

## **SI Materials and Methods**

**Complete Genome Assembly of Strain SGBS001.** Genomic DNA from overnight culture was isolated using a phenol/chloroform extraction protocol and subjected to large-insert PacBio library preparation following the User Bulletin – Guideline for Preparing 20 kb SMRTbell™ Templates (version 2) and Procedure and Checklist – 20kb Template Preparation Using BluePippin Size-Selection (Version 3) (<http://www.pacificbiosciences.com/support/pubmap/documentation.html>). SMRTbell templates were subjected to standard SMRT sequencing using an engineered phi29 DNA polymerase on the PacBio RS system according to the manufacturer's protocol. Subsequently, paired end sequencing reads derived from short-read Illumina HiSeq data were mapped to the newly assembled contig for error correction of the PacBio data (1). Annotation of the closed SGBS001 genome was performed using RAST (2) and BASYS (3) in addition to homologous annotation transfer from the previously sequenced 2603V/R genome (accession number NC\_004116) using CLCBio Genomics Workbench version 7.0.3 (Qiagen). The GenBank accession number for the SGBS001 genome is CP010867.

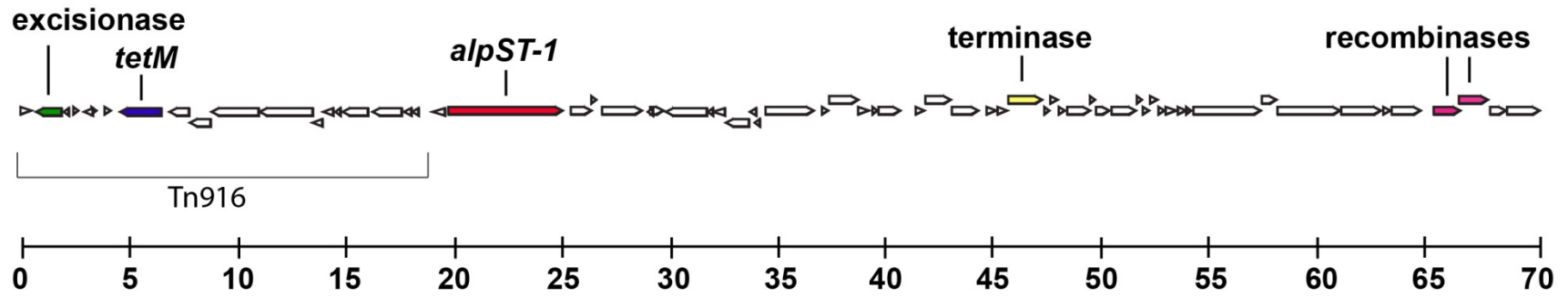
**Identification of Genes with an Excess of Polymorphisms.** To identify genes possibly under positive selection, we sought to determine whether any of the 1,931 core ST-1 genes had a higher number of SNPs than expected for a random distribution as previously described (4). The expected numbers of SNPs was calculated for all core genes and corrected for the size of the gene relative to the core genome. The resulting  $\chi^2$  probabilities were then corrected for multiple testing using the Bonferroni method (Table 1).

**Transcriptome Analysis Using RNASeq.** For RNA-Seq analysis, strains were grown in quadruplicate to mid-exponential phase and RNA was isolated using a RNeasy kit (Qiagen). RNA integrity was confirmed using an Agilent Bioanalyzer. First and second strand cDNA synthesis was performed on 500 ng of rRNA depleted RNA using the Ovation Prokaryotic RNA-Seq System (NuGEN). The resulting cDNA was fragmented to 200 bp (mean fragment size) with the S220 Focused-ultrasonicator (Covaris) and used to make barcoded sequencing libraries on the SPRI-TE Nucleic Acid Extractor (Beckman-Coulter). No size selection was employed. Libraries were quantitated by qPCR (KAPA Systems), multiplexed and 8 samples per lane were sequenced on the HiSeq2000 using 76 bp paired-end sequencing. The raw reads in FASTQ format were aligned to the SGBS001 genome using Mosaik alignment software (version: 1.1.0021, <http://code.google.com/p/mosaik-aligner/>) with the following alignment parameters: 8 percent maximal percentage read length allowed to be errors and 50 percent minimal percentage of the read length aligned. Duplicate fragments with a lower alignment quality were discarded. Next, the overlaps between aligned reads and annotated genes were counted using HTSeq software (<http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html>). If the number of overlapped read of any given gene was less than one per million total mapped read for all samples, this gene was excluded from further analysis. The gene counts were normalized using the scaling factor method (5). A negative binomial generalized linear model was fitted to each gene. Then a likelihood-ratio test was applied to examine if there was a difference in the gene transcript levels among the compared strains (5). The Benjamini-Hochberg method was used to control false discovery rate (FDR) (6). Next, pair-wise comparisons using the likelihood ratio test were performed to compare the gene transcript levels between pairs of strains with the Holm's method used to calculate adjusted *P*-values to correct for multiple testing. Transcript

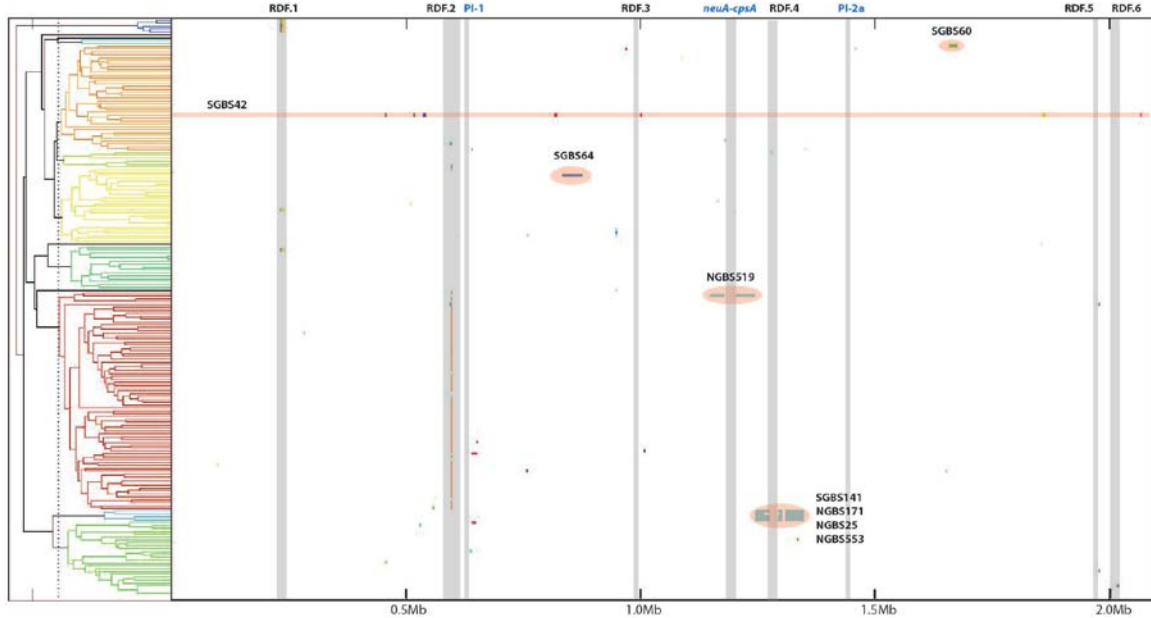
levels were considered significantly different only if the mean transcript level difference was  $\geq$  1.5-fold and the final, adjusted  $P$  value was less than 0.05.

