Rapid Evolution of Virulence and Drug Resistance in the Emerging Zoonotic Pathogen *Streptococcus suis*

Matthew T. G. Holden16, Heidi Hauser1, Mandy Sanders1, Thi Hoa Ngo2, Inna Cherevach1, Ann Cronin1, Ian Goodhead17✉, Karen Mungall1, Michael A. Quail1, Claire Price1, Ester Rabbinowitsch1, Sarah Sharp1, Nicholas J. Croucher1, Tran Bich Chieu2, Nguyen Thi Hoang Mai3, To Song Diep3, Nguyen Tran Chinh3, Michael Kehoe4, James A. Leigh5, Philip N. Ward6, Christopher G. Dowson7, Adrian M. Whatmore7✉, Neil Chanter8, Pernille Iversen9, Marcelo Gottschalk10, Josh D. Slater11, Hilde E. Smith12, Brian G. Spratt13, Jianguo Xu14, Changyun Ye14, Stephen Bentley1, Barclay G. Barrell1, Constance Schultsz2,15, Duncan J. Maskell16, Julian Parkhill1

1 The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge, United Kingdom, 2 Oxford University Clinical Research Unit, Hospital for Tropical Diseases, Ho Chi Minh City, Viet Nam, 3 Hospital for Tropical Diseases, Ho Chi Minh City, Viet Nam, 4 Institute for Cell and Molecular Biosciences, The Medical School, University of Newcastle upon Tyne, Newcastle upon Tyne, United Kingdom, 5 The School of Veterinary Medicine and Science, The University of Nottingham, Sutton Bonington Campus, Sutton Bonington, United Kingdom, 6 Nuffield Department of Clinical Laboratory Sciences, Oxford University, John Radcliffe Hospital, Headington, United Kingdom, 7 Department of Biological Sciences, University of Warwick, Coventry, United Kingdom, 8 Centre for Preventative Medicine, Animal Health Trust, Newmarket, United Kingdom, 9 Department of Biology, University of Copenhagen, Copenhagen, Denmark, 10 Groupe de recherche sur les maladies infectieuses du porc (GREMIP), Université de Montréal, Montréal, Québec, Canada, 11 Royal Veterinary College, Hatfield, United Kingdom, 12 Animal Sciences Group (ASG), Wageningen UR, UR, Lelystad, The Netherlands, 13 Department of Infectious Disease Epidemiology, Imperial College, London, United Kingdom, 14 State Key Laboratory for Infectious Disease Prevention and Control, National Institute for Communicable Disease Control and Prevention, China CDC, Beijing, China, 15 Academic Medical Center-Center for Poverty-Related Communicable Diseases, University of Amsterdam, Amsterdam, The Netherlands, 16 Department of Veterinary Medicine, University of Cambridge, Cambridge, United Kingdom

**Abstract**

**Background:** *Streptococcus suis* is a zoonotic pathogen that infects pigs and can occasionally cause serious infections in humans. *S. suis* infections occur sporadically in human Europe and North America, but a recent major outbreak has been described in China with high levels of mortality. The mechanisms of *S. suis* pathogenesis in humans and pigs are poorly understood.

**Methodology/Principal Findings:** The sequencing of whole genomes of *S. suis* isolates provides opportunities to investigate the genetic basis of infection. Here we describe whole genome sequences of three *S. suis* strains from the same lineage: one from European pigs, and two from human cases from China and Vietnam. Comparative genomic analysis was used to investigate the variability of these strains. *S. suis* is phylogenetically distinct from other *Streptococcus* species for which genome sequences are currently available. Accordingly, ~40% of the ~2 Mb genome is unique in comparison to other *Streptococcus* species. Finer genomic comparisons within the species showed a high level of sequence conservation; virtually all of the genome is common to the *S. suis* strains. The only exceptions are three ~90 kb regions, present in the two isolates from humans, composed of integrative conjugative elements and transposons. Carried in these regions are coding sequences associated with drug resistance. In addition, small-scale sequence variation has generated pseudogenes in putative virulence and colonization factors.

**Conclusions/Significance:** The genomic inventories of genetically related *S. suis* strains, isolated from distinct hosts and diseases, exhibit high levels of conservation. However, the genomes provide evidence that horizontal gene transfer has contributed to the evolution of drug resistance.


**Editor:** Adam J. Ratner, Columbia University, United States of America

**Received** March 11, 2009; **Accepted** April 22, 2009; **Published** July 15, 2009

**Copyright:** © 2009 Holden et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** This work was supported by the Wellcome Trust. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

*✉* E-mail: mh3@sanger.ac.uk

✉ Current address: School of Biological Sciences, University of Liverpool, Liverpool, United Kingdom

✉ Current address: Veterinary Laboratories Agency, Weybridge, United Kingdom
Introduction

*S. suis* is a Gram positive coccus that colonises pigs. While it is generally carried asymptomatically in adult pigs, it can cause severe systemic disease in pigs, manifested as a rapidly fatal sepsis associated with meningitis, polyarthritis and pneumonia. Why adult pigs carry the causative bacteria asymptomatically while piglets develop acute disease is unknown. The main site of carriage in the adult are the tonsils, but bacteria have also been isolated from the nasal cavities, the gastrointestinal tract and genitai tract. The carriage rate in adult pigs can approach 100% and this organism has been a worldwide problem for the pig industry for a number of years. *S. suis* has also been isolated from a range of other mammalian and avian species [1].

*S. suis* is an important zoonotic agent. The first case in humans was described in Denmark in 1968 [2]. Human infection with *S. suis* occurs sporadically in Europe and North-America and case reports suggest that it is almost exclusively related to occupational exposure to pigs or pork products. Incidences of human infection with *S. suis* are greater in S.E. Asia and China. Meningitis is the most common presentation in humans, but septicaemia and endocarditis are also seen. In areas of Vietnam, *S. suis* is the main cause of acute bacterial meningitis in adults [3] and it is the third most common cause of meningitis in Hong Kong. The incidence of *S. suis* infection in humans is almost certainly under-reported. Most cases have been described as being caused by serotype 2 isolates, but other serotypes can cause human disease [4].

Recently two outbreaks of severe acute disease in humans with high morbidity and mortality in humans have been reported in China [5], in both cases due to serotype 2 strains. An outbreak in 1998 killed 14 of 25 patients, and an outbreak in 2005 affected 204 people, killing 38 of them (19%), a mortality rate around two to three times that previously reported. A high proportion of people in these outbreaks had a toxic shock-like syndrome, and most of the deaths occurred in this group rather than in those suffering meningitis. Streptococcal toxic shock syndrome (STSS) was previously associated with *Streptococcus pyogenes*, so the emergence of STSS due to presumptive zoonotic *S. suis* infection is of considerable concern [6]. One isolate from each of these outbreaks has been sequenced [7] and this has identified a proposed pathogenicity island (PI) that may be involved in this particular clinical manifestation of *S. suis* infection, although this remains speculative at this stage. The PI was identified through comparison of the Chinese strains to the unfinished, unannotated genome sequence of strain P1/7 from the Sanger Institute. In this paper we present the completed fully annotated genome sequence of strain P1/7. In addition we have sequenced the genomes of two other *S. suis* strains, SC84 and BM407, which are human isolates from China and Vietnam respectively. Strain SC84 is a representative of the of 2005 outbreak in China [8].

Materials and Methods

Bacterial strains, growth and DNA isolation

*S. suis* strain P1/7 was isolated from an ante-mortem blood culture from a pig dying with meningitis [9], and is ST1 by MLST [10]. *S. suis* strain BM407 is also ST1, and was isolated from CSF from a human case of meningitis in Ho Chi Minh City, Vietnam in 2004 [3]. *S. suis* strain SC84 is ST7, which is closely related to ST1, and was isolated from a case of streptococcal toxic shock-like syndrome in Sichuan Province, China in 2005 [8]. Strain P1/7 is resistant to gentamicin, streptomycin, neomycin, nalidixic acid, and sulfamethoxazole, and sensitive to penicillin, ampicillin, cephalotin, erythromycin, tetracycline, trimethoprim-sulfamethoxazole, ciprofloxacin, and chloramphenicol. Strain BM407 is resistant to trimethoprim-sulfamethoxazole, tetracycline, erythromycin, azithromycin and chloramphenicol and susceptible to penicillin, ceftriaxone and vancomycin. Strain SC84 is resistant to tetracycline, and susceptible to penicillin, ampicillin, cefotaxime, ceftriaxone, cefepime, meropenem, levofloxacin, chloramphenicol, azithromycin, and vancomycin [11].

