

## Supplemental Materials and Methods

All references given correspond to the accompanying publication.

**Bacterial Strain Samples.** The 344 invasive serotype M3 strains (Table S1) that are the basis of this investigation were recovered between January 1992 and December 2007, from a prospective population-based surveillance study of invasive GAS infections conducted in Ontario, Canada (population = 12.2 million in 2006 census). This study has been described in several publications (11-13). The sequence of serotype M3 strain MGAS315, recovered in the late 1980s from a patient with invasive disease, was used as a reference for genome analyses (NCBI ref. seq. NC004070) (14).

**Subclone Assignment.** PCR amplification, gel electrophoresis, and conventional Sanger DNA sequencing was used to determine *emm3* allele, prophage content, and a SNP haplotype, using 15 informative bi-allelic loci distributed around the genome, for all 344 invasive strains. Primers, reagents, and reaction conditions used were as previously described (10). The matrix of these characteristics was used to parse the invasive strains into major subclone lineages providing an initial assessment of the population genetic structure prior to genome sequencing (Fig. 2). Conventional DNA sequencing was performed with fluorescent-terminator chemistry using an Applied Biosystems 3730XL instrument (ABI, Foster City, CA).

**Genome Sequencing and Polymorphism Discovery.** Genome sequence data were generated using the Illumina/Solexa Genome Analyzer System according to the manufacturer's protocols (Illumina, San Diego, CA). Single-end reads were generated for all strains and each strain was sequenced individually in a single flow-cell lane. To avoid sampling bias and assure proportionate representation of each epidemic, we selected every fourth strain in the sequential temporal strain collection dating from January 1992. In the aggregate, 510 million reads constituting 18,360 Mb of high quality sequence data were generated. On average per strain, 5.84 million reads (range: 1.0-to-12.3 x 10<sup>6</sup> reads) or 110 equivalents (range: 18.8-to-232.5 equivalents) of the reference serotype M3 strain MGAS315 genome (1.9 Mbp) were generated. For each strain, ~75% of the reads aligned to the reference sequence providing on average a fairly uniform ~80-fold depth of coverage across the genome. Sequencing reads were mapped and polymorphisms (SNPs and indels) were called relative to the unique portion of the strain MGAS315 reference genome using Variant Ascertainment Algorithm (VAAL) software (16). An inherent limitation of the use of short read length sequence data for polymorphism discovery is that polymorphisms can be accurately identified only relative to regions of unique non-repetitive sequence. Consistent with this only polymorphism identified relative to (that is contained in reads mapping to) the unique endogenous GAS core genome were used herein for genetic analyses. Polymorphisms identified relative to repetitive exogenous sequences, such as prophages, were excluded from the genetic analyses for two reasons: 1<sup>st</sup>) the sequence reads containing them can not be unambiguously uniquely mapped, and 2<sup>nd</sup>) prophages being mobile genetic elements can be transmitted horizontally and therefore polymorphism associated with them may not reflect relationships due to vertical

inheritance. For purposes of analysis and interpretation, we included genome data for eight additional strains from Ontario invasive sample previously obtained by comparative microarray hybridization based sequencing (17). Thus polymorphism discovery and subsequent population genomic analyses are based on genome sequence data for 95 invasive serotype M3 strains from Ontario.

**Mass Spectroscopy SNP Analysis.** One-third ( $n = 280$  of 801, 35%) of the bi-allelic core SNPs identified by full-genome sequencing was analyzed in all 344 invasive *emm3* strains studied and reference strain MGAS315. This random subset of SNP loci was distributed throughout the core genome and reflected the distribution of unique versus informative and coding versus intergenic SNPs of the whole set. Polymorphic loci were assessed by a mass-spectrometry-based method (18, 19) using the iPLEX<sup>®</sup> Gold assay and MassARRAY<sup>®</sup> System according to the manufacturer's instructions (Sequenom, San Diego, CA). Briefly, primers for multiplex PCR amplification, and SNP interrogation by single-base extension reactions were designed using Assay Designer software (Sequenom, Inc.) (Table S8). Templates for extension reactions were amplified by PCR from GAS chromosomal DNA and treated with shrimp alkaline phosphatase. Single-base extension reactions were performed using the extension primers and mass-modified nucleotides at an average multiplex level of 18.1. The resulting extension products were desalted and transferred onto 384 SpectroCHIP<sup>®</sup> solid support (Sequenom, Inc.). Spectra of the extension products were acquired using a MassARRAY Compact Analyzer (Sequenom, Inc.). Genotypes were called (280 loci by 345 strains = 96,600 total calls) using SpectroTYPER V.4 software (Sequenom, Inc.).

**Phylogenetic Reconstruction and Clonal Complex Assignment.** Concatenated SNP and indel loci nucleotide sequences were aligned using ClustalX ([www.clustal.org](http://www.clustal.org)) (24). Phylogenetic relationships based on the aligned sequences were inferred using the Neighbor-Joining method implemented in SplitsTree v.4 ([www.splitstree.org](http://www.splitstree.org)) (25). Schematics of the resultant trees were made using DendroScope (26). Haplotypes were enumerated and genetic distances were calculated using MEGA ([www.megasoftware.net](http://www.megasoftware.net)) (27). Complexes of clonally related strains were systematically inferred by the single-locus-variant-grouping method of eBURST ([eburst.mlst.net](http://eburst.mlst.net)) (20). The concatenated sequence for the 280 core SNP loci (i.e. the SNP genotype for 280 loci = SG<sub>280</sub>) assayed in all 344 of the invasive isolates was parsed into 7 sets each composed of 40 loci, and sequence type numbers were assigned for each of the haplotypes determined for each of the seven concatenated SNP sequences (Fig. S3). This process resulted in a seven-number sequence for each strain, the data format needed by eBURST, analogous to that of standard multilocus sequence typing.

***ropB* Gene Sequencing.** Gene sequencing was done using PCR primers specific for *ropB* (also known as *rgg*), *ropbF*-atcgcatctggctactgacc and *ropbR*-gccctggagctgttgagata. Conventional sequencing was done using an Applied Biosystems 3730XL instrument.

**Statistical Analyses.** As an indice of positive selection, genes having more SNPs than expected for a random distribution were identified by Chi-square test. The expected numbers of SNPs was calculated for all 1549 core genes based on the size of the gene

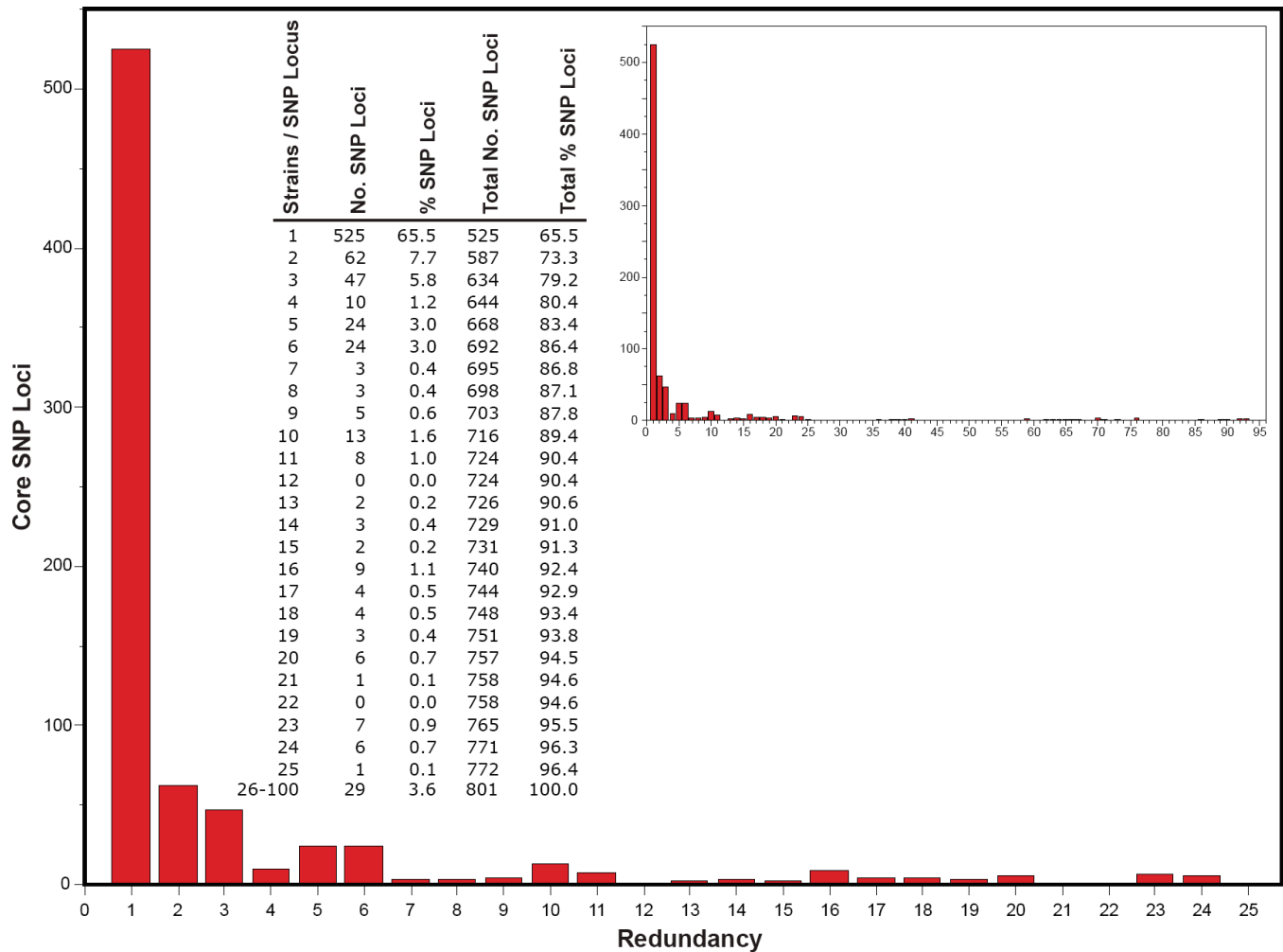
relative to the core genome, and the total number of SNP loci (Table S6). Probabilities were stringently and liberally corrected for multiple testing, using the Bonferroni, and Benjamini-Hochberg methods, respectively.

As an indice of potential bacterial genetic content influencing clinical disease manifestation, association between strain genotype and infection phenotype was assessed for all 21 of the clonal complexes using Fisher's exact test (Table S7).

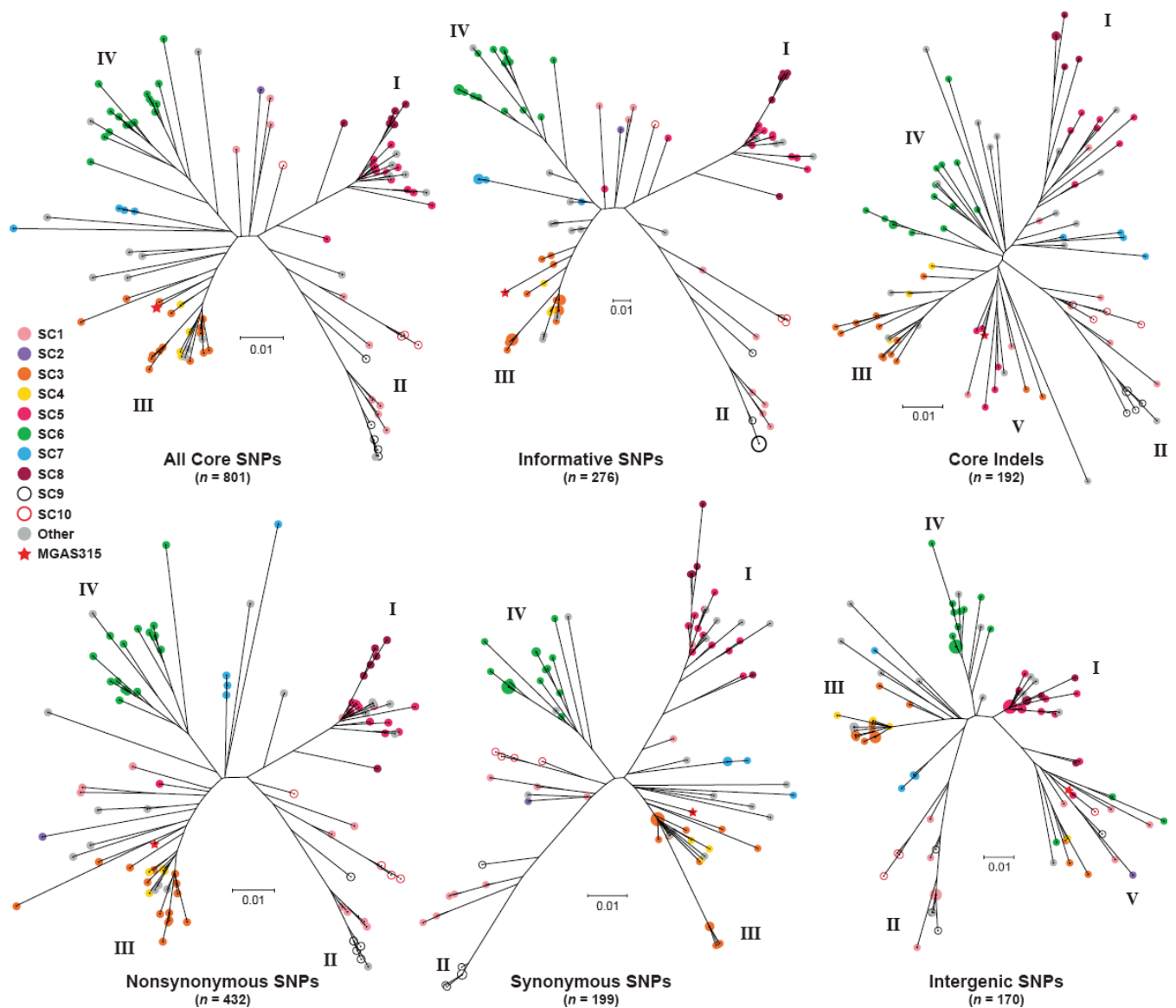
As an indice of phylogeographic structure, the correlation between genetic distance based on nucleotide differences and spatial distance based on postal code was assessed among the 95 sequenced strains using the nonparametric two-tailed Spearman's test. Similarly case-to-case geographic distance, intra- and interclonal complex, was assessed using the nonparametric two-tailed Mann-Whitney test. Various other distribution statistics (means, standard deviations, confidence intervals, etc.) were calculated using Prism ([www.graphpad.com](http://www.graphpad.com)).

**Comparative Expression Microarray Analysis.** Global transcriptional analysis was conducted for four strains representative of four phylogenetically divergent *emm3* clones using a custom GeneChip (Affymetrix, Santa Clara, CA) by methods described previously (28). GAS cells were cultured at 37° C in atmosphere supplemented with 5% CO<sub>2</sub> in Todd Hewitt medium with 0.2% w/v yeast extract. Culture growth was assessed by measuring optical density at 600 nm and by quantitative plating CFU enumeration (Fig S5). Bacteria from three replicate cultures were harvested at mid-exponential (OD<sub>600</sub> = 0.45) and stationary (4 hr post OD<sub>600</sub> = 0.45) phases of growth. Bacteria were lysed using the FastPrep system (MP Biomedicals) and RNA was isolated using the RNeasy kit

(QIAGEN, Valencia, CA). DNA was digested by Turbon DNase treatment (Ambion, Austin, TX). RNA concentration and purity was assessed with a 2100 Bioanalyzer (Agilent, Palo Alto, CA). RNA was reverse transcribed and the resultant cDNA was hybridized to microarray chips according to the manufacturer's protocols (Affymetrix). Hybridized chips were scanned, and image files were processed using GCOS software. Genes significantly differently expressed under the conditions compared were determined by ANOVA analysis with probabilities corrected for multiple testing by the method of Bonferroni. Heat maps were generated with CHIPST2C software ([www.chipST2c.org](http://www.chipST2c.org)).

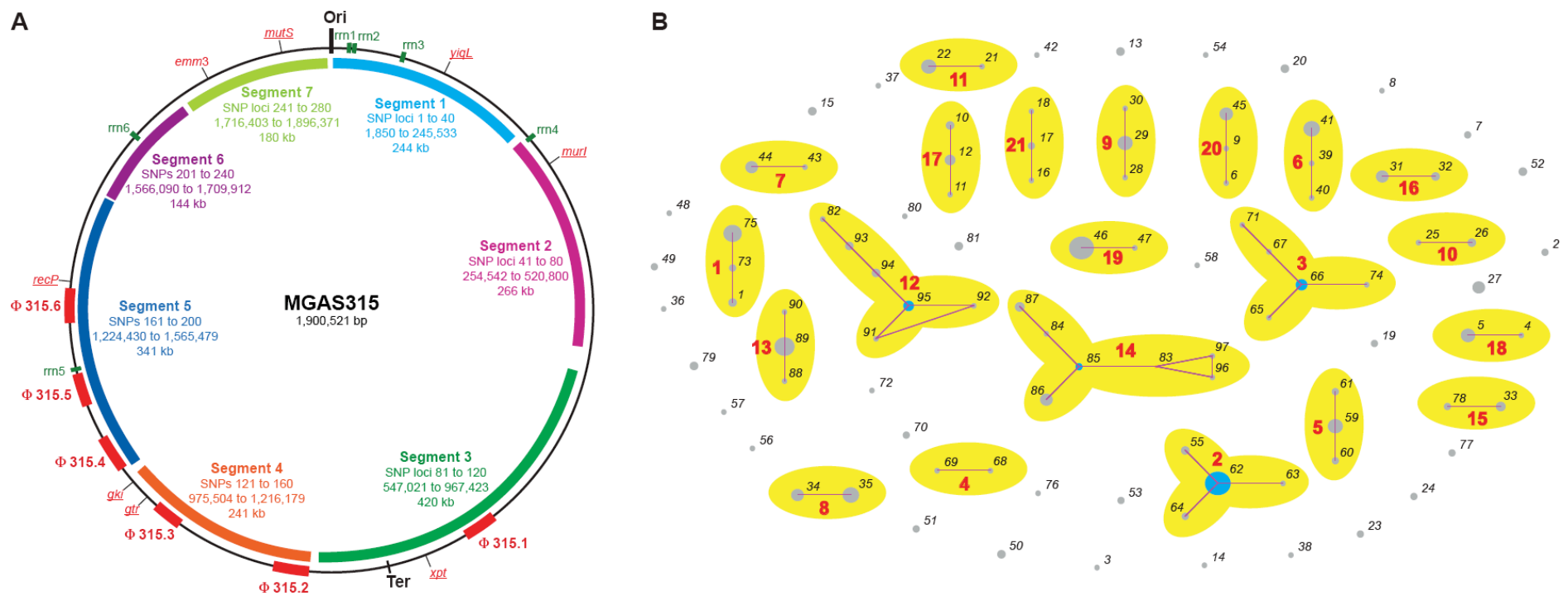


**Fig. S1.** Distribution of SNPs among the 95 sequenced serotype M3 genomes. As a population 65% of SNP loci (525/801) and 10% of the SNPs (525/5243) are strain specific. The vast majority of SNP loci are present in only one or just a few strains. Individually each strain has on average ~6 unique SNPs (range 0 to 47 SNPs), and ~41 informative SNPs (range 10 to 70 SNPs) in the core genome relative to the others. Inset graph shows the complete histogram of SNP loci among the 95 sequenced strains. Inset table lists the SNP loci distribution.

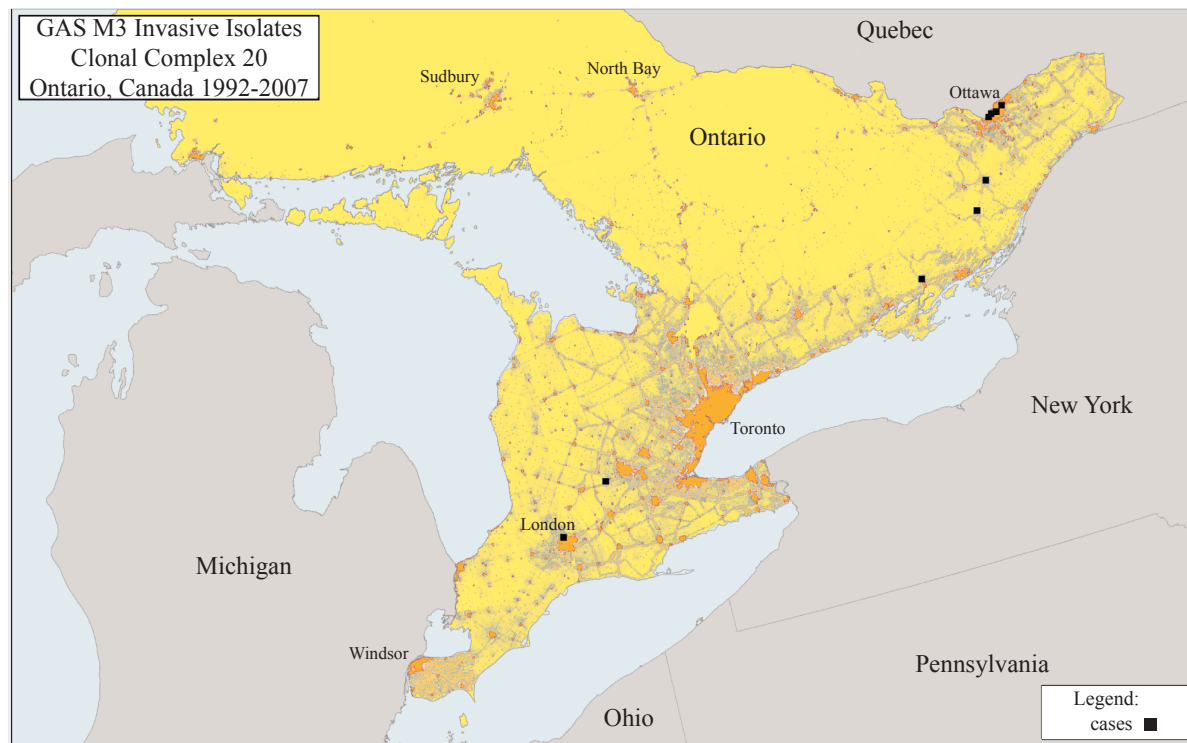
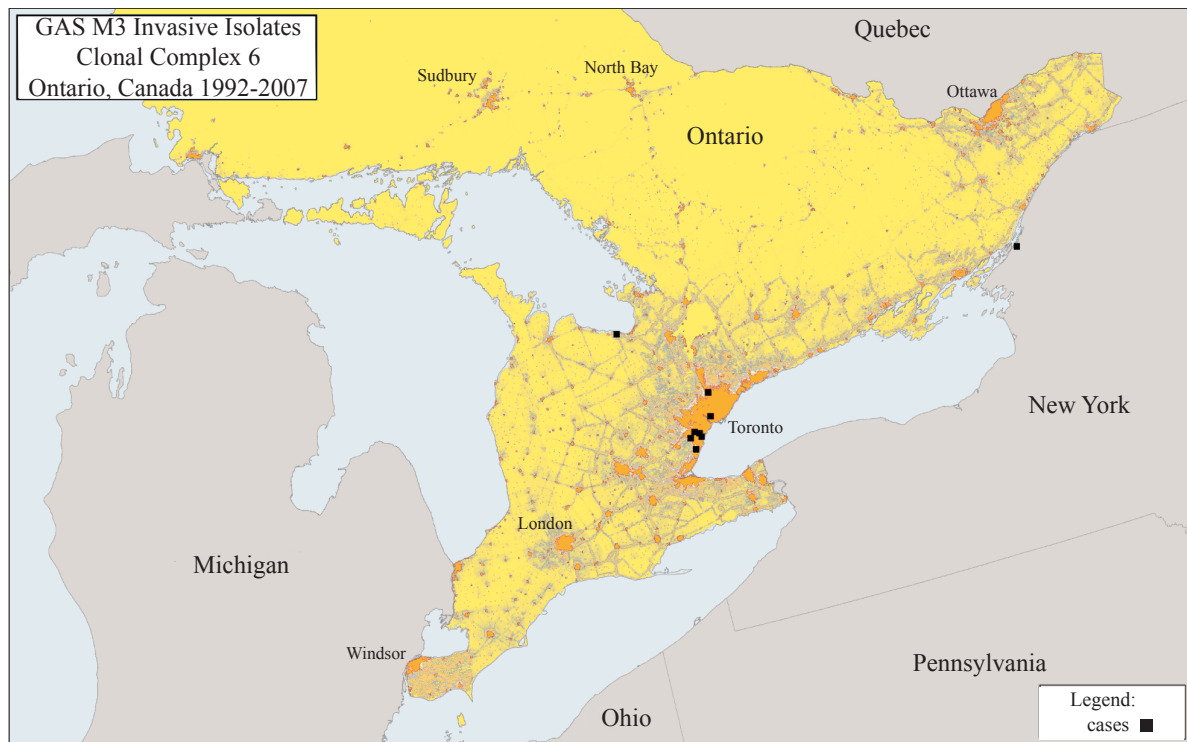


**Fig. S2.** Relationships among serotype M3 strains. Shown for the sequenced strains are unrooted neighbor-joining trees generated using different subsets of the core polymorphisms as indicated. Nodes of the trees are color coded by subclone as shown in the legend. Each tree has a topology with four major branches, a SC5-SC8 branch (I), a SC1-SC9-SC10 branch (II), a SC3-SC4 branch (III), and a SC6 branch (IV). In addition to these four branches, the trees based on intergenic SNPs and indels have a fifth branch (V) that is composed of a mixture of strains coming from the other four main branches. The intermixed composition of the fifth branch is consistent with a subset of the intergenic SNPs and indels being in linkage equilibrium relative to the other polymorphisms. This may indicate convergence for some intergenic/noncoding sequence polymorphisms, and in the case of indels relative to SNPs, might also reflect a different underlying mutational process at work, such as slip-strand mispairing.

