This is a repository copy of *The emergence of successful streptococcus pyogenes lineages through convergent pathways of capsule loss and recombination directing high toxin expression.*

White Rose Research Online URL for this paper:
http://eprints.whiterose.ac.uk/154652/

Version: Published Version

**Article:**
Turner, C.E. orcid.org/0000-0002-4458-9748, Holden, M.T.G., Blane, B. et al. (3 more authors) (2019) The emergence of successful streptococcus pyogenes lineages through convergent pathways of capsule loss and recombination directing high toxin expression. mBio, 10 (6). e02521.

https://doi.org/10.1128/mbio.02521-19

**Reuse**
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here:
https://creativecommons.org/licenses/

**Takedown**
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.
The Emergence of Successful *Streptococcus pyogenes* Lineages through Convergent Pathways of Capsule Loss and Recombination Directing High Toxin Expression

Claire E. Turner, Matthew T. G. Holden, Beth Blane, Carolyne Horner, Sharon J. Peacock, Shiranee Sriskandan

ABSTRACT  Gene transfer and homologous recombination in *Streptococcus pyogenes* has the potential to trigger the emergence of pandemic lineages, as exemplified by lineages of *emm*1 and *emm*89 that emerged in the 1980s and 2000s, respectively. Although near-identical replacement gene transfer events in the *nga* (NADase) and *slo* (streptolysin O) loci conferring high expression of these toxins underpinned the success of these lineages, extension to other *emm* genotype lineages is unreported. The emergent *emm*89 lineage was characterized by five regions of homologous recombination additional to *nga-slo*, including complete loss of the hyaluronic acid capsule synthesis locus *hasABC*, a genetic trait replicated in two other leading *emm* types and recapitulated by other *emm* types by inactivating mutations. We hypothesized that other leading genotypes may have undergone similar recombination events. We analyzed a longitudinal data set of genomes from 344 clinical invasive disease isolates representative of locations across England, dating from 2001 to 2011, and an international collection of *S. pyogenes* genomes representing 54 different genotypes and found frequent evidence of recombination events at the *nga-slo* locus predicted to confer higher toxin genotype. We identified multiple associations between recombination at this locus and inactivating mutations within *hasAB*, suggesting convergent evolutionary pathways in successful genotypes. This included common genotypes *emm*28 and *emm*87. The combination of no or low capsule and high expression of *nga* and *slo* may underpin the success of many emergent *S. pyogenes* lineages of different genotypes, triggering new pandemics, and could change the way *S. pyogenes* causes disease.

IMPORTANT  *Streptococcus pyogenes* is a genetically diverse pathogen, with over 200 different genotypes defined by *emm* typing, but only a minority of these genotypes are responsible for the majority of human infection in high-income countries. Two prevalent genotypes associated with disease rose to international dominance following recombination of a toxin locus that conferred increased expression. Here, we found that recombination of this locus and promoter has occurred in other diverse genotypes, events that may allow these genotypes to expand in the population. We identified an association between the loss of hyaluronic acid capsule synthesis and high toxin expression, which we propose may be associated with an adaptive advantage. As *S. pyogenes* pathogenesis depends both on capsule and toxin production, new variants with altered expression may result in abrupt changes in the molecular epidemiology of this pathogen in the human population over time.

KEYWORDS group A Streptococcus, NADase, *Streptococcus pyogenes*, streptolysin O, 

Citation Turner CE, Holden MTG, Blane B, Horner C, Peacock SJ, Sriskandan S. 2019. The emergence of successful *Streptococcus pyogenes* lineages through convergent pathways of capsule loss and recombination directing high toxin expression. mBio 10:e02521-19. https://doi.org/10.1128/mBio.02521-19.

Editor Paul Keim, Northern Arizona University

Copyright © 2019 Turner et al. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

Address correspondence to Claire E. Turner, c.e.turner@sheffield.ac.uk

Received 23 September 2019
Accepted 29 October 2019
Published 10 December 2019
convergent evolution, homologous recombination, hyaluronic acid capsule, whole-genome sequencing

The capacity for the bacterial human pathogen *Streptococcus pyogenes* to undergo genetic exchange, independent of known bacteriophages or mobile elements, is not well understood, yet recent evidence suggests it underpins the emergence of successful new variants that rapidly rise to international dominance. Homologous recombination of a chromosomal region encompassing the toxin genes *nga* (encoding NADase), *ifs* (encoding the inhibitor for NADase), and *slo* (encoding streptolysin O), which was dated to have occurred in the mid-1980s, is thought to have driven the rise of *emm*1 to almost global dominance (1). The homologous recombination event resulted in increased *nga-slo* expression compared to that of the previous variant, linked to the gain of a highly active *nga-ifs-slo* promoter in the new *emm*1 variant compared to that of the previous variant (2).

A very similar recombination event was recently identified in the genotype *emm*89. A new variant of *emm*89 sequence type (ST) 101 (also referred to as clade 3) emerged, having undergone six regions of predicted homologous recombination compared to its ST101 predecessor (also referred to as clade 2) (3, 4). One of the six regions encompassed the *nga-ifs-slo* locus, comprising a region almost identical to that of *emm*1, which conferred similarly high expression of *nga* and *slo* compared to that of the previous variant. Another recombination region within the emergent ST101 *emm*89 resulted in the loss of the hyaluronic acid capsule. We dated the emergence of this new acapsular high-toxin-expressing ST101 *emm*89 lineage to the mid-1990s, but there was a rapid increase and rise to dominance in the United Kingdom between 2005 and 2010 (3). The lineage is now the dominant form of *emm*89 in the United Kingdom as well as other parts of the world, including Europe, North America, and Japan (4–8).

Given that recombination associated with *nga-ifs-slo* can give rise to new successful *S. pyogenes* variants, we hypothesized that this may be a feature common to other successful *emm* types. To determine if this is the case, we sequenced the genomes of 344 *S. pyogenes* invasive disease isolates originating from hospitals across England between 2001 and 2011 and compared the data with other available historical and contemporary international *S. pyogenes* whole-genome sequence (WGS) data. We identified that recombination of the *nga-ifs-slo* locus has occurred in other leading *emm* types, supporting the hypothesis that it can underpin the emergence and success of new lineages. We also identified an association of *nga-ifs-slo* recombination toward a high-activity promoter variant with inactivating mutations within the capsule locus. This suggests that loss of capsule may also provide an advantage to certain genotypes, either through a direct effect on pathogenesis or an association with the process of recombination.

RESULTS

Genetic characterization of bacteremia isolates. We performed whole-genome sequencing of 344 *S. pyogenes* invasive isolates collected from hospitals across England by the British Society for Antimicrobial Chemotherapy (BSAC) Bacteraemia Resistance Surveillance Program from 2001 to 2011. Forty-four different *emm* types were identified from *de novo* assembly, with the most common being *emm*1 (*n* = 64, 18.6%), *emm*12 (*n* = 34, 9.9%), *emm*89 (*n* = 32, 9.3%), *emm*3 (*n* = 28, 8.1%), *emm*87 (*n* = 22, 6.4%), and *emm*28 (*n* = 15, 4.4%) (see Fig. S1 in the supplemental material). Antimicrobial susceptibilities were typical for *S. pyogenes*, with 100% of isolates susceptible to penicillin and 22% resistant to clindamycin, erythromycin, and/or tetracycline; detailed susceptibilities and associated genotypes are reported in Data Set S1.

The phylogenetic distribution of the 344 isolates based on core genome variation revealed distinct clustering by *emm* type, each forming single lineages with the exceptions of *emm*44, *emm*90, and *emm*101, each of which formed two lineages (Fig. 1A). Pairwise distances between isolates gave a median of just 45 single nucleotide
polymorphisms (SNPs) separating the genomes of isolates of the same emm genotype (range, 0 to 15,137 SNPs) compared to a median of 15,648 SNPs separating the genomes of isolates of different emm types (range, 5,312 to 18,317 SNPs) (Fig. 1B). The genotypes emm44, emm90, and emm101 gave the highest SNP distance for the intra-emm comparison (13,494 to 15,137 SNPs), which approaches the median level observed between emm types. This indicated that while other genotypes represent a relatively conserved chromosomal genetic background, the populations of emm44, emm90, and emm101 exhibit more diverse chromosomal backgrounds despite representing the same emm type, potentially due to emm gene switching.

