

A Controlled Human Infection Model of Group A *Streptococcus* Pharyngitis: Which Strain and Why?

Joshua Osowicki ^{a,b,c}

Kristy I. Azzopardi ^a

Liam McIntyre ^d

Tania Rivera-Hernandez ^e

Cheryl-lynn Y. Ong ^e

Ciara Baker ^a

Christine M. Gillen ^e

Mark J. Walker ^e

Pierre R. Smeesters ^{a,b,f,g}

Mark R. Davies ^d

Andrew C. Steer ^{a,b,c}

a. Tropical Diseases, Murdoch Children's Research Institute, Melbourne, Victoria, Australia.

b. Department of Paediatrics, University of Melbourne, Victoria, Australia

c. Infectious Diseases Unit, Department of General Medicine, The Royal Children's Hospital Melbourne, Victoria, Australia.

d. Department of Microbiology and Immunology, Peter Doherty Institute for Infection and Immunity, The University of Melbourne, Victoria, Australia.

e. School of Chemistry and Molecular Biosciences and Australian Infectious Diseases Research Centre, The University of Queensland, St Lucia, Queensland, Australia.

f. Paediatric Department, Academic Children Hospital Queen Fabiola, Université Libre de Bruxelles, Brussels, Belgium

g. Molecular Bacteriology Laboratory, Université Libre de Bruxelles, Brussels, Belgium.

Running title: Group A *Streptococcus* Human Challenge Strain

Keywords: group A *Streptococcus*, *Streptococcus pyogenes*, controlled human infection, human challenge, vaccine development

Corresponding author:

Dr Joshua Osowicki

joshua.osowicki@rch.org.au

Tropical Diseases

Murdoch Children's Research Institute

50 Flemington Road

Parkville, 3052

Victoria, Australia

Alternate corresponding author:

Professor Andrew C Steer

andrew.steer@rch.org.au

Tropical Diseases

Murdoch Children's Research Institute

50 Flemington Road

Parkville, 3052

Victoria, Australia

54 **Abstract**

55

56 Background: Group A *Streptococcus* (GAS) is a major cause of global infection-related morbidity and
57 mortality. A modern controlled human infection model (CHIM) of GAS pharyngitis can accelerate vaccine
58 development and pathogenesis research. A robust rationale for strain selection is central to meeting ethical,
59 scientific, and regulatory requirements.

60

61 Methods: Multi-faceted characterisation studies were done to compare a preferred candidate *emm75* (M75)
62 GAS strain against three other strains: an alternative candidate *emm12* (M12) strain; an M1 strain used in
63 1970s pharyngitis CHIM studies (SS-496); and a representative (5448) of the globally-disseminated M1T1
64 clone. A range of approaches were used to explore strain growth, adherence, invasion, delivery
65 characteristics, short- and long-term viability, phylogeny, virulence factors, vaccine antigens, resistance to
66 killing by human neutrophils, and lethality in a murine invasive model.

67

68 Results: The strains grew reliably in a medium without animal-derived components, were consistently
69 transferred using a swab method simulating the CHIM protocol, remained viable at -80° Celsius, and carry
70 genes for most candidate vaccine antigens. Considering GAS molecular epidemiology, virulence factors,
71 *in vitro* assays, and results from the murine model, the contemporary strains show a spectrum of virulence
72 with M75 appearing least virulent and 5448 the most. The virulence profile of SS-496, used safely in
73 1970s CHIM studies, was similar to 5448 in the animal model and virulence gene carriage.

74

75 Conclusions: The results of this multi-faceted characterisation confirm the M75 strain as an appropriate
76 choice for initial deployment in the CHIM, with the aim of safely and successfully causing pharyngitis in
77 healthy adult volunteers.

78

79

80 **Importance:**

81 Group A *Streptococcus* (GAS, *Streptococcus pyogenes*) is a leading global cause of infection-related

82 morbidity and mortality. A modern controlled human infection model (CHIM) of GAS pharyngitis could

83 help to accelerate vaccine development and drive pathogenesis research. Challenge strain selection is

84 critical to the safety and success of any CHIM and especially so for an organism such as GAS with its wide

85 strain diversity and potential to cause severe life-threatening acute infections (e.g. toxic shock syndrome,

86 necrotizing fasciitis) and post-infectious complications (e.g. acute rheumatic fever, rheumatic heart disease,

87 acute post-streptococcal glomerulonephritis). In this paper, we outline the rationale for selecting an *emm75*

88 strain for initial use in a GAS pharyngitis CHIM in healthy adult volunteers, drawing on the findings of a

89 broad characterization effort spanning molecular epidemiology, in vitro assays, whole genome sequencing,

90 and animal model studies.

91

92

93 **Introduction**

94 Group A *Streptococcus* (GAS, *Streptococcus pyogenes*) is a major contributor to global infection-related
95 mortality and morbidity. It causes a diverse spectrum of human disease syndromes, from superficial
96 infections (e.g. pharyngitis, impetigo) to invasive disease (e.g. necrotizing fasciitis, toxic shock syndrome)
97 and autoimmune complications (acute rheumatic fever, rheumatic heart disease, glomerulonephritis)(1).
98 Development of a GAS vaccine has been impeded by scientific, regulatory, and commercial obstacles(2).
99 Controlled human infection models (CHIM) are increasingly assuming an important role for vaccine
100 development(3, 4). Drawing on the record of historical CHIM studies that included 172 participants (5-7),
101 a modern pharyngitis CHIM in healthy adult volunteers has been proposed as part of a re-energized global
102 effort to accelerate GAS vaccine development(8). Selection of a thoroughly characterized strain is central
103 for development of a GAS CHIM.

105 A successful CHIM requires that infection and/or symptomatic disease endpoints are reached reliably,
106 safely, and bear sufficient resemblance to a natural state to suggest generalizability. The diverse clinical
107 and microbiological profile of GAS presents challenges for CHIM study design, especially strain selection.
108 There are more than 200 different GAS *emm*-types. This widely-used classification system is based on one
109 part of the gene encoding a single GAS antigen, the M protein. No other antigen has been as closely
110 studied, and the concept of M protein type-specific immunity has been a cornerstone of GAS research.
111 GAS is a highly-adapted human pathogen, and the limitations of *in vitro* assays and animal models have
112 been well described. After more than a century of research, fundamental aspects of pathogenesis and
113 human immune protection against GAS remain unknown. These knowledge gaps are simultaneously an
114 argument for building a CHIM and a source of uncertainty in conceiving its design.

