See related article on page 1522

The American Journal of Pathology, Vol. 180, No. 4, April 2012

DOI: 10.1016/j.ajpath.2012.01.007

Commentary

Population Genomics

An Investigative Tool for Epidemics

Debra E. Bessen

From the Department of Microbiology and Immunology, New York Medical College, Valhalla, New York

Group A streptococcus (GAS) is a highly prevalent bacterial pathogen that has been recovered from its human hosts throughout the world. Most epidemiological studies characterize GAS isolates by their *emm* type, a molecular marker that has ~200 distinct forms. The *emm* gene encodes M protein, a surface fibril that is a major virulence factor and target of protective immunity. Collection of GAS, along with the molecular typing of isolates, has been highly active throughout much of the world for the past several decades. Within any one community over a 1- or 2-year surveillance period, roughly 25 to 40 distinct *emm* types can be detected among the GAS isolates.[1–3] Each host community has a unique array of *emm* types, and the degree of overlap in the spectrum of *emm* types decreases for host populations in rough proportion to their geographical distance. Even within the same community, the predominating *emm* types can shift from year to year.[1]

In this issue of *The American Journal of Pathology*, Fittipaldi et al[4] present a magnified view of the unfolding emergence and epidemic spread of an *emm59* strain of GAS. Before 2004, when the epidemic had its beginnings in western Canada, recovery of *emm59* isolates was rare. Steer at al[5] recently provided a meta-analysis of >100 reports on GAS molecular epidemiology for >38,000 isolates collected from 37 countries since 1980: in every region of the world, *emm59* strains are absent from the list of the 25 most prevalent *emm* types, even when those lists are further stratified by disease. GAS isolates of *emm* type 59 were largely below the radar before 2003, but have since accounted for >540 cases of invasive disease in Canada.[6]

Generation of Epidemic emm59 *GAS*

How does an epidemic begin? One idea is that a preexisting strain finds itself in a new environment or in a new host population, whereupon the conditions encountered are favorable for its reproductive growth and sustained transmission. Either the organism migrates to a new community, or the conditions within a community undergo a drastic change. An alternative hypothesis states that a genetic change in a bacterium leads to an increase in fitness, which in turn leads to rapid spread of the new genetic variant. The genetic change may have resulted in acquisition of a new virulence factor or an antigenic shift, leading to increased disease severity and/or enhanced transmission. The fitness of a microorganism often reflects the host and/or environmental conditions, and each genetic variant can experience a different level of fitness within the same community. For example, if the new bacterium undergoes a change in antigenic structure and is present within an immunologically naïve host population, the organism can experience a leap in fitness whereby the conditions are now ripe for its clonal expansion through enhanced transmission. This occurred with serotype 19A pneumococcus after widespread vaccination of the host population with the heptavalent conjugate vaccine.[7] For another example, if the organism acquires an antibiotic resistance gene and is present in an environment where antibiotic usage is high, an outbreak with the new strain might occur. Positive selection on the new genetic variant is the driving force behind its expansion to epidemic proportions.

Fittipaldi et al[4] present a reasonable argument that a change in host and/or environmental conditions is unlikely to have been a primary cause of the *emm59* epidemic, because none of the other coexisting GAS strains experienced a similar surge in prevalence within the same communities. Thus, a genetic mechanism likely produced the epidemic *emm59* organism. The authors examined the genetic differences that distinguish historic *emm59* isolates from the contemporary epidemic *emm59* isolates. By mapping such genetic differences, they sought to identify gene candidates that may be respon-

Supported by the NIH (R01-AI065572).

Accepted for publication January 19, 2012.

CME Disclosure: The author did not disclose any relevant financial relationships.

Address reprint requests to Debra Bessen, Ph.D., New York Medical College, Department of Microbiology and Immunology, Valhalla, NY 10595. E-mail: debra_bessen@nymc.edu

sible for the shift in virulence or transmissibility that led to the epidemic. To do this, the authors exploited advances in next-generation sequencing technology and determined the nucleotide sequence of the genomes of ~600 isolates of *emm59* GAS. Initially using one representative historic and one epidemic isolate for comparison, ~150 single-nucleotide polymorphisms (SNPs) or indels (insertions or deletions) were uncovered within the core genome. In addition, the epidemic isolate lost one prophage (which harbored two virulence genes, a superantigen and phospholipase) and gained a prophage remnant at another locus, relative to the historic isolate. Might the loss of a virulence gene have led to the epidemic? Do any of the SNPs or indels lie within or adjacent to genes having an established role in GAS virulence? That analysis is currently ongoing at the authors' laboratory.

