Evolutionary Origin and Emergence of a Highly Successful Clone of Serotype M1 Group A *Streptococcus* Involved Multiple Horizontal Gene Transfer Events

Paul Sumby,[1,2] Steve F. Porcella,[1] Andres G. Madrigal,[1,a] Kent D. Barbian,[1] Kimmo Virtaneva,[1] Stacy M. Ricklefs,[1] Daniel E. Sturdevant,[1] Morag R. Graham,[1,a] Jaana Vuopio-Varkila,[3] Nancy P. Hoe,[1] and James M. Musser[1,2]

[1]Laboratory of Human Bacterial Pathogenesis, Rocky Mountain Laboratories, National Institute of Allergy and Infectious Diseases, National Institutes of Health, Hamilton, Montana; [2]Center for Human Bacterial Pathogenesis Research, Department of Pathology, Baylor College of Medicine, Houston, Texas; [3]National Public Health Institute, Helsinki, Finland

To better understand the molecular events involved in the origin of new pathogenic bacteria, we studied the evolution of a highly virulent clone of serotype M1 group A *Streptococcus* (GAS). Genomic, DNA-DNA microarray, and single-nucleotide polymorphism analyses indicated that this clone evolved through a series of horizontal gene transfer events that involved (1) the acquisition of prophages encoding streptococcal pyrogenic exotoxin A and extracellular DNases and (2) the reciprocal recombination of a 36-kb chromosomal region encoding the extracellular toxins NAD^+-glycohydrolase (NADase) and streptolysin O (SLO). These gene transfer events were associated with significantly increased production of SLO and NADase. Virtual identity in the 36-kb region present in contemporary serotype M1 and M12 isolates suggests that a serotype M12 strain served as the donor of this region. Multiple horizontal gene transfer events were a crucial factor in the evolutionary origin and emergence of a very abundant contemporary clone of serotype M1 GAS.

The molecular events involved in the emergence of unusually virulent pathogens are poorly understood. Bacteria can evolve slowly through the accumulation of point mutations or rapidly in an "evolutionary quantum leap" by acquisition of new genetic material through horizontal gene transfer events [1]. New combinations of genes or allelic variants can increase bacterial fitness by conferring unique properties, such as the ability to colonize a previously unexploitable ecological niche or to resist components of the immune system. A number of well-described processes are known to promote horizontal gene transfer events, including prophage transduction, plasmid conjugation, and transformation.

Group A *Streptococcus* (GAS) is responsible for many serious human infections, such as necrotizing fasciitis (NF; also called "the flesh-eating syndrome") and streptococcal toxic shock syndrome (STSS) [2, 3]. However, the most common forms of streptococcal infection are noninvasive, non–life-threatening infections, such as pharyngitis and impetigo. Host and bacterial factors contribute to the outcome of an infection [4]. The importance of GAS factors in disease manifestation was, in part, identified because of the nonrandom distribution of GAS serotypes among particular diseases. For example, in population-based case series, serotype M1 GAS is usually the most common serotype recovered during episodes of invasive disease [5–7].

During the mid-1980s, the frequency and severity of invasive infections caused by serotype M1 GAS (e.g., septicemia, STSS, and, especially, NF) suddenly and precipitously increased (reviewed in Musser and Krause

Received 26 January 2005; accepted 1 April 2005; electronically published 29 July 2005.

Data depositions: GenBank accession number for MGAS5005 genome, CP000017.

Potential conflicts of interest: none reported.

Financial support: National Institutes of Health (grant UO1-AI-060595).

[a] Present affiliations: Boston University School of Medicine, Boston, Massachusetts (A.G.M.); Canadian Centre for Human and Animal Health, Winnipeg, Canada (M.R.G.).

Reprints or correspondence: Dr. James M. Musser, Center for Human Bacterial Pathogenesis Research, Dept. of Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030 (musser@bcm.tmc.edu).

The Journal of Infectious Diseases 2005;192:771–82

0022-1899/2005/19205-0008$15.00

Table 1. Serotype M1 group A *Streptococcus* isolates studied.

Strain	Location	Year	*speA* genotype	*emm* allele	*sic* allele	Disease or source
MGAS1600	England	1920s	*speA1*	1.0	1.07	Scarlet fever
MGAS1374	Germany	1969	Negative	1.0	1.320	…
MGAS1402	Germany	1972	Negative	1.0	1.321	…
MGAS1264	Denmark	1973	*speA2*	1.0	1.249	Sepsis
MGAS253	California	1980s	Negative	1.3	1.3	Invasive
MGAS571	Montreal	1980s	Negative	1.3	1.21	Scarlet fever
MGAS1275	Denmark	1980	*speA2*	1.45	1.319	…
MGAS1301	Germany	1981	*speA2*	1.0	1.58	Scarlet fever
MGAS1508	Czechoslovakia	1985	*speA2*	1.8	1.26	Scarlet fever
SF370	…	1985	Negative	1.6	1.225	Wound infection
MGAS2221	Australia	1988	*speA2*	1.0	1.01	Scarlet fever
MGAS5300	Western Finland	1988	*speA2*	1.2	1.95	Throat
MGAS5305	Northern Finland	1988	*speA2*	1.0	1.01	Throat
MGAS5307	Eastern Finland	1988	*speA2*	1.0	1.01	Blood
MGAS2081	New Zealand	1989	*speA2*	1.0	ND	Cellulitis
MGAS5316	Western Finland	1989	*speA2*	1.0	1.01	Blood
MGAS5322	Eastern Finland	1989	*speA2*	1.0	1.01	Throat
MGAS5337	Western Finland	1989	*speA2*	1.0	1.01	Pus
MGAS5348	Central Finland	1989	*speA2*	1.0	1.01	Blood
MGAS290	Colorado	Late 1980s	*speA2*	1.0	1.36	Invasive
MGAS294	Washington	Late 1980s	*speA2*	1.0	1.78	Invasive
MGAS313	Wyoming	Late 1980s	*speA2*	1.0	1.34	STSS
MGAS337	New York	Late 1980s	*speA2*	1.0	1.17	Throat
MGAS2217	New Zealand	1990s	*speA1*	1.0	1.07	Invasive
MGAS1284	Denmark	1990	*speA2*	1.0	1.01	…
MGAS1573	Germany	1990	*speA2*	1.0	1.01	…
MGAS5393	Central Finland	1990	*speA2*	1.0	1.01	Throat
MGAS5406	Northern Finland	1990	*speA2*	1.0	1.01	Throat
MGAS5411	Central Finland	1990	*speA2*	1.0	1.112	Blood
MGAS5415	Southern Finland	1990	*speA2*	1.0	1.01	Blood
MGAS5432	Western Finland	1991	*speA2*	1.0	1.01	Blood
MGAS4943	Ontario	1992	*speA2*	1.0	1.02	Bacteremia
MGAS5087	Ontario	1993	*speA2*	1.0	1.01	Lower respiratory tract
MGAS5171	Ontario	1994	*speA2*	1.0	1.218	Necrotizing fasciitis
MGAS5456	Southern Finland	1994	*speA2*	1.23	1.01	Throat
MGAS5459	Northern Finland	1994	*speA2*	1.0	1.123	Throat
MGAS5475	Northern Finland	1994	*speA2*	1.0	1.01	Throat
MGAS5508	Southern Finland	1994	*speA2*	1.0	1.01	Pus
MGAS3437	Ontario	1995	*speA2*	1.0	1.01	Soft tissue
MGAS3680	Virginia	1995	*speA2*	1.0	1.01	…
MGAS5531	Eastern Finland	1995	*speA2*	1.0	1.01	Throat
MGAS5579	Western Finland	1995	*speA2*	1.0	1.01	Pus
MGAS6413	Georgia	1995	*speA2*	1.0	1.01	Blood
MGAS5005	Ontario	1996	*speA2*	1.0	1.01	Cerebrospinal fluid
MGAS5669	Southern Finland	1996	*speA2*	1.0	1.01	Pus
MGAS5682	Southern Finland	1996	*speA2*	1.0	1.102	Pus
MGAS5690	Central Finland	1996	*speA2*	1.0	1.01	Blood
MGAS6419	Atlanta	1996	*speA2*	1.0	1.185	…
MGAS6470	Connecticut	1996	*speA2*	1.0	1.33	…
MGAS6499	Connecticut	1996	*speA2*	1.0	1.01	Blood
MGAS5792	Western Finland	1997	*speA2*	1.0	1.01	Blood
MGAS5804	Central Finland	1997	*speA2*	1.0	1.01	Blood
MGAS5807	Central Finland	1997	*speA2*	1.0	1.135	Blood
MGAS5932	Ontario	1997	*speA2*	1.0	1.135	STSS
MGAS6617	Minnesota	1997	*speA2*	1.0	1.01	Blood
MGAS6184	Texas	1997–1998	*speA2*	1.0	1.01	Invasive
MGAS6558	Maryland	1998	*speA2*	1.0	1.244	…
MGAS6700	San Francisco	1998	*speA2*	1.0	1.234	…
MGAS6727	Eastern Finland	1998	*speA2*	1.0	1.01	Blood
MGAS9127	Alberta	Late 1990s	*speA2*	1.0	1.01	Invasive
MGAS9132	Alberta	Late 1990s	*speA2*	1.0	1.78	Invasive
MGAS9138	Alberta	Late 1990s	*speA2*	1.0	1.88	Invasive
MGAS9144	Alberta	Late 1990s	*speA2*	1.0	1.3	Invasive
MGAS9150	Alberta	Late 1990s	*speA2*	1.0	1.163	Invasive
MGAS1922	The Netherlands	…	*speA2*	1.0	1.01	Sepsis