**High level of variation within the nga-ifs-slo locus.** To identify the level of variation within the nga-ifs-slo locus, we extracted the sequence from the 3’ end of nusG (immediately upstream of nga) to the 3’ end of slo (P-nga-ifs-slo), comprising the entire locus and all of the upstream sequence, including the predicted ∼67-bp nga-ifs-slo promoter region (9). We constructed a phylogenetic tree from SNPs within the P-nga-ifs-slo region from the genomes of isolates belonging to the most common emm types and compared it to the phylogeny constructed with SNPs extracted from a whole-genome comparison to a reference emm89 genome, H293 (Fig. 2). Most emm genotypes were associated with a single P-nga-ifs-slo variant that was unique to that genotype. The main exception to this was the P-nga-ifs-slo variant found in modern (post-1980s M1T1) emm1, as this was also found in all emm12, all emm22 (a lineage known to be acapsular) isolates, and 11 of the 32 emm89 isolates. These 11 emm89 isolates represented the emergent acapsular ST101 variant, while the remaining 21 emm89 isolates represented the original encapsulated ST101 variant, with a different
unique P-nga-ifs-slo as previously reported (3). The entire emm75 population and one of the two emm76 isolates were also associated with a P-nga-ifs-slo variant that was closely related to the emm1-like variant. All but two emm87 isolates had a P-nga-ifs-slo variant also found in the acapsular lineage emm4. The presence of multiple P-nga-ifs-slo
variants within the emm76 and emm87 genotypes, where the core chromosome was otherwise relatively conserved, indicated that gene transfer and recombination are responsible for the P-nga-ifs-slo variation in these genotypes rather than extensive genome-wide divergence or emm "switching."

**Variants of the nga-ifs-slo promoter associated with altered expression.** Recombination of P-nga-ifs-slo and surrounding regions in emm1 and emm89 conferred higher activity and expression of NGA (NADase) and SLO (1, 3, 10). This change in expression was linked to the combination of three key residues at –27, –22, and –18 within the nga-ifs-slo promoter. A –27G –22T –18 at these key sites was associated with high nga-ifs-slo promoter activity in emm1 and emergent variant emm89 following recombination (also referred to as Pnga3) compared to low promoter activity of historical emm1 and emm89 isolates has been shown to have low-level promoter activity. A –27T –22T –18 subtype 4 promoter has also been shown to have high activity in emm28 (15). Subtypes 1.2, 1.3, and 2.3 were restricted to emm9, emm88, and emm32 strains, respectively. (8) WebLogo representation of the variability in the 67-bp promoter region of nga-ifs-slo within the 54 different emm types. Key residues –27, –22, and –18 are highlighted (star) and their positions are relative to the –35 and –10 boxes. Figure generated using WebLogo (http://weblogo.berkeley.edu).
States (emm), which are present at a low-activity-associated promoter; these included (in addition to the aforementioned emm89) emm75, emm76, emm77, emm81, emm82, emm87, emm94, and emm108, all of which are emm types frequently identified in the United Kingdom and the United States (12–14). Although one emm28 was found to carry the high-activity-associated A_127G_22T_18 promoter, the rest of the emm28 population was divided between either A_127T_22C_18 or A_127T_22T_18 variants. The data pointed to a switch in P-nga-ifs-slo in all cases rather than an emm switch, except for emm82, where the emm82 gene replaced the emm12 gene in an emm12 genetic background (14).

**High level of mutations within the capsule locus leading to truncations of HasA or HasB.** As well as recombination around the P-nga-ifs-slo region, the emergent ST101
variant of emm89 had also undergone recombination surrounding the hasABC locus, and, in place of the hasABC genes, there was a region of 156 bp that was not found in genotypes with the capsule locus but is found in the acapsular emm4 and emm22 isolates (3). To identify any similar events in other genotypes, we examined the sequences of hasA, hasB, and hasC in the assemblies of isolates from the BSAC collection as well as CUH (11), PHE-2014/15 (12, 13), and ABCs-2015 (14) collections for gene presence as well as premature stop codon mutations (Fig. 4). The hasABC locus was absent in the majority of emm89 isolates, consistent with the previous observations describing the recent emergence of the acapsular emm89 variant (3). Similarly, the hasABC genes were absent in all emm4 and emm22 isolates, as previously identified (16), except for two emm4 isolates and one emm22 isolate which had an intact hasABC locus predicted to encode full-length proteins. We confirmed the genotypes of these isolates by emm typing the assembled genomes. Multilocus sequence typing (MLST) and phylogenetic analysis indicated that they both had a very different genetic background to other isolates.

FIG 4 Nonfunctional mutations within the capsule locus genes. The hasABC genes were extracted from the assembled genomes of BSAC, CUH, PHE-2015/15, and ABCs-2015 isolate genome collections, and polymorphisms or indels leading to nonsense mutations and premature stop codons, as well as gene absence, were identified. The percentage of isolates with full-length (gray), truncated (red), or absent (black) HasA, HasB, or HasC is depicted for each of the 54 emm types. emm types with fewer than 3 isolates were excluded. N = 5,271 isolates genomes shown. Mutations in HasA were detected in more than 50% of isolates belonging to genotypes emm8 (n = 3/4), emm11 (n = 63/108), emm25 (n = 2/3), emm27 (n = 3/3), emm28 (n = 358/363), emm58 (n = 21/33), emm68 (n = 12/14), emm73 (n = 25/27), emm77 (n = 72/80), emm78 (n = 8/8), emm87 (n = 119/121), and emm102 (n = 6/6). Mutations in HasB were detected in 100% of emm94 isolates (n = 54/54) and 60% to 77% of emm63 (n = 3/5), emm81 (n = 50/65), and emm90 (n = 16/26) isolates.
emm4 or emm22 populations, suggesting that these were not typical of these emm types; therefore, they represent examples of emm switching. Interestingly, we also identified a similar replacement of hasABC for the 156-bp region in one emm28 isolate (PHE-2014/15, GASEMM1261 [13]), but phylogenetic analysis suggested this was highly divergent from the rest of the emm28 population, likely to represent another example of emm switching. Isolated examples of individual hasA or hasB gene loss were identified in the genomes of isolates belonging to emm1 (n/H11005), emm3 (n/H11005), emm11 (n/H11005), emm12 (n/H11005), and emm108 (n/H11005).

The majority of genotypes (35/54 [65%]) had isolates without genes or truncation mutations in at least one of the hasABC genes (Fig. 4). Mutations in hasC were rare and only detected in one isolate, an emm77 isolate, which also had a mutation within hasA. Within seven of the eight emm types for which we identified potential P-nga-ifs-slo recombination, a high percentage of isolates had inactivating mutations in hasA and hasB, suggesting a possible association between an acapsular genotype/phenotype and recombination of P-nga-ifs-slo to gain a high-activity promoter. Including the previously identified emm1 and emm89 recombination events, P-nga-ifs-slo recombination to gain a high-activity promoter was detected in 10 genotypes, and in all 10 of these genotypes (100%) were isolates with hasAB gene mutations or gene absence.

Recombination of P-nga-ifs-slo and surrounding regions. To confirm our prediction that genotypes emm28, emm75, emm76, emm77, emm81, emm87, emm94, and emm108 had undergone recombination around P-nga-ifs-slo, we mapped all the genome sequence data for each genotype to the emm89 reference genome H293. Gubbins analysis of SNP clustering predicted regions of recombination spanning the nga-ifs-slo region and varied in length in all eight genotypes (Fig. 5). To further analyze the recombination of these genotypes and potential capsule loss, we studied the population structure of each genotype individually.

