Epidemiological and genomic characterisation of an outbreak of *Streptococcus pyogenes emm*5.23Running title: Epidemiology and genomics of *S. pyogenes emm*5.23

Davide Pagnossin, Andrew Smith, William Weir, Eisin McDonald, Juliana Coelho, Roisin Ure, Katarina Oravcová

PII: S0163-4453(25)00092-1

DOI: https://doi.org/10.1016/j.jinf.2025.106498

Reference: YJINF106498

To appear in: *Journal of Infection*
Accepted date: 25 April 2025

Please cite this article as: Davide Pagnossin, Andrew Smith, William Weir, Eisin McDonald, Juliana Coelho, Roisin Ure and Katarina Oravcová, Epidemiological and genomic characterisation of an outbreak of *Streptococcus pyogenes emm*5.23Running title: Epidemiology and genomics of *S. pyogenes emm*5.23, *Journal of Infection,* (2025) doi:https://doi.org/10.1016/j.jinf.2025.106498

Epidemiological and genomic characterisation of an outbreak of *Streptococcus pyogenes emm*5.23

Running title: Epidemiology and genomics of *S. pyogenes emm*5.23

Davide Pagnossin[a*12], Andrew Smith[a,d3], William Weir[a4], Eisin McDonald[b], Juliana Coelho[c], Roisin Ure[d], Katarina Oravcová[a5]

[a]University of Glasgow, Glasgow, Scotland

[b]Public Health Scotland, Glasgow, Scotland

[c]UK Health Security Agency, UK

[d]Scottish Microbiology Reference Laboratories, Glasgow, Scotland

Email: davide.pagnossin@glasgow.ac.uk.

Email: dpagnoss@ed.ac.uk

[*]**Corresponding author:** Jarrett Building Room 337, Garscube Campus, 464 Bearsden Rd, G61 1QH, Glasgow, UK.

Abstract

Objectives

This retrospective cross-sectional study examined the epidemiology, clinical presentations, and genomics of *Streptococcus pyogenes* genotype *emm*5.23, linked to severe outcomes in Scotland.

Methods

Between 2014 and 2022, 58 cases of invasive Group A *Streptococcus* (iGAS) disease associated with *emm*5.23 were reported in Scotland. Surveillance data from

[1] ORCiD: https://orcid.org/0000-0001-8646-3003
[2] Present address: Easter Bush Campus, Charnock Bradley Building, EH25 9RG, Edinburgh, UK.
[3] ORCiD: https://orcid.org/0000-0003-0580-4078
[4] ORCiD: https://orcid.org/0000-0001-8648-666X
[5] ORCID: https://orcid.org/0000-0001-5930-6803

45 cases were analysed for clinical characteristics and risk factors. Whole-genome sequencing (WGS) included all available *emm*5.23 strains from Scotland (n=58), a subset from England (n=29), and *emm*5 strains of non-5.23 subtypes from Scotland (n=10), England (n=2), and Canada (n=1).

Results

Nearly all cases (96%, 43/45) were hospitalised, of whom 33% (15/45) required intensive care and 20% (9/45) died with iGAS. The most common presentations were bacteraemia (51%, 23/45) and pneumonia (24%, 11/45). WGS identified an emerging *emm*5.23 clade in Scotland, encompassing most isolates, which shared highly similar genomes and three non-synonymous polymorphisms.

Conclusions

Although genomic traits known to increase GAS virulence potential were not found, polymorphisms that may affect the *emm*5.23 phenotype were detected. This suggests this *emm*5.23 genotype was transiently successful rather than hypervirulent, with low population-level immunity contributing to its spread. This study emphasises the need for integration of real-time genomic data in public health surveillance to enhance source attribution and guide interventions.

Keywords: *Streptococcus pyogenes*, Gram-Positive Bacterial Infections, Bacteraemia, Soft Tissue Infections, Disease Outbreaks, Risk Factors, Antimicrobial Drug Resistance, Virulence, Whole Genome Sequencing, Public Health Surveillance.

Introduction

Streptococcus pyogenes, also known as Group A *Streptococcus* (GAS), can colonise the upper respiratory tract of asymptomatic people, with an estimated

overall prevalence of asymptomatic pharyngeal carriage of 7% across all age groups [1]. Occasionally, GAS can cause infections that, in their most severe forms, affect normally sterile body sites such as muscles, the pleural cavity and the bloodstream, with high mortality rates. These forms of infections are referred to as invasive GAS (iGAS) disease [2,3].

Several methods are used to classify GAS strains, including whole-genome classification, multi-locus sequence typing (MLST), and *emm* typing. The latter is widely used due to its ease of implementation and is based on the sequence of a region of the *emm* gene, which encodes the surface M protein. To date, over 275 *emm* types and subtypes have been identified [4], with some being associated with higher incidence of invasive disease than others. GAS *emm*1, for instance, has been frequently involved in invasive disease over the past few decades following the horizontal acquisition of virulence factors Sda2 and SpeA2, as well as a recombination event around the *nga-ifs-slo* operon [5,6]. The emergence of a recent *emm*1 lineage, named M_1UK, is characterised by the acquisition of 27 single nucleotide polymorphisms (SNPs) that confer an increased production of the SpeA toxin [7]. The high incidence and disease severity associated with the modern capsular *emm*89 lineage have been linked to recombination events that resulted in the loss of the capsule locus and the acquisition of a region encoding virulence factors NADase and SLO, with mutations in the *nga* promoter that drive their high expression [6,8]. In the case of *emm4* strains, the recent increase in invasive disease incidence was not attributed to enhanced production of classical virulence factors, such as those described above. Instead, it was associated with mutations leading to improved defences against oxidative stress and enhanced survival within

macrophages [9]. The acquisition of multi-drug resistance, which is often associated with horizontal gene transfer events, is also thought to play a role in the spread of pathogenic GAS clones [10].

In contrast, certain GAS strains have been associated with rapid, self-limiting epidemic waves, characterised by a surge in cases of invasive disease followed by sporadic recurrences or complete disappearance over the course of a few years [11,12]. While this phenomenon has been observed on multiple occasions, the underlying reasons remain unclear, unless they are associated with outbreaks in enclosed settings such as care homes [13,14]. It has been hypothesised that novel strains introduced into naïve populations spread successfully and cause a noticeable disease burden only until herd immunity is achieved. Competition with strains that are 'more fit' could also contribute to the decline in the reported disease burden of certain GAS genotypes [15].

From 2018 to 2022, a newly reported genotype of *S. pyogenes*, *emm* subtype 5.23, was involved in several cases of invasive disease in Scotland. While GAS *emm*5 cases represent less than 2% of the *S. pyogenes* disease burden in high income countries [16], *emm*5.23 in Scotland caused 4.71% and 9.82% of all iGAS cases reported in 2018 and 2019, respectively (SMiRL, unpublished data). Although not commonly isolated, this *emm* subtype was involved in one outbreak of invasive disease in a care home in England [17] and was also associated with increased mortality in the North-West of the country [18]. No clear association between *emm*5.23 infection and known risk factors has been noted to date. A review of the available

literature did not indicate any upsurge of iGAS *emm*5.23 cases in other countries in recent years.

This study constitutes a retrospective investigation into the short-lived epidemic surge of iGAS cases associated with severe clinical outcomes by *emm*5.23 in Scotland. We integrated and analysed epidemiological information collected as part of routine enhanced surveillance together with whole genome sequence (WGS) to comprehensively characterise the upsurge in cases. Epidemiological data, including clinical presentation and risk factors, provided context for the population groups affected, while genomic analyses allowed us to investigate the phylogenetic relationships among strains, gain insight into bacterial virulence and identify evolutionary adaptations.

