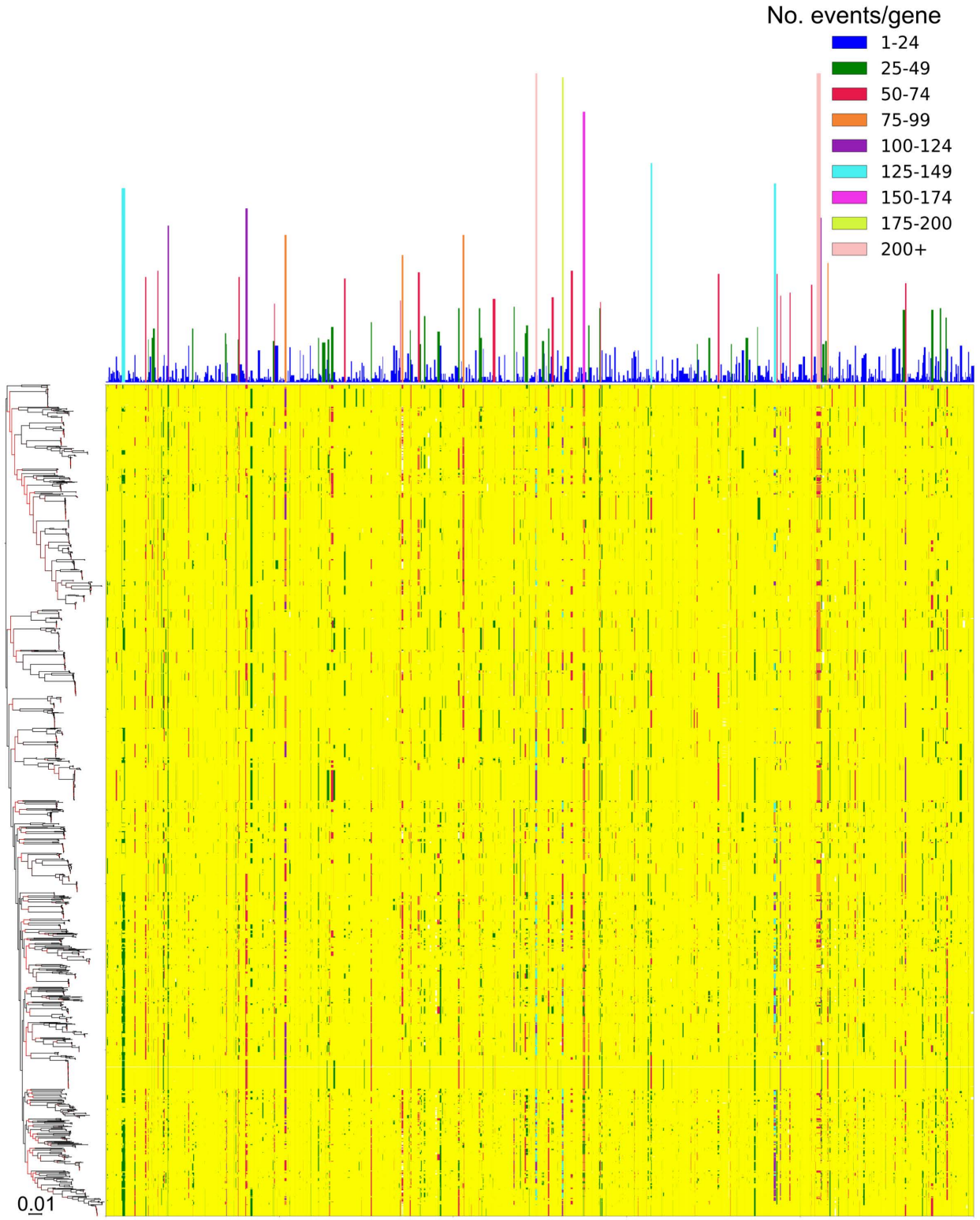
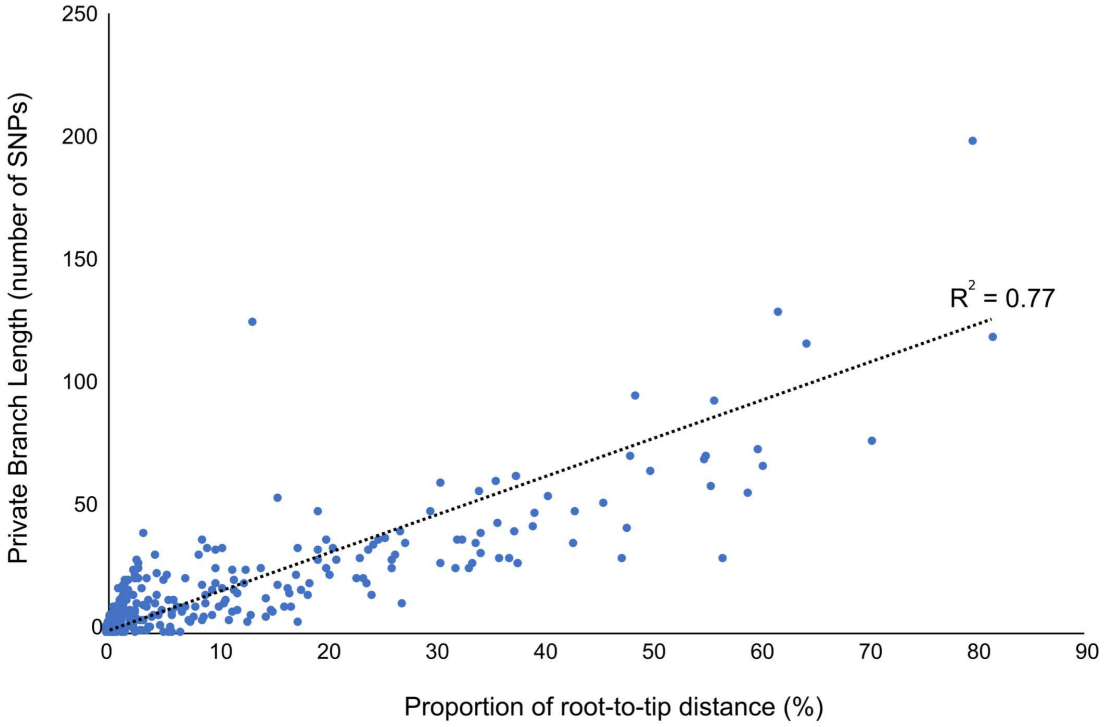
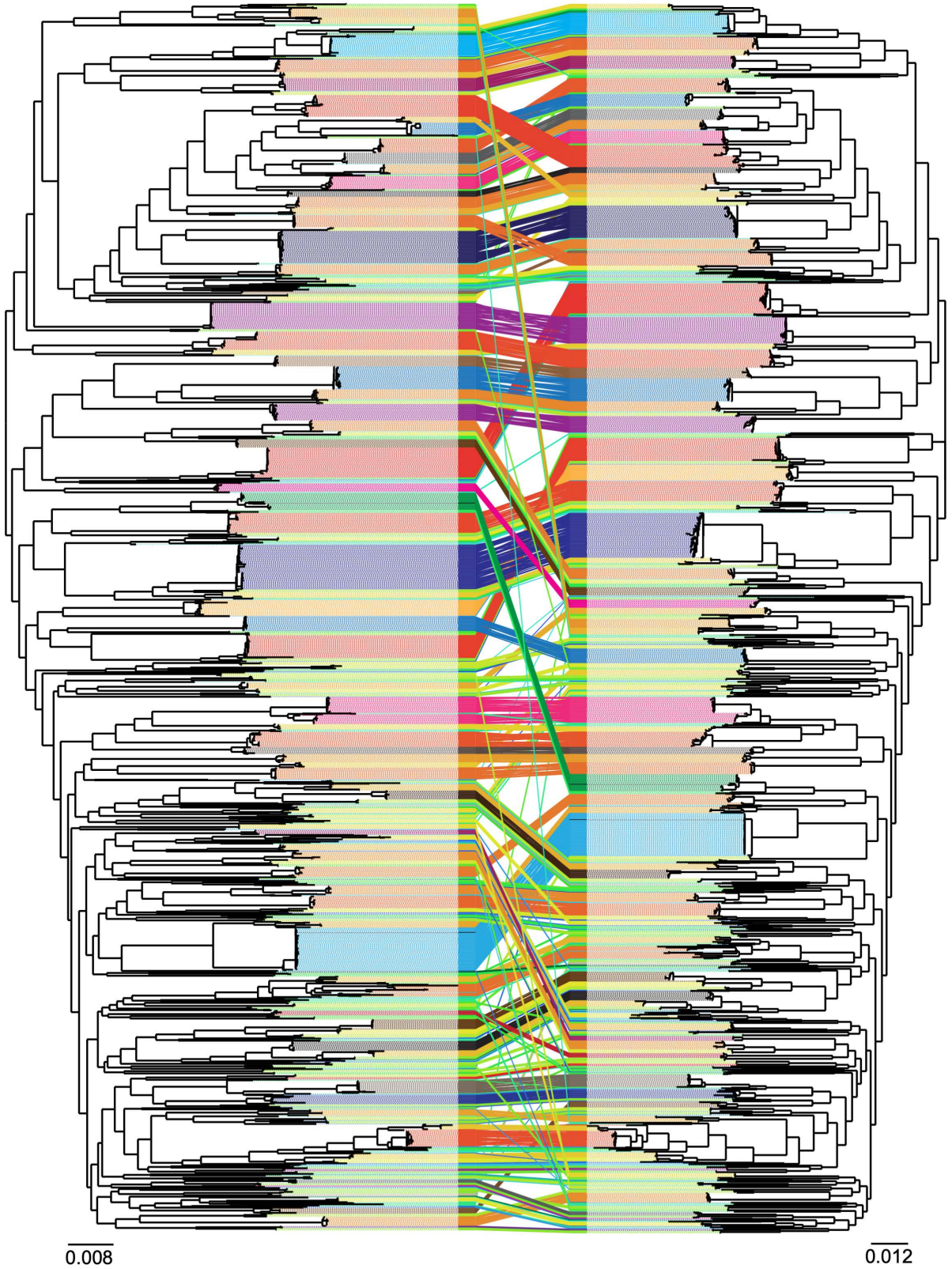