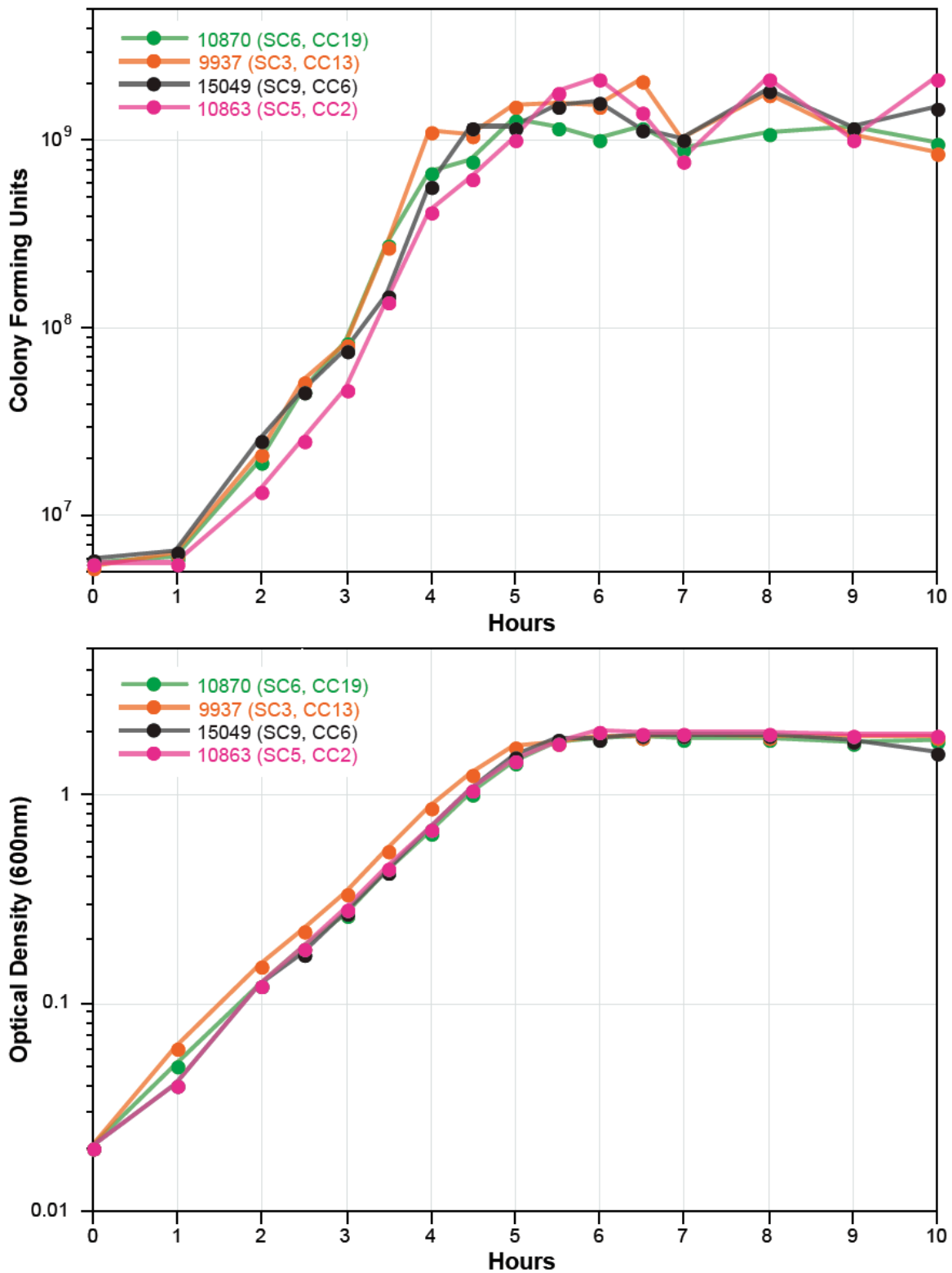




**Fig S3.** Systematic assignment of strains into predicted complexes of clonally related haplotypes. **(A)** Illustrated is the portion of the MGAS315 core genome interrogated by parsing the 280 sequence-validated SNPs into seven equal sets of 40 consecutive SNPs each. This parsing was conducted for all 344 of the epidemic isolates to generate a seven number sequence type analogous to that of standard GAS MLST. In eBURST MLST parlance, allelic profiles were determined for each of the 40 SNP segments and then the allele numbers for the segments of each strain were combined to produce a seven number sequence type. In comparison the portions of the genome interrogated by the seven housekeeping genes used for standard GAS MLST are shown in red and underlined. Given as landmarks on the genome, are the six ribosomal RNA operons (*rrn1-6*) and the six prophages ( $\phi 315.1-6$ ). **(B)** Shown schematically for the 344 epidemic isolates and MGAS315 are the 97 haplotypes defined by the 280 sequence-validated SNPs. Haplotypes are illustrated as circles that are proportional in size to the number of strains represented. Shown in yellow are the 21 clonal complexes of closely related haplotypes determined by the single locus variant grouping method of eBURST. Lines link related haplotypes, and primary founder haplotypes of the complexes are shown in blue if a founder was predicted. Strains encompassed by the 21 clonal complexes account for 80% (275/344) of the invasive isolate study set.



**Fig. S4.** Geographic distributions for strains of two of the 21 clonal complexes. Population density is shown in a gradient from yellow/low to orange/high. Most of the CC6 strains (top panel) are clustered in the Toronto, Ontario area. In contrast many of the CC20 strains (bottom panel) are clustered in Ottawa but none are in Toronto.



**Fig. S5.** Growth curves. Optical density measurements at 600 nm and CFU enumeration by quantitative plating were used to determine growth. RNA for expression microarray analysis was isolated from three replicate cultures harvested at two time points, mid-exponential phase ( $OD_{600} = 0.45$ ) and stationary phase (4 hours post mid-exponential harvest). Growth was virtually identical for all four strains.

**Table S1. Ontario GAS serotype M3 invasive isolate characteristics.**

No.	Strain (MGAS)	Year	Infection Type	<i>emm3</i> Allele	Phage Genotype	SNP <sub>15</sub> Haplotype	Subclone	SNP <sub>280</sub> Haplotype	Clonal Complex
1	3664	1992	lower respiratory	3.1	3.01	7		50	
2	10733	1992	osteomyolitis	3.1	3.01	7		50	
3	3371	1992	necrotizing fasciitis	3.1	3.01	1	1	20	
4	3671	1992	arthritis	3.1	3.01	1	1	27	
5	3370	1992	lower respiratory	3.37	3.01	1		19	
6	3672	1992	osteomyolitis	3.17	3.08	6		24	
7	3673	1992	arthritis	3.17	3.08	6		24	
8	3372	1992	lower respiratory	3.1	3.03	1	2	25	10
9	3373	1992	soft tissue	3.1	3.01	1	1	20	
10	3675	1993	necrotizing fasciitis	3.1	3.03	2	4	86	14
11	3669	1993	bacteremia	3.1	3.01	1	1	22	11
12	3676	1993	soft tissue	3.1	3.09	3		16	21
13	10051	1993	soft tissue	3.1	3.03	2	4	86	14
14	3677	1993	arthritis	3.1	3.03	2	4	86	14
15	3678	1993	lower respiratory	3.1	3.03	2	4	86	14
16	9999	1994	lower respiratory	3.32	3.01	2		77	
17	3374	1994	lower respiratory	3.1	3.01	2	3	94	12
18	10000	1994	bacteremia	3.18	3.01	5		51	
19	10735	1994	soft tissue	3.1	3.03	2	4	84	14
20	3375	1994	lower respiratory	3.20	3.01	1		--	
21	3376	1994	bacteremia	3.1	3.01	2	3	82	12
22	3378	1994	necrotizing fasciitis	3.1	3.03	2	4	86	14
23	10008	1994	soft tissue	3.1	3.01	2	3	94	12
24	10011	1994	soft tissue	3.1	3.01	2	3	95	12
25	10738	1994	bacteremia	3.1	3.07	2		92	12
26	3379	1994	lower respiratory	3.1	3.01	1	1	27	
27	3380	1994	soft tissue	3.1	3.01	2	3	91	12
28	10001	1994	lower respiratory	3.1	3.01	1	1	22	11
29	3382	1994	necrotizing fasciitis	3.1	3.01	2	3	95	12
30	10002	1994	lower respiratory	3.1	3.01	2	3	94	12
31	10009	1994	necrotizing fasciitis	3.1	3.01	2	3	95	12
32	10012	1994	soft tissue	3.1	3.03	1	2	29	9
33	10003	1994	arthritis	3.1	3.01	7		50	
34	10004	1994	lower respiratory	3.19	3.03	2		85	14
35	10005	1994	lower respiratory	3.1	3.03	1	2	29	9
36	10010	1994	soft tissue	3.1	3.01	1	1	22	11
37	10006	1994	soft tissue	3.18	3.07	5		51	
38	10737	1994	soft tissue	3.20	3.01	1		19	
39	3383	1994	necrotizing fasciitis	3.1	3.01	2	3	79	
40	3384	1994	necrotizing fasciitis	3.1	3.01	2	3	--	
41	3385	1995	arthritis	3.1	3.01	4		23	
42	3387	1995	lower respiratory	3.1	3.01	1	1	27	
43	3424	1995	bacteremia	3.1	3.03	1	2	29	9
44	3429	1995	soft tissue	3.1	3.01	1	1	27	
45	10025	1995	arthritis	3.1	3.01	2	3	95	12
46	3390	1995	necrotizing fasciitis	3.1	3.01	1	1	22	11
47	3391	1995	lower respiratory	3.1	3.03	1	2	29	9
48	3452	1995	bacteremia	3.1	3.01	1	1	22	11
49	10399	1995	necrotizing fasciitis	3.1	3.01	2	3	79	
50	3389	1995	necrotizing fasciitis	3.1	3.01	1	1	27	
51	3392	1995	necrotizing fasciitis	3.1	3.01	1	1	21	11
52	3393	1995	necrotizing fasciitis	3.1	3.01	2	3	79	

Table S1 cont.

No.	Strain (MGAS)	Year	Infection Type	<i>emm3</i> Allele	Phage Genotype	SNP <sub>15</sub> Haplotype	Subclone	SNP <sub>280</sub> Haplotype	Clonal Complex
53	3394	1995	necrotizing fasciitis	3.1	3.03	1	2	28	9
54	3455	1995	bacteremia	3.1	3.03	1	2	29	9
55	3465	1995	bacteremia	3.1	3.01	1	1	20	
56	3475	1995	bacteremia	3.1	3.01	2	3	81	
57	3484	1995	soft tissue	3.1	3.03	1	2	30	9
58	3485	1995	soft tissue	3.1	3.01	5		52	
59	10739	1995	necrotizing fasciitis	3.1	3.03	1	2	29	9
60	3395	1995	necrotizing fasciitis	3.1	3.03	2	4	86	14
61	3396	1995	necrotizing fasciitis	3.1	3.03	1	2	29	9
62	3397	1995	soft tissue	3.1	3.01	5		52	
63	3398	1995	necrotizing fasciitis	3.1	3.03	1	2	29	9
64	3399	1995	necrotizing fasciitis	3.1	3.01	1	1	27	
65	10021	1995	soft tissue	3.1	3.01	2	3	80	
66	10022	1995	bacteremia	3.1	3.01	1	1	22	11
67	10023	1995	bacteremia	3.1	3.01	2	3	93	12
68	10740	1995	osteomyolitis	3.1	3.03	1	2	29	9
69	10024	1995	arthritis	3.1	3.03	2	4	86	14
70	10026	1995	peritonitis	3.1	3.01	1	1	22	11
71	10027	1995	gynecological	3.21	3.01	5		52	
72	10028	1995	meningitis	3.1	3.01	1	1	22	11
73	10029	1995	necrotizing fasciitis	3.1	3.01	2	3	93	12
74	10030	1995	upper respiratory	3.1	3.04	2		87	14
75	10031	1995	bacteremia	3.1	3.03	1	2	29	9
76	10741	1995	lower respiratory	3.22	3.01	1		27	
77	10742	1995	arthritis	3.1	3.03	2	4	97	14
78	10034	1995	lower respiratory	3.1	3.03	2	4	96	14
79	10035	1996	necrotizing fasciitis	3.1	3.01	2	3	95	12
80	10037	1996	lower respiratory	3.1	3.03	2	4	83	14
81	10743	1996	lower respiratory	3.1	3.03	1	2	26	10
82	10038	1996	bacteremia	3.1	3.01	1	1	22	11
83	10744	1996	lower respiratory	3.1	3.04	2		87	14
84	10040	1996	surgical	3.1	3.01	2	3	33	15
85	10041	1996	necrotizing fasciitis	3.1	3.04	2		87	14
86	10042	1996	peritonitis	3.1	3.01	2	3	81	
87	10043	1996	bacteremia	3.1	3.01	2	3	81	
88	10044	1996	necrotizing fasciitis	3.1	3.04	2		87	14
89	10400	1996	upper respiratory	3.1	3.01	4	7	23	
90	10045	1996	bacteremia	3.1	3.01	2	3	33	15
91	10046	1996		3.1	3.03	2	4	85	14
92	10047	1996	soft tissue	3.1	3.01	2	3	90	13
93	10048	1996	bacteremia	3.1	3.01	2	3	33	15
94	10049	1997	pharyngeal	3.1	3.01	1	1	22	11
95	10050	1997	necrotizing fasciitis	3.23	3.03	1	2	26	10
96	9977	1998	lower respiratory	3.1	3.03	1	2	26	10
97	9978	1998	arthritis	3.1	3.01	9		37	
98	9979	1998	arthritis	3.1	3.01	2	3	33	15
99	9980	1999	upper respiratory	3.1	3.05	3		10	17
100	9981	1999	lower respiratory	3.1	3.03	2	4	93	12
101	9982	1999	necrotizing fasciitis	3.1	3.01	2	3	89	13
102	9983	1999	necrotizing fasciitis	3.1	3.02	3	6	45	20
103	9984	1999	necrotizing fasciitis	3.1	3.01	2	3	89	13
104	9985	1999	soft tissue	3.1	3.02	3	6	5	18
105	9986	1999	soft tissue	3.1	3.02	3	6	6	20

Table S1 cont.

No.	Strain (MGAS)	Year	Infection Type	<i>emm3</i> Allele	Phage Genotype	SNP <sub>15</sub> Haplotype	Subclone	SNP <sub>280</sub> Haplotype	Clonal Complex
106	9987	1999	soft tissue	3.1	3.01	2	3	88	13
107	9988	1999	soft tissue	3.1	3.02	3	6	10	17
108	9989	1999	arthritis	3.1	3.01	2	3	89	13
109	9990	1999	bacteremia	3.1	3.01	1	1	62	2
110	9878	2000	soft tissue	3.1	3.02	3	6	46	19
111	9879	2000	necrotizing fasciitis	3.1	3.01	2	3	89	13
112	9880	2000	necrotizing fasciitis	3.1	3.01	2	3	89	13
113	9881	2000	arthritis	3.1	3.01	1	1	43	7
114	9882	2000	soft tissue	3.2	3.01	1	5	62	2
115	9883	2000	lower respiratory	3.1	3.02	3	6	46	19
116	9884	2000	upper respiratory	3.31	3.01	1		73	1
117	9885	2000	upper respiratory	3.2	3.01	1	5	75	1
118	9886	2000	upper respiratory	3.1	3.02	3	6	5	18
119	9887	2000	bacteremia	3.2	3.01	1	5	64	2
120	9888	2000	soft tissue	3.1	3.02	3	6	46	19
121	9889	2000	lower respiratory	3.1	3.02	3	6	45	20
122	9890	2000	bacteremia	3.1	3.01	1	1	44	7
123	9892	2000	soft tissue	3.1	3.02	3	6	45	20
124	9893	2000	meningitis	3.2	3.01	1	5	72	
125	9894	2000	necrotizing fasciitis	3.1	3.01	1	1	44	7
126	9895	2000	osteomyolitis	3.1	3.05	3		5	18
127	9896	2000	bacteremia	3.1	3.02	3	6	46	19
128	9897	2000	lower respiratory	3.1	3.01	2	3	89	13
129	9898	2000	peritonitis	3.1	3.02	3	6	45	20
130	9953	2000	lower respiratory	3.1	3.01	4	7	31	16
131	10745	2000	soft tissue	3.30	3.02	10		45	20
132	9899	2000	necrotizing fasciitis	3.1	3.02	3	6	46	19
133	9900	2000	arthritis	3.1	3.02	3	6	46	19
134	9901	2000	arthritis	3.1	3.02	3	6	9	20
135	9902	2000	upper respiratory	3.1	3.01	2	3	89	13
136	9903	2000	lower respiratory	3.1	3.02	3	6	46	19
137	9904	2000	arthritis	3.1	3.02	3	6	45	20
138	9905	2000	soft tissue	3.1	3.02	3	6	46	19
139	9906	2000	upper respiratory	3.1	3.01	2	3	89	13
140	9907	2000	arthritis	3.2	3.01	1	5	62	2
141	9908	2000	bacteremia	3.24	3.01	1	5	56	
142	9909	2000	lower respiratory	3.1	3.01	2	3	89	13
143	9910	2000	arthritis	3.31	3.01	1		73	1
144	9911	2000	necrotizing fasciitis	3.1	3.01	1	1	1	1
145	9927	2000	soft tissue	3.1	3.01	2	3	89	13
146	9967	2000	necrotizing fasciitis	3.1	3.01	1	1	44	7
147	10396	2000	soft tissue	3.2	3.01	1	5	62	2
148	9254	2000	peritonitis	3.2	3.01	1	5	75	1
149	9912	2000	lower respiratory	3.1	3.01	2	3	89	13
150	9913	2000	soft tissue	3.1	3.06	3		46	19
151	9914	2000	soft tissue	3.1	3.01	1	1	62	2
152	9915	2000	bacteremia	3.1	3.02	3	6	10	17
153	9916	2000	soft tissue	3.1	3.02	3	6	46	19
154	9917	2000	soft tissue	3.1	3.05	3		4	18
155	9918	2000	soft tissue	3.1	3.02	3	6	13	
156	9919	2000	necrotizing fasciitis	3.1	3.02	3	6	47	19
157	9921	2000	soft tissue	3.1	3.01	1	1	44	7
158	9922	2000	soft tissue	3.2	3.01	1	5	62	2

Table S1 cont.

No.	Strain (MGAS)	Year	Infection Type	<i>emm3</i> Allele	Phage Genotype	SNP <sub>15</sub> Haplotype	Subclone	SNP <sub>280</sub> Haplotype	Clonal Complex
159	9923	2000	soft tissue	3.1	3.02	3	6	5	18
160	9924	2000	peritonitis	3.2	3.01	1	5	62	2
161	9925	2000	peritonitis	3.1	3.02	3	6	46	19
162	9926	2000	arthritis	3.1	3.01	2	3	89	13
163	9928	2000	lower respiratory	3.2	3.01	1	5	62	2
164	9929	2000	lower respiratory	3.2	3.01	1	5	75	1
165	9930	2000	necrotizing fasciitis	3.1	3.02	3	6	46	19
166	9931	2000	bacteremia	3.1	3.01	2	3	89	13
167	9932	2000	bacteremia	3.2	3.01	1	5	62	2
168	9933	2000	soft tissue	3.1	3.02	3	6	46	19
169	9934	2000	soft tissue	3.1	3.01	2	3	77	
170	9935	2000	surgical	3.1	3.02	6		45	20
171	9936	2000	lower respiratory	3.1	3.01	1	1	62	2
172	9951	2000	arthritis	3.2	3.01	1	5	75	1
173	9937	2000	soft tissue	3.1	3.01	2	3	89	13
174	9939	2000	upper respiratory	3.31	3.01	1		1	1
175	9941	2000	soft tissue	3.2	3.01	1	5		
176	9942	2000	soft tissue	3.1	3.02	3	6	5	18
177	9943	2000	upper respiratory	3.2	3.01	1	5	62	2
178	9944	2000	bacteremia	3.2	3.01	1	5	64	2
179	9945	2000	soft tissue	3.2	3.01	1	5	75	1
180	9947	2000	bacteremia	3.1	3.01	4	7	31	16
181	9948	2000	soft tissue	3.1	3.01	1	1	44	7
182	9954	2000	soft tissue	3.31	3.01	1			
183	9958	2000	upper respiratory	3.1	3.01	1	1	7	
184	10746	2000	upper respiratory	3.2	3.01	1	5	75	1
185	10747	2000	bacteremia	3.1	3.01	4	7	31	16
186	9949	2000	bacteremia	3.2	3.01	1	5	75	1
187	9952	2000	soft tissue	3.2	3.01	1	5	62	2
188	9955	2000	upper respiratory	3.1	3.02	3	6	45	20
189	9956	2000	bacteremia	3.2	3.01	1	5	62	2
190	9957	2000	arthritis	3.2	3.01	1	5	75	1
191	10749	2000	soft tissue	3.24	3.01	1		57	
192	9959	2000	peritonitis	3.1	3.01	4	7	31	16
193	9960	2000	lower respiratory	3.1	3.01	4	7	31	16
194	9961	2000	bacteremia	3.2	3.01	1	5	75	1
195	9962	2000	lower respiratory	3.1	3.01	1	1	62	2
196	9963	2000	soft tissue	3.1	3.01	2	3	89	13
197	9964	2000	arthritis	3.2	3.01	1		8	
198	9965	2000	lower respiratory	3.2	3.01	1	5	62	2
199	9968	2000	soft tissue	3.1	3.01	2	3	89	13
200	10398	2000	adenitis	3.1	3.02	3	6	5	18
201	9966	2000	bacteremia	3.1	3.02	3	6	46	19
202	9969	2000	bacteremia	3.1	3.02	3	6	46	19
203	9971	2000	arthritis	3.1	3.01	1	1	44	7
204	9972	2000	bacteremia	3.1	3.01	2	3	89	13
205	9973	2000	soft tissue	3.1	3.06	6		14	
206	9974	2000	bacteremia	3.1	3.01	4	7	31	16
207	9970	2000	soft tissue	3.2	3.01	1	5	62	2
208	10829	2001	soft tissue	3.24	3.01	1	5	1	1
209	10830	2001	upper respiratory	3.24	3.01	1		53	
210	10831	2001	empyema	3.2	3.01	1	5	69	4
211	10832	2001	arthritis	3.1	3.02	3	6	46	19

Table S1 cont.

No.	Strain (MGAS)	Year	Infection Type	<i>emm3</i> Allele	Phage Genotype	SNP <sub>15</sub> Haplotype	Subclone	SNP <sub>280</sub> Haplotype	Clonal Complex
212	10833	2001	lower respiratory	3.2	3.01	1	5	62	2
213	10834	2001	soft tissue	3.2	3.01	8		70	
214	10835	2001	bacteremia	3.2	3.01	1	5	75	1
215	10836	2001	soft tissue	3.1	3.02	3	6	46	19
216	10837	2001	bacteremia	3.1	3.05	3		46	19
217	10838	2001	soft tissue	3.1	3.01	1	1	7	
218	10839	2001	lower respiratory	3.2	3.01	1	5	53	
219	10840	2001	meningitis	3.1	3.02	3	6	12	17
220	10841	2001	bacteremia	3.2	3.01	1	5	62	2
221	10842	2001	arthritis	3.1	3.01	2	3	89	13
222	10843	2001	soft tissue	3.1	3.02	3	6	5	18
223	10844	2001	necrotizing fasciitis	3.1	3.02	3	6	46	19
224	10845	2001	necrotizing fasciitis	3.1	3.02	3	6	46	19
225	10846	2001	lower respiratory	3.1	3.01	4	7	32	16
226	10847	2001	arthritis	3.1	3.02	3	6	46	19
227	10848	2001	bacteremia	3.1	3.01	4	7	32	16
228	10849	2001	lower respiratory	3.25	3.01	4		32	16
229	10850	2001	soft tissue	3.2	3.01	1	5	62	2
230	10851	2001	lower respiratory	3.2	3.01	1	5	62	2
231	10852	2001	arthritis	3.2	3.01	1	5	62	2
232	10853	2001	upper respiratory	3.2	3.01	1	5	62	2
233	10854	2001	lower respiratory	3.1	3.01	4	7	31	16
234	10855	2001	empyema	3.1	3.02	3	6	5	18
235	10856	2001	soft tissue	3.26	3.01	1		62	2
236	10857	2001	bacteremia	3.2	3.01	1	5	62	2
237	10858	2001	peripartum	3.1	3.02	3	6	46	19
238	10859	2001	necrotizing fasciitis	3.2	3.01	1	5	70	
239	10860	2001	necrotizing fasciitis	3.1	3.02	3	6	46	19
240	10861	2001	arthritis	3.1	3.01	1	1	44	7
241	10863	2001	arthritis	3.2	3.01	1	5	63	2
242	10862	2001	arthritis	3.2	3.01	1	5	75	1
243	10864	2001	soft tissue	3.1	3.02	3	6	46	19
244	10865	2001	arthritis	3.1	3.02	3	6	11	17
245	10866	2001	lower respiratory	3.2	3.01	1	5	75	1
246	10867	2001	soft tissue	3.1	3.02	3	6	13	
247	10868	2001	soft tissue	3.2	3.01	1	5	75	1
248	10869	2001	arthritis	3.1	3.02	3	6	5	18
249	10870	2002	soft tissue	3.1	3.02	3	6	46	19
250	10871	2002	lower respiratory	3.2	3.01	1	5	68	4
251	10872	2002	necrotizing fasciitis	3.1	3.02	3	6	12	17
252	10873	2002	arthritis	3.27	3.01	1		38	
253	10874	2002	bacteremia	3.2	3.01	1	5	62	2
254	10875	2002	necrotizing fasciitis	3.1	3.01	1	1	42	
255	10876	2002	necrotizing fasciitis	3.29	3.02	3		12	17
256	15000	2003	renal PT	3.1	3.02	3	6	46	19
257	15001	2003	necrotizing fasciitis	3.2	3.01	1	5	76	
258	15004	2003	soft tissue	3.1	3.02	3	6	46	19
259	15002	2003	pneumonia	3.1	3.02	3	6	12	17
260	15003	2003	pyrexia	3.1	3.02	3	6	46	19
261	15005	2003	bacteremia	3.2	3.01	1	5	61	5
262	15006	2003	bacteremia	3.2	3.03	1		61	5
263	15008	2003	bacteremia	3.2	3.10	1	8	62	2
264	15007	2003	peritonitis	3.2	3.01	1	5	62	2



Table S1 cont.

