

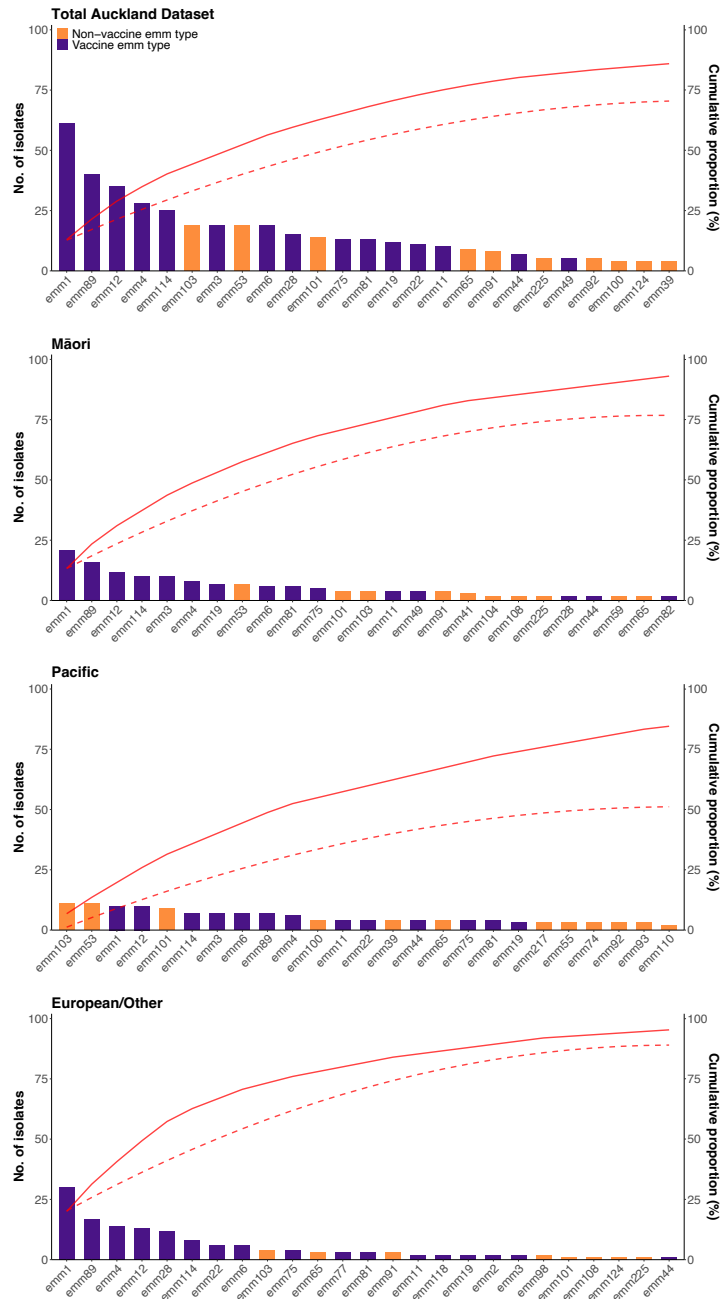
Supplementary information

Supplementary Table 1: Demographics of individuals with GAS isolates included in this study

Demographic	GAS isolate from throat		GAS isolate from skin		Total	
	n=357	%	n=108	%	N=465*	%
Age						
5-9 years	219	61.3	67	62.0	286	61.5
10-15 years	138	38.7	41	38.0	179	38.5
Ethnicity						
Māori	118	34.7	32	29.6	150	33.5
Pacific Peoples	108	31.8	48	44.4	156	34.8
NZ European/Other	114	33.5	28	25.9	142	31.7
Sex						
Male	214	59.9	63	58.3	277	59.6
Female	143	40.1	45	41.7	188	40.4
NZiDep - socio-economic deprivation						
0 (least deprived)	5	1.4	2	1.9	7	1.5
1	13	3.7	6	5.7	19	4.2
2	19	5.4	8	7.5	27	5.9
3	23	6.6	8	7.5	31	6.8
4	18	5.1	10	9.4	28	6.1
5	38	10.9	11	10.4	49	10.7
6	57	16.3	22	20.8	79	17.3
7	66	18.9	13	12.3	79	17.3
8 (most deprived)	111	31.7	26	24.5	137	30.0