Bacteria were cultured in Todd-Hewitt-broth at 37°C for 18 h and pelleted at 10,000 x g. The cells were resuspended in 30 ml of lysis solution (10 mM NaCl, 20 mM Tris HCl pH 8, 1 mM EDTA, 0.5% SDS) and incubated at 50°C overnight. Three ml of 5 M sodium perchlorate was added and incubated for 1 h at ambient temperature. After phenol chloroform extraction the DNA was precipitated with ethanol, spooled into deionised water and stored at −20°C. DNA was also extracted using a genomic DNA extraction kit (G-500, Qiagen).

Whole genome sequencing

The genome of *S. suis* strain P1/7 was sequenced to approximately 8-fold coverage, from pUC18 (insert size 2.8–3.3 kb), and pMAQ1b_SmaI (insert size 3.0–3.5 kb) genomic shotgun libraries using big-dye terminator chemistry on ABI3730 automated sequencers. End sequences from large insert BAC libraries in pBACehr (insert size 10–25 kb) and pBACe3.6 (insert size 12–15 kb) were used as a scaffold. All repeat regions were bridged by read-pairs or end-sequenced polymerase chain reaction (PCR) products.

*S. suis* SC84 and BM407 genomes were sequenced using multiple sequencing technologies to generate contiguous drafts. The bulk of the sequencing for both projects was generated using 454/Roche GS20 and Solexa sequencing platforms. The 454 sequencing of strain SC84 generated 399,145 reads with an average length of 102.7 bp, which assembled de novo into 281 non-redundant contigs using the 454/Roche Newbler assembly program. The 454 sequencing of strain BM407 generated 223,282 reads with an average length of 103.1 bp, which assembled de novo into 1939 non-redundant contigs. The Solexa sequencing of strain SC84 generated 32,016,513 reads with a length of 37 bp, and of strain BM407 generated 1,769,043 reads with a length of 37 bp which were each assembled by alignment to a reference (strain P1/7) sequence using the saha2 [12] alignment program. Data from the 2 technologies, were merged to produce 39 non-redundant contigs in SC84, and 390 non-redundant contigs in BM407.

The merging process involved using the random in silico sheering of the 454 de novo assembly consensus contigs to create 500 bp consensus reads with 250 bp overlaps. This generated a fasta tiling path representing each of the 454 de novo assembly contigs. These were then assembled together with all the capillary shotgun reads using phrap2gap [13] to merge the two technologies.

The contigs were reordered based on BLAST [14] alignments with strain P1/7. The gaps between these contigs were closed by directed PCR and the products sequenced with big dye terminator chemistry on ABI3730 capillary sequencers. Aligned Solexa sequencing was merged to form assembly and also confirm sequence in 454 only regions. For the SC84 project capillary coverage was 5×, 454 coverage was 19× and Solexa coverage was 190×, and for the BM407 project capillary coverage was 5×, 454 coverage was 11× and Solexa coverage was 20×.

The sequences and annotations of the *Streptococcus suis* P1/7, SC84 and BM407 genomes have been deposited in the EMBL database under accession numbers AM946016, FM252031, and FM252032 and FM252033 respectively.
The sequence was annotated using Artemis software [15]. Initial coding sequence (CDS) predictions were performed using Orpheus [16], Glimmer2 [17], and EasyGene [18] software. These predictions were amalgamated, and codon usage, positional base preference methods and comparisons to the non-redundant protein databases using BLAST [14] and FASTA [19] software were used to refine the predictions. The entire DNA sequence was also compared in all six potential reading frames against UniProt, using BLASTX [14] to identify any possible coding sequences previously missed. Protein motifs were identified using Pfam [20] and Prosite [21], transmembrane domains were identified with TMHMM [22], and signal sequences were identified with SignalP version 2.0 [23]. tRNAs were identified using tRNAscan-SE [24]; stable RNAs were identified using BLASTN and TBLASTX comparisons [14]. The complete genome was also compared in all six potential reading frames against the EMBL nucleotide database; tRNAs were identified using tRNAscan-SE [24]; stable RNAs were identified using tRNAscan-SE [24].

Comparative genomics

Comparison of the complete genome sequences was facilitated using the Artemis Comparison Tool (ACT) [26] which enabled the visualization of BLASTN and TBLASTX comparisons [14] between the genomes. Orthologous proteins were identified as reciprocal best matches using FASTA [19] with subsequent manual curation. Pseudogenes had one or more mutations that would prevent correct translation; each of the inactivating mutations was subsequently checked against the original sequencing data. Small scale variation including: SNPs and insertions and deletions, were identified by using the SNP detection pipeline of MUMmer [27]. Clonalframe v.1.1 was used to identify regions of the S. suis chromosomes associated with recombination [28].

Streptococcus sequences used for comparative genomic analysis were: S. suis 05ZYH33 (accession number CP000407) [7], S. suis 98HAH33 (accession number CP000408) [7], S. pyogenes Manfredo (accession number AM295007) [29], Streptococcus equis 4047 (accession number FM204883) [30], Streptococcus uberis 0140 (accession number AM946015) [31], Streptococcus thermophilus CNRZ1066 (accession number CP000024) [32], Streptococcus pneumoniae TIGR4 (accession number AE005672) [33], Streptococcus gordonii SK36 (accession number CP000387) [34], Streptococcus mutans UA159 (accession number AE014133) [35], Streptococcus agalactiae NEM316 (accession number AL732656) [36], and Streptococcus gordoni subsp. Challis str. CH1 (accession number CP000725) [37]. The sequences were also compared with Lactococcus lactis subsp. lactis IL1403 (accession number AE005176) [38].

Phylogenetic analysis

Sequences alignment were made using R-coffee [39] and the alignments inspected and edited using Seaview (v.4.0) [40]. Bayesian analysis was performed using Mr Bayes [41,42]. The model of sequence evolution used was the general time reversible (GTR) model with gamma-distributed rate variation. The model of sequence evolution used was the generalized time-reversible (GTR) model with gamma-distributed rate variation. The model of sequence evolution used was the generalized time-reversible (GTR) model with gamma-distributed rate variation. The model of sequence evolution used was the generalized time-reversible (GTR) model with gamma-distributed rate variation. The model of sequence evolution used was the generalized time-reversible (GTR) model with gamma-distributed rate variation. The model of sequence evolution used was the generalized time-reversible (GTR) model with gamma-distributed rate variation. The model of sequence evolution used was the generalized time-reversible (GTR) model with gamma-distributed rate variation. The model of sequence evolution used was the generalized time-reversible (GTR) model with gamma-distributed rate variation.

Results

The genome of S. suis strain P1/7, a pig disease isolate

The genome of Streptococcus suis P1/7 consists of a single circular chromosome of 2,007,391 bp (Figure 1; accession number AM946016) containing 1,908 predicted protein coding sequences (CDSs; Table 1), 82 of which are pseudogenes or gene fragments (Table S1). S. suis is phylogenetically distinct among the streptococci for which genome sequences are available (Figure 2). Pairwise genomic comparisons reveal very little conservation of gene order with other species (data not shown). Orthologous comparisons against all of the other available streptococcal genomes identified that approximately 60% of the P1/7 genome is orthologous to CDSs in the genomes of other Streptococcus species (Table 2). The highest number of orthologous matches was identified in S. sanguinis (66.1%), followed by S. gordoni (65.6%).

Virulence genes in the P1/7 genome

The mechanisms by which S. suis cause disease in humans and pigs are poorly understood. Many putative virulence functions have been identified, however no single function is crucial for the pathogenesis of S. suis. Many putative virulence genes are only present, or expressed, in some S. suis isolates, suggesting that combinations of virulence factors may be more important than individual functions. Additionally, several of the functions may be important for colonization and carriage, rather than disease. Analysis of the S. suis genomes reveals an array of CDSs encoding potentially important functions.

