AMERICAN SOCIETY FOR MICROBIOLOGY | mBio®

RESEARCH ARTICLE

Population and Whole Genome Sequence Based Characterization of Invasive Group A Streptococci Recovered in the United States during 2015

Sopio Chochua,[a] Ben J. Metcalf,[a] Zhongya Li,[a] Joy Rivers,[a] Saundra Mathis,[a] Delois Jackson,[a] Robert E. Gertz, Jr.,[a] Velusamy Srinivasan,[a] Ruth Lynfield,[b] Chris Van Beneden,[a] Lesley McGee,[a] Bernard Beall[a]

National Center for Immunization and Respiratory Diseases, Centers for Disease Control and Prevention, Atlanta, Georgia, USA[a]; Minnesota Department of Health, St. Paul, Minnesota, USA[b]

ABSTRACT Group A streptococci (GAS) are genetically diverse. Determination of strain features can reveal associations with disease and resistance and assist in vaccine formulation. We employed whole-genome sequence (WGS)-based characterization of 1,454 invasive GAS isolates recovered in 2015 by Active Bacterial Core Surveillance and performed conventional antimicrobial susceptibility testing. Predictions were made for genotype, GAS carbohydrate, antimicrobial resistance, surface proteins (M family, fibronectin binding, T, R28), secreted virulence proteins (Sda1, Sic, exotoxins), hyaluronate capsule, and an upregulated *nga* operon (encodes NADase and streptolysin O) promoter (Pnga3). Sixty-four M protein gene (*emm*) types were identified among 69 clonal complexes (CCs), including one CC of *Streptococcus dysgalactiae* subsp. *equisimilis*. *emm* types predicted the presence or absence of active *sof* determinants and were segregated into *sof*-positive or *sof*-negative genetic complexes. Only one "*emm* type switch" between strains was apparent. *sof*-negative strains showed a propensity to cause infections in the first quarter of the year, while *sof*$^{+}$ strain infections were more likely in summer. Of 1,454 isolates, 808 (55.6%) were Pnga3 positive and 637 (78.9%) were accounted for by types *emm1*, *emm89*, and *emm12*. Theoretical coverage of a 30-valent M vaccine combined with an M-related protein (Mrp) vaccine encompassed 98% of the isolates. WGS data predicted that 15.3, 13.8, 12.7, and 0.6% of the isolates were nonsusceptible to tetracycline, erythromycin plus clindamycin, erythromycin, and fluoroquinolones, respectively, with only 19 discordant phenotypic results. Close phylogenetic clustering of *emm59* isolates was consistent with recent regional emergence. This study revealed strain traits informative for GAS disease incidence tracking, outbreak detection, vaccine strategy, and antimicrobial therapy.

IMPORTANCE The current population-based WGS data from GAS strains causing invasive disease in the United States provide insights important for prevention and control strategies. Strain distribution data support recently proposed multivalent M type-specific and conserved M-like protein vaccine formulations that could potentially protect against nearly all invasive U.S. strains. The three most prevalent clonal complexes share key polymorphisms in the *nga* operon encoding two secreted virulence factors (NADase and streptolysin O) that have been previously associated with high strain virulence and transmissibility. We find that *Streptococcus pyogenes* is phylogenetically subdivided into loosely defined multilocus sequence type-based clusters consisting of solely *sof*-negative or *sof*-positive strains; with *sof*-negative strains demonstrating differential seasonal preference for infection, consistent with the recently demonstrated differential seasonal preference based on phylogenetic clustering of full-length M proteins. This might relate to the differences in GAS strain com-

Received 8 August 2017 **Accepted** 16 August 2017 **Published** 19 September 2017

Citation Chochua S, Metcalf BJ, Li Z, Rivers J, Mathis S, Jackson D, Gertz RE, 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.

Editor James M. Hughes, Emory University School of Medicine

Address correspondence to Bernard Beall, bbeall@cdc.gov.

S.C. and B.J.M. contributed equally to this work.

This article is a direct contribution from a Fellow of the American Academy of Microbiology. Solicited external reviewers: Mark Walker, University of Queensland; Michael Wessels, Boston Children's Hospital.

positions found in different geographic settings and could further inform prevention strategies.

KEYWORDS streptococcal disease seasonality, *Streptococcus pyogenes* vaccines, invasive group A streptococci, streptococcal genetics, streptococcal resistance, streptococcal virulence features

Approximately 1.8 million new severe disease infections attributed to group A streptococci (GAS) (acute rheumatic fever, rheumatic heart disease, poststreptococcal glomerulonephritis, and invasive disease) occur each year worldwide (1). In the United States, 10,600 to 13,400 invasive GAS infections occur annually, of which approximately 12% lead to death (2).

GAS are genetically very diverse, with various complements of virulence factors and adhesins (3). Whole-genome sequence (WGS) analyses have provided insights into the causal relationships of various surface proteins, secreted toxins, hyaluronic acid capsule, and transcriptional features with the success of individual GAS lineages. For example, the emergence of the pandemic M1 strain in the early 1980s and the later emergence of an *emm89* strain in the 2000s have both coincided with variation in the *nga* operon that increases the expression of two important toxins to enhance transmission and virulence (4, 5). Systematic large-scale population-based invasive GAS (iGAS) surveillance employing WGS can provide continued insights into transmission and disease manifestations.

The most intensely studied GAS vaccine candidates in the past 20 years have been multivalent M serotype-specific formulations, although a vaccine based on the conserved C repeat M region has been subjected to a phase 1 clinical trial (6). Geographic differences in M type distributions pose a significant problem for vaccine development; however, cross-opsonization between M types in a 30-valent type-specific vaccine has been demonstrated recently through the use of rabbit antisera in indirect opsonophagocytic killing assays (7, 8). A vaccine combining a multivalent M type-specific protein and a more conserved M-like protein could potentially target the majority of *Streptococcus pyogenes* strains (9).

Previous work associated *emm* locus patterns, based on sequences of peptidoglycan-spanning sequences encoded by *emm* and neighboring *emm*-like genes, with preferences for throat and/or skin reservoirs in human hosts (10). In the United States, common *emm* types such as 1, 3, and 12 were pattern A-C strains that demonstrated a preference for the throat reservoir. M types such as 4, 22, and 77 were pattern E strains that demonstrated no preference between the throat and skin reservoirs. Pattern E strain *emm* types were associated with the presence of the hypervariable *sof* gene or the *sof*-conferred ability to opacify serum, while pattern A-C strains were negative for these features (10). Pattern D strains are also usually *sof* negative and were associated with a preference for the skin disease reservoir (10). Short *emm* type-defining regions segregate into distinct *sof*-negative and *sof*-positive phylogenetic clusters (11). The recently described *emm* clustering scheme is based on phylogenetic comparisons of full-length M protein sequences, predicting structural and binding features of full-length M proteins (12). *emm* cluster types, *emm* locus patterns, and the presence or absence of *sof* are similarly predicted through knowledge of the *emm* type or classical M serotype alone (10, 12–14). Recently, it was shown that *emm* clusters can predict seasonality in that pattern A-C and E clusters display different propensities to cause invasive disease in summer months (15).

We describe here characteristics of iGAS strains obtained from 1 year of population-based surveillance in the United States, including *emm* types, multilocus sequence typing (MLST)-based genotypes, predicted resistosomes, and some key features associated with virulence and adhesion. Regional strain emergence, incidence of *emm* and *emm*-like vaccine candidates, *sof* genetic and seasonal associations, other virulence features, and even the identification of a successful iGAS lineage of *Streptococcus dysgalactiae* subsp. *equisimilis* are presented.

RESULTS

Potential coverage of a combined 30-valent type-specific M and Mrp protein vaccine. Of the 1,454 iGAS isolates in this study, 1,290 (88.7%) had 1 of 27 *emm* types targeted by an experimental 30-valent type-specific vaccine shown to elicit >50% killing of strains expressing individual vaccine M types (7, 8) (Fig. 1A; Table 1). In addition, a single query identifying the three classes of the M-like Mrp vaccine candidate (9) was positive for 140 (85.4%) of the 164 iGAS isolates of *emm* types not covered by the 30-valent vaccine. A majority of the strains (778/1,454 [53.5%]) would theoretically be targeted by both components of a potential combination vaccine (Fig. 1B; Table 1), while 98.3% (1,430/1,454) of the strains may be covered by at least one of the two vaccine components.

Resistance. Of the 1,454 isolates used in this study, 328 (22.6%) were detected with one or more accessory genes or chromosomal signatures associated with resistance or decreased susceptibility to antimicrobials (Table 2). Only 19 isolates (1.3%) with discrepant phenotypes were observed. Among these 19 isolates, there were 14 instances of undetected resistance that included 12 tetracycline-resistant group A *S. dysgalactiae* subsp. *equisimilis* (see Table S1 in the supplemental material for accession numbers) and 2 erythromycin-nonsusceptible (MICs of 0.5 to 1 μg/ml) isolates. Five isolates associated with false predictions of resistance are described in Table 2 footnote *b*.

Although β-lactam antibiotic resistance in *S. pyogenes* has not been reported, resistance to this class of antibiotics in this species would have a profoundly negative public health impact.

For this reason, we incorporated a WGS-based monitoring system. Determination of PBP2x transpeptidase sequence types (STs) of group B streptococci (GBS) and pneumococci is a very sensitive mechanism for detecting potential first-step mutations conferring decreased susceptibility or intermediate resistance to penicillin and other β-lactam antibiotics (16–18). We used this same approach with the corresponding PBP2x region from *S. pyogenes*. The MICs of the six β-lactams for all of the isolates were below the values previously flagged for GBS, which corresponded to 1 of 16 PBP2x types (Table S1). Of the 1,454 isolates, 1,107 (76.1%) shared type PBP2x-1, which served as the reference sequence.

Macrolide resistance was predicted in 210 (14.4%) of the isolates and in most cases (184/210, 87.6%) was associated with *erm* methylase genes (*ermT*, *ermB*, or *ermTR*) and either inducible or constitutive coresistance to clindamycin. Most isolates positive for *ermT* or *ermTR* (>80%) were inducibly clindamycin resistant, while most *ermB*-positive isolates (>90%) were constitutively resistant (data not shown). The most frequently occurring macrolide resistance determinant was *ermT*, because of its association with *emm92* as previously described (19) and possibly more recent associations with the *emm4*/ST39 and *emm77*/ST399 lineages (Table 1). As previously noted in GBS, the *ermT* determinant was detected at an approximately 10-fold greater read depth than other markers, consistent with its presence on a multicopy plasmid (17). Analysis of high-quality assemblies from two randomly selected *ermT*-positive strains (one type *emm4* [isolate 20155033] and one type *emm92* [isolate 20154014]) revealed that the single *ermT*-positive contig in each strain compared to previously described (17) plasmid pRW35 (4,968 bp) had 100% coverage, >99% identity, and a length of ~4,970 bp.

One strain contained the putative efflux determinant *lsaC* and also *ermB*. In GBS, we found that this combination confers decreased susceptibility to quinupristin-dalfopristin, presumably because of streptogramin A resistance conferred by *lsaC* and streptogramin B resistance conferred by *ermB* (17). In this iGAS isolate, the MIC of quinupristin-dalfopristin was somewhat higher (1 μg/ml) than the average (~0.3 μg/ml) but still below the MIC (2 μg/ml) indicative of intermediate resistance (20).

