Characterisation of emergent toxigenic M1$_{UK}$ *Streptococcus pyogenes* and associated sublineages

Ho Kwong Li[1,2]*, Xiangyun Zhi[1,2]*, Ana Vieira[1,2], Harry J Whitwell[3,4], Amelia Schricker[3], Elita Jauneikaite[5,6], Hanqi Li[1], Ahmed Yosef[1], Ivan Andrew[7], Laurence Game[7], Claire E. Turner[8], Theresa Lamagni[5,9], Juliana Coelho[5,9], Shiranee Sriskandan[1,2,5¶]

[1]Department of Infectious Disease, Imperial College London, UK

[2]MRC Centre for Molecular Bacteriology & Infection (CMBI), Imperial College London, UK

[3]National Phenome Centre and Imperial Clinical Phenotyping Centre, Department of Metabolism, Digestion and Reproduction, Imperial College London, UK

[4]Section of Bioanalytical Chemistry, Division of Systems Medicine, Department of Metabolism, Digestion and Reproduction, Imperial College London, UK

[5]NIHR Health Protection Unit in Healthcare-associated Infection and Antimicrobial resistance, Imperial College London, UK

[6]School of Public Health, Imperial College London, UK

[7]Genomics Facility, UKRI-MRC London Institute for Medical Sciences (LMS), Imperial College London, UK

[8]The Florey Institute, School of Biosciences, University of Sheffield, UK

[9]Centre for Infections, UK Health Security Agency, London, UK

*contributed equally

¶for correspondence

Abstract

*Emm*1 *Streptococcus pyogenes* is a successful, globally-distributed epidemic clone that is regarded as inherently invasive. An *emm*1 sublineage, $M1_{UK,}$ that expresses increased SpeA toxin, was associated with increased scarlet fever and invasive infections in England in 2015/2016. Defined by 27 SNPs in the core genome, $M1_{UK}$ is now dominant in England. To more fully characterise $M1_{UK}$, we undertook comparative transcriptomic and proteomic analyses of $M1_{UK}$ and contemporary non-$M1_{UK}$ *emm*1 strains ($M1_{global}$).

Just seven genes were differentially expressed by $M1_{UK}$ compared with contemporary $M1_{global}$ strains. In addition to speA, five genes in the operon that includes glycerol dehydrogenase were upregulated in $M1_{UK}$ (gldA, mipB/talC, pflD, and pts system IIC and IIB components), while aquaporin (glpF2) was downregulated. $M1_{UK}$ strains have a stop codon in gldA. Deletion of the gldA gene in $M1_{global}$ abrogated glycerol dehydrogenase activity, and recapitulated upregulation of gene expression within the operon that includes gldA, consistent with a feedback effect.

Phylogenetic analysis identified two intermediate *emm*1 sublineages in England comprising 13/27 ($M1_{13SNPs}$) and 23/27 SNPs ($M1_{23SNPs}$) respectively, that had failed to expand in the population. Proteomic analysis of these four major phylogenetic *emm*1 groups highlighted sublineage-specific changes in carbohydrate metabolism, protein synthesis and protein processing; upregulation of SpeA was not observed in chemically-defined medium. In rich broth however, transcription and secretion of SpeA was upregulated ~10-fold in both $M1_{23SNPs}$ and $M1_{UK}$ sublineages, compared with $M1_{13SNPs}$ and $M1_{global}$.

We conclude that stepwise accumulation of SNPs led to the emergence of $M1_{UK}$. While increased expression of SpeA is a key indicator of $M1_{UK}$ and undoubtedly important, $M1_{UK}$ strains have outcompeted $M1_{23SNPs}$ and other *emm* types that produce similar or more superantigen toxin. We speculate that an accumulation of adaptive SNPs has contributed to a wider fitness advantage in $M1_{UK}$ on an inherently successful *emm*1 streptococcal background.

Data availability

RNAseq. All new RNAseq data are uploaded to the European Nucleotide Archive under project reference PRJEB58303

Genomic data. All genomes listed are available on the European Nucleotide Archive using accession numbers as listed in the appendix,

Proteomes. Proteomic data are available on FigShare 10.6084/m9.figshare.21777809 and will be uploaded to PRIDE

Impact Summary

Although the major *Streptococcus pyogenes* reservoir is in children with pharyngitis and skin infections, *S. pyogenes* can lead to rarer, invasive infections that are rapidly progressive and associated with high mortality and morbidity. *Emm*1 *S. pyogenes* strains are the single most frequent genotype to cause invasive infections in high income countries and are established worldwide as an epidemic clone. The $M1_{UK}$ *S. pyogenes emm*1 sublineage which is defined by 27 new SNPs in the core genome, and characterised by increased scarlet fever toxin SpeA production, emerged and rose to dominance over a period of 5-6 years since initial recognition, outcompeting other *emm*1 strains in England. Increased dominance of *emm*1 among invasive infections this winter, on a background of already-increased numbers of *S. pyogenes* infections, points to a key shift in host-pathogen interaction. We hypothesize that a combination of pathogen fitness, virulence, and host susceptibility have coalesced to account for the excess of circulating *S. pyogenes* and *emm*1 invasive infections. In this paper we undertake a systems-based evaluation of $M1_{UK}$ in comparison to older non-$M1_{UK}$ *emm*1 strains, and identify a number of pathways that are altered in addition to the previously-reported increased SpeA expression. The emergence of a new sublineage within an already virulent clone requires ongoing surveillance, and more detailed investigation of the likely mechanisms leading to increased fitness. The capacity of *S. pyogenes* to cause outbreaks at national scale highlights a potential need to consider strain-specific public health guidance, underlining the inherent virulence of this exclusively human pathogen.

Introduction

Emm1 Streptococcus pyogenes emerged in the 1980's and spread globally to become the leading cause of invasive *S. pyogenes* infection throughout the developed world (1,2). The lineage expanded following a recombination event that conferred increased expression of the streptolysin O (slo/nga) toxin locus, and was associated with specific phage content, including the phage encoding a superantigen, SpeA (1). More recently, during a period of increased scarlet fever activity in England, a new sublineage of *emm*1 *S. pyogenes* ($M1_{UK}$) was detected and found to have expanded (3). These strains were strongly associated with not only sore throats and scarlet fever, but also increases in invasive infection (3). The earliest $M1_{UK}$ strain detected to date was in a collection of non-invasive isolates from London in 2010, while the first invasive strains were detected in England in 2012. By 2016, the $M1_{UK}$ sublineage represented around 80% of all invasive *emm*1 isolates in England (3); this rose to 91% by end of 2020 (4). Despite differing from older *emm*1 strains by just 27 core genome SNPs, the new sublineage was characterised by a ~ten-fold increase in expression and production of the superantigen SpeA. Since 2019, the $M1_{UK}$ lineage has been identified elsewhere in Europe and North America (5-7).

*Emm*1 strains are the single most dominant cause of invasive *S. pyogenes* infection. In this work, we set out to characterise the wider phenotype of the new sublineage $M1_{UK}$, and to compare $M1_{UK}$ strains with minor sublineages that appeared briefly as intermediates, although did not expand to the extent of $M1_{UK}$. We also examined natural mutants of $M1_{UK}$ and the minor sublineages that provide insight into the cost-benefit balance of the changes in this new highly successful group of *S. pyogenes* M1T1 strains.

Methods

Bacterial strains. S. pyogenes strains used are outlined in Supplementary Tables S1 and S2; strains stored in 20% glycerol were streaked onto Columbia blood agar (CBA) prior to broth culture. *S. pyogenes* were cultured in Todd Hewitt Broth (THB, Oxoid, UK) or chemically defined medium (CDM) comprising iron, phosphate, magnesium, manganese, sodium acetate, calcium, sodium bicarbonate, L-cysteine, bases, vitamins and amino acids, with or without different carbon sources (Supplementary Table S3) at 37^0C in 5% CO_2.

RNA-sequencing. RNA was extracted from four different *S. pyogenes* strains from each lineage (Supplementary Table S1), cultured in THB for six hours corresponding to late-log growth phase using methods as previously described (3). RNA sequencing of $M1_{global}$ and $M1_{UK}$ RNA was undertaken by Novogene, Cambridge, UK and by the MRC London Institute of Medical Sciences (LMS). Data (deposited in project PRJEB58303) were analyzed according to published guidelines (8). Briefly, read quality was accessed using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), filtered and trimmed using trimmomatic (9), and mapped against the MGAS5005 (CP000017) reference genome using bowtie2 (10) with the highest sensitivity options. The resulting alignments were converted to sorted BAM files using vcftools (11). Initial visualizations of the sequencing mapping were performed using the Integrative Genomics Viewer (IGV) (12) including confirmation of *gldA* disruption. The mapped RNA-seq reads were then transformed into a fragment count per gene per sample using HT-seq (13) package. Exploratory data analysis (Principal component analysis and Heatmap of sample-to-sample distances) of the RNAseq data was implemented and plotted using DESeq2 package (14). Differential expression analysis in each dataset was performed using three different R packages (DESeq2 (14), EdgeR (15) and limma (https://bioconductor.riken.jp/packages/3.0/bioc/html/limma.html)) with a $\log_2$fold change of 0.5 and p-adj < 0.05 for $M1_{global}$ vs. $M1_{UK}$, and a $\log_2$fold change of 1 and p-adj < 0.05 for $M1_{H1488\Delta gldA}$ vs. $M1_{H1488}$. Only genes DE in two of the three softwares used were considered as

DE genes and used in analysis. Prophage regions were predicted using phaster (16), and curated by visual assessment and blast alignment.

Gene transcription studies. Specific transcript abundance was evaluated by quantitative RT-PCR using a plasmid standard for each gene and compared with *proS*. For the gldA operon plasmid standard, single amplicons were amplified to create a single linear insert (ProS-gldA-mipB-pflD-pts subunit IIC) that was TA-cloned into plasmid PCR2.1. For *glpF2* and *speA,* the plasmid standard comprised just *glpF2* and *proS*, or *speA* and *proS* respectively. cDNA synthesis from *S. pyogenes* RNA was undertaken as previously reported prior to RT-PCR (3); primers are listed in Supplementary Table S4. Comparisons were subject to analysis in GraphPad Prism v9. Non-parametric (Mann Whitney U) or t-tests were used; $p<0.05$ was considered significant.

