Infection, Genetics and Evolution 11 (2011) 1263–1275

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*

Vincent P. Richards [a], Ping Lang [a,1], Paulina D. Pavinski Bitar [a], Tristan Lefébure [a,2], Ynte H. Schukken [b], Ruth N. Zadoks [b,3], Michael J. Stanhope [a,*]

[a] Department of Population Medicine and Diagnostic Sciences, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA
[b] Quality Milk Production Services, College of Veterinary Medicine, Cornell University, Ithaca, NY 14853, USA

ARTICLE INFO

Article history:
Received 15 February 2011
Received in revised form 12 April 2011
Accepted 16 April 2011
Available online 22 April 2011

Keywords:
Comparative
Genomics
Bovine
Adapted
Streptococcus agalactiae

ABSTRACT

In addition to causing severe invasive infections in humans, *Streptococcus agalactiae*, or group B *Streptococcus* (GBS), is also a major cause of bovine mastitis. Here we provide the first genome sequence for *S. agalactiae* isolated from a cow diagnosed with clinical mastitis (strain FSL S3-026). Comparison to eight *S. agalactiae* genomes obtained from human disease isolates revealed 183 genes specific to the bovine strain. Subsequent polymerase chain reaction (PCR) screening for the presence/absence of a subset of these loci in additional bovine and human strains revealed strong differentiation between the two groups (Fisher exact test: $p < 0.0001$). The majority of the bovine strain-specific genes (~85%) clustered tightly into eight genomic islands, suggesting these genes were acquired through lateral gene transfer (LGT). This bovine GBS also contained an unusually high proportion of insertion sequences (4.3% of the total genome), suggesting frequent genomic rearrangement. Comparison to other mastitis-causing species of bacteria provided strong evidence for two cases of interspecies LGT within the shared bovine environment: bovine *S. agalactiae* with *Streptococcus uberis* (nisin U operon) and *Streptococcus dysgalactiae* subsp. *dysgalactiae* (lactose operon). We also found evidence for LGT, involving the salivaricin operon, between the bovine *S. agalactiae* strain and either *Streptococcus pyogenes* or *Streptococcus salivarius*. Our findings provide insight into mechanisms facilitating environmental adaptation and acquisition of potential virulence factors, while highlighting both the key role LGT has played in the recent evolution of the bovine *S. agalactiae* strain, and the importance of LGT among pathogens within a shared environment.

1. Introduction

Streptococcus agalactiae, also referred to as group B *Streptococcus* (GBS), is an important gram-positive pathogen, which causes severe invasive infections in adults and neonates, as well as asymptomatic infections in women (Balter et al., 2000; Dermer et al., 2004; Schrag et al., 2000), while also representing one of many human commensal organisms colonizing the gastrointestinal and genitourinary tract. In neonates, *S. agalactiae* has been recognized as one of the major causes of pneumonia, meningitis, and septicemia in industrialized nations (Baker, 2000). *S. agalactiae* is also recognized as an important pathogen in immunocompromised individuals and the elderly (Farley, 2001; Henning et al., 2001). It can also survive for extended periods in the human mammary gland causing clinical and subclinical mastitis (Bradley, 2002; Hansen et al., 2004; Schuchat, 1998). In cattle, *S. agalactiae* is a major cause of bovine mastitis, which is the dominant health disorder affecting milk production in the United States dairy industry, with reported annual losses due to mastitis exceeding two billion dollars (Katholm and Rattenborg, 2010; Keefe, 1997; National Mastitis Council, 1996; Wilson et al., 1999). In addition to humans and cattle, *S. agalactiae* has also been isolated from a wide range of other animals including rodents, cats, dogs, elephants, dolphins, amphibians, reptiles, and fish (Bishop et al., 2007).

Initial studies of strains of *S. agalactiae* isolated from human and bovine sources revealed a high degree of distinctiveness. For example, bovine strains frequently could not be serotyped, often possessed antigens absent in human isolates, and showed distinct biochemical characteristics (Finch and Martin, 1984; Pattison et al., 1955; Wanger and Dunny, 1987; Wibawan and Lammler, 1990). Subsequent molecular studies utilizing ribotyping, poly-

* Corresponding author. Tel.: +1 607 253 3859; fax: +1 607 253 3083.
E-mail address: mjs297@cornell.edu (M.J. Stanhope).
[1] Present address: Department of Plant Pathology & Plant-Microbe Biology, Cornell University, Ithaca, NY 14853, USA.
[2] Present address: Université de Lyon, F-69622, Lyon, France; Université Lyon 1, Villeurbanne, CNRS, UMR5023, Laboratoire d'Ecologie des Hydrosystèmes Naturels et Anthropisés, France.
[3] Present address: Moredun Research Institute, Pentlands Science Park, Bush Loan, Penicuik and Royal (Dick) School of Veterinary Studies, University of Edinburgh, Scotland, UK.

1567-1348/$ – see front matter © 2011 Elsevier B.V. All rights reserved.
doi:10.1016/j.meegid.2011.04.019

merase chain reaction (PCR), DNA sequencing, and multilocus sequence typing (MLST), have, by examining the host distribution and evolutionary relationships among these various molecular markers, confirmed the distinctiveness of the human and bovine populations (Bisharat et al., 2004; Bohnsack et al., 2004; Dogan et al., 2005; Sørensen et al., 2010; Sukhnanand et al., 2005). Specifically, the strain sequenced here belongs to clonal complex 67, which has been identified as a putatively cow-adapted subgroup of *S. agalactiae* (Bisharat et al., 2004). However, studies by Martinez et al. (2000) and Brochet et al. (2006), using RAPD and micro-array/MLST approaches respectively, showed less distinction between human and bovine populations.

Comparative genomics can provide insight into the evolutionary processes that influence bacterial populations and species. It can, for example, identify components of the genome that might play a role in virulence and/or niche adaptation and nutrient utilization. To date, three complete and five draft *S. agalactiae* genome sequences, all derived from human disease isolates, have been published (Glaser et al., 2002; Tettelin et al., 2002, 2005). Comparison of these sequences revealed that approximately 20% of the genome is dispensable, and that strain-specific genes tend to cluster into genomic islands (Tettelin et al., 2005), suggesting that lateral gene transfer (LGT) is an important evolutionary force within *S. agalactiae.* Subsequent studies supported this proposal, with 35 putative integrative conjugative elements (ICE) identified within the eight genomes, while a combination of experimental and *in silico* approaches have shown that large genomic segments (up to 334 kbp) can be exchanged via conjugation between *S. agalactiae* strains (Brochet et al., 2008a,b). Furthermore, LGT via ICE may also occur between *S. agalactiae* and other *Streptococcus* species (Davies et al., 2005, 2009; Haenni et al., 2010). Additional mobile genetic elements (MGE) such as phages have also been implicated in LGT within and among various *Streptococcus* species (Beres and Musser, 2007; Domelier et al., 2009; Ferretti et al., 2001; Holden et al., 2004; Salloum et al., 2010).

Here we provide the first genome sequence of a *S. agalactiae* strain isolated from a cow diagnosed with mastitis. In addition to describing the general features of this genome, we compare it to existing genome sequences of eight human isolates and to the genomes of three pathogens known to be major causes of bovine mastitis: *Streptococcus uberis*, *Streptococcus dysgalactiae* subsp. *dysgalactiae*, and *Staphylococcus aureus* (Bradley, 2002; Calvinho et al., 1998; Zadoks and Fitzpatrick, 2009). We also use PCR to screen additional human and bovine *S. agalactiae* isolates for the presence/absence of genes specific to the bovine GBS genome and identify possible virulence factors associated with bovine mastitis. In addition, we look for factors linking the bovine *S. agalactiae* strain to other mastitis-causing pathogens. Our findings reveal a distinctive bovine *S. agalactiae* genome and provide evidence for LGT between bovine *S. agalactiae* and other mastitis-causing pathogens highlighting the importance of LGT as an evolutionary force and agent of ongoing gene exchange among *Streptococcus* species occupying the same habitat.

2. Materials and methods

2.1. Strain selection, sequencing and assembly

S. agalactiae strain FSL S3-026 was isolated from a quarter milk sample obtained from a cow diagnosed with mastitis (Wanger and Dunny, 1987). The sample was collected on July 25th, 2000 from a cow located in Albany County, NY, and was serotype III (Dogan et al., 2005) and ST 67 (Zadoks and Schukken unpublished data). Strain FSL S3-026 was grown overnight on 50 ml Todd Hewitt Broth at 37 °C, and genomic DNA extraction was performed using the Qiagen Gentra Puregene Yeast/Bacteria kit (scaled up from the 5 ml kit protocol). Roche/454 pyrosequencing was used to determine the genome sequence. A total of 121,112 single-end reads and 151,772 paired-end reads were obtained on the GS-FLX sequencer. *De novo* assembly with Newbler Assembler Software version 1.1.03.24 yielded 390 contigs arranged in 23 scaffolds for an average coverage of 22-fold. Whenever possible, the order and orientation of the scaffolds were determined by aligning the 454 scaffold to one of the published *S. agalactiae* genome sequences. Inter and intra-scaffold gaps were then closed by Sanger and Illumina sequencing, and assembled into eight contigs using Sequencher version 4.9 (Gene Codes Corporation: www.genecodes.com). Although the genome sequence is not fully closed, we were able to estimate the size of several of the remaining gaps using PCR and sequencing. The gap between contigs 8 and 1 appeared to be a ~6.5 kbp duplication of a 16S-23S-5S rRNA operon and associated tRNAs. The gap between contigs 1 and 2 was approximately 500 bp. The gaps between contigs 2 and 3 and between contigs 5 and 6 may contain a large direct repeat of 49,517 bp; however, PCR and sequencing produced conflicting results as to the presence or absence of the repeat. We were unable to estimate the size of the gaps between contigs 3 and 4 and between contigs 6 and 7. For the gaps between contigs 4 and 5, and contigs 7 and 8, comparison to *S. agalactiae* strain A909 revealed, at the corresponding positions, the presence of 1.5 kbp and 1 kbp sections respectively, containing imperfect tandem repeats. If similar sequence exists within the bovine genome at these positions, it may have prevented successful sequencing. This Whole Genome Shotgun project has been deposited at DDBJ/EMBL/GenBank under the accession AEXT00000000. The version described in this paper is the first version, AEXT01000000.

2.2. Genome annotation and gene ontology analysis

Annotation of the *de novo* genome assembly of strain FSL S3-026 was performed using the NCBI Prokaryotic Genomes Automatic Annotation Pipeline (PGAAP). Using COG and CMR JCVI role categories we tested for significant enrichment or under-representation of each category within the GBS bovine genome. The eight published *S. agalactiae* genome sequences, derived from human disease isolates, were compared pairwise to the bovine genome. Relative enrichment or under-representation was ranked using odds ratios, and statistical significance was measured using a Fisher exact test. The false discovery rate (FDR) procedure of Benjamini and Hochberg (1995) was used to correct for multiple hypothesis testing (FDR = 0.05).

2.3. Orthology reconstruction

OrthoMCL version 2.0 (Li et al., 2003) was used to delineate orthologous protein sequences among the eight *S. agalactiae* genomes derived from human disease isolates, strain FSL S3-026 (bovine mastitis isolate), and the genomes of three other pathogens also responsible for bovine mastitis (*S. uberis*, *S. dysgalactiae* subsp. *dysgalactiae*, and *S. aureus*) (see Table 1 for accession and strain ID numbers). The first step in the OrthoMCL procedure was to perform a reciprocal BLASTp within and between all genome pairs. The resulting *p* values were then used to build a normalized similarity matrix, which was analyzed using a Markov Cluster algorithm to delineate proteins into clusters containing sets of orthologs and/or recent paralogs. Proteins were considered recent paralogs if they were more similar to each other than to any protein from another genome. Following Li et al. (2003), a *p* value cut-off of 1e−5 was used in the BLASTp. Whenever possible, OrthoMCL clusters were used to help deduce putative functions of genes that were annotated as hypothetical proteins by the PGAAP.

Table 1
Accession numbers for each species and strain used in the orthology reconstruction.