NOTE. *emm*, gene encoding M protein; MGAS, Musser group A *Streptococcus;* ND, not done; *sic,* gene encoding streptococcal inhibitor of complement; STSS, streptococcal toxic shock syndrome.

[3]). However, little is known about the molecular processes associated with this phenotypic shift. In the present study, we sought to understand the molecular events that contribute to the evolutionary origin and emergence of serotype M1 GAS strains through genomic, single-nucleotide polymorphism (SNP), transcriptome, and protein analyses.

MATERIALS AND METHODS

Genome sequencing. The genome of strain MGAS5005 was sequenced by methods previously used to sequence the genome of serotype M6 and M28 GAS strains [8, 9]. Directed sequencing was performed, to increase the minimum consensus-base quality to Q40 for regions of low sequence quality in the assembled genome. The entire genome was tiled by polymerase chain reaction (PCR) after closure, to ascertain the validity of the assembly. Open reading frames (ORFs) were identified by use of proprietary software (Integrated Genomics) and were entered into the ERGO bioinformatics suite for annotation [10]. The genome sequence has been deposited in GenBank (accession number CP000017). Strain MGAS5005 has been deposited in the American Type Culture Collection (number BAA-947).

Bacterial strains. Sixty-five serotype M1 GAS strains were studied (table 1). The sampling of these organisms was based on the genetic analysis of ~2000 serotype M1 strains isolated from patients on several continents. These studies have been conducted during the last 13 years in the laboratory of the senior author (J.M.M.). Importantly, these ~2000 serotype M1 strains include organisms from comprehensive population-based studies conducted over the course of many years in the United States, Canada, Germany, and Finland [11–16]. Sampling of strains for molecular analysis was done randomly on the basis of our knowledge of the distribution of serotype M1 subclones present in the populations studied. Additional strains studied were selected at random from the senior author's strain collection, which includes >12,000 isolates from 36 countries worldwide. The strains studied were not subject to systematic sampling bias. The inclusion criteria were designed to provide a temporally and spatially diverse collection of isolates to study. MGAS5005 was isolated in 1996 from the cerebrospinal fluid of an infected patient in Ontario, Canada, and has been used extensively in studies of GAS pathogenesis [17–19].

GAS DNA-DNA microarray analysis. Two spotted microarrays were used for the DNA-DNA hybridizations, to compare old and contemporary serotype M1 GAS isolates. Twenty-nine isolates, of which 21 had been isolated in Finland as part of a population-based study, were analyzed by use of a DNA microarray containing 1702 ORFs that represented 97% of the genome of serotype M1 strain SF370 [20]. In addition, these isolates were analyzed by use of a GAS microarray composed of the above 1702 M1 ORFs supplemented with ORFs uniquely present (relative to strain SF370) in the serotype M18 strain MGAS8232 [21] and the serotype M3 strain MGAS315 [22]. Most of the supplemental ORFs were encoded by prophages present in the genome of the serotype M3 and M18 strains sequenced. SF370 was isolated in 1985 from the infected wound of a patient, and this isolate produces low NAD^+-glycohydrolase (NADase) activity, similar to all pre-1988 serotype M1 GAS isolates [20, 23].

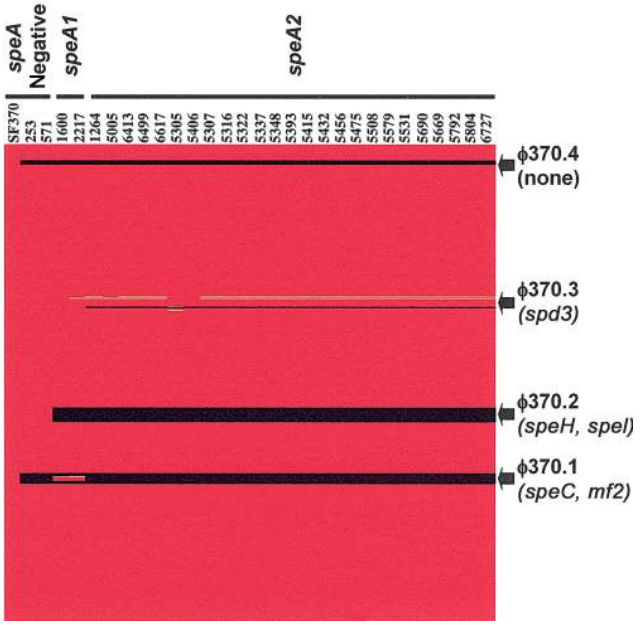

Figure 1. Schematic comparing the gene content of 29 serotype M1 group A *Streptococcus* (GAS) isolates with that of reference strain SF370. The chromosomal position of each gene of SF370 (1702 open reading frames [ORFs]) is shown from top to bottom. The gene content of the 29 serotype M1 isolates relative to strain SF370 is shown from left to right. Isolates are grouped on the basis of the presence or absence of the *speA* gene. Genes were identified as present *(red),* absent *(black),* or divergent *(green)* by microarray analysis, polymerase chain reaction mapping, and targeted sequencing. Below each prophage designation, putative virulence factor genes present in the prophage are indicated.

Bacteria were grown overnight in Todd-Hewitt broth (Difco Laboratories) supplemented with 0.2% yeast extract (THY broth). Chromosomal DNA was isolated by use of the Puregene DNA isolation kit (Gentra Systems). Probe preparation and microarray hybridization experiments were performed as described elsewhere [21, 24]. Differentially hybridizing spots were identified for each strain by use of QuantArray analysis software (Packard BioScience). ORFs defined as "absent" were either not present in the test strain genome or had sequences divergent enough to prevent hybridization with strain SF370 DNA under highly stringent conditions. DNA-DNA microarray results were verified by PCR mapping and targeted DNA sequencing.

SNP analysis. PCR primers were designed to flank regions containing putative SNPs (table 2). PCR amplifications were purified by use of the Qiagen PCR purification kit and then se-

Table 2. Polymerase chain reaction primers used for single-nucleotide polymorphism (SNP) analysis.