No.	Strain (MGAS)	Year	Infection Type	<i>emm3</i> Allele	Phage Genotype	SNP <sub>15</sub> Haplotype	Subclone	SNP <sub>280</sub> Haplotype	Clonal Complex
265	15009	2004	bacteremia	3.2	3.10	1		55	2
266	15010	2004	fever R/O endoc	3.26	3.07	4		15	
267	15011	2004	bacteremia	3.2	3.01	1	5	62	2
268	15012	2004	arthritis	3.2	3.01	1	5	62	2
269	15013	2005	lower respiratory	3.2	3.01	1	5	59	5
270	15014	2005	soft tissue	3.1	3.01	1	1	59	5
271	15015	2005	soft tissue	3.35	3.01	1	9	41	6
272	15016	2005	lower respiratory	3.35	3.01	1	9	41	6
273	15017	2005	toxic shock	3.1	3.02	3	6	48	
274	15018	2005	soft tissue	3.2	3.01	1	5	59	5
275	15019	2005	bacteremia	3.1	3.07	1	10	35	8
276	15020	2005	necrotizing fasciitis	3.2	3.10	1	8	74	3
277	15021	2005	necrotizing fasciitis	3.2	3.10	1	8	66	3
278	15022	2005	lower respiratory	3.2	3.01	1	5	59	5
279	15023	2005	pericarditis	3.1	3.07	1	10	35	8
280	15024	2005	soft tissue	3.1	3.07	1	10	35	8
281	15025	2005	fever, urosepsis	3.26	3.07	4		15	
282	15026	2005		3.2	3.01	1	5	59	5
283	15027	2005	necrotizing fasciitis	3.2	3.07	1	8	58	
284	15028	2005	scarlet fever	3.35	3.01	1	9	41	6
285	15029	2005	soft tissue	3.2	3.10	1	8	65	3
286	15030	2005	bacteremia	3.35	3.01	1	9	41	6
287	15031	2005		3.35	3.01	1	9	41	6
288	15034	2005	fever, lymphangitis	3.35	3.01	1	9	41	6
289	15035	2005	soft tissue	3.1	3.07	1	10	35	8
290	15036	2005	lower respiratory	3.1	3.01	2	3	78	15
291	15037	2005	preterm labor	3.1	3.01	2	3	78	15
292	15038	2005	lower respiratory	3.1	3.02	3	6	49	
293	15040	2005	upper respiratory	3.2	3.01	1	5	59	5
294	15039	2005	lower respiratory	3.1	3.01	1	1	35	8
295	15041	2005		3.2	3.07	1	8	66	3
296	15042	2005		3.26	3.07	4		15	
297	15043	2005		3.35	3.01	1	9	39	6
298	15032	2005	abcess	3.2	3.01	1	5	59	5
299	15033	2005	inguinal	3.2	3.07	1	8	66	3
300	15044	2006	bacteremia	3.1	3.07	1	10	35	8
301	15045	2006	post operative	3.35	3.07	1		41	6
302	15046	2006	bacteremia	3.2	3.01	1	5	59	5
303	15047	2006	bacteremia	3.2	3.07	1	8	67	3
304	15048	2006	lower respiratory	3.1	3.01	1	1	34	8
305	15049	2006	lower respiratory	3.35	3.01	1	9	41	6
306	15050	2006	lower respiratory	3.1	3.01	1	1	36	
307	15051	2006	necrotizing fasciitis	3.1	3.07	1	10	35	8
308	15052	2006	soft tissue	3.35	3.01	1	9	41	6
309	15053	2006	arthritis	3.2	3.07	1	8	66	3
310	15054	2006	bacteremia	3.2	3.01	1	5	62	2
311	15055	2006	bacteremia	3.1	3.07	1	10	35	8
312	15056	2006	fascial cellulitis	3.1	3.07	1	10	34	8
313	15057	2006	lower respiratory	3.2	3.07	1	8	66	3
314	15058	2006	necrotizing fasciitis	3.1	3.02	3	6	3	
315	15059	2006	bacteremia	3.2	3.01	1	5	60	5
316	15060	2006	soft tissue	3.1	3.07	1	10	34	8
317	15061	2006		3.2	3.10	1		55	2

Table S1 cont.

No.	Strain (MGAS)	Year	Infection Type	<i>emm3</i> Allele	Phage Genotype	SNP <sub>15</sub> Haplotype	Subclone	SNP <sub>280</sub> Haplotype	Clonal Complex
318	15062	2006	peritonitis	3.35	3.01	1	9	40	6
319	15063	2006		3.1	3.07	1	10	34	8
320	15184	2006	necrotizing fasciitis	3.2	3.07	1	8	71	3
321	15064	2006	necrotizing fasciitis	3.35	3.01	1	9	41	6
322	15185	2006	soft tissue	3.1	3.07	1	10	35	8
323	15186	2006	necrotizing fasciitis	3.1	3.07	1	10	35	8
324	15187	2007		3.1	3.07	1	10	35	8
325	15188	2007		3.1	3.07	1	10	34	8
326	15189	2007	soft tissue	3.2	3.01	1	5	59	5
327	15190	2007	soft tissue	3.1	3.10	3		17	21
328	15191	2007	necrotizing fasciitis	3.2	3.01	1	5	60	5
329	15192	2007	arthritis	3.1	3.07	1	10	35	8
330	15193	2007	lower respiratory	3.1	3.07	1	10	34	8
331	15194	2007	necrotizing fasciitis	3.1	3.01	1	1	75	1
332	15195	2007		3.1	3.01	1	1	55	2
333	15196	2007	soft tissue	3.1	3.10	3		18	21
334	15197	2007	bacteremia	3.1	3.01	1	1	34	8
335	15198	2007	soft tissue	3.1	3.02	3	6	2	
336	15199	2007		3.1	3.01	1	1	75	1
337	15200	2007	soft tissue	3.35	3.01	1	9	41	6
338	15201	2007		3.1	3.02	1	6	2	
339	15202	2007		3.1	3.02	3	6	13	
340	15203	2007	upper respiratory	3.33	3.02	3		49	
341	15209	2007		3.2	3.01	1	5	59	5
342	15210	2007		3.1	3.02	3	6	12	17
343	15211	2007		3.35	3.01	1	9	41	6
344	15212	2007		3.2	3.07	1	8	66	3

**Table S2. Ontario GAS serotype M3 core single nucleotide polymorphisms.<sup>1</sup>**

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
1	1823	A	G	CDS	0002	DNA polymerase III beta subunit	Non	Ile_I	Val_V	Con	1
2	7975	C	T	IG	IG	IG	--	--	--	--	2
3	10114	G	T	CDS	0009	unknown function	Non	Val_V	Phe_F	Con	66
4	12658	G	C	CDS	0011	Hypoxanthine-guanine phosphoribosyltransferase	Non	Glu_E	Gln_Q	Con	1
5	12730	T	C	CDS	0011	Hypoxanthine-guanine phosphoribosyltransferase	Non	Tyr_Y	His_H	Con	1
6	12845	T	A	CDS	0012	Cell division protein ftsH	Non	Phe_F	Tyr_Y	Con	3
7	13671	C	T	CDS	0012	Cell division protein ftsH	Syn	Asp_D	Asp_D	--	1
8	16387	A	G	CDS	0013	Cationic amino acid transporter	Syn	Leu_L	Leu_L	--	21
9	16645	T	G	IG	IG	IG	--	--	--	--	1
10	28092	T	C	IG	IG	IG	--	--	--	--	1
11	35953	A	T	IG	IG	IG	--	--	--	--	20
12	37285	A	T	CDS	0020	Phosphoribosylformylglycinamide synthase	Non	Tyr_Y	Phe_F	Con	24
13	42490	G	A	CDS	0022	Phosphoribosylformylglycinamide cyclo-ligase	Syn	Leu_L	Leu_L	--	1
14	44364	T	C	CDS	0024	IMP cyclohydrolase	Syn	Asn_N	Asn_N	--	1
15	45232	G	A	CDS	0024	IMP cyclohydrolase	Non	Ala_A	Thr_T	Con	3
16	46653	C	A	CDS	0025	putative choline binding protein	Non	Ser_S	Ile_I	Rad	1
17	47892	G	A	CDS	0026	Phosphoribosylamine--glycine ligase	Syn	Lys_K	Lys_K	--	1
18	48112	A	C	CDS	0026	Phosphoribosylamine--glycine ligase	Non	Ile_I	Leu_L	Con	23
19	49430	G	A	CDS	0028	Phosphoribosylaminoimidazole carboxylase	Syn	Leu_L	Leu_L	--	1
20	61350	G	A	IG	IG	IG	--	--	--	--	1
21	61529	T	C	IG	IG	IG	--	--	--	--	1
22	64980	A	G	IG	IG	IG	--	--	--	--	1
23	66294	T	G	CDS	0041	LSU ribosomal protein L4p	Non	Ser_S	Ala_A	Con	65
24	67203	C	T	CDS	0043	LSU ribosomal protein L4p	Syn	Asn_N	Asn_N	--	1
25	67875	C	T	IG	IG	IG	--	--	--	--	1
26	76867	T	C	CDS	0062	Translation initiation factor 1	Syn	Ser_S	Ser_S	--	1
27	79640	C	T	IG	IG	IG	--	--	--	--	6
28	85069	T	C	IG	IG	IG	--	--	--	--	1
29	85868	A	C	IG	IG	IG	--	--	--	--	1
30	87320	G	A	CDS	--	4-PP-cytidyl-2-C-methyl-D-erythritol kinase	Non	Val_V	Ile_I	Con	1
31	93041	T	C	CDS	0074	Multimodular transpeptidase-transglycosylase	Syn	Tyr_Y	Tyr_Y	--	2
32	98549	A	G	CDS	0076	DNA-directed RNA polymerase beta' subunit	Non	Lys_K	Glu_E	Con	5
33	106561	G	T	CDS	0085	Adenine-specific methyltransferase	Non	Ser_S	Ile_I	Rad	1
34	106604	G	A	CDS	0085	Adenine-specific methyltransferase	Syn	Gln_Q	Gln_Q	--	1
35	106671	C	A	CDS	0085	Adenine-specific methyltransferase	Non	Pro_P	Thr_T	Con	1
36	107126	A	G	CDS	0086	Acetate kinase	Non	Gln_Q	Arg_R	Con	1
37	113828	C	T	CDS	0095	tRNA dihydrouridine synthase	Non	Glu_E	Lys_K	Con	1
38	116911	G	A	CDS	0097	Negative transcriptional regulator	Non	Ala_A	Val_V	Con	1
39	117327	C	T	IG	IG	IG	--	--	--	--	1
40	117536	C	T	IG	IG	IG	--	--	--	--	3
41	118018	C	T	CDS	0098	putative collagen binding protein	Non	Ala_A	Val_V	Con	11
42	119965	T	C	CDS	0099	Signal peptidase I	Syn	Thr_T	Thr_T	--	24
43	122194	C	T	CDS	0101	NPQTN specific sortase B	Non	His_H	Tyr_Y	Con	1
44	127932	T	A	IG	IG	IG	--	--	--	--	1
45	130696	T	C	IG	IG	IG	--	--	--	--	1
46	131210	T	C	CDS	0108	3-ketoacyl-CoA thiolase	Non	Ile_I	Thr_T	Con	9
47	131360	G	A	CDS	0108	3-ketoacyl-CoA thiolase	Non	Gly_G	Asp_D	Con	1
48	131636	C	T	CDS	0108	3-ketoacyl-CoA thiolase	Non	Ala_A	Val_V	Con	10
49	141908	C	T	CDS	0120	V-type ATP synthase subunit A	Non	Thr_T	Ile_I	Con	1
50	143708	C	A	CDS	0121	V-type ATP synthase subunit B	Non	Ala_A	Asp_D	Rad	1
51	144672	T	C	CDS	0122	V-type ATP synthase subunit D	Non	Tyr_Y	His_H	Con	2
52	146152	C	T	CDS	0123	Tellurite resistance protein	Non	Ala_A	Thr_T	Con	1
53	152083	G	A	CDS	0128	Nicotine adenine dinucleotide glycohydrolase	Non	Met_M	Ile_I	Con	1
54	155355	T	C	IG	IG	IG	--	--	--	--	2
55	156470	G	A	IG	IG	IG	--	--	--	--	9
56	157102	G	A	CDS	0133	Cystathionine beta-lyase	Non	Ala_A	Thr_T	Con	1
57	162062	C	T	CDS	0135	PTS system IIC component	Non	Ser_S	Phe_F	Rad	2
58	163387	G	A	CDS	0137	PTS system IIC component	Non	Glu_E	Lys_K	Con	10
59	164253	C	T	CDS	0139	L-xylulose 5-phosphate 3-epimerase	Non	Thr_T	Ile_I	Con	1
60	166685	C	T	CDS	0141	unknown function	Syn	Val_V	Val_V	--	1
61	169236	A	T	IG	IG	IG	--	--	--	--	1
62	170064	T	C	CDS	0143	Glycine betaine ABC transport system	Syn	Asn_N	Asn_N	--	93

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
63	179617	C	T	CDS	0151	Core component BioY of biotin ECF transporter	Syn	Tyr_Y	Tyr_Y	--	1
64	182684	C	T	CDS	0154	unknown function	Non	Val_V	Ile_I	Con	1
65	183592	G	A	CDS	0155	Streptococcal pyrogenic exotoxin G (SpeG)	Non	Ala_A	Thr_T	Con	1
66	184244	C	T	IG	IG	IG	--	--	--	--	1
67	186843	G	A	CDS	0157	Transcriptional regulator, RofA-like Protein	Non	Thr_T	Ile_I	Con	1
68	187289	C	T	CDS	0157	Transcriptional regulator, RofA-like Protein	Non	Met_M	Ile_I	Con	1
69	189525	T	A	CDS	0160	UTP--Glc-1-phosphate uridylyltransferase	Syn	Thr_T	Thr_T	--	1
70	195125	C	T	CDS	0164	Multidrug resistance ABC transporter	Non	Gln_Q	STOP	STOP	5
71	195350	G	T	CDS	0164	Multidrug resistance ABC transporter	Non	Ala_A	Ser_S	Con	1
72	200130	C	T	IG	IG	IG	--	--	--	--	1
73	201055	G	A	CDS	0170	Glutamyl-tRNA synthetase	Non	Glu_E	Lys_K	Con	10
74	201120	G	A	CDS	0170	Glutamyl-tRNA synthetase	Syn	Glu_E	Glu_E	--	1
75	201573	A	G	CDS	0170	Glutamyl-tRNA synthetase	Syn	Leu_L	Leu_L	--	1
76	201820	T	C	IG	IG	IG	--	--	--	--	1
77	204839	T	C	CDS	0174	Response regulator FasA or ComE or BlpR	Syn	Asp_D	Asp_D	--	1
78	205327	C	T	CDS	0174	Response regulator FasA or ComE or BlpR	Non	Ala_A	Val_V	Con	9
79	206116	G	T	CDS	0175	Ribonuclease P protein component	Syn	Val_V	Val_V	--	1
80	213285	C	T	CDS	0184	unknown function	Non	Thr_T	Met_M	Con	1
81	218570	C	T	CDS	0190	Dimethyladenosine transferase	Non	Pro_P	Ser_S	Con	1
82	220205	C	T	CDS	0192	Ribulose-phosphate 3-epimerase	Syn	Tyr_Y	Tyr_Y	--	9
83	221722	C	T	CDS	0194	DNA recombination protein RmuC	Syn	Asp_D	Asp_D	--	1
84	225263	C	T	CDS	0197	Surface exclusion protein	Non	Ala_A	Val_V	Con	1
85	233903	G	T	CDS	0203	Amino acid ABC transporter	Non	Leu_L	Ile_I	Con	1
86	235575	C	T	CDS	0204	unknown function	Non	Thr_T	Ile_I	Con	1
87	238919	G	A	CDS	0206	Negative regulator of genetic competence	Non	Gly_G	Arg_R	Rad	1
88	239812	C	T	CDS	0207	Undecaprenyl-phosphate GlcNAc-1-P transferase	Syn	Asp_D	Asp_D	--	1
89	242152	C	T	CDS	0209	Iron-sulfur cluster assembly protein	Non	Arg_R	Cys_C	Rad	1
90	242944	A	C	CDS	0210	Cysteine desulfurase	Non	Ile_I	Leu_L	Con	14
91	243806	C	T	CDS	0211	Iron-sulfur cluster assembly scaffold protein	Syn	Leu_L	Leu_L	--	1
92	245533	G	A	CDS	0212	Iron-sulfur cluster assembly protein	Non	Ser_S	Asn_N	Con	1
93	248786	A	G	CDS	0215	Oligopeptide transport protein oppA	Syn	Val_V	Val_V	--	1
94	253755	G	T	CDS	0218	Oligopeptide transport protein oppD	Non	Cys_C	Phe_F	Rad	1
95	254542	T	C	CDS	0219	Oligopeptide transport protein oppF	Non	Met_M	Thr_T	Con	5
96	255852	C	T	IG	IG	IG	--	--	--	--	1
97	261079	A	G	IG	IG	IG	--	--	--	--	6
98	262601	C	A	IG	IG	IG	--	--	--	--	1
99	262617	T	C	IG	IG	IG	--	--	--	--	1
100	262888	C	T	CDS	0221	putative transposase - IS1548	Non	Pro_P	Ser_S	Con	1
101	264283	C	T	CDS	0223	Hydrolase, HAD subfamily IIIA	Syn	Ala_A	Ala_A	--	1
102	276064	T	C	CDS	0236	Branched-chain aa transport carrier protein	Non	Thr_T	Ala_A	Con	1
103	277587	A	G	CDS	0237	Dicarboxylate/amino acid:Na+ or H+ symporter	Syn	Ala_A	Ala_A	--	3
104	277878	G	A	CDS	0237	Dicarboxylate/amino acid:Na+ or H+ symporter	Syn	Leu_L	Leu_L	--	1
105	283656	A	T	CDS	0244	TCS response regulator CovR/CsrR	Non	Glu_E	Val_V	Rad	1
106	283739	G	T	CDS	0244	TCS response regulator CovR/CsrR	Non	Glu_E	STOP	STOP	1
107	283982	C	T	CDS	0244	TCS response regulator CovR/CsrR	Non	Arg_R	Cys_C	Rad	1
108	283985	C	T	CDS	0244	TCS response regulator CovR/CsrR	Non	Arg_R	Cys_C	Rad	1
109	284064	C	A	CDS	0244	TCS response regulator CovR/CsrR	Non	Ser_S	STOP	STOP	1
110	284208	A	T	CDS	0244	TCS response regulator CovR/CsrR	Non	Asn_N	Ile_I	Rad	1
111	284369	C	G	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Pro_P	Arg_R	Rad	1
112	284434	G	A	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Ala_A	Thr_T	Con	1
113	284614	C	T	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Arg_R	STOP	STOP	1
114	284719	G	T	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Glu_E	STOP	STOP	1
115	284963	C	T	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Thr_T	Met_M	Con	1
116	284976	G	T	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Leu_L	Phe_F	Con	1
117	285077	A	G	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Glu_E	Gly_G	Con	1
118	285102	G	A	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Met_M	Ile_I	Con	1
119	285188	T	C	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Ile_I	Thr_T	Con	1
120	285260	C	T	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Ala_A	Val_V	Con	1
121	285653	G	A	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Arg_R	Lys_K	Con	1
122	285692	T	G	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Val_V	Gly_G	Rad	64
123	285704	G	A	CDS	0245	TCS sensor kinase CovS/CsrS	Non	Gly_G	Asp_D	Con	1
124	290120	C	T	IG	IG	IG	--	--	--	--	19
125	291438	G	A	CDS	0250	SWF/SNF family helicase	Syn	Ser_S	Ser_S	--	92

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
126	293363	G	A	IG	IG	IG	--	--	--	--	1
127	295278	T	C	CDS	0252	UDP-N-acetylmuramate--alanine ligase	Syn	Gly_G	Gly_G	--	1
128	304058	C	T	CDS	0263	Xanthosine/inosine triphosphate pyrophosphatase	Non	Thr_T	Ile_I	Con	2
129	309506	G	A	CDS	0271	tRNA (cytosine34-2'-O-)-methyltransferase	Syn	Gln_Q	Gln_Q	--	1
130	312102	G	A	CDS	0274	Fe-S OXIDOREDUCTASE	Non	Ser_S	Asn_N	Con	1
131	312174	T	A	CDS	0274	Fe-S OXIDOREDUCTASE	Non	Ile_I	Asn_N	Rad	1
132	313068	A	T	CDS	0276	unknown function	Non	Leu_L	Phe_F	Con	1
133	317708	G	A	CDS	0280	Ferrichrome transport protein	Non	Pro_P	Ser_S	Con	5
134	317815	G	A	CDS	0280	Ferrichrome transport protein	Non	Pro_p	Leu_L	Con	1
135	319262	G	A	CDS	0282	Ferrichrome-binding periplasmic protein	Syn	Leu_L	Leu_L	--	2
136	320910	C	T	IG	IG	IG	--	--	--	--	1
137	321832	C	A	CDS	0284	UDP-MurNAc-D-Glu--2,6-diaminopimelate ligase	Non	Arg_R	Ile_I	Con	6
138	324785	C	G	CDS	0286	Uracil phosphoribosyltransferase	Non	Pro_P	Ala_A	Con	1
139	326376	A	C	CDS	0288	unknown function	Non	Lys_K	Asn_N	Con	1
140	329015	C	T	IG	IG	IG	--	--	--	--	1
141	334609	G	A	CDS	0298	Serine endopeptidase ScpC	Syn	Val_V	Val_V	--	1
142	337817	G	A	CDS	0298	Serine endopeptidase ScpC	Non	Asp_D	Asn_N	Con	1
143	345191	G	A	CDS	0302	Ribonucleotide reduction protein NrdI	Non	Asp_D	Asn_N	Con	4
144	345575	A	G	CDS	0303	Ribonucleotide reductase of class Ib	Non	Gln_Q	Arg_R	Con	2
145	350013	C	A	CDS	0307	unknown function	Non	Glu_E	STOP	STOP	1
146	353069	G	A	CDS	0311	Phosphoglycerate transporter protein	Non	Val_V	Ile_I	Con	1
147	354889	C	T	CDS	0312	GlcNAc-1-phosphate uridylyltransferase	Non	Ala_A	Val_V	Con	1
148	355502	C	T	CDS	0312	GlcNAc-1-phosphate uridylyltransferase	Syn	Ala_A	Ala_A	--	2
149	357387	A	G	CDS	0315	5'-methylthioadenosine nucleosidase	Non	Ile_I	Val_V	Con	2
150	358653	A	T	CDS	0317	Metal transporter of streptococcus regulator MtsR	Non	Ile_I	Asn_N	Rad	1
151	362477	C	T	IG	IG	IG	--	--	--	--	1
152	363501	G	A	CDS	0322	Cell division protein ftsK	Non	Glu_E	Lys_K	Con	1
153	364945	C	T	CDS	0322	Cell division protein ftsK	Non	Ala_A	Val_V	Con	13
154	366105	C	T	IG	IG	IG	--	--	--	--	1
155	371078	C	T	CDS	0331	streptococcal surface immunogenic protein	Non	Pro_P	Ser_S	Con	1
156	371608	A	T	IG	IG	IG	--	--	--	--	1
157	372240	A	C	CDS	0332	67kDa Myosin-crossreactive streptococcal antigen	Non	Glu_E	Asp_D	Con	1
158	373426	G	A	CDS	0332	67kDa Myosin-crossreactive streptococcal antigen	Non	Val_V	Ile_I	Con	1
159	381581	G	T	IG	IG	IG	--	--	--	--	10
160	385465	T	A	CDS	0346	Positive transcriptional regulator, MutR family	Non	Asn_N	Lys_K	Con	1
161	390841	T	A	IG	IG	IG	--	--	--	--	1
162	393294	G	T	CDS	0353	SsrA-binding protein	Non	Ala_A	Ser_S	Con	14
163	393323	C	T	CDS	0353	SsrA-binding protein	Syn	Thr_T	Thr_T	--	1
164	393741	C	T	CDS	0354	Glutaminy-peptide cyclotransferase	Non	Thr_T	Ile_I	Con	2
165	394857	C	A	CDS	0355	Pyrrolidone-carboxylate peptidase	Non	Ala_A	Ser_S	Con	1
166	397063	G	A	IG	IG	IG	--	--	--	--	1
167	398051	C	T	IG	IG	IG	--	--	--	--	1
168	398305	G	A	CDS	0359	Lactoylglycyltransferase	Syn	Ala_A	Ala_A	--	1
169	398983	A	G	CDS	0360	Oxygen-insensitive NAD(P)H nitroreductase	Non	Tyr_Y	His_H	Con	1
170	399627	C	G	CDS	0361	Xaa-Pro dipeptidase	Non	Gly_G	Ala_A	Con	1
171	406412	C	A	IG	IG	IG	--	--	--	--	1
172	409052	C	T	CDS	0368	Multidrug transport system, permease component	Syn	Leu_L	Leu_L	--	5
173	409500	T	C	IG	IG	IG	--	--	--	--	1
174	409569	G	A	IG	IG	IG	--	--	--	--	1
175	413668	G	A	CDS	0372	Two-component response regulator	Syn	Ser_S	Ser_S	--	1
176	420072	T	C	CDS	0376	Chromosome partition protein smc	Syn	Ala_A	Ala_A	--	1
177	425600	G	T	CDS	0381	unknown function	Non	Lys_K	Asn_N	Con	1
178	426890	T	C	CDS	0382	Archaeal S-adenosylmethionine synthetase	Syn	Ile_I	Ile_I	--	1
179	427769	G	A	CDS	0383	unknown function	Non	Gly_G	Glu_E	Con	3
180	429521	G	A	CDS	--	unknown function	Syn	Lys_K	Lys_K	--	1
181	430539	A	G	CDS	0385	UDP-glucose dehydrogenase	Non	Asn_N	Ser_S	Con	1
182	432907	C	T	CDS	0389	unknown function	Non	Ala_A	Thr_T	Con	1
183	433519	C	T	CDS	0389	unknown function	Non	Asp_D	Asn_N	Con	10
184	437736	C	T	CDS	0397	Transposase TnpA	Non	Cys_C	Tyr_Y	Rad	1
185	438447	C	T	CDS	0397	Transposase TnpA	Non	Cys_C	Tyr_Y	Rad	1
186	438462	G	A	IG	IG	IG	--	--	--	--	1
187	439866	G	A	CDS	0399	Hydrolase (HAD superfamily)	Non	Gly_G	Glu_E	Con	2
188	440859	C	T	CDS	0401	SRP receptor protein FtsY	Syn	Asp_D	Asp_D	--	1

