Rapid expansion and international spread of M1UK in the post-pandemic upsurge of Streptococcus pyogenes infections in United Kingdom

Shiranee Sriskandan

s.sriskandan@imperial.ac.uk

Imperial College London https://orcid.org/0000-0002-5214-4941

Ana Vieira

Imperial College London

Yu Wan

Imperial College London

Yan Ryan

UKHSA

Ho Kwong Li

Imperial College London

Rebecca L. Guy

UKHSA

Maria Papangeli

Imperial College London

Kristin Huse

Imperial College London

Lucy C. Reeves

Imperial College London

Valerie WC Soo

Imperial College London

Roger Daniel

UKHSA

Alessandra Harley

UKHSA

Karen Broughton

UKHSA

Chenchal Dhami

UKHSA

Mark Ganner

UKHSA
Marjorie Ganner
UKHSA
Zaynab Mumin
UKHSA
Maryam Razaei
UKHSA
Emma Rundberg
UKHSA
Rufat Mammadox
UKHSA
Ewurabena A. Mills
Imperial College London
Vincenzo Sgro
Imperial College London
Kai Yi Mok
Imperial College London
Xavier Didelot
University of Warwick https://orcid.org/0000-0003-1885-500X
Nicholas Croucher
Imperial College London https://orcid.org/0000-0001-6303-8768
Elita Jauneikaite
Imperial College London https://orcid.org/0000-0002-7075-6896
Theresa Lamagni
UKHSA
Colin S Brown
UKHSA
Juliana Coelho
UKHSA

Article

Keywords:

Posted Date: January 18th, 2024

DOI: https://doi.org/10.21203/rs.3.rs-3842890/v1

Additional Declarations: There is **NO** Competing Interest.

Version of Record: A version of this preprint was published at Nature Communications on May 10th, 2024. See the published version at https://doi.org/10.1038/s41467-024-47929-7.

Rapid expansion and international spread of $M1_{UK}$ in the post-pandemic upsurge of *Streptococcus pyogenes* infections in United Kingdom

Ana Vieira[1,2,3], Yu Wan[1,3,4], Yan Ryan[3,4], Ho Kwong Li[1,2], Rebecca L Guy[4], Maria Papangeli[1,2,] Kristin K Huse[1,2], Lucy C Reeves[1,2], Valerie WC Soo[1,2], Roger Daniel[4], Alessandra Harley[4], Karen Broughton[4], Chenchal Dhami[4], Mark Ganner[4], Marjorie Ganner[4], Zaynab Mumin[4], Maryam Razaei[4], Emma Rundberg[4], Rufat Mammadov[4], Ewurabena A Mills[1,2], Vincenzo Sgro[1], Kai Yi Mok[1], Xavier Didelot[5], Nicholas J Croucher[6,7], Elita Jauneikaite[3,6,7], Theresa Lamagni[3,4], Colin S Brown[3,4], Juliana Coelho[†3.4], Shiranee Sriskandan[†1,2,3]

1.Department of Infectious Disease, Imperial College London, London, UK

2.Centre for Bacterial Resistance Biology, Imperial College London, London, UK

3.NIHR Health Protection Research Unit in Healthcare-associated Infections and AMR, Imperial College London, London, UK

4. Healthcare-Associated Infections, Fungal, AMR, AMU, and Sepsis Division, UK Health Security Agency, London, UK

5. School of Life Sciences and Department of Statistics, University of Warwick, Coventry, UK

6. School of Public Health, Imperial College London, London, UK

7. MRC Centre for Global Infectious Disease Analysis, Imperial College London, London, UK

†Addresses for correspondence.

Shiranee Sriskandan, Section of Adult Infectious Diseases, Faculty of Medicine, Imperial College London, Hammersmith Campus, Du Cane Road, London W12 0NN, UK. s.sriskandan@imperial.ac.uk

Juliana Coelho, Healthcare-Associated Infection, Fungal, AMR, AMU, and Sepsis Division, UK Health Security Agency, 61 Colindale Avenue, London, NW9 5DF, UK. Juliana.coelho@ukhsa.gov.uk

Abstract

The UK observed a marked increase in scarlet fever and invasive group A streptococcal infection in 2022 with severe outcomes in children and similar trends worldwide. Here we report lineage $M1_{UK}$ to be the dominant source of invasive infections in this upsurge and associated with pleural empyema. Compared with ancestral $M1_{global}$ strains, invasive $M1_{UK}$ strains exhibited reduced genomic diversity and fewer mutations in two-component regulator genes *covRS*. The emergence of $M1_{UK}$ was dated to 2008. Following a bottleneck coinciding with the COVID-19 pandemic, three emergent $M1_{UK}$ clades underwent rapid nationwide expansion, despite lack of detection in previous years. All $M1_{UK}$ isolates thus-far sequenced globally have a phylogenetic origin in the UK, with dispersal of the new clades in Europe. While waning immunity may promote streptococcal epidemics, the genetic features of $M1_{UK}$ point to a fitness advantage in pathogenicity, and a striking ability to persist through population bottlenecks.

Introduction

Group A Streptococcus (GAS, *Streptococcus pyogenes*) is a human-restricted pathogen causing diseases ranging from sore throat and scarlet fever to more serious invasive infections, including soft tissue infections, pneumonia, and toxic shock, as well as auto-immune sequelae[1]. Although advanced age and specific presentations such as necrotising fasciitis increase risk of death from invasive infection, the genetic background of *S. pyogenes* strains also contributes to the risk of mortality[2,3] underlining the role of strain genotype and virulence in disease outcome. Among more than 250 recognised *emm* types, the *emm*1 genotype is most frequently associated with invasive infections in high-income countries[4]. *emm1* strains are considered highly virulent[5,6] and often acquire inactivating mutations in the *covRS* two component regulator, which de-represses key virulence factors during invasive infection[7]. In the 1980s, *emm1* emerged as a leading cause of invasive infection following

several genomic changes that altered phage content and streptolysin O (SLO) expression, leading to a new clone that spread globally[8].

In England, prompt notification and antibiotics are advocated for scarlet fever and invasive GAS (iGAS) infections[9], however guidelines that recommend a non-treatment or delayed treatment approach to sore throat were introduced in 2008, to limit unnecessary use of antibiotics[10]. Unexpectedly large seasonal upsurges in scarlet fever were documented annually in England between 2014-2018[11,12] coinciding with the expansion and recognition of a new lineage of *emm*1 termed $M1_{UK}$ among *S. pyogenes* isolates[5]. $M1_{UK}$ differed from other globally circulating *emm1* strains[8] (hereafter referred to as $M1_{global}$) by 27 signature SNPs and was characterised by increased expression of the scarlet fever toxin, streptococcal pyrogenic exotoxin A (*speA*)[5,6,13]. Two intermediate lineages, $M1_{13SNPs}$ and $M1_{23SNPs}$, that share subsets of the 27 SNPs, were also identified[5,6]. $M1_{23SNPs}$ expresses SpeA at the same level typical of $M1_{UK,}$ whereas $M1_{13SNPs}$ does not[6]. By 2016, the $M1_{UK}$ lineage represented 84% of all *emm1* invasive strains in England[5], increasing to 91.5% by 2020[14].

Onset of the COVID-19 pandemic, and implementation of non-pharmaceutical interventions (NPI) to limit SARS-CoV2 transmission triggered a reduction in scarlet fever and iGAS notifications in 2020[12]. However, in late 2022, a highly pronounced out-of-season upsurge in both scarlet fever and iGAS cases was reported in England, with unexpected increase in paediatric pleural empyema and several fatalities[15]. Similar increases in severe paediatric iGAS infections were reported worldwide[16].

In this article we show that the *S. pyogenes* upsurge in England and Wales was predominantly associated with $M1_{UK}$, a lineage we estimate to have emerged around 2008, and, in particular, three emergent clades that are now widely dispersed. The expansion of $M1_{UK}$ occurred following a bottleneck in growth, likely related to reduced transmission during the COVID-19 pandemic.

Results

Trends in *S. pyogenes*-positive samples, England 2016-2023

S. pyogenes identified from non-sterile and sterile site samples are recorded through a national laboratory reporting system (Second Generation Surveillance System, SGSS). The typical pattern of seasonal spring-time peaks (Q1-Q2) in *S. pyogenes* infections was interrupted abruptly in April 2020, coinciding with NPI introduced at the onset of the COVID-19 pandemic (Figure 1). A profound reduction in *S. pyogenes*-positive samples, from both sterile and non-sterile sites, lasted almost two years, ending in Q1 2022. Following cessation of widespread NPI in February 2022, a delayed seasonal increase in microbiologically-confirmed *S. pyogenes* infections returned in April 2022, subsiding only in Q3 2022, in keeping with the UK summer vacation period. Unexpectedly, a second, exponential increase in *S. pyogenes* samples occurred in Q4 of 2022 (Figure 1). This marked increase in microbiologically-confirmed infections peaked in week 49, when 9388 non-sterile site and 241 sterile site *S. pyogenes*-positive samples were recorded (Figure 1), coinciding with increased disease notifications[15,17].

S. pyogenes isolates cultured from iGAS cases are submitted to the national reference laboratory for *emm* typing. Between Q1 of January 2017 and Q1 of 2020, *emm*1 was the leading cause of iGAS, responsible for 16-28% of all iGAS cases; *emm*1 dominance was greater in children than adults (Figure 2). During the period of COVID-19-related NPI, annual iGAS isolates reduced ~6.5-fold in children (274 isolates/year 2017-2019; 44 in 2021) and ~2.5-fold in adults (1944 isolates/year 2017-2019; 785 in 2021) (Figure 2). The proportion of iGAS isolates that were *emm*1 also reduced significantly ($p < 0.00001$), to less than 8% of all iGAS cases. From Q1 of 2022, *emm*1 then showed a sustained quarterly increase in frequency, peaking in Q1 of 2023. For over nine months, *emm*1 accounted for > 50% of all iGAS cases, coinciding with the period of upsurge (Figure 2). Indeed, *emm*1 was the only genotype to expand significantly during this time, increasing from 20% to 55%. In children (<15

years), this increase was more apparent; *emm*1 accounted for 60% and 70% of iGAS in the same period (Figure 2).