Genetic manipulation. The gene encoding gldA was mutated by allelic replacement using the suicide vector pUCMUT. A 541 bp fragment upstream of *gldA* gene was amplified (forward primer: *5′*-AGC*GAATTC*TCGCCCAAGATTACGAAGG-*3′*, reverse primer: *5′*-GG*GGTACC*CGTTGAACTCCTTTATCTGTGATT-*3′*) incorporating 5′ EcoRI and 3′KpnI restriction sites, and cloned into the suicide vector pUCMUT to produce vector $pUCMUT_{gldAUP}$. A 532 bp fragment downstream of the *gldA* gene was amplified (forward primer: *5′*-AA*CTGCAG*CTATTGCAGAGCTGGTGCT-*3′*, reverse primer: *5′* -ACGC*GTCGAC*CGAGTCGATAGGCTAACC-*3′*) incorporating 5′ PstI and 3′ SalI restriction sites and cloned into PstI/SalI digested $pUCMUT_{gldAUP}$ to create $pUCMUTgldA_{KO}$. The construct was introduced into *S. pyogenes* $M1_{global}$ strains H1488 ($M1_{H1488}$), and BHS162 ($M1_{BHS162}$) by electroporation and crossed into the chromosome by homologous recombination. Transformants were selected using kanamycin (400µg/ml). Successful disruption of the gldA gene and insertion of the kanamycin resistance cassette was confirmed

by PCR, DNA sequencing and whole genome sequencing of mutated strains $M1_{H1488DgldA}$ and $M1_{BHS162DgldA}$ (isolate identifiers H1589 and H2151 respectively).

GldA activity assay. Cell free extracts were prepared from bacteria cultured overnight in chemically defined medium containing either 0.5% glucose of 0.5% glycerol to A_{600} of 0.6-0.7 (or as close to this as feasible). Bacteria were washed, centrifuged and kept on ice for 1h within an anaerobic jar, then suspended in 10 mM Tris buffer, pH 9. cells were disrupted by agitation in three 60-second bursts with 0.1 mm glass beads. Beads were allowed to settle, and the supernatant fluid centrifuged in an Eppendorf microcentrifuge for 30 seconds at 14,000 x g. GldA results in conversion of glycerol + NAD to dihydroxyacetone + NADH $+H^+$. GldA activity was derived from the increase in absorbance at 340 nm resulting from the reduction of NAD; one unit reduces one micromole of NAD per minute at 25°C and pH 10.0 under the conditions specific (17).

Phylogenetic analysis. *Emm1* genomes used in phylogenetic analysis were from the UK and are listed in the Supplementary Information. These comprise sequenced non-invasive *emm*1 isolates (n=139) (3); sequenced invasive *emm*1 isolates (n=40) from two studies (3,18); 64 invasive *emm*1 isolates from the British Society for Antimicrobial Chemotherapy (BSAC) collection (19); and 23 *emm*1 isolates from a hospital outbreak study (20). Two new *emm*1 genomes were sequenced from an additional outbreak and are available from the European nucleotide archive (Project PRJEB36425: ERS4267588 and ERS4267589). Raw reads were trimmed using trimmomatic version 0.36 (9) with the default parameters. The SNP calling was performed by mapping trimmed reads to the complete *emm*1.0 MGAS5005 (CP000017) reference genome using Snippy v4.6.0 (https://github.com/tseemann/snippy), with a minimum coverage of 10, minimum fraction of 0.9, and minimum vcf variant call quality of 100. Gubbins version 2.4.1 (21) was used to identify and remove recombinant regions from the resulting full genome alignment file. A maximum likelihood phylogeny was created from core SNPs using the general time-reversible (GTR) model of nucleotide substitution with the gamma distributed

rate heterogeneity implemented in FastTree v2.1.10-4 (22) Phylogenetic trees were visualized using FigTree v1.4.2 (http://tree.bio.ed.ac.uk/software/figtree/) and Microreact (https://microreact.org/showcase) and edited using INKSCAPE (https://inkscape.org/pt/).

SpeA expression. Semi-quantitative analysis of SpeA expression by *S. pyogenes* cultured in THB for 16 hours, was undertaken using cell-free culture supernatants concentrated 5X using Amicon filters, western blotting using a rabbit polyclonal antibody to SpeA and comparison with standard concentrations of rSpeA expressed from *Escherichia coli* as previously reported (3).

Proteomics. In pilot studies, five strains were randomly selected from each of $M1_{UK}$ or $M1_{global}$, cultured in 50mL chemically defined medium (CDM) to A_{600} 1.2-1.4 (6 hours) at 37°C with 5% CO_2, then cytosolic, cell wall, and supernatant fractions prepared for proteomic analysis. For proteomic analysis of sublineages, strains were randomly selected from five phylogenetic branches within each lineage (5 strains per sublineage, 4 sublineages in total). The supernatant fraction was removed, syringe filtered (Minisart 0.2uM filter, Sartorius, Germany) and proteins precipitated overnight at 4°C using 10% Trichloroacetic acid precipitation. Cell-wall- proteins were extracted from the bacterial pellet using 1mL of 30% raffinose, centrifugation at 10,000 RPM for 5 min, followed by resuspending the pellet in 1mL of cell wall extraction buffer (960µL 30% Raffinose, 10µL 1M Tris-HCl pH8, 10µL of 10kU/mL mutanolysin, 10µL 100mg/mL lysozyme, and 10µL protease inhibitor cocktail III (Avantor VWR, USA), followed by incubation at 37°C for 3 hours with occasional turning, and then aspiration of cell wall extract supernatant after centrifugation at 13,000 RPM for 10 minutes. The residual cytosolic fraction was further mechanically lysed via bead beating for 3 cycles for 45 seconds (Lysing Matrix B from MP Bio, USA). The samples of each cellular fraction then underwent centrifugal concentration using 3kDa filters and buffer exchanged (Amicon Ultra-15, Millipore, USA) with 50mM Tris buffer at pH8. The samples were then submitted to the

Proteomics Facility of the National Phenome Centre (London, UK) for LC MSe (Data to be deposited in PRIDE; currently deposited in FigShare). Precipitated samples were dissolved in 8M urea, 100mM ammonium bicarbonate (AmBic) by sonicating for 10 minutes in a water bath. Total protein was determined in all samples by Protein Assay (Protein Assay II, BioRad) according to the manufacturer's instructions. 20µg of protein was digested by the addition of 40mM chloracetamide, 10mM TCEP (Bondbreaker, ThermoScientific) and 0.2µg of trypsin in 100mM AmBic. Proteins in 8M urea were diluted to 1M urea prior to the addition of trypsin and all samples left overnight at 37°C. Desalting was performed by acidifying samples to 0.5% trifluroacetic acid (TFA) and adding them to a pre-equilibrated uElution HLB desalting plate (Waters), washing (3x100µl) with 0.5% TFA and eluding with 80% acetonitrile (3x50µl). All washes were drawn through the plate under vacuum. Desalted peptides were dried completely at 45°C in a vacuum-centrifuge.

For mass-spectrometry analysis, proteins were dissolved in 0.1% formic acid by sonicating in a water bath for 10 minutes. 0.5µg of peptides were analysed by LC-HDMSE (M-class UHPLC (Waters), Synaptic G2S (Waters)). Data was searched and processed using Progenesis QI for Proteomics.

Differentially expressed proteins with a fold change threshold of $\log_2$ 1.5 (p value threshold 0.05) were visualized on volcano plots. Enrichment analysis and protein-protein interactions were performed using STRING (https://string-db.org/), a database able to predict direct (physical) and indirect (functional) associations based on collected data across a range of experimental and *in silico* protein interactions. Proteins with a percentage identity higher than 90% and a "combined interaction score" higher than 0.7 were used to create a protein network in which the interaction between two proteins was inferred based on the information available in the STRING database and colour coded accordingly.

Results

Transcriptome of $M1_{UK}$ S. pyogenes

When comparing broth-cultured $M1_{UK}$ and $M1_{global}$ *S. pyogenes*, significant differential expression of just seven genes was observed (Table 1). As expected, transcription of SpeA was upregulated in all $M1_{UK}$ strains compared with other $M1_{global}$ strains; increased speA transcription by $M1_{UK}$ has previously been confirmed by RT-qPCR of RNA from 135 *emm*1 isolates (3). Unexpectedly, transcription of glpF2, a putative aquaporin (identified as Spy1573 in *emm*1 reference strain MGAS5005), was markedly downregulated in $M1_{UK}$ strains. Bioinformatic analysis of the *S. pyogenes* aquaporin gene glpF2 demonstrated similarity to the glpF3 family of *Lactobacillus plantarum* reported to be associated with both glycerol and water, but also dihydroxyacetone (DHA) transport (23).

The remaining five differentially-expressed transcripts that were upregulated in $M1_{UK}$ represented consecutive open reading frames in an apparent operon that includes glycerol dehydrogenase (gldA), pyruvate formate lyase (pflD), and a transaldolase-like protein (talC or mipB) as well as PTS system IIC and IIB components, annotated as cellobiose-specific.

Table 1. Differentially expressed genes comparing three $M1_{UK}$ and three $M1_{global}$ strains

Gene ID	Gene Name	Description	Average log2foldchange	Average padj	Strand
M5005_Spy0996	speA2	enterotoxin	2.361	3.731 E-09	+
M5005_Spy1573	glpF.2	glycerol uptake facilitator protein	-2.423	5.143E-09	-
M5005_Spy1742	mipB	transaldolase	1.043	0.0002	-
M5005_Spy1741	gldA	glycerol dehydrogenase	1.024	0.0006	
M5005_Spy1743	pflD	formate acetyltransferase	0.963	0.0007	
M5005_Spy1744	NA	PTS system, cellobiose-specific IIC component	0.653	0.0061	-
M5005_Spy1745	NA	PTS system, cellobiose-specific IIB component	0.749	0.0204	

GldA operon.

