Supplementary Materials

This supplementary file contains: 1) Supplementary Methods, 2) three Supplementary Tables, and 3) two Supplementary Figures.

Supplementary Methods

Whole-genome sequencing and closure of reference genomes. Five strains (one from each *emm82*, *emm83*, *emm87*, *emm101*, and *emm114*; strains NGAS596, NGAS327, NGAS743, NGAS638, and NGAS322, respectively) were randomly chosen and their genomes sequenced to closure using SMRT sequencing (Pacific Biosciences, Menlo Park, CA, USA). Briefly, two SMRT cells of sequence were generated for each isolate, generating >220 Mb of data in reads exceeding 3kb in length (Table S3). We next used HGAP v2 (1) to correct the long reads and Celera Assembler 7.0 (2) to assemble the corrected reads, followed by two rounds of polishing with Quiver (https://github.com/PacificBiosciences/GenomicConsensus). The coverage of the final SMRT assemblies in reads >3 kb is presented in Table S3. We did not find evidence of any associated plasmids in any isolate. To assess base-calling accuracy in the Pacific Biosciences assembly, Illumina short-reads for the five strains were generated. Briefly, genomic libraries were prepared using Nextera XT kits (Illumina, San Diego CA) and sequenced as paired-end reads (101 bp +101 bp) in an Illumina HiSeq 2500 instrument (Illumina). Parsing of the multiplexed sequencing reads and removal of barcode information was done using onboard software. Illumina reads from the five isolates were next aligned to their respective assemblies using BLAT (3). All 5 genome assemblies were completely concordant with full length perfectly aligning Illumina short-reads. SMRT assemblies were confirmed using OpGen technology (Figure S3). Briefly, optical maps were prepared according to methods described previously (4, 5) at OpGen Technologies, Inc. (Madison, WI, USA) after digesting genomic DNA of the strains with *KpnI* and were assembled using OpGen's proprietary MapSolver software. All genome assemblies were then formatted to begin at the first nucleotide of the intergenic region immediately preceding the chromosomal replication initiation protein (DnaA, Genbank accession number AAL96837). Prophages were identified in the closed genomes using the programs Phage Finder, PHAST, and Prophage Finder (6-

8), and by manual inspection. Finalized genomes were then annotated using the Prokka pipeline (9). Genome sequences have been deposited in GenBank under accession numbers CP007561 (*emm82* strain NGAS596), CP007562 (*emm83* strain NGAS327), CP007560 (*emm87* strain NGA743), CP010450 (*emm101* strain NGAS638), and CP010449 (*emm114* strain NGAS322). Phylogenetic relationships among newly sequenced GAS genomes and circularized genomes available in GenBank were established as follows: For each strain, single-nucleotide polymorphisms (SNPs) relative to the genome of *emm59* strain MGAS15252, selected the arbitrary reference, were identified using Nucmer (10). A matrix file containing the genotype of all strains at each polymorphic locus was then created from the Nucmer polymorphism output data using a custom script. Then, for each individual strain, SNPs were concatenated in order of occurrence relative to the reference strain and converted to a multiFASTA sequence using custom scripts. A neighbor joining phylogenetic tree was then created using SplitsTree4 (11). The GenBank accession numbers for the genome sequences not generated in this study were as follows: AE009949.1, AE014074.1, AM295007.1, CP000003.1, CP000017.2, CP000056.1, CP000260.1, CP000261.1, CP000262.1, CP000829.1, CP003068.1, CP003116.1, CP006366.1, CP007240.1, and HG316453.2. Full genomes were aligned using progressive Mauve (12). The genomes of the remaining 96 GAS strains used in this study were sequenced as paired-end reads with either Illumina HiSeq (101 bp +101 bp) or MiSeq (150 bp +150 bp) instruments. Genomic libraries for Illumina sequencing were prepared as described above. The average number of 101 bp and 151 bp reads per strain was 4,584,045 and 2,749,103, respectively (maximum 8,372,724 and 4,163,997; minimum 1,570,382 and 1,270,823, respectively) and the coverage considering an average GAS genome size of 1.8 Mbp was 162 X, on average (maximum 302 X, minimum 57 X). Table S1 lists the National Center for Biotechnology Short-Sequence Archive accession numbers for each sequenced genome.

