In silico prediction of conserved vaccine targets in *Streptococcus agalactiae* strains isolated from fish, cattle, and human samples

U.P. Pereira[1,2], S.C. Soares[3], J. Blom[4], C.A.G. Leal[1,5], R.T.J. Ramos[6],
L.C. Guimarães[3], L.C. Oliveira[3], S.S. Almeida[3], S.S. Hassan[3], A.R. Santos[3],
A. Miyoshi[3], A. Silva[6], A. Tauch[4], D. Barh[7], V. Azevedo[3] and H.C.P. Figueiredo[1,5]

[1]AQUAVET - Laboratório de Doenças de Animais Aquáticos,
Departamento de Medicina Veterinária Preventiva,
Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brasil
[2]Departamento de Medicina Veterinária, Universidade Federal de Lavras,
Lavras, MG, Brasil
[3]Instituto de Ciências Biológicas, Universidade Federal de Minas Gerais,
Belo Horizonte, MG, Brasil
[4]Center for Biotechnology, Bielefeld University, Bielefeld,
Nordrhein-Westfalen, Germany
[5]AQUACEN - Laboratório Oficial Central do Ministério da Pesca e Aquicultura,
Ministério da Pesca e Aquicultura, Universidade Federal de Minas Gerais,
Belo Horizonte, MG, Brasil
[6]Instituto de Ciências Biológicas, Universidade Federal do Pará, Belém,
PA, Brasil
[7]Centre for Genomics and Applied Gene Technology,
Institute of Integrative Omics and Applied Biotechnology,
Nonakuri, Purba Medinipur, West Bengal, India

Corresponding author: H.C.P. Figueiredo
E-mail: figueiredoh@yahoo.com

ABSTRACT. *Streptococcus agalactiae* (Lancefield group B; group B streptococci) is a major pathogen that causes meningoencephalitis

in fish, mastitis in cows, and neonatal sepsis and meningitis in humans. The available prophylactic measures for conserving human and animal health are not totally effective and have limitations. Effective vaccines against the different serotypes or genotypes of pathogenic strains from the various hosts would be useful. We used an *in silico* strategy to identify conserved vaccine candidates in 15 genomes of group B streptococci strains isolated from human, bovine, and fish samples. The degree of conservation, subcellular localization, and immunogenic potential of *S. agalactiae* proteins were investigated. We identified 36 antigenic proteins that were conserved in all 15 genomes. Among these proteins, 5 and 23 were shared only by human or fish strains, respectively. These potential vaccine targets may help develop effective vaccines that will help prevent *S. agalactiae* infection.

Key words: *Streptococcus agalactiae*; Vaccine targets; Comparative genomics

INTRODUCTION

Streptococcus agalactiae (Lancefield group B; GBS) is a major bacterial pathogen that causes diseases in humans, cattle, and fish (Brochet et al., 2006; Mian et al., 2009). In humans, it is frequently associated with neonatal sepsis and meningitis, but can also affect immunocompromised adults, besides also being a common colonizer of the gastrointestinal and genitourinary tracts (Park et al., 2012). In dairy cattle, GBS is an important pathogen of clinical and subclinical mastitis, affecting milk quality and production (Richards et al., 2011). In fish, *S. agalactiae* is an emerging pathogen that causes septicemia and meningoencephalitis with high mortality in wild and cultured species worldwide (Mian et al., 2009; Chen et al., 2012).

Preventive measures are taken to prevent GBS infection in humans, such as intrapartum antibiotic prophylaxis. But this practice is only effective in early-onset disease (0 to 7 days of life), while having no impact on late-onset disease (7 to 90 days postpartum), where the infection can still occur, and it is not practical in nonpregnant adults (Edmond et al., 2012). In animal production, the extensive use of antibiotics is not permitted in many countries and may compromise the economic viability of the practice and safety of those products (Mitchell et al., 1998; Cañada-Cañada et al., 2009).

Thus, immunoprophylaxis strategy is the most viable option to reduce the damage caused by this pathogen in human health and animal production (Maione et al., 2005; Nuccitelli et al., 2011). Several types of vaccine formulations have been tested for use in humans, such as whole-cell inactivated bacteria, capsular carbohydrates, and recombinant proteins present on the bacterial surface. However, there have been relatively few clinical trials with GBS vaccines in the last years (Heath, 2011). Capsular polysaccharide (CPS)-based vaccines, developed using serotypes II and III conjugated with tetanus toxoid, have been shown to be safe and immunogenic in humans. However, this formulation has shown little or no cross-protection against other serotypes (Johri et al., 2006). Additionally, fur-

ther studies are needed to investigate vaccine efficacy when more than two GBS CPS types are simultaneously administered (Paoletti and Madoff, 2002). Another limiting factor in CPS vaccination is the significant prevalence of nontypeable CPS strains, which may compromise its efficacy (Baker and Edwards, 2003).