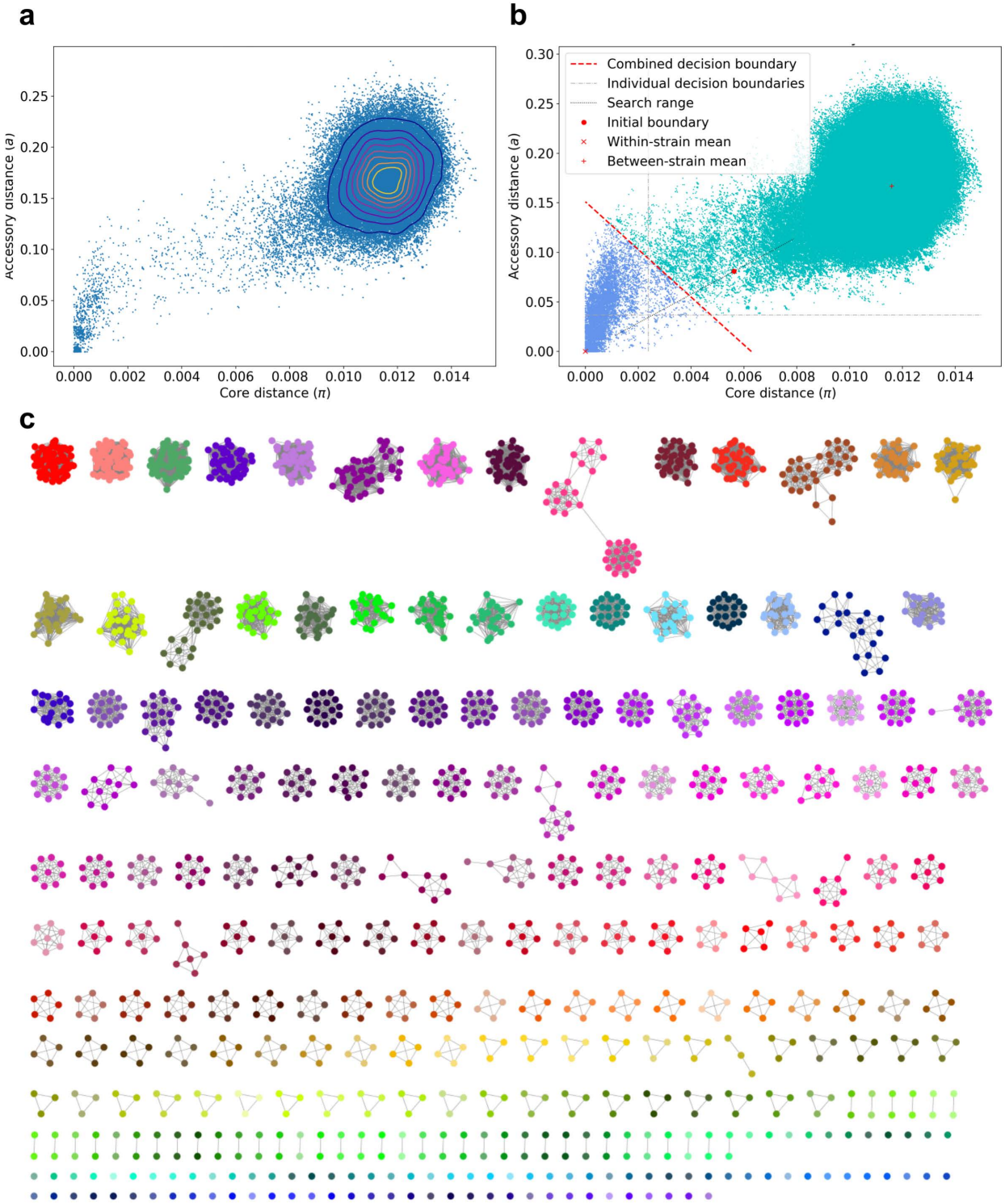
Supplementary Figure 1. Recombination analysis of 890 recombinogenic core GAS genes. Maximum-likelihood phylogenetic tree of the core global GAS genome (416 'non-recombinogenic' genes; 30,738 SNPs) is shown on the left. Middle panel shows fastGEAR[16] outputs where each gene was analysed independently with recombinations coloured by donor lineage for each gene. Yellow is used to represent the most frequent lineage for each gene to optimise visualisation. Homologous recombination blocks are represented as gene fragments harboring multiple 'colors'. The plot on top shows the number of homologous recombination events detected per gene.



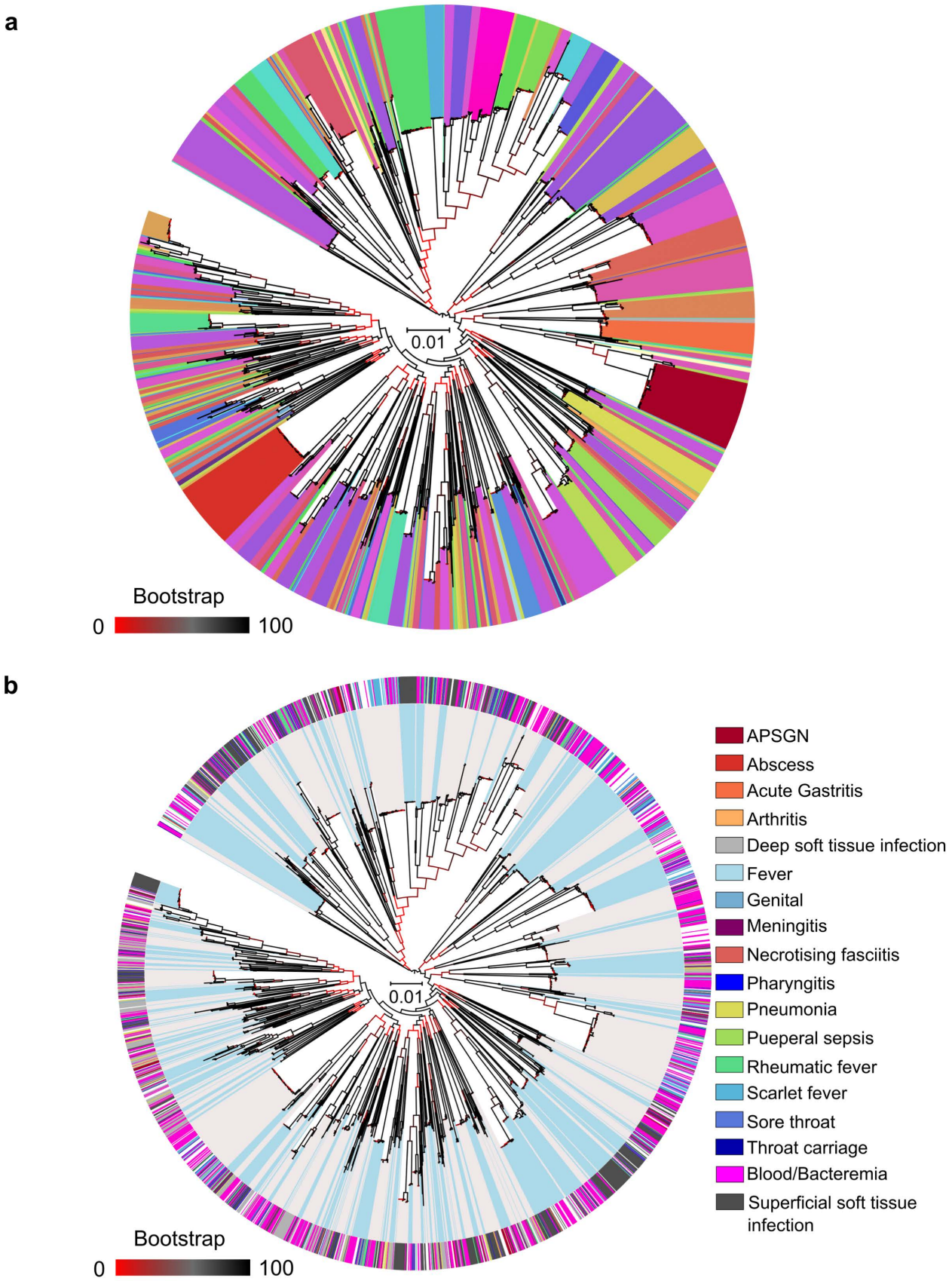
Supplementary Figure 2. Support for the core genome phylogeny. Correlation between private SNPs (i.e. SNPs unique to each genome) and the length of the branch leading to that genome in the maximum-likelihood phylogenetic tree (416 genes), displayed as a proportion of the total length to the root. This correlation indicates that the raw SNP data supports the deep branching observed in the phylogeny and this is not an artefact of enforcing a tree-like structure onto the data.



