Supporting information for Tettelin *et al.* (2002) *Proc. Natl. Acad. Sci. USA*, 10.1073/pnas.182380799

Table 2. *S. agalactiae* predicted and experimentally characterized surface and secreted proteins

ORF	Size, aa	Signal peptide	Sortase motif	Lipo-protein	Other	Western blot	FACS	GBS specific	Annotation
SAG0017	447	+							pcsB
SAG0031	299	+							peptidase, M23/M37 family
SAG0032	434	+				+	+		group B streptococcal surface immunogenic protein
SAG0034	438	+		+		+	+		sugar ABC transporter, sugar-binding protein
SAG0051	126	+				+	+		MORN motif family protein
SAG0079	212				+	+	+		adenylate kinase
SAG0086	85			+				+	lipoprotein, putative
SAG0093	250	+				+	+		D-alanyl-D-alanine carboxypeptidase family protein
SAG0094	191	+							*N*-acetylmuramoyl-L-alanine amidase, family 4 protein
SAG0108	308	+							conserved hypothetical protein
SAG0114	322	+		+					ribose ABC transporter, periplasmic D-ribose-binding protein
SAG0124	356	+							sensor histidine kinase
SAG0132	294	+				+	+		SPFH domain/Band 7 family protein
SAG0134	96	+						+	hypothetical protein
SAG0146	395	+							penicillin-binding protein 4, putative
SAG0147	411	+							D-alanyl-D-alanine carboxypeptidase family protein
SAG0148	551			+		+	–		oligopeptide ABC transporter, substrate-binding protein, putative
SAG0166	123	+							conserved domain protein
SAG0176	94	+							conserved hypothetical protein
SAG0187	542	+		+		+	+		oligopeptide ABC transporter, oligopeptide-binding protein
SAG0206	60			+				+	lipoprotein, putative
SAG0213	39	+						+	hypothetical protein
SAG0231	135	+							hypothetical protein
SAG0242	308			+		+	–		amino acid ABC transporter, amino acid-binding protein
SAG0245	152			+		+	–	+	protein of unknown function/lipoprotein, putative
SAG0255	315	+							conserved hypothetical protein
SAG0257	53			+				+	lipoprotein, putative
SAG0265	235	+				+	–	+	conserved hypothetical protein
SAG0290	270	+				+	+		ABC transporter, substrate-binding protein
SAG0298	750	+							penicillin-binding protein 1A
SAG0306	535	+							KH domain protein
SAG0321	339	+							sensor histidine kinase, putative
SAG0329	106	+							PTS system, cellobiose-specific IIB component
SAG0368	435	+				+	+		protein of unknown function
SAG0371	167	+						+	hypothetical protein
SAG0383	334	+		+		+	–		protein of unknown function/lipoprotein, putative
SAG0392	521	+	+			+	+		cell wall surface anchor family protein
SAG0394	345				+				sensor histidine kinase
SAG0405	347	+		+		+	+		protein of unknown function/lipoprotein, putative
SAG0406	299	+							UTP-glucose-1-phosphate uridylyltransferase
SAG0407	338	+							glycerol-3-phosphate dehydrogenase (NAD(P)+)
SAG0416	1233	+	+			+	+		protease, putative
SAG0421	1055		+			+	–		cell wall surface anchor family protein
SAG0433	1389		+						surface protein Rib
SAG0437	123			+					lipoprotein, putative
SAG0451	149	+		+				+	bacteriocin transport accessory protein,putative
SAG0455	357	+							conserved hypothetical protein
SAG0472	126	+				+	–		rhodanese-like family protein
SAG0482	84	+							YGGT family protein
SAG0499	275				+				hemolysin A
SAG0503	279	+				+	+		lipase/acylhydrolase
SAG0504	200	+							conserved hypothetical protein
SAG0506	65	+						+	hypothetical protein
SAG0521	236	+							carboxymethylenebutenolidase-related protein
SAG0535	506	+				+	+		zinc ABC transporter, zinc-binding adhesion liprotein
SAG0596	670				+				prophage LambdaSa1, pblA protein, internal deletion
SAG0603	111				+				conserved hypothetical protein
SAG0604	239				+				prophage LambdaSa1, lysin, putative