* Missing data, n=4

Supplementary Figure 1: Distribution of the 25 most common *emm*-types in the total dataset (n = 469) and by prioritised ethnic groupings. The *emm*-types are plotted in descending order of frequency (shown as number of isolates) and coloured by whether they are included in the 30-valent StreptAnova vaccine. Secondary y-axis shows the cumulative total proportion of isolates (red line) and the cumulative proportion of vaccine *emm*-type from the total isolates (red dashed line).



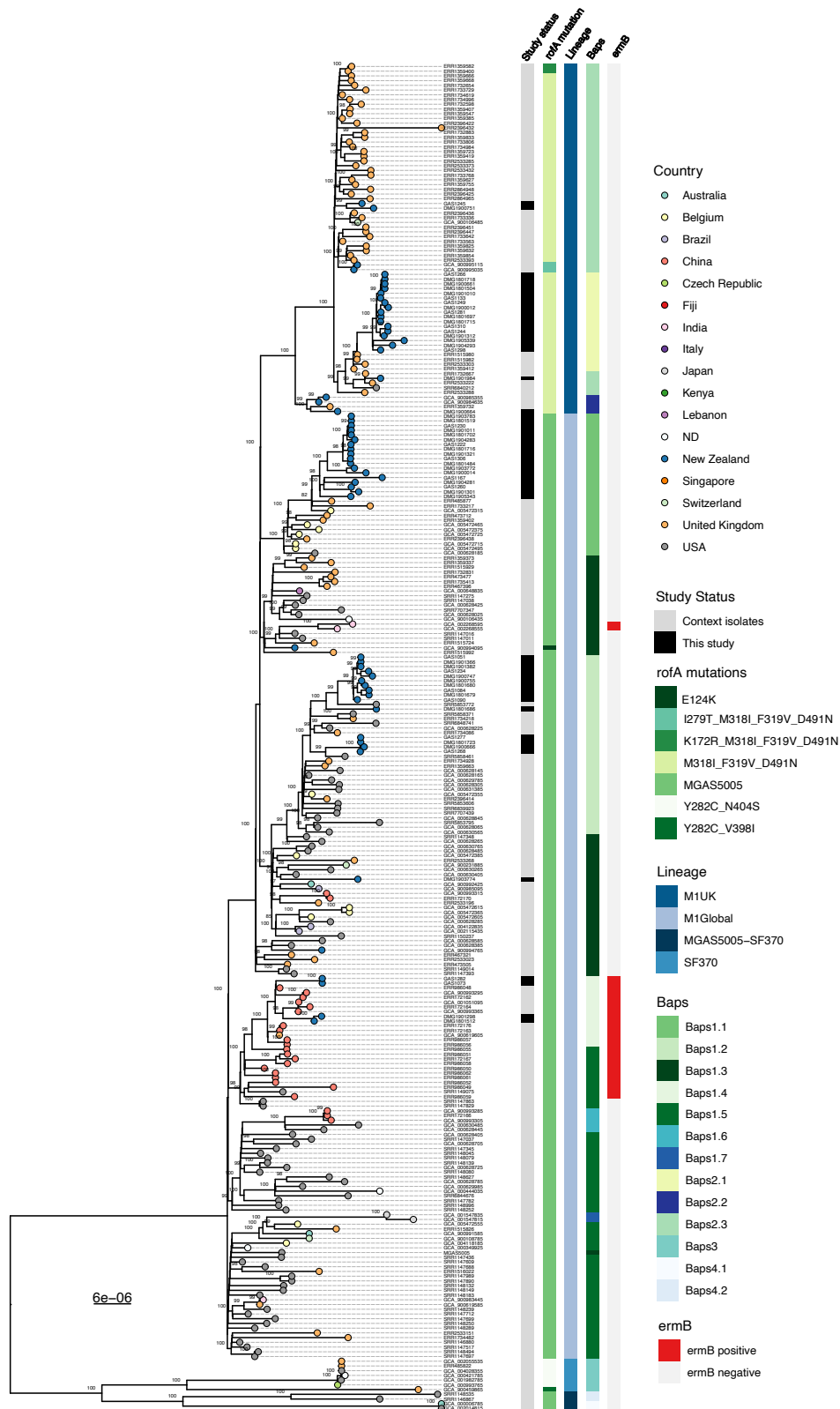
Supplementary Table 2: Contingency table for *emm*-type isolate count data and Chi-squared goodness-of-fit tests results