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
189	442045	A	C	CDS	0401	SRP receptor protein FtsY	Non	Ser_S	Arg_R	Rad	15
190	447750	C	T	CDS	0405	6-P-beta-glucosidase	Syn	Asp_D	Asp_D	--	1
191	447822	T	C	CDS	0405	6-P-beta-glucosidase	Syn	Asn_N	Asn_N	--	4
192	448608	C	T	CDS	0406	ABC transporter, substrate-binding protein	Syn	Pro_P	Pro_P	--	1
193	453672	G	A	IG	IG	IG	--	--	--	--	1
194	455064	C	T	CDS	0413	Prolipoprotein diacylglycerol transferase	Non	Ala_A	Val_V	Con	92
195	458239	C	T	CDS	0418	peptidase, U32 family large subunit	Syn	Gly_G	Gly_G	--	3
196	465455	C	A	CDS	0425	Phage lysin glycosyl hydrolase	Non	Ala_A	Ser_S	Con	1
197	467972	T	G	CDS	0429	Oligoendopeptidase F	Syn	Ala_A	Ala_A	--	5
198	468760	C	A	CDS	0429	Oligoendopeptidase F	Non	Ala_A	Ser_S	Con	1
199	468848	G	C	CDS	0429	Oligoendopeptidase F	Syn	Ala_A	Ala_A	--	5
200	470370	G	A	CDS	0430	Phosphoenolpyruvate carboxylase	Non	Gly_G	Asp_D	Con	6
201	472699	C	T	CDS	0431	Cell division protein ftsW	Non	Thr_T	Ile_I	Con	2
202	476851	C	T	CDS	0434	MurN protein	Non	Val_V	Met_M	Con	1
203	477854	T	C	CDS	0435	PG serine/alanine adding enzyme	Non	Thr_T	Ala_A	Con	62
204	481574	T	C	IG	IG	IG	--	--	--	--	1
205	484867	G	A	CDS	0440	Calcium-transporting ATPase	Non	Val_V	Ile_I	Con	6
206	485738	C	T	CDS	0441	Transcriptional regulator, LacI family	Non	Ala_A	Thr_T	Con	1
207	491836	A	G	IG	IG	IG	--	--	--	--	20
208	491954	G	T	IG	IG	IG	--	--	--	--	1
209	492838	G	A	CDS	0448	2-deoxy-D-gluconate 3-dehydrogenase	Syn	Gly_G	Gly_G	--	1
210	493581	C	T	CDS	0449	Gal-6-P isomerase LacB subunit	Non	Pro_P	Leu_L	Con	1
211	493657	T	C	CDS	0450	2-dehydro-3-deoxygluconate kinase	Non	Phe_F	Ser_S	Rad	1
212	494333	G	A	CDS	0450	2-dehydro-3-deoxygluconate kinase	Syn	Gln_Q	Gln_Q	--	6
213	495679	C	T	CDS	0452	Beta-phosphoglucomutase	Non	Thr_T	Ile_I	Con	1
214	501388	A	T	CDS	0457	Hydroxyacylglutathione hydrolase	Non	Ser_S	Thr_T	Con	2
215	503347	A	T	CDS	0459	DinG family ATP-dependent helicase YoaA	Non	Tyr_Y	Phe_F	Con	5
216	504376	C	A	CDS	0459	DinG family ATP-dependent helicase YoaA	Non	Pro_P	His_H	Con	1
217	505864	C	T	CDS	0460	Aspartate aminotransferase	Non	Thr_T	Ile_I	Con	1
218	506901	G	T	CDS	0460	Aspartate aminotransferase	Non	Ala_A	Ser_S	Con	1
219	509549	C	T	CDS	0462	unknown function	Non	Ala_A	Val_V	Con	6
220	512693	A	G	CDS	0465	Dipeptidase	Non	Thr_T	Ala_A	Con	23
221	513888	G	A	CDS	0466	Candidate zinc-binding lipoprotein ZinT	Non	Val_V	Ile_I	Con	1
222	515097	G	C	CDS	0467	Transcriptional regulator, GntR family	Non	Glu_E	Gln_Q	Con	1
223	520800	A	G	IG	IG	IG	--	--	--	--	3
224	524342	G	A	CDS	0476	DNA gyrase subunit B	Syn	Lys_K	Lys_K	--	1
225	530558	C	T	CDS	0481	Streptolysin S biosynthesis protein B (SagB)	Syn	Tyr_Y	Tyr_Y	--	1
226	530719	C	T	CDS	0481	Streptolysin S biosynthesis protein B (SagB)	Non	Thr_T	Ile_I	Con	1
227	531205	G	A	CDS	0482	Streptolysin S biosynthesis protein C (SagC)	Non	Met_M	Ile_I	Con	1
228	539377	T	C	CDS	0489	Endonuclease/Exonuclease/phosphatase	Syn	Val_V	Val_V	--	1
229	540894	A	G	CDS	0489	Endonuclease/Exonuclease/phosphatase	Non	His_H	Arg_R	Con	3
230	541819	T	G	CDS	0490	predicted queuosine-regulated ECF transporter	Non	Ile_I	Ser_S	Rad	2
231	543888	G	A	CDS	0491	DNA ligase	Non	Gly_G	Asp_D	Con	1
232	546139	G	A	CDS	0494	ATP synthase A chain	Non	Met_M	Ile_I	Con	1
233	547021	G	C	CDS	0496	ATP synthase delta chain	Non	Gly_G	Ala_A	Con	70
234	553152	T	C	CDS	0502	UDP-GlcNAc 1-carboxyvinyltransferase	Syn	Gly_G	Gly_G	--	5
235	557039	G	A	CDS	0507	Phenylalanyl-tRNA synthetase beta chain	Syn	Ser_S	Ser_S	--	1
236	560370	C	T	CDS	0510	ABC transporter permease protein	Non	Ala_A	Val_V	Con	1
237	561889	C	T	CDS	0512	Neutral zinc metallopeptidase family	Syn	Leu_L	Leu_L	--	1
238	562452	C	T	IG	IG	IG	--	--	--	--	1
239	569809	C	T	CDS	0515	Amino acid ABC transporter	Non	Pro_P	Leu_L	Con	1
240	570302	T	C	CDS	0515	Amino acid ABC transporter	Syn	Asp_D	Asp_D	--	1
241	571094	C	T	CDS	0517	Large-conductance mechanosensitive channel	Non	Gly_G	Ser_S	Con	1
242	571553	C	T	CDS	0518	DNA primase	Syn	Leu_L	Leu_L	--	2
243	573336	A	C	CDS	0519	RNA polymerase sigma factor RpoD	Non	Asp_D	Ala_A	Rad	1
244	577347	C	T	CDS	0523	Alpha-L-Rha alpha-1,3-L-rhamnosyltransferase	Syn	Asp_D	Asp_D	--	1
245	579033	T	C	CDS	0525	ABC-transporter	Non	Ile_I	Thr_T	Con	1
246	580636	G	C	CDS	0526	putative glycosyltransferase	Syn	Ser_S	Ser_S	--	1
247	585698	A	G	CDS	0529	Glycosyltransferase involved in cell wall biogenesis	Non	Gln_Q	Arg_R	Con	1
248	589433	G	A	IG	IG	IG	--	--	--	--	1
249	597031	C	T	CDS	0542	Methyltransferase	Non	Thr_T	Ile_I	Con	1
250	597618	T	G	CDS	0542	Methyltransferase	Non	Ser_S	Ala_A	Con	2
251	606726	T	C	CDS	0551	Capsule biosynthesis protein capA	Syn	Leu_L	Leu_L	--	11

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
252	611831	G	C	CDS	0558	Uracil phosphoribosyltransferase	Syn	Leu_L	Leu_L	--	1
253	612785	G	A	CDS	0559	Uracil permease	Syn	Val_V	Val_V	--	3
254	614402	T	C	IG	IG	IG	--	--	--	--	16
255	615671	G	C	IG	IG	IG	--	--	--	--	3
256	618026	T	A	CDS	0562	Carbamoyl-phosphate synthase large chain	Non	Asp_D	Glu_E	Con	41
257	625921	T	C	CDS	0569	Cell surface protein	Syn	His_H	His_H	--	2
258	626474	A	G	CDS	0569	Cell surface protein	Non	Thr_T	Ala_A	Con	18
259	627910	A	G	CDS	0569	Cell surface protein	Syn	Ala_A	Ala_A	--	1
260	629448	G	A	IG	IG	IG	--	--	--	--	1
261	629450	G	A	IG	IG	IG	--	--	--	--	1
262	629459	A	G	IG	IG	IG	--	--	--	--	1
263	629785	G	A	CDS	0571	Cobalt-zinc-cadmium resistance protein czcD	Non	His_H	Tyr_Y	Con	4
264	637987	A	G	CDS	0580	PTS system, fructose-specific IIA component	Non	Ile_I	Val_V	Con	1
265	638853	C	T	CDS	0580	PTS system, fructose-specific IIA component	Syn	Gly_G	Gly_G	--	1
266	640536	A	G	CDS	0582	Peptidoglycan hydrolase, Autolysin2	Syn	Glu_E	Glu_E	--	2
267	644431	T	A	CDS	0586	DegV family protein	Non	Tyr_Y	Asn_N	Rad	1
268	650988	A	G	CDS	0591	5'-nucleotidase	Syn	Val_V	Val_V	--	24
269	651247	C	A	CDS	0591	5'-nucleotidase	Non	Gly_G	Val_V	Rad	1
270	655514	C	T	CDS	0595	Mevalonate kinase	Syn	Ile_I	Ile_I	--	1
271	660832	T	C	CDS	0600	Hydroxymethylglutaryl-CoA synthase	Non	Asp_D	Gly_G	Con	23
272	665402	G	T	CDS	0606	unknown function	Non	Trp_W	Cys_C	Rad	1
273	672424	C	T	CDS	0612	Purine nucleoside phosphorylase	Non	Ser_S	Phe_F	Rad	1
274	677902	T	C	CDS	0618	6-aminohexanoate-cyclic-dimer hydrolase	Syn	His_H	His_H	--	1
275	678501	A	G	CDS	0618	6-aminohexanoate-cyclic-dimer hydrolase	Non	Asp_D	Gly_G	Con	1
276	680903	T	C	CDS	0621	Uracil-DNA glycosylase	Syn	Thr_T	Thr_T	--	1
277	682514	A	G	CDS	0622	Dihydroorotase	Non	Thr_T	Ala_A	Con	1
278	685931	A	G	CDS	0625	Topoisomerase	Non	Asp_D	Gly_G	Con	1
279	686415	C	T	CDS	0625	Topoisomerase	Syn	Ile_I	Ile_I	--	1
280	690093	G	A	CDS	0628	SSU ribosomal protein S1p	Non	Val_V	Ile_I	Con	2
281	692292	C	T	CDS	0630	Rhodanese-related sulfurtransferases	Non	Ser_S	Leu_L	Rad	11
282	700559	C	T	CDS	0639	Single-stranded-DNA-specific exonuclease recJ	Syn	Ala_A	Ala_A	--	1
283	701833	C	T	CDS	0639	Single-stranded-DNA-specific exonuclease recJ	Non	Thr_T	Ile_I	Con	2
284	703184	C	T	CDS	0641	DNA replication protein dnaD	Non	Thr_T	Ile_I	Con	1
285	705471	C	A	CDS	0644	unknown function	Non	Gln_Q	Lys_K	Con	20
286	710274	C	T	CDS	0649	Mutator mutT protein	Non	Arg_R	Cys_C	Rad	13
287	712833	T	C	IG	IG	IG	--	--	--	--	1
288	712926	A	G	CDS	0652	Fibronectin/fibrinogen-binding protein	Non	Val_V	Ala_A	Con	1
289	714894	A	G	CDS	0653	ABC transporter substrate-binding protein	Syn	Lys_K	Lys_K	--	2
290	716905	C	T	CDS	0655	ABC transporter permease protein	Non	Thr_T	Ile_I	Con	1
291	719266	G	C	CDS	0657	Zn-dependent hydrolase	Non	Val_V	Leu_L	Con	1
292	720812	T	C	CDS	0659	unknown function	Syn	Asp_D	Asp_D	--	1
293	722755	C	T	CDS	0660	ABC transporter ATP-binding protein	Non	Thr_T	Ile_I	Con	1
294	723656	G	T	CDS	0661	Acetoin dehydrogenase	Non	Ala_A	Ser_S	Con	10
295	724702	C	T	CDS	0662	Acetoin dehydrogenase	Non	Ala_A	Val_V	Con	67
296	725529	C	T	CDS	0663	Dihydroliipoamide acetyltransferase	Syn	Thr_T	Thr_T	--	1
297	727093	G	A	CDS	--	unknown function	Non	Arg_R	STOP	STOP	15
298	728929	C	A	CDS	0664	Dihydroliipoamide dehydrogenase	Non	Pro_P	Thr_T	Con	1
299	729009	G	A	IG	IG	IG	--	--	--	--	1
300	735618	A	T	CDS	0669	ABC transporter, permease protein	Syn	Ala_A	Ala_A	--	3
301	737189	G	A	CDS	0670	unknown function	Syn	Pro_P	Pro_P	--	1
302	740997	T	A	CDS	0674	Acyl-ACP thioesterase	Syn	Pro_P	Pro_P	--	1
303	741001	A	G	CDS	0674	Acyl-ACP thioesterase	Non	Lys_K	Glu_E	Con	1
304	741281	T	C	CDS	0674	Acyl-ACP thioesterase	Non	Ile_I	Thr_T	Con	1
305	743034	A	G	IG	IG	IG	--	--	--	--	1
306	748981	C	A	CDS	0680	unknown function	Non	Phe_F	Leu_L	Con	3
307	749005	A	T	IG	IG	IG	--	--	--	--	1
308	789053	C	T	IG	IG	IG	--	--	--	--	1
309	790332	A	G	CDS	0737	Translation elongation factor LepA	Syn	Gln_Q	Gln_Q	--	1
310	790681	G	C	CDS	0737	Translation elongation factor LepA	Non	Ala_A	Pro_P	Con	1
311	791207	A	T	IG	IG	IG	--	--	--	--	1
312	791209	C	T	IG	IG	IG	--	--	--	--	1
313	791498	C	A	CDS	0738	collagen-like protein SclB	Non	Ala_A	Glu_E	Rad	2
314	791510	G	A	CDS	0738	collagen-like protein SclB	Non	Gly_G	Glu_E	Con	1

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
315	791567	A	G	CDS	0738	collagen-like protein SclB	Non	Asp_D	Gly_G	Con	1
316	791623	A	G	CDS	0738	collagen-like protein SclB	Non	Lys_K	Glu_E	Con	2
317	798474	T	C	CDS	0745	Two-component sensor kinase YesM	Syn	Gly_G	Gly_G	--	2
318	801392	T	C	CDS	0748	Succinate-semialdehyde dehydrogenase [NAD]	Non	Leu_L	Ser_S	Rad	1
319	801548	T	C	CDS	0748	Succinate-semialdehyde dehydrogenase [NAD]	Non	Leu_L	Pro_P	Con	1
320	801965	C	T	CDS	0748	Succinate-semialdehyde dehydrogenase [NAD]	Non	Ala_A	Val_V	Con	1
321	806701	T	C	CDS	0751	Xaa-His dipeptidase	Syn	Asp_D	Asp_D	--	1
322	806863	G	A	CDS	0751	Xaa-His dipeptidase	Syn	Glu_E	Glu_E	--	1
323	806899	T	C	CDS	0751	Xaa-His dipeptidase	Syn	Ala_A	Ala_A	--	1
324	807058	G	A	CDS	0752	GTPase and tRNA-U34 5-formylation enzyme	Non	Ala_A	Val_V	Con	1
325	809931	G	T	CDS	0755	D-alanyl-D-alanine carboxypeptidase	Syn	Gly_G	Gly_G	--	1
326	815182	G	T	CDS	0760	Dihydropteroate synthase	Non	Asp_D	Tyr_Y	Rad	1
327	816115	G	A	IG	IG	IG	--	--	--	--	1
328	816196	T	C	IG	IG	IG	--	--	--	--	1
329	817265	C	T	CDS	0764	Spermidine Putrescine transport	Non	Leu_L	Phe_F	Con	1
330	822556	C	T	CDS	0769	Two-component sensor histidine kinase	Syn	Glu_E	Glu_E	--	1
331	823184	T	G	CDS	0769	Two-component sensor histidine kinase	Non	Lys_K	Thr_T	Con	1
332	825681	T	C	CDS	0771	NAD-dependent malic enzyme	Non	Phe_F	Ser_S	Rad	1
333	830478	C	T	IG	IG	IG	--	--	--	--	1
334	832287	G	T	CDS	0777	DNA repair protein RadC	Non	Arg_R	Ile_I	Con	1
335	832888	A	T	CDS	0778	Glutamine amidotransferase	Non	Leu_L	STOP	STOP	16
336	840698	T	G	CDS	0787	Phosphate acetyltransferase	Non	Ile_I	Met_M	Con	1
337	841828	T	C	CDS	0790	Na+ driven multidrug efflux pump	Non	Ile_I	Val_V	Con	1
338	842041	G	T	CDS	0790	Na+ driven multidrug efflux pump	Non	Arg_R	Ser_S	Rad	1
339	844336	T	G	CDS	0792	ABC transporter protein proX	Non	Ser_S	Arg_R	Rad	1
340	846951	A	C	CDS	0794	Xanthine phosphoribosyltransferase	Syn	Ala_A	Ala_A	--	2
341	849465	G	A	CDS	0796	unknown function	Syn	Asn_N	Asn_N	--	1
342	850454	C	T	CDS	0798	Thymidine kinase	Syn	Asn_N	Asn_N	--	8
343	850721	C	T	CDS	0799	Peptide chain release factor 1	Non	Thr_T	Ile_I	Con	4
344	853669	G	A	CDS	0803	Serine hydroxymethyltransferase	Non	Ala_A	Thr_T	Con	6
345	853735	G	A	CDS	0803	Serine hydroxymethyltransferase	Non	Gly_G	Arg_R	Rad	1
346	855352	C	T	CDS	0804	Pseudouridylyl synthases, 23S RNA-specific	Non	Gln_Q	STOP	STOP	1
347	855603	T	A	CDS	0804	Pseudouridylyl synthases, 23S RNA-specific	Non	Asp_D	Glu_E	Con	1
348	859856	G	T	IG	IG	IG	--	--	--	--	1
349	860866	G	A	CDS	0808	NADH oxidase H2O-forming	Non	Gly_G	Ser_S	Con	1
350	870839	C	T	CDS	0818	Ribonuclease HII	Syn	Gly_G	Gly_G	--	1
351	872144	G	A	CDS	0819	Rossmann fold nt-binding protein Smf	Non	Gly_G	Glu_E	Con	1
352	872275	G	T	CDS	0819	Rossmann fold nt-binding protein Smf	Non	Gly_G	Cys_C	Rad	2
353	872314	C	T	CDS	0819	Rossmann fold nt-binding protein Smf	Non	Arg_R	STOP	STOP	1
354	874111	G	C	CDS	0820	DNA topoisomerase I	Non	Val_V	Leu_L	Con	1
355	874336	C	T	CDS	0820	DNA topoisomerase I	Non	His_H	Tyr_Y	Con	1
356	876852	A	G	CDS	0822	Oxaloacetate decarboxylase	Non	Asp_D	Gly_G	Con	1
357	881273	G	A	CDS	0827	Transcriptional regulator, GntR family	Non	Pro_P	Leu_L	Con	5
358	883268	C	A	IG	IG	IG	--	--	--	--	1
359	884070	C	A	IG	IG	IG	--	--	--	--	1
360	886217	C	T	CDS	0833	Citrate lyase beta chain	Non	Ala_A	Val_V	Con	1
361	891596	T	C	CDS	0838	unknown function	Non	Ile_I	Val_V	Con	1
362	892506	G	A	IG	IG	IG	--	--	--	--	1
363	893301	G	A	CDS	0839	Tyrosine recombinase xerC	Non	Leu_L	Phe_F	Con	1
364	893651	C	T	IG	IG	IG	--	--	--	--	1
365	893674	T	A	IG	IG	IG	--	--	--	--	1
366	901307	C	T	CDS	0846	UDP-MurNAc-pp-lysine N(6)-alanyltransferase	Syn	Lys_K	Lys_K	--	10
367	903712	G	A	CDS	0847	ABC transporter ATP-binding protein	Non	Ala_A	Val_V	Con	39
368	908783	G	A	CDS	0852	Cardiolipin synthetase	Syn	Tyr_Y	Tyr_Y	--	1
369	911132	C	T	CDS	0853	Formate--tetrahydrofolate ligase	Non	Met_M	Ile_I	Con	1
370	913036	T	C	CDS	0855	chromosome architecture/replication ATPase	Syn	Gln_Q	Gln_Q	--	86
371	914240	C	T	CDS	0856	chromosome architecture/replication ATPase	Syn	Leu_L	Leu_L	--	3
372	916032	G	A	CDS	0859	Putative trimethylamine dehydrogenase	Non	His_H	Tyr_Y	Con	5
373	920897	A	G	CDS	0864	Phosphoglucomutase	Non	Ile_I	Val_V	Con	1
374	924493	A	G	CDS	0867	Putative deoxyribose-specific ABC transporter	Non	Ser_S	Pro_P	Con	23
375	926557	T	C	CDS	0868	Nucleoside-binding protein	Non	Ser_S	Gly_G	Con	1
376	927384	A	G	IG	IG	IG	--	--	--	--	1
377	927398	A	G	IG	IG	IG	--	--	--	--	1



Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
378	930663	C	T	CDS	0873	Two component system sensor histidine kinase CiaH	Non	Val_V	Ile_I	Con	1
379	933863	C	T	CDS	0875	Lysyl-aminopeptidase; aminopeptidase N	Non	Asp_D	Asn_N	Con	10
380	934254	T	C	CDS	0875	Lysyl-aminopeptidase; aminopeptidase N	Syn	Thr_T	Thr_T	--	1
381	935891	A	T	CDS	0877	Phosphate transport protein pstB	Non	Leu_L	STOP	STOP	1
382	936462	G	A	CDS	0877	Phosphate transport protein pstB	Non	Pro_P	Ser_S	Con	1
383	939321	C	T	CDS	0881	Phosphate ABC transporter protein pstS	Syn	Val_V	Val_V	--	1
384	940431	A	T	CDS	0882	tRNA and rRNA cytosine-C5-methylases	Syn	Gly_Q	Gly_G	--	4
385	941901	T	G	CDS	0883	Myo-inositol-1(or 4)-monophosphatase	Non	Gln_Q	Pro_P	Con	1
386	943568	G	A	CDS	0886	Riboflavin kinase	Non	Ser_S	Leu_L	Rad	1
387	949710	G	A	CDS	0891	ABC transporter permease protein	Non	Ala_A	Val_V	Con	36
388	959456	A	T	IG	IG	IG	--	--	--	--	1
389	959976	T	C	IG	IG	IG	--	--	--	--	2
390	961312	T	G	IG	IG	IG	--	--	--	--	1
391	966666	G	A	CDS	0910	Glucosamine--fructose-6-P aminotransferase	Non	Thr_T	Ile_I	Con	1
392	967157	G	A	CDS	0910	Glucosamine--fructose-6-P aminotransferase	Syn	Gly_Q	Gly_G	--	1
393	967423	T	C	IG	IG	IG	--	--	--	--	14
394	970425	T	C	CDS	0913	6-phosphofructokinase	Syn	Ala_A	Ala_A	--	1
395	970889	A	G	IG	IG	IG	--	--	--	--	1
396	971026	A	G	CDS	0914	DNA polymerase III	Syn	Ile_I	Ile_I	--	1
397	972816	C	T	CDS	0914	DNA polymerase III	Non	Val_V	Ile_I	Con	1
398	973734	G	A	CDS	0914	DNA polymerase III	Non	Leu_L	Phe_F	Con	1
399	974807	G	A	CDS	0916	ABC-type multidrug transport system	Non	Arg_R	His_H	Con	1
400	975292	A	T	CDS	0916	ABC-type multidrug transport system	Non	Thr_T	Ser_S	Con	23
401	975504	T	C	CDS	0917	ABC transporter, ATP-binding protein	Syn	Phe_F	Phe_F	--	5
402	1019065	G	A	CDS	0979	lipoprotein, putative	Syn	Leu_L	Leu_L	--	24
403	1019314	G	A	IG	IG	IG	--	--	--	--	1
404	1020687	G	A	CDS	0980	Glycogen phosphorylase	Syn	Thr_T	Thr_T	--	1
405	1022760	C	T	CDS	0981	4-alpha-glucanotransferase	Non	Ala_A	Thr_T	Con	6
406	1022824	T	A	CDS	0981	4-alpha-glucanotransferase	Syn	Thr_T	Thr_T	--	2
407	1023643	G	A	CDS	0982	Maltose operon transcriptional repressor MalR	Syn	Ser_S	Ser_S	--	10
408	1023900	G	A	CDS	0982	Maltose operon transcriptional repressor MalR	Non	Gln_Q	STOP	STOP	1
409	1026493	A	G	CDS	0984	Maltose/maltodextrin ABC transporter protein MalF	Non	Ser_S	Gly_G	Con	1
410	1026908	C	T	CDS	0984	Maltose/maltodextrin ABC transporter protein MalF	Non	Thr_T	Ile_I	Con	2
411	1028537	A	C	CDS	0987	putative transposase N-terminal portion	Syn	Ser_S	Ser_S	--	1
412	1028567	C	G	CDS	0987	putative transposase N-terminal portion	Syn	Arg_R	Arg_R	--	1
413	1030601	A	G	CDS	0990	Esterase	Syn	Asn_N	Asn_N	--	1
414	1033471	C	T	CDS	0993	D-alanyl transfer protein DltB	Non	Gly_G	Arg_R	Rad	2
415	1035354	T	A	IG	IG	IG	--	--	--	--	1
416	1037803	A	T	CDS	0997	Glutamine ABC transporter	Syn	Thr_T	Thr_T	--	1
417	1039955	C	T	CDS	0998	Glutamine transport protein glnQ	Non	Ala_A	Val_V	Con	6
418	1040577	C	T	CDS	0998	Glutamine transport protein glnQ	Syn	His_H	His_H	--	3
419	1040731	G	A	IG	IG	IG	--	--	--	--	1
420	1040775	G	A	CDS	0999	unknown function	Non	Arg_R	Cys_C	Rad	1
421	1041721	C	T	CDS	1000	PTS system, cellobiose-specific IIC	Non	Arg_R	His_H	Con	2
422	1044396	C	T	CDS	1005	Transcriptional antiterminator of lichenan operon	Non	Ala_A	Thr_T	Con	1
423	1044702	C	T	CDS	1005	Transcriptional antiterminator of lichenan operon	Non	Asp_D	Asn_N	Con	1
424	1045382	T	C	CDS	1005	Transcriptional antiterminator of lichenan operon	Non	Asp_D	Gly_G	Con	1
425	1049136	C	T	CDS	1009	unknown function	Syn	Leu_L	Leu_L	--	1
426	1051557	A	G	CDS	1011	putative transposase associated protein	Syn	Arg_R	Arg_R	--	1
427	1051565	C	T	CDS	1011	putative transposase associated protein	Non	Ala_A	Val_V	Con	1
428	1051569	A	G	CDS	1011	putative transposase associated protein	Syn	Lys_K	Lys_K	--	1
429	1051686	C	T	CDS	1012	Transposase	Non	Gln_Q	STOP	STOP	1
430	1051782	T	C	CDS	1012	Transposase	Non	Cys_C	Arg_R	Rad	1
431	1051842	C	T	CDS	1012	Transposase	Syn	Leu_L	Leu_L	--	1
432	1051851	T	C	CDS	1012	Transposase	Non	Cys_C	Arg_R	Rad	1
433	1051866	T	C	CDS	1012	Transposase	Non	Cys_C	Arg_R	Rad	1
434	1054883	C	T	CDS	1014	Malonate permease	Syn	Pro_p	Pro_P	--	2
435	1055067	G	A	CDS	1014	Malonate permease	Non	Val_V	Met_M	Con	1
436	1055202	A	G	IG	IG	IG	--	--	--	--	1
437	1056106	G	A	CDS	1017	putative transposase - IS904A	Non	Arg_R	Gln_Q	Con	3
438	1056126	C	T	CDS	1017	putative transposase - IS904A	Non	Arg_R	Trp_W	Rad	3
439	1056136	A	G	CDS	1017	putative transposase - IS904A	Non	Tyr_Y	Cys_C	Rad	1
440	1056146	G	T	CDS	1017	putative transposase - IS904A	Non	Arg_R	Ser_S	Rad	3