Bioinformatics analysis. Multilocus sequence-typing (MLST) was determined for all strains directly from the short-read Illumina data using the SRST2 software (13). SRST2 and its resistance gene database, which contains sequences for 1913 genes associated with resistance to antibiotics was used to detect presence or absence of genes encoding antibiotic resistance in iGAS strains (13). Illumina short-reads of each strain were aligned against reference genomes generated in this study (*emm82* strain NGAS596, *emm83* strain NGAS327, *emm87* strain NGA743, *emm101* strain NGAS638 and *emm114* strain NGAS322), or available in GenBank (*emm59* strain MGAS15252) using Mosaik (https://code.google.com/p/mosaik-aligner/). Coverage relative to the reference genomes was calculated using the same software. Polymorphisms such as single-nucleotide polymorphisms (SNPs) and short insertion/deletions (indels) were identified against the corresponding reference genomes generated in this study or available at GenBank using VAAL (14). A matrix file containing the genotype of all strains at each polymorphic locus was then created from the VAAL polymorphism output data using a custom script. Then, for each individual strain, SNPs were concatenated in order of occurrence relative to the genome of the corresponding reference strain and converted to a multiFASTA sequence. Neighbor-joining phylogenetic trees (1,000 bootstrap replications) were then generated with SplitsTree4 (11). The A5 pipeline was used for *de novo* assembly of Illumina sequenced GAS strains (15). Contigs were annotated with Prokka (9). Presence of recombination was evaluated using BRATNextGen (16). Genome visualizations were created using BRIG (17) and edited using Adobe Illustrator.

References.

1. **Chin CS, Alexander DH, Marks P, Klammer AA, Drake J, Heiner C, Clum A, Copeland A, Huddleston J, Eichler EE, Turner SW, Korlach J.** 2013. Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. Nat Methods **10:**563-569.

2. **Myers EW, Sutton GG, Delcher AL, Dew IM, Fasulo DP, Flanigan MJ, Kravitz SA, Mobarry CM, Reinert KH, Remington KA, Anson EL, Bolanos RA, Chou HH, Jordan CM, Halpern AL, Lonardi S, Beasley EM, Brandon RC, Chen L, Dunn PJ, Lai Z, Liang Y, Nusskern DR, Zhan M, Zhang Q, Zheng

X, Rubin GM, Adams MD, Venter JC. 2000. A whole-genome assembly of *Drosophila*. Science **287:**2196-2204.

3. **Kent WJ.** 2002. BLAT--the BLAST-like alignment tool. Genome Res **12:**656-664.

4. **Reslewic S, Zhou S, Place M, Zhang Y, Briska A, Goldstein S, Churas C, Runnheim R, Forrest D, Lim A, Lapidus A, Han CS, Roberts GP, Schwartz DC.** 2005. Whole-genome shotgun optical mapping of *Rhodospirillum rubrum*. Appl Env Microbiol **71:**5511-5522.

5. **Riley MC, Kirkup BC, Jr., Johnson JD, Lesho EP, Ockenhouse CF.** 2011. Rapid whole genome optical mapping of *Plasmodium falciparum*. Malaria journal **10:**252.

6. **Fouts DE.** 2006. Phage_Finder: automated identification and classification of prophage regions in complete bacterial genome sequences. Nucleic Acids Res **34:**5839-5851.

7. **Zhou Y, Liang Y, Lynch KH, Dennis JJ, Wishart DS.** 2011. PHAST: a fast phage search tool. Nucleic Acids Res **39:**W347-352.

8. **Bose M, Barber RD.** 2006. Prophage Finder: a prophage loci prediction tool for prokaryotic genome sequences. In silico biology **6:**223-227.

9. **Seemann T.** 2014. Prokka: rapid prokaryotic genome annotation. Bioinformatics **30:**2068-2069.

10. **Kurtz S, Phillippy A, Delcher AL, Smoot M, Shumway M, Antonescu C, Salzberg SL.** 2004. Versatile and open software for comparing large genomes. Genome Biol **5:**R12.

11. **Huson DH, Bryant D.** 2006. Application of phylogenetic networks in evolutionary studies. Mol Biol Evol **23:**254-267.

12. **Darling AE, Mau B, Perna NT.** 2010. progressiveMauve: multiple genome alignment with gene gain, loss and rearrangement. PLoS One **5:**e11147.