Methods

Enhanced surveillance of iGAS *emm*5.23 cases

Confirmed iGAS infections in Scotland are notifiable to Public Health Scotland (PHS) under the Public Health etc. (Scotland) Act 2008. Suspected invasive cases are defined as potential GAS infections localised in normally sterile body sites or in non-sterile body sites of patients with severe clinical manifestations [3,19]. A case is confirmed once GAS is isolated from a specimen originating from a suspected case, after which the isolate is submitted to the Scottish Microbiology Reference Laboratory (SMiRL) for *emm* typing. Additionally, an enhanced questionnaire on clinical outcome and risk factors is sent from local health protection teams to PHS by secure email. Since not all health protection teams submit this form of information, only 45 of the 58 *emm*5.23 cases had enhanced questionnaires available. As there

are no definitions for each risk factor category, the interpretation of risk factor items in the enhanced questionnaires can be subjective. However, the definition of the term 'death' with iGAS is specified as the patient dying within seven days of a positive specimen, regardless of the actual cause of death. Questionnaire data were consulted and summarised. The burden of cases per 100,000 people in different age groups was calculated using mid-year population estimates published by the National Records of Scotland for 2019, the year when most cases occurred [20].

Ethics statement

The Public Health Scotland Order 2019 in Article 9(2)(i) places an obligation on Public Health Scotland (PHS) to engage in the control of spread of infectious diseases in accordance with section 43 of the National Health Service (Scotland) Act 1978. In accordance with sections 15, 16(5), and 21(2) of the Public Health etc. (Scotland) Act 2008, the Scottish Microbiology Reference Laboratory (SMiRL), Glasgow and PHS are obliged to process data in relation to notifiable diseases, health risk states of patients, notifiable organisms, and carrying out public health investigations, and therefore, individual patient consent is not required. Permissions to process and share relevant patient information data was approved by the Public Benefit and Privacy Panel for Health and Social Care (Reference 2223-0204 Smith).

Bacterial isolates

Invasive GAS isolates from Scotland were genotyped according to the CDC guidelines [4]. In the present study, we included all available iGAS *emm*5.23 strains from Scotland up until 2022 (n=58). To enable a more insightful assessment of genomic differentiation among *emm*5.23 strains, a sample of *emm*5.23 isolates

obtained from patients with invasive (n=22) and non-invasive (n=7) disease between 2012 and 2020 in England was supplied by the UK Health Security Agency (UKHSA) and included in the MGE identification, variant calling and phylogenetic analyses. Similarly, we sequenced and analysed genomes of ten randomly selected invasive *emm*5 isolates of non-5.23 subtypes (*emm*5.132, *emm*5.16, *emm*5.165, *emm*5.166, *emm*5.167, *emm*5.175, *emm*5.176, *emm*5.188, *emm*5.3, *emm*5.6) collected in Scotland in 2019 (n=8) and 2020 (n=2). Additionally, we searched for publicly available whole genome sequence reads for non-5.23 *emm*5 isolates, identifying three genomes: an *emm*5.14 isolate from Canada (2010-2013, accession: SRR1105943), and two *emm*5.3 isolates from England (2019, accession: SRR14668868 and SRR14668867), which were incorporated into the dataset.

Antimicrobial susceptibility testing

Antimicrobial susceptibility to a panel of antibiotics commonly used to treat Gram positive infections, namely ampicillin, amoxicillin, clindamycin, ceftriaxone, cefotaxime, doxycycline, erythromycin, levofloxacin, meropenem, moxifloxacin, oxacillin, penicillin G, tetracycline and vancomycin, was measured for a randomly selected subset of the *emm*5.23 isolates (n=25). The Micronaut-S kit (Merlin, Berlin, Germany), which employs the broth microdilution method, was used to determine the isolates' minimum inhibitory concentrations (MIC) for the tested antibiotics as per EUCAST guidelines v 5.0 [21] and manufacturer's instructions. MIC values were interpreted according to the EUCAST breakpoint table v 14.0, 2024 [22].

Whole genome sequencing

Bacterial genomic DNA was extracted using the DNeasy 96 Blood and Tissue Kit (Qiagen, Hilden, Germany) according to the manufacturer's instructions. DNA extractions were performed on a QIAsymphony automated instrument (Qiagen, Hilden, Germany). Genomic DNA was quantified using a Qubit dsDNA High Sensitivity kit on a Qubit 3 Fluorometer (Thermo Fisher Scientific, Waltham, MA USA). Paired-end libraries were generated using a Nextera XT DNA Sample Preparation Kit with Index Kit v2 and sequenced using a MiSeq reagent kit V3 with a 2 x 250 bp output on a MiSeq platform (all from Illumina, San Diego, CA, USA) at SMiRL, Glasgow. Ten of the isolates from England had previously undergone whole-genome sequencing (WGS) by UKHSA using Illumina HiSeq technology. Those sequences were shared with the University of Glasgow. For the remaining isolates from England that had not been previously sequenced (n=19), bacterial samples were delivered to SMiRL, and WGS was performed as described for the isolates from Scotland.

Sequence assembly and quality control

Paired-end reads were trimmed using ConDeTri v3.11.1 with default settings [23]. Trimmed reads were *de novo* assembled using SPAdes v3.11.1 [24]. Assembly quality was assessed with QUAST v5.0.2 [25]. Assemblies where the cumulative contig length was greater than 2.25 Mbp and/or having a GC% content higher or lower than two standard deviations from the mean were considered potentially contaminated and removed from further analyses.

Mobile genetic elements, virulence and AMR genes

The presence of mobile genetic elements (MGEs) was detected using SRST2 v0.2.0 [26] coupled with a published database of GAS phage and integrative and conjugative element (ICE) integrase and virulence genes, as previously described [12]. Since all strains, including *emm*5.23 and other *emm*5 subtypes, appeared to share the same combination of integrase genes, it was initially assumed that they also carried the same MGEs. To investigate this, the complete sequences of these MGEs were aligned to each genome using BLAST [27]. MGE sequences were obtained from the closed *emm*5.23 genome iGAS376 (accession: CP067010) which originated from the same isolate used for sequence S.376 in the present study [28]. MGE sequences were also queried against the NCBI non-redundant database using BLAST [27].

The presence of virulence and antimicrobial resistance (AMR) genes was established using ARIBA [29] coupled with the virulence factors database (VFDB) [30] and the comprehensive antibiotic resistance database (CARD) [31], respectively.

Polymorphism detection and phylogenetic analysis

WGS data of all strains in our dataset were mapped for variant calling using Snippy [32] against two chosen reference strains: iGAS426 (accession: CP067008), which is the oldest *emm*5.23 isolate from Scotland (2015) and the same isolate used for sequence S.426 in this study, and Manfredo (accession: AM295007), which is a historical *emm*5 strain isolated in 1952 in the United States [33]. Variant calling using Manfredo as a reference served two main purposes. First, it allowed us to assess the acquisition of polymorphisms shared by all recent *emm*5 isolates. Second, it helped determine whether the *emm*5.23 genotype had developed mutations specific to this *emm* subtype. Additionally, variant calling using iGAS426 as a reference was performed to investigate whether *emm*5.23 strains isolated in Scotland during the

2018–2019 surge in invasive disease cases had acquired mutations relative to previously circulating *emm*5.23 strains.

