S. zooepidemicus* isolates (101 of 140 isolates tested), including *Sz*H70, encode a 131 kDa putative sortase-processed surface protein, SZO_08560, which contains a C-terminal LPXTG motif (Holden *et al.*, 2009). SZO_08560 contains four Listeria-Bacteroides repeat Pfam domains (PF09479) with structural similarity to mucin-binding proteins (Ebbes *et al.*, 2011), but the function of this protein remains unknown. The *Se*4047 genome encodes only the final 112 amino acids of the orthologous protein (SEQ_1307a) and lacks an

orthologue of an adjacent gene, SZO_08550, which is predicted to encode a serine recombinase (pfam00239), named PinR (COG1961). Examination of the *Sz*H70 genome sequencing data revealed five of fifty sequence reads that positioned 114 bp of the promoter region of SZO_08560 (-170 bp to -55 bp) in the inverted 'B' orientation as opposed to the annotated reference 'A' orientation. This sequence is bordered by GTAGACTTTA and TAAAGTCTAC inverted repeats and it is proposed that inversion of this sequence by PinR switches transcription of SZO_08560 on or off, thereby modulating the production of the SZO_08560 surface protein in a manner akin to phase variation in Gram-negative bacteria such as *Escherichia coli* or *Bacteroides fragilis* (Abraham *et al.*, 1985; Cerdeno-Tarraga *et al.*, 2005; Coyne *et al.*, 2003).

We constructed a series of *S. zooepidemicus* deletion mutants to determine if PinR mediates the inversion of the SZO_08560 promoter and investigate the wider recombinase-mediated regulation of protein production in *S. zooepidemicus*.

Methods:

Bacterial isolates

Full details of all of the isolates examined in this study are available in Table S1 and on the MLST database (http://pubmlst.org/szooepidemicus/). *Sz*H70 was isolated from a nasopharyngeal swab taken from a healthy Thoroughbred racehorse in Newmarket, UK during 2000 and is ST-1 (Holden *et al.*, 2009). *S. zooepidemicus* strain 2329 (*Sz*2329) is an ST-118 strain that was isolated from a tracheal wash recovered from a healthy Welsh mountain pony in the UK during 1996. *S. zooepidemicus* strain 1770 (*Sz*1770) was recovered from a case of acute fatal hemorrhagic pneumonia in a greyhound from Kent in 2008 and is ST-18. *S. zooepidemicus* strain B260863 (*Sz*B260863) was isolated from an aborted fetus of equine origin in the UK during 2006 and is ST-13. *S. zooepidemicus* strain H050840501 (*Sz*H050840501) is an ST-195 strain that was recovered from the blood of a man who died of septicemia in the UK during 2005. Unless otherwise stated, *S.*

zooepidemicus strains were grown on COBA strep select plates (bioMérieux), on Todd Hewitt Agar (THA) (Oxoid) or in Todd Hewitt Broth (THB) (Oxoid) at 37 °C in an atmosphere containing 5 % CO_2.

Allelic replacement mutagenesis

Internal gene deletions and rearrangements were introduced into *Sz*H70 or *Sz*1770 through an allelic replacement strategy using the pG$^+$host9 plasmid (Maguin *et al.*, 1996), which has been described previously for the deletion of *prtM* in *Se*4047 (Hamilton *et al.*, 2006). Briefly, approximately 500 bp fragments of DNA that flanked the desired sequence to be modified were generated by PCR using the primers listed in Table S2, and cloned into the p$^+$Ghost9 plasmid via *Eco*RI and *Sal*I restriction sites. To complement gene disruptions, full copies of *pinR* or SZO_08560 under the control of their native promoters, were cloned into the *Age*I and *Pvu*I restriction sites of the pGHost9ΔSZO07770 construct that was previously utilized to insert a novel control qPCR target sequence into *Sz*H70 (Webb *et al.*, 2013). The sequences of the insertions into each plasmid were obtained on both strands using an ABI3100 DNA sequencer with BigDye fluorescent terminators and the primers listed in Table S2.

In order to generate each modified strain, *Sz*H70 or *Sz*1770 was transformed with the relevant pG$^+$host9 plasmid and transformants were subjected to two rounds of homologous recombination as described previously (Hamilton *et al.*, 2006). The first recombination event, leading to the integration of the plasmid into the bacterial chromosome, was achieved by growing transformants in THB containing erythromycin at 0.5 µg ml^{-1} (THBE) at 28 ^{o}C overnight and then increasing the temperature to 37 ^{o}C for 3 hours. Integrants were selected following growth on Todd Hewitt agar containing erythromycin at 0.5 µg ml^{-1} (THAE) overnight at 37 ^{o}C. Integrants were inoculated into THB and grown at 37 ^{o}C overnight followed by dilution into THB and incubation at 28 ^{o}C for a further 48 hours. Incubation at the permissive temperature (28 ^{o}C) allowed plasmid replication and facilitated the second recombination event. Bacteria were plated on THA and grown at 37 ^{o}C to promote the loss of free plasmid. Putative mutant colonies were sub-cultured onto fresh THA and THAE plates to confirm their erythromycin sensitivity. The presence of the relevant mutant allele in the chromosome of putative mutants was

determined by PCR using the primers listed in Table S2 followed by DNA sequencing on an ABI3100 DNA sequencer with BigDye fluorescent terminators. A schematic of the mutants generated in this study is shown in Figure 1.

Preparation of chromosomal DNA

Chromosomal DNA was purified from a single colony using GenElute spin columns according to manufacturer's instructions (Sigma).

Isolation of total bacterial RNA and preparation of cDNA

An overnight culture was diluted 1/20 in fresh THB and grown to an $OD_{600\ nm}$ of 0.3. The culture was mixed with two volumes of RNA protect (Qiagen) and cells were harvested by centrifugation at 4 °C at 5000 x g for 10 minutes followed by 8000 x g for 10 minutes. Supernatant was poured off and the pellet re-suspended in 200 µl tris-EDTA buffer (Fluka), 3 mg lysozyme (Sigma) and 500 U of mutanolysin (Sigma). The cells were vortexed repeatedly for 45 minutes, 700 µl of RLT buffer (Qiagen) was added and the sample vortexed for 10 seconds. 0.05 g of acid washed glass beads (Sigma) was added and the sample vortexed for 5 minutes to complete cell lysis. The sample was centrifuged at 16100 x g and RNA was extracted from the supernatant using an RNeasy midi kit (Qiagen) with the inclusion of two on-column DNase 1 treatment steps according to the manufacturer's instructions. RNA was quantified using a NanoDrop 1000 V3.7.1 spectrophotometer and reverse transcribed using a Verso cDNA kit according to the manufacturer's instructions (Thermo Scientific).