	No. isolates			Total isolates	<i>p</i> -value
	Māori	Other	Pacific Peoples		
<i>emm1</i>	21	29	11	61	0.003*
<i>emm100</i>	0	0	4	4	0.022*
<i>emm101</i>	4	1	9	14	0.041*
<i>emm103</i>	4	4	11	19	0.099
<i>emm104</i>	2	0	1	3	0.380
<i>emm108</i>	2	1	0	3	0.366
<i>emm11</i>	4	2	4	10	0.721
<i>emm110</i>	0	0	2	2	0.149
<i>emm114</i>	10	8	7	25	0.738
<i>emm116</i>	0	0	2	2	0.149
<i>emm118</i>	0	2	1	3	0.342
<i>emm12</i>	12	13	10	35	0.720
<i>emm124</i>	1	1	2	4	0.807
<i>emm15</i>	1	0	1	2	0.626
<i>emm183</i>	0	0	1	1	0.387
<i>emm19</i>	7	2	3	12	0.186
<i>emm2</i>	0	2	0	2	0.118
<i>emm217</i>	0	0	3	3	0.058
<i>emm22</i>	1	6	4	11	0.154
<i>emm225</i>	2	1	2	5	0.849
<i>emm28</i>	2	12	1	15	<0.001*
<i>emm3</i>	10	2	7	19	0.091
<i>emm39</i>	0	0	4	4	0.022*
<i>emm4</i>	8	14	6	28	0.107
<i>emm41</i>	3	0	0	3	0.052
<i>emm42</i>	0	0	1	1	0.387
<i>emm44</i>	2	1	4	7	0.411
<i>emm49</i>	4	0	1	5	0.078
<i>emm53</i>	7	1	11	19	0.026*
<i>emm55</i>	0	1	3	4	0.191
<i>emm56</i>	1	1	0	2	0.590
<i>emm58</i>	1	1	1	3	0.998
<i>emm59</i>	2	0	2	4	0.391
<i>emm6</i>	6	6	7	19	0.973
<i>emm65</i>	2	3	4	9	0.736
<i>emm71</i>	0	1	2	3	0.384
<i>emm74</i>	1	0	3	4	0.195
<i>emm75</i>	5	4	4	13	0.929
<i>emm76</i>	1	0	2	3	0.395
<i>emm77</i>	0	3	0	3	0.041*
<i>emm81</i>	6	3	4	13	0.613
<i>emm82</i>	2	0	2	4	0.391
<i>emm87</i>	1	1	0	2	0.590
<i>emm89</i>	15	17	7	40	0.073
<i>emm91</i>	4	3	1	8	0.399
<i>emm92</i>	1	1	3	5	0.486
<i>emm93</i>	0	0	3	3	0.058
<i>emm95</i>	1	0	1	2	0.626
<i>emm98</i>	1	2	1	4	0.740
Total Isolates	157	149	163	469	