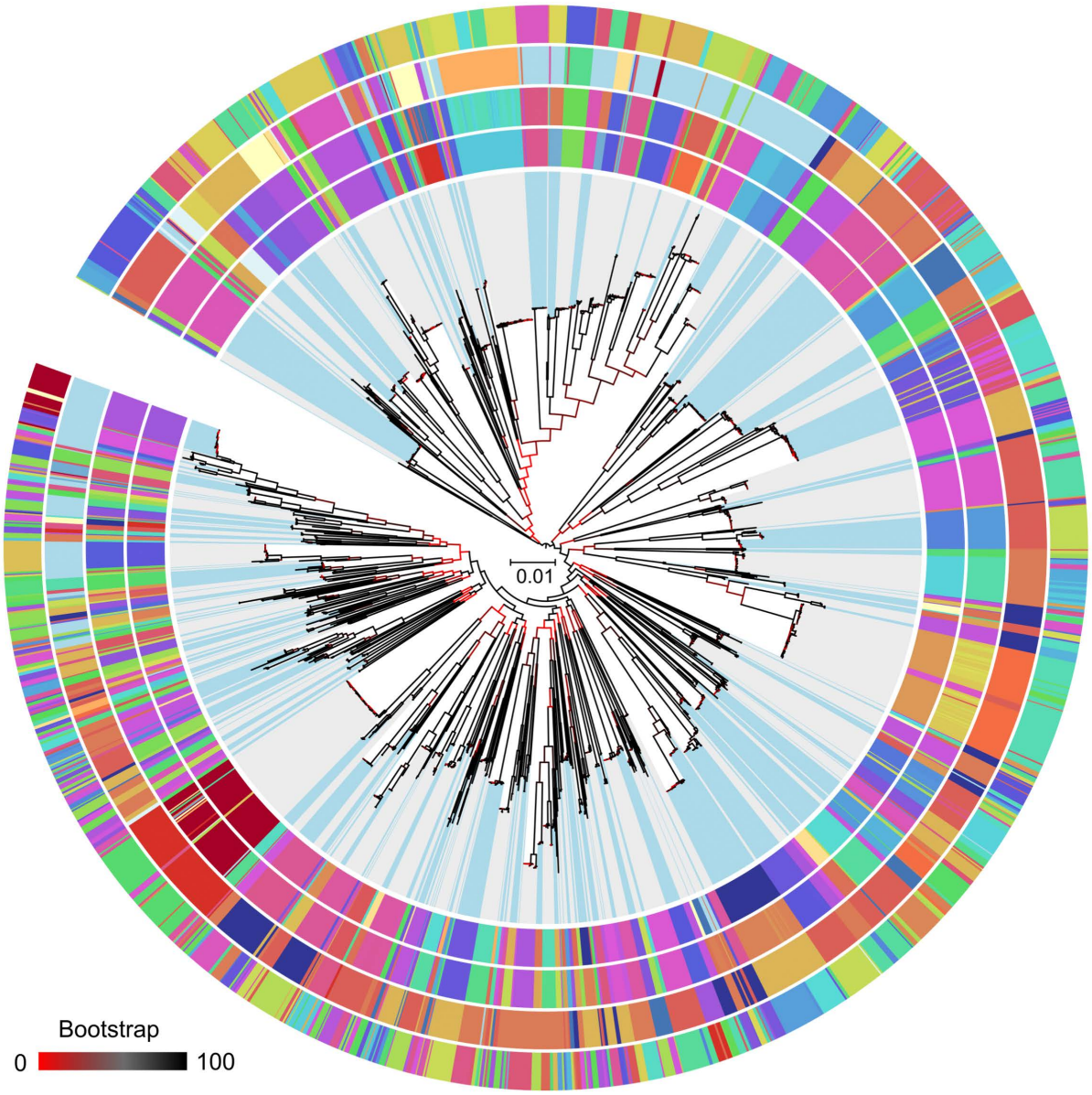
Supplementary Figure 3. Tanglegram comparison of maximum-likelihood phylogenetic tree topologies based on the 1,306 gene core genome (left: 170,653 SNPs) and the 416 gene core genome (right: 30,738 SNPs) after the removal of 'recombinogenic' genes. Colours refer to 299 different PopPUNK[17] phylogroupings. While the terminal clustering of the trees has not changed after removal of 'recombinogenic' genes, recombination has distorted the ancestral evolutionary signal within the core GAS genome.



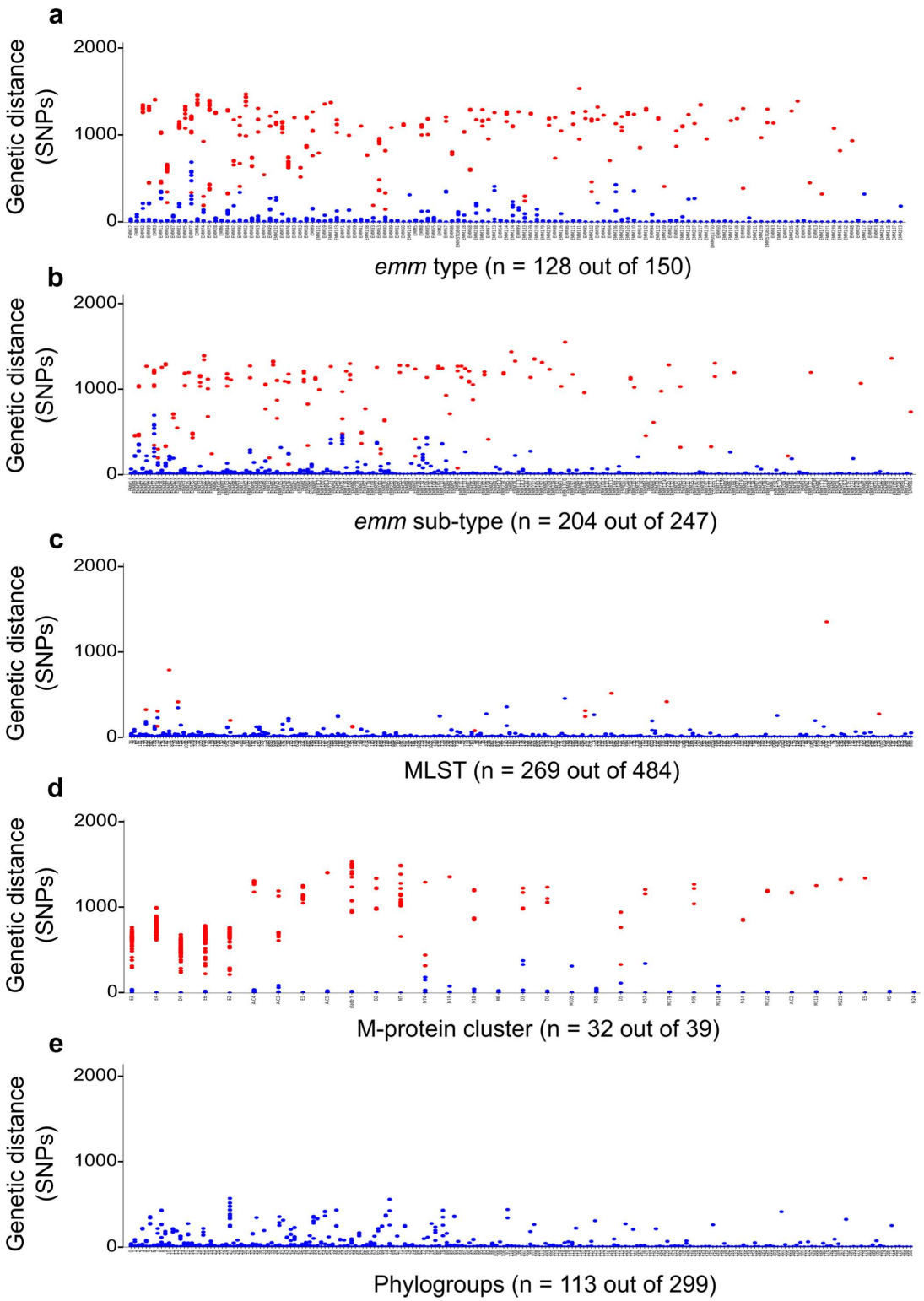
Supplementary Figure 4. PopPUNK[17] model fitting and refined network clustering of 2,083 GAS genomes. (**a**) Scatter plot showing core (π) and accessory (a) distance between all pairs of isolates with density contours displayed. (**b**) Same scatter plot as (a) after network score fit refinement for within cluster boundary. Blue points are distances comparing genomes within the same phylogroup, turquoise points are distances comparing genomes in different phylogroups. The red dashed line separates these assignments, grey dashed lines are the fit using core and accessory distances only. These data suggest low within-strain recombination (dense cluster of points near the origin of the graph), but high between-strain recombination (single broad cluster of between-strain points). (**c**) Network assignment of refined 299 PopPUNK clusters where samples are nodes (coloured by assigned cluster) and edges are pairwise links judged to be within the same cluster using the refined fit shown in (b).



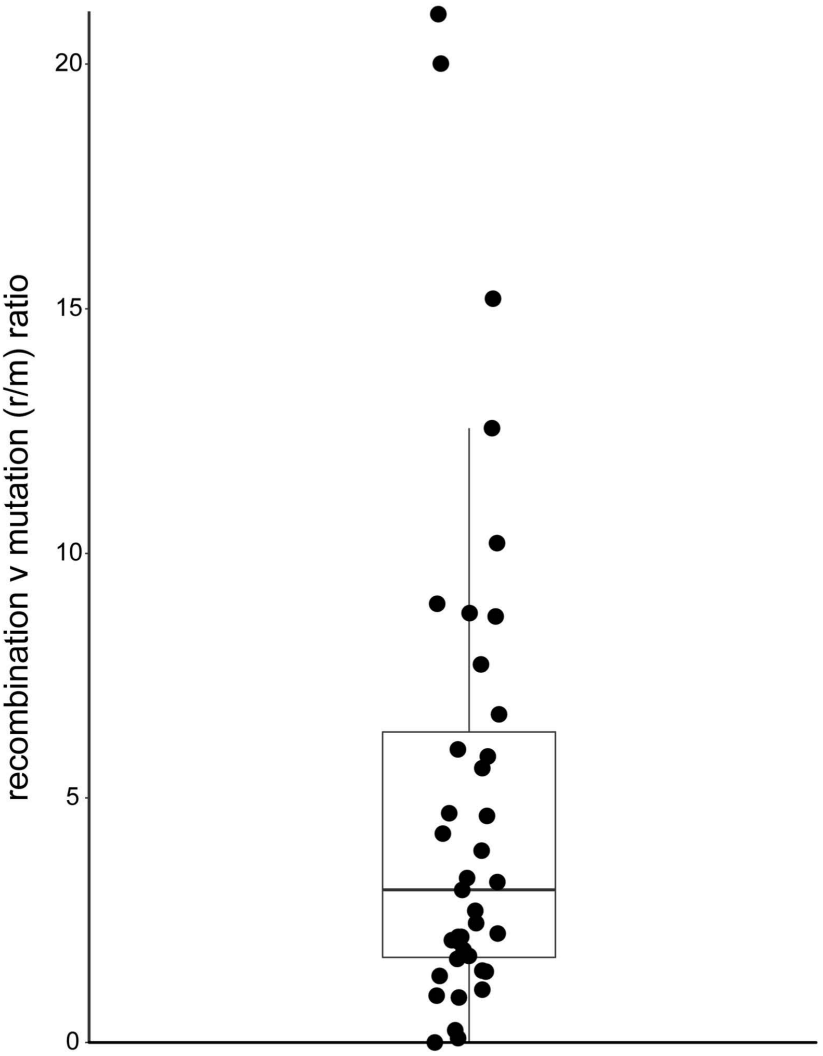
Supplementary Figure 5. Population structure of 2,083 GAS genomes based on (**a**) 299 phylogroups and (**b**) their clinical association. Maximum-likelihood phylogenetic tree of core global GAS genome (416 genes) as displayed in Fig. 1. Branch colours indicate bootstrap support according to the legend. Distinct genetic phylogroups (n = 299) are assigned a unique colour to aid in visual designation of clusters in panel (**a**) while for each phylogroup in panel (**b**) two alternate colours (blue and grey) are assigned (as in Fig. 1).