A core SNP phylogeny of *emm*5.23 strains was generated using identified SNPs with the iGAS426 genome as a reference. A second core SNP phylogeny, including all *emm*5 strains in our dataset, was constructed using the Manfredo genome as a reference. In order to limit the phylogenetic inference to vertically acquired genomic variations, SNPs located in MGE regions were excluded from the alignment used for phylogenetic tree construction. IQ-TREE was used to build maximum likelihood phylogenetic trees with standard non-parametric bootstrap analysis using the best-fit substitution model function [34]. The first tree was midpoint rooted, whilst the second was rooted on the Manfredo reference genome. Both trees were visualised with iTOL [35].

Results

Overview of clinical cases

A total of 58 cases of iGAS disease were associated with the genotype *emm*5.23 between 2014 and 2022 in Scotland. The surge in *emm*5.23 infections was rapid but short-lived, with most cases (83%, n=48/58) occurring between 2018 (n=21) and 2019 (n=27). Eight cases were reported in 2020 and only a single case was reported in 2015 and another in 2022. Invasive strains were isolated from blood (n=40/58, 69%), cutaneous (n=10/58, 17%) and respiratory specimens (n=8/58, 14%). The mean age of the patients affected was 50, and the median age was 55, with a range spanning from 1 day to 93 years.

Demographics, clinical presentation and severity of disease

Enhanced surveillance data were available for 45 cases. Twenty-six of the individuals infected were females (58%) and 19 were males (42%). The majority of individuals were aged between 45 and 64 (42%, n=19/45), 14 were 65 or older (31%), six were in the age group 15-44 (13%) and six were 14 or younger (13%). The burden of *emm*5.23 cases per 100,000 people by age group over the entire study period was as follows: 0.7 for people aged 14 or younger, 0.3 for people aged 15-44, 1.3 for people aged 45-64 and 1.3 for people aged 65 or older.

Nearly all cases (96%, n=43/45) were hospitalised, and one third of them (33%, n=15/45) were admitted to an intensive care unit (ICU). One in five patients (20%, n=9/45) died with iGAS. The highest case-fatality rate by clinical presentation was observed in cases of necrotising fasciitis (33%, n=1/3), followed by those of pneumonia (27%, n=3/11) and bacteraemia with unknown focus (17%, n=4/23). The majority of cases (51%, n=23/45) presented as bacteraemia with unknown focus and nearly one in four cases had pneumonia (n=11/45). Less common clinical presentations included necrotising fasciitis (n=3/45), surgical wound infection (n=3/45), non-iatrogenic skin and soft tissue infection (n=2/45), septic arthritis (n=2/45), meningitis (n=1/45), puerperal sepsis (n=1/45) and osteomyelitis (n=1/45). Six cases were classified with multiple clinical presentations.

Risk factors

Twenty-one patients (47%) had at least one of the following risk factors indicative of chronic disease: chronic pulmonary disease, chronic cardiac disease, diabetes,

malignancy and immunosuppression (Table 1). Only six cases were reported as not being associated with any risk factors.

Table 1. Breakdown of the cases of *emm*5.23 infection and death by known invasive Group A *Streptococcus* risk factors. Some patients had more than one risk factor

Risk factor	**Cases n/45**	**Percent of total**	**Deaths**	**Case fatality rate**
Skin lesion or wound (Group A *Streptococcus* secondary infection not confirmed)	11	24%	1	9%
Chronic pulmonary disease	10	22%	3	30%
Diabetes	6	13%	1	17%
Respiratory tract infection	6	13%	2	30%
Chronic cardiac disease	5	11%	1	20%
Sore throat (Group A *Streptococcus* infection not confirmed)	5	11%	0	0%
Contact of person with Group A *Streptococcus* infection	3	7%	0	0%
Immunosuppression	3	7%	1	33%
Malignancy	3	7%	1	33%
Alcohol abuse	2	4%	0	0%
Surgery	2	4%	0	0%
Use of non-steroidal anti-inflammatory drugs	2	4%	0	0%
Childbirth	1	2%	0	0%
Non-invasive Group A *Streptococcus* infection confirmed	1	2%	0	0%
Steroid use	1	2%	0	0%
Injecting drug user	0	0%	0	0%
Varicella infection	0	0%	0	0%
Homeless	0	0%	0	0%
Other risk factors (comorbidities, e.g., obesity, liver failure, cellulitis not confirmed to be caused by Group A *Streptococcus*)	4	9%	0	0%

Antimicrobial susceptibility

All isolates tested were susceptible to β-lactams, fluoroquinolones, glycopeptides, lincosamides and macrolides. All isolates, however, displayed phenotypic resistance to tetracyclines. MIC values for the panel of antibiotics tested are reported in Supplementary Table 1.

Whole genome sequencing

Fifty-four WGS from Scotland and all WGS from England showed good quality control metrics, while four sequences from Scotland showed signs of contamination and were removed from further analyses. Sequence characteristics of the remaining 64 genomes from Scotland, and of the 29 genomes from England, are presented in Supplementary Table 2.

Mobile genetic elements, virulence and AMR genes

The same five MGEs were found in each *emm*5.23 genome, as shown in Figure 1. Three of these were prophages, one was a *S. pyogenes* Chromosomal Island (SpyCI), while the last one was a composite element made of a prophage and an ICE. Although it is unusual for an AMR-associated element to integrate into a phage carrying the virulence gene *spd3* [36], we found that the ICE in this composite element carries the *tetM* gene, which confers resistance to tetracyclines. Six phage-associated virulence genes were identified in the *emm5.23* genomes: four encode DNases, one encodes the superantigen *speC*, and one encodes a hyaluronidase. The location of each MGE found in the reference sequence iGAS.376 is reported in Supplementary Table 3.