Polysaccharides and capsule. Like most other pathogens that can cause meningitis, S. suis expresses a polysaccharide capsule. P1/7 is serotype 2 and contains a locus (SSU0539 to SSU0544) for the biosynthesis of type 2 capsule [43,44]. The structure of the S. suis serotype 2 capsule is unknown, but it is composed of glucose, galactose, N-acetylgalactosamine, rhamnose and sialic acid in a ratio of 1:3:1:1:1 [45]. Although 35 serotypes of S. suis have been described, type 2 is most frequently associated with disease in pigs and humans [3,46–48]. Capsule plays a role in the virulence of S. suis; type 2 capsule mutants are avirulent in murine and pig models of infections [43,49]. The capsule itself can
mediate prostaglandin E2 (PGE2) and matrix metalloproteinase 9 (MMP-9) production by human macrophages [50] suggesting a role in pathogenesis of disease.

The P1/7 genome also contains a cluster of genes that appears to encode the components required for the biosynthesis of a putative exopolysaccharide (SSU1111 to SSU1133; Figure S1a). Included in this are CDSs encoding enzymes for synthesis of the exopolysaccharide building blocks, for processing and for export. Analysis of the cluster suggests that it is involved in the synthesis of a rhamnose-glucose-like polysaccharide (RGP), and is similar to components of RGP biosynthesis identified in S. mutans (Figure S1b). Four loci have been identified in S. mutans that encode proteins necessary for RGP synthesis: RmlA, RmlC, RmlB, and RmlD direct the synthesis of dTDP-L-rhamnose from D-glucose-
monocyes in vivo to release inflammatory cytokines, including tumor necrosis factor alpha and interleukin-1β [58], induce NO synthase activity [59], and confer resistance to phagocytosis by human polymorphonuclear leukocytes [60]. Elevated NO production has been observed in animal models of bacterial meningitis and also in human patients with the disease [61, 62]. In a mouse model of S. suis infection, high levels of inflammatory cytokines were observed both systemically and in the central nervous system [63]. Distinct from the type 2 capsule, it is possible that the exopolysaccharide encoded by this cluster may have an important role in virulence and may contribute to the inflammatory response during S. suis infection.

**Protein secretion and secreted proteins.** In the publication by Chen et al. describing the genomes of two Chinese S. suis isolates [7], they presented possible virulence-associated factors or pathways in S. suis, including strain P1/7. This analysis identified thirteen components of a Type II secretion system and one component of a Type III secretion system in the genome of strain P1/7. Our analysis of the genome failed to identify components of either secretion system, although components of Sec and SRP pathways were identified. Neither Type II, nor Type III secretion systems have been identified in Gram-positive bacteria, and it is therefore very unlikely that S. suis secretes proteins via these pathways. It would appear that the description of these secretion system components in S. suis is a case of mistaken identity due to sub-optimal transitive annotation. The S. suis genome contains CDSs that have weak similarity to components of these Gram-negative protein secretion systems, but these genes in S. suis encode proteins that are involved in the transport of other molecules, or are components of functionally distinct systems. The observed similarity may more to distant evolutionary relationships rather than conservation of function [64].

Suilysin (SSU1231) is a secreted thiol-activated cytolysin [65] produced by some strains of S. suis [66]. This protein is part of a family of cytolysins that play a role in disease caused by several bacterial pathogens [67]. The role that suilysin plays in S. suis disease is unclear. Immunization of pigs with suilysin provided partial protection from S. suis infection [68]. However, while a suilysin mutant was avirulent in a mouse infection model, it was only slightly attenuated in a model of infection in pigs [69].

**Surface proteins.** Several of the previously characterized S. suis virulence factors are cell wall-anchored proteins. These sortase-processed proteins contain a C-terminal LPXTG-type processing motif. The genome of strain P1/7 contains a large number (28) of proteins of this type: all but five contained C-terminal LPXTG motifs followed by non-polar and basic amino acid motifs. The other five putative surface proteins (SSU1828, SSU1186, SSU1888 and SSU1889) contained LPXTS, terminal LPXTG motifs followed by non-polar and basic amino acid motifs. The other five putative surface proteins (SSU1828, SSU1186, SSU1888 and SSU1889) contained LPXTS, terminal LPXTG motifs followed by non-polar and basic amino acid motifs. The other five putative surface proteins (SSU1828, SSU1186, SSU1888 and SSU1889) contained LPXTS, terminal LPXTG motifs followed by non-polar and basic amino acid motifs. These genes encode proteins that are involved in glucose side chain formation [56], therefore it is likely that the polysaccharide produced by P1/7 has a different repeat unit structure to that produced by S. mutans.

The cluster does not appear to encode an obvious repeat unit polymerase. Downstream of the flippase is a membrane protein (SSU1118) that has weak similarity matches to proteins of unknown function. Due to the sequence diversity of repeat unit polymerases, it is not always possible to annotate them by sequence similarity. Hydrophobicity plot profile comparison has been used to identify tentatively these polymerases [57]. A comparison of the hydrophobicity plot of SSU1118 with those of known polymerases identified a similar profile (data not shown).

RGP s constitute serospecific-antigens that form the basis of some streptococcal serotyping. These molecules are associated with modulating host immune cell interactions, and may have a role in eliciting sepsis. Streptococcal RGPs stimulate human

| Table 1. General properties of the genomes of S. suis strains P1/7, SC84 and BM407. |
|---------------------------------|--------|--------|--------|
|                                | P1/7   | SC84   | BM407  |
| Size (bp)                      | 2,007,491 | 2,095,898 | 2,146,229 |
| G+C content                    | 41.3%  | 41.1%  | 41.1%  |
| CDSs                            | 1,908  | 1,985  | 2,040  |
| Coding %                       | 85.1%  | 84.8%  | 83.9%  |
| Av. gene length (bp)           | 931    | 933    | 932    |
| rRNA (16S-23S-5S)              | 4      | 4      | 4      |
| tRNA                            | 56     | 56     | 56     |
| miscellaneous RNA              | 13     | 13     | 13     |
| Pseudogenes and partial genes  | 82     | 87     | 108    |
| IS elements                    | 27     | 28     | 32     |
| Genomic islands                | 4      | 4      | 3      |
| ICE regions                     | 0      | 1      | 2      |
| Plasmid                         | -      | -      | 1*     |

doi:10.1371/journal.pone.0006072.t001
thought to lead to the proteolytic activation of this matrix remodelling protein [72]. Matrix metalloproteinases can contribute to blood-brain barrier (BBB) disruption during bacterial meningitis [75,76]. MMP-9 produced by leukocytes during meningitis is essential to enable them to migrate through the BBB. A consequence of the increase in the permeability of the BBB mediated by MMP-9, is that bacterial cells are able to migrate to the central nervous system and cause further tissue damage.

IgA1 protease mediates the site-specific cleavage of human IgA1 in the hinge region [73]. IgA is the major class of immunoglobulin in mucosal secretions of the upper respiratory tract, and functions by interfering with microbial attachment to host tissues. The activity of IgA1 protease may help pathogens to subvert the antigen specificity of the humoral immune response, and facilitate adhesion and persistence at the mucosal surface [77].

Figure 2. Phylogenetic relationships of S. suis to the other genome sequenced streptococci. Unrooted majority-rule tree of Bayesian analysis of combined 16S rRNA and RNase P RNA gene datasets. The tree was built from 16S rRNA and rnpB gene sequences from the genomes of the Streptococcus species: S. suis P1/7, S. mutans UA159, S. gordonii Challis CH1, S. sanguinis SK36, S. pyogenes Manfredo, S. equi 4047, S. agalactiae NEM316, S. uberis 0140J, S. pneumoniae TIGR4, and S. thermophilus CNRZ1066. Lactococcus lactis subsp. lactis IL1403 was included as an outgroup. The numbers at the branches are posterior probabilities indicating the support for the branch. The bar indicates the number of substitutions per site (4 per 100 sites).

doi:10.1371/journal.pone.0006072.g002
terminal signal sequence, and the 3-prime reading frame contains a LPXTG sortase-processing motif. The EF+ phenotype of P1/7 is probably due to the expression of the 5-prime reading frame.