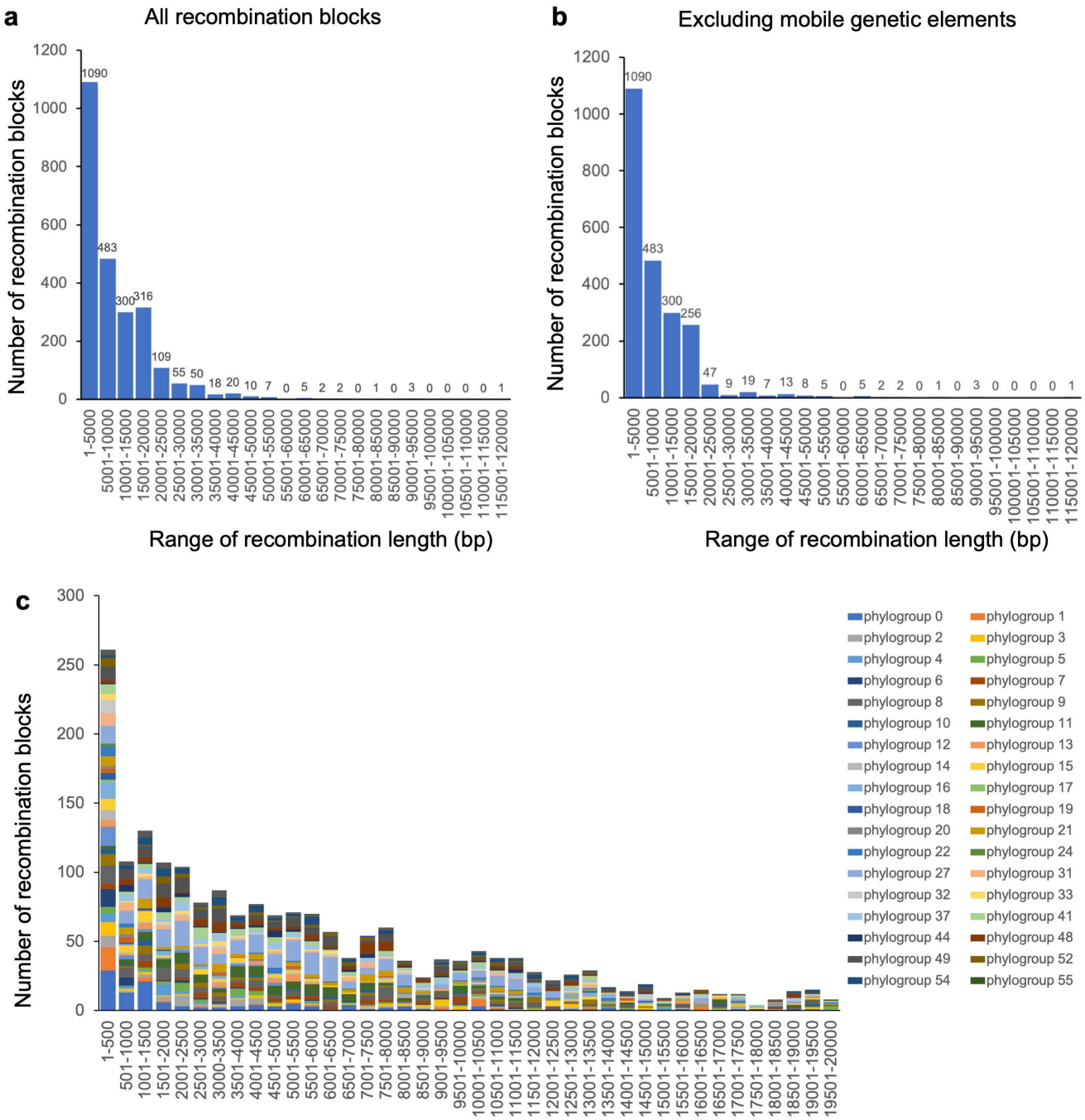
Supplementary Figure 6. Population structure of 2,083 GAS genomes and their association with primary GAS epidemiological markers. Maximum-likelihood phylogenetic tree of core global GAS genome (416 genes) as displayed in Fig. 1. Branch colours indicate bootstrap support according to the legend. Distinct genetic lineages (n = 299) are highlight in alternating colours (blue and grey) from the tips of the tree. Represented from inner to outer rings of the epidemiological data are *emm*-type (n = 150), *emm* sub-type (n = 347), M-cluster (n = 39), and MLST (n = 484).



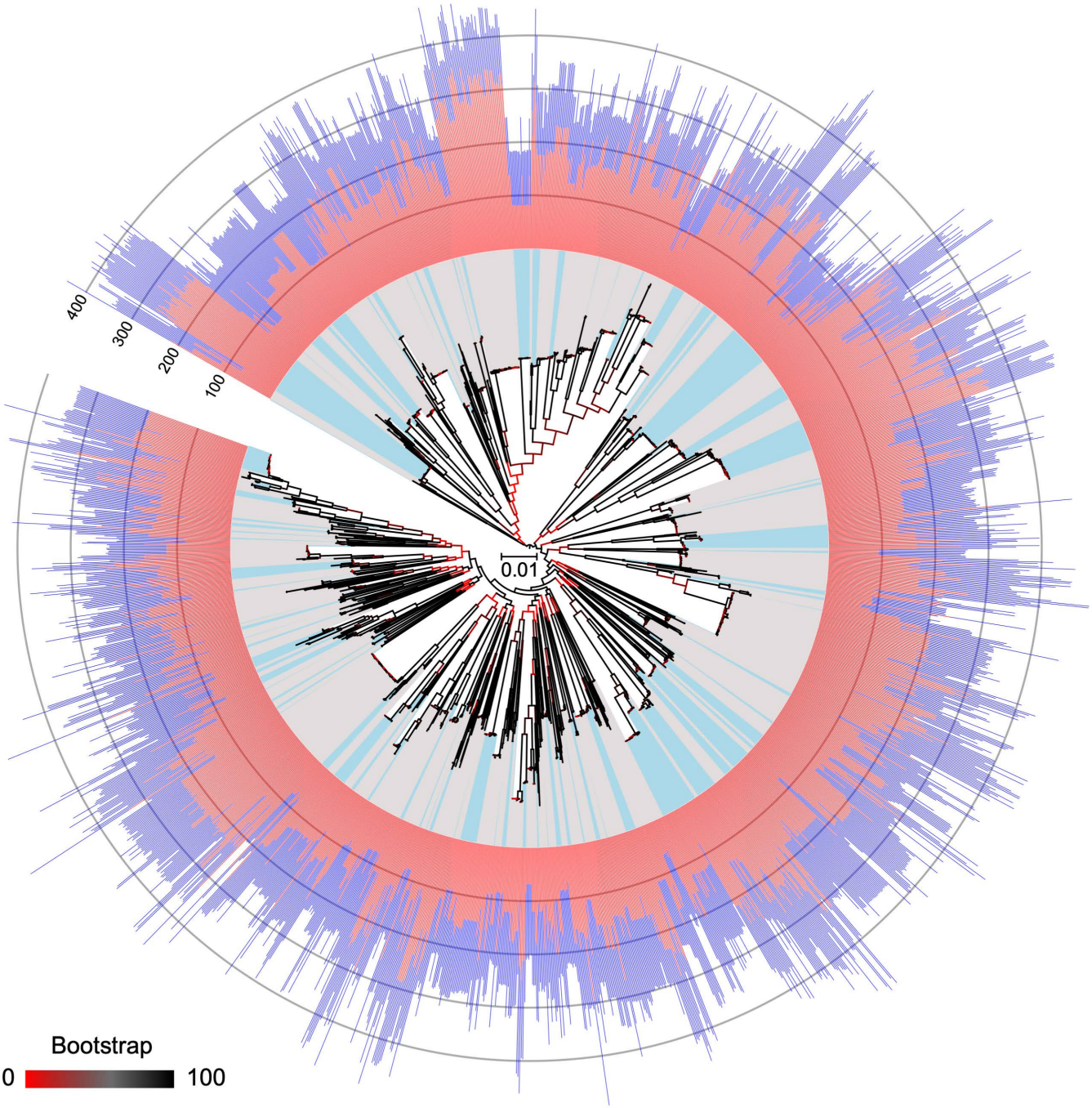
Supplementary Figure 7. Pairwise genetic (hamming) distance of the non-recombinogenic GAS core genome (416 genes) based on isolates being of the same same; (a) *emm* type; (b) *emm* sub-type; (c) MLST; (d) M-protein cluster and (e) core genome PopPUNK[17] phylogroups. Only groups which contain multiple isolates are represented on the X-axis. For each group, one reference was chosen based on having the mimimum median SNP distance from all other samples in the group. Each dot indicates the genetic distance (number of nucleotide SNPs, Y-axis) of samples from this reference, blue (same phylogroup [evolutionary lineage] as the reference), red (different phylogroup to the reference).