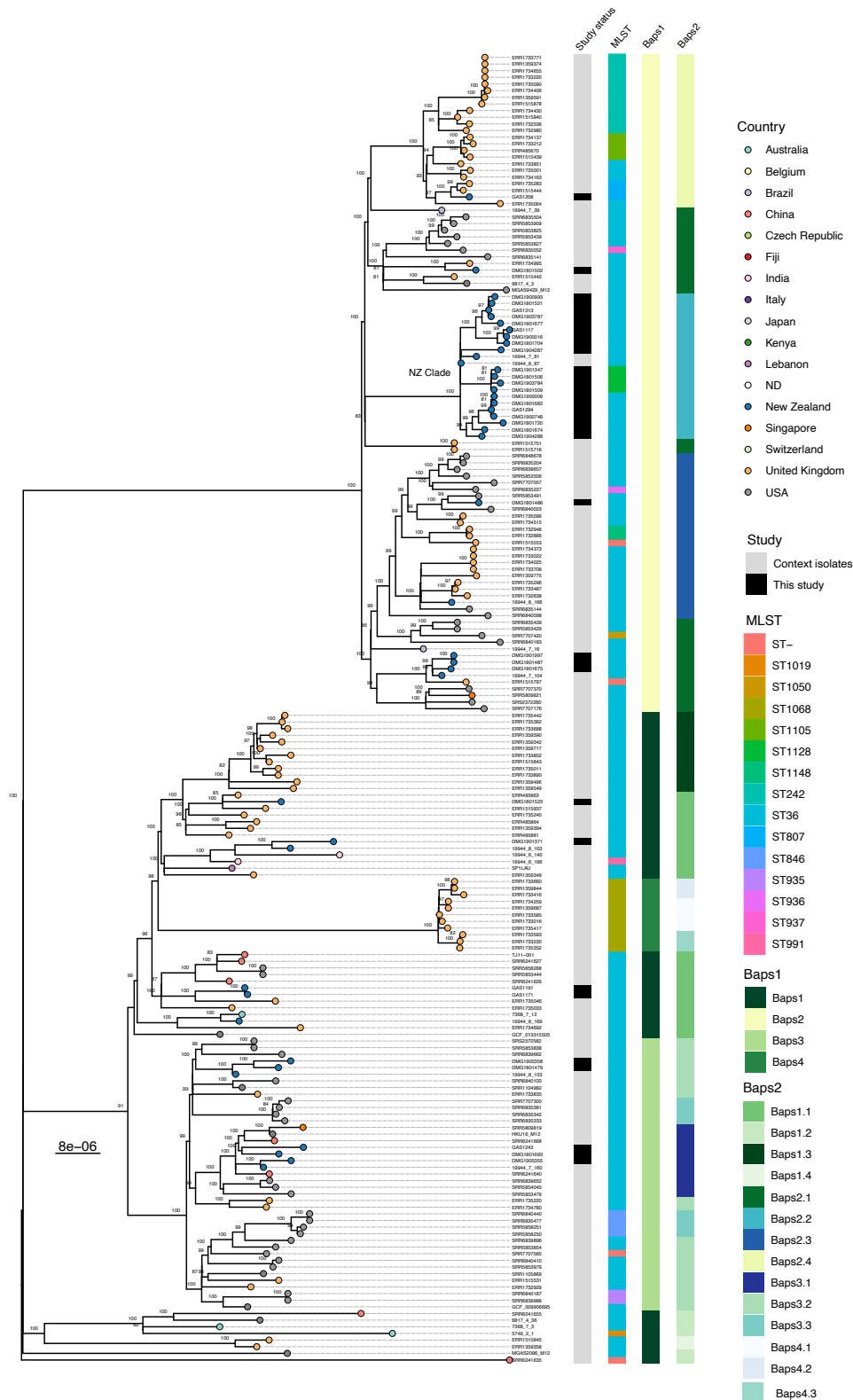
*Indicates a statistically significant values from the Chi-squared goodness-of-fit test. Statistically significant associations based on a low sample number (n<5) are indicated by italics.

Supplementary Table 3: Summary of datasets used for lineage-based phylogenetic analysis of the four most prevalent *emm*-types in the study population.