Approximately 95% of the GAS isolates tested had a ciprofloxacin MIC of $\leq$2 μg/ml and a levofloxacin MIC of $\leq$1 μg/ml. For this reason, we considered the 2-μg/ml MIC of both antibiotics an indicator of reduced susceptibility to fluoroquinolones. We found nine substitutions in ParC and/or GyrA that were highly associated with reduced susceptibility or nonsusceptibility to fluoroquinolones (Tables 2 and 3). Current guide-

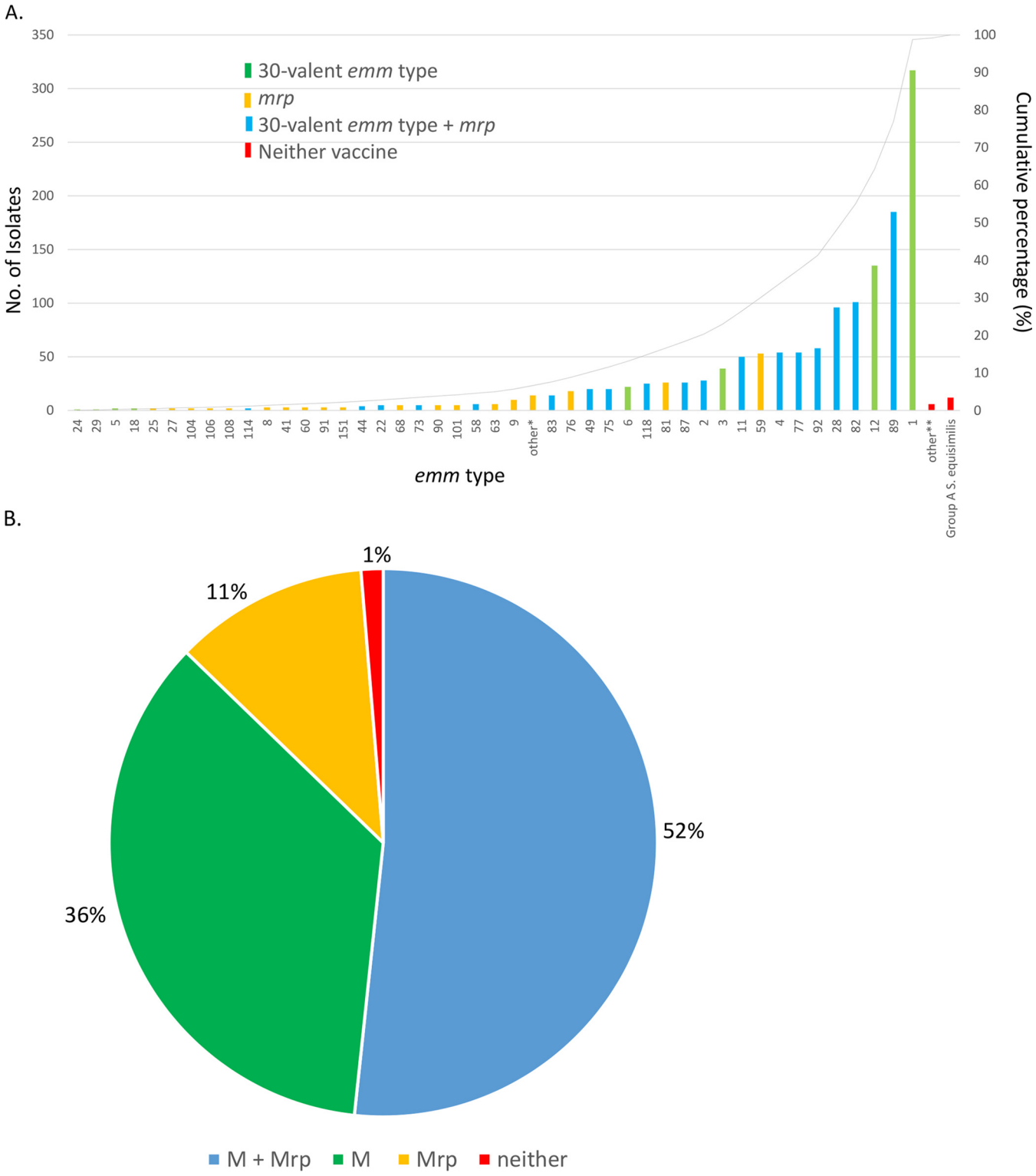

FIG 1 (A) *emm* type distribution data for the 1,454 iGAS isolates recovered through 2015 ABCs and their potential coverage by two experimental vaccines. Twenty-five *emm* types are shown in green as part of the 30 included in an experimental 30-valent M type-specific vaccine (7). Thirty-two types (orange) are non-30-valent vaccine *emm* types but are associated with one of three classes of the *emm*-like *mrp* gene included in an experimental trivalent vaccine (9). Eighteen *emm* types (blue) are included in the 30-valent vaccine and also associated with one of the *mrp* gene classes. Other*, 14 isolates of 13 *emm* types (not shown) were *mrp* positive. Other**, six isolates of four *emm* types (60, 227, 234, and 238; red) are not included in the 30-valent vaccine and also lack an *mrp* gene. Three additional *emm* types (not shown) were from 12 ST128 group A *S. dysgalactiae* subsp. *equisimilis* isolates (red). (B) Summary of *emm* type distribution data from 1,454 iGAS recovered in 2015 shown in four mutually exclusive categories.

TABLE 1 Cumulative bioinformatics pipeline data for 1,454 iGAS isolates recovered in 2015[a]

emm type	No. of isolates	MC	*tee* type	*emm*-like genes	Sof	*emm* locus pattern/*emm* cluster	Other fibronectin-binding repeat gene(s)	R28	*cpsA*	*sda1*	*sic*	Pnga3/ NADase 330G	Exotoxin genes
1	317	28	1	−	−	A-C/AC3	*fbaA*	−	+	+	+	+/+	A, c, G, J, k, s, Z
12	135	36	12	−	−	A-C/AC4	*prtF2, sfb1*	−	+	+	+	+/+	C, G, H, I, s, Z
3	39	15[c]	3	−	−	A-C/AC5	*prtF2*	−	+	−	−	+/+	A, c, G, K, S, z
6	22	382	6	−	−	A-C/M6	*sfb1*	−	+	−	−	−/+	A, C, G, h, i, K, Z
18	1	41[c]	49	*enn*	−	A-C/M18	*prtF2*	−	+	−	−	−/−	G, L, M, Z
18	1	535	3	*mrp, enn*	−	A-C/M18	*fbaA, prtF2, sfb1*	−	+	−	−	−/−	A, G, K, Z
238	2	867	−	−	−	A-C/AC3	*fbaA, sfb1*	−	+	−	+	−/−	A, G, J, Z
5	2	99	5	−	−	A-C/M5	*prtF2*	−	+	−	−	−/−	C, G, Z
24	1	70	−	−	−	A-C/M24	*fbaA, sfb1*	−	+	+	−	−/−	A, G, K, Z
29	1	12	91	−	−	A-C/M29	*prtF2*	−	+	−	−	−/−	G, L, M, Z
227[b]	1	28	1	−	−	A-C/AC3	*fbaA*	−	+	+	+	+/+	A, G, J, Z
234	1	914	−	−	−	A-C/ND	*fbaA, sfb1*	−	+	−	−	−/−	G
83	14	5	83	*mrp, enn*	−	D/D4	*fbaA, prtF2*	−	+	−	−	−/−	G, k, L, M, Z
101[d]	5	182	13	*mrp, enn*	−	D/D4	*fbaA, prtF2*	−	+	+	−	−/−	G, J, K, Z
41	3	579	83	*mrp, enn*	−	D/D4	*fbaA, prtF2*	−	+	−	−	−/−	C, G, L, M, Z
91	3	12	91	*mrp, enn*	−	D/D4	*fbaA, prtF2*	−	+	−	−	−/−	G, S, Z
108	2	14	−	*mrp, enn*	−	D/D4	*fbaA, sfb1*	−	+	−	−	+/−	C, G, J, L, M, S, Z
42	1	80	3	*mrp, enn*	−	D/E6	*fbaA, prtF2*	−	+	−	−	−	G, H, I, Z
54	1	302	91	*mrp, enn*	−	D/D1	*fbaA, prtF2*	−	+	−	−	−/−	G, Z
80	1	8	49	*mrp, enn*	−	D/D4	*fbaA, prtF2*	−	+	−	−	−/−	A, G, K, Z
111	1	214	91	*mrp, enn*	−	D/M111	*fbaA, prtF2*	−	+	−	−	−/−	G, J, S, Z
216	1	83	−	*mrp, enn*	−	D/ND	*fba*	−	+	−	−	−/−	A, S, Z
stG245[e]	3	se128	−	−	−	ND/ND	*sfb1* (1 isolate)	−	−	−	−	−/−	−
stG485[e]	3	se128	−	−	−	ND/ND	Neg	−	−	−	−	−/−	−
stG652[e]	6	se128	−	−	−	ND/ND	*sfb1* (2 isolates)	−	−	−	−	−/−	−
Negative	1	36	12	−	−	NA	*prtF2, sfb*	−	+	+	+	+/+	C, G, H, I, Z
15	1	872	3	*mrp, enn*	−	E/E3	*fbaA, prtF2, sfb1*	−	+	−	−	−/−	A, C, G, Z
89	185	101	89, 11[c]	*mrp, enn*	+	E/E4	*fbaA, prtF2, sfb1*	−	−	−	−	+/+	C, G, k, Z
82	99	334	5	*mrp, enn*	+	E/E3	*fbaA, prtF2, sfb1*	−	+	−	−	−/−	C, G, H, I, l, m, Z
82	2	36	12	*mrp*	+	E/E3	*fbaA, prtF2*	−	+	+	−	+/+	A, C, G, H, I, Z
28	95	52	28	*mrp, enn*	+	E/E4	*fbaA, prtF2, sfb1*	+	+	−	−	−/−	a, C, G, h, i, J, k, Z
92	58	82	92	*mrp, enn*	+	E/E2	*fbaA, prtF2*	−	+	−	−	−/−	C, G, J, Z
4	52	39	4	*mrp, enn*	+	E/E1	*fbaA, sfb1*	−	−	−	−	+/+	a, C, S, Z
77	35	63	13	*mrp, enn*	+	E/E4	*fbaA, prtF2, sfb1*	+	+	−, +	−	+/+	a, c, k, l, m, Z
77	12	399	59	*mrp, enn*	+	E/E4	*fbaA, prtF2*	−	+	−	−	−/+	c, G, h, i, j, k, l, m, Z
77	7	133	5	*mrp, enn*	+	E/E4	*fbaA, prtF2, sfb1*	−	+	−	−	−/−	G, M, Z
59	53	172	59	*mrp, enn*	+	E/E6	*fbaA, prtF2, sfb1*	−	+	−	−	−/+	c, G, J, k, Z
11	50	403	11	*mrp, enn*	+	E/E6	*fbaA, prtF2, sfb1*	−	+	−	−	−/+	C, G, H, I, k, s, Z
2	28	55	2	*mrp, enn*	+	E/E4	*fbaA*	+	+	−	−	−/+	C, G, h, i, K
118	25	167	13	*mrp, enn*	+	E/E3	*fbaA, prtF2, sfb1*	−	+	−	−	−/−	C, G, Z, h
87	26	62	28	*mrp, enn*	+	E/E3	*fbaA, prtF2, sfb1*	−	+	−	−	+/+	C, G, J, k, S, Z
81	24	624	12	*mrp/mrp, enn*	+	E/E6	*fbaA, prtF2*	−	+	−	−	+/+	c, G, H, I, m
81	1	903	6	*mrp, enn*	+	E/E6	*fbaA, sfb1*	−	+	−	−	−/−	G, J, Z
81	1	901	81	*mrp, enn*	+	E/E6	*fbaA, prtF2, sfb1*	−	+	−	−	−/−	G, Z
75	20	49	25	*mrp, enn*	+	E/E6	*fbaA, sfb1*	−	+	−	−	−, +/+	C, G, H, i, k, L, M, Z
49	20	433	49	*mrp, enn*	+	E/E3	*fbaA, prtF2*	−	+	−	−	−/+	a, c, G, H, I
76	16	50	−/12	*mrp, enn*	+	E/E2	*fbaA, sfb1*	−	+	−	−	−, +/+	c, G, h, i, J, l, m, Z
76	2	631	−	*mrp, enn*	+	E/E2	*fbaA, sfb1*	−	+	−	−	−/+	G, H, K, Z
9	10	75	9	*mrp, enn*	+	E/E3	*fbaA, prtF2, sfb1*	−	+	−	−	−/+	c, G, l, m, S, Z
22	5	46	12	*mrp, enn*	+	E/E4	*fbaA, prtF2, sfb1*	−	−	−	−	+/+	a, C, G, k, S, Z
58	7	176	58	*mrp, enn*	+	E/E3	*fbaA, sfb1*	−	+	−	−	−/−	c, G, h, S, Z
63	5	385	−	*mrp, enn*	+	E/E6	*fbaA, sfb1*	−	+	−	−	−/+	Z
63	1	210	−	*mrp, enn*	+	E/E6	*fbaA, sfb1*	−	+	−	−	−/+	G, Z
60	5	53	−	*mrp, enn*	+	E/E1	*fbaA, sfb1*	−	+	−	−	−/+	−
68	3	247	89	*mrp, enn*	+	E/E2	*fbaA, prtF2, sfb1*	−	+	−	−	−/+	C, G, J, L, M, Z
68	2	894	12	*mrp, enn*	+	E/E2	*fbaA, prtF2, sfb1*	+	+	−	−	−/+	C, G, J, Z
73	5	331	13	*mrp, enn*	+	E/E4	*fbaA, prtF2, sfb1*	−	+	−, +	−	−/−	c, G, H, I, Z
90	5	184	13	*mrp, enn*	+	E/E2	*fbaA, prtF2, sfb1*	−	+	−	−	−/+	c, G, j, Z
44	3	25	5	*mrp, enn*	+	E/E3	*fbaA, prtF2, sfb1*	−	+	−	−	−/−	G, J, S, Z
44	1	641	−	*mrp, enn*	+	E/E3	*fbaA, sfb1*	−	+	−	−	−/+	G, H, J, K

(Continued on next page)