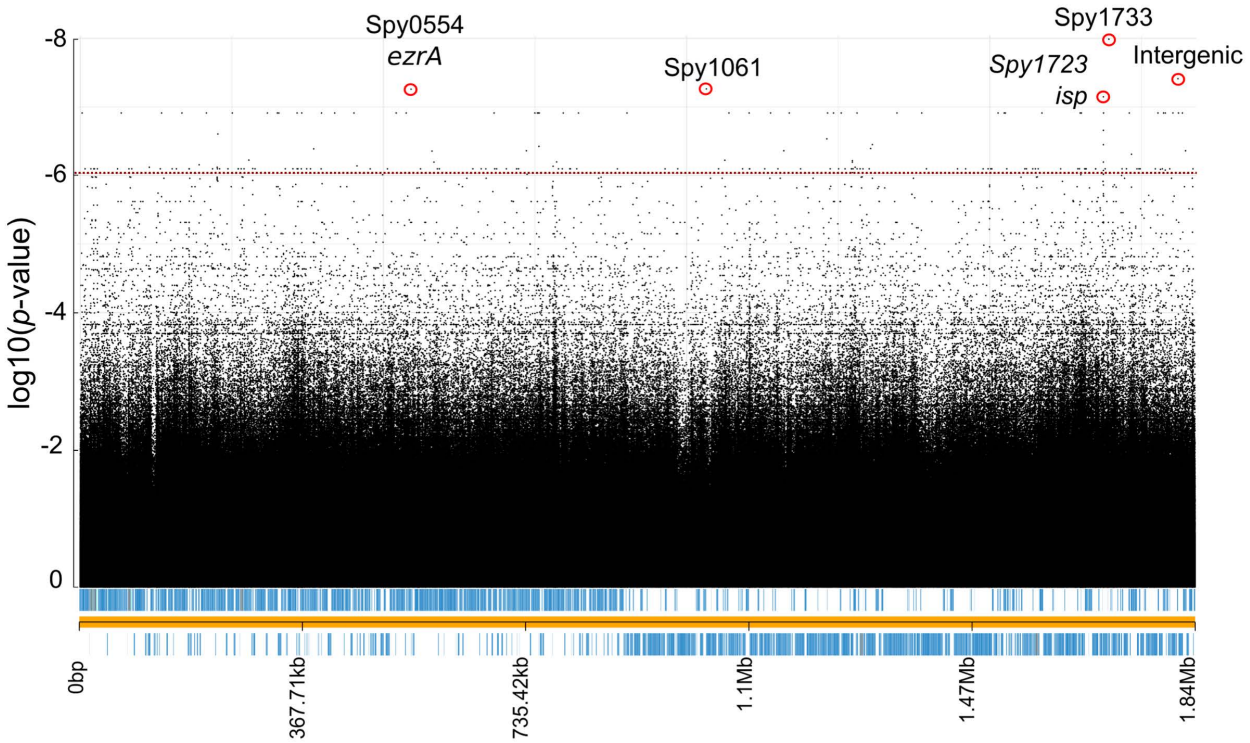
Supplementary Figure 8. Box and whisker plot showing the ratio of recombination derived mutation versus vertically inherited mutation (r/m) for the 36 most sampled phylogroups. Overall the mean r/m across these phylogroups was 4.95 (median of 3.12), and notably, is significantly greater than 1 (p-value = 7×10^{-7}), using a one-sample Wilcoxon test.



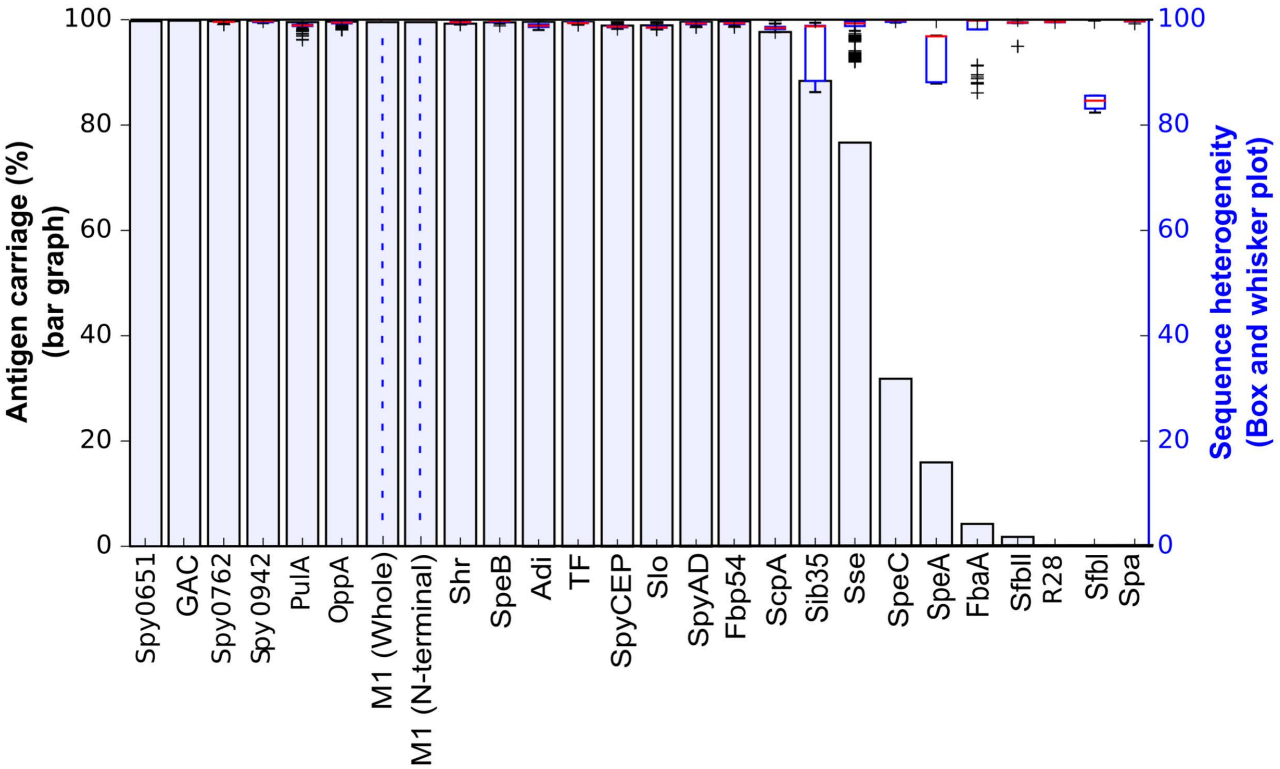
Supplementary Figure 9. Size distribution of intra-phylogroup recombination lengths. Lengths of putative recombination blocks in a subset of the most highly sampled PopPUNK[17] phylogroups (n = 36 [total of 1,062 genomes]). Recombination blocks were defined using the sliding window approach of Gubbins[19] based on intra-phylogroup mapping. Cumulative frequency of all recombination blocks within the 36 phylogroups including putative mobile genetic elements (MGE) (**a**) or excluding putative MGEs (**b**) within a 5000 base pair (bp) range. (**c**) Distribution of homologous recombination blocks (excluding MGE) within each of the 36 phylogroups (uniquely coloured) plotted from a 500 base pair sliding window (up to 20,000 bp). Collectively, evolution of GAS phylogroups is linked to high rates of small (<5000 bp) homologous recombination blocks.



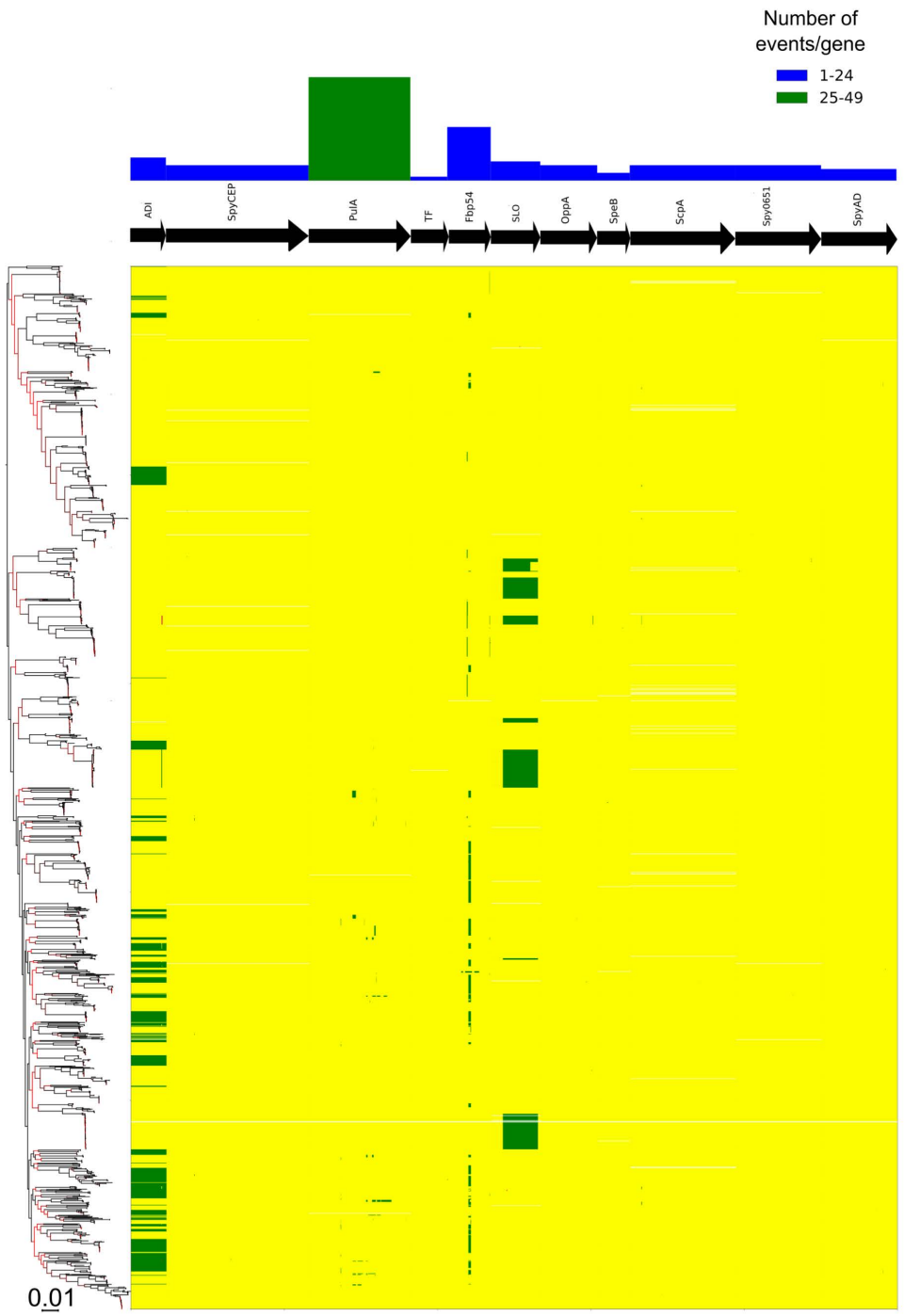
Supplementary Figure 10. Variation in the size and prophage content of the GAS accessory genome. Counts of accessory genes per genome are overlaid against the maximum-likelihood phylogenetic tree of core global GAS genome (416 genes) as displayed in Fig. 1. Branch colours indicate bootstrap support according to the legend. Distinct genetic lineages (n = 299 phylogroups) are highlight in alternating colours (blue and grey) from the tips of the tree. Red bars relate to the total count of phage-related genes based on PHASTER[69] analysis of the draft genome assemblies and blue bars relate to 'other' genes. Accessory gene scale refers to the number of genes (in 100 gene increments).