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
441	1056189	G	T	CDS	1017	putative transposase - IS904A	Non	Ala_A	Ser_S	Con	3
442	1056211	G	A	CDS	1017	putative transposase - IS904A	Non	Arg_R	Lys_K	Con	3
443	1056592	G	A	CDS	1018	Ribosomal small subunit pseudouridine synthase A	Syn	Gly_G	Gly_G	--	1
444	1059231	G	A	CDS	1021	ATP-dependent Zn protease	Syn	Lys_K	Lys_K	--	1
445	1060275	T	C	CDS	1021	ATP-dependent Zn protease	Syn	Pro_P	Pro_p	--	1
446	1060414	G	A	CDS	1021	ATP-dependent Zn protease	Non	Glu_E	Lys_K	Con	1
447	1063938	C	T	CDS	1025	Putative PBP synthesis repressor	Non	Val_V	Ile_I	Con	3
448	1067702	C	T	CDS	1028	Ribonuclease	Non	Arg_R	Gln_Q	Con	1
449	1068710	G	A	CDS	1029	Methionine aminopeptidase	Syn	His_H	His_H	--	1
450	1071276	A	G	CDS	1032	GRAB protein	Syn	Asp_D	Asp_D	--	1
451	1076895	C	T	CDS	1035	internalin	Non	Met_M	Ile_I	Con	1
452	1077899	G	A	IG	IG	IG	--	--	--	--	17
453	1077928	C	G	IG	IG	IG	--	--	--	--	1
454	1078726	G	T	CDS	1036	Biotin-protein ligase	Syn	Ala_A	Ala_A	--	5
455	1078777	C	T	CDS	1036	Biotin-protein ligase	Syn	Arg_R	Arg_R	--	1
456	1079364	A	G	CDS	1038	DNA polymerase III	Syn	Val_V	Val_V	--	5
457	1084572	C	T	CDS	1043	ATP-dependent RNA helicase	Non	Thr_T	Ile_I	Con	1
458	1084815	A	G	CDS	1044	Peptidoglycan GlcNAc deacetylase	Syn	His_H	His_H	--	1
459	1086464	G	A	CDS	1045	glyceraldehyde-3-phosphate dehydrogenase	Non	Ala_A	Val_V	Con	1
460	1087785	T	A	CDS	1046	Phosphoenolpyruvate-protein phosphotransferase	Non	Glu_E	Val_V	Rad	1
461	1089991	G	A	pbs	IG	NrdR_Proteobacteria	--	--	--	--	1
462	1091393	A	G	CDS	1049	Ribonucleotide reductase of class Ib	Syn	Glu_E	Glu_E	--	6
463	1092744	C	A	IG	IG	IG	--	--	--	--	17
464	1094245	G	A	CDS	1051	Chloride channel protein	Non	Trp_W	STOP	STOP	19
465	1095918	T	A	IG	IG	IG	--	--	--	--	1
466	1098505	C	T	CDS	1058	Alanyl-tRNA synthetase	Syn	Val_V	Val_V	--	1
467	1101887	G	T	CDS	1059	Foldase protein prsA	Non	Pro_P	Thr_T	Con	16
468	1104382	A	G	CDS	1061	Major facilitator:Oxalate:Formate Antiporter	Syn	Pro_P	Pro_p	--	2
469	1104478	C	T	CDS	1061	Major facilitator:Oxalate:Formate Antiporter	Non	Trp_W	STOP	STOP	1
470	1106038	A	C	CDS	1062	Oligoendopeptidase F	Syn	Val_V	Val_V	--	1
471	1106386	A	G	CDS	1062	Oligoendopeptidase F	Syn	Asp_D	Asp_D	--	1
472	1108998	C	A	CDS	1065	Glucosamine-6-phosphate deaminase	Non	Ala_A	Ser_S	Con	23
473	1113430	C	T	CDS	1071	Manganese superoxide dismutase	Non	Ala_A	Thr_T	Con	2
474	1114921	C	T	CDS	1072	DNA polymerase III	Non	Asp_D	Asn_N	Con	3
475	1115629	T	C	CDS	1073	late competence protein comEC	Non	Tyr_Y	Cys_C	Rad	70
476	1119152	G	A	CDS	1076	COG4123: Predicted O-methyltransferase	Non	Gly_G	Ser_S	Con	1
477	1121564	C	T	CDS	1078	Kup system potassium uptake protein	Syn	Leu_L	Leu_L	--	1
478	1121783	C	T	CDS	1078	Kup system potassium uptake protein	Syn	Leu_L	Leu_L	--	3
479	1122228	A	C	IG	IG	IG	--	--	--	--	1
480	1126295	C	T	CDS	1082	membrane protein, putative	Non	Ala_A	Thr_T	Con	16
481	1128304	G	A	IG	IG	IG	--	--	--	--	1
482	1131764	C	T	IG	IG	IG	--	--	--	--	11
483	1131903	A	G	CDS	1088	predicted niacin ECF transporter	Non	Val_V	Ala_A	Con	1
484	1135308	G	C	IG	IG	IG	--	--	--	--	1
485	1135396	C	A	CDS	1093	Pb, Cd, Zn and Hg transporting ATPase	Syn	Thr_T	Thr_T	--	3
486	1135601	G	A	CDS	1093	Pb, Cd, Zn and Hg transporting ATPase	Non	Thr_T	Ile_I	Con	1
487	1172322	A	T	IG	IG	IG	--	--	--	--	3
488	1174118	A	G	CDS	1149	unknown function	Non	Ile_I	Thr_T	Con	5
489	1174752	G	A	IG	IG	IG	--	--	--	--	1
490	1179891	C	T	CDS	1155	Exodeoxyribonuclease VII	Non	Gly_G	Arg_R	Rad	1
491	1181412	T	C	IG	IG	IG	--	--	--	--	16
492	1183038	C	T	CDS	1158	putative phosphomannomutase	Non	Gly_G	Ser_S	Con	1
493	1184490	G	A	CDS	1159	Deoxyribodipyrimidine photolyase	Non	Leu_L	Phe_F	Con	1
494	1186147	G	A	CDS	1160	Glutamine transport ATP-binding protein glnQ	Syn	Leu_L	Leu_L	--	25
495	1186151	T	A	CDS	1160	Glutamine transport ATP-binding protein glnQ	Syn	Ser_S	Ser_S	--	1
496	1186289	G	A	CDS	1160	Glutamine transport ATP-binding protein glnQ	Syn	Ser_S	Ser_S	--	1
497	1187925	T	C	IG	IG	IG	--	--	--	--	4
498	1187947	C	T	IG	IG	IG	--	--	--	--	1
499	1195058	T	C	CDS	1167	Cell division initiation protein DivIVA	Non	Asn_N	Asp_D	Con	1
500	1199100	A	T	CDS	1173	Cell division protein ftsA	Syn	Ile_I	Ile_I	--	11
501	1200565	G	A	IG	IG	IG	--	--	--	--	1
502	1203073	G	A	CDS	1176	UDP-MurNAc-alanine--D-glutamate ligase	Non	Arg_R	Cys_C	Rad	40
503	1205969	G	A	CDS	1178	GTP-binding protein TypA/BipA	Non	Ala_A	Val_V	Con	1

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
504	1206168	G	A	CDS	1178	GTP-binding protein TypA/BipA	Non	Pro_P	Ser_S	Con	1
505	1209268	G	A	IG	IG	IG	--	--	--	--	1
506	1215091	G	C	CDS	1188	Phosphopantetheine adenylyltransferase	Non	Thr_T	Ser_S	Con	20
507	1216179	A	G	CDS	1190	Aspartate--ammonia ligase	Syn	Asn_N	Asn_N	--	6
508	1217049	C	T	IG	IG	IG	--	--	--	--	1
509	1217650	C	T	CDS	1191	Carbamate kinase	Non	Gly_G	Asp_D	Con	1
510	1218814	C	T	CDS	1192	Xaa-His dipeptidase	Syn	Lys_K	Lys_K	--	1
511	1219476	G	A	IG	IG	IG	--	--	--	--	1
512	1220774	G	A	CDS	1193	Arginine/ornithine antiporter ArcD	Syn	Gly_G	Gly_G	--	5
513	1224182	T	A	IG	IG	IG	--	--	--	--	7
514	1224183	T	A	IG	IG	IG	--	--	--	--	7
515	1224184	T	C	IG	IG	IG	--	--	--	--	7
516	1224430	T	G	CDS	1197	Transcriptional regulator ArcR	Non	Met_M	Leu_L	Con	1
517	1229090	T	C	CDS	1201	Two-component sensor kinase yesM	Non	Asp_D	Gly_G	Con	8
518	1229620	G	A	CDS	1202	Two-component response regulator yesN	Non	Ala_A	Val_V	Con	1
519	1272025	G	C	CDS	1267	Peptide methionine sulfoxide reductase	Syn	Gly_G	Gly_G	--	18
520	1273604	G	A	CDS	1268	Thiol:disulfide oxidoreductase	Non	Thr_T	Ile_I	Con	17
521	1274506	T	C	IG	IG	IG	--	--	--	--	1
522	1276450	T	C	CDS	1271	Oxidoreductase, Gfo/Idh/MocA family	Non	Asp_D	Gly_G	Con	3
523	1282510	A	G	CDS	1277	unknown function	Syn	Ile_I	Ile_I	--	73
524	1283071	A	G	IG	IG	IG	--	--	--	--	1
525	1283090	C	T	IG	IG	IG	--	--	--	--	1
526	1287161	G	A	CDS	1283	Ubiquinone biosynthesis methyltransferase	Non	Thr_T	Ile_I	Con	1
527	1292747	A	G	CDS	1286	Two-component response regulator yesN	Non	Val_V	Ala_A	Con	1
528	1296749	C	A	CDS	1289	Sugar-binding protein	Non	Glu_E	STOP	STOP	1
529	1298852	C	T	CDS	1291	Multiple sugar ABC transporter MsmF	Non	Met_M	Ile_I	Con	1
530	1300321	C	T	CDS	1292	N-acetylmannosamine kinase	Syn	Ala_A	Ala_A	--	1
531	1300624	C	T	IG	IG	IG	--	--	--	--	16
532	1301424	T	C	CDS	1293	Beta-glucosidase	Syn	Val_V	Val_V	--	1
533	1307932	C	T	CDS	1297	Alpha-mannosidase	Non	Ala_A	Val_V	Con	1
534	1309701	A	G	IG	IG	IG	--	--	--	--	9
535	1310694	A	G	CDS	1298	Putative TCS histidine kinase	Syn	Asn_N	Asn_N	--	1
536	1311420	G	A	IG	IG	IG	--	--	--	--	1
537	1311968	G	A	CDS	1299	RNA methyltransferase, TrmA family	Non	Ala_A	Val_V	Con	1
538	1354603	G	A	IG	IG	IG	--	--	--	--	1
539	1362922	C	T	CDS	1361	comF operon protein A, DNA transporter ATPase	Non	Ala_A	Thr_T	Con	2
540	1365099	G	A	CDS	1364	General stress protein 13	Non	Pro_P	Leu_L	Con	1
541	1365959	A	G	CDS	1365	Streptococcal lipoprotein rotamase A	Non	Ile_I	Thr_T	Con	6
542	1368463	A	G	CDS	1368	Transporter yvqF	Syn	Ile_I	Ile_I	--	6
543	1380032	G	C	CDS	1377	Transcriptional regulators, LysR family	Non	Pro_P	Arg_R	Rad	16
544	1384410	A	G	CDS	1383	D-beta-hydroxybutyrate permease	Syn	Ala_A	Ala_A	--	1
545	1386291	G	A	CDS	1385	unknown function	Non	Pro_P	Leu_L	Con	1
546	1386987	G	A	CDS	1385	unknown function	Non	Pro_P	Leu_L	Con	1
547	1389549	A	G	CDS	1388	unknown function	Non	Tyr_Y	His_H	Con	1
548	1398545	C	T	CDS	1394	Amino acid permease	Non	Ala_A	Thr_T	Con	1
549	1403116	G	A	CDS	1399	ATP-dependent RNA helicase YqfR	Non	Arg_R	Cys_C	Rad	1
550	1408991	A	G	CDS	1406	Gamma-glutamyl phosphate reductase	Non	Val_V	Ala_A	Con	2
551	1456631	A	T	CDS	1464	Trans-acting positive regulator	Non	Cys_C	Ser_S	Rad	1
552	1459415	G	A	CDS	1466	Glycerol uptake facilitator protein	Non	Ala_A	Val_V	Con	1
553	1459752	C	T	CDS	1466	Glycerol uptake facilitator protein	Non	Ala_A	Thr_T	Con	1
554	1459819	G	C	CDS	1466	Glycerol uptake facilitator protein	Syn	Ala_A	Ala_A	--	2
555	1459955	C	T	CDS	1466	Glycerol uptake facilitator protein	Non	Gly_G	Asp_D	Con	10
556	1460518	G	A	CDS	1467	Aerobic glycerol-3-phosphate dehydrogenase	Non	Pro_P	Ser_S	Con	1
557	1465235	T	G	CDS	1471	Glycyl-tRNA synthetase beta chain	Non	Lys_K	Gln_Q	Con	18
558	1465401	G	A	CDS	1471	Glycyl-tRNA synthetase beta chain	Syn	Phe_F	Phe_F	--	1
559	1466966	T	C	IG	IG	IG	--	--	--	--	63
560	1469148	A	G	CDS	1474	oxidoreductase	Non	Ser_S	Pro_P	Con	89
561	1470224	G	A	CDS	1475	GlcNAc-6-phosphate deacetylase	Syn	Thr_T	Thr_T	--	1
562	1471633	G	A	CDS	1476	Sodium-dependent phosphate transporter	Non	Arg_R	Cys_C	Rad	16
563	1471920	T	A	CDS	1476	Sodium-dependent phosphate transporter	Non	Tyr_Y	Phe_F	Con	1
564	1473301	T	C	CDS	1477	unknown function	Syn	Phe_F	Phe_F	--	1
565	1475087	T	C	IG	IG	IG	--	--	--	--	1
566	1475088	T	C	IG	IG	IG	--	--	--	--	1

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
567	1477774	C	T	CDS	1482	Tagatose 1,6-diphosphate aldolase	Syn	Val_V	Val_V	--	1
568	1478114	A	G	CDS	1482	Tagatose 1,6-diphosphate aldolase	Non	Val_V	Ala_A	Con	70
569	1481314	C	T	CDS	1486	PTS system, galactose-specific IIC component	Non	Gly_G	Glu_E	Con	2
570	1482194	G	A	CDS	1488	PTS system, galactose-specific IIA component	Syn	Asp_D	Asp_D	--	1
571	1482327	G	A	CDS	1488	PTS system, galactose-specific IIA component	Non	Thr_T	Ile_I	Con	1
572	1482416	C	T	IG	IG	IG	--	--	--	--	1
573	1484947	C	T	CDS	1491	Copper-translocating P-type ATPase	Syn	Val_V	Val_V	--	6
574	1487337	T	C	CDS	1493	putative esterase	Non	Tyr_Y	His_H	Con	6
575	1487474	A	G	CDS	1493	putative esterase	Syn	Lys_K	Lys_K	--	1
576	1488312	C	T	CDS	1495	Translation initiation factor 2	Non	Glu_E	Lys_K	Con	3
577	1489915	A	G	CDS	1495	Translation initiation factor 2	Syn	Asp_D	Asp_D	--	1
578	1497764	G	A	CDS	1505	unknown function	Non	Ala_A	Thr_T	Con	1
579	1498551	T	C	CDS	1506	Transcriptional regulator, LytR family	Non	Thr_T	Ala_A	Con	1
580	1500872	A	G	CDS	1509	Guanine-hypoxanthine permease	Non	Ile_I	Thr_T	Con	93
581	1501014	C	A	CDS	1509	Guanine-hypoxanthine permease	Non	Glu_E	STOP	STOP	3
582	1501703	C	T	CDS	1509	Guanine-hypoxanthine permease	Non	Gly_G	Asp_D	Con	3
583	1501884	A	G	CDS	1509	Guanine-hypoxanthine permease	Non	Phe_F	Leu_L	Con	1
584	1504386	G	A	CDS	1511	PTS system, mannose-specific IIB component	Non	Ala_A	Thr_T	Con	3
585	1507086	C	T	CDS	1515	putative transposase - IS1548	Non	Pro_P	Ser_S	Con	1
586	1508031	A	T	IG	IG	IG	--	--	--	--	1
587	1509415	C	G	CDS	1516	Seryl-tRNA synthetase	Syn	Thr_T	Thr_T	--	1
588	1509527	G	T	CDS	1516	Seryl-tRNA synthetase	Non	Ala_A	Ser_S	Con	1
589	1514831	T	C	CDS	1522	3-oxoacyl-[acyl-carrier-protein] synthase	Non	Ile_I	Val_V	Con	1
590	1515192	A	G	CDS	1523	3-oxoacyl-[acyl-carrier-protein] reductase	Non	Val_V	Ala_A	Con	3
591	1518653	C	A	CDS	1527	3-oxoacyl-[acyl-carrier-protein] synthase	Syn	Pro_P	Pro_P	--	38
592	1519050	C	T	CDS	1527	3-oxoacyl-[acyl-carrier-protein] synthase	Non	Gly_G	Glu_E	Con	3
593	1519054	A	T	CDS	1527	3-oxoacyl-[acyl-carrier-protein] synthase	Non	Ser_S	Thr_T	Con	3
594	1519058	A	G	CDS	1527	3-oxoacyl-[acyl-carrier-protein] synthase	Syn	Thr_T	Thr_T	--	3
595	1519059	G	A	CDS	1527	3-oxoacyl-[acyl-carrier-protein] synthase	Non	Thr_T	Ile_I	Con	1
596	1525730	C	T	CDS	1533	Heat-inducible transcription repressor HrcA	Syn	Gly_G	Gly_G	--	41
597	1528442	A	C	CDS	1537	Integral membrane protein	Non	Phe_F	Leu_L	Con	1
598	1534713	T	C	IG	IG	IG	--	--	--	--	1
599	1536050	C	T	CDS	1545	Aspartate aminotransferase	Syn	Ala_A	Ala_A	--	6
600	1536233	C	T	IG	IG	IG	--	--	--	--	16
601	1540610	C	T	CDS	1550	putative ABC transporter	Non	Val_V	Ile_I	Con	1
602	1544580	G	A	CDS	1553	Duplicated ATPase component	Syn	Ser_S	Ser_S	--	11
603	1549042	C	T	CDS	1557	Putative ABC transporter	Syn	Glu_E	Glu_E	--	6
604	1551111	C	A	CDS	1559	Streptococcal heme ABC transporter	Non	Ser_S	Ile_I	Rad	1
605	1557279	C	T	CDS	1561	Streptococcal surface hemoprotein receptor Shr	Non	Asp_D	Asn_N	Con	1
606	1561493	G	T	CDS	1565	SecA ATPase RNA helicase	Non	Gln_Q	Lys_K	Con	1
607	1563662	C	A	CDS	1565	SecA ATPase RNA helicase	Non	Val_V	Leu_L	Con	1
608	1565479	C	A	CDS	1567	Fructokinase	Non	Val_V	Phe_F	Con	5
609	1566090	G	A	IG	IG	IG	--	--	--	--	1
610	1567143	G	A	CDS	1568	Endo-beta-N-acetylglucosaminidase EndoS	Non	His_H	Tyr_Y	Con	1
611	1568122	G	A	CDS	1568	Endo-beta-N-acetylglucosaminidase EndoS	Syn	Asp_D	Asp_D	--	1
612	1568511	C	T	CDS	1568	Endo-beta-N-acetylglucosaminidase EndoS	Non	Asp_D	Asn_N	Con	2
613	1569141	T	G	CDS	1568	Endo-beta-N-acetylglucosaminidase EndoS	Non	Thr_T	Pro_p	Con	2
614	1570545	A	T	CDS	1569	PTS system, sucrose-specific IIB component	Syn	Ala_A	Ala_A	--	4
615	1573272	T	C	CDS	1571	Sucrose operon repressor ScrR	Non	Leu_L	Ser_S	Rad	24
616	1577764	G	A	CDS	1577	Excinuclease ABC subunit A	Non	Ala_A	Val_V	Con	1
617	1579234	G	A	CDS	1577	Excinuclease ABC subunit A	Non	Ser_S	Phe_F	Rad	3
618	1580901	C	A	CDS	1578	Magnesium and cobalt transport protein CorA	Syn	Ser_S	Ser_S	--	19
619	1581937	T	C	CDS	1579	Integral membrane protein	Non	Leu_L	Ser_S	Rad	4
620	1583782	C	T	CDS	1583	unknown function	Non	Trp_W	STOP	STOP	1
621	1588379	G	T	CDS	1588	Recombination inhibitory protein MutS2	Non	Arg_R	Ser_S	Rad	6
622	1591647	T	A	CDS	1592	Signal peptidase I	Syn	Gly_G	Gly_G	--	1
623	1592075	A	C	CDS	1593	Exodeoxyribonuclease V alpha chain	Non	Lys_K	Thr_T	Con	1
624	1592856	A	G	CDS	1593	Exodeoxyribonuclease V alpha chain	Syn	Lys_K	Lys_K	--	1
625	1592955	C	G	CDS	1593	Exodeoxyribonuclease V alpha chain	Syn	Thr_T	Thr_T	--	1
626	1594459	T	C	CDS	1593	Exodeoxyribonuclease V alpha chain	Non	Ser_S	Pro_p	Con	1
627	1595060	G	A	IG	IG	IG	--	--	--	--	1
628	1596008	C	T	CDS	1595	DNA polymerase IV	Non	Ala_A	Thr_T	Con	1
629	1597641	C	T	CDS	1596	Pyruvate formate-lyase	Non	Ala_A	Val_V	Con	2

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
630	1601536	C	T	CDS	1599	Integral membrane protein	Syn	Phe_F	Phe_F	--	1
631	1605096	T	C	CDS	1602	Transcriptional regulator, CrP/Fnr family	Syn	Tyr_Y	Tyr_Y	--	1
632	1611736	C	T	CDS	1610	NAD(FAD)-utilizing dehydrogenases	Non	His_H	Tyr_Y	Con	1
633	1613345	T	C	CDS	1611	Deoxyribose-phosphate aldolase	Syn	Leu_L	Leu_L	--	5
634	1613488	G	T	CDS	1612	Nucleoside permease nupC	Syn	Ala_A	Ala_A	--	3
635	1616776	A	T	CDS	1615	SSU ribosomal protein S14p	Non	Tyr_Y	Phe_F	Con	1
636	1618236	C	G	CDS	1617	Ribosomal-protein-S18p-Ala acetyltransferase	Non	Gln_Q	His_H	Con	3
637	1619537	A	T	IG	IG	IG	--	--	--	--	1
638	1620754	G	A	CDS	1620	Zn-dependent hydrolase	Non	Gly_G	Asp_D	Con	2
639	1627571	C	T	CDS	1626	membrane protease	Syn	Arg_R	Arg_R	--	10
640	1628980	G	A	CDS	1627	kinase related to hydroxyacetone kinase	Syn	Gly_G	Gly_G	--	1
641	1630376	A	G	IG	IG	IG	--	--	--	--	1
642	1630588	G	A	IG	IG	IG	--	--	--	--	1
643	1631973	T	C	CDS	1631	Hydrolases of the alpha/beta superfamily	Non	Lys_K	Arg_R	Con	2
644	1633944	C	T	CDS	1632	CTP synthase	Non	Gly_G	Arg_R	Rad	1
645	1634345	T	C	CDS	1632	CTP synthase	Non	Tyr_Y	Cys_C	Rad	1
646	1634682	C	T	IG	IG	IG	--	--	--	--	5
647	1638027	T	C	CDS	1635	Potassium efflux mechanosensitive channel	Syn	Tyr_Y	Tyr_Y	--	1
648	1638438	A	T	CDS	1636	unknown function	Non	Val_V	Asp_D	Rad	1
649	1640751	A	G	IG	IG	IG	--	--	--	--	1
650	1647055	A	T	IG	IG	IG	--	--	--	--	76
651	1647056	A	T	IG	IG	IG	--	--	--	--	76
652	1647075	G	C	IG	IG	IG	--	--	--	--	76
653	1647120	G	A	IG	IG	IG	--	--	--	--	71
654	1651729	A	T	CDS	1643	Type I restriction system, specificity subunit S	Non	Asn_N	Tyr_Y	Rad	6
655	1652180	T	C	CDS	1643	Type I restriction system, specificity subunit S	Non	Leu_L	Ser_S	Rad	2
656	1652301	A	G	CDS	1643	Type I restriction system, specificity subunit S	Syn	Arg_R	Arg_R	--	1
657	1656383	G	C	CDS	1646	putative SaK homologue	Non	Pro_p	Ala_A	Con	6
658	1659152	T	G	CDS	1648	putative salivaricin A ABC transporter	Non	Thr_T	Pro_P	Con	3
659	1661654	T	G	CDS	1650	Salivaricin A modification enzyme	Non	Glu_E	Ala_A	Rad	11
660	1667220	G	A	CDS	1654	PTS system, lactose-specific IIB component	Non	Gln_Q	STOP	STOP	1
661	1667466	C	A	CDS	1654	PTS system, lactose-specific IIB component	Non	Gly_G	Cys_C	Rad	2
662	1667998	G	C	CDS	1654	PTS system, lactose-specific IIB component	Syn	Gly_G	Gly_G	--	1
663	1668714	G	A	CDS	1655	PTS system, lactose-specific IIA component	Syn	Asp_D	Asp_D	--	1
664	1668761	G	A	CDS	1655	PTS system, lactose-specific IIA component	Non	His_H	Tyr_Y	Con	1
665	1668828	C	T	CDS	1655	PTS system, lactose-specific IIA component	Syn	Ala_A	Ala_A	--	6
666	1669155	A	T	CDS	1656	Tagatose 1,6-diphosphate aldolase	Non	Phe_F	Tyr_Y	Con	1
667	1669604	G	A	CDS	1656	Tagatose 1,6-diphosphate aldolase	Syn	Cys_C	Cys_C	--	1
668	1669949	A	G	CDS	1657	Tagatose-6-phosphate kinase	Syn	Ile_I	Ile_I	--	1
669	1671163	T	C	CDS	1658	Galactose-6-phosphate isomerase, LacB subunit	Non	Arg_R	Gly_G	Rad	1
670	1671883	A	G	CDS	1659	Galactose-6-phosphate isomerase, LacA subunit	Non	Met_M	Thr_T	Con	5
671	1672060	T	A	IG	IG	IG	--	--	--	--	11
672	1672204	A	T	IG	IG	IG	--	--	--	--	1
673	1678204	G	C	CDS	1667	unknown function	Syn	Thr_T	Thr_T	--	17
674	1684826	T	G	CDS	1676	Polyribonucleotide nucleotidyltransferase	Non	STOP	Ser_S	STOP	1
675	1686676	G	A	CDS	1676	Polyribonucleotide nucleotidyltransferase	Syn	Asp_D	Asp_D	--	1
676	1691812	C	T	IG	IG	IG	--	--	--	--	2
677	1694199	G	A	CDS	1685	unknown function	Non	Ala_A	Val_V	Con	1
678	1697897	T	C	CDS	1687	DNA polymerase III	Non	Glu_E	Gly_G	Con	1
679	1698882	G	A	CDS	1687	DNA polymerase III	Non	Leu_L	Phe_F	Con	1
680	1698973	C	T	CDS	1687	DNA polymerase III	Syn	Glu_E	Glu_E	--	1
681	1699412	C	T	CDS	1687	DNA polymerase III	Non	Arg_R	His_H	Con	1
682	1700339	G	A	CDS	1688	Proyl-tRNA synthetase	Syn	Phe_F	Phe_F	--	3
683	1703042	C	T	IG	IG	IG	--	--	--	--	59
684	1703346	T	C	CDS	1690	Phosphatidate cytidyltransferase	Non	Ile_I	Val_V	Con	1
685	1709402	C	A	IG	IG	IG	--	--	--	--	1
686	1709912	C	T	CDS	1695	Glucan 1,6-alpha-glucosidase	Non	Ala_A	Thr_T	Con	3
687	1712890	G	A	CDS	1697	Leucine rich protein	Non	His_H	Tyr_Y	Con	1
688	1716146	G	C	CDS	1701	GTP pyrophosphokinase	Non	Ala_A	Gly_G	Con	1
689	1716403	A	G	CDS	1701	GTP pyrophosphokinase	Syn	Ile_I	Ile_I	--	59
690	1718958	A	G	CDS	1703	Collagen-like protein SclA	Syn	Asn_N	Asn_N	--	1
691	1719733	C	T	CDS	1704	putative transposase - IS1548	Non	Pro_P	Ser_S	Con	1
692	1721556	C	T	IG	IG	IG	--	--	--	--	1