Recombination within emm28 and emm87 around P-nga-ifs-slo and the capsule locus. The genotypes emm28 and emm87 were the sixth and fifth most common in the BSAC collection, respectively, and emm28 was previously noted to be a major cause of infection in high-income countries [17]. We focused attention on emm28 and emm87, as there has been little genomic work on these genotypes so far.

All BSAC emm28 isolates carried the A$\_27T\_22C\_18$ low-activity-associated promoter, but inclusion of international genomic data identified A$\_27T\_22T\_18$ variant-carrying isolates. These two promoter variants were associated with different major
lineages within the entire population of 379 international emm28 isolates, including one newly sequenced English isolate originally isolated in 1938. The majority of isolates (n = 373) clustered either with the reference MGAS6180 strain (United States) (18) or with the reference MEW123 strain (United States) (19) (Fig. 6A). Gubbins analysis for core SNP clustering predicted that the two lineages were distinguished by a single 28,200-bp region of recombination, between positions 142,426 bp (mtpE; M28_Spy0126) and 170,625 bp (M28_Spy0153) of the MGAS6180 chromosome. This suggests the emergence of one lineage from the other through a single recombination event followed by expansion of both lineages (Fig. 6B). Within the recombination region was the P-nga-ifs-slo locus, which differed between the two lineages; although unique in the MGAS6180-like lineage and with low-activity-associated promoter residues A-27T-22C-18P, the MEW123-like lineage had a P-nga-ifs-slo identical to that found in emm78 isolates, with the three key residues of A-27T-22T-18P. This is supported by recent findings identifying two main lineages within emm28 and that the A-27T-22T-18P promoter variant conferred greater toxin expression than A-27T-22C-18P (15).

Although we identified an A-27G-22T-18 high-activity variant of P-nga-ifs-slo within emm28, this was only associated with the highly divergent GASEMM1261 isolate that may represent an "emm" switching event. This isolate, along with three other PHE-2014/15 isolates (GASEMM2648, GASEMM1396, and GASEMM1353), also representing highly divergent lineages, were removed from the phylogenetic analysis.

All emm28 isolates, regardless of lineage and including MGAS6180 (originally isolated in the 1990s), had the same insertion mutation within hasA of an A residue after 219 bp. This insertion was predicted to lead to a frameshift and a premature stop codon after 72 amino acids (aa) instead of full-length 420 aa, rendering hasA a pseudogene.
Some isolates also had additional mutations in hasA: a deletion of an A residue in a septa-A tract leading to a frameshift and a stop codon after 7 aa \((n = 1)\); a deletion of a T residue in a septa-T tract leading to a frameshift and a stop codon after 15 aa \((n = 2)\); an insertion of an A residue after 57 bp leading to a frameshift and a stop codon after 46 aa \((n = 3)\). The loss of full-length HasA would render the isolates acapsular.

In emm28, there were just two exceptions where hasA was found to be intact: the historical emm28 isolate from 1938 had an intact hasABC capsule operon, and BSAC bs2099, which appeared to have undergone recombination to acquire a 22,316-bp region surrounding the hasABC genes that was 99% identical to the same region in emm2 isolate MGAS10270, suggesting emm2 might be the donor for this recombination. Both isolates were predicted to express full-length HasA and synthesize capsule. Taken together, in comparison with the oldest isolates from the United Kingdom and North America, including one historical NCTC UK isolate from ~1970 to 1980 (NCTC12065, GenBank accession number GCA_900460075.1), Gubbins analysis predicted a single 20,506-bp region of recombination surrounding the P-nga-ifs-slo region that distinguished the main population from the oldest BSAC isolates from 2001 and the historical 1970 to 1980 NCTC isolate (Fig. 6C). While the two 2001 BSAC isolates and the NCTC isolate had a P-nga-ifs-slo variant with low-activity-associated promoter residues, G\(_{27/T-22T-18g}\) all other emm28 isolates had a P-nga-ifs-slo region with high-activity-associated promoter residues, A\(_{27/G-22T-18g}\) identical to that found in emm4 and some emm77 isolates. This suggested the emergence of a new lineage through a single recombination event followed by expansion within the population, redolent of that previously observed in emm89 (Fig. 6D).

Similar to emm28, all emm87 isolates, bar four, had an insertion of an A residue after 57 bp that resulted in a frameshift mutation in hasA and the introduction of a premature stop codon after 46 aa of HasA. This mutation was also identified within the historical NCTC isolate but was not found in the two 2001 BSAC isolates that had an intact hasABC locus. This mutation was also absent in two PHE-2014/15 isolates that had undergone an additional recombination event (32,243 bp) surrounding the hasABC locus; although, as this region shared 100% DNA identity to emm28 isolate MGAS6180, HasA is truncated. Overall the data showed that, like emm89 isolates, contemporary emm87 isolates are acapsular with a high-activity nga-ifs-slo promoter, suggesting that this emm lineage may have recently shifted toward this genotype/phenotype.

**Recombination within different multilocus sequence types of emm75.** The emm75 genotype is of interest as a common cause of noninvasive infection in the United Kingdom; it is also used in models of nasopharyngeal infection (20, 21). Eleven emm75 isolates were present in the BSAC collection, all multilocus sequence type (ST) 150. When we incorporated other available genome sequence data for emm75 \((n = 174)\), including two newly sequenced historical English emm75 isolates from 1937 and 1938, two major lineages were identified, characterized by two different MLSTs: ST49 or ST150 (Fig. 7A). Although the two historic English isolates were ST49, like the majority of modern North American isolates, the modern England/Wales isolates were predominantly ST150.

Although these two ST lineages differed in the P-nga-ifs-slo region, there was a high level of predicted recombination across the genomes of both STs, perhaps indicative of historic emm switching or extensive genetic exchange. ST49 isolates had the...
subtype 1.1 low-activity $A_{27}T_{22}C_{18}$ promoter, whereas all ST150 isolates had the $A_{27}G_{22}T_{18}$ subtype 3.1 high-activity promoter variant, identical to that of emm1 and emm89. Modern ST49 isolates did, however, differ from historic 1930s isolates by ten distinct regions of predicted recombination (Fig. 7B), including a region spanning the nga-ifs-slo locus, although this did not include the promoter region. We did not detect any mutations affecting the capsule region in emm75. Taken together, emm75 was characterized by two major MLST lineages differing in P-nga-ifs-slo promoter activity genotypes but without evidence of recent recombination or loss of capsule.

**Lineages associated with recombination in emm76, emm77, and emm81.** The phylogeny of all available genome data for emm76, emm77, and emm81 confirmed the presence of diverse lineages associated with different MLSTs (Fig. 8A to C). In all genotypes, however, there was a dominant MLST lineage representing the majority of isolates: ST50 emm76, ST63 emm77, and ST624 emm81. Within the dominant MLST lineages of emm76 and emm77, there were sublineages that were associated with different P-nga-ifs-slo variants as well as loss of functional HasA through mutation.

We identified five different MLSTs within emm76 (Fig. 8A), but the majority of isolates (30/38) belonged to ST50, including both BSAC isolates. Recombination analysis of the ST50 lineage identified a sublineage that differed from other ST50 isolates by 19 regions of recombination (Fig. 7B), including a region spanning the nga-ifs-slo locus, although this did not include the promoter region. We did not detect any mutations affecting the capsule region in emm75. Taken together, emm75 was characterized by two major MLST lineages differing in P-nga-ifs-slo promoter activity genotypes but without evidence of recent recombination or loss of capsule.
FIG 8 Variants of P-nga-ifs-slo and capsule mutations associated with lineages of emm76, emm77, and emm81. Maximum likelihood phylogeny identified multiple sequence type (ST) lineages within the populations of emm76 (A), emm77 (B), and emm81 (C). Collection indicates either BSAC or CUH (dark red), PHE-2014/15 isolates (red), ABCs-2015 (blue), or English historical (brown). Dates for BSAC, CUH, or historical are shown; other isolates were from 2014/2015. STs are indicated on the right and major lineages are shaded gray.