A single SNP in the glycerol dehydrogenase gene *gldA* among all $M1_{UK}$ strains is known to introduce a premature stop codon at position 175 of the 362 residue enzyme and is predicted to result in a truncated protein with abrogated enzyme activity (3). GldA is the final open reading frame in the sequence of genes that was found to be differentially expressed (Figure 1A). Differential expression of genes comprising the apparent operon was confirmed using RT-qPCR (Figure 1B-E). Transcription of the aquaporin gene was evaluated in three strains from each lineage, and although non significant, there was a 2-fold reduction in transcription in $M1_{UK}$. (Supplementary Figure S1)

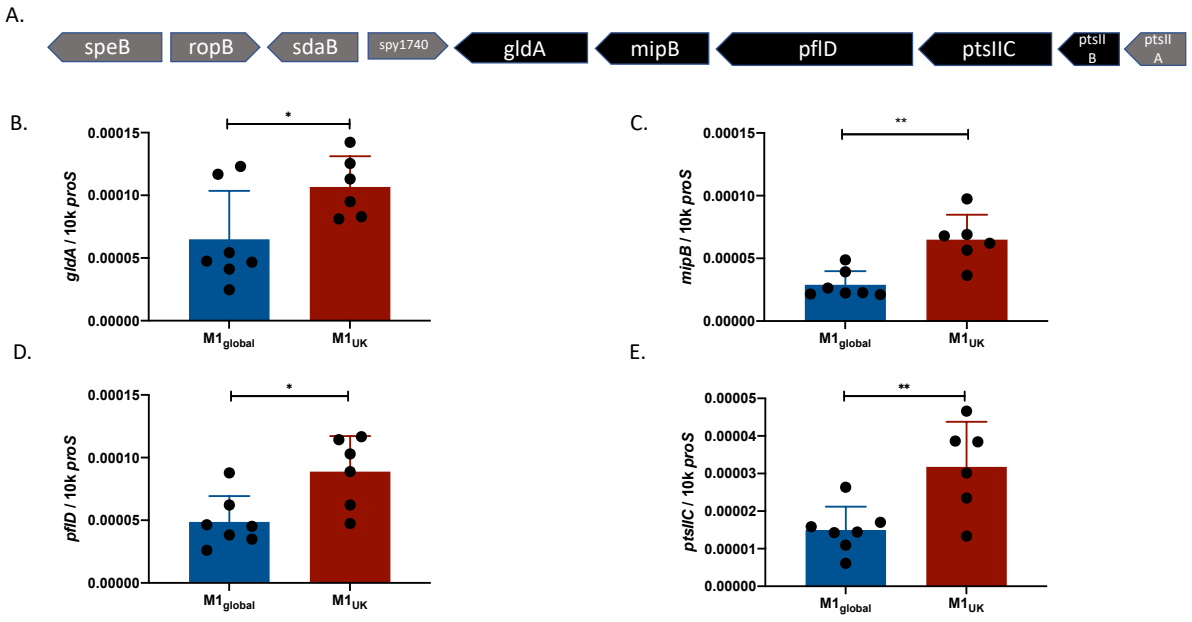


Figure 1. The genes within the *pflD-mipB-gldA* operon are upregulated in $M1_{UK}$. Five adjacent genes were found to be upregulated in $M1_{UK}$ compared to $M1_{global}$ (A). Genes upregulated in RNAseq are shown in black and include two components of a PTS system annotated as a PTS system (cellobiose) subunits IIC and IIB. Quantitative real time PCR using RNA from $M1_{global}$ (n=7) and $M1_{UK}$ (n=6) strains indicating transcription of gldA (B); mipB, also known as talC (C); pflD (D); and PTS subunit IIC (E). Data points (black dots) represent individual strains tested as technical triplicates and expressed as copies per 10,000 copies proS. Error bars show SD of the mean.**p<0.01 using unpaired t-test; *p<0.05.

We hypothesised that the loss of GldA enzyme activity may in some way feedback on transcription of the adjacent PTS subunit EII genes, as well as *mipB* and *pflD*. To determine the impact of isolated loss of GldA function in *S. pyogenes*, *gldA* was disrupted through allelic replacement in $M1_{global}$ strain $M1_{H1488}$ to create $M1_{H1488\Delta gldA}$. A GldA enzyme activity assay was undertaken, in the presence of glycerol and glucose, to confirm that enzyme function was present in the parent strain, but abrogated in the mutant (Figure 2A-B); this was replicated using a second pair of isogenic $M1_{global}$ strains ($M1_{BHS162}$ and $M1_{BHS162\Delta gldA}$). By comparison, $M1_{UK}$ strain BHS581 demonstrated barely detectable gldA activity, similar to the knockouts. RNA from $M1_{H1488}$ and the isogenic M1 $_{H1488\Delta gldA}$ was subject to RNAseq to compare the wider transcriptome of *S. pyogenes* in the absence of a functional gldA gene. Surprisingly there were almost no changes in the transcriptome except in the genes of the putative 'gldA' operon; deletion of *gldA* abrogated transcription of *gldA* as expected, but was associated with a clear increase in transcription of *pflD*, *mipB*, and the adjacent PTS system cellobiose-specific IIC genes. (Table 2). Upregulation of two adjacent genes Spy0123 and Spy0124 (including sloR) was also observed.

Significant downregulation of gldA transcription, and upregulation of the adjacent genes was confirmed by RTqPCR (Figure 2 C- F). Taken together, the data suggested that loss of gldA activity led to upregulation of the entire operon that is concerned with metabolism of dihydroxyacetone (DHA), fructose 1, 6, phosphate, and pyruvate. *S. pyogenes* has been reported to use a number of carbon sources, however, under conditions where *emm1 S. pyogenes* grew well in CDM supplemented with glucose, we were unable to demonstrate any growth in CDM supplemented with glycerol alone, consistent with other reports (24) (not shown). Informatic analysis of publicly available genomes from a range of bacterial species demonstrated remarkable conservation of the genes and organisation of this region in all members of the streptococcaciae. Whereas other bacterial species possessed the three genes mipB (also annotated as talC, for transaldolase), pflD and gldA, the organisation of genes differed widely (Figure 2G).

307

308 Table 2. RNAseq comparison of gldA-mutant *S. pyogenes* and parent strain

Gene ID	Gene Name	Description	Average log2foldchange *	Average padj	Strand
M5005_Spy0008	divIC	cell division protein	-1.0026	0.032	+
M5005_Spy0123	NA	translation initiation inhibitor	1.152	0.016	+
M5005_Spy0124	sloR	transcriptional regulator	1.250	0.0009	
M5005_Spy1166	NA	hypothetical protein	-1.094	0.0001	-
M5005_Spy1258	NA	putative cytosolic protein	-1.361	0.019	-
M5005_Spy1541	NA	hypothetical protein	-1.030	0.018	-
M5005_Spy1741	gldA	glycerol dehydrogenase	-9.212	1.111E-05	-
M5005_Spy1742	mipB	transaldolase	1.366838263	1.91E-03	
M5005_Spy1743	pflD	formate acetyltransferase	1.393501703	0.0019	
M5005_Spy1744	NA	PTS system cellobiose-specific IIC	1.03639459	0.0036	-

309 *Comparison is made between $M1_{H1488\Delta gldA}$ and parent strain $M1_{H1488}$; only genes differentially
310 expressed by at least log_2 fold value of 1.0 are shown, $p<0.05$. Genes from the same predicted
311 operon are shaded in grey.
312
313

314

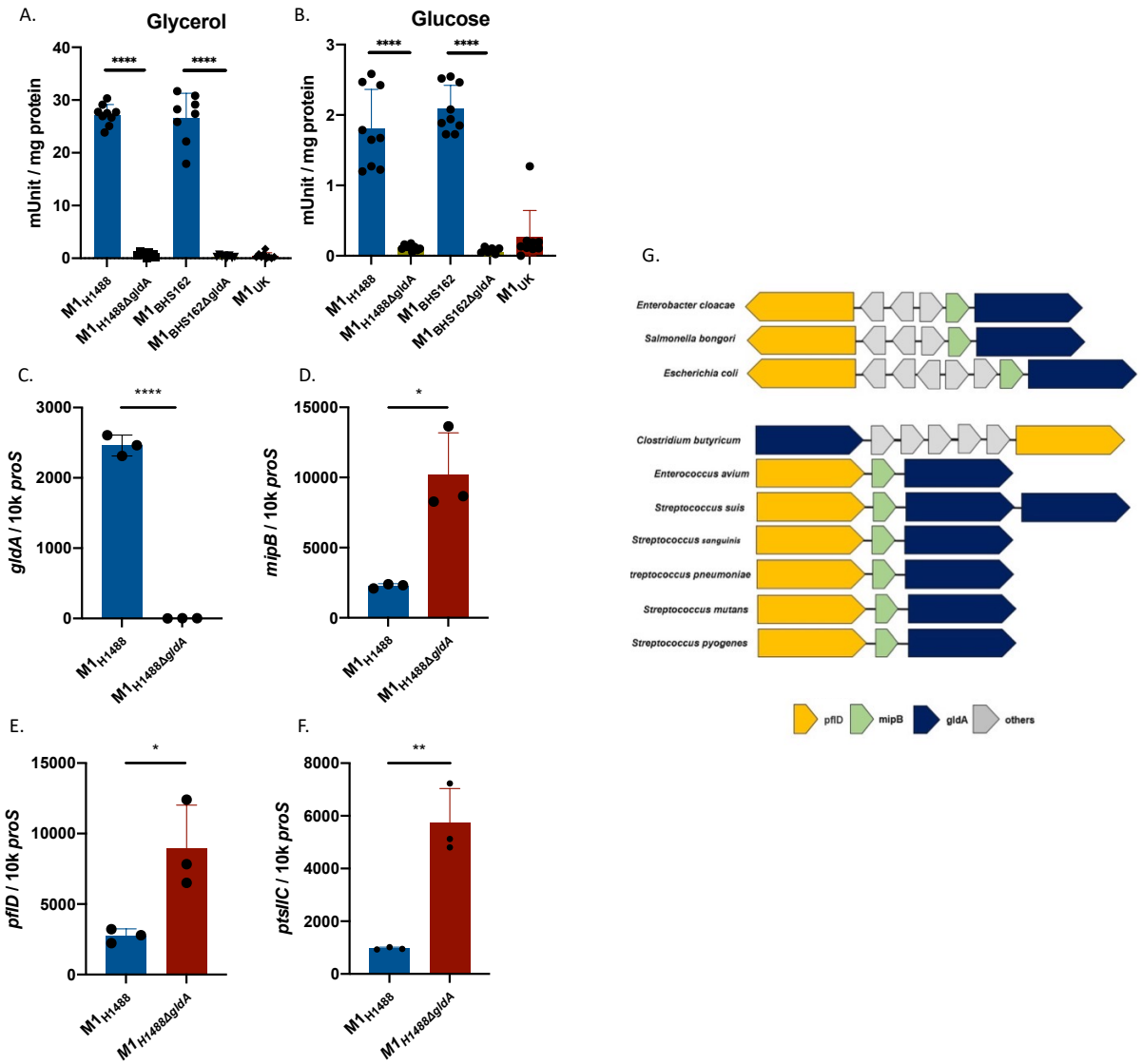


Figure 2. Loss of gldA function is associated with upregulated expression of adjacent genes. Glycerol dehydrogenase activity in *S. pyogenes* $M1_{global}$ is abrogated following inactivation of the gldA gene, to the level observed in $M1_{UK}$. Activity in the parent strain was greatest when cultured in the presence of glycerol (A) and less in glucose (B). Data show 8 or 9 individual reactions for each strain. Deletion of the gldA gene resulted in markedly reduced expression of gldA, (C), upregulated transcription of mipB (D); pflD (E); and PTS component IIC (F). Data show three biological replicates per strain. Error bars show SD of the mean *p<0.05; **p<0.01; ****p<0.0001 using Mann-Whitney U (A-B) & unpaired t-test (C-F). (G) Arrangement of pflD, mipB, gldA genes in different bacterial species.

