JCI The Journal of Clinical Investigation

A molecular trigger for intercontinental epidemics of group A *Streptococcus*

Luchang Zhu, Randall J. Olsen, Waleed Nasser, Stephen B. Beres, Jaana Vuopio, Karl G. Kristinsson, Magnus Gottfredsson, Adeline R. Porter, Frank R. DeLeo, James M. Musser

J Clin Invest. 2015;125(9):3545-3559. https://doi.org/10.1172/JCI82478.

Research Article **Infectious disease**

The identification of the molecular events responsible for strain emergence, enhanced virulence, and epidemicity has been a long-pursued goal in infectious diseases research. A recent analysis of 3,615 genomes of serotype M1 group A *Streptococcus* strains (the so-called “flesh-eating” bacterium) identified a recombination event that coincides with the global M1 pandemic beginning in the early 1980s. Here, we have shown that the allelic variation that results from this recombination event, which replaces the chromosomal region encoding secreted NADase and streptolysin O, is the key driver of increased toxin production and enhanced infection severity of the M1 pandemic strains. Using isoallelic mutant strains, we found that 3 polymorphisms in this toxin gene region increase resistance to killing by human polymorphonuclear leukocytes, increase bacterial proliferation, and increase virulence in animal models of pharyngitis and necrotizing fasciitis. Genome sequencing of an additional 1,125 streptococcal strains and virulence studies revealed that a highly similar recombinational replacement event underlies an ongoing intercontinental epidemic of serotype M89 group A *Streptococcus* infections. By identifying the molecular changes that enhance upper respiratory tract fitness, increased resistance to innate immunity, and increased tissue destruction, we describe a mechanism that underpins epidemic streptococcal infections, which have affected many millions of people.

A molecular trigger for intercontinental epidemics of group A *Streptococcus*

Luchang Zhu,[1] Randall J. Olsen,[1,2] Waleed Nasser,[1] Stephen B. Beres,[1] Jaana Vuopio,[3,4] Karl G. Kristinsson,[5,6] Magnus Gottfredsson,[5,6] Adeline R. Porter,[7] Frank R. DeLeo,[7] and James M. Musser[1,2]

[1]Center for Molecular and Translational Human Infectious Diseases Research, Department of Pathology and Genomic Medicine, Houston Methodist Research Institute and Houston Methodist Hospital, Houston, Texas, USA. [2]Department of Pathology and Laboratory Medicine, Weill Cornell Medical College, New York, New York, USA. [3]Bacterial Infections Unit, National Institute for Health and Welfare, Turku, Finland. [4]Department of Medical Microbiology and Immunology, Medical Faculty, University of Turku, Turku, Finland. [5]Departments of Clinical Microbiology and Infectious Disease, Landspitali University Hospital, Reykjavik, Iceland. [6]Faculty of Medicine, School of Health Sciences, University of Iceland, Reykjavik, Iceland. [7]Laboratory of Bacteriology, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, NIH, Hamilton, Montana, USA.

The identification of the molecular events responsible for strain emergence, enhanced virulence, and epidemicity has been a long-pursued goal in infectious diseases research. A recent analysis of 3,615 genomes of serotype M1 group A *Streptococcus* strains (the so-called "flesh-eating" bacterium) identified a recombination event that coincides with the global M1 pandemic beginning in the early 1980s. Here, we have shown that the allelic variation that results from this recombination event, which replaces the chromosomal region encoding secreted NADase and streptolysin O, is the key driver of increased toxin production and enhanced infection severity of the M1 pandemic strains. Using isoallelic mutant strains, we found that 3 polymorphisms in this toxin gene region increase resistance to killing by human polymorphonuclear leukocytes, increase bacterial proliferation, and increase virulence in animal models of pharyngitis and necrotizing fasciitis. Genome sequencing of an additional 1,125 streptococcal strains and virulence studies revealed that a highly similar recombinational replacement event underlies an ongoing intercontinental epidemic of serotype M89 group A *Streptococcus* infections. By identifying the molecular changes that enhance upper respiratory tract fitness, increased resistance to innate immunity, and increased tissue destruction, we describe a mechanism that underpins epidemic streptococcal infections, which have affected many millions of people.

Introduction

One elusive goal in infectious diseases research is identification of the genetic changes and molecular mechanisms responsible for strain emergence, enhanced virulence, and epidemicity. This goal has practical importance because of its impact on humans, livestock, and plant health and on economies. However, with the exception of certain phenomena, such as antimicrobial agent resistance, the mechanisms of pathogen emergence and enhanced virulence are not understood for most microbes.

Group A *Streptococcus* (GAS), a strict human pathogen, has served as a powerful model organism for studying epidemic bacterial disease (1–3). For approximately 100 years, GAS strains have been categorized on the basis of serologic diversity caused by amino acid changes in the amino terminus of an antiphagocytic cell surface molecule known as M protein (4). Classification of strains based on this serologic typing scheme led to the generally accepted concept that M types represent genetically uniform groups. One important consequence of this notion was that well-described temporal fluctuations in frequency of infections caused by strains of particular streptococcal M protein serotypes (5, 6) were interpreted as being caused by essentially identical organisms. However, large-scale comparative genome sequencing has revealed that streptococcal epidemics involve clonal replacement events rather than reemergence of previously extant clones (1–3).

Moreover, advances in DNA sequencing now allow delineation of key molecular events responsible for generating new epidemic-producing clones. This was recently shown for streptococci by sequencing the genomes of 3,615 serotype M1 strains obtained from comprehensive population-based studies conducted in the US, Canada, Denmark, Finland, Iceland, Norway, and Sweden (3). The culminating and final major molecular genetic event that produced a new streptococcal M1 clone was allelic replacement by recombination of a 36-kb region encoding two actively secreted and potent toxins, NAD^+-glycohydrolase (*Streptococcus pyogenes* NADase [SPN]) and streptolysin O (SLO) (3). This horizontal gene transfer event was estimated by evolutionary genetic dating methods to have occurred in approximately 1983, a time that immediately preceded the onset of a serotype M1 global pandemic that has affected millions of humans (3). However, the molecular pathogenesis processes underlying how this ultimate genetic event triggered global spread of progeny of a single bacterial cell and produced a striking increase in infection frequency and severity remain unknown.

Of particular interest, in attempting to understand streptococcal strain emergence and epidemicity, is the replaced genomic area encoding *nga* and *slo* the genes for SPN and SLO, respectively. Multiple roles in invasive infections have been attributed to SPN.

Conflict of interest: The authors have declared that no conflict of interest exists.
Submitted: April 22, 2015; **Accepted:** July 2, 2015.
Reference information: *J Clin Invest*. 2015;125(9):3545–3559. doi:10.1172/JCI82478.

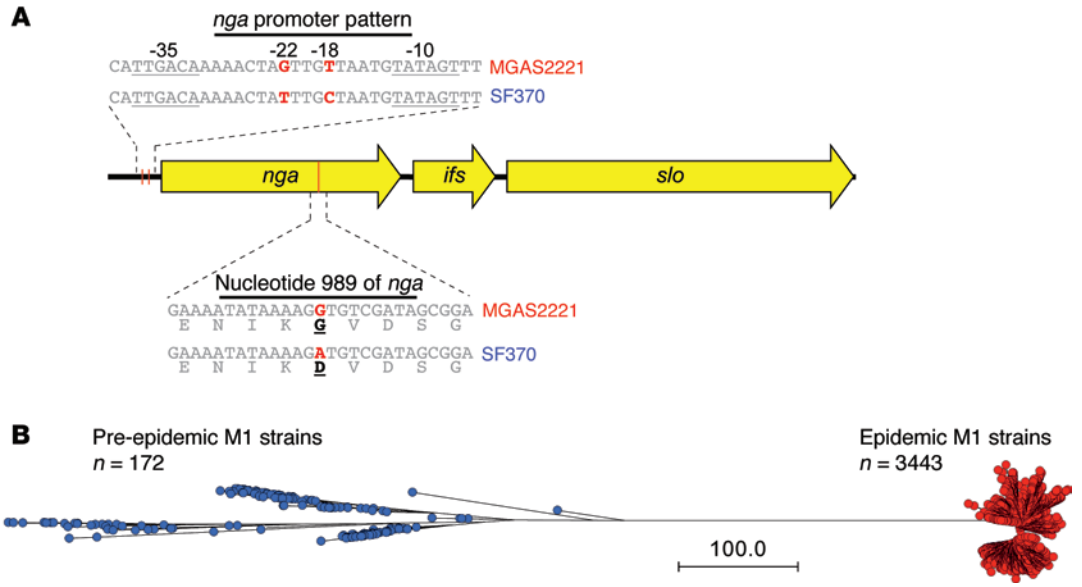

Figure 1. *nga*-region target SNPs and their pattern distribution among GAS M1 pre-epidemic and epidemic strains. (**A**) Schematic of the noncoding SNPs at positions –22 and –18 of the promoter spacer region of the *nga*, *ifs*, and *slo* gene cluster and the nonsynonymous coding SNP located at nucleotide 989 of *nga* that results in either glycine or aspartate at residue 330. (**B**) Genetic relationships among 3,615 M1 strains inferred by neighbor joining based on 13,221 core chromosomal SNPs as previously described (3). All pre-epidemic strains have the same *nga-slo* target SNP pattern as SF370, and all epidemic strains have the same pattern as MGAS2221.

SPN enhances GAS survival by inhibiting pathogen internalization by host cells and also augments SLO cytotoxicity (7–10). SLO is a potent oxygen-sensitive cytolytic toxin that forms pores in host-cell membranes (10). The coordinated activities of SPN and SLO prevent maturation of phagolysosomes and thereby decrease phagocytic killing of GAS (9, 11).

Several lines of evidence suggest that the *nga* and *slo* genes may play a critical role in M1 epidemicity. First, the transcript levels of *nga* and *slo* are significantly higher in epidemic strains than in pre-epidemic M1 GAS strains (12). Second, epidemic M1 strains produce more SPN and SLO activity than pre-epidemic strains (12, 13). By comparing the genome sequences of a genetically representative epidemic (MGAS2221) strain and a pre-epidemic (SF370) strain, we discovered that the two strains differ by 59 SNPs and 2 indels in the *nga-slo* region of recombination (3). Among these SNPs, 3 were of particular interest because they are located in regions that, in principle, might influence *nga* and *slo* transcript level and strain phenotype. Two SNPs are located in the –35 to –10 spacer region of the *nga* promoter, at positions –22 and –18, designated as G-22T and T-18C, respectively (Figure 1A). The third SNP of interest (G in epidemic strain MGAS2221 and A in pre-epidemic strain SF370) is located at nucleotide 989 in *nga* and results in a Gly330Asp amino acid replacement in SPN (Figure 1A). This amino acid replacement has been reported to alter NADase activity (14).

We hypothesized that one or more of these 3 SNPs described above bestowed a fitness advantage in the upper respiratory tract as well as increased virulence (9), thereby triggering the contemporary epidemic of severe invasive M1 GAS infections. Isoallelic mutant strains and in vivo and ex vivo virulence assays were used to test this hypothesis. The sum of the resulting data is fully consistent with our hypothesis. Unexpectedly, based on genome sequencing and pathogenesis studies, we discovered that an analogous recombinational molecular genetic event resulted in upregulation of production of extracellular NADase and SLO toxins and recently stimulated an epidemic of serotype M89 GAS infections caused by a recently emerged clone.