Unlike MRF and EF, opacity factor (OFS) is a large surface-associated protein that may play a role in the virulence of S. suis [82]. The N-terminus of OFS is homologous to the N-terminal regions of the fibronectin-binding protein A (FnBA) of Streptococcus dysgalactiae and the serum opacity factor of S. pyogenes. Mutational studies have suggested a role for OFS in virulence, but not colonization. However in strain P1/7 OFS (SSU1474) is mutated, containing two frameshift mutations.

Two of the LPXTG proteins were found in a cluster similar to the pilus islands identified in S. agalactiae [83], and comprise the main pilus subunit (SSU0427) and ancillary subunit of this pilus (SSU0425). In addition the cluster contains a signal peptidase (SSU0424) and a sortase (SSU0420). The genome contains a second pilus cluster, containing 3 sortases, srtD, srtC, srtB (SSU1881, SSU1882, SSU1883), ancillary protein 1 (SSU1885), the main pilus subunit (SSU1886), ancillary protein 2 (SSU1888) and ancillary protein 3 (SSU1889). In silico analysis of the P1/7 genome suggests that this strain does not express complete pili, as both of the clusters contain pseudogenes; SSU0425 contains a nonsense mutation, and SSU1886 and SSU1888 contain frameshift mutations. To this end it is worth noting that orthologous CDS in the other sequenced S. suis strains are also pseudogenes.

The genome also contains an incomplete pilus cluster that contains a signal peptidase (SSU0450) and sortase SrtE (SSU0453), but lacks pilus subunits and ancillary subunits. Between the signal peptidase and sortase there are a putative exported protein (SSU0451) and a transposase fragment (SSU0452). It is possible that the putative exported protein represents the N-terminal region of a pilus protein that has been truncated by a deletion event that generated the partial transposase sequence. Evidence for the potential role of pili in S. suis pathogenesis has come from in

<table>
<thead>
<tr>
<th>Table 2. Orthologues of S. suis P1/7 CDSs in other streptococci.</th>
</tr>
</thead>
<tbody>
<tr>
<td>% of CDSs with orthologous matches</td>
</tr>
<tr>
<td><strong>S. sanguinis</strong> SK36</td>
</tr>
<tr>
<td>S. gordonii CH1</td>
</tr>
<tr>
<td>S. agalactiae nem316</td>
</tr>
<tr>
<td>S. agalactiae 2603v</td>
</tr>
<tr>
<td>S. agalactiae A909</td>
</tr>
<tr>
<td>S. uberis 0140J</td>
</tr>
<tr>
<td>S. pneumoniae D39</td>
</tr>
<tr>
<td>S. pneumoniae TIGR4</td>
</tr>
<tr>
<td>S. pneumoniae R6</td>
</tr>
<tr>
<td>S. mutans UA159</td>
</tr>
<tr>
<td>S. equi 4047</td>
</tr>
<tr>
<td>S. pyogenes Manfredo</td>
</tr>
<tr>
<td>S. thermophilus CNRZ1066</td>
</tr>
<tr>
<td>S. thermophilus LMG 18311</td>
</tr>
</tbody>
</table>

Orthologues matches are displayed as a percentage of the total CDSs in the P1/7 genome.
doi:10.1371/journal.pone.0006072.t002

Figure 3. Functional distribution of Streptococcal orthologues in the S. suis P1/7 genome. Functional groups are displayed clockwise on the web graph, in decreasing range of orthologue matches for a single category.
doi:10.1371/journal.pone.0006072.g003
in vitro work looking for preferentially expressed genes in S. suis upon interaction with porcine brain microvascular endothelial cells [71]. The signal peptidase (SSU0424) of the S. agalactiae-like cluster is upregulated, as is srlE, the srtE belonging to the incomplete pilus cluster.

Comparison of cell wall-anchored proteins among streptococcal species shows that they are one of the most variable components of the genome; 13 out of the 29 proteins in P1/7 were unique to S. suis (Table S2), the rest being intermittently distributed. Interestingly, orthologues of several of the S. suis cell wall-anchored proteins, or proteins associated with their processing, were conserved within the meningitis-associated streptococci (Table 2). Six of the cell wall-anchored proteins, and four of the sortases were found in S. agalactiae and/or S. pneumoniae, and were absent in other streptococci [84].

In addition to LPXTG-type cell wall-anchored proteins, the P1/7 genome also contains other surface-exposed proteins that may play an important role in modulating host-cell interactions. The P1/7 genome encodes a protein, SSU0496, which contains a Mac 1 domain (PF09028), an N-terminal signal sequence and a C-terminal transmembrane domain. The Mac 1 domain is found in a small number of streptococcal proteins that modulate the immune response. Mac, also known as IdeS, is a cysteine protease secreted by S. pyogenes [85,86]. Two allelic families of Mac have been identified in S. pyogenes; Mac I and Mac II [87] exhibit sequence divergence in the middle third of the proteins (~50% amino acid identity in this region). Mac was originally named because it had limited sequence homology to the ζ-subunit of the leukocyte β2 integrin, Mac-1 [85]. S. pyogenes Mac proteins block polymorphonuclear leukocyte opsonophagocytosis and inhibit the production of reactive oxygen species [87]. The mechanisms of action of the S. pyogenes Mac allelic families differ; Mac I binds to, and cleaves, the hinge region of human IgG [88] whereas Mac-2 exhibits lower levels of endopeptidase activity, and does not bind IgG, but instead to the IgG receptor, Fcy [89]. Homologues of Mac have been identified in Streptococcus equi subsp. equi (349 amino acids), the causative agent of strangles, and the related pathogen Streptococcus zooepidemicus [90]. In S. equi the Mac homologue is expressed during infection, and reduces the antiphagocytic activity of equine neutrophils [91].

SSU0496 is similar (~29% sequence identity) in the N-terminal region to the S. pyogenes Mac I and II proteins. In comparison, the S. suis protein is considerably larger than the S. pyogenes Mac proteins, comprising 1141 amino acids as opposed to ~340 amino acids. The enzymatic activities of S. pyogenes Mac proteins have been characterized, and they belong to a novel family of cysteine proteases [88]. Protein alignment of the S. suis protein with the S. pyogenes proteins shows that the amino acids of the catalytic triad (Cys94, His262 and Asp284) are conserved. The S. suis protein however, is missing an Arg-Gly-Asp (RGD) integrin-binding motif present in the other streptococcal proteins, suggesting that while it may possess the peptidase activity of its homologues it may lack the ability to bind integrin.

Comparative genomics of streptococci that can cause meningitis

Several species of Streptococcus have been associated with meningitis in humans: S. pneumoniae is an important cause of bacterial meningitis [92], and S. agalactiae is important in this regard [93]. Comparative genomic analysis was used to identify components of the S. suis genome shared with S. pneumoniae and S. agalactiae that were absent in other streptococci. Three strains of S. agalactiae, NEM316 [94], 2603V/R [95] and A909 [96] and three strains of S. pneumoniae, R6 [97], TIGR4 [33] and D39 [98], have been sequenced thus far. Whilst S. pneumoniae strain TIGR4 is from a case of invasive disease, and S. agalactiae strains NEM316, 2603V/R and A909 are all clinical isolates, none of the strains appears to come from a meningitis case. 71 CDSs were identified that were present in at least one of the strains of S. pneumoniae or S. agalactiae, but were absent in other Streptococcus species (Table S3). Of these 12 were found in MGEs in the P1/7 genome. The distribution of orthologue matches is varied between species and between strains; only 5 CDSs (SSU0685, SSU0686, SSU0687, SSU1050 and SSU1403) have orthologues matches in all of the strains for both species. Amongst these common CDSs are 3 CDSs that form part of an operon and encode a phosphomethylpyrimidine kinase, a hydroxethylhiazole kinase and a thiamine-phosphate pyrophosphorylase. These 3 CDSs catalyse the final steps in the biosynthesis of thiamine monophosphate from pyrimidine and thiazole molecules. The other CDSs in this operon, SSU0684, encodes a putative phosphatase which does not have any orthologue matches. Another of the CDSs common to S. suis, S. agalactiae and S. pneumoniae encodes a PadR family regulatory protein.