Primer	Sequence (5′→3′)	Gene containing SNP	Within 36-kb region?a
SNP1F	ATGGTGCCTGGAGTATTTGC	M5005_Spy0158	Yes
SNP1R	CCAGCTTTGGCTTTTTCTTG	M5005_Spy0158	Yes
SNP2F	ACATGCCTCGTCCAGTCTTC	M5005_Spy1612	No
SNP2R	TACCGCCATCAAACTCAACA	M5005_Spy1612	No
SNP3F	CGCTCACCTTACCATCCATT	M5005_Spy1159	No
SNP3R	TGTCACGATTCACGCTAAAA	M5005_Spy1159	No
SNP4F	AACGAGCAGAGCGTGGTATT	M5005_Spy0691	No
SNP4R	ACTGGTAAGCGGCCAATAA	M5005_Spy0691	No
SNP5F	GGAGAAGAAGGGCTAAATCATC	M5005_Spy0663	No
SNP5R	TAAGTGCCTGTTAAGGTTGCAG	M5005_Spy0663	No
SNP6F	TGGGCTAGTAGCTGCCTTGT	M5005_Spy0572	No
SNP6R	CCAAAGAAGTAAAGTTCGATTGC	M5005_Spy0572	No
SNP7F	AGAGGGACTCTTCTCTCAATGC	M5005_Spy1070	No
SNP7R	GTTTGATGCAGGTGTGTTGG	M5005_Spy1070	No
SNP8F	GCTGGAAAAACTACGGTTGC	M5005_Spy0446	No
SNP8R	CCCAACTCCTTTGGGAAAAT	M5005_Spy0446	No
SNP9F	TTCCGAATTCTCAACTGAACG	M5005_Spy0452	No
SNP9R	TTCCTGTACCATCATCATCTGG	M5005_Spy0452	No
SNP10F	CGAAGTGGTGAATACCTATCCAT	M5005_Spy1367	No
SNP10R	GAGATAACTCATTTACTTCTTGCT	M5005_Spy1367	No
SNP11F	CATTTTCCAGTAGGAGCAGCTT	M5005_Spy0943	No
SNP11R	TGGTCGGCAAAGGTAACAC	M5005_Spy0943	No
SNP12F	ACGCTGACGACTGATCCTCT	M5005_Spy1811	No
SNP12R	ACAAAGGCTTGTGGGTCTG	M5005_Spy1811	No
SNP13F	AAACCATCAACAACCCCTTG	M5005_Spy0146	Yes
SNP13R	ACCGTCAAATCCGCTCTT	M5005_Spy0146	Yes
SNP14F	TGCTGTGGGTCACGAAGTAG	M5005_Spy0151	Yes
SNP14R	TTCACCCCAAGTTTCTCCAG	M5005_Spy0151	Yes
SNP15F	ATTGTGGCTAAAATTTTGGAT	M5005_Spy0159	Yes
SNP15R	CCAGATTAATCATAGCGATTTTG	M5005_Spy0159	Yes
SNP16F	TGTGGTCGAAGGTGATTTGA	M5005_Spy0153	Yes
SNP16R	TCAAGCGATTCATCCTAGCC	M5005_Spy0153	Yes
SNP17F	AACAGCATGTGGCAATGGTA	M5005_Spy0149	Yes
SNP17R	AGACCAACCAAGTGGCCTTT	M5005_Spy0149	Yes
SNP18F	CCTTTGGACTTCCTGCTGAG	M5005_Spy0147	Yes
SNP18R	CTGGCCAGTCAACTTCTTCC	M5005_Spy0147	Yes
SNP19F	TGTTAGCTGCACCAAGTCG	M5005_Spy1319	No
SNP19R	CGGAAATTCTATGGGGACAA	M5005_Spy1319	No
SNP20F	AACATTCCGCTCTCGCTAAA	M5005_Spy1215	Yes
SNP20R	GCGATAAAGGTGAGCGATT	M5005_Spy1215	Yes
SNP21F	GTTTTATGGGATGGATCCTTGT	M5005_Spy0008	No
SNP21R	CATTTCTCCTTCACGCGATA	M5005_Spy0008	No
SNP22F	AAAGAAGATGTCCTACCA	M5005_Spy0288	No
SNP22R	CACAGATTTGCCTCAATCG	M5005_Spy0288	No
SNP23F	GTTCCTCACCTGGGAAAATCTC	M5005_Spy0162	Yes
SNP23R	ATCTATTCGCCACCACACTAGG	M5005_Spy0162	Yes
SNP24F	CCCTCAGCGTGATATCCTTC	M5005_Spy0732	No
SNP24R	CTGTCAGCATCTCCTGTGC	M5005_Spy0732	No
SNP25F	ATGCCTTTAGGGGCGTAACT	M5005_Spy0499	No
SNP25R	TGCTATTTTTGCTGCATTGG	M5005_Spy0499	No
SNP26F	CGGCACAACCGGTAAACTAT	M5005_Spy0609	No
SNP26R	TGTTCCCAGAATAAGCAAAA	M5005_Spy0609	No
SNP27a	CCCAAGCAGCCTTTATCCCTG	M5005_Spy1566	No
SNP27b	TCACCACTTGAAACTCACTCCC	M5005_Spy1566	No
SNP28a	TTGTTGGCTCGCTTCGTGAAG	M5005_Spy1670	No
SNP28b	ATAGAGAAGTTGTAAACTGGTG	M5005_Spy1670	No
SNP31a	GTTGGTGACTTTGCAGATCGTG	M5005_Spy1724	No
SNP31b	TCTGGTCGATAAAACGGTTGG	M5005_Spy1724	No
SNP32a	GATATGATCATGATAAAGGTAGAAG	M5005_Spy0385	No

(continued)

Table 2. ***(Continued.)***

Primer	Sequence (5′→3′)	Gene containing SNP	Within 36-kb region?[a]
SNP32b	GGTCACAAAACCAAGATGAGG	M5005_Spy0385	No
SNP33a	TAGGAGCAGGTATCTCTTTGG	M5005_Spy0322	No
SNP33b	ATAATATTAAGGAAATGGGCTTG	M5005_Spy0322	No
SNP34a	CGACCTCACTATCGCCCATCTC	M5005_Spy0759	No
SNP34b	CCCTAGTAATTGGAAGCCACC	M5005_Spy0759	No
SNP35a	ATGCAGCGCAAAGTAACCATC	M5005_Spy0517	No
SNP35b	AATAATAAAACCATCAACACC	M5005_Spy0517	No
SNP36a	TTTTAGAACGTGATGTTGAGG	M5005_Spy0538	No
SNP36b	GTGGGAATGGAATAGCTTCG	M5005_Spy0538	No
SNP37a	GGGCAAGAGGAGGAAGAATG	M5005_Spy1094	No
SNP37b	GATGATGCTAATAAAAAGAGG	M5005_Spy1094	No
SNP38a	ATTGGACCAGCTTCCTTTACTC	M5005_Spy1162	No
SNP38b	CCAATCAAGCCTAGCATAGACC	M5005_Spy1162	No
SNP39a	TATCTTTTGCCATCAGCTCAAAGC	M5005_Spy0080	No
SNP39b	CTCATCCATCAACTCCTTGTC	M5005_Spy0080	No
SNP40A	GAAGATATTGTGTTTGATGGTGG	M5005_Spy0097	No
SNP40B	GAACTTGGTGGTTGAGACGTG	M5005_Spy0097	No
SNP43A	ATTAAAGACGGTAAGCATGAATTG	M5005_Spy1530	No
SNP43B	AGCTTTTGTTCCAAGATGAAG	M5005_Spy1530	No
SNP44A	TTGGTTGGCTGTTTGATGCTGG	M5005_Spy0897	No
SNP44B	GGCGTGCATGAGTAAGAAATTAC	M5005_Spy0897	No
SNP45A	GATCAATTGCAGGCAGTTGAAG	M5005_Spy0863	No
SNP45B	GGATCTTTAGGTAATAGAAGG	M5005_Spy0863	No
SNP47A	AAGGAGCCTGGGGAGAAAGAAG	M5005_Spy0555	No
SNP47B	ATTGCGTGGTCGGTATGAAGG	M5005_Spy0555	No
SNP48A	CAAACAATCAGTTGAACACTATTC	M5005_Spy1291	No
SNP48B	TACTTGACTGTCCCAGCTATCAC	M5005_Spy1291	No
SNP49A	AGATGAGATTGTCGGAGGAGTG	M5005_Spy1743	No
SNP49B	ATCTTGGGTTGCTTTTGAGAG	M5005_Spy1743	No
SNP51A	GTCAGGTGTTTCATGATCGTG	M5005_Spy1803	No
SNP51B	GCGTAAAGGTAAATGACAAGAG	M5005_Spy1803	No
SNP52A	GGAACTTCTGGGTATGAAGAGG	M5005_Spy1838	No
SNP52B	GCGGTCATTGGTTTCTTTAATCGG	M5005_Spy1838	No

[a] The 36-kb region corresponds to nt 146,000–182,000 in the MGAS5005 genome.

quenced with both forward and reverse primers. The SNP data for each strain were concatenated, resulting in a single character string (nucleotide sequence). Phylogenetic analysis was performed by use of MEGA (version 2.1; available at: http://www.megasoftware.net/) with the neighbor-joining method, in which 1000 bootstrap replicates are used and distance is calculated on the basis of the number of different SNPs.