TABLE 1 (Continued)

emm type	No. of isolates	MC	*tee* type	*emm*-like genes	Sof	*emm* locus pattern/*emm* cluster	Other fibronectin-binding repeat gene(s)	R28	*cpsA*	*sda1*	*sic*	Pnga3/ NADase 330G	Exotoxin genes
8	2	59	−	*mrp*, *enn*	+	E/E4	*fbaA*, *sfb1*	−	+	−	−	−/+	G, H, L, M, Z
8	1	869	−	*mrp*, *enn*	+	E/E4	*fbaA*, *prtF2*, *sfb1*	−	+	−	−	−/+	G, L, M, Z
25	2	192	−	*mrp*, *enn/mrp*	+	E/E3	*fbaA*, *sfb1*	−	+	−	−	−/+	G, Z
27	2	308	5	*mrp*, *enn*	+	E/E2	*fbaA*, *prtF2*, *sfb1*	−	+	+	−	−/+	C, G, J, Z
104	2	137	11	*mrp*, *enn*	+	E/E2	*fbaA*, *prtF2*, *sfb1*	−	+	−	−	−/+	L, M, Z
106	1	900	59	*mrp*, *enn*	+	E/E2	*fbaA*, *prtF2*, *sfb1*	−	+	−	−	−/+	G, Z
106	1	338	3	*mrp*, *enn*	+	E/E2	*fbaA*, *prtF2*, *sfb1*	−	+	−	−	−/+	G, Z
114	1	188	81	*mrp*, *enn*	+	E/E4	*fbaA*, *prtF2*, *sfb1*	−	+	−	−	−/+	G, Z
114	1	220	13	*mrp*, *enn*	+	E/E4	*fbaA*, *prtF2*, *sfb1*	−	+	−	−	−/+	G, Z
165	1	768	−	*mrp*, *enn*	+	E/E1	*fbaA*, *sfb1*	−	+	−	−	−/+	H, I, L, M, Z
85	1	849	59	*mrp*, *enn*	+	E/E6	*fbaA*, *prtF2*, *sfb1*	−	+	−	−	−/+	G, I, J, Z
94	1	89	12	*mrp*, *enn*	+	E/E6	*fbaA*, *prtF2*	−	+	−	−	−/+	G, H, Z
102	1	376	−	*mrp*, *enn*	+	E/E4	*fbaA*, *sfb1*	−	+	−	−	−/+	C, G, J, L, M, Z
113	1	148	12	*mrp*, *enn*	+	E/E3	*fbaA*, *prtF2*, *sfb1*	−	+	−	−	−	G, Z
232	1	185	11	*mrp*, *enn*	+	E/E4	*fbaA*, *prtF2*, *sfb1*	−	+	−	−	−	G, H, I, J, Z
151[b]	2	433	49	*mrp*, *enn*	+	E/E3	*fbaA*, *prtF2*	−	+	−	−	−	G, H, I

[a]*sof*-negative isolates were added first, and then isolates were added in order of frequency according to *emm* locus pattern (10) or *emm* cluster (12). For some parameters (e.g., *sda1*) there was >25% variability in a given strain complex. Exotoxin genes for which ≥75% of the isolates were positive (*spe* genes and *smeZ* are represented by the last letter, and *ssa* is represented by S) are capitalized. For all of the parameters whose presence (+) or absence (−) is indicated, the symbols indicate ≥90% positivity or negativity, respectively.

[b]In-frame deletion derivative of precursor *emm* type. *emm227* is a 17-codon deletion derivative of *emm1.0*. *emm151* is an 11-codon deletion derivative of *emm49*. All *emm* subtype sequences are available at ftp://ftp.cdc.gov/pub/infectious_diseases/biotech/tsemm/.

[c]Contains a *rocA* null mutation that is conserved in the lineage of *emm3* and *emm18* strains indicated (3).

[d]There were six single-locus variants of ST101 (five ST407 and one ST910) that differed from the other isolates in this strain complex in being T11, *hasA* positive, and Pnga3 negative.

[e]This *S. dysgalactiae* subsp. *equisimilis* isolate contains the *gacI* gene required for group A carbohydrate synthesis (28). Se128 refers to ST128 in the *S. dysgalactiae* MLST database (https://pubmlst.org/sdysgalactiae/).

lines do not assign cutoff values for ciprofloxacin, although we found that isolates with ciprofloxacin MICs of ≥4 μg/ml contained uncommon substitutions in ParC. All instances of intermediate levofloxacin resistance (MIC of 4 μg/ml) contained one of these three substitutions in ParC. All three levofloxacin-resistant strains contained the GyrA-S81F substitution, two of which additionally contained the ParC S79F or S79Y substitution (Table 3).

Four instances of chloramphenicol resistance, corresponding to the presence of a *cat* gene, were found (Table 2). We found no resistance to rifampin, gentamicin, or vancomycin, consistent with finding no previously described *rpoB* substitutions associated with rifampin resistance (17, 21, 22) and no genes for gentamicin or vancomycin resistance (23, 24). All isolates were also susceptible to daptomycin and linezolid.

***S. pyogenes* strain diversity as assessed by MLST and *emm* type.** The 1,454 isolates comprised 70 different MLST complexes (MCs) and 64 different *emm* types (Table 1). On the basis of a looser criterion than MCs (described in Materials and Methods), there were 15 clonal groups and 28 singletons (Fig. 2). There were generally strong associations of genetic features with the MCs depicted in Table 1.

***emm* types.** Type *emm1* was the most frequently occurring type overall (Table 1), accounting for 21.7% of the isolates, and was among the three most common *emm* types at each ABCs site (https://www.cdc.gov/abcs/reports-findings/survreports/gas15.html). Types *emm89* and *emm12* were also frequent (12.9 and 9.3% of the isolates tested, respectively) and widely distributed. With the exception of *emm82*, the five most frequently occurring *emm* types in 2015 (*emm1*, *emm89*, *emm12*, *emm82*, and *emm28*) have each been among the five most frequently occurring types each year since 2012 (*emm1* was the most frequent each year).

Recent regional emergence of *emm59*. Regional instances of *emm* type emergence were apparent. For example, *emm59* was predominant in New Mexico in 2015 but infrequent elsewhere (https://www.cdc.gov/abcs/reports-findings/survreports/gas15.html). *emm59* was also predominant in New Mexico in 2014 (https://www.cdc.gov/abcs/reports-findings/survreports/gas14.html). In 2015, an *emm59* clone had also

TABLE 2 Resistance features of 1,454 iGAS recovered in 2015[a]

Determinant(s) found in ≥1 isolates	Predicted resistance to:					No. of isolates	No. of isolates with discordant phenotypes (relevant MIC [μg/ml])[b]	*emm*/ST complex associated with resistance determinant(s) when >1 isolate (no. of isolates)
	Tet	Ery	Cli	Fq	Other			
None						1,126	11 (9 Tet, 4–8; 2 Ery, 0.5–1)	
ermT, *tetM*	58	58	58			58	0	92/82 (51), 77/399 (4)
ermT		25	25			25	1 (Ery + Cli, <0.5)	4/39 (25)
ermT, *ermTR*, *tetM*	1	1	1			1	0	
ermB, *tetM*	52	52	52			52	0	11/403 (31), 9/75 (4), 76/50 (4), 12/36 (3), 73/331 (2), 77/399 (2)
ermB, *ermTR*, *tetM*	3	3	3			3	0	11/403 (3)
ermB, *lsaC*, *tetM*	1	1	1		1 (Syn)	1	1 (Syn, 1.0)	
ermB, *cat*, *tetM*	1	1	1		1 (Chl)	1	0	
ermB, ParC 79F, *tetM*	1	1	1	1		1	0	
ermB		1	1			1	0	
ermTR, *tetO*	18	18	18			18	0	77/63 (10)
ermTR		14	14			14	0	12/36 (8), 89/101 (3)
ermTR, *tetM*	13	13	13			13	1 (Tet, <4)	83/853 (9), 58/176 (3)
ermTR, *tetT*	2	2	2			2	0	8/59 (2)
ermTR, ParC 83G, *tetM*	1	1	1	1		1	0	
ermTR, ParC 79F, GyrA 81F		1	1	1		1	0	
ermTR, ParC 83N		1	1	1		1	0	
ermTR, ParC 78A		1	1	1		1	0	
mef, *msrD*		12				12	3 (Tet, 4–8)	12/36 (5), stG245/128 (3), 75/49 (2)
tetM, *mef*, *msrD*, *cat*	3	3			3 (Chl)	3	0	3/15 (3)
mef, *msrD*, GyrA 81F		1		1		1	0	
tetM	61					61	2 (Tet, <4)	76/50 (12), 77/133 (6), 90/184 (5), 60/53 (5), 92/82 (4), 77/399 (3), 68/247 (3), 1/28 (2), 82/334 (3), 25/192 (2), 89/101 (2), 101/182 (2), 104/137 (2), 108/14 (2)
tetM, ParC 79F	1			1		1	0	
tetM, ParC 83G	1			1		1	0	
tetO	15					15	0	77/63 (15)
ParC 79A				23		23	0	6/382 (22)
ParC 79F				7		7	0	89/101 (2), 28/52 (2)
ParC 83N				6		6	0	89/101 (4)
ParC 83G				2		2	0	1/28 (2)
ParC 79Y, Gyr A81F				1		1	0	
ParC 79Y				1		1	0	
Total	223	200	184	48[c]		1,454	19	

[a]Abbreviations: Tet, tetracycline; Ery, erythromycin; Cli, clindamycin; Fq, ciprofloxacin and levofloxacin; Syn, quinupristin plus dalfopristin (Synercid); Chl, chloramphenicol.

[b]These 18 isolates consisted of the following. There were 12 tetracycline-nonsusceptible (MICs, 2 to 6 μg/ml) group A *S. dysgalactiae* subsp. *equisimilis* isolates without detected tetracycline resistance determinants (9 lacked any resistance determinants, and 3 were positive for *mef* and *msrD*). There were two erythromycin-nonsusceptible isolates (MICs, 0.5 and 1 μg/ml) without detected erythromycin resistance determinants (isolates 20152764 and 20161074 lacked detectable resistance determinants). Three *tetM*-positive isolates (20154605, 20154042, and 20160491) were tetracycline susceptible (two had *tetM*$^+$, and one had *tetM*$^+$ and *ermTR*$^+$). One *ermT*-positive isolate (20160450) was susceptible to erythromycin and clindamycin. One isolate (*ermB lsaC tetM*; 20156709) predicted to be intermediately resistant to quinupristin-dalfopristin had an E-test value of 1 μg/ml (below the cutoff of 2 μg/ml).

[c]Results of phenotypic testing of these 48 isolates are shown in Table 3.

emerged in Arizona, a non-ABCs state (25). We were unable to directly compare the phylogeny of New Mexico primary and secondary *emm59* clades with recent Arizona *emm59* isolates (25), since WGS data from these 18 isolates were not provided. However, two genomic sequences closely related to the Arizona cluster were available from two New Mexico isolates recovered in 2011 and 2012 (SRR1574573 and SRR1574608; a single Colorado outlier from this study is also shown) and included in our phylogenetic analysis of ABCs 2015 type *emm59* isolates (Fig. 3). The close similarity of the two recently described New Mexico isolates to the 2015 clade 1 isolates (clade 1 is ST864, while clade 2 is ST172 and *speC* positive, with the exception of 20154161, which was used to root the tree) indicates very close relatedness (compare Fig. 3 with Fig. 2 in reference 25; however, these two Arizona isolates differed from the ABCs 2015

TABLE 3 Summary of fluoroquinolone MICs for 48 isolates containing ParC and/or GyrA substitutions

ParC and/or GyrA substitution	MIC (μg/ml) of Cip/Lev[a] at:			
	2/2 μg/ml	4/2 μg/ml	4/4 μg/ml	>8/>8 μg/ml
ParC S79A	12	9	2	0
ParC S79F	2	3	4	0
ParC D78A	0	0	1	0
ParC D83N	4	3	0	0
ParC D83G	3	1	0	0
ParC S79Y	0	1	0	0
ParC S79F, GyrA S81F	0	0	0	1
ParC S79Y, GyrA S81F	0	0	0	1
GyrA S81F	0	0	0	1

[a]Abbreviations: Cip, ciprofloxacin; Lev, levofloxacin.

New Mexico clade 1 isolates by the presence of *speC*. The five New Mexico clade 2 isolates from the 2015 ABCs also form a very tight phylogenetic cluster.