qPCR for quantification of transcripts and the orientation of the invertible region

The number of copies of DNA or cDNA of interest were quantified by qPCR with the primers listed in Table S2. Reactions contained 10 µl Kapa SYBR fast (Kapa Biosystems), 0.3 µM forward primer, 0.3 µM reverse primer, 6 µl 1/10 dilution of DNA or cDNA. Reactions were made up to 20 µl with water and thermocycled on an ABI StepOnePlus instrument at 95 °C for 3 minutes followed by 40 cycles of 95 °C for 30 seconds and 60 °C for 10 seconds with a SYBR read taken at the end of each cycle, then

95 °C for 15 seconds. A melt curve was performed from 60 °C to 95 °C with SYBR reads every 0.3 °C to differentiate potential non-specific amplification products and data analyzed using StepOnePlus Software v2.1. No template and no reverse transcription controls were used as negative controls and standard curves with a DNA reference were performed for each primer pair. The experiments were repeated in triplicate and data were normalized by comparison with the house-keeping gene *gyrA*. Amplified FimIV DNA fragments were purified using a PCR purification kit (Qiagen), and the sequences obtained on both strands using an ABI3100 DNA sequencer with BigDye fluorescent terminators using the original PCR primers. Sequence data were assembled using SeqMan 5.03 (DNAstar Inc.).

Quantification of *in vitro* growth rate

Mutant strains were inoculated into THB containing 10 % fetal calf serum (THBS) in triplicate and the growth of each strain was monitored by measuring the OD_{600nm}.

Air-interface infection model

Air-interface respiratory tract organ cultures were constructed using explants of equine trachea as described previously (Hamilton *et al.*, 2006). The trachea used in this study were recovered from six ponies that were euthanized for reasons unrelated to this project and processed on the same day to maximize cell viability. Trachea were washed in Dulbecco's modified Eagle's medium supplemented with 2 mM L-glutamine (DMEM) containing penicillin 100 U ml^{-1}; streptomycin 50 µg ml^{-1}; gentamicin 100 µg ml^{-1} and amphotericin-B 2.5 µg ml^{-1} (PAA) for four hours to remove commensal flora. Following further washing in DMEM to remove residual antibiotics and amphotericin-B, the trachea were dissected into pieces approximately 5 mm^2 and mounted on agarose platforms surrounded by 3 ml DMEM supplemented with 2 mM L-glutamine, in six-well cell culture plates. Organ cultures were maintained in a humidified 5 % CO_2 incubator at 37 ^{o}C. The viability of the air-interface organ cultures was assessed using 1 µm polystyrene bead (Park Scientific) clearance. Contamination was monitored by running a bacteriology loop around all four edges of the culture pieces and streaking onto strep

select plates. Any tissue pieces in which contamination was detected were discarded. Organ culture pieces were infected with a 10 µl suspension containing 1 x 10^6 colony forming units (c.f.u.) of *Sz*H70 mutants, or were mock-infected with THB. Attachment of bacteria to the organ culture pieces was quantified by measuring viable counts (six organ culture pieces per time point) of adherent bacteria at two hours post-infection. Organ culture pieces were vortexed for 5 seconds in phosphate buffered saline (PBS) to remove non-adherent bacteria and then homogenized before plating serial ten-fold dilutions onto THA and enumerating colonies. Data are presented from six independent experiments.

Whole genome sequencing

*Sz*2329 has previously been shown to lack *pinR* and SZO_08560 by PCR screening of a diverse population of *S. zooepidemicus* (Holden *et al.*, 2009). Lack of *pinR* raised the possibility that novel invertible sequences could be fixed in the genome, facilitating their identification. Therefore, the genome of *Sz*2329 was sequenced to 25-fold coverage using a Genome Sequencer-FLX (454 Life Sciences, Roche Applied Sciences, IN, USA). Two sequencing libraries were prepared from genomic DNA, the first a fragment (~250 bp read length) and a second 3,000 bp insert, long-tag paired end library (~100 bp) to provide scaffolding. The reads were assembled with Newbler (v2.0.01.14) using default assembly parameters. Comparison with the genome sequence of *Sz*H70 (FM204884) (Holden *et al.*, 2009) was facilitated by using the Artemis Comparison Tool (ACT) (Carver *et al.*, 2005). The sequence and annotation of the *Sz*2329 genome has been deposited in the EMBL database under accession number JTJH00000000.

Statistical analysis

A two-sided student's *t*-test was used to compare continuous data where assumptions of a normal distribution and equal variance were satisfied. A Kruskal Wallis test was performed to determine the significance of growth curve data. A paired student's *t*-test was used to compare the attachment of wild-type and mutant strains of *Sz*H70 to explants of equine trachea, accounting for variation between the six different trachea.

Results:

PinR is responsible for inversion of the region upstream of SZO_08560

We generated a series of mutant and complementation strains in *Sz*H70 (Figure 1) to determine if PinR mediates inversion of the promoter of SZO_08560. Each deletion was confirmed by PCR and sequencing across the deletion site. The amount of the invertible promoter region in both the A and B orientations was quantified by qPCR. Wild-type *Sz*H70 contained 96 % of SZO_08560 promoter copies in the A orientation and 4 % in the B orientation (Figure 2). Deletion of *pinR* (*ΔpinR* A) fixed the promoter in the A orientation, no copies of the promoter in the B orientation were identified in this mutant. Both *pinR* and the SZO_08560 promoter were deleted (*ΔpinR*) and then the SZO_08560 promoter was re-introduced in the B orientation to produce a mutant strain (*ΔpinR* B) that only contained the SZO_08560 promoter in the B orientation. Complementation of the *pinR* deletion in strains *ΔpinR* A and *ΔpinR* B by insertion of a copy of *pinR* under the control of its native promoter into the pseudogene SZO_07770 (*ΔpinR* A c and *ΔpinR* B c, respectively) restored inversion of the promoter of SZO_08560 in the *ΔpinR* A c strain such that 0.8 % of promoter copies were in the B orientation. However, inversion of the promoter of SZO_08560 was not restored in the *ΔpinR* B c strain (Figure 2). The *ΔpinR* A and *ΔpinR* A c strains were found to have a significantly reduced growth rate when compared with *Sz*H70 and the other mutant strains ($P = 0.006$) (Figure S1).