#### SI Material and Methods References

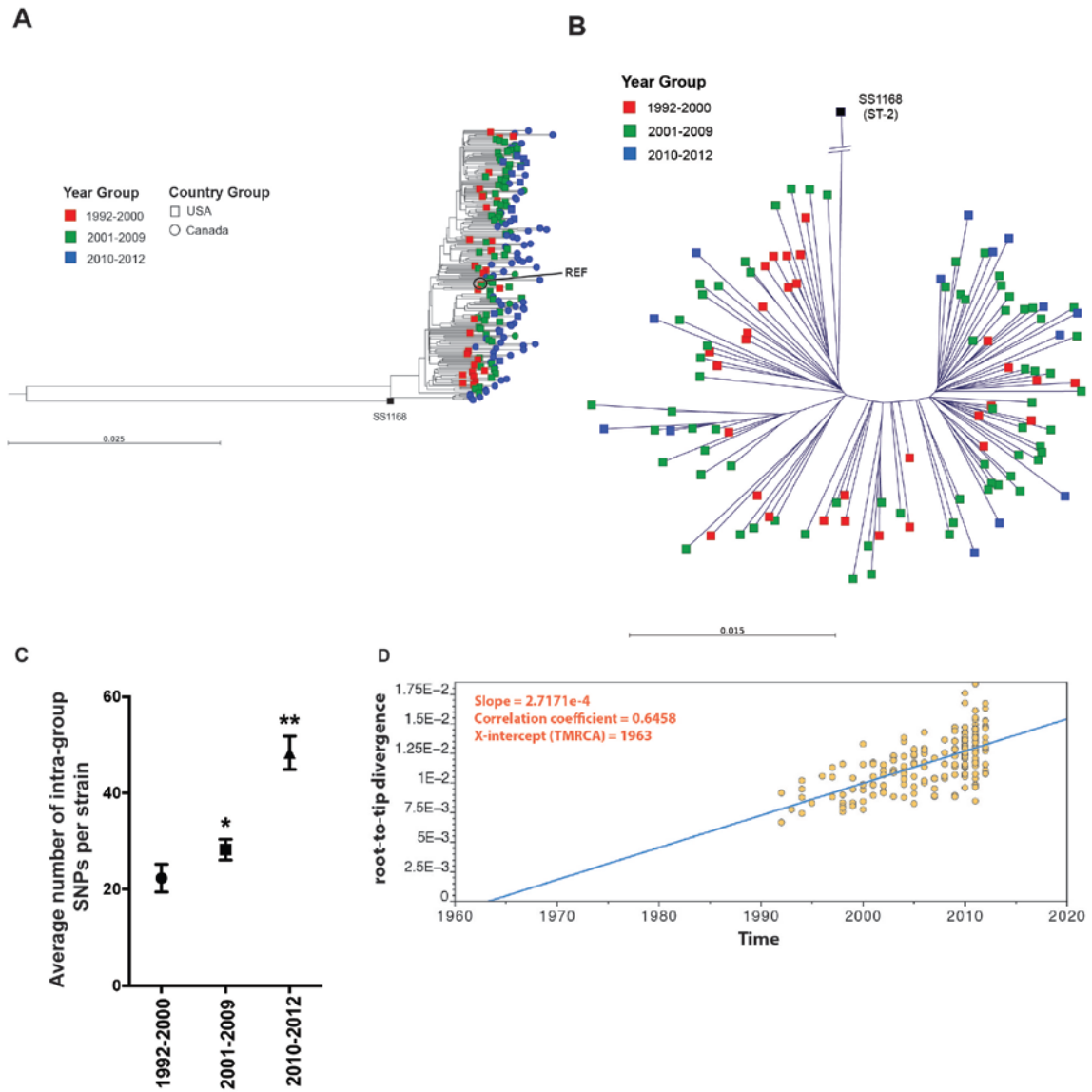
1. Koren S, *et al.* (2013) Reducing assembly complexity of microbial genomes with single-molecule sequencing. *Genome Biol* 14(9):R101.
2. Aziz RK, *et al.* (2008) The RAST Server: rapid annotations using subsystems technology. *BMC Genomics* 9:75.
3. Van Domselaar GH, *et al.* (2005) BASys: a web server for automated bacterial genome annotation. *Nucleic Acids Res* 33(Web Server issue):W455-459.
4. Beres SB, *et al.* (2010) Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc Natl Acad Sci U S A* 107(9):4371-4376.
5. Anders S & Huber W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11(10):R106.
6. Reiner A, Yekutieli D, & Benjamini Y (2003) Identifying differentially expressed genes using false discovery rate controlling procedures. *Bioinformatics* 19(3):368-375.