Potential *emm* type switching. There were only four instances of a single *S. pyogenes* ST (ST28, ST433, ST12, and ST36) associated with more than a single *emm* type, only one of which, *emm82* in the common *emm12* genotype ST36, appeared representative of an *emm* gene switching event (described below). The predominant global *emm1* lineage is ST28. The single *emm227*/ST28 strain is a deletion derivative of *emm1* (predicted to lack processed M protein residues 17 to 24). Similarly, the three *emm151* isolates were ST433, as expected of a derivative of M49/ST433 with mature M protein residues 3 to 13 deleted. The relationship between the two unrelated *emm* types, *emm29* (one isolate) and *emm91* (three isolates), in ST12 is not straightforward. ST12

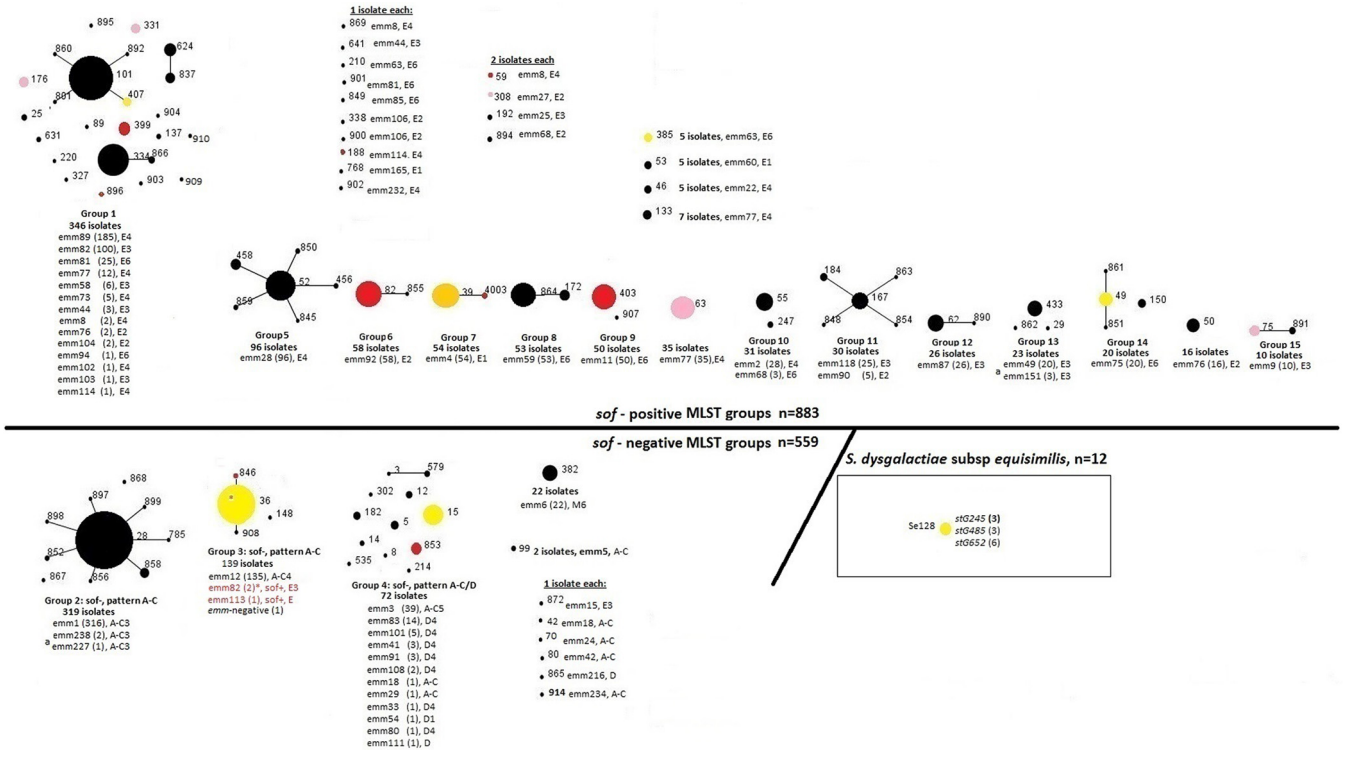

FIG 2 eBURST analysis (53) of 1,454 iGAS isolates recovered in 2015. The 15 eBURST groups consist of isolates that share four or more MLST alleles with at least one other member of the set. Twenty-eight STs are singletons representing as few as 1 or as many as 35 isolates. Singleton STs do not share four or more MLST alleles with other GAS isolates included in this study. Of the 883 *sof*-positive isolates, only 3 are closely genetically associated (≥4 identical MLST alleles) with any *sof*-negative strains. These three exceptions, shown in group 3 (red font), include two *emm82* serotype-switching variant isolates of ST36 (explained in the text and in Fig. 4) and one *emm113*/ST148 isolate. STs shaded yellow, orange, pink, and red indicate macrolide resistance proportions of 10 to 25, 26 to 50, 51 to 75, and 76 to 100%, respectively. A small asterisk is indicated within the main ST36 circle to represent the single *emm12* deletion strain.

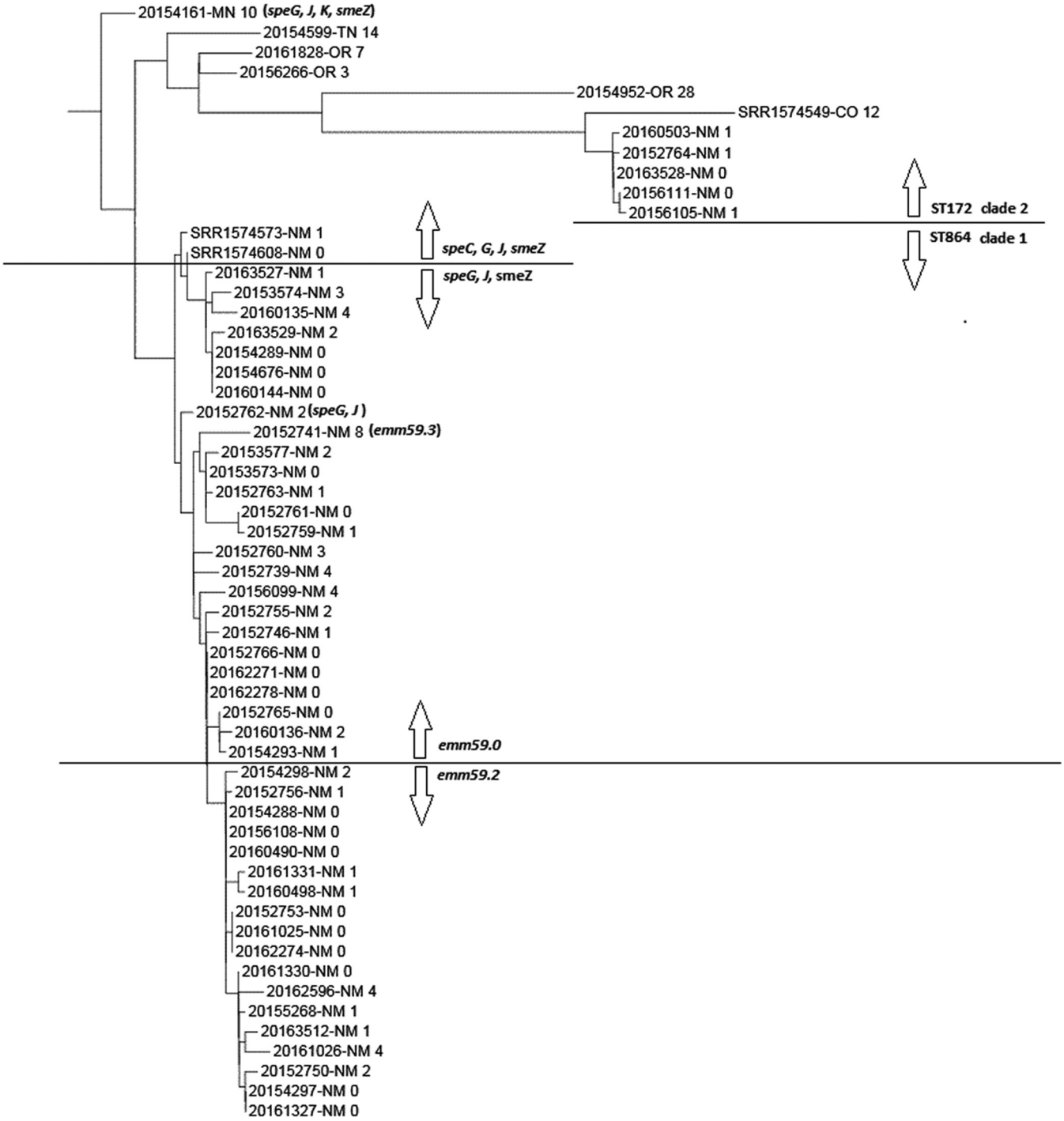

FIG 3 Phylogenetic relationships of *emm59* ABCs isolates recovered in 2015. All but 5 of these 53 isolates were recovered in New Mexico. The values shown are SNP differences. Three additional isolates (recovered in 2011 and 2012, two in New Mexico and one in Colorado) were obtained from the GenBank database from reference 25. Clade 1 is defined by ST864, while clade 2 is defined by ST172. All isolates above or below the arrows share the features indicated, with three exceptions (20154161, 20152762, and 20152741).

was first documented in an *emm91.0* strain recovered in 1943 and subsequently documented among five *emm29* isolates recovered from 1997 to 2004 in the United States and Germany and three *emm91* isolates recovered from 1990 to 1994 in Chile and Australia.

Unusual *emm* switching and *emm* deletion in the ST36 lineage. There were two *emm82* isolates (20154051 [*emm82.0*] and 201624915 [*emm82.7*]) that were ST36, the

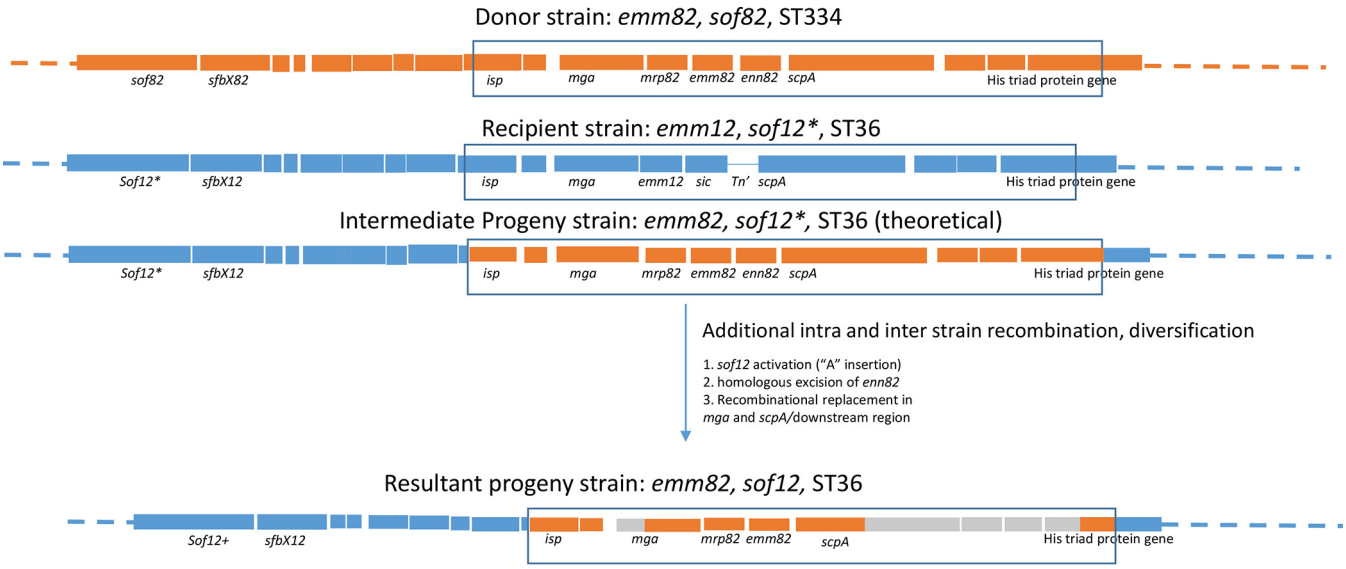

FIG 4 Gene replacement event indicative of replacement of the *emm12-sic* region with the *mrp82-emm82* region. The parental donor (orange) and recipient (blue) strain chromosomal regions are indicated on the basis of observed sequence homologies. The homologies observed suggest that the gene replacement event involved a large *isp*–His triad-encoding gene region, followed by a deletion event facilitated by homologous sequences in the tandem *emm82* and *enn82* 3′ regions. The inactivating frameshift in *sof12* (indicated by *sof12**) was removed by a single base insertion (indicated by *sof12*$^{+}$). Additional recombination events involving unknown donors are indicated by gray zones in the *isp* and *scpA* downstream region.

MLST lineage of the global *emm12* strain. Both *emm82*/ST36 strains were positive for *sof* and *mrp* and negative for the *sic* determinant. Comparison of the approximately 22-kb *sof-emm* region of the two *emm82*/ST36 strains with those of the putative parental *emm12*/ST36 (recipient) and donor (*emm82*/ST334) strains predicted their descent from the same progenitor (Fig. 4). Both strains revealed the same likely crossover point in the *isp* gene 5.8 kb upstream of *emm82* at (base 1682938 relative to the *emm12*/ST36 reference genome [accession no. CP000259]). The downstream crossover point for the switching event appears to have been in a downstream histidine triad gene approximately 8.4 kbp downstream of the *emm82* gene (base 1665610 relative to the *emm12*/ST36 reference genome). Following this double-crossover event, we predict that there was homologous excision of the *enn82* gene that normally lies downstream of *emm82*, facilitated by the near sequence identity between the *emm82* and *enn82* 3′ regions (see GenBank accession number CP007561 from an *emm82*/ST334 strain for comparison). All of the *emm12*/ST36 strains analyzed to date contain a conserved single base deletion predicted to result in a truncated nonfunctional 746-residue protein lacking its fibronectin-binding repeats (26). Both *emm82*/ST36 strains shared the same highly conserved *sof* gene shared by *emm12* strains (14, 26); however, their *sof12* allele no longer contained the inactivating single base deletion and was predicted instead to encode a full-length 1,019-residue Sof12 protein inclusive of fibronectin-binding repeats and the C-terminal membrane anchor. Consistent with this observation, the two *emm82*/*sof12* recombinant progeny strains were found to be serum opacity factor positive.