Population genomics of *emm*1 *S. pyogenes* strains comprising the upsurge

To investigate any genetic basis for the increase in *emm*1 iGAS cases, genomes of all 1092 iGAS *emm1* isolates submitted to the reference laboratory from January 2022 to March 2023 were whole genome sequenced. Phylogenetic analysis revealed clustering of *emm*1 genomes into expected lineages. The vast majority (1001/1092, 91.5%) of isolates were $M1_{UK}$, 4.1% (44/1092) were derivatives of $M1_{UK}$ having lost the phi5005.3 phage (and therefore lacking the phage portal protein SNP that is typical but not essential to $M1_{UK}$) and 4.2% (46/1092) were $M1_{global.}$ Taken together, 95.7% of all *emm*1 strains from the upsurge period were $M1_{UK}$ or a derivative thereof, representing overall expansion of the lineage since 2020 (Figure 3A). Isolates from 2022/2023 were further compared to 723 *emm*1 iGAS strains sequenced in the same reference laboratory between 2013-2021 to determine evidence for recent genomic change. Phylogenetic analysis of these 1815 *emm*1 *S. pyogenes* genomes associated with iGAS showed $M1_{UK}$ isolates from 2022/2023 to be broadly distributed across the pre-existing $M1_{UK}$ population, with three emergent dominant clades and several small clades formed almost exclusively of isolates from 2022/2023 (Figure 3A). Three clades accounted for over half (54.8%) of all $M1_{UK}$ from 2022/2023. Clade 1 comprised 123 invasive strains exclusively from 2022/2023 and was characterised by two SNPs (Ext.Data Table 1). Clade 2 comprised 166 invasive strains exclusively from 2022/2023 and was characterised by 6 SNPs, including three non-synonymous mutations (in *sic1*.01, *pyrC* and M5005_Spy1146). Clade 3 comprised 284 strains from 2022/2023, plus a single strain collected in February 2020, and was defined by 3 non-synonymous mutations (in *xerD*, *huTu* and *secA*). Clade 3 was enriched by invasive strains collected in southern England (70%), consistent with regional transmission. In contrast, Clades 1 and 2 had similar proportions of strains from northern (26% and 35%), southern (43% and 35%), and central regions including Wales (23% and 28%) consistent with a wider national outbreak (Figure 3A). The average genetic distance between any two strains from

Clade 1 was just 2 SNPs, while for Clades 2 and 3, the average was just 3 SNPs (Ext.Data Table 2). The low diversity was consistent with rapid emergence and dispersion through the year and across the country from a recent common ancestor.

Among the 1815 *emm*1 genomes associated with iGAS from 2013-2023, the clinical sources of isolates were known for most strains: 67.8% (1231/1815) were blood isolates; 6.9% (124/1815) were lower respiratory tract isolates, of which 71.8% (89/124) were pleural sample isolates, indicative of empyema (Ext.Data Table 3). Overall, a significantly higher proportion of $M1_{UK}$ (5.0%) isolates were associated with pleural samples compared to $M1_{global}$ (2.6%), in particular Clade 3 (8.4%) (Ext.Data Table 4). Considering only disease occurring in 2022/2023, inter-lineage differences were not significant, however $M1_{global}$ isolate numbers were very low (Ext.Data Table 4). Pleural sample isolates were notably more frequent at the time of the upsurge. Despite the notable impact of the upsurge on children, no single clade was uniquely associated with a specific age group, and closely related strains (<3 SNPs apart) caused invasive infections in both adults and children (Ext.Data Figure 1).

The average pairwise distance within $M1_{UK}$ increased from 16 SNPs in 2013-2021 to 22 SNPs in 2022/2023, while the average pairwise distance within the $M1_{global}$ lineage increased from 39 SNPs in 2013-2021 to 55 SNPs in 2022/2023 (Ext.Data Table 2). Despite the recent increase in the genetic diversity of both lineages ($M1_{global}$ and $M1_{UK}$), $M1_{UK}$ showed greater genomic stability (point mutations) than $M1_{global}$. Most mutations (excluding the 27 $M1_{UK}$ signature SNPs) were unique to individual strains outside the main clades (Ext.Data Figure 2) consistent with a rapid population size expansion. The four indels previously reported[13] were present in 99% of $M1_{UK}$ isolates but were not lineage specific (Supplementary Data 1).

Recombination and pangenome analyses showed little evidence of gain or loss of transferable elements between $M1_{UK}$ and $M1_{global}$, and no genomic feature(s) associated only with $M1_{UK}$ from 2022/2023, or $M1_{global}$ from 2022/2023, or the three $M1_{UK}$ clades previously described.

Most strains had three prophages typical of *emm*1: Φ5005.1 that encodes *speA*; Φ5005.2 that encodes *spd3* or *spd4*; and Φ5005.3 that encodes another DNase, *sdaD2/sda1*, reported to contribute virulence to modern $M1_{global}$ strains[8]. Although $M1_{UK}$ strains are characterised by increased SpeA expression, 9/1552 (0.6%) invasive $M1_{UK}$ strains had a partial deletion of phage Φ5005.1 including *speA* (Ext.Data Table 5). Furthermore, 43/1552 (2.8%) invasive $M1_{UK}$ strains had lost Φ5005.3 and consequently cannot express *sdaD2/sda1*. Prophage Φ370.1 containing *speC* and *spd1* was present in ~10% (174/1815) of *emm*1 strains; 9% (139/1552) in $M1_{UK}$ and 16% (31/189) in $M1_{global}$. Only 4/1815 *emm*1 strains (one $M1_{global}$ and three $M1_{UK}$) from 2014-2020 had the ΦSP1380.vir phage (with *speC, ssa, spd1*) reported in Australia[13] and Hong Kong[18].

*Emm*1 invasiveness has been associated with regulatory gene mutations *in vivo*[7]. Among iGAS clinical isolates, mutations in the two component regulatory genes *covR* and *covS* were significantly more frequent in $M1_{global}$ (8.7% and 22% respectively) than $M1_{UK}$ (3% and 7%) (*covR* p-value < 0.00001; *covS* p-value < 0.00001) (Figure 3B), pointing to greater selection pressure on $M1_{global}$ strains during invasive infection. This difference in the frequency of *covS* mutations could not be replicated by *in vivo* passage of pharyngitis (non-invasive) $M1_{global}$ and $M1_{UK}$ isolates in mice, although only five strains from each lineage were tested using intramuscular inoculation (Ext.Data Figure 3). Mutations in *rgg*1 and *rgg*4 were frequent, but not different between lineages (Ext.Data Figure 4). The frequency of resistance to common antimicrobials among *emm*1 isolates was low (<0.5%); furthermore, *pbp2x* missense mutations (T553K and P601L)[19,20] were absent in our dataset (Supplementary Data 2).

Relationship between non-invasive and invasive *emm*1 isolates

To extend our analysis to include the reservoir of non-invasive pharyngitis *S. pyogenes* isolates, we sequenced 133 *emm*1 strains collected sequentially from pharyngitis cases in west London in 2022-2023. 14.3% (19/133) of *emm*1 throat isolates were $M1_{global}$ while 85.7% (112/133) were either $M1_{UK}$ or $M1_{UK}$ without the phi5005.3 phage (2/133). Interestingly, the

proportion of non-invasive and invasive $M1_{global}$ isolates was higher in London than observed nationally during the same period. Phylogenetic analysis of invasive and non-invasive isolates showed that non-invasive $M1_{UK}$ isolates from west London clustered mostly within Clade 3 (62/111, 55.9%), with other isolates scattered throughout the wider $M1_{UK}$ population, including Clade 1 (8/111, 7.2%) and Clade 2 (4/111, 3.6%) (Figure 4). The average number of mutations between two isolates from the same clade (Clade 1, 2 or 3) was 2-3 SNPs. 48% (64/133) of non-invasive isolates were found to be identical to at least one invasive isolate (0 SNPs apart, Figure 4). Point mutations in bacterial regulatory genes in non-invasive *emm*1 isolates were rare (<5%), in comparison to invasive isolates. 5/133 (4%) of non-invasive isolates collected in London in 2022 had the ΦSP1380.vir phage.

Time and place of emergence of $M1_{UK}$ and intermediate lineages

To elucidate the origin and time of emergence of the $M1_{UK}$ lineage, a dated phylogenetic tree was constructed using a newly sequenced $M1_{UK}$ reference strain H1490 (NCTC14935). The tree comprised 2365 $M1_{UK}$ and intermediate ($M1_{13SNPs}$, $M1_{23SNPs}$ and $M1_{26SNPs}$) genomes collected from Europe (Denmark[21], Iceland[21], Netherlands[22]), United Kingdom[5,23], plus the isolates from the current study, North America (Canada[24] and USA[25]), and Australia[13] between March 2005 and July 2023. This showed $M1_{13SNPs}$ and $M1_{23SNPs}$ to share a common ancestor with the $M1_{UK}$ lineage, while $M1_{26SNPs}$ are derivatives of $M1_{UK}$ that have lost the Φ5005.3 phage (Figure 5A). According to the inferred ancestral dates in the tree, the $M1_{13SNPs}$ lineage diverged in 2002 (95% confidence interval (CI): 2000–2004), followed by $M1_{23SNPs}$ in 2006 (95% CI: 2004–2007), and $M1_{UK}$ in 2008 (95% CI: 2006–2009), prior to rapid expansion. The genome-wide mutation rate was estimated to be 1.49 nucleotide substitutions per year.

Ancestral state reconstruction of geographical locations was limited to those regions that undertake and report sequencing of *S. pyogenes*; this revealed that $M1_{UK}$, $M1_{13SNPs}$ and $M1_{23SNPs}$ originated in the UK and then dispersed, with multiple independent introductions into Australia, North America, Netherlands, Iceland, and Denmark (Figures 5A-B). Denmark and

UK strains collected in 2022-2023 were dispersed within the $M1_{UK}$ circulating population, including Clade 3, while almost all 2023 Iceland isolates grouped together in Clade 2.