Accession number	Species and strain ID
AEXT01000000	*Streptococcus agalactiae* FSL S3-026
NC_004116	*Streptococcus agalactiae* 2603V/R
NC_004368	*Streptococcus agalactiae* NEM316
NC_007432	*Streptococcus agalactiae* A909
NZ_AAJP00000000	*Streptococcus agalactiae* 515
NZ_AAJQ00000000	*Streptococcus agalactiae* CJB111
NZ_AAJR00000000	*Streptococcus agalactiae* COH1
NZ_AAJS00000000	*Streptococcus agalactiae* H36B
NZ_AAJO00000000	*Streptococcus agalactiae* 18RS21
NC_012004	*Streptococcus uberis* 0140J
AEGO01000000	*Streptococcus dysgalactiae* subsp. *dysgalactiae* ATCC 27957
NC_007622	*Staphylococcus aureus* RF122

Fragmented protein sequences, such as those that span separate contigs or insertion sequences, can be erroneously categorized as distinct orthologs. To correct for this, clusters containing single proteins were not considered distinct orthologs (rather fragments of the same protein) if they met the following criteria: (i) showed strong homology with another cluster (i.e. could potentially group together to form a single orthologous cluster), (ii) failed to group together because the protein clustering independently showed no reciprocal BLASTp hit with one of the proteins in the second cluster, (iii) the two proteins showing no reciprocal BLASTp hit originated from the same genome. Proteins that were larger than 30 amino acids and had no BLASTp hit with any other protein were considered strain specific (*E*-value $\leq$ 1e−10).

Candidate bovine strain-specific genes not annotated as hypothetical proteins were then searched against NCBI databases (nr/nt and nr) using BLASTn and BLASTp to confirm their specificity. For the genes that remained specific, internal PCR primers were designed and 20 bovine and 20 human source isolates were screened for presence/absence of these genes (strain FSL S3-026 was included in the screening as a positive control). Although nucleotide polymorphism at the PCR priming site could potentially yield a false negative, the relatively low level of nucleotide divergence, typically observed at the intra-specific level of *Streptococcus* taxa, should render this a rare enough event that it would not significantly impact our overall conclusions. The *S. agalactiae* strain FSL S3-026 genome contained a high frequency of insertion sequences, often in close proximity. To avoid the confounding effect of multiple priming sites, bovine strain-specific insertion sequences were not screened. The bovine strains were obtained from quarter milk samples collected from 20 herds in 15 counties throughout New York, USA (Table 2). The human samples came from New York State, Germany, Sweden, Belgium, and Norway.

2.4. Deferred antagonism test

Production of the lantibiotic nisin by bovine *S. agalactiae* strains was detected using a deferred antagonism test. The test was performed following the protocol of Tagg and Bannister (1979)

Table 2
Collection data for *Streptococcus agalactiae* strains PCR screened for the presence/absence of genes specific to bovine strain FSL S3-026.

Strain	Host	Source	Collection date	County/city	State/country	Serotype	Sequence type/clonal group
FSL S3-026	Bovine	1/4 milk sample	7/25/00	Albany	NY	III	Clonal group AH
FSL S3-271	Bovine	1/4 milk sample	11/14/00	Steuben	NY	NT	Clonal group AB
FSL C1-487	Bovine	1/4 milk sample	5/16/00	Allegany	NY	NT	Clonal group AC
FSL S3-062	Bovine	1/4 milk sample	6/15/00	Cattaraugus	NY	II	Clonal group AG
FSL S3-342	Bovine	1/4 milk sample	12/6/00	Livingston	NY	III	Clonal group AI
FSL S3-586	Bovine	1/4 milk sample	3/3/01	Wayne	NY	II	Clonal group AJ
FSL S3-591	Bovine	1/4 milk sample	3/14/01	Cattaraugus	NY	III	Clonal group AK
FSL S3-071	Bovine	1/4 milk sample	7/14/00	Montgomery	NY	III	Clonal group AL
FSL S3-034	Bovine	1/4 milk sample	5/28/00	Wyoming	NY	II	Clonal group N
FSL S3-222	Bovine	1/4 milk sample	10/4/00	Montgomery	NY	III	Clonal group N
FSL S3-277	Bovine	1/4 milk sample	11/10/00	Ontario	NY	III	Clonal group O
FSL S3-105	Bovine	1/4 milk sample	6/5/00	Steuben	NY	III	Clonal group P
FSL C1-494	Bovine	1/4 milk sample	5/16/00	Erie	NY	NT	Clonal group Q
FSL S3-229	Bovine	1/4 milk sample	10/9/00	Saratoga	NY	NT	Clonal group R
FSL S3-027	Bovine	1/4 milk sample	7/25/00	Schoharie	NY	II	Clonal group S
FSL S3-043	Bovine	1/4 milk sample	5/31/00	Columbia	NY	III	Clonal group T
FSL S3-603	Bovine	1/4 milk sample	4/11/01	Ontario	NY	III	Clonal group U
FSL S3-114	Bovine	1/4 milk sample	7/28/00	Wyoming	NY	III	Clonal group V
FSL S3-128	Bovine	1/4 milk sample	Unknown	Onondaga	NY	II	Clonal group X
FSL S3-568	Bovine	1/4 milk sample	2/15/01	Yates	NY	NT	Clonal group Z
FSL S3-023	Human	Blood	6/12/00	Greene	NY	V	Clonal group A
FSL S3-337	Human	Blood	11/00	Schenectady	NY	III	Clonal group B
CCUG 91	Human	Cerebrospinal fluid	07/11/68	Göteborg	Sweden	NA	NA
CCUG 24810	Human	Ear	06/11/89	Göteborg	Sweden	NA	NA
CCUG 29376	Human	Neonatal blood, complete recovery	11/11/91	Göteborg	Sweden	NA	NA
CCUG 29782	Human	Septicemia	Unknown	Unknown	Unknown	III	NA
CCUG 37740	Human	Joint, arthritis, diabetes	08/21/96	Göteborg	Sweden	NT	NA
CCUG 39096 A	Human	Urine	02/98	Linköping	Sweden	NA	NA
CCUG 44110	Human	Vagina, fornix, healthy	06/17/99	Göteborg	Sweden	NA	NA
CCUG 44186	Human	Cervix	08/18/00	Göteborg	Sweden	NA	NA
CCUG 45061	Human	Blood, neonatal	03/15/01	Tromsö	Norway	NA	NA
ATCC BAA-1175	Human	Throat	Unknown	New York	NY	II	NA
LMG 15090	Human	Urinary tract infection	1994	Liège	Belgium	Ib	NA
LMG 15091	Human	Urinary tract infection	1994	Liège	Belgium	IV	NA
LMG 15093	Human	Vagina	1994	Liège	Belgium	V	NA
LMG 15094	Human	Urinary tract infection	1994	Liège	Belgium	III	NA
Spellerberg-252	Human	Respiratory tract CF patient	2004	Münster	Germany	V	ST 1
Spellerberg-248	Human	Respiratory tract CF patient	2004	Münster	Germany	Ib	ST 12
Spellerberg-188	Human	Vagina	2003	Aachen	Germany	Ia	ST 2
Spellerberg-165	Human	Vagina	2003	Aachen	Germany	II	ST 28

FSL strains characterized previously by Sukhnanand et al. (2005) (clonal groups were defined as unique combinations of sub-types [EcoRI ribotypes and *hylB* and *sodA* allelic types]). NA = not available.

with the following modifications: (i) strains were grown on Columbia Agar with 5% sheep's blood, (ii) producer strains were grown in a streak at 37 °C and 7% CO_2 overnight, (iii) after the indicator strain was applied, plates were left overnight at 28 °C. Producer strain: *S. uberis* ATCC 27958, indicator strain: *Kocuria rhizophila* ATCC 9341.

3. Results and discussion

3.1. General features of the bovine Streptococcus agalactiae strain FSL S3-026 genome

The total length of the eight concatenated strain FSL S3-026 contigs was 2,455,848 bp. Fig. 1 shows the proposed position of the contigs and unsequenced gaps. Our assembly shows the characteristic pattern of either positive or negative GC skew in each half of the genome, as well as the same clustering of 16S-23S-5S rRNA operons and associated tRNAs immediately after the origin of replication, as seen for other *S. agalactiae* genome sequences. The three complete genomes derived from human disease isolates (A909, 2603V/R, and NEM316) ranged in size from 2,127,839 bp to 2,211,485 bp. Consequently, the concatenation of the bovine GBS contigs was approximately 290 kbp (12%) larger than the mean of the three human GBS genomes. GC content of the bovine GBS genome was 36.1%, slightly higher than the three human GBS genomes, which were all 35.6%. The bovine GBS genome contained 2334 CDS. Of these, 15 appeared to be smaller fragments of other protein sequences. The human GBS genomes contained on average 248 fewer CDS (A909 = 1996, 2603V/R = 2124, NEM316 = 2094).

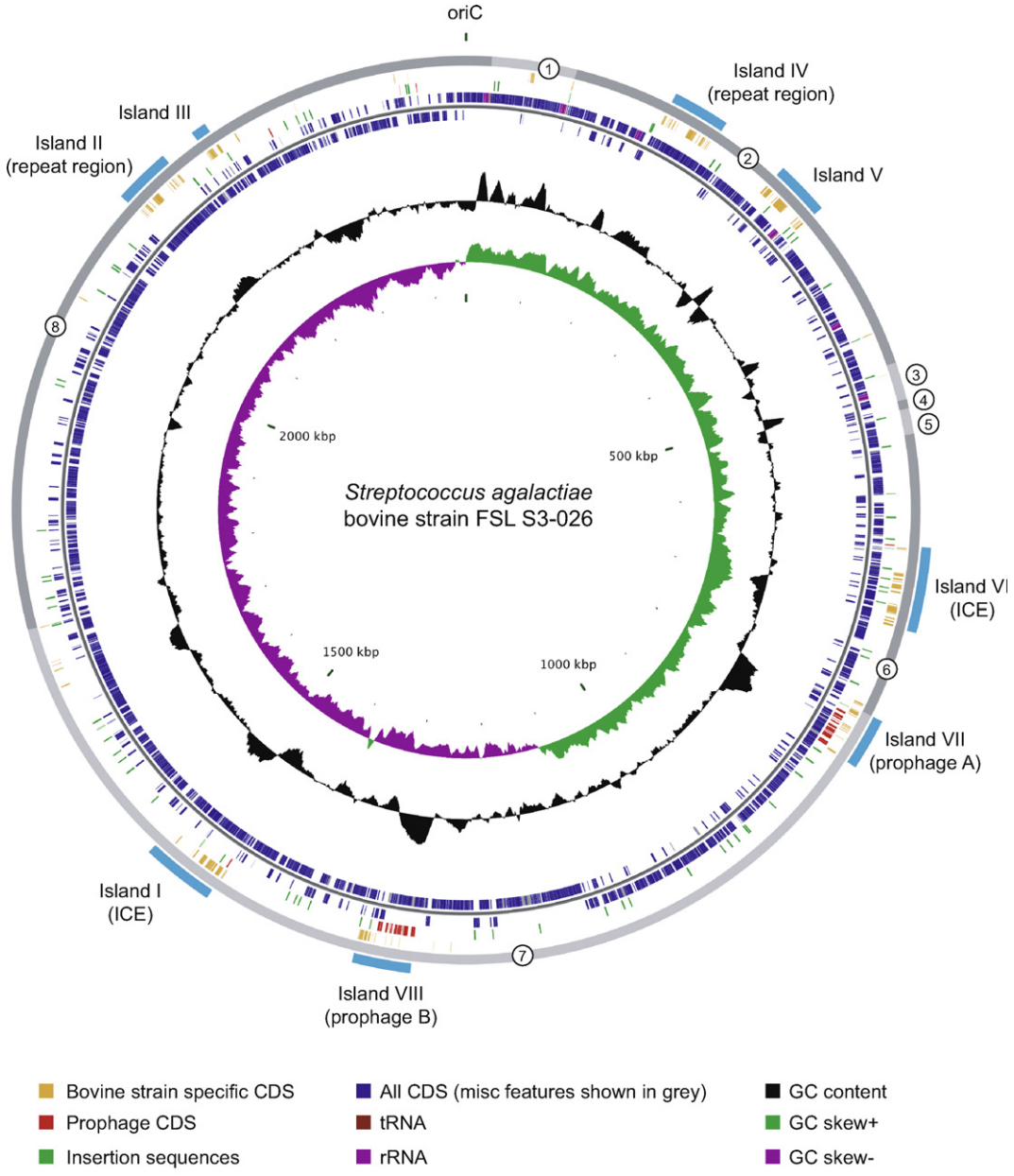

Fig. 1. Genome map of *Streptococcus agalactiae* bovine strain FSL S3-026. Starting from the outermost ring and moving inwards, rings show the location of: (1) eight genomic islands (see text for detailed description), (2) eight assembled contigs (numbers circled), (3) bovine strain specific CDS, (4) prophage CDS and insertion sequences, (6) all annotated CDS on the leading strand, and (7) all annotated CDS on the lagging strand. Two innermost rings show GC content and GC skew. Map was created using the software CGView (Stothard and Wishart, 2005).

