

Insights gained from sequencing Australian non-invasive and invasive *Streptococcus pyogenes* isolates

Trent A.J. Butler^{1,*†}, Chloe Story^{2,†}, Emily Green¹, Kirsten M. Williamson³, Peter Newton², Frances Jenkins⁴, Hemalatha Varadhan¹ and Sebastiaan van Hal^{4,5}

Abstract

Epidemiological data have indicated that invasive infections caused by the Gram-positive cocci *Streptococcus pyogenes* (group A streptococcus, GAS) have increased in many Australian states over the past two decades. In July 2022, invasive GAS (iGAS) infections became nationally notifiable in Australia via public-health agencies. Surveillance for *S. pyogenes* infections has been sporadic within the state of New South Wales (NSW). This has led to a lack of genetic data on GAS strains in circulation, particularly for non-invasive infections, which are the leading cause of GAS's burden on the Australian healthcare system. To address this gap, we used whole-genome sequencing to analyse the genomes of 318 *S. pyogenes* isolates collected within two geographical regions of NSW. Invasive isolates were collected in 2007–2017, whilst non-invasive isolates were collected in 2019–2021. We found that at least 66 different *emm*-types were associated with clinical disease within NSW. There was no evidence of any Australian-specific clones in circulation. The M1_{UK} variant of the *emm1* global pandemic clone (M1_{global}) has been detected in our isolates from 2013 onwards. We detected antimicrobial-resistance genes (mainly *tetM*, *ermA* or *ermB* genes) in less than 10% of our 318 isolates, which were more commonly associated with non-invasive infections. Superantigen virulence gene carriage was reasonably proportionate between non-invasive and invasive infection isolates. Our study adds rich data on the genetic makeup of historical *S. pyogenes* infections within Australia. Ongoing surveillance of invasive and non-invasive GAS infections within NSW by whole-genome sequencing is warranted to inform on outbreaks, antimicrobial resistance and vaccine coverage.

DATA SUMMARY

Raw sequence reads have been deposited into the National Center for Biotechnology Information (NCBI) database under BioProject PRJNA996294. Accession numbers for each isolate and their associated metadata are provided in Table S1, available with the online version of this article. All bioinformatics tools used for data analysis have been reported in Methods.

INTRODUCTION

Streptococcus pyogenes (group A streptococcus, GAS) is a strict human pathogen with a wide variety of clinical manifestations, from self-limiting infections of the skin and throat causing impetigo and pharyngitis, respectively, to invasive infections including septicaemia, necrotizing fasciitis and streptococcal toxic shock syndrome [1]. The bacteria are transmitted from person to person, via droplet or direct contact from colonized or infected individuals. Specific M proteins located on the bacterial cell surface

Received 10 August 2023; Accepted 22 November 2023; Published 10 January 2024

Author affiliations: ¹Microbiology, NSW Health Pathology, John Hunter Hospital, New Lambton Heights, New South Wales, Australia; ²Microbiology, NSW Health Pathology, Wollongong Hospital, Wollongong, New South Wales, Australia; ³Hunter New England Population Health, Hunter New England Local Health District, Newcastle, New South Wales, Australia; ⁴Department of Infectious Diseases and Microbiology, NSW Health Pathology, Royal Prince Alfred Hospital, Sydney, New South Wales 2050, Australia; ⁵Central Clinical School, University of Sydney, Sydney, New South Wales 2006, Australia.

*Correspondence: Trent A.J. Butler, trent.butler@health.nsw.gov.au

Keywords: group A streptococcus.

Abbreviations: GAS, group A streptococcus; HNELHD, Hunter New England local health district; iGAS, invasive group A streptococcus; ISLHD, Illawarra Shoalhaven local health district; MLST, multilocus sequence typing; NCBI, National Center for Biotechnology Information; NSW, New South Wales; NSWHP, New South Wales Health Pathology; ST, sequence type.

†These authors contributed equally to this work

Data statement: All supporting data, code and protocols have been provided within the article or through supplementary data files. One supplementary table is available with the online version of this article.

001152 © 2024 Crown Copyright



This is an open-access article distributed under the terms of the Creative Commons Attribution License.

Impact Statement

Our study has described the molecular epidemiology of 136 non-invasive and 182 invasive *Streptococcus pyogenes* isolates collected from two communities in New South Wales (NSW), Australia. This work is vital as our healthcare system is encumbered by the breadth of clinical infection types and persistence of *S. pyogenes* within the Australian population. The results showed that at least 66 different *S. pyogenes* strains caused infections within NSW. There was substantial overlap in the *S. pyogenes* strains that caused non-invasive and invasive diseases. We found no evidence of a clone specific to the Australian continent. Notably, *emm*-types of global concern, including *emm1*, *emm3.1*, *emm4*, *emm12*, *emm28* and *emm89*, were associated with most invasive infections, whilst many *emm*-types were sporadically detected. To our knowledge, we provide the first data on the *emm1* global sub-lineage M1_{UK} within NSW. Fifty-eight per cent of our invasive disease isolates in 2017 were the M1_{UK} sub-lineage. This finding is important because other epidemiological studies have demonstrated that in 2017, there was a substantial increase in case numbers of invasive group A streptococcus (iGAS) diseases within the Hunter region of NSW, where these isolates were collected. We also found antimicrobial-resistance genes in less than 10 % of study isolates. These genes were mainly *tetM*, *ermA* or *ermB*. Since invasive *S. pyogenes* infections have become notifiable by public-health units in NSW in 2022, and prior studies on *S. pyogenes* within NSW are sparse, our data help to address critical public-health needs. It will be a valuable reference for future Australian studies investigating outbreaks or strain emergence, and for potential vaccination programmes.

mediate adherence and internalization of GAS into epithelial cells [2, 3]. Additional pathogen virulence gene expression, as well as resistance to host immune mechanisms, may result in microbial entry into sterile sites (e.g. blood, bone, joint fluid or deep tissue) causing invasive group A streptococcus (iGAS) infection [4, 5]. Patient risk factors for developing iGAS include older age, immune disorders, diabetes and intravenous drug use [6].

GAS infections are strongly associated with post-infection immune-mediated diseases, including acute rheumatic fever, rheumatic heart disease and post-streptococcal glomerulonephritis [7]. These diseases represent autoimmune states involving cross-reactive antibodies and T cell dysregulation with repeated episodes of GAS infection, compounding tissue injury [1, 8]. Social determinants of health strongly influence the risk of developing GAS infection and subsequent immune sequelae [9]. Due to the preventable nature of antecedent infections, the overall incidence of immune-mediated complications is low in high-income countries [10]. However, First Nations peoples are still disproportionately affected; with an estimated incidence of approximately 374 cases of acute rheumatic fever per 100000 persons per year in the Pacific and Indigenous Australia and New Zealand combined [10]. Although, the true overall incidence of GAS infection in Australia is unknown [7, 11, 12], the health impact is substantial with iGAS, rheumatic heart disease and GAS-associated kidney disease estimated to cost approximately AU\$185.1 million (£97.4 million; £1=AU\$1.9) and cause 110 deaths annually [13].