Results

Pre-epidemic and epidemic serotype M1 strains. Mouse invasive infection and nonhuman primate pharyngitis and invasive infection studies show that the MGAS2221 strain, genetically representative of the epidemic ("new") clone, is more virulent than reference strain SF370, genetically representative of pre-epidemic ("old") serotype M1 strains (3). Strain MGAS2221 lacks polymorphisms in regulatory genes known to influence virulence, such as *covRS*, *ropB*, and *mga*. A previous expression microarray analysis of the transcriptomes of pre-epidemic and epidemic M1 strains found that only 8 core chromosomal genes were differentially transcribed between the two strain groups in early logarithmic growth phase (12). Given the clear evidence that pre-epidemic and epidemic M1 strains differ significantly in virulence, we tested the hypothesis that the expression microarray analysis had failed to identify critical virulence genes differentially regulated between the two strain groups. We reassessed global transcriptome differences in a panel of pre-epidemic and epidemic M1 strains using RNA sequencing (RNA-Seq) analysis, a more sensitive and accurate method of transcript analysis. RNA-Seq analysis detected 5 differentially expressed genes, confirming our previous expression microarray findings of very limited gene transcript differences between pre-epidemic and epidemic strains. Of particular interest, all 5 differentially regulated genes were located in the 36-kb region of recombinational replacement. These genes include *nga*, *ifs* (encoding Ifs, an endogenous inhibitor of SPN; refs. 15, 16), *slo*, and two small genes of uncertain function (Spy0142 and Spy0144;

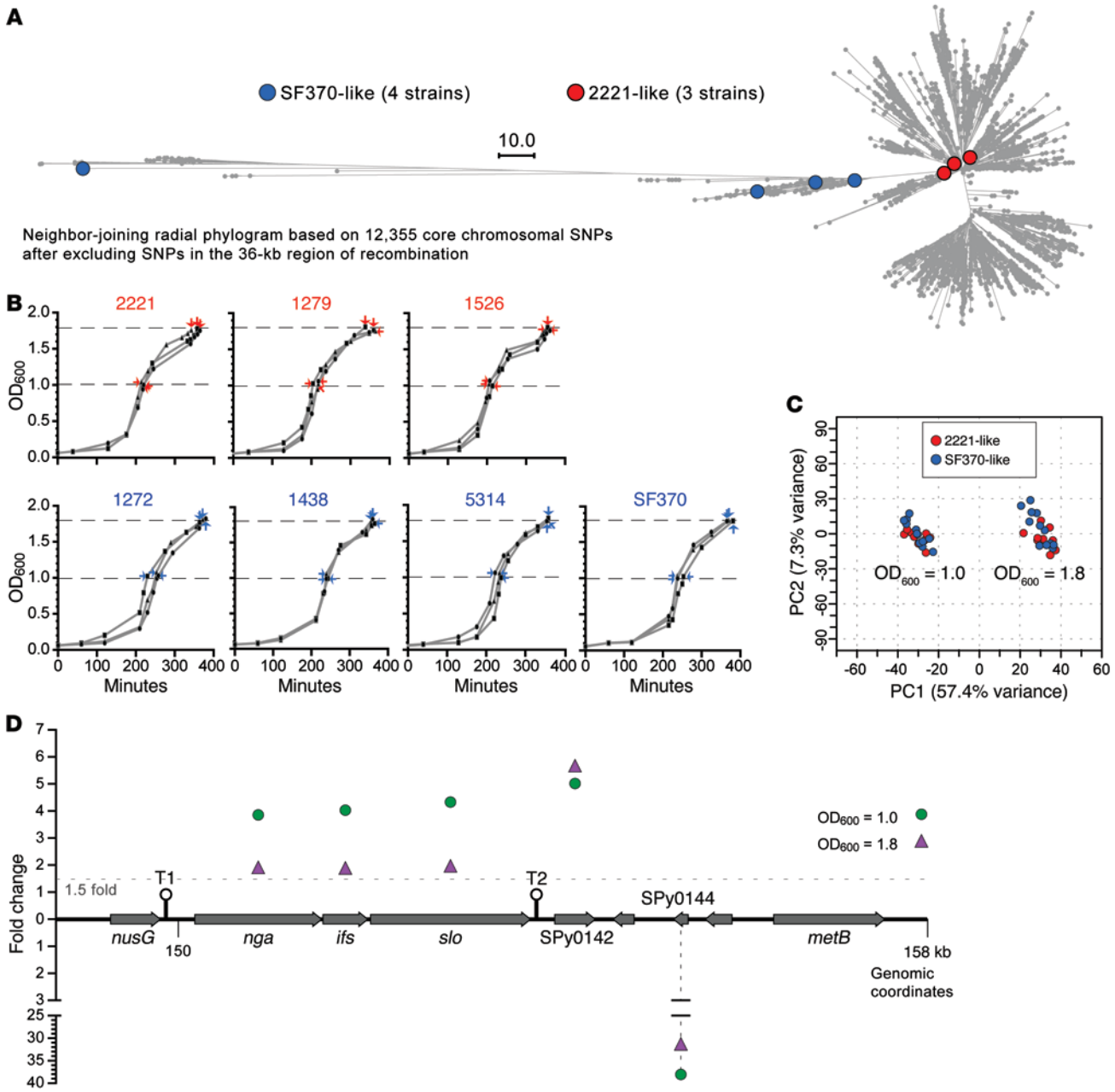

Figure 2. RNA-seq analysis of GAS serotype M1 pre-epidemic and epidemic strains. (**A**) Phylogenetic tree showing genetic relationships among strains used in the RNA-seq experiments. (**B**) Growth curves of strains used for RNA-seq experiments. The RNA-seq analyses were done in triplicate, and growth curves for the strains are similar. Cells were harvested at OD_{600} values of 1.0 and 1.8. MGAS strain collection strain designations are given above the growth curves in blue for pre-epidemic strains and in red for epidemic strains. (**C**) RNA-seq transcriptome data principal component analysis plot. Plotted are expression relationships among the strains based on principal component 1 (PC1) and principal component 2 (PC2), which account for the two largest unrelated variances in the data, 57.4% and 7.3% respectively. The pre-epidemic and epidemic strains are closely clustered, consistent with very limited transcriptome variance between them at both growth points. (**D**) Schematic showing the 5 genes markedly differentially expressed between pre-epidemic and epidemic stains at both OD_{600} of 1.0 (green dots) and 1.8 (purple triangles). All 5 differentially expressed genes are located in the 36-kb region of recombinational replacement. Differences in expression are plotted as the fold change of mean normalized transcript levels of 3 biological replicates for the 3 epidemic strains relative to the 4 pre-epidemic strains assessed. Transcript levels for *nga*, *ifs*, and *slo* are all increased by 1.5-fold or greater for the epidemic strains relative to the pre-epidemic strains at both stages of growth. Inferred transcriptional terminators downstream of *nusG* and *slo* are indicated as T1 and T2, respectively (17).

Figure 2D). *nga*, *ifs*, and *slo* comprise a single transcription unit (17). A transcriptional terminator (denoted T2 in Figure 2D) that separates the *nga*, *ifs*, and *slo* operon and Spy0142 (17) is located downstream of *slo*. All differentially expressed genes were upregulated in the epidemic strains, with the exception of Spy0144, which was downregulated (Figure 2D).

Inasmuch as *nga* and *slo* encode secreted toxin virulence factors, and none of the other differentially regulated genes

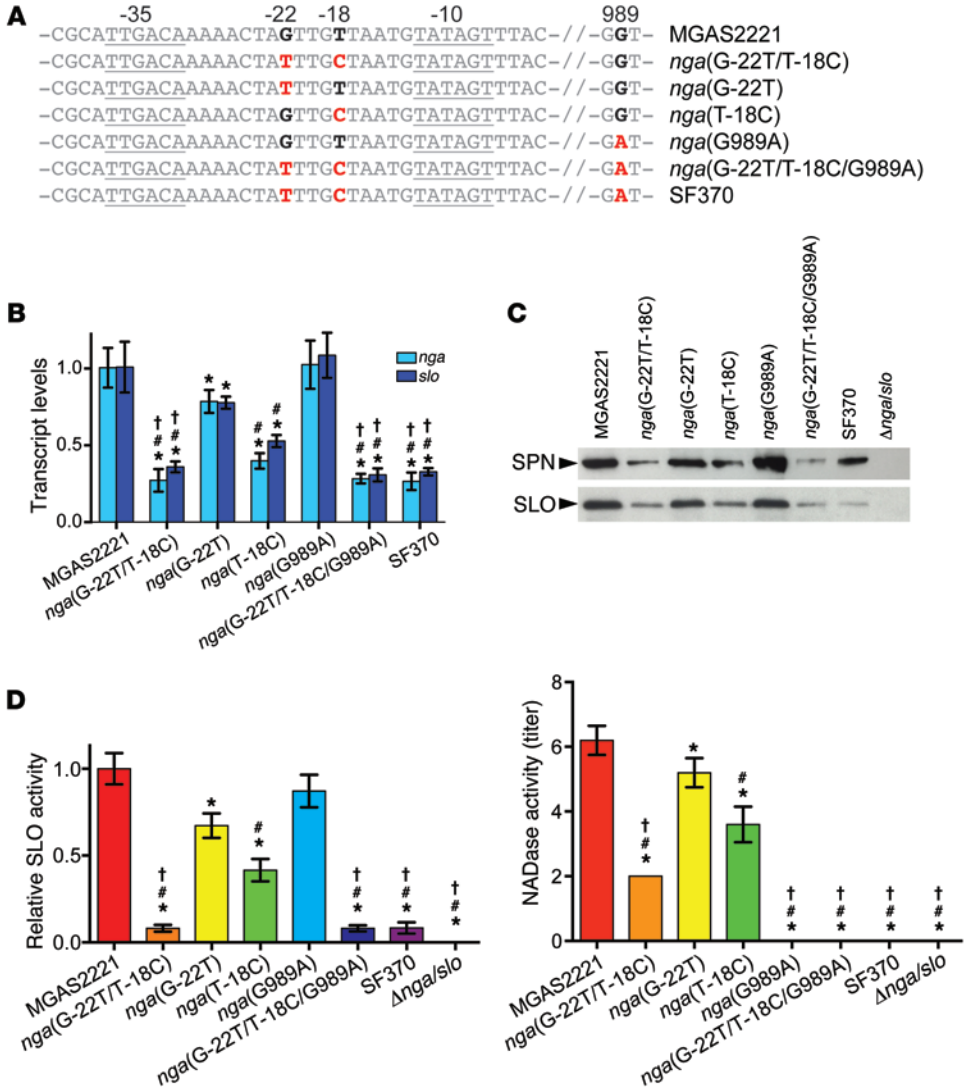

Figure 3. In vitro characteristics of GAS M1 reference and isoallelic mutant strains. (**A**) Schematic showing the target SNP genotype of the strains assayed. (**B**) TaqMan qRT-PCR analysis of *nga* and *slo* transcript levels relative to MGAS2221 levels. (**C**) Western immunoblot analysis of secreted SPN and SLO in culture supernatants. (**D**) SLO cytolytic activity relative to MGAS2221 and SPN NADase activity. Transcript, cytolysis, and NADase assays were performed in triplicate on 3 separate occasions. Replicate data are expressed as the mean ± SD in **B** and **D**. $^{*}P < 0.05$ vs. MGAS2221; $^{\#}P < 0.05$ vs. *nga*(G-22T); $^{\dagger}P < 0.05$ vs. *nga*(T-18C), 2-tailed Student's *t* test; $n = 3$.

encodes a product known to be involved in GAS pathogenesis, we sought the molecular cause underlying upregulation of these two genes. The *nga*, *ifs*, and *slo* transcripts were upregulated to an equivalent magnitude, consistent with the genes being expressed from a single polycistronic transcript (15–17). In principle, it is possible that nucleotide sequence changes in a regulatory region upstream of *nga* are responsible for the increased transcription observed in the epidemic strains. Alternatively, insertional inactivation of a hypothetical negative regulator or acquisition of an unknown trans-acting positive regulator could cause a temporal shift in SPN-producing phenotype (13). A genomic rearrangement that affects expression of the *slo* gene also has been reported (18).

To differentiate among these possibilities, we interrogated the 3,615 serotype M1 genomes for polymorphisms that might be expected to inactivate a negative gene regulator. We also searched the genomes for evidence of acquisition of a trans-acting positive regulator gene or novel genomic rearrangements that might alter gene expression. We identified no evidence of these types of genetic alterations that would account for the shift to increased expression of *nga*, *ifs*, and *slo*.

We next analyzed the 3,615 genome sequences for polymorphisms in the upstream promoter region of the *nga* gene that, in principle, could alter transcription. Regardless of geographic origin, all epidemic strains had the same allele (SNPs G-22T and T-18C) located in the promoter (Figure 1). Inasmuch as sequence variation in promoter spacer regions may alter transcript levels in bacteria (19, 20), we hypothesized that these two SNPs caused the difference in *nga*, *ifs*, and *slo* transcript levels between pre-epidemic and epidemic strains.