**Fig. S1.** Map of RDF.2 from SGBS001. Scale is in kilobases.

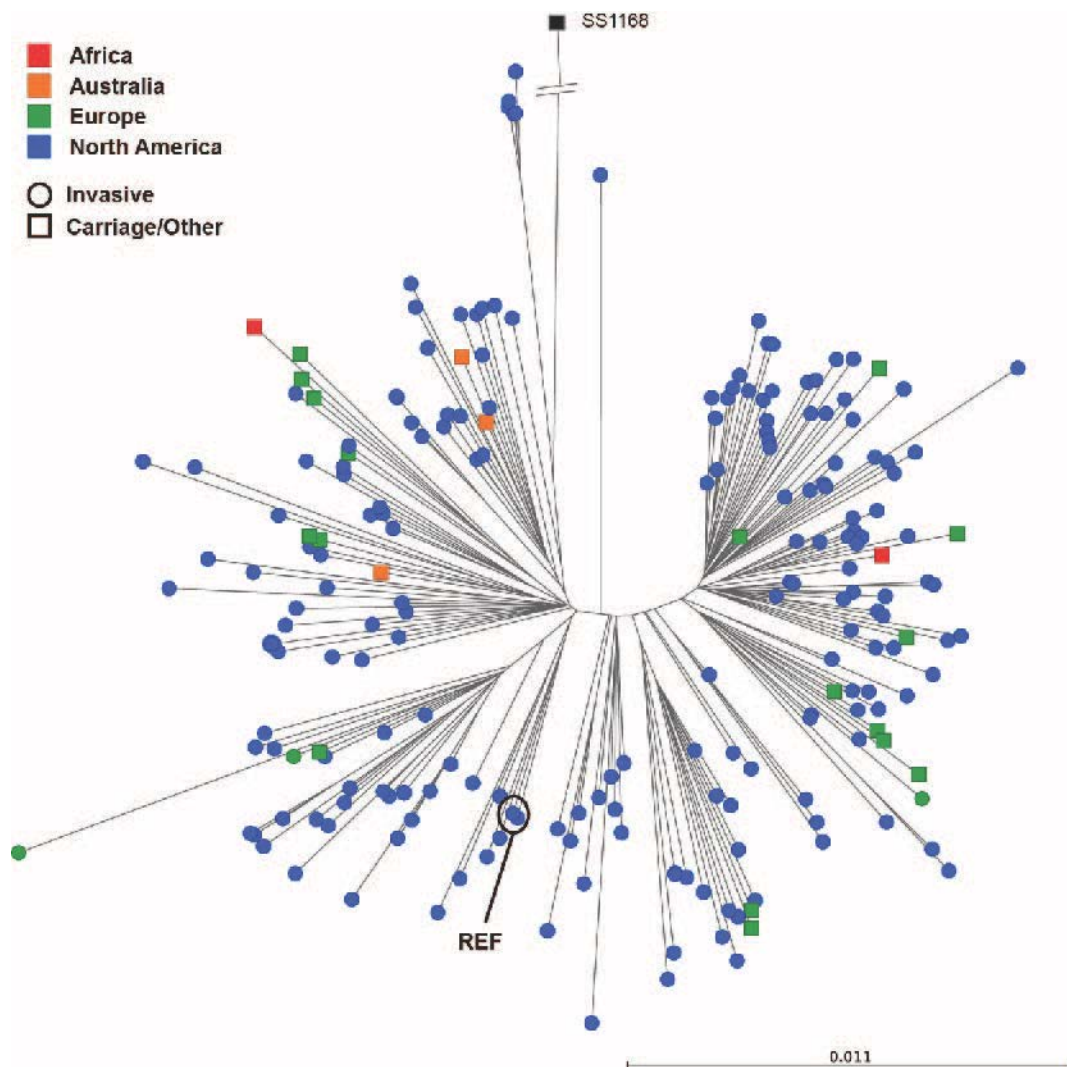


**Fig. S2.** Identified regions of potential recombination in ST-1 GBS. Proportion of shared ancestry (PSA) tree and recombinogenic segments were determined by BratNextGen. PSA clusters based on 202 GBS ST-1 pseudogenomes are displayed on left. The SGBS001 genome position is given on the bottom of the figure. Highlighted (gray) are regions of the reference genome including mobile genetic elements (RDF.1 – RDF.6), pilus islands (PI-1, PI-2a), and the capsule biosynthesis locus (*neuA-cpsA*). Also highlighted (salmon) are the recombinogenic segments of GBS strains showing greater phylogenetic distance from the core ST1 cluster (Fig. 3B).



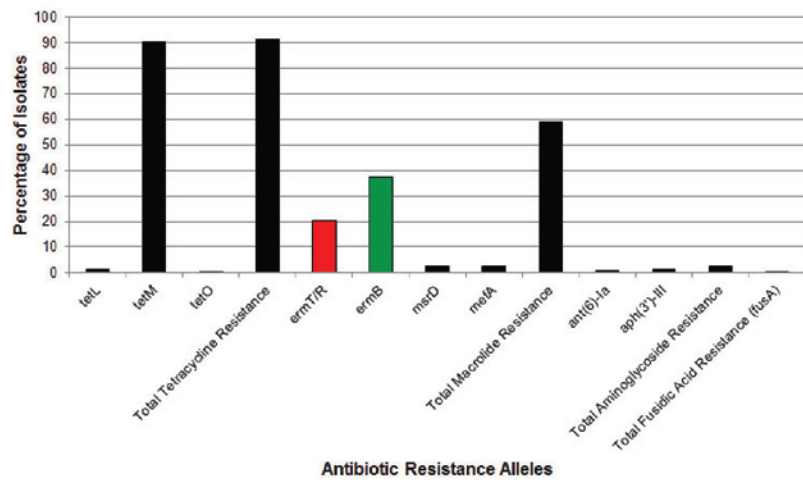
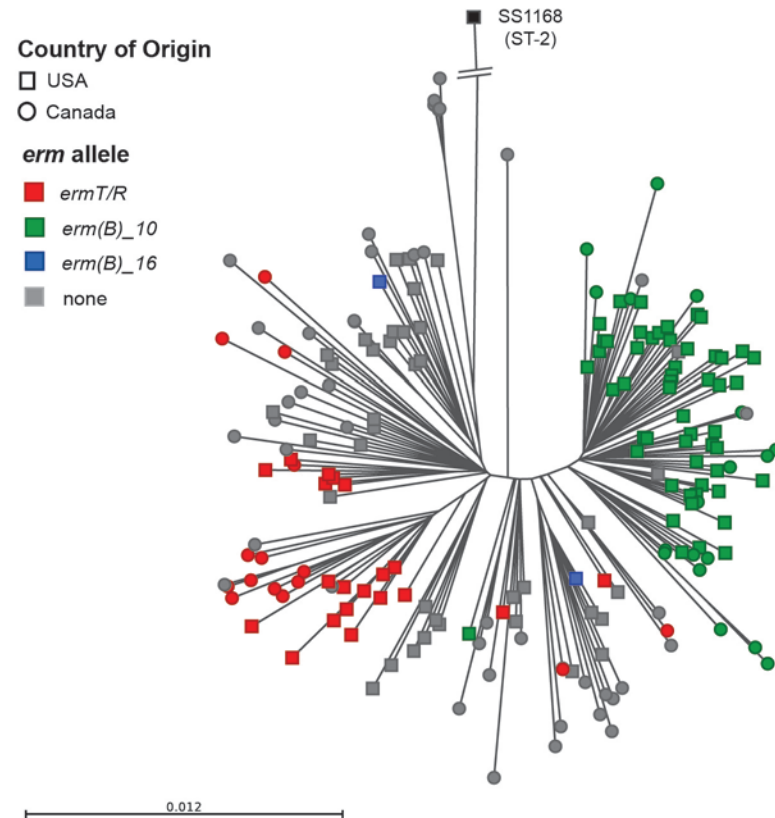
**Fig. S3.** (A) Phylogram of neighbor-joining tree of 194 ST-1 strains lacking evidence of recombination. Shape and color of strains are as indicated in legend. Genetic distance is indicated by in-laid bar. Data are the same as for Fig. 3D except using a phylogram rather than a radial tree depiction. (B) Neighbor-joining tree of all Houston isolated strains after excluding strains with potential recombination. The year of isolation of