The incidence of GAS bacteraemia in the Hunter New England local health district (HNELHD) of New South Wales (NSW) has gradually increased over the last two decades, with a peak incidence of 9.00 cases per 100000 population in 2017, which was double the previous 5 year mean (4.18 cases per 100000 population per year for 2012–2016) [14]. Clinical cases of iGAS and/or laboratory isolation of GAS from sterile sites were made nationally notifiable through local Public Health Units in Australia in 2022 [15, 16]. This decision, in part, was made to better understand the dynamics of iGAS including the recent increases in Australia [7, 11, 12, 14], detect outbreaks and assess preventative strategies. Current control guidelines for iGAS within NSW include molecular typing for suspected outbreaks [16]. However, there remains a scarcity of reports on the molecular epidemiology of both invasive [17] and non-invasive GAS isolates found within NSW communities. Due to the dynamically changing number of GAS infections within NSW, updated reporting on strains in circulation is a public-health need. Whole-genome sequencing can be used to investigate GAS strains and obtain data on the carriage of genetic antimicrobial-resistance and virulence genes [18]. Furthermore, it is useful to monitor new infection sub-lineages occurring globally such as the recent M1_{UK} clone, which emerged from the parent global pandemic *emm1* (M1_{global}) clone [19]. To better understand GAS infections occurring within NSW and address public-health needs, we undertook sequencing of both invasive and non-invasive GAS isolates from two distinct hospital isolate collections within NSW.

METHODS

Isolate collections

A total of 318 group A streptococcal isolates from 318 patients (one isolate per patient) were retrieved for the study. Of these, 182 represented invasive isolates [sources include blood cultures ($n=171$), body fluids ($n=10$) and cerebrospinal fluid ($n=1$)] collected and stored between 2007 and 2017 (Fig. 1a), obtained from eleven public hospitals [serving a population of ~676 556 (2021)] within the Greater Newcastle and Hunter regions of the HNELHD. These hospitals were serviced by the NSW Health

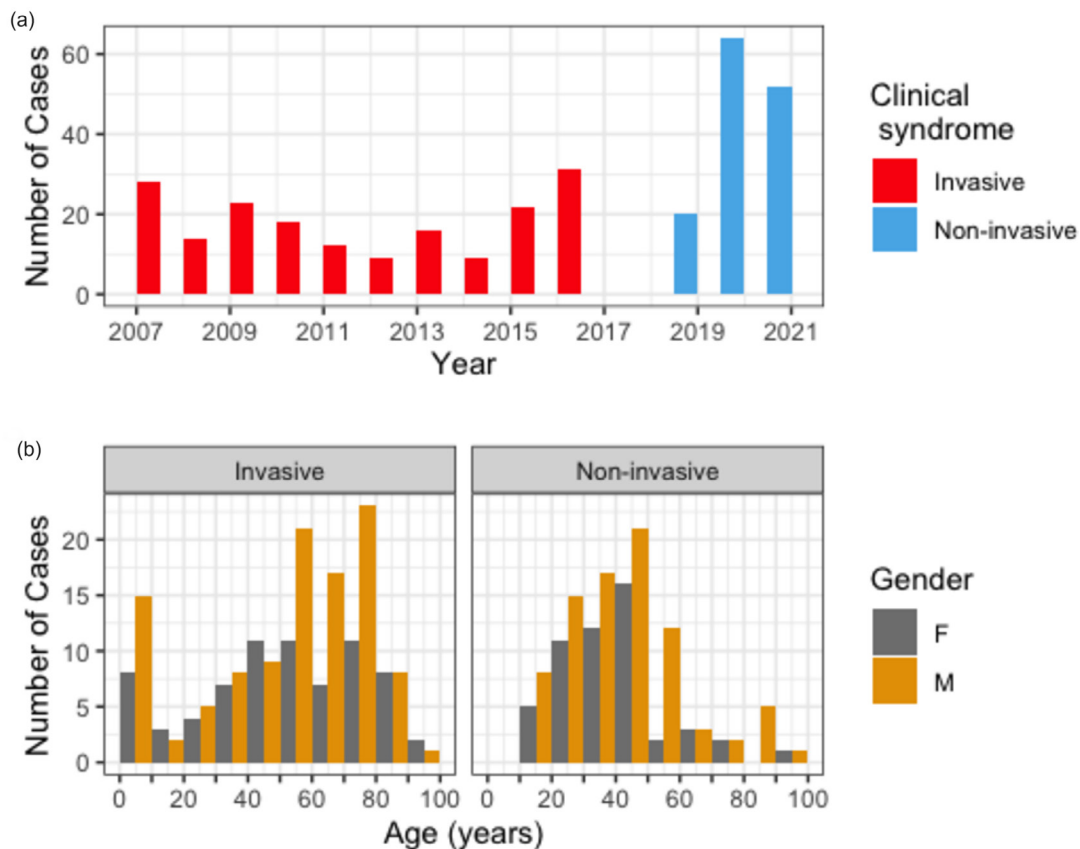


Fig. 1. Epidemiology characteristics of sequenced isolates. (a) The distribution by year of collection. Invasive isolates were collected only in years 2007 to 2017 from the Greater Newcastle and Hunter regions of NSW and are shown in red. Non-invasive isolates were collected only in years 2019 to 2021 from the ISLHD and are shown in blue. Note that GAS isolates were not collected as part of this study during 2018. (b) The age distribution grouped by gender of patients presenting with *S. pyogenes* infection split by source of infection. F, Female; M, male. Note that non-invasive isolates were not collected from patients under 18 years of age.

Pathology (NSWHP) John Hunter Hospital microbiology laboratory. Gender and age were known for 181 out of these 182 isolates. The remaining 136 isolates were obtained from patients presenting to Illawarra Shoalhaven local health district (ISLHD) hospitals [serving a population of ~404 000 people (2021)] with microbiology services performed at the NSWHP–Wollongong microbiology laboratory. These isolates represented non-invasive skin and throat isolates collected from patients presenting with cellulitis and pharyngitis, respectively, between 2019 and 2021 (Fig. 1a). Gender and age were known for all 136 of these isolates. Due to restrictions in the ethical approvals for this study, non-invasive isolates were collected only from people older than 18 years. Both laboratories are greater than 100 km from Metropolitan Sydney, cover metropolitan, inner regional and outer regional areas, service a lower socio-economic demographic, and have higher proportions of residents that identify as Aboriginal and/or Torres Strait Islander (5.9% in HNEHLD and 3.5% in ISLHD against a mean of 2.9% across NSW census data) [20]. Non-overlapping time frames occurred as result of isolate collection criteria at each of the laboratories.