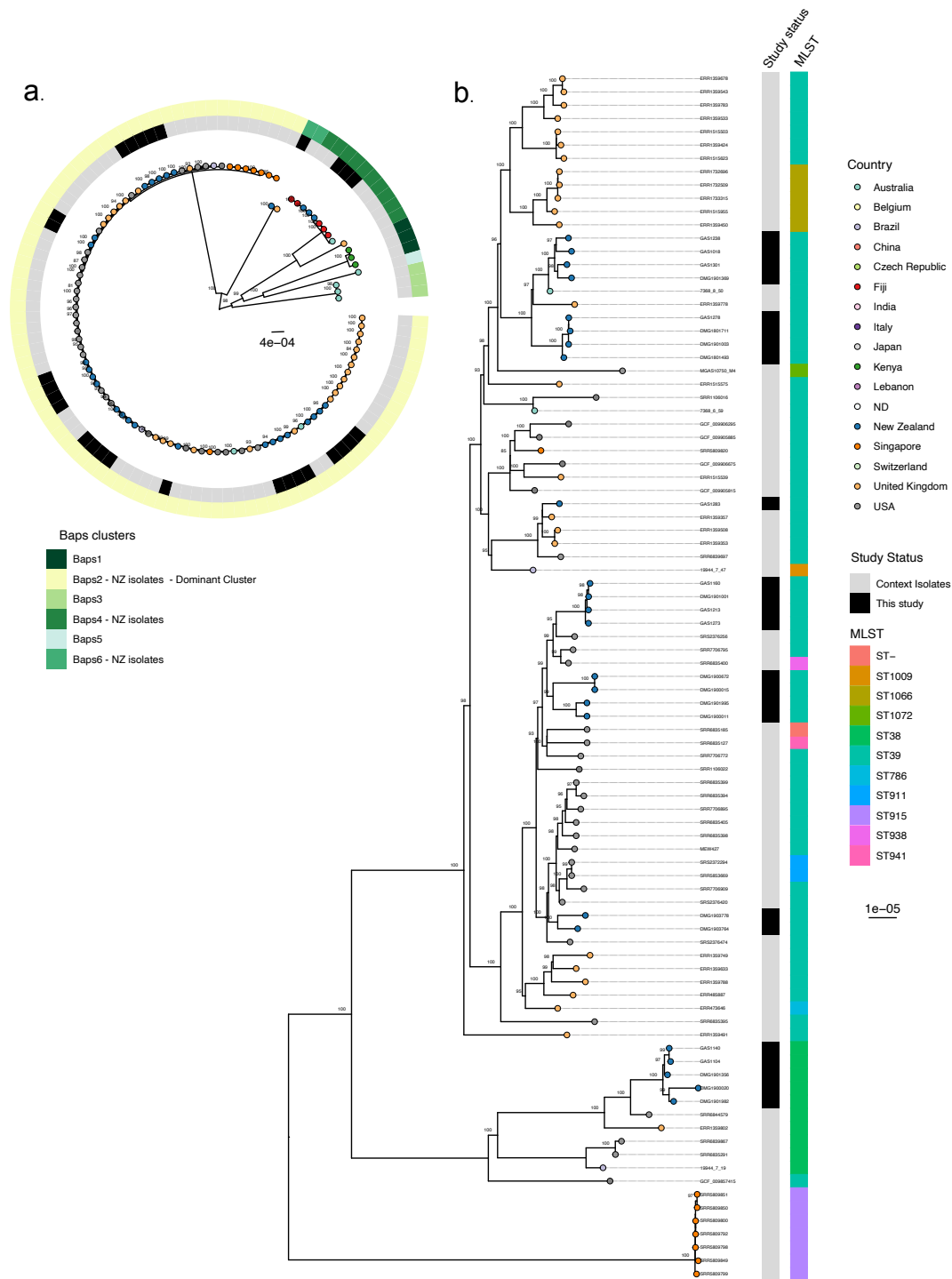
	Isolates (this study)	Context isolates (public)	Context isolates (NZ)	Number of Countries	MLST (n)	Subtypes (n)	Reference strain
<i>emm1</i>	59	226	6	13	4	14	MGAS5005
<i>emm89</i>	39	97	8	11	13	8	MGAS23530
<i>emm12</i>	35	162	8	13	15	24	HKU16
<i>emm4</i>	28	82	0	10	21	4	MGAS10750