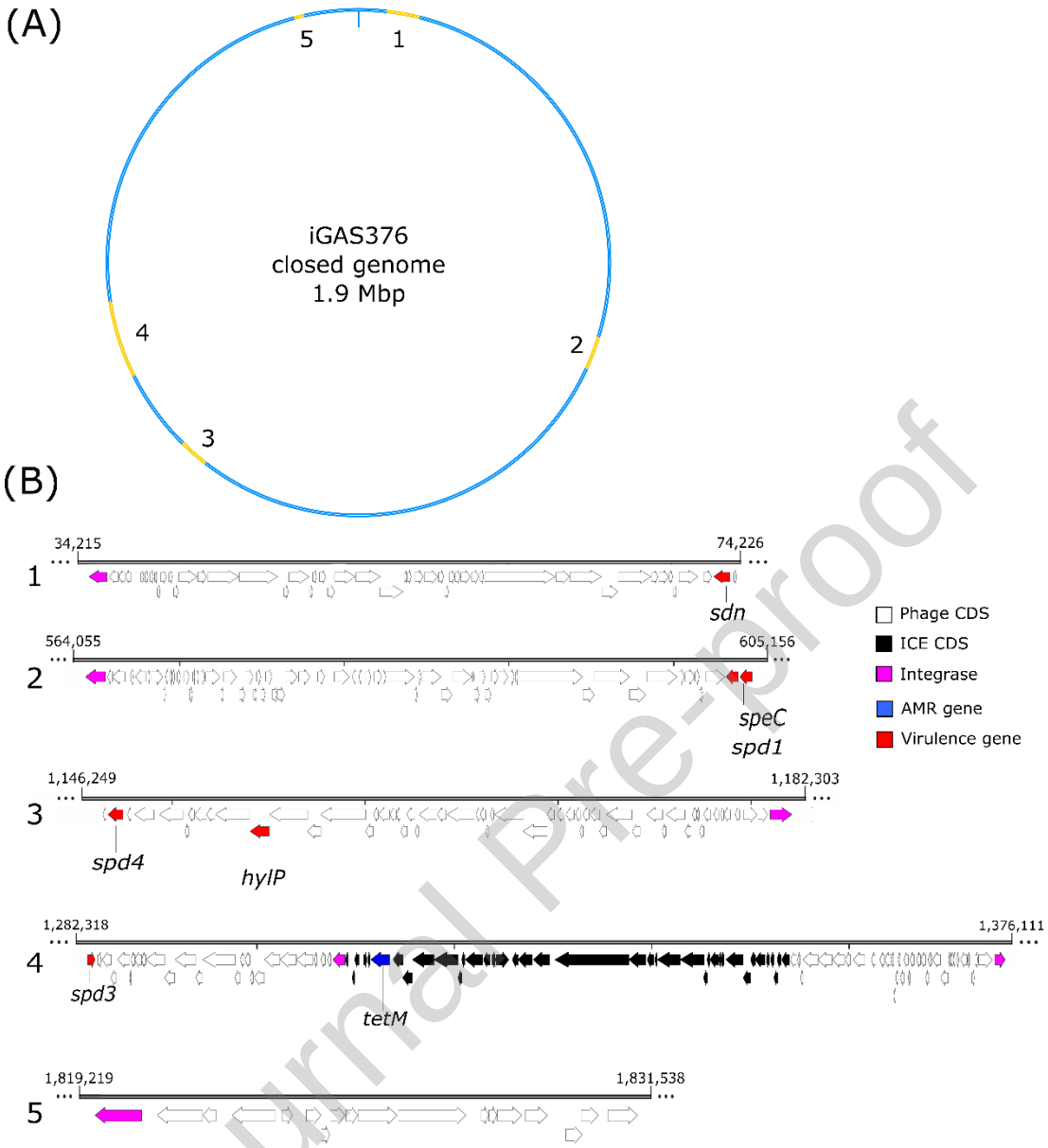

Figure 1. A - The five mobile genetic elements (MGEs) detected in all *emm*5.23 strains, location shown in yellow on the reference strain iGAS.376. B - Characteristics of the coding sequences (CDS) belonging to each MGE. MGEs 1, 2, 3 and 5 consist entirely of phage-origin sequences, while MGE 4 has both phage and integrative and conjugative element-derived sequences. MGE orientation in the figure is 5′-3′. Integrase, antimicrobial resistance and virulence genes are highlighted. Nucleotide positions in the reference genome indicating the beginning and end of each MGE are also shown.

Although all other *emm*5 genomes, other than *emm*5.23, carried the same MGE integrase genes initially detected in the *emm*5.23 isolates, five lacked the ICE located within MGE 4. However, they showed no significant differences in the sequence of the remaining four MGEs. These five isolates comprised of an *emm*5.14 strain from Canada, an *emm*5.6 strain from Scotland, an *emm*5.3 strain from Scotland, and two *emm*5.3 strains from England. The remaining eight isolates of a non-5.23 subtype all carried the previously described ICE identified in our *emm*5.23 strains.

Comparison of the MGEs detected in the *emm*5.23 and other non-5.23 *emm*5 genomes with those integrated into the Manfredo strain revealed a high degree of similarity to four of the five Manfredo MGEs, with identical insertion sites [33]. Specifically, MGE 2 matched Man 4, MGE 3 matched Man 2, the phage in MGE 4 matched Man 1, and MGE 5 matched Man 5. In contrast, the analysed genomes had acquired MGE 1, which contains the virulence gene *spd*, but lacked phage Man 3, harbouring the virulence gene *speH* in the Manfredo strain.

The same AMR genes were shared by all *emm*5.23 genomes, namely *lmrP* and *tetM*. Based on the AMR gene content, the *emm*5.23 population had the potential to express phenotypic resistance to four antibiotic classes: lincosamides (*lmrP*), macrolides (*lmrP*), streptogramin (*lmrP*) [37] and tetracyclines (*lmrP* and *tetM*) [37,38]. All isolates analysed also shared the same 44 virulence genes (including the ones integrated into phages) responsible for the synthesis of 34 virulence factors (Supplementary Table 4). The absence of virulence genes, including *sda2*, *speA2* [5],

speK, and *ssa* [39], which are associated with high virulence in *emm* types 1 and 3, was confirmed in the *emm*5.23 genomes analysed. Additionally, the presence of full-length *hasA*, *hasB* and *hasC* genes, which encode the hyaluronic acid capsule and whose absence or truncation has been linked with increased virulence in *emm*89 and other *emm* types, was also confirmed [8,40].

Polymorphism detection and phylogenetic analysis

All recent *emm*5 isolates, including *emm*5.23 (n=83) and other subtypes (n=13), shared 37 non-synonymous SNPs and 10 indels compared to the core genome of the 1952 Manfredo isolate. Given previous reports of mutations in virulence genes (e.g., *speA* and *has* genes[5,40]) and transcriptional regulators (*covR*, *covS*, and *rocA*[8,41]) associated with increased invasive disease incidence in certain strains of different *emm* types, we focused on mutations in these genes. In light of the recent association between the expansion of an emergent *emm*4 genotype and the acquisition of a mutation in a putative carbonic anhydrase-encoding gene[9], along with evidence that strains carrying this mutation develop improved phenotypic survival strategies during interaction with macrophages, we also examined polymorphisms in genes related to metabolic pathways. We found that all recent *emm*5 isolates carried three non-synonymous SNPs in transcription regulation genes, two in the *mga* pseudogene (G952A, Leu309Phe; G1430A, Thr477Ile) and one in *ropB* (G505A, Val169Ile). In addition to the non-synonymous mutations shared by all recent *emm*5 genomes, *emm*5.23 strains specifically carried 16 non-synonymous SNPs and five indels compared to the core genome of the Manfredo strain. None of these mutations were found in virulence or transcription regulation genes, however, a SNP was identified in gene *adhE* (aldehyde-alcohol

dehydrogenase, G2425A, Val809Ile), which is involved in bacterial metabolism. Finally, we did not detect any non-synonymous polymorphisms that were uniquely shared by all *emm*5.23 strains isolated in Scotland during the 2018–2019 surge in invasive disease compared to the earliest *emm*5.23 strain collected in Scotland in 2015.

The majority of the *emm*5.23 genomes from Scotland (85%, n=46/54), however, carried the same three polymorphisms compared to both the Manfredo and iGAS426 reference sequences. One is a SNP identified in the lactate oxidase gene *lctO* (C54A - Phe18Leu). Another is a nucleotide insertion (T -> AT) in a non-coding region upstream of the gene *greA*, at position -37. The third mutation is a SNP that caused an early stop codon in the gene encoding the enzyme Acetyl-CoA C-Acyltransferase (C112T - Gln38*). When queried in BLASTn against the NCBI database, the three polymorphisms were not identified in any other published sequences.

Phylogenetic analysis of *emm*5.23 strains suggests that most isolates from Scotland (85%, n=46/54) along with one isolate from England form a monophyletic group, in which all strains share the three previously discussed polymorphisms (Figure 2). We refer to this group as the 'emerging clade.' Strains within this clade exhibit close genomic relatedness, with an average of four SNP differences (range: 0–16). In contrast, the remaining eight *emm*5.23 isolates from Scotland and nearly all *emm*5.23 isolates from England (n=28) show greater genomic divergence, with an average of 18 core SNP differences (range: 0–66). The tree topology suggests that the *emm*5.23 isolates in the emerging clade likely originated from a relatively recent common ancestor, represented by a central cluster of six isolates with no core SNP

differences. This central cluster, in turn, appears to have emerged from *emm*5.23 strains already circulating in England and Scotland. Of the 47 isolates in the emerging clade, 36 (77%) differ from the central cluster genotype by only 0–3 SNPs, while the remaining 11 have no more than 7 SNP differences from this putative ancestral genome.

