Integration of Genomic and Other Epidemiologic Data to Investigate and Control a Cross-Institutional Outbreak of *Streptococcus pyogenes*

Victoria J. Chalker, Alyson Smith, Ali Al-Shahib, Stella Botchway, Emily Macdonald, Roger Daniel, Sarah Phillips, Steven Platt, Michel Doumith, Rediat Tewolde, Juliana Coelho, Keith A. Jolley, Anthony Underwood, Noel D. McCarthy

Single-strain outbreaks of *Streptococcus pyogenes* infections are common and often go undetected. In 2013, two clusters of invasive group A *Streptococcus* (iGAS) infection were identified in independent but closely located care homes in Oxfordshire, United Kingdom. Investigation included visits to each home, chart review, staff survey, microbiologic sampling, and genome sequencing. *S. pyogenes emm* type 1.0, the most common circulating type nationally, was identified from all cases yielding GAS isolates. A tailored whole-genome reference population comprising epidemiologically relevant contemporaneous isolates and published isolates was assembled. Data were analyzed independently using whole-genome multilocus sequencing and single-nucleotide polymorphism analyses. Six isolates from staff and residents of the homes formed a single cluster that was separated from the reference population by both analytical approaches. No further cases occurred after mass chemoprophylaxis and enhanced infection control. Our findings demonstrate the ability of 2 independent analytical approaches to enable robust conclusions from nonstandardized whole-genome analysis to support public health practice.

The reported annual incidence of invasive group A *Streptococcus* (iGAS) infection in industrialized countries is ≈3 cases per 100,000 persons per year (*1*–*3*). Incidence is 3-fold higher among persons ≥70 years of age, particularly among the very elderly (*1*,*2*,*4*). Older persons have also been shown to have higher case-fatality rates, including in studies considering particular clinical syndromes, suggesting that age is a risk for death independent of the clinical form of iGAS (*1*,*5*,*6*). Population-based study data estimate incidence among long-term care facility (LTCF) residents as 3.4-fold (*7*), 6.0-fold (*8*), and 7.8-fold (*9*) higher than that among elderly persons living outside institutional settings. Annual incidence estimates among LTCF residents range from 27 (*7*) to 74 (*9*) cases per 100,000 persons. Case-fatality rates are also higher among LTCF residents (*7*–*9*).

Although iGAS mainly occurs sporadically, outbreaks have been recognized in hospitals, particularly in association with surgery and maternity care (*10*,*11*), and in LTCFs, where they appear to be increasing (*12*,*13*). Many iGAS outbreaks go unidentified. Investigations that reviewed residents' medical records in 7 LTCFs in which 1 case of iGAS was reported identified missed outbreaks in 4 of the facilities (*4*). Furthermore, a surveillance study reported that LTCF staff members were unaware of hospital-diagnosed iGAS among residents other than through study feedback (*7*). Three surveillance studies, including 2 that used bacterial subtyping (*7*,*8*), identified iGAS clusters that had not been identified as outbreaks outside the surveillance scheme (*7*–*9*). The delayed identification of outbreaks in LTCFs (*14*,*15*) and clinical geriatric care settings (*16*,*17*) has also been described. Prospective surveillance in Ontario, Canada, identified 20 hospital outbreaks that were of smaller average size, often shorter duration, and more often outside the surgical and maternity settings than was expected based on findings in the nosocomial outbreak literature (*10*). Thus, sporadic disease may include many unidentified small outbreaks. In an LTCF surveillance study (*8*), 40 of 383 isolates were members of 18 clusters of indistinguishable strains, and in another study (*7*), 34 of 134 isolates were associated with 13 clusters, suggesting that 10%–25% of culture-confirmed cases in LTCFs may be associated with outbreaks.

Author affiliations: Public Health England, Colindale, UK (V.J. Chalker, A. Al-Shahib, R. Daniel, S. Phillips, S. Platt, M. Doumith, R. Tewolde, J. Coelho, A. Underwood); South East England Public Health England Centre, Chilton, UK (A. Smith, S. Botchway, E. Macdonald, N.D. McCarthy); University of Oxford Department of Zoology, Oxford, UK (K.A. Jolley, N.D. McCarthy); University of Warwick, Warwick Medical School, Coventry, UK (N.D. McCarthy)

DOI: http://dx.doi.org/10.3201/eid2206.142050

Genomic data are increasingly available to support the identification and investigation of outbreaks (*18*,*19*), including 2 iGAS outbreaks associated with maternity units (*20*,*21*). In 1 of these studies, isolates collected on 2 consecutive days at a hospital were highly similar, and they were distinguishable from 2 other isolates of the same M protein gene (*emm*) type collected at a later date from 2 other hospitals (*21*). In the second study (*20*), genome sequencing confirmed the relatedness of isolates from 2 patients on a maternity ward with fatal disease and isolates subsequently obtained from another patient, household contacts, and healthcare workers; the study also discriminated these isolates from 9 epidemiologically and geographically separated isolates of identical *emm* type. Genome sequencing may, therefore, separate *Streptococcus pyogenes* isolates with close epidemiologic relationships from the background population, as suggested from findings from some other species of bacteria (*19*).

The purpose of this study was to integrate genomic data with other epidemiologic data in the investigation and control of a cross-institutional outbreak of *S. pyogenes* infection. We assessed approaches to enable robust inferences in the absence of standard analytical methods.

Methods

We used standard epidemiologic and microbiologic approaches to investigate clusters of GAS infection in 2 managerially independent but closely located LTCFs (home A and home B) in Oxfordshire, United Kingdom, in 2013. We also applied genomic sequencing to available isolates, analyzed the data using 2 independent approaches, and performed a systematic literature review to describe evidence for the efficacy of different control strategies.

Literature Search

On February 21, 2014, we searched PubMed, using the terms: ("pyogenes" OR "group A streptococc*") AND ("care" OR "nursing" OR "residential") AND ("home" OR "homes" OR "facility" OR "facilities" OR "setting" OR "settings"). One author (N.D.M.) reviewed the 131 abstracts retrieved by this search to identify those that referred to outbreaks or LTCFs (27 papers) and to population-level surveillance (3 papers). Identified papers were reviewed, references were searched for similar papers, and information was extracted on the control methods used, outcomes, and whether reported outbreaks were each due to a single strain of *S. pyogenes* or involved multiple strains.

Epidemiologic Investigation of GAS Cases in LTCFs

We reviewed medical records for all possible cases of GAS infection among residents of the 2 LTCFs in 2013. In addition to cases already notified to public health authorities, additional possible cases of GAS were identified for review through interviews by public health staff with senior LTCF staff. Healthcare staff used case definitions from the UK national guidance (*22*) to assess residents or staff with symptoms suggestive of GAS infection. Clinically indicated samples were obtained, as were skin and soft tissue infection samples and throat swab samples from residents or staff reporting sore throats. A web-based survey of staff enabled anonymous reporting of symptoms and infection-control practices.