To test this hypothesis, we constructed a series of isoallelic mutant strains from wild-type epidemic reference strain MGAS2221 (Supplemental Figure 1; supplemental material available online with this article; doi:10.1172/JCI82478DS1). The MGAS2221 parental strain was used because it lacks function-altering polymorphisms in known regulatory genes, its genome has been sequenced to very high accuracy, and it has been used in animal virulence studies (3). In addition, the strain is located at the root of the phylogenetic tree branch, giving rise to all new epidemic M1 strains (Figure 2A). We expected to find that introduction of either or both of the spacer region SNPs into strain MGAS2221 would significantly decrease the level of *nga* and *slo*

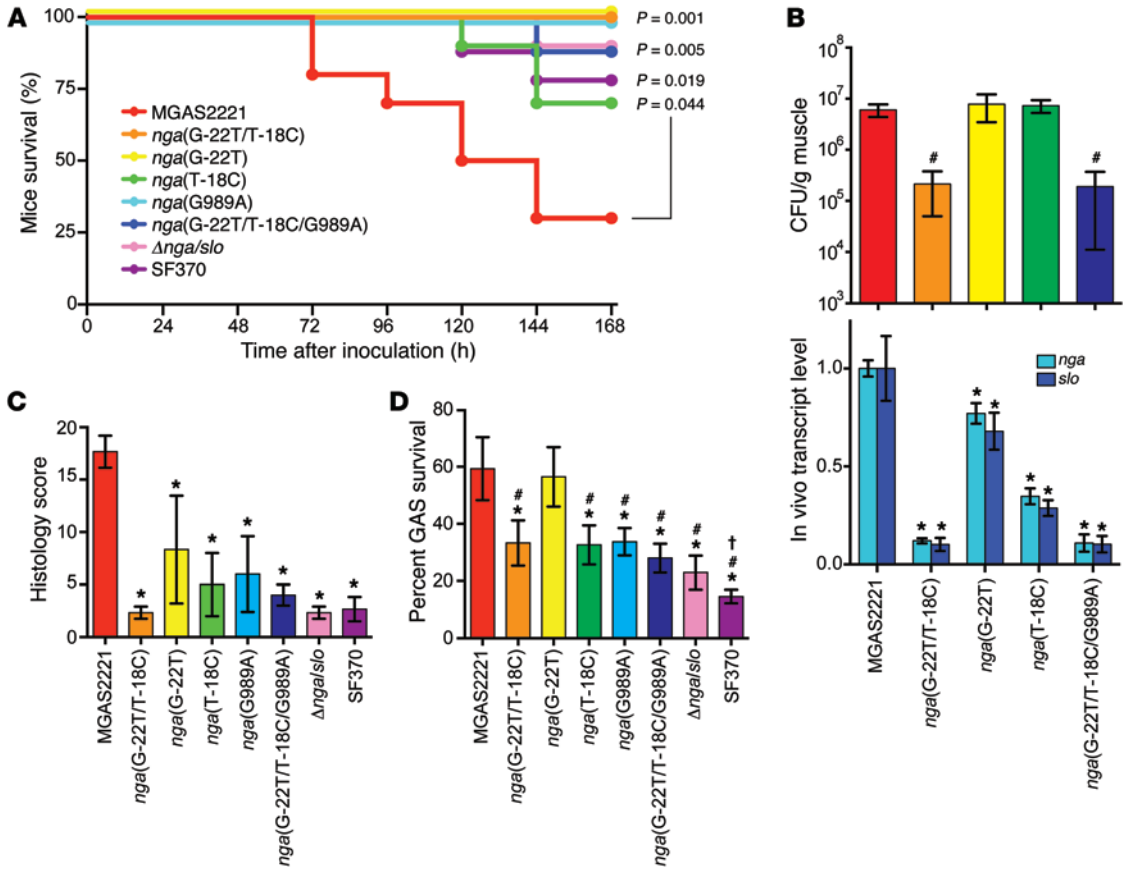

Figure 4. Virulence features of GAS M1 reference and isoallelic mutant strains. (**A**) Kaplan-Meier survival curves for mice (n = 10 per strain) inoculated intramuscularly with reference and mutant strains as indicated. P values were determined using the log-rank test. (**B**) GAS CFUs per gram of infected hind limb at 96 hours after infection (n = 20 per strain) and relative transcript levels (n = 3 per strain) of *nga* and *slo* extracted from infected hind limbs at 96 hours after infection. CFU data are expressed as the mean ± SEM. Transcript data are expressed as the mean ± SD. $^{\#}P < 0.0044$, Mann-Whitney test. $^{*}P < 0.05$, Student's t test. (**C**) Tissue damage histology scores derived from microscopic examination by a pathologist blinded to infecting strain. Replicate scores are expressed as mean ± SEM (n = 3 per strain). $^{*}P < 0.05$, Mann-Whitney test. (**D**) GAS resistance to the bactericidal activity of purified human PMNs was assessed with a suspension phagocytosis assay using PMNs from 6 donors. Shown are percentages of GAS survival after a 3-hour PMN exposure. Survival data are expressed as the mean ± SEM. $^{*}P < 0.05$ vs. MGAS2221; $^{\#}P < 0.05$ vs. *nga*(G-22T); $^{\dagger}P < 0.05$ vs. *nga*(G-22T/T-18C), repeated-measures 1-way ANOVA and Tukey's post test.

transcripts made by the mutated strain to that made by the pre-epidemic strains and result in decreased strain virulence.

Allelic variation in the promoter spacer region alters toxin gene transcript levels and toxin production. Generation of the G-22T and T-18C SNPs together in parental strain MGAS2221 significantly decreased the transcript level of *nga* and *slo* (Figure 3, A and B, and Supplemental Figure 1). The transcript levels of *nga* and *slo* made by isoallelic mutant strain *nga*(G-22T/T-18C) and strain SF370 (which naturally contains these two polymorphisms) did not differ significantly (Figure 3B).

We next studied the ability of each of these two SNPs in isolation to significantly decrease *nga* and *slo* transcription by creating isoallelic mutant strains from strain MGAS2221 that contain either G-22T or T-18C individually (Figure 3A and Supplemental Figure 1). Compared with parental strain MGAS2221, the transcript levels of *nga* and *slo* in strain *nga*(G-22T) were significantly decreased (Figure 3B). Moreover, the transcript levels of *nga* and *slo* in mutant strain *nga*(T-18C) were decreased to a significantly greater magnitude (Figure 3B). Introduction of the G989A SNP into the *nga*-coding region did not affect *nga* and *slo* transcript levels (Figure 3B). Taken together, these results indicated that (a) the two SNPs (G-22T and T-18C) in the *nga* promoter spacer region significantly influence the transcript levels of *nga* and *slo*, (b) lack of significant difference in *nga* and *slo* transcript levels between mutant strain *nga*(G-22T/T-18C) and strain SF370 demonstrates that these two nucleotide changes contribute to the increased transcript level of *nga* and *slo* in epidemic strains, and (c) the SNPs G-22T and T-18C each contribute to the altered expression of *nga* and *slo*, with the T-18C SNP exerting a greater effect than G-22T on transcript abundance.

Western immunoblot analysis of SPN and SLO in the culture supernatant confirmed that the G-22T and T-18C SNPs in the –35 and –10 spacer region substantially altered the amount of production of these two toxins (Figure 3C). In contrast, the amount of immunoreactive SPN and SLO in the culture supernatant was not influenced by the G989A SNP in the *nga*-coding region (Figure 3C). Consistent with previous data (12), strain SF370 produced very low levels of secreted SPN and SLO (Figure 3C). Moreover, the presence of G-22T and T-18C SNPs in the promoter spacer region resulted in significantly decreased SPN and SLO activity in

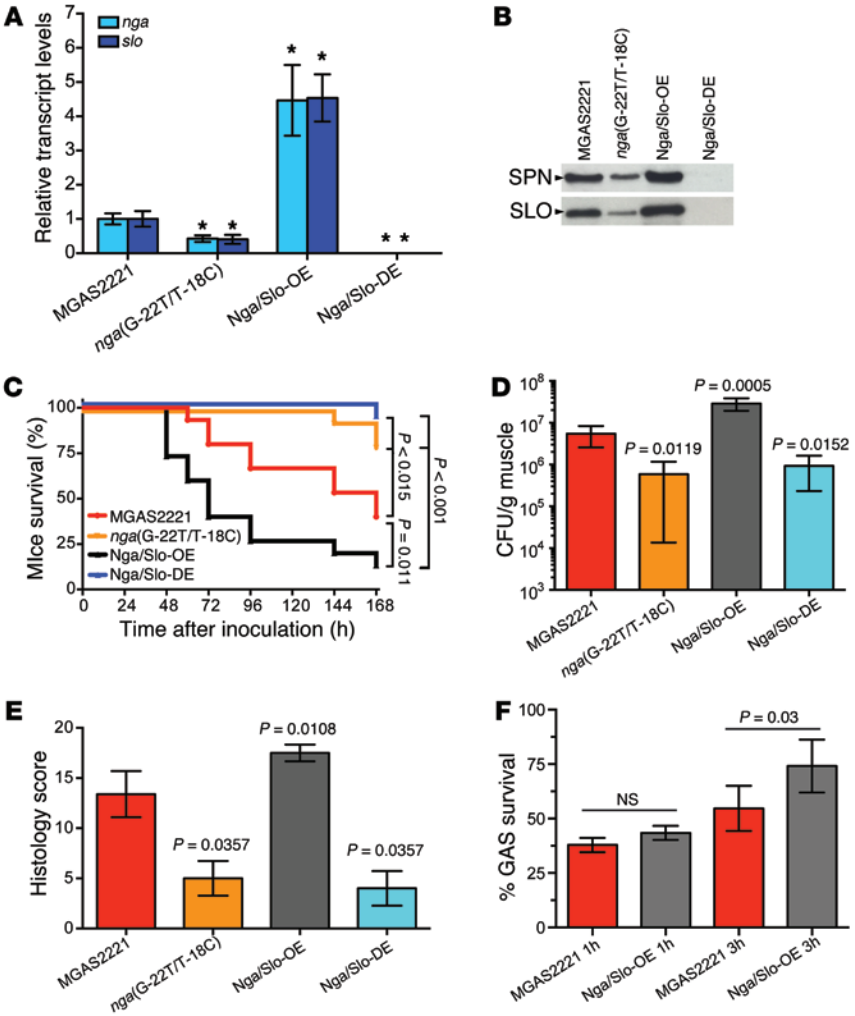

Figure 5. An isoallelic mutant GAS strain that overexpresses NADase and SLO is significantly more virulent than MGAS2221 in a mouse model of necrotizing fasciitis. (**A**) Transcript levels of *nga* and *slo* in MGAS2221, 2-SNP mutant *nga*(G-22T/T-18C), Nga/Slo-OE (overexpressing), and Nga/Slo-DE (2-nucleotide deletion) as determined by TaqMan qRT-PCR (n = 3 per strain). Data are expressed as the mean ± SD. *P < 0.05 vs. MGAS2221, Student's t test. (**B**) Western immunoblot analysis of NADase and SLO present in the culture supernatant of each strain. (**C**) Kaplan-Meier survival curves of mice (n = 15 per strain) infected with strain MGAS2221, *nga*(G-22T/T-18C), Nga/Slo-OE (overexpression), or Nga/Slo-DE (2-nucleotide deletion mutant). P values were determined using the log-rank test. (**D**) Bacterial load in mice infected with each strain. Shown are CFUs of GAS per gram of infected hind limb tissue at 96 hours after infection. Data are shown as mean ± SEM (n = 20 per strain). P values were calculated using the Mann-Whitney test. (**E**) Histopathology scores of mouse limbs infected with each strain (n = 3 per strain). Data are shown as mean ± SEM. P values were calculated using the Mann-Whitney test. (**F**) Percentage GAS survival after treatment with purified human PMNs. At 3 hours, the isoallelic mutant strain overexpressing SPN and SLO exhibited significantly greater survival than wild-type parental strain MGAS2221. Results are presented as the mean ± SEM for GAS survival using 13 separate human PMN donors. P values were determined by paired 2-tailed Student's t test.

the culture supernatant (Figure 3D). Introducing the G989A SNP (resulting in Gly330Asp replacement) into the *nga*-coding region abolished SPN activity, consistent with other reports (refs. 14, 21, and Figure 3D). However, this amino acid replacement did not alter SLO activity (Figure 3D). Compared with G-22T, the T-18C SNP contributed more to the attenuation of SPN and SLO activity. Thus, the amount of functional SPN activity secreted by M1 strains is influenced both by the coding region SNP (G989A SNP) and the two spacer region SNPs in the *nga* promoter. The isoallelic mutant strains did not differ from the epidemic parental strain MGAS2221 or the pre-epidemic strain SF370 in growth in rich medium, amount of capsular hyaluronic acid produced, or casein proteolytic activity (Supplemental Figure 2).