The number of rRNA and tRNA genes in the bovine GBS genome sequence was 18 and 70 respectively, whereas the human GBS genomes all contained 21 rRNA genes and 80 tRNA genes. However, if the proposed replication of the 16S-23S-5S rRNA operon and associated tRNAs between contigs 8 and 1 is correct, this could reconcile the discrepancy. When compared to 42 genome sequences from 10 other *Streptococcus* species (see Table S1 in the supplementary material), the largest genome (concatenated contigs and number of genes) belonged to the bovine *S. agalactiae* strain, possibly reflecting a wide range of niches already available to *S. agalactiae*, followed by subsequent adaptation to a bovine host. In addition, all of the *S. agalactiae* genomes contained considerably more RNA genes than other *Streptococcus* species. A possible explanation for this observation is selection for rapid growth in *S. agalactiae,* as genes expressed at high levels often show codon usage bias (Kanaya et al., 1999), and Sharp et al. (2005) demonstrated that codon usage bias was strongly positively correlated with the number of rRNA operons within a bacterial genome. Approximately one third of the complete and draft GBS human genome sequences consist of CDS annotated as hypothetical proteins (Tettelin et al., 2005). Similarly, 34.7% of the CDS within the bovine GBS genome were annotated as hypothetical proteins.

3.2. Insertion sequences

A distinctive feature of the bovine GBS genome, compared to all eight of the human GBS genomes, was the high frequency of insertion sequences (IS). The bovine GBS genome contained 97 IS that grouped into 14 orthologous clusters. The human GBS genomes contained an average of 10.1 IS orthologous clusters, with a mean number of IS within the genome of 20.4 (Fig. 2). These findings are concordant with the results of the gene ontology analysis, which showed significant enrichment of the JCVI transposon function category for each of the pairwise comparisons ($p \ll 0.05$). With the exception of comparisons to strains NEM316 and COH1, which showed FSL S3-026 enrichment for prophage functions ($p \ll 0.05$), no other category was significantly enriched (see Fig. S1 in the supplementary material).

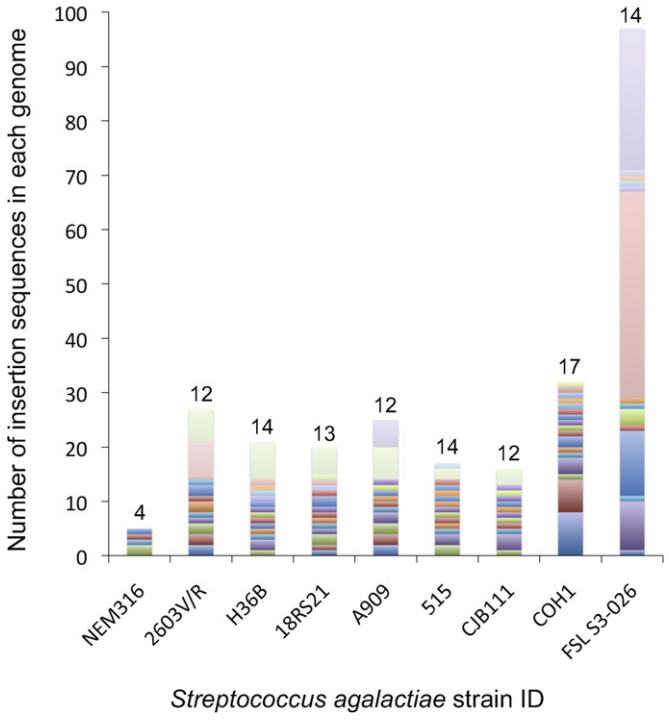

Fig. 2. Number of insertion sequences (IS) in eight human source *Streptococcus agalactiae* genomes and the bovine source *S. agalactiae* genome (FSL S3-026). Each different fill color in the bars represents one of 47 separate orthologous clusters of IS. Numbers over bars indicate the number of orthologous IS clusters for each genome. IS from the human genomes fell into 38 clusters. IS from the bovine genome fell into 14 clusters.

A total of 87.6% of the bovine IS grouped into four orthologous clusters, with counts of 38, 26, 12, and 9 in each cluster. Average sequence identity among sequences within each cluster was 99.0%, with sequence identity within the most frequent cluster particularly high (99.9%), suggesting a recent proliferation of IS. Furthermore, almost one third (29%) of the bovine IS occurred in tandem, an arrangement that can potentially lead to high transposase expression (Mahillon and Chandler, 1998). Collectively, this suggests elevated mobility of genetic elements within the bovine GBS genome, many of which may have been acquired recently. Indeed, seven protein sequences within the genome appear to have been split into two by the insertion of an IS (see Table S2 in the supplementary material). This type of insertion has previously been shown to inactivate proteins within strains of *S. agalactiae* (Granlund et al., 1998; Spellerberg et al., 2000), and could serve as an important evolutionary mechanism, since the inactivation of a gene within an operon can occasionally result in alternative functionality and adaptation of the organism (Phelps and Neely, 2007). The IS may also contribute to elevated mobility of genes among *S. agalactiae* isolates. With a few rare exceptions (Parkhill et al., 2001, 2003), the proportion of IS within a bacterial genome generally falls below 3% (Siguier et al., 2006). Consequently, the proportion of IS within the bovine GBS genome (4.3%) is unusually high. A possible explanation for high IS proportions is an evolutionary bottleneck resulting from population isolation (Parkhill et al., 2003). This is because IS, which are often associated with selectively disadvantageous traits and competitively removed in typical populations, are able to accumulate in very small populations, where they are less likely to be removed by competition (Parkhill et al., 2003; Siguier et al., 2006). A bottleneck or founder effect origin for bovine GBS is consistent with the analysis of Bisharat et al. (2004), which suggests more limited genetic diversity in the bovine population compared to human.

3.3. Strain-specific genes

Compared to the human GBS genomes, 183 orthologs were identified as specific to the bovine GBS genome. This included 103 orthologs (56%) that were annotated as "hypothetical proteins." Of the remaining 80, seven were insertion sequences and were excluded from the subsequent PCR screening of 20 bovine and human GBS strains. Of the 73 FSL S3-026 strain-specific genes screened (Fig. 3), half (37) were observed exclusively within bovine strains. Slightly fewer (36) were observed in both bovine and human strains; however, these were far more common in bovine than in human strains: average frequency of occurrence within each group was 60% (bovine) as opposed to 14% (human). These results, combined with that of a Fisher exact test ($p < 0.0001$), clearly show strong differentiation between bovine and human strains of *S. agalactiae*, confirming, at the genomic level, the findings of previous molecular studies suggesting differentiation between bovine and human populations (Bisharat et al., 2004; Bohnsack et al., 2004; Dogan et al., 2005; Sukhnanand et al., 2005). Collectively, these findings suggest a low potential for transmission of *S. agalactiae* between bovine and human hosts. However, rare transmission may nonetheless be possible, particularly if there is prolonged bovine–human contact (Manning et al., 2010).

Numerous *S. agalactiae* membrane proteins have been implicated in virulence (Lindahl et al., 2005), and almost half (17) of the

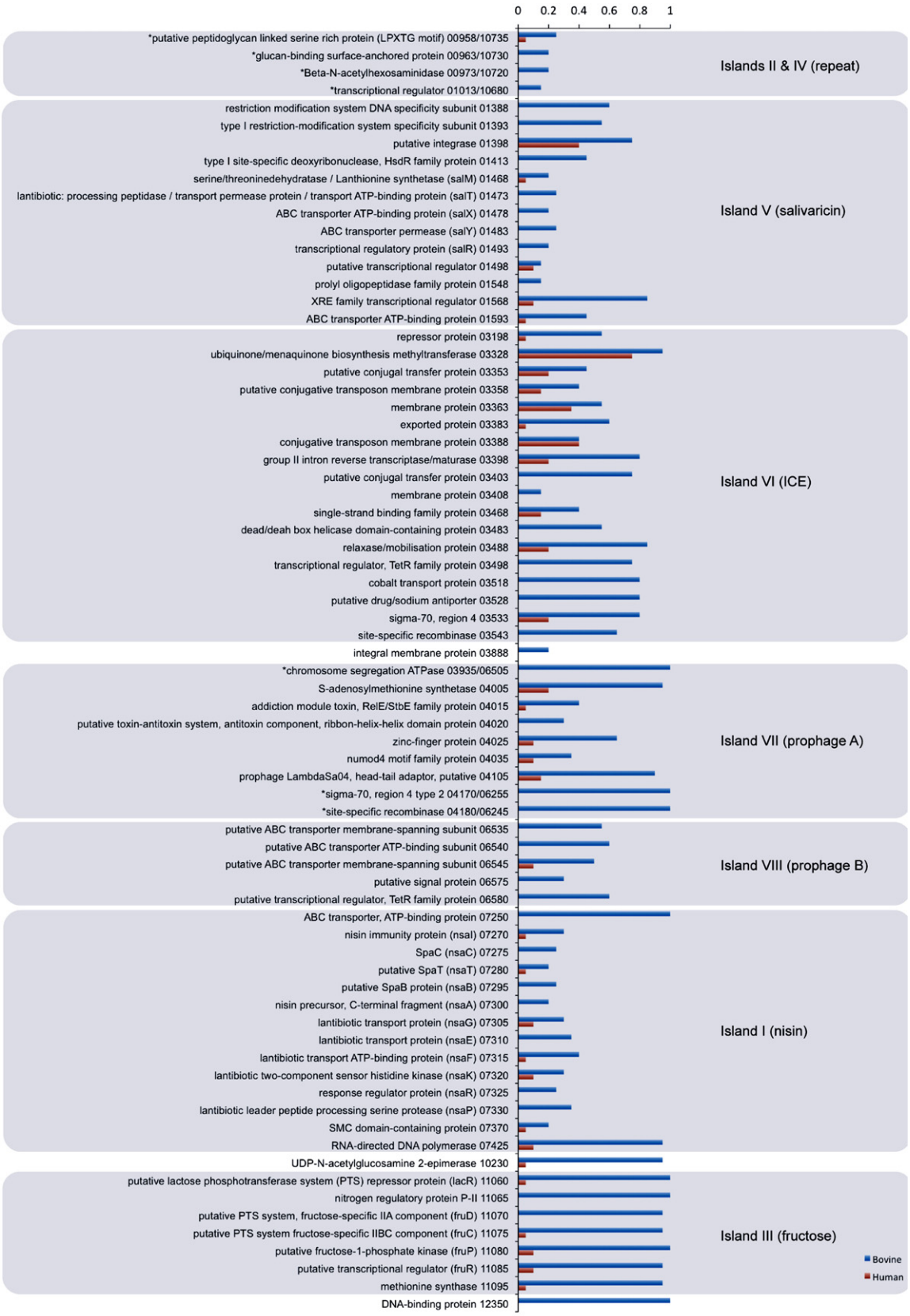

Fig. 3. Results of PCR screening for presence/absence of 73 orthologs identified as specific to *Streptococcus agalactiae* bovine strain FSL S3-026 when compared to eight human GBS strains; 20 bovine and human GBS strains were screened (bovine strains included FSL S3-026 as a positive control). Blue bars show frequency of occurrence in bovine GBS strains and red bars show frequency for human GBS strains. CDS annotations are followed by locus IDs. Orthologs belonging to the separate genomic islands described in text are boxed. Seven orthologs (indicated with an asterisk) existed as duplicate copies (four in the repeats and three in the prophages), and although the screening could determine general presence/absence for these orthologs, it could not determine which particular copy was present. The three prophage orthologs are only shown within genomic island VII.

genes exclusive to the bovine strains were annotated as producing some type of membrane protein (typically membrane transport or binding proteins). More specifically, six of the bovine-specific genes were ATP-binding cassette (ABC) transporter proteins, which in addition to *S. agalactiae* (Jones et al., 2000), have also been shown to be important in the virulence of other bacteria, for example, *Streptococcus pneumoniae* and *S. aureus* (Basavanna et al., 2009; Mei et al., 1997). These bovine-specific membrane proteins could be relevant for future vaccine development.