The use of conserved antigenic proteins has been emerging as a useful alternative in the development of vaccines against many bacterial pathogens including GBS (Maione et al., 2005; He et al., 2010; Soares et al., 2012). The bioinformatic tools currently available allow us to predict many features of all proteins present in the genome such as subcellular localization, conservation degree (when other genomes of the same species are analyzed), and antigenic potential (MHC I and II epitope predictions) (He et al., 2010).

A previous study evaluated the genome sequences of eight human strains of *S. agalactiae* to identify potential surface-exposed proteins in core and accessory genomes to be used as vaccine candidates. Only 4 of 312 candidates tested showed protection in *in vivo* experiments, and from these, only 1 belonged to the core genome (Maione et al., 2005). The *in silico* immunogenicity of these proteins was not evaluated, which may explain the low efficiency of the majority of tested targets. Currently, other GBS genomes were deposited in GenBank, including strains isolated from other hosts such as cattle and fish (Richards et al., 2011; Delannoy et al., 2012; Liu et al., 2012; Wang et al., 2012; Pereira et al., 2013). Furthermore, new tools for vaccine target prediction are available that estimate the immunogenic potential of proteins, which may improve the choice (and reduce the number) of vaccine targets to be tested (He et al., 2010). Thus, the prediction of conserved immunogenic proteins may help in finding new and efficient candidates for vaccine development.

There are few studies of vaccines against mastitis caused by *S. agalactiae*, because this is usually easy to control through hygiene and treatment measures (Duarte et al., 2004). However, a study that used the conserved Sip protein along with one *S. aureus* protein (ClfA protein) showed higher efficacy of this chimeric vaccine when compared with the inactivated vaccine of each one of these pathogens in experimental challenge in lactating mice (Xu et al., 2011). In fish, bacterial vaccines against GBS have already been tested and are commercially available. However, vaccine protection was demonstrated to be variable between strains of the same serotype and different genotypes, suggesting that the capsule is not the single antigen to elicit immune response in fish (Chen et al., 2012).

This study aimed to perform a comparative genomic analysis to identify the core genes of 15 *S. agalactiae* strains isolated from humans, cattle, and fish, and moreover, to find conserved antigenic proteins to be used in the development of vaccines against GBS infection in these hosts.

MATERIAL AND METHODS

Genome sequences

The complete and draft genome sequences of 15 *S. agalactiae* strains were retrieved from the NCBI database (http://www.ncbi.nlm.nih.gov/genbank/): 10 strains isolated from humans (three complete genomes), 1 strain isolated from cow (draft genome), and 4 strains isolated from fish (2 complete genomes) (Table 1). The human strains were

isolated in USA, Italy, Canada, United Kingdom, or unknown country. These strains were isolated from clinical cases of septicemia and meningitis or from body sites of healthy person (colonizing the oral cavity or vagina). The bovine strain was isolated from clinical mastitis in the USA, and the fish strains were isolated from meningoencephalitis outbreaks in China, Brazil, and Honduras (Table 1).

Table 1. General information about the 15 *Streptococcus agalactiae* strains used in this study.

Strain	Serotype	ST by MLST	Host	Isolate source or clinical description	Country of isolation	Number of genes	GenBank accession No.	Reference
A909	Ia	7	Human	Septicemia	USA	2136	CP00114	Tettelin et al., 2005
2603V/R	V	110	Human	Septicemia	Italy	2276	AE009948	Tettelin et al., 2002
NEM316	III	23	Human	Meningitis	Unknown	2235	AL732656	Glaser et al., 2002
18RS21	II	19	Human	Septicemia	USA	2421	AAJO01000000	Tettelin et al., 2005
515	Ia	23	Human	Septicemia (cerebrospinal fluid)	USA	2287	AAJP01000000	Tettelin et al., 2005
ATCC13813	II	61	Human	Oral cavity	United Kingdom	2243	AEQQ00000000	-
CJB111	V	1	Human	Septicemia	USA	2208	AAJQ01000000	Tettelin et al., 2005
COH1	III	17	Human	Septicemia	USA	2412	AAJR01000000	Tettelin et al., 2005
GB00112	III	17	Human	Vagina	Canada	1994	AKXO00000000	Singh et al., 2012
H36B	Ib	6	Human	Septicemia	USA	2396	AAJS01000000	Tettelin et al., 2005
FSLS3-026	III	67	Bovine	Mastitis	USA	2422	AEXT01000000	Richards et al., 2011
SA20-06	Ib	553	Fish	Meningoencephalitis (kidney)	Brazil	1872	CP003919	Pereira et al., 2013
GD201008-001	Ia	7	Fish	Meningoencephalitis	China	2088	CP003810	Liu et al., 2012
STIR-CD-17	Ib	260	Fish	Meningoencephalitis (heart)	Honduras	1737	ALXB00000000	Tettelin et al., 2005
ZQ0910	Ia	7	Fish	Meningoencephalitis	China	2003	AKAP00000000	Delannoy et al., 2012

ST = sequence type; MLST = multilocus sequence type.