The 3 target SNPs significantly alter virulence. All isoallelic mutant strains with downregulated levels of SPN and/or SLO activity were significantly less virulent than wild-type strain MGAS2221 in a mouse model of necrotizing fasciitis (Figure 4). Like pre-epidemic strain SF370, all isoallelic mutant strains were significantly attenuated in virulence compared with wild-type reference strain MGAS2221. However, there were no significant differences in virulence among the mutant strains relative to each other or strain SF370 (Figure 4A). Compared with wild-type strain MGAS2221, the isoallelic mutant strains caused markedly smaller gross lesions at the hind limb injection site, with less tissue destruction (Supplemental Figure 3). The histopathological phenotypes of the *nga*(G-22T/T-18C) and *nga*(G-22T/T-18C/G989A) strains were indistinguishable from those of the Δ*nga/slo* mutant strain or pre-epidemic reference strain SF370 (Figure 4C). Strain *nga*(G-22T), which secretes an intermediate level of SPN and SLO, correspondingly had an intermediate histopathology phenotype (Figure 4C).

One explanation for the significantly decreased virulence of the isoallelic mutant strains posits decreased bacterial burden at the infection site. Thus, we measured the number of CFUs present in infected hind limbs. Bacterial loads in mouse limbs infected with isoallelic mutant strains *nga*(G-22T/T-18C) or *nga*(G-22T/T-18C/G989A) were significantly lower than in mice infected with wild-type strain MGAS2221 (Figure 4B). Interestingly, the bacterial loads in mice infected with single-SNP mutant strains *nga*(G-22T) and *nga*(T-18C) were not statistically different from animals given strain MGAS2221. A likely explanation is that strains with either SNP alone retain some level of in vivo production of SPN and SLO, as inferred from the transcript data (Figure 4B).

The data are consistent with the idea that the promoter region polymorphisms directly alter the transcript levels of *nga* and *slo* and thereby directly change the amount of extracellular SPN and SLO toxin production. The data imply that increasing the amount of SPN and SLO production by wild-type strain MGAS2221 would increase strain virulence. To test this idea, we constructed strain

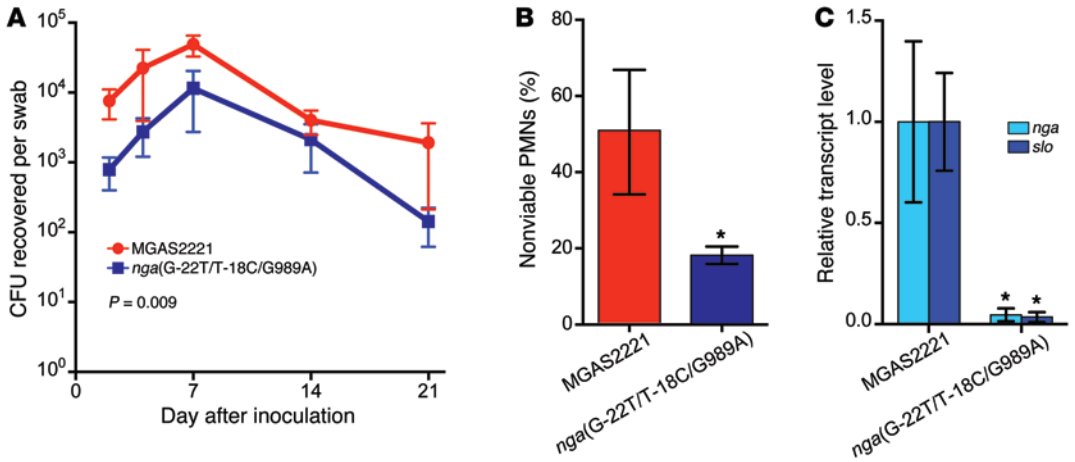

Figure 6. Epidemic parental strain MGAS2221 is significantly more virulent than a triple-SNP isoallelic mutant strain in a nonhuman primate model of upper respiratory tract infection. (**A**) Cynomolgus macaques (n = 4 per strain) were inoculated in the upper respiratory tract with epidemic parental wild-type strain MGAS2221 or triple-SNP isoallelic mutant strain *nga*(G-22T/T-18C/G989A). P values were assessed by repeated-measures 2-way ANOVA. Mean CFUs recovered from the oropharynx on each day are shown with SEM. (**B**) Viability of PMNs in throat swab samples of each strain group determined by flow cytometry. Data are shown as average percentage of nonviable PMNs ± SEM (n = 4 per strain). *P < 0.029, Mann-Whitney test. (**C**) Relative transcript levels of *nga* and *slo* in animals infected with strain MGAS2221 or *nga*(G-22T/T-18C/G989A) during pharyngitis, as determined by TaqMan qRT-PCR. Data are shown as mean ± SD (n = 4 per strain). *P < 0.003, Student's t test.

Nga/SLO-OE, an isogenic derivative strain that makes increased *nga* and *slo* transcript, resulting in increased production of secreted SPN and SLO toxins. This isogenic mutant strain was constructed by inserting the strong constitutive promoter of the GAS *tufA* gene directly upstream and adjacent to the native *nga* promoter (Supplemental Figure 1H). This isogenic mutant strain (Supplemental Figure 1H), with upregulation of the SPN and SLO virulence toxins, had increased mouse virulence and necrotizing fasciitis tissue destruction (Figure 5, A–E).

We next assessed the relative transcript levels of *nga* and *slo* during animal infection by performing TaqMan quantitative RT-PCR (qRT-PCR) on lesion samples. Animals infected with wild-type strain MGAS2221 had the highest levels of *nga* and *slo* transcripts (Figure 4B). All 4 isoallelic mutant strains had significantly decreased *nga* and *slo* transcript levels in vivo compared with the wild-type parental strain. The transcript levels of *nga* and *slo* in the *nga*(G-22T/T-18C) strain were only 12% and 10% of those made by wild-type strain MGAS2221 (Figure 4B). Single-SNP strains *nga*(G-22T) and *nga*(T-18C) had intermediate levels of *nga* and *slo* transcripts, with transcript levels made by strain *nga*(G-22T) being comparatively higher than strain T-18C (Figure 4B). In parallel with our in vitro data, *nga* and *slo* transcript levels in the triple-SNP mutant strain *nga*(G-22T/T-18C/G989A) were not statistically different from those of double-SNP mutant strain *nga*(G-22T/T-18C) (Figure 4B).

Inasmuch as the 3 target SNPs (G-22T, T-18C, and G989A) influence both the level of production and functional activities of SPN and SLO, toxins that detrimentally alter polymorphonuclear leukocyte (PMN) function (22, 23) and contribute to virulence, we hypothesized that these 3 SNPs influence killing of GAS by purified human PMNs. We found that wild-type strain MGAS2221 had the highest capacity to resist PMN-mediated killing (Figure 4D). With the exception of the strain with the G-22T SNP, which has an intermediate phenotype in level of *nga* and *slo* transcript expression, secreted toxin activity, and animal virulence, all isoallelic mutants were significantly more susceptible to PMN-mediated killing. Survival of mutant strains *nga*(G-22T/T-18C), *nga*(T-18C), and *nga*(G989A) was decreased to similar levels (Figure 4D). The survival of triple-SNP mutant strain *nga*(G-22T/T-18C/G989A) was slightly lower than that of *nga*(G-22T/T-18C) or *nga*(G989A), but the difference did not reach statistical significance (Figure 4D). As expected, the Δ*nga*/*slo* mutant strain and strain SF370 had significantly attenuated survival in these assays (Figure 4D). Consistent with these findings and our key hypothesis, an isoallelic mutant strain with significantly increased production of SPN and SLO toxins had significantly increased survival in the presence of human PMNs (Figure 5F).

We recently reported that pre-epidemic reference strain SF370 was significantly less virulent than epidemic reference strain MGAS2221 in nonhuman primate models of pharyngitis and necrotizing fasciitis, a finding consistent with human clinical and epidemiologic data (3, 24). However, because strains SF370 and MGAS2221 have many genetic differences other than the 3 target SNPs (3), we could not conclude that these SNPs are sufficient to recapitulate all differences in virulence. The cynomolgus macaque has been used extensively for upper respiratory tract studies because it closely mimics human-GAS molecular interactions (3). Next, we directly tested the role of the 3 target SNPs in pharyngitis in this model using the triple-SNP isoallelic mutant strain. We found that strain MGAS2221 was significantly more virulent than the triple-SNP isoallelic mutant strain *nga*(G-22T/T-18C/G989A), as assessed by bacterial burden recovered over the course of the experiment (Figure 6A). In addition, animals infected with the isoallelic mutant strain had a significantly lower level of nonviable PMNs in the posterior oropharynx (Figure 6B). TaqMan qRT-PCR confirmed that the level of *nga* and *slo* transcript present in situ in throat swab specimens was significantly lower in animals inoculated with the isoallelic mutant strain (Figure 6C).

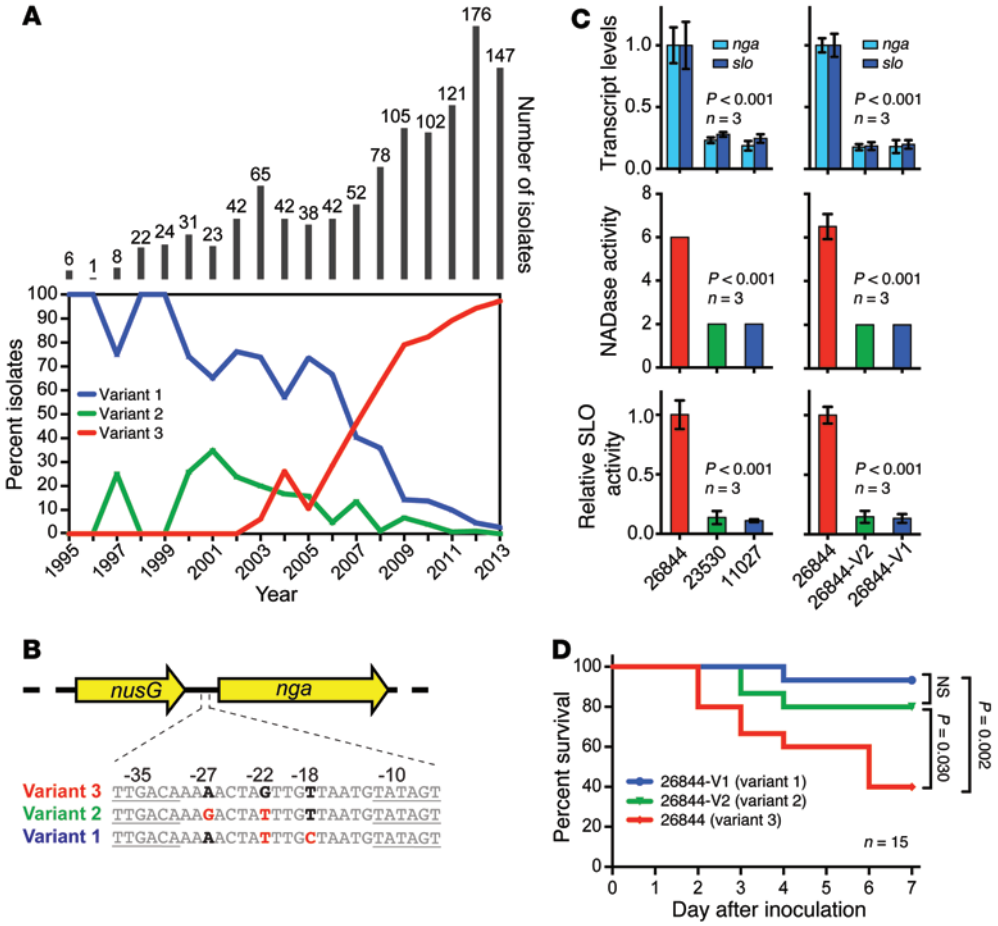

Figure 7. An analogous molecular event has triggered an intercontinental epidemic of infections caused by serotype M89 GAS. (**A**) Epidemic curve showing increase in frequency of infections caused by progeny of a new clone (variant 3) of serotype M89 strains. Data are from comprehensive population-based studies in the US, Finland, and Iceland. (**B**) Three *nga* promoter variants in the M89 strains. The genetically representative strains used (**C**) have variant 1 (MGAS11027), variant 2 (MGAS23530), and variant 3 (MGAS26844) patterns. (**C**) TaqMan qRT-PCR analysis of in vivo *nga* and *slo* transcript levels in the 3 strains. Transcript levels are displayed relative to the strain MGAS26844 (with variant 3) transcript level. Cultures were analyzed in triplicate on 3 separate occasions. NADase and SLO activity data for wild-type strain MGAS26844 (variant 3) and isogenic mutant strains with variant 2 (26844-V2) and variant 1 (26844-V1) *nga* promoter. Transcript data and NADase and SLO activity data are shown as mean ± SD (n = 3). Statistical significance was determined by 2-tailed Student's t test. (**D**) Mouse survival data. Shown are Kaplan-Meier survival curves. Mice (n = 15 per strain) were inoculated in the hind limb with epidemic parental wild-type strain MGAS26844 or isoallelic mutant strains. P values were determined using the log-rank test.