The orientation of the invertible region determines SZO_08560 transcription

To determine if the promoter of SZO_08560 was more active in the A or B orientation, total RNA isolated from each mutant strain was reverse transcribed and used to quantify the transcription of SZO_08560 by qPCR. Data were normalized based on the number of *gyrA* transcripts in each triplicate sample. The transcription of SZO_08560 in wild-type *Sz*H70 was found to be equivalent to that of *gyrA* in this strain (Figure 3). Deletion of *pinR* such that the promoter of SZO_08560 was fixed in the A orientation (*ΔpinR* A) caused a reduction of SZO_08560 transcription to 0.7 % of wild-type levels ($P < 0.0001$). However, fixation of the promoter of SZO_08560 in the B orientation increased

SZO_08560 transcription to 189 % of wild-type levels ($P < 0.0001$). Deletion of SZO_08560 abolished its transcription, which was restored to 50 % of wild-type transcription levels by complementation through the insertion of a copy of SZO_08560 under the control of its native promoter in the B orientation into the pseudogene SZO_07770 (strain Δ08560 c).

Deletion or increased transcription of SZO_08560 alone did not significantly affect attachment of *S. zooepidemicus* to equine tissues

The number of bacteria recovered from explants of equine trachea two hours post-infection with the wild-type *Sz*H70 strain did not significantly differ from the number recovered from those infected with the mutant strains (Figure 4). However, the reduction in the amount of Δ08560 and Δ08560 c strains recovered relative to *Sz*H70 approached statistical significance ($P = 0.0859$ and $P = 0.0883$, respectively). A higher number of bacteria were recovered from those explants infected with the *ΔpinR* B mutant, which transcribes the most SZO_08560 relative to *Sz*H70, although this was also not statistically significant ($P = 0.67$).

PinR inverts sequences distant to the promoter of SZO_08560

Analysis of the draft genome sequence of *Sz*2329 using the ACT confirmed that this strain contained a deletion of *pinR* and the majority of its SZO_08560 homologue, which was identical to that previously identified in *Se*4047 (Holden *et al.*, 2009). The assembled *Sz*2329 draft genome contained one example of altered locus architecture consisting of an inversion of a 7,137 bp region containing the major and minor pilin genes, but not the AraC-like regulator or associated sortases of FimIV (Figure 5). The inversion occurred in 100 % of the sequencing reads covering this region, which was represented in a single contig. The inverted region of FimIV is flanked by a six-base inverted repeat (TAGAAA), which partially (TAGA) matches the 10 base inverted repeat (G**TAGA**CTTTA) that flanks the invertible promoter region upstream of SZO_08560 in *Sz*H70 (Holden *et al.*, 2009).

To determine if inversion of the FimIV locus was actively occurring in other strains of *S. zooepidemicus*, PCR primers were designed to amplify a product when the FimIV region was in either the original orientation as annotated in the *Sz*H70 genome, or inverted orientation. A collection of ten FimIV-containing strains were screened by PCR for the occurrence of amplification products suggesting the presence of DNA in both orientations (Table S1). Active inversion of the FimIV sequence was identified in *S. zooepidemicus* strains *Sz*1770, *Sz*B260863 and *Sz*H050840501. Only the inverted FimIV PCR product was amplified from strain *Sz*2329. The PCR products were purified and sequenced, confirming that the inverted region in FimIV was flanked by the same inverted repeat (TAGAAA) in all strains.

To determine if PinR was mediating FimIV inversion, *pinR* was deleted from *Sz*1770 by allelic replacement mutagenesis. Deletion of *pinR* was confirmed by PCR and sequencing across the deletion site. The number of original and inverted copies of FimIV in the *ΔpinR* mutant, wild-type *Sz*1770 and *Sz*2329 were quantified by qPCR and normalized to *gyrA*. Wild-type *Sz*1770 contained 0.01 % (1:10,000) of FimIV copies in the inverted orientation (Figure 6). Deletion of *pinR* from strain *Sz*1770 prevented inversion of the FimIV region, yielding 100 % of qPCR products in the original orientation.

Discussion

The surface architecture of *S. zooepidemicus* is likely to be crucial to its ability to adapt and interact with mammalian hosts and the wider environment in order to fulfill the requirements of its opportunistic lifestyle. The organism must survive outside a host, in drinking water or on soil, grass and other surfaces in competition with a vast array of other micro-organisms, whilst remaining in a state of readiness to infect a susceptible new host should the opportunity arise. The population of *S. zooepidemicus* infects many different mammalian hosts and tissues. Indeed, individual strains are themselves capable of infecting multiple hosts and zoonotic transmission, for example from an infected dog to a veterinary nurse, has been demonstrated (Abbott *et al.*, 2010). *S. zooepidemicus*

persists in the tonsils or on the mucosal surfaces of recovered horses in the face of a mature immune response, increasing the likelihood of onward transmission (Lindahl *et al.*, 2013). Therefore, the ability of *S. zooepidemicus* to modulate its surface is likely to be essential to its long-term survival.

Here we present evidence that the inversion of the promoter of SZO_08560 is performed by PinR and demonstrate that inversion acts as a switch, controlling transcription of SZO_08560. SZO_08560 contains an N'-terminal signal sequence, C'-terminal LPXTG sortase-processing motif and four Listeria-Bacteroides repeat Pfam domains (PF09479) with structural similarity to mucin-binding proteins (Ebbes *et al.*, 2011). Whilst the exact receptor bound by SZO_08560 remains unknown, the reduced ability of SZO_08560 mutants to attach to explants of equine trachea, which approached statistical significance, suggests that SZO_08560 is likely to play a role in the attachment of *S. zooepidemicus* to host tissue.

The *ΔpinR* A mutant lacks *pinR* with the SZO_08560 promoter orientated in the A direction and had a slow growth rate. One explanation for the slow growth of this strain is interference of the transcription of SZO_08540 or SZO_08530 by the SZO_08560 promoter, which could be enhanced by its closer proximity to these coding sequences following the deletion of *pinR*. SZO_08540 encodes a conserved hypothetical protein, whilst SZO_08530 encodes RpsP, the 30S ribosomal protein S16. Interestingly, the *ΔpinR* B mutant, which lacks *pinR*, whilst orientating the SZO_08560 promoter in the B direction had a normal growth rate, as did the *ΔpinR* mutant, which lacks both *pinR* and the SZO_08560 promoter (Figure S1). Therefore, the inversion of the SZO_08560 promoter from the A to the B orientation in strain *ΔpinR* A c, which contains a complementing copy of *pinR* may be preferred as it is likely to yield strains with a normal growth rate. However, the inversion of the SZO_08560 promoter from the B to the A orientation in strain *ΔpinR* B c was not detected, most likely as the resultant mutants would have a slower growth rate.