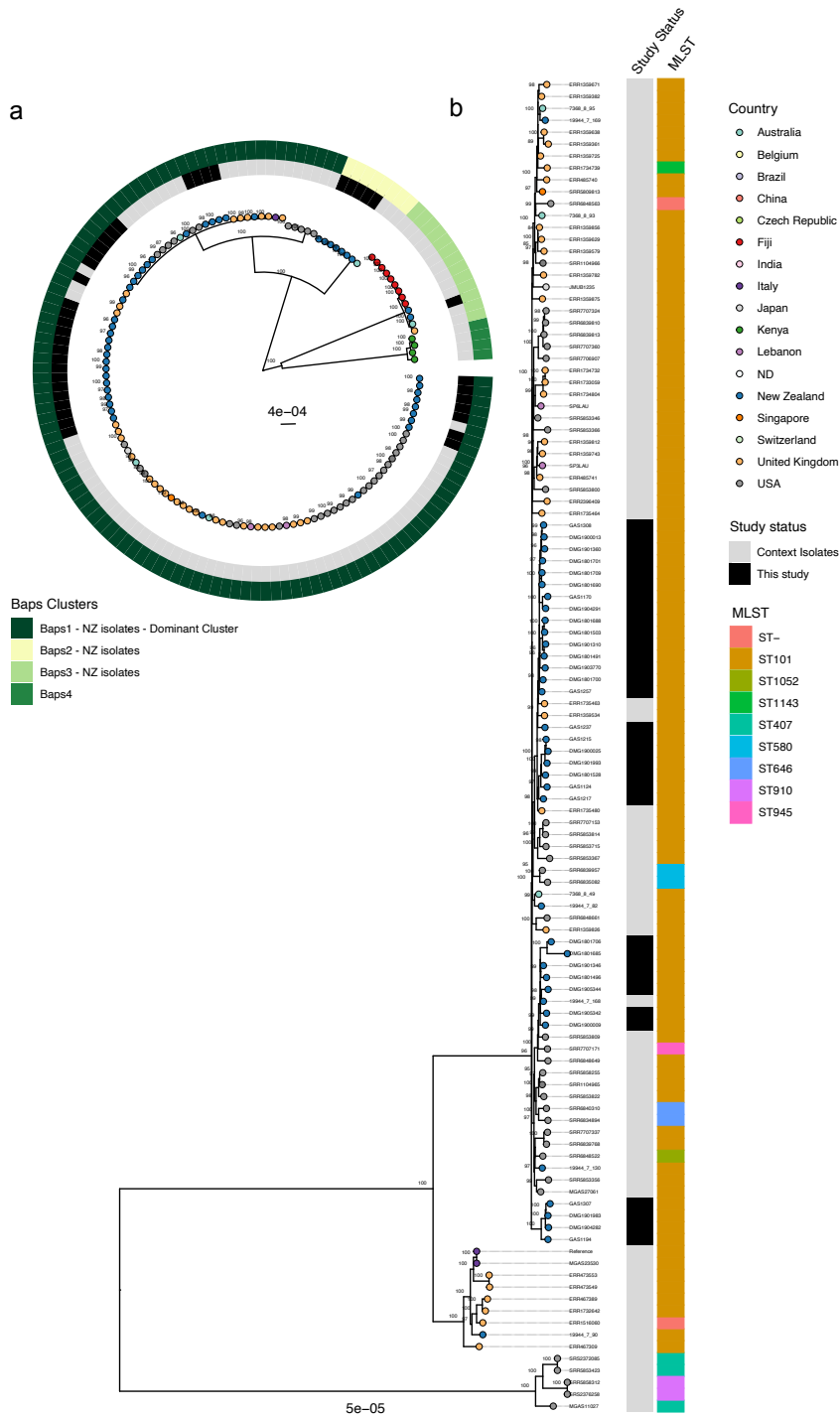
Supplementary Figure 2: Maximum likelihood phylogeny of 59 Auckland *emmI* genomes from this study with 226 global context isolates. The tree was inferred from 766 parsimony informative sites from a core genome alignment against the MGAS5005 reference genome. Annotation as follows; tip colour refers to country of origin for each genome “dark blue” = Aotearoa/New Zealand. Heatmaps include (a) Context “black” = this study, “grey” = context isolates, (b) *rofA* mutations in shades of green, (c) Baps grouping and Lineage in shades of blue, (d) baps2 groupings and (e) *ermB* gene presence “red” = positive, “grey” = negative.