Microbiology

Samples from residents and staff were cultured to detect *S. pyogenes* using standard methods. The Public Health England National Streptococcal Reference Laboratory (Bacterial Reference Department) performed *emm* gene sequence typing on each isolate obtained as previously described (*23*,*24*) using DNA prepared by using the Wizard Genomic DNA Purification Kit (Promega, Madison, Wisconsin, USA); quality was determined by using a NanoDrop 2000 Spectrophotometer (Thermo Scientific, Waltham, MA, USA), and quantity was determined by using a Qubit 3.0 Fluorometer and quantitation assays (Thermo Fisher Scientific, Waltham, MA, USA). For sequencing preparation, we used a Nextera XT DNA Library Preparation Kit (Illumina, San Diego, CA, USA), and for sequencing, we used a HiSeq 2500 System (Illumina) and the 2 × 100-bp paired-end mode. As a reference dataset to represent the background population, we used published genomes (*20*) and contemporaneous isolates of the same *emm* type from 3 clusters in other areas of the United Kingdom (Table).

Bioinformatic Processing

We used 2 independent approaches to process 39 genome sequences, of which 6 were for isolates from the LTCFs, 9 were for isolates from contemporaneous confirmed cases, and 24 were published genomes (*20*). We used Burrows-Wheeler Aligner software (http://bio-bwa.sourceforge.net/) (*25*) to map reads to *emm* sequence type 1.0 (*emm*1) of *S. pyogenes* MGAS5505 (GenBank accession no. NC_007297-2). Single-nucleotide polymorphisms (SNPs) were discovered by using GATK2 software (Genome Analysis Toolkit v2) (*26*) and filtered by using the following parameters: genotype quality ≥40, ratio of consensus/nonconsensus base ≥0.8, distance to nearest SNP >15, root mean square of the mapping quality of the reads ≥50, number of reads with 0 mapping quality = 0. Reads were independently assembled by using the Velvet algorithm package (*27*), and loci were annotated with the genome annotation of *S. pyogenes* MGAS5005, which resulted in identification of 1,514 non-paralogous genetic loci with sequence data, enabling whole-genome multilocus sequence type (wgMLST) analysis (*28*).

Table. Clinical and demographic characteristics of patients with isolates sequenced in a study integrating genomic and other epidemiologic data to investigate and control a cross-institutional outbreak of *Streptococcus pyogenes**

Area, laboratory no.	Healthcare setting	Age/sex	Source of isolate	Clinical presentation	Outcome
1					
H131441217	Care home	95 y/F	Blood culture	Bilateral periorbital cellulitis, sepsis	Died
H131520646	Care home	91 y/M	Blood culture	Facial cellulitis	Died, ANP
H131640460	Care home	65 y/M	Nasal swab sample	Rash, fever	Recovered
H131620455	Care home	84 y/F	Arm wound swab sample	Arm cellulitis	Recovered
H131720333	Care home	101 y/F	Ear swab sample	Weeping ear	Recovered
H132060515	Care home (CW)	19 y/F	Throat swab sample	Sore throat	Recovered
2					
H131280521	Care home	94 y/F	Blood culture	Severe soft tissue infection	Died
H131100707	Care home	84 y/F	Blood culture	Fever, leg cellulitis, diarrhea, vomiting	NR
H131220725	Care home	93 y/F	Blood culture	Fever, severe cellulitis	Died
H131620436	Care home	78 y/F	Blood culture	Emergency room admission	NR
H131020872	Maternity service	7 d/F	Umbilical wound swab sample	NR	NR
3					
H131180727	Hospital	87 y/M	Blood culture	NR	NR
H130500483	Hospital	60 y/M	Blood culture	Rash, sepsis, suspected CAP	NR
4					
H130620575	Hospital	39 y/M	Pus swab sample	NR	NR
H130620574	Hospital	71 y/M	Cannula site swab sample	NR	NR

*ANP, acute necrotizing pneumonitis diagnosed postmortem; CAP, community-acquired pneumonia; CW, care worker, NR not recorded.

Bioinformatic Analysis

We constructed a maximum-likelihood phylogenetic tree by using RAxML (*29*), a multiple FASTA file of concatenated SNPs. We searched for 68 known superantigen and antimicrobial resistance genes, which were considered present if matches were found with >95% coverage and <10% SNPs difference from known variants.

We summarized allelic relationships by using the Genome Comparator tool within BIGSdb (*30*) and used pairwise differences to estimate a neighbor-joining tree by using MEGA5 software (*31*). The distribution of pairwise allelic differences among isolates was compared in each of 3 categories: within the current suspected outbreak, within an outbreak reported by Turner et al. (*20*), and comparing the current suspected outbreak isolates with all other isolates in the assembled dataset.

Results

Literature Review

Our literature review identified 72 LTCF-associated iGAS outbreaks from individual outbreak reports, summary data identified in surveillance studies, reviews, and laboratory studies (online Technical Appendix, http://wwwnc.cdc.gov/EID/article/22/6/14-2050-Techapp1.pdf). These outbreaks included 31 clusters, which were defined as outbreaks on the basis of shared subtypes among ≥2 cases occurring at a facility within 1 year. Subtyping results were available for 22 other outbreaks that were identified by other means. A single or dominant strain was identified for 19 of these outbreaks; multiple strains were reported for the other 3 outbreaks.

Interventions were varied but encompassed 3 broad approaches and showed limited evidence of different outcomes. First, treatment restricted to patients with GAS or to patients with GAS and to their direct contacts with laboratory confirmed infection was associated with disease control in some outbreaks. However, in 1 outbreak, disease recurred after 2 rounds of this selective treatment, so mass chemoprophylaxis was administered within the LTCF. Second, in 2 reported outbreaks, all staff and residents were screened for GAS and if positive, they were offered chemoprophylaxis. This screening and treatment was associated with disease recurrence and repeated screening and treatment. Additional cases of iGAS were diagnosed between decisions to screen and provide chemoprophylaxis and to actually implement chemoprophylaxis. Third, in all identified reports, chemoprophylactic treatment of all staff and residents was associated with control of iGAS, although in 1 incident, persistent infection was shown in a resident who had a gastrostomy tube.

The literature review showed that screening detected carriage rates of <10% among LTCF residents, with 2 exceptions, for which carriage rates were 20% and 16%. As previously described, carriage rates were lower among LTCF staff than residents (*12*,*13*).

Epidemiology of Cases in Home A and Home B

After 1 iGAS case each was identified in homes A and B, advice was given to the LTCF managers by Public Health England on infection control and how to identify other GAS-compatible infections in residents and staff. Although the 2 LTCFs were geographically close to each

other, neither home could initially identify any links with the other. Mass chemoprophylaxis was initiated at home A the day after a second case was reported and at home B after GAS (not iGAS) was confirmed in another resident who had cellulitis (Figure 1). Further investigation identified 3 staff members who had worked in both of the managerially independent homes; 2 of these staff members had GAS-compatible symptoms. Of 41 staff who responded to the anonymous survey, 38 identified their roles as a healthcare assistant (18 persons), cleaner (6 persons), manager or administrator (6 persons), nurse (3 persons), or other (5 persons). One respondent reported working at both homes, and 1 reported working at home B and in the community. Of the 41 respondents, 32 reported no symptoms, 7 reported sore throats, 1 reported a rash on the neck, and 1 reported illness without specifying symptoms.