The phylogenetic tree, constructed using all available *emm*5 genomes, shows that certain strains of non-5.23 subtypes cluster within the same phylogenetic group as *emm*5.23 strains (Supplementary Figure 1). In contrast, isolates forming distinct phylogenetic groups separate from *emm*5.23 include the reference strain Manfredo, the *emm*5.14 isolate from Canada, the *emm*5.6 isolate from Scotland, the *emm*5.3 isolates from Scotland and England, and the *emm*5.165, *emm*5.166, and *emm*5.167 isolates, all from Scotland.

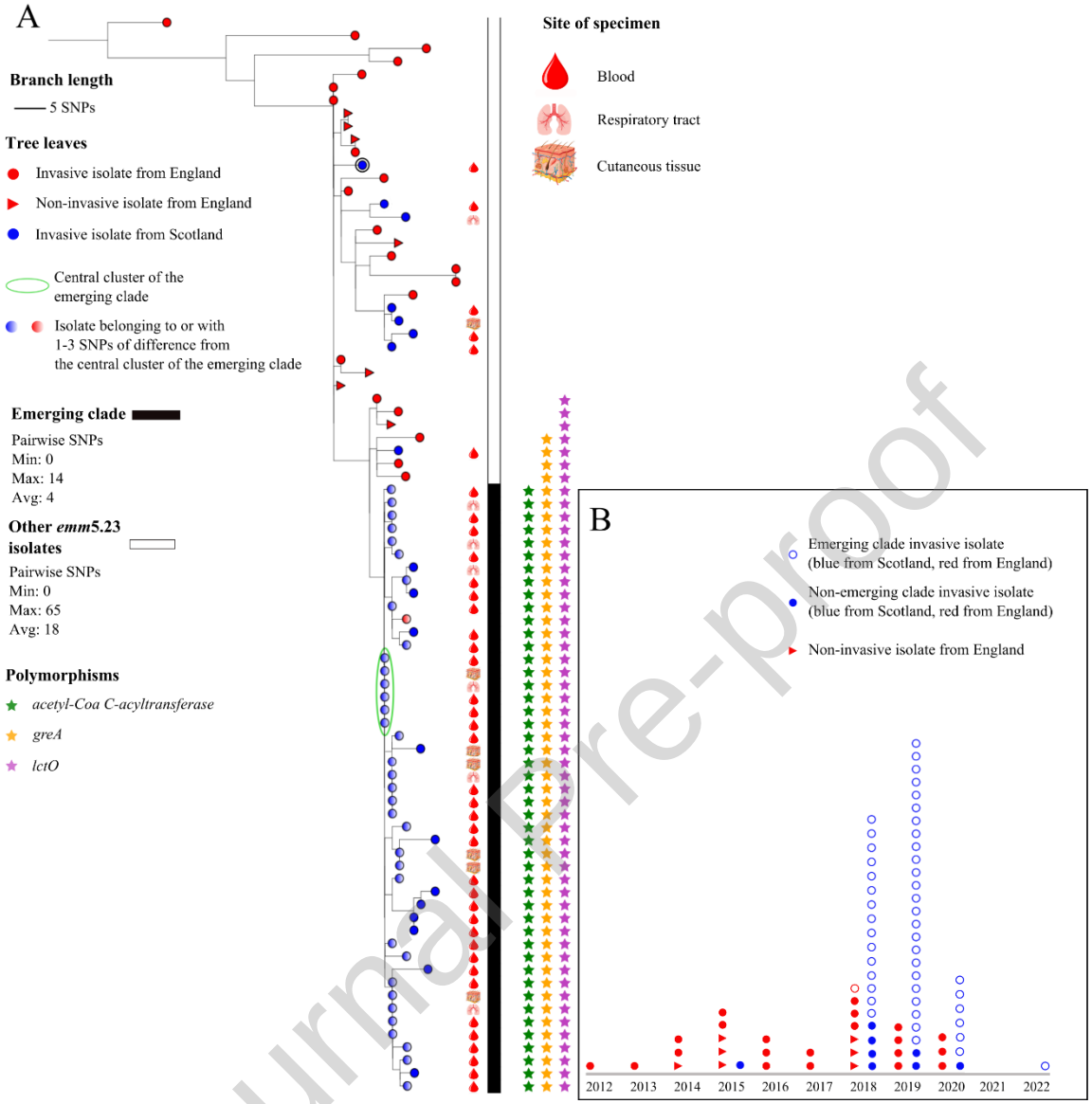

Figure 2. Core genome maximum likelihood single nucleotide polymorphism (SNP)-phylogeny showing the evolutionary relatedness of *emm*5.23 isolates from Scotland and a randomly selected subset of isolates from England. The tree in panel A was midpoint rooted. The isolate used as reference, iGAS.426 (accession number: CP067008), is highlighted on the tree by a black outer circle. For ease of interpretation, the part of the tree dominated by isolates from Scotland sharing the same three polymorphisms compared to the reference is referred to as the "emerging clade". The specimen site is reported next to each isolate from Scotland. Panel B shows temporal distribution of *emm*5.23 strain isolation from Scotland and England.

The strains from Scotland represent nearly all (93%, 54/58) of the invasive isolates confirmed since 2015, providing a reliable representation of the incidence of iGAS disease in Scotland in this figure. In contrast, the strains from England include non-invasive isolates and only a subsample of invasive ones, and are therefore not combined with the strains from Scotland, as they are not representative of disease incidence.

Discussion

Enhanced surveillance data for 45 cases of iGAS *emm*5.23 infection has revealed that the highest disease burden is among individuals aged 45 or older. This only partially corresponds with the typical age distribution observed in epidemiological surveillance of iGAS disease, which usually shows a “J” shaped curve, with one minor peak in infants and one major peak in the elderly aged 75 or older [2]. Clinical presentations indicate that infections tended to be severe, with approximately one-third of cases requiring admission to ICU. The overall mortality rate of 20% was higher than that observed in Scotland for all *emm* types, which ranged from 5.6 to 12.0% between 2018-2023 [42], agreeing with the previously reported higher-than-average mortality rate of *emm*5.23 in England [18]. Nearly 25% of cases developed pneumonia. Although *emm*5 strains are classified as *emm* pattern A-C, which are associated with throat tropism [43], the high rate of pneumonia among *emm*5.23 invasive disease cases is concerning. A possible, though unconfirmed, explanation could be the elevated incidence of influenza during the winter months of 2018–2019 [44]. Risk factors were predominantly health-related, with a significant proportion associated with chronic diseases. Based on the available data, *emm*5.23 cases

tended to be severe, predominantly affecting middle-aged or elderly individuals with underlying health conditions.

