Rapid selection of complement-inhibiting protein variants in group A *Streptococcus* epidemic waves

NANCY P. HOE[1], KAZUMITSU NAKASHIMA[1], SLAWOMIR LUKOMSKI[1], DIANA GRIGSBY[1], MENGYAO LIU[1], PARICHHER KORDARI[1], SHU-JUN DOU[1], XI PAN[1], JAANA VUOPIO-VARKILA[2], SAARA SALMELINNA[2], ALLISON MCGEER[3], DONALD E. LOW[3], BENJAMIN SCHWARTZ[4], ANNE SCHUCHAT[4], STEVEN NAIDICH[5], DSVID DE LORENZO[6], YUN-XIN FU[6] & JAMES M. MUSSER[1]

[1]*Institute for the Study of Human Bacterial Pathogenesis, Department of Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, Texas 77030, USA*
[2]*Department of Bacteriology, National Public Health Institute, Mannerheimintie 166, Helsinki FIN-00300, Finland*
[3]*Department of Microbiology, Mount Sinai Hospital, 600 University Avenue, and University of Toronto, Ontario M5G 1X5, Canada*
[4]*Centers for Disease Control and Prevention, 1600 Clifton Road, NE, Atlanta, Georgia 30333, USA*
[5]*eGenomics, New York, New York 10013, USA*
[6]*Human Genetics Center, The University of Texas Health Sciences Center, Houston, Texas 77030, USA*
J.M.M. present address: Office of the Chief, Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, 903 South 4th Street, Hamilton, Montana 59840, USA
Correspondence should be addressed to J.M.M.; email: jmusser@bcm.tmc.edu

Serotype M1 group A *Streptococcus* strains cause epidemic waves of human infections long thought to be mono- or pauciclonal. The gene encoding an extracellular group A *Streptococcus* protein (streptococcal inhibitor of complement) that inhibits human complement was sequenced in 1,132 M1 strains recovered from population-based surveillance of infections in Canada, Finland and the United States. Epidemic waves are composed of strains expressing a remarkably heterogeneous array of variants of streptococcal inhibitor of complement that arise very rapidly by natural selection on mucosal surfaces. Thus, our results enhance the understanding of pathogen population dynamics in epidemic waves and infectious disease reemergence.

Many infectious diseases suddenly increase in frequency and severity, a state known as an epidemic, outbreak or emergence[1]. However, knowledge of the microbial and host molecular factors that contribute to the development and perpetuation of epidemics is limited. Molecular analyses[2–7] have led to the prevailing view that most epidemics are caused by a pathogen genotype with special fitness properties.

Group A streptococci (GAS) are human-adapted bacteria that infect the upper respiratory tract and inflame the tonsils and pharynx ('strep throat'). GAS are also important causes of severe invasive infections, and are responsible for post-infection sequelae such as rheumatic heart disease and glomerulonephritis. GAS are divided into serologic subtypes on the basis of M protein, a highly polymorphic surface molecule that contributes to virulence through its antiphagocytic function[8]. Although more than 100 M protein serotypes have been identified, GAS expressing relatively few are responsible for a disproportionate number of invasive infections[9]. Serotype M1 organisms that produce streptococcal pyrogenic exotoxin A (scarlet fever toxin, a superantigen) are the most frequent cause of invasive episodes on several continents[9]. Epidemics of M1 invasive infections occur that develop precipitously and usually last 1–3 years[9–11]. Assessment of genetic variation in M1 organisms has failed to identify substantial genomic diversity[9–13], indicating that waves of M1 infections, like those caused by other bacteria, are mono- or pauciclonal[9–13].

Streptococcal inhibitor of complement (Sic) is an extracellular protein that inhibits the normal host-protective cytolytic function of the complement cascade *in vitro*[14]. The exact mechanism of inhibition has not been completely elucidated, but Sic is incorporated into the complement membrane-attack complex (C5b–C9) responsible for target killing. Sic is one of a heterogeneous array of unrelated proteins produced by diverse pathogens that inhibit the human complement system and presumably promote survival of the infecting organism[15]. Unexpectedly, preliminary analysis of variation in the *sic* gene in M1 GAS strains identified a level of polymorphism far exceeding that of other genes in these organisms and demonstrated that natural selection has contributed to Sic variation[16,17]. Moreover, the level of Sic variation was unprecedented for any bacterial gene comparatively sequenced in organisms that are very closely allied in chromosomal character. Study of very few strains recovered from two consecutive M1 epidemics in the former East Germany found that each epidemic wave was composed of many new Sic variants[16]. Thus, our goal was to evaluate Sic variation in M1 epidemic waves by analyzing organisms obtained in comprehensive, population-based surveillance of invasive disease in widespread localities[11,18,19].