Twelve possible GAS infections were not confirmed by culture. One of the infections was on the scalp of a resident at home A; the patient, who had a temperature >39°C and hallucinations, was treated with antimicrobial drugs and tested negative on subsequent samples. Another possible infection was on the eyelid of a resident at home B. Eight possible infections were in persons who worked at home A and who reported symptoms consistent with pharyngitis; 1 of the 8 staff members also worked at home B. Another person who worked at both LTCFs reported having paronychia that was treated with antimicrobial drugs, and 1 staff member at home B reported a recurrent skin infection.

Microbiology

S. pyogenes isolates were obtained from 6 of 13 specimens from symptomatic residents and staff at homes A and B. Two patients in home A, 1 of whom died, had *S. pyogenes*–positive blood cultures and facial cellulitis. One patient in home B had periorbital cellulitis and GAS bacteremia. Two other patients in home B had confirmed GAS; 1 of these patients had cellulitis and systemic symptoms, including fever, but no sample from a normally sterile site, and the other patient had an outer ear infection. One staff member working in both institutions had GAS pharyngitis.

Isolate Relatedness, Antimicrobial Resistance, and Virulence Genes

All 6 isolates from the cross-institutional *S. pyogenes* outbreak were T-type 1 (phenotypic typing), *emm*1, and *sic* (streptococcal inhibitor of complement gene) sequence type 1.02. Nine isolates from 3 contemporaneous putative clusters in the United Kingdom were available for comparison. Whole-genome sequencing yielded, on average, 85-fold depth for these 15 isolates. All shared 7-locus MLST sequence type 28 with 24 published genomes from the United Kingdom (*20*). Analysis of all 39 genomes against the *S. pyogenes* MGAS5005 genome demonstrated 334 SNPs. Separation of the isolates from the cross-institutional *S. pyogenes* outbreak from all other analyzed isolates was strongly supported (98% bootstrap value). Isolates from the staff member and 3 residents of the 2 homes were positioned in the same monophylogenetic clade and, in each case, differed from the isolates from the remaining 2 infected residents by a single SNP (Figure 2). Gene-by-gene analysis (wgMLST) also clustered the LTCF outbreak (Figure 2). The 6 isolates showed pairwise differences from each other at 1–9 (median 4) of the 1,514 loci and pairwise differences from all other isolates at ≥13 (median 37) loci (Figure 3). The outbreak reported by Turner et al. (*20*) showed similar within-outbreak pairwise differences at 0–9 (median 5) loci. Putative contemporaneous clusters X and Y were also supported by genome sequencing.

We determined the presence or absence of 68 virulence or antimicrobial drug–resistance–associated genes in the genomes sequenced for this study (online Technical Appendix Table). With the following exceptions, all isolates had the same genotype: *speC* was present only in isolates Y1 and Y2 (Figure 2), MF4 was present only in isolate TR7, and *spd*3 and *spy*1438 were present in all isolates except TR7. This varied presence of these 4 bacteriophage-associated genes may reflect phage mobility within *S. pyogenes* (*32*). Genes and mutations associated with macrolide and tetracycline resistance (homologs of *mefA*, *tetM,* and *tetO*) and fluoroquinolone-resistance mutations (in genes *parC* and *gyrA*) were not found in comparison with the fluoroquinolone-susceptible reference strain ATCC700294 (GenBank accession no. AF220946.1);

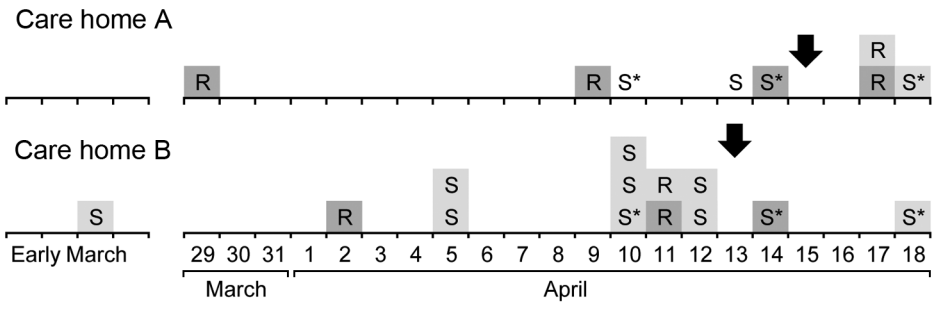

Figure 1. Onset dates of group A *Streptococcus* infection in 2 long-term care facilities in Oxfordshire, United Kingdom, 2013. Arrows indicate initiation of chemoprophylaxis; S, staff; R, resident; * indicates staff who worked in both homes. Dark gray shading indicates laboratory-confirmed infections; light gray shading indicates nonconfirmed infections.