Recombinational replacement ignited a second, ongoing intercontinental epidemic. Our discovery that upregulation of the *nga*, *ifs*, and *slo* cytolytic toxin-encoding genes underpinned the contemporary serotype M1 epidemic raised the possibility that this is a common theme triggering GAS epidemics. Inspection of data generated in comprehensive population-based studies conducted over many years in Finland, Iceland, and the US (refs. 6, 25; http://www.cdc.gov/abcs/index.html) found that invasive infections caused by serotype M89 GAS increased substantially in frequency over the last 15 years (Figure 7A). We therefore hypothesized that, analogous to serotype M1 strains, upregulation of *nga* and *slo* may have contributed to the very recent upsurge in M89 infection frequency in these disparate countries. To test this idea, we first sequenced the promoter region upstream of *nga* in all 866 serotype M89 strains deposited in the US Centers for Disease Control and Prevention's Active Bacterial Core Surveillance Isolate Bank (http://www.cdc.gov/abcs/pathogens/isolatebank/). We identified (Figure 7B) 3 main variants of the *nga* promoter sequence in these strains. These 3 variants were also identified by analysis of comprehensive population-based M89 strain samples (n = 259 isolates) from human invasive infections in Finland and Iceland (Figure 7A). Importantly, one of these 3 variants (variant 3) is identical to the variant we identified in the epidemic serotype M1 strains (Figure 7B). Serotype M89 strains with this promoter spacer region variant increased substantially in frequency commensurate with the upsurge in M89 infections in Finland, Iceland, and the US (Figure 7A). Strains with this variant (that is, variant 3) had significantly increased amounts of NADase and SLO activity in culture supernatants (Figure 7C). In contrast with pre-epidemic serotype M1 strains, all M89 strains lack the "G989A" SNP in *nga* gene that inactivates the NADase activity of SPN. That is, serotype M89 strains produce enzymatically active SPN. Put another way, in the *nga-slo* region, variation in the *nga* promoter is the only factor that affects production of secreted NADase and SLO activity of M89 strains. Isoallelic mutant strains with each of the 3 promoter spacer region variants confirmed that the SNPs in the *nga-slo* upstream regulatory region are responsible for the increased gene transcripts, toxin production, and mouse virulence (Figure 7, C and D). A subsequent analysis of 1,125 M89 genome sequences revealed the presence of 3 major clades of M89 strains, designated clade-1, clade-2, and clade-3. Strains in each of the 3 clades have a unique *nga* promoter region sequence (Supplemental Figure 7). All clade-1 strains have variant 1 promoter region, all clade-2 strains have variant 2 promoter region, and all clade-3 strains have variant 3 promoter region (Supplemental Figure 7).

As with the serotype M1 strains, full-genome analysis of all 1,125 M89 strains found no other potential mechanism to account for increased SPN and SLO production. Only a small minority of the epidemic clade-3 strains (variant 3 *nga* promoter) have poly-

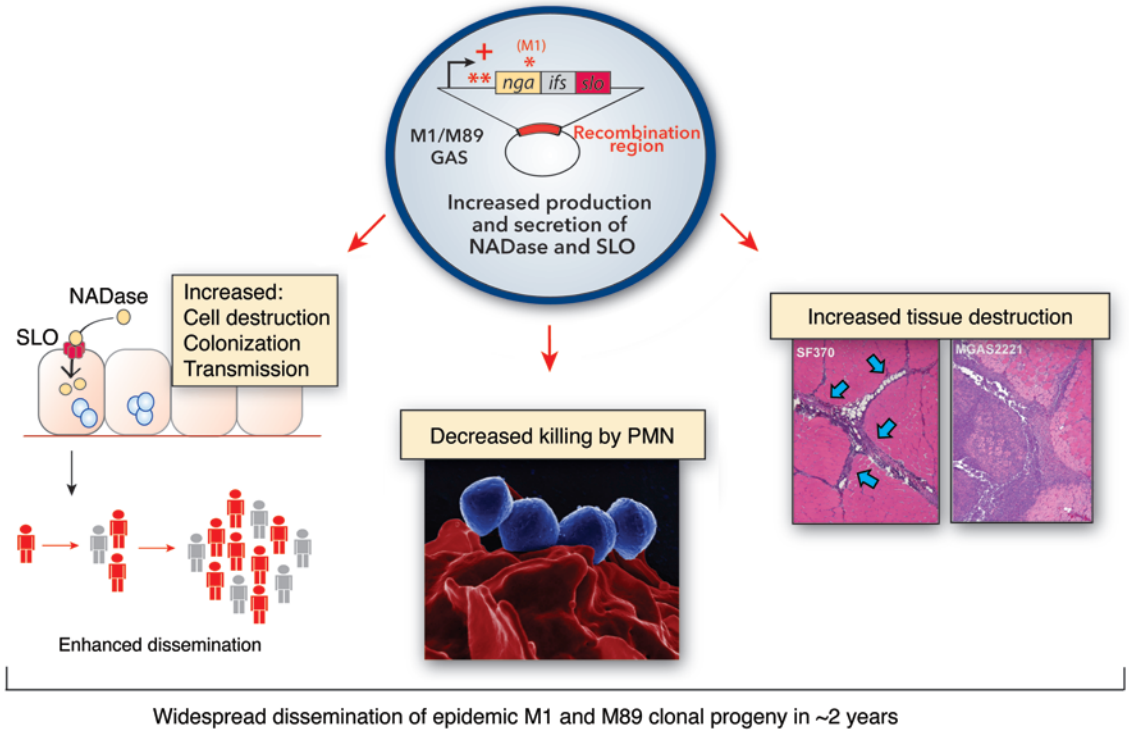

Figure 8. Summary of molecular events underpinning serotype M1 and M89 GAS virulent clone emergence and epidemics of bacterial pathogens. The key molecular event for both M1 and M89 epidemics was recombinational replacement of a segment of chromosomal DNA with SNPs that cause upregulation and increased production of two potent secreted toxins. The asterisk above the *nga* gene denotes that the recombined segment in the new epidemic M1 clone contains a nonsynonymous SNP that results in production of enzymatically active extracellular NADase. All serotype M89 strains produce enzymatically active NADase. The two asterisks upstream of the *nga* gene denote SNPs in the promoter space region that result in significant upregulation of transcription of this operon in M1 and M89 epidemic strains. The blue arrows in the histopathology panel identify the small lesion observed in muscle infected with the SPN/SLO-deficient pre-epidemic *emm1* strain SF370 (left) compared with the much larger lesion with extensive tissue destruction caused by the epidemic *emm1* strain MGAS2221 (right). Original magnification, ×4.

morphisms in transcriptional regulators that can potentially alter expression levels of *nga* and *slo* (Supplemental Figure 7). Moreover, we found that the M89 strains with variant 3 *nga* promoter (clade-3 strains) emerged very recently from a single precursor cell and spread rapidly intercontinentally (Supplemental Figure 7). In this regard, we note that serotype M89 strains have caused a substantial increase in infections in recent years in Canada, New Zealand, Lebanon, and many areas of Europe (26–29). As with the M1 strains, the pre-epidemic and epidemic M89 strains differ by more than just a few SNPs. But importantly, identical to that in the M1 strains, in M89 strains it is the *nga-slo* region recombinational replacement event and commensurate upregulation of SPN and SLO production that temporally coincides with the increase in M89 infection frequency. Thus, in the aggregate, the M89 data remarkably parallel all of our findings for the M1 epidemic strains. These data emphasize the critical role that SNPs in the promoter space region upstream of the *nga*, *ifs*, and *slo* operon play in enhancing pathogen fitness and dissemination.

Discussion

Many diverse factors can contribute to pathogen emergence and epidemic spread, including microbial, ecological, host immunologic and susceptibility factors, and political and sociological events. As illustrated by Ebola, SARS, MERS, and other viruses, a common theme in pathogen emergence is introduction of an existing strain into a new susceptible host species (species "jump") (30–33). In addition, new viruses are sometimes generated by recombinational reassortment, a process well described for influenza viruses (34, 35). However, despite some progress (34, 36), identification of the precise molecular genetic events and pathogenic mechanisms that lead to pathogen strain emergence, altered virulence, and epidemicity generally has been an elusive goal in infectious disease research. Information of this type is particularly lacking for pathogenic human bacteria. In this work, using an integrative approach and a model organism, the power of high-throughput sequencing of the genome of approximately 5,000 bacterial isolates, comprehensive population-based strain samples, molecular genetics, and animal infection models, we addressed an important question in infectious diseases research. Unexpectedly, for both serotype M1 and M89 streptococcal human epidemics, independent recombinational replacement events involved acquisition of a promoter spacer region variant characterized by the same two SNPs in the *nga*, *ifs*, and *slo* promoter. The end result of both genetic events was significantly increased production of the same two potent secreted toxins, resulting in clone emergence, rapid intercontinental dispersal, and increased virulence (Figure 8). Importantly, for the M1 epidemic clone, the recombination event also "repaired" a nonsynonymous SNP in

the *nga* gene, resulting in the production of enzymatically active extracellular NADase toxin by the epidemic strain. In the case of the M89 strains, regardless of year of isolation, all M89 strains have an *nga* allele encoding enzymatically active NADase. Thus, the crucial event in spurring the M89 epidemic was significantly increased transcription of the *nga*, *ifs*, and *slo* operon caused by the promoter-region SNPs, resulting in significantly increased production of two key extracellular toxins and significantly increased virulence of the resulting clonal progeny strains (Figure 7 and summarized in Figure 8).

We believe that increasing the production of SPN and SLO activity is an especially nefarious way to ignite an explosive epidemic because it tips the balance in favor of the microbe in at least three critical steps in human pathogenesis. First, upregulation of these virulence factors produces an organism with enhanced proliferation and fitness in the upper respiratory tract. This is a key early step in the interaction between pathogen and host and one prerequisite for successful rapid and widespread transmission. Based on extensive human studies, Hamburger et al. (37) found that patients with higher GAS burden in the upper respiratory tract are more likely to have clinical illness (i.e., increased virulence) and transmit the pathogen to others (increased epidemic potential). Our data suggest that enhanced production of SPN and SLO creates a pathogen with increased fitness in the upper respiratory tract, resulting in ability to disseminate in a very short time frame and cause abundant human infections in susceptible persons (Figure 8).