All bacterial isolates preserved in sterile nutrient broth with 15% glycerol at -80°C were rejuvenated following subculture and overnight incubation on horse blood agar in 5% CO_2 at 35°C . GAS identity was confirmed using MALDI-TOF MS Biotyper (Bruker Daltronics) prior to DNA extraction.

Genomic DNA extraction and whole-genome sequencing

DNA from each specimen was extracted following a lysis step. Extraction was undertaken for invasive isolates using Zymbiomics DNA microprep kits (Zymbiomics) and for non-invasive isolates using the Qiagen-EZ1 DSP virus kit on the Qiagen-EZ1 Advanced XL platform (Qiagen). Qubit high-sensitivity dsDNA fluorometry was used for DNA quantitation to ensure adequate quality of material post-extraction prior to library preparation. Whole-genome sequencing was performed on the Illumina MiSeq platform after paired-end libraries were prepared using the Illumina DNA prep kit.

Sequencing metrics were optimized to aim for a target sequencing depth of 100× and minimum Q30 score across >90% obtained reads with quality visualization performed using FastQC v0.11.9 [21]. Contamination and species identification was performed using a *k*-mer based approach using Kraken2 v2.1.2 with exclusion of sequences with <90% matching GAS isolates [22, 23].

Global context and bioinformatic analysis

To put our isolates into a global context, the dataset ($n=318$) was augmented by including 117 Illumina sequences of group A streptococcal genomes with the raw reads downloaded from the National Center for Biotechnology Information (NCBI) Sequence Read Archive (SRA) [24] (see Table S1). These sequences represented isolates from a range of (i) invasive infections, (ii) *emm*-types including *emm1* sub-lineages (i.e. M1_{UK} and M1_{global}) and (iii) those representing several different geographical regions.

Raw reads were trimmed using FASTP v0.22.0 [25] prior to generating assemblies using Spades v3.12.0 [26]. Multilocus sequence typing (MLST) was used to characterize sequence types (STs), and was performed using the PubMLST database [27]. Virulence gene(s) and antimicrobial-resistance gene(s) were detected on the contigs using ABRicate v1.0.1 [28] against the virulence factor database (VFDB) [29] and amrFinder v3.11.14, respectively [30]. Greater than 85% identity and 90% coverage were required for detection of genes using these tools. *emm*-typing was performed using emmtyper with BLASTN v2.12.0 [31], against the downloaded Centers for Disease Control and Prevention (CDC) *emm*-type specific database (https://ftp.cdc.gov/pub/infectious_diseases/biotech/tsemm/). The two *emm*-like gene homologues (*enn* and *mrp*) were filtered with EMM clustering designation based on previous functional grouping of types into 48 groups [32, 33].

The trimmed reads were mapped against the *S. pyogenes* type strain reference sequence (NCBI accession no. CP028841.1) [34] using BWA-MEM v0.7.17-r1188 [35]. Resulting BAM files were sorted with SNPs detected using freebayes v1.3.5 [36]. Consensus genomes were subsequently determined using bcftools v1.14 ignoring indels [37].

Recombination was determined using Gubbins v3.3.10 [38] with default parameters. A maximum-likelihood phylogenetic tree was reconstructed using IQ-TREE2 from the masked SNP multifasta file [39]. The GTR+F+G4 model was used for tree reconstruction and 100 bootstrap trees were included. Tree visualization was performed using the web-based program iTOL, which allows for mapping epidemiological data to the phylogeny [40]. M1_{UK} variants were identified in the dataset by screening all invasive and non-invasive *emm1* isolate genomes for the 27 SNPs defined by Lynskey *et al.* [19].

RESULTS

Epidemiology of isolates

The iGAS isolates sequenced from the HNELHD between 2007 and 2017 showed a bimodal age distribution; 12.7% of isolates were from children under 10 years old ($n=23$), 26.5% of isolates were from people 10–49 years of age ($n=48$), and 60.8% of isolates were from adults 50 years of age and over ($n=110$). Overall, 109 cases occurred in males (60%), and 72 cases occurred in females (40%). Cases in males predominated after 50 years of age (70 out of 110 cases, 63.6%) (Fig. 1b).

Of the 136 non-invasive GAS isolates sequenced from adults in ISLHD between 2019 and 2022, 27.9, 19.9 and 27.9% of cases were in the 18–29 year ($n=38$), 30–39 year ($n=27$) and 40–49 year ($n=38$) age groups, respectively. Only 33 cases (24.3%) were from adults aged 50 years and over (Fig. 1b). Overall, 84 cases occurred in males (62%), and 52 cases occurred in females (38%). Cases in males predominated across all age groups (range 55–73%), and occurred to a greater extent in adults aged 50 years or over (24 out of 33 cases, 73%).

Distribution of *emm*-types and MLST characterization

A total of 53 *emm*-types and 57 different STs were detected in the 182 iGAS isolates. The top seven STs, which accounted for approximately 60% of invasive isolates, were ST-28 ($n=40$); ST-52 ($n=20$); ST-101 ($n=14$); ST-15 ($n=14$); ST-36 ($n=10$); ST39 ($n=7$); ST-182 ($n=5$). A total of 33 *emm*-types and 34 different STs were detected in the 136 non-invasive isolates. The top seven STs, which accounted for approximately 61% of non-invasive isolates, were ST-11 ($n=22$); ST-172 ($n=22$); ST-101 ($n=14$); ST-12 ($n=6$); ST-36 ($n=6$); ST-922 ($n=6$).

Across all 318 isolates, there were 66 *emm*-types and at least 75 different STs in total (Table S1). Seven *emm*-types represented 53.8% of the entire database [*emm1* ($n=42$), *emm89.0* ($n=30$), *emm53* ($n=26$), *emm28* ($n=24$), *emm59.1* ($n=22$), *emm3.1* ($n=14$) and *emm12* ($n=13$)]. All isolates were able to be assigned an *emm*-type; however, three isolates were novel and unable to be matched to an existing ST. These have been assigned STs 1443–1445 by PubMLST. *emm*-types were not restricted to any one genomic backbone, with 16 out of the 66 *emm*-types linked to multiple STs.