(A) Genome data for emm76 was mapped to the de novo assembled sequence of BSAC bs448 from 2002, selected as the oldest isolate (Continued on next page)
Two sublineages were also identified within the dominant emm77 lineage ST63 (Fig. 8B), and one was associated with the high-activity cluster P-nga-ifs-slo variant compared to predicted low-activity variants found in the other emm77 lineages. Recombination analysis predicted only two regions of recombination distinguishing the two sublineages: a region of 17,954 bp surrounding P-nga-ifs-slo, and a 173-bp region within a hypothetical gene (SPYH293_00394) (see Fig. S4). While all BSAC emm77 isolates (years 2001 to 2009) were ST63 with low-activity P-nga-ifs-slo, PHE isolates from 2014 to 2015 were almost evenly divided between the two sublineages, indicating a potential recent change in England/Wales. All ST63 isolates except two had a deletion of a T residue within a septa-poly(T) tract at 458 bp in hasA, predicted to truncate the HasA protein after 154 aa. The two exceptions were predicted to encode full-length HasA and were associated with low-activity P-nga-ifs-slo promoter variants. Although also not associated with high-activity P-nga-ifs-slo promoter variants, other lineages of emm77 also carried mutations within hasA that would truncate HasA; ST399 isolates carried an insertion of a T residue at 71 bp of the hasA gene resulting in a premature stop codon after 46 aa, and two ST133 isolates carried a G894A substitution resulting in a premature stop codon after amino acid residue 297.

The emm81 population (n = 68) was more diverse with nine different sequence types (Fig. 8C), but the majority of isolates (41/68) were ST624 or the single locus variant ST837 (9/68; one SNP in recP allele) within the same lineage. ST171 was restricted to three historical isolates originally collected in 1938 and 1939. We did not detect any hasABC variations that would disrupt translation in emm81 lineages except for the dominant group of ST624/ST837, where we identified an A residue insertion at 128 bp in hasB resulting in a frameshift and premature stop codon after 50 aa. All ST624/ST837 isolates carried the high-activity cluster P-nga-ifs-slo variant identical to that seen in emm3 compared to all other lineages associated with other low-activity P-nga-ifs-slo variants. Recombination analysis identified extensive recombination had occurred within emm81 leading to the different levels of diversity, but we identified one region of recombination that distinguished the ST624/ST837 lineage from the closely related ST909 and ST117 populations (see Fig. S5). This region surrounded the P-nga-ifs-slo locus, suggesting ST624/ST837 gained the high-activity cluster P-nga-ifs-slo variant through recombination, like other emm types, potentially from emm3. The emergence of the high-activity P-nga-ifs-slo variant and truncated HasB ST624/ST837 lineage may be recent in England/Wales, as all BSAC isolates obtained prior to 2009 were outside this lineage.

High-activity cluster P-nga-ifs-slo variants gained by recombination in emm94 and emm108. Within emm94, we identified a P-nga-ifs-slo identical to that found in emm1 with high-activity promoter variant subtype 3.1. Phylogenetic analysis of 51 emm94 isolates identified a dominant lineage among England/Wales isolates separate to the single US isolate and two England/Wales isolates (see Fig. S6A) that belonged to

---

**FIG 8 Legend (Continued)**

representing the genotypes. Genome data from a total of 38 isolates were used: BSAC (n = 2), PHE-2014/15 (n = 18 [12, 13]), ABCs-2015 (n = 18 [14]). Predicted prophage regions were removed and a maximum likelihood phylogenetic tree constructed from 30,264 core SNPs. Five STs were identified (indicated on right of tree), but the main lineage was ST50. (B) All emm77 genome data were mapped to the de novo assembled sequence of BSAC bs229 from 2001. Genome data from a total of 68 isolates were used: BSAC (n = 9), CUH (n = 1 [11]), PHE-2014/15 (n = 29 [12, 13]), ABCs-2015 (n = 26 [14]), and English historical 1930s (n = 3). Predicted prophage regions were removed, and a maximum likelihood phylogenetic tree was constructed from 42,258 core SNPs. Nine STs were identified but the main lineage was ST624 with and minor (single base change in recP) ST variant ST387. We identified variants of P-nga-ifs-slo (P) associated with one of three combinations of key promoter residues, including the high-activity-associated A_{-27}G_{-25}T_{-18} (P; black). For emm76 (A) and emm77 (B), mutations were detected in hasA predicted to truncate HasA (H; black). (C) All emm81 isolates were predicted to express full-length HasA, but the ST624/ST837 lineage carried a mutation within hasB leading to a truncated HasB (H; gray). Branches are colored based on bootstrap support (scale bar provided). Scale bars represent substitutions per site. Isolates used as references for mapping indicated with black circles. Branches for lineages outside main lineages were shortened for presentation purposes (indicated by line breaks).
ST89. Gubbins analysis predicted 11 regions of recombination in all lineage-associated isolates compared to the three outlying isolates, including one (22,648 bp) that encompassed *P*-nga-ifs-slo, transferring a high-activity A$_{−27}$G$_{−22}$T$_{−18}$ P-nga-ifs-slo variant. All *emm*94 isolates contained an indel within *hasB* compared to the reference (H293), losing 6 bp and gaining 13 bp between 127 and 133 bp. This variation causes a frameshift and would truncate the HasB protein after 45 aa.

We identified a similar high-activity cluster P-nga-ifs-slo variant within a single *emm*108 genome originating from the United States. Within the 9 isolates from PHE-2014/15 (n = 7) and ABCs-2015 (n = 2), there were two sequence types, ST1088 and ST14. ST14 was represented by the only two ABCs-2015 isolates, and we identified that both had lost the entire *hasB* gene, although *hasA* and *hasC* were still present (Fig. S6B). Additionally, one of the ABCs-2015 isolates had undergone recombination of a single ~29,683-bp region surrounding the P-nga-ifs-slo, replacing P-nga-ifs-slo for one identical to that found in *emm*3 with high-activity promoter variant A$_{−27}$G$_{−22}$T$_{−18}$ subtype 3.1.

**Mobile genetic elements and antimicrobial resistance.** The acquisition of mobile genetic elements such as prophages and transposons may also be influenced by capsule expression and can also influence the expansion and success of new lineages. We therefore determined the presence of prophage-associated superantigen and DNase genes as well as antimicrobial resistance genes to estimate the number of mobile genetic elements present within each isolate of the genotypes *emm*28, *emm*75, -76, -77, -81, -87, -94, and -108 (Fig. S3 to S5; Data Set S3). On average, there were 4.4 elements present in isolates predicted to express full-length *HasABC* compared to 2.5 elements present in isolates with *hasABC* gene mutations or gene absence, suggesting that the presence of capsule does not hinder mobile genetic elements. We also detected no link between lineages within these genotypes that had undergone P-nga-ifs-slo recombination and mobile factors, except within *emm*76 and *emm*77. Isolates belonging to the *emm*76 ST50 sublineage associated with *HasA* mutation and P-nga-ifs-slo recombination all carried the prophage-associate superantigen genes *speH* and *speI* as well as a diverse variant of the DNase *spd3* and the erythromycin resistance gene *ermB* (Fig. S3). This differed from the other ST50 isolates that carried another variant of *spd3* and multiple different resistance genes. The sublineage of ST63 *emm*77 associated with P-nga-ifs-slo recombination also carried *spd3*, and all, except one isolate, carried the erythromycin resistance gene *ermTR*; both genes were not common in other ST63 *emm*77 isolates (Fig. S4).

**DISCUSSION**

The emergence of new internationally successful lineages of *S. pyogenes* can be driven by recombination-related genome remodeling, as demonstrated by *emm*1 and *emm*89. The transfer of a P-nga-ifs-slo region conferring increased expression to the new variant was common to both genotypes. In the case of *emm*89, five other regions of recombination were identified in the emergent variant, one resulting in the loss of the hyaluronic acid capsule. Although, potentially, all six regions of recombination combined underpinned the success of the emergent *emm*89, we have shown here that recombination of P-nga-ifs-slo has occurred in other leading *emm* types as well as a high frequency of capsule loss through mutation. These data point to an association between genetic change affecting capsule and recombination affecting the P-nga-ifs-slo locus, conferring increased production of nga-ifs-slo; in some cases (notably *emm*87, *emm*89, and *emm*94), this has further been associated with an apparent fitness advantage and expansion within the population.