116 A thorough and explicitly stated rationale for strain selection is an important step in minimizing potential
117 harm to participants and maximizing scientific impact. We considered desirable characteristics in selecting
118 an initial strain to establish a GAS pharyngitis CHIM and surveyed available collections for suitable
119 strains, focusing on an *emm75* strain (GAS M75 611024, ‘M75’) isolated in 2011 from the throat of a 5-

120 year-old girl with acute symptomatic pharyngitis in Melbourne (Table 1 and Supplementary Table S1)(9).
121 In this paper, we present a multifaceted characterisation of the preferred CHIM candidate M75 strain,
122 compared to three others: GAS M12 611025, ‘M12’, an alternative challenge candidate; M1T1 5448
123 (‘5448’), representative of the M1T1 clone recently responsible for most pharyngitis and invasive disease
124 globally(10); and CDC SS-496, a M1 strain administered to 88 subjects in 1970s pharyngitis CHIM studies
125 (‘SS-496’)(5, 7).

126

127 **Methods**

128 *Bacterial isolates*

129 M75 611024 and M12 611025 were isolated in 2011 from throat swabs collected from children with acute
130 pharyngitis in Melbourne, Australia, and stored at the Murdoch Children’s Research Institute(9). Professor
131 Mark Walker at the University of Queensland supplied 5448(10, 11). The US Centers for Disease Control
132 and Prevention (CDC) *Streptococcus* Laboratory provided the SS-496 strain, submitted in 1958 from Duke
133 University.

134

135 *Growth and viability*

136 For administration to human volunteers, an animal-free medium must sustain sufficient strain growth. A
137 chemically-defined medium was developed (‘VR broth’) consisting of RPMI-1640 (Gibco) and 2% (w/v)
138 Veggietone GMO-free soya-peptone (Oxoid). Eight-hour growth assays were done comparing growth of
139 M75 in this medium and in Todd-Hewitt broth (Oxoid) with 1% (w/v) yeast extract (Bacto)(THY).
140 Bacteria were grown in 125ml Erlenmeyer flasks containing 25ml of VR or THY broth and agitated gently
141 75 RPM. To simulate manufacturing processes, M75 growth in VR broth was examined after seven days
142 repeated *in vitro* passage, using frozen cultures of three post-passage clones and the pre-passage parent
143 isolate.

144

145 *Hyaluronic acid (HA) capsule assay*

146 The HA capsule is a GAS virulence factor, resisting opsonophagocytosis(12). Capsular HA levels were
147 quantified as previously described, using a test kit (Corgenix)(10).

148

149

150

151 *Attachment properties*

152 Cell culture lines have been used to study GAS adherence(13, 14). We used Detroit 562 (D562) human
153 pharyngeal cells and skin HaCaT cells, simulating natural sites of infection. As previously described, total
154 cell-associated GAS (percentage of original inoculum) and invasiveness (intracellular fraction of total cell-
155 associated GAS) were determined using GAS grown to mid-exponential phase ($OD_{600nm} \sim 0.5$) in VR broth
156 and diluted in 500 μ l of assay media (MEM with 5% Fetal Bovine Serum; Gibco) to a multiplicity of
157 infection of 5:1 (GAS:cells)(10). Inoculated trays were centrifuged for five minutes at 200 $\times g$ and
158 incubated for one hour at 37°C in 5% CO_2 , then washed three times with PBS to remove non-adherent
159 bacteria. Cell-associated GAS (adherent plus invasive) were detached using 200 μ l 0.25% trypsin, lysed
160 with 0.025% Triton-X (Sigma) in dH_2O , and enumerated by track dilution on horse blood agar (HBA). To
161 measure invasive bacteria cells were washed once after incubation in assay media for one hour then
162 incubated for another hour in media containing 100 μ g/ml gentamicin and enumerated as before.

163

164 *Delivery characteristics*

165 To assess M75 viability following storage at -80°C, bacteria were grown in VR broth ($OD_{600nm}=0.5$),
166 centrifuged and suspended in broth containing 10 % (v/v) glycerol (Sanofi). Vials containing 10^5 , 10^6 , 10^7 ,
167 and 10^8 CFU/ml were thawed at intervals and immediately tested without washing (mimicking the
168 challenge protocol) for: 1) growth in solid and liquid media; 2) viability by enumeration; and 3) attachment
169 properties, as above.

170

171 For the challenge procedure, swab uptake and release of the GAS inoculum should be consistent. We
172 simulated direct oropharyngeal application using four Copan swabs compared to cotton:

173 FLOQSwab™(nylon), Dacron™ (polyester), and small (S) and large (L) rayon swabs. For uptake, vials of
174 broth were weighed before and after dipping of swabs for 10 seconds. Swab release of GAS was measured
175 by dipping swabs in 1ml vials containing $1-3 \times 10^5$ CFU of M75 for 10 seconds followed by transferring to
176 1ml of PBS for 10 seconds, then enumerated by spread plate dilutions.

177

178 *Antibiotic susceptibility testing*

179 Minimum inhibitory concentrations (MIC) were determined by E-test, and double disk diffusion (D-zone
180 test) used to detect inducible clindamycin resistance. Interpretive breakpoints of the Clinical and
181 Laboratory Standards Institute (penicillin, erythromycin, azithromycin, clindamycin, levofloxacin) and
182 European Committee on Antimicrobial Susceptibility Testing (rifampicin) were used(15, 16).

183

184 *Whole genome sequencing and phylogenetic analyses*

185 The complete M75 611024 genome sequence was determined using long-read single molecule real-time
186 sequencing on the Pacific Biosciences RS II platform. Filtering of the long reads identified 104,694 reads
187 with an average polymerase read length of 4.1kb. A single circular assembly was generated using SMRT
188 analysis v2.3.0 (Pacific Biosciences), HGAP v3 and polished using Quiver at an average read depth of 96-
189 fold. To aid assembly validation, M75 was also sequenced on an Illumina Next-seq 500 to produce paired-
190 end reads with a read length of 150 bases. The M75 611024 genome sequence has been submitted to
191 GenBank (accession number CP033621). The genomes of M12 611025 and CDC SS-496 were sequenced
192 by Illumina Next-seq 500 with a paired-end read length of 150 bases. Draft genome assemblies were
193 generated using SPAdes v3.12.0. Illumina short reads of M12 611025 (accession number SRR8217179)
194 and CDC SS-496 (SRR8217180) have been submitted to the Short Read Archive (PRJNA504701).

195

196 To study M75 genomic stability, three clones were sequenced by Illumina Next-seq 500, 150 bp paired-end
197 reads, after seven days of repeated *in vitro* passage. These sequences were aligned with the pre-passage
198 parent M75 reference sequence to identify single nucleotide polymorphisms (SNP).

199

Phylogenetic analysis of a global dataset of *emm75* isolates was determined by mapping short read sequences of 131 global *emm75* genomes from the UK (n=124), USA(n=4), and France (n=3) (17-19) to the M75 611024 reference genome with BWA MEM (v0.7.16). Single nucleotide polymorphisms (SNPs) with a Phred quality score ≥ 30 were identified in each isolate using SAMtools pileup with a minimum coverage of 30x. Prophage sequences within M75 611024 were identified using the Phaster server with SNPs located within these prophage excluded as they represent evolutionary confounders. A maximum likelihood phylogenetic tree was built from 1,046 concatenated SNP sites using RAxML v8.2.8 with the general time-reversible model and gamma correction with 100 bootstrap re-samplings to assess phylogenetic support.