Hypervariation in Sic in invasive isolates

We sequenced the *sic* gene in 892 isolates from Finland; Ontario, Canada; and four sites in the United States (Table). We identified a total of 162 *sic* alleles (sequence types) that would encode 158

Sic variants. This variation was caused by in-frame insertions and deletions and by nonsynonymous nucleotide substitutions (those resulting in amino-acid replacements). Compilation of *sic* sequence data generated here and in previous studies[16] identified 236 *sic* alleles encoding 230 Sic variants. Most nucleotide changes (77 of 86; 90%) occurred in the first or second codon position and produced amino-acid replacements. This result permits us to reject the null hypothesis of selective neutrality ($P < 0.01$). Moreover, 65 of 77 amino acid changes (84%) were radical replacements[20]. (Because almost all *sic* alleles encode distinct Sic protein variants, we use 'variant' interchangeably to refer to *sic* genes and Sic proteins).

Table Serotype M1 group A *Streptococcus* isolates ($n = 1,132$) analyzed from prospective, population-based surveillance studies

Locality	Population (in millions)[a]	Years[b]	Number of M1 isolates	Number of Sic variants
Finland	Nationwide (5.1)	1988–1998	524	92
	Sterile Site		284	59
	Pharyngitis		240	58
Canada	Ontario province (9.0)	1992–1998	349	68
United States		1994–1998	259	78
Georgia	1994–1996, 8-county metro Atlanta area (2.7) 1997–1998, 20-county metro Atlanta area (3.5)	1994–1998	47	18
Connecticut	Statewide (3.3)	1995–1998	87	36
Minnesota	Several-county Minneapolis/St. Paul area (2.5)	1995–1998	80	21
California	3-county San Francisco area (2.9)	1994–1998	45	23

[a]Based on 1990s census data. [b]Includes through August 1998 for Finland, November 1998 for Ontario and April 1998 for the United States.

Sic variation in individual localities

The identification of extensive Sic variation ruled out the prevailing idea that the epidemic waves of invasive infections represented spread of one or a few related clones[9,12,13]. Finland had peaks of M1 invasive disease in 1989–1990 ($n = 78$ cases) and 1995–1997 ($n = 156$ cases). Each epidemic wave was composed of an extremely heterogeneous array of Sic variants, and only seven variants were shared between the two epidemics (Table). In 1989–1990, there were prominent changes in the frequency of occurrence of Sic1.01, 1.02, and 1.121 (Fig. 1*a*). Similarly, the frequency of infections caused by strains with Sic1.102 increased significantly between 1995 and 1997 ($\chi^2 = 5.66$; $P = 0.017$).

In Ontario, strains expressing M1 protein increased in the early 1990s and have been abundant since then. As in Finland, many ($n = 68$) Sic variants were identified, and only seven variants were shared between the 1994 and 1998 isolates. The frequency of occurrence of two of the four more abundant Sic variants (Sic1.01 and Sic1.135) changed considerably during the study period (Fig. 1*b*). The commonalities in Finland and Ontario were confirmed in the United States (Table; Fig. 1*c* and *d*).

Sic evolution

The *sic* data indicated that most variants could be linked to each other by one or very few molecular changes. All evolutionary trees generated from the entire *sic* data set and for each locality had a star-like topology characteristic of a population with rapidly generated variants (Fig. 2*a*). There was a considerable excess of singleton changes (alterations present in only one *sic* allele)(refs. 21–24) with respect to their expected frequency of occurrence under evolutionary neutrality (Fig. 2*b*). The molecular data indicate that new Sic variants are selected very rapidly in the course of epidemic waves.

Analysis of other variable chromosomal loci

If the hypothesis of rapid Sic selection is correct, then we anticipate a relative lack of variation in other genetic markers, because sufficient time would not elapse to permit accumulation of changes at the other loci. The failure to identify allelic variation in ten other genes (including eight genes encoding virulence factors) in M1 strains supports this[9,12,16,17]. We next studied variation in two types of genetic marker systems that generally accumulate polymorphisms faster than simple nucleotide substitutions. We found no significant differences over time in the restriction fragment length polymorphism profile of insertion element IS*1548* or direct repeat subtype[17] (data not shown).