13. **Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE.** 2014. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. Genome medicine **6:**90.

14. **Nusbaum C, Ohsumi TK, Gomez J, Aquadro J, Victor TC, Warren RM, Hung DT, Birren BW, Lander ES, Jaffe DB.** 2009. Sensitive, specific polymorphism discovery in bacteria using massively parallel sequencing. Nat Methods **6:**67-69.

15. **Tritt A, Eisen JA, Facciotti MT, Darling AE.** 2012. An integrated pipeline for *de novo* assembly of microbial genomes. PLoS One **7:**e42304.

16. **Marttinen P, Hanage WP, Croucher NJ, Connor TR, Harris SR, Bentley SD, Corander J.** 2012. Detection of recombination events in bacterial genomes from large population samples. Nucleic Acids Res **40:**e6.

17. **Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA.** 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. BMC Genomics **12:**402.

Table S1. Group A *Streptococcus* strains used in this study.

Strain name	Year	Isolation source	Geographical area[a]	*emm* type	MLST[b]								*mrp*	*sof*	SRA[c] acccesion number
					ST	*gki*	*gtr*	*murI*	*mutS*	*recP*	*xpt*	*yqiL*			
NGAS298	2011	Blood	C	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424640
NGAS299	2011	Blood	H	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424641
NGAS300	2011	Synovial Fluid	C	*emm83*	5	2	2	2	3	2	3	2	+	-	SRX424642
NGAS302	2011	Blood	H	*emm28*	52	11	6	14	5	9	17	19	+	+	SRX424643
NGAS303	2011	Blood	D	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765545
NGAS304	2011	Blood	D	*emm28*	52	11	6	14	5	9	17	19	+	+	SRX424644
NGAS305	2011	Synovial Fluid	G	*emm1*	28	4	3	4	4	4	2	4	-	-	SRX424645
NGAS306	2011	Synovial Fluid	B	*emm83*	103*	2	2	2	3*	37	3	2	+	-	SRX424646
NGAS308	2011	Blood	H	*emm18*	NF	6	5	7	1	5	0	3	-	-	SRX424647
NGAS310	2011	Blood	NA	*emm114*	188	2	31	8	25	52	2	27	+	+	SRX765546
NGAS311	2011	Blood	D	*emm118*	167	29	32	2	5	48	5	21	+	+	SRX424648
NGAS313	2011	Soft Tissue	C	*emm83*	103*	2	2	2	3*	37	3	2	+	-	SRX765547
NGAS314	2011	Soft Tissue	E	*emm114*	188	2	31	8	25	52	2	27	+	+	SRX765548
NGAS315	2011	Soft Tissue	A	*emm83*	5	2	2	2	3	2	3	2	+	-	SRX765549
NGAS320	2011	Synovial Fluid	C	*emm83*	103*	2	2	2	3*	37	3	2	-*	-	SRX424653
NGAS321	2011	Synovial Fluid	NA	*emm59*	172	56	24	39	7	30	2	33	+	+	SRX424654
NGAS322	2011	Blood	H	*emm114*	188	2	31	8	25	52	2	27	+	+	SRX424655
NGAS323	2011	Soft Tissue	B	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424656
NGAS324	2011	Soft Tissue	B	*emm83*	103*	2	2	2	3*	37	3	2	+	-	SRX765550
NGAS325	2011	Blood	D	*emm22*	46*	9*	8	1	1	1	3	4	+	+	SRX424657
NGAS327	2011	Blood	C	*emm83*	5	2	2	2	3	2	3	2	+	-	SRX765551
NGAS328	2011	Soft Tissue	NA	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424658
NGAS329	2011	Soft Tissue	B	*emm83*	103*	2	2	2	3*	37	3	2	+	-	SRX765552
NGAS330	2011	Blood	D	*emm41*	579	2	2	32	2	2	2	2	+	-	SRX424659
NGAS332	2011	Soft Tissue	C	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424660
NGAS335	2011	Blood	B	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424662
NGAS336	2011	Blood	E	*emm1*	28	4	3	4	4	4	2	4	-	-	SRX424663
NGAS337	2011	Blood	E	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424664
NGAS338	2011	Blood	NA	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424665