The final common CDS encodes a hyaluronidase. Hyaluronidase has been described as a virulence factor of S. pneumoniae; the addition of hyaluronidase to an intra-nasal inoculum was necessary to cause invasive disease in an infant mouse model of meningitis [99]. Hyaluronidase mutants do not appear to be attenuated in a murine model of meningitis after intracerebral infection [100], suggesting that this degradative enzyme may be important for the bacteria to cross the mucosa or the BBB during the early stages of pathogenesis. Notably in all five S. suis strains sequenced, including the human meningitis strains this CDS is present as a pseudogene. A study investigating the distribution of hyaluronidases in a diverse collection of S. suis isolates identified fewer than 30% exhibited hyaluronate lyase activity in vitro, although all the isolates tested (n. 309) contained the hyaluronidase gene (hly) [101]. In many cases mutations were identified that disrupted the expression of hly. Notably hyaluronate lyase activity was absent in the majority of isolates associated with invasive disease (meningitis and septicemia), whereas activity was detected more often in isolates associated with pneumonia.

Given the unknown potential to cause meningitis of the S. agalactiae and S. pneumoniae strains sequenced, variably distributed matches should not be ruled out as having important roles in causing meningitis. Some of these CDSs encode functions that are important for pathogenesis, such as the zinc metalloproteinase previously mentioned. Orthologues of this protein are present in all of the S. pneumoniae strains compared but absent in S. agalactiae (Table 3). Components of the two pilus clusters are amongst the differentially distributed matches.

Comparative genomics of two geographically distinct human disease strains SC84 and BM407

S. suis strain SC84 is a human isolate, and was isolated from the same outbreak as Chinese S. suis strain 05ZYH33 [8]. Strains from this outbreak have shown to belong to MLST sequence type 7 [7]. Strain P1/7 is MLST sequence type 1 and, as a single locus variant, belongs to the same clonal complex as SC84. Previous comparisons of strain P1/7 with S. suis strains 98HAIH33 and 05ZYH33 showed that only 83.6% and 83.3% respectively of the P1/7 CDSs had orthologous matches. It was also reported that 4.9% of the P1/7 genome was unique in comparison to S. suis 05ZYH33 and 98HAIH33 [7]. However, according to our analysis, all of the CDSs in the strain P1/7 genome have equivalent coding regions in the S. suis 05ZYH33 and 98HAIH33 sequences. The difference in the number of orthologues identified can be
Table 3. CDSs in the genome of *S. suis* strain P1/7 that have orthologues in *S. pneumoniae* and/or *S. agalactiae*, but not other streptococci.

<table>
<thead>
<tr>
<th>ID</th>
<th>Product</th>
<th><em>S. agalactiae</em></th>
<th><em>S. pneumoniae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>2603V/R</td>
<td>A909</td>
</tr>
<tr>
<td>SSU0003</td>
<td>diacylglycerol kinase protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0103*</td>
<td>replication initiation factor</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0104*</td>
<td>conserved hypothetical protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0106*</td>
<td>FtsK/SpoIIIE family protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0110</td>
<td>conserved hypothetical protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0173</td>
<td>putative membrane protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0174</td>
<td>ABC transporter ATP-binding protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0200</td>
<td>conserved hypothetical protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0226</td>
<td>putative transcriptional regulator</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0228</td>
<td>putative membrane protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0415</td>
<td>Fic protein family</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0417</td>
<td>conserved hypothetical protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0418</td>
<td>putative DNA-binding protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0423</td>
<td>hypothetical protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0424</td>
<td>putative signal peptidase I 2</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0428</td>
<td>sortase</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0440</td>
<td>acetyltransferase (GNAT) family protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0452*</td>
<td>putative transposase (fragment)</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0520</td>
<td>putative rhamnosyl transferase</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0524</td>
<td>oligosaccharide repeat unit polymerase</td>
<td>+</td>
<td>−</td>
</tr>
<tr>
<td>SSU0533</td>
<td>putative Lipoooligosaccharide sialytransferase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0535</td>
<td>putative N-acetyleneuraminic acid synthase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0537</td>
<td>putative transferase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0538</td>
<td>N-acetylneuraminic cytidylytransferase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0541*</td>
<td>putative transposase</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0544*</td>
<td>putative transposase</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0549*</td>
<td>putative transposase (fragment)</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0566*</td>
<td>putative transposase</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0601</td>
<td>conserved hypothetical protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0685</td>
<td>phosphomethylpyrimidine kinase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0686</td>
<td>hydroxyethylthiazole kinase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0687</td>
<td>thiamine-phosphate pyrophosphorylase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0706</td>
<td>muramidase-released protein precursor</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0722</td>
<td>putative L-threonine aldolase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU0829</td>
<td>plasmid addiction system, toxin protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0830</td>
<td>conserved hypothetical protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0879</td>
<td>zinc metalloproteinase</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU0885</td>
<td>acetyltransferase (GNAT) family protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU1036*</td>
<td>putative transposase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1050</td>
<td>hyaluronidase precursor (pseudogene)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1054</td>
<td>putative membrane protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU1067</td>
<td>conserved hypothetical protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1074</td>
<td>putative membrane protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU1163</td>
<td>conserved hypothetical protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU1201</td>
<td>putative surface-anchored protein</td>
<td>−</td>
<td>+</td>
</tr>
<tr>
<td>SSU1242</td>
<td>VanZ like family protein</td>
<td>−</td>
<td>+</td>
</tr>
</tbody>
</table>
accounted for by differences in gene prediction. The only large scale region of difference is associated with the 89 kb putative PI that is present in the two sequenced ST7 strains [7], and absent in P1/7.

The genome of *S. suis* strain SC84 consists of a single circular chromosome of 2,095,898 bp (accession number FM252031) and containing 1,985 CDSs (Table 1), of which 87 are pseudogenes or gene fragments (Table S1). Strain BM407 was isolated from a case of human meningitis in Vietnam in 2004 and, like strain P1/7, has MLST sequence type 1. The genome of *S. suis* strain SC84 consists of a single circular chromosome of 2,146,229 bp (accession number FM252032) containing 1,952 CDSs (Table 1), of which 108 pseudogenes or gene fragments, and a plasmid (pBM407) (accession number FM252033) which contains 2,040 CDSs (Table 1).

**Genome structure**

The chromosomes of strains P1/7 and SC84 have a conserved structure and are collinear along their length (Figure 4). The BM407 chromosome is collinear, with the exception of a large inversion (nucleotide 600371 to 1416050). The rearrangement is an ongoing process (data not shown).

Comparison of the three *S. suis* genomes reveals discrete regions of difference that distinguish these closely related strains. For strains P1/7 and SC84, the regions of difference are exclusively associated with MGEs; all of the genes identified in strain P1/7 outside MGEs have homologs in strain SC84. Strain BM407 also contains variation associated with MGEs, but additionally there are two regions that appear to be deleted, resulting in a reduction in the number of genes on the chromosome of this strain in comparison to the other two. The first is a small region of 268 bp, flanked by 30 bp direct repeats, that encodes a putative membrane protein (in P1/7, SSU1245). The second absent region is 3.3 kb and contains 2 complete CDSs encoding subunits of an ATP-dependent exonuclease RexAB and the