Lack of M1 protein variation

M1 protein variants differing by single amino-acid replacements in the hypervariable amino terminus are thought to be escape mutants that arise by host antibody selection[9,12,17,25]. In addition,

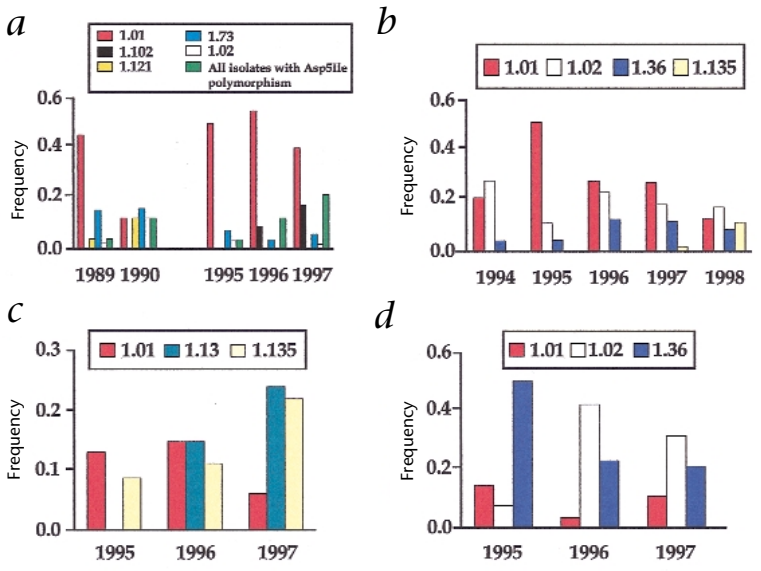

Fig. 1 Frequency distribution of abundant Sic variants over time in four study localities. ***a***, Finland, 1989–1990 and 1995–1997; the years 1991–1994 are excluded because M1 GAS rarely caused invasive episodes in that period. Green, all Sic variants with a signal sequence amino-acid replacement (Asp5Ile) also found in Sic1.102 but never detected in areas outside of Finland. ***b***, Ontario, 1994–1998. ***c***, Connecticut, 1995–1997. ***d***, Minnesota, 1995–1997. Only Sic1.01 is abundant in all areas.

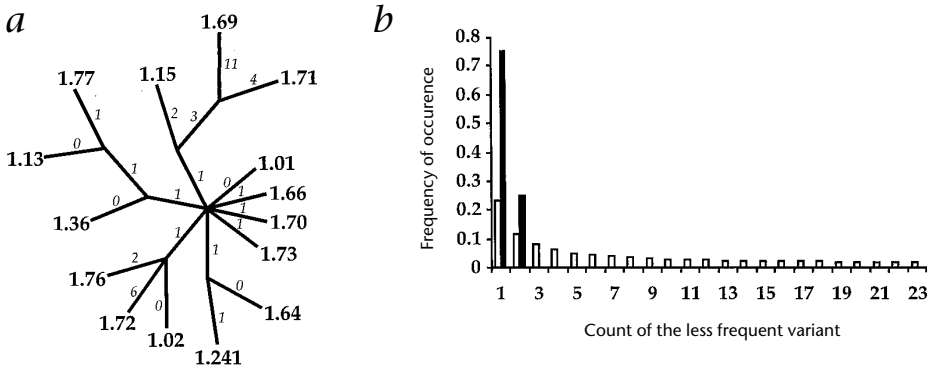

Fig. 2 ***a***, Evolutionary tree for *sic* alleles in a representative local area. This unrooted parsimony[21,22] tree is for *sic* alleles identified in the Ontario sample in 1996. Branch lengths are not proportional to the numbers of differences between alleles. The number at the end of each branch is the arbitrary *sic* allele designation; italicized numbers indicate the number of molecular differences for each branch. ***b***, Observed (■) and expected (□) frequency distribution of molecular changes and deviation from neutral expectation. The graph is based on the 15 *sic* alleles identified in the 47 strains found in the 1996 Ontario sample. Horizontal axis, possible counts of the less-frequent variant of a nucleotide polymorphism in a sample size of $n = 47$; vertical axis, expected and observed frequency of occurrence of the variants: 75% of the total number of nucleotide changes were observed in only one *sic* allele and 25% were present in two alleles. The comparison of the observed and expected frequency of occurrence in a neutral Wright-Fisher model[23,24] indicates a great excess of variant sites present in only a few alleles.

human opsonic antibody may be strain-specific rather than M type-specific[26]. Thus, variation in the amino terminus of M1 protein perhaps mediated or contributed to the change in frequency of GAS invasive episodes. However, sequence analysis of the variable part of the gene encoding the M1 protein (*emm1*) in a random sample of 492 strains (94%) of the invasive and pharyngitis isolates from Finland, and 213 isolates from Ontario and the United States found that 97% had allele *emm1.0*.