NGAS339	2011	Soft Tissue	D	*emm59*	172	56	24	39	7	30	2	33	+	+	SRX424666
NGAS340	2011	Soft Tissue	B	*emm118*	167	29	32	2	5	48	5	21	+	+	SRX424667
NGAS341	2011	Blood	H	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424668
NGAS344	2011	Blood	B	*emm75*	49	11	2	1	3	12	3	7	+	+	SRX424669
NGAS345	2011	Soft Tissue	NA	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424670
NGAS346	2011	Soft Tissue	E	*emm41*	579	2	2	32	2	2	2	2	+	-	SRX424671
NGAS347	2011	Soft Tissue	H	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424672
NGAS592	2012	Soft Tissue	NA	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424709
NGAS594	2012	Soft Tissue	NA	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424710
NGAS595	2012	Soft Tissue	C	*emm89*	101	16	2	8	3	1	13	3	+	+	SRX424711
NGAS596	2012	Blood	C	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424712
NGAS597	2012	Blood	A	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424713
NGAS599	2012	Soft Tissue	C	*emm114*	188	2	31	8	25	52	2	27	+	+	SRX424714
NGAS600	2012	Soft Tissue	NA	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765554
NGAS602	2012	Soft Tissue	C	*emm114*	188	2	31	8	25	52	2	27	+	+	SRX765555
NGAS603	2012	Blood	C	*emm6*	382	5	52	5	5	5	4	3	-	-	SRX424716
NGAS604	2012	Blood	B	*emm75*	49	11	2	1	3	12	3	7	+	+	SRX424717
NGAS605	2012	Soft Tissue	A	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424718
NGAS606	2012	Blood	D	*emm80*	538	11	2	69	3	84	13	83	+	-	SRX424719
NGAS608	2012	Soft Tissue	F	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424720
NGAS609	2012	Blood	I	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424721
NGAS610	2012	Synovial Fluid	H	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424722
NGAS612	2012	Soft Tissue	H	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424723
NGAS613	2012	Blood	D	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424724
NGAS615	2012	Soft Tissue	D	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424725
NGAS616	2012	Soft Tissue	D	*emm22*	46*	9*	8	1	1	1	3	4	+	+	SRX424726
NGAS618	2012	Blood	H	*emm6*	382	5	52	5	5	5	4	3	-	-	SRX424727
NGAS621	2012	Soft Tissue	C	*emm83*	103*	2	2	2	3*	37	3	2	+	-	SRX765556
NGAS624	2012	Other[d]	C	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765557
NGAS625	2012	Soft Tissue	B	*emm114*	188	2	31	8	25	52	2	27	+	+	SRX424730
NGAS626	2012	Soft Tissue	D	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765558
NGAS628	2012	Blood	B	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424731

NGAS630	2012	Soft Tissue	NA	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424733
NGAS631	2012	Blood	B	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765559
NGAS632	2012	Blood	D	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765560
NGAS634	2012	Blood	A	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424735
NGAS638	2012	Blood	D	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX424736
NGAS639	2012	Other	C	*emm59*	172	56	24	39	7	30	2	33	+	+	SRX424737
NGAS641	2012	Soft Tissue	A	*emm114*	188	2	31	8	25	52	2	27	+	+	SRX424739
NGAS739	2013	Blood	B	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424764
NGAS742	2013	Soft Tissue	A	*emm80*	538	11	2	69	3	84	13	83	+	-	SRX765561
NGAS743	2013	Soft Tissue	A	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424765
NGAS746	2013	Synovial Fluid	D	*emm4*	39	5	11	8	5	15	2	1	+	+	SRX424766
NGAS747	2013	Blood	B	*emm87*	62	13	3	10	12	20	2	3	+	+	SRX424767
NGAS748	2013	Blood	C	*emm1*	28	4	3	4	4	4	2	4	-	-	SRX765562
NGAS749	2013	Blood	H	*emm12*	36	5	2	2	6	6	2	2	-	+	SRX424768
NGAS750	2013	Other	H	*emm1*	28	4	3	4	4	4	2	4	-	-	SRX424769
NGAS755	2013	Soft Tissue	C	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX424771
NGAS756	2013	Blood	D	*emm4*	39	5	11	8	5	15	2	1	+	+	SRX765563
NGAS757	2013	Soft Tissue	C	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX424772
NGAS758	2013	Soft Tissue	D	*emm4*	39	5	11	8	5	15	2	1	+	+	SRX424773
NGAS768	2013	Synovial Fluid	D	*emm4*	39	5	11	8	5	15	2	1	+	+	SRX765568
NGAS769	2013	Other	D	*emm4*	39	5	11	8	5	15	2	1	+	+	SRX765569
NGAS772	2013	Synovial Fluid	I	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765572
NGAS774	2013	Soft Tissue	D	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765574
NGAS776	2013	Blood	A	*emm82*	334	84	2	21	16	17	3	1	+	+	SRX765576
NGAS778	2013	Soft Tissue	C	*emm4*	39	5	11	8	5	15	2	1	+	+	SRX765578
NGAS779	2013	Soft Tissue	D	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765579
NGAS781	2013	Soft Tissue	D	*emm101*	182	2	2	37	2	2	13	1	+	-	SRX765581
NGAS785	2013	Soft Tissue	A	*emm68*	247	11	9	1	7	2	8	3	+	+	SRX765585
NGAS786	2013	Other	I	*emm1*	28	4	3	4	4	4	2	4	-	-	SRX765586
NGAS787	2013	Blood	A	*emm12*	36	5	2	2	6	6	2	2	-	+	SRX765587
NGAS788	2013	Soft Tissue	B	*emm53*	347	4	31	2	11	34	3	2	+	+	SRX765588
NGAS789	2013	Blood	C	*emm80*	538	11	2	69	3	84	13	83	+	-	SRX765589

