Supplementary Materials

Figure S1.

Minimum Spanning tree of the 368 STs in the *S. suis* MLST database. Any STs containing Chinese STs were marked red. STs selected for this study were marked with a blue star. Clonal complexes are shaded with different colours. The distance between STs are denoted by lines of different styles: thick solid black lines = 1 locus difference; thin solid black lines= 2 locus difference; dashed black line = 3 locus difference; dark grey lines = 4 locus difference; and light grey lines > 4 locus difference. (The figure is best viewed when zoomed in).

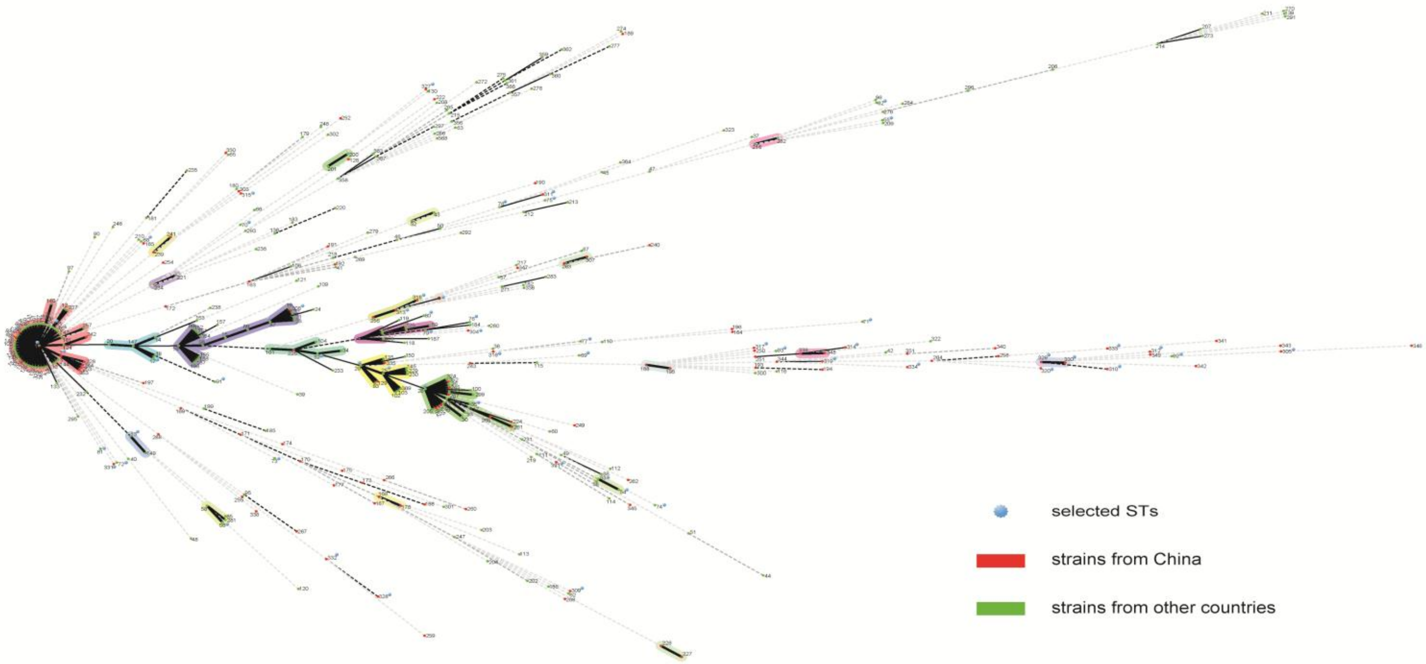
selected STs
strains from China
strains from other countries

Figure S2. The percentage of core genome genes (MCG genes) and accessory genome genes (AG genes) in *S. suis* GZ1 according to function categories.