Previously described virulence gene profiles associated with hypervirulence in epidemic GAS strains were not detected in the *emm*5.23 population. However, the lack of both transcriptomic data and *in vitro* characterisation of the phenotype of these isolates precludes definitive conclusions being drawn as to the virulence of the *emm*5.23 isolates [5,8,39,45]. All *emm*5.23 and most non-5.23 isolates examined shared the same MGEs, four of which resembled MGEs already described in the historical strain Manfredo. All recent *emm*5 strains appear to have lost the superantigen *speH*-encoding phage Man 3, while acquiring an *sdn*-encoding phage and a *tetM*-encoding ICE, which is integrated into a phage with high sequence similarity to Man 1. The only detected difference in MGE content among the recent *emm*5 strains was that five strains lacked the aforementioned ICE and, consequently, did not carry the tetracycline resistance encoding gene *tetM*. The only other AMR-associated gene, *lmrP*, was found to be present in every isolate. Together, *tetM* and *lmrP* have been associated with potential resistance to macrolides, lincosamides and tetracyclines [37,38]. All *emm*5.23 isolates tested displayed phenotypic resistance to tetracycline but full susceptibility to lincosamides and macrolides, indicating that the presence of the *lmrP* gene did not confer phenotypic resistance to these classes of antimicrobials.

All analysed genomes displayed three non-synonymous polymorphisms in the transcriptional regulators *mga* and *ropB* compared to the historical Manfredo isolate. Although mutations in transcriptional regulators have been linked to an enhanced ability to cause invasive disease in certain GAS *emm* types [41], there is no evidence

of a sustained increase in invasive infections attributed to *emm*5 isolates over the past few decades. However, in other *emm* types, the acquisition of hypervirulent traits and a fitter phenotype with enhanced survival in experimental infections has been linked to the progressive accumulation of multiple genomic mutations [7,9]. Therefore, while the impact of the detected polymorphisms remains uncertain in *emm*5.23, we cannot rule out their potential role in future evolutionary changes that may contribute to increased virulence.

The majority of the *emm*5.23 isolates from Scotland (46/54) also carried three shared polymorphisms with the potential to induce phenotypic changes. The first one was in the *lctO* gene, which is involved in L-lactate metabolism [46]. The second was located in the gene encoding Acetyl-CoA C-Acyltransferase, which is implicated in fatty acid metabolism [47]. Finally, we detected a nucleotide insertion upstream of the *greA* gene, which encodes a transcription elongation factor [48]. These genes are not known to be directly associated with bacterial virulence; hence, their role in pathogenesis remains uncertain. Since the *lctO* gene and the gene encoding Acetyl-CoA C-Acyltransferase are both involved in metabolic pathways, we cannot exclude the possibility that the identified polymorphisms confer a biological fitness advantage. Consistent with this hypothesis, we detected a non-synonymous SNP in the *adhE* gene in all *emm*5.23 isolates from both Scotland and England. Given that this gene also participates in metabolic pathways [49] and is known to promote bacterial adherence in *Streptococcus pneumoniae* strains [50], further investigation is needed to determine its potential functional impact.

Phylogenetic analysis of all *emm*5 genomes revealed that isolates of different subtypes can cluster within the same clade, highlighting the limitations of *emm* subtyping and the importance of whole genome sequencing for determining the relatedness of isolates and investigating transmission events. The *emm*5.23 phylogeny suggests that most of the isolates from Scotland form a monophyletic group of highly similar strains, which all share the *lctO*, *greA* and Acetyl-CoA C-Acyltransferase polymorphisms. Of the Scottish strains that are not part of this emerging clade, most were collected in 2018, a period marked by a general increase in iGAS incidence in Scotland, potentially linked to the high incidence of influenza during the 2018-2019 winter [44]. Conversely, only 30% of the *emm*5.23 isolates from Scotland in the emerging clade were collected in 2018, with the remainder being collected from 2019 onward. Together with the tree topology, this suggests that the emerging clade genotype may have originated from *emm*5.23 strains previously circulating in England and Scotland, gradually replacing the earlier *emm*5.23 genotype in Scotland. Taken together, these findings indicate that 46 of 54 *emm*5.23 isolates responsible for invasive disease in Scotland are phylogenetically closely related and share a relatively recent common ancestor. The tree topology indicated that the central cluster of six strains (circled in green in Figure 2) with an identical core genome and differing by no more than 7 SNPs from the remaining isolates of the clade, can be considered the ancestral genotype of the emerging clade. It should be noted, however, that due to the small number of SNPs informing the phylogeny and the limited number of available sequences it is impossible to resolve the emerging clade to an actual single common ancestral isolate.

One hypothesis to explain the high incidence of *emm*5.23 invasive infections in 2018-2019 could be a lack of herd immunity to this genotype. While there was one invasive disease case in 2015, there was no evidence of its circulation until 2018, when cases began to rise. We hypothesise that before 2018, this genotype had not become widely established due to a combination of its recent introduction, competition with other genotypes, or potentially a low transmission rate. In 2018, it began to be detected more widely, potentially due a combination of factors, including the decline of other *emm* types, increased contact rates, and the concurrent surge in influenza cases [44]. The previously limited exposure of the population to this genotype may have also contributed to its successful spread and disease occurrence. However, as immunological correlates of protection for GAS have not been conclusively demonstrated, the role of acquired immunity in this context requires further investigation[51]. Despite the lack of obvious genomic changes that could account for the high incidence in invasive disease cases in Scotland in 2018-2019, the *emm*5.23 genotype caused high mortality, particularly among “at-risk” individuals. Typically, highly virulent strains cause invasive disease across a broad spectrum of individuals, including those without known risk factors and who are more likely to recover. In contrast, we hypothesise that new strains spreading widely due to low population-level immunity may cause invasive disease primarily in the most susceptible individuals, who are more likely to die from the infection and therefore act as sentinels. Under these circumstances, a strain with average virulence potential could be associated with invasive disease and high mortality rates, at least until a sufficient level of herd immunity is reached. It is also possible that a combination of increased biological fitness associated with the described polymorphisms and a lack of herd immunity contributed to the spread of the

*emm*5.23 genotype. The incidence of *emm*5.23 invasive infections declined after the first national lockdown in March 2020, which may have been due to reduced direct contact among individuals. When social distancing measures were lifted, new genotypes may have replaced *emm*5.23, explaining the lack of new cases in 2023 and 2024. A notable limitation of this study, however, is the lack of concurrent surveillance on non-invasive disease cases in Scotland, which precludes a more detailed understanding of strain transmission, invasiveness, and cross-strain competition.

This investigation serves as a stark reminder of the virulence and high case-fatality rates associated with GAS infections, as well as the potential for novel *emm* types to cause significant public health challenges. The rise in cases was possibly linked to increased influenza transmission in the community during 2018-2019, along with possible higher carriage rates of *emm*5.23 and the exploitation of gaps in immunity or colonisation resistance, leading to severe disease; the absence of links to enclosed settings and the demographics of the cases support this hypothesis. Although limited by its retrospective nature, this investigation highlights the critical importance of an integrated bacterial genomics and public health epidemiology system operating as close to real-time as possible to maximise opportunities for public health interventions and the protection of vulnerable groups.

Funding

This work was funded by the James Herriot Scholarship, School of Biodiversity, One Health and Veterinary Medicine, University of Glasgow.

Acknowledgments

We thank Sarah Rankin and Michael Coyne for guiding us through the AMR testing. We also thank Fiona Johnston for extracting and validating the enhanced surveillance data from the PHS database.

Data availability

Genomic data for the isolates sequenced to conduct this study are available under accession PRJNA1114706. Isolate-specific read accessions are listed in Supplementary Table 2.

References

1 Oliver Jane, Malliya Wadu Erandi, Pierse Nevil, Moreland Nicole J., Williamson Deborah A., Baker Michael G. Group A Streptococcus pharyngitis and pharyngeal carriage: A meta-analysis. *PLoS Neglected Tropical Diseases* 2018;**12**(3). Doi: 10.1371/journal.pntd.0006335.