Analysis of the *Sz*2329 genome sequence, which lacks *pinR*, identified a disruptive internal inversion of the FimIV locus, which was bordered by short inverted repeats that shared a four-base motif (TAGA) with the SZO_08560 promoter. The FimIV locus encodes an AraC-like regulator, three putative sortase enzymes, a putative exported protein and three putative sortase-processed proteins that are predicted to form a surface pilus structure (Holden *et al.*, 2009). Screening of a panel of *S. zooepidemicus* isolates identified three strains with active FimIV inversion. The deletion of *pinR* in one of these strains, *Sz*1770, stopped FimIV inversion revealing a wider role for PinR in the global regulation of bacterial surface components and highlighting redundancy in the DNA sequences of the inverted repeats. Our data suggest that the ancestor of *Sz*2329 contained a functional copy of *pinR* and was actively inverting the FimIV region until the loss of *pinR* fixed this region in the position that it was in at the time. FimIV was present in 81 (58%) of 140 isolates of *S. zooepidemicus* that were tested and is missing from the *S. equi* genome [1], indicating that its loss from the genomes of some strains may be beneficial in the particular environments that they occupy. It is interesting that inversion of the FimIV region was not observed in strain *Sz*H70, despite this strain actively inverting the SZO_08560 promoter via PinR, suggesting that co-factors may assist PinR to invert alternative substrates. Variation in the sequence of inverted repeats and size of the inverted regions of DNA confounds the *in silico* identification of substrate sites and further research is required to identify the range of PinR substrates and the consequences of inversion on the properties of the variants produced.

The data presented here suggest that PinR plays an important role in modulating the surface architecture of *S. zooepidemicus* forming a mixture of distinct phenotypes, which provides this organism with a bet-hedging solution to survival in fluctuating environments (Stewart & Cookson, 2012). PinR of *S. zooepidemicus* shares >60 % predicted amino acid identity with putative resolvases including those encoded by strains of *Streptococcus anginosus*, *Streptococcus pneumonia* (strain GA17545), *Streptococcus constellatus*, *Streptococcus suis*, *Streptococcus mitis, Streptococcus ovis*, *Streptococcus pseudopneumoniae*, *Peptoniphilus indolicus*, *Eubacterium saphenum*, *Parvimonas*

micra, *Eggerthia catenaformis*, *Gemella bergeri*, *Gamella cuniculi*, *Bulleidia extructa*, *Enterococcus faecium*, *Erysipelotrichaceae bacterium*, *Clostridiales bacterium*, *Gardnerella vaginalis*, *Coprobacillus* sp., *Catenibacterium* sp. and *Mogibacterium* sp., suggesting that serine recombinase-mediated modulation of surface architecture is a mechanism that is widely adopted by other Gram-positive bacteria. It is intriguing to note that PinR shares 27% amino acid sequence identity and conserved serine residue with the site-specific recombinase of *Bacteroides fragilis*, FinA (also known as Mpi), which modulates the production of several surface components in this Gram-negative bacterium by inversion of promoter sequences (Cerdeno-Tarraga *et al.*, 2005; Coyne *et al.*, 2003). Our data provide the first evidence to suggest that the reversible ON-OFF phenotype known as phase variation can be mediated by a recombinase in streptococci.

Acknowledgements

This project was funded by a project grant from the Horserace Betting Levy Board (vet/prj/751). *Sz*H050840501 was kindly provided by Dr. Androulla Efstratiou of Public Health England.

References

Abbott, Y., Acke, E., Khan, S., Muldoon, E. G., Markey, B. K., Pinilla, M., Leonard, F. C., Steward, K. & Waller, A. (2010). Zoonotic transmission of *Streptococcus equi* subsp. *zooepidemicus* from a dog to a handler. *J Med Microbiol* **59**, 120-123.

Abraham, J. M., Freitag, C. S., Clements, J. R. & Eisenstein, B. I. (1985). An invertible element of DNA controls phase variation of type 1 fimbriae of *Escherichia coli*. *Proc Natl Acad Sci U S A* **82**, 5724-5727.

Balter, S., Benin, A., Pinto, S. W. & other authors (2000). Epidemic nephritis in Nova Serrana, Brazil. *Lancet* **355**, 1776-1780.

Carver, T. J., Rutherford, K. M., Berriman, M., Rajandream, M. A., Barrell, B. G. & Parkhill, J. (2005). ACT: the Artemis Comparison Tool. *Bioinformatics* **21**, 3422-3423.

Cerdeno-Tarraga, A. M., Patrick, S., Crossman, L. C. & other authors (2005). Extensive DNA inversions in the *B. fragilis* genome control variable gene expression. *Science* **307**, 1463-1465.

Chalker, V. J., Brooks, H. W. & Brownlie, J. (2003). The association of *Streptococcus equi* subsp. *zooepidemicus* with canine infectious respiratory disease. *Vet Microbiol* **95**, 149-156.

Coyne, M. J., Weinacht, K. G., Krinos, C. M. & Comstock, L. E. (2003). Mpi recombinase globally modulates the surface architecture of a human commensal bacterium. *Proc Natl Acad Sci U S A* **100**, 10446-10451.

Ebbes, M., Bleymuller, W. M., Cernescu, M., Nolker, R., Brutschy, B. & Niemann, H. H. (2011). Fold and function of the InlB B-repeat. *J Biol Chem* **286**, 15496-15506.

Hamilton, A., Robinson, C., Sutcliffe, I. C., Slater, J., Maskell, D. J., Davis-Poynter, N., Smith, K., Waller, A. & Harrington, D. J. (2006). Mutation of the maturase lipoprotein attenuates the virulence of *Streptococcus equi* to a greater extent than does loss of general lipoprotein lipidation. *Infect Immun* **74**, 6907-6919.

Heather, Z., Holden, M. T., Steward, K. F., Parkhill, J., Song, L., Challis, G. L., Robinson, C., Davis-Poynter, N. & Waller, A. S. (2008). A novel streptococcal integrative conjugative element involved in iron acquisition. *Mol Microbiol* **70**, 1274-1292.

Holden, M. T., Heather, Z., Paillot, R. & other authors (2009). Genomic evidence for the evolution of *Streptococcus equi*: host restriction, increased virulence, and genetic exchange with human pathogens. *PLoS Pathogens* **5**, e1000346.