ORF	Size, aa	Signal peptide	Sortase motif	Lipo-protein	Other	Western blot	FACS	GBS specific	Annotation
SAG0617	439				+				sensor histidine kinase VncS
SAG0624	574	+							septation ring formation regulator EzrA, putative
SAG0629	354	+							conserved domain protein
SAG0635	245	+				+	–		acid phosphatase, class B
SAG0638	109	+							cell wall surface anchor family protein, interruption-N
SAG0645	554		+			+	+		cell wall surface anchor family protein
SAG0646	307	+	+			+	–		cell wall surface anchor family protein
SAG0647	305	+							sortase family protein
SAG0649	890		+			+	+		cell wall surface anchor family protein, putative
SAG0658	383	+		+					lipoprotein, putative
SAG0675	171	+							putative secreted protein
SAG0676	885				+				proteinase, putative
SAG0677	1062		+						hypothetical protein
SAG0679	343	+		+		+	–		protein of unknown function
SAG0680	339	+				+	–		protein of unknown function
SAG0681	353	+							conserved domain protein
SAG0686	261	+				+	+		DNA-entry nuclease, putative
SAG0714	188	+						+	conserved hypothetical protein
SAG0717	266	+				+	+		amino acid ABC transporter, amino acid-binding protein
SAG0720	449				+				sensory box histidine kinase
SAG0738	132	+							conserved hypothetical protein
SAG0739	143	+							conserved hypothetical protein
SAG0742	428				+	+	+		peptidase, U32 family
SAG0755	282	+							peptidase, U32 family
SAG0757	129	+		+		+	–		protein of unknown function/lipoprotein, putative
SAG0764	230				+	+	+		phosphoglycerate mutase family protein
SAG0765	681	+							penicillin-binding protein 2b
SAG0771	512	+	+			+	+	+	cell wall surface anchor family protein
SAG0776	276	+		+					YaeC family protein, putative
SAG0777	528				+	+	+		ATP-dependent RNA helicase, DEAD/DEAH box family
SAG0785	330	+							conserved hypothetical protein
SAG0808	309	+		+		+	+		protease maturation protein, putative
SAG0824	417	+							polysaccharide deacetylase family protein
SAG0832	753	+				+	+		protein of unknown function
SAG0833	181	+						+	hypothetical protein
SAG0867	63	+							conserved hypothetical protein
SAG0868	285	+				+	–		DNA-entry nuclease
SAG0886	319	+				+	+		protein of unknown function
SAG0904	56	+						+	hypothetical protein
SAG0907	877	+		+		+	–		protein of unknown function/lipoprotein, putative
SAG0926	333	+							Tn916, NLP/P60 family protein
SAG0942	185	+				+	+		signal peptidase I, putative
SAG0949	276	+		+		+	+		amino acid ABC transporter, amino acid-binding protein
SAG0954	349			+		+	–		protein of unknown function/lipoprotein, putative
SAG0961	247	+				+	–		sortase SrtA
SAG0963	320	+							conserved hypothetical protein
SAG0971	282	+		+		+	–		protein of unknown function/lipoprotein, putative
SAG0973	320	+						+	nisin-resistance protein, putative
SAG0977	312				+				sensor histidine kinase
SAG0979	553	+		+		+	–		ABC transporter, substrate-binding protein
SAG0984	437	+							sensor histidine kinase CiaH
SAG0992	286	+		+		+	+		phosphate ABC transporter, phosphate-binding protein
SAG1007	342	+		+		+	–		iron-compound ABC transporter, iron-compound-binding protein
SAG1014	190	+				–	–		conserved hypothetical protein
SAG1018	40			+				+	lipoprotein, putative
SAG1024	183	+		+					lipoprotein, putative
SAG1029	101	+							hypothetical protein
SAG1030	304	+				+	+		protein of unknown function
SAG1037	157	+						+	hypothetical protein
SAG1052	47		+					+	cell wall surface anchor family protein, putative
SAG1072	200	+							conserved hypothetical protein
SAG1094	278				+	+	+		conserved hypothetical protein
SAG1108	357	+				+	–		spermidine/putrescine ABC transporter, spermidine/putrescine-binding prot.