One isolate (20160179), also of major *emm12* lineage ST36, was not *emm* typeable. Surprisingly, this strain lacked both the *emm12* and *sic* genes, corresponding to a precise deletion of a 4,892-bp region between *mga* and *scpA* (relative to the recipient strain shown in Fig. 4, bases 1673889 to 1678781 relative to the *emm12*/ST36 reference genome [GenBank accession number CP000259]).

***emm* types present among multiple genetic backgrounds in *S. pyogenes*.** While only one instance of *emm* gene switching between different lineages was clear (two *emm82*/ST36 strains described above), there were 10 additional instances of the same *emm* type distributed between two and three completely unrelated MLSTs (Table 1). There is insufficient information to track the origins of these different *emm*/ST combi-

nations. An example that shows the genetic diversity in certain *emm* types is type *emm77*, which is shared among three unrelated multi-isolate MLST-based lineages. The differences between these distinct MLST-based lineages are reflected in their different genetic features (Table 1). The common *emm77.0*/ST63 lineage (35 isolates found in nine ABCs states) has long been documented in Germany, Poland, and the United States (https://pubmlst.org/spyogenes/). The *emm77.0*/ST399 lineage (12 isolates, includes one single-locus variant, ST904) was also found in 2015 in nine ABCs states, with the only known documentation of ST399 being a single *emm77* isolate recovered in Thailand (https://pubmlst.org/spyogenes/). Finally, the sole association of *emm77*/ST133, shared by seven isolates recovered in Tennessee (Table S1), is actually with the original *tee* (T) type 5 Lancefield M27 reference strain recovered more than 60 years ago. Briefly, this type was subsequently redesignated *emm77* by the CDC *emm* database curator in the 1990s (B. Beall, unpublished data) because of its *emm* sequence identity to the prevalent *emm77*/ST63 lineage and its sequence dissimilarity from the original Griffiths M27 strain, which is the current established type *emm27* (13, 27).

Surface protein determinants. In general, MCs were strongly associated with specific *emm* and T types and the presence or absence of additional surface protein genes. These included fibronectin-binding repeat motif-containing genes (*sof*, *fbaA*, *prtF2*, *sfb1*) (48) and the *emm*-like *mrp* and *enn* virulence genes that flank *emm* in many strains (9). These two *emm*-like genes show much less interstrain variation than *emm* genes and are also virulence factors. All *sof*-negative *S. pyogenes* strains were associated with previously described *emm* cluster A-C or D types (or patterns) (10), with the exception of a single *emm15* (pattern E/cluster E3) isolate. Type *emm15* is the only cluster/pattern E *emm* type in this study that is also historically associated with the serum opacification-negative phenotype (13).

GAS pili are important virulence factors that function in epithelial adhesion (reviewed in reference 3. Most (1,388/1,454, 95.5%) isolates had 1 of 21 different pilus (*tee*) types, corresponding to different pilus backbone protein subunit genes (29) and classical T agglutination types (13). In individual MCs, most *tee* types (based on 120- to 240-bp gene segment queries) were generally predictive of highly conserved 950- to 1,800-bp open reading frames that shared >95% sequence identity within the type (data not shown). The single exception included the *tee* gene of an *emm106*/ST338 isolate that shared only 82.5% sequence identity with the previously described reference *tee3* gene. As previously described (29), *tee* genes were highly diverse but exhibited differing regions of inter-*tee* gene homology and all contained signal sequence motifs situated near the 5′ ends with SrtB sortase family wall attachment motifs near the 3′ ends.

The R28 antigen gene, which is a close homolog of a group B streptococcal adhesin (30), was found in four strain complexes (*emm28*/ST52, *emm77*/ST63, *emm2*/ST55, *emm68*/ST894) that included 157 (10.8%) of the 1,454 isolates.

Extracellular virulence determinants. The *sda1* (virulence-associated DNase) (49) determinant was found in 450 (30.9%) of the 1,454 isolates in seven strain complexes (*emm1*/ST28, *emm12*/ST36, *emm77*/ST63, *emm101*/ST182, *emm81*/ST624, *emm27*/ST308, and *emm24*/ST70).

A query for the streptococcal inhibitor of complement (*sic*) gene (31, 32) employed a short sequence targeting the derivatives found in the *emm* region of the prevalent *emm1*/MC28 and *emm12*/MC36 lineages, which were primarily positive for the query (309/316 for *emm1*, 114/135 for *emm12*). In addition, an *emm227*/ST28 deletion derivative of *emm1* (Table 1) and two *emm228* isolates (both double-locus variants of ST28) were *sic* positive.

Exotoxin gene patterns were generally highly associated with GAS *emm*/ST-defined lineages. For example, this was evident in *emm1*/MC28, where 306/316 isolates were positive for *speA*, *speG*, *speJ*, and *smeZ*. Three *emm1*/ST28 isolates were positive for additional exotoxin genes besides these four. One *emm1*/ST28 isolate was additionally positive for *speK*, and another was additionally positive for *speC*. Finally, *emm1*/ST28

isolate 20156011 was additionally positive for *speC* and *ssa*. Examination of the genome revealed that *spd1* (DNase gene), *speC*, and *ssa* were tandemly situated in a prophage sequence, as in the highly related ϕHKU488.vir prophage from an antimicrobial-resistant *emm1* strain causing scarlet fever in Hong Kong (33). Unlike the Hong Kong *emm1* strain, ABCs strain 20156011 was susceptible to antibiotics and lacked *ermB* and *tetM* determinants.

Capsular biosynthetic locus. The *emm89*/ST101, *emm4*/ST39, and *emm22*/ST46 lineages lacked the *hasA* hyaluronic acid synthetase determinant, as previously described (5, 35, 55), accounting for 233 (16.2%) of the 1,442 *S. pyogenes* isolates. Single *hasA*-negative *emm1*, *emm11*, and *emm12* isolates were also in the isolate set. The 12 group A *S. dysgalactiae* subsp. *equisimilis* isolates (described below) also lacked a sequence similar to the *hasA* query.

***nga* operon markers previously associated with strain emergence.** The variant 3 *nga* promoter (Pnga3) has been associated with increased transcription of the *nga* operon (4). Pnga3 is a promoter sequence associated with increased transcriptional activity relative to the previously described less active *emm89* clade 1 and 2 promoters situated upstream of the genes (*nga* and *slo*) encoding the extracellular toxins NADase and streptolysin O (5). The presence of Pnga3 was invariably linked to the putatively active NADase 330G query among the study isolates (4). Consistent with Pnga3 and NADase 330G being associated with transmissibility and virulence, these two features were evident in the three most frequently occurring strain complexes. These three strain complexes (*emm1*/MC28, *emm89*/MC101, and *emm12*/MC36) accounted for 44% (636 isolates) of the entire iGAS sample. Overall, about 56% of the isolates (808/1,454) contained these two *nga* operon markers. NADase 330G was found in 326 isolates unlinked to Pnga3, with 320 isolates containing the inactive NADase (G330D substitution) and also lacking the more active Pnga3 promoter (Table 1).

***emm89* emergence.** Of the 185 *emm89* isolates recovered in 2015, 178 were acapsular (*hasA* negative) and positive for the previously described clade 3 promoter Pnga3 (5). These data are consistent with recent studies that also employed ABCs *emm89* isolates (recovered from 1995 to 2013), as well as *emm89* isolates recovered in Finland, Iceland, and the United Kingdom (4, 5, 35). Recent studies correlated the acquisition of Pnga3 and the acapsular genotype with the increase in infections caused by *emm89* GAS recovered through ABCs in the mid-2000s (4, 5).

It is interesting that the only *emm89* isolates found in ABCs from 1995 to 1999 were serologically T type 11 (Fig. 5). WGS obtained from two T11/*emm89* isolates recovered in 1995 and 2000 revealed that both were single-locus variants of *emm89*/ST101 (*emm89*/ST407), were *hasA*$^+$, and contained Pnga1 (5) (Fig. 5). As recently described (36), we found that the emergent clade 3 *emm89* also acquired a *tee* gene distinct from that of clade 1 strains. The T89 genetic marker, which we find to be associated with the serological T13 type, was first detected in 2000 in ABCs *emm89* isolates (Fig. 5), associated with less active Pnga2 (5), and was also *hasA*$^+$ (Fig. 5). In our 2015 isolate set, we found that only 6 of the 185 *emm89* isolates were of the *emm89*/ST407/T11(tee11) lineage, were *hasA*$^+$, and contained Pnga1 (Fig. 5; Table 1), corresponding to previously described clade 1 (5).

***sof* gene relationships with different *emm* clusters/patterns, *emm*-like genes, and MCs.** The majority (884/1,454, 60.8%) of the isolates tested corresponded to pattern E strains on the basis of 3′ sequences of *emm* family genes at the *mga* locus (10). Nearly all of the pattern E strains contained *emm* genes of different E clusters according to the recently described clustering scheme (12). Only one pattern E *emm15* isolate (*emm* cluster E3) was *sof* negative (Table 1), consistent with previous *emm15* associations (13). There were 39 *emm* types corresponding to E clusters or patterns. Nearly all of the other strains contained 1 of 18 A-C or D cluster *emm* types (equating to patterns C and D), all of which were *sof* negative or historically serum opacity factor negative. Unlike pattern D and E strains, pattern A-C strains were negative for the

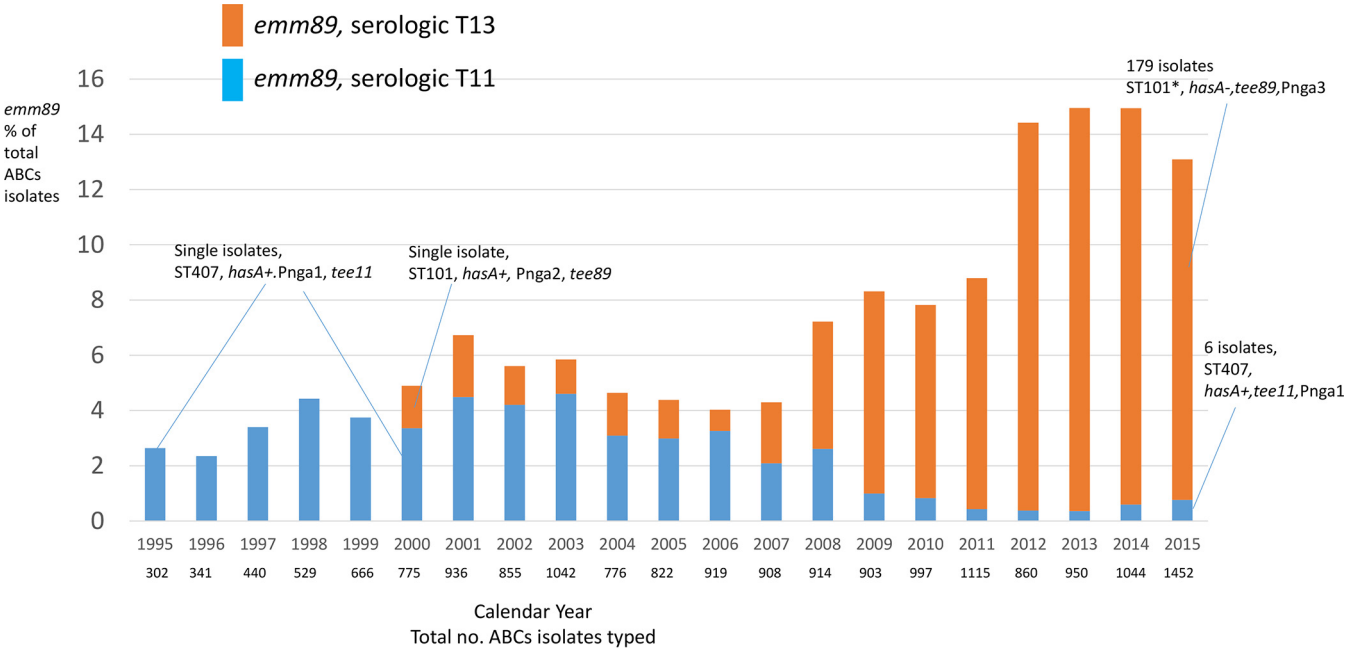

FIG 5 Emergence of *emm89* in ABCs. Notes inserted into the graph describe the numbers of isolates sequenced and relevant pipeline results. The data are consistent with previous observations (4, 5) associating the emergence of *emm89* with the appearance of the acapsular clade 3 strain containing the upregulated *nga* operon promoter Pnga3 and a different *tee* gene (36). According to our data, the clade 3 strain is T serotype 13 (corresponding to the *tee89* gene) and associated with the decline of the T serotype 11 (corresponding to the *tee11* gene) clade 1 *emm89* strain (5) that expressed capsule and contained the less active *nga* promoter Pnga1. Although only two T11 (*tee11*) isolates (recovered in 1995 and 2000) and one T13 (*tee89*) isolate (recovered in 2000) were sequenced in this study prior to 2015, where all 185 isolates were sequenced, our data are entirely consistent with the previous reports and publicly available WGS data that include a large number (870) of these isolates recovered from 1995 to 2013 (4, 5, 36). Compare Fig. 1 in reference 36, which describes the international emergence of clade 3 *emm89* to the data presented here.

emm-like genes *mrp* and *enn*, with the sole exceptions of the two type *emm18* isolates (Table 1).