2 Lamagni Theresa L., Darenberg Jessica, Luca-Harari Bogdan, Siljander Tuula, Efstratiou Androulla, Henriques-Normark Birgitta, et al. Epidemiology of severe *Streptococcus pyogenes* disease in Europe. *Journal of Clinical Microbiology* 2008;**46**(7):2359–67. Doi: 10.1128/JCM.00422-08.

3 UKHSA. UK guidelines for the management of contacts of invasive group A streptococcus (iGAS) infection in community settings 2023.

4 CDC. emm Typing Overview and Guidelines. Streptococcus Laboratory. Available at https://www.cdc.gov/strep-lab/php/group-a-strep/emm-typing.html. Accessed March 30, 2025.

5 Nasser Waleed, Beres Stephen B., Olsen Randall J., Dean Melissa A., Rice Kelsey A., Long S. Wesley, et al. Evolutionary pathway to increased virulence and epidemic group A Streptococcus

disease derived from 3,615 genome sequences. *Proceedings of the National Academy of Sciences of the United States of America* 2014;**111**(17). Doi: 10.1073/pnas.1403138111.

6 Zhu Luchang, Olsen Randall J., Lee Jessica D., Porter Adeline R., DeLeo Frank R., Musser James M. Contribution of Secreted NADase and Streptolysin O to the Pathogenesis of Epidemic Serotype M1 *Streptococcus pyogenes* Infections. *The American Journal of Pathology* 2017;**187**(3):605–13. Doi: 10.1016/j.ajpath.2016.11.003.

7 Lynskey NN, Jauneikaite E, Li HK, Zhi X, Turner CE, Mosavie M, Pearson M, Asai M, Lobkowicz L, Chow JY, Parkhill J. Emergence of dominant toxigenic M1T1 *Streptococcus pyogenes* clone during increased scarlet fever activity in England: a population-based molecular epidemiological study. *The Lancet Infectious Diseases* 2019;19(11):1209-18.

8 Beres Stephen B., Kachroo Priyanka, Nasser Waleed, Olsen Randall J., Zhu Luchang, Flores Anthony R., et al. Transcriptome remodeling contributes to epidemic disease caused by the human pathogen *Streptococcus pyogenes*. *mBio* 2016;**7**(3). Doi: 10.1128/mBio.00403-16.

9 Odo Chioma M., Vega Luis A., Mukherjee Piyali, DebRoy Sruti, Flores Anthony R., Shelburne Samuel A. Emergent emm4 group A Streptococcus evidences a survival strategy during interaction with immune effector cells. *Infection and Immunity* 2024;**92**(7):e00152-24. Doi: 10.1128/iai.00152-24.

10 Davies MR, Holden MT, Coupland P, Chen JH, Venturini C, Barnett TC, Zakour NL, Tse H, Dougan G, Yuen KY, Walker MJ. Emergence of scarlet fever *Streptococcus pyogenes* emm 12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance. *Nature genetics* 2015;47(1):84-7.

11 Athey Taryn B.T., Teatero Sarah, Sieswerda Lee E., Gubbay Jonathan B., Marchand-Austin Alex, Li Aimin, et al. High incidence of invasive group a streptococcus disease caused by strains of uncommon emm types in Thunder Bay, Ontario, Canada. *Journal of Clinical Microbiology* 2016;**54**(1):83–92. Doi: 10.1128/JCM.02201-15.

12 Southon Sara B, Beres Stephen B, Kachroo Priyanka, Ojeda Saavedra Matthew, Erlendsdóttir Helga, Haraldsson Gunnsteinn, Yerramilli Prasanti, Pruitt Layne, Zhu Luchang, Musser James M, Kristinsson Karl G. Population Genomic Molecular Epidemiological Study of Macrolide-Resistant *Streptococcus pyogenes* in Iceland, 1995 to 2016: Identification of a Large Clonal Population with a pbp2x Mutation Conferring Reduced In Vitro-Lactam Susceptibility. *Journal of Clinical Microbiology* 2020;**58**(9):10-1128.

13 Harkness Gail A, Bentley David W, Mottley Milwood, Lee June. *Streptococcus pyogenes* outbreak in a long-term care facility. *American Journal of Infection Control* 1992;20(3):142–8.

14 Ruben Frederick L, Norden Carl W, Heisler Barbara, Korica Yvonne. An Outbreak of *Streptococcus pyogenes* Infections in a Nursing Home. *Annals of Internal Medicine* 1984;101(4):494-496.

15 Beres SB, Carroll RK, Shea PR, Sitkiewicz I, Martinez-Gutierrez JC, Low DE, McGeer A, Willey BM, Green K, Tyrrell GJ, Goldman TD. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proceedings of the National Academy of Sciences* 2010;107(9):4371-6.

16 Steer Andrew C, Law Irwin, Matatolu Laisiana, Beall Bernard W, Carapetis Jonathan R. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. *Lancet Infectious Diseases* 2009:611–6.

17 Gossain S, Chalker V, Kapatai G, Coelho J, Mohamed H, Martin K, Tahir M. A protracted outbreak of invasive group a streptococcal infection at a UK long-term facility investigated using whole genome sequencing. InOpen Forum Infectious Diseases 2016 (Vol. 3, No. suppl_1, p. 1411). Oxford University Press.

18 Blagden S., Watts V., Verlander N. Q., Pegorie M. Invasive group A streptococcal infections in North West England: epidemiology, risk factors and fatal infection. *Public Health* 2020;**186**:63–70. Doi: 10.1016/j.puhe.2020.06.007.

19 Scottish Microbiology Reference Laboratories (SMiRL) https://www.nhsggc.scot/staff-recruitment/staff-resources/laboratory-medicine/scottish-microbiology-reference-laboratories/. Accessed March 30, 2025.

20 National Records of Scotland. Scotland's Population 2019. National Records of Scotland. Available at https://www.nrscotland.gov.uk/publications/rebased-population-estimates-scotland-mid-2011-to-mid-2021/. Accessed March 30, 2025.

21 EUCAST. eucast: MIC determination. Available at https://www.eucast.org/ast_of_bacteria/mic_determination. Accessed March 30, 2025.

22 EUCAST. eucast: Clinical breakpoints and dosing of antibiotics. Available at https://www.eucast.org/clinical_breakpoints. Accessed March 30, 2025.

23 Smeds Linnéa, Künstner Axel. ConDeTri - a content dependent read trimmer for illumina data. *PLoS ONE* 2011;**6**(10). Doi: 10.1371/journal.pone.0026314.

24 Bankevich A, Nurk S, Antipov D, Gurevich AA, Dvorkin M, Kulikov AS, Lesin VM, Nikolenko SI, Pham S, Prjibelski AD, Pyshkin AV. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *Journal of computational biology* 2012;19(5):455-77.

25 Gurevich Alexey, Saveliev Vladislav, Vyahhi Nikolay, Tesler Glenn. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 2013;**29**(8):1072–5. Doi: 10.1093/bioinformatics/btt086.

26 Inouye M, Dashnow H, Raven LA, Schultz MB, Pope BJ, Tomita T, Zobel J, Holt KE. SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome medicine* 2014;6:1-6.

27 Altschul Stephen F., Gish Warren, Miller Webb, Myers Eugene W., Lipman David J. Basic local alignment search tool. *Journal of Molecular Biology* 1990;**215**(3):403–10. Doi: 10.1016/S0022-2836(05)80360-2.

28 Pagnossin D, Smith A, Weir W, Crestani C, Lindsay D, Ure R, Oravcova K. Complete Genome Sequences of Three Invasive Strains of Streptococcus pyogenes Subtype emm 5.23 Isolated in Scotland. *Microbiology Resource Announcements* 2021;10(15):10-128.