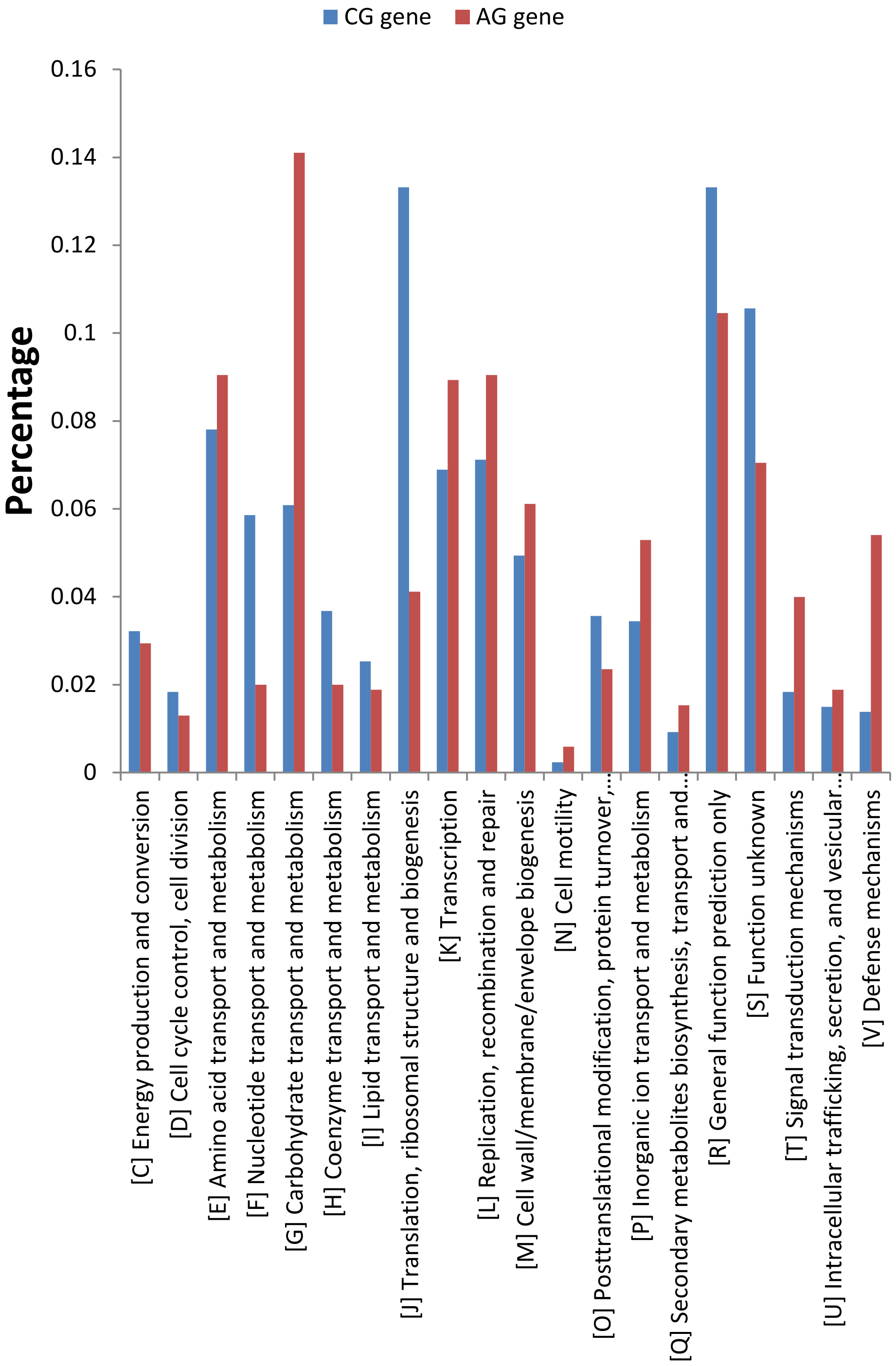

Figure S3. (A) The distribution of whole genome SNPs, core genome SNPs and accessory genome SNPs. X-axis denotes the SNP density in one kb for each gene and y-axis represents the density of genes in each window. (B) Relationship of all SNPs and core SNPs. X and y coordinate represent the density of whole genome SNPs and core genome SNPs, respectively. (C) Relationship of whole genome SNPs and absent genes. X and y coordinate represent the number of SNPs and absent genes of each strain comparing to reference genome *S. suis* GZ1, respectively.