Hong, C. B., Donahue, J. M., Giles, R. C., Jr. & other authors (1993). Etiology and pathology of equine placentitis. *J Vet Diagn Invest* **5**, 56-63.

Lindahl, S. B., Aspan, A., Baverud, V., Paillot, R., Pringle, J., Rash, N. L., Soderlund, R. & Waller, A. S. (2013). Outbreak of upper respiratory disease in horses caused by *Streptococcus equi* subsp. *zooepidemicus* ST-24. *Vet Microbiol* **166**, 281-285.

Maguin, E., Prevost, H., Ehrlich, S. D. & Gruss, A. (1996). Efficient insertional mutagenesis in lactococci and other gram-positive bacteria. *J Bacteriol* **178**, 931-935.

Pesavento, P. A., Hurley, K. F., Bannasch, M. J., Artiushin, S. & Timoney, J. F. (2008). A clonal outbreak of acute fatal hemorrhagic pneumonia in intensively housed (shelter) dogs caused by *Streptococcus equi* subsp. *zooepidemicus*. *Vet Pathol* **45**, 51-53.

Rasmussen, C. D., Haugaard, M. M., Petersen, M. R., Nielsen, J. M., Pedersen, H. G. & Bojesen, A. M. (2013). *Streptococcus equi* subsp. *zooepidemicus* isolates from equine infectious endometritis belong to a distinct genetic group. *Vet Res* **44**, 26.

Smith, K. C., Blunden, A. S., Whitwell, K. E., Dunn, K. A. & Wales, A. D. (2003). A survey of equine abortion, stillbirth and neonatal death in the UK from 1988 to 1997. *Equine Vet J* **35**, 496-501.

Stewart, M. K. & Cookson, B. T. (2012). Non-genetic diversity shapes infectious capacity and host resistance. *Trends Microbiol* **20**, 461-466.

Velineni, S., Desoutter, D., Perchec, A. M. & Timoney, J. F. (2014). Characterization of a mucoid clone of *Streptococcus zooepidemicus* from an epizootic of equine respiratory disease in New Caledonia. *Vet J*.

Webb, K., Jolley, K. A., Mitchell, Z., Robinson, C., Newton, J. R., Maiden, M. C. & Waller, A. (2008). Development of an unambiguous and discriminatory multilocus sequence typing scheme for the *Streptococcus zooepidemicus* group. *Microbiology* **154**, 3016-3024.

Webb, K., Barker, C., Harrison, T., Heather, Z., Steward, K. F., Robinson, C., Newton, J. R. & Waller, A. S. (2013). Detection of *Streptococcus equi* subspecies *equi* using a triplex qPCR assay. *Vet J* **195**, 300-304.

Wood, J. L., Burrell, M. H., Roberts, C. A., Chanter, N. & Shaw, Y. (1993). Streptococci and Pasteurella spp. associated with disease of the equine lower respiratory tract. *Equine Vet J* **25**, 314-318.

Wood, J. L., Newton, J. R., Chanter, N. & Mumford, J. A. (2005). Association between respiratory disease and bacterial and viral infections in British racehorses. *J Clin Microbiol* **43**, 120-126.

Figure 1. Schematic of the modified strains produced during this study. A) The structure of the pinR/SZO_08560 region for wild-type *Sz*H70 is shown. The *ΔpinR* A mutant lacks *pinR*, which was predicted to fix the invertible region in the A orientation. The *ΔpinR* mutant lacks both *pinR* and the invertible region. The *ΔpinR* B mutant was generated from the *ΔpinR* mutant by introducing the invertible region in the B orientation. The *Δ*08560 mutant lacks SZO_08560. The direction of the SZO_08560 promoter is indicated by the black arrow. B) The structure of the SZO_07770 region for wild-type *Sz*H70 is shown. The *ΔpinR* A c and *ΔpinR* B c mutants were generated by inserting a copy of *pinR* under the control of its native promoter into the *ΔpinR* A and *ΔpinR* B mutants, respectively. The *Δ*08560 c mutant was generated by inserting a copy of SZO_08560 downstream of the invertible region in the B orientation into the *Δ*08560 mutant. The top DNA strand is shown by the solid black line and bottom strand by the broken black line. The inverted repeats are shown in red and blue boxes. The direction of the SZO_08560 promoter is indicated by the black arrow.

Figure 2. Orientation of the invertible region as determined by qPCR. The Log10 of the number of copies of the invertible region in the A or B orientation are shown following normalisation of the samples based on the amount of *gyrA*. Error bars indicate the standard deviation.

Figure 3. Transcription of SZO_08560 in the mutant strains. The number of transcript copies of SZO_08560 were quantified by qPCR and normalised relative to the amount of *gyrA*. Error bars indicate the standard deviation.

Figure 4. Attachment of *Sz*H70 and mutant strains to explants of equine trachea. Error bars indicate 95% confidence intervals.

Figure 5. Partial inversion of the FimIV locus in strain *Sz*2329 relative to the *Sz*H70 reference genome visualised using the Artemis Comparison Tool (Carver *et al.*, 2005). The coloured bars separating each genome (blue and red) represent similarity matches identified by reciprocal TBLASTX

505 analysis, with a score cut-off of 100. Blue lines link matches in the same orientation; red lines link
506 matches in the reverse orientation.
507
508 **Figure 6. Inversion of FimIV pre- and post-deletion of *pinR*.** Graph showing the Log10 of mean
509 copies of FimIV in the original annotated orientation of the *Sz*H70 genome, and disrupted inverted
510 orientation normalised to the number of copies of *gyrA*. Error bars indicate the standard deviation.
511
512 **Table S1. List of strains used in this study.** AFHP: acute fatal haemorrhagic pneumonia.
513
514 **Table S2. List of oligonucleotide primers used in this study.** Restriction sites in primers used to
515 clone target sites for gene deletion are underlined.
516
517 **Figure S1. Growth of *Sz*H70, *ΔpinR* A, *ΔpinR*, *ΔpinR* B, *Δ*8560, *ΔpinR* A c and *ΔpinR* B c strains**
518 **in Todd Hewitt Broth.** Error bars indicate the standard deviation.
519