29 Hunt M, Mather AE, Sánchez-Busó L, Page AJ, Parkhill J, Keane JA, Harris SR. ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial genomics* 2017;3(10):e000131.

30 Chen Lihong, Xiong Zhaohui, Sun Lilian, Yang Jian, Jin Qi. VFDB 2012 update: Toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Research* 2012;**40**(D1). Doi: 10.1093/nar/gkr989.

31 Alcock BP, Raphenya AR, Lau TT, Tsang KK, Bouchard M, Edalatmand A, Huynh W, Nguyen AL, Cheng AA, Liu S, Min SY. CARD 2020: antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic acids research* 2020;48(D1):D517-25.

32 Seemann Torsten. Snippy 2014.

33 Holden MT, Scott A, Cherevach I, Chillingworth T, Churcher C, Cronin A, Dowd L, Feltwell T, Hamlin N, Holroyd S, Jagels K, Moule S, Mungall K, Quail MA, Price C, Rabbinowitsch E, Sharp S, Skelton J, Whitehead S, Barrell BG, Kehoe M, Parkhill J. Complete Genome of Acute Rheumatic Fever-Associated Serotype M5 Streptococcus pyogenes Strain Manfredo. *Journal of Bacteriology* 2007;**189**(4):1473–7. Doi: 10.1128/jb.01227-06.

34 Kalyaanamoorthy Subha, Minh Bui Quang, Wong Thomas K.F., Von Haeseler Arndt, Jermiin Lars S. ModelFinder: Fast model selection for accurate phylogenetic estimates. *Nature Methods* 2017;**14**(6):587–9. Doi: 10.1038/nmeth.4285.

35 Letunic Ivica, Bork Peer. Interactive tree of life (iTOL) v5: An online tool for phylogenetic tree display and annotation. *Nucleic Acids Research* 2021;**49**(W1):W293–6. Doi: 10.1093/nar/gkab301.

36 Kondo Kohei, Kawano Mitsuoki, Sugai Motoyuki. Distribution of Antimicrobial Resistance and Virulence Genes within the Prophage-Associated Regions in Nosocomial Pathogens. *mSphere* 2021;**6**(4):10.1128/msphere.00452-21. Doi: 10.1128/msphere.00452-21.

37 Putman M, van Veen HW, Degener JE, Konings WN. The lactococcal secondary multidrug transporter LmrP confers resistance to lincosamides, macrolides, streptogramins and tetracyclines. *Microbiology* 2001;147(10):2873-80.

38 Morse SA, Johnson SR, Biddle JW, Roberts MC. High-level tetracycline resistance in Neisseria gonorrhoeae is result of acquisition of streptococcal tetM determinant. *Antimicrobial agents and chemotherapy* 1986;30(5):664-70.

39 Beres SB, Sylva GL, Barbian KD, Lei B, Hoff JS, Mammarella ND, Liu MY, Smoot JC, Porcella SF, Parkins LD, Campbell DS. Genome sequence of a serotype M3 strain of group A Streptococcus: phage-encoded toxins, the high-virulence phenotype, and clone emergence. *Proceedings of the national academy of sciences* 2002;99(15):10078-83.

40 Turner Claire E., Holden Matthew T. G., Blane Beth, Horner Carolyne, Peacock Sharon J., Sriskandan Shiranee. The Emergence of Successful *Streptococcus pyogenes* Lineages through Convergent Pathways of Capsule Loss and Recombination Directing High Toxin Expression. *mBio* 2019;**10**(6):10.1128/mbio.02521-19. Doi: 10.1128/mbio.02521-19.

41 Ikebe T, Ato M, Matsumura T, Hasegawa H, Sata T, Kobayashi K, Watanabe H. Highly frequent mutations in negative regulators of multiple virulence genes in group A streptococcal toxic shock syndrome isolates. *PLoS pathogens* 2010 Apr 1;6(4):e1000832.

42 PHS. Group A Streptococcus (GAS) quarterly surveillance report. Available at https://publichealthscotland.scot/publications/group-a-streptococcus-gas-quarterly-surveillance-report/group-a-streptococcus-gas-quarterly-surveillance-report-january-to-april-2024/#section-2-4. Accessed July 16, 2024, 2024.

43 Bessen Debra E., Smeesters Pierre R., Beall Bernard W. Molecular Epidemiology, Ecology, and Evolution of Group A Streptococci. *Microbiology Spectrum* 2018;**6**(5). Doi: 10.1128/microbiolspec.cpp3-0009-2018.

44 PHE. Annual flu reports. GOV.UK. Available at https://webarchive.nationalarchives.gov.uk/ukgwa/20220401215804mp_/https://www.gov.uk/government/statistics/annual-flu-reports. Accessed March 30, 2025.

45 Kachroo P, Eraso JM, Beres SB, Olsen RJ, Zhu L, Nasser W, Bernard PE, Cantu CC, Saavedra MO, Arredondo MJ, Strope B. Integrated analysis of population genomics, transcriptomics and

virulence provides novel insights into *Streptococcus pyogenes* pathogenesis. *Nature genetics* 2019;51(3):548-59.

46 Taniai H, Iida KI, Seki M, Saito M, Shiota S, Nakayama H, Yoshida SI. Concerted action of lactate oxidase and pyruvate oxidase in aerobic growth of *Streptococcus pneumoniae*: role of lactate as an energy source. *Journal of bacteriology* 2008;190(10):3572-9.

47 Röttig Annika, Steinbüchel Alexander. Acyltransferases in Bacteria. *Microbiology and Molecular Biology Reviews* 2013;**77**(2):277–321. Doi: 10.1128/mmbr.00010-13.

48 Yuzenkova Y, Gamba P, Herber M, Attaiech L, Shafeeq S, Kuipers OP, Klumpp S, Zenkin N, Veening JW. Control of transcription elongation by GreA determines rate of gene expression in *Streptococcus pneumoniae. Nucleic acids research* 2014 Sep 29;42(17):10987-99.

49 Pony Pauline, Rapisarda Chiara, Terradot Laurent, Marza Esther, Fronzes Rémi. Filamentation of the bacterial bi-functional alcohol/aldehyde dehydrogenase AdhE is essential for substrate channeling and enzymatic regulation. *Nat Commun* 2020;**11**(1):1426. Doi: 10.1038/s41467-020-15214-y.

50 Luong TT, Kim EH, Bak JP, Nguyen CT, Choi S, Briles DE, Pyo S, Rhee DK. Ethanol-induced alcohol dehydrogenase E (AdhE) potentiates pneumolysin in *Streptococcus pneumoniae*. *Infection and immunity* 2015 Jan;83(1):108-19.

51 Frost Hannah, Excler Jean-Louis, Sriskandan Shiranee, Fulurija Alma. Correlates of immunity to Group A Streptococcus: a pathway to vaccine development. *Npj Vaccines* 2023;**8**(1):1–7. Doi: 10.1038/s41541-022-00593-8.

Declaration of Interest Statement

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Highlights

- Newly emerged invasive *S. pyogenes emm*5.23 caused high mortality rates in Scotland
- Concerningly, 24% of patients with iGAS *emm*5.23 developed pneumonia in Scotland
- Most *emm*5.23 iGAS strains are highly clonal and share a recent common ancestor
- The shared mutations in *emm*5.23 iGAS strains have unclear relevance to virulence
- Increased fitness of *emm*5.23 or lack of herd immunity may have driven its spread