NGAS791	2013	Blood	D	*emm11*	NF	3	4	6	7	1	5	0	+	+	SRX765591
NGAS793	2013	Soft Tissue	H	*emm80*	538	11	2	69	3	84	13	83	+	-	SRX765593
NGAS794	2013	Soft Tissue	H	*emm9*	75	15	14	7	18	19	3	1	+	+	SRX765594
NGAS795	2013	Soft Tissue	D	*emm11*	NF	3	4	6	7	1	5	0	+	+	SRX765595
NGAS796	2013	Blood	C	*emm80*	538	11	2	69	3	84	13	83	+	-	SRX765596
NGAS797	2013	Soft Tissue	B	*emm68*	86	21	31	1	3	28	3	4	+	+	SRX765597
NGAS798	2013	Blood	A	*emm114*	188	2	31	8	25	52	2	27	+	+	SRX765598
NGAS799	2013	Blood	I	Untyped	182	2	2	37	2	2	13	1	+	-	SRX765599
NGAS800	2013	Soft Tissue	D	*emm4*	39	5	11	8	5	15	2	1	+	+	SRX765600

---

[a] Geographical areas are Thunder Bay District regions with same three first digits of Canadian Postal Code.

[b] As determined by deriving information from short-read whole-genome sequencing (see methods). NF: Not found in MLST database. *: mismatches relative to the listed allele were found.

[c] National Center for Biotechnology information Sequence Read Archive.

[d] Other: undefined normally sterile anatomical site.

Table S2. Presence of genes putatively conferring antibiotic resistance among invasive Group A *Streptococcus* strains.