Second, we have shown that enhanced production of SPN and SLO results in more extensive tissue destruction, which mirrors clinical findings in infected patients (24). Third, the observation of increased resistance to killing by human PMNs, key mediators of innate immune protection, is consistent with other reports bearing on the role of these two potent secreted toxins in pathogenesis. Thus, increased production of SPN and SLO is an elegant mechanism to create pathogens that are especially adept at transmitting dynamically in epidemic waves throughout the human population and causing unusually aggressive infections. The discovery that the same molecular process has triggered two distinct epidemics indicates that production of SPN and SLO must bestow on the pathogen a substantial selective advantage and survival premium. Although we do not know the precise survival premium operative, in part because of a lack of understanding about the host immune processes at work in the upper respiratory tract, it is reasonable to think that a relative lack of sufficient neutralizing antibody levels or deficit in other host defenses attacked by SPN and SLO may contribute. A second potential contributing factor is the recombinational replacement process used to generate the new clones that give rise to the epidemics. That is, when the *nga*, *ifs*, and *slo* region from a donor GAS strain is introduced into a recipient of a different M protein serotype, or a genetically divergent clone of the same M type, a recombinant bacterium is created that not only has enhanced production of SPN and SLO, but also likely expresses an array of other extracellular products against which the host may be immunologically naive or relatively so. This would enhance the probability that the recombinant cell might gain a foothold in the human host and spread throughout the population.

Our findings have implications for basic, clinical, and translational research, including development of vaccines, therapeutics, diagnostics, and public health endeavors. For example, the data imply that inclusion of these two secreted toxins in a human GAS vaccine may contribute to protective immunity. In this regard, immunization with recombinant native or genetically toxoided SLO protects mice against infection (38, 39). It may also be useful to consider developing small-molecule inhibitors (40) that target SPN and SLO. To summarize, our study provides an interdisciplinary investigational framework whereby other bacterial epidemics can be examined and their molecular causes deconvoluted and perhaps prevented.

Methods

Bacterial strains. Strains SF370 and MGAS2221 are representative of GAS M1 pre-epidemic and epidemic strains, respectively. Eight isoallelic mutant strains were derived from epidemic parental strain MGAS2221 (Supplemental Figure 1). Strains *nga*(G-22T) and *nga* (T-18C) each have a single SNP in the *nga* promoter of MGAS2221, G-22T, and T-18C, respectively. Strain *nga*(G-22T/T-18C) has 2 SNPs in the *nga* promoter region. Strain *nga*(G989A) has a nonsynonymous mutation in the *nga*-coding region, changing an amino acid critical to NADase activity. Strain *nga*(G-22T/T-18C/G989A) has all 3 SNPs. Δ*nga/slo* is an *nga*, *ifs*, and *slo* deletion mutant. Two analogous isoallelic mutant strains were also derived from M89 parental strain MGAS26844, genetically representative of contemporary epidemic M89 strains. MGAS26844 was isolated from a case of invasive disease in California, USA, in 2003. Strains MGAS2221 and MGAS26844 were selected as parentals in part because they lack function-altering mutations in major regulatory genes (*mga*, *covRS*, *ropB*, etc.) known to alter the global transcriptome and virulence. In addition, as shown by genome sequencing of 3,615 strains, MGAS2221 is genetically representative of the epidemic serotype M1 clone, and a very-high-quality genome sequence is available for it.

The clinical strains studied were all collected as part of comprehensive population-based studies of invasive infections. The *emm1* strains have been described previously (3). The *emm89* strains were obtained from culture collections housed at the CDC, in Finland, and Iceland. The US isolates (1995–2013) were obtained during large epidemiologic studies of invasive infections conducted by the CDC (Active Bacterial Core Surveillance Program; http://www.cdc.gov/abcs/index.html). The Iceland *emm89* isolates were collected during country-wide surveillance studies of invasive infections between 2003 and 2013 (25). The Finnish *emm89* GAS strains (all blood isolates) were collected between 2003 and 2013 as part of population-based, national infectious disease surveillance conducted by National Institute for Health and Welfare (THL) (http://www3.thl.fi/stat/).

Construction of double-SNP strain nga(G-22T/T-18C). SNPs G-22T and T-18C were introduced into the –10 and –35 spacer region of the MGAS2221 *nga* promoter by overlap-extension PCR (Supplemental Table 1 and Supplemental Figure 4). Primers ngapro1 and ngapro4 (see Supplemental Table 1 for primer sequences) were used to amplify a 1,097-bp fragment of the MGAS2221 chromosome containing the 5′ end of *nga* and its upstream promoter flanked by BamHI restrictions sites. This product was used in PCR with primers ngapro2 and ngapro3 (see Supplemental Table 1 for primer sequences) to generate an amplicon with both the G-22T and T-18C SNPs. The double-SNP–containing amplicon was digested with BamHI and ligated into the BamHI site of plasmid pBBL740. The ligated product was transformed into

E. coli strain JM109. Chloramphenicol-resistant transformants were analyzed by DNA sequencing, and all were found to have acquired spontaneous spurious mutations within the *nga* promoter, suggesting that the *nga* fragment or its translated product is toxic to *E. coli* JM109. A resultant plasmid, pBBL740-ngapro, with a spontaneous 2-bp deletion in the *nga* promoter, was used as the template for inside-out PCR using phosphorylated primers ngapro5 and ngapro6 (see Supplemental Table 1 for primer sequences). These primers were designed to correct the spurious spontaneous mutations and to amplify the recombinant plasmid (Supplemental Figure 5). The resultant linear amplicon was circularized and closed using T4 DNA ligase, column purified, and electroporated into strain MGAS2221. Chloramphenicol-resistant transformants were selected and verified by PCR and DNA sequencing to ascertain that (a) the plasmid had integrated via single-crossover homologous recombination adjacent to *nga* and (b) the transformed integrant now contained two copies of the *nga* promoter region: one with the native promoter and one with the two introduced SNPs. One such integrant was passaged on nonselective Todd-Hewitt broth containing yeast extract (THY) agar plates to allow for a second crossover event looping-out the integrated plasmid, completing the allelic replacement. Strains from which the plasmid was cured were identified on the basis of chloramphenicol susceptibility by replicate patching on THY agar plates with and without chloramphenicol. To identify allelic replacement strains in which the native MGAS2221 sequence "GTTGT" was replaced with the derived 2-SNP sequence "TTTGC" in the –10 and –35 spacer region of the *nga* promoter, we sequenced the *nga* promoter region of chloramphenicol-sensitive strains with primer ngapro7. Resultant strain *nga*(G-22T/T-18C) was whole-genome sequenced as final confirmation of the allelic replacement and to check for the spontaneous generation of spurious unwanted polymorphisms potentially influencing *nga/slo* expression/activity.

Construction of isoallelic single-SNP mutant strains nga(G-22T) and nga(T-18C). Analogous to the procedure described to introduce the G-22T and T-18C SNPs together into the *nga* promoter, inside-out PCR was used to introduce these SNPs individually into pBBL740-ngapro. Phosphorylated primers ngapro5 and ngaproGT were used to introduce SNP G-22T into the *nga* promoter. Similarly, phosphorylated primers ngapro5 and ngaproTC (see Supplemental Table 1 for primer sequences) were used to introduce SNP T-18C into the *nga* promoter. The respective PCR products were ligated, purified, and individually electroporated into MGAS2221. Chloramphenicol-resistant transformants were selected and single-crossover integrants were verified by PCR and DNA sequencing. Integrants were serially passaged on nonselective THY agar plates to allow for second crossover plasmid looping-out events. Patch plating onto THY agar with and without antibiotic was used to identify chloramphenicol-sensitive strains. PCR and DNA sequencing of chloramphenicol-sensitive strains using primer ngapro7 (see Supplemental Table 1 for primer sequences) was used to identify strains with the single-introduced G-22T SNP and strains with the single-introduced T-18C SNP. Whole-genome sequencing was used for final confirmation of the resultant mutant strains *nga*(G-22T) and *nga*(T-18C).

Construction of coding-SNP mutant strain nga(G989A). Overlap-extension PCR using primer sets ngaGA1/2 and ngaGA3/4 was performed to generate a 1,349-bp amplicon spanning the 989th base of *nga* coding sequence flanked with BamH1 sites. Primers ngaGA2 and ngaGA3 (see Supplemental Table 1 for primer sequences) were used to introduce the nonsynonymous G989A SNP into the *nga* coding sequence. The PCR product was digested with BamHI and cloned into the BamHI site of pBBL740 to construct recombinant plasmid pBBL740-G330D. The recombinant plasmid was electroporated into strain MGAS2221. Chloramphenicol-resistant single-crossover integrants were screened for by PCR and DNA sequencing. Integrants were passaged on nonselective THY agar to allow for a second crossover event and looping-out resolution of the integrated plasmid. Chloramphenicol-sensitive strains that had lost the integrated plasmid were selected by patching colonies onto THY agar plates with and without chloramphenicol. Sanger sequencing was performed to screen for strains with the "G989A" polymorphism using primer ngaG330D1.

Construction of triple-SNP isoallelic mutant nga(G-22T/T-18C/G989A). Triple-SNP strain *nga*(G-22T/T-18C/G898A) was generated by electroporating coding-SNP G989A encoding plasmid pBBL740-G330D into double-SNP isoallelic mutant strain *nga*(G-22T/T-18C). Selection, identification, and verification of strain *nga*(G-22T/T-18C/G989A) were accomplished as described for the other isoallelic mutants.

Construction of the nga/slo double knockout mutant Δnga/slo. Primer sets ngasloΔ1/2 and ngasloΔ3/4 (see Supplemental Table 1 for primer sequences) were used to amplify a 770-bp fragment upstream of *nga* and a 784-bp fragment that overlaps with 650 bp of *slo*. The two fragments were fused together by overlap-extension PCR. The PCR product was then digested with BamHI and cloned into the BamHI site of pBBL740. The recombinant plasmid was then electroporated into strain MGAS2221. Chloramphenicol-resistant single-crossover integrants were verified by PCR and sequencing. A verified integrant was passaged on nonselective THY agar to allow for a second crossover event and looping-out resolution of the integrated plasmid. This process results in deletion of 2,607 bp of the *nga-slo* region that includes the promoter upstream of *nga*, *nga*, and *ifs*, and the first 630 bp of *slo*. Chloramphenicol-sensitive strains were selected by patch plating on THY agar with and without chloramphenicol. Primers ngasloΔ1 and ngasloΔ4 were used to screen for mutants with the *nga-slo* deletion event by PCR.

Construction of the nga/slo-overexpressing strain Nga/Slo-OE. An *nga/slo*-overexpressing strain was constructed by inserting the strong constitutive promoter of the GAS gene *tufA* directly upstream adjacent to the native *nga* promoter. Primer sets tuf-nga-1/2 and tuf-nga-5/6 were used to amplify 1-kb fragments flanking the site of insertion, and primer set tuf-nga-3/4 was used to amplify the *tufA* promoter (see Supplemental Table 1 for primer sequences). The 3 DNA fragments were fused into a single amplicon by overlap-extension PCR. BamHI restriction sites were introduced into the two ends of the PCR product by primers tuf-nga-1 and tuf-nga-6. The PCR product was digested with BamHI and ligated into BamHI site of pBBL740. After overnight incubation at 6°C, the ligation product was column purified and electroporated into MGAS2221. A single-crossover integrant and then a double-crossover mutant with the *tufA-nga* combined promoter introduced upstream of *nga* were obtained as described above. Primers tuf-nga-1 and tuf-nga-6 were used for PCR screening, and tug-nga-3 was used for Sanger sequencing confirmation.

Construction of the nga/slo-downregulated strain Nga/Slo-DE. An *nga/slo*-downregulated strain was constructed by introducing a 2-bp deletion into the *nga* promoter. Plasmid pBBL740-ngapro, which has a 2-bp deletion in the *nga* promoter, was used to transform serotype M1 parental strain MGAS2221. One verified transformant was subjected to serial passages on nonselective THY agar. Strains that looped out

the integrated plasmid were selected to screen for the 2-bp deletion in *nga* promoter by Sanger sequencing.