Western immunoblot analysis of culture supernatant proteins. Bacteria were grown to exponential phase (OD_{600}, 0.25) or stationary phase (overnight), and culture supernatants were obtained by centrifugation and filtering through a 0.45-μm filter. Protein was concentrated by ethanol precipitation and was assayed for presence of streptolysin O (SLO) and NADase by techniques described elsewhere [19].

NADase enzyme-activity assay. The NADase activity secreted from old and contemporary isolates was assayed as described elsewhere [25]. Briefly, supernatants from overnight cultures were centrifuged and filtered through a 0.45-μm filter. Two-fold serial dilutions of supernatants were performed in microtiter plates with PBS. NAD^+ (Sigma), diluted in PBS, was added to a concentration of 0.67 mmol/L before incubation for 1 h at 37°C. Reactions were developed by the addition of NaOH to 2N, were incubated for 1 h in the dark, and were read macroscopically by exciting the samples with 360-nm light. Results are reported as the highest 2-fold dilution capable of hydrolyzing 100 nmol of exogenous NAD^+.

SLO enzyme-activity assay. The SLO activity secreted from old and contemporary isolates was assayed as described elsewhere [26], with slight modifications. Briefly, strains were grown to an OD_{600} of 0.25 $\pm$ 0.05, and supernatants were clarified by centrifugation and filtration (pore size, 0.22 μm). The cleared supernatant was incubated with 20 mmol/L dithiothreitol for 10 min at room temperature and was aliquoted (500 μL) into 2 tubes. Water-soluble cholesterol (25 μg) (cholesterol/methyl-β-cyclodextrin; Sigma-Aldrich), a specific SLO inhibitor, was added to one sample, and the samples were incubated at 37°C for 30 min. A 2% sheep erythrocyte/PBS suspension (250 μL) was added to each sample, was mixed by inversion, and was

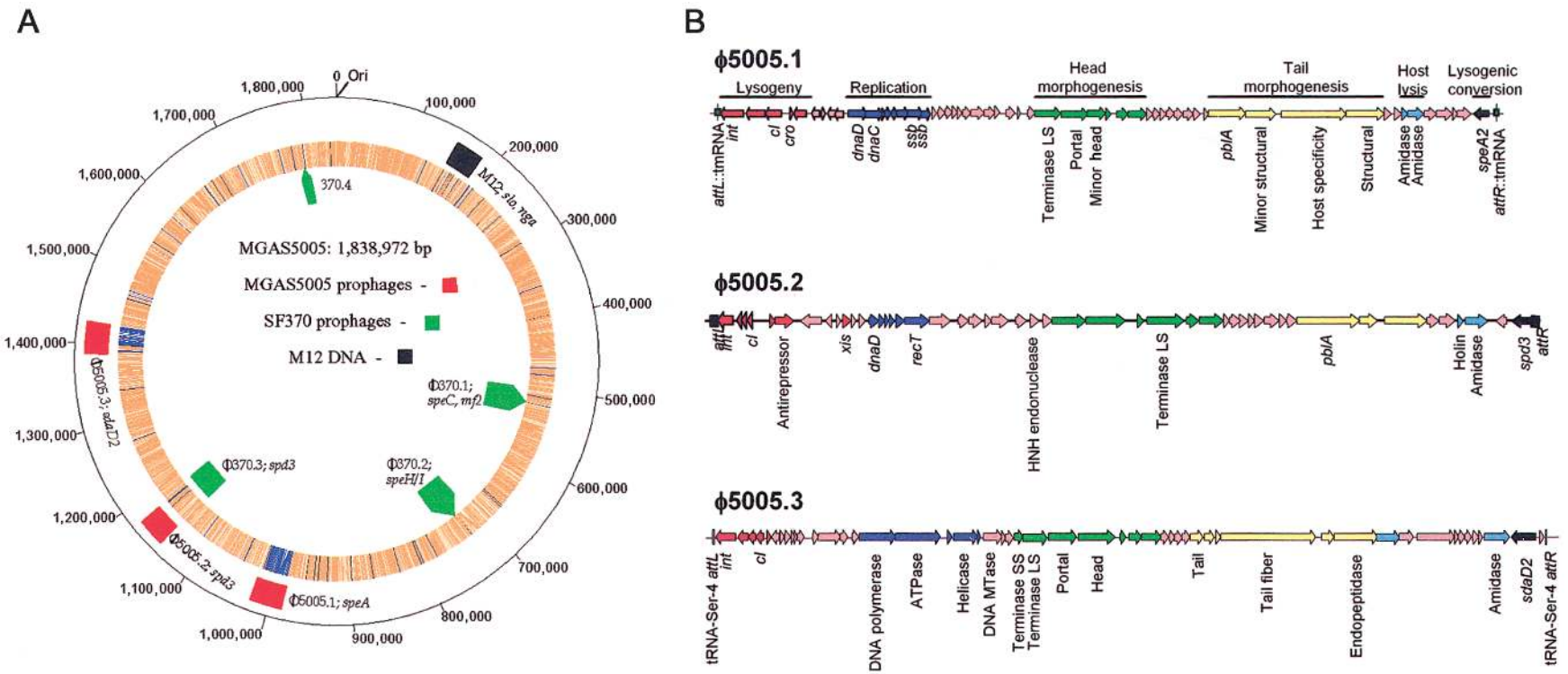

Figure 2. *A,* Atlas comparing the chromosomes of strains SF370 and MGAS5005. A BLAST analysis was performed comparing MGAS5005 open reading frames (ORFs) with those of SF370. The colored circle represents all MGAS5005 ORFs, with high-homology ORFs ($\geq e^{-10}$) being color-coded yellow and low-homology or unique ORFs ($< e^{-10}$) being color-coded blue. The red blocks outside the circle show the locations of MGAS5005 prophages; the green blocks and arrows inside the circle show the locations and insertion sites, respectively, of SF370 prophages. The location of the 36-kb M12-like region is also marked. *B,* Organization and ORF map of the 3 prophages present in the genome of strain MGAS5005. Putative ORFs are indicated by arrows that show the direction of transcription. Groups of genes whose protein products are functionally related are color coded.

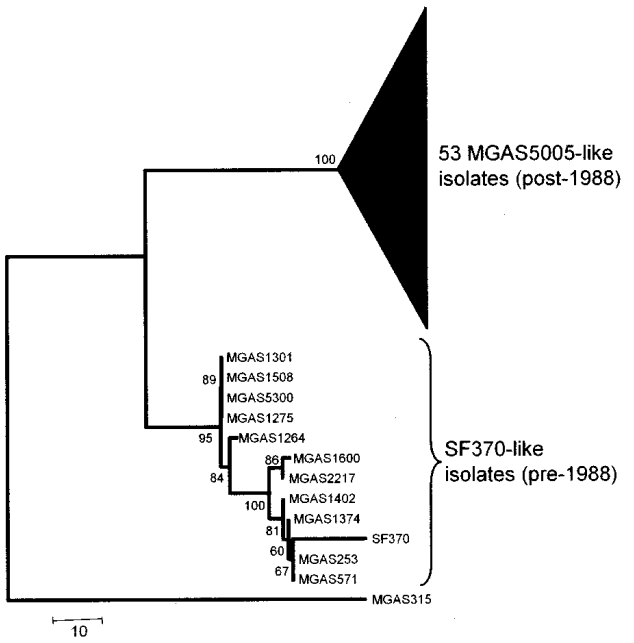

Figure 3. Genetic relationships among 65 serotype M1 group A *Streptococcus* (GAS) isolates, according to single-nucleotide polymorphism (SNP) analysis. Thirty-seven SNPs were analyzed in each of the 65 strains. The 37 SNPs are located, on average, 50 kb from each other on the GAS core chromosome. Serotype M3 strain MGAS315 was used as an outgroup.

incubated for 30 min at 37°C. PBS (500 μL) was added to each sample, and unlysed erythrocytes were removed by centrifugation at 3000 *g* for 5 min. The amount of hemoglobin present in the supernatant was measured by analysis of the optical density at 541 nm. Erythrocytes incubated in water acted as a positive control (100%) for lysis, and fresh THY broth was used as a negative control. Controls were treated exactly as experimental samples, except that medium rather than PBS was added to the water sample for the final step.