Is it possible that the genetic change hypothesis is incorrect? It is possible, but another strong argument weighs heavily in favor of this hypothesis. Fittipaldi et al[4] used animal models to test the virulence of representative historic and contemporary *emm59* isolates. Under three different models for infection of normally sterile soft tissue after injection of bacteria at a subcutaneous or intramuscular site, the epidemic *emm59* isolates displayed significantly greater virulence than the historic isolate. The animal findings showing differences in virulence provide supporting evidence that changes in virulence capacity evolved only recently. Thus, it appears that there may be a genetic change (yet to be defined) underlying the *emm59* epidemic.

Tissue-Specific Properties of the Epidemic emm59 *Strain*

Invasive disease accounts for a large measure of morbidity and mortality due to GAS; however, infection at superficial epithelial tissue sites is far more prevalent.[8] It is from the superficial tissue sites that most person-to-person transmission initiates. The most highly prevalent infections caused by GAS are pharyngitis and nonbullous impetigo. GAS is also associated with asymptomatic carriage at the throat; among school-aged children, the carriage rate often exceeds 20%. Compared with understanding of oropharyngeal carriage, there is less of an understanding of GAS colonization of normal skin. However, in one seminal prospective surveillance study involving children living on the Red Lake Indian Reservation (in northern Minnesota), where GAS pyoderma was endemic, GAS was frequently isolated from normal skin before the development of impetigo lesions.[9] Thus, the primary reservoirs for GAS are the oropharyngeal mucosa and skin, and these tissues accommodate both carriage and infectious processes.

Within just a few years, *emm59* GAS in Canada rose from being rare to becoming the most common cause of invasive GAS infections, accounting for 13% of all invasive isolates collected over a 4-year period.[6] However, no parallel rise in *emm59* GAS isolates associated with pharyngitis occurred, with *emm59* accounting for <0.5% of total pharyngitis isolates.[4] Thus, if the *emm59* strain is transmitted primarily through a respiratory route, either it is extraordinarily virulent and has a very high invasive index, or there is a very strong tendency toward an asymptomatic course when colonizing the oropharynx. The far more likely explanation is that the primary mode of transmission for *emm59* strains is via skin contact.

Fittipaldi et al[4] have begun to investigate the tissue-specific properties of the epidemic *emm59* strain by comparison with an *emm1* GAS strain (MGAS5005); *emm1* is among the most prevalent *emm* types recovered from cases of pharyngitis in the United States and Canada in recent years.[1] The *emm1* and *emm59* strains were compared for survival and growth in human saliva,[10] where essential nutrients may be limiting and antimicrobial peptides may be active. The *emm1* strain exhibited significantly greater survival in human saliva than the epidemic *emm59* strain. The *emm1* strain also showed significantly higher levels of colonization in a mouse oropharyngeal model. The findings support the notion that the MGAS5005 strain has a significantly higher tropism for the throat, compared with the epidemic *emm59* strain.

In contrast to the experimental models for throat infection, in a newly developed model for mouse skin lesions the *emm1* and *emm59* isolates showed equivalent levels of skin-to-skin transmission.[4] Thus, the collective data suggest that the *emm1* strain is highly effective at infecting both throat and skin; the latter finding has strong epidemiological support in the high association of *emm1* strains with soft-tissue infection. However, the MGAS5005 strain (*emm1*) under study is a sterile-site isolate with a mutation in the global transcriptional regulator gene *covS* and, consequently, exhibits the invasive transcriptome profile (TP), as opposed to the pharyngitis TP.[11] Perhaps the invasive TP of MGAS5005 provides an explanation for why this isolate effectively transmits in the mouse skin lesion model (a full-thickness incision), particularly if infection is established by introduction of bacteria into subepithelial tissue.[12] However, why an invasive TP strain such as MGAS5005 has a high capacity for oropharyngeal colonization is not entirely clear. Both the invasive TP and the pharyngitis TP are associated with invasive disease, but mounting evidence suggests that the pharyngitis TP is essential for infection of the throat.[11–13] In some ways, the findings of Fittipaldi et al[4] raise more questions than they provide answers, and the streptococcal field as a whole may need to reconcile these data in the future. The transcriptome profiles of the epidemic and historic *emm59* strains were not reported,[4] and this points to another ripe avenue for future investigations.