Serum opacity factor is an important hypervariable virulence factor expressed by a large percentage of GAS strains (37). Positivity for the *sof* gene fragment query was predictable by previously established associations of *sof* genes with *emm* types or of previously established associations of the opacity factor phenotype with M serotypes and/or *emm* types (13, 14). The single exception was that type *emm12* strains were positive for the *sof* query; however, *emm12* strains are invariably opacity factor negative according to decades of published data (13). The *emm12*/ST36 lineage contains a conserved frameshift mutation in *sof12* that prematurely truncates the protein (found in all 20 randomly selected *emm12*/ST36 strains in this study and the CDC M12 reference strain isolated >60 years ago). Otherwise, the presence or absence of *sof* completely conformed to previously observed associations (13, 14).

A loose definition of MLST groups allowing any ST related by four or more alleles to any other in the group divided the 1,442 *S. pyogenes* isolates into 15 groups of 2 to 24 STs (10 to 346 isolates each) and 28 "singleton" STs (1 to 35 isolates each) not related to other STs by four or more alleles (Fig. 2). Also indicated are the 12 ST128 (*S. dysgalactiae* MLST scheme) *S. dysgalactiae* subsp. *equisimilis* strains included in the 2015 ABCs. It is striking that 14 of the 15 groups consisted solely of either *sof*-positive or *sof*-negative strains, with the exception of group 4, which consisted of 136 *sof*-negative isolates and 3 *sof*-positive isolates. These three *sof*-positive isolates consisted of the two unusual *emm* switching *emm82*/ST36 strains (described above and in Fig. 4) and the single *emm113*/ST148 isolate. The single ST148 isolate recorded at https://pubmlst.org/spyogenes/ was an *emm113* strain recovered in New Zealand in 1997.

Contrasting seasonality of infections shown by *sof*-negative and *sof*-positive strains. Recently, it was observed that infections due to *emm* AC cluster (or pattern A-C) strains peaked in the winter (first quarter, from January to March), while E cluster

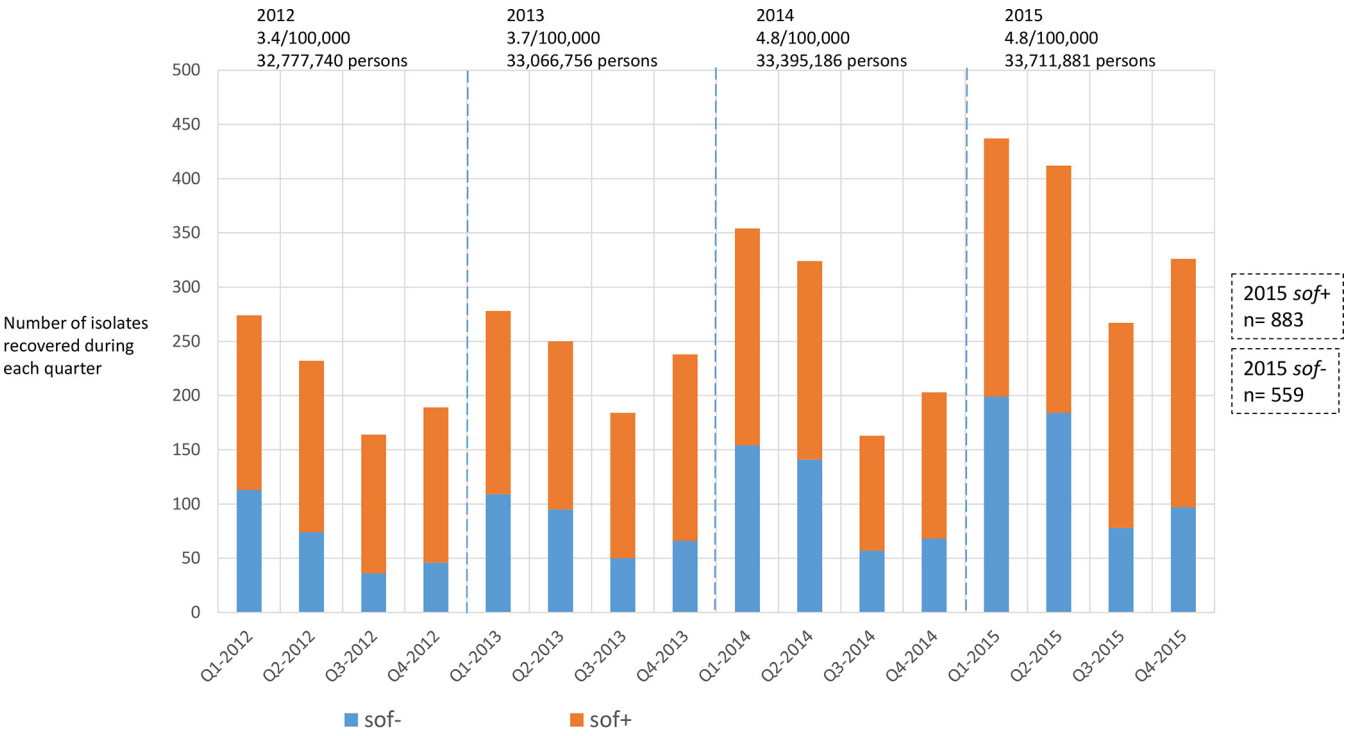

FIG 6 Different seasonalities of infections with *sof*-positive and *sof*-negative *S. pyogenes* strains. For the 1,442 isolates obtained in 2015, the data are based on the presence or absence of the *sof* gene. As shown, there is a disproportionate number of *sof*-positive isolates versus *sof*-negative isolates that is most evident in quarters 3. In 2012 to 2014, the presence or absence of *sof* was based solely on historic associations with the *emm* type or M serotype (13, 14). Because of differences in isolate collection from 2012 to 2015, the numbers of isolates shown (solid line) do not completely correlate with differences in disease incidence between different years. Rates per 100,000 population and surveillance populations are shown for each year.

(or pattern E) strain cases were disproportionally represented in the summer (third quarter, from July to September) (15). Figure 6 shows that the seasonal relationship of *emm* clustering (or *emm* locus patterns) is reflected by the presence or absence of an active *sof* determinant. Among the *S. pyogenes* isolates studied, *sof*-negative isolates accounted for only 38.8% (559/1442) of the total yet accounted for 45.5% (199/437) of the *S. pyogenes* cases in quarter 1 ($P < 0.0005$). In quarter 3, *sof*-negative isolates accounted for only 29.2% (78/267) of the cases ($P < 0.0005$). This marked fluctuation of *sof*-negative iGAS incidence between quarters 1 and 3 contrasts with the relatively stable incidence of *sof*-positive iGAS in these periods (Fig. 6).

As with *emm* clusters and *emm* patterns, the presence or absence of *sof* is nearly always predicted by identification of the *emm* type (13, 14). We used *emm* typing to determine that *emm* type-based predictions of *sof* presence/absence resulted in the same seasonality pattern in 2012 to 2014 that was seen in 2015 on the basis of the actual presence of an intact *sof* gene (Fig. 6).

Invasive group A *S. dysgalactiae* subsp. *equisimilis*. The ABCs program is based on the identification of iGAS isolates without identification to the species level. Almost all of the isolates (1,442/1,454; 99.2%), including 12 *S. dysgalactiae* subsp. *equisimilis* isolates, reported to ABCs in 2015 were *gacI* positive, which is predictive of group A carbohydrate production (28). A single *gacI*-negative isolate of *emm* type *stG643.0* was subsequently found to be serogroup G *S. dysgalactiae* subsp. *equisimilis* and was removed from the study. Twelve *S. pyogenes* isolates (0.8%) were negative for the *gacI* query; however, these were found to be serogroup A.

All 12 group A *S. dysgalactiae* subsp. *equisimilis* isolates identified through phylogenetic analysis (38) were of ST128 according to the *S. dysgalactiae* MLST scheme at https://pubmlst.org/sdysgalactiae/. The recovery of these 12 ST128 isolates of three different *emm* types from four different states suggests that this is a long-standing group A lineage of this species. The single group G 2015 ABCs isolate of this species was

found to be ST48 (*S. dysgalactiae* MLST scheme). Analysis of the *gac* (group A carbohydrate) operon from these strains revealed a hybrid structure with an upstream crossover point in *gacE* and a downstream crossover point in the second open reading frame immediately downstream of the *gacA-gacL* operon (data not shown). This approximately 11,500-bp recombinational fragment apparently originating from *S. pyogenes* corresponds to coordinates 609389 to 620916 of the *S. pyogenes* sequence with GenBank accession number CP000017. This fragment encompasses the *gacI*, *gacJ*, and *gacK* genes, which shared 99.4 to 99.7% sequence identity with counterparts in *S. pyogenes*. These three genes were recently shown to be essential for expression of the immunodominant *N*-acetylglucosamine side chain of the Lancefield group A carbohydrate (28).

DISCUSSION

While *emm* typing and antimicrobial resistance phenotyping have served as the basis of ABCs iGAS strain surveillance for the past 2 decades (2, 43, 44), the addition of WGS-based strain characterization to this population-based surveillance system encompassing nearly 34 million individuals provides much more insight into underlying strain features and strain emergence. We found in invasive GBS that PBP2x typing was actually more reliable and sensitive for detecting first-step mutations leading to β-lactam nonsusceptibility (17), and having this system in place for iGAS allows us greater vigilance for this potential threat. We now see that *ermT*, discovered in GAS only in the last decade (19), actually accounts for the major percentage of emerging GAS resistance to macrolides and lincosamides. Through our current WGS pipeline data, we have several additional parameters to evaluate in association with disease manifestations, virulence, and as vaccine components. Nearly all (~99%) of the study isolates would be covered by a combination M-Mrp vaccine (7–9), with more than half of this isolate set putatively targeted by both vaccine components. This is an important observation, since recent work indicated that the combination vaccine would provide more effective opsonization than either vaccine alone (9). We were able to quantitate MLST-defined diversity and to determine the extent of *emm* type switching in the same manner that pneumococcal strains have been assessed for capsular serotype switching in the past 15 to 20 years. From the results shown here, it appears that *emm* type switching is rare and might not be a significant immune escape mechanism should an M protein-based vaccine be implemented. In the entire sample set, we detected only one example of a past switching event, represented in two isolates, where the *emm12* gene in the ST36 genetic background was replaced with the *emm82* gene. In addition, we detected only one *emm*-negative iGAS isolate (also in the ST36 background), consistent with the M protein's historical role as an essential virulence factor. Nonetheless, the detection of these three unusual invasive isolates does present the possibility that such variant strains could emerge as successful pathogens in the presence of selection exerted by an M protein-based vaccine.

Increased documentation of GAS strain parameters may hasten the understanding of features that affect pathogenic potential. In particular, the association of the three major iGAS lineages (*emm1*/ST28, *emm89*/ST101, *emm12*/ST36) with an upregulated *nga* operon is compelling, especially when this feature directly correlated with the marked emergence of *emm89* in ABCs isolates over the past decade (4, 5, 36). Individual strain parameters may provide greater understanding of GAS tissue tropism and disease manifestations. For example, *emm28* has been shown to be significantly associated with postpartum iGAS infections (39, 40). Vaginal tissue tropism could be influenced by expression of the R28 determinant, detected primarily in the *emm28* isolates in this study. This possibility is further suggested by the existence of a close R28 homolog in group B streptococci that commonly colonize the vaginal epithelium (30). The acquisition of the *tee89* gene in emergent clade 3 *emm89* may have conferred new functional adherence or immune evasion properties (3, 36, 41). The recent increased superantigen complement in *emm1* subclones described in China is reason for increased awareness of the enhanced virulence potential imposed by already impactful iGAS strains, wherein *emm12* strains facilitated the horizontal transfer of scarlet fever-

associated mobile elements carrying *speC* and *ssa* to the *emm1*/ST28 lineage (33). In our strain set, we observed a single *emm1* isolate (20156011) that was positive for *speC* and *ssa* in addition to the usual *emm1*/ST28 superantigen complement (*speA*, *speG*, *speJ*, *speZ*) that was situated on a prophage highly related to previously described ϕHKU488.vir (33).