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
693	1721852	T	C	CDS	1707	Exodeoxyribonuclease III	Non	Asp_D	Gly_G	Con	1
694	1735135	G	A	CDS	1719	Dipeptide transport system protein dppB	Non	Ala_A	Thr_T	Con	1
695	1738301	C	T	CDS	1723	unknown function	Non	Glu_E	Lys_K	Con	1
696	1739068	C	T	CDS	1724	Streptococcal histidine triad protein	Non	Gly_G	Asp_D	Con	1
697	1742009	C	T	CDS	1725	Laminin-binding surface protein	Non	Ala_A	Thr_T	Con	5
698	1742699	A	G	CDS	1726	C5a peptidase	Non	Val_V	Ala_A	Con	1
699	1742705	T	C	CDS	1726	C5a peptidase	Non	Asp_D	Gly_G	Con	1
700	1744981	G	A	CDS	1726	C5a peptidase	Syn	Ala_A	Ala_A	--	1
701	1745193	G	A	CDS	1726	C5a peptidase	Non	Pro_P	Ser_S	Con	1
702	1745700	C	T	CDS	1726	C5a peptidase	Non	Val_V	Ile_I	Con	1
703	1745873	C	T	IG	IG	IG	--	--	--	--	2
704	1745943	G	A	IG	IG	IG	--	--	--	--	4
705	1746410	G	A	CDS	1727	Antiphagocytic M protein	Non	Thr_T	Ile_I	Con	6
706	1747604	T	C	CDS	1727	Antiphagocytic M protein	Non	Asp_D	Gly_G	Con	2
707	1747652	C	T	CDS	1727	Antiphagocytic M protein	Non	Gly_G	Asp_D	Con	1
708	1747653	C	T	CDS	1727	Antiphagocytic M protein	Non	Gly_G	Ser_S	Con	3
709	1747683	C	T	CDS	1727	Antiphagocytic M protein	Non	Asp_D	Asn_N	Con	1
710	1747692	C	G	CDS	1727	Antiphagocytic M protein	Non	Gly_G	Arg_R	Rad	1
711	1747701	C	T	CDS	1727	Antiphagocytic M protein	Non	Ala_A	Thr_T	Con	1
712	1747734	T	C	CDS	1727	Antiphagocytic M protein	Non	Thr_T	Ala_A	Con	1
713	1747737	A	G	CDS	1727	Antiphagocytic M protein	Non	Tyr_Y	His_H	Con	2
714	1747738	T	C	CDS	1727	Antiphagocytic M protein	Syn	Leu_L	Leu_L	--	5
715	1747754	T	C	CDS	1727	Antiphagocytic M protein	Non	Asp_D	Gly_G	Con	1
716	1747889	G	A	CDS	1727	Antiphagocytic M protein	Non	Ala_A	Val_V	Con	1
717	1749825	A	G	IG	IG	mga promoter	--	--	--	--	1
718	1749870	T	C	IG	IG	mga promoter	--	--	--	--	23
719	1750069	T	G	IG	IG	IG	--	--	--	--	1
720	1755630	G	A	CDS	1734	ABC transporter permease protein	Non	Pro_P	Ser_S	Con	1
721	1755860	T	C	CDS	1735	ABC transporter ATP-binding protein	Non	Asp_D	Gly_G	Con	2
722	1760061	G	A	CDS	1740	Peptidylproline cis-trans-isomerase RopA	Non	Thr_T	Ile_I	Con	1
723	1761030	G	A	IG	IG	IG	--	--	--	--	3
724	1761173	A	C	CDS	1741	Streptopain inhibitor	Non	Phe_F	Cys_C	Rad	3
725	1762318	C	A	CDS	1742	Streptococcal cysteine protease SpeB	Non	Gly_G	STOP	STOP	1
726	1762807	C	T	IG	IG	IG	--	--	--	--	1
727	1762893	T	C	CDS	1743	unknown function	Syn	Leu_L	Leu_L	--	1
728	1763133	T	C	IG	IG	IG	--	--	--	--	1
729	1763553	G	A	CDS	1744	RopB; Rgg-like transcription regulator	Non	Val_V	Ile_I	Con	20
730	1763599	G	A	CDS	1744	RopB; Rgg-like transcription regulator	Non	Cys_C	Tyr_Y	Rad	1
731	1763788	G	A	CDS	1744	RopB; Rgg-like transcription regulator	Non	Cys_C	Tyr_Y	Rad	5
732	1763841	C	T	CDS	1744	RopB; Rgg-like transcription regulator	Non	Pro_P	Ser_S	Con	90
733	1763985	A	T	CDS	1744	RopB; Rgg-like transcription regulator	Non	Asn_N	Tyr_Y	Rad	1
734	1763996	G	A	CDS	1744	RopB; Rgg-like transcription regulator	Non	Met_M	Ile_I	Con	1
735	1764086	T	A	CDS	1744	RopB; Rgg-like transcription regulator	Non	Asn_N	Lys_K	Con	1
736	1764204	T	C	CDS	1744	RopB; Rgg-like transcription regulator	Non	Tyr_Y	His_H	Con	2
737	1764211	G	A	CDS	1744	RopB; Rgg-like transcription regulator	Non	Arg_R	Gln_Q	Con	2
738	1764244	G	T	CDS	1744	RopB; Rgg-like transcription regulator	Non	Gly_G	Val_V	Rad	1
739	1764267	G	A	CDS	1744	RopB; Rgg-like transcription regulator	Non	Ala_A	Thr_T	Con	1
740	1764337	T	G	CDS	1744	RopB; Rgg-like transcription regulator	Non	Phe_F	Cys_C	Rad	1
741	1771377	C	A	CDS	1750	PTS system, cellobiose-specific IIC	Non	Ala_A	Ser_S	Con	1
742	1772884	G	A	IG	IG	IG	--	--	--	--	1
743	1779638	G	C	CDS	1760	unknown function	Non	Glu_E	Gln_Q	Con	1
744	1780162	C	T	CDS	1761	unknown function	Non	Ala_A	Val_V	Con	6
745	1790135	G	A	CDS	1768	Transcriptional regulator of class III stress genes	Syn	Asp_D	Asp_D	--	2
746	1791355	G	A	IG	IG	IG	--	--	--	--	4
747	1794512	C	T	CDS	1772	Imidazolonepropionase	Non	Ala_A	Thr_T	Con	3
748	1795623	G	A	CDS	1773	Urocanate hydratase	Non	Asp_D	Asn_N	Con	1
749	1796147	G	T	CDS	1773	Urocanate hydratase	Non	Met_M	Ile_I	Con	3
750	1798264	A	G	CDS	1774	Glutamate formiminotransferase	Non	Asn_N	Ser_S	Con	1
751	1806778	G	A	CDS	1781	Transcriptional regulator, LuxR family	Syn	Gly_G	Gly_G	--	18
752	1812955	A	G	CDS	1784	Neutral endopeptidase O	Non	Tyr_Y	His_H	Con	3
753	1815075	C	T	IG	IG	IG	--	--	--	--	3
754	1815253	A	T	CDS	1786	PTS system, trehalose-specific IIB component	Non	Tyr_Y	Asn_N	Rad	2

Table S2 cont.

No.	Locus	Ref	Test	CDS /IG	Locus Tag	Annotation	Syn /Non	Ref	Test	Con /Rad	No. of Strains
755	1817279	C	T	IG	IG	IG	--	--	--	--	5
756	1818197	T	C	IG	IG	IG	--	--	--	--	3
757	1821510	G	A	CDS	1792	Acetyltransferase	Non	Pro_p	Leu_L	Con	1
758	1823995	T	C	CDS	1794	Ribonucleotide reductase	Non	Asp_D	Gly_G	Con	1
759	1824906	G	A	CDS	1794	Ribonucleotide reductase	Syn	Asn_N	Asn_N	--	1
760	1829072	T	C	IG	IG	IG	--	--	--	--	2
761	1834388	A	G	CDS	1805	DNA mismatch repair protein MutL	Syn	Ile_I	Ile_I	--	2
762	1838947	A	G	CDS	1807	biofilm formation protein YmcA	Non	Ser_S	Pro_P	Con	1
763	1839008	C	T	IG	IG	IG	--	--	--	--	1
764	1841544	G	A	CDS	1809	Arginyl-tRNA synthetase	Non	Gly_G	Asp_D	Con	1
765	1841704	C	T	CDS	1810	unknown function	Syn	Ser_S	Ser_S	--	20
766	1842608	C	T	CDS	1811	Transporter	Non	Gly_G	Glu_E	Con	1
767	1850661	A	C	CDS	1820	unknown function	Non	Thr_T	Pro_P	Con	2
768	1852034	T	C	CDS	1822	unknown function	Non	Val_V	Ala_A	Con	1
769	1852385	G	A	IG	IG	IG	--	--	--	--	2
770	1852396	G	A	IG	IG	IG	--	--	--	--	2
771	1860162	A	G	CDS	1832	Transcriptional regulator, TetR family	Non	Lys_K	Glu_E	Con	1
772	1865795	T	A	CDS	1837	Phosphoesterase, DHH family protein	Non	Leu_L	Phe_F	Con	1
773	1868362	T	C	IG	IG	IG	--	--	--	--	24
774	1868487	A	G	IG	IG	IG	--	--	--	--	1
775	1868573	C	T	IG	IG	IG	--	--	--	--	1
776	1871539	A	G	CDS	1842	L-serine dehydratase	Non	Thr_T	Ala_A	Con	1
777	1875003	G	A	CDS	1845	ATPase component of ECF transporters	Non	Thr_T	Met_M	Con	1
778	1875799	G	A	CDS	1846	ATPase component of ECF transporters	Syn	Asn_N	Asn_N	--	1
779	1875805	T	G	CDS	1846	ATPase component of ECF transporters	Syn	Gly_G	Gly_G	--	1
780	1876848	A	C	CDS	1848	Transcriptional regulator	Syn	Ser_S	Ser_S	--	1
781	1878245	T	C	CDS	1849	peptidase, M16 family	Non	Asn_N	Ser_S	Con	1
782	1880381	A	T	IG	IG	IG	--	--	--	--	1
783	1880442	A	G	IG	IG	promoter of hasABC operon	--	--	--	--	1
784	1880446	G	T	IG	IG	promoter of hasABC operon	--	--	--	--	1
785	1880453	G	A	IG	IG	promoter of hasABC operon	--	--	--	--	2
786	1880456	G	T	IG	IG	promoter of hasABC operon	--	--	--	--	1
787	1880457	C	T	IG	IG	promoter of hasABC operon	--	--	--	--	2
788	1880458	C	T	IG	IG	promoter of hasABC operon	--	--	--	--	1
789	1881838	G	A	CDS	1852	UDP-glucose dehydrogenase hasB	Non	Gly_G	Arg_R	Rad	10
790	1881840	A	T	CDS	1852	UDP-glucose dehydrogenase hasB	Syn	Gly_G	Gly_G	--	10
791	1881985	T	C	CDS	1852	UDP-glucose dehydrogenase hasB	Syn	Leu_L	Leu_L	--	1
792	1881986	T	G	CDS	1852	UDP-glucose dehydrogenase hasB	Non	Leu_L	STOP	STOP	1
793	1882099	G	C	CDS	1852	UDP-glucose dehydrogenase hasB	Non	Glu_E	Gln_Q	Con	1
794	1882120	C	T	CDS	1852	UDP-glucose dehydrogenase hasB	Non	His_H	Tyr_Y	Con	1
795	1884495	T	G	CDS	1854	unknown function	Non	Phe_F	Val_V	Con	1
796	1889178	A	G	CDS	1858	Tryptophanyl-tRNA synthetase	Non	Tyr_Y	His_H	Con	1
797	1891225	C	T	IG	IG	IG	--	--	--	--	1
798	1896371	C	T	CDS	1862	putative transposase	Non	Arg_R	Gln_Q	Con	8
799	1898296	A	G	IG	IG	IG	--	--	--	--	1
800	1900461	C	T	CDS	1865	Chromosome partitioning protein ParB	Non	Leu_L	Phe_F	Con	1
801	1900477	C	T	CDS	1865	Chromosome partitioning protein ParB	Non	Ser_S	Phe_F	Rad	1

1. Abbreviations: Ref, reference strain MGAS315; CDS, coding sequence; IG, intergenic; Syn, synonymous; Non, nonsynonymous; Con, conservative amino acid change; Rad, radical amino acid change.

**Table S3. Nonsense single nucleotide polymorphisms inferred to cause prematurely terminated products.**

Locus	Gene (SpyM3_)	Function	SNP		Codon		Ref. (aa)	Product	
			Ref.	Test	Ref.	Test		Test (aa)	Truncated %
195,125	0164	Multidrug resistance ABC transporter ATP-binding and permease protein	C	T	CAA	TAA	594	457	23.1
283,739	0244	TCS response regulator CovR/CsrR	G	T	GAA	TAA	228	36	84.2
284,064	0244	TCS response regulator CovR/CsrR	C	A	TCA	TAA	228	144	36.8
284,614	0245	TCS sensor kinase CovS/CsrS	C	T	CGA	TGA	500	97	80.6
284,719	0245	TCS sensor kinase CovS/CsrS	G	T	GAA	TAA	500	132	73.6
350,013	0307	Unknown function	C	A	GAA	TAA	176	113	35.8
727,093	--	Unknown function	G	A	CGA	TGA	95	19	80.0
832,888	0778	Glutamine amidotransferase, class I	A	T	TTA	TAA	231	103	55.4
855,352	0804	Pseudouridylate synthases, 23S RNA-specific	C	T	CAA	TAA	325	184	43.4
872,314	0819	Rossmann fold nucleotide-binding protein	C	T	CGA	TGA	278	239	14.0
935,891	0877	Phosphate transport ATP-binding protein PstB	A	T	TTA	TAA	253	194	23.3
1,023,900	0982	Maltose operon transcriptional repressor MalR	G	A	CAA	TAA	339	153	54.9
1,051,686	1012	Transposase	C	T	CAA	TAA	268	21	92.2
1,094,245	1051	Chloride channel protein	G	A	TGG	TAG	437	162	62.9
1,104,478	1061	Major facilitator:Oxalate:Formate Antiporter	C	T	TGG	TGA	398	38	90.5
1,296,749	1289	Sugar-binding protein	C	A	GAA	TAA	481	275	42.8
1,501,014	1509	Guanine-hypoxanthine permease	C	A	GAA	TAA	486	322	33.7
1,583,782	1583	Unknown function	C	T	TGG	TAG	97	42	56.7
1,667,220	1654	PTS system, lactose-specific IIB component	G	A	CAA	TAA	565	461	18.4
1,762,318	1742	Streptococcal cysteine protease SpeB (Streptopain)	C	A	GGA	TGA	398	92	76.9
1,881,986	1852	UDP-glucose dehydrogenase HasB	T	G	TTA	TGA	402	64	84.1



**Table S4. Insertions and deletions inferred to cause reading frame shifts resulting in prematurely terminated products.<sup>1</sup>**

Locus	Gene (SpyM3_)	Function	Insertions and Deletions				Ref. (aa)	Product		
			Indel	Sequence	Size (nt)	Type		RFS (aa)	Test (aa)	Truncated %
9,109	0007	Ribosome-associated heat shock protein	Ins	A	1	HPNT	90	69	73	18.9
45,781	0024	IMP cyclohydrolase	Del	--	42	--	515	512	515	0.0
62,130	0037	Alcohol dehydrogenase	Ins	A	1	HPNT	338	184	186	45.0
130,209	0107	Transcriptional regulator, LysR family	Ins	T	1	HPNT	303	137	159	47.5
236,940	0204	Unknown function	Del	G	1	HPNT	641	521	571	10.9
283,972	0244	TCS response regulator CovR/CsrR	Ins	G	1	--	228	113	122	46.5
284,344	0245	TCS sensor kinase CovS/CsrS	Ins	A	1	HPNT	500	7	11	97.8
284,405	0245	TCS sensor kinase CovS/CsrS	Del	T	1	HPNT	500	27	35	93.0
284,425	0245	TCS sensor kinase CovS/CsrS	Ins	--	11	VNTR	500	34	39	92.2
284,753	0245	TCS sensor kinase CovS/CsrS	Ins	C	1	--	500	143	165	67.0
285,537	0245	TCS sensor kinase CovS/CsrS	Del	GAAAA	5	VNTR	500	405	407	18.6
302,388	0260	Membrane protein	Ins	A	1	--	229	224	228	0.4
358,676	0317	Metal transporter of streptococcus regulator MtsR	Ins	T	1	HPNT	215	5	13	94.0
373,410	0332	67kDa Myosin-crossreactive streptococcal antigen	Del	T	1	HPNT	590	499	565	4.2
423,167	0378	Shikimate 5-dehydrogenase	Del	T	1	HPNT	290	271	277	4.5
425,140	0381	Unknown function	Ins	A	1	HPNT	328	117	121	63.1
426,253	0382	Archaeal S-adenosylmethionine synthetase	Del	C	1	HPNT	399	156	164	58.9
428,141	0384	Dolichyl-phosphate transferase, involved in cell wall biogenesis	Del	T	1	HPNT	329	50	50	84.8
431,132	0386	ABC transporter membrane-spanning permease - macrolide efflux	Del	A	1	HPNT	391	62	75	80.8
435,966	0394	Plasmid stabilization system toxin protein	Ins	A	1	HPNT	101	74	85	15.8
438,538	0398	Probable transposase TnpA	Del	T	1	HPNT	106	95	104	1.9
444,105	0403	Beta-glucoside bgl operon antiterminator, BglG family	Ins	A	1	HPNT	280	34	37	86.8
450,256	0408	Unknown function	Del	T	1	HPNT	65	51	53	18.5
493,748	0450	2-dehydro-3-deoxygluconate kinase	Del	T	1	HPNT	333	31	42	87.4
501,845	0457	Hydroxyacylglutathione hydrolase	Del	A	1	HPNT	232	9	13	94.4
520,333	0473	Chloride channel protein	Ins	A	1	--	420	268	279	33.6
520,439	0473	Chloride channel protein	Del	A	1	HPNT	420	304	304	27.6
624,022	0566	Glycerophosphoryl diester phosphodiesterase	Del	A	1	HPNT	577	475	532	7.8
651,248	0591	5'-nucleotidase	Del	C	1	HPNT	670	182	188	71.9
652,457	0592	GTP pyrophosphokinase	Del	A	1	HPNT	216	147	152	29.6
665,479	0606	Unknown function	Ins	A	1	--	248	74	75	69.8
665,981	0606	Unknown function	Ins	AA	2	HPNT	248	241	247	0.4
695,107	0633	Unknown function	Del	T	1	HPNT	58	31	55	5.2

Table S4 cont.

Locus	Gene (SpyM3_)	Function	Insertions and Deletions				Product			
			Indel	Sequence	Size (nt)	Type	Ref. (aa)	RFS (aa)	Test (aa)	Truncated %
712,189	0651	TPR-repeat-containing protein	Ins	T	1	HPNT	415	204	204	50.8
791,356	0738	Streptococcal collagen-like protein ScIB	Del	CAAAA	5	VNTR	573	8	9	98.4
856,016	0805	Pneumococcal vaccine antigen A homolog	Del	A	1	HPNT	199	79	92	53.8
857,798	0806	ABC transporter, ATP-binding/permease protein	Del	T	1	HPNT	536	432	439	18.1
877,877	0823	Unknown function	Del	A	1	HPNT	104	42	59	43.3
903,155	0847	ABC transporter ATP-binding protein	Ins	T	1	HPNT	514	294	295	42.6
927,833	0870	16S rRNA m(2)G 1207 methyltransferase	Del	T	1	HPNT	197	161	181	8.1
1,064,894	1025	Putative PBP 5 synthesis repressor	Del	T	1	HPNT	483	37	37	92.3
1,070,306	1031	Ribosomal-protein-S5p-alanine acetyltransferase	Del	T	1	HPNT	180	148	153	15.0
1,104,584	1061	Major facilitator:Oxalate:Formate Antiporter	Ins	T	1	HPNT	398	5	30	92.5
1,135,220	1092	Unknown function	Del	TTCTATTT	8	VNTR	57	17	24	57.9
1,185,432	1159	Deoxyribodipyrimidine photolyase	Del	T	1	HPNT	469	116	171	63.5
1,188,542	1163	ATP-dependent Clp protease, ATP-binding subunit ClpE	Ins	A	1	HPNT	760	181	183	75.9
1,380,210	1377	Transcriptional regulators, LysR family	Ins	T	1	HPNT	329	73	77	76.6
1,462,554	1468	Glycerol kinase	Del	--	19	VNTR	508	296	316	37.8
1,550,607	1558	Heme ABC transporter (Streptococcus), ATP-binding protein	Del	A	1	HPNT	278	112	127	54.3
1,610,634	1608	DNA polymerase III alpha subunit	Del	C	1	HPNT	208	38	69	66.8
1,755,860	1735	ABC transporter ATP-binding protein	Del	T	1	--	224	221	222	0.9
1,881,860	1852	UDP-glucose dehydrogenase HasB	Ins	T	1	--	402	22	28	93.0
1,881,975	1852	UDP-glucose dehydrogenase HasB	Ins	A	1	--	402	60	61	84.8

1. Abbreviations: Ref, reference strain MGAS315; RFS, reading frame shift; Ins, insertion; Del, deletion; HPNT, homopolymeric nucleotide tract; VNTR, variable number tandem repeat.

**Table S6. Core genes significantly overabundant in single nucleotide polymorphisms.**

Locus Tag	Gene	Function	Size (bp)	Exp. SNPs <sup>1</sup>		Obs. SNPs	$\chi^2$	$P^2$	$P_{Bon}^3$	$P_{B\&H}^4$
				$n = 1$	$n = 801$					
SpyM3_1744	<i>ropB/rgg</i>	Rgg-like transcription regulator	843	0.000506	0.405	12	331.58	$4.36 \times 10^{-74}$	$6.75 \times 10^{-71}$	$6.75 \times 10^{-71}$
SpyM3_1017	IS904_A	transposase ORF-A	324	0.000195	0.156	6	219.18	$1.36 \times 10^{-49}$	$2.11 \times 10^{-46}$	$1.06 \times 10^{-46}$
SpyM3_0245	<i>covS/csrS</i>	TCS sensor histidine kinase	1503	0.000902	0.723	13	208.52	$2.90 \times 10^{-47}$	$4.49 \times 10^{-44}$	$1.50 \times 10^{-44}$
SpyM3_1727	<i>emm3</i>	antiphagocytic M protein	1746	0.001048	0.840	12	148.32	$4.03 \times 10^{-34}$	$6.25 \times 10^{-31}$	$1.56 \times 10^{-31}$
SpyM3_0244	<i>covR/csrR</i>	TCS response regulator	687	0.000412	0.330	6	97.29	$6.00 \times 10^{-23}$	$9.30 \times 10^{-20}$	$1.86 \times 10^{-20}$
SpyM3_1012	IS861A	transposase ORF-A	807	0.000485	0.388	5	54.80	$1.33 \times 10^{-13}$	$2.07 \times 10^{-10}$	$3.44 \times 10^{-11}$
SpyM3_1655	<i>lacF</i>	phosphotransferase system lactose specific IIA component	318	0.000191	0.153	3	53.00	$3.34 \times 10^{-13}$	$5.17 \times 10^{-10}$	$7.38 \times 10^{-11}$
SpyM3_1852	<i>hasB</i>	HA capsule synthesis UDP-glucose dehydrogenase	1209	0.000726	0.581	6	50.49	$1.20 \times 10^{-12}$	$1.85 \times 10^{-9}$	$2.31 \times 10^{-10}$
SpyM3_1527	<i>fabH</i>	fatty acid biosynthesis H, beta-keto-acyl-ACP synthase	975	0.000585	0.469	5	43.78	$3.76 \times 10^{-11}$	$5.68 \times 10^{-8}$	$6.32 \times 10^{-9}$
SpyM3_1466	<i>glpF.1</i>	glycerol uptake facilitator protein	702	0.000422	0.338	4	39.73	$2.92 \times 10^{-10}$	$4.52 \times 10^{-7}$	$4.52 \times 10^{-8}$
SpyM3_1011	IS861B	transposase ORF-B	468	0.000281	0.225	3	34.21	$4.95 \times 10^{-9}$	$7.66 \times 10^{-6}$	$6.97 \times 10^{-7}$
SpyM3_0987	--	transposase	246	0.000148	0.118	2	29.93	$4.49 \times 10^{-8}$	$6.95 \times 10^{-5}$	$5.79 \times 10^{-6}$
SpyM3_1160	<i>glnQ</i>	glutamine transport ATP-binding protein	735	0.000441	0.353	3	19.81	$8.54 \times 10^{-6}$	$1.32 \times 10^{-2}$	$1.02 \times 10^{-3}$
SpyM3_0674	--	fatty acid biosynthesis A-like, acyl-ACP thioesterase	753	0.000452	0.362	3	19.21	$1.17 \times 10^{-5}$	$1.81 \times 10^{-2}$	$1.29 \times 10^{-3}$
SpyM3_0819	<i>smf</i>	Smf family DNA processing protein	837	0.000503	0.403	3	16.76	$4.24 \times 10^{-5}$	$6.57 \times 10^{-2}$	$4.38 \times 10^{-3}$
SpyM3_1509	--	guanine-hypoxanthine permease	1461	0.000877	0.703	4	15.47	$8.37 \times 10^{-5}$	$1.30 \times 10^{-1}$	$8.10 \times 10^{-3}$
SpyM3_0085	--	adenine-specific methyltransferase	954	0.000573	0.459	1	14.07	$1.76 \times 10^{-4}$	$2.72 \times 10^{-1}$	$1.60 \times 10^{-2}$
SpyM3_0353	<i>smpB</i>	binds to tmRNA, required for ribosome binding	468	0.000281	0.225	2	14.00	$1.83 \times 10^{-4}$	$2.84 \times 10^{-1}$	$1.58 \times 10^{-2}$
SpyM3_1488	--	phosphotransferase system IIA component	474	0.000285	0.228	1	13.77	$2.06 \times 10^{-4}$	$3.19 \times 10^{-1}$	$1.68 \times 10^{-2}$
SpyM3_0738	<i>sclB</i>	streptococcal collagen-like protein B	1722	0.001034	0.828	3	12.15	$4.92 \times 10^{-4}$	$7.61 \times 10^{-1}$	$3.81 \times 10^{-2}$
SpyM3_0999	--	hypothetical protein of unknown function	153	0.000092	0.074	3	11.66	$6.37 \times 10^{-4}$	$9.87 \times 10^{-1}$	$4.70 \times 10^{-2}$
SpyM3_0011	<i>hpt</i>	hypoxanthine-guanine phosphoribosyltransferase	543	0.000326	0.261	1	11.58	$6.67 \times 10^{-4}$	1.03	$4.70 \times 10^{-2}$

1. The mean chance expectation for a polymorphism occurring in a given gene for a single mutation event is the size of the gene relative to the core genome (for *ropB*, 843 bp/ 1,665,462 bp = 0.000506), and for multiple events, is the expectation for a single event multiplied by the number of total events (for *ropB*,  $0.000506 \times 801 = 0.405$ ).