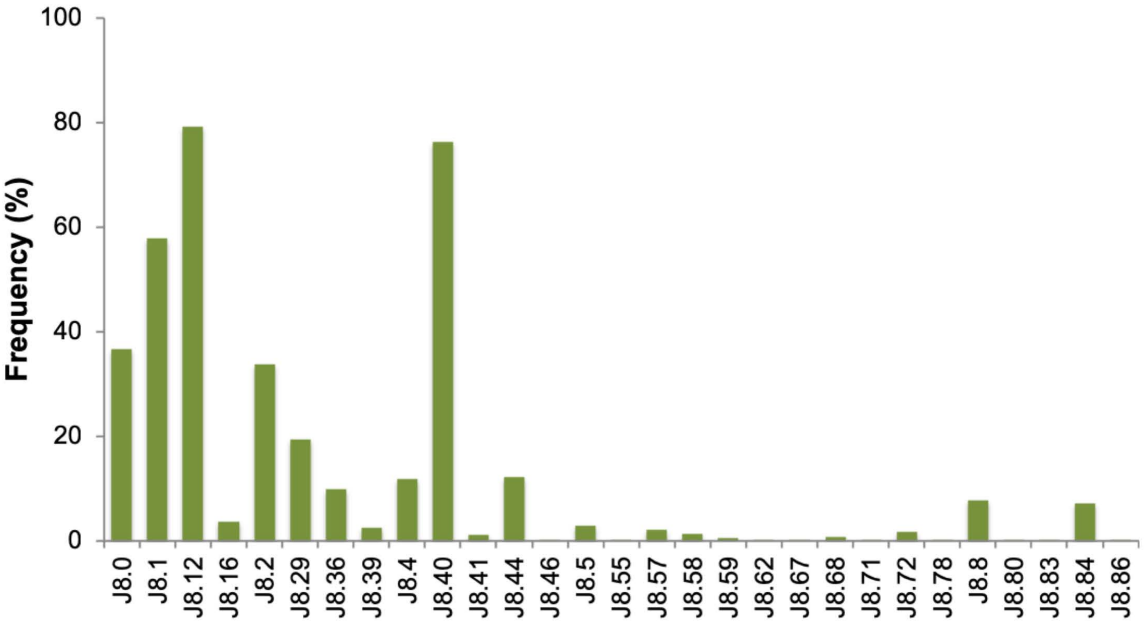
Supplementary Figure 11. Manhattan plot of SNPs associated with invasive GAS infection. The significance (y-axis) of each SNPs association with severe infection against its relative position within the MGAS5005 genome (x-axis). The red line denotes a significance cutoff of $p < 9x10^{-7}$. Top five loci reaching significance (Supplementary Table 7) are annotated. Associations were investigated by pyseer[27] using k-mers with a minimum minor allele frequency of 1%.



Supplementary Figure 12. GAS vaccine antigen carriage (gene product) and sequence variation within the 2,083 genome database. Left vertical axis refers to the bar graph showing frequency of antigen carriage of 25 GAS vaccine antigens and the group A carbohydrate (GAC) operon (X-axis) within 2,083 GAS genomes and the right vertical axis refers to the box and whisker plot showing quartile range, median (red line) and minimum/maximum values of each antigen as inferred by BlastP (as per Fig. 2b). Generally, genes that are 'core' have less sequence heterogeneity than genes that are variably carried (accessory genes).



Supplementary Figure 13. Recombination analysis of 11 conserved GAS vaccine antigens within the context of 2,083 GAS genomes. Maximum-likelihood phylogenetic tree of core GAS genome based on 416 non-recombinogenic genes of the 2,083 genomes is shown on the left. Middle panel shows fastGEAR[16] outputs per gene, with colors representing gene lineages (as per Supplementary Fig. 1). The plot above shows the number of homologous recombination events detected per gene.



Supplementary Figure 14. Frequency of J8 alleles in the 2,083 GAS genome database.

Supplementary Table 1. Vaccine antigen candidates examined in this study^ and published status of vaccine development (shaded boxes).

ANTIGEN	Pre-clinical	Phase I	Phase II	Proof of concept	Phase III	Reference
M protein: N-terminal peptide (30-valent)						Dale et al 2011[30]
M protein: C-terminal peptide (J8)						Batzloff et al 2003[31]
M1 protein: (whole protein)						Fox et al 1973[79]
M protein: C-terminal peptide (StreptInCor)						Guilherme et al 2006[32]
Trigger factor (TF)#						Henningham et al 2012[10]
Group A carbohydrate (GAC)						Sabharwal et al 2006[8]
C5a peptidase (ScpA)+#						Ji et al 1997[51]
Fibronectin-binding protein A (FbaA)						Ma et al 2009[80]
Fibronectin-binding protein 54 (Fbp54)						Kawabata et al 2001[81]
Streptococcal fibronectin binding protein I (SfbI)						Guzman et al 1999[82]
Serum opacity factor (SfbII/SOF)						Courtney et al 2003[83]
Streptococcal pyrogenic exotoxin A (SpeA)						Ulrich et al 2008[84]
Streptococcal pyrogenic exotoxin C (SpeC)						McCormick et al 2000[85]
Cysteine protease (SpeB)						Kapur et al 1994[86]
Serine protease (SpyCEP)*#						Zingaretti C et al 2010[87]
Serine protease (SpyCEP): S2 peptide						Pandey et al 2016[33]
Adhesion and division protein (SpyAD)+*						Bensi et al 2012[50]
Streptolysin O (SLO)*#						Bensi et al 2012[50]
Serine esterase (Sse)						Liu et al 2007[88]
Arginine deiminase (ADI)#						Henningham et al 2012[10]

Rib-like cell wall protein (R28)						Stalhammar-Carlemalm et al 1999[89]
Streptococcal hemoprotein receptor (Shr)						Huang et al 2011[90]
Streptococcal immunoglobulin-binding protein 35 (Sib35)						Okamoto et al 2005[91]
Streptococcal protective antigen (Spa)						Dale et al 1999[92]
Oligopeptide-binding protein (OppA)+						Reglinski et al 2016[93]
Putative pullulanase (PulA)+						Reglinski et al 2016[93]
Nucleoside-binding protein (Spy0942)+						Reglinski et al 2016[93]
Hypothetical membrane associated protein (Spy0762) +						Reglinski et al 2016[93]
Cell surface protein (Spy0651) +						Reglinski et al 2016[93]

Footnotes:
^All query sequences were based on the M1 GAS strain MGAS5005 as a query reference sequence (if present). Otherwise, query sequences from original published reference were used.
*Components of the Novartis (GSK) combination vaccine (Bensi et al 2012)[50]
+Components of the Spy7 combination vaccine (Reglinski et al., 2016)[93]
#Components of the Combo#5 vaccine (Rivera-Hernandez et al., 2016)[94]

Supplementary Table 2. GAS strains used in this study.

See separate Excel file

Supplementary Table 3. List of 890 core GAS genes identified as having recombinogenic signatures as defined by fastGEAR[16].

See separate Excel file

Supplementary Table 4. List of 416 "non-recombinogenic" core GAS genes and markers of selection pressue (ratio of non-synonymous [d_N] to synonymous [d_S] codon subsititutions [d_N/d_S]).

See separate Excel file

Supplementary Table 5: Strain and genome characteristics of 30 new globally sampled GAS reference genomes.

Strain ID	Country of isolation	Site	*emm*-subtype	Other *emm*-subtype	M-cluster	MLST	genome size (bp)	CDS (no.)	plasmid size (bp)	Prophage (no.)	GenBank Accession
GAS13475	New Zealand	Throat	197.0	-	AC2	998	1797172	1800		3	CP035455
NS178	Australia	Skin	54.1	166.2	D1	302	1742565	1708		1	CP035454
20123V1I1	Fiji	Blood	100.0	167.0	D2	119	1839531	1864		3	CP035453
31010V3S1	Fiji	Skin	123.0	205.0	D3	325	1768816	1708		10	CP035452
NS5694	Australia	Skin	230.0	-	D4	205	1826832	1813		4	CP035451
31165V2S1	Fiji	Skin	93.4	174.1, 156.0	D4	814	1701466	1642		10	CP035450
NS5958	Australia	Skin	56.0	205.0	D4	115	1825427	1833		3	CP035449
31041V2S1	Fiji	Skin	70.0	174.1	D4	10	1826467	1818		3	CP035448
K23890	Kenya	Soft Tissue	97.1	-	D5	283	1812090	1774		1	CP035447
Bra006	Brazil	Throat	68.2	-	E2	989	1747924	1691		1	CP035446
30109V1T1	Fiji	Throat	92.0	-	E2	1026	1758778	1718	3453	1	CP035445
14GA0958	New Zealand	Blood	90.5	-	E2	184	1764969	1703		8	CP035444
NS365	Australia	Blood	58.0	236.1	E3	176	1888806	1902		4	CP035443
31132V1S1	Fiji	Skin	25.0	159.0	E3	1032	1835714	1832		2	CP035442
A1268	India	Blood	1.0	-	E3	28	1834762	1841		3	CP035441
NS7124	Australia	Throat	124.0	-	E4	199	1790668	1759		1	CP035440
NS5128	Australia	Throat	77.0	149.2	E4	588	1806314	1782		1	CP035439
A995	India	Skin	22.8	-	E4	360	1950616	1960	3626	4	CP035438
GAS02198	New Zealand	Throat	78.3	-	E1	1000	1806521	1767		1	CP035437
31143V3S1	Fiji	Skin	89.14	236.2	E4	380	1806344	1797	3043	2	CP035436
Bra010	Brazil	Throat	64.3	205.1	E5	1008	1779766	1769		2	CP035435
NS20	Australia	Skin	75.1	170.0	E6	607	1887700	1870		2	CP035434

K3534	Kenya	Blood	65.0	-	E6	716	1789855	1734		1	CP035433
GAS11291	New Zealand	Throat	11.0	202.1	E6	547	1809631	1793		2	CP035432
31034V1S1	Fiji	Skin	105.0	-	M105	954	1800116	1757	2645	1	CP035431
NS4972	Australia	Skin	55.0	-	M55	100	1899479	1908		4	CP035430
NS7259	Australia	Throat	NA	138.0	NA	612	1788166	1788		3	CP035429
K17300	Kenya	Soft Tissue	stg866.1	166.1	NT	450	1816007	1786		1	CP035428
14GA0287	New Zealand	Throat	74.0	156.0	M74	120	1861037	1873		5	CP035427
31034V4S1	Fiji	Skin	57.0	166.1	M57	1025	1756622	1718	6485	1	CP035426

Supplementary Table 6: Frequency, size (length) and relative rates of recombination within 36 PopPUNK phylogroups.