3.4. Genomic islands

With the exception of three genes (locus IDs: 03888, 10230, 12350), all genes specific to bovine strain FSL S3-026, not annotated as hypothetical proteins, were tightly clustered into eight regions, to which we refer here, as genomic islands (Figs. 1 and 3). A total of 25 of the bovine strain-specific hypothetical proteins (20%) occurred outside these genomic islands. This suggests that the majority of these strain-specific genes were acquired via LGT of these genomic islands, a pattern also reported for the eight published *S. agalactiae* genomes arising from human sources (Glaser et al., 2002; Tettelin et al., 2005).

3.5. Genomic island I (nisin)

Genomic island I contained 24 strain FSL S3-026 specific genes, including seven hypothetical proteins (Figs. 1 and 3) and the region surrounding this island was characteristic of an ICE, encompassing 59 genes (Fig. 4). Typically, ICE have (i) a site-specific recombinase such as an integrase gene at their 3′ end, which is used to catalyze recombination, (ii) relaxase and conjugation genes involved in the conjugation process, and (iii) recombination sites known as *attL* and *attR* that flank the element. ICEs usually insert at the 3′ end of a tRNA or protein-encoding gene (often ribosomal) (Burrus et al., 2002; Burrus and Waldor, 2004). The putative ICE surrounding genomic island I had all of these general characteristics (Fig. 4). The integrase gene was a site-specific recombinase and the ICE was inserted at the 3′ end of the ribosomal gene *rplL*. In addition, it was located in approximately the same genomic position as ICEs in four other human source *S. agalactiae* genomes (2603V/R, NEM316, 18RS21, and H36B) (Brochet et al., 2008a). These ICEs shared the same general characteristics as the putative bovine ICE and were also inserted at the 3′ end of *rplL*. Following the nomenclature protocol of Brochet et al. (2008a), where an ICE is named according to its insertion site, we designated this putative bovine strain ICE as ICE_FSL S3-026_*rplL*. Putative recombination sites *attL* and *attR* were identified as follows: The *attL* site was an A + T rich (62.7%) region of 198 bp located immediately downstream of *rplL*, which shared 97.5% sequence identity with regions of the same length and position for three of the ICE mentioned above (2603V/R, 18RS21, and H36B). The *attR* site, also A + T rich (66.7%), was a region of 195 bp located immediately upstream of the integrase gene. This region shared 81.6% sequence identity with regions of the same length and position for the same three ICE (2603V/R, 18RS21, and H36B). Finally, *attR* contained a 15 bp motif that was a direct repeat of 15 bp at the 3′ end of *rplL*, which was the ICE insertion site. The complete ICE_FSL S3-026_*rplL* shared 46–65% identity to the ICE in the four human GBS strains.

A major feature within ICE_FSL S3-026_*rplL* was the presence of an 11-gene operon (14,973 bp) (Fig. 4) which showed close sequence similarity (78.4%) and identical gene order to the nisin U operon of *S. uberis* strain 42 (Wirawan et al., 2006). Exclusion of an IS in the middle of the bovine GBS operon (see Fig. 4) increased the sequence identity to 86.3%. Although some of the genes within the operon showed close sequence similarity to genes from several other *Streptococcus* species, as well as *Lactococcus lactis*, *S. uberis* was the only species that matched the entire operon (Fig. 4). Earlier molecular characterization of the nisin U operon in *S. uberis* strain 42 focused exclusively on the operon and provided no information regarding flanking regions of DNA (Wirawan et al., 2006). Furthermore, the only published *S. uberis* genome (strain 0140J) (Ward et al., 2009) lacks the operon. Consequently, it is not possible to confirm, with currently available data, whether the

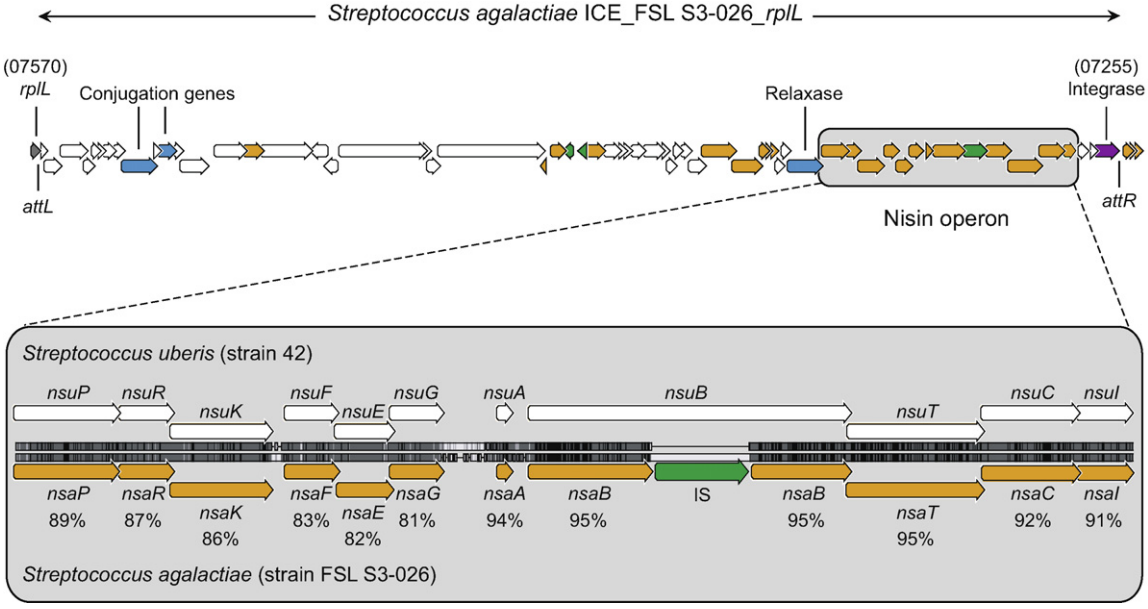

Fig. 4. Gene organization within putative integrative conjugative element ICE_FSL S3-026 (genomic island I). Locus IDs for the *rplL* and integrase genes are shown in parentheses. Orange genes are specific to the bovine *Streptococcus agalactiae* strain FSL S3-026 when compared to eight other human *S. agalactiae* strains. Grey shaded boxes depict the 11 gene nisin operon. Bottom box shows an alignment between the nisin U operon from *Streptococcus uberis* (strain 42) (white genes) and the operon found within the putative ICE (ICE_FSL S3-026_*rplL*) of *S. agalactiae* (strain FSL S3-026). The nisin genes in *S. agalactiae* are designated here as *nsa*. Nucleotide sequence identities are shown below each gene. The green gene is an insertion sequence (IS) that has fragmented *nsaB*. Two horizontal bars are a generalized representation of the aligned nucleotide sequences, with black shading representing 100% identity. Figure created using the programs Geneious v5.1 (Drummond et al., 2010) and Adobe Illustrator.

nisin U operon in *S. uberis* is carried within an ICE. However, prior to the report of nisin within *S. uberis*, nisin was only reported in *L. lactis*, and within this species it occurs within a MGE described as a conjugative transposon, which might be more appropriately described as an ICE due to its site-specific insertion (Burrus and Waldor, 2004; Rauch and De Vos, 1992; Wirawan et al., 2006). Recent studies have shown LGT mediated via ICEs for numerous *Streptococcus* species (Davies et al., 2009; Holden et al., 2009), and collectively, the observations and findings reported above support the hypothesis that the nisin U operon was laterally exchanged between *S. uberis* and *S. agalactiae* via an ICE.

Environmental proximity is likely a dominant factor determining LGT among bacterial species. For example, Delorme et al. (2007) reported LGT between two Salivarius *Streptococcus* species (*Streptococcus salivarius* and *Streptococcus vestibularis*) that share the human oral cavity, and Bolotin et al. (2004) reported LGT among *Streptococcus thermophilus*, *Lactobacillus bulgaricus*, and two sub species of *L. lactis*, which are all found in milk. When compared to the eight human *S. agalactiae* genomes, the nisin U operon was specific to the bovine strain FSL S3-026 genome. In addition, the PCR screening results showed the operon to be present in one fifth of the 20 bovine isolates screened and none of the 20 human GBS isolates screened. Consequently, the presence of the nisin U operon appears correlated with a bovine environment, and its presence in two mastitis-causing pathogens (*S. uberis* and bovine *S. agalactiae*) also suggests LGT within a shared environment.

Wirawan et al. (2006) PCR screened 45 *S. uberis* strains for the presence of the nisin U structural gene (*nsuA*) and its two flanking genes (*nsuB* and *nsuG*). Similar to our results for bovine *S. agalactiae* strains, the genes were only present in 20% of strains, suggesting that only a small proportion of strains within *S. uberis* possessed the ability to produce nisin. Pryor et al. (2009) examined the effect of simultaneously infecting cows with both nisin producer and non-producer strains of *S. uberis*. Results showed that producer strains usually predominated in intramammary infection, suggesting that the ability to produce nisin may have provided these strains with a competitive advantage during infection. The ability to produce nisin might confer a similar advantage to bovine *S. agalactiae* strains. However, results of our deferred antagonism test showed that none of the *S. agalactiae* strains possessing the nisin operon could produce nisin. Truncation of the *nsaB* gene by the internal IS is likely an important factor contributing to strain FSL S3-026's inability to produce nisin (see Fig. 4). Nonetheless, it is possible that some bovine *S. agalactiae* strains may possess functional nisin operons (i.e. free of IS disruption), providing them with a competitive advantage during mastitis infection.

3.6. Genomic island III (fructose–lactose)

Genomic island III contained a four-gene operon involved in fructose utilization: (i) phosphotransferase system (PTS) fructose-specific IIA component, (ii) fructose-specific IIBC PTS component, (iii) fructose-1-phosphate kinase, and (iv) transcriptional regulator (designated *fruD*, *fruC*, *fruP*, and *fruR* respectively) (Figs. 3 and 5). Five other bovine strain FSL S3-026 specific genes flanked this operon. Annotated products were a putative lactose PTS repressor protein, a nitrogen regulatory protein P-II, a fragmented methionine synthase, and a hypothetical (designated *lacR*, *N*, *meth*, and *H* respectively) (Fig. 5). The PCR screening results showed the complete fructose operon to be present in 95% of the bovine samples and none of the human samples. The lactose repressor, nitrogen regulatory, and methionine synthase genes were similarly differentiated with a significant bovine bias (Fig. 3). Regarding the fructose operon, this differentiation suggests a unique fructose utilization pathway for the bovine population of *S. agalactiae*, which might facilitate survival in potential extramammary reservoirs such as the bovine rumen. The nitrogen regulatory gene was also strongly differentiated, present in all bovine samples screened and none of the human samples. The product of this gene belongs to a family of signal transducing adaptor proteins that control nitrogen metabolism, suggesting that nitrogen metabolic processes might be distinct in the bovine *S. agalactiae* population.