Affymetrix GeneChip expression microarray analysis. A recently described custom-made Affymetrix GeneChip was used for expression microarray studies, to compare old and contemporary genotype M1 GAS isolates [27]. The chip consisted of an antisense oligonucleotide array (18-micron size) that represented >400,000 25-mer probes (16 pairs/probe set). It contained probe sets (42,351 features) for 2662 predicted GAS ORFs that represented a composite superset of 6 GAS genome sequences (M1, M3, M5, M12, M18, and M49). The chip also contained 1925 redundant probe sets that together represented >95% of the nonredundant predicted coding regions in the genome of strain SF370.

The transcriptomes of 6 serotype M1 GAS isolates with wild-type *covRS* genes, which encode a 2-component signal transduction system that regulates ~15% of the GAS genome [28], were compared by expression microarray analysis. Four of these isolates (MGAS2221, MGAS5322, MGAS1284, and MGAS5087) had a phylogenetic lineage genotype similar to that of strain MGAS5005, whereas the other 2 isolates (MGAS1264 and MGAS1508) had a phylogenetic lineage genotype similar to that of strain SF370 (see Results). Isolates were grown overnight in THY broth at 37°C in 5% CO_2. The next morning, 2 cultures of each isolate were seeded with a 1:100 dilution of the overnight cultures in fresh, prewarmed THY broth and were incubated at 37°C in 5% CO_2. RNA was obtained from each of the 12 cultures, which were grown to an OD_{600} of 0.14 (early exponential phase). RNA isolation, cDNA synthesis, labeling, and hybridization was performed as described elsewhere [27]. Gene expression estimates were calculated for each chip by use of GCOS (version 1.1.1; Affymetrix). GAS-specific chip data were normalized across samples, to minimize discrepancies due to experimental variables (e.g., probe preparation and hybridization). A 2-sample Student's *t* test (unequal variance) was applied to the data by use of Partek Pro (version 5.1; Partek), followed by a false-discovery-rate correction ($P < .05$), to account for multiple testing [27].

Promoter comparisons. Promoter regions from genes of interest in MGAS5005 and SF370 were aligned by use of ClustalW, available at the Network Protein Sequence Analysis Web site [29].

RESULTS

Gene content variation in serotype M1 GAS strains. DNA-DNA microarray analysis revealed restricted variation in gene content among 30 serotype M1 isolates from 6 countries (figure 1). All strains had very similar core genomes corresponding to 93% of the ORFs present in the genome of the reference serotype M1 strain SF370 [20]. A maximum of 113 ORFs (7%) were absent in any one test strain, compared with those in the reference strain, and virtually all variably present ORFs were associated with prophages (figure 1). The majority of strains lacked 3 of the 4 prophages present in the reference strain. Two additional prophages (ϕ5005.1 and ϕ5005.3; see below) were identified in recent isolates by use of a DNA microarray supplemented with ORFs present in the sequenced genomes of serotypes M3 and M18 [21, 22]. One genomic profile predominated among contemporary isolates (figure 1), consistent with the hypothesis of a recent global spread of a distinct serotype M1 clone [11, 12, 30].

Together with the results of previous studies [11–14], these data suggest that reference strain SF370 is genetically distinct from the serotype M1 strains responsible for most recent human infections. To study this issue in more detail, we sequenced the genome of strain MGAS5005, a serotype M1 organism that is genetically representative of contemporary isolates and that has been used extensively in pathogenesis research [17, 18, 31]. Consistent with the data from the DNA-DNA microarray analysis, the genomes of strains MGAS5005 and SF370 were very similar (figure 2A). However, the genomes differed in terms of

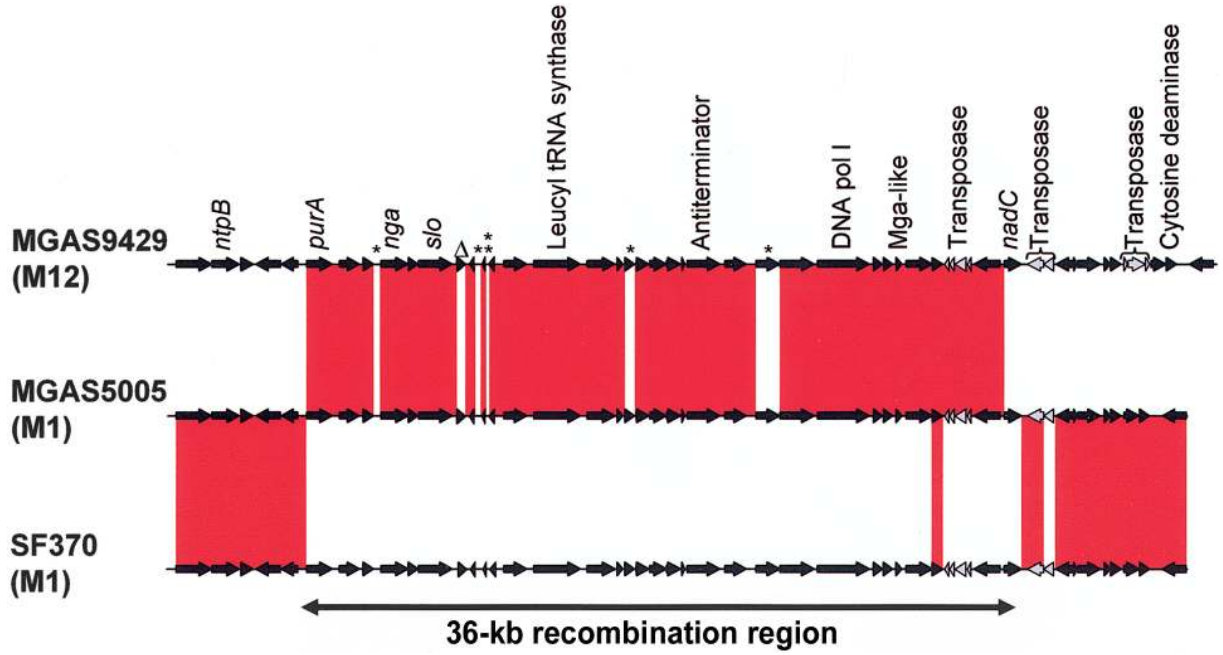

Figure 4. Open reading frame (ORF) map of a 51-kb region in the genomes of strains MGAS9429 (serotype M12), MGAS5005 (serotype M1), and SF370 (serotype M1). The schematic shows the boundaries of a presumed inter-M–type horizontal gene transfer involving ~36 kb of DNA. The red shading between ORFs indicates 100% identity at the nucleotide level. The 36-kb regions of MGAS5005 and MGAS9429 are identical with the exception of 6 single-nucleotide polymorphisms (SNPs; indicated by asterisks) and 1 in-frame deletion (Δ). In comparison, the 36-kb regions of MGAS5005 and SF370 differ by 435 SNPs and 5 deletions. Select genes of interest are labeled. Transposase ORFs are shown in gray.

prophage content, small insertions and deletions, and many SNPs. Prophages accounted for the majority of strain-specific ORFs. The genome of strain MGAS5005 contained 3 prophages (figure 2*B*) encoding a total of 3 proven or putative extracellular virulence factors—a potent pyrogenic toxin superantigen (scarlet fever toxin, SpeA2 variant [32]) and 2 DNases (Spd3 and SdaD2 [19, 33]).