See separate Excel file

Supplementary Table 7: Top 5 k-mers associated with invasiveness as determined by pyseer[27].

Gene/Locus Tag	Product	k-mer Coordinates	Log10(p-value)
M5005_Spy1733	Hypothetical protein	1696410..1696509	7.987
Intergenic		1810451..1810550	7.410
M5005_Spy1061	LacI family transcriptional regulator	1032656..1032755	7.261
M5005_Spy0554 (*ezrA*)	Cell division regulation	545543..545642	7.257
M5005_Spy1723 (*isp*)	immunogenic secreted protein	1687402..1687501	7.146

Supplementary Table 8: Position of amino acid variants within the Streptolysin O protein (SLO) and the consensus sequence of the SLO mature protein (as plotted in Fig 3a and 3c).

See separate Excel file

Supplementary Table 9: Position of amino acid variants within the C5a peptidase (ScpA) protein and the consensus sequence of the ScpA mature protein (as plotted in Fig 3a and 3d).

See separate Excel file

Supplementary Table 10: Mutation sensitivity analysis of amino acid variants within the mature Streptolysin O protein.

	Amino acid position within Streptolysin O protein				
hotspot[&]	HS1	HS2	HS3	HS4	HS5
aa position	172	182	324	450	470
major aa[$]	R(1242)	N(1305)	E(1204)	T(1843)	Q(1260)
minor aa[$]	M(841)	D(778)	D(879)	S(240)	R(823)
mutational sensitivity (minor aa)[#]	M=6	D=2	D=1	S=2	R=2

Footnotes:
aa (amino acid) .
[&] Diversity hotspot as determined by a minor amino acid frequency of >10% in the mature enzymatic protein (amino acids 103-501). Relative location is plotted in Figure 3c.
[$] Number in brackets refers to the total number of 2,083 GAS genomes carrying the respective amino acid.
[#] Determined using the SuSPect[37] platform (ranked between 1-9; 1 representing "very low" and 9 as "very high" mutational sensitivity). Sensitivity being a measure of likely functional consequence of the observed amino acid mutation.

Supplementary Table 11: Mutation sensitivity analysis of amino acid variants within the mature C5a peptidase protein.

Amino acid position within C5a peptidase protein												
hotspot	HS1	HS2	HS3	HS4	HS5	HS6	HS7	HS8	HS9	HS10	HS11	HS12
aa position	110	146	247	346	348	350	376	448	450	451	605	637
major aa[$]	Q(1745)	T(1634)	R(1921)	A(1367)	Q(1229)	D(1711)	M(1553)	D(1534)	P(1351)	Q(1381)	K(1709)	H(1948)
minor aa[$]	H(338)	A(360) S(89)	I(150) K(12)	D(622) E(93) V(1)	H(420) K(434)	A(346) G(19) N(7)	T(530)	E(549)	S(617) L(98) R(11) F(6)	K(692) P(10)	T(374)	L(131) Y(4)
mutational sensitivity (minor aa)[#]	H=2	A=2 S=2	I=1 K=3	D=3 D=2 V=2	H=2 K=2	A=2 G=3 N=3	T=3	E=1	S=5 L=5 R=6 F=6	K=2 P=3	T=1	L=1 Y=2

Amino acid position continued								
hotspot[&]	HS13	HS14	HS15	HS16	HS17	HS18	HS19	HS20
aa position	665	669	671	679	697	942	959	999
major aa[$]	V(1878)	A(1963)	R(1804)	Q(1952)	T(1111)	T(1671)	V(1775)	A(1679)
minor aa[$]	I(200) A(5)	V(120)	Q(279)	P(131)	K(972)	A(412)	I(308)	G(404)
mutational sensitivity (minor aa)[#]	I=1 A=3	V=2	Q=1	P=1	K=2	A=2	I=1	G=1

Footnotes:
aa (amino acid)
[&] Diversity hotspot as determined by a minor amino acid frequency of >10% in the mature enzymatic protein (amino acids 97-1032). Relative location is plotted in Figure 3d
[$] Number in brackets refers to the total number of 2,083 genomes carrying the respective amino acid

[#] Determined using the SuSPect[37] platform (ranked between 1-9; 1 representing “very low” and 9 as “very high” mutational sensitivity). Sensitivity being a measure of likely functional consequence of the observed amino acid mutation.

Supplementary Table 12: Theoretical global coverage of combination vaccines based on the genome database presented in this study.

		Theoretical Vaccine Coverage						
Vaccine	**Vaccine antigens**	Europe	Oceania	North America	South America	Asia	East Africa	**TOTAL**
		(n = 242)	*(n = 906)*	*(n = 474)*	*(n = 51)*	*(n = 79)*	*(n = 328)*	***(n = 2080)***
3-component (GSK)[50]	SLO$^+$	98%	>99%	>99%	>99%	97%	98%	**>99%**
	SpyCEP$^+$	99%	>99%	99%	>99%	99%	>99%	**>99%**
	SpyAD$^+$	>99%	>99%	>99%	>99%	>99%	>99%	**>99%**
	any antigen	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**
Spy7[93]	Spy0651$^+$	>99%	>99%	>99%	>99%	>99%	>99%	**>99%**
	Spy0762$^+$	>99%	>99%	>99%	>99%	>99%	>99%	**>99%**
	Spy0942$^+$	>99%	>99%	>99%	>99%	>99%	>99%	**>99%**
	PulA$^+$	>99%	>99%	>99%	>99%	99%	99%	**>99%**
	OppA$^+$	>99%	>99%	99%	>99%	99%	99%	**>99%**
	SpyAD$^+$	>99%	>99%	99%	>99%	99%	99%	**>99%**
	ScpA$^+$	90%	>99%	96%	>99%	99%	99%	**98%**
	any antigen	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**
Combo #5[94]	TF$^+$	>99%	>99%	>99%	98%	>99%	>99%	**>99%**
	ScpA$^+$	90%	>99%	96%	>99%	99%	99%	**98%**
	SpyCEP$^+$	99%	>99%	99%	>99%	99%	>99%	**>99%**
	ADI$^+$	>99%	>99%	99%	>99%	99%	99%	**>99%**
	SLO$^+$	98%	>99%	>99%	>99%	97%	98%	**>99%**
	any antigen	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**
StreptInCor[32]	B cell epitope$^@$	39%	25%	15%	12%	34%	14%	**22%**
	T cell epitope$^@$	8%	17%	3%	2%	4%	9%	**11%**
	common epitope$^@$	35%	16%	14%	10%	30%	11%	**18%**
	any epitope	**39%**	**26%**	**15%**	**12%**	**34%**	**17%**	**23%**
S2 - J8.0[33]	S2$^@$	>99%	>99%	>99%	>99%	>99%	>99%	**>99%**
	J8$^\$$	45%	41%	31%	31%	57%	25%	**37%**
	any epitope	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**	**>99%**
30-valent[30]	30 *emm* families$^\&$	**71%**	**33%**	**75%**	**53%**	**73%**	**28%**	**48%**
30-valent with Mrp[40]	30 *emm* families$^\&$	71%	33%	75%	53%	73%	28%	**48%**
	MrpI$^@$	8%	9%	12%	4%	5%	4%	**8%**
	MrpII$^@$	6%	13%	6%	18%	6%	3%	**9%**
	MrpIII$^@$	5%	4%	7%	4%	9%	6%	**5%**
	any antigen	**77%**	**51%**	**83%**	**59%**	**83%**	**33%**	**60%**

Footnotes:
[+] Defined by BlastN as 70% homology over 70% length of the nucleotide sequence.
[@] Peptide sequence carriage is defined by BlastP at 95% homology over 95% of query length.
[$] Defined as clustering at 90% of the J8 allelic database (encompasses J8.0, J8.57 or J8.59) by CD-HIT EST.
[&] Defined at the *emm* family level (irrespective of *emm* sub-type).

SUPPLEMENTARY REFERENCES

79. Fox, E.N., Waldman, R.H., Wittner, M.K., Mauceri, A.A. & Dorfman, A. Protective study with a group A streptococcal M protein vaccine. Infectivity challenge of human volunteers. *J Clin Invest* **52**, 1885-92 (1973).
80. Ma, C.Q. *et al.* Similar ability of FbaA with M protein to elicit protective immunity against group A *Streptococcus* challenge in mice. *Cell Mol Immunol* **6**, 73-7 (2009).
81. Kawabata, S. *et al.* Systemic and mucosal immunizations with fibronectin-binding protein FBP54 induce protective immune responses against *Streptococcus pyogenes* challenge in mice. *Infect Immun* **69**, 924-30 (2001).
82. Guzman, C.A., Talay, S.R., Molinari, G., Medina, E. & Chhatwal, G.S. Protective immune response against *Streptococcus pyogenes* in mice after intranasal vaccination with the fibronectin-binding protein SfbI. *J Infect Dis* **179**, 901-6 (1999).
83. Courtney, H.S., Hasty, D.L. & Dale, J.B. Serum opacity factor (SOF) of *Streptococcus pyogenes* evokes antibodies that opsonize homologous and heterologous SOF-positive serotypes of group A streptococci. *Infect Immun* **71**, 5097-103 (2003).
84. Ulrich, R.G. Vaccine based on a ubiquitous cysteinyl protease and streptococcal pyrogenic exotoxin A protects against *Streptococcus pyogenes* sepsis and toxic shock. *J Immune Based Ther Vaccines* **6**, 8 (2008).
85. McCormick, J.K. *et al.* Development of streptococcal pyrogenic exotoxin C vaccine toxoids that are protective in the rabbit model of toxic shock syndrome. *J Immunol* **165**, 2306-12 (2000).
86. Kapur, V. *et al.* Vaccination with streptococcal extracellular cysteine protease (interleukin-1 beta convertase) protects mice against challenge with heterologous group A streptococci. *Microb Pathog* **16**, 443-50 (1994).
87. Zingaretti, C. *et al. Streptococcus pyogenes* SpyCEP: a chemokine-inactivating protease with unique structural and biochemical features. *FASEB J* **24**, 2839-48 (2010).
88. Liu, M., Zhu, H., Zhang, J. & Lei, B. Active and passive immunizations with the streptococcal esterase Sse protect mice against subcutaneous infection with group A streptococci. *Infect Immun* **75**, 3651-7 (2007).
89. Stalhammar-Carlemalm, M., Areschoug, T., Larsson, C. & Lindahl, G. The R28 protein of *Streptococcus pyogenes* is related to several group B streptococcal surface

proteins, confers protective immunity and promotes binding to human epithelial cells. *Mol Microbiol* **33**, 208-19 (1999).