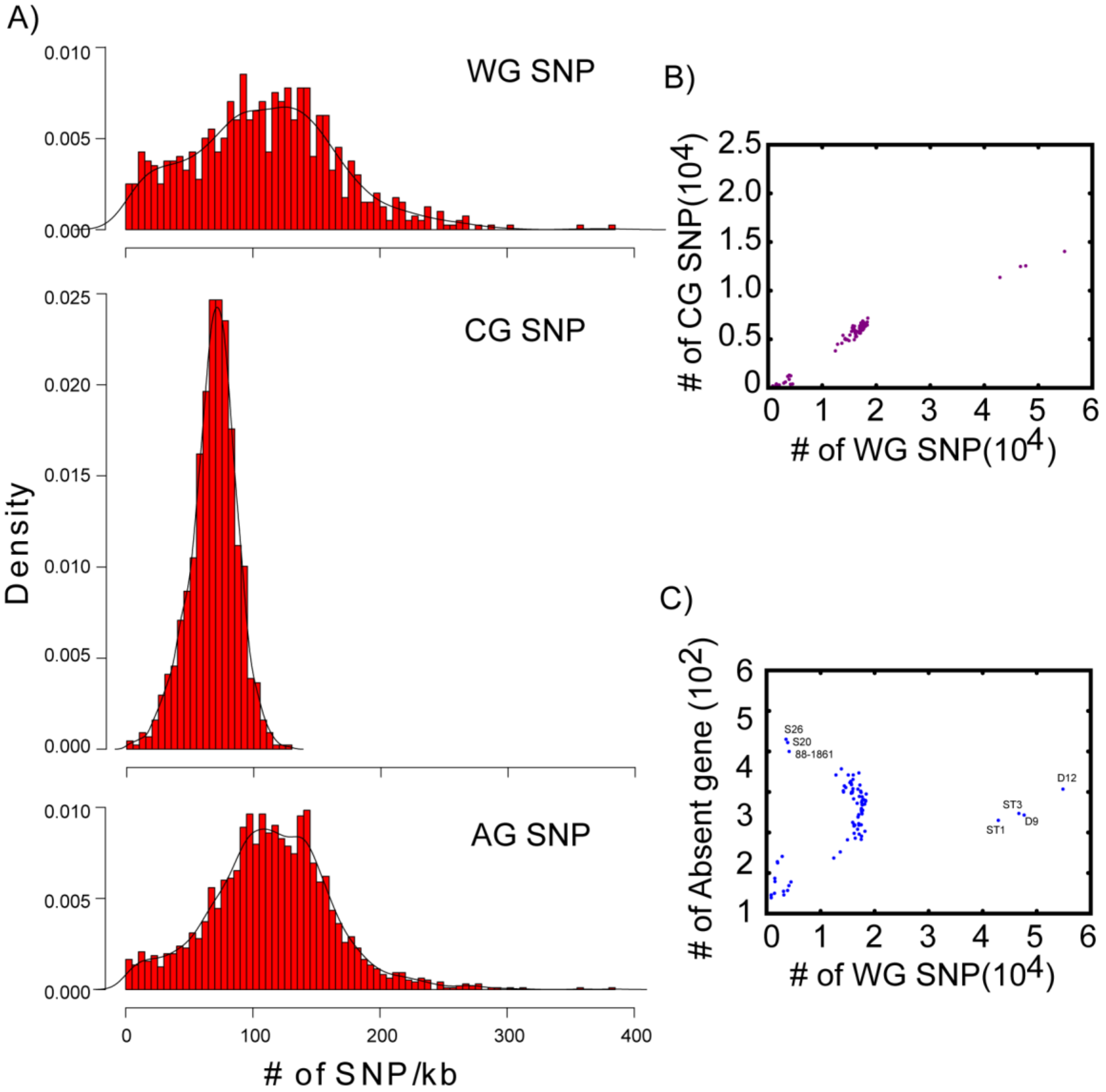

Figure S4. The function categories of three groups of core genes based on SNP density.