Experimental data suggest that the *emm1* strain has a higher affinity for the throat than does the *emm59* strain. This finding gains additional support from epidemiological data showing that the *emm59* strain is recovered only rarely from cases of pharyngitis. Superficial tissue site preferences for infection (ie, throat versus skin) for individual GAS strains have long been noted.[14] A useful biomarker for preferred tissue sites of infection is found in the *emm* pattern genotype, which is based on the 3′ ends of *emm* genes. There are three distinct genotype-defined groups: throat specialists causing pharyngitis (patterns A, B, and C), skin specialists causing impetigo (pattern D), and generalists, which as a group have no overriding preference (pattern E).[15]

The *emm* pattern was determined for three non-Canadian *emm59*/ST172 isolates examined by other investigators in earlier studies; all were pattern D (skin specialists).[16,17] In a recent study comparing the accessory gene regions of GAS strains of >90 different *emm* types,[15] all but a few pattern D strains are associated with a single tight, genetically related cluster. The atypical pattern D exceptions include *emm59*, *emm81*, and *emm85* strains, each of which have the *sof* gene, which is otherwise a distinguishing feature of pattern E strains. In fact, the *emm59*, *emm81*, and *emm85* pattern D strains tend to group with the majority of pattern E strains and harbor pilus-encoding fibronectin-collagen-T (FCT) genetic regions that are also more typical of those of pattern E strains. Recently, *emm85* isolates have been recovered from impetigo lesions in a tropical Australian community.[2] In a London (UK) hospital, *emm81* isolates obtained from a variety of wound and soft-tissue infections in patients had the highest prevalence among all *emm* types.[18] An *emm81* strain was recently associated with an outbreak of ulcerated skin lesions (ecthyma) among Israeli soldiers.[19] The atypical pattern D strains having a pattern E-like genotype (ie, *emm59*, *emm81*, and *emm85*) may possess factors that are required to initiate infection at the skin. However, unlike the typical pattern D strains, which tend to be associated with impetigo and are seemingly uncommon causes of invasive disease, the atypical pattern D strains may have an enhanced capacity for gaining access to deeper tissue. Future studies that examine the genetic similarities and differences between atypical pattern D strains and the typical pattern D and E strains may uncover the virulence factors critical for the initial skin infection and those necessary for subsequent invasion into deeper tissue.

Most of the Canadian cases of invasive disease caused by the *emm59* strain involved bacteremia or cellulitis, although there were also several instances of necrotizing fasciitis.[6] The age distribution of invasive disease due to *emm59* was strikingly different from nonbullous impetigo, which affects primarily children (although adults also can have impetigo).[8] In the Canadian epidemic, there was a peak incidence of invasive disease involving infants, which dropped sharply by 2 years of age and rose again to a high level for patients 25 to 60 years of age.[6] In addition, there were several notable risk factors for invasive disease caused by the *emm59* strain, which were significant compared with all other GAS strains causing invasive disease within the very same communities. The risk factors for *emm59* invasive disease included alcohol abuse and illicit drug use, homelessness, and hepatitis C virus infection.[6] Thus, the *emm59* strain epidemic may have originated and/or been amplified within a disadvantaged host population.

Additional Genetic Data on emm*59 Emergence*

Data on multilocus sequencing typing (MLST) of GAS are available in a valuable resource compiled by investigators from throughout the world. Five *emm59*/ST172 isolates are listed in the MLST database (http://spyogenes.mlst.net; last accessed January 2, 2012). Of these, the oldest *emm59* isolate was recovered from an impetigo lesion in the U.S. in 1969. Three other *emm59* isolates originated from normally sterile tissue sites. The fifth *emm59*/ST172 isolate (strain 149405, a blood isolate from the Czech Republic collected in 2004) was tested for resistance to antibiotics; it is reported to be susceptible to erythromycin and clindamycin, but resistant to tetracycline. Fittipaldi et al[4] also report historic *emm59*/ST172 strains from Japan and Spain.

Another plausible hypothesis for the *emm59* epidemic is migration of an established clone into a new community, whereby the new host population and/or the new environment have risk factors that, in turn, enable the new migrant to grab a foothold and quickly spread. Tetracycline resistance in GAS is widespread among numerous *emm* types and typically arises via newly acquired resistance genes (as opposed to mutation of core genes). Nonetheless, no such genomic distinctions in antibiotic resistance genes were reported for the historic versus Canadian epidemic *emm59* strains,[4] raising the reasonable possibility that all of the *emm59* strains are tetracycline-resistant, like the sole *emm59* isolate reportedly tested after its recovery in the Czech Republic. Because tetracycline is not a recommended treatment for GAS infections, there are few recognized treatment failures, and resistance to this antibiotic often goes unnoticed. However, if a tetracycline-resistant strain migrates into a community having unusually high tetracycline usage, there may be enough positive selection pressure for the strain to reproduce and increase in numbers to exceed a critical threshold, thereby yielding sufficient penetration in the community to support its continual transmission to new hosts. Macrolide consumption is a key factor leading to increased prevalence of macrolide-resistant GAS,[20] which is often rooted in a heterogeneous set of genetically distinct clones. However, a sharp increase in prevalence of macrolide-resistant GAS can also be due to a single dominant clone.[21] Knowledge of the antibiotic resistance profiles of the Canadian *emm59* strains is critical information that can aid in a more thorough consideration of the possible mechanisms by which the epidemic arose, because the selective pressure exerted by antibiotics is among the strongest to which bacteria are subject.