Sequence variation that results in changes in opsonic epitopes also has been identified in other regions of the M protein[12,25,27–30]. In addition, some amino-acid replacements in non-amino-terminal regions of the M1 protein correlate with altered opsonophagocytosis[30], indicating that structural changes in non-amino-terminus regions may allow the organism to escape acquired host defenses and thereby participate in the epidemic waves. We sequenced nucleotides 394–1,188 of the *emm1* gene[25] in 55 strains representing each locality, but only one new *emm1* allele was identified, a result ruling out this idea.

Sic variation among pharyngitis isolates

Rapid selection of Sic variants could occur on mucosal surfaces or in a normally sterile site after the mucosal barrier was breached. If selection occurred on mucosal surfaces, then many Sic variants would be expected among M1 isolates recovered from patients with pharyngitis. Alternatively, if selection happened only or mainly after the organism penetrated into a sterile site, relatively few Sic variants would be expected among mucosal isolates. We identified 58 Sic variants in 240 random M1 isolates recovered from patients in Finland with pharyngitis. The ratio of distinct Sic variants to strains studied was almost identical for the pharyngitis and invasive M1 isolates (Table). Notwithstanding the possibility that limited selection may have occurred in sterile sites, the data strongly indicated that most or all selection occurred on the mucosal surface. This observation is consistent with the failure to identify Sic variants after intraperitoneal passage in mice[16].

Selection of Sic variants on the mucosal surface

To test the hypothesis that variants could be selected on mammalian mucosal surfaces, we studied the *sic* gene in GAS cultured sequentially from the throats of eight CD-1 outbred mice

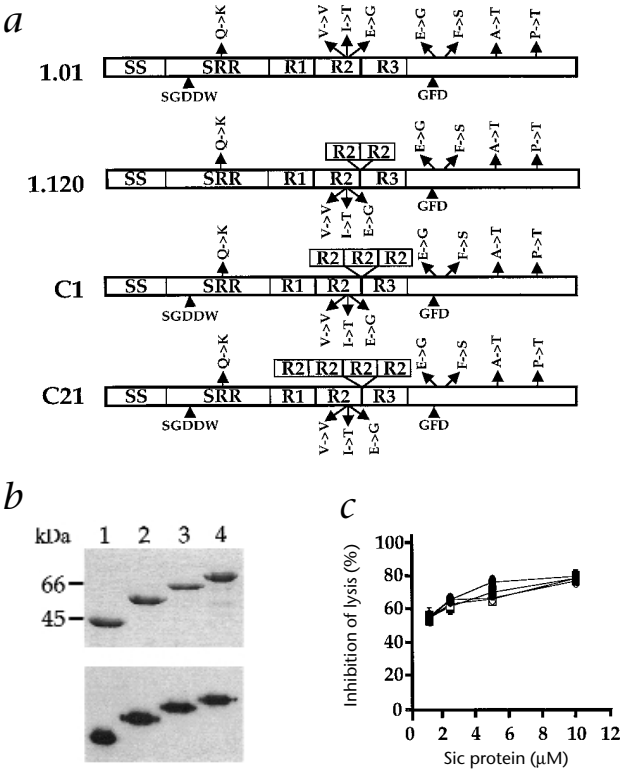

Fig. 3 Inhibition of complement-mediated target cell lysis by Sic variants. ***a***, The four Sic variants used in the inhibition studies. Sic1.01 and Sic1.120 are variants identified among human GAS isolates. Variants C1 and C21 were identified after mouse throat passage of MGAS5005 expressing variant Sic1.01. SS, signal sequence; SRR, short repeat region; R1, R2 and R3, repeat regions. The single-letter amino-acid code is used. ***b***, SDS–PAGE and Coomassie blue (top) and western blot analysis (bottom) of the four purified Sic variants. The amino-terminal sequence for each purified protein was ETYTSRNFDW, the correct Sic sequence. Hyperimmune rabbit antisera raised against purified Sic1.01 was used for western blot analysis. Lane 1, Sic1.01; lane 2, Sic1.120; lane 3, C1; lane 4, C21. ***c***, Inhibition of complement lysis by the Sic variants. The four Sic variants were purified to apparent homogeneity and tested for inhibition of complement-mediated red blood cell lysis. There was no substantial difference in the four Sic variants in complement lysis-inhibiting ability.