2. Chi square probability for one degree of freedom.

3. Probability corrected for multiple tests ( $n = 1549$ ) by the method of Bonferroni (uncorrected probability multiplied by the number of tests).

4. Probability corrected for multiple tests ( $n = 1549$ ) by the method of Benjamini and Hochberg (uncorrected probability multiplied by the number of tests divided by the rank).

**Table S5. Single nucleotide polymorphism accretion rate estimation.**

<b>Progenitor-Descendant CC Pairs Compared<sup>1</sup></b>	<b>Genetic Distance Between Strains (<math>\Delta nt</math>)<sup>2</sup></b>	
	<b>~5yr</b>	<b>~1yr</b>
<b>Subclones (Epidemic) Clonal complexes Number of sequenced strains Number of comparisons</b>	<b>Range Mean Net Mean</b>	<b>Mean Net Mean</b>
SC5 (II) $\leftrightarrow$ SC8 (III) CC2 $\leftrightarrow$ CC3 $n = 7 \leftrightarrow n = 5$ 35	13 $\leftrightarrow$ 33 20.429 9.724	4.086 1.945
SC5 (II) $\leftrightarrow$ SC5 (III) CC2 $\leftrightarrow$ CC5 $n = 7 \leftrightarrow n = 3$ 21	6 $\leftrightarrow$ 42 20.095 1.525	4.419 0.305
SC1 (II) $\leftrightarrow$ SC9 (III) CC7 $\leftrightarrow$ CC6 $n = 3 \leftrightarrow n = 6$ 18	18 $\leftrightarrow$ 45 32.833 11.265	6.567 2.253
SC3 (I) $\leftrightarrow$ SC3 (II) CC12 $\leftrightarrow$ CC13 $n = 7 \leftrightarrow n = 5$ 35	12 $\leftrightarrow$ 30 19.943 11.467	3.989 2.293

1. Progenitor-descendant clonal complex pairs were identified based on relationships in the 280 sequence-validated SNP neighbor-joining tree (Fig. 3B).
2. Genetic distance between strains, mean SNPs and net mean SNPs between clonal complexes were determined using Molecular Evolutionary Genetic Analysis (MEGA) software ([www.megasoftware.net](http://www.megasoftware.net)).

Estimated mean SNPs accumulated per strain per year :

$$(4.086 + 4.419 + 6.567 + 3.989) / 4 = \underline{4.765}$$

Estimated net mean SNPs accumulated per strain per year :

$$(1.945 + 0.305 + 2.253 + 2.293) / 4 = \underline{1.699}$$

**Table S7. GAS serotype M3 SLVG clonal complexes and invasive disease types associations.<sup>1</sup>**

SLVG CC	<u>ART</u> (n = 35)		<u>BAC</u> (n = 54)		<u>LRT</u> (n = 53)		<u>NF</u> (n = 52)		<u>STI</u> (n = 77)		<u>URT</u> (n = 17)		<u>Other</u> (n = 55)	
	Obs	Exp <sup>2</sup> P	Obs	Exp P	Obs	Exp P	Obs	Exp P	Obs	Exp P	Obs	Exp P	Obs	Exp P
<b>1</b> (n = 20)	4	2.03 0.0924	3	3.14 0.6139	2	3.08 0.3788	2	3.02 0.3930	3	4.48 0.3071	<b>4</b>	<b>0.99</b> <b>0.0106</b>	2	3.26 0.3383
<b>2</b> (n = 36)	4	3.66 0.4769	<b>12</b>	<b>5.65</b> <b>0.0046</b>	6	5.55 0.4996	<b>0</b>	<b>5.44</b> <b>0.0019</b>	8	8.06 0.5857	2	1.78 0.4372	4	5.86 0.2669
<b>3</b> (n = 10)	1	1.02 0.7298	1	1.57 0.5167	1	1.54 0.5273	3	1.51 0.1317	1	2.24 0.3042	0	0.49 0.5982	3	1.63 0.1499
<b>4</b> (n = 2)	0	0.2 0.8066	0	0.31 0.7103	1	0.31 0.2848	0	0.3 0.7202	0	0.45 0.6019	0	0.1 0.9035	1	0.33 0.2995
<b>5</b> (n = 14)	0	1.42 0.2159	4	2.2 0.1970	2	2.16 0.6312	1	2.12 0.3467	3	3.13 0.6149	1	0.69 0.3676	3	2.28 0.5305
<b>6</b> (n = 14)	0	1.42 0.2159	1	2.2 0.3246	2	2.16 0.6312	1	2.12 0.3467	3	3.13 0.6149	0	0.69 0.4850	7	2.28 0.0029
<b>7</b> (n = 8)	<b>3</b>	<b>0.81</b> <b>0.0332</b>	1	1.26 0.6347	0	1.23 0.2583	2	1.21 0.2419	2	1.79 0.7438	0	0.4 0.6638	0	1.3 0.2375
<b>8</b> (n = 19)	1	1.93 0.4034	4	2.98 0.3551	3	2.93 0.6671	2	2.87 0.4305	4	4.25 0.5738	0	0.94 0.3717	5	3.09 0.1416
<b>9</b> (n = 12)	0	1.22 0.2699	3	1.88 0.3095	2	1.85 0.7233	4	1.81 0.0682	2	2.69 0.4733	0	0.59 0.5388	1	1.95 0.3910
<b>10</b> (n = 4)	0	0.41 0.6497	0	0.63 0.5034	<b>3</b>	<b>0.62</b> <b>0.0124</b>	1	0.6 0.3725	0	0.9 0.3611	0	0.2 0.8158	0	0.65 0.4896
<b>11</b> (n = 11)	0	1.12 0.3016	4	1.73 0.0591	1	1.69 0.4738	2	1.66 0.4527	1	2.46 0.2524	0	0.54 0.5678	3	1.79 0.3106
<b>12</b> (n = 14)	1	1.42 0.5733	3	2.2 0.3065	3	2.16 0.3069	4	2.12 0.1970	3	3.13 0.6149	0	0.69 0.4850	0	2.28 0.0788
<b>13</b> (n = 20)	3	2.03 0.3086	2	3.14 0.3650	3	3.08 0.6285	4	3.02 0.3524	6	4.48 0.2772	2	0.99 0.1913	0	3.26 0.0256
<b>14</b> (n = 17)	3	1.73 0.3159	0	2.67 0.0508	5	2.62 0.1245	5	2.57 0.1240	2	3.81 0.2249	1	0.84 0.3843	1	2.77 0.2027
<b>15</b> (n = 6)	1	0.61 0.3611	2	0.94 0.1875	1	0.92 0.7677	0	0.91 0.3711	0	1.34 0.2159	0	0.3 0.7361	2	0.98 0.1963
<b>16</b> (n = 10)	0	1.02 0.3369	<b>4</b>	<b>1.57</b> <b>0.0443</b>	<b>5</b>	<b>1.54</b> <b>0.0099</b>	0	1.51 0.1896	0	2.24 0.0763	0	0.49 0.5982	1	1.63 0.4957
<b>17</b> (n = 9)	1	0.92 0.7698	1	1.41 0.5742	0	1.39 0.2176	2	1.36 0.4893	1	2.01 0.3644	1	0.44 0.3023	3	1.47 0.1246
<b>18</b> (n = 10)	1	1.02 0.7298	0	1.57 0.1768	0	1.54 0.1831	0	1.51 0.1896	<b>5</b>	<b>2.24</b> <b>0.0486</b>	1	0.49 0.3198	3	1.63 0.1499
<b>19</b> (n = 29)	3	2.95 0.6609	4	4.55 0.5098	2	4.47 0.1427	6	4.38 0.2866	10	6.49 0.0719	0	1.43 0.2155	4	4.72 0.4739
<b>20</b> (n = 10)	2	1.02 0.1998	0	1.57 0.1768	1	1.54 0.5273	1	1.51 0.5381	3	2.24 0.5356	1	0.49 0.3198	2	1.63 0.4573
<b>21</b> (n = 3)	0	0.31 0.7241	0	0.47 0.5981	0	0.46 0.6044	0	0.45 0.6107	<b>3</b>	<b>0.67</b> <b>0.0109</b>	0	0.15 0.8586	0	0.49 0.5858
<b>Other</b> (n = 66)	7	6.72 0.4753	5	10.36 0.0278	10	10.17 0.5605	12	9.98 0.2801	17	14.77 0.3272	4	3.26 0.5409	11	10.74 0.6194

1. Abbreviations: ART, arthritis; BAC, bacteremia; LRT, lower respiratory tract; NF, necrotizing fasciitis; STI, soft tissue infection; URT, upper respiratory tract; SLVG, single locus variant group; CC, clonal complex Obs, observed; Exp, expected; P, probability.

2. Expected values for a given infection type are based on the relative proportion of that infection type to all invasive cases in the study set. For CC2 and NF, the relative proportion of NF is 52/344 = 0.151 and for CC2 the expected number of NF cases is 36 × 0.151 = 5.4.

**Table S8. Primers used for mass spectroscopy based single nucleotide polymorphism analysis.<sup>1</sup>**

No.	Locus	Amplification Primer 1	Amplification Primer 2	Single Base Extension Primer
1	1850	GCCATTCCTATTCTTTCATC	GATATTTGACCGTTAGACCC	GACAGTTACTCCTGTAGGAGTGACTT
2	10114	GCCATTTGAGTGGCATGAAC	CTTAGGCTTTATTTCAGCCCC	CTAGGTACGCAAACAATTCCTTCTTCT
3	16387	GGCAATAGCAAAAGCAATCC	CCATTCCTACCAATGTTGGC	GGTTGGCAATTATCATTTGTCT
4	37285	GGATATTGCAGAAGCAGAGC	GCGGCAAAGTAATGTCCTTG	AATCAACAGGGTTCAAAAGA
5	44364	ATCTCTGGGAATCCAGTCAC	AACTGGTGGGACTAAGGTTG	CCGAAGGTTGCCCTTGATAA
6	45232	TAGCTAGCGCTTCTTCTGAG	CGACAGCTAAGAAGATGCAC	CACCCTATCTTCTTGAAATCATCATC
7	46653	GACCGTAAGCAGCATTGAC	GCGAAAAATCAGGGTTATCC	ATCCTGTGAATAAGACACCAA
8	48112	GAGTTCAACTCACGTTTCGG	GAGCGAAATCAGAAGTCAGG	AGTCAGGCGAGGGAGGATAA
9	61350	ACTACATTTTCTGACACTGG	CCAGGACGACGTAAGTGATT	TTTTAATGTTGATACCAGGAAAT
10	66294	GACTTCTCAGTGATAACTGG	CGTGAACGCAGACAACTTC	GGACAATTGAGGAGGTTCTT
11	67203	GATACCTTTCCAACGGCTAC	TACATGCCGTGCAACTATCG	ACTATCGGTACTGTTGGTAA
12	67875	ACGAGCCATTGCTTTAGCTG	GCAGCTGACGACAAGAAAAC	TTCGTTAATAGGAGGAGGA
13	79640	CCTCCAAATTTAAAAGAGGAT	TGCATCGCAAACAGATGCTC	CAAACCTATATTGTTAGTTTTATCCC
14	93041	AGAACCTGGCGATCTAACTG	GTGTTAGGTTTTGTTGGTGG	ATTTGTTGGTGGTAGAGATTA
15	98549	TTTGTGCTAAGATGGGTGC	ATTTCAGCGATTTTCAGCTGCC	TATGCCAAGTCCACACGTT
16	107126	ACAGCTTCTGTATGATCGTG	CGTATCGGTCTCAAGGATTC	TTGGCAAAAAAGAAGAGC
17	113828	GTCATTAGCGTACTGTTTCG	TGCCAAAGTTCGTGGTGCTG	TTAGACTCTAGCTGAGGTT
18	118018	TCTGACTTGGATGGAAAGTG	GCTACCCAGACTATAGTCCG	AGATGGAAGTAAGGAGTATCAAG
19	119965	GGCTATTTGGTCGTAAAGGG	GCTTTCAATAGTGGATTCTC	CTCGCATTTAAGCGATTTAGGTAATT
20	130696	CCTCTTTTCGATAACGTTTAG	CACCACTTCTTTAGTCATAG	CTTTTTTCATTATTGAAACGCTTAC
21	131210	GGGTTGAGTGATGCTTTTGC	CTTCTCTGGTCAGTCCATAC	GGTGCACAATATTTTCGGCTGTT
22	131636	CATCAATTGTCAGCTGAGCC	GACCCAAGTATTATGGGCTG	CCCCCTACCTGCTACGAAAAAGG
23	139513	GTTGGCATCGCTAGTCAAAC	AAGAAGAAAGCTCCCGCATC	CTCCCGCATCATTTGATA
24	144672	CATGTCCAGTCTAATACGGC	GTCCATCTCACTATTTGAGC	ACCTCACTATTTGAGCTGAGGT
25	157102	AATAGCAGCCATACCAGATG	CTGAAGCTTTAGAAGACGGG	GACTGTTTAGAAAAAGCGACTTAT
26	163387	CCTGCCAAAGTCCATAAGAG	GCATTCAAGTCCATACCTTC	GATCTTTAGATTCTTCTACCATTT
27	164253	GTCTTTTGCAAATTGCAAGCG	TTTATGGCGCGTCCATTGG	CAGCAACGCCAAAACAGTTTA
28	170064	AACACGCTGCTGCATCCAC	CCAGCGATTGGCAGAAAAAG	TTAAAGACCAGTACCCAAA
29	183592	GTGTTTTTAGCTATGGAAG	CATTTTCATAATCGCATGGG	TCTTTTAAATTTTCATCTGCATAAG
30	184244	CTCATGCTATCTTCTTAAC	CTTAAAAACGTACTAGTGC	TAAAAACGTACTAGTGCTATTTCTC
31	186843	TGACGGCCTGTGTCATAAAG	TCAAGGCAACATGCTACTAC	CATGCTACTACTAGATAAGCTTGAAA
32	195125	GTCATGAATATGGGTGCGAC	ATTCGTTTTGGTGATCAGAC	TTGGTGATCAGACCATTAGT
33	201055	GGAAATCTTCTCTCGTGAGC	TTCTGGTCAAAGGCAGCTGG	TGCTTGCTAAGGCGATTTT
34	201573	GGGACGTGATAAATCCATTG	TGTCGTGACTTTTTAACTCC	GGGAAAGGTGAACGATACAGCCTTA
35	205327	GAAATGTTGCTTTTTTCGTCC	TGAAATGTCTGTTTGAACGG	CTGTCTGTTTGAACGGTGGGAGG
36	220205	TGCCCCATCTATTCTAGCAG	GTGAACGTATTTCGGCATCTG	AGAGCTAGCTCAGAAGCAAAATTAGC
37	225263	AGAGACTGCATTGTCAGAAC	GACGTTTTGACTTGTTCAC	CACTTGTTCCACTAAATCCTGA
38	233903	GCTGGTAAAATGTTGCGAAC	GCAACTGCTCTTGGCTTTAC	GGCCCATGCGTAAAATCGTG
39	242944	ACATAGACTAACCTTGCTCC	CGATTATGGAGCATCATGCC	TATTATGGAGCATCATGCCAATATC
40	245533	GTGGCGTTAGAACATGAAGC	ATCTCTGTGGCTTCGCTTTC	TTTCTGATAAGCCACGG
41	254542	GAAGCCAATGAGCTCATTCG	GGTTGCCCGTTTCATTCAAAC	GGGCGGGATCTTGAAAAATC
42	261079	GCTCGAGGACTTATCCAAA	GCTATTTTCATAACCCACTC	TTCGAGTCTAACAAGAATCCGTAA

No.	Locus	Amplification Primer 1	Amplification Primer 2	Single Base Extension Primer
43	262601	ATCATCGTAGTCACCTCATC	GTATATTTAACTATCACGCC	CCCTATCACGCCAAACGCATG
44	264283	ATTAGCCGTGCTTTGAAGCC	ACCTCGTTGCGATCAAATCC	ATCGATCAAATCCATAACGAGCAAT
45	277587	ATTGGCTGCAGAACTTCGTG	CCTTATCCCCTTGTCTTGAG	TTGAGTGAGTCAGGTCCTTACAGC
46	283656	GTTGCAGCTCAAGAGAAACG	GATTAAGAGGATAAGGGTTGG	AATGACAAAAGAAAATTTTAATTATTG
47	284064	CCTAGTCAAGGCATTTATCG	AGAGAAATCTCATCGTCGCC	CAATCTCATCGTCGCCACGATTAAC
48	284719	GCGCGATATCACAAATACTC	GATAGGACCATGCAAGCCAG	GGGATGCAAGCCAGGAGATGATT
49	285188	CTGTCATCTTTACCCCAACG	GCCATGAATTACGAACACCG	AACACCGGTTGCGATTA
50	285692	TGAACGCTTTTATAGAACGG	TGCTTGAGAATAGACAAGCC	GAGACAAGCCAATCCCCTAAT
51	290120	GGGTGTAGACGGAAGATTAG	CTAGCCATCATTTTTCCACC	TAGCAAAAACCGACTGATATTTCTT
52	291438	GCGCTCAAACATGGAAC	ACTTGCTTTTGGCTTCTCGC	ATCTCGCCACAATTTTACTC
53	312102	AGCTGGTCACAAATAATCG	GATGACCAATACTAGAATGC	GGAGGACGCTTAAAGTTGTTGA
54	317708	TTGTTACAACAGCACCAAGC	C'TTTGGCTATTGTGGTCTCC	GCTCATATCAGCCATTACTTG
55	319262	TGTCGTCATTGCTGCAGAAG	TCACAGCTGGAATGGATTGC	TGCCATAGCTTGCTTTTCATATA
56	326376	TACAGGTAAACGACGAGTCC	GACCC'TTATTTGGTCTCTAC	GGAAAAGGATTTAGAATGTTTCAAAA
57	334609	CCGAAAACGTTCTGATTTG	GTTAAAACATTTAGGAGGG	TTGAAACATTTAGGAGGGTATCAGGT
58	345575	AGTGCATTCATCACAGCACG	TGCCTTAAAAACCTTGGACG	TAACCTTGGACGGTAAGC
59	364945	CATCATGAGACGTCAGGTTT	GGAAGAAGCAGGTGTTATCG	AGGTGTTATCGGTCCAG
60	381581	GTCTTCTATCTGCTATACTC	TAGTCTAACAAAAACCCG	CTATATAAATCATATAAAATGACGAG
61	393294	AGACAATCCCTGCTTCAACG	GTGAAGGTCATATACTGGCG	CTGGCGCAAAAATAAAAAG
62	398051	CTGACGATTGGGACAATCTC	GGGACAATGATAGATAGTAG	CTCTTAGGGCAATTGTAACACAAAAG
63	409052	AGAGTCTCTCTAGGAACCTG	TACTGGT'TTTGGCCTTGTCG	GGCTTAAGCTTTGGAAATGGGGC
64	427769	ACTATCTGTTCTCGTACCCG	TGGACAGGTTGCATGTTTTT	TTTTGCATGTTTTCAATCAGATAATG
65	433519	ACCAGGTTGGAGATAAAGGC	CACGATACAAGGAACGTCTG	GCACGGAACGCTGAAAGATAAGAGC
66	439866	TACCGTCTAAATCAAGCGCC	ATTGGCCACTGTAATGACGG	CAGCGCAGTACTAACGTATTTAAAAG
67	442045	GGCCGTGCCATCAATTTTAG	C'TTAGATGCGTCAACAGGTC	CCCAGGTCAAAATGCGCTT
68	458239	GGATCTGAAACAATGACCGC	CACGAAGGGAACGAAAT'TGG	AAGGGAACGAAAT'TGGTGC
69	468848	CGAAAATGCTGAGATTTCGTG	AACCTTGGCTAAGTAAGCCG	GGGACCTTTTAAAAAGAGCGGAA
70	470370	TCCCCATGGTAATGGGTTTG	TGACCAAGTTGACAACAGAG	CCCTTAGCCCCAAAAACACG
71	476851	CCATTTTCTTTGAAGAGTACGG	CGGCC'TCTGTTTGGTATAG	AGACTTCCC GCAAGGAACA
72	477854	TCAAAGCTAGACTCAACCC	CTAAGCTTGAACAAGGCCG	CAAGGAATGTTGAATTCACCTG
73	484867	GACACACTTGACGATCACAC	CCAAGTGATCGTTGAACAGC	GAACAGCTTTCACCAATTCAA
74	491836	TCGCAAAATGTCCATGTCCC	AGAATTTTGAAGCGCTAC	CGGTAAACCAAAAGGAG
75	494333	TGGGCATCATTTGAACGCTC	ATTAGAGAACGCTTGGCAGG	TGGCGTCTGGCAGGACTTTATCA
76	503347	TTGAGCTGCCTAGATAGGTG	GCCCAACTATTTTTCCCTCG	CCCTCGCTTTGAAAAGT
77	504376	GACAAGCTTTCCAATAAGCG	TCATCAGCTTAACCTGACGG	CCCTACAGGCCAAGTTAAGCAACC
78	509549	CGACGATTTTGATCACGATG	TTGCCTTTGCCATTGCTTG	TTGCTTGGCAGAAAAGTCTGG
79	512693	GTGTTGGCAAAATAATCTGG	GATGGCTTATGGGTCAATGC	TCCCATGGTGCCTTTTTTTC
80	520800	CATGACGGTAAATCAGATGG	GTGTGTCCATTACACACGAG	CCATTTATTCCTAACAAAAAGCTTCT
81	547021	TGTTCACTTGCTACCTCAAC	C'TTGATGACTTAGGAGAAGC	GCGGCTCTTATCGAGCAATATG
82	553152	TGCTCGTGAGCCTGAAATTG	CGCCTGTTATCGTTAACGTC	TTTCAACGTCCTGTTCACAGC
83	570302	CTTTGGCATCTCCATCACTG	AGCTGCTGGTTTTTTTAGCCC	TTGGTTTTTTTAGCCCAACATAAGGA
84	606726	GCCTCCATGCCGTTATAACC	GGCATCTATGACAAGGATCG	GATCGTTCAAAAAGAAATCCTTC
85	612785	ACCCCGTAGGTTAAGAAAGG	AATGTGGCTCAGGCCAAATG	GGGGTTTAGTATCCCATCAGT

No.	Locus	Amplification Primer 1	Amplification Primer 2	Single Base Extension Primer
86	614402	CCTTCAAAAATAGTGCCGTC	GAAGCCATTTTGAATGGTAG	GGATTAAAGCATGTTTAGAAGG
87	615671	GAAGGCAAGGAGAACAAG	CGGTAAAGTCAGCAC'TTCTC	CCCACAAGAGCTTTATCTCTTTGAAAA
88	618026	ACACCTGCTCTTTTCGATGTG	AAGGAGTGTGAAGTGGATGC	GTGGATGCCATTTTCAGA
89	626474	TCAAGCAGCAGATGGAACCTC	GCCCTACTGGTATATTTCTGC	ATATTTCTGCAATTGCCG
90	629785	CTATATGAGGAAGTTGGCTG	CCAAAGTTAGTAGCAACCGC	ACTATTGCGCTTTGCAC
91	650988	GCTCATGTTCCCTGCAACAAG	TCGGGATAGATTTGGTTCAC	CGATAGATTTGGTTCAC'TTTTTCCAT
92	651247	GTCACAACACCAATGAAGCC	CCATCGTCATTGC'TAATGTC	AAAAAAACTAATGATATCCCC'TATG
93	660832	AGGATCTGGTAGTATTGCTC	ATATCCCTAGTTT'GAGCCAC	TTAGAGCAAGAATACGAGGG
94	692292	GAAGGAGAAAAAATGTCAG	TGTTTTCGCAGCGTACTCTTG	GCTCTTGCGCATTTTCAATT
95	701833	CGTCACAAAAGACAAC'TCTTC	TCCTGAATTTGATTTACGGC	ATTTGATTTACGGCATAAAATTGA
96	705471	AAGATTGCTATCTGTGGCGG	GTAATATAAACGTCTGCGCC	CGTCTGCGCCTTTTGCACAGCATCCT
97	710274	CTTGGGAAGGTGATTATGAC	CCATGAGTTGGTTGTTT'TGG	CATTTGGCAGAGAAAAAAGGAC
98	714894	GATTGACATGAAATGTGAGG	GTTAAAACAAGGAGGGTAGC	AGGGTAGCTATTTAAACTCTTGTT
99	723656	TTACGCTCTTGCAGTTGGTG	TTCGTTAGTAGCGCCATCAC	ATTAGCGCCATCACCTGAGAAAAG
100	724702	AGCATCGTTGGCATTACCAG	TGCGCAACACTCACAATCTC	CTCACATTCAGGAATCAAAG
101	727093	CGTAGACGTCTGAAC'TTTCC	TTGATTTTCGTTGACGCACCC	GTTTCGTTGACGCACCCACGTCAGGA
102	790332	CAAGCCTAAAAAGCCGCATC	TGAACGATGCCAGTTTACAG	AAAAGTTTGAACCCGAAACGTCACA
103	801548	AGATCTACTTCACCTTGGGC	TGACCGCGATAAATACGCTG	GGACCAAAGACATGGGAAAAAC
104	823185	GATAAACTGGCCATAAACG	GCAGGTCATTTCTTATCGTG	GGAAATAAAGAGGAGACTTTATGAAA
105	832888	AGAGCGAGACACATTTGAGC	TGTGTACCACGACAAATCCC	TCCCTAAGATTGGTTTTTTTT
106	842041	TTTGCCAAC'TACTGATAACG	AACCCGAAGGAAGTAGGAAG	GGCTTTACAAGGGTTATCTGATAAAC
107	850454	ACTGTAGTAAGAAAGCAACC	CTGATCACCTTCATAGACAG	ACTTCATAGACAGTTT'TCC
108	853669	CTGGATAGCCTTCTGCATAC	GATTGCATCTGAAAACATGG	GGGTTTCAAAGCTGTTCATG
109	860866	TTATCCTCGCCGTTGGTTTC	CAAGGAAAGCACCATTACGG	TTACGGAAAAGATCAATCTTAC
110	872314	CCCGTCCAAAATGTTTCTTG	CGTTCAGGTTCC'TTGATTAC	GTTGGGATAGAAGAAGGC
111	874111	TCGGCAGTAAAAGCAACATC	C'TTTGAGCCAAC'TGAGCTTG	GCTGAGCTTGGGGAAATT
112	881273	TTCCAACAGGAATTTGTTCGC	GAATCCTATTATCGAAGCCG	TTGACCTATCAAGAAATCTCC
113	901307	ACTTCCTCAGGATATTGGTC	TATTACTCGCACCCAAGCTC	TGCCCAAGCTCATCAAAGAA
114	903712	CCGAGAATCGTATTAATGAG	GCTCACCTACTTCTTCCATC	CTCAGCTTTATCATCTGCCATTGAC
115	913036	AGGGGAATATGGTTTGTGGC	TTTGACGGATGACCGTATCG	AGTTTGCTGATGGCAGTG
116	933863	CACGAGATAGCACACCAATG	CCATCATGTTGGCAAACCTC	TCAATCTCATTTAACCACAGATCGT
117	940431	GATTACCTATACCTCTTGCC	CGTTTCTTCTTGAAGACGCC	AAGGCCATCTAAAGCAGGAAG
118	941901	AGTTTGCAGTGATGATTGCC	CAGCTAGTAATTGATCTGCC	ATCACATCATAAATAAGCCCAAAT
119	949710	GGGTCGAGCGATTATAAGTC	GGAAAAGCAGTTAGAGCAGG	AGTTAGAGCAGGTGCAGG
120	967423	CGTTAACTTGCTACTTTCTC	TGTTTGACAAGACATTGTGG	CATTGTGGATATTATATACTGTAAAC
121	975504	GTTGATAAGAGGGAACCAGC	CCAATAAGATGTTGCCCTCTG	CTCATCTGACATTAGGTTTAACTTT
122	1019065	TAGCCAAGCGTCATCTAGTG	TAGCTCCGCTTGTATCTGTC	CAGCCACTAACC'CAAATCGCTTG
123	1022760	AGCTTCTAGCTGGTTAACTG	ACAGTTTCGTCGTCCCAATC	TCGTCCCAATCTTGTAAGG
124	1023643	ACCAACAACAACCAAGGCAG	TCGTAAGGATCCTGTTAGCG	AAATATTGATGGAAACTATATCTTC
125	1026493	TCTGCCTAGTCTTTGCTTAC	GAAGCTTGGTATCTTCTGAC	TCGATAAAGGTTACGAGCAC
126	1030601	ACTTG'CATCCC'GTTACATC	GGGTTGAGACGCAGCGTAA	ACGCAGCGTAATCGTAA
127	1039955	AGAGGCTTAACTAATGACAG	AAACGACATCACCTTCGTAG	TTTTCATCACCTTCGTAGAATTTG
128	1041721	GGCTTTATTAACCGCAGGTG	GTTTTGTGTATGCGGCTTTG	TGCGGCTTTGACCAAAC