In the context of clinical presentations for the 318 isolates, 10 out of the 14 STs with at least five isolates caused both invasive and non-invasive infections (Table S1). To limit sampling bias, we restricted the analysis to those ST types that occurred at least once for 3 or more years and represented at least 5% of the isolates in the dataset ($n>15$). We found that certain strains were

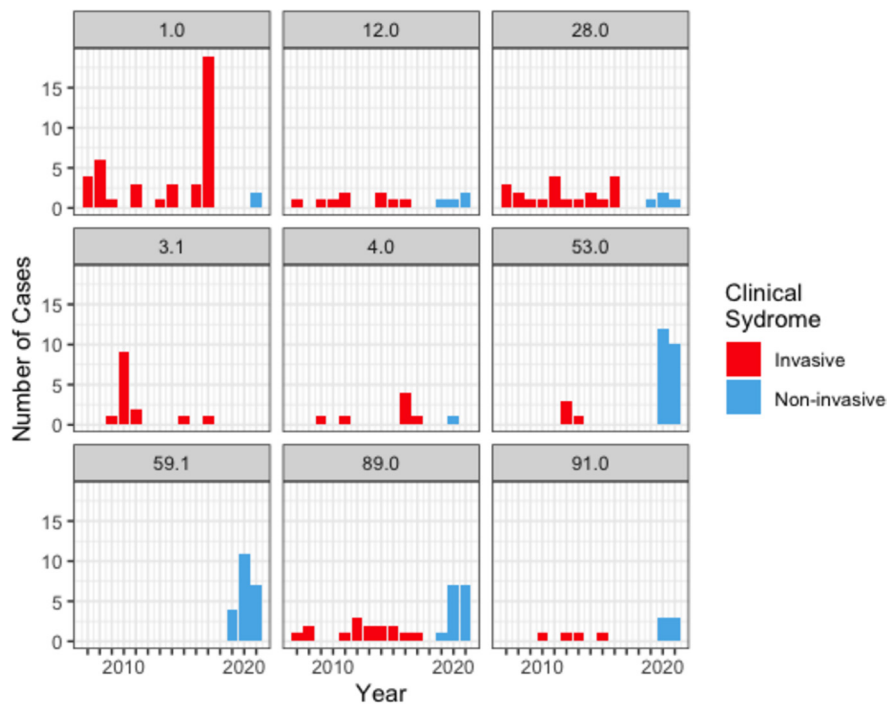


Fig. 2. Histogram of sequenced *S. pyogenes* infection isolates grouped by *emm*-type and by year of isolation. Stacked columns are coloured by clinical syndrome presentation (invasive isolates are shown in red, and non-invasive isolates are shown in blue). To circumvent a potential sampling bias, the analysis was restricted to *emm*-types that represented $\geq 5\%$ of the dataset and occurred at least once in 3 or more years. Note that invasive isolates were collected only in years 2007 to 2017 from the HNELHD. Non-invasive isolates were collected only in years 2019 to 2022 from the ISLHD. GAS isolates were not collected as part of this study during 2018.

more commonly detected in patients with invasive infection (e.g. *emm*1.0 ST-28 and *emm*28.0 ST-52) compared to non-invasive infections (e.g. *emm*53.0 ST-11 and *emm*59.1 ST-172) (Fig. 2).

Phylogeny of GAS in NSW, Australia

We reconstructed a maximum-likelihood phylogenetic tree to understand better the interrelationships between GAS found in NSW, Australia, and those seen in other countries. To make this tree, we masked recombination sites and added 117 international characterized sequences from a diversity of *emm*-types to our 318 GAS sequences for a total of 435 sequences (Table S1). International sequences were spread across the phylogeny, and we found no evidence of an Australian-restricted clone. The tree demonstrated a population structure with a large genetic diversity with multiple *emm*-types. There were several clonal expansion events observed, with little diversity within the clones (Fig. 3). As expected from the analysis of *emm*-types and STs, invasive and non-invasive disease manifestations were spread across the phylogeny, with a considerable number of *emm*-types and STs causing both invasive and non-invasive infections.

In our study, the within *emm*-type genomic diversity of our 318 isolates increased as circulation time increased. For example, genetically identical isolates were common within *emm*59.1, which had a mean diversity of 5 (IQR (interquartile range) 1–8) SNPs across a 2 year time frame compared to *emm*1, where the mean SNP difference was 33 (IQR 18–47) from isolates representing a 14 year time frame.

Characterization of *emm*1 sequences of M1_{UK} clone within NSW, Australia

A total of 42 *emm*1 types were identified from invasive ($n=40$) and non-invasive ($n=2$) isolates. Based on 27 defining SNPs, these *emm*1 sequences were further characterized into M1_{UK} or M1_{global} clones. M1_{global} clones comprised all *emm*1 isolates collected before 2013. The first M1_{UK} clone was seen in our dataset in 2013, followed by a large expansion event almost replacing the M1_{global} clone in 2017 (Fig. 4).

Antimicrobial-resistance genes and virulome

Antimicrobial resistance was more frequently encountered in non-invasive isolates than invasive isolates. Thirty-two (invasive, 10; non-invasive, 22; 10% of total, $n=318$) isolates harboured the *tetM* gene associated with tetracycline resistance. Nineteen