Supplementary Figure 3: Maximum likelihood phylogeny of 35 Auckland *emm12* genomes from this study with 162 global context isolates. The tree was inferred from 1,239 parsimony informative sites from a core genome alignment against the HKU16 M12 reference genome. Phylogenetic tree is annotated as follows; tip colour refers to country of origin for each genome “dark blue” = Aotearoa/New Zealand. Heatmaps include (a) Context “black” = this study, “grey” = context isolates, (b) MLST typing, (c) Baps1 grouping, (d) Baps2 groupings.



Supplementary Figure 4. Analysis of the *emm4* genomes (A) Maximum likelihood phylogeny of 28 Auckland *emm4* genomes from this study with 82 global context isolates. The tree was inferred from 26,868 parsimony informative sites from a core genome alignment against the MGAS10750 M4 reference genome. Phylogenetic tree is annotated as follows; tip colour refers to country of origin for each genome “dark blue” = Aotearoa/New Zealand. Inner ring refers to study context “black” = this study, “grey” = context isolates, and outer ring refers to Baps clusters. **(B)** Maximum likelihood phylogeny of 24 Auckland *emm4* genomes in the dominant sequence cluster (Baps2) from this study with 67 global contexts. The tree was inferred from 1,227 parsimony informative sites from a core genome alignment against the MGAS10750 M4 reference genome. Phylogenetic tree is annotated as follows; tip colour refers to country of origin for each genome “dark blue” = Aotearoa/New Zealand. Heatmaps include (a) Study context “black” = this study, “grey” = context isolates, (b) MLST typing.



Supplementary Figure 5: Analysis of the *emm89* genomes. (A) Maximum likelihood phylogeny of 39 Auckland *emm89* genomes from this study with 97 global context isolates. The tree was inferred from 21,052 parsimony informative sites from a core genome alignment against the MGAS23530 reference genome. Phylogenetic tree is annotated as follows; tip colour refers to country of origin for each genome “dark blue” = Aotearoa/New Zealand. Inner ring refers to study context “black” = this study, “grey” = context isoaltes, and outer ring refers to Baps clusters. **(B)** Maximum likelihood phylogeny of 33 Auckland *emm4* genomes isolates of the dominant sequence cluster (Baps1) from this study with 78 global contexts. The tree was inferred from 2,106 parsimony informative sites from a core genome alignment against the MGAS23530 reference genome. Phylogenetic tree is annotated as follows; tip colour refers to country of origin for each genome “dark blue” = Aotearoa/New Zealand. Heatmaps include (a) Context “black” = this study, “grey” = context isolates, (b) MLST typing.

Supplementary Table 4: Summary of virulence gene prevalence within Auckland GAS isolates.

Class	Gene	No isolates positive* (%)	Unique variants (amino acid)
Streptodornase/DNAse	<i>spd1/mf2</i>	187 (40%)	5
	<i>spd2/mf1</i>	469 (100%)	13
	<i>spd3/mf3</i>	206 (44%)	8
	<i>spd4/mf4</i>	17 (4%)	1
	<i>sda</i>	94 (20%)	2
	<i>sdn</i>	69 (15%)	4
Superantigens	<i>speG</i>	349 (75%)	17
	<i>smeZ</i>	157 (34%)	3
	<i>speC</i>	157 (34%)	3
	<i>speJ</i>	157 (34%)	3
	<i>speA</i>	82 (17%)	6
	<i>ssa</i>	76 (16%)	1
	<i>speL</i>	75 (16%)	7
	<i>speK</i>	72 (15%)	6
	<i>speH</i>	71 (15%)	5
	<i>speI</i>	41 (9%)	1
	<i>speR</i>	21 (4%)	4
	<i>speM</i>	14 (3%)	4
	<i>speQ</i>	12 (3%)	3
	Capsule	<i>hasA</i>	469 (100%)
<i>hasB</i>		371 (80%)	41
<i>hasC</i>		289 (62%)	36
Regulators	<i>mga1</i>	136 (29%)	12
	<i>mga2</i>	315 (67%)	40
	<i>rofA</i>	328 (70%)	42
	<i>nra</i>	107 (22%)	11
Other	<i>sla</i>	97 (20%)	2
	<i>ska1</i>	170 (36%)	14
	<i>ska2</i>	297 (63%)	48

Supplementary Table 5: Summary of antimicrobial resistance gene prevalence within Auckland GAS isolates.