Phylogeny of 141 *emm12* genomes, including the Illumina reads of M12 611025 and sequences from Australia, USA, Hong Kong and mainland China, was analyzed by mapping to the reference genome HKU16 (strain QMH11M0907901 (GenBank AFRY01000001) from 1,452 vertically inherited SNPs as previously described (20). Illumina reads of the M1 genome sequence CDC SS-496 were mapped to MGAS5005 (GenBank NC_007297) and other M1 reference genomes with phylogeny inferred from 780 vertically inherited SNPs.

Virulence factors and vaccine antigens

Virulence gene carriage was determined for M75, M12, 5448, and SS-496, by blastN screening assemblies against the virulence factor database (VFDB)(21). Gene presence was defined by an 80% nucleotide cut-off over 80% of gene length.

Protection in animal models has been shown for more than twenty-five candidate GAS vaccine protein antigens and several peptide-based antigens(22). For protein antigens, presence was defined by an 80% nucleotide cut-off over 80% of gene length. For sequence-constrained peptide-based vaccine epitopes: J8.0 (SREAKKQVEKAL)(23); the StreptInCor sequence (KGLRRDLASREAKKQLEAEQQKLEEQNKIS-EASRKGLRRDLASREAKKQVEKA)(24) and associated T-cell (KGLRRDLASREAKKQLEAEQQ),

227 B-cell (EASRKGLRRDLASREAKKQVEKA), and common B-T-cell epitopes (KGLRRDLASREA-
228 KKQ); a 100% nucleotide sequence match was taken to define presence, although 100% homology may
229 not be required to induce production of broadly cross-reactive antibodies and vaccine protection.

230

231 *Neutrophil killing assay*

232 Survival of M75, M12, and 5448 incubated with human neutrophils *in vitro* was assayed as previously
233 described(25). Experiments were performed in triplicate using mid-exponential phase GAS at a
234 multiplicity of infection of 10:1. Differences in neutrophil survival were analysed using 1-way analysis of
235 variance (Graphpad Prism).

236

237 *Murine invasive model*

238 Strain virulence was compared in a humanized plasminogen transgenic *AlbPLG1* mouse model(26). In
239 separate experiments, groups (n=10) of *AlbPLG1* (+/-) mice were administered subcutaneous doses of
240 either M75 (7×10^7 CFU), M12 (8×10^7 CFU), 5448 (5×10^7 CFU), or SS-496 (3×10^7 CFU), and survival
241 monitored for ten days, as previously described(27, 28).

242

243 *Ethics statement*

244 Animal procedures followed the Australian Code for the Care and Use of Animals for Scientific Purposes
245 and were approved by the University of Queensland Animal Ethics Committee
246 (SCMB/140/16/NHMRC)(29). An initial dose-ranging CHIM study has been approved by The Alfred
247 Hospital Ethics Committee (500/17) and is registered at ClinicalTrials.gov (NCT03361163).

248

249 **Results**

250 *Growth in an animal-free medium*

251 Compared to THY broth, no detrimental effect on growth of M75, M12, and 5448 was observed in the
252 animal-free media (Figure 1A). Eight-hour growth curves for M75 clones tested after seven days repeated
253 *in vitro* passage were similar to the non-passaged parent (data not shown).

254

255 *Attachment properties*

256 M75 had the highest adherence to D562 (75%) and HaCaT (81%) cells (Figure 1B). M12 (53%) and 5448
257 (51%) were similarly adherent to D562 cells. M12 preferentially adhered to D562 over HaCaT cells
258 ($p=0.005$), whereas M75 and M1T1 showed no preference. The affinity of M12 to pharyngeal over skin
259 cells matches its designation as an A-C pattern strain, associated with throat tissue tropism(30).
260 Invasiveness of M75 and M12 was low for both cell lines ($\leq 0.45\%$). Invasion by 5448 of HaCaT cells
261 (10%) was greater than for D562 cells (0.2%) (Figure 1B).

262

263 *Capsule production*

264 M75 produced 74 ng/ml of HA capsule, compared to M12 which produced 7506 ng/ml (Figure 1C).
265 Capsule production by 5448 matched previous findings(31).

266

267 *Delivery characteristics and viability*

268 The Dacron™ swab was considered most suitable for delivery of the challenge inoculum (Supplementary
269 Figure S1). Mean broth uptake by Dacron™ (105mg) and Rayon S (108mg) swabs was comparable to
270 cotton (129mg) and uptake variance was lowest (7.7mg) for the Dacron™ swab. Superior release was
271 noted for the Dacron swab with a mean of 1.8×10^3 CFU of M75 recovered after swab dipping (Figure 2B).
272 Recovery from M75 vials frozen for four months did not fall below 95% of the original inoculum at T=0.
273 Adherence and invasion were similarly unaffected by storage (data not shown).

274

275 *Antibiotic susceptibility*

276 M75 was susceptible to all tested antibiotics, while M12 was resistant to macrolides and fluoroquinolones
277 (Table 2). All strains were susceptible to clindamycin, and inducible resistance was not detected.

278

279 *Whole genome sequencing and phylogenetic analyses*

280 The complete genome of M75 611024 is comprised of a single chromosome of 1,852,894 bp (Figure 2A).
281 M75 has the multi-locus sequence type (MLST) ST150 and contains the *emm75.0* allele, and *mrp24* and
282 *enn334* alleles corresponding to the *emm*-like genes *mrp* and *enn* (P. Smeesters, personal communication,
283 July 2018). Three putative prophage sequences were identified in M75 harboring the endonuclease
284 streptodornase 3 (*spd3*); pyrogenic exotoxins *speL* and *speM*; and the endonuclease *sdn*. M75 shared a
285 hypothetical ancestral relationship with a UK *emm75* cluster (Figure 2B) yet represents a distinct evolving
286 lineage, suggesting an ancestral relationship to modern day ST150 *emm75* clones.

287

288 One SNP was found for each of three M75 clones sequenced after seven days repeated *in vitro* passage,
289 compared to the non-passaged parent strain. Each SNP was intergenic and different, suggestive of random
290 mutations of unlikely functional consequence (data not shown).

291

292 M12 611025 belongs to MLST ST36 and carries the *emm12.0* allele. It shares a high degree of genome
293 conservation with other *emm12* genome sequences, varying in prophage and integrative conjugative
294 element content relative to the reference genomes HKU16 and MGAS9429 (Figure 2C). Phylogenetic
295 analysis alongside 141 extant *emm12* isolates showed an evolutionary relationship with other modern ST36
296 strains (Figure 2D) including recent scarlet fever outbreak strains(20).

297

298 The historical challenge strain SS-496 shares a higher degree of genetic and evolutionary similarity with
299 the ancestral M1 reference strain SF370, relative to the modern M1T1 strains MGAS5005 and 5448
300 (Figures 2E, 2F)(32). SS-496 contains the historical SF370-like *purA* to *nadC* genomic region encoding
301 Streptolysin O.