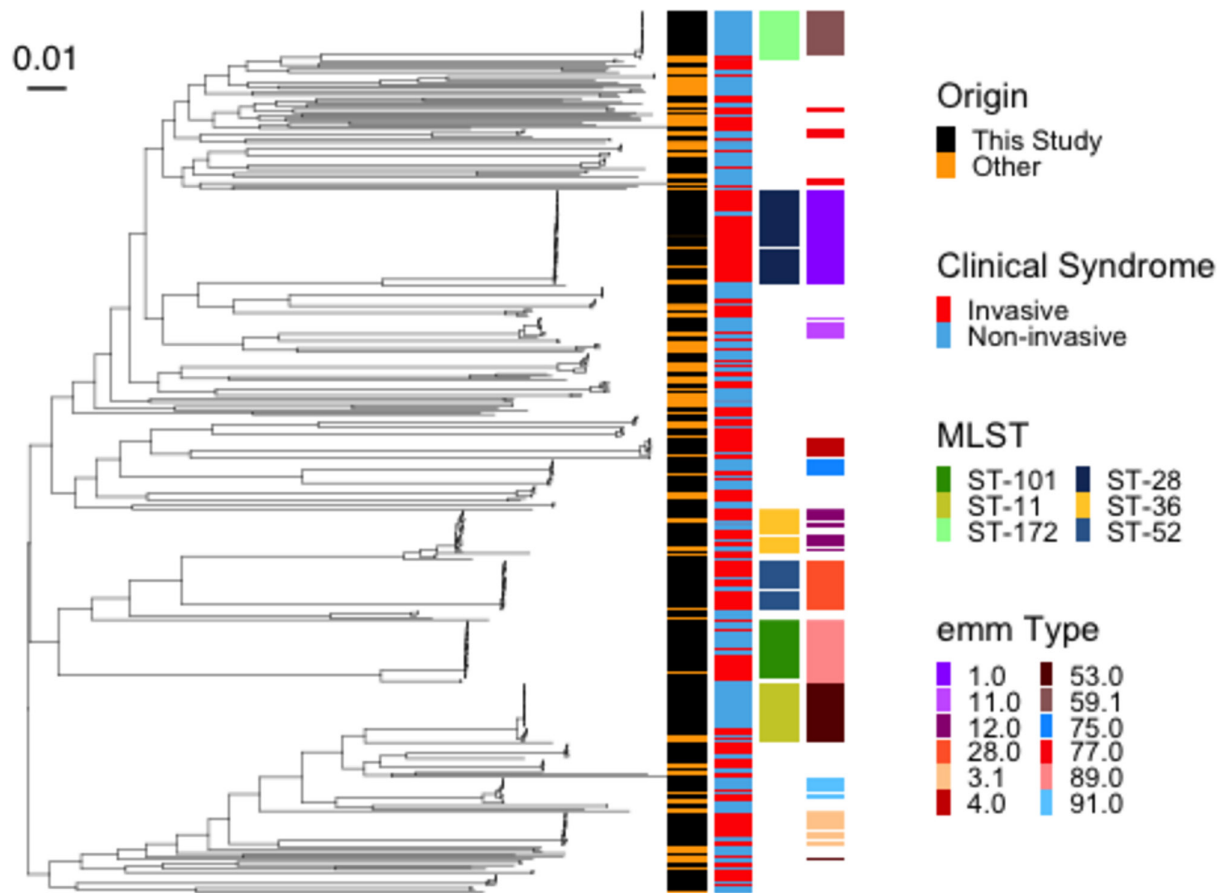


Fig. 3. Maximum-likelihood phylogeny following masking of recombination of 438 *S. pyogenes* sequences. This included 117 internationally characterized sequences from invasive and non-invasive disease cases that were collected across nine countries. Associated metadata are depicted by coloured tiles next to the tree as indicated in the key. Only MLST and *emm*-types that represented ~5% of the dataset and occurred in more than 1 year are shown. Scale bar shows phylogenetic distance.

isolates (invasive, 4; non-invasive, 15; 6% of total, $n=318$) harboured *ermA* or *ermB* genes, and one isolate (<1% of total, $n=318$) harboured *mefA* genes, which are all associated with macrolide resistance. Two isolates (<1% of total, $n=318$) harboured *dfzG* associated with sulfamethoxazole resistance. DNA gyrase mutations were not detected in any isolate.

Although numerous virulence genes have been associated with GAS, we restricted the analysis to only the 11 chromosomal and phage-associated superantigens present in the virulence factor database that was used. A minimum of one and maximum of six of these superantigen genes were detected in each isolate in our collection. Approximately 85 % of all of our isolates harboured between three and five of these superantigens. The proportions of the total number of superantigens detected remained similar when isolates were grouped by clinical presentation (Table S1).

DISCUSSION

Our study has described the molecular epidemiology of iGAS isolates collected over 11 years from the Hunter region and non-invasive GAS isolates collected over 3 years from the Illawarra region of NSW, Australia. Whilst the 66 different *emm*-types observed indicate that many GAS strains cause infections in these communities, a larger Australian dataset focused solely on iGAS cases from Victoria ($n=1202$; sampling period 2007–2017) found 140 *emm*-types [41]. Our findings regarding the most common *emm*-types for iGAS cases were similar to the Victorian study [41], and a study performed in the Western Sydney local health district, which found 27 *emm*-types ($n=55$; sampling period 2008 and 2010) [17]. Given the lack of evidence for Australian-specific clones, it is of little surprise that during the study period internationally significant strains (e.g. *emm1*, *emm3.1*, *emm12*, *emm28* and *emm89*) that make up the bulk of iGAS disease in Western countries also caused the majority of iGAS cases within these eastern Australian states [17, 41, 42].

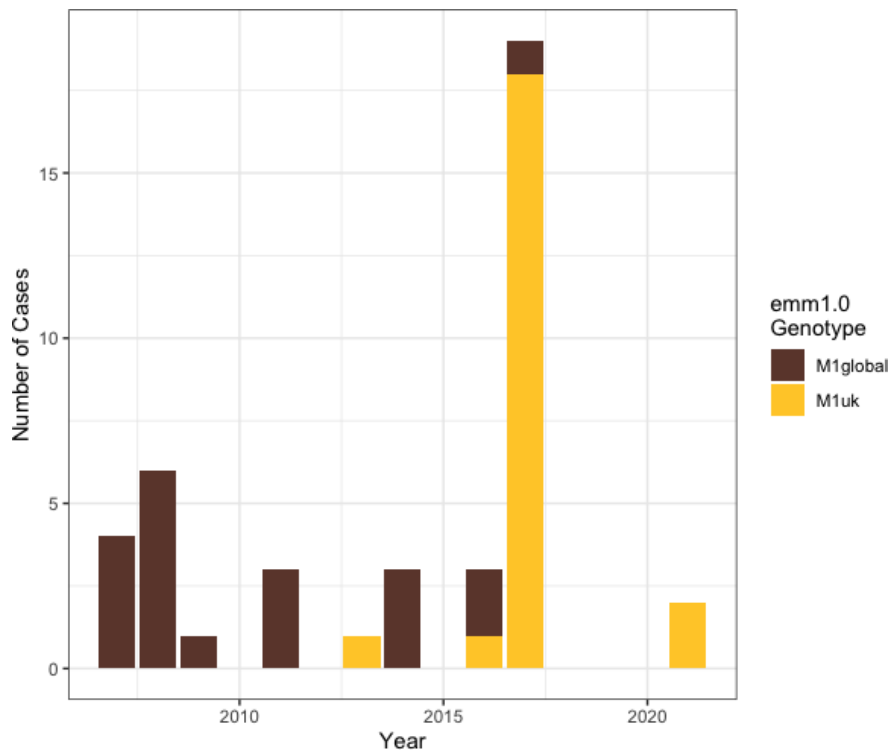


Fig. 4. Characterization of our *emm1* isolates as M1_{global} or M1_{UK}. Stacked bar graph of current study group A *S. pyogenes* infection isolates ($n=42$) grouped by year of isolation with columns coloured by *emm1* genotype. Note that comparisons from 2007 to 2017 only include iGAS isolates collected within the HNELHD, and comparisons from 2019 to 2021 only include non-invasive isolates collected within the ISLHD. GAS isolates were not collected as part of this study during 2018.

Previous studies found that certain *emm*-types were more frequently associated with invasive infections [43–46]. This led to the concept of an invasive index, which estimates the propensity for a strain to be associated with invasive infection compared to non-invasive infection [46]. Our data are consistent with these prior studies, and nominally support the invasive index concept (Fig. 2). However, due to the non-overlapping periods and different sampling across the two different time periods, firm conclusions cannot be drawn on the association of certain *emm*-types with invasiveness.