SNP analysis of 65 serotype M1 GAS strains. The availability of 2 genome sequences facilitated high-resolution analysis of genetic relationships among serotype M1 GAS strains via a genomewide study of SNPs. Genetic relationships among 65 geographically and temporally diverse serotype M1 GAS isolates were investigated by analysis of 37 sequence-confirmed SNPs that are distributed throughout the genome (tables 1 and 2). Isolates recovered before the mid-1980s were genetically heterogeneous, as was shown by variation in SNP genotype (figure 3). In striking contrast, virtually all recent isolates had the same SNP genotype, regardless of their place of origin, thus confirming that they were clonally related as a consequence of recent common descent (figure 3).

Comparison of a 36-kb chromosomal region between genotype M1 and M12 isolates. Most of the SNPs that differentiate strains SF370 and MGAS5005 were distributed around the genome in apparently random fashion. However, a 36-kb region located between *purA* and *nadC* had an excessive number of SNPs (figure 4). This unusual pattern of SNP variability suggests that the 36-kb regions from these 2 strains have an evolutionary history quite different from those of the rest of the core genomes. Comparison of the 36-kb region present in strain MGAS5005 with our in-house GAS genome database revealed its virtual identity with the analogous chromosomal segment present in 2 serotype M12 strains (figure 4 and data not shown). In striking contrast, however, the sequence of the 36-kb region in strain SF370 differed considerably (figure 4). This observation was confirmed by resequencing 10,081 bp arrayed across the 36-kb region present in strains SF370 (M1), MGAS5005 (M1), and MGAS9429 (M12). Moreover, analysis of 45 SNPs in the 36-kb region in each of 65 serotype M1 strains from diverse geographic sources confirmed the key finding that old (pre-1988) and more recent (post-1988) serotype M1 strains differ in this chromosomal segment. Together, the data strongly suggest that the 36-kb region in contemporary M1 strains was very recently acquired from a serotype M12 GAS strain by horizontal gene transfer and recombinational replacement.

Comparison of SLO and NADase activity between old and contemporary M1 isolates. Two prominent extracellular toxins are encoded by genes present in this 36-kb region: SLO and NADase. SLO is a membrane-lytic toxin [26]. NADase inhibits internalization of GAS by host cells, induces apoptosis, and has potent detrimental effects on human polymorphonuclear leu-

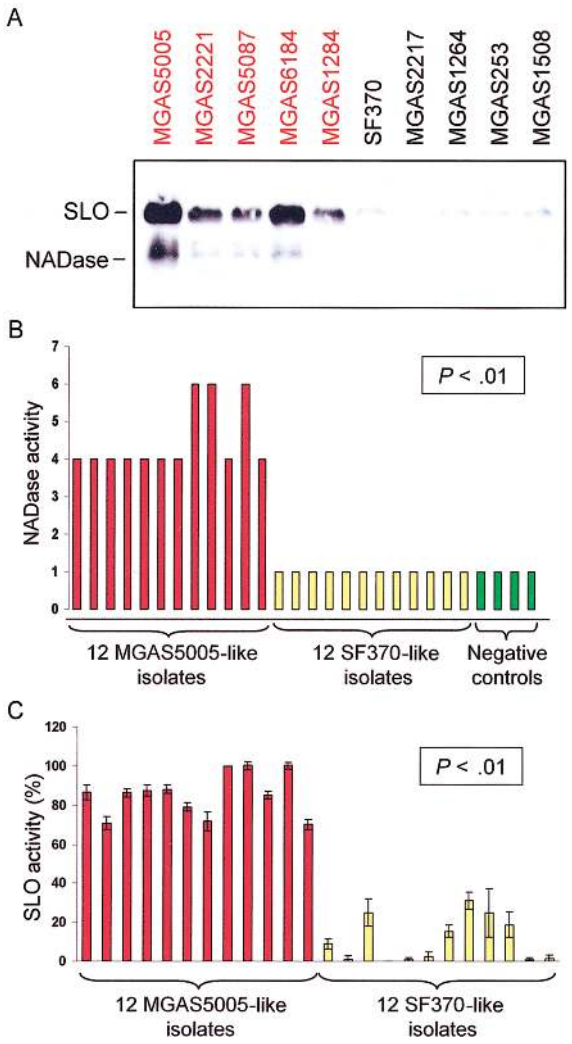

Figure 5. Detection of streptolysin O (SLO) and NAD^+-glycohydrolase (NADase) in the culture supernatant of disease-matched old (pre-1988) and contemporary (post-1988) serotype M1 group A *Streptococcus* (GAS) isolates. *A,* SDS-PAGE of concentrated culture supernatants from MGAS5005-like *(red)* or SF370-like *(black)* GAS probed with anti-SLO antibody. This antibody is known to cross-react with both SLO and NADase, because of the physical association of these proteins in the immunizing antigen preparation. *B,* NADase enzyme-activity assay, with activity reported as the highest 2-fold dilution capable of hydrolyzing 100 nmol of exogenous NAD^+. The experiment was performed in triplicate, with results identical to those shown obtained on each occasion. *C,* SLO enzyme-activity assay, with activity reported as percentage of activity relative to MGAS5005. The experiment was performed in triplicate; shown are mean values, with error bars indicating SDs.

kocytes, thereby enhancing GAS survival [23, 25, 34]. It is important to note that serotype M1 strains isolated before the mid-1980s reportedly do not produce NADase, although they do contain an intact gene *(nga)* for this enzyme [23, 34–38]. In contrast, serotype M12 strains (regardless of year of isolation) and serotype M1 strains isolated after the mid-1980s do produce NADase [23, 26, 34–37]. Furthermore, patients infected by serotype M12 strains are known to seroconvert to NADase, which indicates in vivo production in humans [34, 39, 40]. Therefore, we hypothesized that contemporary serotype M1 isolates (MGAS5005-like SNP profile with an M12-like 36-kb region) could produce extracellular NADase. This hypothesis was confirmed by Western immunoblot analysis (figure 5*A*). Immunologically reactive protein also was present in the supernatants of older isolates (SF370-like), albeit at substantially lower levels. A concomitant increase in NADase activity in the supernatants of contemporary isolates versus older isolates also was observed ($P < .01$) (figure 5*B*). SLO had a similar pattern of increased protein and activity levels in the supernatants of contemporary isolates ($P < .01$) (figure 5*A* and 5*C*).

Comparison of global gene expression between old and contemporary M1 isolates. Our genetic data indicated that serotype M1 strains with a clonal genotype similar to that of strain MGAS5005 were responsible for the unusually severe infections occurring after the mid-1980s. This was in keeping with our previous finding that, in mice, strain MGAS5005 is more virulent than strain SF370 [31]. Although we considered it to be likely that differences in expression of SLO, NADase, and prophage-encoded virulence factors contributed to the high-virulence phenotype, it was also possible that members of the more recent M1 subclone differed from the older strains in their expression of many other genes. To test this hypothesis, the transcriptomes of 4 strains with a phylogenetic lineage genotype similar to that of strain MGAS5005 and 2 strains with a phylogenetic lineage genotype similar to that of strain SF370 GAS were compared by use of an Affymetrix GeneChip. These

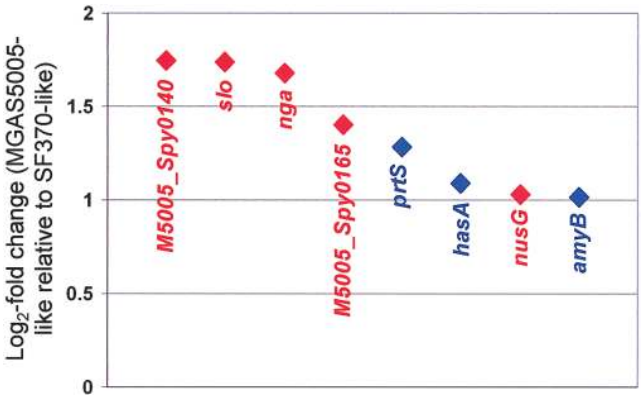

Figure 6. Expression microarray analysis of the transcriptomes of SF370-like and MGAS5005-like isolates. Shown are $\log_2$ values of the fold change in the mean transcript levels of core chromosomal genes that are significantly different ($P < .05$, Student's *t* test followed by a false-discovery-rate correction, to account for multiple comparisons), with at least a 2-fold change in expression between SF370-like and MGAS5005-like isolates. Red and blue data points relate to the presence or absence of these genes within the horizontally transferred 36-kb region, respectively.