SAG1121	295	+							polysaccharide deacetylase family protein
SAG1126	228	+				+	+		protein of unknown function
SAG1127	446	+						+	conserved domain protein
SAG1130	49	+						+	hypothetical protein
SAG1138	64	+							conserved hypothetical protein

ORF	Size, aa	Signal peptide	Sortase motif	Lipo-protein	Other	Western blot	FACS	GBS specific	Annotation
SAG1139	193	+							conserved hypothetical protein
SAG1149	207	+		+					lipoprotein, putative
SAG1184	236	+							conserved hypothetical protein
SAG1186	553				+				metallo-beta-lactamase superfamily protein
SAG1189	334	+							conserved hypothetical protein
SAG1190	551				+				adherence and virulence protein A
SAG1197	1072	+							hyaluronidase
SAG1201	367	+							iminodiacetate oxidase, putative
SAG1206	854	+							conserved domain protein
SAG1214	58	+							hypothetical protein
SAG1216	1252		+			+	–		pullulanase, putative
SAG1227	198	+				+	–		protein of unknown function
SAG1233	822	+				+	–		streptococcal histidine triad family protein
SAG1234	306	+		+		+	+		laminin-binding surface protein
SAG1238	202	+							hypothetical protein
SAG1283	1631		+			+	+		agglutinin receptor
SAG1313	56	+							conserved hypothetical protein
SAG1327	409	+							sensor histidine kinase
SAG1331	979	+	+			+	+		R5 protein
SAG1333	690	+	+			+	+		5'-nucleotidase family protein
SAG1350	544	+							surface antigen-related protein
SAG1361	414	+							conserved hypothetical protein
SAG1371	392	+							conserved hypothetical protein
SAG1393	310			+					iron compound ABC transporter, substrate-binding protein
SAG1404	308	+	+			+	–		cell wall surface anchor family protein
SAG1405	294	+			+	+	+		sortase family protein
SAG1406	293	+							sortase family protein
SAG1407	705	+	+			+	+		cell wall surface anchor family protein
SAG1408	901		+						cell wall surface anchor family protein
SAG1419	577			+				+	lipoprotein, putative
SAG1431	268			+					amino acid ABC transporter, amino acid-binding protein
SAG1433	375	+							conserved hypothetical protein
SAG1441	415	+				+	+		maltose/maltodextrin ABC transporter, maltose/maltodextrin-binding protein
SAG1462	970		+						cell wall surface anchor family protein
SAG1473	192	+	+					+	cell wall surface anchor family protein
SAG1474	680	+	+						amidase family protein
SAG1483	78	+							preprotein translocase, SecG subunit
SAG1488	195	+				+	+		dephospho-CoA kinase
SAG1491	530	+						+	hypothetical protein
SAG1508	590				+	+	–		67 kDa Myosin-crossreactive streptococcal antigen
SAG1518	538	+		+					peptide ABC transporter, peptide-binding protein
SAG1530	267	+		+		+	–		peptidyl-prolyl cis-trans isomerase, cyclophilin-type
SAG1533	308	+		+		+	–		manganese ABC transporter, manganese-binding adhesion liprotein
SAG1544	232	+							gluconate 5-dehydrogenase, putative
SAG1551	67	+						+	hypothetical protein
SAG1552	719	+							conserved hypothetical protein
SAG1553	477	+						+	hypothetical protein
SAG1562	280	+							conserved hypothetical protein
SAG1582	388	+		+		+	–		branched-chain amino acid ABC transporter, amino acid-binding protein
SAG1590	449				+	+	+		potassium uptake protein, Trk family
SAG1601	79	+							conserved hypothetical protein
SAG1610	285			+		+	–		amino acid ABC transporter, substrate-binding protein
SAG1618	1032				+	+	+		Snf2 family protein
SAG1624	501	+							sensor histidine kinase CsrS
SAG1628	184	+							lemA protein
SAG1631	223	+				+	–		potassium uptake protein, Trk family, putative
SAG1641	274	+				+	–		YaeC family protein
SAG1642	277	+		+		+	–		ABC transporter, substrate-binding protein
SAG1683	512	+							immunogenic secreted protein, putative
SAG1706	238	+							conserved hypothetical protein
SAG1745	148	+						+	hypothetical protein