No.	Locus	Amplification Primer 1	Amplification Primer 2	Single Base Extension Primer
129	1044702	GGTCCTGCAACTTTTCATCG	TTGCACTGCATCTAAAGGTC	ATTATGTTGGTAAGCATTTGGCAA
130	1054883	GAGGACTTTGAGCACTGATG	GTCAC'TCTTTAGTTCCACCC	GCCCCAACTCTCTTTCC
131	1060414	TGGTTCAC'TTTGAGCCCTTG	CAGCTGTTATCATCAATGCG	CCCCACCAC'TCATTTGATTTGAACCC
132	1063938	TTCCAGATGACCTGCGAATG	CTACCATCCATGCGTTGTTT	GGTTTGGTTGGGAACAA
133	1076895	AATACCCTCACCATGACCAC	GCATGTGGATCTGGAATGTC	TTTCCGATTTCAATATCAGATAA
134	1077899	AGCTCCAAC'TCGTCAACAAC	TCGTTTTTCA'TTCAACCCCTCC	GCAGTCAACCCCTCCTTTTTATTTAACT
135	1078726	CTTTTTTGTGACCGACTTCC	CGCGTGATAGTTCGGTTTTTTC	GAGCACCTGCC'TTTTGTAGCAAT
136	1079364	CCCTTCTGGAATATCAAGCG	TGCTCAGCAAATGAAATCGC	AGATGAAATCGCAAAAAGACAGTGT
137	1084815	TCCATAACTGTGCGCAAAGC	AAACCAATTGCAACCGGGAG	GGAGTTGTTTTTAATGCATGATATTC
138	1087785	GAACATTCTGTAAGCGCACG	AAC'TTCAGTACTTCGTACGC	GGGTGACTCTGCTAAGATGG
139	1091393	CTCGTGATCTTGAAACAGAG	TTTCCATCGATAGGATTAGC	GATGATATAAGGGTAACCAGA
140	1092744	CGCCTAAACCACTCACTAAC	ACCAGTAGAACCACCAAAAG	CCCTAACCGACCAAAAGTTCTAAC
141	1094245	AAGTCGCTGTTGCCTTGACG	GCCAGAAGCGCAAGCTAATA	CAAGTTACATCTTTGGCTC
142	1101887	ATGAGGGTGGCATATCAGAC	GCTTTTTTGTGACCTTAAAC	GAAC'TTCTTTTGATAAGAAGTTG
143	1108998	TGTTCTTCTTGCC'TTTGGAC	TGGGAGGTCTTCTGTGATTG	AACCATTCCCTTGATCG
144	1113430	GGGAGCATGCTTACTACTTG	AGCGGCTTGGTAAAGTTCTG	TCCAATTAATGATTTCAAAGAAAG
145	1114921	CTTTGATATGTCTGAGGCCG	CTTCTGCTCAGCAAAGAAGG	AAGGGTAGGCTCAC'TAGAT
146	1115629	GGCAGTTTAAGCCATCCAAG	TGGTCTCCAGCTGCTAAAAC	CCCCCAGCTGCTAAAACCCTAACGTGA
147	1119152	TCAAACATTACCTCAAGCC	GCCAACATCAC'TTGAAAAGAG	GAGTTGATCAATACGTTCCG
148	1121783	TATTGAGGGCTTGAAAGCCG	CGCTGAATACCAAACAAGAC	GTGATTTTGGTAGATGTGACT
149	1126295	GGTTTGATTTACCTCTTGCG	GCGTACTACTGATAGCAGTC	ATAGCAGCTTTAACTGACTGG
150	1135396	AGTTAACTTGCCGCTTGGTG	AAAGAAAGTAAACGCAAGCC	CCAAAAGCCATTTAAAATCACTAAAAT
151	1174118	TACACCAATGTAGAGCTCTG	AGCGGGTGCTAGTTTGAATG	TTGGCTGCTGTTCAAGCTA
152	1179891	GAAATAGCTTCTTCCAAAGGG	GCGTGATGGACAATTAGATG	CCCCCTATTGTGAACAAACTTGAAAAT
153	1181412	ATCTGCTCTCCGTAGTACTC	AAAAGGAAGCCATTGCTGAC	GCTGACTTCTTATATTCCTTATG
154	1186289	GACTAGCAGGATAAGCCTTG	AGGGAATGGCTAAAGCAGAG	TGACAAAACAGCCTTGAG
155	1199100	CAACAAGCGTCACAAACAGC	CCAAATATGCCACGAACACG	CCACGAACACGTTCTGA
156	1203073	TTGACGATTTGGTGCCAGAC	TCTGCTTTGTTAGCAGCTCG	GGTTAGCAGCTCGCTTCATAC
157	1205969	GGTTTCTTTCGGTTTGGAACG	TCTTCCATACC'TGAAACAGC	GGCAGCAATCAAGTCACCT
158	1206168	CCAGCGCCAGTTGATAATTC	CCGATACGACCTACGAAATC	GGGGACACTTGGAA'TGCAAAG
159	1215091	TCAGCCAGAGCTTGAGTAAG	GGGCAAGTGGCCTATTTGAT	GTTGGGATTTTTTGATAATCCTA
160	1216179	TGAGCCAAGTTGGTCATTCC	CCTGACTATGATGACTGGAC	GGACAACAGAAACTGAAAA
161	1224430	AATCTCTGGCGATTTTGCTG	GTCGTCCCAC'TTATTTGAGC	CCTGTATCACTGTCTTTACCCA
162	1229090	TCGGTGCTGGCGAAAAATTC	AGCTGAAGAACC'TGCTTGTGTC	TCTTGATGAATTTGACTAGCAAAG
163	1273604	ACTAACTGCTTGCTCAACCC	TAAC'TGCCATGCTCATCTTG	CCCAC'TGCCATGCTCATCTTGCTCCTGA
164	1274506	TAGTGGCGCAAACCATAGAC	GAGGTGGGTTTTTACATGC	AACCATTTATCTGATAGCGTATC
165	1282510	AGCCAGTTGTAGCTGGTATC	CCACTCCTTAAAAGCTTGCG	ATCAAAGCTTGCCTGAATG
166	1292747	GCTCTAATTGGTGACACTTC	AGAAACAGTTGATGACTGGG	TTATGAAAGCTACAATCAGG
167	1300624	TACCAAGAACACAAGGAGGC	AGATGTCAGAGGCGATTAGG	AAATAGGGGATAAGAGGGATTTAGC
168	1309701	GGGACTATTGTCTCAGACTC	GAAATAGCTAAGCCTGACTG	GACTGATGTTAAAGGTATGAA
169	1362922	CTTTGGACGAAATCAATCAG	TTAGTTGACCGCTCCAAACC	TAAAATGCATAGGTGGGATAG
170	1365099	ATTGGCTATAGGCAACCAGG	TTCTGCTAAGGTCCGCATAG	ATAAGGTCCGCATAGATAAACTT
171	1368463	AGGGATTGGTTATCCGTTTC	TTCCATATACGGAAGCGTTG	TTTGCGCAACGAATCTAT

No.	Locus	Amplification Primer 1	Amplification Primer 2	Single Base Extension Primer
172	1380032	CACAGTAAAGGGTTAAAGGG	GGCATTACTTCTGAGAAAAC	AAAACAAC TGATAAAAATAAGAGGA
173	1398545	ATAGATGGTCTCTGTTAAGGC	GCGATTGCAGGAACAATTGG	GAGTGGAACAGGTCTCTTTTTTAGGG
174	1408991	CTGCATTGCAAACACTTGGG	GGATGCTCAAGCTGACTTAG	TGACTTAGATATGGCTACTAAGATTG
175	1459415	AGGACATGACTTGGAAAAGC	TGGTCTTATGCTTGGATACC	CCCTGTTGTAGGACCAATTATTGGAG
176	1459955	CTTGCCCAAACGAAAACCC	CTGATAAAGACTGCCACAGC	CCAACTGCAATTCCCCAG
177	1465235	CCCAATGGATAAGCTGGTAG	CACGAGCACGGATAAAGTTC	CCCGTTCATCACATCCGCCT
178	1466966	AGGCTGGCTAACTAGGAATG	TTAGGTAGGGGTACGCAAGG	GTGTACGCAAGGTGAAATTAT
179	1469148	ACCTTTGTCAAACCGTTTTTC	CGCTATCTCTCAGGAATGTC	CAGGAATGTCAGCGATT
180	1471633	CCAGTGTTAAACAATGCGTC	GGCATTCTAATTGAAGGCATC	TGGTTTTTCAATATCTGCT
181	1475088	TATCAGTAGCAAAAACCGAC	TGAGGCTCCTGATAACTAAC	CCACCTTTAGTTGACAAACTAAGACT
182	1478114	ATGGGTTGCTGGCTACTAAG	CGACTAGTGGTTGTAGCATC	GATCATAACCTGTTTTTTTCGTAA
183	1481314	GGTAATTGTTGCAAGCGGTG	AGTCTTTGCAGGCCTTTGTTC	GCGCCTTTGTCAAAAATACAG
184	1484947	TACCAACCAAACACCCCTG	TGCTGATAGAGTCTCAGGTG	GTCTCAGGTGCTTTTGTTCGGGT
185	1487337	GGTGGTAGATAAGTGATCGG	GGGATTA AACAGAGTGGGAG	GGTGGGAGAATTGTGGGCT
186	1497764	GCTGTTATCCATCAGTAGAC	GACTGTGAATGAGGTCAAGG	ATGTTATCTAAATCAAGTTGGG
187	1500872	TAAGAAGCATACTGCGGTGG	AGCAGCAGGAATTGCTGAAG	TGCTGAAGGTGGACGTA
188	1501014	CGTCGAACAGGTATTTTCTC	AGCATCTGCAAAAAGCGCAC	TTTTGAACTAAAGCCTATACATTTTT
189	1509415	ACGGTAGGTGTTTTGTGCAG	TTTGTACTGGAGACATGGGC	GCTTCTCAGCTGCTAAAAAC
190	1509527	GCAAGTCCTGACCCATTAAG	GACGTGCTCAAATCCGTTAC	TCCGTTACCGTGATGAA
191	1515192	CCGCCATCAATAGCGATAAC	AAGCTCAAGAAGTTGCTCAC	TCACTTAGCAAGCTTCTTAG
192	1518653	AGTCGCTATGGCTTTTTTGC	TTACACACAGATGGTGCAG	CCGAAACTAGTCTTTCATCTCC
193	1519059	GACCTAGCTATTCAGGTTGC	GGCATCAGGGCTGATAGTTG	TCCCAGCTTTTAACCCAGACTGA
194	1525730	GGCGCTACTAGAATCAATCG	AAGGGAGTGGTTGGTTTGTG	CCCCACACAAACTCATGAACCAGTAGG
195	1528442	TCTCATAGGGTCAGCTGTTG	GGACGGGAAAATAGCCATAC	TGAATTCCCCAATCCCC
196	1536050	AGGGCCTGTTTTGGATGAAG	AAAAGCAGCAGGATTGCCAG	CCGCAGAATTTTTTCACCACT
197	1536233	GAGTCATTTTGAAGCGATCC	AAGCTTAGGACAGCTAATCG	ATTAGGACAGCTAATCGTTTTAA
198	1544580	GCAGCAAAACTAAGCTGTTT	CAACATTATCGCCAGATGC	CATCTAACAGTTAGAGCGGGAAG
199	1563662	ACCTGAAGCTTTCGCAGTTG	CCTCCCATAATTTGAACACG	AGAGGGAACAAGCCCAACA
200	1565479	TCCCTTGCCATATCATCAG	GAGCTATTCAACATGGCCAC	AAGAAGCGGGTCATACCTAT
201	1566090	CACTTCTGGTAGTCTAACG	CTGAGATGAATAGTCCCAGA	GATGAATAGTCCCAGATTTTTTAGT
202	1569141	TTGTGCTGCAACGTTGATGG	GTCTTCTCCTCCGCTTTTAC	GAGTGAATCATGGTGGG
203	1570545	TGGA AAAC TGGTCCAATCAC	TGGACTTATTGGTACACCG	CCCCACC GTTTTTAACGTTTGC
204	1573272	CGTTCTTTGACGGGATTATG	TTGGGCTTATCTTCCCCAAC	TGAACATCTGGA AATTGAAT
205	1579234	AGCATTGGCTCCTTGG AATC	ATATCCACCCCAAAC TCGC	ACATGGTTGGGTAGTAATTGGAG
206	1580901	TCAGCTTGAAGCAGCTACTC	GCGTTCGTTGAATTTCAAGG	GGGAGAAGCCTTTAAGTAGACAAT
207	1583782	CTTCTTTGAGCGGACTACAG	ACCATGGCACATACGATACC	CTTGCGTCATGATCAGC
208	1588379	TCTAATGACAGCACGTTAGC	TGTAAGTCATTAGCCACAGG	ACAGGATTACTTAAAAGAGGATGAC
209	1592955	CTAGCAACATCTCTCGTCTC	TTGAACGGTTTGGATGGTTG	GGGTTTGGATGGTTGTTTGAAT
210	1597641	GGTCAAGTGGTGACACACAA	CCACAAGCAC TCTTCTATCC	AAAGGTGTGACAACAATGG
211	1605096	TTCTCTGACTCGATGCGAAC	GAGATATGTTACAGGACTTG	CCTTCTGTTCTCAACACTGGTTA
212	1613345	GACGTAAGAGGCTGGAATAC	ACGTGGCCAGAAATCCAAAC	ACCAGAAATCCAAACAATTTT
213	1618236	TAGTAGTCTGGTCTTTTGGC	TCTAATCATAGAGCACAGGG	GGGAGCACAGGGTTTATATCA
214	1620754	TCGTATCGCTAATGGCACTC	CAGGGATTGGACTTGATGAG	TTGATGAGAAAATCACAGTATCA

No.	Locus	Amplification Primer 1	Amplification Primer 2	Single Base Extension Primer
215	1627571	AAGGCGGACATGAATACCAC	TTCGTCAACAATCTGTTGCC	AGTGGAACGTTTTGGTAG
216	1628980	GGAGTCGCTTTAAAGTCTGC	TGATTCAGGCGGTCAAGGTC	TCCTATCTGC'TTTGAATGG
217	1630376	GCTAGCACCACATACTTATC	GTGAAAAATGAGTACGTAGAC	AACTGGGAAATACCTTGTCAAA
218	1631973	CTGGTATTGTTCTGGGTTTG	TATAAGGCCACGAAAGGTCC	ACGT'TAAAGGGGCAAACACGCCA
219	1634682	GAATAATGCCAGTTTGCCTC	TCCCTATTAGTTGGTTCCAC	GTTCCACATTGGTTAGGTAT
220	1647120	GTGCGTCTCCTTCTTCATAG	CCTTATGCTATTGATAGCTC	TTGTGTAGCATTCA'TTTGGCGG
221	1651729	GACACGCGGAAAAACATATC	GGGCCAATCTCCATCAAGTC	CTTCTGATTATAT'TCTAGTACAAGGA
222	1652180	TTCTCTTCCCACTCACCATC	CTCAGAAAGATGTTCCAGC	AGAGTCCCTGAAATCCGTT
223	1655280	TGCCCTCTCTCTATAGATTG	TGGAGATAGCAAAAGAACTC	AAGAACTCATAACAGTCTACTC
224	1656383	CCTAACCATCTTTGGGTAAC	GCCTTTAATAA'TTTGGGATGG	TTGGGATGGTATCTGTTT
225	1661654	GAAACAAATACCATGACAGG	GCTTCTATCTACGGAGTTAC	AATTATCGTTTTTCCAAATATCCCG
226	1667220	CCGGCACAAAGAACGAGAAC	GCCCTCAAAGAAAAAGTAGC	TAAGAAACCAAAGAAATCACTGAT
227	1667466	AGCACTTGGAAGTTTGTCCC	ACTGCAAAC'TTGCCTTGGAC	TCCAACACCAGCTCCATTA
228	1668714	TGACCATGCATTAAGTGAC	GCTCACAATGCTCAGACAAG	AGAAGCAGCTGGAGATGA
229	1668828	AATCGTAGCCTATGCTGGTG	TCTGCTTTAGCATAATCCCC	AGCATAATCCCC'TTCTTG
230	1669155	GTCCACATAAGACACCGTTG	TTGAGTGCTGGTGT'TTCTGC	GTGTTTCTGCCAAAT'TGT
231	1669604	GCTCCAGCTTCTTTAATACG	GGATACGATGCAACAACAAC	AGTCGCTTACCTGATTG
232	1669949	ACTGGGCATGTTAACATGGC	GCTTTTACGTTTGT'TTTCAG	TCATTTTTTATACCTCTTTGACTTT
233	1678204	TTGCCACTATCTCTGGCTTC	TGTTGATACCTTGGCTGCTG	TGAGGAGAAGGCTATCTAAC
234	1691812	GCCACCTTTTGGAGAACTTG	GAGTCCAATTGTCTCACGTC	CGTCTCACGTCATTTTTGACTG
235	1694883	CTGCGTGGCTGTTAAATCTG	GGTGGTGT'TTCGTAAAGCTC	AACGCTAGATGCTTTTTGG
236	1697897	CACAGGCTGTGTGATTTTGG	ACAACGCCAGTTT'TGACGTG	AATGGGCTTTATGAACGCCAATTATG
237	1699412	GGACCAGATAGAAATGCCAC	ACCGAATGCACCTTTACCTC	GATGAAAAGGCAC'TTGAA
238	1700339	TGGTTTCCATGTGACAGGAG	TCCCC'TTCTTTGACTTTCAG	TCCCGAATATCTACATATTCAGCCTG
239	1703042	GCAACAACGCGACAAAAC'TC	CCAAATTCATGGACAATCAC	TAAAGGTTAT'TATCC'TAACATAGAT
240	1709912	TCTAAGGACACACATGCTGG	TCGGCCACAT'TGATGTCTTG	GGCAACCAAGTCTCTTGGG
241	1716403	TTCACGGCGCTCTTTTTCAG	AAGAATCGCTCTATGCAGCC	CCGTTGGGTTTGGTGACAT
242	1739068	GATCCACGAGATGTTTGGC	TGAACTGCCATCTGCTCTTG	TTGGGATGCTAAAGGAG
243	1742009	GGCTATGGTCGTGAGTTTAG	AACTGACATGCC'TTGCCTAG	CTGACTTATCACACCCTG
244	1745943	CAAATAACCCCTGTTTGCTC	ACGATTCTCTCCTAGCTGTC	TGTGCGTAGAGAAATAAGGATAAC
245	1746410	TGGTGCTTGACCTTTACCTG	GCTGGAAAAGCATCAGACTC	AATCAGACTCACAAATCCCTGATA
246	1747737	TTGTCTCAGGTCAAGTCTGC	TGAGAACTTGT'TAGATCAGG	TTTGGTTAGATCAGGTTACACAATTA
247	1747738	TTGTCTCAGGTCAAGTCTGC	TGAGAACTTGT'TAGATCAGG	GTTAGATCAGGTTACACAATT
248	1749870	TAAATGCGTGATCGTACTG	TTCGTTAAGTCAGTACGACC	TCCATAAAAATGC'TTTCAAAT
249	1750069	ATGGTTGTACCATAACAGTC	GCCATTTATGCTCTTTTTTG	TAACCTAATTTGTATTTTGTACAC
250	1755630	AGACAGTTGATTGGTGGCAG	TTAGCATTTGACTTGTCTGG	TCATTTGACTTGTCTGGAATTTAG
251	1760061	GGCAGACTTTGCAGCTATTG	CGCCTGAGTCAAACGTATAG	GGTTTTTATCGGCTGCAATG
252	1761030	CAAACACTTGCTTAGGTCGG	CCTCACCAC'TCGTTAAGAC	ATACGACCGTGT'TTCTAAAATA
253	1761173	GCTTCTATCATACTCCTGAC	AGAGCGCACACCATTTTAGG	TCCCACCATTTTAGGCTATTCTT
254	1763553	TGCCTAGTGAGATAATCACC	TAGGAAGGCTTGACATATGG	GAGATGGAAAT'TGGTGAAACC
255	1764204	AAAGCTTCTGAGATTTTGGC	ATGAGGCCAATAATGCCATC	CACACAACACCGTTTCAT
256	1771377	ATAACTCACCTGCTGCCATC	TGGGCAATCTGTGGTCAATG	TCTGTGGTCAATGGAAT'TGTCAC'T
257	1780162	ACACGGCAATGATCCAACAG	GATGAACCACATAGCAGGTC	CTAGGTCTGTCCCCTGTAA

No.	Locus	Amplification Primer 1	Amplification Primer 2	Single Base Extension Primer
258	1790135	GATATTTCCCTCACGCTCTG	CAGTGAGCAAGTTTTTCACGG	CTGACTCGATACAATTGCTTTTTGA
259	1796147	CATAGTTGGTGACACCCATC	ACAAATGGCCTTTTGGTGGG	TTTGGTGGGTGAATACGACAACAT
260	1798264	GCATGCCATTGGTTGCTTTC	TTTGTAGCCACCACCTGAAC	AAATTTTATGAGCAATGTCGATA
261	1801826	GCAAAGTTTAGCTTGAGCGG	TAAATGGCCCTAGGTAAGAG	ACAAGAGCCGATAACGGCATTGAT
262	1806778	ACCGTGATTTCCAGCAGTTG	TCTTGATACCTCCCTGTTTCG	ACCCTGTTCGAATATCATAACG
263	1812955	GGAGTAAGTCCAATTCTCCC	CCTGCCAATTCATATTTTCGC	CTGAAAAAGATTTTAAGGACTGAT
264	1815075	GGAAGTGAAGATCAGTTCCCTG	TTCCAGTCGTGTCCCTTATAG	CTTTTTTATCAATTGTCATGAGTTTT
265	1815253	CAGGTTAAGGCTGGACAAAC	ATTAGTCACCACAAGAGGCG	GGAGGCGTTTCAGTGGCGT
266	1817279	ATAATGGCTTTTTTGGCAC	GATAAAATCTGACTGACACC	CCTCGATTGCAAATTTGTCTG
267	1818197	GCGCTCATCTCACGAATATG	GTGTGCACTTACTTGACCTG	CTTGACCTGAGCTTCTG
268	1825921	ACAAGACTATCGCCCCCTATC	CCATACTGATAGCCAACAGG	TGATAGCCAACAGGAATAGCAGT
269	1834388	GATTGCATCTGGTGTCTATG	TAATAGCTAACTCTGCTCGG	AGAGCTAACTCTGCTCGGTAAGTTTT
270	1841704	AGCTCGAAGATTGCTTGACG	TGCCATGCCATCAGAAAAAG	TTAGATGCTCTTAGTAAAGTTTATAG
271	1842608	TCCAAGCCAAGTTCGATAGG	AGCGTTCCTTGGCTCTAACC	TCTGTGGCTCTAACCATCTATG
272	1860162	GGACAAGTTCAGTACCGAAG	GACACCAAAAAAGCATGGG	CTAAGTAGATACTTTGGTATTCCT
273	1868362	AACAATCAAGCACTAGCTCG	GAGATGGCTAATAGGACTGG	GGAATAGGACTGGAATTTAGAATC
274	1871539	GGGAACAGCCTTTCTGATAC	TAAGCATCCTGCACTTCCCTG	GCATCCTGCACTTCCCTGCAGTGGGTG
275	1880453	CTCTATGTCTAGTTGACATT	GGCACAATTACACCTCTTTC	GGGACTTGAAAGTAAGGGCCT
276	1881838	AGTGACTTCGTTTTGAAGTG	ATAGCAGTTGCTGGATCAGG	GGAGGATATGTTGGATTATCACTA
277	1881840	AGTGACTTCGTTTTGAAGTG	ATAGCAGTTGCTGGATCAGG	GATATGTTGGATTATCACTAGG
278	1881986	TGGCTTATCACCAATTCAAG	GCAATAATGACCAGTTCCGC	CCTTATAAGCTGCTTTGCTATCT
279	1891225	GCGTAGTGATACATCAGAAAC	GCCTATGTAGTGATTAGCCC	GCTTAGATAAAATAAAATGAGGTGA
280	1896371	CAGAGGTTAAGCTGTTTGGG	CATCAATGAGCGCCTAAGAC	CCATCGCCTAAGACGACTTCTTC

1. All amplification primers were modified at the 5 prime end with the following 10 mer sequence: ACG TTG GAT G.