Bayesian inference of the $M1_{UK}$ effective population size through time in the UK demonstrated rapid population growth of $M1_{UK}$ from 2008 until 2015, followed by a progressive decline until 2019, and then a sharp decline in early 2020 (Figure 5C). Strikingly, the population dynamics suggested a transmission bottleneck in $M1_{UK}$ during implementation of severe NPI designed to limit spread of COVID-19 (April 2020 –March 2021). The mean effective population size over this period dropped to one-fifth of the pre-pandemic maximum and then rose steeply after lifting of lockdown and other NPI measures. Importantly the inferred patterns of population growth and decline were not driven by any variation in the number of sequenced $M1_{UK}$ isolates in the UK through time (Ext.Data Figure 5).

Discussion

The marked increase in bacteriologically confirmed *S. pyogenes* infections in England in late 2022-2023 coincided with the reported national upsurge in notifications of both scarlet fever and iGAS[15,16]. The upsurge in invasive infections was clearly associated with a significant increase in *emm*1 *S. pyogenes* only, the vast majority (95.7%) of which belonged to the emergent $M1_{UK}$ lineage or its derivatives. No substantial genomic changes in $M1_{UK}$ were observed during the upsurge, but three new clades emerged and expanded within $M1_{UK}$, accounting for 53% of *emm*1 iGAS in 2022-2023.

Several countries have now reported similar iGAS upsurges in the period 2022-2023, chronologically associated with the end of mitigation strategies implemented during the COVID-19 pandemic[21, 26-29]. The marked increase in iGAS observed in the second half of 2022 in England was accounted for by *emm*1*,* however *emm*12 infections were prominent in early 2022[13], similar to Portugal[29]. The dominance of *emm*1 among invasive isolates (>50% overall,

and almost 70% in children) is unprecedented in UK records. In contrast, during the period of the COVID-19 pandemic-related NPI in 2020-2022, bacteriologically-confirmed *S. pyogenes* infections were rare. While reduction in non-invasive infection detection might be explained by a reduction in consultations, this would not explain the reduction in sterile site isolates. Furthermore, during the period of COVID-19 NPI, invasive infections due to *emm*1 were exceedingly rare with no *emm*1 isolates identified during some quarters of 2020-2021 in either adults or children. We posit this points to differential modes of transmission, whereby 'throat specialist' strains[30] such as *emm*1 require respiratory transmission in order to circulate, while others may spread via skin contact.

The reported increase in iGAS in late 2022 was particularly evident in children, with complicated clinical presentations including meningitis[28] and, specifically, rapidly progressive pleural empyema in countries where such data are collected[14,29]. Isolates from empyema are often not cultured due to antibiotic pre-treatment, hence the pleural sample isolates in the current study represent a subset of all pleural empyema cases. Regardless, pleural isolates were significantly associated with $M1_{UK}$ overall, and with clade 3 in particular. Timing of the upsurge in Q3 2022 is very likely to have contributed to the pleural empyema phenotype; respiratory viral infections were identified in 25% of paediatric cases of empyema[14], playing a potential role in progression to lower respiratory tract infection.

$M1_{UK}$ is increasingly dominant in the UK. Our findings are mirrored to different degrees in other countries, where the proportion of *emm*1 isolates that are $M1_{UK}$ ranges from 44%-82%[21,28,29]. The fitness of $M1_{UK}$ has been attributed to its ability to express SpeA, a superantigen that can promote pharyngeal infection[5]. Increased SpeA is associated with a SNP in the leader sequence of *ssrA*[13], which is present in not only $M1_{UK}$ but also the near-extinct intermediate $M1_{23SNPs}$ lineage. The contraction of the $M1_{23SNPs}$ lineage suggests that additional fitness advantages prevail in $M1_{UK}$[6]. Genome stability appeared greater in $M1_{UK}$ than $M1_{global}$, suggesting the accumulated 27 SNPs in $M1_{UK}$ may be sufficient to confer a fitness advantage

during human infection, including increased transmissibility. Indeed, in one study, mean secondary attack rate was 40% among asymptomatic contacts of $M1_{UK}$ infection in two classes of schoolchildren, compared with 22.8% in classroom outbreaks involving different *emm* types[31], supporting a potential transmission advantage. In the current study, $M1_{UK}$ invasive isolates were significantly less likely to exhibit mutations in *covRS* than $M1_{global}$ strains, suggesting a fitness advantage in invasive infection as well. Although we were unable to reproduce this difference experimentally, the intramuscular route of infection in mice does not reflect the bottleneck of natural mucosal infection in humans, and was necessarily limited to just five strains per group.

Comparison of non-invasive *emm*1 isolates from London and invasive *emm*1 isolates nationally revealed both groups to be interspersed and clustered tightly in the phylogenetic tree, indicating a common genetic pool. The analysis showed that individual invasive isolates can be derived repeatedly from the population of pharyngitis strains. The identical nature of strains underlines the route of direct transmission from cases of pharyngitis and scarlet fever to dangerous invasive infections, often unnoticed. We found that diversifying selection in the invasive population, especially in $M1_{global}$, drives the accumulation of mutations in *covRS*, as reported[31].

Our study evaluated the origin, dispersion, and population dynamics of $M1_{UK}$ by assembling the most comprehensive global collection of $M1_{UK}$ strains to date. The analysis showed $M1_{UK}$ to be globally distributed, with nearly identical strains found all over the world, and multiple introductions from the UK population. The 2022 upsurge in the UK was characterised by rapid expansion of three clades within $M1_{UK}$, of which two showed swift dispersal to at least two other European countries. In Iceland, a single introduction event appeared responsible for reported $M1_{UK}$ cases, whereas in Denmark multiple introductions seemed likely. We found no evidence of importation of a new lineage recently reported in Denmark ($M1_{DK}$)[21].

The origin of $M1_{UK}$ was estimated to date from 2008, the year in which national guidelines to reduce swab testing and unnecessary antibiotic treatment of sore throat were introduced in England[10]. An exponential increase in the $M1_{UK}$ population commenced around 2010. Given the propensity for $M1_{UK}$ to spread readily in classrooms[32], it is conceivable that new lineages can emerge and rapidly expand if active *S. pyogenes* throat infections are not detected and treated with antibiotics, and transmission is not controlled. Antecedent intermediate lineages emerged in 2002 ($M1_{13SNPs}$) and 2006 ($M1_{23SNPs}$), during which time secular changes in sore throat management were ongoing in the UK[33,34].

Our dataset is limited to the UK and other high-income temperate countries, hence no inferences about the importation of $M1_{UK}$ into low-income countries were possible. This underlines the importance of global surveillance to monitor evolution and epidemiology of emerging variants with increased capacity for pathogenicity. Although $M1_{UK}$ geographic origin was identified as the UK, this was the only country with genomes available from the time of emergence, hence we cannot exclude an alternate origin.

The phylodynamic analysis of $M1_{UK}$ in the UK showed a decline in population size between 2015-2019 after the initial rapid rise, consistent with the cyclical changes in *S. pyogenes* populations known to occur[35], however population size plummeted in early 2020, when NPI to combat spread of COVID-19 were introduced. The marked $M1_{UK}$ population bottleneck was followed by rapid expansion in 2022 and 2023, raising the question of how such strains are maintained during periods of such low *S. pyogenes* population activity. Global reductions in other bacterial respiratory pathogens were seen during the period of COVID-19 NPI[36], however the scale of resurgence in invasive *S. pyogenes* following relaxation of NPI thus-far appears unique, perhaps related to the lack of a vaccine for *S. pyogenes* compared with other pathogens studied[36]. The observed magnitude and severity of the upsurge could be explained by the coincidence of enhanced $M1_{UK}$ pathogenicity and diminished human population immunity. The role of exposure-driven human immunity in shaping cyclical and post-COVID-

19 changes in *S. pyogenes* epidemiology is the subject of ongoing research. Scarlet fever affects children in their first year of school[37], an experience that was delayed for many during two years of COVID-19-related NPI. We hypothesise this resulted in a ~3-fold increase in susceptible children returning to school in Q3 2022, with similar reduction in immunity in siblings and adults. We posit that the transmissibility and invasiveness of $M1_{UK}$ facilitated the exponential and unprecedent increase in invasive *S. pyogenes* infections.

Material and Methods

Surveillance of *S. pyogenes* detection in clinical samples in England

All reports of *S. pyogenes*-positive clinical samples, including post-mortem, from ISO week 1 2016 to ISO week 30 2023[38] reported by English laboratories were extracted from the UK Health Security Agency (UKHSA) Second Generation Surveillance System (SGSS) on 7 December 2023. SGSS captures approximately 98% of electronically supplied hospital microbiology laboratory data in England, however, is the primary route for statutory reporting of laboratory confirmed invasive *S. pyogenes* infections[38]. Invasive *S. pyogenes* samples are defined as culture-positive samples (or positive by molecular detection) obtained from a normally sterile site. *S. pyogenes*-positive samples were deduplicated where patients had more than one positive *S. pyogenes* similar specimen type taken on the same date.

Invasive *Streptococcus pyogenes* isolates

S. pyogenes isolates from invasive disease (iGAS) cases in England, Wales, and Northern Ireland are routinely submitted to the national reference laboratory for *emm* genotyping using standard methods (https://www.cdc.gov/streplab/groupa-strep/emm-typing-protocol.html). The percentage of invasive isolates that were determined to be *emm*1 was determined compared with the overall total number of isolates genotyped. As part of the investigation into the upsurge of *S. pyogenes*, all *S. pyogenes* isolates from 2022 were whole genome sequenced (WGS). For this study, we included all *emm*1 isolates from invasive infections that had been genome sequenced at the reference laboratory from 2014-2023 including a small number from other regions. This included *emm1* isolates from 2014-2015, previously reported (n=516)[39]; *emm1* isolates from 2016-2021 (n= 207) intermittently sequenced as part of service delivery; and all *emm1* strains (n= 1092) submitted to the reference laboratory from January 2022-March 2023 that were sequenced as part of this outbreak investigation. Metadata and accessions for all isolate genome sequences are listed in Supplementary Data 3. Isolate WGS

was linked to reported clinical sample type. Differences in the proportion of *emm*1 between time points were evaluated using the proportion test (https://www.socscistatistics.com/tests/ztest/).