Intermediate sublineages of emm1 S. pyogenes

$M1_{UK}$ strains are distinguished from older *emm*1 strains by the presence of 27SNPs (3) (Table 3). Although a number of additional indels are common in $M1_{UK}$, only the 27SNPs define the new lineage. When analysing genomes from *S. pyogenes* strains isolated in the United Kingdom, we identified small numbers of strains with either 13 of the 27SNPs, or 23 of the 27SNPs (3). All *emm*1 sublineages bar $M1_{global}$ possessed three SNPs in the transcriptional regulator RofA, however the gldA stop codon is present only in strains with 23SNPs or 27SNPs. (Table 3). We analysed our original non-invasive *S. pyogenes* WGS alongside other sequenced UK *emm*1 strains (Supplementary Table 2) and enriched for sublineages by including 10 invasive isolates from each of the following groups; $M1_{global}$; $M1_{13SNPs}$; $M1_{23SNPs}$; $M1_{UK}$. (Figure 3). As reported before, the earliest $M1_{UK}$ strain identified was 2010, however the earliest $M1_{13SNPs}$ strain was 2005, from the BSAC collection.

Table 3. SNPs in sublineages.

Position in MGAS5005	Gene locus	Gene	Product	S/NS	Ref	SNP	$M1_{13\ SNPs}$	$M1_{19SNPs}$¶	$M1_{23SNPs}$	$M1_{27\ SNPs}$‡
115646	M5005_Spy0106	rofA	Transcriptional regulator	NS	C	T	T	T	T	T
116162	M5005_Spy0106	rofA	Transcriptional regulator	NS	A	C	C	C	C	C
116163	M5005_Spy0106	rofA	Transcriptional regulator	NS	C	A	A	A	A	A
250832	M5005_Spy0243		ABC transporter-associated protein	S	T	C	T	T	C	C
513254	M5005_Spy0525		galactose-6-phosphate isomerase LacB	NS	G	T	T	T	T	T
528360	Intergenic		-	-	A	T	A	T	T	T
563631	M5005_Spy0566	sagE	streptolysin S putative self-immunity protein	NS	G	A	G	G	A	A
613633	M5005_Spy0609		phosphoglycerol transferase	NS	T	C	C	C	C	C
626494	M5005_Spy0623		methyltransferase	S	G	A	A	A	A	A
661707	M5005_Spy0656	trmD	tRNA (guanine-N(1)-)-methyltransferase	NS	G	A	G	A	A	A
730823	M5005_Spy0727	recJ	single-stranded-DNA-specific exonuclease	NS	C	T	T	T	T	T
784467	M5005_Spy0779		putative membrane spanning protein	S	T	C	T	T	C	C
819098	M5005_Spy0825	murB	UDP-N-acetylenolpyruvoylglucosamine reductase	NS	G	A	G	A	A	A
923079	M5005_Spy0933		putative NADH-dependent flavin oxidoreductase	NS	G	A	G	A	A	A
942633	M5005_Spy0951	pstB	phosphate transport ATP-binding protein	NS	G	T	G	G	G	T
983438	Intergenic/within ssrA		-	-	G	C	G	G	C	C
1082253	M5005_Spy1108	metK2	S-adenosylmethionine synthetase	NS	C	T	T	T	T	T
1238124	M5005_Spy1282	msrA	peptide methionine sulfoxide reductase	NS	G	A	A	A	A	A
1238673	M5005_Spy1283	tlpA	thiol:disulfide interchange protein	NS	G	A	A	A	A	A
1251193	M5005_Spy1293		hypothetical protein	NS	G	A	G	G	G	A
1373176	M5005_Spy1400		PTS system, galactose-specific IIB component	NS	C	A	C	C	C	A
1407497	M5005_Spy1439		portal protein	NS	C	T	T	T	T	T
1446116	M5005_Spy1490		3-oxoacyl-[acyl-carrier protein] reductase	S	C	T	T	T	T	T
1535209	Intergenic		-	-	A	G	A	A	A	G
1702540	M5005_Spy1714	gldA	glycerol dehydrogenase	STOP	C	T	C	T	T	T
1734749	M5005_Spy1772		glutamate formimidoyltransferase	NS	G	A	A	A	A	A
1828734	M5005_Spy1860		putative membrane spanning protein	NS	G	A	G	A	A	A

¶Single strain with 19 of the 27SNPs that characterize $M1_{UK}$ (not a sublineage); ‡Lineage with 27SNPs is equivalent to $M1_{UK}$

SpeA expression by sublineages.

Previous comparison had demonstrated ~10-fold greater speA gene transcription by non-invasive $M1_{UK}$ isolates compared to non-invasive $M1_{global}$ strains (3); we first established that SpeA protein expression was similarly elevated in the same large panel of non-invasive isolates (Supplementary Figure S2). There was an indication that SpeA expression was not increased in a small number of strains from intermediate lineages. To better understand the impact of the step-wise changes in SNP content, we examined SpeA gene transcription and protein expression in a new set of strains. To include sufficient numbers of intermediate sublineage isolates, we used 40 strains from a larger national collection of invasive *emm*1 *S. pyogenes* that had been submitted to the reference laboratory and were previously sequenced (3, 18).

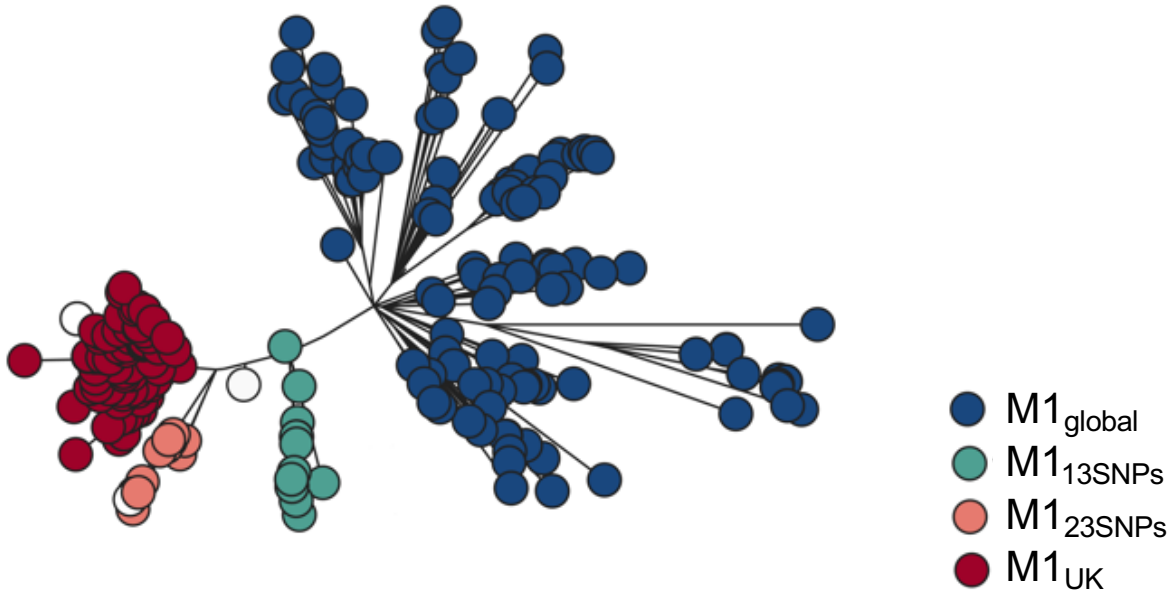


Figure 3. $M1_{UK}$, $M1_{global}$ and two intermediate sublineages Maximum likelihood phylogenetic tree constructed from core single-nucleotide polymorphisms (without recombination regions) of 269 invasive and non-invasive *emm*1 *S. pyogenes* strains representative of four main groups ($M1_{global}$, $M1_{13SNPs}$, $M1_{23SNPs}$, $M1_{UK}$). The phylogenetic tree is coloured as described in the legend. White bubbles represent isogenic strains from two distinct outbreaks with 26 and

22 SNPs, respectively and one invasive strain with 19 SNPs. Strains used in the phylogenetic tree are listed in supplementary table S2.

SpeA transcription was low in all $M1_{global}$ and $M1_{13SNPs}$ strains, except for the occasional strain with a mutation in covRS, a two component system regulator known to suppress virulence factors, but which can undergo mutation to confer a more invasive phenotype in *emm*1 and other *S. pyogenes* strains. In contrast, transcription of SpeA was high in all invasive strains with 23 or 27SNPs (Figure 4A). Likewise, SpeA protein production differed markedly between the sublineages; again SpeA production was greatest in all invasive strains with 23 or 27SNPs and was hard to detect in all $M1_{global}$ and $M1_{13SNPs}$ (Figure 4B). Indeed, the amount of SpeA produced routinely by $M1_{UK}$ strains was similar to that produced by $M1_{global}$ strains with mutations in CovRS, that is known to repress SpeA in *emm1* (25). We did not detect a difference in expression of other virulence factors such as SpyCEP, SPEB, or M protein, in broth culture (not shown). We concluded that the genetic changes required for basal increased SpeA expression in $M1_{UK}$ resided in $M1_{23SNPs}$ but not $M1_{13SNPs}$.