The gene for the lactose PTS repressor protein (*lacR*) was within a nine-gene lactose utilization operon (*lacXGEFDCBAR*) just downstream of the fructose operon. The combined sequence of the lactose operon, the fructose operon, the methionine synthase, the nitrogen regulatory protein, and three genes downstream of the lactose operon (two hypotheticals, and an integrase) forms a sequence of 16,031 bp that shares 99.0% sequence identity with a contiguous sequence in *S. dysgalactiae* subsp. *dysgalactiae* (strain ATCC 27957) (Fig. 5). This sequence identity excluded two IS that were inserted between the integrase gene and a hypothetical in *S. agalactiae* (FSL S3-026) (Fig. 5). An approximately 60 kbp genomic region surrounding this 16 kbp sequence in *S. dysgalactiae* subsp. *dysgalactiae* was characteristic of an ICE, with an integrase gene at its 3′ end, a relaxase gene, and four conjugation genes. Excluding the highly similar 16 kbp region shared between *S. dysgalactiae* subsp. *dysgalactiae* and bovine strain FSL S3-026, this putative ICE had a largely similar gene content and gene order to ICE_FSL S3-026_*rplL* (67.2% sequence identity). Therefore, we used an alignment of ICE_FSL S3-026_*rplL* with the putative ICE in *S. dysgalactiae* subsp. *dysgalactiae* to help identify the limits of the element in this subspecies. The ICE appears to have been inserted

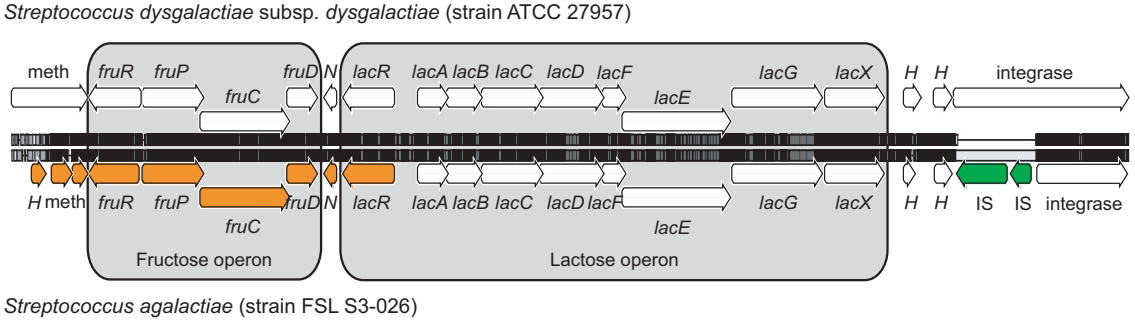

Fig. 5. Gene organization within highly similar 16kbp sequence shared between bovine *Streptococcus agalactiae* strain FSL S3-026 and *Streptococcus dysgalactiae* subsp. *dysgalactiae* strain ATCC 27957. Orange genes are specific to the bovine *S. agalactiae* strain when compared to eight other human *S. agalactiae* strains and correspond to genomic island III. Fructose and lactose gene designations follow Wen et al., 2001 and Siezen et al., 2005 respectively. Green genes are insertion sequence (IS). Two horizontal bars are a generalized representation of the aligned nucleotide sequences, with black shading representing 100% identity.

into a 986 bp A + T rich (70.1%) region of non-coding DNA. The putative *attL* and *attR* sites were 896 bp and 90 bp respectively, and contained a direct repeat of 23 bp. The region surrounding the 16 kbp sequence in the bovine strain FSL S3-026 genome was not characteristic of an ICE. However, this section was flanked by four IS (two upstream and two downstream). This section of the genome, therefore, may have been relocated from an ICE or other MGE. Indeed, another putative ICE was located approximately 51 kbp downstream of this location. This particular ICE (corresponding to locus IDs 11385–11760) was flanked by $tRNA^{Lys}$ and the ribosomal gene *rpmG*. Both genes are potential insertion sites, and numerous MGE have been reported inserted at these two genomic locations in other *S. agalactiae* genomes (Brochet et al., 2008a; Haenni et al., 2010). The very high sequence identity of the 16 kbp section of DNA shared between *S. dysgalactiae* subsp. *dysgalactiae* and bovine strain FSL S3-026 is strongly suggestive of LGT between these two species, and the characteristics of the DNA flanking this section in both species make transportation via an ICE the most likely explanation. Consequently, this appears to be a second example of LGT between S3-026 and another mastitis-causing *Streptococcus* species, sharing the same bovine environment.

Comparisons of the 16 kbp sequence from strain FSL S3-026 and *S. dysgalactiae* subsp. *dysgalactiae* to the genome sequences of the human strains of *S. agalactiae* identified a similar lactose operon in strain NEM316 that contained an additional gene between *lacF* and *lacD*, an antiterminator (*lacT*), which may function to regulate the expression of the operon (Alpert and Siebers, 1997). With the exception of *lacX*, the gene order and content of this lactose operon within strain FSL S3-026 corresponded to the lactose operon designated Lac.2 in *Streptococcus pyogenes* (Ferretti et al., 2001) (*lacX* was absent in *S. pyogenes*). Other human GBS strains, however, do not possess this Lac.2 operon, but instead have a different lactose operon containing 10 genes: *lacXDCBA*-Neuraminidase-IICIIBIIA*lacR* (IIC, IIB, IIA refer to PTS components C, B, and A respectively). With the exception of *lacX* and neuraminidase, this second lactose operon is also shared with *S. pyogenes* (*lacX* and neuraminidase were absent in *S. pyogenes*), and is referred to as Lac.1.

Using a combination of BLASTn and DNA sequence alignments, *Streptococcus* genomes in the NCBI database were searched for the presence of Lac.1 and Lac.2. Lac.2, which was absent in all human strains of *S. agalactiae* except NEM316, was detected in a wide diversity of 12 additional species representing four of the six streptococcal taxonomic groups: Pyogenic, Mitis, Mutans, and Bovis (Kawamura et al., 1995) (see Table S3 and Fig. S4 in the supplementary material). In contrast, Lac.1 was only detected in four species, in addition to *S. agalactiae* (Table S3) (all five species belonged to the Pyogenic group). Three of these taxa (*S. pyogenes*, *S. uberis*, and *S. dysgalactiae* subsp. *dysgalactiae*) also contained Lac.2. Although gene content of each lactose operon was generally conserved across all species, there were some differences (Figs. S3 and S4). Interestingly, three of the four *Streptococcus* species that possessed both Lac.1 and Lac.2 (*S. agalactiae*, *S. dysgalactiae* subsp. *dysgalactiae*, and *S. uberis*) also cause bovine mastitis. However, in FSL S3-026 Lac.1 was fragmented by two IS, and two genes (*lacX* and *lacD*) were missing, possibly rendering the operon non-functional. This appears to suggest, therefore, that bovine and human strains of *S. agalactiae* may have distinct lactose utilization pathways involving Lac.1 in human strains and Lac.2 in bovine strains. However, Loughman and Caparon (2006, 2007) have shown that for *S. pyogenes* Lac.1 no longer participates in lactose utilization. Rather, the operon (specifically *lacD.1*) has adapted to become a metabolic sensor that regulates the transcription of a virulence gene (*speB*). The Lac.1 operon in *S. agalactiae* may have similarly adapted to a role in virulence. More specifically, it may be utilizing the neuraminidase-related protein. This locus showed 58% sequence identity to the neuraminidase gene *nanA* of *S. pneumoniae*, which has been shown to be important in colonization and persistence in the nasopharynx and middle ear (Tong et al., 2000); neuraminidase has also been implicated as a virulence factor in a number of other bacteria (Soong et al., 2006). Lac.2, within strain NEM316, was located within ICE_NEM316_*rplL* (Brochet et al., 2008a), and given the prevalence of Lac.2 within other *Streptococcus* species, and its absence in the other human strains of *S. agalactiae*, it seems likely that strain NEM316 acquired Lac.2 via LGT from a species of *Streptococcus* other than *S. agalactiae*. Sequence identity for Lac.2 among *Streptococcus* species provides some support for this hypothesis, as the highest identity for strain NEM316 was with *Streptococcus gordonii* (*lacGEFTDCBA* = 76.8%), whereas sequence identity between strains NEM316 and FSL S3-026 was 59.6%.

3.7. Genomic islands VII and VIII (prophages)

In addition to ICEs, other MGEs such as phages have been shown to be important vehicles for LGT within and among *Streptococcus* species (Boyd and Brüssow, 2002; Brenciani et al., 2010; Canchaya et al., 2004; Davies et al., 2007). Two regions within the bovine strain FSL S3-026 genome, corresponding to genomic islands VII and VIII, were characteristic of prophages (Fig. 1). Each region displayed the distinctive modular arrangement of tailed prophage structural genes described for lactic acid bacteria (Canchaya et al., 2003), and shared high sequence identity with prophage LambdaSa04 of *S. agalactiae* strain A909 (Brenciani et al., 2010; Tettelin et al., 2005). The putative prophages corresponding to genomic islands VII and VIII were designated prophages A and B respectively. Each prophage was located on an opposite replichore and orientated in the opposite direction (Fig. 1). End points of these putative prophages were determined by identifying putative *attL*/*attR* sites (see Fig. S2 in the online supplementary material). Putative prophage B contained 61 genes (53,465 bp), with the *attR* end containing a span of 41 genes (38,325 bp) including 22 structural prophage loci (Fig. S2). Within this section, there were 15 hypothetical proteins, three of which were bovine strain FSL S3-026 specific. This 41-gene section shared 80.3% sequence identity with prophage LambdaSa04 of *S. agalactiae* strain A909. The *attL* end of prophage B contained 20 genes (15,140 bp), of which 13 were FSL S3-026 strain specific. Of the strain-specific genes, eight were annotated as hypothetical proteins. The remaining five genes encoded three putative ABC membrane transporter subunits, a putative signal protein, and a putative transcription regulator (*TetR*) (Fig. 3). On average, the PCR screening detected each of these genes in approximately half of the bovine isolates, with only two occurrences of one of the ABC subunits in the human isolates (Fig. 3). ABC transporter proteins have been shown to be important for the virulence of several *Streptococcus* species including *S. agalactiae* (Basavanna et al., 2009; Jones et al., 2000; Mei et al., 1997), and LGT of virulence factors via prophages is an important evolutionary mechanism in many bacterial species including *Streptococcus* (Boyd and Brüssow, 2002; Brenciani et al., 2010; Canchaya et al., 2003). The 38,325 bp *attR* end of prophage B aligned well with the 38,660 bp *attR* end of prophage A. All the prophage structural genes were contained within this section and the sequences shared 84.6% sequence identity. We were not able to locate the 13 bp motif for the *attL* site corresponding to prophage A, and consequently could not determine the precise starting point for this prophage. It is possible the *attL* site may have been located within the unsequenced gap between contigs 6 and 7. This putatively incomplete *attL* end of prophage A contained 10 intact genes, eight of which were bovine strain specific. Two were annotated as chromosome segregation ATPase and an IS, the remainder were annotated as hypothetical proteins. The structural

region of prophage A contained 11 bovine strain-specific genes. These were annotated as hypothetical proteins or were associated with prophage function.

3.8. Genomic island VI

The region encompassing genomic island VI contained a number of genes characteristic of an ICE: a site-specific recombinase, relaxase, and six conjugative genes (Fig. 3). However, we were not able to locate putative *attL* and *attR* sites, which could be due, in part, to the high frequency of IS within this region. There were 12 IS with 10 occurring in tandem, suggesting this region may have experienced considerable rearrangement. Genomic island VI contained 66 CDS. Of these, 31 were bovine strain FSL S3-026 specific, including 12 annotated as hypothetical proteins and one IS. Six of the 18 remaining strain-specific genes were characteristic of functional ICEs. Of the remaining 12 genes, five had annotations associated with cell membrane processes (e.g. binding and transport), a class of proteins implicated in *S. agalactiae* virulence (Lindahl et al., 2005). PCR screening results showed the strongest bovine bias for a single-strand binding family protein, cobalt transport protein, and a putative drug/sodium antiporter (Fig. 3).