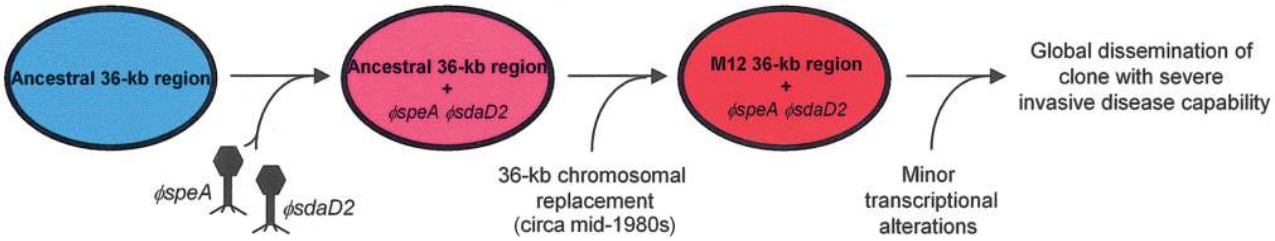

Figure 7. Reconstruction of the molecular evolutionary events that resulted in an abundant clone of group A *Streptococcus.* The hypothesis takes into account data presented here and in previous work [3, 11–13, 30, 41]. Key events include acquisition of prophages encoding SpeA and SdaD2 (an extracellular DNase virulence factor) and a horizontal gene transfer event involving a 36-kb chromosomal region encoding streptolysin O and NAD^+-glycohydrolase. On the basis of near identity in DNA sequence, we hypothesize that a serotype M12 strain served as the donor of this chromosomal region. All serotype M1 strains containing the *speA1* allele were recovered before 1988, suggesting that this allele was ancestral to the *speA2* allele characteristic of post-1988 M1 isolates. These 2 alleles differ by 1 nucleotide change [3].

6 strains all have a wild-type allele of *covR* and *covS*. Only 8 core chromosomal genes were found to be differentially transcribed (≥2-fold change in transcript level) between the SF370-like and MGAS5005-like isolates (figure 6). All 8 genes were more highly expressed in the MGAS5005-like isolates, with 5 of these 8 genes present in the 36-kb region of DNA involved in the horizontal gene transfer event (figure 6). Consistent with the findings from the extracellular protein analysis (figure 5), 2 of these genes were *slo* and *nga*.

DISCUSSION

Taken together, our genomic, SNP, transcriptomal, and protein analyses suggest that clonal replacement was a key factor associated with the recent increase in the frequency and severity of invasive infections caused by serotype M1 GAS. A genetically distinct serotype M1 clone, apparently more fit than other serotype M1 isolates, emerged during the mid-1980s and rapidly rose to dominance among disease isolates [11, 12, 30]. Moreover, our aggregate data suggest that a clear series of molecular changes account for the major evolutionary events that created the newer M1 clone, which is now highly abundant (figure 7). In this scenario, at least 3 molecular processes are implicated: acquisition of prophages encoding the putative or proven virulence factors SpeA and DNase SdaD2; reciprocal recombination of a chromosomal segment encoding 2 extracellular toxins; and accumulation of 1 or more mutations that increase the expression of a small number of other chromosomal genes. Evidence in support of this order of events is the knowledge that *speA2* and *sdaD2* have been identified in M1 GAS strains isolated during the 1970s (table 1 and authors' unpublished data) and that acquisition of the M12-like 36-kb region appears to have occurred during the mid-1980s.

The prophage-encoded genes *speA* and *sdaD2* are associated with the great majority of contemporary genotype M1 isolates (table 1 and figure 1) [11, 12, 41]. Although the exact contribution of SpeA to disease in mouse models of GAS infection is not clear, production of SpeA in humans, coupled with its well-studied superantigen activity, suggests that this protein participates in GAS pathogenesis [17, 42–45]. Recently, we discovered that SdaD2 increases virulence through enhanced evasion of the innate immune system, likely through degradation of neutrophil extracellular traps, structures that are composed of chromatin and granule proteins and that are released from polymorphonuclear leukocytes [46].

Although it is clear that lateral gene transfer was involved in the acquisition of the M12-like 36-kb region, our experiments here did not address the molecular mechanism responsible. The size of this element and the polylysogenic nature of GAS strains lead us to favor the view that the transfer was due to generalized transduction from a serotype M12 strain to a serotype M1 strain. This hypothesis is strongly supported by the observation that prophage ϕ5005.3, present in the genome of serotype M1 strain MGAS5005, is virtually identical to prophage ϕ9429.3, present in the genome of a serotype M12 strain that has been sequenced (authors' unpublished data).

The finding that only 8 core chromosomal genes were differentially expressed (≥2-fold) between genetically representative old and contemporary genotype M1 isolates was unexpected (figure 6). Five of the 8 differentially expressed genes are located within the 36-kb region that distinguishes pre- and post-1988 serotype M1 GAS isolates (figure 4), providing important circumstantial evidence supporting that this horizontal gene transfer event contributed to the emergence of the contemporary M1 clone. The increased *slo* and *nga* transcription in contemporary isolates correlates with the increased protein and activity levels observed in the supernatants of contemporary strains (figures 5 and 6).

In principle, the differences in the levels of gene transcripts between strains of the 2 serotype M1 phylogenetic lineages may be explained by polymorphisms that affect promoter activity. Hence, we compared the promoter-region sequences of the 8 differentially expressed genes in strains MGAS5005 and SF370.

The −10 and −35 promoter regions of *nusG* (which encodes the putative transcriptional antitermination protein NusG) and *nga* described by Kimoto et al. [47] in the homologous genes of group C *Streptococcus* were used. We found that both *nusG* and *nga* had 2 nucleotide changes within the spacer region separating the −10 and −35 promoter sequences. The 6 other differentially expressed genes had identical upstream putative regulatory sequences in strains SF370 and MGAS5005.

Nucleotide polymorphisms located within the spacer region can alter promoter activity [48, 49], providing support for the hypothesis that the increased number of *nusG* and *nga* transcripts made by strains of the MGAS5005-like M1 lineage is due to enhanced promoter activity. Inasmuch as *nga*, M5005_Spy0140, and *slo* are cotranscribed [47], transcription of all 3 genes would be increased in strains containing a stronger *nga* promoter, which is consistent with our expression microarray data (figure 6).

An inverse relationship between disease severity and SpeB expression by contemporary genotype M1 isolates has been postulated [50]. This raises the possibility that the increase in the severity of invasive infections since the mid-1980s may be due, in part, to down-regulation of SpeB production by the contemporary M1 clone. Our expression microarray data were generated with GAS cells harvested during the exponential phase, and it is possible that additional genes, such as *speB*, are differentially expressed during the stationary phase. However, in this regard we note that Western immunoblot analysis of SpeB in stationary-phase culture supernatants failed to identify a uniform difference in the level of immunoreactive SpeB between old and contemporary genotype M1 isolates (data not shown).

In conclusion, our findings have implications for understanding the molecular events that underlie the emergence of other bacterial clones that are highly virulent or have other characteristic disease phenotypes. Our results stress the importance of deploying integrated genomewide analyses in such endeavors.

Acknowledgments

We thank M. Gutacker, for help with single-nucleotide polymorphism analysis; K. Pflughoeft, for technical assistance; S. Shelburne, for critical comments; and J. Richard, for editorial assistance.