It remains possible that the epidemic *emm59* clone existed elsewhere in the world for a long period of time, emerging to cause an epidemic in Canada because of unknown host or environmental risk factors that are uniquely specific to the *emm59* clone. Alternatively, the initial rise in prevalence of a migrant may be due to entirely random factors. It is also plausible that the contemporary *emm59* clone is not rare but, rather, is abundant in an undersampled region. Although surveillance for GAS is extensive, in many pockets of the world, particularly much of Africa and parts of Asia, there is a dearth of surveillance reports. Of the >38,000 isolates evaluated in a meta-analysis of GAS surveillance studies,[5] the vast majority of organisms (>32,000) originate from countries with market economies. Another possible scenario is that the *emm59* clone persisted in a region that does indeed

undergo frequent surveillance, but that it tended not to cause an extensive degree of severe disease because the host and/or environmental risk factors did not align. If the *emm59* clone that caused the Canadian epidemic was a migrant, it may not be ancestral to the historic isolates from Japan, Spain, and the U.S. that were analyzed by Fittipaldi et al.[4] However, from the standpoint of using genomic analysis to map genetic differences that may have given rise to differences in virulence properties, the direction of the ancestral-descendant pathway probably does not matter.

The genome sequencing of ~600 *emm59* GAS isolates clearly shows the epidemic spread of a single clone over a very short period of evolutionary time.[4] Genes previously identified in organisms of other *emm* types as being highly variable, such as *emm* type, which is often under intense immune selection favoring genetic variants display unusual genetic stability among the *emm59* isolates. The genes *covS* and *rgg*, which encode global transcriptional regulators that affect the pharyngitis TP to invasive TP transition are also highly variable in other GAS strains, yet display unusual genetic stability among the *emm59* isolates. On the other hand, genetic changes in clustered regularly interspaced short palindromic repeat elements (CRISPR elements) helped confirm the genotype that was most likely present at the onset of the epidemic. Another surprise was the relatively high number of SNPs in another transcriptional regulatory gene (*mga*) of GAS. The phylogenetic tree based on concatenated SNPs reveals deep branches between each of the historic isolates collected from three distant regions of the world and the recent Canadian epidemic isolates. Compared with the genomes of GAS of numerous additional *emm* types, however, the relative genetic distance between the historic and contemporary *emm59* isolates is miniscule.

Concluding Remarks

If a genetic change helped spark the *emm59* epidemic in Canada, the genetic differences between the historic and contemporary isolates require closer evaluation. At least during initial analysis, the strongest candidates will be variant genes whose role in GAS virulence is already well established. For an optimal experimental test, genetic swap mutants could be constructed whereby the genetic variation present in the epidemic strain is introduced into a historic strain and vice versa. Using appropriate animal models, a finding of increased virulence by the modified historic strain, and decreased virulence by the modified contemporary strain, would support cause and effect and thus would close the circle on the genetic change hypothesis.