Prediction of core genes

The identification of all genes of the *S. agalactiae* species and conserved genes (core genes) between the 15 genomes analyzed was performed using the EDGAR (version 1.2) software (Blom et al., 2009). EDGAR is a framework for the comparative analysis of prokaryotic genomes that performs homology analyses based on a specific cut-off that is automatically adjusted to the query dataset. Briefly, the genome sequences of *S. agalactiae* were retrieved from GenBank; the homogenization of the genome annotations was conducted by GenDB (version 2.4) (Meyer et al., 2003); an EDGAR project was created; and homology calculations based on BLAST Score Ratio Values at the protein level were performed resulting in all genes representative of the species (pan-genome), genes shared by all (or a group) strains (core genes), accessory and strain-specific genes.

Prediction of vaccine candidates

The prediction of cross-reactive vaccine candidates was performed similarly as with the criteria described by Soares et al. (2012). The potential targets had to present the following attributes: rule I) to be exposed to the host immune system, such as secreted, surface-exposed, and membrane proteins (Rappuoli, 2001); rule II) to show MHC I and II binding properties with adhesion probability greater than 0.51 and no similarity to human

proteins when analyzed by the Vaxign software (He et al., 2010); rule III) protein conservation among different genomes of *S. agalactiae* in all 15 strains or in genomes of human or fish strains (He et al., 2010; Soares et al., 2012).

The subcellular localization of proteins (rule I) was predicted by an *in silico* analysis using the SurfG+ 1.0 tool (Barinov et al., 2009), where a multi-FASTA file of the pangenome of 15 genomes of *S. agalactiae* was analyzed. SurfG+ pipeline searches protein motifs, including SignalP, LipoP, and TMHMM, which are related to subcellular localization. It also creates novel HMMSEARCH profiles to predict cell wall retention signals. Afterwards, it looks for retention signals, lipoproteins, SEC (secreted proteins) pathway export motifs, and transmembrane motifs. The proteins were characterized as cytoplasmic (CYT) if none of these motifs were found in their sequences. Moreover, SurfG+ has the ability to better distinguish between MEM (membrane proteins) and PSE (potentially surface-exposed proteins), which may assist in the choice of possible vaccine targets. Proteins could thereby be classified into four different subcellular locations: CYT, MEM, PSE, or SEC. The SurfG+ software uses cell wall thickness to better differentiate integral membrane proteins from PSE proteins. This measure was performed for the NEM316 human strain and SA20-06 fish strain by transmission electron microscopy EM10A equipment (Zeiss) according to Barinov et al. (2009). Briefly, *S. agalactiae* strains were streaked on blood agar and incubated at 30°C for 48 h. Subsequently, these strains were grown in 100 mL brain heart infusion broth at 30°C for 24 h. After centrifugation, the pellet was fixed in 2.5% glutaraldehyde in 0.1 M sodium cacodylate buffer, pH 7.2, for 18 h at 4°C. The samples were post-fixed in 1% osmium tetroxide in 0.1 M sodium cacodylate buffer, pH 7.2, + 1.5% potassium ferrocyanide for 90 min, dehydrated in a graded ethanol series (30-100%), and embedded in Eponate-Araldite resin. Ultrathin sections were obtained using uranyl acetate and lead citrate and later examined in a Zeiss-EM-10A. Microphotographs were obtained using a CCD Mega view III camera. The means of the cell wall of human strain NEM316 and fish strain SA20-06 were performed by more than 100 measurements in at least five photographs of each strain.

To apply rule II, the proteins predicted by SurfG+ as SEC, PSE, and MEM were analyzed by the Vaxign software (He et al., 2010). The software evaluates adhesion probability and protein conservation between different genomes in the OrthoMCL database, and excludes sequences of nonpathogenic strains or host-similar proteins. *S. agalactiae* proteins with high similarity to human proteins were excluded from the analysis.

As one of the aims of this study was to search for conserved vaccine candidates, the proteins predicted by the EDGAR software in the core genome of all strains or proteins conserved only in human or fish strains were considered to conform to rule III. However, the conservation of proteins (rule III) does not exclude the antigenic target, since a universal vaccine described in a previous study (Maione et al., 2005) had consisted of proteins of the accessory genome. Additionally, some antigenic proteins of the accessory genome can be good vaccine targets against more prevalent serotypes in a given specific geographic region or host.