It is very interesting that in the two recombinant (*emm* type switching) *emm82*/ST36 strains, the normally inactive *sof12* gene reverted to an active allele upon the insertion of a single nucleotide. This observation is compatible with previously established *emm* type associations with *sof* (13, 14). It is plausible that some biologically defined barrier prevents the presence of an active *sof* gene in association with cluster A-C and D *emm* types. The reverse association also seems to be indicated, in that the combination of an active *sof* gene and most cluster E *emm* types might be essential for strain success. The association of the *emm* type with the presence or absence of the multifunctional *sof* virulence gene (37) appears to have an underlying clonal basis, since MLST divides isolate sets into defined *sof*-negative and *sof*-positive groups. The observed differences in seasonality between *sof*-negative and *sof*-positive strains could be based on the presence or absence of *sof* or could be based on other, unknown, clonal features.

The appearance of a specific group A lineage of the diverse subspecies *S. dysgalactiae* subsp. *equisimilis* widely spread among different ABCs sites is indirectly indicative of the considerable disease burden attributable to this subspecies (42), which is almost always associated with group C or G carbohydrates (38). The acquisition of the ability to express the group A antigen, itself a virulence factor (28), is reason for continued close monitoring of this iGAS subspecies. We provide evidence here that the ST128 iGAS lineage arose through a single interspecies gene replacement event. The association of group A *S. dysgalactiae* subsp. *equisimilis* ST128 with three distinct *emm* types is indicative of a successful longstanding lineage. We have previously shown that the association of multiple *emm* types in a single ST is not unusual in *S. dysgalactiae* subsp. *equisimilis* (38), although our data indicate that it is extremely rare in *S. pyogenes*.

An important aspect of WGS-based strain surveillance in ABCs is the ability to deduce close temporal and geographic relatedness between GAS isolates. The predominance of *emm59* in New Mexico in 2015 shows the potential of the use of WGS to elucidate disease transmission patterns and therefore to potentially guide efforts to control disease. We are working toward faster identification of such clusters in ABCs and trying to identify potential outbreaks for which public health intervention may be effective.

To summarize, through WGS, we have examined several aspects of iGAS strains that we were previously unable to explore in a systematic population-based manner. We provide our basic WGS genetic data in association with the genomic accession data from a full year (2015) of ABCs isolates, along with lab identifiers, in Table S1. These isolates and some accompanying epidemiological data can be acquired at https://www.cdc.gov/abcs/pathogens/isolatebank/overview.html for further investigation.

MATERIALS AND METHODS

Isolates. ABCs conducts active laboratory and population-based surveillance for iGAS infections (including necrotizing fasciitis, streptococcal toxic shock syndrome, and other infections associated with GAS isolated from a normally sterile site) in geographic areas of 10 states, representing 33.7 million persons. The 1,454 available isolates, representing 89.6% of the cases that occurred in 2015, were subjected to WGS and antimicrobial susceptibility testing. Key features of ABCs iGAS surveillance from 1997 to 2015 have been described previously (2, 43, 44; https://www.cdc.gov/abcs/reports-findings/surv-reports.html).

Whole-genome sequencing. GAS chromosomal DNA preparation, library construction, and WGS generation for the 1,454 isolates were performed as previously described (16).

Conventional MIC determinations. Isolates were subjected to broth dilution testing (BDT) for determination of MICs with the panel previously described for GBS that included a well containing both erythromycin and clindamycin to detect inducible clindamycin resistance (17). Discordant results where WGS-based predictions differed from BDT results by ≥2 dilutions (≥4-fold MIC differences) were retested by E test as described by the manufacturer (BioMérieux) or by D test (20).

Serum opacity factor determination. Serum opacity factor determination was performed with bacterial supernatants from specific isolates as previously described (45).

WGS GAS typing pipeline. Bioinformatics methods are described and updated at https://github.com/BenJamesMetcalf. *emm* subtypes were obtained on the basis of a database of defined 180-bp sequences maintained at the CDC (ftp://ftp.cdc.gov/pub/infectious_diseases/biotech/tsemm/). This subtyping scheme is based on a sequence that consists of 10 codons corresponding to the C-terminal end of the M protein signal sequence and 50 codons corresponding to the N terminus of the mature M protein (46). The WGS *emm* typing scheme employs *de novo* assembly and queries sequences closely linked to 21-bp *emm* typing primer 1 (27) situated adjacent to the *emm* type-specific region.

A PBP2x transpeptidase amino acid sequence type was generated for each isolate as described for GBS PBP2x for detection of first-step mutations leading to β-lactam resistance (17). Additionally, the ARG-ANNOT and ResFinder databases were incorporated (23, 24). Sequence targets for detection of the presence/absence of 21 T antigen backbone (*tee*) genes (29), the *gacI* glycosyl transferase specific for the group A antigen (28), the hyaluronic acid synthetic locus *hasA* (47), *emm*-like genes that flank *emm* (9), four different fibronectin-binding domain repeat proteins (48), the R28 surface antigen (30), the *sda1*-encoded DNase (49), sequence polymorphisms associated with the *ngo* operon (4, 5), two conserved *rocA* null mutations (50, 51), 12 exotoxin genes (*speA* to *speC*, *speG* to *speM*, *ssa*, *smeZ*) (52), and the streptococcal inhibitor of complement (31, 32) were obtained through the references indicated.

MLST. MLST relied upon SRST2 and the database at http://pubmlst.org/spyogenes/.

MLST-based CCs and groups. MCs were defined as isolates sharing at least five alleles with the reference ST, which represented the major ST found in an *emm* type. An eBurst (53) group was defined as an ST set where each member shared at least four alleles with one or more other members of the set.

Phylogenetic analysis. kSNP3.0 analysis was performed as previously described (54).

Statistical analyses. A chi-square test was performed to evaluate differences in seasonality between *sof*-negative and *sof*-positive groups.

Accession number(s). Accession numbers for the 1,454 fastq files used in this work are provided in Table S1, along with lab identifiers, WGS-generated genetic data, and quality metrics.

SUPPLEMENTAL MATERIAL

Supplemental material for this article may be found at https://doi.org/10.1128/mBio.01422-17.

TABLE S1, PDF file, 0.7 MB.

ACKNOWLEDGMENTS

We are indebted to all of the hospitals and laboratories participating in the Active Bacterial Core Surveillance component of the Emerging Infections Programs network, a collaboration of the CDC, state health departments, and universities. We are grateful to the Minnesota Department of Public Health laboratory for pneumococcal serotyping and susceptibility testing of all of the isolates recovered in Minnesota. We thank the following individuals from the following programs and institutions for their contributions to the establishment and maintenance of the ABCs system: California Emerging Infections Program, A. Reingold, S. Brooks, and H. Randel; Colorado Emerging Infections Program, L. Miller, B. White, D. Aragon, M. Barnes, and J. Sadlowski; Connecticut Emerging Infections Program, S. Petit, M. Cartter, C. Marquez, and M. Wilson; Georgia Emerging Infections Program, M. Farley, S. Thomas, A. Tunali, and W. Baughman; Maryland Emerging Infections Program, L. Harrison, J. Benton, T. Carter, R. Hollick, K. Holmes, and A. Riner; Minnesota Emerging Infections Program, A. Glennon, C. Holtzman, K. Como-Sabetti, R. Danila, and K. MacInnes; New Mexico Emerging Infections Program, K. Scherzinger, K. Angeles, J. Bareta, L. Butler, S. Khanlian, R. Mansmann, and M. Nichols; New York Emerging Infections Program, N. Bennett, S. Zansky, S. Currenti, and S. McGuire; Oregon Emerging Infections Program, A. Thomas, M. Schmidt, J. Thompson, and T. Poissant; Tennessee Emerging Infections Program, W. Schaffner, B. Barnes, K. Leib, K. Dyer, and L. McKnight; CDC, R. Gierke, K.-A. Toews, E. Weston, L. McGlone, and G. Langley.

This study used the *S. pyogenes* MLST website (http://pubmlst.org/spyogenes/) at the University of Oxford (K. A. Jolley and M. C. J. Maiden, BMC Bioinformatics 11:595, 2010, https://doi.org/10.1186/1471-2105-11-595). The development of this site has been funded by the Wellcome Trust.

Major funding for this work was provided through support from the CDC Advanced Molecular Detection (AMD) initiative and the CDC Emerging Infection Program.

The findings and conclusions in this report are those of the authors and do not necessarily represent the official position of the Centers for Disease Control and Prevention.