Figure 1
Click here to download Figure: Figure 1.pptx

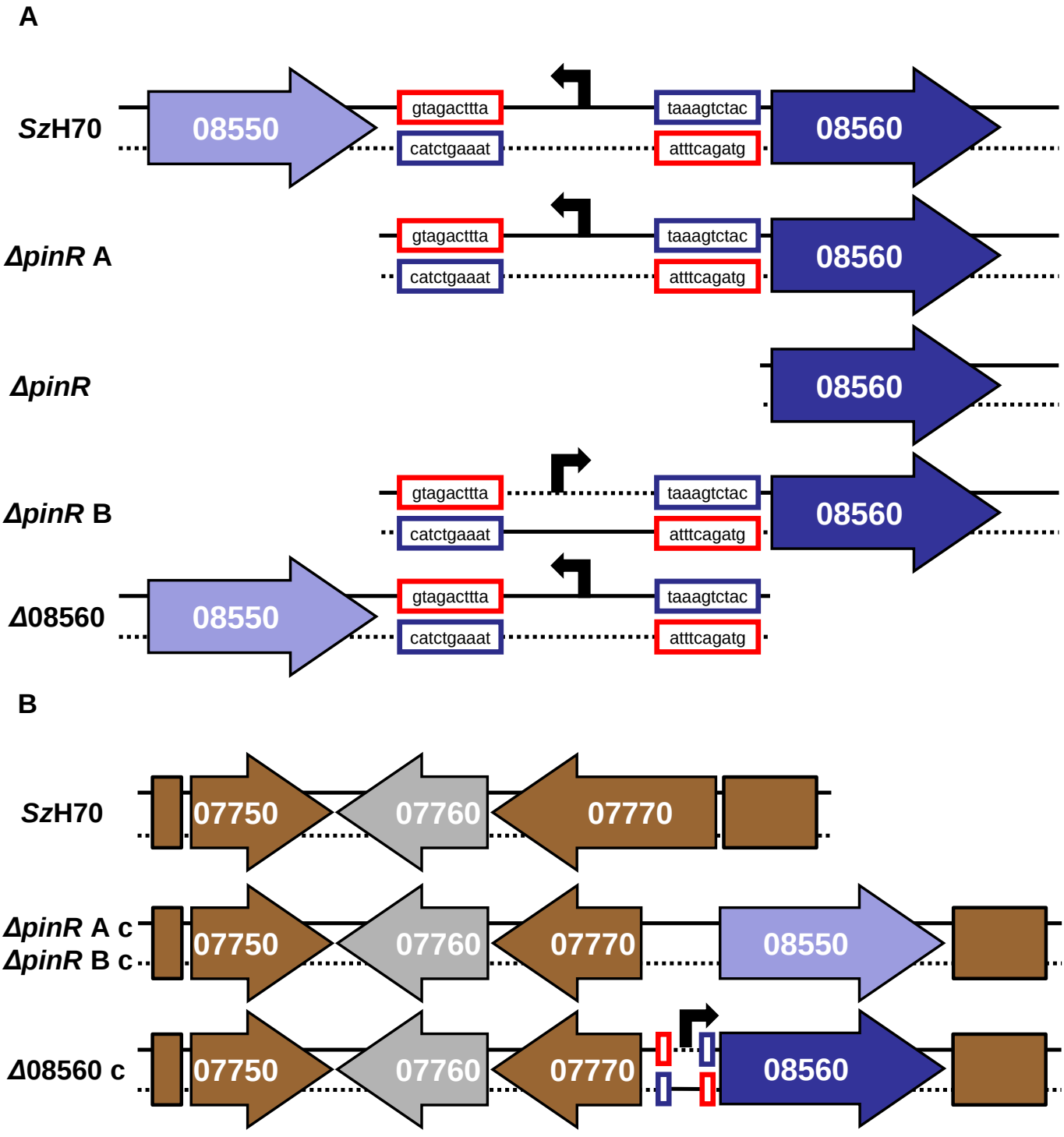


Figure 2
Click here to download Figure: Figure 2.pptx

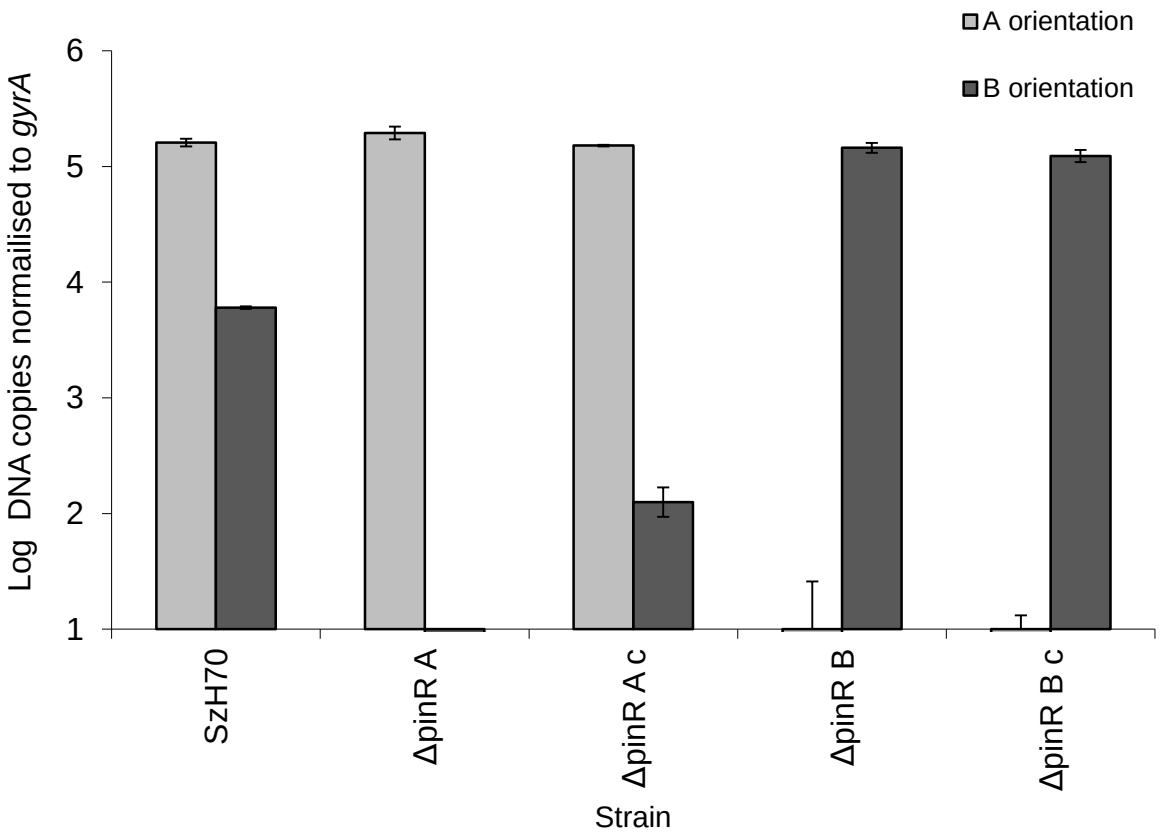


Figure 3
Click here to download Figure: Figure 3.pptx