Non-invasive *S. pyogenes* isolates

Non-invasive *S. pyogenes* isolates were identified by MALDI-Biotyper (Bruker) from swabs submitted to the Diagnostic Laboratory at Imperial College Healthcare NHS Trust (London, UK) during 2022 (1 January - 31 December). This laboratory serves northwest London, a population of ~2 million people, representing ~3.5% of the population of England. *S. pyogenes* isolates were cultured on Columbia Blood Agar (CBA, Oxoid, Basingstoke, UK) or in Todd Hewitt broth (Oxoid) at 37°C with 5% CO_2. Demographic and clinical data were linked to all isolates and anonymised in accordance with research ethics approval (06/Q0406/20). All *emm*1 pharyngitis isolates (from throat swabs) were genome sequenced at the National Reference laboratory (Supplementary data 3).

Genomic data contextualisation

Three different genomic datasets were included in this study. The first contains 1815 (1092 newly sequenced from 2022-2023; and 723 from 2013-2021) *emm*1 strains associated with invasive infections collected at national level and sequenced at the UKHSA national reference laboratory from 2013 to 2023 (Supplementary Data 3); 12 isolates were from outbreak investigations. The second dataset contained the above 1815 invasive strains described above plus 133 newly sequenced non-invasive *emm*1 isolate whole genome sequences (WGS) collected in London during 2022 as part of this study (1 January to 31 December), yielding a total of 1948 *S. pyogenes* isolate WGS (Supplementary Data 3). The third dataset was created to provide phylogenetic context for the $M1_{UK}$ global population and intermediate strains only. This dataset included an additional 385 previously-published $M1_{UK}$ or intermediate WGS from the UK sequenced at the Wellcome Trust Sanger Institute dating from 2005-2018[5,23]; 166 $M1_{UK}$ or intermediate WGS collected in Australia 2010-2022[11]; 16 $M1_{UK}$ or

intermediate WGS collected in Canada 2016-2019[23]; 138 $M1_{UK}$ or intermediate WGS collected in Denmark 2018-2023[21]; 18 $M1_{UK}$ or intermediate WGS from Iceland, 2023[21]; 27 $M1_{UK}$ or intermediate WGS collected in the Netherlands 2019[22]; and 11 $M1_{UK}$ or intermediate WGS from USA collected in 2015-2018[25]. Data collection finished in July 2023, and therefore genomes reported after that time point were not included. The final global dataset contained 2365 $M1_{UK}$ and intermediate strains (Supplementary Data 3).

Generation of new $M1_{UK}$ reference genome: Reference strain NCTC14935

Genomic DNA from *S. pyogenes* $M1_{UK}$ isolate H1490 and $M1_{global}$ isolate H1499 (both sore throat isolates) was sheared using a Megaruptor to prepare 20-22 kb PacBio SMRT libraries, following the manufacturer´s recommendations. The libraries were sequenced using one Single Molecule Real-Time (SMRT) cell in a PacBio RSII platform (Pacific Biosciences of California, Inc., Menlo Park, CA, USA) at the University of Edinburgh. The data was demultiplexed using Lima v2.2.0 (https://lima.how/). The demultiplexed CLR data was converted to CCS using ccs tool v6.3.0 and further HiFi reads (CCS >Q20) were extracted using extract hi fi tool from the same package. The genome assemblies were generated from the HiFi reads using Redbean v 2.25[40] and Trycycler v0.5.3[41]. The assembly quality was assessed using QUAST v5.0.2[42] and BUSCO v5.3.0[43]. The annotation was performed using prokka v1.14.6[44]. PacBio sequencing reads, and complete genome assemblies are deposited in the European Nucleotide Archive under BioProject accession PRJEB68198 ($M1_{UK,}$ H1490 - ERR12378139 and $M1_{global,}$ H1499 - ERR12378140). The two isolates have been deposited in the National Collection of Type Cultures (NCTC) with the accessions NCTC14935 ($M1_{UK}$, H1490) and NCTC14936 ($M1_{global}$, H1499).

Illumina genome sequencing, assembly, and annotation

For this study, whole genome sequencing of all clinical isolates (invasive and non-invasive) was performed by the UKHSA reference laboratory using Illumina NextSeq 1000 platform with 100 base paired-end chemistry. Reads were trimmed to remove adaptor sequences and low-

quality bases with Trimmomatic v0.39[45]. Contamination was assessed based on Kraken2[46] classification of reads mapped against standard database for bacteria. Genomes with less 90% of the reads mapped against *S. pyogenes* were excluded. Draft genomes were generated using SPAdes v3.15.4[47]. The assembly quality was assessed using QUAST v5.0.2[42] and poor assemblies were filtered out if the genome size was higher than 2.1 Mbp and/or have more than 400 contigs. Genome annotation was performed with prokka v1.14.6[44].

Identification of single nucleotide variations and phylogenetic analysis

Core genome alignment was obtained by mapping trimmed reads of *S. pyogenes* genomes to MGAS5005 (GenBank accession: CP000017.2) reference genome using snippy v4.6.0 (https://github.com/tseemann/snippy), with a minimum coverage of 10, minimum fraction of 0.9 and minimum vcf variant call quality of 100. The SNP distance matrix was obtained using snp-dist (https://github.com/tseemann/snp-dists). SNPs identified were classified as non-coding, missense or synonymous according to their location in the genome and their effect on protein sequence using Snippy. Gubbins v3.3.0[48] was used to identify and remove recombinant regions. A maximum-likelihood (ML) phylogenetic tree was constructed from the multi-sequence alignment using RAxML-NG v1.0.1[49] implemented in Gubbins v3.3.0 (substitution and rate variation model: GTR + Gamma). The ML tree was rooted on NCTC8198 (GenBank accession: GCA_002055535.1, reference genome of old *emm1* lineage). Phylogenetic trees and associated data were visualised using iTOLv6.8.1[50].

Characterisation of genomic features of interest

The presence of AMR genes was predicted combining the results from ABRicate (https://github.com/tseemann/abricate), Ariba[51] and srst2[52]. The *pbp* genes sequences (*pbp1a*, *pbp1b*, and *pbp2x*) were obtained using a BLASTN (NCBI BLAST+ v2.7.1) search. The nucleotide sequences were converted to amino acids and examined for the presence of non-synonymous mutations. None of the non-synonymous mutations previously associated with penicillin resistance in *S. pyogenes* were identified. A similar approach was used to

identify non-synonymous mutations in *S. pyogenes* regulatory genes (*covR*, *covS*, *fasA*, *fasB*, *fasC*, *rgg1*, *rgg2*, *rgg3*, *rgg4*, *rivR*, *rofA* and *rocA*). The presence of superantigens (*smeZ*, *speA2*, *speC*, *speG*, *speH*, *speI*, *speJ*, *speK*, *speL*, *speM*, *speN*, *speO*, *speP*, *speQ*, *speR*, *ssa*) and DNAses (*sda2*, *sdn1*, *spdn1*, *spd3*, *spd4*, *spdB*, *spnA*) was accessed with a BLASTN (NCBI BLAST+ v2.7.1) analysis with the default parameters. Differences between lineages ($M1_{global}$ and $M1_{UK}$) regarding the number/type of mutations found in regulatory genes and *pbp* genes were evaluated using the proportion test (https://www.socscistatistics.com/tests/ztest/).

Pangenome analysis

A pangenome graph was constructed from annotated genome assemblies of MGAS5005 and 1815 *emm1* isolates collected from across the UK between 2013 and 2023 using Panaroo v1.3.0[53] under its moderate decontamination mode. Clusters of orthologous genes (COGs) were defined by a minimum nucleotide identity of 98%, and core genes were defined by a minimum frequency of 95%. The resulting gene presence-absence matrix was filtered to remove pseudo and fragmented genes as well as those of unusual lengths. The pangenome graph was simplified with the MGAS5005 genome as a reference using Panaroo's helper script *reference_based_layout.py* for visualisation in Cytoscape v3.10.1[54]. Presence-absence of COGs was compared between $M1_{UK}$ and $M1_{global}$ and between pre-2022/2023 and 2022/2023 groups using Python v 3.11.6.

Phylodynamic analysis of $M1_{UK}$

A maximum-likelihood (ML) phylogenetic tree corrected for recombination events was constructed from the multi-sequence alignment of global $M1_{UK}$ and intermediate genomes (against the $M1_{UK}$ reference genome H1490) using RAxML-NG v1.0.1[49] as implemented in Gubbins v3.3.0[48] (model: GTR + Gamma). The ML tree was rooted on $M1_{Global}$ isolate ERS17508611, which was the most closely related to $M1_{UK}$ and intermediate lineages according to SNP distances. A dated phylogenetic tree was generated from the ML tree using the least-squares dating method implemented in the LSD2 module of IQ-Tree v2.2.2.7 (model:

GTR+I+G4)[55,56]. Ancestral geographical locations were inferred from the dated tree and isolate information using the MPPA method and F81 model as implemented in PastML[57].

To reconstruct the population dynamics of the $M1_{UK}$ lineage in the UK, a UK-specific subtree of $M1_{UK}$ genomes was extracted from the dated tree, and the $M1_{UK}$ effective population size (N_e) was thereby modelled through time using a skygrowth model[58] implemented in R package mlesky (with 60-time intervals as determined using the package's parameter-optimisation algorithm based on the Akaike Information Criterion)[59]. Furthermore, the same model was iteratively fitted on 40 subtrees of randomly sampled UK $M1_{UK}$ genomes (with a maximum of 76, 22, and 14 genomes per year, respectively, based on sample sizes between 2019 and 2021) to evaluate if the variation in sample size over time could impact the inference of N_e.