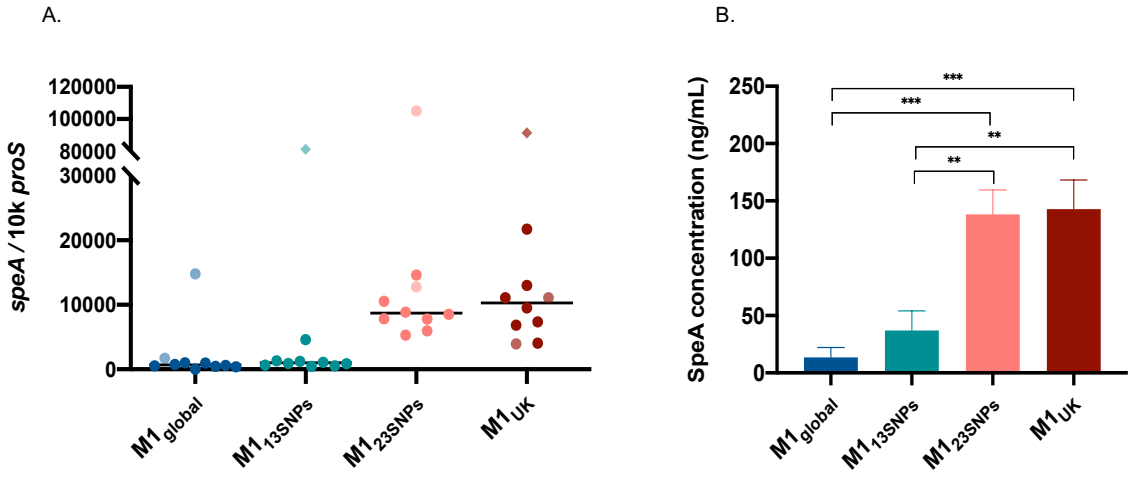


Figure 4. SpeA expression is increased in $M1_{23SNPs}$ and $M1_{UK}$ sublineages. SpeA transcription (A) using 10 strains from each sublineage is shown (total n=40). Each dot represents a single strain with lighter shading indicating the presence of *covRS* mutation (all outliers). Two isolates

possess both *covRS* and *rgg4* mutations (diamond shape). Solid line represents the median. There was no statistically significant difference between the sublineages in speA transcription, largely related to the outlying covRS mutants in each sublineage. Excluding isolates with *covRS* mutation a difference was observed between $M1_{global}$ and $M1_{23SNPs}$ or $M1_{UK}$ ($p<0.0001$), and a difference between between $M1_{13SNPs}$ and $M1_{23SNPs}$ or $M1_{UK}$. ($p=0.0002$). SpeA protein expression (B) comprising 40 isolates (10 in each sublineage) inclusive of *covRS* mutations. Bar chart shows mean, and SEM. Multiple comparisons test made using one-way ANOVA (Tukey's).

Isogenic isolates that differed by just single SNPs were available from two outbreak settings. Interestingly, in both settings, a single isolate was identified wherein a single SNP from the 27 SNPs that define $M1_{UK}$ reverted to wild type. In one daycare outbreak, a non invasive isolate exhibited only 26 of the 27SNPs but was otherwise identical to an invasive isolate from the same cluster; in this case, the SNP in trmD, a tRNA (guanine-N(1)-)-methyltransferase, had reverted to wildtype. This isolate made as much SpeA as the isolate with 27SNPs.

In a separate hospital outbreak associated with a fatal case of invasive infection caused by the $M1_{23SNPs}$ sublineage, (20), one isolate from a healthcare worker was identical to 5 other isolates in the cluster, bar one single SNP. This single SNP represented one of the 23SNPs but is present in both $M1_{13SNPs}$ and $M1_{23SNPs}$, a phage portal protein (Spy1439). This isolate also produced the same amount of SpeA as the parent $M1_{23SNPs}$ strain, demonstrating the SNPs that were dispensible for increased SpeA expression.

Review of published UK *emm*1 genome sequences (19) identified a single strain with 19 of the 27SNPs among *emm*1 bloodstream isolates. Unlike the sublineage that possessed 23SNPs, this $M1_{19SNPs}$ strain did not produce detectable quantities of SpeA, pointing to an influential role for the four SNPs that differentiate $M1_{19snp}$ and the $M1_{23SNP}$ sublineage in SpeA expression. Of these four SNPs, two were synonymous SNPs and felt to be unlikely to affect phenotype; one was a non-synonymous SNP in sagE; while the final change was a SNP that appeared to be intergenic in annotated *emm*1 *S. pyogenes* genomes, but lies within the start

406 of the tmRNA ssrA (26) upstream of the phage insertion and start site of SpeA (Spy0996 in
407 MGAS5005). RNAseq read abundance in this region did not show a difference between
408 $M1_{global}$ and $M1_{UK}$ strains, with the exception of the gene encoding SpeA. Abundance of reads
409 in the 'paratox' (Spy0995) gene, which is transcribed on the opposite strand to SpeA, was
410 increased in two of four $M1_{UK}$ strains, but this finding was not consistent.
411
412 *Proteomic analysis of S. pyogenes emm1 sublineages*
413 To screen for lineage-specific difference in proteomes, cell wall, cytosolic, and supernatant
414 fractions of five randomly selected $M1_{UK}$ strains were compared with five $M1_{global}$ following
415 culture in CDM. Though SpeA was detected, a significant difference between $M1_{UK}$ and
416 $M1_{global}$ supernatants was not observed when strains were cultured in CDM, in contrast to
417 results (reported above) in Todd Hewitt broth, pointing to a major role for specific culture
418 conditions in induction of SpeA. CDM supernatant from $M1_{UK}$ strains demonstrated increased
419 phage-encoded DNase (spd3), acid phosphatase (lppC), and a DNA binding protein. CDM
420 supernatant from $M1_{global}$ however demonstrated increased phosphoglycerate mutase and
421 phosphofructokinase, both of which are linked to carbohydrate utilisation pathways in *S.*
422 *pyogenes* (Figure 5A & Supplementary Figure S3A) (27). Cell wall fractions demonstrated a
423 small number of proteins that were differentially expressed in $M1_{UK}$ strains. These included a
424 more than 3-fold increase in PrsA2 (Spy1732, AAZ52350.1), which controls protein folding
425 and may operate at the ExPortal (28), and almost 2-fold increases in GAPDH and the 10kDa
426 chaperonin groS. (Figure 5B and Supplementary Figure S3B).
427 In $M1_{global}$ strains, a number of cytosolic proteins were increased compared to $M1_{UK}$, including
428 adjacent genes Spy0438 (rnc, Ribonuclease III) and Spy0439 (smc) as well as mecA, an
429 adapter protein and negative regulator of competence; the greatest fold changes were
430 however seen in infC, encoding Initiation Factor 3, satD, and a number of proteins linked to
431 protein secretion (secA), maintenance of ribosomal function and RNA. String analysis
432 highlighted a number of links between phosphoentomutase (DeoB), protein synthesis
433 pathways (gidA), and acid tolerance (satD). (Figure 5C and Supplementary Figure S3C).

To screen for differences between all four sublineages (two intermediate and two major sublineages), five strains from each intermediate sublineage ($M1_{13SNPs}$ and $M1_{23SNPs}$) were randomly selected from appropriate phylogenetic branches as well as 5 new strains from each of $M1_{UK}$ and $M1_{global}$. Fresh cytosolic fractions of the four phylogenetic groups (20 strains) were prepared and subject to new proteomic analysis. The data were then analysed by comparing groups in different combinations. When cytosolic preparations from all four sublineages were compared with one another, fruR expression by $M1_{23SNPs}$ was increased in comparison to other lineages, and lowest in $M1_{global,}$ while a network of ribosomal proteins was increased in $M1_{13snps}$ (Supplementary Figure S4A). Comparison of cytosolic preparations from new $M1_{UK}$ and $M1_{global}$ strains did not identify the same DE features seen previously; however a negative regulator of competence, mecA, again was increased in $M1_{global}$ strains although only by 1.3-fold (Supplementary Figure 4B). The biggest fold change was a 3.6-fold upregulation of fruR and 5.87-fold upregulation of mur1.2 a potential autolysin (adjacent to a PTS fructose-specific IIABC system and fruR) in $M1_{UK}$ (Figure 6A). As $M1_{UK}$ and $M1_{23SNPs}$ strains had demonstrated comparable SpeA production, we proceeded to determine if there was commonality between these two sublineages by comparing cytosolic proteomes of [$M1_{UK}$ and $M1_{23SNPs}$] with [$M1_{global}$ and $M1_{13SNPs}$]. NtpA and B, a V type ATPase, was increased in [$M1_{global}$ and $M1_{13SNPs}$]; genes linked to ligase activity were found to be enriched in string analysis and highest in [$M1_{global}$ and $M1_{13SNPs}$](Figure 6B and Supplementary Figure S4C). When considering $M1_{global}$ compared with all other 3 'new' lineages, carbohydrate metabolism genes were further highlighted, specifically Phosphotransferase system (PTS) and disaccharide metabolic processes (Supplementary figure S4D). FruR was four-fold increased in non-$M1_{global}$ strains, with increased FruA in $M1_{global}$ strains; a similar pattern was seen for LacR and lacA1/lacA2 (Figure 6C, and Supplementary Figure S4D). A glutamate formiminotransferase (MGAS5005_Spy1772) was also increased in $M1_{global}$ strains compared with non-$M1_{global}$. Finally, comparing cytosolic proteins in $M1_{UK}$ with all other lineages, just one protein was

clearly upregulated in $M1_{UK}$, and this was Spy0848 (ppnK), an ATP-NAD kinase. (Figure 6D and Supplementary Figure S4E).