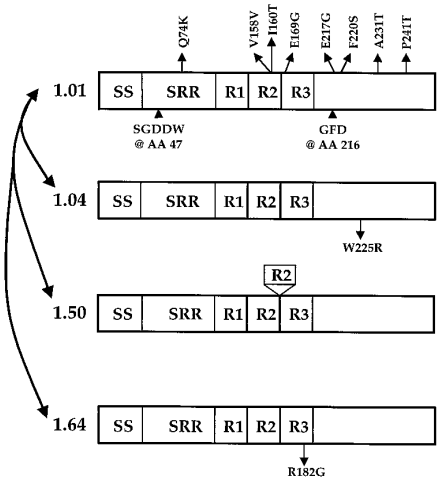

Fig. 4 Sic variants identified among associated isolates. All five events identified with isolates expressing distinct Sic variants involved Sic1.01. Two events had Sic1.01 and Sic1.50, and one event each had Sic1.01 and Sic1.04 or Sic1.64. The polymorphisms shown in the Sic1.01 variant are relative to the Sic variant originally described[14]. Sic1.04, Sic1.50 and Sic1.64 are identical to Sic1.01 except for the polymorphisms shown. The single-letter amino acid code is used. SRR, short repeat region; R1, R2 and R3, repeat regions; AA, amino acid.

persistently colonized with MGAS 5005 expressing Sic1.01. During the 80 days of the experiment, we analyzed 296 colonies by PCR for size variation in the *sic* gene. In the first 45 days, we assessed 198 colonies; only two had a *sic* PCR product that differed in size from the expected approximately 1-kb fragment. One colony had an additional R2 insert and one had two additional R2 inserts. PCR analysis showed that 73 of 98 colonies (74%) obtained from four of the five colonized mice on day 54 had *sic* size variants. Next, we sequenced *sic* genes from 20 colonies selected to represent the breadth of size variation and from all four colonized animals. Size variation was due to insertions in the R2 region that involved addition of the R2 repeat. Most colonies had *sic* with two or three additional R2 repeats, but some colonies had as many as ten additional R2 repeats (data not shown).

To test the hypothesis that the interaction of host and pathogen contributed to the high frequency of Sic size variation observed in the model, we analyzed by PCR 596 colonies grown *in vitro*. Only one colony had a variant *sic* PCR product. There was a highly significant ($\chi^2 = 488$; $P < 0.0001$) difference in the proportion of colonies with Sic size variation between the GAS grown *in vitro* and those recovered after 54 days of mammalian mucosal colonization.

Inhibition of complement by Sic variants

The identification of organisms expressing Sic structural variants in the throats of persistently colonized mice indicated that these isolates successfully survived host defenses operative in that niche; thus, the Sic variants produced by the surviving organisms should be competent inhibitors of complement-mediated cell lysis. We purified two of the Sic variants (called C1 and C21; Fig. 3) recovered from colonized mice and compared their ability to inhibit lysis of sensitized red blood cells[14] with that of Sic1.01 and Sic1.120. Sic1.01 is made by the strain used to initially colonize the mice and has one R2 insert. Sic1.120 is made by a strain that caused human invasive disease and has three R2 inserts. The two Sic variants made by the strains that successfully competed against host defenses inhibited complement as well as the Sic1.01 and Sic1.120 variants (Fig. 3).

Sic variation in epidemiologically related episodes

Because all evidence indicated rapid selection was involved in producing new Sic variants, we considered that it might be possible to identify Sic variation in epidemiologically associated strains. We therefore sequenced *sic* in 170 M1 strains cultured from individuals known through intensive investigation to be linked in short time periods in 43 associated events. Of the 43 distinct situations studied, five had M1 isolates with different Sic variants. These epidemiologically related variants always differed from one another by only one molecular event such as an insertion, deletion or single-nucleotide change resulting in an amino-acid replacement (Fig. 4).

Discussion

The ecological and biomedical factors underlying microbial epidemic waves have generally been studied with data obtained from relatively few strains thought to be representative of all organisms causing infection in the period investigated. These studies have led to the conventional wisdom that most microbial epidemic waves are caused by clones having little or no genetic variation, although RNA viruses with unusually high mutation rates[1–7,31–33] are important exceptions. However, M1 GAS epidemic waves clearly are composed of a highly polymorphic array of subclones defined by distinct Sic variants that arise by natural selection on mucosal surfaces. Although many pathogen and host factors contribute to GAS epidemic waves, our data favor a general model in which selection of new Sic structural variants on mucosal surfaces generates a very large 'pool' of subclones in the course of epidemic waves. This process may help to sustain and enlarge the epidemic waves and contribute to M1 GAS emergence and reemergence.

We recovered Sic size variants with significantly increased frequency after approximately 5–6 weeks of mouse throat colonization, a time frame indicating that acquired immunity was involved. All size variants had gained one or more R2 repeat units, a type of variation that was also abundantly represented among human invasive infections and pharyngitis. The data indicate that host antibody is involved in rapid Sic diversification on mucosal surfaces, and provide a plausible explanation for their rapid appearance and abundance in humans. Implicit in this model is the idea that the infected host makes antibody against polymorphic regions of Sic. Preliminary study of the serological response in humans and mice infected with M1 strains indicates that this is the case (N.P.H. *et al.*, unpublished data).