In vivo screening for *covR/S* mutations using five representative strains of $M1_{UK}$ and $M1_{global}$

Five $M1_{global}$ and five $M1_{UK}$ strains were used in this study (ExtendedData figure 3). Experimental soft tissue infections were performed using female BALB/c mice aged 6 weeks (Charles River, UK). Bacteria were cultured on CBA overnight and resuspended in sterile PBS. Mice were infected with 5 x 10^8 CFU of one of the 10 strains (3 mice per strain) into thigh muscle. 24 h after infection mice were sacrificed and 150µl heparinized blood obtained by cardiac puncture from each mouse was plated onto CBA prior to euthanisation. Each spleen was removed, homogenised using FastPrep-24™ 5G in 1 ml PBS and plated on CBA for enumeration. Agar-based casein digestion assay was used to determine SpeB activity to infer *covS* mutations. 50 colonies cultured on CBA from spleens were patched on to 2% w/v skim milk Todd Hewitt agar (THA) to determine SpeB activity. One spleen sample with only a single colony was excluded from analysis; three samples with 16, 33 and 36 colonies were included. Fifty colonies from the inoculum of each strain were patched onto skim milk THA to rule out *covS* mutations occurring before introduction to the mice. SpeB (caseinolytic) activity was determined by comparing zones of clearance from *S. pyogenes* isolates to positive controls on the same plates and repatched to confirm the phenotype. Statistical analysis was

performed with GraphPad Prism 10. Comparison of the two groups was carried out using two-tailed nested t-test.

Data availability

The complete annotated genome sequences (PacBio sequences) of $M1_{UK}$ and $M1_{global}$ generated in this study have been deposited in ENA database under the BioProject PRJEB68198. Illumina short reads of all 1815 *emm*1 *S. pyogenes* used in this study from invasive disease cases (from the UK, 2013 to 2023) were deposited under the BioProject PRJEB68199. Illumina short reads of *emm*1 non-invasive disease pharyngitis isolates collected in London in 2022 were deposited under the BioProject PRJEB71329. Source data are provided with this paper as Supplementary Data 3. Genome assemblies and metadata of 2365 global $M1_{UK}$ isolates analysed in this study are available as a collection on Pathogen Watch (pathogen.watch/collection/czvgald6pluq-vieira-et-al-2024).

Ethics

The collection and genomic analysis of fully anonymised bacterial isolates previously linked to routine data in west London was approved by a national research ethics committee (06/Q0406/20). UK Health Security Agency surveillance of infections for health protection purposes is approved under Regulation 3 of The Health Service (Control of Patient Information) Regulations 2020 and under Section 251 of the NHS Act 2006. All animal experiments were undertaken using protocols approved by the Imperial College Animal Welfare Advisory Board (AWERB) and authorised by a UK Home Office Project Licence.

Funding

This work was supported by grants from the UK Medical Research Council (MR/P022669/1); the UK National Institute for Health Research (NIHR) Imperial College Biomedical Research Centre (BRC); the NIHR Health Protection Unit in Healthcare Associated Infections and

Antimicrobial Resistance; the Conor Kerin Foundation (PacBio sequencing); Illumina sequencing was funded by the UK Health Security Agency.

Acknowledgements

The authors are grateful to the local diagnostic laboratories and microbiologists who submit isolates to the reference laboratory for genotyping, and to the Imperial College Healthcare Trust diagnostic laboratory for bioresourcing isolates for this study. SS acknowledges support from the NIHR Imperial Biomedical Research Centre, which also supports the BRC Leonard and Dora Colebrook Laboratory. HKL is a Medical Research Council (MRC) CMBI Clinical Training Fellow; VWCS is an MRC Career Development Fellow; EJ is an Imperial College Research Fellow, funded by Rosetrees Trust and the Stoneygate Trust; NJC and EJ acknowledge the MRC Centre for Global Infectious Disease Analysis funded by the MRC and Department for International Development (grants MR/R015600/1 and MR/T016434/1). AV, YW, EJ, JC, TL, CSB, and SS are affiliated with the NIHR Health Protection Research Unit in Healthcare Associated Infections and Antimicrobial Resistance at Imperial College London in partnership with the UK Health Security Agency (formerly, Public Health England), in collaboration with, Imperial Healthcare Partners, University of Cambridge and University of Warwick; XD is affiliated with the NIHR Health Protection Research Unit in Genomics and Enabling Data at the University of Warwick. The views expressed in this publication are those of the author(s) and not necessarily those of the NHS, the National Institute for Health Research, the Department of Health and Social Care or the UK Health Security Agency.

Author contributions

Conceived and designed the study SS, JC; Data analysis AV, YW, YR, VWCS; Data collection; HKL, MP, KKH, LCR, RD, AH, KB, CD, MG, MG, ZM, MR, ER, RM, EAM, VS, KYM; Data visualisation AV, YW, RLG; Data analysis tools XD, NJC; Project supervision SS, JC, TL;

559 CSB; Analysis supervision EJ, XD, NJC. Writing, first draft AV, YW, SS; Editing and approval
560 of final manuscript, All.
561

562 ## Competing interests None declared
563

564 ## References

565 1. Brouwer, S. *et al.* Pathogenesis, epidemiology and control of Group A Streptococcus
566 infection. *Nat Rev Microbiol* **21**, 431–447 (2023).
567 2. Nelson, G. E. *et al.* Epidemiology of Invasive Group A Streptococcal Infections in the
568 United States, 2005–2012. *Clin Infect Dis.* **63**, 478–486 (2016).
569 3. Lamagni, T. L. *et al.* Predictors of Death after Severe *Streptococcus pyogenes*
570 Infection. *Emerg. Infect. Dis.* **15**, 1304–1307 (2009).
571 4. Steer, A. C., Law, I., Matatolu, L., Beall, B. W. & Carapetis, J. R. Global *emm* type
572 distribution of group A streptococci: systematic review and implications for vaccine
573 development. *The Lancet Infectious Diseases* **9**, 611–616 (2009).
574 5. Lynskey, N. N. *et al.* Emergence of dominant toxigenic M1T1 *Streptococcus pyogenes*
575 clone during increased scarlet fever activity in England: a population-based molecular
576 epidemiological study. *The Lancet Infectious Diseases* **19**, 1209–1218 (2019).
577 6. Li, H. K. *et al.* Characterization of emergent toxigenic M1UK Streptococcus pyogenes
578 and associated sublineages. *Microbial Genomics* **9**, (2023).
579 7. Sumby, P., Whitney, A. R., Graviss, E. A., DeLeo, F. R. & Musser, J. M. Genome-Wide
580 Analysis of Group A Streptococci Reveals a Mutation That Modulates Global Phenotype
581 and Disease Specificity. *PLoS Pathog* **2**, e5 (2006).
582 8. Nasser, W. *et al.* Evolutionary pathway to increased virulence and epidemic group A
583 Streptococcus disease derived from 3,615 genome sequences. Proc Natl Acad Sci USA
584 **111,**17 (2014).

9. National Institute for Health and Care Excellence (NICE). Management of Scarlet Fever. 2023. https://cks.nice.org.uk/topics/scarlet-fever/management/management/

10. Tan, T., Little, P. & Stokes T. Antibiotic prescribing for self-limiting respiratory tract infections in primary care: Summary of NICE guidance. British Medical Journal **337** a437 doi:10.1136/bmj.a437 (2008).

11. Lamagni, T. *et al.* Resurgence of scarlet fever in England, 2014–16: a population-based surveillance study. *The Lancet Infectious Diseases* **18**, 180–187 (2018).

12. UK Health Security Agency. *Group A streptococcal infections: 15th update on seasonal activity in England* https://www.gov.uk/government/publications/group-a-streptococcal-infections-activity-during-the-2022-to-2023-season/group-a-streptococcal-infections-15th-update-on-seasonal-activity-in-england (2023).

13. Davies, M. R. *et al.* Detection of *Streptococcus pyogenes* M1UK in Australia and characterization of the mutation driving enhanced expression of superantigen SpeA. *Nat Commun* **14**, 1051 (2023).

14. Zhi, X. *et al.* Emerging Invasive Group A Streptococcus $M1_{UK}$ Lineage Detected by Allele-Specific PCR, England, 2020. *Emerg Infect Dis* **29**, 1007-1010 (2023).

15. Guy, R. *et al.* Increase in invasive group A streptococcal infection notifications, England, 2022. *Eurosurveillance* **28**, (2023).

16. WHO. Increased incidence of scarlet fever and invasive Group A Streptococcus infection - multi-country. https://www.who.int/emergencies/disease-outbreak-news/item/2022-DON429

17. https://www.gov.uk/government/publications/group-a-streptococcal-infections-report-on-seasonal-activity-in-england-2023-to-2024/group-a-streptococcal-infections-report-on-seasonal-activity-in-england-2023-to-2024

18. Chen, M. *et al.* Increase of emm1 isolates among group A Streptococcus strains causing scarlet fever in Shanghai, China. *International Journal of Infectious Diseases* **98**, 305–314 (2020).

19. Vannice, K. S. *et al. Streptococcus pyogenes* pbp2x Mutation Confers Reduced Susceptibility to β-Lactam Antibiotics. *Clinical Infectious Diseases* **71**, 201–204 (2020).

20. Musser, J. M. *et al.* Reduced *In Vitro* Susceptibility of *Streptococcus pyogenes* to β-Lactam Antibiotics Associated with Mutations in the *pbp2x* Gene Is Geographically Widespread. *J Clin Microbiol* **58**, e01993-19 (2020).

21. Johannesen, T. B. *et al.* Increase in invasive group A streptococcal infections and emergence of novel, rapidly expanding sub-lineage of the virulent *Streptococcus pyogenes* M1 clone, Denmark, 2023. *Eurosurveillance* **28**, (2023).

22. Rümke, L. W. *et al.* Dominance of M1UK clade among Dutch M1 *Streptococcus pyogenes*. *The Lancet Infectious Diseases* **20**, 539–540 (2020).

23 Turner, C. E. et al. The Emergence of Successful *Streptococcus pyogenes* Lineages through Convergent Pathways of Capsule Loss and Recombination Directing High Toxin Expression. *Mbio* 10, 6 (2019)

24. Demczuk, W., Martin, I., Domingo, F. R., MacDonald, D. & Mulvey, M. R. Identification of *Streptococcus pyogenes* M1UK clone in Canada. *The Lancet Infectious Diseases* **19**, 1284–1285 (2019).

25. Li, Y., Nanduri, S. A., Van Beneden, C. A. & Beall, B. W. M1UK lineage in invasive group A *streptococcus* isolates from the USA. *The Lancet Infectious Diseases* **20**, 538–539 (2020).