REFERENCES

1. Carapetis JR, Steer AC, Mulholland EK, Weber M. 2005. The global burden of group A streptococcal diseases. Lancet Infect Dis 5:685–694. https://doi.org/10.1016/S1473-3099(05)70267-X.
2. Nelson GE, Pondo T, Toews KA, Farley MM, Lindegren ML, Lynfield R, Aragon D, Zansky SM, Watt JP, Cieslak PR, Angeles K, Harrison LH, Petit S, Beall B, Van Beneden CA. 2016. Epidemiology of invasive group A streptococcal infections in the United States, 2005–2012. Clin Infect Dis 63:478–486. https://doi.org/10.1093/cid/ciw248.
3. Walker MJ, Barnett TC, McArthur JD, Cole JN, Gillen CM, Henningham A, Sriprakash KS, Sanderson-Smith ML, Nizet V. 2014. Disease manifestations and pathogenic mechanisms of group A *Streptococcus*. Clin Microbiol Rev 27:264–301. https://doi.org/10.1128/CMR.00101-13.
4. 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.
5. Zhu L, Olsen RJ, Nasser W, de la Riva Morales IR, 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.
6. Steer AC, Carapetis JR, Dale JB, Fraser JD, Good MF, Guilherme L, Moreland NJ, Mulholland EK, Schodel F, Smeesters PR. 2016. Status of research and development of vaccines for *Streptococcus pyogenes*. Vaccine 34:2953–2958. https://doi.org/10.1016/j.vaccine.2016.03.073.
7. Dale JB, Penfound TA, Chiang EY, Walton WJ. 2011. New 30-valent M protein-based vaccine evokes cross-opsonic antibodies against non-vaccine serotypes of group A streptococci. Vaccine 29:8175–8178. https://doi.org/10.1016/j.vaccine.2011.09.005.
8. Dale JB, Penfound TA, Tamboura B, Sow SO, Nataro JP, Tapia M, Kotloff KL. 2013. Potential coverage of a multivalent M protein-based group A streptococcal vaccine. Vaccine 31:1576–1581. https://doi.org/10.1016/j.vaccine.2013.01.019.
9. Courtney HS, Niedermeyer SE, Penfound TA, Hohn CM, Greeley A, Dale JB. 2017. Trivalent M-related protein as a component of next generation group A streptococcal vaccines. Clin Exp Vaccine Res 6:45–49. https://doi.org/10.7774/cevr.2017.6.1.45.
10. Bessen DE, Sotir CM, Readdy TL, Hollingshead SK. 1996. Genetic correlates of throat and skin isolates of group A streptococci. J Infect Dis 173:896–900. https://doi.org/10.1093/infdis/173.4.896.
11. Facklam RF, Martin DR, Lovgren M, Johnson DR, Efstratiou A, Thompson TA, Gowan S, Kriz P, Tyrrell GJ, Kaplan E, Beall B. 2002. Extension of the Lancefield classification for group A streptococci by addition of 22 new M protein gene sequence types from clinical isolates: *emm103* to *emm124*. Clin Infect Dis 34:28–38. https://doi.org/10.1086/324621.
12. Sanderson-Smith M, De Oliveira DM, Guglielmini J, McMillan DJ, Vu T, Holien JK, Henningham A, Steer AC, Bessen DE, Dale JB, Curtis N, Beall BW, Walker MJ, Parker MW, Carapetis JR, Van Melderen L, Sriprakash KS, Smeesters PR, M Protein Study Group. 2014. A systematic and functional classification of *Streptococcus pyogenes* that serves as a new tool for molecular typing and vaccine development. J Infect Dis 210:1325–1338. https://doi.org/10.1093/infdis/jiu260.
13. Johnson DR, Kaplan EL, VanGheem A, Facklam RR, Beall B. 2006. Characterization of group A streptococci (*Streptococcus pyogenes*): correlation of M-protein and *emm*-gene type with T-protein agglutination pattern and serum opacity factor. J Med Microbiol 55:157–164. https://doi.org/10.1099/jmm.0.46224-0.
14. Beall B, Gherardi G, Lovgren M, Facklam RR, Forwick BA, Tyrrell GJ. 2000. *emm* and *sof* gene sequence variation in relation to serological typing of opacity-factor-positive group A streptococci. Microbiology 146:1195–1209. https://doi.org/10.1099/00221287-146-5-1195.
15. Smeesters PR, Laho D, Beall B, Steer AC, Van Beneden CA. 2017. Seasonal, geographic, and temporal trends of *emm* clusters associated with invasive group A streptococcal infections in US multistate surveillance. Clin Infect Dis 64:694–695. https://doi.org/10.1093/cid/ciw807.
16. Metcalf BJ, Chochua S, Gertz RE, Jr, Li Z, Walker H, Tran T, Hawkins PA, Glennen A, Lynfield R, Li Y, McGee L, Beall B, Active Bacterial Core Surveillance Team. 2016. Using whole genome sequencing to identify resistance determinants and predict antimicrobial resistance phenotypes for year 2015 invasive pneumococcal disease isolates recovered in the United States. Clin Microbiol Infect 22:1002.e1–1002.e8. https://doi.org/10.1016/j.cmi.2016.08.001.
17. Metcalf BJ, Chochua S, Gertz RE, Jr, Hawkins PA, Ricaldi J, Li Z, Walker H, Tran T, Rivers J, Mathis S, Jackson D, Glennen A, Lynfield R, McGee L, Beall B. 2017. Short-read whole genome sequencing for determination of antimicrobial resistance mechanisms and capsular serotypes of current invasive *Streptococcus agalactiae* recovered in the USA. Clin Microbiol Infect 23:574.e7–574.e14. https://doi.org/10.1016/j.cmi.2017.02.021.
18. Li Y, Metcalf BJ, Chochua S, Li Z, Gertz RE, Jr, Walker H, Hawkins PA, Tran T, Whitney CG, McGee L, Beall BW. 2016. Penicillin-binding protein transpeptidase signatures for tracking and predicting β-lactam resistance levels in *Streptococcus pneumoniae*. mBio 7:e00756-16. https://doi.org/10.1128/mBio.00756-16.
19. Woodbury RL, Klammer KA, Xiong Y, Bailiff T, Glennen A, Bartkus JM, Lynfield R, Van Beneden C, Beall BW, Active Bacterial Core Surveillance Team. 2008. Plasmid-borne *erm*(T) from invasive, macrolide-resistant *Streptococcus pyogenes* strains. Antimicrob Agents Chemother 52:1140–1143. https://doi.org/10.1128/AAC.01352-07.
20. Clinical and Laboratory Standards Institute (CLSI). 2014. Performance standards for antimicrobial susceptibility testing; twenty-third informational supplement. CLSI document M100-S22. CLSI, Wayne, PA.
21. Herrera L, Salcedo C, Orden B, Herranz B, Martinez R, Efstratiou A, Sáez Nieto JA. 2002. Rifampin resistance in *Streptococcus pyogenes*. Eur J Clin Microbiol Infect Dis 21:411–413. https://doi.org/10.1007/s10096-002-0732-2.
22. Aubry-Damon H, Galimand M, Gerbaud G, Courvalin P. 2002. *rpoB* mutation conferring rifampin resistance in *Streptococcus pyogenes*. Antimicrob Agents Chemother 46:1571–1573. https://doi.org/10.1128/AAC.46.5.1571-1573.2002.
23. Gupta SK, Padmanabhan BR, Diene SM, Lopez-Rojas R, Kempf M, Landraud L, Rolain JM. 2014. ARG-Annot, a new bioinformatic tool to discover antibiotic resistance genes in bacterial genomes. Antimicrob Agents Chemother 58:212–220. https://doi.org/10.1128/AAC.01310-13.
24. Zankari E, Hasman H, Cosentino S, Vestergaard M, Rasmussen S, Lund O, Aarestrup FM, Larsen MV. 2012. Identification of acquired antimicrobial resistance genes. J Antimicrob Chemother 67:2640–2644. https://doi.org/10.1093/jac/dks261.
25. Engelthaler DM, Valentine M, Bowers J, Pistole J, Driebe EM, Terriquez J, Nienstadt L, Carroll M, Schumacher M, Ormsby ME, Brady S, Livar E, Yazzie D, Waddell V, Peoples M, Komatsu K, Keim P. 2016. Hypervirulent *emm59* clone in invasive group A *Streptococcus* outbreak, southwestern United States. Emerg Infect Dis 22:734–738. https://doi.org/10.3201/eid2204.151582.
26. Jeng A, Sakota V, Li Z, Datta V, Beall B, Nizet V. 2003. Molecular genetic analysis of a group A *Streptococcus* operon encoding serum opacity factor and a novel fibronectin-binding protein, SfbX. J Bacteriol 185:1208–1217. https://doi.org/10.1128/JB.185.4.1208-1217.2003.
27. Whatmore AM, Kapur V, Sullivan DJ, Musser JM, Kehoe MA. 1994. Non-congruent relationships between variation in *emm* gene sequences and the population structure of group A streptococci. Mol Microbiol 14:619–631. https://doi.org/10.1111/j.1365-2958.1994.tb01301.x.
28. van Sorge NM, Cole JN, Kuipers K, Henningham A, Aziz RK, Kasirer-Friede A, Lin L, Berends ETM, Davies MR, Dougan G, Zhang F, Dahesh S, Shaw L, Gin J, Cunningham M, Merriman JA, Hütter J, Lepenies B, Rooijakkers SHM, Malley R, Walker MJ, Shattil SJ, Schlievert PM, Choudhury B, Nizet V. 2014. The classical Lancefield antigen of group A *Streptococcus* is a virulence determinant with implications for vaccine design. Cell Host Microbe 15:729–740. https://doi.org/10.1016/j.chom.2014.05.009.
29. Falugi F, Zingaretti C, Pinto V, Mariani M, Amodeo L, Manetti AG, Capo S, Musser JM, Orefici G, Margarit I, Telford JL, Grandi G, Mora M. 2008. Sequence variation in group A *Streptococcus* pili and association of pilus backbone types with Lancefield T serotypes. J Infect Dis 198:1834–1841. https://doi.org/10.1086/593176.
30. Stålhammar-Carlemalm M, Areschoug T, Larsson C, Lindahl G. 1999. The R28 protein of *Streptococcus pyogenes* is related to several group B streptococcal surface proteins, confers protective immunity and promotes binding to human epithelial cells. Mol Microbiol 33:208–219. https://doi.org/10.1046/j.1365-2958.1999.01470.x.
31. Akesson P, Sjöholm AG, Björck L. 1996. Protein SIC, a novel extracellular protein of *Streptococcus pyogenes* interfering with complement function. J Biol Chem 271:1081–1088. https://doi.org/10.1074/jbc.271.2.1081.
32. Hartas J, Sriprakash KS. 1999. *Streptococcus pyogenes* strains containing *emm12* and *emm55* possess a novel gene coding for distantly related SIC

protein. Microb Pathog 26:25–33. https://doi.org/10.1006/mpat.1998.0244.
33. Ben Zakour NL, Davies MR, You Y, Chen JH, Forde BM, Stanton-Cook M, Yang R, Cui Y, Barnett TC, Venturini C, Ong CL, Tse H, Dougan G, Zhang J, Yuen KY, Beatson SA, Walker MJ. 2015. Transfer of scarlet fever-associated elements into the group A *Streptococcus* M1T1 clone. Sci Rep 5:15877. https://doi.org/10.1038/srep15877.
34. Reference deleted.
35. Turner CE, Abbott J, Lamagni T, Holden MT, David S, Jones MD, Game L, Efstratiou 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. https://doi.org/10.1128/mBio.00622-15.
36. Beres SB, Kachroo P, Nasser W, Olsen RJ, Zhu L, Flores AR, de la Riva I, Paez-Mayorga J, Jimenez FE, Cantu C, Vuopio J, Jalava J, Kristinsson KG, Gottfredsson M, Corander J, Fittipaldi N, Di Luca MC, Petrelli D, Vitali LA, Raiford A, Jenkins L, Musser JM. 2016. Transcriptome remodeling contributes to epidemic disease caused by the human pathogen *Streptococcus pyogenes*. mBio 7:e00403-16. https://doi.org/10.1128/mBio.00403-16.
37. Courtney HS, Pownall HJ. 2010. The structure and function of serum opacity factor: a unique streptococcal virulence determinant that targets high-density lipoproteins. J Biomed Biotechnol 2010:956071. https://doi.org/10.1155/2010/956071.
38. Ahmad Y, Gertz RE, Jr, Li Z, Sakota V, Broyles LN, Van Beneden C, Facklam R, Shewmaker PL, Reingold A, Farley MM, Beall BW. 2009. Genetic relationships deduced from *emm* and multilocus sequence typing of invasive *Streptococcus dysgalactiae* subsp. *equisimilis* and *S. canis* recovered from isolates collected in the United States. J Clin Microbiol 47:2046–2054. https://doi.org/10.1128/JCM.00246-09.
39. Colman G, Tanna A, Efstratiou A, Gaworzewska ET. 1993. The serotypes of *Streptococcus pyogenes* present in Britain during 1980–1990 and their association with disease. J Med Microbiol 39:165–178. https://doi.org/10.1099/00222615-39-3-165.
40. Chuang I, Van Beneden C, Beall B, Schuchat A. 2002. Population-based surveillance for postpartum invasive group A *Streptococcus* infections, 1995–2000. Clin Infect Dis 35:665–670. https://doi.org/10.1086/342062.
41. Tsai JC, Loh JM, Clow F, Lorenz N, Proft T. 2017. The group A *Streptococcus* serotype M2 pilus plays a role in host cell adhesion and immune evasion. Mol Microbiol 103:282–298. https://doi.org/10.1111/mmi.13556.
42. Broyles LN, Van Beneden C, Beall B, Facklam R, Shewmaker PL, Malpiedi P, Daily P, Reingold A, Farley MM. 2009. Population-based study of invasive disease due to beta-hemolytic streptococci of groups other than A and B. Clin Infect Dis 48:706–712. https://doi.org/10.1086/597035.
43. O'Brien KL, Beall B, Barrett NL, Cieslak PR, Reingold A, Farley MM, Danila R, Zell ER, Facklam R, Schwartz B, Schuchat A. 2002. Epidemiology of invasive group A *Streptococcus* disease in the United States, 1995–1999. Clin Infect Dis 35:268–276. https://doi.org/10.1086/341409.
44. O'Loughlin RE, Roberson A, Cieslak PR, Lynfield R, Gershman K, Craig A, Albanese BA, Farley MM, Barrett NL, Spina NL, Beall B, Harrison LH, Reingold A, Van Beneden C, Active Bacterial Core Surveillance team. 2007. The epidemiology of invasive group A streptococcal infections and potential vaccine implications, United States, 2000–2004. Clin Infect Dis 45:853–862. https://doi.org/10.1086/521264.
45. Maxted WR, Widdowson JP, Fraser CAM, Ball LC, Bassett DCJ. 1973. The use of the serum opacity reaction in the typing of group-A streptococci. J Med Microbiol 6:83–90. https://doi.org/10.1099/00222615-6-1-83.
46. Li Z, Sakota V, Jackson D, Franklin AR, Beall B, Active Bacterial Core Surveillance/Emerging Infections Program Network. 2003. Array of M protein gene subtypes in 1064 recent invasive group A *Streptococcus* isolates recovered from the Active Bacterial Core Surveillance. J Infect Dis 188:1587–1592. https://doi.org/10.1086/379050.
47. Ashbaugh CD, Albertí S, Wessels MR. 1998. Molecular analysis of the capsule gene region of group A *Streptococcus*: the *hasAB* genes are sufficient for capsule expression. J Bacteriol 180:4955–4959.
48. Yamaguchi M, Terao Y, Kawabata S. 2013. Pleiotropic virulence factor—*Streptococcus pyogenes* fibronectin-binding proteins. Cell Microbiol 15:503–511. https://doi.org/10.1111/cmi.12083.
49. Aziz RK, Ismail SA, Park HW, Kotb M. 2004. Post-proteomic identification of a novel phage-encoded streptodornase, Sda1, in invasive M1T1 *Streptococcus pyogenes*. Mol Microbiol 54:184–197. https://doi.org/10.1111/j.1365-2958.2004.04255.x.
50. Lynskey NN, Goulding D, Gierula M, Turner CE, Dougan G, Edwards RJ, Sriskandan S. 2013. RocA truncation underpins hyper-encapsulation, carriage longevity and transmissibility of serotype M18 group A streptococci. PLoS Pathog 9:e1003842. https://doi.org/10.1371/journal.ppat.1003842.
51. Lynskey NN, Turner CE, Heng LS, Sriskandan S. 2015. A truncation in the regulator RocA underlies heightened capsule expression in serotype M3 group A streptococci. Infect Immun 83:1732–1733. https://doi.org/10.1128/IAI.02892-14.
52. Commons RJ, Smeesters PR, Proft T, Fraser JD, Robins-Browne R, Curtis N. 2014. Streptococcal superantigens: categorization and clinical associations. Trends Mol Med 20:48–62. https://doi.org/10.1016/j.molmed.2013.10.004.
53. Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG. 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J Bacteriol 186:1518–1530. https://doi.org/10.1128/JB.186.5.1518-1530.2004.
54. Gardner SN, Slezak T, Hall BG. 2015. kSNP3.0: SNP detection and phylogenetic analysis of genomes without genome alignment or reference genome. Bioinformatics 31:2877–2878. https://doi.org/10.1093/bioinformatics/btv271.
55. 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.