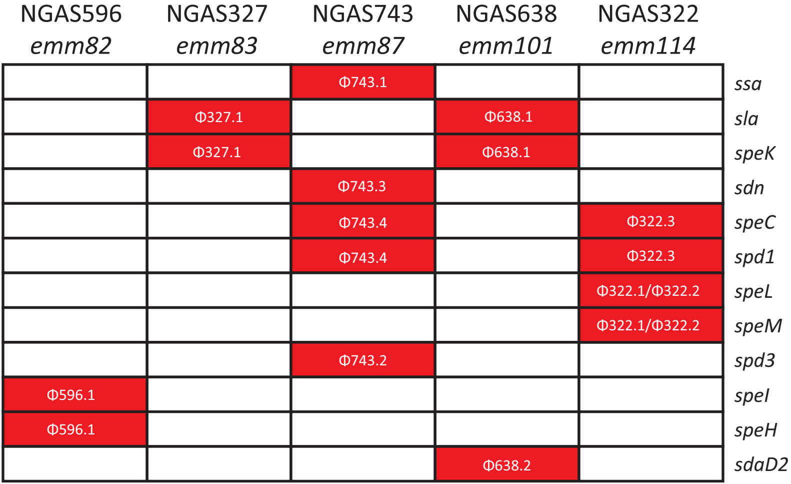
Strain name	*emm* type	MLST ST	Gene associated with antimicrobial resistance									
			aph(3')-III	*dfrG*	*dfrK*	*erm(A)*	*erm(B)*	*lsa(A)*	*mef(A)*	*msr(D)*	*tet(M)*	*tet(S)*
NGAS298	87	62	-	-	-	-	-	-	-	-	-	-
NGAS299	87	62	-	-	-	-	-	-	-	-	-	-
NGAS300	83	5	-	-	-	-	-	-	-	-	-	-
NGAS302	28	52	-	-	-	-	-	-	-	-	-	-
NGAS303	101	182	-	-	-	-	-	-	-	-	-	-
NGAS304	28	52	-	-	-	-	-	-	-	-	-	-
NGAS305	1	28	-	-	-	-	-	-	-	-	-	-
NGAS306	83	103	+	-	-	+	-	-	-	-	+	-
NGAS308	18	NF	-	-	-	-	-	-	-	-	-	-
NGAS310	114	188	+	-	-	+	-	-	-	-	+	+
NGAS311	118	167	-	-	-	-	-	-	-	-	-	-
NGAS313	83	103	+	-	-	+	-	-	-	-	+	+
NGAS314	114	188	+	-	-	+	-	-	-	-	+	+
NGAS315	83	5	-	-	-	-	-	-	-	-	-	-
NGAS320	83	103	+	-	-	+	-	-	-	-	+	-
NGAS321	59	172	-	-	-	-	-	+	-	-	-	-
NGAS322	114	188	+	-	-	+	-	-	-	-	+	+
NGAS323	82	334	-	-	-	-	-	-	-	-	-	-
NGAS324	83	103	+	-	-	+	-	-	-	-	+	+
NGAS325	22	46	-	-	-	-	-	-	-	-	-	-
NGAS327	83	5	-	-	-	-	-	-	-	-	-	-
NGAS328	82	334	-	-	-	-	-	-	-	-	-	-
NGAS329	83	103	+	-	-	+	-	-	-	-	+	+
NGAS330	41	579	-	-	-	-	-	-	-	-	-	-
NGAS332	82	334	-	-	-	-	-	-	-	-	-	-
NGAS335	87	62	-	-	-	-	-	-	-	-	-	-
NGAS336	1	28	-	-	-	-	-	-	-	-	-	-
NGAS337	82	334	-	-	-	-	-	+	-	-	-	-
NGAS338	82	334	-	-	-	-	-	-	-	-	-	-
NGAS339	59	172	-	-	-	-	-	-	-	-	-	-
NGAS340	118	167	-	-	-	-	-	-	+	+	-	-
NGAS341	87	62	-	-	-	-	-	-	-	-	-	-
NGAS344	75	49	-	-	-	-	-	-	-	-	-	-

NGAS345	82	334	-	-	-	-	-	-	-	-	-
NGAS346	41	579	-	-	-	-	-	-	-	-	-
NGAS347	82	334	-	-	-	-	-	+	-	-	-
NGAS592	82	334	-	-	-	-	-	-	-	-	-
NGAS594	82	334	-	-	-	-	-	-	-	-	-
NGAS595	89	101	-	-	-	-	-	-	-	-	-
NGAS596	82	334	-	-	-	-	-	-	-	-	-
NGAS597	82	334	-	-	-	-	-	-	-	-	-
NGAS599	114	188	+	-	-	+	-	-	-	-	+	+
NGAS600	101	182	-	-	-	-	-	-	-	-	-	-
NGAS602	114	188	+	-	-	+	-	-	-	-	+	+
NGAS603	6	382	-	-	-	-	-	-	-	-	-	-
NGAS604	75	49	-	-	-	-	-	-	-	-	-	-
NGAS605	82	334	-	-	-	-	-	-	-	-	-	-
NGAS606	80	538	-	-	-	-	-	-	-	-	-	-
NGAS608	82	334	-	-	-	-	-	-	-	-	-	-
NGAS609	82	334	-	-	-	-	-	-	-	-	-	-
NGAS610	87	62	-	-	-	-	-	-	-	-	-	-
NGAS612	87	62	-	-	-	-	-	-	-	-	-	-
NGAS613	82	334	-	-	-	-	-	-	-	-	-	-
NGAS615	82	334	-	-	-	-	-	-	-	-	-	-
NGAS616	22	46	-	-	-	-	-	-	-	-	-	-
NGAS618	6	382	-	-	-	-	-	-	-	-	-	-
NGAS621	83	103	+	-	-	+	-	-	-	-	+	+
NGAS624	101	182	-	-	-	-	-	-	-	-	-	-
NGAS625	114	188	+	-	-	+	-	-	-	-	+	-
NGAS626	101	182	-	-	-	-	-	-	-	-	-	-
NGAS628	87	62	-	-	-	-	-	-	-	-	-	-
NGAS630	87	62	-	-	-	-	-	-	-	-	-	-
NGAS631	101	182	-	-	-	-	-	-	-	-	-	-
NGAS632	101	182	-	-	-	-	-	-	-	-	-	-
NGAS634	87	62	-	-	-	-	-	-	-	-	-	-
NGAS638	101	182	-	-	-	-	-	-	-	-	-	-
NGAS639	59	172	-	-	-	-	-	-	-	-	-	-
NGAS641	114	188	+	-	-	+	-	-	-	-	+	-
NGAS739	87	62	-	-	-	-	-	-	-	-	-	-