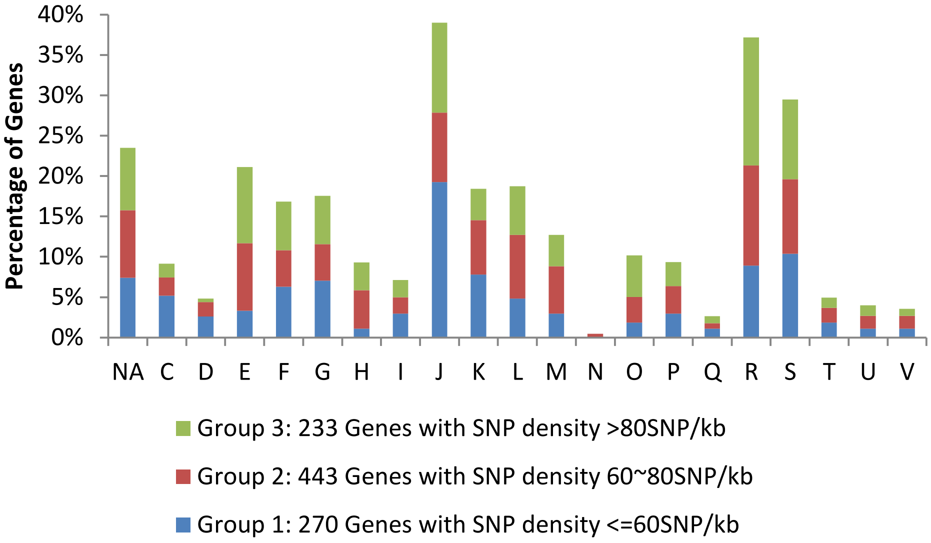

Figure S5. The NJ tree of present *S. suis* strains in 7 MCG groups with outgroup strain *Streptococcus pneumonia* R6.