Class	Gene	No isolate positive (%)
Aminoglycoside	<i>aph(3')-IIIa</i>	1 (<1%)
Macrolide	<i>erm(A)</i>	3 (1%)
Macrolide	<i>erm(B)</i>	7 (1%)
Aminoglycoside	<i>mef(A)</i>	2 (<1%)
Aminoglycoside	<i>msr(D)</i>	2 (<1%)
Tetracycline	<i>tet(M)</i>	74 (16%)
Tetracycline	<i>tet(O)</i>	1 (<1%)

Supplementary Table 6: Theoretical coverage provided by the 30-valent StreptAnova vaccine. Values are presented assuming cross-opsonisation or no cross-opsonisation in the prioritised ethnic groups, VE – Vaccine *emm*-type, NVE – Non-vaccine *emm*-type, CO – cross opsonisation.

	Total		Māori		Pacific Peoples		Other	
	No. <i>emm</i> types (%)	No. isolates (%)	No. <i>emm</i> types (%)	No. isolates (%)	No. <i>emm</i> types (%)	No. isolates (%)	No. <i>emm</i> types (%)	No. isolates (%)
Total <i>emm</i> types and isolates	49 (100)	469 (100)	36 (100)	157 (100)	43 (100)	163 (100)	32 (100)	149 (100)
Assuming no cross-opsonisation								
VE in dataset	21 (42.9)	329 (70.2)	18 (50.0)	117 (74.7)	18 (41.9)	84 (51.2)	19 (59.4)	128 (86.0)
NVE in dataset	28 (57.1)	140 (29.8)	18 (50.0)	40 (25.3)	25 (58.1)	79 (48.8)	13 (40.6)	21 (14.0)
Assuming cross-opsonisation								
CO positive and VE in dataset	29 (59.2)	372 (79.4)	22 (61.1)	128 (81.6)	26 (60.5)	111 (67.9)	22 (68.8)	133 (89.3)
CO negative and NVE in dataset	20 (40.8)	97 (20.6)	14 (38.9)	29 (18.4)	17 (39.5)	52 (32.1)	10 (31.3)	16 (10.7)

Supplementary Table 7: Theoretical coverage provided by the TeeVax vaccine. Values are presented assuming cross-opsonisation or no cross-opsonisation and prioritised ethnic groups. VT – Vaccine *tee*-type, NVT – Non-vaccine *tee*-type, CR – cross reactivity only observed in two additional *tee*-types in this dataset *tee*18.2 and *tee*28.1

	Total		Māori		Pacific Peoples		Other	
	No. <i>tee</i> types (%)	No. isolates (%)	No. <i>tee</i> types (%)	No. isolates (%)	No. <i>tee</i> types (%)	No. isolates (%)	No. <i>tee</i> types (%)	No. isolates (%)
Total <i>tee</i> types and isolates	35 (100)	469 (100)	27 (100)	157 (100)	31 (100)	163 (100)	25 (100)	149 (100)
Assuming No Cross-reactivity								
VT in dataset	18 (51.4)	353 (75.2)	18 (66.6)	114 (72.6)	17 (54.8)	119 (73.0)	14 (56.0)	120 (80.5)
NVT in dataset	17 (48.6)	116 (24.8)	16 (33.4)	43 (27.4)	14 (45.2)	44 (27.0)	11 (44.0)	29 (19.5)
Assuming Cross-reactivity								
CR positive and VT in dataset	20 (57.2)	381 (81.2)	18 (66.6)	126 (80.2)	19 (61.2)	128 (78.5)	15 (60.0)	127 (85.2)
CR negative and NVT in dataset	15 (42.8)	88 (18.8)	9 (33.4)	31 (19.8)	12 (38.8)	35 (21.4)	10 (40.0)	22 (14.8)