individual strains is noted in the legend. In-laid bar represents genetic distance. (C) Graph plotting the mean  $\pm$  standard deviation of the number of intra-group SNPs per strain stratified by indicated time group. Asterisks indicate  $P < 0.0001$  relative to 1992-2000 (\*) and  $P < 0.0001$  relative to both 1992-2000 and 2001-2010 (\*\*) as determined Mann-Whitney U-test. (D) Estimation of the year of origin inferred from 194 ST-1 strains based on 5,880 concatenated core SNPs by neighbor joining. Illustrated is the best fit of the root-to-tip divergence for each of the strains as determined by Path-O-Gen. The X-intercept (1963) is the best estimate of the time of origin of the most recent common serotype V, ST-1 ancestor.



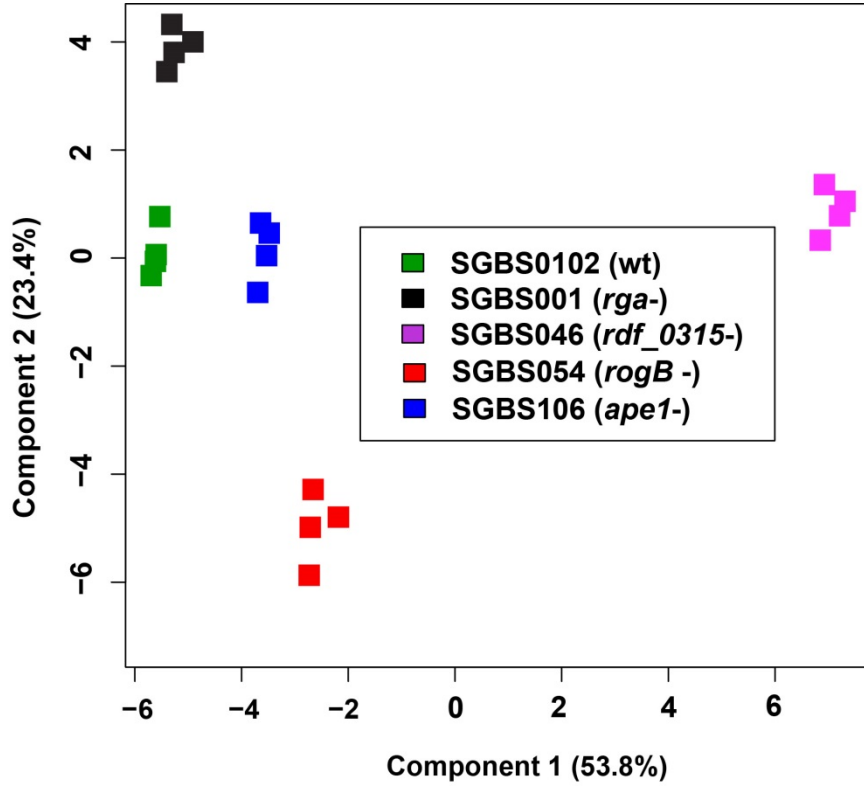
**Fig. S4.** Rooted neighbor joining tree of 194 North American ST-1 strains from this study (as depicted in Fig. 3D) along with 24 serotype V, ST-1 strains from (10). The strains from this study were all isolated in North America and are in blue whereas strains from (10) were isolated in Europe, Africa, or Australia as indicated in the legend. All strains from this study were invasive and thus are depicted by a circle. The strains from (10) were either invasive (circle) or colonizers/unknown clinical status (squares).



**A****B**

**Fig. S5.** Summary of antimicrobial resistance elements identified in ST-1 strains. (A) Antimicrobial resistance determinants were determined among sequenced serotype V, ST-1 strains using RESFINDER. Y-axis indicates percent of isolates with given

antimicrobial resistance determinant. (B) *erm* alleles are indicated along with phylogenetic relationships as determined in Fig. 3D. Country of origin of strains and *erm* alleles are as indicated in legend. Note clustering of strains with *ermT/R* and *ermB\_10* alleles suggesting single insertion events. *ermB\_10* containing strains had the *ermB* gene located in Tn3872 which appeared to have occurred from a single recombination event of Tn917 into Tn916 to create Tn3872 as described by Da Cunha et al. (10).



**Fig. S6.** Principal components analysis (PCA) of RNAseq data. PCA showing inter-sample variation of quadruplicate replicates of indicated strains grown to mid-exponential in TH broth. Strain numbers and phenotype are indicated in the legend. % values refer to amount of variation explained by each axis.

Table S1. List of strains used in this study.