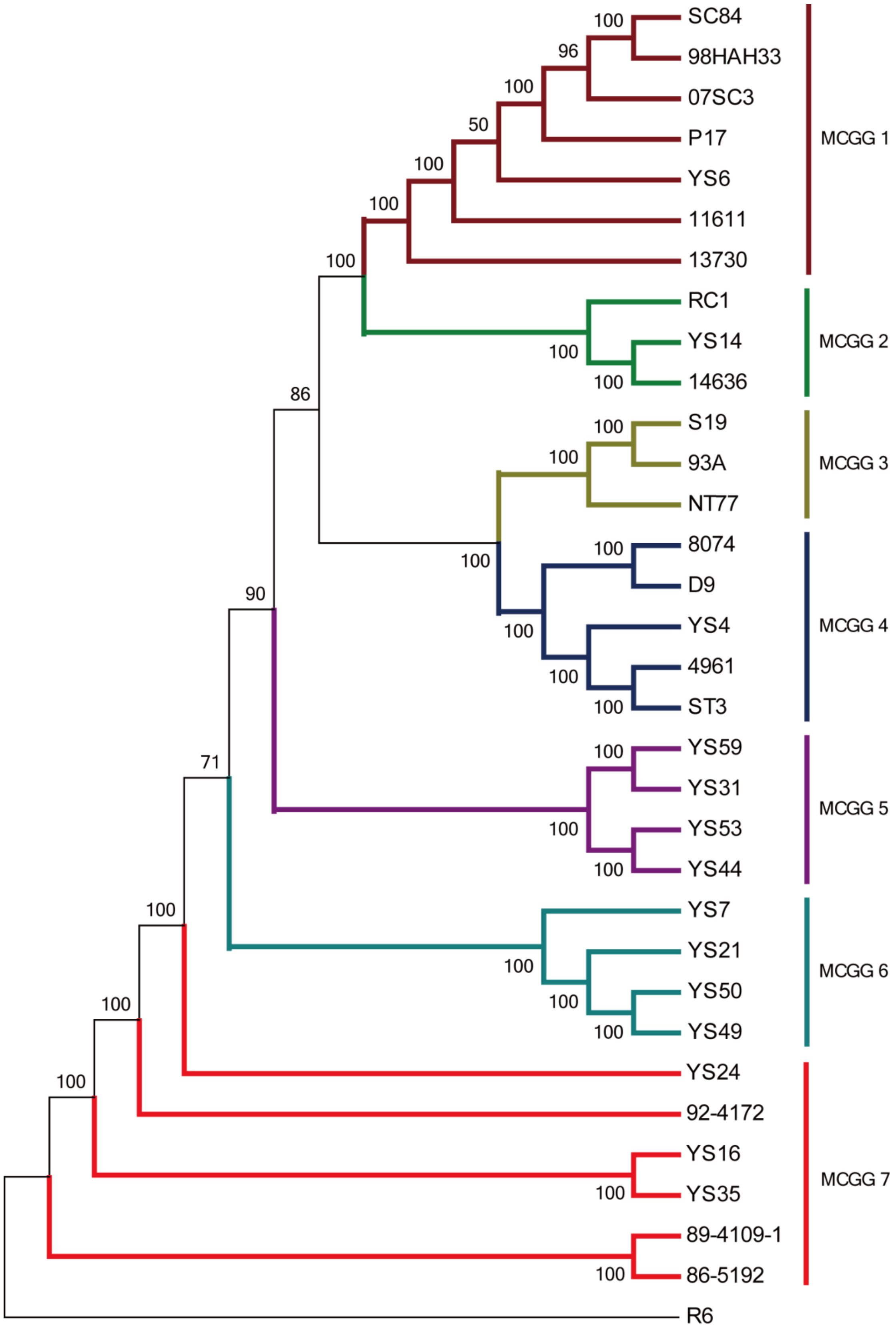

Figure S6. Estimated population structure (K=7). Each individual is represented by a thin vertical line, which is partitioned into K colored segments that represent the individual's estimated membership fractions.

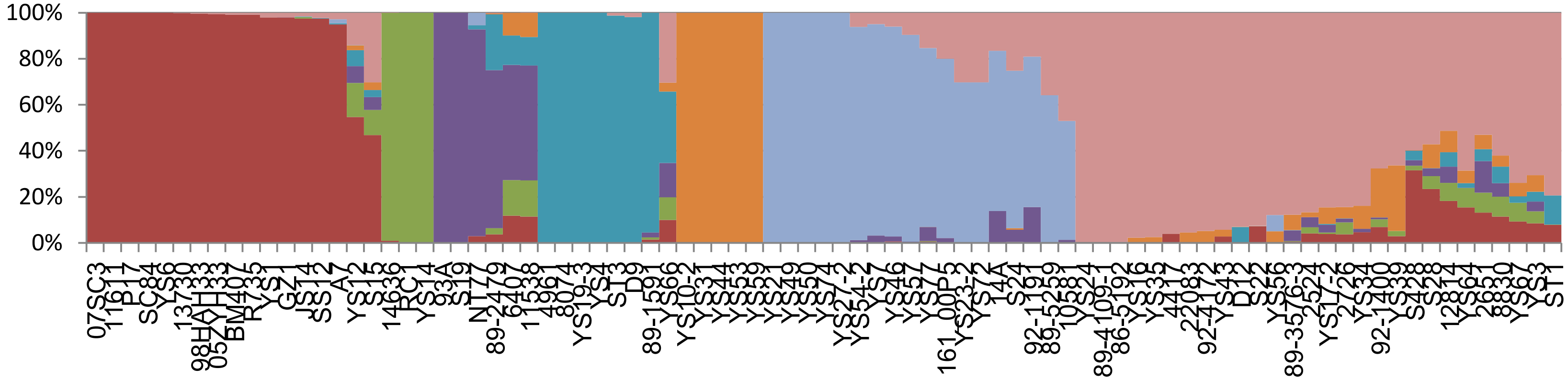

Table S1. Sequencing information of 72 strains.

Strain	#Scafflod	#ScfLen	CDS#	GC(%)
11611	102	2149119	2250	41.08
07SC3	24	2039810	1958	41.21
161_00P5	26	2062220	1972	41.05
89-1591	83	2270244	2037	41.39
RC1	35	2198129	2218	41.05
S428	76	1987921	2036	41.42
2651	37	2224018	2231	41.4
4417	70	2069567	1996	41.13
12814	33	2198386	2304	41.07
8830	28	1984673	1992	40.81
10581	51	2483953	2512	41.06
13730	34	1962186	1954	41.33
S15	29	2015465	1963	41.3
2726	45	2224039	2282	41.06
93A	154	2328965	2240	41.06
NT77	125	2318985	2445	41.03
S19	150	2293667	2318	41.18
R735	35	1950587	1904	41.36
86-5192	39	2111528	2121	39.94
14A	94	2350341	2360	41.37
S22	72	2275838	2379	39.68
89-2479	38	2169000	2259	40.95