26. De Gier, B. *et al.* Increase in invasive group A streptococcal (*Streptococcus pyogenes*) infections (iGAS) in young children in the Netherlands, 2022. *Eurosurveillance* **28**, (2023).

27. Lassoued, Y. *et al.* Unexpected Increase in Invasive Group A Streptococcal Infections in Children After Respiratory Viruses Outbreak in France: A 15-Year Time-Series Analysis. *Open Forum Infectious Diseases* **10**, ofad188 (2023).

28. Rodriguez-Ruiz, J. P. *et al.* Increase in bloodstream infections caused by *emm*1 group A *Streptococcus* correlates with emergence of toxigenic M1UK, Belgium, May 2022 to August 2023. *Eurosurveillance* **28**, (2023).

639 29. Gouveia, C. *et al.* Sustained increase of paediatric invasive *Streptococcus pyogenes*
640 infections dominated by M1UK and diverse *emm*12 isolates, Portugal, September 2022 to
641 May 2023. *Eurosurveillance* **28**, (2023).
642 30. Bessen, D. E. Population biology of the human restricted pathogen, *Streptococcus*
643 *pyogenes*. *Infection, Genetics and Evolution* **9**, 581–593 (2009).
644 31. Mayfield, J. A. *et al.* Mutations in the Control of Virulence Sensor Gene from
645 Streptococcus pyogenes after Infection in Mice Lead to Clonal Bacterial Variants with
646 Altered Gene Regulatory Activity and Virulence. *PLoS ONE* **9**, e100698 (2014).
647 32. Cordery, R. *et al.* Frequency of transmission, asymptomatic shedding, and airborne
648 spread of *Streptococcus pyogenes* in schoolchildren exposed to scarlet fever: a
649 prospective, longitudinal, multicohort, molecular epidemiological, contact-tracing study in
650 England, UK. *The Lancet Microbe* **3**, e366–e375 (2022).
651 33. Gulliford, M. *et al.* Selective decrease in consultations and antibiotic prescribing for
652 acute respiratory tract infections in UK primary care up to 2006. *Journal of Public Health*
653 **31**, 512–520 (2009).
654 34. Yap, D., Harris, A. & Clarke, J. Serious tonsil infections versus tonsillectomy rates in
655 Wales: A 15-year analysis. *annals* **99**, 31–36 (2017).
656 35. Tamayo, E., Montes, M., García-Arenzana, J. M. & Pérez-Trallero, E. *Streptococcus*
657 *pyogenes emm*-types in northern Spain; population dynamics over a 7-year period. *Journal*
658 *of Infection* **68**, 50–57 (2014).
659 36. Shaw. D., *et al*. Trends in invasive bacterial diseases during the first 2 years of the
660 COVID-19 pandemic: analyses of prospective surveillance data from 30 countries and
661 territories in the IRIS Consortium. *Lancet Digit Health* **9**, e582-e593 (2023).
662 37. Herdman, M. T. *et al.* Clinical management and impact of scarlet fever in the modern
663 era: findings from a cross-sectional study of cases in London, 2018–2019. *BMJ Open* **11**,
664 e057772 (2022).
665 38. ISO. ISO 8601:2004: data elements and interchange formats— information
666 interchange—representation of dates and times. Geneva: (2021).

667 39. Kapatai, G., Coelho, J., Platt, S. & Chalker, V. J. Whole genome sequencing of group
668 A *Streptococcus*: development and evaluation of an automated pipeline for *emm* gene
669 typing. *PeerJ* **5**, e3226 (2017).
670 40. Ruan, J. & Li, H. Fast and accurate long-read assembly with wtdbg2. *Nat Methods*
671 **17**,155–158 (2020).
672 41. Wick, R. R. *et al.* Trycycler: consensus long-read assemblies for bacterial genomes.
673 *Genome Biol* **22**, 266 (2021).
674 42. Mikheenko, A., Prjibelski, A., Saveliev, V., Antipov, D. & Gurevich, A. Versatile genome
675 assembly evaluation with QUAST-LG. *Bioinformatics* **34**, i142–i150 (2018).
676 43. Manni, M., Berkeley, M. R., Seppey, M., Simão, F. A. & Zdobnov, E. M. BUSCO
677 Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic
678 Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Molecular Biology*
679 *and Evolution* **38**, 4647–4654 (2021).
680 44. Seemann, T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* **30**, 2068–
681 2069 (2014).
682 45. Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
683 sequence data. *Bioinformatics* **30**, 2114–2120 (2014).
684 46. Wood, D. E., Lu, J. & Langmead, B. Improved metagenomic analysis with Kraken 2.
685 *Genome Biol* **20**, 257 (2019).
686 47. Bankevich, A. *et al.* SPAdes: A New Genome Assembly Algorithm and Its Applications
687 to Single-Cell Sequencing. *Journal of Computational Biology* **19**, 455–477 (2012).
688 48. Croucher, N. J. *et al.* Rapid phylogenetic analysis of large samples of recombinant
689 bacterial whole genome sequences using Gubbins. *Nucleic Acids Research* **43**, (2015).
690 49. Kozlov, A. M., Darriba, D., Flouri, T., Morel, B. & Stamatakis, A. RAxML-NG: a fast,
691 scalable and user-friendly tool for maximum likelihood phylogenetic inference.
692 *Bioinformatics* **35**, 4453–4455 (2019).
693 50. Letunic, I. & Bork, P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic
694 tree display and annotation. *Nucleic Acids Research* **49**, W293–W296 (2021).

51. Hunt, M. *et al.* ARIBA: rapid antimicrobial resistance genotyping directly from sequencing reads. *Microbial Genomics* **3**, (2017).

52. Inouye, M. *et al.* SRST2: Rapid genomic surveillance for public health and hospital microbiology labs. *Genome Med* **6**, 90 (2014).

53. Tonkin-Hill, G. *et al.* Producing polished prokaryotic pangenomes with the Panaroo pipeline. *Genome Biol* **21**, 180 (2020).

54. Shannon, P. *et al.* Cytoscape: A Software Environment for Integrated Models of Biomolecular Interaction Networks. *Genome Res.* **13**, 2498–2504 (2003).

55. Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution* **37**, 1530–1534 (2020).

56. To, T.-H., Jung, M., Lycett, S. & Gascuel, O. Fast Dating Using Least-Squares Criteria and Algorithms. *Syst Biol* **65**, 82–97 (2016).

57. Ishikawa, S. A., Zhukova, A., Iwasaki, W. & Gascuel, O. A Fast Likelihood Method to Reconstruct and Visualize Ancestral Scenarios. *Molecular Biology and Evolution* **36**, 2069–2085 (2019).

58. Didelot, X., Franceschi, V., Frost, S. D. W., Dennis, A. & Volz, E. M. Model design for nonparametric phylodynamic inference and applications to pathogen surveillance. *Virus Evolution* **9**, vead028 (2023).

59. Volz, E. M. & Didelot, X. Modeling the Growth and Decline of Pathogen Effective Population Size Provides Insight into Epidemic Dynamics and Drivers of Antimicrobial Resistance. *Systematic Biology* **67**, 719–728 (2018)

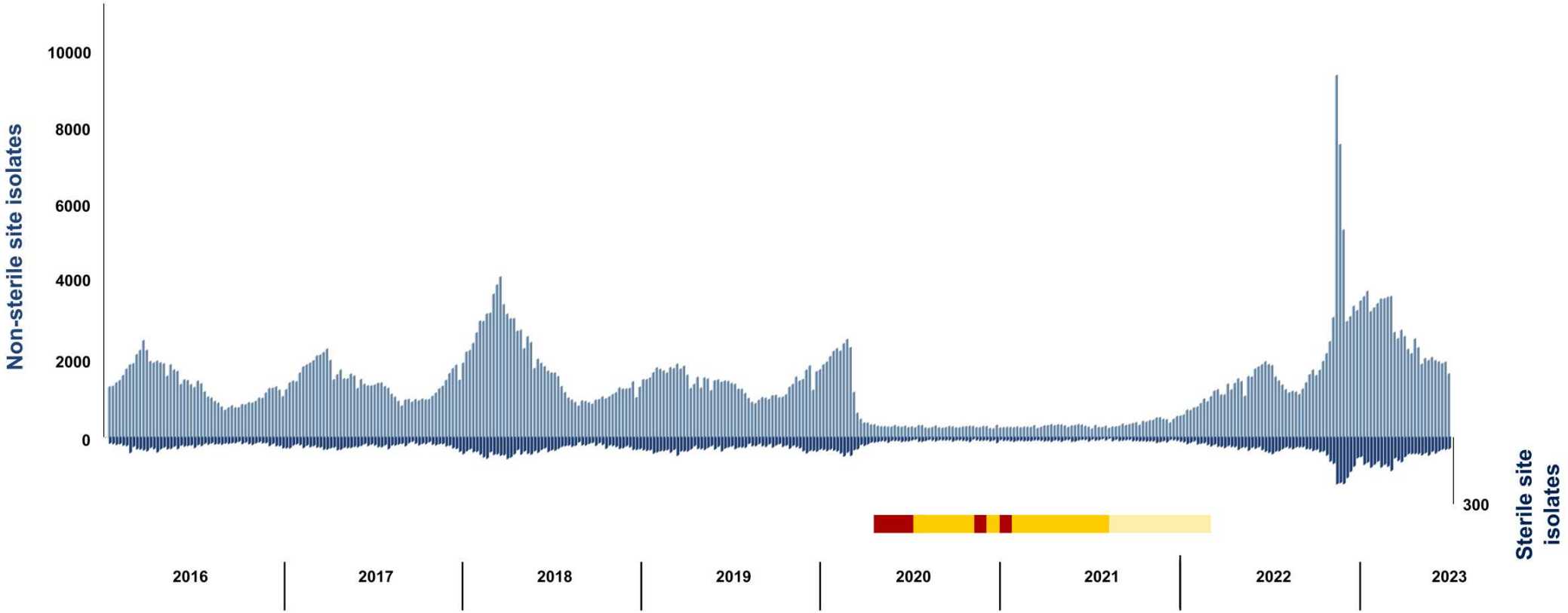

Figure 1. Trend in *S. pyogenes*-positive samples, England 2016-2023. Data show absolute numbers of weekly *S. pyogenes*-positive samples from non-sterile sites (light blue bars, left hand, positive axis) and sterile sites (dark blue bars, right hand, negative axis) recorded by the Second Generation Surveillance System (SGSS) in England, by year. Timing of non-pharmaceutical interventions (NPI) related to COVID-19 in England are indicated by the horizontal bar: red, lockdown periods; orange legally enforced NPI including no mixing; yellow, non-severe NPI. Schools were closed during lockdown periods, and between the two later lockdown periods except for children of keyworkers and vulnerable children.