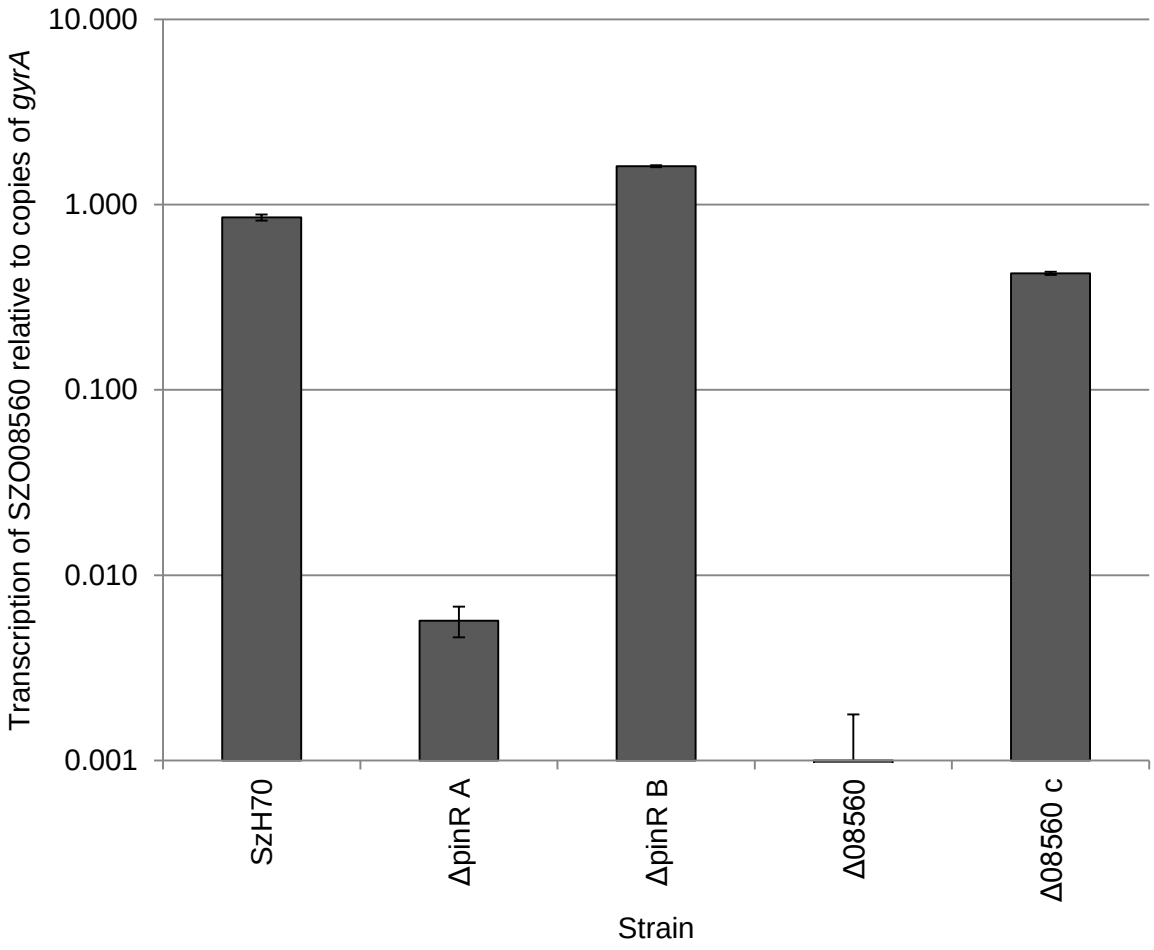


Figure 4
Click here to download Figure: Figure 4.pptx

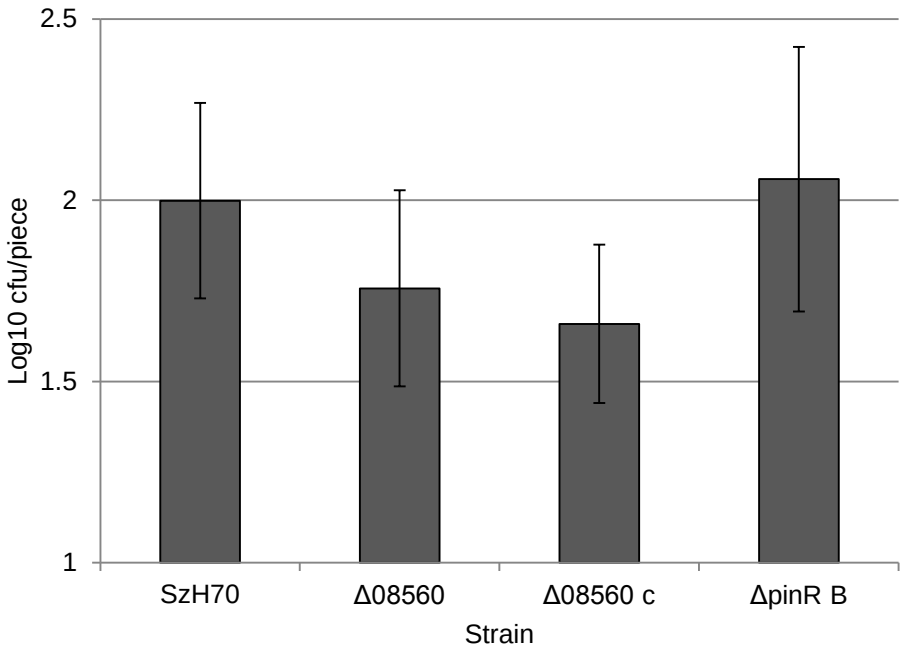


Figure 5
Click here to download Figure: Figure 5.pptx