An increasing body of molecular evidence, including recent whole-genome sequencing data, has reinforced the concept that within each GAS *emm*-type, genetically similar clones cause both non-invasive and invasive infections; indeed, comparison of individual lineages by infection type has usually revealed that any genetic differences were minimal, and sometimes no genetic differences were found [45, 47–56]. For example, an *emm1* outbreak lasting approximately 20 days that affected nine people (two invasive cases and seven non-invasive cases) associated with a maternity unit in the UK displayed zero SNP differences between all but one isolate [53]. Our data on within *emm*-type genomic diversity within NSW are similar to those from international studies, which have noted minimal divergence in clones over years [48, 50, 53, 57–63]. The genetic similarity indicates that many epidemiologically unrelated clones circulating within the community would be indistinguishable over short time frames thereby posing a false notion of an outbreak at a molecular level. Therefore, spatial and temporal relatedness in addition to ‘molecular’ similarity is important to confirm an outbreak.

In our study, non-invasive and invasive isolates of the same *emm*-type and ST clustered together in the phylogeny. Reports by Shea *et al.* [51] and Hoe *et al.* [47] indicate that invasive infections could be sporadic events that repeatedly arise from the circulating pool of clones. Whilst likely, whether the invasive clones in our study arose from the circulating pool of colonizing non-invasive isolates was unable to be definitively determined due to the non-overlapping time periods of this study. Recent studies from school classrooms within the UK and indigenous communities within the Northern Territory of Australia have demonstrated that asymptomatic throat carriage of temporary duration is a common reservoir for many GAS strains [64, 65]. Failure to investigate asymptomatic carriage is a limitation of our study. Future studies should better account for this and investigate whether asymptomatic carriage provides a pool of clones from which invasive infections can arise.

The M1_{UK} strain is of international concern as it produces more superantigen SpeA than its progenitor M1_{global} strains [19, 66]. SpeA is a virulence factor that prolongs survival against the immune system [67]. The oldest known M1_{UK} isolates were collected

in the UK in 2010 [19]. By 2013, M1_{UK} was the UK's most commonly sequenced *emm1* clone [19, 68]. Recently, Davies *et al.* [66] showed that 50–75% of sequenced *emm1* isolates in Victoria and Queensland during 2017–2020 were M1_{UK} strains, and that it has been present in Australia since at least 2013. We had similar findings, including detection of M1_{UK} amongst the invasive isolates that were collected in 2013 within NSW. The notable expansion of M1_{UK} isolates in 2017 aligns with a significant peak in iGAS case numbers in the HNELHD during the same year [14], which also occurred in other states [41].

A key aim of our study was to understand the molecular epidemiology of invasive and non-invasive *S. pyogenes* isolates in NSW. Whole-genome sequencing incorporated into surveillance programmes could assist our public-health teams in outbreak investigation, specifically to exclude outbreaks. Whole-genome sequencing has advantages over *emm*-typing and MLST for outbreak investigations for GAS because small genetic differences can be used to 'rule out' links in appropriate epidemiological contexts [48, 50, 62]. Key findings from an Australian healthcare and economic modelling study revealed that non-invasive (and locally invasive infection) accounted for approximately 90% of healthcare presentations for *S. pyogenes* infections and 80% of healthcare-associated costs [13]. Public-health measures to reduce the burden of non-invasive infections, in addition to invasive infections, could positively impact on the overall healthcare costs [13, 69].

Finally, molecular epidemiology by whole-genome sequencing of invasive and non-invasive *S. pyogenes* isolates could provide rich information on vaccine coverage in our communities. The vaccine candidates undergoing clinical trials, such as the 30-valent vaccine, are the most promising tools to reduce GAS infection numbers [69–71]. On this note, under the assumption that all subtype *emm* alleles of a parent *emm*-type were covered by the 30-valent vaccine candidate [72], approximately 83% of our invasive isolates and 61% of our non-invasive isolates were covered or had documented evidence of cross-reactivity [70, 72]. With the inclusion of iGAS in the list of Public Health notifiable conditions and embedding whole-genome sequencing in continuous genomic surveillance, there will be a greater ability to track outbreaks, and inform on vaccine coverage.

Funding information

This work was supported by internal funding provided by NSWHP. This included NSWHP, John Hunter Hospital, funding provided to Dr Varadhan.

Conflicts of interest

The authors declare that there are no conflicts of interest.

Ethical statement

Study isolates and associated metadata were obtained following ethics approvals 2019/ETH12208 (HNELHD isolates) and 2022/ETH02409 (ISLHD isolates).