RESULTS

The total number of genes found in the 15 genomes of *S. agalactiae* strains iso-

lated from humans, cattle, and fish was 5143. A total of 1111 genes were shared by all strains, while 1297 and 1523 genes were shared by human or fish strains, respectively. The mean cell wall measure of fish strain SA20-06 and human strain NEM316 was 18.67 and 19.12 nm, respectively. Thus, due to the high similarity of these values in both human and fish strains, an approximate value of 19 nm was used to perform the SurfG+ analysis (Figure 1). The SurfG+ analyses of 5143 predicted proteins from the pan-genome of all the 15 strains of *S. agalactiae* identified 156 SEC proteins, 398 PSE proteins, 773 MEM proteins, and 3816 CYT proteins.
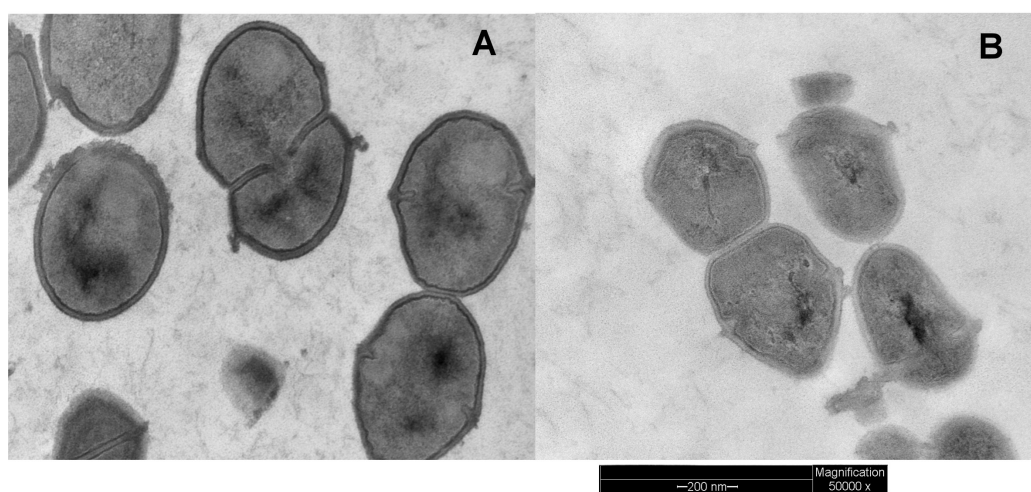

Figure 1. Transmission electron microscopy of *Streptococcus agalactiae* human strain NEM316 (**A**) and *S. agalactiae* fish strain SA20-06 (**B**).

The 1327 proteins classified as SEC, PSE, or MEM (rule I) were evaluated by the Vaxign software, and 150 showed an adhesion probability greater than 0.51 (Table S1). Eight of these 150 proteins (SAK_1493, SAI_0146, SAK_0722, SAK_0477, SAK_0477, SAK_2135, SAK_0084, and SAK_0084) were shown to be similar to mammalian proteins and were excluded from the analyses (Table S1).

Of those 142 proteins, 36 were present in the core genome of the 15 *S. agalactiae* strains fulfilling all rules of vaccine candidates proposed here. A total of 41 candidates were found to be conserved in genomes of human isolates (Table 2). Of these, 5 proteins (SAK_0050, SAK_1293, SAK_1319, SAK_066, and SAK_1074) were exclusively found in the strains of this host. Fifty-nine protein candidates were found in core genes of fish strains. Of these, 23 proteins were shared only by fish strains (Table 2).

DISCUSSION

The development of broad protective vaccines is highly needed to control GBS infections mainly in humans and fish, since prophylactic measures are frequently effective in bovine infections. Additionally, different pathogenesis mechanisms may be related to disease

Table 2. Subcellular location and adhesion probability of conserved antigenic proteins identified by reverse vaccinology strategy in core genes of all strains (upper in table), shared genes only by human (gray lines), or fish (bottom of the table) strains.