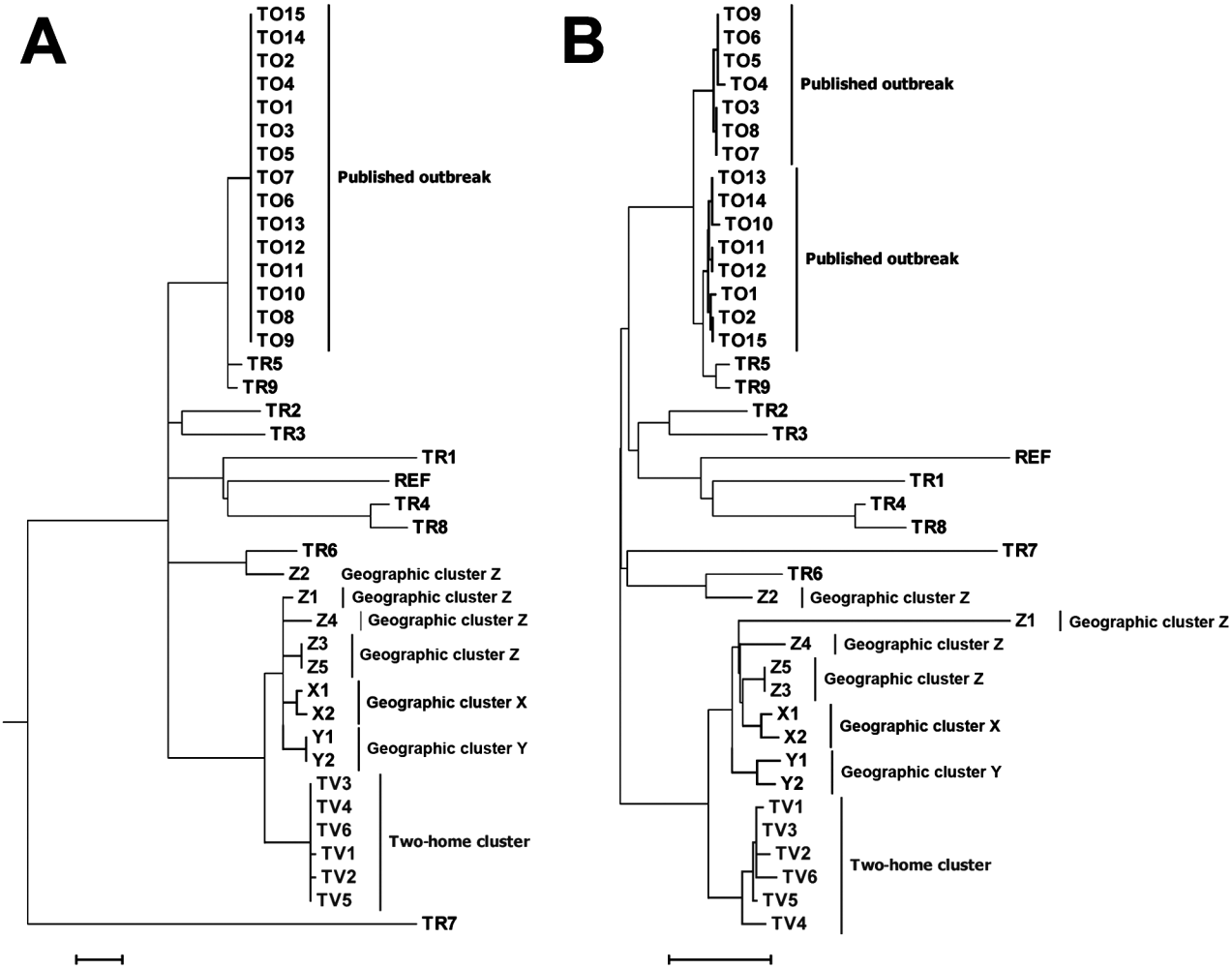

Figure 2. Genetic relatedness of isolates from a cross-institutional *Streptococcus pyogenes* outbreak in Oxfordshire, United Kingdom (indicated by TV plus isolate number); an outbreak described by Turner et al. (*20*) (indicated by TR and TO plus isolate number for reference and outbreak isolates, respectively); and 3 geographic outbreak clusters in the United Kingdom around the time of the TV outbreak (indicated by X, Y, or Z plus isolate number). Dendrograms are based on a single-nucleotide polymorphism maximum-likelihood phylogenetic tree constructed by using RAxML (*29*) (A) and on a neighbor-joining tree constructed from the allelic differences distance matrix (B). Scale bars indicate 10 single-nucleotide polymorphism differences (A) and 15 allelic differences (B).

however, a nonsynonymous SNP (T2393C) of the *par*C gene was identified in all strains. Mutations in covR/S (the *S. pyogenes* regulatory system), which are reported to be associated with strain hyperinvasiveness, were seen in all 15 isolates, in common with strain MGAS5005. No unique mutations were identified in any of the isolates.

Discussion

We have described an iGAS outbreak across 2 managerially independent LTCFs, in which some staff worked in both facilities on separate employment contracts. Clustering in the temporal and genomic dimensions supported GAS transmission within and between these settings. Shared staff across care settings without managerial awareness may be common in this sector, in which low pay and part-time working patterns are common. Loss of pay while absent from work has been recognized as a risk factor in LTCF outbreaks (*33*,*34*). Loss of pay when excluded from work in 1 setting may also contribute to spread of infection to other settings, unless staff members understand that this exclusion applies to similar work in other settings and comply with the exclusion from work in all settings.

We found 1 other report of an iGAS outbreak across 2 LTCFs; the article explored the use of subtyping but did not include epidemiologic details of the outbreak (*35*). Spread of infection between institutions, resulting in apparently sporadic cases or small outbreaks in each, may be difficult to identify as a single outbreak. Our genomic data, triangulated with other descriptive epidemiologic data, supported our conclusion that this was a single outbreak. However,

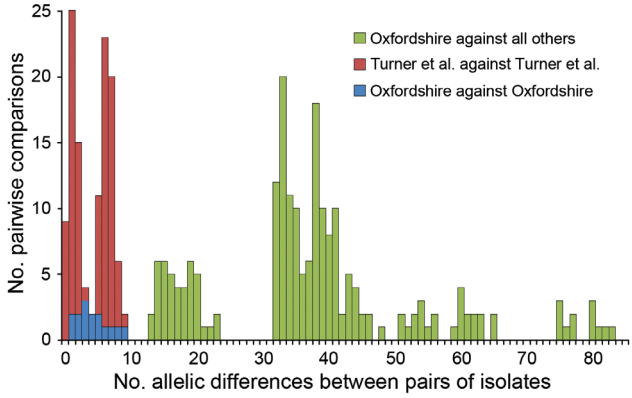

Figure 3. Pairwise allelic differences (across 1,514 genetic loci) among 6 isolates from a cross-institutional *Streptococcus pyogenes* outbreak in Oxfordshire, United Kingdom, and other isolates. Green indicates differences between each of the 6 Oxfordshire outbreak isolates and each of the other 33 isolates that occurred in other geographic areas in the United Kingdom around the time of the Oxfordshire outbreak or were reported by Turner et al. in 2013 (*20*). Red indicates differences between outbreak isolates from the cluster described by Turner et al. (*20*). Blue indicates differences between each isolate in the Oxfordshire outbreak compared with each of the other 5 isolates in the outbreak.

the application of these techniques is relatively untried: neither standardized methods nor extensive national, genome-sequenced reference populations are in place for *S. pyogenes*, and the population biology of this species at the whole-genome level is not fully described.

We dealt with the lack of an established analytical method, which would confirm that isolates are part of a single outbreak, and uncertain *S. pyogenes* population genomics by using 2 types of analyses (i.e., whole-genome multilocus sequence and single-nucleotide polymorphism analyses) and bioinformatics pipelines that did not rely on shared assumptions about the population biology of this species. The identification of core SNPs from a reference-based assembly, exclusion of SNPs that are nearby to avoid an excessive effect from single recombination events, and subsequent generation of an SNP-based phylogeny echo the techniques used in other analyses of this species (*20*,*21*). In our analysis of wgMLST data, we used a reference-free assembly method and assessed the number of shared and discordant alleles across the dataset without making assumptions on processes giving rise to allelic variation (*28*). The replication of clustering by different and independent approaches adds credibility to the epidemiologic inferences drawn. In the absence of extensive genome-sequenced reference populations relevant to the incident under investigation, we used a set of isolates that were identical to the isolates of interest by conventional *emm* sequence typing methods (*36*); some of the isolates from this set were also from the same time period and country as the isolates of interest. Discriminating isolates from other isolates of the same *emm* type can thus demonstrate discrimination from the population of *S. pyogenes* strains as a whole and support outbreak management when appropriate population-based, genome-sequenced reference datasets are unavailable.