S24	78	2339565	2268	41.29
89-3576-3	63	2126721	2059	41.31
89-4109-1	59	2178521	2203	39.88
89-5259	111	2462624	2393	41.28
S28	46	2163749	2178	41.2
92-1191	135	2356931	2312	41.27
4961	53	2018439	1996	41.22
92-1400	62	2308516	2266	40.79
92-4172	87	2305152	1685	41.26
6407	42	2250266	2293	40.91
11538	27	2166871	1567	40.93
2524	51	1883386	1792	41.5
8074	88	2060773	2112	41.19
14636	26	2081361	2111	41.16
22083	26	2026680	1955	41.49
YS1	20	2031282	1972	41.26
YS10-2	44	2221121	2274	40.95
YS12	42	2080241	1970	41.12
YS14	30	2172902	1542	40.97
YS16	33	2332845	2344	40.98
YS17-2	46	2339153	2314	41.03
YS19-3	50	2110290	2125	41.12
YS21	91	2474394	2445	41.31
YS23-2	52	2348283	1582	41.4

YS24	48	2107854	1999	41.2
YS27-2	90	2430841	2350	41.25
YS3	43	2172487	2118	41.02
YS31	29	2230390	2204	41.04
YS34	45	2376987	2301	41.02
YS35	50	2321968	2223	40.92
YS39	53	2243368	1546	40.96
YS4	54	2110742	1995	41.09
YS43	48	2054213	1624	41.21
YS44	37	2208333	2211	41.11
YS46	114	2430062	2460	41.21
YS49	92	2412706	2246	41.3
YS50	89	2419800	2400	41.29
YS53	35	2208787	2264	41.11
YS54-2	87	2318900	2246	41.24
YS56	61	2234172	2414	41.17
YS57	114	2521029	1736	41.15
YS59	44	2202033	2141	41
YS6	26	2039819	2057	41.13
YS64	41	2067782	2102	41.11
YS66	29	2060760	2081	41.18
YS67	27	2182719	2139	41.1
YS7	77	2354331	2311	41.26
YS72	48	2361733	2276	41.43

YS74	89	2471261	2543	41.19
YS77	85	2264602	2421	41.44

Table S2. The 50 lowest SNP density MCG genes.

Gene	Gene_Length	COG	SNP#(1kb)	Gene_Annotation
SSGZ1_1335	1308	G	45.87	putative Enolase
SSGZ1_1289	1374	R	45.12	tRNA modification GTPase TrmE
SSGZ1_0081	399	J	45.11	ribosomal protein S8
SSGZ1_0002	1137	L	43.98	DNA polymerase III, beta subunit
SSGZ1_0987	1254	S	43.86	hypothetical protein
SSGZ1_1471	693	G	43.29	Phosphoglycerate mutase 1
SSGZ1_1829	813	U	43.05	in regions AI 121: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_1128	489		42.94	conserved hypothetical protein
SSGZ1_0001	1374	L	42.94	chromosomal replication initiator protein, DnaA
SSGZ1_1697	210		42.86	in regions AI 110: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0030	1548	F	42.64	Bifunctional phosphoribosylaminoimidazolecarboxamide formyltransferase / IMP cyclohydrolase
SSGZ1_0078	306	J	42.48	ribosomal protein L24
SSGZ1_0142	471	J	42.46	in regions AI 9: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_1621	477	I	41.93	Acetyl-CoA biotin carboxyl carrier
SSGZ1_0521	1215	J	41.15	in regions AI 35: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0451	537	K	40.97	in regions AI 32: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0351	708	K	40.96	Bacterial regulatory protein, GntR
SSGZ1_1914	759	S	40.84	hypothetical protein
SSGZ1_1277	1128	H	40.78	in regions AI 81: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0085	495	J	40.40	ribosomal protein S5