**Table 3. cont.**

<table>
<thead>
<tr>
<th>ID</th>
<th>Product</th>
<th><em>S. agalactiae</em></th>
<th><em>S. pneumoniae</em></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>2603V/R</td>
<td>A909</td>
<td>NEM316</td>
</tr>
<tr>
<td>SSU1272</td>
<td>type I restriction-modification system S protein</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1273</td>
<td>type I restriction-modification system M protein</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1274</td>
<td>type I restriction-modification system R protein</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1383</td>
<td>putative acyl carrier protein phosphodiesterase</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1402</td>
<td>putative membrane protein</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>SSU1403</td>
<td>PadR-like family regulator protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1405*</td>
<td>putative transposase</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1422*</td>
<td>putative transposase</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1546</td>
<td>putative membrane protein</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1631*</td>
<td>putative transposase</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1718</td>
<td>putative membrane protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1773</td>
<td>putative surface-anchored serine protease</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1795</td>
<td>toxin-antitoxin system, toxin protein</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>SSU1796</td>
<td>toxin-antitoxin system, antitoxin protein</td>
<td>+</td>
<td>–</td>
</tr>
<tr>
<td>SSU1852</td>
<td>putative amino-acid ABC transporter permease protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1866</td>
<td>metal cation ABC transporter membrane protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1872</td>
<td>CAAX amino terminal protease family protein</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1881</td>
<td>sortase SrtD</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1882</td>
<td>sortase SrtC</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1883</td>
<td>sortase SrtB</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1885</td>
<td>putative surface-anchored protein</td>
<td>–</td>
<td>–</td>
</tr>
<tr>
<td>SSU1886</td>
<td>putative surface-anchored protein (pseudogene)</td>
<td>+</td>
<td>+</td>
</tr>
<tr>
<td>SSU1901</td>
<td>putative nucleotidase</td>
<td>–</td>
<td>+</td>
</tr>
<tr>
<td>SSU1934</td>
<td>putative exported protein</td>
<td>+</td>
<td>+</td>
</tr>
</tbody>
</table>

Orthologues identified using reciprocal Fasta, against the *S. pneumoniae* strains R6, TIGR4 and D39, and *S. agalactiae* strains NEM316, 2603V/R and A909. Orthologues with matches to *S. mutans* UA159, *S. gordonii* CHallis CH1, *S. sanguinis* SK36, *S. pyogenes* Manfredo, *S. equi* 4047, *S. uberis* O140J, or *S. thermophilus* CNRZ1066 were not included in the list. * contained on a putative MGE.

doi:10.1371/journal.pone.0006072.t003
C-terminal region of Mrp (SSUBM407_1126; equivalent to bases 733785 to 743796 in strain P1/7). Immediately downstream of mrp is a partial IS\textsubscript{1216} transposase. The 3-prime region of mrp (pSSUBM407_006) is found on the plasmid pBM407, along with rex\textsubscript{AB} (pSSUBM407_015 and pSSUBM407_016), the genes for the missing chromosomal ATP-dependent exonuclease. The plasmid also contains three copies of IS\textsubscript{1216}, and it is likely that the chromosomal mrp and rex\textsubscript{AB} locus was deleted by IS\textsubscript{1216}-mediated recombination.

**Small-scale variation.** The three strains also exhibited small-scale variation at the nucleotide level including SNPs, insertions and deletions. Of the three strains BM407 appears to be the most diverse: comparison of strain BM407 to P1/7 identified 977 SNPs, 201 insertions, and 122 deletions in the core chromosome (excluding ICE regions); and comparison to SC84 identified 1017 SNPs, 232 insertions, and 133 deletions, whereas a comparison of strain P1/7 to SC84 identified 85 SNPs, 10 insertions, and 15 deletions.

The relative levels of variation are, on the face of it, at odds with the relatedness of the strains as defined by MLST: P1/7 belongs to the same sequence type as BM407, whereas SC84 is a single locus variant of P1/7. MLST is based on variation at only seven loci in the genome however, so it should not be surprising that it does not necessarily reflect the overall diversity at the whole genome level. Analysis of the distribution and density of SNPs in the genomes suggest that recombination has generated allelic variation at certain loci in the BM407 chromosome, thereby increasing the apparent level of diversity. Eight loci in the BM407 chromosome contained a high density of SNPs and were predicted by Clonalframe to be regions that have undergone recombination (Table 4). In total the SNPs in these regions account for \( \sim 60\% \) of the SNPs in the comparison with P1/7 and SC84.

Several of the putatively recombined loci in BM407 encode functions or proteins that potentially have a role in virulence or survival in the host. Two of the loci encode components of transport systems, that include an ABC-type glycine betaine
transport system protein (SSUBM407_1490) homologous (56.1% identity at the amino acid level) to a bile exclusion system protein BilEB from *Listeria monocytogenes* shown to have a role in virulence [103]; and a binding protein-dependent transport system (SSUBM407_1447 to SSUBM407_1449) orthologous to a system from *S. mutans* responsible for the uptake of melibiose, raffinose, and isomaltotriose and the metabolism of melibiose, sucrose, and isomaltosaccharides [104]. One of the loci is composed of three CDSs that are part of the putative exopolysaccharide biosynthesis cluster previously described. Two of the loci contain cell wall anchored surface proteins, encoding a zinc carboxypeptidase (SSUBM407_0646) and a N-acetylmuramoyl-L-alanine amidase (SSUBM407_0662). One of the loci is composed of three CDSs that are part of the putative exopolysaccharide biosynthesis cluster previously described. Two of the loci contain cell wall anchored surface proteins, encoding a zinc carboxypeptidase (SSUBM407_0646) and a N-acetylmuramoyl-L-alanine amidase (SSUBM407_0662). A single regulatory CDS (SSUBM407_0840) is found in the loci that appear to have undergone recombination. This CDS encodes a sensor kinase which is part of a two component regulatory system operon, with the SNPs being found in the N-terminal region of the sensor kinase protein.  

### Table 4. Putative regions of recombination identified in the BM407 genome.

<table>
<thead>
<tr>
<th>CDS</th>
<th>Product</th>
<th>Coordinates</th>
<th>Number of SNPs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Region 1</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSUBM407_0131</td>
<td>acetate kinase</td>
<td>124591..125778</td>
<td>62</td>
</tr>
<tr>
<td>SSUBM407_0132</td>
<td>membrane protein</td>
<td>126092..126643</td>
<td>48</td>
</tr>
<tr>
<td>SSUBM407_0133</td>
<td>folypolyglutamate synthase</td>
<td>126699..127955</td>
<td>48</td>
</tr>
<tr>
<td>SSUBM407_0134</td>
<td>putative glutamyl-aminopeptidase</td>
<td>129061..128000</td>
<td>19</td>
</tr>
<tr>
<td>Region 2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSUBM407_0646</td>
<td>putative surface-anchored zinc carboxypeptidase</td>
<td>715332..712147</td>
<td>33</td>
</tr>
<tr>
<td>SSUBM407_0647</td>
<td>putative cardiolipin synthetase</td>
<td>715548..717095</td>
<td>16</td>
</tr>
<tr>
<td>Region 3</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSUBM407_0662</td>
<td>N-acetylmuramoyl-L-alanine amidase</td>
<td>730766..733573</td>
<td>41</td>
</tr>
<tr>
<td>Region 4</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSUBM407_0668</td>
<td>polysaccharide export system ATP-binding protein</td>
<td>739785..740999</td>
<td>18</td>
</tr>
<tr>
<td>SSUBM407_0669</td>
<td>rhamman synthesis protein F family protein</td>
<td>741020..742744</td>
<td>40</td>
</tr>
<tr>
<td>SSUBM407_0670</td>
<td>exopolysaccharide biosynthesis protein</td>
<td>742777..744036</td>
<td>16</td>
</tr>
<tr>
<td>Region 5</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSUBM407_0840</td>
<td>sensor histidine kinase</td>
<td>742777..744036</td>
<td>37</td>
</tr>
<tr>
<td>Region 6</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSUBM407_1445</td>
<td>glucan 1,6-alpha-glucosidase</td>
<td>153333..1531717</td>
<td>16</td>
</tr>
<tr>
<td>SSUBM407_1446</td>
<td>sucrose phosphorylase</td>
<td>1534855..1533407</td>
<td>32</td>
</tr>
<tr>
<td>SSUBM407_1447</td>
<td>multiple sugar-binding transport system permease protein</td>
<td>1535755..1534925</td>
<td>33</td>
</tr>
<tr>
<td>SSUBM407_1448</td>
<td>multiple sugar-binding transport system permease protein</td>
<td>1536641..1535766</td>
<td>23</td>
</tr>
<tr>
<td>SSUBM407_1449</td>
<td>multiple sugar-binding transport system permease protein</td>
<td>1538008..1536773</td>
<td>16</td>
</tr>
<tr>
<td>Region 7</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSUBM407_1489</td>
<td>conserved hypothetical protein</td>
<td>1572295..1571534</td>
<td>31</td>
</tr>
<tr>
<td>SSUBM407_1490</td>
<td>ABC-type glycine betaine transport system protein</td>
<td>1573816..1572305</td>
<td>38</td>
</tr>
<tr>
<td>Region 8</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SSUBM407_1989</td>
<td>acetyltransferase (GNAT) family protein</td>
<td>2097869..2097417</td>
<td>39</td>
</tr>
<tr>
<td>SSUBM407_1990</td>
<td>aldo/keto reductase family protein</td>
<td>2098790..2097879</td>
<td>14</td>
</tr>
</tbody>
</table>