<b>Strain</b>	<b>Year of isolation</b>	<b>Location of isolation</b>	<b>MLST</b>	<b>Genotype<sup>a</sup> (if applicable)</b>
NGBS008	2009	Toronto, CA	1	
NGBS009	2010	Toronto, CA	1	
NGBS010	2009	Toronto, CA	1	
NGBS021	2009	Toronto, CA	1	
NGBS022	2009	Toronto, CA	1	
NGBS025	2009	Toronto, CA	1	
NGBS028	2009	Toronto, CA	1	
NGBS030	2010	Toronto, CA	1	
NGBS035	2010	Toronto, CA	1	
NGBS044	2010	Toronto, CA	1	
NGBS054	2010	Toronto, CA	1	
NGBS063	2010	Toronto, CA	1	
NGBS068	2010	Toronto, CA	1	
NGBS092	2010	Toronto, CA	1	
NGBS093	2010	Toronto, CA	1	
NGBS094	2010	Toronto, CA	1	
NGBS099	2010	Toronto, CA	1	
NGBS107	2010	Toronto, CA	1	
NGBS110	2010	Toronto, CA	1	
NGBS117	2010	Toronto, CA	1	
NGBS134	2010	Toronto, CA	1	
NGBS164	2010	Toronto, CA	1	
NGBS171	2010	Toronto, CA	1	
NGBS172	2010	Toronto, CA	1	
NGBS177	2010	Toronto, CA	1	
NGBS180	2010	Toronto, CA	1	
NGBS200	2010	Toronto, CA	1	
NGBS210	2011	Toronto, CA	1	
NGBS229	2011	Toronto, CA	1	
NGBS234	2010	Toronto, CA	1	
NGBS241	2011	Toronto, CA	453	
NGBS244	2011	Toronto, CA	1	
NGBS246	2011	Toronto, CA	1	
NGBS267	2010	Toronto, CA	1	
NGBS272	2011	Toronto, CA	1	
NGBS273	2011	Toronto, CA	1	
NGBS275	2010	Toronto, CA	1	
NGBS279	2010	Toronto, CA	1	
NGBS283	2010	Toronto, CA	1	
NGBS287	2010	Toronto, CA	1	
NGBS288	2010	Toronto, CA	1	
NGBS298	2011	Toronto, CA	1	

NGBS303	2010	Toronto, CA	1
NGBS321	2011	Toronto, CA	1
NGBS323	2011	Toronto, CA	1
NGBS325	2011	Toronto, CA	1
NGBS330	2011	Toronto, CA	1
NGBS331	2011	Toronto, CA	1
NGBS332	2011	Toronto, CA	1
NGBS338	2011	Toronto, CA	1
NGBS348	2011	Toronto, CA	1
NGBS357	2011	Toronto, CA	1
NGBS359	2011	Toronto, CA	1
NGBS360	2011	Toronto, CA	1
NGBS372	2011	Toronto, CA	1
NGBS380	2011	Toronto, CA	1
NGBS381	2011	Toronto, CA	1
NGBS411	2011	Toronto, CA	1
NGBS418	2011	Toronto, CA	1
NGBS425	2011	Toronto, CA	1
NGBS434	2011	Toronto, CA	1
NGBS444	2011	Toronto, CA	1
NGBS462	2011	Toronto, CA	1
NGBS492	2012	Toronto, CA	1
NGBS494	2012	Toronto, CA	1
NGBS497	2012	Toronto, CA	1
NGBS499	2012	Toronto, CA	1
NGBS513	2012	Toronto, CA	1
NGBS519	2012	Toronto, CA	1
NGBS536	2012	Toronto, CA	1
NGBS553	2012	Toronto, CA	1
NGBS558	2012	Toronto, CA	1
NGBS561	2012	Toronto, CA	1
NGBS571	2012	Toronto, CA	1
NGBS579	2012	Toronto, CA	1
NGBS580	2012	Toronto, CA	1
NGBS586	2012	Toronto, CA	1
NGBS604	2012	Toronto, CA	1
NGBS624	2012	Toronto, CA	1
NGBS630	2012	Toronto, CA	1
NGBS633	2012	Toronto, CA	531

595\_596insT RDF\_1357 (AP1\_2a encoding gene)  
1195delT RDF\_1410 (*rga*)

SGBS001	1993	Houston, TX, USA	1
SGBS002	1993	Houston, TX, USA	1
SGBS003	1994	Houston, TX, USA	1
SGBS004	1995	Houston, TX, USA	1
SGBS005	1996	Houston, TX, USA	1
SGBS006	1998	Houston, TX, USA	1

SGBS007	1997	Houston, TX, USA	1	
SGBS008	1998	Houston, TX, USA	1	
SGBS010	1999	Houston, TX, USA	1	
SGBS011	2000	Houston, TX, USA	24	
SGBS012	2000	Houston, TX, USA	452	
SGBS013	2001	Houston, TX, USA	1	
SGBS014	2002	Houston, TX, USA	1	
SGBS015	2003	Houston, TX, USA	1	
SGBS016	2003	Houston, TX, USA	1	
SGBS019	2005	Houston, TX, USA	26	
SGBS020	2006	Houston, TX, USA	1	
SGBS021	2007	Houston, TX, USA	1	
SGBS022	2008	Houston, TX, USA	1	
SGBS023	2009	Houston, TX, USA	86	
SGBS024	2010	Houston, TX, USA	370	
SGBS025	2010	Houston, TX, USA	1	
SGBS026	2011	Houston, TX, USA	1	
SGBS027	2011	Houston, TX, USA	1	
SGBS028	2011	Houston, TX, USA	19	
SGBS029	2011	Houston, TX, USA	1	
SGBS030	2012	Houston, TX, USA	1	
SGBS031	1992	Houston, TX, USA	1	
SGBS032	1992	Houston, TX, USA	1	
SGBS033	1994	Houston, TX, USA	1	
SGBS034	1994	Houston, TX, USA	1	
SGBS035	1994	Houston, TX, USA	1	
SGBS036	1997	Houston, TX, USA	1	
SGBS037	1997	Houston, TX, USA	1	
SGBS038	1998	Houston, TX, USA	1	
SGBS039	1998	Houston, TX, USA	1	435delG RDF_0659 (AP1-2a encoding gene)
SGBS040	1998	Houston, TX, USA	1	
SGBS041	1998	Houston, TX, USA	1	
SGBS042	1998	Houston, TX, USA	1	
SGBS043	1998	Houston, TX, USA	1	Wild-type
SGBS044	1999	Houston, TX, USA	1	
SGBS045	1999	Houston, TX, USA	1	
SGBS046	1999	Houston, TX, USA	1	Y222X RDF_0655 (BP1 encoding gene) E155X RDF_0315
SGBS047	2000	Houston, TX, USA	1	
SGBS048	2000	Houston, TX, USA	1	672_673insA RDF_1161 ( <i>cpsO</i> )
SGBS049	2000	Houston, TX, USA	1	
SGBS050	2000	Houston, TX, USA	1	
SGBS051	2000	Houston, TX, USA	1	
SGBS052	2000	Houston, TX, USA	1	
SGBS053	2000	Houston, TX, USA	1	
SGBS054	2001	Houston, TX, USA	1	983delT RDF_1359 ( <i>rogB</i> )
SGBS056	2001	Houston, TX, USA	1	