SSGZ1_0718	375	F	40.00	Anaerobic ribonucleoside-triphosphate reductase
SSGZ1_1850	450	K	40.00	regulatory protein, MarR
SSGZ1_0530	783		39.59	Predicted membrane protein
SSGZ1_0645	2283	R	39.42	ComEC/Rec2-related protein: DNA internalization-related competence protein ComEC/Rec2
SSGZ1_0395	534	J	39.33	Protein of unknown function DUF402
SSGZ1_0083	537	J	39.11	ribosomal protein L6
SSGZ1_1931	1689	J	39.08	Arginyl-tRNA synthetase
SSGZ1_1514	276	S	36.23	in regions AI 96: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0603	498	S	36.14	hypothetical protein
SSGZ1_1463	636	F	34.59	in regions AI 94: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0309	819	G	34.19	in regions AI 24: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0367	618	R	33.98	Recombination protein U
SSGZ1_1847	444	F	33.78	DeoxyUTP pyrophosphatase
SSGZ1_1559	303	K	33.00	hypothetical protein
SSGZ1_0960	243	J	32.92	ribosomal protein S20
SSGZ1_1182	426	J	32.86	ribosomal protein L11
SSGZ1_1721	1137	G	32.54	Sulfate-transporting ATPase
SSGZ1_0079	543	J	31.31	ribosomal protein L5
SSGZ1_1413	321		31.15	conserved hypothetical protein
SSGZ1_1096	1047	O	28.65	PpiC-type peptidyl-prolyl cis-trans isomerase
SSGZ1_1399	282	L	28.37	Excinuclease ABC, C subunit, N-terminal
SSGZ1_1799	177	U	28.25	in regions AI 117: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_1620	303	I	26.40	Beta-hydroxyacyl-(acyl-carrier-protein) dehydratase FabZ
SSGZ1_0441	231	S	25.97	in regions AI 31: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0644	663	L	25.64	Competence protein ComEA helix-hairpin-helix region
SSGZ1_1830	402	J	24.88	RNaseP protein

SSGZ1_1089	567		22.93	in regions AI 68: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0076	261	J	19.16	in regions AI 5: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0870	459	J	13.07	Ribosomal protein L7/L12
SSGZ1_0084	366	J	10.93	ribosomal protein L18

Table S3. The top 50 SNPs density MCG genes

Gene	Gene_Length	COG	SNP#(1kb)	Gene_Annotation
SSGZ1_0746	669	S	127.06	putative membrane protein
SSGZ1_1596	1074	P	120.11	Sulfate-transporting ATPase
SSGZ1_0185	747		112.45	Methyl-accepting chemotaxis protein
SSGZ1_1010	303		112.21	hypothetical protein
SSGZ1_0828	834	J	111.51	Modification methylase HemK
SSGZ1_1810	876	R	109.59	Predicted GTPase
SSGZ1_1107	870	E	109.20	Shikimate 5-dehydrogenase
SSGZ1_0339	744	QR	108.87	SAM-dependent methyltransferase
SSGZ1_0733	765		107.19	in regions AI 49: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_1108	1068	E	106.74	3-dehydroquinate synthase
SSGZ1_0873	930	LU	106.45	SMF protein
SSGZ1_0321	303	J	105.61	Glutamyl-tRNA(Gln) amidotransferase C subunit
SSGZ1_0271	267	T	104.87	in regions AI 21: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_1929	2544	L	104.56	MutS-1 protein
SSGZ1_0151	231	S	103.90	Uncharacterized conserved small protein
SSGZ1_1184	387	E	103.36	putative lactoylglutathione lyase
SSGZ1_0338	360	S	102.78	Iojap-related protein
SSGZ1_1486	216	L	101.85	Exonuclease VII, small subunit
SSGZ1_1947	1035	M	101.45	Putative glycosyl transferase
SSGZ1_1900	444	J	101.35	in regions AI 127: Region of the GZ1 chromosome not homologous