90. Huang, Y.S., Fisher, M., Nasrawi, Z. & Eichenbaum, Z. Defense from the Group A *Streptococcus* by active and passive vaccination with the streptococcal hemoprotein receptor. *J Infect Dis* **203**, 1595-601 (2011).
91. Okamoto, S., Tamura, Y., Terao, Y., Hamada, S. & Kawabata, S. Systemic immunization with streptococcal immunoglobulin-binding protein Sib 35 induces protective immunity against group A *Streptococcus* challenge in mice. *Vaccine* **23**, 4852-9 (2005).
92. Dale, J.B., Chiang, E.Y., Liu, S., Courtney, H.S. & Hasty, D.L. New protective antigen of group A streptococci. *J Clin Invest* **103**, 1261-8 (1999).
93. Reglinski, M., Lynskey, N.N., Choi, Y.J., Edwards, R.J. & Sriskandan, S. Development of a multicomponent vaccine for *Streptococcus pyogenes* based on the antigenic targets of IVIG. *J Infect* **72**, 450-9 (2016).
94. Rivera-Hernandez, T. *et al.* Differing efficacies of lead group A streptococcal vaccine candidates and full-length M protein in cutaneous and invasive disease models. *MBio* **7**(2016).
95. Bowen, A.C. *et al.* Whole genome sequencing reveals extensive community-level transmission of group A *Streptococcus* in remote communities. *Epidemiol Infect* **144**, 1991-8 (2016).
96. Bao, Y.J. *et al.* Genomic characterization of a pattern D *Streptococcus pyogenes emm*53 isolate reveals a genetic rationale for invasive skin tropicity. *J Bacteriol* **198**, 1712-24 (2016).
97. Athey, T.B. *et al.* High incidence of invasive group A *Streptococcus* disease caused by strains of uncommon *emm* types in Thunder Bay, Ontario, Canada. *J Clin Microbiol* **54**, 83-92 (2016).
98. Fiebig, A. *et al.* Comparative genomics of *Streptococcus pyogenes* M1 isolates differing in virulence and propensity to cause systemic infection in mice. *Int J Med Microbiol* **305**, 532-43 (2015).
99. Teatero, S. *et al.* Canada-Wide epidemic of *emm*74 group A *Streptococcus* invasive disease. *Open Forum Infect Dis* **5**, ofy085 (2018).
100. Fittipaldi, N. *et al.* Full-genome dissection of an epidemic of severe invasive disease caused by a hypervirulent, recently emerged clone of group A *Streptococcus*. *Am J Pathol* **180**, 1522-34 (2012).

101. Soriano, N. *et al.* Complete genome sequence of *Streptococcus pyogenes* M/*emm*44 strain STAB901, isolated in a clonal outbreak in French Brittany. *Genome Announc* **2**(2014).
102. Soriano, N. *et al.* Closed genome sequence of noninvasive *Streptococcus pyogenes* M/*emm*3 strain STAB902. *Genome Announc* **2**(2014).
103. Soriano, N. *et al.* Full-Length genome sequence of type M/*emm*83 group A *Streptococcus pyogenes* strain STAB1101, isolated from clustered cases in Brittany. *Genome Announc* **3**(2015).
104. Meygret, A. *et al.* Genome sequence of the uncommon *Streptococcus pyogenes* M/*emm*66 strain STAB13021, isolated from clonal clustered cases in French Brittany. *Genome Announc* **4**(2016).
105. Longo, M. *et al.* Complete genome sequence of *Streptococcus pyogenes emm*28 clinical isolate M28PF1, responsible for a puerperal fever. *Genome Announc* **3**(2015).
106. de Andrade Barboza, S. *et al.* Complete genome sequence of noninvasive *Streptococcus pyogenes* M/*emm*28 Strain STAB10015, isolated from a child with perianal dermatitis in French Brittany. *Genome Announc* **3**(2015).
107. Rochefort, A. *et al.* Full Sequencing and genomic analysis of three *emm*75 group A *Streptococcus* strains recovered in the course of an epidemiological shift in French Brittany. *Genome Announc* **5**(2017).
108. Ben Zakour, N.L. *et al.* Transfer of scarlet fever-associated elements into the group A *Streptococcus* M1T1 clone. *Sci Rep* **5**, 15877 (2015).
109. Tse, H. *et al.* Molecular characterization of the 2011 Hong Kong scarlet fever outbreak. *J Infect Dis* **206**, 341-51 (2012).
110. Sagar, V. *et al.* Variability in the distribution of genes encoding virulence factors and putative extracellular proteins of *Streptococcus pyogenes* in India, a region with high streptococcal disease burden, and implication for development of a regional multisubunit vaccine. *Clin Vaccine Immunol* **19**, 1818-25 (2012).
111. Haggar, A. *et al.* Clinical and microbiologic characteristics of invasive *Streptococcus pyogenes* infections in north and south India. *J Clin Microbiol* **50**, 1626-31 (2012).
112. Hertzog, B.B. *et al.* A Sub-population of group A *Streptococcus* elicits a population-wide production of bacteriocins to establish dominance in the host. *Cell Host Microbe* **23**, 312-323 e6 (2018).
113. Beres, S.B. *et al.* Transcriptome remodeling contributes to epidemic disease caused by the human pathogen *Streptococcus pyogenes*. *MBio* **7**(2016).

114. Nakagawa, I. *et al.* Genome sequence of an M3 strain of *Streptococcus pyogenes* reveals a large-scale genomic rearrangement in invasive strains and new insights into phage evolution. *Genome Res* **13**, 1042-55 (2003).
115. Miyoshi-Akiyama, T., Watanabe, S. & Kirikae, T. Complete genome sequence of *Streptococcus pyogenes* M1 476, isolated from a patient with streptococcal toxic shock syndrome. *J Bacteriol* **194**, 5466 (2012).
116. Yoshida, H. *et al.* Comparative genomics of the mucoid and nonmucoid strains of *Streptococcus pyogenes*, isolated from the same patient with streptococcal meningitis. *Genome Announc* **3**(2015).
117. Watanabe, S. *et al.* Complete genome sequence of *Streptococcus pyogenes* strain JMUB1235 isolated from an acute phlegmonous gastritis patient. *Genome Announc* **4**(2016).
118. Minogue, T.D. *et al.* Complete genome assembly of *Streptococcus pyogenes* ATCC 19615, a group A beta-hemolytic reference strain. *Genome Announc* **2**(2014).
119. Port, G.C., Paluscio, E. & Caparon, M.G. Complete genome sequence of *emm* type 14 *Streptococcus pyogenes* strain HSC5. *Genome Announc* **1**(2013).
120. Port, G.C., Paluscio, E. & Caparon, M.G. Complete genome sequences of *emm*6 *Streptococcus pyogenes* JRS4 and parental strain D471. *Genome Announc* **3**(2015).
121. Bao, Y. *et al.* Unique genomic arrangements in an invasive serotype M23 strain of *Streptococcus pyogenes* identify genes that induce hypervirulence. *J Bacteriol* **196**, 4089-102 (2014).
122. McShan, W.M. *et al.* Genome sequence of a nephritogenic and highly transformable M49 strain of *Streptococcus pyogenes*. *J Bacteriol* **190**, 7773-85 (2008).
123. Suvorova, M.A. *et al.* Complete genome sequences of *emm*111 Type *Streptococcus pyogenes* strain GUR, with antitumor activity, and its derivative strain GURSA1 with an inactivated *emm* gene. *Genome Announc* **5**(2017).
124. Zheng, P.X. *et al.* Complete genome sequence of *emm*1 *Streptococcus pyogenes* A20, a strain with an intact two-component system, CovRS, isolated from a patient with necrotizing fasciitis. *Genome Announc* **1**(2013).
125. Beres, S.B. & Musser, J.M. Contribution of exogenous genetic elements to the group A *Streptococcus* metagenome. *PLoS One* **2**, e800 (2007).
126. Bessen, D.E. *et al.* Whole-genome association study on tissue tropism phenotypes in group A *Streptococcus*. *J Bacteriol* **193**, 6651-63 (2011).

127. Beres, S.B. *et al.* Genome sequence of a serotype M3 strain of group A *Streptococcus*: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proc Natl Acad Sci U S A* **99**, 10078-83 (2002).
128. Sichtig, H. *et al.* FDA-ARGOS: A public quality-controlled genome database resource for infectious disease sequencing diagnostics and regulatory science research. *bioRxiv*, 482059 (2018).
129. Ferretti, J.J. *et al.* Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. *Proc Natl Acad Sci U S A* **98**, 4658-63 (2001).
130. Jacob, K.M., Spilker, T., LiPuma, J.J., Dawid, S.R. & Watson, M.E., Jr. Complete genome sequence of *emm*28 type *Streptococcus pyogenes* MEW123, a atreptomycin-resistant derivative of a clinical throat isolate suitable for investigation of pathogenesis. *Genome Announc* **4**(2016).
131. Jacob, K.M., Spilker, T., LiPuma, J.J., Dawid, S.R. & Watson, M.E., Jr. Complete genome sequence of *emm*4 *Streptococcus pyogenes* MEW427, a throat isolate from a child meeting clinical criteria for pediatric autoimmune neuropsychiatric disorders associated with *Streptococcus* (PANDAS). *Genome Announc* **4**(2016).
132. Smoot, J.C. *et al.* Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. *Proc Natl Acad Sci U S A* **99**, 4668-73 (2002).
133. Long, S.W., Kachroo, P., Musser, J.M. & Olsen, R.J. Whole-genome sequencing of a human clinical isolate of *emm*28 *Streptococcus pyogenes* causing necrotizing fasciitis acquired contemporaneously with hurricane Harvey. *Genome Announc* **5**(2017).
134. Green, N.M. *et al.* Genome sequence of a serotype M28 strain of group A *Streptococcus*: potential new insights into puerperal sepsis and bacterial disease specificity. *J Infect Dis* **192**, 760-70 (2005).
135. Holden, M.T. *et al.* Complete genome of acute rheumatic fever-associated serotype M5 *Streptococcus pyogenes* strain manfredo. *J Bacteriol* **189**, 1473-7 (2007).
136. Banks, D.J. *et al.* Progress toward characterization of the group A *Streptococcus* metagenome: complete genome sequence of a macrolide-resistant serotype M6 strain. *J Infect Dis* **190**, 727-38 (2004).