A number of genotypes were found to be associated with multiple variants of P-nga-ifs-slo. The majority of genotypes had P-nga-ifs-slo variants with the low-activity-promoter associated with three key residue variants: G$_{−27}$T$_{−22}$T$_{−18}$ or A$_{−27}$G$_{−22}$C$_{−18}$. Only *emm*1, *emm*3, and *emm*12 were exclusively associated with the high-activity A$_{−27}$G$_{−22}$T$_{−18}$ variant. We have shown that the same high-activity promoter variant is present in isolates belonging to twelve other *emm* types, notably, *emm*76, *emm*77, *emm*81, *emm*87, and *emm*94, although this is not a consistent feature in these geno-
types due to emm switching or recombination. We identified four combinations of the three key promoter residues and several subtypes of the 67-bp promoter that varied in bases other than those at the −27, −22, and −18 key positions. Although some subtypes were restricted to single genotypes, variation in the −40 base led to the subtype 2.2 of G_{−22}T_{−22}T_{−18} and subtype 3.2 of A_{−22}G_{−22}T_{−18}. We measured the activity of NADase in representative strains and genotypes of these promoter variants and found that variation in the −40 base did not impact the activity conferred by the −27, −22, and −18 bases. Although we predicted the level of nga and slo expression based on the promoter variant, this may not relate to actual expression given the level of other genetic variation between genotypes. However, our consistent findings of lineages emerging following acquisition of the high-activity promoter variant supports the hypothesis that this confers some benefit that may relate to increased toxin expression.

Intriguingly, where we identified an acquisition of the high-activity promoter variant through recombination, these genotypes also had a genetic change in the capsule locus, likely rendering the organism unable to make capsule (hasA mutation) or only low levels of capsule (hasB mutation). To date, only emm4, emm22, and the emergent emm89 lineages are known to lack all three genes required to synthesize capsule. Here, we identified mutations that would truncate HasA and HasB in 35% of all isolates and 65% (35/54) of all genotypes. As the majority of isolates included in this study were invasive or sterile-site isolates, the findings further challenge the dogma that the hyaluronan capsule is required for full virulence of S. pyogenes and, in addition, lend credence to the possibility that the increased expression of NADase and SLO may in some way compensate for the lack of capsule (22). While capsule has been shown to underpin resistance to opsonophagocytic killing in the most constitutively hyperencapsulated genotypes such as emm18 (23, 24), there is less evidence that it contributes measurably to opsonophagocytosis killing resistance in other genotypes (3). Whether the loss of capsule synthesis is of benefit to S. pyogenes is uncertain; the capsule may shield several key adhesins used for interaction with host epithelium and fomites but may also act as a barrier to transformation with DNA. An accumulation of hasABC inactivating mutations has been identified during long-term carriage (25); although for some genotypes, capsule loss reduced survival in whole human blood, a high number of acapsular hasA mutants have also recently been found to be causing a high level of disease in children, including emm1, emm3, and emm12 (26).

The process of recombination in S. pyogenes is not well understood, and natural competence has only been demonstrated once and under conditions of biofilm or nasopharyngeal infection (27). We do not know if the six regions of recombination that led to the emergence of the new ST101 emm89 variant occurred simultaneously, although no intermediate isolates have been identified. The loss of the hyaluronic acid capsule in the new emergent emm89, along with our consistent findings of inactivating mutations associated with P-nga-ifs-slo transfer, indicates either (i) the process of recombination requires the inactivation of capsule, (ii) capsule-negative S. pyogenes requires high expression of nga-ifs-slo for survival, or (iii) that a capsule-negative phenotype combined with high expression of nga-ifs-slo provides a greater selective advantage to S. pyogenes.

The phylogeny of emm28, emm87, emm77, emm94, and emm108 indicated that mutations in hasA or hasB occurred prior to recombination of P-nga-ifs-slo, supporting the first hypothesis that prior capsule inactivation is required for recombination. There is no evidence, however, to suggest this was required for recombination in the emm1 population. It could be hypothesized that capsule acts as a barrier to genetic exchange, but there has also been a positive genetic association of capsule to recombination rates (28). A positive association may, however, be related only to species expressing antigenic capsule, whereby recombination is required to introduce variation for immune escape.

The hasC gene is not essential for capsule synthesis (29), because a paralog of hasC exists within the S. pyogenes genome. A paralog for hasB (hasB.2) also exists elsewhere.
in the \textit{S. pyogenes} chromosome and can act in the absence of \textit{hasB} to produce low levels of capsule (30), but \textit{hasA} is absolutely essential for capsule synthesis (29). The mutations in \textit{hasA} in \textit{emm28} and \textit{emm87} have been previously noted and confirmed to render the isolates acapsular (26, 31). Not all acapsular isolates were found to carry the high-activity promoter of \textit{nga-ifs-slo}, despite being invasive, perhaps refuting the hypothesis that the high-activity \textit{nga-ifs-slo} promoter is essential for the survival of acapsular \textit{S. pyogenes}. High expression of \textit{nga-ifs-slo} may also occur through other mechanisms, for example, through mutation in regulatory systems. We looked at the sequences of \textit{covRS} and \textit{rocA}, known to negatively regulate \textit{nga-ifs-slo}, in all isolates (see Data Set S2 in the supplemental material) and identified some \textit{emm}-type specific variants, consistent with our previous findings (11). We did not identify any other genotypes where all isolates carried truncation mutations in \textit{rocA}, such as \textit{emm3} and \textit{emm18} that were previously confirmed to affect function and increase expression of \textit{rocA-covR}-regulated virulence factors (23, 32), consistent with other findings (14). It is unclear as to whether the amino acid changes in found in other genotypes would affect function of \textit{rocA} as well as \textit{covR} and \textit{covS}, and this requires further work.

Interestingly, we identified that the capsule locus is also a target for recombination as, similarly to \textit{emm89}, isolates within \textit{emm28} and \textit{emm87} had undergone recombination of this locus and surrounding regions, varying in length and restoring capsule synthesis in \textit{emm28}. Isolated examples of \textit{hasA} or \textit{hasB} gene loss were identified in some genotypes, such as \textit{emm108}, possibly due to internal recombination and deletion.

Only two \textit{emm4} isolates and one \textit{emm22} isolate were found to have \textit{P-nga-ifs-slo} variants that were not $\textit{A}_{-22}\text{T}_{-22}\text{G}_{-18}$ high-activity promoter variants, and interestingly, these isolates carried the \textit{hasABC} genes, typically absent in \textit{emm4} and \textit{emm22}. The high genetic distance of these isolates to other \textit{emm4} and \textit{emm22} genomes indicated potential \textit{emm} switching of the \textit{emm4} or \textit{emm22} genes into different genetic backgrounds. The single \textit{emm28} with a high-activity \textit{P-nga-ifs-slo} variant may also be an example of this, and was one of four \textit{emm28} isolates that did not cluster with the two main \textit{emm28} lineages. Although we excluded them from our analysis, as we focused on recombination within the two main lineages, the presence of highly diverse variants within genotypes and the potential for \textit{emm}-switching warrants further investigation, particularly as the most promising current vaccine is multivalent toward common M types (33).

All other genotypes carrying the high-activity \textit{P-nga-ifs-slo} variant were found to have undergone recombination of this region: \textit{emm28}, \textit{emm75}, \textit{emm76}, \textit{emm77}, \textit{emm81}, \textit{emm87}, \textit{emm94}, and \textit{emm108}, as well as the previously described \textit{emm1} and \textit{emm89}.

Within \textit{emm87}, we identified three isolates outside the main population lineage that represented the oldest isolates in the collection: two from 2001 (different geographical locations within England) and one NCTC strain from $\sim$1970 to 1980 (NCTC12065). A single region of recombination surrounding the \textit{P-nga-ifs-slo} locus distinguished the main population lineage from the three older isolates, consistent with a recombination event, but due to a lack of earlier isolates of \textit{emm87}, we could not confirm a recombination-related shift in the population, as reported previously for \textit{emm89} and \textit{emm1}.