302

303 *Virulence factors and vaccine antigens*

304 M75, M12, and SS-496 carry genes for an array of adhesion and invasion factors common to many *emm*-
305 types (Table 3). M75 contains a frameshift mutation in the fibronectin binding protein Sfb1 within the FCT
306 locus. M12 carries the streptococcal superantigen A (*ssa*) gene recently reported in scarlet fever-associated

isolates in China and the UK(33). M12 does not carry the multidrug-resistant integrative conjugative element ICE-*emm12* or the *ssa*-carrying prophage Φ HKU.vir linked to the emergence of scarlet fever clades(20). The virulence profile of SS-496 is similar to pre-1980 M1 strains such as SF370, with *speH* and *speI* exotoxins and absence of the *speA* exotoxin typical of modern isolates such as 5448 (Table 2). M75, M12, SS-496 and 5448 all possess wildtype *covR/S* and *ropB* two-component virulence regulators.

High carriage (>60%) of protein and peptide candidate vaccine antigens was observed for M75 and M12 using a homology-based genome approach (Table 4).

Human neutrophil killing assay

M75 was most susceptible to *in vitro* killing when incubated with human neutrophils, although killing was observed for all strains (Figure 1D).

Mouse lethal invasive model

Compared to M12, 5448, and SS-496, M75 was avirulent in the humanized mouse invasive disease model (Figure 1E).

Discussion

We have described the rationale for selecting M75 for initial use in a new GAS pharyngitis CHIM in healthy adults, including results of diverse pre-clinical studies assessing its fitness for purpose. For context and comparison, we have presented results for three other strains (M12, 5448, SS-496).

M75 is compatible with critical protocol points: reliable growth in an animal-free medium; retains growth and attachment properties after prolonged storage at -80° C; consistent delivery using a commercially available swab; and susceptibility to antibiotics used to treat GAS pharyngitis. M75 looks to have an acceptable virulence profile, with capacity to cause pharyngitis and low potential for invasive disease. M75 has attractive attachment properties for immortalized human pharyngeal and skin cell lines, with limited

cellular invasion. M75 was highly susceptible to *in vitro* killing by human neutrophils, possibly due to its minimal capsular HA production. In a humanized mouse model of invasive infection, M75 was avirulent whereas M12 and both M1 strains were lethal. Whole genome sequencing placed the strains in the context of epidemiologically-related phylotypes, found broad representation of candidate vaccine antigens, and a relatively restricted array of virulence factor genes in M75.

CHIM strain selection has been guided by varied general and pathogen-specific considerations, all with the goal of safely and reliably reproducing relevant and generalisable asymptomatic (infection/carriage) or symptomatic (disease) study endpoints(3, 4, 34). Suitable well-characterized strains may already exist, such as the *Salmonella* Typhi Quail strain(35). Patients with mild-moderately severe uncomplicated disease may be a source of ‘naturally attenuated’ new strains. Multiple strains, sometimes from different locations, may be required to represent natural strain diversity and/or enable heterologous re-challenge(36). If mechanisms of severe infection and/or complications are known, the implicated virulence factor(s) may be avoided (e.g. Shiga toxin-producing *Escherichia coli*(37) and *Campylobacter jejuni* inducing cross-reactive antibodies to GM1 and GQ1b gangliosides(38)). Pathogens may be modified for use, such as the propagation of single sex *Schistosoma mansoni* cercariae to prevent chronic schistosomiasis(39). For vaccine studies, target antigen(s) must be present in the challenge strain(s). In every instance, strains must be characterized and be compatible with protocols for manufacturing, inoculation, and techniques to measure organism and host responses.

The limitations of this characterization effort are inherent in the rationale for pursuing a GAS CHIM. *In vitro* assays, genomics, and animal models do not fully capture or predict the dynamic elements and sequelae of human infection by GAS, a highly-adapted and human-restricted pathogen. Even advanced non-human primate models produce a pharyngitis syndrome with important differences from human disease. A single contemporary clone, represented here by 5448, is simultaneously the most common cause in urbanized settings of both the mildest and most severe disease syndromes, with the basis for tissue-tropism and bacterial-human genotype-phenotype relationships still relatively obscure. These uncertainties

dictate a cautious approach extending beyond strain selection, including: strain manufacture following principles of Good Manufacturing Practice; initial inclusion of healthy adults only without risk factors for severe GAS disease; a dose-ranging study to establish attack rate and safety; inpatient admission at a trials facility supported by a tertiary hospital; universal antibiotic treatment; outpatient follow-up; and echocardiography at screening and final visits.

367

A generic limitation of CHIM studies is the uncertain degree to which data from healthy adults experiencing a single syndrome (pharyngitis) caused by one strain (M75) can be generalized to other subjects, syndromes, strains, and settings (e.g. children with GAS skin infections due to other *emm*-types in low- and middle-income countries). While inclusion of other strains and even a skin infection CHIM are conceivable extensions, model findings must be interpreted alongside knowledge derived from more naturalistic studies. For vaccine development, a GAS pharyngitis CHIM has dual scientific and strategic purposes, aiming to serve as a bridge to field trials with a more natural distribution of subjects, syndromes, and strains.

376

With a view to the very high priority given to participant safety and risk minimization, findings from these strain characterization studies reinforce the appropriateness of M75 for initial use in a GAS pharyngitis CHIM.

380

(3081 words)

382

383 **Funding**

384 This work is supported by the Australian National Health and Medical Research Council [GNT1099183;
385 GNT1133299 to J.O; GNT1126805 to M.R.D; GNT1102621, GNT1071659, GNT1074899 to M.J.W; and
386 GNT1127077 to A.S]; the Australian Heart Foundation [to A.S]; the Garnett Passe and Rodney Williams
387 Memorial Foundation [to C.L.Y.O]; and the Belgian National Fund for Scientific Research (FNRS) [PDR
388 T.0255.16, CDR J.0019.17 to P.R.S].

389

390 **Transparency declaration**

391 The lead author affirms that this manuscript is an honest, accurate, and transparent account of the study
392 being reported; that no important aspects of the study have been omitted; and that any discrepancies from
393 the study as planned have been explained. The authors have no financial or other conflicts to declare.

394

395 **Acknowledgements**

396 The Controlled Human Infection for Vaccination against *Streptococcus pyogenes* (CHIVAS) study group
397 includes Andrew Steer, Anneke Grobler (Murdoch Children’s Research Institute); Michael Batzloff,
398 Michael Good, and Manisha Pandey (Griffith University); Jonathan Carapetis (Telethon Kids Institute),
399 Pierre Smeesters (Université Libre de Bruxelles); Claire Waddington (University of Cambridge); Jim Dale
400 (University of Tennessee); Tibor Schuster (McGill University); Allen Cheng (Monash University);
401 Andrew Pollard (Oxford University); James McCarthy (QIMR Berghofer); and Mark Walker (University
402 of Queensland).