Variation in M1 protein structure does not contribute to changes in the frequency of invasive disease or pharyngitis. This failure is perplexing, given that almost all nucleotide substitutions in *emm1* result in amino-acid replacements[12,13,17,25], which indicates that natural selection is also operative. Notwithstanding structural constraints, the simplest interpretation is that natural selection contributes to the diversification of both Sic and M1, but that the selective pressure to vary Sic exceeds the pressure to alter M1 protein. The findings emphasize the need for renewed examination and delineation of the pertinent selective forces that function on human mucosal surfaces. In addition, it will also be essential for future studies to deter-

mine if Sic variation enhances the relative fitness of M1 GAS, or otherwise alters the character of host-pathogen interaction.

Our demonstration of profound variation in an extracellular bacterial protein that arises very rapidly by natural selection may be relevant to other pathogens causing epidemic waves thought to be mono- or pauciclonal. Identification of the host forces responsible for rapid selection will contribute new understanding of pathogen epidemics and may provide new strategies to predict and prevent pathogen reemergence.

Methods

Bacterial strains. Serotype M1 GAS strains cultured from patients (n = 892) with invasive disease episodes in Finland; Ontario, Canada; and four United States sites were studied (Table). The strains were recovered from prospective, population-based surveillance of invasive GAS disease and represent almost all M1 infections in the times studied[11,18,19,34]. Invasive disease was defined broadly as infection of any normally sterile body site, generally blood. M1 isolates (n = 240) randomly selected from strains causing pharyngitis in Finland in 1992–1997 and collected partly in a population-based countrywide surveillance in 1994–1997 were also studied. Because M1 pharyngitis cases far exceed the number of invasive episodes, the isolates are not a comprehensive collection, but instead represent the pharyngitis cases occurring in Finland in the same time period as the invasive organisms analyzed. GAS strains were identified as M1 either by conventional serologic techniques or by genotypic methods[11,18,19,34].

Molecular strain characterization. The *sic* and *emm1* gene sequencing, and IS*1548* profiling and PCR and sequence analysis of the direct repeat region were done as described[12,16,17,25].

Variation in *sic* during laboratory and mouse strain passage. Adult CD-1 outbred mice were inoculated intranasally with 3.1×10^7 colony-forming units of an arbitrarily chosen M1 strain (MGAS 5005) expressing Sic variant 1.01. This variant was selected for analysis because it is one of the more abundant Sic proteins identified in all localities studied. Bacteria were grown overnight in Todd-Hewitt broth, collected by centrifugation, and washed once with sterile ice-cold, pyrogen-free PBS. The absorbance at 600 nm was measured and the culture volume was adjusted to give the required inoculum in a 0.05-ml volume. The number of colony-forming units in the inoculum were verified by colony counts on tryptose agar plates containing 5% sheep blood (Becton Dickinson, San Jose, California). Mouse nasopharynges were swabbed periodically and blood agar plates were inoculated and cultured overnight at 37 °C in an atmosphere supplemented with 5% CO_2.

Purification of Sic protein and complement inhibition studies. Sic was purified from culture supernatants of GAS grown in Todd-Hewitt broth supplemented with 0.2% yeast extract (THY medium). Bacteria were grown overnight at 37 °C in 300 ml of THY medium, the culture supernatant was decanted, and 60 ml of fresh THY medium was added. Aliquots (10 ml) were used to inoculate six 1-liter vessels with THY medium, which were then incubated at 37 °C until the absorbance at 600 nm reached 0.9–1.0. The bacteria were removed by centrifugation and the 'pooled' supernatant was filter sterilized. The supernatant was concentrated to 400–500 ml with a 10-kDa spiral ultrafiltration cartridge (Amicon, Houston, Texas). Ammonium sulfate was added to 20% saturation and the solution was stirred for 15 min at 4 °C. The precipitate was collected by centrifugation, dissolved in 20 ml 20 mM Tris-HCl, pH 7.0 (buffer A) at 4 °C, and dialyzed overnight at 4 °C against buffer A. The dialyzed material was loaded onto a DEAE–Sepharose Fast Flow column (1.5 cm × 10 cm) equilibrated with 50 mM NaCl. The column was washed with 30 ml of 50 mM NaCl and 50 ml of 200 mM NaCl. The Sic protein was eluted with 80 ml of 300 mM, 400 mM or 500 mM NaCl, depending on the Sic variant being purified. Fractions (6 ml) containing the Sic protein were identified by SDS–PAGE. The protein was more than 98% pure, as assessed by SDS-PAGE and Coomassie blue staining. Amino-terminal sequencing confirmed the presence of the Sic protein. Electrospray mass spectroscopy analysis with a VG (Micromass, Weston, Bath, UK) Quattro tandem quadrupole mass spectrometer of QhQ geometry confirmed that the entire Sic protein had been purified.