References

- Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, *et al.* Disease manifestations and pathogenic mechanisms of group A *Streptococcus*. *Clin Microbiol Rev* 2014;27:264–301.
- Dombek PE, Cue D, Sedgewick J, Lam H, Ruschkowski S, *et al.* High-frequency intracellular invasion of epithelial cells by serotype M1 group A streptococci: M1 protein-mediated invasion and cytoskeletal rearrangements. *Mol Microbiol* 1999;31:859–870.
- Purushothaman SS, Wang B, Cleary PP. M1 protein triggers a phosphoinositide cascade for group A *Streptococcus* invasion of epithelial cells. *Infect Immun* 2003;71:5823–5830.
- Cole JN, Barnett TC, Nizet V, Walker MJ. Molecular insight into invasive group A streptococcal disease. *Nat Rev Microbiol* 2011;9:724–736.
- Brouwer S, Rivera-Hernandez T, Curren BF, Harbison-Price N, De Oliveira DMP, *et al.* Pathogenesis, epidemiology and control of Group A *Streptococcus* infection. *Nat Rev Microbiol* 2023;21:431–447.
- Factor SH, Levine OS, Schwartz B, Harrison LH, Farley MM, *et al.* Invasive group A streptococcal disease: risk factors for adults. *Emerg Infect Dis* 2003;9:970–977.
- Wright CM, Langworthy K, Manning L. The Australian burden of invasive group A streptococcal disease: a narrative review. *Intern Med J* 2021;51:835–844.
- Gray L-A, D'Antoine HA, Tong SYC, McKinnon M, Bessarab D, *et al.* Genome-wide analysis of genetic risk factors for rheumatic heart disease in aboriginal Australians provides support for pathogenic molecular mimicry. *J Infect Dis* 2017;216:1460–1470.
- Coffey PM, Ralph AP, Krause VL. The role of social determinants of health in the risk and prevention of group A streptococcal infection, acute rheumatic fever and rheumatic heart disease: a systematic review. *PLoS Negl Trop Dis* 2018;12:e0006577.
- Carapetis JR, Steer AC, Mulholland EK, Weber M. The global burden of group A streptococcal diseases. *Lancet Infect Dis* 2005;5:685–694.
- Wright CM, Moorin R, Pearson G, Dyer JR, Carapetis JR, *et al.* Increasing incidence of invasive group A streptococcal disease in Western Australia, particularly among indigenous people. *Med J Aust* 2021;215:36–41.
- Thomson TN, Campbell PT, Gibney KB. The epidemiology of invasive group A streptococcal disease in Victoria, 2007–2017: an analysis of linked datasets. *Aust NZ J Public Health* 2022;46:878–883.
- Cannon JW, Jack S, Wu Y, Zhang J, Baker MG, *et al.* An economic case for a vaccine to prevent group A streptococcus skin infections. *Vaccine* 2018;36:6968–6978.
- Williamson KM, Varadhan H, Taylor K, Crooks K, Brett K. Epidemiology of group A streptococcal bacteraemia in Hunter New England Local Health district, 2008 to 2019. *Commun Dis Intell* 2008;47:cdi.2023.47.49.
- Australian Government. *Group A Streptococcal Disease – Invasive (iGAS)*. Canberra: Australian Government, Department of Health and Aged Care; 2022.
- NSW Government, NSW Health, Invasive Group A Streptococcus control guideline; 2022. <https://www.health.nsw.gov.au/Infectious/controlguideline/Pages/invasive-group-astrep.aspx>
- Sivagnanam S, Zhou F, Lee AS, O'Sullivan MVN. Epidemiology of invasive group A *Streptococcus* infections in Sydney, Australia. *Pathology* 2015;47:365–371.
- Li Y, Rivers J, Mathis S, Li Z, Velusamy S, *et al.* Genomic surveillance of *Streptococcus pyogenes* strains causing invasive disease, United States, 2016–2017. *Front Microbiol* 2020;11:1547.
- Lynskey NN. 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;19:1209–1218.
20. NSW Government, Department of Communities and Justice, District Profiles, ABS 2016 Census Data; 2020. <https://www.facs.nsw.gov.au/resources/statistics/dcj-districts>
 21. Andrews S. FastQC: a quality control tool for high throughput sequence data; 2010. <http://www.bioinformatics.babraham.ac.uk/projects/fastqc>
 22. Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol* 2019;20:257.
 23. Souvorov A, Agarwala R, Lipman DJ. SKESA: strategic k-mer extension for scrupulous assemblies. *Genome Biol* 2018;19:153.
 24. Davies MR, McIntyre L, Mutreja A, Lacey JA, Lees JA, et al. Atlas of group A streptococcal vaccine candidates compiled using large-scale comparative genomics. *Nat Genet* 2019;51:1035–1043.
 25. Chen S, Zhou Y, Chen Y, Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics* 2018;34:i884–i890.
 26. Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol* 2012;19:455–477.
 27. Seeman T. MLST; 2022. <https://github.com/tseemann/mlst>
 28. Seemann T. Abricate; 2020. <https://github.com/tseemann/abricate>
 29. Chen L, Zheng D, Liu B, Yang J, Jin Q. VFDB 2016: hierarchical and refined dataset for big data analysis – 10 years on. *Nucleic Acids Res* 2016;44:D694–D697.
 30. Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, et al. Validating the AMRFinder tool and resistance gene database by using antimicrobial resistance genotype-phenotype correlations in a collection of isolates. *Antimicrob Agents Chemother* 2019;63:e00483-19.
 31. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009;10:421.
 32. Frost HR, Davies MR, Delforge V, Lakhoulfi D, Sanderson-Smith M, et al. Analysis of global collection of group A *Streptococcus* genomes reveals that the majority encode a trio of M and M-Like proteins. *mSphere* 2020;5:e00806-19.
 33. Sanderson-Smith M, De Oliveira DMP, Guglielmini J, McMillan DJ, Vu T, et al. A systematic and functional classification of *Streptococcus pyogenes* that serves as a new tool for molecular typing and vaccine development. *J Infect Dis* 2014;210:1325–1338.
 34. Salvà-Serra F, Jaén-Luchoro D, Jakobsson HE, Gonzales-Siles L, Karlsson R, et al. Complete genome sequences of *Streptococcus pyogenes* type strain reveal 100%-match between PacBio-solo and Illumina-Oxford Nanopore hybrid assemblies. *Sci Rep* 2020;10:11656.
 35. Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013:1303.3997.
 36. Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv* 2012:1207.3907.
 37. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, et al. Twelve years of SAMtools and BCFtools. *Gigascience* 2021;10:giab008.
 38. Croucher NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, et al. Rapid phylogenetic analysis of large samples of recombinant bacterial whole genome sequences using gubbins. *Nucleic Acids Res* 2014;43:e15.
 39. Minh BQ, Schmidt HA, Chernomor O, Schrempf D, Woodhams MD, et al. IQ-TREE 2: new models and efficient methods for phylogenetic inference in the genomic era. *Mol Biol Evol* 2020;37:1530–1534.
 40. Letunic I, Bork P. Interactive tree of life (iTOL) v3: an online tool for the display and annotation of phylogenetic and other trees. *Nucleic Acids Res* 2016;44:W242–W245.
 41. Oliver J, Wilmot M, Strachan J, St George S, Lane CR, et al. Recent trends in invasive group A *Streptococcus* disease in Victoria. *Commun Dis Intell* 2018;43:cdi.2019.43.8.
 42. Gherardi G, Vitali LA, Creti R. Prevalent *emm* types among invasive GAS in Europe and North America since year 2000. *Front Public Health* 2018;6:59.
 43. Ekelund K, Darenberg J, Norrby-Teglund A, Hoffmann S, Bang D, et al. Variations in *emm* type among group A streptococcal isolates causing invasive or noninvasive infections in a nationwide study. *J Clin Microbiol* 2005;43:3101–3109.
 44. Shea PR. Group A *Streptococcus emm* gene types in pharyngeal isolates. *Emerg Infect Dis* 2002;17:2010–2017.
 45. Li Y, Dominguez S, Nanduri SA, Rivers J, Mathis S, et al. Genomic characterization of group A streptococci causing pharyngitis and invasive disease in Colorado, USA, June 2016– April 2017. *J Infect Dis* 2022;225:1841–1851.
 46. Shulman ST, Tanz RR, Kabat W, Kabat K, Cederlund E, et al. Group A streptococcal pharyngitis serotype surveillance in North America, 2000–2002. *Clin Infect Dis* 2004;39:325–332.
 47. Hoe NP, Vuopio-Varkila J, Vaara M, Grigsby D, De Lorenzo D, et al. Distribution of streptococcal inhibitor of complement variants in pharyngitis and invasive isolates in an epidemic of serotype M1 group A *Streptococcus* infection. *J Infect Dis* 2001;183:633–639.
 48. Bergin SM, Periaswamy B, Barkham T, Chua HC, Mok YM, et al. An outbreak of *Streptococcus pyogenes* in a mental health facility: advantage of well-timed whole-genome sequencing over *emm* typing. *Infect Control Hosp Epidemiol* 2018;39:852–860.
 49. Turner CE, Bedford L, Brown NM, Judge K, Török ME, et al. Community outbreaks of group A *Streptococcus* revealed by genome sequencing. *Sci Rep* 2017;7:8554.
 50. Plainvert C, Longo M, Seringe E, Saintpierre B, Sauvage E, et al. A clone of the emergent *Streptococcus pyogenes emm89* clade responsible for a large outbreak in a post-surgery oncology unit in France. *Med Microbiol Immunol* 2018;207:287–296.
 51. Shea PR, Beres SB, Flores AR, Ewbank AL, Gonzalez-Lugo JH, et al. Distinct signatures of diversifying selection revealed by genome analysis of respiratory tract and invasive bacterial populations. *Proc Natl Acad Sci USA* 2011;108:5039–5044.
 52. Alamarat ZI, Sanson MA, McNeil JC, Baker CJ, Flores AR. Use of whole-genome sequencing to detect transmission of group A *Streptococcus* in Houston, TX. *Access Microbiol* 2022;4:acmi000351.
 53. Turner CE, Dryden M, Holden MTG, Davies FJ, Lawrenson RA, et al. Molecular analysis of an outbreak of lethal postpartum sepsis caused by *Streptococcus pyogenes*. *J Clin Microbiol* 2013;51:2089–2095.
 54. Cockerill FR, MacDonald KL, Thompson RL, Roberson F, Kohner PC, et al. An outbreak of invasive group A streptococcal disease associated with high carriage rates of the invasive clone among school-aged children. *JAMA* 1997;277:38–43.
 55. Musser JM, Kapur V, Peters JE, Hendrix CW, Drehner D, et al. Real-time molecular epidemiologic analysis of an outbreak of *Streptococcus pyogenes* invasive disease in US Air Force trainees. *Arch Pathol Lab Med* 1994;118:128–133.
 56. Muotiala A, Seppälä H, Huovinen P, Vuopio-Varkila J. Molecular comparison of group A streptococci of T1M1 serotype from invasive and noninvasive infections in Finland. *J Infect Dis* 1997;175:392–399.
 57. Teatero S, McGeer A, Tyrrell GJ, Hoang L, Smadi H, et al. Canada-wide epidemic of *emm74* group A *Streptococcus* invasive disease. *Open Forum Infect Dis* 2018;5:ofy085.
 58. Fittipaldi N, Tyrrell GJ, Low DE, Martin I, Lin D, et al. Integrated whole-genome sequencing and temporospatial analysis of a continuing group A *Streptococcus* epidemic. *Emerg Microbes Infect* 2013;2:1–8.
 59. Athey TBT, Teatero S, Sieswerda LE, Gubbay JB, Marchand-Austin A, et al. High incidence of invasive group A *Streptococcus* disease caused by strains of uncommon *emm* types in Thunder Bay, Ontario, Canada. *J Clin Microbiol* 2016;54:83–92.
 60. Mosites E, Frick A, Gounder P, Castrodale L, Li Y, et al. Outbreak of invasive infections from subtype *emm26.3* group A *Streptococcus* among homeless adults – Anchorage, Alaska, 2016–2017. *Clin Infect Dis* 2017;66:1068–1074.