Locus tag*	Subcellular location	Core global	Core human	Core fish	Protein note	Adhesion probability	Similar human protein
SAK_2105	PSE	YES	YES	YES	Transglycosylase-like domain-containing protein	0.751	NO
SAK_0370	PSE	YES	YES	YES	Penicillin-binding protein	0.700	NO
SAK_0442	SECRETED	YES	YES	YES	Hypothetical protein	0.695	NO
SAK_0065	PSE	YES	YES	YES	Group B streptococcal surface immunogenic protein	0.678	NO
SAK_2073	PSE	YES	YES	YES	Hypothetical protein	0.666	NO
SAK_0337	SECRETED	YES	YES	YES	Hypothetical protein	0.662	NO
SAK_1271	PSE	YES	YES	YES	Hypothetical protein	0.642	NO
SAK_0064	SECRETED	YES	YES	YES	Zoocin A	0.642	NO
SAK_0553	SECRETED	YES	YES	YES	Hypothetical protein	0.638	NO
SAK_1497	PSE	YES	YES	YES	Polar amino acid ABC transporter permease/substrate-binding protein	0.617	NO
SAK_0932	PSE	YES	YES	YES	Foldase prsA	0.613	NO
SAK_1394	SECRETED	YES	YES	YES	RND family efflux transporter MFP subunit	0.607	NO
SAK_1994	SECRETED	YES	YES	YES	Hypothetical protein	0.588	NO
SAK_1009	SECRETED	YES	YES	YES	Hypothetical protein	0.584	NO
SAK_1656	PSE	YES	YES	YES	Amino acid ABC transporter amino acid-binding protein	0.575	NO
SAK_0604	SECRETED	YES	YES	YES	GDSL family lipase/acylhydrolase	0.574	NO
SAK_1158	PSE	YES	YES	YES	Hypothetical protein	0.568	NO
SAK_1087	PSE	YES	YES	YES	Phosphate ABC transporter substrate-binding protein	0.567	NO
SAK_1784	SECRETED	YES	YES	YES	CHAP domain-containing protein	0.565	NO
SAK_0556	SECRETED	YES	YES	YES	Hypothetical protein	0.565	NO
SAK_0457	PSE	YES	YES	YES	GDXG lipolytic enzyme family protein	0.561	NO
SAK_1870	SECRETED	YES	YES	YES	Mannosyl-glycoprotein endo-beta-N-acetylglucosamidase	0.556	NO
SAK_1625	PSE	YES	YES	YES	Polar amino acid ABC transporter polar amino acid-binding protein	0.556	NO
SAK_2007	SECRETED	YES	YES	YES	Hypothetical protein	0.554	NO
SAK_0854	PSE	YES	YES	YES	Hypothetical protein	0.548	NO
SAK_1235	PSE	YES	YES	YES	Hypothetical protein	0.545	NO
SAL_1416	PSE	YES	YES	YES	PTS system, fructose specific IIABC components	0.541	NO
SAK_0222	PSE	YES	YES	YES	Penicillin-binding protein 1B	0.539	NO
SAK_1109	SECRETED	YES	YES	YES	Hypothetical protein	0.539	NO
SAK_1503	PSE	YES	YES	YES	Cell wall surface anchor family protein	0.536	NO
SAK_1580	SECRETED	YES	YES	YES	Hypothetical protein	0.534	NO
SAK_1927	PSE	YES	YES	YES	Phosphate ABC transporter substrate-binding protein	0.532	NO
SAK_0321	PSE	YES	YES	YES	Lipoprotein	0.532	NO
SAK_0301	PSE	YES	YES	YES	Quaternary amine ABC transporter amino acid-binding protein	0.531	NO
SAK_0685	SECRETED	YES	YES	YES	Zinc ABC transporter substrate-binding protein	0.517	NO
SAK_1426	PSE	YES	YES	YES	Ferrichrome ABC transporter substrate-binding protein	0.512	NO
SAK_0050	SECRETED	NO	YES	NO	PcsB protein	0.748	NO
SAK_1293	SECRETED	NO	YES	NO	Hydrophobic W repeat-containing protein	0.632	NO
SAK_1319	PSE	NO	YES	NO	Laminin-binding surface protein	0.589	NO
SAK_0166	PSE	NO	YES	NO	Ribose ABC transporter ribose-binding protein	0.561	NO
SAK_1074	SECRETED	NO	YES	NO	ABC transporter substrate-binding protein	0.514	NO
SAK_1407	MEMBRANE	NO	NO	YES	Hypothetical protein	0.744	NO
SAK_0814	SECRETED	NO	NO	YES	DNA-entry nuclease	0.692	NO
SAK_2106	PSE	NO	NO	YES	LysM domain-containing protein	0.691	NO
SAK_1440	PSE	NO	NO	YES	PI-2b backbone protein	0.659	NO
SAK_0220	SECRETED	NO	NO	YES	Hypothetical protein	0.654	NO

Table 2. Continued.

Locus tag*	Subcellular location	Core global	Core human	Core fish	Protein note	Adhesion probability	Similar human protein
SAK_1983	SECRETED	NO	NO	YES	CAMP factor	0.629	NO
SAK_1437	PSE	NO	NO	YES	PI-2b sortase	0.621	NO
SAK_1441	PSE	NO	NO	YES	PI-2b ancillary protein 1	0.617	NO
SAK_1419	PSE	NO	NO	YES	Hypothetical protein	0.613	NO
SAG1127	SECRETED	NO	NO	YES	Hypothetical protein	0.610	NO
SAK_1213	SECRETED	NO	NO	YES	Hypothetical protein	0.606	NO
SAK_1313	SECRETED	NO	NO	YES	Hypothetical protein	0.598	NO
SAK_1971	MEMBRANE	NO	NO	YES	Hypothetical protein	0.595	NO
SAK_1454	PSE	NO	NO	YES	Hypothetical protein	0.571	NO
SAK_1556	PSE	NO	NO	YES	Metal ABC transporter substrate-binding protein	0.570	NO
SAK_0866	MEMBRANE	NO	NO	YES	Hypothetical protein	0.568	NO
SAK_1058	SECRETED	NO	NO	YES	Hypothetical protein	0.563	NO
SAK_0206	PSE	NO	NO	YES	Oligopeptide ABC transporter oligopeptide-binding protein	0.559	NO
SAK_0896	PSE	NO	NO	YES	Cell wall surface anchor family protein	0.558	NO
SAK_0546	MEMBRANE	NO	NO	YES	Hypothetical protein	0.534	NO
SAJ_0744	SECRETED	NO	NO	YES	Conserved hypothetical protein	0.530	NO
SAK_1655	SECRETED	NO	NO	YES	NLPA family lipoprotein	0.528	NO
SAK_0478	PSE	NO	NO	YES	BMP family protein	0.516	NO