The existence of two lineages within the contemporary \textit{emm28} suggests that one has not yet displaced the other, although the MEW123-like lineage was predominantly US isolates, consistent with recent findings (15). The \textit{P-nga-ifs-slo} region with the high-activity-associated $\textit{A}_{-22}\text{T}_{-22}\text{T}_{-18}$ and acquired through recombination by the MEW123-like lineage was identical to that found in \textit{emm78}, indicating \textit{emm78} as the potential genetic donor. We found \textit{emm78} to have high levels of NADase activity, as predicted, and interestingly, similarly to \textit{emm28}, all eight \textit{emm78} isolates were acapsular due to a deletion within the \textit{hasABC} promoter region extending into \textit{hasA}. This again may support the hypothesis that capsule-negative \textit{S. pyogenes} requires high expression of \textit{nga-ifs-slo} for survival.

A strength of this study was the systematic longitudinal sampling over a 10-year
period; as expected, this again identified the shift in the emm89 population. Other emm types exhibited lineages with different P-nga-ifs-slo variants, and those with the more active promoter variant did appear to become dominant over time, similarly to emm1 and the emergent emm89 lineages. For example, the high-activity P-nga-ifs-slo ST63 lineage of emm77 was not detected in England/Wales isolates prior to 2014 and 2015. Similarly, the high-activity P-nga-ifs-slo variant emm81 ST646/ST837 lineage was represented by only a single isolate (of six) collected between 2001 and 2009 but became dominant by 2014 to 2015 in England/Wales and the United States. emm75 was the 6th most common genotype in England/Wales in 2014 to 2015 and dominated by high-activity P-nga-ifs-slo variant ST150 lineage yet was less common in the United States, where ST49 with low-activity P-nga-ifs-slo is dominant. A high prevalence of emm94 was also found in England/Wales between 2014 and 2015 but was rare in the United States (only 1 isolate). Our analysis of this genotype indicated there has been a recombination-related change in the population, as we detected 11 regions of predicted recombination, including P-nga-ifs-slo, potentially conferring high toxin expression. The other ten regions of recombination may also provide advantages to this lineage along with a potential low level of capsule through hasB mutation.

Other factors may also contribute to the success of emergent new lineages, including mobile prophage-associated virulence factors and antimicrobial resistance genes. Acquisition of mobile genetic elements did not appear to be affected by capsule loss; indeed, fewer mobile genetic element-associated factors were detected in isolates with capsule gene mutations than in isolates with functional capsule genes. A number of bacteriophages that target S. pyogenes encode a hyaluronidase thought to allow the bacteriophage to access the bacterial surface by degrading the outer capsule layer (34); therefore, recombination of these elements is likely to be different from gene transfer of core genetic regions, such as P-nga-ifs-slo.

We did, however, identify an association in the lineages of emm76 and emm77 with prophage-associated virulence factors and antimicrobial resistance genes. It is possible that the superantigens speH, speI, and DNase spd3 may also have contributed to the success of the lineages that had undergone P-nga-ifs-slo recombination. Of concern is that both emm76 and emm77 carried genes for resistance to tetracycline and erythromycin, which were rarer in other genotypes. If the acapsular/high-toxin-expressing lineages do expand in the population, it will be important to monitor the levels of antimicrobial resistance in these lineages. This is also true for emm108, as tetM was detected in all isolates, but the presence of antimicrobial resistance genes was rare in emm28, emm75, emm81, emm87, and emm94, regardless of lineage.

The development and boosting of circulating antibodies to SLO is often used as a diagnostic biomarker of recent S. pyogenes infection and is known to be more specific to throat rather than skin infections. The genomic analysis provides explanation for this historic and well-recognized association between anti-streptolysin O (ASO) titers and disease patterns, due to known tissue tropism of S. pyogenes emm types. Whether the alteration of SLO activity in different S. pyogenes strains might render such a test more or less specific will be of interest, although it may explain observed differences in ASO titers between genotypes (35). There is also the possibility that other beta hemolytic streptococci might acquire similarly active SLO production, reducing the specificity of ASO titer to S. pyogenes.

Our genomic analysis has uncovered convergent evolutionary pathways toward capsule loss and recombination-related remodeling of the P-nga-ifs-slo locus in leading contemporary genotypes. This suggests that a combination of capsule loss and gain of high nga-ifs-slo expression provide a greater selective advantage than either of these phenotypes alone. Acquisition of the high-activity promoter led to pandemic emm1 and emm89 clones that are dominant and highly successful. Active surveillance of the lineages comprising emm76, emm77, emm81, emm87, emm94, and emm108 is required to determine if capsule loss/reduction and recombination of P-nga-ifs-slo toward high expression will trigger expansion toward additional pandemic clones in the next few years.
MATERIALS AND METHODS

Isolates. Three hundred forty-four isolates of *S. pyogenes* associated with bloodstream infections and submitted to the British Society for Antimicrobial Chemotherapy (BSAC; [http://www.bsacsurv.org](http://www.bsacsurv.org)) from 11 different sites across England between 2001 and 2011 were subjected to whole-genome sequencing (see Data Set S1 in the supplemental material). All BSAC isolates were tested for antibiotic susceptibility using the BSAC agar dilution method to determine MICs (36).

A further six isolates were sequenced from a historical collection of *S. pyogenes* originally collected in the 1930s from puerperal sepsis patients at Queen Charlotte's Hospital, London, UK; one emm28 isolate from 1938 (ERR485803), two emm75 isolates from 1937 (ERR485807) and 1939 (ERR485820), and three emm81 isolates from 1938 (ERR485805) and 1939 (ERR485801, ERR485802).

Genome sequencing. Streptococcal DNA was extracted using the QiAextractor instrument according to the manufacturer’s instructions (Qiagen, Hilden, Germany) or manually using a phenol-chloroform method (37). DNA library preparation was conducted according to the Illumina protocol, and sequencing was performed on an Illumina HiSeq 2000 with 100-cycle paired-end runs.

Genomes were de novo assembled using Velvet with the pipeline and improvements found at [https://github.com/sanger-pathogens/vr-codebase](https://github.com/sanger-pathogens/vr-codebase) and [https://github.com/sanger-pathogens/assembly improvement](https://github.com/sanger-pathogens/assembly improvement) (38). Annotation was performed using Prokka. *emm* genotypes were determined from the assemblies, and multilocus sequence types (MLSTs) were identified using the MLST database ([https://pubmlst.org/spyogenes/](https://pubmlst.org/spyogenes/)) and an in-house script ([https://github.com/sanger-pathogens/mlst_check](https://github.com/sanger-pathogens/mlst_check)). New MLSTs were submitted to the database ([https://pubmlst.org/](https://pubmlst.org/)).

Genome sequence analysis. Sequence reads were mapped using SMALT ([https://www.sanger.ac.uk/science/tools/smalt-0](https://www.sanger.ac.uk/science/tools/smalt-0)) to the completed *emm*89 reference genome H293 (HG316453.2) (3), as this genome contains no known prophage regions. Other reference genomes were also used where indicated with predicted prophage regions (Table S1) excluded to obtain “core” SNPs. Maximum likelihood phylogenetic trees were generated from aligned core SNPs using RAxML (39) with the GTR substitution model and 100 bootstraps. Regions of recombination were predicted using Gubbins analysis with the default parameters (40). Branches of phylogenetic trees were colored according to bootstrap support using iTOL (41).