Each approach clearly identified the outbreak cluster and differentiated the isolates in this outbreak from other contemporaneous isolates within the same *emm* gene sequence type (*emm*1) and MLST type, similar to findings by Turner et al. (*20*). SNP analysis indicated isolates from homes A and B were separated from all other clades by at least 14 SNPs, but they differed from each other by a maximum of 2 SNPs. Gene-by-gene analysis showed a median of 4 allelic differences in pairwise comparisons within the outbreak across 1,514 genetic loci; this finding contrasts with a median of 37 allelic differences between isolates from the current outbreak and 33 other genomes of the same *emm* type. Gene-by-gene analysis of data from the outbreak reported by Turner et al. (*20*) showed almost identical within-outbreak variation (median of 5 differences, range 0–9) (Figure 3). Thus a similar amount of allelic variation was present in the cross-institutional cluster and the cluster reported by Turner et al. (*20*). These findings contrast with much greater variation when compared with other isolates from the same *emm* type. These 2 incidents had epidemiologic data indicating likely transmission over a period of days to a few weeks. This range of variation may therefore be an estimate for expected variation in the transmission systems generating small, short-lived GAS outbreaks.

Additional well-characterized outbreaks with genome-sequenced isolates will enable fuller empirical validation of these results. More long-lived outbreaks may be associated with greater variation, and, indeed, multistrain streptococcal outbreaks can arise where a mobile genetic element can support the increase of several pathogenic lineages acquiring it (*37*). In contrast to the clear and similar discrimination of the cross-institutional outbreak isolates from other *emm*1 isolates by each analytical approach, no marked structure was shown by either analysis among the outbreak isolates. The limited extent of evolution occurring, as indexed and analyzed by our approaches across the available genome, does not enable inference on particular transmission pathways within this short-lived, single-clone outbreak.

Our literature review showed that most reported LTCF iGAS outbreaks have been caused by a single strain, and many outbreaks go unrecognized. The use of genome sequence analysis may distinguish epidemiologic clusters from background isolates, enabling detection of a large proportion of currently missed LTCF outbreaks. Given literature-based estimates that 10%–25% of apparently sporadic iGAS cases in LTCFs may belong to outbreaks (*7*,*8*), the integration of genome sequencing into iGAS surveillance in this population might be justified by this purpose alone.

Detection of missed outbreaks may enable identification of risk factors that contribute to the outbreaks. Genomic surveillance could also support detection and investigation of transmission events between LTCFs and other settings. The relatively low population-level incidence of iGAS (*1*–*3*) may make genome sequence surveillance financially feasible in countries that have microbiologic iGAS surveillance systems in place.

This outbreak was controlled after use of mass chemoprophylaxis and standard infection-control measures. Infection-control measures are widely reported as part of iGAS outbreak control methods; chemoprophylaxis shows more variation. There was, on average, better control in outbreaks in which mass chemoprophylaxis was undertaken, and there were reports of further cases during the wait for screening test results (*38*,*39*). However, the evidence base is limited and carriage rates are low among residents and staff (*12*,*13*), and some argue that mass chemoprophylaxis is inappropriate (*40*). The ability to identify large numbers of outbreaks early, as appears possible by genomic surveillance, may offer a sampling frame to generate more reliable data for the role of chemoprophylaxis by analyzing outcomes across a large number of outbreaks by the approach used or through trials of different approaches.

In summary, we investigated an iGAS outbreak in 2 institutions by integrating pathogen genomic epidemiology to infer epidemiologic relatedness. Independent bioinformatic and population genetic approaches enabled credible conclusions in the absence of a standardized approach. The excess burden of iGAS among elderly residents of LTCFs, the large proportion of cases that are associated with undetected outbreaks, and the consequent opportunity to improve the evidence base for control support consideration of genomic surveillance of iGAS.

This work was supported by core UK government public health funding through Public Health England.

Dr. Chalker is the head of the Respiratory and Systemic Bacteria Section at the Respiratory and Vaccine Preventable Bacteria Reference Unit of Public Health England. Her main interests are in streptococcal, *Legionella* and *Leptospira* spp., and mycoplasma infections in humans and other animals.

References

1. Lamagni TL, Darenberg J, Luca-Harari B, Siljander T, Efstratiou A, Henriques-Normark B, et al. Epidemiology of severe *Streptococcus pyogenes* disease in Europe. J Clin Microbiol. 2008;46:2359–67. http://dx.doi.org/10.1128/JCM.00422-08
2. Lepoutre A, Doloy A, Bidet P, Leblond A, Perrocheau A, Bingen E, et al. Epidemiology of invasive *Streptococcus pyogenes* infections in France in 2007. J Clin Microbiol. 2011;49:4094–100. http://dx.doi.org/10.1128/JCM.00070-11
3. O'Loughlin RE, Roberson A, Cieslak PR, Lynfield R, Gershman K, Craig A, et al. The epidemiology of invasive group A streptococcal infection and potential vaccine implications: United States, 2000–2004. Clin Infect Dis. 2007;45:853–62. http://dx.doi.org/10.1086/521264
4. Davies HD, McGeer A, Schwartz B, Green K, Cann D, Simor AE, et al. Invasive group A streptococcal infections in Ontario, Canada. Ontario Group A Streptococcal Study Group. N Engl J Med. 1996 ;335:547–54. http://dx.doi.org/10.1056/NEJM199608223350803
5. Lamagni TL, Neal S, Keshishian C, Powell D, Potz N, Pebody R, et al. Predictors of death after severe *Streptococcus pyogenes* infection. Emerg Infect Dis. 2009;15:1304–7. http://dx.doi.org/10.3201/eid1508.090264
6. Muller MP, Low DE, Green KA, Simor AE, Loeb M, Gregson D, et al. Clinical and epidemiologic features of group a streptococcal pneumonia in Ontario, Canada. Arch Intern Med. 2003;163:467–72. http://dx.doi.org/10.1001/archinte.163.4.467
7. Rainbow J, Jewell B, Danila RN, Boxrud D, Beall B, Van Beneden C, et al. Invasive group a streptococcal disease in nursing homes, Minnesota, 1995–2006. Emerg Infect Dis. 2008; 14:772–7.
8. Thigpen MC, Richards CL Jr, Lynfield R, Barrett NL, Harrison LH, Arnold KE, et al. Invasive group A streptococcal infection in older adults in long-term care facilities and the community, United States, 1998–2003. Emerg Infect Dis. 2007;13:1852–9. http://dx.doi.org/10.3201/eid1312.070303
9. Zurawski CA, Bardsley M, Beall B, Elliott JA, Facklam R, Schwartz B, et al. Invasive group A streptococcal disease in metropolitan Atlanta: a population-based assessment. Clin Infect Dis. 1998;27:150–7. http://dx.doi.org/10.1086/514632
10. Daneman N, Green KA, Low DE, Simor AE, Willey B, Schwartz B, et al. Surveillance for hospital outbreaks of invasive group A streptococcal infections in Ontario, Canada, 1992 to 2000. Ann Intern Med. 2007;147:234–41. http://dx.doi.org/10.7326/0003-4819-147-4-200708210-00004
11. Steer JA, Lamagni T, Healy B, Morgan M, Dryden M, Rao B, et al. Guidelines for prevention and control of group A streptococcal infection in acute healthcare and maternity settings in the UK. J Infect. 2012;64:1–18. http://dx.doi.org/10.1016/j.jinf.2011.11.001
12. Jordan HT, Richards CL Jr, Burton DC, Thigpen MC, Van Beneden CA. Group A streptococcal disease in long-term care facilities: descriptive epidemiology and potential control measures. Clin Infect Dis. 2007;45:742–52. http://dx.doi.org/10.1086/520992
13. Cummins A, Millership S, Lamagni T, Foster K. Control measures for invasive group A streptococci (iGAS) outbreaks in care homes. J Infect. 2012;64:156–61. http://dx.doi.org/10.1016/j.jinf.2011.11.017
14. Centers for Disease Control and Prevention. Invasive group A streptococcus in a skilled nursing facility—Pennsylvania, 2009–2010. MMWR Morb Mortal Wkly Rep. 2011;60:1445–9.
15. Greene CM, Van Beneden CA, Javadi M, Skoff TH, Beall B, Facklam R, et al. Cluster of deaths from group A streptococcus in a long-term care facility—Georgia, 2001. Am J Infect Control. 2005;33:108–13. http://dx.doi.org/10.1016/j.ajic.2004.07.009
16. Rahman M. Outbreak of *Streptococcus pyogenes* infections in a geriatric hospital and control by mass treatment. J Hosp Infect. 1981;2:63–9. http://dx.doi.org/10.1016/0195-6701(81)90007-4
17. Reid RI, Briggs RS, Seal DV, Pearson AD. Virulent *Streptococcus pyogenes*: outbreak and spread within a geriatric unit. J Infect. 1983;6:219–25. http://dx.doi.org/10.1016/S0163-4453(83)93549-1
18. McCarthy N. An epidemiological view of microbial genomic data. Lancet Infect Dis. 2013;13:104–5. http://dx.doi.org/10.1016/S1473-3099(12)70324-9
19. Walker TM, Lalor MK, Broda A, Saldana Ortega L, Morgan M, Parker L, et al. Assessment of *Mycobacterium tuberculosis* transmission in Oxfordshire, UK, 2007–12, with whole pathogen genome sequences: an observational study. Lancet Respir Med. 2014;2:285–92. http://dx.doi.org/10.1016/S2213-2600(14)70027-X