61. Hammond-Collins K, Strauss B, Barnes K, Demczuk W, Domingo M-C, *et al.* Group A *Streptococcus* outbreak in a Canadian armed forces training facility. *Mil Med* 2019;184:e197–e204.
62. Chalker VJ, Smith A, Al-Shahib A, Botchway S, Macdonald E, *et al.* Integration of genomic and other epidemiologic data to investigate and control a cross-institutional outbreak of *Streptococcus pyogenes*. *Emerg Infect Dis* 2016;22:973–980.
63. Fittipaldi N, Olsen RJ, Beres SB, Van Beneden C, Musser JM. Genomic analysis of emm59 group A *Streptococcus* invasive strains, United States. *Emerg Infect Dis* 2012;18:650–652.
64. Cordery R, Purba AK, Begum L, Mills E, Mosavie M, *et al.* Frequency of transmission, asymptomatic shedding, and airborne spread of *Streptococcus pyogenes* in schoolchildren exposed to scarlet fever: a prospective, longitudinal, multicohort, molecular epidemiological, contact-tracing study in England, UK. *Lancet Microbe* 2022;3:e366–e375.
65. Lacey JA. Evaluating the role of asymptomatic throat carriage of *Streptococcus pyogenes* in impetigo transmission in remote Aboriginal communities in Northern territory, Australia: a retrospective genomic analysis. *Lancet Microbe* 2023;4:e524–e533.
66. Davies MR, Keller N, Brouwer S, Jespersen MG, Cork AJ, *et al.* Detection of *Streptococcus pyogenes* M1_{UK} in Australia and characterization of the mutation driving enhanced expression of superantigen SpeA. *Nat Commun* 2023;14:1051.
67. Kasper KJ, Zeppa JJ, Wakabayashi AT, Xu SX, Mazzuca DM, *et al.* Bacterial superantigens promote acute nasopharyngeal infection by *Streptococcus pyogenes* in a human MHC class II-dependent manner. *PLoS Pathog* 2014;10:e1004155.
68. Zhi X, Li HK, Li H, Loboda Z, Charles S, *et al.* Emerging invasive group A *Streptococcus* M1(UK) lineage detected by allele-specific PCR, England, 2020. *Emerg Infect Dis* 2023;29:1007–1010.
69. Vekemans J, Gouvea-Reis F, Kim JH, Excler J-L, Smeesters PR, *et al.* The path to group A *Streptococcus* vaccines: World Health Organization research and development technology roadmap and preferred product characteristics. *Clin Infect Dis* 2019;69:877–883.
70. Dale JB, Penfound TA, Chiang EY, Walton WJ. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. *Vaccine* 2011;29:8175–8178.
71. Walkinshaw DR, Wright MEE, Mullin AE, Excler J-L, Kim JH, *et al.* The *Streptococcus pyogenes* vaccine landscape. *NPJ Vaccines* 2023;8:16.
72. Dale JB, Penfound TA, Tamboura B, Sow SO, Nataro JP, *et al.* Potential coverage of a multivalent M protein-based group A streptococcal vaccine. *Vaccine* 2013;31:1576–1581.

The Microbiology Society is a membership charity and not-for-profit publisher.

Your submissions to our titles support the community – ensuring that we continue to provide events, grants and professional development for microbiologists at all career stages.

Find out more and submit your article at microbiologyresearch.org