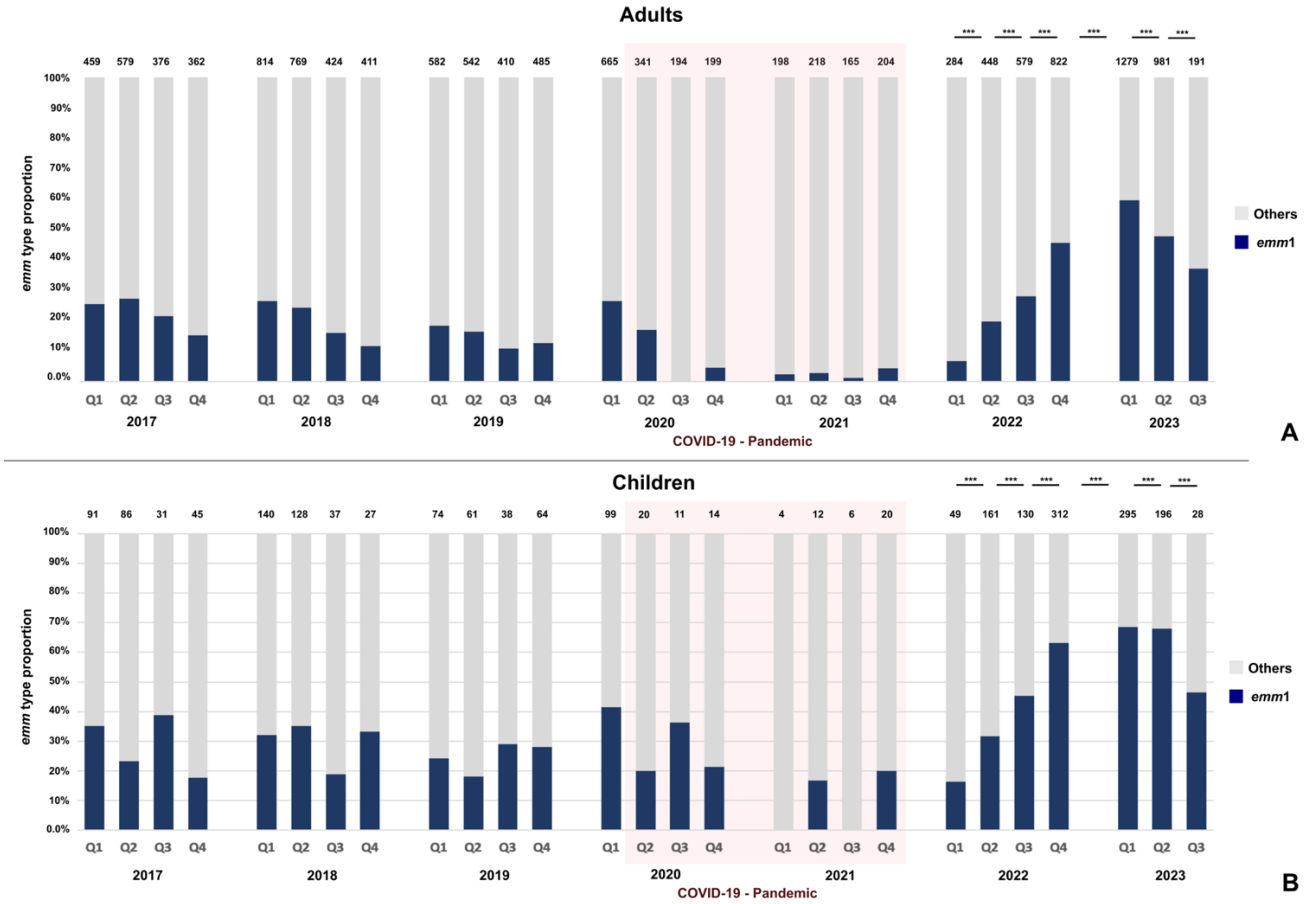

Figure 2. Contribution of *emm*1 *S. pyogenes* to invasive group A streptococcal (iGAS) infections 2017-2023. *emm*1 isolates are shown as proportions of the total number of isolates from iGAS cases submitted to, and genotyped at, the national reference laboratory for each quarter of each year. A, adults (≥15 years); B, children <15 years. The total number of isolates from iGAS cases received by the reference laboratory and genotyped in each quarter are shown on top of each bar; *emm1* proportions are shown in navy blue. Pink shaded region highlights the period of COVID-19 non pharmaceutical interventions. Q1, January-March; Q2, April-June; Q3, July- September; Q4, October-December. Statistical analysis applied to 2022-2023: Proportion of *emm1* from Q1 2022 to Q3 2023 significantly different between each quarter are indicated by asterisks. p-value < 0.0001

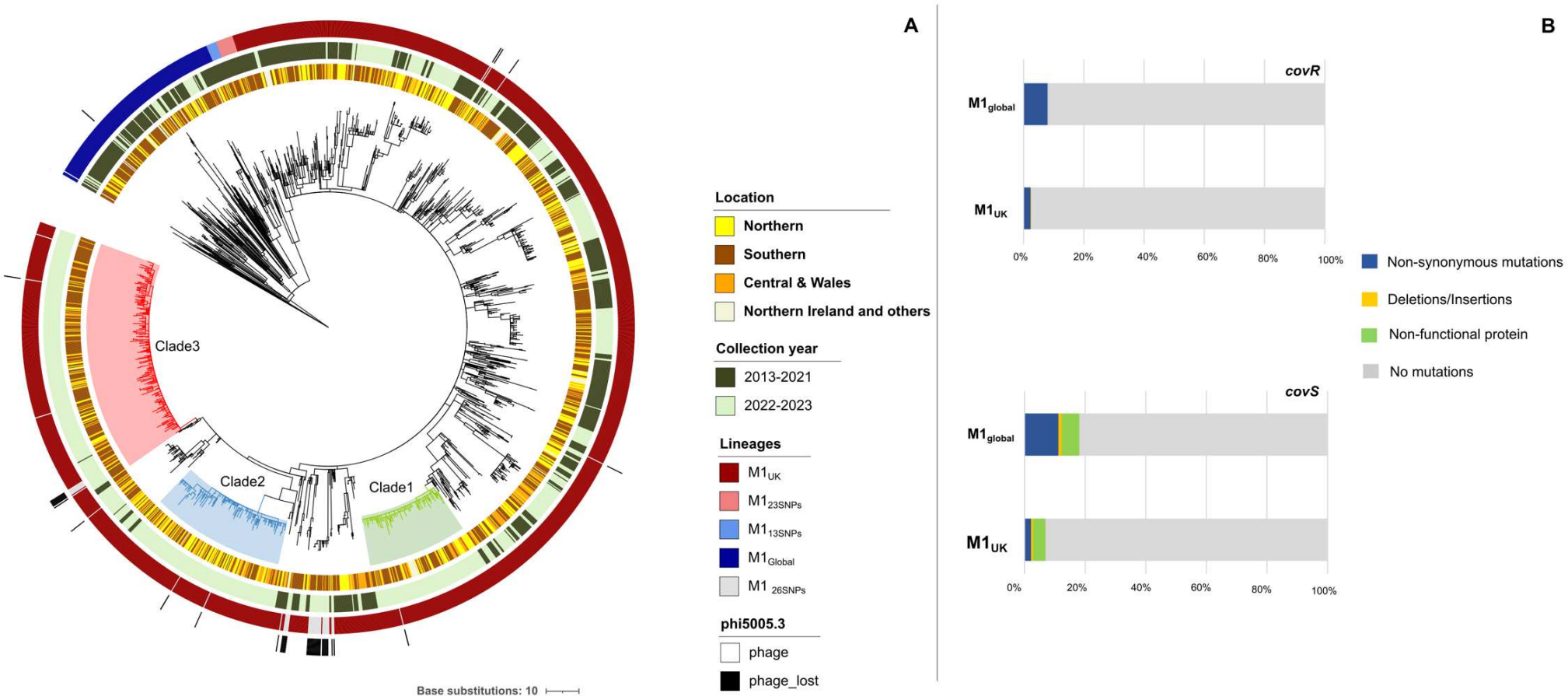

Figure 3. Genetic analysis of 1815 *emm1 S. pyogenes* isolates from invasive group A streptococcal (iGAS) infections 2013-2023.

A. Phylogenetic tree comprising sequenced *emm1* isolates associated with invasive infections (iGAS) from 2013-2023 sequenced at reference laboratory: Maximum likelihood phylogenetic tree constructed from 278 core SNPs (excluding recombination regions) extracted after mapping 1815 *emm1* isolates to the MGAS5005 reference genome. The tree was drawn in a circular layout and rooted on outgroup genome NCTC8198. Bars in concentric circles represent (from inside to outside) regional location of isolate; collection period (pre-upsurge 2013-2021 or upsurge 2022-2023); *emm*1 lineage, and presence/absence of the phi5005.3 phage. Regional data have been grouped for purpose of data

visualisation as follows: Northern (North-East England, North-West England, Yorks & Humber); Central and Wales (East Midlands, West Midlands, Wales); Southern (South-East England, South-West England, London); and Northern Ireland and others (comprises regions with less than 5 isolates including Scotland, Eire, Jersey, Malta). B) Frequency of *covR* and *covS* non-synonymous and other mutations within $M1_{UK}$ and $M1_{global}$ invasive isolates. Percentage of strains with non-synonymous mutations, deletions/insertions, or an inactive protein in 1552 $M1_{UK}$ and 189 $M1_{global}$ isolates is shown. Mutation types are indicated by coloured bars. Differences in *covR* and *covS* mutation frequency between $M1_{global}$ (covR 15/189; *covS* 34/189) and $M1_{UK}$ (*covR* 38/1552; *covS* 106/1552) are significant (*covR* p=0.0006; *covS* p=0.0002). Ten $M1_{global}$ isolates formed a previously unrecognised clade with covRS mutations. If all strains from this cluster are removed, the *covS* mutation frequency within $M1_{global}$ remains significantly greater than $M1_{UK}$ strains.