References

1. Shulman ST, Tanz RR, Dale JB, Beall B, Kabat W, Kabat K, Cederlund E, Patel D, Rippe J, Li Z, Sakota V; North American Streptococcal Pharyngitis Surveillance Group: Seven-year surveillance of North American pediatric group A streptococcal pharyngitis isolates. Clin Infect Dis 2009, 49:78–84
2. McGregor K, Bilek N, Bennett A, Kalia A, Beall B, Carapetis J, Currie B, Sriprakash K, Spratt B, Bessen D: Group A streptococci from a remote community have novel multilocus genotypes but share emm types and housekeeping alleles with isolates from worldwide sources. J Infect Dis 2004, 189:717–723
3. Sakota V, Fry AM, Lietman TM, Facklam RR, Li ZY, Beall B: Genetically diverse group A streptococci from children in far-western Nepal share high genetic relatedness with isolates from other countries. J Clin Microbiol 2006, 44:2160–2166
4. Fittipaldi N, Beres SB, Olsen RJ, Kapur V, Shea PR, Watkins ME, Cantu CC, Laucirica DR, Jenkins L, Flores AR, Lovgren M, Ardanuy C, Liñares J, Low DE, Tyrrell GJ, Musser JM: Full-genome dissection of an epidemic of severe invasive disease caused by a hypervirulent, recently emerged clone of group A Streptococcus. Am J Pathol 2012, 180:1522–1534
5. Steer AC, Law I, Matatolu L, Beall BW, Carapetis JR: Global emm type distribution of group A streptococci: systematic review and implications for vaccine development. Lancet Infect Dis 2009, 9:611–616
6. Tyrrell GJ, Lovgren M, St Jean T, Hoang L, Patrick DM, Horsman G, Van Caeseele P, Sieswerda LE, McGeer A, Laurence RA, Bourgault AM, Low DE: Epidemic of group A Streptococcus M/emm59 causing invasive disease in Canada. Clin Infect Dis 2010, 51:1290–1297
7. Brueggemann AB, Pai R, Crook DW, Beall B: Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog 2007, 3:e168
8. Carapetis JR, Steer AC, Mulholland EK, Weber M: The global burden of group A streptococcal diseases. Lancet Infect Dis 2005, 5:685–694
9. Ferrieri P, Dajani AS, Wannamaker LW, Chapman SS: Natural history of impetigo. I. Site sequence of acquisition and familial patterns of spread of cutaneous streptococci. J Clin Invest 1972, 51:2851–2862
10. Shelburne SA 3rd, Granville C, Tokuyama M, Sitkiewicz I, Patel P, Musser JM: Growth characteristics of and virulence factor production by group A Streptococcus during cultivation in human saliva. Infect Immun 2005, 73:4723–4731
11. Sumby P, Whitney AR, Graviss EA, DeLeo FR, Musser JM: Genome-wide analysis of group a streptococci reveals a mutation that modulates global phenotype and disease specificity. PLoS Pathog 2006, 2:e5
12. Cole JN, Barnett TC, Nizet V, Walker MJ: Molecular insight into invasive group A streptococcal disease. Nat Rev Microbiol 2011, 9:724–736
13. Ikebe T, Ato M, Matsumura T, Hasegawa H, Sata T, Kobayashi K, Watanabe H: Highly frequent mutations in negative regulators of multiple virulence genes in group A streptococcal toxic shock syndrome isolates. PLoS Pathog 2010, 6:e1000832
14. Wannamaker LW: Differences between streptococcal infections of the throat and of the skin. I. N Engl J Med 1970, 282:23–31
15. Bessen DE, Kumar N, Hall GS, Riley DR, Luo F, Sabharwal H, Tengra FK, Lizano S, Ford CN, McShan WM, Nyugen SV, Dunning Hotopp JC, Tettelin H: Genetic linkage mapping of determinants of tissue tropisms in group A Streptococcus. J Bacteriol 2011, 193:6651–6663
16. McGregor KF, Spratt BG, Kalia A, Bennett A, Bilek N, Beall B, Bessen DE: Multi-locus sequence typing of Streptococcus pyogenes representing most known emm types and distinctions among subpopulation genetic structures. J Bacteriol 2004, 186:4285–4294
17. Bessen DE, McGregor KF, Whatmore AM: Relationships between emm and multilocus sequence types within a global collection of Streptococcus pyogenes. BMC Microbiol 2008, 8:59
18. McGregor KF, Spratt BG: Identity and prevalence of multilocus sequence typing-defined clones of group A streptococci within a hospital setting. J Clin Microbiol 2005, 43:1963–1967
19. Wasserzug O, Valinsky L, Klement E, Bar-Zeev Y, Davidovitch N, Orr N, Korenman Z, Kayouf R, Sela T, Ambar R, Derazne E, Dagan R, Zarka S: A cluster of ecthyma outbreaks caused by a single clone of invasive and highly infective Streptococcus pyogenes. Clin Infect Dis 2009, 48:1213–1219
20. SeppälÄ H, Klaukka T, Vuopio-Varkila J, Muotiala A, Helenius H, Lager K, Huovinen P: The effect of changes in the consumption of macrolide antibiotics on erythromycin resistance in group A streptococci in Finland. Finnish Study Group for Antimicrobial Resistance. N Engl J Med 1997, 337:441–446
21. Freeman AF, Shulman ST: Macrolide resistance in group A Streptococcus. Pediatr Infect Dis J 2002, 21:1158–1160