Inhibition of complement-mediated lysis of sensitized erythrocytes was assayed in a microtiter format with Low-Tox Guinea-Pig Complement (Cedarlane Laboratories, Hornby, Ontario, Canada) and sheep erythrocytes sensitized with hemolysin (BioWhittaker, Walkersville, Maryland). Purified Sic was serially diluted in IX veronal buffer (BioWhittaker, Walkersville, Maryland) with 0.1% gelatin (weight/volume) and incubated for 30 min at 37 °C with the amount of complement activity required to lyse 30% of the erythrocytes. Sensitized erythrocytes (1.6×10^6) were then added to each well of a 96-well microtiter plate. The plate was incubated for 20 min at 37 °C and centrifuged at 298*g* for 5 min at 4 °C. Aliquots (50 µl) of the supernatants were transferred to a new plate, and the absorbance was measured at 415 nm with a Spectramax PLUS instrument (Molecular Devices, Sunnyvale, California).

Phylogenetic and statistical analyses. Unrooted evolutionary trees were reconstructed with the maximum parsimony method using *sic* nucleotide sequences[21–24]. The trees obtained are unrooted because the ancestral sequence of the sample is not known. Insertions and deletions were weighted equally as nucleotide substitutions. All statistical tests for significance used chi square analysis and Bonferroni correction for multiple comparisions when appropriate.

Acknowledgments

We thank W.A. Keitel and J. Dale for critical reading of the manuscript, R.S. Orkiszewski for the mass spectroscopy analysis, I. Eckstrand for support, and S. Ruusunen, R. Scotford and E. Siren for technical assistance. We thank A. Muotiala, H. Seppala, P. Huovinen and the Finnish Study Group for Antimicrobial Resistance for their work and support during collection of the Finnish strains. Specimens from the United States were provided by the Emerging Infections Program Network/Active Bacterial Core Surveillance, supported by a cooperative agreement between the Centers for Disease Control and Prevention and the State Health Departments of California, Connecticut, Georgia, and Minnesota. This study was supported by United States Public Health Service Grants AI-33119 to J.M.M. and GM-50428 to Y.-X. F., and the Canadian Bacterial Disease Network.

RECEIVED 8 MARCH; ACCEPTED 25 MAY 1999

1. Krause, R.M. & Fauci, A. in *Emerging Infections* (Academic Press, San Diego, 1998).
2. Hennessy, T.W. *et al.* A national outbreak of *Salmonella enteritidis* infections from ice cream. *N. Engl. J. Med.* **334**, 1281–1286 (1996).
3. Armstrong, G.L., Hollingsworth, J. & Morris, J.G.Jr. Emerging foodborne pathogens: *Escherichia coli* O157:H7 as a model of entry of a new pathogen into the food supply of the developed world. *Epidemiol. Rev.* **18**, 29–51 (1996).
4. Piffaretti, J.-C. *et al.* Genetic characterization of clones of the bacterium *Listeria monocytogenes* causing epidemic disease. *Proc. Natl. Acad. Sci. USA* **86**, 3813–3822 (1989).
5. Wachsmuth, I.K., Blake, P.A. & Olsvik, O. *Vibrio cholerae and Cholera* (American Society for Microbiology Press, Washington, D.C., 1994).
6. Morelli, G. *et al.* Clonal descent and microevolution of *Neisseria meningitidis* during 30 years of epidemic spread. *Mol. Microbiol.* **25**, 1047–1064 (1997).
7. Webster, R.G. in *Emerging Infections* (eds. Krause, R.M & Fauci, A.) 275–300 (Academic Press, San Diego, 1998).
8. Fischetti, V.A. Streptococcal M protein: molecular design and biologic behavior. *Clin. Microbiol. Rev.* **2**, 285–314 (1989).
9. Musser, J.M. & Krause, R.M. in *Emerging Infections* (eds. Krause, R.M. & Fauci, A.) 185–218 (Academic Press, San Diego, 1998).
10. Martin, D.R. & Single, L.A. Molecular epidemiology of group A streptococcus M type 1 infections. *J. Infect. Dis.* **167**, 1112–1117 (1993).
11. Muotiala, A., Seppala, H., Huovinen, P. & Vuopio-Varkila, J. Molecular comparison of group A streptococci of T1M1 serotype from invasive and noninvasive infections in Finland. *J. Infect. Dis.* **175**, 392–399 (1997).
12. Musser, J.M. *et al.* Genetic diversity and relationships among *Streptococcus pyogenes* strains expressing serotype M1 protein: recent continental spread of a subclone causing episodes of invasive disease. *Infect. Immun.* **63**, 994–1003 (1995).
13. Cleary, P.P. *et al.* Clonal basis for the resurgence of serious *Streptococcus pyogenes* disease in the 1980s. *Lancet* **339**, 518–521 (1992).
14. Akesson, P., Sjoholm, A.G. & Bjorck, L. Protein SIC-a novel extracellular protein of *Streptococcus pyogenes* interfering with complement function. *J. Biol. Chem.* **271**, 1081–1088 (1996).