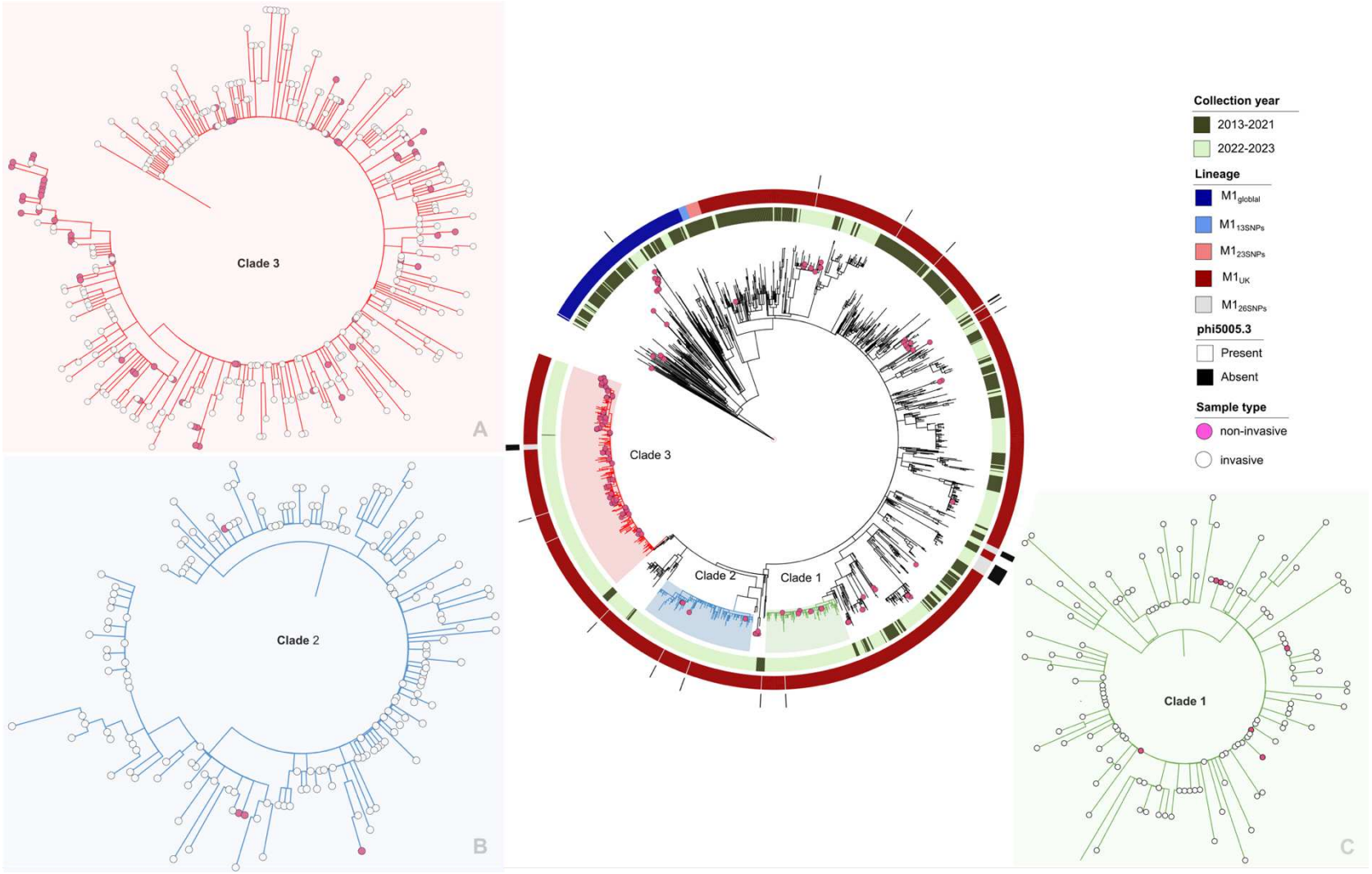

Figure 4 – *emm1* phylogenetic tree showing non-invasive isolates collected in London in 2022 with invasive isolates from UK 2013-2023. Maximum likelihood phylogenetic tree constructed with the core alignment of 274 SNPs extracted after mapping 1815 *emm1* invasive isolates and 133 non-invasive isolates against MGAS5005. A-C) Relationship between invasive and non-invasive isolates within Clades1-3.

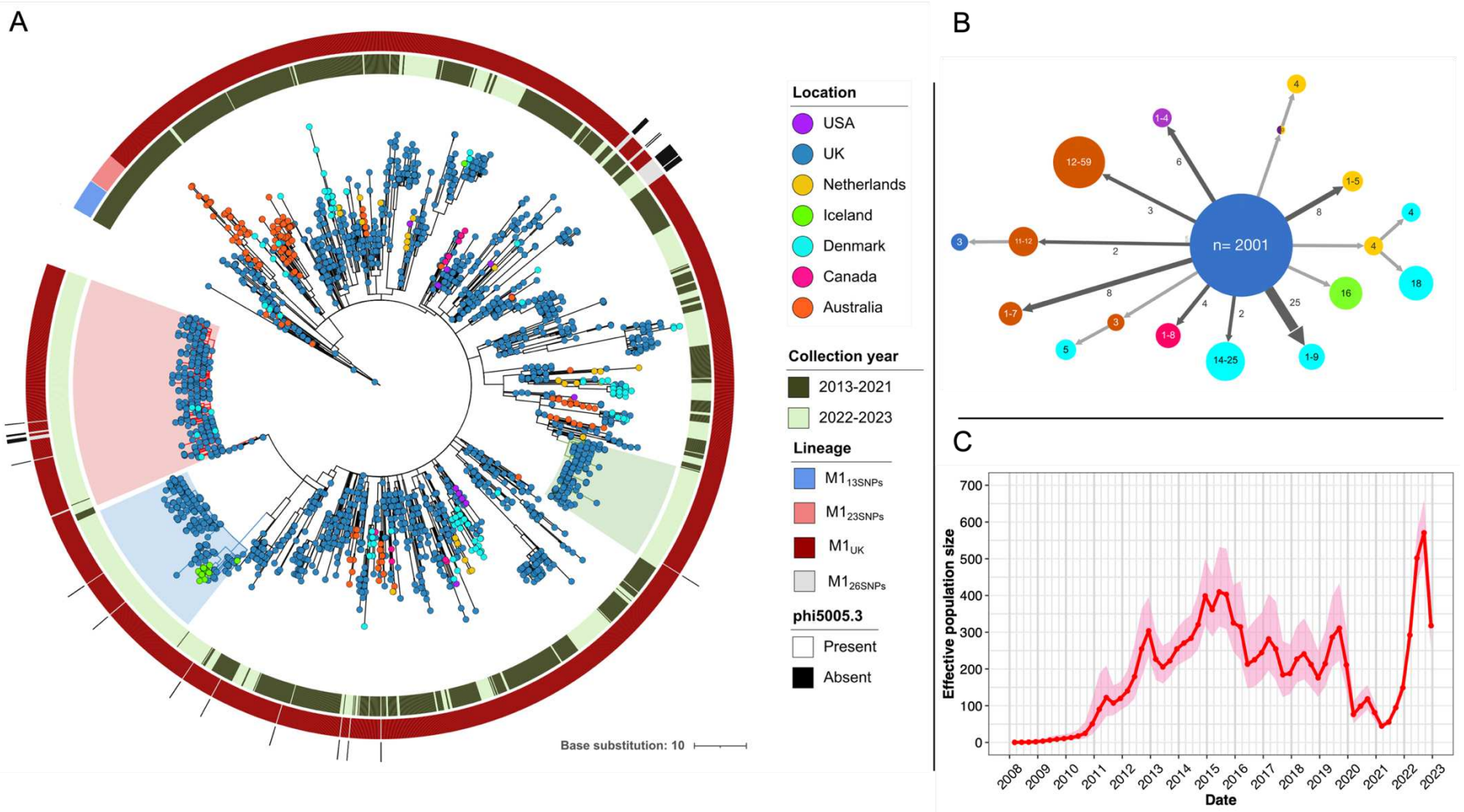

Figure 5 – Global distribution and potential introduction events of $M1_{UK}$ and intermediate populations. - A) Phylogenetic tree of 2365 $M1_{UK}$ and intermediate strains collected globally March 2005 - July 2023. The tree was built based on 3406 SNPs from a core genome alignment relative to $M1_{UK}$ (H1490/NCTC14935) reference genome and

rooted on a closely related $M1_{global}$ genome ERR12372446. Leaves are coloured based on country where samples were collected. Shading indicates the 3 emergent clades (Clade 1, green; Clade 2, blue; Clade 3 red). Bars in concentric circles represent (from inside to outside) are coloured by collection years (pre-upsurge 2013-2021 and upsurge 2022-2023); *emm*1 lineage; and presence/absence of the phi5005.3 phage. B) Simplified transmission tree by PastML showing the ancestral epidemic location of $M1_{UK}$ and intermediate lineages. Each node represents a cluster of leaves sharing the same probable ancestral location and is labelled by the range of leaves numbers. Each arrow indicates inferred international transmission events; arrow width and labels indicate the number of identical origin-destination transmission events. For example, the arrow labelled “6” pointing at the node “1–4” (USA) indicates six clusters of 1 to 4 leaves were present in the USA that were likely imported from the UK. C) Estimated effective population size (N_e) of $M1_{UK}$ in UK through time. The red line and pink shading at each time point indicate the mean and 95% confidence interval of N_e, respectively.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- M1UKSupplementaryData123.xlsx
- M1UKExtendedData07.01.23.pdf