*Locus tags of table belong to human strain A909 (SAK), human strain 2603 V/R (SAG) and human strain 18RS21 (SAJ). PSE = potentially surface-exposed proteins.

in cattle when compared with humans and fish, since the bovine disease is limited to mammary glands (Richards et al., 2011). Furthermore, the genome sequence of only one strain is not enough to understand the biology of this GBS bovine subpopulation and to perform the prediction of vaccine candidates specifically for this host. Similarly, differences in the immune response between humans and fish may also result in a variable efficacy of vaccine targets in these different hosts.

Thirty-six shared proteins in all 15 GBS genomes is a discrepant number when compared with the study of Maione et al. (2005), which identified 396 vaccine candidate proteins shared by genomes of 8 human strains. It is mainly due to the adhesion probability score of the Vaxign software, which predicts the presence of immunogenic epitopes for MHC I and II in the proteins analyzed, eliminating protein candidates with low scores. Another factor is the greater number of genomes currently available, which results in a lower number of core genes. The large number of vaccine targets to be tested is expensive and tedious. Thus, the use of *in silico* prediction of immunogenicity of proteins may provide better selection and reduce the number of vaccine targets. Among the 36 vaccine candidates found here, *in silico* prediction identified a well-characterized immunogenic protein Sip (GBS surface immunogenic protein). Penicillin-binding protein 1A and zoocin A are two other identified proteins that deserve attention, due to their high potential as vaccine candidates. In addition, the genes conserved only in human or fish strains may be evaluated as possible candidates to host specific vaccines (Table 2).

The Sip protein (SAK_0065) was the only protein of the core genome that resulted in protection in experimental challenge in the study of Maione et al. (2005). Here, this protein was also classified among the best five vaccine targets according to adhesion probability predicted by the Vaxign software. The penicillin-binding proteins have a traditional function in the biosynthesis of cell wall peptidoglycan. In addition, the penicillin-binding protein 1A (SAK_0370) has been related to resistance of GBS to phagocytosis and antimicrobial peptides (Hamilton et al., 2006). The zoocin A protein (SAK_0064) is a bacteriocin-like inhibitory

substance reported to be produced by *Streptococcus equi* ssp *zooepidemicus*. It is associated with peptidoglycan hydrolysis of sensitive *Streptococcus* species, such as *S. pyogenes* (Akesson et al., 2007).

In core genes of human *S. agalactiae* strains, the PcsB protein (SAK_0050) is a secreted protein with pivotal function in cell division and antibiotic tolerance in GBS. The inactivation of this gene results in abnormal cell division and reduction of tolerance to osmotic pressure and antibiotics (Reinscheid et al., 2001). The virulence factor laminin-binding protein (SAK_1319) was also predicted as a vaccine target in core genome of human strains. In *S. agalactiae*, this protein is related to adhesion and invasion of host tissues, and due to its 3-D structure features, was suggested as a good vaccine target (Ragunathan et al., 2009).

Currently, three distinct pilus islands (PIs) have been characterized in *S. agalactiae* and the presence of these PIs is variable in the genome of different strains. Pilus proteins have been described as having a key role in *S. agalactiae* pathogenesis (Maisey et al., 2007; Konto-Ghiorghi et al., 2009) and have been proposed as good vaccine candidates (Sharma et al., 2013). Although the pilus proteins do not belong to the core genome of all 15 strains of *S. agalactiae*, promising results in development of universal vaccines have been proposed by a combination of immunogenic pilus proteins of three PIs (Margarit et al., 2009). Additionally, the PI-2b was present in all four fish strains, suggesting that a vaccine developed by proteins of this PI may be effective against the disease in fish, mainly using the backbone protein or ancillary protein 1, which showed higher adhesion probability (Table 2). A previous study demonstrated that backbone protein and ancillary protein 1 of PI-2b conferred protection in experimental challenge (Margarit et al., 2009), and the *in silico* prediction carried out here suggests that these proteins are good vaccine candidates (Table 2). Additionally, in this previous study, ancillary protein 2 was not found on the bacterial surface and did not confer protection in experimental challenge. Similarly, this protein (SAK_1438 - ancillary protein 2) was not suggested by *in silico* analysis performed here as a good vaccine candidate due to the low adhesion probability score observed (Table S1).