20. Turner CE, Dryden M, Holden MT, Davies FJ, Lawrenson RA, Farzaneh L, et al. Molecular analysis of an outbreak of lethal postpartum sepsis caused by *Streptococcus pyogenes*. J Clin Microbiol. 2013;51:2089–95. http://dx.doi.org/10.1128/JCM.00679-13
21. Ben Zakour NL, Venturini C, Beatson SA, Walker MJ. Analysis of a *Streptococcus pyogenes* puerperal sepsis cluster by use of whole-genome sequencing. J Clin Microbiol. 2012;50:2224–8. http://dx.doi.org/10.1128/JCM.00675-12
22. Health Protection Agency, Group A Streptococcus Working Group. Interim UK guidelines for management of close community contacts of invasive group A streptococcal disease. Commun Dis Public Health. 2004;7:354–61.
23. Johnson DR, Kaplan EL, Sramek J, Bicova R, Havlicek JHH, Havlickova H, et al. Laboratory diagnosis of group A streptococcal infections. Geneva: World Health Organization; 1996 [cited 2014 Dec 20]. http://apps.who.int/iris/bitstream/10665/41879/1/9241544953_eng.pdf
24. Bidet P, Lesteven E, Doit C, Liguori S, Mariani-Kurkdjian P, Bonacorsi S, et al. Subtyping of *emm*1 group A streptococci causing invasive infections in France. J Clin Microbiol. 2009;47:4146–9. http://dx.doi.org/10.1128/JCM.00866-09
25. Li H, Durbin R. Fast and accurate long-read alignment with Burrows–Wheeler transform. Bioinformatics. 2010;26:589–95. http://dx.doi.org/10.1093/bioinformatics/btp698
26. McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. Genome Res. 2010;20:1297–303. http://dx.doi.org/10.1101/gr.107524.110
27. Zerbino DR, Birney E. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. Genome Res. 2008;18:821–9. http://dx.doi.org/10.1101/gr.074492.107
28. Maiden MC, van Rensburg MJ, Bray JE, Earle SG, Ford SA, Jolley KA, et al. MLST revisited: the gene-by-gene approach to bacterial genomics. Nat Rev Microbiol. 2013;11:728–36. http://dx.doi.org/10.1038/nrmicro3093
29. Stamatakis A. RAxML-VI-HPC: maximum likelihood–based phylogenetic analyses with thousands of taxa and mixed models. Bioinformatics. 2006;22:2688–90. http://dx.doi.org/10.1093/bioinformatics/btl446
30. Jolley KA, Maiden MCJ. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinformatics. 2010;11:595. http://dx.doi.org/10.1186/1471-2105-11-595
31. Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S. MEGA5: Molecular Evolutionary Genetics Analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol. 2011;28:2731–9. http://dx.doi.org/10.1093/molbev/msr121
32. Banks DJ, Beres SB, Musser JM. The fundamental contribution of phages to GAS evolution, genome diversification and strain emergence. Trends Microbiol. 2002;10:515–21. http://dx.doi.org/10.1016/S0966-842X(02)02461-7
33. Schwartz B, Ussery XT. Group A streptococcal outbreaks in nursing homes. Infect Control Hosp Epidemiol. 1992;13:742–7. http://dx.doi.org/10.2307/30146492
34. Thigpen MC, Thomas DM, Gloss D, Park SY, Khan AJ, Fogelman VL, et al. Nursing home outbreak of invasive group A streptococcal infections caused by 2 distinct strains. Infect Control Hosp Epidemiol. 2007;28:68–74.
35. Stanley J, Desai M, Xerry J, Tanna A, Efstratiou A, George R. High-resolution genotyping elucidates the epidemiology of group A streptococcus outbreaks. J Infect Dis. 1996;174:500–6. http://dx.doi.org/10.1093/infdis/174.3.500
36. Athey TB, Teatero S, Li A, Marchand-Austin A, Beall BW, Fittipaldi N. Deriving group A *Streptococcus* typing information from short-read whole-genome sequencing data. J Clin Microbiol. 2014;52:1871–6. http://dx.doi.org/10.1128/JCM.00029-14
37. Davies MR, Holden MT, Coupland P, Chen JH, Venturini C, Barnett TC, et al. Emergence of scarlet fever *Streptococcus pyogenes* emm12 clones in Hong Kong is associated with toxin acquisition and multidrug resistance. Nat Genet. 2015;47:84–7. http://dx.doi.org/10.1038/ng.3147
38. Smith A, Li A, Tolomeo O, Tyrrell GJ, Jamieson F, Fisman D. Mass antibiotic treatment for group A streptococcus outbreaks in two long-term care facilities. Emerg Infect Dis. 2003;9:1260–5. http://dx.doi.org/10.3201/eid0910.030130
39. Arnold KE, Schweitzer JL, Wallace B, Salter M, Neeman R, Hlady WG, et al. Tightly clustered outbreak of group A streptococcal disease at a long-term care facility. Infect Control Hosp Epidemiol. 2006;27:1377–84. http://dx.doi.org/10.1086/508820
40. Milne LM, Lamagni T, Efstratiou A, Foley C, Gilman J, Lilley M, et al. *Streptococcus pyogenes* cluster in a care home in England April to June 2010. Euro Surveill. 2011;16:20021.