Construction of the nga promoter mutant 26844-V1 (variant 1). Phosphorylated primers ngapro5 and ngapro-AG (see Supplemental Table 1 for primer sequences) were used for inside-out PCR to amplify plasmid pBBL740-ngapro and to introduce SNP "A-27G" and "G-22T" into the spacer region of *nga* promoter. Column-purified PCR product was circularized with T4 ligase and electroporated into *emm89* parental strain MGAS26844. A chloramphenicol-resistant transformant that contained a single-crossover was subjected to allelic exchange as described above.

Construction of the nga promoter mutant 26844-V2 (variant 2). Analogous to mutant strain 26844, phosphorylated primers ngapro5 and ngapro6 were used for inside-out PCR to amplify plasmid pBBL740-ngapro and to introduce SNP "G-22T" and "T-18C" into the spacer region of the *nga* promoter. The PCR product was circularized and transformed into strain MGAS26844. Allelic exchange was performed as described above.

Transcriptome (RNA-Seq) analysis. Four pre-epidemic and three epidemic invasive serotype M1 GAS strains were analyzed (Figure 2). These strains lack mutations in known major regulator genes (e.g., *mga*, *ropB*, *covR*, *covS*) that would alter the global transcriptomes. The strains are located at the base of all major branches of the phylogenetic tree. Strains were grown in Todd-Hewitt broth (Bacto Todd-Hewitt broth [BD] supplemented with 0.2% yeast extract [THY medium]). The bacteria were grown overnight in 10 ml THY medium at 37°C with 5% CO_2. Aliquots (0.6 ml) from overnight cultures were used to inoculate 30 ml prewarmed THY in 50-ml nuclease-free conical tubes. Each strain was grown in triplicate, and samples were collected when the absorbance at 600 nm reached ODs reading of 1.0 and 1.8, corresponding to midlogarithmic and early-stationary phases, respectively (Figure 2B). The growth curves of the 7 strains analyzed were superimposable (Figure 2). Bacteria were added to RNAprotect (RNAprotect Bacteria Reagent, Qiagen) and centrifuged at 3,220 *g* for 10 minutes. The supernatant was discarded, and the bacterial pellet was frozen in liquid nitrogen and stored at –80°C.

To extract RNA, the frozen cell pellets were thawed on ice and suspended in 100 μl of RNase-free TE buffer (TE pH 7.0, Ambion). The bacteria were added to tubes containing 0.1-mm silica beads (Lysis matrix B, MP Bio) and lysed via cell disruption (FastPrep-96, MP Bio). The RNA was extracted using an RNeasy-96 Kit (Qiagen) per the manufacturer's guidelines. The extracted RNA was subjected to rigorous DNase treatment (DNA-Free Kit, Ambion), and the quality and quantity of total RNA was evaluated using a Bioanalyzer instrument (Agilent Bioanalyzer RNA 6000 Nano Total RNA Kit) and fluorometry (Qubit RNA Assay Kit, Life Technologies), respectively. Ribosomal RNA was depleted with a bacterial Ribo-Zero Kit (Ribo-Zero Magnetic Kit [Bacteria], Epicentre). The ribosomal RNA was removed from total RNA using the magnetic beads provided with the kit per the manufacturer's instructions. The purified rRNA-depleted RNA was used to synthesize di-tagged cDNA using the ScriptSeq RNA-Seq Library Preparation Kit (Epicentre Biotechnologies). The directional cDNA library was prepared using the ScriptSeq RNA-Seq Library Preparation Kit with indexed reverse primers obtained from the ScriptSeq Index PCR Primers Kit (Epicentre Biotechnologies). The cDNA library for each sample was produced via 15 cycles of PCR, and the generated libraries were purified with AMPureXP beads (Agencourt AMPUre XP, Beckman Coulter Inc.). The cDNA library for each sample was quantified fluorometrically with Qubit dsDNA HS Assay Kits (Invitrogen). The library size was analyzed using a High-Sensitivity DNA Analysis Kit (Agilent). The 96-barcode cDNA libraries were diluted, pooled, and analyzed with an Illumina HiSeq 2500 instrument.

RNA-seq data were analyzed with the CLC bio software suite. Briefly, FASTQ sequence files were mapped to the genome of reference strain MGAS5005. Total aligned reads in reads per kilobase per million ORF read values were normalized to total mRNA (excluding rRNA and tRNA). Normalized gene expression values from each strain and from each condition were compared with each other. The Baggerly test and Bonferroni correction were used to test for differential transcript expression. We considered a gene to be differentially expressed when the fold change in transcript level was at least 1.5-fold different between strains and the *P* value was less than 0.05.

Principal component analysis of the RNA-seq data, which were first compiled into a matrix of samples by gene expression levels, was implemented and plotted using prcomp function in R package stats. RNA-seq data were deposited in the NCBI Gene Expression Omnibus database under accession number GSE70496.

qRT-PCR analysis of nga and slo transcripts in GAS strains grown in THY broth. To determine relative abundance of *nga* and *slo* mRNA, we performed qRT-PCR with TaqMan primers and probes. RNA from GAS cultures (grown to OD_{600} of 0.5) was extracted with an RNAeasy Mini Kit (Qiagen) and converted into cDNA using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems). Quantitative PCR was performed with TaqMan Fast Universal PCR Master Mix (Applied Biosystems) and an ABI 7500 Fast System (Life Technologies) instrument. Sequences of TaqMan primers and probes used are shown in the Supplemental Table 2. Each experiment was performed in triplicate with mean values (± SD) shown.

Western immunoblot analysis of NADase and SLO in the culture supernatant. GAS strains were grown in THY broth at 37°C to OD_{600} of 0.5. Cell-free supernatants were collected by centrifugation at 3,220 *g* for 10 minutes and filtered through a 0.2-μm filter. Proteins in the supernatant were concentrated with 2-ml centrifugal filters (Amicon) and assayed for presence of immunoreactive SLO and NADase with anti-SLO (American Research Product, 12-3017R) and anti-NADase antibodies (Abcam, ab127942).

Measurement of SLO activity and NADase activity in the culture supernatant. SLO activity in the culture supernatant was assayed as described previously (12). NADase activity in culture supernatants was assayed as described elsewhere (12, 41) with minor modifications. Briefly, GAS strains were grown in THY broth at 37°C to OD_{600} of 0.5. Cell-free supernatants were prepared by centrifuging bacterial cells at 3,220 *g* for 10 minutes and filtering them through a 0.2-μm membrane. 100 μl cell-free supernatant was pooled and subjected to 2-fold serial dilutions with fresh THY broth in 96-well plates. 100 nmol NAD^+ (in 100 μl PBS) was added to each well and incubated at 37°C for 1 hour. To determine the extent of NAD^+ hydrolysis, the phenazine methosulphate-*p*-iodonitrotetrazolium violet colorimetric method was used to assess the presence or absence of NAD^+ in each well. Briefly, 50 l master mix was added to each well with 10 U/ml of alcohol dehydrogenase, 1% ethanol, 2.7 mg/ml of *p*-iodonitrotetrazolium violet (PITZ), and 0.07 mg/ml of phenazine methosulfate. NAD^+ is the cofactor of alcohol dehydrogenase. If NAD^+ is not completely hydrolyzed by streptococcal NADase, alcohol dehydrogenase reduces PITZ into formazan, which turns the color of the

reaction into rusty red. If NAD^+ is completely hydrolyzed, there will be no color change. The NADase activity of each strain was reported as the highest 2-fold dilution that could completely hydrolyze 100 nmol NAD^+.

Capsule assay and protease activity assay. The capsule assay (hyaluronic acid assay) and protease activity assay were performed according to previous descriptions (42, 43).

Mouse and nonhuman primate virulence experiments. Immunocompetent 6-week-old female CD1 mice (Harlan Laboratories) were used for the necrotizing fasciitis studies as described previously (43). Mice were randomly assigned to treatment groups and inoculated in the right hind limb to a uniform depth with 1×10^7 CFUs of M1 GAS strains or 2.5×10^8 CFUs of M89 strains in 100 μl PBS. Stocks of each strain were prepared at known CFUs and stored at –80°C. Inocula were prepared immediately before infection by diluting frozen stocks in PBS to the desired number of CFUs. For survival experiments ($n = 10$ mice per strain), near-mortality was determined by observation using predefined criteria, with statistical differences between strain groups determined with the log-rank test. For histology evaluation ($n = 3$ mice per strain), lesions were excised, visually inspected, and photographed. The tissue was fixed in 10% phosphate buffered formalin, decalcified, and embedded in paraffin using standard automated instruments. Histology was scored by a pathologist blinded to the strain treatment groups as described previously (3, 43). For the quantitative culture experiment ($n = 20$ mice per strain at each time point), limbs were homogenized (Omni International) and CFUs were determined by plating serial dilutions of the homogenate. Statistical differences between strain groups were determined using the Mann-Whitney test. For the nonhuman primate experiments, the virulence of wild-type strain MGAS2221 and the 3-SNP isoallelic mutant strain was assessed in a pharyngitis model using GAS-naive, 4- to 6-year-old cynomolgus macaques (*Macaca fasicularis*) (Charles River BRF) as previously described (3). Monkeys were randomly assigned to treatment groups ($n = 4$ animals per strain), anesthetized, inoculated in the upper respiratory tract with 1×10^8 CFUs, and sampled sequentially for 21 days.

Measurement of transcript levels of nga and slo during mouse and monkey infection. To determine in vivo transcript levels of *nga* and *slo*, 4 mouse lesions per strain were collected at indicated time points and immediately placed in RNAlater (Qiagen). GAS RNA was isolated from limbs or throat swabs using a previously described method (3, 43), with minor modifications. Briefly, tissue samples containing the lesion areas were excised from the limbs, placed in 700 μl lysis buffer, and homogenized using FastPrep Lysing Matrix F (MP Biomedicals). The homogenized tissue samples were transferred to FastPrep Lysing Matrix B (MP Biomedicals) to further disrupt bacterial cells. RNA from each sample was extracted and purified with an RNAeasy Mini Kit (Qiagen). Reverse transcription and quantitative PCR were performed according to the methods described above for the in vitro qRT-PCR assay.

Measurement of nonviable PMNs in the nonhuman primate oropharynx. To assess viability of PMNs, cells were eluted from throat swabs, washed twice with PBS, stained with 7-aminoactinomycin D, and analyzed by flow cytometry.

Human PMN isolation and GAS survival assays. Human PMNs were purified from heparinized blood of healthy donors using a standard method (44). Donors were informed of the procedure risks, and written informed consent was obtained before phlebotomy. Purity and viability of purified PMNs was typically 98%–99%.

Bactericidal activity assays with human PMNs and GAS were performed using a standard phagocytosis method with modifications (45). Bacteria were cultured to an early midexponential phase of growth (OD_{600} ~ 0.150–0.200) in THY and washed once in phosphate-buffered saline. 10^6 CFUs were combined with 10^6 PMNs in a 1.5-ml tube containing 10% autologous human serum in HEPES-buffered RPMI 1640 medium (600 μl final volume). Assay mixtures were rotated gently at 37°C for 1 hour or 3 hours, at which time saponin (0.1% final concentration) was added to each assay mixture and tubes were placed on ice for 15 minutes. An aliquot of each tube was diluted appropriately and plated on THY agar plates for enumeration of CFUs the following day. Percentage GAS survival was determined by comparison of CFUs from assays with PMNs to CFUs from assays without PMNs (CFU data from Figure 4D and Figure 5F are presented in Supplemental Figure 6). Data were analyzed with repeated-measures 1-way ANOVA and Tukey's post test to correct for multiple comparisons (Prism 6 for Mac OS X, GraphPad Prism, GraphPad Software Inc.).

Whole-genome sequencing and analysis of emm89 isolates. The 1,125 *emm89* GAS isolates were recovered from the US, Finland, and Iceland, between January 1995 and December 2013, as part of comprehensive population based surveillance studies. Chromosomal DNA was extracted from each of the strains using the DNeasy 96 Blood & Tissue Kit (Qiagen). Multiplexed libraries for sequencing were generated using the Nextera XT DNA Sample Preparation Kit and the Nextera XT Index Kit V2 (Illumina). Library sequence data were predominantly obtained with an Illumina HiSeq2500 instrument, and some library sequence data were obtained with an Illumina MiSeq instrument. Polymorphisms among the strains were identified relative to assembled reference genomes of 3 *emm89* GAS strains (MGAS11027, variant 1; MGAS23530, variant 2; MGAS27061, variant 3) using SMALT, and FreeBayes software. *emm89* whole-genome sequencing data were deposited in the Sequence Read Archive (NCBI) under accession number SRP059971.