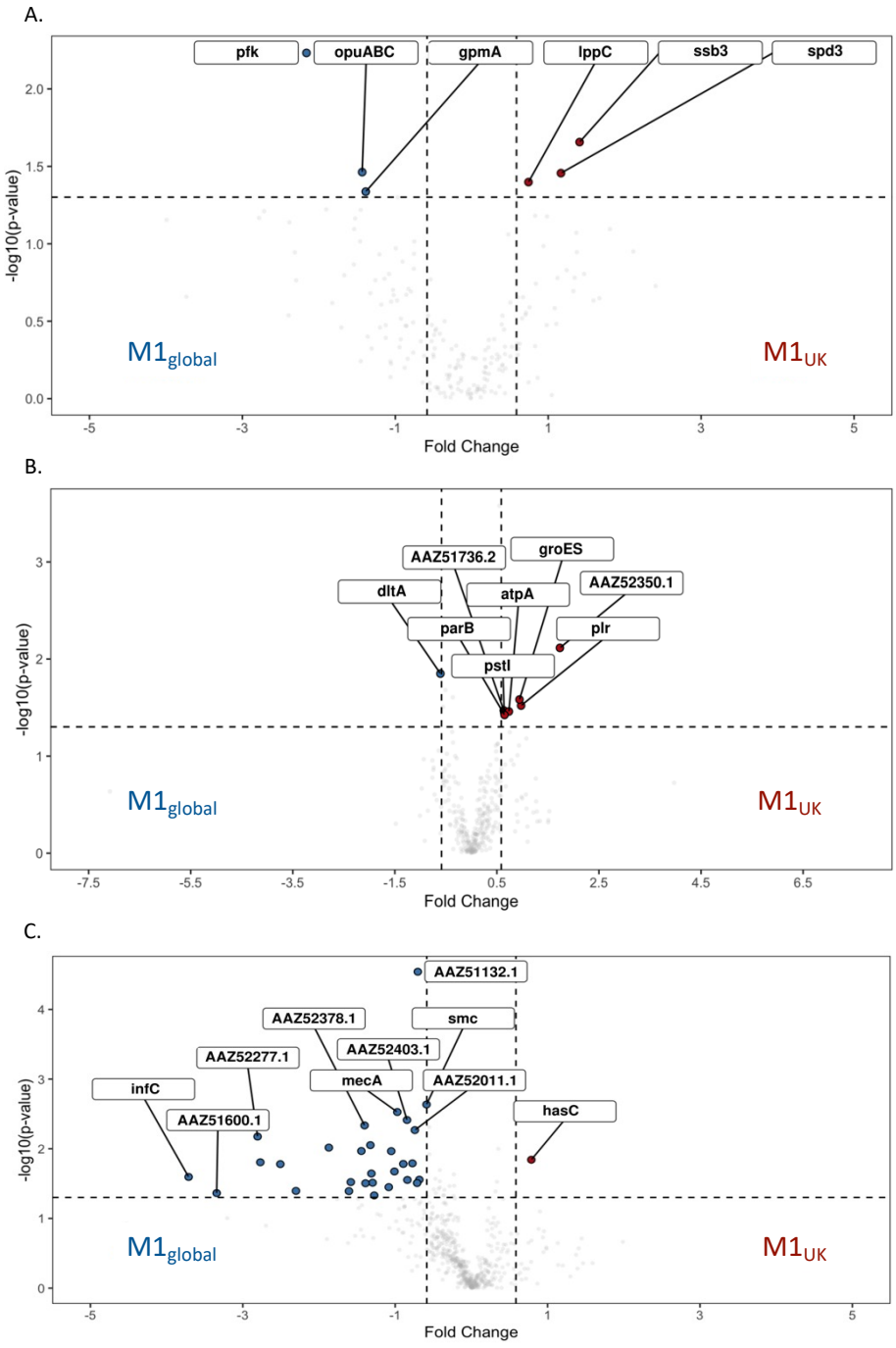
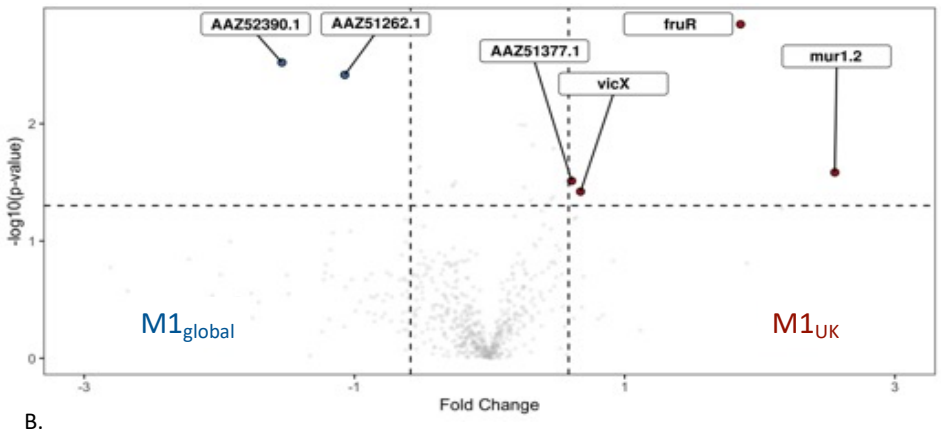
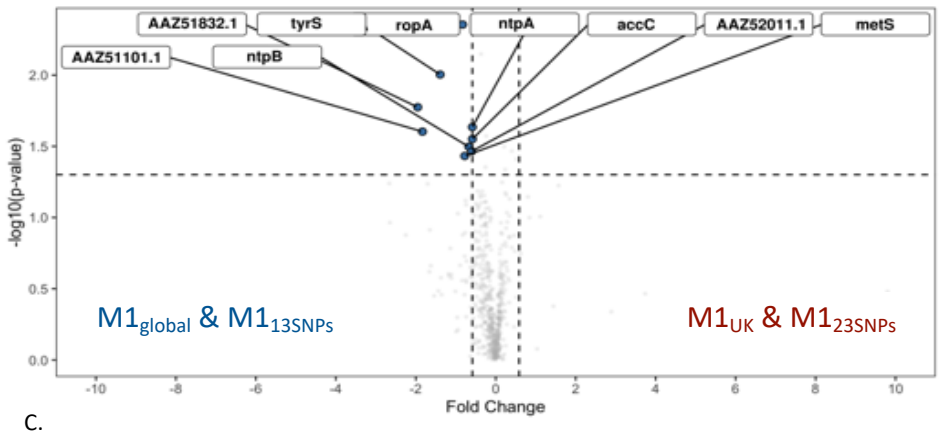


Figure 5: Volcano plots comparing proteins differentially expressed by $M1_{UK}$ vs. $M1_{global}$ cultured in CDM. Specific fractions examined were Supernatant (A); Cell wall (B); and Cytosol (C). Proteins upregulated in $M1_{UK}$ are shown on the right in red. Those upregulated in $M1_{global}$ are shown in the left in blue.

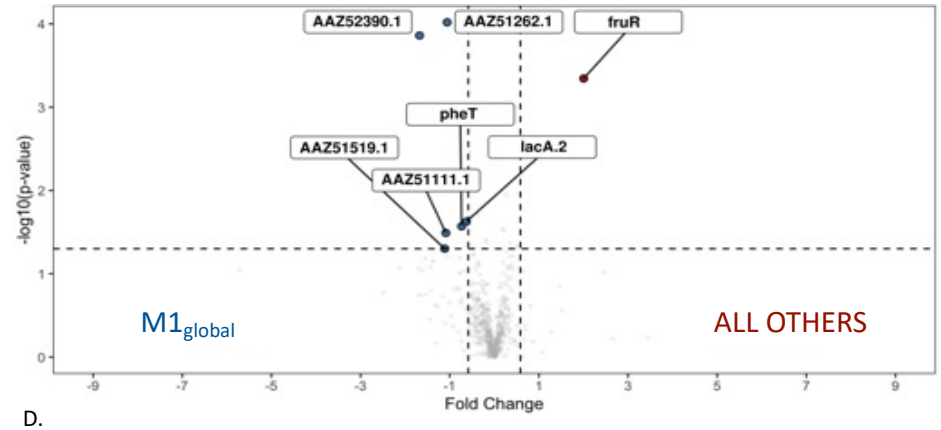
A.



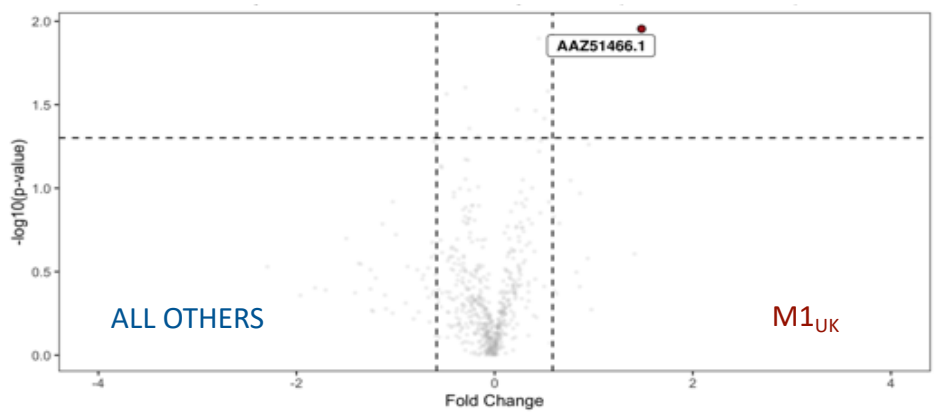
B.



C.



D.



470

Figure 6: Volcano plots comparing proteins differentially expressed by different pairings of $M1_{global}$, $M1_{13SNPs}$, $M1_{23SNPs}$, and $M1_{UK}$ cultured in CDM. Cytosolic fractions only were compared as follows (A) $M1_{UK}$ vs. $M1_{global}$ (B) [$M1_{UK}$ + $M1_{23SNPs}$] vs. [$M1_{global}$ + $M1_{13SNPs}$] (C) All other sublineages vs. $M1_{global}$, and (D) $M1_{UK}$ vs. all other sublineages.