Address for correspondence: Noel D. McCarthy, Warwick Medical School, University of Warwick, Coventry, UK; email: n.d.mccarthy@warwick.ac.uk

PubMed Central

PubMed

NATIONAL LIBRARY OF MEDICINE

Find *Emerging Infectious Diseases* content in the digital archives of the National Library of Medicine

www.pubmedcentral.nih.gov

Article DOI: http://dx.doi.org/10.3201/eid2206.142050

Integration of Genomic and Other Epidemiologic Data to Investigate and Control Cross-Institutional *Streptococcus* p*yogenes* Outbreak

Technical Appendix

Full Literature Review Results

Seventy-two iGAS outbreaks in LTCFs were identified, from individual outbreak reports (*1–14*) or summary data identified in surveillance studies or reviews (*6*,*13*,*15–20*), or laboratory studies (*16*). These 72 include 31 clusters defined as outbreaks based on shared subtypes among 2 or more cases occurring at a facility within 1 year (*15*,*18*). One of these found 13 clusters which shared sub-type, institution and time period among 134 cases, and eight instances of two isolates from the same time period and institution but with distinguishable subtypes which did not meet their subtype based outbreak definition. Among outbreaks identified by other means 22 reported subtyping. Nineteen of these identified a single or dominant strain (*1*,*3*,*5–13*,*16*), and three multiple strains (*2*,*4*,*14*).

Treatment restricted to those ill or to cases and direct contacts of cases with laboratory confirmed infection was associated with control in some outbreaks (*5*,*7*,*13*), although in one mass chemoprophylaxis was applied following recurrence after two rounds of this selective treatment (*1*). Screening all staff and residents with chemoprophylaxis if positive (*2–4*,*6*,*8*,*9*,*11–14*) was associated with recurrence and repeated screening and treatment in two reports (*2*,*8*). Further cases of iGAS were diagnosed between decisions to screen and chemoprophylax and the implementation of chemoprophylaxis (*1*,*11*). Mass chemoprophylaxis of all staff and residents was associated with control of iGAS (*1*,*3*,*10*,*13*) although in one incident persistent infection was shown in one resident with a gastrostomy tube (*1*). Screening detected carriage rates were below 10% among residents with two exceptions (20% (*5*) and 16% (*11*)) with lower rates among staff than residents as reviewed elsewhere (*6*,*13*).

References

1. Smith A, Li A, Tolomeo O, Tyrrell GJ, Jamieson F, Fisman D. Mass antibiotic treatment for group A streptococcus outbreaks in two long-term care facilities. Emerg Infect Dis. 2003;9:1260–5. PubMed http://dx.doi.org/10.3201/eid0910.030130

2. Control CfD. Invasive group A streptococcus in a skilled nursing facility—Pennsylvania, 2009–2010. MMWR Morb Mortal Wkly Rep. 2011;60:1445–9. PubMed

3. Control CfD. Nursing home outbreaks of invasive group A streptococcal infections—Illinois, Kansas, North Carolina, and Texas. MMWR Morb Mortal Wkly Rep. 1990;39:577–9. PubMed

4. Schwartz B, Elliott JA, Butler JC, Simon PA, Jameson BL, Welch GE, et al. Clusters of invasive group A streptococcal infections in family, hospital, and nursing home settings. Clin Infect Dis. 1992;15:277–84. PubMed http://dx.doi.org/10.1093/clinids/15.2.277

5. Ruben FL, Norden CW, Heisler B, Korica Y. An outbreak of Streptococcus pyogenes infections in a nursing home. Ann Intern Med. 1984;101:494–6. PubMed http://dx.doi.org/10.7326/0003-4819-101-4-494

6. Jordan HT, Richards CL Jr, Burton DC, Thigpen MC, Van Beneden CA. Group A streptococcal disease in long-term care facilities: descriptive epidemiology and potential control measures. Clin Infect Dis. 2007;45:742–52. PubMed http://dx.doi.org/10.1086/520992

7. Harkness GA, Bentley DW, Mottley M, Lee J. *Streptococcus pyogenes* outbreak in a long-term care facility. Am J Infect Control. 1992;20:142–8. PubMed http://dx.doi.org/10.1016/S0196-6553(05)80181-6

8. Greene CM, Van Beneden CA, Javadi M, Skoff TH, Beall B, Facklam R, et al. Cluster of deaths from group A streptococcus in a long-term care facility—Georgia, 2001. Am J Infect Control. 2005;33:108–13. PubMed http://dx.doi.org/10.1016/j.ajic.2004.07.009

9. Barnham M, Hunter S, Hanratty B, Kirby P, Tanna A, Efstratiou A. Invasive M-type 3 *Streptococcus pyogenes* affecting a family and a residential home. Commun Dis Public Health. 2001;4:64–7. PubMed

10. Auerbach SB, Schwartz B, Williams D, Fiorilli MG, Adimora AA, Breiman RF, et al. Outbreak of invasive group A streptococcal infections in a nursing home. Lessons on prevention and control. Arch Intern Med. 1992;152:1017–22. PubMed http://dx.doi.org/10.1001/archinte.1992.00400170099019

11. Arnold KE, Schweitzer JL, Wallace B, Salter M, Neeman R, Hlady WG, et al. Tightly clustered outbreak of group A streptococcal disease at a long-term care facility. Infect Control Hosp Epidemiol. 2006;27:1377–84. PubMed http://dx.doi.org/10.1086/508820

12. Milne LM, Lamagni T, Efstratiou A, Foley C, Gilman J, Lilley M, et al. *Streptococcus pyogenes* cluster in a care home in England April to June 2010. Euro Surveill. 2011;16:20021. PubMed

13. Cummins A, Millership S, Lamagni T, Foster K. Control measures for invasive group A streptococci (iGAS) outbreaks in care homes. J Infect. 2012;64:156–61. PubMed http://dx.doi.org/10.1016/j.jinf.2011.11.017

14. Thigpen MC, Thomas DM, Gloss D, Park SY, Khan AJ, Fogelman VL, et al. Nursing home outbreak of invasive group A streptococcal infections caused by 2 distinct strains. Infect Control Hosp Epidemiol. 2007;28:68–74. PubMed