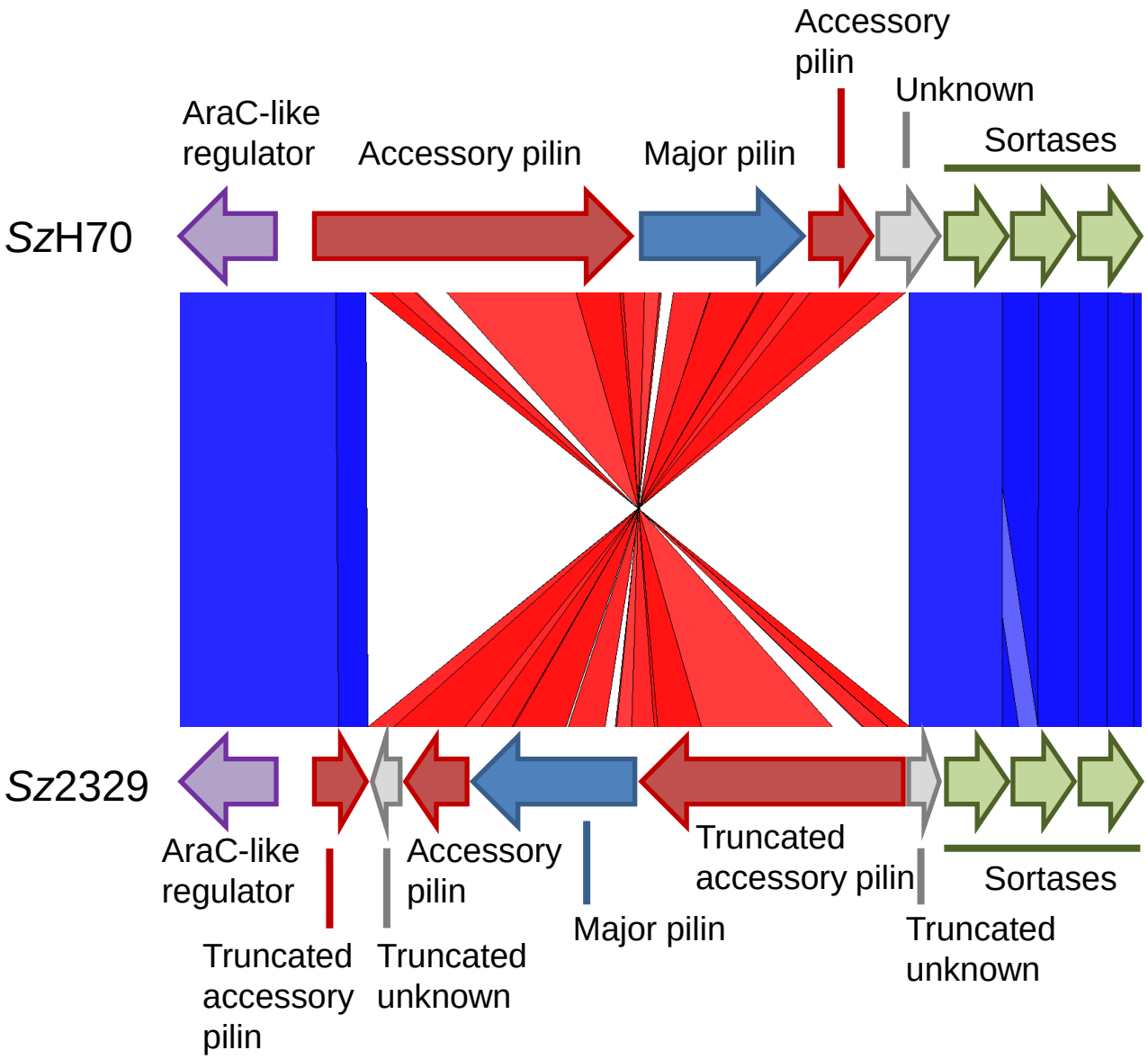
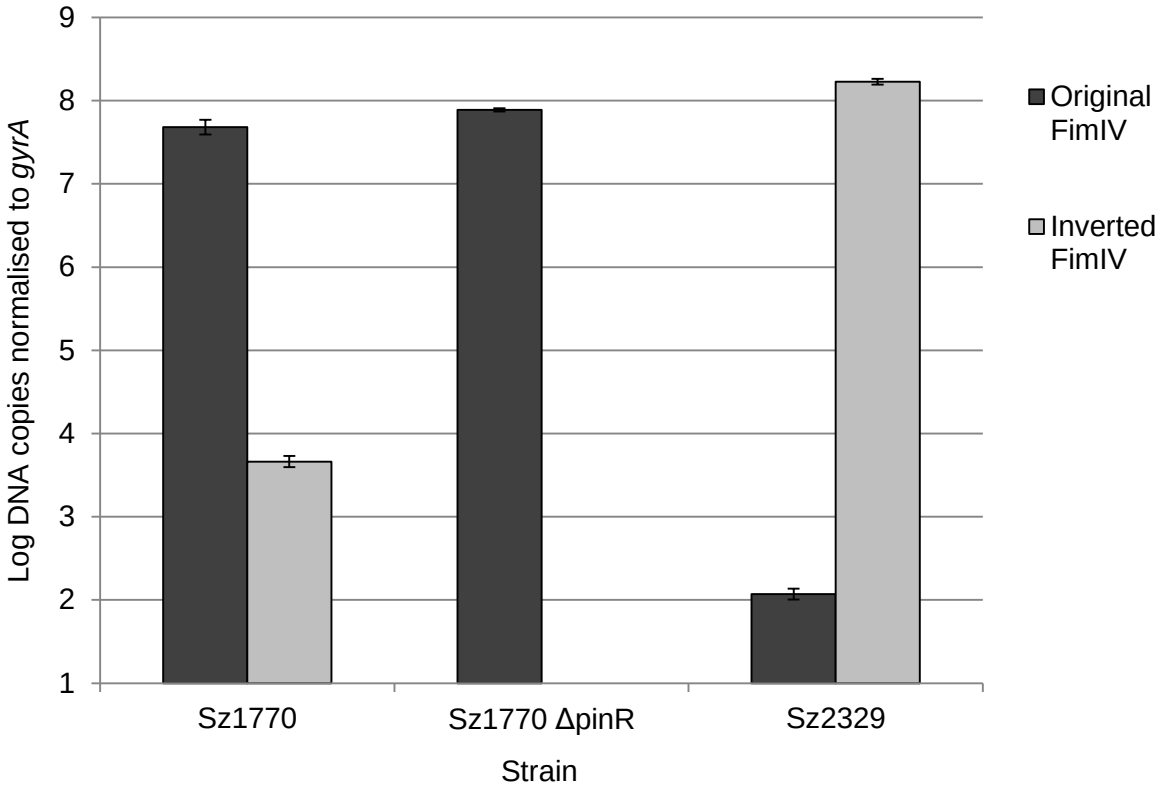


Figure 6
Click here to download Figure: Figure 6.pptx



Supplementary Material Files

Click here to download Supplementary Material Files: Supplementary data.pdf