				to 89/1591
SSGZ1_1917	471		99.79	hypothetical protein
SSGZ1_0625	705	G	99.29	Glucosamine/galactosamine-6-phosphate isomerase
SSGZ1_1392	1161	E	99.05	aminotransferase
SSGZ1_0735	759	M	98.81	D-alanyl-D-alanine carboxypeptidase
SSGZ1_0417	1848	T	97.94	GTP-binding protein TypA
SSGZ1_1817	909	J	97.91	Dimethyladenosine transferase
SSGZ1_1284	675	G	97.78	Ribose 5-phosphate isomerase
SSGZ1_1275	618	F	97.09	NUDIX hydrolase
SSGZ1_0868	1335	J	96.63	Gid protein
SSGZ1_0322	1467	J	96.11	Glutamyl-tRNA(Gln) amidotransferase A subunit
SSGZ1_1759	594	E	95.96	3-isopropylmalate dehydratase small subunit
SSGZ1_0599	606	R	95.71	NADPH-dependent FMN reductase
SSGZ1_1321	1035	R	95.65	ABC-type uncharacterized transport system, periplasmic component
SSGZ1_0027	1455	F	94.85	Amidophosphoribosyl transferase
SSGZ1_1756	1257	GEPR	94.67	Permease of the major facilitator superfamily
SSGZ1_1241	402	M	94.53	in regions AI 77: Region of the GZ1 chromosome not homologous to 89/1591
SSGZ1_0827	1095	J	94.06	Peptide chain release factor 1
SSGZ1_0216	2334	L	93.83	DNA structure-specific ATPase involved in suppression of recombination, MutS family
SSGZ1_1093	1194	J	93.80	Alanyl-tRNA synthetase, class IIc
SSGZ1_0360	984	R	93.50	Rhodanese-like protein
SSGZ1_0815	840	F	92.86	Thymidylate synthase
SSGZ1_1736	399	S	92.73	Uncharacterized conserved protein
SSGZ1_0220	939	O	92.65	Peptide methionine sulfoxide reductase
SSGZ1_1197	1545	J	92.56	Peptide chain release factor 3
SSGZ1_0208	996	R	92.37	proposed virulence factor
SSGZ1_1502	555	QR	91.89	Putative rRNA methylase
SSGZ1_1332	654	R	91.74	HAD-superfamily hydrolase
SSGZ1_1274	774	G	91.73	HAD-superfamily subfamily IIA hydrolase

SSGZ1_1295	579	L	91.54	Uracil-DNA glycosylase superfamily
SSGZ1_1281	810	F	91.36	Purine nucleoside phosphorylase 1

Table S4. Ln probability of 85 strains using Structure from K=2 to K15 (BURNIN=50000 and NUMREPS=150000).

	Ln Prob					
K	**1**	**2**	**3**	**4**	**5**	**Avg**
3	-773633	-778824	-775478	-779350	-781946	-777846
4	-722604	-717157	-715467	-717361	-722655	-719049
5	-701979	-699990	-690599	-690914	-688266	-694350
6	-662034	-666211	-659236	-657441	-658641	-660713
7	-638266	-629659	-630083	-634596	-632543	-633029
8	-722285	-634885	-627486	-620956	-675057	-656134
9	-987961	-892512	-699480	-617733	-611452	-761828
10	-1.6E+07	-7883639	-4.1E+07	-1.3E+07	-5.1E+07	-2.6E+07
11	-2.2E+08	-3336368	-9.3E+07	-3408706	-1.3E+07	-6.7E+07
12	-1.2E+08	-1.1E+07	-2.9E+07	-3.5E+07	-2.2E+07	-4.3E+07
13	-3.5E+07	-1.1E+08	-1.1E+08	-2.6E+08	-2.4E+08	-1.5E+08
14	-4.5E+08	-5.5E+07	-8.8E+07	-1.8E+08	-1.5E+08	-1.8E+08
15	-2.4E+08	-1.6E+08	-7.4E+07	-1.3E+08	-1899634	-1.2E+08

Table S5. MCG groups and their serotype

CGG	**# of Strain**	**Serotype(# of Strain)**
1	17	1/2(1); 2(11); 7(1); 14(3); 15(1)
2	3	8(3)
3	6	4(1); 5(1); 17(1); 18(1); 19(1); 23(1)
4	8	1/2(1); 2(2); 3(2); 7(3);
5	5	11(1); NA(4)
6	18	2(1);13(1);21(1);24(1);27(1);29(4);NA(9)
7	19	6(1);9(2);10(2); 16(1);20(1);22(1);25(1);26(1);30(1);31(1);NA(7)

Ungroupable	9	1/2(1); 1(2);11(1);12(1);28(1);NA(3)

Table S6. Genetic Distance within and between minimal core genome groups[#]

	MCGG 1	MCGG 2	MCGG 3	MCGG 4	MCGG 5	MCGG 6	MCGG 7
MCGG 1	0.023						
MCGG 2	0.115	0.003					
MCGG 3	0.131	0.133	0.087				
MCGG 4	0.145	0.135	0.121	0.042			
MCGG 5	0.143	0.145	0.149	0.149	0.043		
MCGG 6	0.158	0.163	0.155	0.168	0.167	0.125	
MCGG 7	0.159	0.161	0.160	0.165	0.156	0.172	0.155

[#] MCGGs: minimal core genome groups. The P-distance was used as genetic distance and calculated using MEGA.