Other genome sequence data were obtained from the short read archive. We combined data collected across England and Wales through Public Health England during 2014 and 2015 (PHE-2014/15) supplied by Kapatai et al. (13) and Chalker et al. (12) from invasive and noninvasive *S. pyogenes* isolates. We also used data supplied by Chochua et al. (14) collected by Active Bacterial Core Surveillance USA in 2015 (ABCs-2015) from invasive *S. pyogenes* isolates. ABCs-2015 sequence data were preprocessed by Trimmomatic (42) to remove adapters and low-quality sequences. PHE-2014/15 had already been preprocessed (12, 13). Genome data from these collections were assembled de novo using Velvet (assembly statistics provided in Data Set S2), and any isolates with greater than 2.2 Mbp total assembled length and/or more than 500 contig numbers were excluded. We also used data from Turner et al. (11) of invasive and noninvasive isolates from the Cambridgeshire region, UK, and collected through Cambridge University Hospital (CUH). We relied on the *emm* types determined during the original studies and excluded any data where the *emm* type was uncertain or negative. The genes hasA, hasB, hasC, covR, covS, and rocA, and the P-*nga*-ifs-slo were extracted from the assembled genome using *in silico* PCR ([https://github.com/simonharris/in_silico_pcr](https://github.com/simonharris/in_silico_pcr)). Capsule locus and P-*nga*-ifs-slo variants were also confirmed through manual inspection of mapping data where genotype could not be accurately determined from the assembly.

Mapping of *emm76*, *emm77*, and *emm81* sequence data was performed using *de novo* assembled genomes from one BSAC collection isolate representing the equivalent genotype. Prophage regions were predicted using PHASTER (43) and removed before SNP extraction.

Antimicrobial resistance genes were identified by srst2 (44) using the ARG-ANNOT database ([ARGannot.r2.fasta](https://github.com/BenJamesMetcalf/srastock)). The presence of prophage-associated superantigen genes speA, speC, speH, speI, speM, and ssa was determined using srst2 and the feature database previously used by Chochua et al. (14) available at [https://github.com/BenJamesMetcalf](https://github.com/BenJamesMetcalf). The presence of prophage-associated DNases genes sda, sdn, spd1, spd3, spd3v6, and spd4 was also determined using srst2 by adding regions of these genes to the feature database. Representative alleles of these DNase genes were taken from previous analysis (46) to identify regions that would detect all variants of each DNase, except we included spd3v6 separate from spd3 as it represents a divergent allele to spd3. Sequences used are available at Mendeleev ([https://data.mendeley.com/datasets/hzwjk2gt9p/1](https://data.mendeley.com/datasets/hzwjk2gt9p/1)).

**NADase activity.** Activity of NADase was measured in culture supernatants as previously described (3). Activity was determined as the highest dilution capable of hydrolyzing NAD+. Isolates were selected from the BSAC collection to represent different promoter variants for which there were three or more isolates available that were lacking mutations in regulatory genes.

**Data availability.** Sequence data have been submitted to the European Nucleotide Archive (ENA) ([www.ebi.ac.uk/ena](http://www.ebi.ac.uk/ena)) as accession numbers ERS361826 to ERS379364, ERR1359331 to ERR485881, ERS361826 to ERS379364, and SRR5853328 to SRR585742 (listed in Data Sets S1 and S2 in the supplemental material).

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at [https://doi.org/10.1128/mBio.02521-19](https://doi.org/10.1128/mBio.02521-19).

**FIG S1**, TIF file, 0.2 MB.

**FIG S2**, TIF file, 0.1 MB.
FIG S3, TIF file, 1.8 MB.
FIG S4, TIF file, 2.5 MB.
FIG S5, TIF file, 2.4 MB.
FIG S6, TIF file, 0.9 MB.
TABLE S1, DOCX file, 0.1 MB.
DATASET S1, XLSX file, 0.1 MB.
DATASET S2, XLSX file, 0.6 MB.
DATASET S3, XLSX file, 0.2 MB.

ACKNOWLEDGMENTS
This publication presents independent research supported by the Health Innovation Challenge Fund (WT098600, HICF-T5-342), a parallel funding partnership between the Department of Health and the Wellcome Trust. The work was also funded by the UK Clinical Research Collaboration (UKCRC, National Centre for Infection Prevention & Management) and the National Institute for Health Research Biomedical Research Centre awarded to Imperial College London.

The views expressed in this publication are those of the author(s) and not necessarily those of the Department of Health, NIHR, or Wellcome Trust. C.E.T. was an Imperial College junior Research Fellow and is a Royal Society & Wellcome Trust Sir Henry Dale Fellow (208765/2/17/2). S.J.P. is a consultant to Specific and Next Gen Diagnostics.

REFERENCES
1. Nasser W, Beres SB, Olsen RJ, Dean MA, Rice KA, Long SW, Kristinsson KG, Gottfredsson M, Vuopio J, Raisanen K, Caugant DA, Steinbak M, Low DE, McGeer A, Darenberg J, Henriques-Normark B, Van Beneden CA, Hoffmann S, Musser JM. 2014. Evolutionary pathway to increased virulence and epidemic group A Streptococcus disease caused by 3,615 genome sequences. Proc Natl Acad Sci U S A 111:E1768–E1776. https://doi.org/10.1073/pnas.1403138111.
2. Zhu L, Olsen RJ, Nasser W, Beres SB, Vuopio J, Kristinsson KG, Gottfredsson M, Porter AR, DeLeo FR, Musser JM. 2015. A molecular trigger for intercontinental epidemics of group A Streptococcus. J Clin Invest 125:3545–3559. https://doi.org/10.1172/JCI82478.
3. Turner CE, Abbott J, Lamagni T, Holden MT, David S, Jones MD, Game L, Elsatriou A, Sriskandan S. 2015. Emergence of a new highly successful acapsular group A Streptococcus clade of genotype emm89 in the United Kingdom. mBio 6:e00622-15. https://doi.org/10.1128/mBio.00622-15.
4. Zhu L, Olsen RJ, Nasser W, de la Riva Morales I, Musser JM. 2015. Trading capsule for increased cytotoxin production: contribution to virulence of a newly emerged clade of emm89 Streptococcus pyogenes. mBio 6:e01378-15. https://doi.org/10.1128/mBio.01378-15.
5. Beres SB, Olsen RJ, Ojeda Saavedra M, Ure R, Reynolds A, Lindsay DSJ, Smith AJ, Musser JM. 2017. Genome sequence analysis of emm89 Streptococcus pyogenes strains causing infections in Scotland, 2010-2016. J Med Microbiol 66:1765–1773. https://doi.org/10.1099/jmm.0.006622.
6. Friaes A, Machado MP, Pato C, Carrico J, Melo-Cristino J, Ramirez M. 2015. Emergence of the same successful acapsular clade among distinct populations of emm89 Streptococcus pyogenes in multiple geographic regions. mBio 6:e01780-15. https://doi.org/10.1128/mBio.01780-15.
7. Latronico F, Nasser W, Puhakainen K, Ollgren J, Hyrylainen HL, Beres SB, Lyytikainen O, Jalava J, Musser JM, Vuopio J. 2016. Genomic characteristics behind the spread of bacteremic group A Streptococcus type emm89 in Finland, 2004–2014. J Infect Dis 214:1987–1995. https://doi.org/10.1093/infdis/jiw468.
8. Hasegawa T, Hata N, Matsui H, Isaka M, Tatsuno I. 2017. Characterisation of clinically isolated Streptococcus pyogenes from balanoposthitis patients, with special emphasis on emm89 isolates. J Med Microbiol 66:511–516. https://doi.org/10.1099/jmm.0.000460.
9. Kimoto H, Fuji Y, Yokota Y, Taketo A. 2005. Molecular characterization of NADase-streptolysin O operon of hemolytic streptococci. Biochim Biophys Acta 1681:134–149. https://doi.org/10.1016/j.bbapap.2004.10.011.
10. Sumby P, Porcella SF, Madrigal AB, Barbian KD, Virtaneva K, Ricklefs SM, Sturdevant DS, Graham MR, Vuopio-Varkila J, Hoe NP, Musser JM. 2005. Evolutionary origin and emergence of a highly successful clone of serotype M1 group A Streptococcus involved multiple horizontal gene transfer events. J Infect Dis 192:771–782. https://doi.org/10.1086/432514.
11. Turner CE, Bedford L, Brown NM, Judge K, Torok ME, Parkhill J, Peacock SJ. 2017. Community outbreaks of group A Streptococcus revealed by genome sequencing. Sci Rep 7:8554. https://doi.org/10.1038/s41598-017-08914-x.
12. Chalker V, Jirinkin A, Coelho J, Al-Shalib A, Platt S, Kapatai G, Daniel R, Dhami C, Laranjeira M, Chambers T, Guy R, Lamagni T, Harrison T, Chamberlain J, Johnson AP, Underwood A, Scarlet Fever Incident Management Team. 2017. Genome analysis following a national increase in scarlet fever in England 2014. BMC Genomics 18:224. https://doi.org/10.1186/s12864-017-3603-z.
13. Kapatai G, Coelho J, Platt S, Chalker VJ. 2017. Whole genome sequencing of group A Streptococcus: development and evaluation of an automated pipeline for emm gene typing. PeerJ 5:e3226. https://doi.org/10.7717/peerj.3226.
14. Chochua S, Metcalf BJ, Li Z, Rivers J, Mathis S, Jackson D, Gertz RE, Jr, Srinivasan V, Lynfield R, Van Beneden C, McGee L, Beall B. 2017. Population and whole-genome sequence based characterization of invasive group A streptococci recovered in the United States during 2015. mBio 8:e01422-17. https://doi.org/10.1128/mBio.01422-17.
15. Kachroo P, Eraso JM, Beres SB, Olsen RJ, Zhu L, Nasser W, Bernard PE, Cantu CC, Saavedra MO, Arredondo MJ, Strope B, Do H, Kumaraswami M, Vuopio J, Grondahl-Yli-Hannukela K, Kristinsson KG, Gottfredsson M, Persen M, Pensar J, Davenport ER, Clark AG, Corander J, Caugant DA, Gaini S, Magnusen MD, Kubiak SL, Nguyen HAT, Long SW, Porter AR, DeLeo FR, Musser JM. 2019. Integrated analysis of population genomics, transcriptomics and virulence provides novel insights into Streptococcus pyogenes pathogenesis. Nat Genet 51:548–559. https://doi.org/10.1038/s41588-018-0343-1.
16. Flores AR, Jewell BE, Fittipaldi N, Beres SB, Musser JM. 2012. Human disease isolates of serotype m4 and m22 group A Streptococcus lack genes required for hyaluronic acid capsule biosynthesis. mBio 3:e00413-12. https://doi.org/10.1128/mBio.00413-12.
17. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR. 2009. Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. Lancet Infect Dis 9:611–616. https://doi.org/10.1016/S1473-3099(09)70178-1.
18. Green NM, Zhang S, Porcella SF, Nagiec MJ, Barbian KD, Beres SB, LeFebvre RB, Musser JM. 2005. Genome sequence of a serotype M28 strain of group A Streptococcus: potential new insights into puerperal sepsis and bacterial disease specificity. J Infect Dis 192:760–770. https://doi.org/10.1086/430618.
19. Jacob KM, Spilker T, LiPuma JJ, Dawid SR, Watson ME, Jr. 2016. Complete genome sequence of emm28 type Streptococcus pyogenes MEW123, a streptolysin-resistant derivative of a clinical throat isolate suitable for...
investigation of pathogenesis. Genome Announc 4:e00136-16. https://doi.org/10.1128/genomeA.00136-16.