SGBS057	2001	Houston, TX, USA	1	
SGBS058	2001	Houston, TX, USA	1	
SGBS059	2001	Houston, TX, USA	1	
SGBS060	2001	Houston, TX, USA	1	
SGBS061	2002	Houston, TX, USA	1	
SGBS062	2002	Houston, TX, USA	1	
SGBS063	2002	Houston, TX, USA	1	
SGBS064	2002	Houston, TX, USA	1	
SGBS065	2002	Houston, TX, USA	1	
SGBS066	2002	Houston, TX, USA	1	
SGBS067	2002	Houston, TX, USA	1	
SGBS068	2003	Houston, TX, USA	1	
SGBS069	2003	Houston, TX, USA	1	
SGBS070	2003	Houston, TX, USA	1	
SGBS071	2003	Houston, TX, USA	1	
SGBS072	2003	Houston, TX, USA	1	
SGBS073	2003	Houston, TX, USA	26	
SGBS074	2004	Houston, TX, USA	1	
SGBS075	2004	Houston, TX, USA	1	
SGBS076	2004	Houston, TX, USA	1	
SGBS077	2004	Houston, TX, USA	1	
SGBS078	2004	Houston, TX, USA	1	
SGBS079	2004	Houston, TX, USA	1	
SGBS080	2004	Houston, TX, USA	1	
SGBS081	2004	Houston, TX, USA	1	
SGBS082	2004	Houston, TX, USA	1	
SGBS083	2004	Houston, TX, USA	1	
SGBS084	2005	Houston, TX, USA	1	
SGBS085	2005	Houston, TX, USA	1	
SGBS086	2005	Houston, TX, USA	1	
SGBS087	2005	Houston, TX, USA	1	Wild-type
SGBS088	2005	Houston, TX, USA	1	
SGBS089	2005	Houston, TX, USA	1	
SGBS090	2005	Houston, TX, USA	190	
SGBS092	2005	Houston, TX, USA	1	
SGBS093	2005	Houston, TX, USA	1	
SGBS094	2005	Houston, TX, USA	1	
SGBS095	2005	Houston, TX, USA	1	
SGBS096	2005	Houston, TX, USA	1	
SGBS097	2005	Houston, TX, USA	1	
SGBS098	2005	Houston, TX, USA	1	
SGBS099	2005	Houston, TX, USA	19	
SGBS100	2006	Houston, TX, USA	1	
SGBS101	2006	Houston, TX, USA	1	
SGBS102	2006	Houston, TX, USA	1	Wild-type
SGBS103	2006	Houston, TX, USA	1	
SGBS104	2006	Houston, TX, USA	1	

SGBS105	2006	Houston, TX, USA	1	
SGBS106	2006	Houston, TX, USA	1	433delA RDF_0654 (ape1)
SGBS107	2006	Houston, TX, USA	1	159delA RDF_0659 (AP1_1 encoding gene)
SGBS108	2006	Houston, TX, USA	1	
SGBS109	2006	Houston, TX, USA	1	
SGBS110	2007	Houston, TX, USA	1	252_253insG RDF_1165 ( <i>cpsG</i> )
SGBS111	2007	Houston, TX, USA	1	
SGBS114	2007	Houston, TX, USA	1	
SGBS115	2007	Houston, TX, USA	1	
SGBS116	2007	Houston, TX, USA	1	
SGBS117	2008	Houston, TX, USA	26	
SGBS118	2008	Houston, TX, USA	1	
SGBS119	2008	Houston, TX, USA	1	
SGBS120	2008	Houston, TX, USA	1	
SGBS121	2009	Houston, TX, USA	190	
SGBS122	2009	Houston, TX, USA	1	
SGBS123	2009	Houston, TX, USA	55	
SGBS124	2009	Houston, TX, USA	49	
SGBS125	2009	Houston, TX, USA	1	
SGBS126	2009	Houston, TX, USA	1	
SGBS127	2009	Houston, TX, USA	1	
SGBS128	2009	Houston, TX, USA	19	
SGBS129	2009	Houston, TX, USA	1	
SGBS130	2009	Houston, TX, USA	19	
SGBS131	2009	Houston, TX, USA	19	
SGBS132	2009	Houston, TX, USA	1	
SGBS133	2009	Houston, TX, USA	1	
SGBS134	2009	Houston, TX, USA	452	
SGBS135	2007	Houston, TX, USA	1	
SGBS136	2007	Houston, TX, USA	1	
SGBS137	2010	Houston, TX, USA	19	
SGBS138	2010	Houston, TX, USA	1	
SGBS139	2008	Houston, TX, USA	1	
SGBS140	2008	Houston, TX, USA	1	
SGBS141	2009	Houston, TX, USA	1	
SGBS142	2009	Houston, TX, USA	1	
SGBS143	2010	Houston, TX, USA	1	
SGBS144	2010	Houston, TX, USA	1	
SGBS145	2010	Houston, TX, USA	1	
SGBS146	2010	Houston, TX, USA	1	751_delGAACTTTTAAATAAAC RDF_1159 ( <i>cpsK</i> )
SGBS147	2010	Houston, TX, USA	1	940delT RDF_1357 (AP1-2a encoding gene)
SGBS148	2011	Houston, TX, USA	1	930delT RDF_1357 (AP1-2a encoding gene)
SGBS149	2012	Houston, TX, USA	19	
SGBS150	2012	Houston, TX, USA	1	
SGBS151	2012	Houston, TX, USA	1	
SGBS152	2012	Houston, TX, USA	1	
SGBS153	2013	Houston, TX, USA	26	



SGBS154	2013	Houston, TX, USA	19
SGBS155	2013	Houston, TX, USA	1

---

<sup>a</sup>Genotype refers to strains used in functional assays in figure 5

**Table S2. Examples of recombination of at least 5 kb identified in ST-1, serotype V GBS.**

Strain	SGBS001 Start <sup>a</sup>	SGBS001 End	SGBS001 CDS	Length (bp)	Best BLAST hit <sup>b</sup>	Hit Start <sup>c</sup>	Hit End	Identity (%)
SGBS060	1658415	1674424	RDF_1615 to RDF_1635	15875	2603V/R (V)	1686219	1702221	99.94
SGBS064	834287	875209	RDF_0812 to RDF_0851	40922	NEM316 (III)	845210	885592	99.92
SGBS141	1245926	1274722	RDF_1206 to RDF_1232	28796	09mas018883 (V)	1217847	1246654	99.99
	1294760	1300559	RDF_1253 to RDF_1258	5799	NEM316 (III)	1425086	1430886	99.1
	1310117	1347812	RDF_1270 to RDF_1304	37695	2603V/R (V)	1328302	1362368	99.04