The CAMP factor protein was shared only by fish strains. It is a known virulence factor described for *S. agalactiae*, and is a secreted protein that forms pores in the host cell, promoting damage and tissue invasion (Rajagopal, 2009). This protein was absent only in a draft genome of *S. agalactiae* human strain CJB111, suggesting that the sequence fragment that harbors this protein might not have been represented in genome assembly. Thus, CAMP factor probably may be a good vaccine target against GBS infection in both hosts.

Concerning the proteins with unknown function or not still tested as a vaccine target, future studies have to be carried out to evaluate the protection elicited by those candidates in experimental challenge in mice, cattle, and/or fish. Furthermore, these proteins may be related to the virulence of *S. agalactiae* in these hosts, and their characterization could provide new insights into GBS infection and pathogenesis. Therefore, the vaccine targets predicted here open doors to new studies of functional characterization of proteins (mainly related to interaction with host) and vaccine development.

ACKNOWLEDGMENTS

Supplementary material

REFERENCES

Akesson M, Dufour M, Sloan GL and Simmonds RS (2007). Targeting of streptococci by zoocin A. *FEMS Microbiol. Lett.* 270: 155-161.

Baker CJ and Edwards MS (2003). Group B streptococcal conjugate vaccines. *Arch. Dis. Child.* 88: 375-378.

Barinov A, Loux V, Hammani A, Nicolas P, et al. (2009). Prediction of surface exposed proteins in *Streptococcus pyogenes*, with a potential application to other Gram-positive bacteria. *Proteomics* 9: 61-73.

Blom J, Albaum SP, Doppmeier D, Puhler A, et al. (2009). EDGAR: a software framework for the comparative analysis of prokaryotic genomes. *BMC Bioinformatics* 10: 154.

Brochet M, Couvé E, Zouine M, Vallaeys T, et al. (2006). Genomic diversity and evolution within the species *Streptococcus agalactiae*. *Microbes Infect.* 8: 1227-1243.

Cañada-Cañada F, Muñoz de la Peña A and Espinosa-Mansilla A (2009). Analysis of antibiotics in fish samples. *Anal. Bioanal. Chem.* 395: 987-1008.

Chen M, Wang R, Li LP, Liang WW, et al. (2012). Screening vaccine candidate strains against *Streptococcus agalactiae* of tilapia based on PFGE genotype. *Vaccine* 30: 6088-6092.

Delannoy CM, Zadoks RN, Lainson FA, Ferguson HW, et al. (2012). Draft genome sequence of a nonhemolytic fish-pathogenic *Streptococcus agalactiae* strain. *J. Bacteriol.* 194: 6341-6342.

Duarte RS, Miranda OP, Bellei BC, Brito MAVP, et al. (2004). Phenotypic and molecular characteristics of *Streptococcus agalactiae* isolates recovered from milk of dairy cows in Brazil. *J. Clin. Microbiol.* 42: 4214-4222.

Edmond KM, Kortsalioudaki C, Scott S, Schrag SJ, et al. (2012). Group B streptococcal disease in infants aged younger than 3 months: systematic review and meta-analysis. *Lancet* 379: 547-556.

Glaser P, Rusniok C, Buchrieser C, Chevalier F, et al. (2002). Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol. Microbiol.* 45: 1499-1513.

Hamilton A, Popham DL, Carl DJ, Lauth X, et al. (2006). Penicillin-binding protein 1a promotes resistance of group B streptococcus to antimicrobial peptides. *Infect. Immun.* 74: 6179-6187.

He Y, Xiang Z and Mobley HL (2010). Vaxign: the first web-based vaccine design program for reverse vaccinology and applications for vaccine development. *J. Biomed. Biotechnol.* 2010: 297505.

Heath PT (2011). An update on vaccination against group B streptococcus. *Expert. Rev. Vaccines* 10: 685-694.

Johri AK, Paoletti LC, Glaser P, Dua M, et al. (2006). Group B *Streptococcus*: global incidence and vaccine development. *Nat. Rev. Microbiol.* 4: 932-942.

Konto-Ghiorghi Y, Mairey E, Mallet A, Dumenil G, et al. (2009). Dual role for pilus in adherence to epithelial cells and biofilm formation in *Streptococcus agalactiae*. *PLoS Pathog.* 5: e1000422.

Liu G, Zhang W and Lu C (2012). Complete genome sequence of *Streptococcus agalactiae* GD201008-001, isolated in China from tilapia with meningoencephalitis. *J. Bacteriol.* 194: 6653.