403

404 Thanks to Dr. Deborah Williamson, Dr. Kate Worthing and staff at the Microbiological Diagnostic Unit
405 Public Health Laboratory, The Peter Doherty Institute for Infection and Immunity, The University of
406 Melbourne. Thanks to Dr. Bernard Beall and staff at the US Centers for Disease Control and Prevention
407 *Streptococcus* Laboratory. We acknowledge the assistance of the sequencing and pathogen informatics
408 core teams at the Wellcome Trust Sanger Institute and support for the genome sequencing components
409 from The Wellcome Trust, UK. JO and AS are members of the Human Infection Challenge Network for

Vaccine Development (HIC-Vac), which is funded by the UK Global Challenges Research Fund (GCRF) Networks in Vaccines Research and Development, which was co-funded by the Medical Research Council (MRC) and Biotechnology and Biological Sciences Research Council (BBSRC).

References

1. Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, Sriprakash KS, Sanderson-Smith ML, Nizet V. 2014. Disease manifestations and pathogenic mechanisms of Group A Streptococcus. *Clin Microbiol Rev* 27:264-301.
2. Steer AC, Carapetis JR, Dale JB, Fraser JD, Good MF, Guilherme L, Moreland NJ, Mulholland EK, Schodel F, Smeesters PR. 2016. Status of research and development of vaccines for Streptococcus pyogenes. *Vaccine* 34:2953-8.
3. Roestenberg M, Hoogerwerf MA, Ferreira DM, Mordmuller B, Yazdanbakhsh M. 2018. Experimental infection of human volunteers. *Lancet Infect Dis* doi:10.1016/S1473-3099(18)30177-4.
4. Anonymous. 2018. Controlled Human Infection Model Studies: Summary of a workshop held on 6 February 2018. . The Academy of Medical Sciences., Sciences. TAOm,
5. Polly SM, Waldman RH, High P, Wittner MK, Dorfman A. 1975. Protective studies with a group A streptococcal M protein vaccine. II. Challenge of volunteers after local immunization in the upper respiratory tract. *J Infect Dis* 131:217-24.
6. D'Alessandri R, Plotkin G, Kluge RM, Wittner MK, Fox EN, Dorfman A, Waldman RH. 1978. Protective studies with group A streptococcal M protein vaccine. III. Challenge of volunteers after systemic or intranasal immunization with Type 3 or Type 12 group A Streptococcus. *J Infect Dis* 138:712-8.
7. Fox EN, Waldman RH, Wittner MK, Mauceri AA, Dorfman A. 1973. Protective study with a group A streptococcal M protein vaccine. Infectivity challenge of human volunteers. *J Clin Invest* 52:1885-92.
8. Osowicki J, Vekemans J, Kaslow DC, Friede MH, Kim JH, Steer AC. 2018. WHO/IVI global stakeholder consultation on group A Streptococcus vaccine development: Report from a meeting held on 12-13 December 2016. *Vaccine* 36:3397-3405.
9. Dunne EM, Marshall JL, Baker CA, Manning J, Gonis G, Danchin MH, Smeesters PR, Satzke C, Steer AC. 2013. Detection of group A streptococcal pharyngitis by quantitative PCR. *BMC Infect Dis* 13:312.
10. Hollands A, Pence MA, Timmer AM, Osvath SR, Turnbull L, Whitchurch CB, Walker MJ, Nizet V. 2010. Genetic switch to hypervirulence reduces colonization phenotypes of the globally disseminated group A streptococcus M1T1 clone. *J Infect Dis* 202:11-9.
11. Chatellier S, Ihendyane N, Kansal RG, Khambaty F, Basma H, Norrby-Teglund A, Low DE, McGeer A, Kotb M. 2000. Genetic relatedness and superantigen expression in group A streptococcus serotype M1 isolates from patients with severe and nonsevere invasive diseases. *Infect Immun* 68:3523-34.
12. Dale JB, Washburn RG, Marques MB, Wessels MR. 1996. Hyaluronate capsule and surface M protein in resistance to opsonization of group A streptococci. *Infect Immun* 64:1495-501.
13. Loh JMS, Tsai JC, Proft T. 2017. The ability of Group A streptococcus to adhere to immortalized human skin versus throat cell lines does not reflect their predicted tissue tropism. *Clin Microbiol Infect* 23:677 e1-677 e3.
14. Ryan PA, Juncosa B. 2016. Group A Streptococcal Adherence. *In* Ferretti JJ, Stevens DL, Fischetti VA (ed), Streptococcus pyogenes : Basic Biology to Clinical Manifestations, Oklahoma City (OK).