NGAS742	80	538	-	-	-	-	-	-	-	-	-	-
NGAS743	87	62	-	-	-	-	-	-	-	-	-	-
NGAS746	4	39	-	-	-	-	-	-	-	-	-	-
NGAS747	87	62	-	-	-	-	-	-	-	-	-	-
NGAS748	1	28	-	-	-	-	-	-	-	-	-	-
NGAS749	12	36	-	-	-	-	-	-	+	+	-	-
NGAS750	1	28	-	-	-	-	-	-	-	-	-	-
NGAS755	101	182	-	-	-	-	-	-	-	-	-	-
NGAS756	4	39	-	-	-	-	-	-	-	-	-	-
NGAS757	82	334	-	-	-	-	-	-	-	-	-	-
NGAS758	4	39	-	-	-	-	-	-	-	-	-	-
NGAS768	4	39	-	-	-	-	-	-	-	-	-	-
NGAS769	4	39	-	-	-	-	-	-	-	-	-	-
NGAS772	101	182	-	-	-	-	-	-	-	-	-	-
NGAS774	101	182	-	-	-	-	-	-	-	-	-	-
NGAS776	82	334	-	-	-	-	-	-	-	-	-	-
NGAS778	4	39	-	-	-	-	-	-	-	-	-	-
NGAS779	101	182	-	-	-	-	-	-	-	-	-	-
NGAS781	101	182	-	-	-	-	-	-	-	-	-	-
NGAS785	68	247	-	-	-	-	-	-	-	-	+	-
NGAS786	1	28	-	-	-	-	-	-	-	-	-	-
NGAS787	12	36	-	-	-	-	-	-	-	-	-	-
NGAS788	53	347	-	+	+	-	-	-	-	-	+	-
NGAS789	80	538	-	-	-	-	-	-	-	-	-	-
NGAS791	11	NF	-	-	-	+	-	-	-	-	-	-
NGAS793	80	538	-	-	-	-	-	-	-	-	-	-
NGAS794	9	75	-	-	-	-	+	-	-	-	+	-
NGAS795	11	NF	-	-	-	+	-	-	-	-	-	-
NGAS796	80	538	-	-	-	-	-	-	-	-	-	-
NGAS797	68	86	-	-	-	-	-	-	-	-	-	-
NGAS798	114	188	+	-	-	+	-	-	-	-	+	-
NGAS799	Untyped	182	-	-	-	-	-	-	-	-	-	-
NGAS800	4	39	-	-	-	-	-	-	-	-	-	-

Table S3. Whole-genome sequencing statistics.