15. Lubinski, J.M. *et al.* Herpes simplex virus type 1 glycoprotein gC mediates immune evasion *in vivo*. *J. Virol.* **72**, 8257–8263 (1998).
16. Stockbauer, K.E. *et al.* Hypervariability generated by natural selection in an extracellular complement-inhibiting protein of serotype M1 strains of group A *Streptococcus*. *Proc. Natl. Acad. Sci. USA* **95**, 3128–3133 (1998).
17. Hoe, N. *et al.* Rapid subtyping of serotype M1 Group A *Streptococcus* strains by automated sequencing of the *sic* gene encoding streptococcal inhibitor of complement and a chromosomal region with multiple direct repeat sequences. *Emerg. Infect. Dis.* **5**, 254–263 (1999).
18. Davies, H.D. *et al.* Invasive group A streptococcal infections in Ontario, Canada. *N. Engl. J. Med.* **335**, 547–554 (1996).
19. Zurawski, C.A. *et al.* Invasive group A streptococcal disease in metropolitan Atlanta: a population-based assessment. *Clin. Infect. Dis.* **27**, 150–157 (1998).
20. Ramsay Taylor, W. The classification of amino acid conservation. *J. Theor. Biol.* **119**, 205–218 (1986).
21. Fitch, W.M. On the problem of discovering the most parsimonious trees. *Am. Nat.* **3**, 223–257 (1977).
22. Swofford, D.L. in *PAUP: Phylogenetic Analysis Using Parsimony* (Illinois Natural History Survey, Champaign, Illinois, 1993).
23. Tajima, F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. *Genetics* **123**, 585–595 (1989).
24. Fu, Y.-X. Statistical properties of segregating sites. *Theor. Popul. Biol.* **48**, 172–197 (1995).
25. Harbaugh, M. P., Podbielski, A., Hugl, S. & Cleary, P.P. Nucleotide substitutions and small-scale insertion produce size and antigenic variation in group A streptococcal M1 protein. *Mol. Microbiol.* **8**, 981–991 (1993).
26. de Malmanche, S.A. & Martin, D.R. Protective immunity to the group A *Streptococcus* may be only strain specific. *Med. Microbiol. Immunol.* **183**, 299–306 (1994).
27. Jones, K.F., Hollingshead, S.K., Scott, J.R. & Fischetti, V.A. Spontaneous M6 protein size mutants of group A streptococci display variation in antigenic and opsonogenic epitopes. *Proc. Natl. Acad. Sci. USA* **85**, 8271–8275 (1988).
28. Fischetti, V.A., Jarymowycz, M., Jones, K.F. & Scott, J.R. Streptococcal M protein size mutants occur at high frequency within a single strain. *J. Exp. Med.* **164**, 971–980 (1986).
29. Hollingshead, S.K., Fischetti, V.A. & Scott, J.R. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. *Mol. Gen. Genet.* **207**, 196–203 (1987).
30. Villasenor, A., McShan, W.M., Salmi, D. & Stevens, D.L. Variation in susceptibility to opsono-phagocytosis by strains of a M-1 clone of group A Streptococcus (GAS) could be associated with diversity in the sequence of the *emm1* gene. Abstract E-61 (American Society for Microbiology 98th General Meeting, Atlanta, Georgia, 1998).
31. Ou, C-Y. *et al.* Molecular epidemiology of HIV transmission in a dental practice. *Science* **256**, 1165–1171 (1992).
32. Garcia, O. *et al.* Evolutionary pattern of human respiratory syncytial virus (subgroup A): cocirculating lineages and correlation of genetic and antigenic changes in the G glycoprotein. *J. Virol.* **68**, 5448–5459 (1994).
33. Domingo, E. & Holland, J.J. RNA virus mutations and fitness for survival. *Annu. Rev. Microbiol.* **51**, 151–178 (1997).
34. Beall, B., Facklam, R., Hoenes, T. & Schwartz, B. Survey of gene sequences and T-antigen types from systemic *Streptococcus pyogenes* infection isolates collected in San Francisco, California; Atlanta, Georgia; and Connecticut in 1994 and 1995. *J. Clin. Microbiol.* **35**, 1231–1235 (1997).