15. Clinical and Laboratory Standards Institute. 2016. Performance Standards for Antimicrobial Susceptibility Testing: Twenty-sixth Informational Supplement M100-S26, CLSI, Wayne, PA, USA.
16. Anonymous. The European Committee on Antimicrobial Susceptibility Testing. Breakpoint tables for interpretation of MICs and zone diameters. Version 6.0, 2016., <http://www.eucast.org>.
17. Kapatai G, Coelho J, Platt S, Chalker VJ. 2017. Whole genome sequencing of group A *Streptococcus*: development and evaluation of an automated pipeline for emm gene typing. *PeerJ* 5:e3226.
18. Chochua S, Metcalf BJ, Li Z, Rivers J, Mathis S, Jackson D, Gertz RE, Jr., Srinivasan V, Lynfield R, Van Beneden C, McGee L, Beall B. 2017. Population and Whole Genome Sequence Based Characterization of Invasive Group A *Streptococci* Recovered in the United States during 2015. *MBio* 8.
19. Rochefort A, Boukthir S, Moullec S, Meygret A, Adnani Y, Lavenier D, Fali A, Kayal S. 2017. Full Sequencing and Genomic Analysis of Three emm75 Group A *Streptococcus* Strains Recovered in the Course of an Epidemiological Shift in French Brittany. *Genome Announc* 5.
20. Davies MR, Holden MT, Coupland P, Chen JH, Venturini C, Barnett TC, Zakour NL, Tse H, Dougan G, Yuen KY, Walker MJ. 2015. Emergence of scarlet fever *Streptococcus pyogenes* emm12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance. *Nat Genet* 47:84-7.
21. Chen L, Zheng D, Liu B, Yang J, Jin Q. 2016. VFDB 2016: hierarchical and refined dataset for big data analysis--10 years on. *Nucleic Acids Res* 44:D694-7.
22. Henningham A, Gillen CM, Walker MJ. 2013. Group a streptococcal vaccine candidates: potential for the development of a human vaccine. *Curr Top Microbiol Immunol* 368:207-42.
23. Batzloff MR, Hayman WA, Davies MR, Zeng M, Pruksakorn S, Brandt ER, Good MF. 2003. Protection against group A streptococcus by immunization with J8-diphtheria toxoid: contribution of J8- and diphtheria toxoid-specific antibodies to protection. *J Infect Dis* 187:1598-608.
24. Guilherme L, Fae KC, Higa F, Chaves L, Oshiro SE, Freschi de Barros S, Puschel C, Juliano MA, Tanaka AC, Spina G, Kalil J. 2006. Towards a vaccine against rheumatic fever. *Clin Dev Immunol* 13:125-32.
25. Buchanan JT, Simpson AJ, Aziz RK, Liu GY, Kristian SA, Kotb M, Feramisco J, Nizet V. 2006. DNase expression allows the pathogen group A *Streptococcus* to escape killing in neutrophil extracellular traps. *Curr Biol* 16:396-400.
26. Sun H, Ringdahl U, Homeister JW, Fay WP, Engleberg NC, Yang AY, Rozek LS, Wang X, Sjobring U, Ginsburg D. 2004. Plasminogen is a critical host pathogenicity factor for group A streptococcal infection. *Science* 305:1283-6.
27. Walker MJ, Hollands A, Sanderson-Smith ML, Cole JN, Kirk JK, Henningham A, McArthur JD, Dinkla K, Aziz RK, Kansal RG, Simpson AJ, Buchanan JT, Chhatwal GS, Kotb M, Nizet V. 2007. DNase Sda1 provides selection pressure for a switch to invasive group A streptococcal infection. *Nat Med* 13:981-5.
28. Maamary PG, Sanderson-Smith ML, Aziz RK, Hollands A, Cole JN, McKay FC, McArthur JD, Kirk JK, Cork AJ, Keefe RJ, Kansal RG, Sun H, Taylor WL, Chhatwal GS, Ginsburg D, Nizet V, Kotb M, Walker MJ. 2010. Parameters governing invasive disease propensity of non-M1 serotype group A streptococci. *J Innate Immun* 2:596-606.
29. Anonymous. National Health and Medical Research Council. 2013. Australian code for the care and use of animals for scientific purposes, 8th ed. National Health and Medical Research Council, Canberra, Australia.
30. Bessen DE, Smeesters PR, Beall BW. 2018. Molecular Epidemiology, Ecology, and Evolution of Group A *Streptococci*. *Microbiol Spectr* 6.
31. Cole JN, Pence MA, von Kockritz-Blickwede M, Hollands A, Gallo RL, Walker MJ, Nizet V. 2010. M protein and hyaluronic acid capsule are essential for in vivo selection of covRS mutations characteristic of invasive serotype M1T1 group A *Streptococcus*. *MBio* 1.
32. Nasser W, Beres SB, Olsen RJ, Dean MA, Rice KA, Long SW, Kristinsson KG, Gottfredsson M, Vuopio J, Raisanen K, Caugant DA, Steinbakk M, Low DE, McGeer A, Darenberg J, Henriques-

- Normark B, Van Beneden CA, Hoffmann S, Musser JM. 2014. Evolutionary pathway to increased virulence and epidemic group A Streptococcus disease derived from 3,615 genome sequences. *Proc Natl Acad Sci U S A* 111:E1768-76.
33. You Y, Davies MR, Protani M, McIntyre L, Walker MJ, Zhang J. 2018. Scarlet Fever Epidemic in China Caused by Streptococcus pyogenes Serotype M12: Epidemiologic and Molecular Analysis. *EBioMedicine* 28:128-135.
34. Darton TC, Blohmke CJ, Moorthy VS, Altmann DM, Hayden FG, Clutterbuck EA, Levine MM, Hill AV, Pollard AJ. 2015. Design, recruitment, and microbiological considerations in human challenge studies. *Lancet Infect Dis* 15:840-51.
35. Waddington CS, Darton TC, Jones C, Haworth K, Peters A, John T, Thompson BA, Kerridge SA, Kingsley RA, Zhou L, Holt KE, Yu LM, Lockhart S, Farrar JJ, Sztein MB, Dougan G, Angus B, Levine MM, Pollard AJ. 2014. An outpatient, ambulant-design, controlled human infection model using escalating doses of Salmonella Typhi challenge delivered in sodium bicarbonate solution. *Clin Infect Dis* 58:1230-40.
36. Stanistic DI, McCarthy JS, Good MF. 2018. Controlled Human Malaria Infection: Applications, Advances, and Challenges. *Infect Immun* 86.
37. Harro C, Chakraborty S, Feller A, DeNearing B, Cage A, Ram M, Lundgren A, Svennerholm AM, Bourgeois AL, Walker RI, Sack DA. 2011. Refinement of a human challenge model for evaluation of enterotoxigenic Escherichia coli vaccines. *Clin Vaccine Immunol* 18:1719-27.
38. Tribble DR, Baqar S, Carmolli MP, Porter C, Pierce KK, Sadigh K, Guerry P, Larsson CJ, Rockabrand D, Ventone CH, Poly F, Lyon CE, Dakdouk S, Fingar A, Gilliland T, Daunais P, Jones E, Rymarchyk S, Huston C, Darsley M, Kirkpatrick BD. 2009. Campylobacter jejuni strain CG8421: a refined model for the study of Campylobacteriosis and evaluation of Campylobacter vaccines in human subjects. *Clin Infect Dis* 49:1512-9.
39. Janse JJ, Langenberg MCC, Kos-Van Oosterhoud J, Ozir-Fazalalikhani A, Brienens EAT, Winkel BMF, Erkens MAA, van der Beek MT, van Lieshout L, Smits HH, Webster BL, Zandvliet ML, Verbeek R, Westra IM, Meij P, Visser LG, van Diepen A, Hokke CH, Yazdanbakhsh M, Roestenberg M. 2018. Establishing the Production of Male Schistosoma mansoni Cercariae for a Controlled Human Infection Model. *J Infect Dis* 218:1142-1146.
40. O'Loughlin RE, Roberson A, Cieslak PR, Lynfield R, Gershman K, Craig A, Albanese BA, Farley MM, Barrett NL, Spina NL, Beall B, Harrison LH, Reingold A, Van Beneden C. 2007. The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000-2004. *Clin Infect Dis* 45:853-62.
41. Nelson GE, Pondo T, Toews KA, Farley MM, Lindegren ML, Lynfield R, Aragon D, Zansky SM, Watt JP, Cieslak PR, Angeles K, Harrison LH, Petit S, Beall B, Van Beneden CA. 2016. Epidemiology of Invasive Group A Streptococcal Infections in the United States, 2005-2012. *Clin Infect Dis* 63:478-86.
42. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): simple prokaryote genome comparisons. *BMC Genomics* 12:402.