20. Alam FM, Turner CE, Smith K, Wiles S, Sriskandand S. 2013. Inactivation of the CovR/S virulence regulator impairs infection in an improved murine model of Streptococcus pyogenes naso-pharyngeal infection. PLoS One 8:e61655. https://doi.org/10.1371/journal.pone.0061655.

21. Osoiwckii J, Azzopardi KL, McIntyre L, Rivera-Hernandez T, Ong CY, Baker C, Gillen CM, Walker MJ, Smeezers PR, Davies MR, Steer AC. 2019. A controlled human infection model of group A Streptococcus pharyngitis; which strain and why? mSphere 4:e00647-18. https://doi.org/10.1128/mSphere.00647-18.

22. Sierig G, Gywee C, Wessels MR, Ashbaugh CD. 2003. Cytotoxic effects of streptolysin O and streptolysin S enhance the virulence of poorly encapsulated group A streptococci. Infect Immun 71:446–455. https://doi.org/10.1128/iai.71.2.446-455.2003.

23. Lynskey SN, Goulding D, Gierula M, Turner CE, Dougan G, Edwards RJ, Sriskandand S. 2013. ROCa truncation underpins hypers-encapsulation, cartridge longevity and transmissibility of serotype M18 group A streptococci. PLoS Pathog 9:e1003842. https://doi.org/10.1371/journal.ppat.1003842.

24. Moses AE, Wessels MR, Zalcman K, Alberti S, Natanson-Yaron S, Menes T, Hanski E. 1997. Relative contributions of hyaluronic acid capsule and M protein to virulence in a mucoid strain of the group A Streptococcus. Infect Immun 65:64–71.

25. Flores AR, Jewell BE, Olsen RJ, Shelburne SA, III, Fittipaldi N, Beres SB, Moses AE, Wessels MR, Zalcman K, Alberti S, Natanson-Yaron S, Menes T, Hanski E. 1997. Relative contributions of hyaluronic acid capsule and M protein to virulence in a mucoid strain of the group A Streptococcus. Infect Immun 65:64–71.

26. Flores AR, Chase McNeil J, Shah B, Van Beneden C, Shelburne SA, III. 2018. Capsule-negative emm types are an increasing cause of pediatric infections. J Clin Microbiol 56:e00136-16. https://doi.org/10.1128/JCM.00136-16.

27. Turner et al. 2018. The DNases of pathogenic Lancefield streptococci. Microbiology 164:224–250. https://doi.org/10.1099/micro.0.006612.

28. Athey TB, Teatoro S, Li A, Marchand-Austin A, Beall BW, Fittipaldi N. 2014. Deriving group A Streptococcus typing information from short-read whole-genome sequencing data. J Clin Microbiol 52:1871–1876. https://doi.org/10.1128/JCM.00299-14.

29. Ibrahim J, Eisen JA, Jospin G, Coil DA, Khazen G, Tokajian S. 2016. Asymptomatic carriage of group A Streptococcus: basic biology to clinical manifestations. University of Oklahoma Health Sciences Center, Oklahoma City, OK. https://www.ncbi.nlm.nih.gov/books/NBK333409/.

30. Johnson DR, Kurlan R, Leckman J, Kaplan EL. 2010. The human immune response to streptococcal extracellular antigens: clinical, diagnostic, and potential pathogenic implications. Clin Infect Dis 50:481–490. https://doi.org/10.1086/650167.

31. Reynolds R, Hope R, Williams L, BSAC Working Parties on Resistance Surveillance. 2008. Survey, laboratory and statistical methods for the BSAC resistance surveillance programmes. J Antimicrob Chemother 62(Suppl 2):ii15–ii28. https://doi.org/10.1093/jac/dkn349.

32. Pospiech A, Neumann B. 1995. A versatile quick-prep of genomic DNA from gram-positive bacteria. Trends Genet 11:217–218. https://doi.org/10.1016/0168-9525(95)00060-2.

33. Page AJ, De Silva N, Hunt M, Quail MA, Parkhill J, Harris SR, Otto TD, Keane JA. 2016. Robust high-throughput prokaryote de novo assembly and improvement pipeline for Illumina data. Microb Genom 2:e000083. https://doi.org/10.1099/mgen.0.000083.

34. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313. https://doi.org/10.1093/bioinformatics/btu333.

35. Turner NJ, Page AJ, Connor TR, Delaney AJ, Keane JA, Bentley SD, Parkhill J, Harris SR. 2015. Rapid phylogenetic analysis of large samples of recombined bacterial whole genome sequences using Gubbins. Nucleic Acids Res 43:e15. https://doi.org/10.1093/nar/gku1196.

36. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. Nucleic Acids Res 47:W236–W259. https://doi.org/10.1093/nar/gkz239.

37. Pospiech A, Neumann B. 1995. A versatile quick-prep of genomic DNA from gram-positive bacteria. Trends Genet 11:217–218. https://doi.org/10.1016/0168-9525(95)00060-2.