SNP were identified in a comparison with the other ST 1 strain, P1/7. doi:10.1371/journal.pone.0006072.t004

There are fewer SNPs in the comparisons between P1/7 and SC84 than were described in the publication by Chen et al. (2007) which compared 98HAH33, 05ZYH33 and P1/7. Interestingly, a comparison of SC84 with 05ZYH33, both strains from the same outbreak, identified 2602 SNPs, 596 insertions, and 304 deletions. This is in marked contrast to the low number of SNPs identified in the comparison between P1/7 and SC84, two strains isolated from different hosts, on different continents over 20 years apart.

Strains from the Sichuan outbreak have been found to be ST7 [8]. Examination of the sequence at MLST loci of 05ZYH33 revealed that they do not match the expected alleles for those of ST7. In comparison to SC84, which itself belongs to ST7, SNPs were identified at two of the seven MLST loci in the 05ZYH33 sequence: the *aroA* locus has 29 SNPs, and *gki* has 2 SNPs. Although, it is possible that 05ZYH33 is actually a different sequence type, albeit a double locus variant of ST7. The predicted sequence type of 98HAH33 is more divergent still, being a triple locus variant of ST7; the *aroA* has 6 SNPs, *cpn* has 1 single base deletion, and *dpr* has 1 single base deletion. The most likely explanation for these observations is that the sequence data for the 05ZYH33 and 05HAH33 strains is incorrect.

---

**Streptococcus suis Genomics**

**[PLoS ONE](http://www.plosone.org)**

July 2009 | Volume 4 | Issue 7 | e6072
**Mobile genetic elements.** Four regions of the P1/7 genome were identified that had properties indicative of recent acquisition, which include 1 prophage-like region and 3 miscellaneous islands, which have been designated genomic islands. None of these genomic islands carried characterized or putative virulence factors. Equivalent genomic islands are present in the SC84 genome. BM407 contains three of the genomic islands present in P1/7, but is lacking genomic island 3 (in strain P1/7 nucleotides 835135 to 846313; Figure 4). In contrast to some streptococcal species such as *S. pyogenes* [106], the proportion of the genome comprised by MGEs is relatively small, at 1.8% of the P1/7 genome, suggesting that the pan-genome of this species is smaller than other streptococci [96,107], and that there may be some limitations or restrictions to the amount of recent horizontal gene transfer in this species.

Previous sequencing of the two *S. suis* strains associated with meningitis outbreaks in China in 1998 and 2005 identified highly similar 89 kb regions present in both genomes which were designated a candidate PI [7]. This region is absent in strain P1/7, but strain SC84 contains an almost identical region. This island has a composite structure, contains several regions that appear to be integrated MGEs and has a structure similar to integrative conjugative elements (ICE) found in other streptococci [108]. Accordingly this region has been designated ICE*SuSC84*.

**Integrative conjugative elements.** ICE are commonly found in streptococcal genomes. The most common are Tn916-type elements, which almost invariably carry the tetracycline resistance determinant tetM, often accompanied by other drug resistance genes, and Tn5252-type elements, which are larger and more variable in sequence. In many cases, Tn5252 transposons are carrying Tn916 elements; varying arrangements of the two are observed in different genomes, suggesting this association has arisen independently on a number of occasions. Analogous with the bacterial chromosome as a whole, the ICE have core and accessory components to their gene complement. The conjugative machinery appears to be relatively well conserved between the ICE, with all of the streptococcal Tn5252-like elements sharing a large recombination protein, as well as VirB4-type and VirD4-type proteins. Furthermore, a genus-wide comparison of streptococcal ICE reveals a surprising level of similarity in their complement of cargo genes. The *pezAT* addiction toxin system, first identified in the chromosome of *S. pneumoniae*, is found on elements present in *S. suis*, *S. pneumoniae* and *S. agalactiae*, presumably aiding fixation of the transposon following integration. Bacteriocins, either alone or as part of a gene cluster encoding the accompanying processing machinery, are present on all the sequenced *S. suis* and *S. pneumoniae* [109] elements. Many of the other cargo coding sequences appear to be involved in increasing the stress tolerance of the host: abi genes are common, and the ICE of *S. dysgalactiae* subsp. *equinum* and *S. agalactiae* 2603V/R encode multiple heavy metal resistance genes. Of greater clinical importance, sequenced streptococcal ICE can carry genes for resistance to tetracycline, erythromycin, chloramphenicol, trimethoprim, aminoglycosides and streptomycin.

ICE*SuSC84* is integrated into the SC84 *S. suis* genome in the 3-prime region of the gene encoding the 50S ribosomal protein L7/L12 (SSUSC84_0891). ICE*SuSC84* has extended identity to a Tn2424 region (Figure 5) found in *S. agalactiae* strain NEM316 [94], with three regions of difference. The first of these contains a putative bacteriocin biosynthesis cluster that is disrupted by a putative integron carrying five CDSs including an aminoglycoside resistance gene (aminoglycoside 6-adenylyltansferase; SSUSC84_0863). The second contains MGE-associated CDSs, including a Tn916 insertion that carries a tetracycline resistance gene (tetM; SSUSC84_0827). The third region contains genes associated with putative lantibiotic export/resistance and an asparagine synthetase. The only component of the 89kb island that is an obvious candidate

---

**Figure 5. Comparisons of the ICE regions from *S. suis* strains SC84 and *S. agalactiae* NEM315.** The results of a TBLASTX comparison of the ICE*SuSC84* from strain SC84 and *S. agalactiae* NEM315 Tn2424 is displayed using the Artemis Comparison Tool (ACT) [26].
doi:10.1371/journal.pone.0006072.g005

---

*Streptococcus suis* Genomics
for a pathogenicity determinant is a surface-anchored protein that contains a LPXTG motif that is annotated as an agglutinin receptor (SSUSC84_0873). This protein is found in the \textit{S. agalactiae} element (and other streptococci) and contains a glucan-binding domain.

The BM407 genome contains two regions with extended similarity to ICE \textit{Ssu} SC84. The first of these, ICE \textit{Ssu} BM4071, is inserted into a luciferase-like monooxygenase (SSUBM407_0455), resulting in the disruption of this CDS. The second, ICE \textit{Ssu} BM4072, is found at the same locus as ICE \textit{Ssu} SC84; the element is inserted into the 3-prime region of the gene encoding the 50S ribosomal protein L7/L12 (SSUBM407_0933). Although ICE \textit{Ssu} BM4072 and ICE \textit{Ssu} SC84 are inserted at the same attachment site it is likely that they have arisen in these two strains by independent insertion events. Not only do the two elements contain regions of difference, the conserved DNA in the ICEs is more diverse than the core chromosome DNA (the ICEs exhibit \(95\% - 98\%\) identity, in contrast to \(99\% - 100\%\) identity for chromosomal DNA). ICE \textit{Ssu} BM4072 and ICE \textit{Ssu} SC84 are inserted at the same attachment site. Comparison of the BM407 and SC84 ICEs shows that CDSs associated with conjugation are conserved (Figure 6). The ICEs contain regions of sequence divergence with ICE \textit{Ssu} BM4071 appearing to be most diverse, containing a greater proportion of unique sequence, and an inverted region. All three elements contain variants of Tn916; ICE \textit{Ssu} BM4071 and ICE \textit{Ssu} SC84 have Tn916 with the same gene content, whilst ICE \textit{Ssu} BM4072 contains additional genes encoding the tetracycline resistance protein TetL (SSUBM407_0984) and chloramphenicol acetyltransferase (SSUBM407_0980). In addition to the drug resistance genes associated with Tn916, ICE \textit{Ssu} BM4072 also carries genes encoding the tetracycline resistance protein TetO (SSUBM407_0954), erythromycin ribosome methylase ErmB (SSUBM407_0952) and dihydrofolate reductase (SSUBM407_0957), in a region of difference located elsewhere on the element.