Maione D, Margarit I, Rinaudo CD, Masignani V, et al. (2005). Identification of a universal Group B *Streptococcus* vaccine by multiple genome screen. *Science* 309: 148-150.

Maisey HC, Hensler M, Nizet V and Doran KS (2007). Group B streptococcal pilus proteins contribute to adherence to and invasion of brain microvascular endothelial cells. *J. Bacteriol.* 189: 1464-1467.

Margarit I, Rinaudo CD, Galeotti CL, Maione D, et al. (2009). Preventing bacterial infections with pilus-based vaccines: the group B *streptococcus* paradigm. *J. Infect. Dis.* 199: 108-115.

Meyer F, Goesmann A, McHardy AC, Bartels D, et al. (2003). GenDB - an open source genome annotation system for prokaryote genomes. *Nucleic Acids Res.* 31: 2187-2195.

Mian GF, Godoy DT, Leal CA, Yuhara TY, et al. (2009). Aspects of the natural history and virulence of *S. agalactiae* infection in Nile tilapia. *Vet. Microbiol.* 136: 180-183.

Mitchell JM, Griffiths MW, McEwen SA, McNab WB, et al. (1998). Antimicrobial drug residues in milk and meat: causes, concerns, prevalence, regulations, tests, and test performance. *J. Food Prot.* 61: 742-756.

Nuccitelli A, Cozzi R, Gourlay LJ, Donnarumma D, et al. (2011). Structure-based approach to rationally design a chimeric protein for an effective vaccine against Group B *Streptococcus* infections. *Proc. Natl. Acad. Sci. U. S. A.* 108: 10278-

10283.

Paoletti LC and Madoff LC (2002). Vaccines to prevent neonatal GBS infection. *Semin. Neonatol.* 7: 315-323.

Park SE, Jiang S and Wessels MR (2012). CsrRS and environmental pH regulate group B *Streptococcus* adherence to human epithelial cells and extracellular matrix. *Infect. Immun.* 80: 3975-3984.

Pereira UP, Santos AR, Hassan SS, Aburjaile FF, et al. (2013). Complete genome sequence of *Streptococcus agalactiae* strain SA20-06, a fish pathogen associated to meningoencephalitis outbreaks. *Stand. Genomic Sci.* 8: 188-197.

Ragunathan P, Spellerberg B and Ponnuraj K (2009). Structure of laminin-binding adhesin (Lmb) from *Streptococcus agalactiae*. *Acta Crystallogr. D. Biol. Crystallogr.* 65: 1262-1269.

Rajagopal L (2009). Understanding the regulation of Group B Streptococcal virulence factors. *Future Microbiol.* 4: 201-221.

Rappuoli R (2001). Reverse vaccinology, a genome-based approach to vaccine development. *Vaccine* 19: 2688-2691.

Reinscheid DJ, Gottschalk B, Schubert A, Eikmanns BJ, et al. (2001). Identification and molecular analysis of PcsB, a protein required for cell wall separation of group B streptococcus. *J. Bacteriol.* 183: 1175-1183.

Richards VP, Lang P, Bitar PD, Lefebure T, et al. (2011). Comparative genomics and the role of lateral gene transfer in the evolution of bovine adapted *Streptococcus agalactiae*. *Infect. Genet. Evol.* 11: 1263-1275.

Sharma P, Lata H, Arya DK, Kashyap AK, et al. (2013). Role of pilus proteins in adherence and invasion of *Streptococcus agalactiae* to the lung and cervical epithelial cells. *J. Biol. Chem.* 288: 4023-4034.

Singh P, Springman AC, Davies HD and Manning SD (2012). Whole-genome shotgun sequencing of a colonizing multilocus sequence type 17 *Streptococcus agalactiae* strain. *J. Bacteriol.* 194: 6005.

Soares SC, Trost E, Ramos RT, Carneiro AR, et al. (2012). Genome sequence of *Corynebacterium pseudotuberculosis* biovar *equi* strain 258 and prediction of antigenic targets to improve biotechnological vaccine production. *J. Biotechnol.* 167: 135-141.

Tettelin H, Masignani V, Cieslewicz MJ, Eisen JA, et al. (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.

Tettelin H, Masignani V, Cieslewicz MJ, Donati C, et al. (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.

Wang B, Jian J, Lu Y, Cai S, et al. (2012). Complete genome sequence of *Streptococcus agalactiae* ZQ0910, a pathogen causing meningoencephalitis in the GIFT strain of Nile tilapia (*Oreochromis niloticus*). *J. Bacteriol.* 194: 5132-5133.

Xu H, Hu C, Gong R, Chen Y, et al. (2011). Evaluation of a novel chimeric B cell epitope-based vaccine against mastitis induced by either *Streptococcus agalactiae* or *Staphylococcus aureus* in mice. *Clin. Vaccine Immunol.* 18: 893-900.