Discussion

$M1_{UK}$ is now the dominant *S. pyogenes emm*1 lineage in the United Kingdom, having expanded during an earlier upsurge in scarlet fever 2014-2016 (3,4). Importantly, *emm*1 strains are inherently invasive and represent the single most frequent *emm* type to cause invasive infections in the UK (29, 30). As such, any change in the *emm*1 lineage that results in increased fitness is of relevance to public health. In this first systematic study to characterize the changes in $M1_{UK}$ and its associated lineages, we have confirmed the SpeA over-expression phenotype, and demonstrated that increased SpeA production is restricted to $M1_{UK}$ and an increasingly rare sublineage $M1_{23SNPs}$. The phenotype is manifest in broth culture but not CDM. $M1_{UK}$ is defined by just 27 SNPs in the core genome including 3 SNP in rofA and a stop codon in glycerol dehydrogenase, gldA. RNA sequencing demonstrated a difference in expression of the operon that includes gldA and a phosphotransfer system (PTS) EIIC and B which represents a combined phosphate and sugar transporter, pointing to a potential shift in metabolism in the new lineage. This was accompanied by a sharp reduction in transcripts for the aquaporin gene glpF2. Preliminary proteomic analysis of strains by sublineage identified altered carbohydrate pathways related to fructose that may well be important.

Alterations in expression of gldA, mipB, pflD, and the adjacent PTS system impact on the glycolytic Embden-Meyerhof Parnas pathway (27) which, in *S. pyogenes,* relies on the phospho-enolpyruvate PTS system for acquisition of sugars other than glucose, and for transfer of phosphate ions required for carbon catabolite repression and gene regulation (31).

The results indicate that both the stop codon in gldA present in $M1_{UK}$ and allelic replacement of gldA impact glycerol dehydrogenase activity and result in upregulation of the other genes in the operon, mipB and pflD. These are involved in the glycolytic pathway required for generation and metabolism of pyruvate from glucose; the changes in carbohydrate metabolism are supported by preliminary proteomic findings that indicate alterations in fructose pathways. Interestingly, increased transcription in this operon was accompanied by increased transcription of an adjacent PTS components IIC and IIB when comparing $M1_{UK}$ with $M1_{global}$, and following experimental deletion of gldA. This PTS is annotated as being a cellobiose transporter; systematic experimental disruption of PTS EII systems in *S. pyogenes* has not shown a key role for these genes, however the precise sugar transported is not known (31).

The role of gldA in *S. pyogenes* has not been experimentally examined previously; gldA is reported to catalyse the conversion of glycerol to dihydroxyacetone (DHA) under microaerophilic or anaerobic conditions, however it is clear that gldA may undertake a reverse role, which is to catalyse DHA to glycerol. This may be of importance since an absence of gldA activity may lead to a build-up of DHA, which when converted to methylglyoxal can be toxic (32). The upregulation of the PTS system is of interest since these are recognised to be key players in a phosphorelay process that maintains central carbon catabolite repression of many virulence systems in *S. pyogenes* (31, 33).

The marked ~8-10 fold downregulation of aquaporin glpF2 (Spy1573) transcription, was unexpected but may represent an adaptation to the metabolic changes that have arisen in $M1_{UK}$. There are few reports, if any, relating to glpF2 in *S. pyogenes* but there is evidence of functional links to the pflD containing operon in enterococci (34). Notably one of the intergenic SNPs that defines $M1_{UK}$ is 39 bp from the start of the Spy1573 gene, though the significance of this is not yet known. Aquaporins are membrane proteins that function as channels for water and other uncharged solutes in all forms of life. While mostly considered as channels for water or glycerol in bacteria, potentially important to osmoregulation, aquaporins also can function as a channel for DHA. Indeed, there are similarities between glpF2 of streptococci and glpf3

of *Lactobacillus plantarum* that points to a possibility for action as a channel for DHA or similar molecules (23). Research undertaken in related *Lactococcus lactis* has also identified marked downregulation of glpF2 following osmotic stress (35). Taken together it would seem that the downregulation of glpF2 may be a necessary adaptation for $M1_{UK}$ *S. pyogenes*, although it may also confer as-yet unknown advantage.

The upregulation of SpeA expression by $M1_{UK}$ is clearly of importance to virulence particularly in interaction with the human host, and those who have not yet mounted an immune response to the secreted toxins of this species. There is good evidence that superantigens such as SpeA undermine development of the adaptive host immune response to *S. pyogenes* through promotion of a dysregulated T cell response associated with B cell death (36, 37). SpeA has also been shown to promote carriage of *S. pyogenes* in the nasopharynx of transgenic mice (38). To date, the expression of SpeA has only been measured in broth culture and we do not know if the upregulation in $M1_{UK}$ might differ *in vivo*. Recent epidemiological studies found a high (44%) secondary infection rate in schoolchildren and household contacts of a case of scarlet fever caused by $M1_{UK}$, pointing to a potential transmission advantage compared with other *S. pyogenes* lineages (39). We identified sublineage-specific altered expression of SpeA allowing us to highlight the genetic changes likely to account for this. Importantly, the three SNPs identified in the major regulator RofA do not alone account for the SpeA phenotype since these SNPs are present in $M1_{13SNPs}$ although we cannot discount a role for these in the wider success of this lineage. While the genetic changes required for increased SpeA expression do not reside in $M1_{13SNPs}$, they do reside in $M1_{23SNPs}$, and strains with reversion of single SNPs pointed to a potential key role for the SNP in ssrA in SpeA upregulation. The amount of SpeA made by $M1_{UK}$ and $M1_{23SNPs}$ was augmented to the level of $M1_{global}$ covRS mutants yet presumably without the fitness burden of covRS mutation that might impair pharyngeal carriage (40).

There are a number of limitations to our study. Firstly, investigation of the gldA operon is in its early stages; it is possible that the stop codon mutation in gldA confers an additional phenotype that is not recapitulated by gldA gene deletion, while the metabolic pathways that include gldA, mipB, and pflD are not fully understood. The roles of glpF2 and the PTS EII system that is upregulated are also not understood; any role in transfer of DHA for example has not been experimentally addressed. The proteomic studies are preliminary and require both validation and repetition using richer media, but have provided a rationale for further study of the role of sugar metabolism in *emm*1 *S. pyogenes*. Finally, the role of specific SNPs would necessarily require experimental proof.

Several European countries are, at the time of writing, affected by epidemic waves of invasive *S. pyogenes* disease, notably in England, where the leading cause of invasive infection is *emm*1 underlining the importance of understanding pathogenicity and transmission (30, 41). Importantly however, despite the enhanced production of SpeA by $M1_{23SNPs}$, this intermediate sublineage did not expand in the manner seen for $M1_{UK}$ in England, and was not detected at all in a 2020 systematic evaluation of >300 invasive *emm*1 isolates from England (4). This suggests that the fitness of $M1_{UK}$ has required the additional acquisition of four further SNPs. These include three non-synonymous SNPs in phosphate transport ATP binding protein, pstB; a PTS galactose-specific IIB component gene; a hypothetical protein; as well as the intergenic SNP adjacent to glpF2. The amount of SpeA produced by $M1_{UK}$ strains remains an order of magnitude lower than the amount produced by the historic *emm*1 strain NCTC8198 (42). Despite this, the new $M1_{UK}$ lineage has outcompeted $M1_{23SNPs}$ and has replaced older strains suggesting that the added fitness of $M1_{UK}$ may lie beyond the ability to make SpeA.

Funding

This work was supported by the UK Medical Research Council (grant number MR/P022669/1); a UKRI (MRC) Research Training Fellowship (HKL); and the NIHR Imperial Biomedical Research Centre.

Acknowledgements

SS acknowledges support from the UK National Institute for Health Research (NIHR) Health Protection Unit in Healthcare Associated Infections and Antimicrobial Resistance and the NIHR Imperial Biomedical Research Centre.

The authors acknowledge the support of colleagues in the UKRI-MRC LMS and National Phenome Centre in facilitating sequencing and proteomics.

Disclosures

None

REFERENCES

1. 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. Evolutionary pathway to increased virulence and epidemic group A Streptococcus disease derived from 3,615 genome sequences. Proc Natl Acad Sci U S A. 2014 Apr 29;111(17):E1768-76. doi: 10.1073/pnas.1403138111.

2. Sumby, P., Porcella, S.F., Madrigal, A.G., Barbian, K.D., Virtaneva, K., Ricklefs, S.M., Sturdevant, D.E., Graham, M.R., Vuopio-Varkila, J., Hoe, N.P. and Musser, J.M., 2005. Evolutionary origin and emergence of a highly successful clone of serotype M1 group A Streptococcus involved multiple horizontal gene transfer events. *The Journal of infectious diseases*, *192*(5), pp.771-782.

3. Lynskey NN, Jauneikaite E, Li HK, Zhi X, Turner CE, Mosavie M, Pearson M, Asai M, Lobkowicz L, Chow JY, Parkhill J, Lamagni T, Chalker VJ, Sriskandan S. Emergence of dominant toxigenic M1T1 Streptococcus pyogenes clone during increased scarlet fever activity in England: a population-based molecular epidemiological study. Lancet Infect Dis. 2019 Nov;19(11):1209-1218. doi: 10.1016/S1473-3099(19)30446-3.

4. Zhi X, Li HK, Li H, Loboda Z, Charles S, Vieira A, Huse KK, Jauneikaite E, Coelho J, Lamagni T, Sriskandan S. Ongoing emergence of $M1_{UK}$ lineage among invasive group A streptococcus isolates in 2020 and use of allele-specific PCR https://biorxiv.org/cgi/content/short/2022.12.18.520871v1.

5. Rümke LW, de Gier B, Vestjens SMT, van der Ende A, van Sorge NM, Vlaminckx BJM, Witteveen S, van Santen M, Schouls LM, Kuijper EJ. Dominance of $M1_{UK}$ clade among Dutch M1 Streptococcus pyogenes. Lancet Infect Dis. 2020 May;20(5):539-540. doi: 10.1016/S1473-3099(20)30278-4.

6. Demczuk W, Martin I, Domingo FR, MacDonald D, Mulvey MR. Identification of Streptococcus pyogenes $M1_{UK}$ clone in Canada. Lancet Infect Dis. 2019 Dec;19(12):1284-1285. doi: 10.1016/S1473-3099(19)30622-X.

7. Li Y, Nanduri SA, Van Beneden CA, Beall BW. $M1_{UK}$ lineage in invasive group A streptococcus isolates from the USA. Lancet Infect Dis. 2020 May;20(5):538-539. doi: 10.1016/S1473-3099(20)30279-6.

8. Conesa A, Madrigal P, Tarazona S, Gomez-Cabrero D, Cervera A, McPherson A, Wojciech Szcześniak M, Gaffney DJ, Elo LL, Zhang X, Mortazavi A. A survey of best practices for RNA-seq data analysis. 2016, 17(13).

9. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30(15):2114-20. doi.org/10.1093/bioinformatics/btu170.

10. Clausen PT, Aarestrup FM, Lund O. Rapid and precise alignment of raw reads against redundant databases with KMA. BMC bioinformatics. 2018;19(1):1-8.

11. Danecek P, Auton A, Abecasis G, Albers CA, Banks E, DePristo MA, et al. The variant call format and VCFtools. Bioinformatics. 2011;27(15):2156-8.

12. Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, et al. Integrative genomics viewer. Nature biotechnology. 2011;29(1):24-6.

13. Anders S, Pyl PT, Huber W. HTSeq—a Python framework to work with high-throughput sequencing data. bioinformatics. 2015;31(2):166-9.

14. Love M, Anders S, Huber W. Differential analysis of count data–the DESeq2 package. Genome Biol. 2014;15(550):10-1186.

15. Robinson MD, McCarthy DJ, Smyth GK. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics. 2010;26(1):139-40.

16. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. Nucleic acids research. 2016;44(W1):W16-W21.

17. Lin EC, Magasanik B. The activation of glycerol dehydrogenase from Aerobacter aerogenes by monovalent cations. J Biol Chem. 1960 Jun;235:1820-3. PMID: 14417009.

18. 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 https://doi.org/10.7717/peerj.3226.

19. Turner CE, Holden MTG, Blane B, Horner C, Peacock SJ, Sriskandan S. The Emergence of Successful Streptococcus pyogenes Lineages through Convergent Pathways of Capsule Loss and Recombination Directing High Toxin Expression. mBio. 2019 Dec 10;10(6):e02521-19. doi: 10.1128/mBio.02521-19. PMID: 31822586; PMCID: PMC6904876.

20. Sharma H, Ong MR, Ready D, Coelho J, Groves N, Chalker V, Warren S. Real-time whole genome sequencing to control a Streptococcus pyogenes outbreak at a national orthopaedic hospital. J Hosp Infect. 2019 Sep;103(1):21-26. doi: 10.1016/j.jhin.2019.07.003.

21. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using Gubbins. Nucleic Acids Res 43:e15.

22. Price, M.N., Dehal, P.S., and Arkin, A.P. (2009) FastTree: Computing Large Minimum-Evolution Trees with Profiles instead of a Distance Matrix. Molecular Biology and Evolution 26:1641-1650, doi:10.1093/molbev/msp077.

23. Bienert GP, Desguin B, Chaumont F, Hols P. Channel-mediated lactic acid transport: a novel function for aquaglyceroporins in bacteria. Biochem J. 2013 Sep 15;454(3):559-70. doi: 10.1042/BJ20130388.

24. Sundar GS, Islam E, Braza RD, Silver AB, Le Breton Y, McIver KS. Route of Glucose Uptake in the Group a Streptococcus Impacts SLS-Mediated Hemolysis and Survival in Human Blood. Front Cell Infect Microbiol. 2018 Mar 14;8:71. doi: 10.3389/fcimb.2018.00071.

25. Sumby P, Whitney AR, Graviss EA, DeLeo FR, Musser JM. Genome-wide analysis of group a streptococci reveals a mutation that modulates global phenotype and disease specificity. PLoS Pathog. 2006 Jan;2(1):e5. doi: 10.1371/journal.ppat.0020005.

679 26. Rosinski-Chupin I, Sauvage E, Fouet A, Poyart C, Glaser P. Conserved and specific
680 features of Streptococcus pyogenes and Streptococcus agalactiae transcriptional landscapes.
681 BMC Genomics. 2019 Mar 22;20(1):236. doi: 10.1186/s12864-019-5613-5.
682
683 27. Pancholi V, Caparon M. *Streptococcus pyogenes* Metabolism. 2016 Feb 10. In: Ferretti
684 JJ, Stevens DL, Fischetti VA, editors. *Streptococcus pyogenes*: Basic Biology to Clinical
685 Manifestations [Internet]. Oklahoma City (OK): University of Oklahoma Health Sciences
686 Center; 2016–. PMID: 26866220.
687
688 28. Wu ZY, Campeau A, Liu CH, Gonzalez DJ, Yamaguchi M, Kawabata S, Lu CH, Lai CY,
689 Chiu HC, Chang YC. Unique virulence role of post-translocational chaperone PrsA in
690 shaping *Streptococcus pyogenes* secretome. Virulence. 2021 Dec;12(1):2633-2647. doi:
691 10.1080/21505594.2021.1982501.
692
693 29. Nelson GE, Pondo T, Toews KA, Farley MM, Lindegren ML, Lynfield R, Aragon D, Zansky
694 SM, Watt JP, Cieslak PR, Angeles K, Harrison LH, Petit S, Beall B, Van Beneden CA.
695 Epidemiology of Invasive Group A Streptococcal Infections in the United States, 2005-2012.
696 Clin Infect Dis. 2016 Aug 15;63(4):478-86. doi: 10.1093/cid/ciw248.
697
698 30. UK Health Security Agency Group A Streptococcal infections: second update on seasonal
699 activity in England 2022 to 2023 Updated 15 December 2022.
700 https://www.gov.uk/government/publications/group-a-streptococcal-infections-activity-during-
701 the-2022-to-2023-season/group-a-streptococcal-infections-second-update-on-seasonal-
702 activity-in-england-2022-to-2023 (Accessed 16th December 2022).
703
704 31. Sundar GS, Islam E, Gera K, Le Breton Y, McIver KS. A PTS EII mutant library in Group
705 A Streptococcus identifies a promiscuous man-family PTS transporter influencing SLS-
706 mediated hemolysis. Mol Microbiol. 2017 Feb;103(3):518-533. doi: 10.1111/mmi.13573.
707
708 32. Subedi KP, Kim I, Kim J, Min B, Park C. Role of GldA in dihydroxyacetone and
709 methylglyoxal metabolism of Escherichia coli K12. FEMS Microbiol Lett. 2008 Feb;279(2):180-
710 7. doi: 10.1111/j.1574-6968.2007.01032.x.
711
712 33. Rom JS, Hart MT, McIver KS. PRD-Containing Virulence Regulators (PCVRs) in
713 Pathogenic Bacteria. Front Cell Infect Microbiol. 2021 Oct 19;11:772874. doi:
714 10.3389/fcimb.2021.772874.
715

716 34. Doi Y, Ikegami Y. Pyruvate formate-lyase is essential for fumarate-independent anaerobic
717 glycerol utilization in the Enterococcus faecalis strain W11. J Bacteriol. 2014
718 Jul;196(13):2472-80. doi: 10.1128/JB.01512-14.
719
720 35. van der Meulen SB, de Jong A, Kok J. Early Transcriptome Response of *Lactococcus*
721 *lactis* to Environmental Stresses Reveals Differentially Expressed Small Regulatory RNAs and
722 tRNAs. Front Microbiol. 2017 Sep 14;8:1704. doi: 10.3389/fmicb.2017.01704.
723
724 36. Davies FJ, Olme C, Lynskey NN, Turner CE, Sriskandan S. Streptococcal superantigen-
725 induced expansion of human tonsil T cells leads to altered T follicular helper cell phenotype,
726 B cell death and reduced immunoglobulin release. Clin Exp Immunol. 2019 Jul;197(1):83-94.
727 doi: 10.1111/cei.13282.

728 37. Dan JM, Havenar-Daughton C, Kendric K, Al-Kolla R, Kaushik K, Rosales SL, Anderson
729 EL, LaRock CN, Vijayanand P, Seumois G, Layfield D, Cutress RI, Ottensmeier CH,
730 Lindestam Arlehamn CS, Sette A, Nizet V, Bothwell M, Brigger M, Crotty S. Recurrent group
731 A *Streptococcus* tonsillitis is an immunosusceptibility disease involving antibody deficiency
732 and aberrant T_{FH} cells. Sci Transl Med. 2019 Feb 6;11(478):eaau3776. doi:
733 10.1126/scitranslmed.aau3776.

734 38. Kasper KJ, Zeppa JJ, Wakabayashi AT, Xu SX, Mazzuca DM, Welch I, Baroja ML, Kotb
735 M, Cairns E, Cleary PP, Haeryfar SM, McCormick JK. Bacterial superantigens promote acute
736 nasopharyngeal infection by Streptococcus pyogenes in a human MHC Class II-dependent
737 manner. PLoS Pathog. 2014 May 29;10(5):e1004155. doi: 10.1371/journal.ppat.1004155.

738 39. Cordery R, Purba AK, Begum L, Mills E, Mosavie M, Vieira A, Jauneikaite E, Leung RCY,
739 Siggins MK, Ready D, Hoffman P, Lamagni T, Sriskandan S. Frequency of transmission,
740 asymptomatic shedding, and airborne spread of Streptococcus pyogenes in schoolchildren
741 exposed to scarlet fever: a prospective, longitudinal, multicohort, molecular epidemiological,
742 contact-tracing study in England, UK. Lancet Microbe. 2022 May;3(5):e366-e375. doi:
743 10.1016/S2666-5247(21)00332-3.

744 40. Alam FM, Turner CE, Smith K, Wiles S, Sriskandan S. Inactivation of the CovR/S virulence
745 regulator impairs infection in an improved murine model of Streptococcus pyogenes naso-
746 pharyngeal infection. PLoS One. 2013 Apr 25;8(4):e61655. doi:
747 10.1371/journal.pone.0061655.

748 41. World Health Organization, Increase in invasive Group A streptococcal infections among
749 children in Europe, including fatalities. https://www.who.int/europe/news/item/12-12-2022-
750 increase-in-invasive-group-a-streptococcal-infections-among-children-in-europe--including-
751 fatalities#:~:text=The%20observed%20increases%20reported%20to,during%20the%20COV
752 ID%2D19%20pandemic. (Accessed, 14th December 2022).

753 42. Sriskandan S, Moyes D, Buttery LK, Krausz T, Evans TJ, Polak J, Cohen J. Streptococcal
754 pyrogenic exotoxin A release, distribution, and role in a murine model of fasciitis and
755 multiorgan failure due to Streptococcus pyogenes. J Infect Dis. 1996 Jun;173(6):1399-407.
756 doi: 10.1093/infdis/173.6.1399.

757