Unlike ICE \textit{Ssu} BM4072, ICE \textit{Ssu} BM4071 does not carry as many CDSs associated with drug resistance. Amongst its cargo are a lactose utilization operon (SSUBM407_0510 to SSUBM407_0519), including a gluconokinase-like protein, a galactose-6-phosphate isomerase, a tagatose-6-phosphate kinase, a tagatose-1,6-diphosphate aldolase, a 6-phospho-beta-galactosidase, and a PTS system lactose-specific transporter, which are absent in ICE \textit{Ssu} BM4072 and ICE \textit{Ssu} SC84. The chromosome of BM407 (and P1/7 and SC84) already contains homologues of these CDSs at two loci (SSUBM407_0879 to SSUBM407_0889 and SSUBM407_1060). Interestingly the chromosomal copy of the tagatose-6-phosphate kinase \textit{lacC} (SSUBM407_0883) is mutated in BM407, therefore the provision of an additional copy of this CDS on ICE \textit{Ssu} SC84 complements the mutation. Complete lactose utilization operons have been found on other ICEs, including the \textit{S. agalactiae} NEM315 ICE.
Contained on ICE$_\text{SuSC84}$ is a two component regulator, salK/salR (SSUSC04_0049 and SSUSC04_0050), that has been implicated in the virulence of the Chinese isolates in an animal model [110]. These regulatory components are part of a larger operon encoding lantibiotic biosynthesis, similar to the salicinacin cluster of Streptococcus salivarius [111]. Neither of the ICEs in BM407 contains this cluster (Figure 6); these elements contain alternative cargo CDSs in equivalent regions on these elements. Upstream of the operon is a CDS encoding a lantibiotic precursor (SSUSC04_0066). In SC34, and also in 05ZYH33 and 08HAH33, the operon has been disrupted by the insertion of a 7.9 kb element containing 9 CDSs, one of which encodes an aminoglycoside 6-adenylyltransferase (SSUSC34_00963), into the homologue of the S. salivarius salicacin C biosynthesis gene salM (SSUSC04_0856). The precise regulatory function of SalK/SalR remains unclear, but their operonic association with the lantibiotic cluster suggests that their primary role is associated with this function. However it is possible that regulatory components of this operon may have been adapted to other functions. $S$. pyogenes with a mutation in salT, a gene encoding a putative ABC transporter permease of a lantibiotic operon, was attenuated for virulence in a zebrafish invasive-disease model [112]. Analysis of the $S$. pyogenes lantibiotic cluster showed that it contained a homologue of salM that was mutated, suggesting that like the lantibiotic cluster in S. suis SC34, this cluster also no longer produced a functional lantibiotic, but it was still was essential for full virulence.

Screening of 91 Vietnamese $S$. suis serotype 2 strains of MLST ST1 isolated from patients with meningitis, for the presence of ICE$_\text{SuSC84}$ using PCR targeted at the insertion sites [7], indicated that this ICE is not absolutely required for virulence in humans as 16 (17.6%) strains lacked the element (manuscript in preparation). In addition, only 41 (45%) carried sequences homologous to the salK/salR genes, suggesting that the presence of these genes cannot be used as a marker of virulence. Similar to strains BM407 and SC34, all strains containing ICE-associated sequences were resistant to tetracycline whereas those that were negative in the screening were sensitive, indicating the common presence of ICE carrying tet genes in these isolates.

**Discussion**

The sequencing and detailed annotation of the $S$. suis P1/7, SC34 and BM407 genomes provides valuable data for studying the evolutionary events that are shaping the virulence and drug resistance in this emerging zoonotic pathogen.

There is strong evidence that genetic background influences the ability of $S$. suis to cause zoonotic infections as human disease isolates are almost always serotype 2 and are very closely related by MLST. Of the 121 STs in the $S$. suis MLST data base (http://ssuis.mlst.net/), only five STs have representative isolated from humans, and four (ST1, ST6, ST7 and ST34) out of the five are all part of the same clonal complex. Two of the strains in this study (P1/7 and BM407) introduce a greater degree of diversity into the comparisons, insofar as they are from different geographical areas, and belong to different STs.

A major feature of the previously sequenced $S$. suis strains’ genomes hypothesised to be important for pathogenicity, is an 89 kb ICE region designated as a candidate PI [7]. BM407 also contains an almost identical ICE with some notable differences that encompass drug resistance genes, and a two-component regulatory system associated with a bacteriocin cluster. Sequence variation in the BM407 ICE suggest that although it is present at an orthologous site, it has been independently acquired.

The comparative analysis suggests that ICEs have been the major contributor to the evolution of drug resistance in the strains of $S$. suis examined. Although it is not possible to pinpoint the origins of the $S$. suis ICEs, due to the limited number of these elements in the databases, it is apparent from the analyses that they are closely related to ICEs found in phylogenetically distant streptococci. It is worth noting that some of these species occupy the same niches as $S$. suis, suggesting that inter-species genetic exchange is taking place. This has implications for defining the pan-genomes of some streptococcal species, since it may not be possible to define their boundaries if the pool of MGEs they exchange is effectively communal.

Aside from the ICEs, the virulence determinant inventories of P1/7, SC34 and BM407 strains are identical. Small nucleotide differences resulting from point mutation and recombination were observed in the genome comparisons. This variation may have subtle effects on the expression and function of these genes, and as a corollary, generate differences in the virulence phenotypes. To this extent it is possible that some of the small nucleotide differences observed in the SC34 and BM407 genomes in comparison to P1/7 may influence the ability of these strains to cause disease. However, as the capacity of P1/7 to cause human disease is unknown, it is not possible to speculate on the impact of genomic differences identified in this study on host-association. Future studies intending to increase the depth of sequencing in this zoonotic lineage will help unravel the pathogenic significance of sequence variation.

**Supporting Information**

**Table S1** Pseudogenes in the genome of $S$. suis strains P1/7, SC34 and BM407.

Found at: doi:10.1371/journal.pone.0006072.s001 (0.15 MB DOC)

**Table S2** Streptococcus suis orphan CDSs in the genome of $S$. suis strain P1/7.

Found at: doi:10.1371/journal.pone.0006072.s002 (0.24 MB DOC)

**Figure S1** Rhamnose-based polysaccharide cluster of strain P1/7

A) Structure and functional organisation of the Rhamnose-based polysaccharide cluster of strain P1/7. B) Comparison of the Rhamnose-based polysaccharide cluster from $S$. suis P1/7 with the RGP mrl and rpg gene clusters of Streptococcus mutans UA159. The comparison of the RGP clusters from the $S$. suis P1/7 (top) and S. mutans UA159 and (bottom) is displayed using the Artemis Comparison Tool (ACT) [26]. The red bars separating each genome represent similarity matches identified by TBLASTX analysis [14].

Found at: doi:10.1371/journal.pone.0006072.s003 (0.35 MB PDF)

**Acknowledgments**

We would like to thank the Sanger Institute's Pathogen Production Group for shotgun and finishing sequencing, and the core Informatics Group for support.

**Author Contributions**

Conceived and designed the experiments: MTGH MK JAL PNW CGD AMW NC MG JDS HS BG5 JX CY SDB BB CS DJM JP. Performed the experiments: MTGH HH MS THN IC AC IG KM MAQ CP ER SS TBC. Analyzed the data: MTGH HH MS NJC PJ CS. Contributed reagents/materials/analysis tools: THN TBC NTHM TSD NTC JX CY. Wrote the paper: MTGH HH MS NJC JAL MG HS BG5 SDB CS DJM JP.
References


edema, and meningeal inflammation are linked to the production of nitric oxide. Ann Neurol 37: 313–323.