			SMRT sequencing				Illumina sequencing				
Strain name	NCBI Accession	*emm* type	Number of > 3kb reads	Average length of reads >3 kB	Total bases (Mb)	Coverage	Number of reads	Read length (nt)	Total bases (Mb)	Coverage	Genome size (Mbps)
NGAS596	CP007561	82	67,797	5,823	394.8	220	4943624	101	499.3	279	1,791,306
NGAS327	CP007562	83	55,567	5,984	332.5	195	3858344	101	389.7	229	1,702,054
NGAS743	CP007560	87	62,014	6,070	376.4	196	5915890	101	597.5	312	1,915,554
NGAS638	CP010450	101	90,846	6,072	551.6	308	4498930	101	454.4	254	1,791,401
NGAS322	CP010449	114	44,361	4,991	221.4	114	4920574	101	497.0	255	1,950,469



NGAS596 *emm82*	NGAS327 *emm83*	NGAS743 *emm87*	NGAS638 *emm101*	NGAS322 *emm114*	
		Φ743.1			*ssa*
	Φ327.1		Φ638.1		*sla*
	Φ327.1		Φ638.1		*speK*
		Φ743.3			*sdn*
		Φ743.4		Φ322.3	*speC*
		Φ743.4		Φ322.3	*spd1*
				Φ322.1/Φ322.2	*speL*
				Φ322.1/Φ322.2	*speM*
		Φ743.2			*spd3*
Φ596.1					*speI*
Φ596.1					*speH*
			Φ638.2		*sdaD2*

Figure S1. Prophage content of newly sequenced genomes of *emm82*, *emm83*, *emm87*, *emm101* and *emm114* iGAS strains. All of the sequenced iGAS strains had one or multiple prophages, all of which encoded one or two proven or putative secreted virulence factors, such as secreted pyrogenic-toxin-superantigens: *speC, speH, speI, speK, speL, speM*, and *ssa*; secreted DNAses: *sdaD2, sdn, spd1, spd3*; secreted phospholipase *sla*.

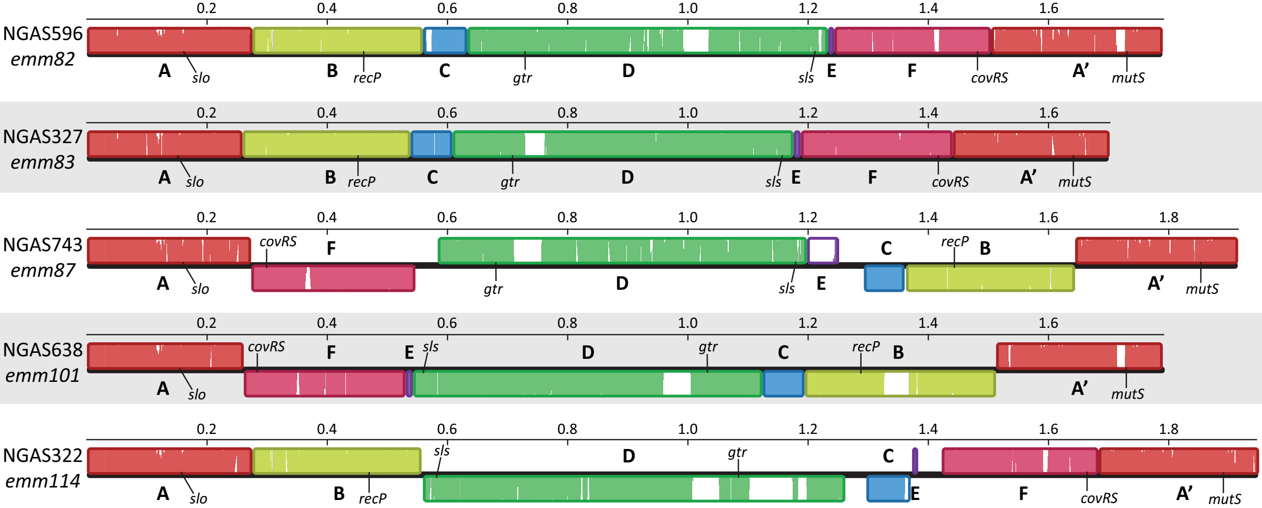


Figure S2. Genomic rearrangements in newly closed GAS genomes. Genomes were aligned using progressive Mauve as described in the Supplementary Methods. Locally collinear blocks (LCB), or conserved genome segments that appear to be internally free from rearrangements are represented in like colors (arbitrarily labelled A-G). LCBs above the center line are in the forward orientation relative to the first depicted genome while regions below the center line are in the reverse orientation. White spaces and gaps represent regions that are unique to a particular genome, prophages, and other mobile genetic elements. A few landmark genes are depicted to provide reference.