3.9. Genomic island V (salivaricin)

Genomic island V contained 21 bovine strain FSL S3-026 specific genes of which nine were annotated as hypothetical proteins and three as ABC transporter genes. However, the most notable feature of this island was an operon involved in the production and immunity to the lantibiotic salivaricin A. First reported in *S. salivarius*, the complete operon (SalA) contains seven genes: *salAMTXYKR* (Ross et al., 1993; Wescombe et al., 2006). The operon in the bovine FSL S3-026 genome lacked the structural *salA* gene and the first 124 bp at the 5′ end of the *salM* gene. This latter aspect was most likely due to the insertion of an IS located immediately upstream of the truncated *salM* locus and perhaps renders the SalA operon in the FSL S3-026 genome non-functional. However, the deletions to the SalA operon seen in the FSL S3-026 genome may not be typical of *S. agalactiae*. Wescombe et al. (2006) detected two strains of *S. agalactiae* that could produce salivaricin. The ability to produce salivaricin, however, may be restricted to the bovine population as no human GBS strains, in our screening, contained the operon (three individuals each possessed a single gene: either *salM* or *salR*). The production of salivaricin may nonetheless be rare in the bovine population, since (i) the PCR screening results revealed only two additional bovine strains that contained five of the seven SalA genes, and (ii) Wescombe et al. (2006) screened a total of 16 *S. agalactiae* strains, all isolated from bovine mastitis sources (John Tagg, personal communication), and only the two strains mentioned above could produce salivaricin.

Despite the incomplete nature of the salivaricin operon in the FSL S3-026 genome, the findings of Phelps and Neely (2007) suggest a possible alternative function. They proposed that the salivaricin operon in an M14 serotype strain of *S. pyogenes*, which possessed a mutated and non-functional *salA* gene, and consequently lacked the ability to produce salivaricin, had shifted its immunity function from salivaricin to the host immune system. No longer required for immunity to salivaricin, they demonstrated that the *salY* gene of this bacterium was now required for survival within macrophages. If a similar adaptation has occurred for the bovine FSL S3-026 strain, it might be able to evade the bovine immune system and persist for an extended period of time within the udder, which is consistent with the persistent nature of bovine intramammary *S. agalactiae* infections (Li et al., 2008). Similarly, Li et al. (2008) showed alternative function for salivaricin genes, demonstrating that the *salK*/*salR* genes were required for full virulence in *Streptococcus suis*. In this case the genes were not contained within a salivaricin operon and were surmised to be functioning as a two-component signal transduction system controlling virulence by regulating distant gene expression.

The presence of the salivaricin operon as a bovine FSL S3-026 strain-specific genomic island suggests LGT. Like the *fru* operon, the salivaricin operon was flanked by IS and may have been relocated from a MGE elsewhere. Suggesting other instances of LGT involving these loci, the salivaricin genes *salK*/*salR* from *S. suis* are carried on an ICE (Holden et al., 2009). The closest BLASTn matches to the FSL S3-026 salivaricin operon (9608 bp) included three *S. pyogenes* strains (closest match: coverage = 100%, identity = 98.3%) followed by two *S. salivarius* strains (closest match: coverage = 100%, identity = 94.4%). Therefore, despite the strong differentiation detected between bovine and human strains of *S. agalactiae*, these findings suggest the possibility of LGT between the bovine FSL S3-026 strain of *S. agalactiae* and another *Streptococcus* species from a human host. At the same time, it is also possible that this LGT history involves *S. salivarius* within the bovine environment, as evidenced by occasional isolation of *S. salivarius* from bovine bulk tank milk (Zadoks et al., 2004) and clinical bovine mastitis (Nam et al., 2009).

3.10. Genomic islands II and IV (repeat regions)

Contigs 2 and 8 contained a 49,517 bp exact repeat (Fig. 1). With the exception of 270 bp obtained via Sanger sequencing, the repeat sequence was represented by a single 454 sequencing scaffold that had an average coverage of 91-fold, whereas the remainder of the genome had an average coverage of 22-fold, suggesting that the sequence had been repeated four times. Positioning of the repeat scaffold within the genome required supplementary Sanger sequencing, and we could unambiguously position the repeat within contigs 2 and 8. The positioning of the remaining two repeats between contigs 2 and 3 and contigs 5 and 6 (Fig. 1) remains tentative due to conflicting PCR and sequencing results regarding presence or absence of the repeat.

Interestingly, the human *S. agalactiae* strain NEM316 also contained a direct repeat of similar size to the one in the bovine strain. Designated pNEM316-1, the NEM316 repeat was 47,068 bp and repeated three times (Glaser et al., 2002). Overall, the bovine GBS and NEM316 repeats were moderately related (65.2%). However, they showed a mosaic pattern of similarity, with numerous genes sharing much higher sequence identity (average identity: 76.9%); eight of these shared genes were similar to proteins involved in plasmid replication and transfer. One of these genes was the putative integrase identified by Glaser et al. (2002). The terminal 24 bp at each end of the bovine strain repeat (designated FSL S3-026-S20) contained imperfect inverted repeats that were virtually identical to the inverted repeats at the ends of repeat pNEM316-1 (the regions differed by one base pair) (Fig. S5 supplementary material). In addition, immediately exterior to pNEM316-1 and FSL S3-026-S20 there were direct repeat motifs (putative *attL*/*R* sites) of either 9 or 10 bp for pNEM316-1 and 9 bp for FSL S3-026-S20. For each insertion of a repeat sequence, there was a unique 9/10 bp repeat motif. Glaser et al. (2002) showed that pNEM316-1 could be present in a circular form, and proposed that this sequence was a type of integrative plasmid. Given the similarity between pNEM316-1 and FSL S3-026-S20, it is likely that FSL S3-026-S20 is a similar type of genetic element. The possibility that the repeat could be located between contigs 2 and 3 and contigs 5 and 6 is supported by the presence of 8 bp direct repeats located at the terminal region of the contigs flanking the two gaps. FSL S3-026-S20 contained a total of 51 genes and virtually all (47)

were annotated as hypothetical proteins. A total of 25 genes were bovine FSL S3-026 strain specific, with four having an annotation other than hypothetical protein (Fig. 3). Two of these four proteins contained the LPXTG anchoring motif. Proteins with this motif are usually covalently linked to the cell surface and are implicated in pathogen survival within the host (Navarre and Schneewind, 1999; Schneewind et al., 1995). In the case of mastitis caused by *S. uberis*, experimental challenge studies have demonstrated a role for sortase and sortase anchored proteins in virulence, most of which carry the LPXTG anchoring motive (Leigh et al., 2010). Consequently, each of the four copies of FSL S3-026-S20 may contribute to virulence of the bovine *S. agalactiae* strain, and given the nature of the repeat sequence, these potential virulence factors may have been exchanged with other bacteria via LGT. PCR screening of these four genes indicates they occur in fairly low frequency and are largely restricted to the bovine population (Fig. 3).

4. Conclusion

Our findings corroborate, at a genomic level, previous studies showing that strains of *S. agalactiae* isolated from cows diagnosed with bovine mastitis are distinct from human disease isolates. A major factor responsible for this distinctiveness appears to be LGT, and here we report multiple examples involving a number of different mobile genetic elements, including ICE, IS, phages, and plasmids for the bovine mastitis isolate FSL S3-026. Importantly, many of the examples of LGT involve the exchange of potential virulence factors. Furthermore, we provide strong evidence for two independent cases of LGT between bovine *S. agalactiae* and two other *Streptococcus* species within a shared bovine environment: (i) *S. uberis* and (ii) *S. dysgalactiae* subsp. *dysgalactiae*. This exchange of genetic material between separate mastitis-causing species may have aided the continued adaptation of *S. agalactiae* to the bovine environment and incorporated potential virulence factors into the genome. Although bovine mastitis caused by *S. agalactiae* is currently treated effectively with penicillin (Keefe, 1997), the prevalence of genetic exchange between *S. agalactiae* and other mastitis-causing pathogens reported here is cause for concern. Specifically, other mastitis-causing pathogens are less responsive to antibiotics (Wilson et al., 1999), and evidence is accumulating that LGT is an important mechanism contributing to increased antibiotic resistance. For example, *S. pneumoniae* and *Streptococcus dysgalactiae* subsp. *equisimilis* have tetracycline resistance genes associated with IS and ICE (Doherty et al., 2000; Liu et al., 2008). Consequently, in addition to possibly facilitating adaptation to the bovine host, LGT between bovine *S. agalactiae* and other pathogens within this environment may result in the acquisition of increased antibiotic resistance. As with numerous other bacterial species, LGT has clearly played a significant role in the recent evolution of bovine *S. agalactiae*. Our findings highlight the importance of this type of genetic exchange among pathogens within a shared environment.

Acknowledgements

We are grateful to Haruo Suzuki for his assistance with the analysis and preparation of the supplementary figure on JCVI and COG categories and the *Streptococcus* genome characterization table, and also Barbara Spellerberg and Martin Wiedmann for providing samples. This work was supported by the National Institute of Allergy and Infectious Disease, US National Institutes of Health, under grant number AI073368-01A2, and by USDA grant 2006-35204-17422, both awarded to MJS.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2011.04.019.

References

Alpert, C.A., Siebers, U., 1997. The lac operon of *Lactobacillus casei* contains *lacT*, a gene coding for a protein of the Bg1G family of transcriptional antiterminators. J. Bacteriol. 179, 1555–1562.

Baker, C., 2000. Group B streptococcal infections. In: Stevens, D., Kaplan, E. (Eds.), Streptococcal Infections. Clinical Aspects, Microbiology, and Molecular Pathogenesis. Oxford Univeristy Press, New York, pp. 222–237.

Balter, S., Whitney, C., Schuchat, A., 2000. Epidemiology of group B streptococcal infections. In: Fischetti, V., Novick, R., Ferreti, J., Portnoy, D., Rood, J. (Eds.), Gram-Positive Pathogens. ASM Press, Washington, D.C..

Basavanna, S., Khandavilli, S., Yuste, J., Cohen, J.M., Hosie, A.H., Webb, A.J., Thomas, G.H., Brown, J.S., 2009. Screening of *Streptococcus pneumoniae* ABC transporter mutants demonstrates that LivJHMGF, a branched-chain amino acid ABC transporter, is necessary for disease pathogenesis. Infect. Immun. 77, 3412–3423.

Benjamini, Y., Hochberg, Y., 1995. Controlling the false discovery rate: a practical and powerful approach to multiple testing. J. R. Stat. Soc. Ser. B Stat. Methodol. 57, 289–300.

Beres, S.B., Musser, J.M., 2007. Contribution of exogenous genetic elements to the group A *Streptococcus* metagenome. PLoS One 2, e800.

Bisharat, N., Crook, D.W., Leigh, J., Harding, R.M., Ward, P.N., Coffey, T.J., Maiden, M.C., Peto, T., Jones, N., 2004. Hyperinvasive neonatal group B *Streptococcus* has arisen from a bovine ancestor. J. Clin. Microbiol. 42, 2161–2167.

Bishop, E.J., Shilton, C., Benedict, S., Kong, F., Gilbert, G.L., Gal, D., Godoy, D., Spratt, B.G., Currie, B.J., 2007. Necrotizing fasciitis in captive juvenile *Crocodylus porosus* caused by *Streptococcus agalactiae*: an outbreak and review of the animal and human literature. Epidemiol. Infect. 135, 1248–1255.

Bohnsack, J.F., Whiting, A.A., Martinez, G., Jones, N., Adderson, E.E., Detrick, S., Blaschke-Bonkowsky, A.J., Bisharat, N., Gottschalk, M., 2004. Serotype III *Streptococcus agalactiae* from bovine milk and human neonatal infections. Emerg. Infect. Dis. 10, 1412–1419.

Bolotin, A., Quinquis, B., Renault, P., Sorokin, A., Ehrlich, S.D., Kulakauskas, S., Lapidus, A., Goltsman, E., Mazur, M., Pusch, G.D., Fonstein, M., Overbeek, R., Kyprides, N., Purnelle, B., Prozzi, D., Ngui, K., Masuy, D., Hancy, F., Burteau, S., Boutry, M., Delcour, J., Goffeau, A., Hols, P., 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. Nat. Biotechnol. 22, 1554–1558.

Boyd, E.F., Brüssow, H., 2002. Common themes among bacteriophage-encoded virulence factors and diversity among the bacteriophages involved. Trends Microbiol. 10, 521–529.