Tables

Table 1. Preferred strain characteristics for a controlled-human infection model of GAS pharyngitis*

Desirable strain characteristics	Rationale	M75 611024
<i>A definite but uncommon contemporary cause of symptomatic pharyngitis</i>	Pharyngitis is the critical early target for GAS vaccine development. Historical CHIM studies offer a template for a reliable and safe protocol. GAS pharyngitis is most common in childhood and adolescence, suggesting previous exposure and immune memory could prevent experimentally-induced disease in adult volunteers.	<ul style="list-style-type: none"> • From a child with symptomatic GAS pharyngitis in Melbourne, 2011 • Pre-existing immunity in adults is unknown (no correlate of protection) • $\leq 5\%$ of strains in most recent pharyngitis studies are <i>emm75</i>
<i>Should cause skin infection</i>	Common GAS skin infections (e.g. impetigo) will also be important in initial vaccine field trials. Ideally, the pharyngitis CHIM strain(s) should also be suitable for use in a potential future human model of cutaneous GAS infection.	<ul style="list-style-type: none"> • E pattern ‘generalist’ (throat and skin infections) • Cluster E6 - linked phylogenetically to D pattern skin isolates
<i>An uncommon cause of invasive GAS disease and immunological sequelae</i>	GAS pharyngitis can lead to locally-invasive infectious complications (e.g. retropharyngeal abscess), severe invasive infection (e.g. streptococcal toxic-shock syndrome, STSS), acute rheumatic fever and glomerulonephritis.	<ul style="list-style-type: none"> • $\leq 5\%$ of isolates in recent reports of invasive GAS are <i>emm75</i>; from 2000-2016, 403/17,002 (2.4%) typeable invasive isolates reported to the US CDC’s Active Bacterial Core surveillance were <i>emm75</i> (Chris A. Van Beneden, Personal Communication, September 11th, 2018) (40, 41) • <i>emm75</i> strains rarely associated with ARF/RHD or APSGN (1)
<i>Should have predictable and limited virulence, and be suitable for use in animal models</i>	Whole genome sequencing, in vitro assays and animal models may inform understanding of a GAS strain’s relative virulence, although none fully predict human disease patterns.	<ul style="list-style-type: none"> • <i>covR/S</i> virulence regulator wildtype (non-mutant) • Does not bind plasminogen and fibrinogen • <i>emm75</i> strains have been used in animal nasopharyngitis and invasive disease models
<i>Should have limited antibiotic resistance</i>	Ideally, the challenge strain should be eradicated from the pharynx by antibiotic treatment. Resistance to penicillin has not been documented in GAS, however it does not reliably eradicate GAS from the pharynx. Resistance is variably observed to other drugs.	<ul style="list-style-type: none"> • See text
<i>The challenge strain should possess a wide array of candidate vaccine antigens</i>	For greatest impact, a GAS pharyngitis CHIM should be suitable for early use as a preliminary testing ground for vaccines.	<ul style="list-style-type: none"> • See text

* See web-only Supplementary Table S1 for a detailed and referenced version of this table

ARF: acute rheumatic fever; APSGN: acute post-streptococcal glomerulonephritis; CDC: Centers for Disease Control; CHIM: controlled human infection model; GAS: group A *Streptococcus*; RHD: rheumatic heart disease

Table 2. Antibiotic susceptibility of contemporary group A streptococcal strains M75 611024, M12 611025, and M1T1 5448

Antibiotic	Breakpoints [#] (mg/L)			E-test MIC (mg/L)		
	S	I	R	M75	M12	5448
Penicillin	≤0.12	-	-	0.012	0.016	0.012
Erythromycin	≤0.25	0.5	≥1	0.094	16	0.125
Clindamycin	≤0.25	0.5	≥1	0.125*	0.125*	0.125*
Azithromycin	≤0.5	1	≥2	1	64	1.5
Levofloxacin	≤2	4	≥8	0.5	4	0.5
Rifampicin	≤0.06		>0.5	0.064	0.064	0.125

* Inducible clindamycin resistance (D test) not detected

All CLSI breakpoints except rifampicin (EUCAST)

CLSI: Clinical & Laboratory Standards Institute; EUCAST: European Committee on Antimicrobial Susceptibility Testing; I: intermediate susceptibility; MIC: minimum inhibitory concentration; R: resistant; S: susceptible

Table 3. Group A *Streptococcus* virulence factor genomic screen

gene	function	M75 611024	M12 611025	M1T1 5448	CDC SS-496
<i>cfa/cfb</i>	CAMP factor	✓	✓	✓	✓
<i>tee (cpa)</i>	T-pilus antigen		✓	✓	✓
<i>cppA</i>	putative C3-degrading proteinase	✓	✓	✓	✓
<i>emm</i>	M-protein	✓	✓	✓	✓
<i>endoS</i>	endo-beta-N-acetylglucosaminidase F2 precursor	✓	✓	✓	✓
<i>fbp54</i>	fibrinogen-binding protein	✓	✓	✓	✓
<i>fctA</i>	major pilin Ap1 (FctA)			✓	✓
<i>fctB</i>	minor pilin Ap2 (FctB)			✓	✓
<i>grab</i>	protein G-related alpha 2M-binding protein		✓	✓	✓
<i>hasA</i>	hyaluronate synthase capsule	✓	✓	✓	✓
<i>hasB</i>	UDP-glucose 6-dehydrogenase capsule	✓	✓	✓	✓
<i>hasC</i>	putative UDP-glucose pyrophosphorylase	✓	✓	✓	✓
<i>htrA/degP</i>	serine protease	✓	✓	✓	✓
<i>htsA</i>	putative ABC transporter periplasmic binding protein	✓	✓	✓	✓
<i>htsB</i>	putative ABC transporter permease	✓	✓	✓	✓
<i>htsC</i>	putative ABC transporter ATP-binding protein	✓	✓	✓	✓
<i>hyl</i>	hyaluronoglucosaminidase	✓	✓	✓	✓
<i>hylA</i>	hyaluronate lyase precursor		✓	✓	✓
<i>hylP</i>	hyaluronoglucosaminidase	✓	✓	✓	✓
<i>ideS/mac</i>	IgG-degrading protease	✓	✓	✓	✓
<i>lepA</i>	signal peptidase I			✓	✓
<i>lmb</i>	laminin binding protein	✓	✓	✓	✓
<i>mf/spd</i>	deoxyribonuclease	✓	✓	✓	✓
<i>mf3</i>	deoxyribonuclease	✓		✓	✓
<i>plr/gapA</i>	glyceraldehyde-3-phosphate dehydrogenase	✓	✓	✓	✓
<i>prtF2</i>	collagen adhesion protein		✓		
<i>psaA</i>	manganese-binding protein	✓	✓	✓	✓
<i>sagA</i>	streptolysin S precursor	✓	✓	✓	✓
<i>sclA</i>	collagen-like surface protein A	✓	✓	✓	✓
<i>sclB</i>	putative collagen-like protein			✓	✓
<i>scpA</i>	C5A peptidase precursor	✓	✓	✓	✓
<i>sda</i>	phage-encoded streptodornase Sda		✓	✓	
<i>sdn</i>	phage-encoded endonuclease Sdn	✓			
<i>sfbII/sof</i>	fibronectin-binding protein	✓	✓		
<i>sfbX</i>	fibronectin-binding protein	✓	✓		
<i>shp</i>	hypothetical protein	✓	✓	✓	✓
<i>shr</i>	Fe3+-siderophore transporter	✓	✓	✓	✓
<i>sic</i>	streptococcal inhibitor of complement			✓	
<i>ska</i>	streptokinase precursor	✓	✓	✓	✓
<i>slo</i>	streptolysin O	✓	✓	✓	✓
<i>smeZ</i>	enterotoxin	✓	✓	✓	✓
<i>speB</i>	cysteine protease	✓	✓	✓	✓
<i>speA</i>	exotoxin A			✓	
<i>speG</i>	exotoxin G	✓	✓	✓	✓
<i>speH</i>	exotoxin H		✓		✓
<i>speI</i>	exotoxin I		✓		✓
<i>speJ</i>	exotoxin J			✓	✓
<i>speL</i>	exotoxin L	✓			
<i>speM</i>	exotoxin M	✓			
<i>spyA</i>	C3 family ADP-ribosyltransferase	✓	✓	✓	✓
<i>srtC1</i>	sortase			✓	✓
<i>ssa</i>	streptococcal superantigen A		✓		
<i>tig/ropA</i>	trigger factor	✓	✓	✓	✓