15. Thigpen MC, Richards CL Jr, Lynfield R, Barrett NL, Harrison LH, Arnold KE, et al. Invasive group A streptococcal infection in older adults in long-term care facilities and the community, United States, 1998–2003. Emerg Infect Dis. 2007;13:1852–9. PubMed http://dx.doi.org/10.3201/eid1312.070303

16. Stanley J, Desai M, Xerry J, Tanna A, Efstratiou A, George R. High-resolution genotyping elucidates the epidemiology of group A streptococcus outbreaks. J Infect Dis. 1996;174:500–6. PubMed http://dx.doi.org/10.1093/infdis/174.3.500

17. Zurawski CA, Bardsley M, Beall B, Elliott JA, Facklam R, Schwartz B, et al. Invasive group A streptococcal disease in metropolitan Atlanta: a population-based assessment. Clin Infect Dis. 1998;27:150–7. PubMed http://dx.doi.org/10.1086/514632

18. Rainbow J, Jewell B, Danila RN, Boxrud D, Beall B, Van Beneden C, et al. Invasive group a streptococcal disease in nursing homes, Minnesota, 1995–2006. Emerg Infect Dis. 2008;14:772–7. PubMed

19. Muller MP, Low DE, Green KA, Simor AE, Loeb M, Gregson D, et al. Clinical and epidemiologic features of group a streptococcal pneumonia in Ontario, Canada. Arch Intern Med. 2003;163:467–72. PubMed http://dx.doi.org/10.1001/archinte.163.4.467

20. Davies HD, McGeer A, Schwartz B, Green K, Cann D, Simor AE, et al. Invasive group A streptococcal infections in Ontario, Canada. Ontario Group A Streptococcal Study Group. N Engl J Med. 1996;335:547–54. PubMed http://dx.doi.org/10.1056/NEJM199608223350803

Technical Appendix Table. Genes of interest, presence and absence

Gene name	Accession no. used for comparison	Protein and Function	Presence/absence in study isolates
Capsule			
HasA		Hyaluronase, production of hyaluronic acid capsule	+
HasB		Production of hyaluronic acid capsule	+
HasC		Production of hyaluronic acid capsule	+
PrtS		Cell envelope proteinase	+
Phage			
AbiR		Abortive infection phage resistance protein	+
MF4		Mitogenic factor - phage associated	– (except ERR027369 +)
Spd3		DNase (Similar to mitogenic factor), phage associated)	+ (except ERR027369 –)
Spy1438		Phage associated cell wall hydrolase	+ (except ERR027369 –)
SpyM3–1302	SpyM3_1302, gi|21909536:c1315318–1314461 *Streptococcus pyogenes* MGAS315 chromosome, complete genome	Phage protein	–
Regulatory proteins			
CcpA		Regulatory protein	+
CovR	gb|EU726239.1|:241–927 *S. pyogenes* strain H342 CovR	Two-component regulation	+
CovS	gb|EU726239.1|:933–2435 *S. pyogenes* strain H342 CovS	Two-component regulation	+
MsmR	gb|DQ984656.1|:6744–7949 *S. pyogenes* strain ALAB49 MsmR	Operon regulatory protein, msm	–
MtsR		Transcriptional regulator of metal ABC transporter MtsR-like Metalloregulator – manganese	+
Nra	gb|DQ984656.1|:126–1661 *S. pyogenes* strain ALAB49 Nra	Negative transcriptional regulator	–
PerR		Ferric uptake regulation protein	+
Rgg		Rgg regulator, production of SpeB Cysteine protease	+
RopB		Rgg regulator, production of SpeB Cysteine protease	+
Attachment			
Cpa	gb|DQ984656.1|:2092–3675 *S. pyogenes* strain ALAB49 Cpa	Fimbrial minor sturctural protein Cpa/FctA	–
FctA	gb|DQ984656.1|:4211–5158 *S. pyogenes* strain ALAB49 FctA	Fimbrial minor sturctural protein Cpa/FctA	–
FctB	gb|DQ984656.1|:6014–6583 *S. pyogenes* strain ALAB49 Nra FctB	Fimbrial minor structural protein	–
Toxins/pathogenicity factors			
ABC-transporter	gb|AF227521.1|:4607–6070	ABC-transporter	–
Emm1		M protein	+
EndoS		Endoglycoside, immunoglobulin degrading enzyme	+
Fba		Fructose-bisphosphate aldolase	+
GRAB		G protein-related a2-macroglobulin binding protein	+
IdeS		Immunoglobulin G endopeptidase, CD11b homologue	–
LaCD		Tagatose 1,6-diphosphate aldolase	+
Mac		Immunoglobulin G-endopeptidase (IdeS)	–
PrtF2	gb|DQ984656.1|:8334–10535 *S. pyogenes* strain ALAB49 PrtF2	Fibrinonectin binding protein	–
SagA		Streptolysin S precursor	+
SagB		Streptolysin S production	+
SagC		Streptolysin S production	+
Scl1		Collagen-like surface protein	–
Scl2		Collagen-like surface protein	–
ScpA		C5a peptidase	+
Sda1		Extracellular streptodornase S, DNase	+
SpeA		Exotoxin A superantigen	+
SpeB		Streptococcal pyrogenic exotoxin B, cysteine protease	+

Gene name	Accession no. used for comparison	Protein and Function	Presence/absence in study isolates
SpeC		Exotoxin C	– (except H130620574, H130620575 +)
SpeJ		Exotoxin J	+
Sic		Streptococcal inhibitor of complement	+
Ska		Streptokinase	–
Slo		Streptolysin O	+
Sof		Serum opacity factor	–
Spd		DNase	+
SpeG		Pyrogenic exotoxin	+
Spy1063		Iron(III) binding protein	+
Spy1064		Unknown	+
Spy1066		Hydrolase (HAD superfamily)	+
Spy1065		O-Acetyltranseferase (cell wall biosynthesis)	+
SpyCEP	gi|94989509:345557–351976 *S. pyogenes* MGAS10270 chromosome, complete genome	Interleukin-8 protease, cell envelope protein	–
SpyA	gi|71909814:c358588–357610 *S. pyogenes* MGAS5005 chromosome, complete genome	C3 family ADP-ribosyltransferase	+
SrtC2	gb|DQ984656.1|:5254–5979 *S. pyogenes* strain ALAB49 SrtC2	Sortase	–
UmuC-MucB	gb|AF227521.1|:6839–8128	Lesion-replicating DNA polymerases	–
Antimicrobial resistance			
FbaA		Fibrinonectin binding protein	+
GyrA		Fluoroquinalone resistance	+
Mef		Macrolide resistance	–
MefA	gb|AF227521.1|:3270–4487	Macrolide resistance macrolide-efflux protein A (mefA)	–
MefMB56Spyo029		Macrolide resistance	–
MefV1		Macrolide resistance	–
MefV2		Macrolide resistance	–
MefV3		Macrolide resistance	–
Nga		NAD glycohydrolase	+
ParC		Fluoroquinalone resistance	+
TetM		Tetracycline resistance	–
TetO		Tetracycline resistance	–