Statistics. Standard statistical methods were performed as follows. Baggerly's test with Bonferroni correction was used to evaluate differences in genome-wide transcript analysis (CLC bio, Qiagen). Results of TaqMan qRT-PCR gene expression, NADase/SLO enzymatic activity, SpeB-secreted protease activity, and hyaluronic acid capsule production are expressed as mean ± SD, with statistically significant differences determined using 2-tailed Student's *t* test. Results of the PMN bactericidal assay were analyzed with a repeated-measures 1-way ANOVA and Tukey's post test to correct for multiple comparisons (Prism 6 for Mac OS X, GraphPad Prism, GraphPad Software Inc.). Results of CFU recovery and histology scoring of limbs harvested from mice are expressed as mean ± SEM, with statistically significant differences determined using the Mann-Whitney test (Prism 6, Graphpad Software); a nonparametric test was used since the data were shown to not follow a normal distribution using the Shapiro-Wilk test (Prism 6). Mouse survival data are shown as Kaplan-Meier survival curves, with statistically significant differences determined using the log-rank test (Prism 6). In all cases, a *P* value of less than 0.05 was considered statistically significant.

Study approval. All animal experiments were approved by the Institutional Animal Care and Use Committee of Houston Methodist Research Institute. Human PMN isolation was performed in accordance with a protocol approved by the Institutional Review Board for Human Subjects of the National Institute of Allergy and Infectious Diseases, NIH.

Acknowledgments

This study was supported in part by the Fondren Foundation, Houston Methodist Hospital and Research Institute, the Academy of Finland (grant 255636), and the Intramural Research Program of the National Institute of Allergy and Infectious Diseases, NIH. We thank Neal Copeland, David Ginsburg, Nancy Jenkins, David Huston, Vivek Kapur, David Morens, and James Versalovic for critical comments and suggestions for improving the manuscript; Kathryn Stockbauer for critical comments and editorial assistance; Concepcion Cantu, Helga Erlendsdottir, Yu Joe, Jesus Paez Mayorga, and Audrey Ponce De Leon for technical assistance; Leslie Jenkins and Annessa Raiford for veterinary assistance; Laura Lindholm, Jari Jalava, and the Finnish Study Group for Antimicrobial Resistance (FiRe); and Chris Van Beneden, Bernard Beall, and the Active Bacterial Core Surveillance of the CDC's Emerging Infections Programs network.

Address correspondence to: James M. Musser, Houston Methodist Hospital, 6565 Fannin, Suite B490, Houston, Texas 77030, USA. Phone: 713.441.5890; E-mail: jmmusser@houstonmethodist.org.

1. Beres SB, et al. Molecular complexity of successive bacterial epidemics deconvoluted by comparative pathogenomics. *Proc Natl Acad Sci U S A*. 2010;107(9):4371–4376.
2. Fittipaldi N, et al. 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(4):1522–1534.
3. Nasser W, et al. Evolutionary pathway to increased virulence and epidemic group A Streptococcus disease derived from 3,615 genome sequences. *Proc Natl Acad Sci U S A*. 2014;111(17):E1768–E1776.
4. Lancefield RC. Current knowledge of type-specific M antigens of group A streptococci. *J Immunol*. 1962;89(3):307–313.
5. Colman G, Tanna A, Efstratiou A, Gaworzewska ET. The serotypes of Streptococcus pyogenes present in Britain during 1980–1990 and their association with disease. *J Med Microbiol*. 1993;39(3):165–178.
6. Lamagni TL, et al. Epidemiology of severe Streptococcus pyogenes disease in Europe. *J Clin Microbiol*. 2008;46(7):2359–2367.
7. Bricker AL, Carey VJ, Wessels MR. Role of NADase in virulence in experimental invasive group A streptococcal infection. *Infect Immun*. 2005;73(10):6562–6566.
8. Bricker AL, Cywes C, Ashbaugh CD, Wessels MR. NAD^+-glycohydrolase acts as an intracellular toxin to enhance the extracellular survival of group A streptococci. *Mol Microbiol*. 2002;44(1):257–269.
9. Bastiat-Sempe B, Love JF, Lomayesva N, Wessels MR. Streptolysin O and NAD-glycohydrolase prevent phagolysosome acidification and promote group A streptococcus survival in macrophages. *mBio*. 2014;5(5):e01690–e01614.
10. Walker MJ, et al. Disease manifestations and pathogenic mechanisms of group A Streptococcus. *Clin Microbiol Rev*. 2014;27(2):264–301.
11. O'Seaghdha M, Wessels MR. Streptolysin O and its co-toxin NAD-glycohydrolase protect group A Streptococcus from Xenophagic killing. *PLoS Pathog*. 2013;9(6):e1003394.
12. Sumby P, et al. Evolutionary origin and emergence of a highly successful clone of serotype M1 group A Streptococcus involved multiple horizontal gene transfer events. *J Infect Dis*. 2005;192(5):771–782.
13. Stevens DL, Salmi DB, McIndoo ER, Bryant AE. Molecular epidemiology of nga and NAD glycohydrolase/ADP-ribosyltransferase activity among Streptococcus pyogenes causing streptococcal toxic shock syndrome. *J Infect Dis*. 2000;182(4):1117–1128.
14. Tatsuno I, et al. Characterization of the NAD-glycohydrolase in streptococcal strains. *Microbiology*. 2007;153(pt 12):4253–4260.
15. Meehl MA, Pinkner JS, Anderson PJ, Hultgren SJ, Caparon MG. A novel endogenous inhibitor of the secreted streptococcal NAD-glycohydrolase. *PLoS Pathog*. 2005;1(4):e35.
16. Kimoto H, Fujii Y, Hirano S, Yokota Y, Taketo A. Genetic and biochemical properties of streptococcal NAD-glycohydrolase inhibitor. *J Bio Chem*. 2006;281(14):9181–9189.
17. Kimoto H, Fujii Y, Yokota Y, Taketo A. Molecular characterization of NADase-streptolysin O operon of hemolytic streptococci. *Biochim Biophys Acta*. 2005;1681(2–3):134–149.
18. Savic DJ, Ferretti JJ. Novel genomic rearrangement that affects expression of the Streptococcus pyogenes streptolysin O (slo) gene. *J Bacteriol*. 2003;185(6):1857–1869.
19. Hook-Barnard IG, Hinton DM. The promoter spacer influences transcription initiation via sigma70 region 1.1 of Escherichia coli RNA polymerase. *Proc Natl Acad Sci U S A*. 2009;106(3):737–742.
20. Singh SS, Typas A, Hengge R, Grainger DC. Escherichia coli sigma(7)(0) senses sequence and conformation of the promoter spacer region. *Nucleic Acids Res*. 2011;39(12):5109–5118.
21. Chandrasekaran S, Ghosh J, Port GC, Koh EI, Caparon MG. Analysis of polymorphic residues reveals distinct enzymatic and cytotoxic activities of the Streptococcus pyogenes NAD+ glycohydrolase. *J Bio Chem*. 2013;288(27):20064–20075.
22. Bernheimer AW, Lazarides PD, Wilson AT. Diphosphopyridine nucleotidase as an extracellular product of streptococcal growth and its possible relationship to leukotoxicity. *J Exp Med*. 1957;106(1):27–37.
23. Hirsch JG, Bernheimer AW, Weissmann G. Motion picture study of the toxic action of streptolysins on leucocytes. *J Exp Med*. 1963;118(2):223–228.
24. Stevens DL, et al. Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A. *N Engl J Med*. 1989;321(1):1–7.
25. Olafsdottir LB, et al. Invasive infections due to Streptococcus pyogenes: seasonal variation of severity and clinical characteristics, Iceland, 1975 to 2012. *Euro Surveill*. 2014;19(17):5–14.
26. Shea PR, et al. Group A Streptococcus emm gene types in pharyngeal isolates, Ontario, Canada, 2002–2010. *Emerging Infect Dis*. 2011;17(11):2010–2017.
27. Karaky NM, Araj GF, Tokajian ST. Molecular characterization of Streptococcus pyogenes group A isolates from a tertiary hospital in Lebanon. *J Med Microbiol*. 2014;63(pt 9):1197–1204.
28. Williamson DA, et al. Increasing incidence of invasive group A streptococcus disease in New Zealand, 2002–2012: a national population-based study. *J Infect*. 2014;70(2):127–134.
29. Luca-Harari B, et al. Clinical and microbiological characteristics of severe Streptococcus pyogenes disease in Europe. *J Clin Microbiol*. 2009;47(4):1155–1165.
30. Woolhouse M, Scott F, Hudson Z, Howey R, Chase-Topping M. Human viruses: discovery and emergence. *Philo Trans R Soc Lond B Biol Sci*. 2012;367(1604):2864–2871.
31. Gire SK, et al. Genomic surveillance elucidates Ebola virus origin and transmission during the 2014 outbreak. *Science*. 2014;345(6202):1369–1372.
32. Poon LL, Guan Y, Nicholls JM, Yuen KY, Peiris JS. The aetiology, origins, and diagnosis of severe acute respiratory syndrome. *Lancet Infect Dis*. 2004;4(11):663–671.
33. Azhar EI, et al. Evidence for camel-to-human transmission of MERS coronavirus. *N Engl J Med*. 2014;370(26):2499–2505.
34. Hatta M, Gao P, Halfmann P, Kawaoka Y. Molecular basis for high virulence of Hong Kong H5N1 influenza A viruses. *Science*. 2001;293(5536):1840–1842.
35. Smith GJ, et al. Origins and evolutionary genomics of the 2009 swine-origin H1N1 influenza A epidemic. *Nature*. 2009;459(7250):1122–1125.
36. DeLeo FR, et al. Molecular differentiation of historic phage-type 80/81 and contemporary epidemic Staphylococcus aureus. *Proc Natl Acad Sci U S A*. 2011;108(44):18091–18096.
37. Hamburger M Jr, Green MJ, Hamburger VG. The problem of the dangerous carrier of hemolytic streptococci; spread of infection by individuals with strongly positive nose cultures who expelled large numbers of hemolytic streptococci. *J Infect Dis*. 1945;77:96–108.
38. Bensi G, et al. Multi high-throughput approach for highly selective identification of vaccine candidates: the Group A Streptococcus case. *Mol Cell Proteomics*. 2012;11(6):M111.015693.
39. Chiarot E, et al. Targeted amino acid substitutions impair streptolysin O toxicity and group A Streptococcus virulence. *mBio*. 2013;4(1):e00387-12.

40. Sun H, et al. Inhibitor of streptokinase gene expression improves survival after group A streptococcus infection in mice. *Proc Natl Acad Sci U S A*. 2012;109(9):3469–3474.
41. Ajdic D, McShan WM, Savic DJ, Gerlach D, Ferretti JJ. The NAD-glycohydrolase (nga) gene of Streptococcus pyogenes. *FEMS Microbiol Lett*. 2000;191(2):235–241.
42. Flores AR, Jewell BE, Fittipaldi N, Beres SB, Musser JM. Human disease isolates of serotype m4 and m22 group A streptococcus lack genes required for hyaluronic acid capsule biosynthesis. *mBio*. 2012;3(6):e00413-12.
43. Olsen RJ, et al. Decreased necrotizing fasciitis capacity caused by a single nucleotide mutation that alters a multiple gene virulence axis. *Proc Natl Acad Sci U S A*. 2010;107(2):888–893.
44. Kobayashi SD, Voyich JM, Buhl CL, Stahl RM, DeLeo FR. Global changes in gene expression by human polymorphonuclear leukocytes during receptor-mediated phagocytosis: cell fate is regulated at the level of gene expression. *Proc Natl Acad Sci U S A*. 2002;99(10):6901–6906.
45. Lu T, Porter AR, Kennedy AD, Kobayashi SD, DeLeo FR. Phagocytosis and killing of Staphylococcus aureus by human neutrophils. *J Innate Immun*. 2014;6(5):639–649.