<sup>a</sup> Beginning and ending genome position of the reference genome (SGBS001)

<sup>b</sup> Best BLAST hit determined by blastn alignment of de novo assembled query contig against all completed GBS genomes. with serotype of identified strain noted in parentheses

<sup>c</sup> Start and end points refer to genome identified as source of recombination

**Table S3. Gene regions of strain SGBS001 > 1 kb missing from at least one serotype V, ST-1 strain**

<b>RDF#</b>	<b># strains affected</b>	<b>Region of difference<sup>a</sup></b>
RDF_0167-0169	1	No
RDF_0217-0229	1	Yes, 1.2
RDF_0230-0238	10	Yes, 1.2
RDF_0240-0246	1	Yes, 1.3
RDF_0407-0408	1	No
RDF_0547-0592	8	Yes, 1.4
RDF_0948-0959	1	Yes, 5.4
RDF_1226-1234	2	Yes, 1.13
RDF_1240-1250	1	Yes, 2.6
RDF_1299-1301	1	Yes, 1.13
RDF_1440-1442	5	Yes, 1.17
RDF_1900-1906	1	Yes, 1.21
RDF_1877-1890	1	Yes, 2.6
RDF_1895-1916	1	No
RDF_1972-1985	2	Yes, 2.9
RDF_1986-1988	4	No

<sup>a</sup>Refers to regions previously identified in Tettelin et al, PNAS 2005;102:13950-13955.

**Table S4.** Genes with significantly increased genetic variation compared to the remainder of the ST-1 core genome.

Gene location	Gene name	<i>P</i> value for polymorphisms	SNPs <sup>a</sup> + Indels <sup>b</sup> (n)	NON/ SYN <sup>c</sup>	Putative function of encoded protein
<i>rdf_0707</i>		3.71E-20	6	3/1	hypothetical protein CDS
<i>rdf_1161</i>	<i>cpsO</i>	1.26E-19	19	17/0	Capsule synthesis protein
<i>rdf_0654</i>	<i>ape1</i>	5.25E-16	20	15/2	Regulation of pilus 1 expression
<i>rdf_1162</i>	<i>cpsN</i>	2.10E-12	15	13/2	Capsule synthesis protein
<i>rdf_1167</i>	<i>cpsE</i>	8.46E-10	18	17/0	Capsule synthesis protein
<i>rdf_1124</i>		6.85E-09	11	10/1	GntR family transcriptional regulator
<i>rdf_0315</i>		8.97E-08	10	7/1	LuxR family response regulator
<i>rdf_1131</i>		5.45E-07	7	3/3	Hypothetical protein
<i>rdf_0836</i>	<i>udk</i>	5.78E-07	10	6/1	Uridine kinase
<i>rdf_1873</i>		1.35E-06	9	4/1	Hypothetical protein
<i>rdf_1166</i>	<i>cpsF</i>	1.98E-06	8	7/0	Capsule synthesis protein
<i>rdf_1410</i>	<i>rga</i>	2.40E-05	13	9/0	Regulation of pilus 2 expression
<i>rdf_0863</i>	<i>glgC</i>	6.24E-05	13	7/4	Glucose-1-phosphate adenylyltransferase
<i>rdf_0652</i>		1.46E-04	4	1/0	Chaperonin
<i>rdf_0973</i>		1.50E-04	16	12/3	Oligopeptide transporter
<i>rdf_1561</i>	<i>trkA</i>	2.21E-04	9	3/3	Potassium uptake protein
<i>rdf_0257</i>		2.44E-04	3	3/0	Hypothetical protein
<i>rdf_1359</i>	<i>rogB</i>	3.14E-04	15	10/2	Regulation of pilus 2 expression
<i>rdf_1660</i>		3.77E-04	7	6/0	Universal stress protein family
<i>rdf_1181</i>	<i>estA</i>	4.66E-04	8	6/0	Tributylin esterase
<i>rdf_1652</i>		1.26E-03	7	5/0	Putative membrane protein
<i>rdf_0497</i>		1.37E-03	9	2/2	Glutamate transport protein
<i>rdf_1119</i>	<i>skzL</i>	2.02E-03	13	10/3	Streptokinase B
<i>rdf_0087</i>		4.41E-03	5	5/1	Hypothetical protein
<i>rdf_0314</i>		5.30E-03	11	4/1	Sensor histidine kinase

<i>rdf_1481</i>		8.22E-03	10	5/5	NAD-dependent oxidoreductase
<i>rdf_1169</i>	<i>cpsC</i>	1.45E-02	8	4/1	Capsule synthesis protein
<i>rdf_1309</i>		2.19E-02	8	4/4	ATP-binding transport protein
<i>rdf_0055</i>	<i>adhP</i>	2.40E-02	10	5/1	Alcohol dehydrogenase
<i>rdf_0747</i>		2.80E-02	3	1/0	Hypothetical protein
<i>rdf_1793</i>		3.14E-02	11	9/0	Glucuronyl hydrolase

---

**Table S5. Summary of RNAseq data.**

<b>Gene</b>	<b>Fold-change in SGBS001 (<i>rga-</i>)<sup>a</sup></b>	<b>Fold-change in SGBS046 (<i>rdf_0315 -</i>)</b>	<b>Fold change in SGBS054 (<i>rogB -</i>)</b>	<b>Fold-change in SGBS106 (<i>ape1 -</i>)</b>	<b>Protein function</b>
RDF_0032	<sup>b</sup>	0.47			Putative peptidase
RDF_0035		2.09			ABC transporter
RDF_0054		5.69			Alcohol dehydrogenase
RDF_0055		3.60			Alcohol dehydrogenase
RDF_0117		2.02			Ribose transport protein
RDF_0119		2.07			Ribokinase
RDF_0120		3.67			Ribose operon repressor
RDF_0125		0.43			Argininosuccinate synthase
RDF_0126		0.43			Argininosuccinate lyase
RDF_0191	0.48				Dipeptide-binding ABC transporter
RDF_0192	0.42	0.42		0.47	Oligopeptide transport system permease
RDF_0193	0.39	0.45		0.48	Dipeptide transport system permease protein DppC
RDF_0194	0.41	0.41		0.49	Dipeptide transport ATP-binding protein DppD
RDF_0195	0.43	0.49		0.48	Dipeptide transport ATP-binding protein DppF
RDF_0231				2.62	conserved hypothetical protein
RDF_0232				2.25	Rep protein
RDF_0233				2.12	FtsK/SpoIIIE family protein
RDF_0235				2.48	hypothetical protein
RDF_0279		2.06			glpK
RDF_0421		2.36			nrdF-1
RDF_0422		2.07			nrdI-1
RDF_0423		2.09			nrdE-1
RDF_0425		2.12			hypothetical protein
RDF_0438		0.26	0.33		AraC family transcriptional regulator
RDF_0439		0.04	0.05		Alp3 protein
RDF_0470	0.10				BioY
RDF_0471	0.14				BioB
RDF_0569		3.30			Streptococcal extracellular nuclease
RDF_0580	4.81	4.99	5.00	3.49	hypothetical protein