Bradley, A., 2002. Bovine mastitis: an evolving disease. Vet. J. 164, 116–128.

Brenciani, A., Bacciaglia, A., Vignaroli, C., Pugnaloni, A., Varaldo, P.E., Giovanetti, E., 2010. Phim46.1, the main *Streptococcus pyogenes* element carrying *mef*(A) and *tet*(O) genes. Antimicrob. Agents Chemother. 54, 221–229.

Brochet, M., Couve, E., Glaser, P., Guedon, G., Payot, S., 2008. Integrative conjugative elements and related elements are major contributors to the genome diversity of *Streptococcus agalactiae*. J. Bacteriol. 190, 6913–6917.

Brochet, M., Couve, E., Zouine, M., Vallaeys, T., Rusniok, C., Lamy, M.C., Buchrieser, C., Trieu-Cuot, P., Kunst, F., Poyart, C., Glaser, P., 2006. Genomic diversity and evolution within the species *Streptococcus agalactiae*. Microbes Infect. 8, 1227–1243.

Brochet, M., Rusniok, C., Couve, E., Dramsi, S., Poyart, C., Trieu-Cuot, P., Kunst, F., Glaser, P., 2008. Shaping a bacterial genome by large chromosomal replacements, the evolutionary history of *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. U.S.A. 105, 15961–15966.

Burrus, V., Pavlovic, G., Decaris, B., Guedon, G., 2002. Conjugative transposons: the tip of the iceberg. Mol. Microbiol. 46, 601–610.

Burrus, V., Waldor, M.K., 2004. Shaping bacterial genomes with integrative and conjugative elements. Res. Microbiol. 155, 376–386.

Calvinho, L.F., Almeida, R.A., Oliver, S.P., 1998. Potential virulence factors of *Streptococcus dysgalactiae* associated with bovine mastitis. Vet. Microbiol. 61, 93–110.

Canchaya, C., Fournous, G., Brussow, H., 2004. The impact of prophages on bacterial chromosomes. Mol. Microbiol. 53, 9–18.

Canchaya, C., Proux, C., Fournous, G., Bruttin, A., Brussow, H., 2003. Prophage genomics. Microbiol. Mol. Biol. Rev. 67, 238–276.

Davies, M.R., McMillan, D.J., Van Domselaar, G.H., Jones, M.K., Sriprakash, K.S., 2007. Phage 3396 from a *Streptococcus dysgalactiae* subsp. *equisimilis* pathovar may have its origins in *Streptococcus pyogenes*. J. Bacteriol. 189, 2646–2652.

Davies, M.R., Shera, J., Van Domselaar, G.H., Sriprakash, K.S., McMillan, D.J., 2009. A novel integrative conjugative element mediates genetic transfer from group G *Streptococcus* to other {beta}-hemolytic Streptococci. J. Bacteriol. 191, 2257–2265.

Davies, M.R., Tran, T.N., McMillan, D.J., Gardiner, D.L., Currie, B.J., Sriprakash, K.S., 2005. Inter-species genetic movement may blur the epidemiology of streptococcal diseases in endemic regions. Microbes Infect. 7, 1128–1138.

Delorme, C., Poyart, C., Ehrlich, S.D., Renault, P., 2007. Extent of horizontal gene transfer in evolution of Streptococci of the salivarius group. J. Bacteriol. 189, 1330–1341.

Dermer, P., Lee, C., Eggert, J., Few, B., 2004. A history of neonatal group B streptococcus with its related morbidity and mortality rates in the United States. J. Pediatr. Nurs. 19, 357–363.

Dogan, B., Schukken, Y.H., Santisteban, C., Boor, K.J., 2005. Distribution of serotypes and antimicrobial resistance genes among *Streptococcus agalactiae* isolates from bovine and human hosts. J. Clin. Microbiol. 43, 5899–5906.

Doherty, N., Trzcinski, K., Pickerill, P., Zawadzki, P., Dowson, C.G., 2000. Genetic diversity of the *tet*(M) gene in tetracycline-resistant clonal lineages of *Streptococcus pneumoniae*. Antimicrob. Agents Chemother. 44, 2979–2984.

Domelier, A.S., van der Mee-Marquet, N., Sizaret, P.Y., Hery-Arnaud, G., Lartigue, M.F., Mereghetti, L., Quentin, R., 2009. Molecular characterization and lytic activities of *Streptococcus agalactiae* bacteriophages and determination of lysogenic-strain features. J. Bacteriol. 191, 4776–4785.

Drummond, A., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Heled, J., Kearse, M., Moir, R., Stones-Havas, S., 2010. Geneious v5.1. Available from http://www.geneious.com.

Farley, M., 2001. Group B streptococcal disease in nonpregnant adults. Clin. Infect. Dis. 33, 556–561.

Ferretti, J.J., McShan, W.M., Ajdic, D., Savic, D.J., Savic, G., Lyon, K., Primeaux, C., Sezate, S., Suvorov, A.N., Kenton, S., Lai, H.S., Lin, S.P., Qian, Y., Jia, H.G., Najar, F.Z., Ren, Q., Zhu, H., Song, L., White, J., Yuan, X., Clifton, S.W., Roe, B.A., McLaughlin, R., 2001. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. Proc. Natl. Acad. Sci. U.S.A. 98, 4658–4663.

Finch, L.A., Martin, D.R., 1984. Human and bovine group B streptococci: two distinct populations. J. Appl. Bacteriol. 57, 273–278.

Glaser, P., Rusniok, C., Buchrieser, C., Chevalier, F., Frangeul, L., Msadek, T., Zouine, M., Couve, E., Lalioui, L., Poyart, C., Trieu-Cuot, P., Kunst, F., 2002. Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. Mol. Microbiol. 45, 1499–1513.

Granlund, M., Oberg, L., Sellin, M., Norgren, M., 1998. Identification of a novel insertion element, IS1548, in group B streptococci, predominantly in strains causing endocarditis. J. Infect. Dis. 177, 967–976.

Haenni, M., Saras, E., Bertin, S., Leblond, P., Madec, J.Y., Payot, S., 2010. Diversity and mobility of integrative and conjugative elements in bovine isolates of *Streptococcus agalactiae*, *S. dysgalactiae* subsp. dysgalactiae, and *S. uberis*. Appl. Environ. Microbiol. 76, 7957–7965.

Hansen, S.M., Uldbjerg, N., Kilian, M., Sorensen, U.B., 2004. Dynamics of *Streptococcus agalactiae* colonization in women during and after pregnancy and in their infants. J. Clin. Microbiol. 42, 83–89.

Henning, K., Hall, E., Dwyer, D., Billmann, L., Schuchat, A., Johnson, J., Harrison, L., 2001. Invasive group B streptococcal disease in Maryland nursing home residents. J. Infect. Dis. 183, 1138–1142.

Holden, M.T., Feil, E.J., Lindsay, J.A., Peacock, S.J., Day, N.P., Enright, M.C., Foster, T.J., Moore, C.E., Hurst, L., Atkin, R., Barron, A., Bason, N., Bentley, S.D., Chillingworth, C., Chillingworth, T., Churcher, C., Clark, L., Corton, C., Cronin, A., Doggett, J., Dowd, L., Feltwell, T., Hance, Z., Harris, B., Hauser, H., Holroyd, S., Jagels, K., James, K.D., Lennard, N., Line, A., Mayes, R., Moule, S., Mungall, K., Ormond, D., Quail, M.A., Rabbinowitsch, E., Rutherford, K., Sanders, M., Sharp, S., Simmonds, M., Stevens, K., Whitehead, S., Barrell, B.G., Spratt, B.G., Parkhill, J., 2004. Complete genomes of two clinical *Staphylococcus aureus* strains: evidence for the rapid evolution of virulence and drug resistance. Proc. Natl. Acad. Sci. U.S.A. 101, 9786–9791.

Holden, M.T., Hauser, H., Sanders, M., Ngo, T.H., Cherevach, I., Cronin, A., Goodhead, I., Mungall, K., Quail, M.A., Price, C., Rabbinowitsch, E., Sharp, S., Croucher, N.J., Chieu, T.B., Mai, N.T., Diep, T.S., Chinh, N.T., Kehoe, M., Leigh, J.A., Ward, P.N., Dowson, C.G., Whatmore, A.M., Chanter, N., Iversen, P., Gottschalk, M., Slater, J.D., Smith, H.E., Spratt, B.G., Xu, J., Ye, C., Bentley, S., Barrell, B.G., Schultsz, C., Maskell, D.J., Parkhill, J., 2009. Rapid evolution of virulence and drug resistance in the emerging zoonotic pathogen *Streptococcus suis*. PLoS One 4, e6072.

Jones, A.L., Knoll, K.M., Rubens, C.E., 2000. Identification of *Streptococcus agalactiae* virulence genes in the neonatal rat sepsis model using signature-tagged mutagenesis. Mol. Microbiol. 37, 1444–1455.

Kanaya, S., Yamada, Y., Kudo, Y., Ikemura, T., 1999. Studies of codon usage and tRNA genes of 18 unicellular organisms and quantification of *Bacillus subtilis* tRNAs: gene expression level and species-specific diversity of codon usage based on multivariate analysis. Gene 238, 143–155.

Katholm, J., Rattenborg, E., 2010. The surviellance program of *Streptococcus agalactiae* in Danish dairy herds 1989–2008. In: Hillerton, J. (Ed.), Proceedings of the 5th IDF Mastitis Conference: Mastitis Research into Practice. New Zealand Veterinary Association, Vet Learn, Christchurch, New Zealand, pp. 241–246.

Kawamura, Y., Hou, X.G., Sultana, F., Miura, H., Ezaki, T., 1995. Determination of 16S rRNA sequences of *Streptococcus mitis* and *Streptococcus gordonii* and phylogenetic relationships among members of the genus *Streptococcus*. Int. J. Syst. Bacteriol. 45, 406–408.

Keefe, G.P., 1997. *Streptococcus agalactiae* mastitis: a review. Can. Vet. J. 38, 429–437.

Leigh, J.A., Egan, S.A., Ward, P.N., Field, T.R., Coffey, T.J., 2010. Sortase anchored proteins of *Streptococcus uberis* play major roles in the pathogenesis of bovine mastitis in dairy cattle. Vet. Res. 41, 63.

Li, L., Stoeckert Jr., C.J., Roos, D.S., 2003. OrthoMCL: identification of ortholog groups for eukaryotic genomes. Genome Res. 13, 2178–2189.

Li, M., Wang, C., Feng, Y., Pan, X., Cheng, G., Wang, J., Ge, J., Zheng, F., Cao, M., Dong, Y., Liu, D., Lin, Y., Du, H., Gao, G.F., Wang, X., Hu, F., Tang, J., 2008. SalK/SalR, a two-component signal transduction system, is essential for full virulence of highly invasive *Streptococcus suis* serotype 2. PLoS One 3, e2080.

Lindahl, G., Stalhammar-Carlemalm, M., Areschoug, T., 2005. Surface proteins of *Streptococcus agalactiae* and related proteins in other bacterial pathogens. Clin. Microbiol. Rev. 18, 102–127.

Liu, L.C., Tsai, J.C., Hsueh, P.R., Tseng, S.P., Hung, W.C., Chen, H.J., Teng, L.J., 2008. Identification of *tet*(S) gene area in tetracycline-resistant *Streptococcus dysgalactiae* subsp. *equisimilis* clinical isolates. J. Antimicrob. Chemother. 61, 453–455.

Loughman, J.A., Caparon, M.G., 2006. A novel adaptation of aldolase regulates virulence in *Streptococcus pyogenes*. EMBO J. 25, 5414–5422.

Loughman, J.A., Caparon, M.G., 2007. Comparative functional analysis of the lac operons in *Streptococcus pyogenes*. Mol. Microbiol. 64, 269–280.

Mahillon, J., Chandler, M., 1998. Insertion sequences. Microbiol. Mol. Biol. Rev. 62, 725–774.

Manning, S.D., Springman, A.C., Million, A.D., Milton, N.R., McNamara, S.E., Somsel, P.A., Bartlett, P., Davies, H.D., 2010. Association of group B *Streptococcus* colonization and bovine exposure: a prospective cross-sectional cohort study. PLoS One 5, e8795.