Table 4. Group A *Streptococcus* candidate vaccine antigen genomic screen*

gene/antigen	gene identifier ⁺	function	M75 611024	M12 611025	MIT1 5448
		M-protein: N-terminal (30-valent vaccine)	✓	✓	✓
		M-protein: C-terminal (J8.0)		✓	✓
		M-protein: C-terminal (StreptInCor T-cell epitope)			
		M-protein: C-terminal (StreptInCor B-cell epitope)			
		M-protein: C-terminal (StreptInCor common epitope)		✓	
adi	MGAS5005_spy1275	Arginine deaminase	✓	✓	✓
fbaA	MGAS5005_spy1714	Fibronectin-binding protein A			
fbp54	AAA57236	Fibronectin-binding protein 54	✓	✓	✓
oppA	M5005_spy0249	Oligopeptide-binding protein	✓	✓	✓
GAC	MGAS5005 [#]	Group A carbohydrate	✓	✓	✓
pulA	SF370_spy1972	Putative pullulanase	✓	✓	✓
r28	AF091393	Rib-like cell wall protein			
scpA	MGAS5005_spy1715	C5a peptidase	✓	✓	✓
sfbI	X67947	Streptococcal fibronectin binding protein I			
sfbII/sof	X83303	Serum opacity factor	✓		
shr	SPY1530	Streptococcal hemoprotein receptor	✓	✓	✓
sib35	AB254157	Streptococcal immunoglobulin-binding protein 35	✓	✓	✓
slo	M5005_spy0124	Streptolysin O	✓	✓	✓
spa	MGAS8232_spyM18_2046	Streptococcal protective antigen			
speA	X03929	Streptococcal pyrogenic exotoxin A			✓
speB	M5005_spy1735	Cysteine protease	✓	✓	✓
speC	SF370_spy0711	Streptococcal pyrogenic exotoxin C			
spy0651	MGAS5005_spy0651	Cell surface protein	✓	✓	✓
spy0762	MGAS5005_spy0762	Hypothetical membrane associated protein	✓	✓	✓
spy0942	MGAS5005_spy0942	Nucleoside-binding protein	✓	✓	✓
spyAD	MGAS5005_spy0229	Adhesin and division protein	✓	✓	✓
spyCEP	MGAS5005_spy0341	IL-8 serine protease	✓	✓	✓
sse	SF370_spy1407	Serine esterase		✓	✓
tee	MGAS5005_spy0109	T-antigen	✓	✓	✓
tif	SF370_spy1612	Trigger factor	✓	✓	✓

* BLAST analyses at a homology level of 80% for protein antigens and 100% for peptide-derived sequences.

+ Nucleotide gene sequences derived from completely sequenced genomes or listed GenBank identifiers. Accession numbers for genome sequences; MGAS5005 (CP000017), SF370 (AE004092) and MGAS8232 (AE009949).

GAC operon (~14.2 kb) refers to MGAS5005 genome coordinates 604873 to 619151.

Figure legends

Figure 1. *In vitro* characterisation of contemporary candidate strains for human challenge

A) Growth kinetics of candidate strains in RPMI-1640 supplemented with 2% Veggietone (filled) and Todd-Hewitt broth with 1% yeast extract (open). Means and standard deviation (SD) are representative of three separate experiments done in triplicate. **B)** Strain attachment and cellular invasion. Means and SD are from three separate experiments with triplicate wells. **C)** Capsular hyaluronic acid quantification. Mean and SD derived from a single experiment. **D)** Resistance of M75, M12, and 5448 to killing by human neutrophils, means and SD are from three separate experiments using different blood donors, with seven biological replicates. **E)** Strain lethality in a humanized plasminogen transgenic AlbPLG1 murine invasive disease model (n=10 for each strain).

Figure 2. Comparative genomics of M75 611024, M12 611025, M1T1 5448, and M1 CDC SS-496

(A) Circular schematic of GAS M75 611024 showing: GC plot (inner ring) with GC content above the genome average (black) and below (grey); predicted prophage sequences shown in red with associated prophage virulence determinants annotated; and relative position of predicted coding sequences on the forward strand (blue) and reverse strand (gold). **(B)** Unrooted maximum likelihood tree of 131 *emm75* strains from UK, US and France based on 1,046 SNPs relative to the M75 611024 reference genome. Tips of the tree are colour coded based on country of isolation. Location of genomes corresponding to M75 611024 and the completely sequenced *emm75* strains from France STAB090229 (CP020027); STAB120304 (CP020082) and STAB14018 (CP014542) are annotated. **(C)** Comparative BlastN analysis of M12 611025 (blue ring) and MGAS9429 (purple ring) relative to the *emm12* reference genome HKU16 (inner black circle). HKU16 GC content and GC skew is indicated in the inner ring while annotated around the outside is the genomic position of known HKU16 mobile genetic elements. **(D)** Maximum likelihood phylogenetic relationship of strain 611025 with 141 *emm12* *S. pyogenes* from other geographical regions based on 1,452 SNP sites from the core genome of the HKU16 reference genome. Tips of the tree are colour coded based on country of isolation of each isolate. Genomes from completely sequenced *emm12* strains MGAS9429 (CP000259), MGAS2096 (CP000261) and HKU16 (QMH11M0907901; AFRY01000001) are annotated. **(E)** Comparative BlastN analysis of CDC SS-496 and other GAS M1 reference genomes AP1, 5448 and MGAS5005 relative to the SF370 M1 GAS reference genome (inner black circle). **(F)** Mid-point rooted maximum likelihood phylogenetic relationship of M1 GAS reference genomes based on 780 SNP sites. Tips of the tree are annotated by strain name and colour coded by ring colour from (E) and annotated by strain name. Genomes belonging to SF370-like and MGAS5005-like lineages(32) are clustered by grey shading. Comparative BlastN analyses were generated using BRIG(42).