RDF_0587			31.82		Anti-restriction protein
RDF_0658	2.10				Pilus1a associated sortase
RDF_0683			2.66		Secreted protein
RDF_0701		2.71			ABC transporter
RDF_0702		3.10		2.06	Transmembrane helix protein
RDF_0708			0.17		Gluconoride permease
RDF_0758		0.46			RibD
RDF_0800		6.48			Beta-glucoside anti-terminator
RDF_0801		5.62			Beta-glucoside PTS
RDF_0802		2.64			BglA
RDF_0809	3.15	4.52	2.36	2.69	hypothetical protein
RDF_0810	2.22	2.51			NagB
RDF_0816	3.06	2.61	3.60	2.80	Hydrolase
RDF_0907	2.14				hypothetical protein
RDF_0908		0.45			Nucleoside diphosphate kinase
RDF_1009		2.37			Autolysis regulator LytR
RDF_1080		0.46			GuaC
RDF_1135		0.36		2.00	Sodium/glycine symporter
RDF_1353		17.96	12.85	4.16	Pilus 2a accessory protein 2
RDF_1354		19.72	15.36	4.14	Pilus 2a associated sortase
RDF_1355		26.17	18.45	4.23	Pilus 2a associated sortase
RDF_1356		17.92	12.23	4.91	Pilus 2a backbone protein
RDF_1357		34.89	29.74	6.60	Pilus 2a accessory protein 1
RDF_1409	0.08		4.91		Srr-1 protein
RDF_1410	2.04		3.00		Rga protein
RDF_1471		2.06			Permease protein
RDF_1489	2.14				hypothetical protein
RDF_1490	2.37				hypothetical protein
RDF_1491	2.08				hypothetical protein
RDF_1509		10.06			hypothetical protein
RDF_1510		10.49			ABC transporter
RDF_1511		11.17			Branched chain amino acid transporter
RDF_1512			2.15		Branched chain amino acid transporter
RDF_1513	3.71	2.23	5.62	2.94	Branched chaing amino acid permease

RDF_1567	4.71	2.29	Adh2
RDF_1578		0.05	PEP-dihydroxyacetone PTS
RDF_1579		0.05	PEP-dihydroxyacetone PTS
RDF_1580		0.04	PEP-dihydroxyacetone PTS
RDF_1581		0.07	GlpF-2
RDF_1691		0.48	CAMP factor
RDF_1709		2.26	Acid phosphatase
RDF_1739	0.39	0.44	Glycine transporter
RDF_1740	0.44	0.42	Glycine transporter
RDF_1742		9.44	Xfp protein
RDF_1743		7.41	L-ascorbate-6-phosphatase
RDF_1747		2.76	hypothetical protein
RDF_1752		4.52	Transaldolase
RDF_1753		5.27	L-ribulose-5-phosphate-4-epimerase
RDF_1754		6.38	L-xyulose-5-phosphate-3-epimerase
RDF_1755		10.50	3.19 3-keto-L-gulonate 6-phosphate decarboxylase CDS
RDF_1756		11.20	Ascorbate-specific PTS system, EIIA component
RDF_1757		10.62	Ascorbate-specific PTS system, EIIB component
RDF_1758		19.35	2.26 Ascorbate-specific PTS system
RDF_1762		3.46	PfoR
RDF_1764		2.55	Mobile element protein
RDF_1774			2.43 Transcriptional regulator CtsR
RDF_1790		2.37	PTS system, N-acetylgalactosamine-specific IID component
RDF_1792		2.56	PTS system, N-acetylgalactosamine-specific IIB component
RDF_1794		2.58	PTS system, N-acetylgalactosamine- and galactosamine-specific IIA component
RDF_1795		4.14	2-deoxy-D-gluconate 3-dehydrogenase
RDF_1796		3.88	predicted 4-deoxy-L-threo-5-hexosulose-uronate ketol-isomerase
RDF_1797		3.68	2-dehydro-3-deoxygluconate kinase
RDF_1798		3.64	Eda-2
RDF_1820		2.00	LacC
RDF_1840		3.54	PTS system, mannose-specific IID component
RDF_1841		4.58	PTS system, mannose-specific IIC component
RDF_1842		4.19	PTS system, mannose/fructose family IIB component
RDF_1855	11.89		Exodeoxyribonuclease III



RDF_1856	14.41	2.64			Maltose-glucose PTS
RDF_1881		2.17			hypothetical protein
RDF_1887	5.72				ESAT-6/Esx family secreted protein
RDF_1925	3.67	3.21		3.09	BibA protein
RDF_1977			3.37		hypothetical protein
RDF_1978			3.26		FtsK/SpoIIIE family protein
RDF_1979			2.98		hypothetical protein CDS
RDF_1980			4.65		Replication protein repB
RDF_1992		2.14			Two-component response regulator
RDF_1998		0.46			hypothetical protein
RDF_1999		0.40		0.49	Phage infection protein
RDF_2006			2.33		hypothetical protein
RDF_2013			4.27		hypothetical protein
RDF_2017	36.64	43.77	28.29	31.36	Phage anti-repressor protein
RDF_2033	0.35	0.27			Immuodominant antigen A protein
RDF_2034	0.21	0.12			Aggregation promoting factor
RDF_2049		6.28			arcA
RDF_2050		5.92			Acetyltransferase
RDF_2051		5.61			argF-2
RDF_2052		6.61			arcD
RDF_2053		6.47			arcC-2

<sup>a</sup>Numbers indicate fold change in transcript level relative to strain SGBS102.

<sup>b</sup>Absence of number indicates transcript level was not significantly different for that particular gene in indicated strain relative to strain SGBS102.