SAG1752	390	+							conserved hypothetical protein TIGR00275
SAG1759	230				+	+	+		protein of unknown function
SAG1762	169	+							conserved hypothetical protein
SAG1767	289	+		+					acid phosphatase
SAG1768	336				+	+	+		glyceraldehyde 3-phosphate dehydrogenase
SAG1774	424	+							conserved hypothetical protein
SAG1786	130	+				+	–		protein of unknown function

ORF	Size, aa	Signal peptide	Sortase motif	Lipo-protein	Other	Western blot	FACS	GBS specific	Annotation
SAG1787	420	+							dltD protein
SAG1791	395	+							sensor histidine kinase
SAG1822	272	+				+	–		protein of unknown function
SAG1823	418				+	+	+		protein of unknown function
SAG1837	468				+				prophage LambdaSa2, lysin, putative
SAG1838	109	+							prophage LambdaSa2, holin, putative
SAG1839	136	+							conserved hypothetical protein
SAG1842	1224				+				prophage LambdaSa2, PblB, putative
SAG1912	194	+							*N*-acetylmuramoyl-L-alanine amidase, family 4 protein
SAG1921	508	+							sensor histidine kinase
SAG1932	816	+							neuraminidase-related protein
SAG1938	307	+		+		+	–		adhesion lipoprotein
SAG1941	800	+	+			+	–		2`,3`-cyclic-nucleotide 2`-phosphodiesterase
SAG1945	345	+							iron ABC transporter, iron-binding protein
SAG1947	549				+				conserved hypothetical protein
SAG1960	551				+	+	+		sensor histidine kinase
SAG1966	293			+		+	–		hemolysin precursor, putative
SAG1996	263	+	+						cell wall surface anchor family protein, putative
SAG1997	182	+							hypothetical protein
SAG1998	457	+							hypothetical protein
SAG2021	826		+						cell wall surface anchor family protein
SAG2043	255	+							cAMP factor
SAG2053	1570	+	+						serine protease, subtilase family, putative
SAG2055	462				+				sensor histidine kinase
SAG2056	202	+						+	chromosome assembly-related protein
SAG2063	630	+	+						pathogenicity protein, putative
SAG2078	320	+		+		+	–		protein of unknown function/lipoprotein, putative
SAG2094		+				+	+		competence/damage-inducible protein CinA, authentic frameshift
SAG2121	223	+						+	hypothetical protein
SAG2123	454	+							sensor histidine kinase
SAG2141	660	+				+	–		DHH family protein
SAG2147	234	+		+		+	+		protein of unknown function/lipoprotein, putative
SAG2148	179	+							LysM domain protein
SAG2174	409	+							serine protease
SAG0013	428	+				+	–		protein of unknown function

Candidate signal peptides and lipoprotein motifs were predicted with PSORT [Nakai, K. & Horton, P. (1999) *Trends Biochem. Sci.* **24,** 34-36] and other methods (see *Methods*). Sortase motifs (LPxTG) were detected by using the FINDPATTERNS program of the GCG Package [Devereux, J., Haeberli, P. & Smithies, O. (1984) *Nucleic Acids Res.* **12,** 387-395] and hidden Markov models. "Other" indicates proteins carrying other motifs (e.g., integrin-binding motif RGD) or that are similar to characterized surface-exposed proteins. Western blot results were considered positive when the antibodies revealed a predominant band of the expected molecular weight on the total protein extracts of *S. agalactiae* strain 2603 V/R. ORFs without + or – in this column were not tested in Western blot. FACS analyses were performed for Western blot-positive proteins only. Western blot and FACS data are displayed only for proteins carrying at least one of the other motifs shown in the table. Column "GBS specific" indicates genes unique to *S. agalactiae* (when compared to other completely sequenced genomes) that are present in all the *S. agalactiae* strains tested in comparative genome hybridization analyses. Finally, only proteins carrying fewer than three predicted transmembrane domains are shown in the table; other proteins are likely to be embedded in the cytoplasmic membrane and are probably not exposed on the organism's surface.