Martinez, G., Harel, J., Higgins, R., Lacouture, S., Daignault, D., Gottschalk, M., 2000. Characterization of *Streptococcus agalactiae* isolates of bovine and human origin by randomly amplified polymorphic DNA analysis. J. Clin. Microbiol. 38, 71–78.

Mei, J.M., Nourbakhsh, F., Ford, C.W., Holden, D.W., 1997. Identification of *Staphylococcus aureus* virulence genes in a murine model of bacteraemia using signature-tagged mutagenesis. Mol. Microbiol. 26, 399–407.

Nam, H., Lim, S., Kang, H., Kim, J., Moon, J., Jang, K., Joo, Y., Kang, M., Jung, S., 2009. Antimicrobial resistance of streptococci isolated from mastitic bovine milk samples in Korea. J. Vet. Diagn. Invest. 21, 698–701.

National Mastitis Council, 1996. Current Concepts in Bovine Mastitis, Madison, WI.

Navarre, W.W., Schneewind, O., 1999. Surface proteins of gram-positive bacteria and mechanisms of their targeting to the cell wall envelope. Microbiol. Mol. Biol. Rev. 63, 174–229.

Parkhill, J., Sebaihia, M., Preston, A., Murphy, L.D., Thomson, N., Harris, D.E., Holden, M.T., Churcher, C.M., Bentley, S.D., Mungall, K.L., Cerdeno-Tarraga, A.M., Temple, L., James, K., Harris, B., Quail, M.A., Achtman, M., Atkin, R., Baker, S., Basham, D., Bason, N., Cherevach, I., Chillingworth, T., Collins, M., Cronin, A., Davis, P., Doggett, J., Feltwell, T., Goble, A., Hamlin, N., Hauser, H., Holroyd, S., Jagels, K., Leather, S., Moule, S., Norberczak, H., O'Neil, S., Ormond, D., Price, C., Rabbinowitsch, E., Rutter, S., Sanders, M., Saunders, D., Seeger, K., Sharp, S., Simmonds, M., Skelton, J., Squares, R., Squares, S., Stevens, K., Unwin, L., Whitehead, S., Barrell, B.G., Maskell, D.J., 2003. Comparative analysis of the genome sequences of *Bordetella pertussis*, *Bordetella parapertussis* and *Bordetella bronchiseptica*. Nat. Genet. 35, 32–40.

Parkhill, J., Wren, B.W., Thomson, N.R., Titball, R.W., Holden, M.T., Prentice, M.B., Sebaihia, M., James, K.D., Churcher, C., Mungall, K.L., Baker, S., Basham, D., Bentley, S.D., Brooks, K., Cerdeno-Tarraga, A.M., Chillingworth, T., Cronin, A., Davies, R.M., Davis, P., Dougan, G., Feltwell, T., Hamlin, N., Holroyd, S., Jagels, K., Karlyshev, A.V., Leather, S., Moule, S., Oyston, P.C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens, K., Whitehead, S., Barrell, B.G., 2001. Genome sequence of *Yersinia pestis*, the causative agent of plague. Nature 413, 523–527.

Pattison, I., Matthews, P., Howell, D., 1955. The type classification of group-B streptococci, with special reference to bovine strains apparently lacking in type polysaccharide. J. Pathol. Bacteriol. 69, 51–60.

Phelps, H.A., Neely, M.N., 2007. SalY of the *Streptococcus pyogenes* lantibiotic locus is required for full virulence and intracellular survival in macrophages. Infect. Immun. 75, 4541–4551.

Pryor, S.M., Cursons, R.T., Williamson, J.H., Lacy-Hulbert, S.J., 2009. Experimentally induced intramammary infection with multiple strains of *Streptococcus uberis*. J. Dairy Sci. 92, 5467–5475.

Rauch, P.J., De Vos, W.M., 1992. Characterization of the novel nisin-sucrose conjugative transposon Tn5276 and its insertion in *Lactococcus lactis*. J. Bacteriol. 174, 1280–1287.

Ross, K.F., Ronson, C.W., Tagg, J.R., 1993. Isolation and characterization of the lantibiotic salivaricin A and its structural gene *salA* from *Streptococcus salivarius* 20P3. Appl. Environ. Microbiol. 59, 2014–2021.

Salloum, M., van der Mee-Marquet, N., Domelier, A.S., Arnault, L., Quentin, R., 2010. Molecular characterization and prophage DNA contents of *Streptococcus agalactiae* strains isolated from adult skin and osteoarticular infections. J. Clin. Microbiol. 48, 1261–1269.

Schneewind, O., Fowler, A., Faull, K.F., 1995. Structure of the cell wall anchor of surface proteins in *Staphylococcus aureus*. Science 268, 103–106.

Schrag, S., Zywicki, S., Farley, M., Reingold, A., Harrison, L., Lefkowitz, L., Hadler, J., Danila, R., Cieslak, P., Schuchat, A., 2000. Group B Streptococcal disease in the era of Intrapartum antibiotic prophylaxis. N. Engl. J. Med. 342, 15–20.

Schuchat, A., 1998. Epidemiology of group B streptococcal disease in the United States: shifting paradigms. Clin. Microbiol. Rev. 11, 497–513.

Sharp, P.M., Bailes, E., Grocock, R.J., Peden, J.F., Sockett, R.E., 2005. Variation in the strength of selected codon usage bias among bacteria. Nucleic Acids Res. 33, 1141–1153.

Siezen, R.J., Renckens, B., van Swam, I., Peters, S., van Kranenburg, R., Kleerebezem, M., de Vos, W.M., 2005. Complete sequences of four plasmids of *Lactococcus lactis* subsp. *cremoris* SK11 reveal extensive adaptation to the dairy environment. Appl. Environ. Microbiol. 71, 8371–8382.

Siguier, P., Filee, J., Chandler, M., 2006. Insertion sequences in prokaryotic genomes. Curr. Opin. Microbiol. 9, 526–531.

Soong, G., Muir, A., Gomez, M.I., Waks, J., Reddy, B., Planet, P., Singh, P.K., Kaneko, Y., Wolfgang, M.C., Hsiao, Y.S., Tong, L., Prince, A., 2006. Bacterial neuraminidase facilitates mucosal infection by participating in biofilm production. J. Clin. Invest. 116, 2297–2305.

Sørensen, U.B., Poulsen, K., Ghezzo, C., Margarit, I., Kilian, M., 2010. Emergence and global dissemination of host-specific *Streptococcus agalactiae* clones. mBio 1, 1–9.

Spellerberg, B., Martin, S., Franken, C., Berner, R., Lutticken, R., 2000. Identification of a novel insertion sequence element in *Streptococcus agalactiae*. Gene 241, 51–56.

Stothard, P., Wishart, D.S., 2005. Circular genome visualization and exploration using CGView. Bioinformatics 21, 537–539.

Sukhnanand, S., Dogan, B., Ayodele, M.O., Zadoks, R.N., Craver, M.P., Dumas, N.B., Schukken, Y.H., Boor, K.J., Wiedmann, M., 2005. Molecular subtyping and characterization of bovine and human *Streptococcus agalactiae* isolates. J. Clin. Microbiol. 43, 1177–1186.

Tagg, J.R., Bannister, L.V., 1979. ''Fingerprinting'' beta-haemolytic streptococci by their production of and sensitivity to bacteriocine-like inhibitors. J. Med. Microbiol. 12, 397–411.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Donati, C., Medini, D., Ward, N.L., Angiuoli, S.V., Crabtree, J., Jones, A.L., Durkin, A.S., Deboy, R.T., Davidsen, T.M., Mora, M., Scarselli, M., Margarit y Ros, I., Peterson, J.D., Hauser, C.R., Sundaram, J.P., Nelson, W.C., Madupu, R., Brinkac, L.M., Dodson, R.J., Rosovitz, M.J., Sullivan, S.A., Daugherty, S.C., Haft, D.H., Selengut, J., Gwinn, M.L., Zhou, L., Zafar, N., Khouri, H., Radune, D., Dimitrov, G., Watkins, K., O'Connor, K.J., Smith, S., Utterback, T.R., White, O., Rubens, C.E., Grandi, G., Madoff, L.C., Kasper, D.L., Telford, J.L., Wessels, M.R., Rappuoli, R., Fraser, C.M., 2005. Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial ''pan-genome''. Proc. Natl. Acad. Sci. U.S.A. 102, 13950–13955.

Tettelin, H., Masignani, V., Cieslewicz, M.J., Eisen, J.A., Peterson, S., Wessels, M.R., Paulsen, I.T., Nelson, K.E., Margarit, I., Read, T.D., Madoff, L.C., Wolf, A.M., Beanan, M.J., Brinkac, L.M., Daugherty, S.C., DeBoy, R.T., Durkin, A.S., Kolonay, J.F., Madupu, R., Lewis, M.R., Radune, D., Fedorova, N.B., Scanlan, D., Khouri, H., Mulligan, S., Carty, H.A., Cline, R.T., Van Aken, S.E., Gill, J., Scarselli, M., Mora, M., Iacobini, E.T., Brettoni, C., Galli, G., Mariani, M., Vegni, F., Maione, D., Rinaudo, D., Rappuoli, R., Telford, J.L., Kasper, D.L., Grandi, G., Fraser, C.M., 2002. Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. Proc. Natl. Acad. Sci. U.S.A. 99, 12391–12396.

Tong, H.H., Blue, L.E., James, M.A., DeMaria, T.F., 2000. Evaluation of the virulence of a *Streptococcus pneumoniae* neuraminidase-deficient mutant in nasopharyngeal colonization and development of otitis media in the chinchilla model. Infect. Immun. 68, 921–924.

Wanger, A.R., Dunny, G.M., 1987. Identification of a *Streptococcus agalactiae* protein antigen associated with bovine mastitis isolates. Infect. Immun. 55, 1170–1175.

Ward, P.N., Holden, M.T., Leigh, J.A., Lennard, N., Bignell, A., Barron, A., Clark, L., Quail, M.A., Woodward, J., Barrell, B.G., Egan, S.A., Field, T.R., Maskell, D., Kehoe, M., Dowson, C.G., Chanter, N., Whatmore, A.M., Bentley, S.D., Parkhill, J., 2009. Evidence for niche adaptation in the genome of the bovine pathogen *Streptococcus uberis*. BMC Genomics 10, 54.

Wen, Z.T., Browngardt, C., Burne, R.A., 2001. Characterization of two operons that encode components of fructose-specific enzyme II of the sugar:phosphotransferase system of *Streptococcus mutans*. FEMS Microbiol. Lett. 205, 337–342.

Wescombe, P.A., Upton, M., Dierksen, K.P., Ragland, N.L., Sivabalan, S., Wirawan, R.E., Inglis, M.A., Moore, C.J., Walker, G.V., Chilcott, C.N., Jenkinson, H.F., Tagg, J.R., 2006. Production of the lantibiotic salivaricin A and its variants by oral streptococci and use of a specific induction assay to detect their presence in human saliva. Appl. Environ. Microbiol. 72, 1459–1466.

Wibawan, I.W., Lammler, C., 1990. Properties of group B streptococci with protein surface antigens X and R. J. Clin. Microbiol. 28, 2834–2836.

Wilson, D.J., Gonzalez, R.N., Case, K.L., Garrison, L.L., Grohn, Y.T., 1999. Comparison of seven antibiotic treatments with no treatment for bacteriological efficacy against bovine mastitis pathogens. J. Dairy Sci. 82, 1664–1670.

Wirawan, R.E., Klesse, N.A., Jack, R.W., Tagg, J.R., 2006. Molecular and genetic characterization of a novel nisin variant produced by *Streptococcus uberis*. Appl. Environ. Microbiol. 72, 1148–1156.

Zadoks, R., Fitzpatrick, J., 2009. Changing trends in mastitis. Ir. Vet. J. 62S, 59–70.

Zadoks, R., González, R., Boor, K.J., Schukken, Y.H., 2004. Mastitis-causing *Streptococci* are important contributors to bacterial counts in raw bulk tank milk. J. Food Prot. 67, 2644–2650.