References

1. Ochman H, Lawrence JG, Groisman EA. Lateral gene transfer and the nature of bacterial innovation. Nature **2000**; 405:299–304.
2. Cunningham MW. Pathogenesis of group A streptococcal infections. Clin Microbiol Rev **2000**; 13:470–511.
3. Musser JM, Krause RM. The revival of group A streptococcal diseases, with a commentary on staphylococcal toxic shock syndrome. In: Krause RM, ed. Emerging infections. New York: Academic Press, **1998**:185–218.
4. Kotb M, Norrby-Teglund A, McGeer A, et al. An immunogenetic and molecular basis for differences in outcomes of invasive group A streptococcal infections. Nat Med **2002**; 8:1398–404.
5. Eriksson BK, Andersson J, Holm SE, Norgren M. Epidemiological and clinical aspects of invasive group A streptococcal infections and the streptococcal toxic shock syndrome. Clin Infect Dis **1998**; 27:1428–36.
6. Tyrrell GJ, Lovgren M, Forwick B, Hoe NP, Musser JM, Talbot JA. M types of group A streptococcal isolates submitted to the National Centre for *Streptococcus* (Canada) from 1993 to 1999. J Clin Microbiol **2002**; 40: 4466–71.
7. Zurawski CA, Bardsley M, Beall B, et al. Invasive group A streptococcal disease in metropolitan Atlanta: a population-based assessment. Clin Infect Dis **1998**; 27:150–7.
8. Banks DJ, Porcella SF, Barbian KD, et al. Progress toward characterization of the group A *Streptococcus* metagenome: complete genome sequence of a macrolide-resistant serotype M6 strain. J Infect Dis **2004**; 190:727–38.
9. Green NM, Zhang S, Porcella SF, et al. Genome sequence of a serotype M28 strain of group A *Streptococcus:* potential new insights into puerperal sepsis and bacterial disease specificity. J Infect Dis **2005**; 192: 000–00 (in this issue).
10. Overbeek R, Larsen N, Walunas T, et al. The ERGO genome analysis and discovery system. Nucleic Acids Res **2003**; 31:164–71.
11. Musser JM, Hauser AR, Kim MH, Schlievert PM, Nelson K, Selander RK. *Streptococcus pyogenes* causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. Proc Natl Acad Sci USA **1991**; 88:2668–72.
12. Musser JM, Kapur V, Szeto J, Pan X, Swanson DS, Martin DR. Genetic diversity and relationships among *Streptococcus pyogenes* strains expressing serotype M1 protein: recent intercontinental spread of a subclone causing episodes of invasive disease. Infect Immun **1995**; 63:994–1003.
13. Hoe NP, Nakashima K, Lukomski S, et al. Rapid selection of complement-inhibiting protein variants in group A *Streptococcus* epidemic waves. Nat Med **1999**; 5:924–9.
14. Hoe NP, Nakashima K, Grigsby D, et al. Rapid molecular genetic subtyping of serotype M1 group A *Streptococcus* strains. Emerg Infect Dis **1999**; 5:254–63.
15. Hoe NP, Vuopio-Varkila J, Vaara M, et al. Distribution of streptococcal inhibitor of complement variants in pharyngitis and invasive isolates in an epidemic of serotype M1 group A *Streptococcus* infection. J Infect Dis **2001**; 183:633–9.
16. 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 **1997**; 175:392–9.
17. Virtaneva K, Graham MR, Porcella SF, et al. Group A *Streptococcus* gene expression in humans and cynomolgus macaques with acute pharyngitis. Infect Immun **2003**; 71:2199–207.
18. Voyich JM, Sturdevant DE, Braughton KR, et al. Genome-wide protective response used by group A *Streptococcus* to evade destruction by human polymorphonuclear leukocytes. Proc Natl Acad Sci USA **2003**; 100:1996–2001.
19. Sumby P, Barbian KD, Gardner DJ, et al. Extracellular deoxyribonuclease made by group A *Streptococcus* assists pathogenesis by enhancing evasion of the innate immune response. Proc Natl Acad Sci USA **2005**; 102:1679–84.
20. Ferretti JJ, McShan WM, Ajdic D, et al. Complete genome sequence of an M1 strain of *Streptococcus pyogenes*. Proc Natl Acad Sci USA **2001**; 98:4658–63.
21. Smoot JC, Barbian KD, Van Gompel JJ, et al. Genome sequence and comparative microarray analysis of serotype M18 group A *Streptococcus* strains associated with acute rheumatic fever outbreaks. Proc Natl Acad Sci USA **2002**; 99:4668–73.
22. Beres SB, Sylva GL, Barbian KD, et al. Genome sequence of a serotype M3 strain of group A *Streptococcus:* phage-encoded toxins, the high-virulence phenotype, and clone emergence. Proc Natl Acad Sci USA **2002**; 99:10078–83.
23. 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**; 182:1117–28.

24. Fitzgerald JR, Sturdevant DE, Mackie SM, Gill SR, Musser JM. Evolutionary genomics of *Staphylococcus aureus:* insights into the origin of methicillin-resistant strains and the toxic shock syndrome epidemic. Proc Natl Acad Sci USA **2001**; 98:8821–6.
25. 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**; 44:257–69.
26. Limbago B, Penumalli V, Weinrick B, Scott JR. Role of streptolysin O in a mouse model of invasive group A streptococcal disease. Infect Immun **2000**; 68:6384–90.
27. Graham MR, Virtaneva K, Porcella SF, et al. Group A *Streptococcus* transcriptome dynamics during growth in human blood reveals bacterial adaptive and survival strategies. Am J Pathol **2005**; 166:455–65.
28. Graham MR, Smoot LM, Migliaccio CA, et al. Virulence control in group A *Streptococcus* by a two-component gene regulatory system: global expression profiling and in vivo infection modeling. Proc Natl Acad Sci USA **2002**; 99:13855–60.
29. Combet C, Blanchet C, Geourjon C, Deléage G. NPS@: Network Protein Sequence Analysis. Trends Biochem Sci **2000**; 25:147–150.
30. Cleary PP, Kaplan EL, Handley JP, et al. Clonal basis for resurgence of serious *Streptococcus pyogenes* disease in the 1980s. Lancet **1992**; 339: 518–21.
31. Lukomski S, Nakashima K, Abdi I, et al. Identification and characterization of the *scl* gene encoding a group A *Streptococcus* extracellular protein virulence factor with similarity to human collagen. Infect Immun **2000**; 68:6542–53.
32. Weeks CR, Ferretti JJ. The gene for type A streptococcal exotoxin (erythrogenic toxin) is located in bacteriophage T12. Infect Immun **1984**; 46:531–6.
33. Aziz RK, Ismail SA, Park HW, Kotb M. Post-proteomic identification of a novel phage-encoded streptodornase, Sda1, in invasive M1T1 *Streptococcus pyogenes*. Mol Microbiol **2004**; 54:184–97.
34. 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**; 106:27–37.
35. Lazarides PD, Bernheimer AW. Association of production of diphosphopyridine nucleotidase with serological type of group A *Streptococcus*. J Bacteriol **1957**; 74:412–3.
36. Lutticken R, Lutticken D, Johnson DR, Wannamaker LW. Application of a new method for detecting streptococcal nicotinamide adenine dinucleotide glycohydrolase to various M types of *Streptococcus pyogenes*. J Clin Microbiol **1976**; 3:533–6.
37. Ajdic D, McShan WM, Savic DJ, Gerlach D, Ferretti JJ. The NAD-glycohydrolase *(nga)* gene of *Streptococcus pyogenes*. FEMS Microbiol Lett **2000**; 191:235–41.
38. Karasawa T, Yamakawa K, Tanaka D, Gyobu Y, Nakamura S. NAD^+-glycohydrolase productivity of haemolytic streptococci assayed by a simple fluorescent method and its relation to T serotype. FEMS Microbiol Lett **1995**; 128:289–92.
39. Bernhard GC, Stollerman GH. Serum inhibition of streptococcal diphosphopyridine nucleotidase in uncomplicated streptococcal pharyngitis and in rheumatic fever. J Clin Invest **1959**; 38:1942–9.
40. Kellner A, Freeman EB, Carlson AS. Neutralizing antibodies to streptococcal diphosphopyridine nucleotidase in the serum of experimental animals and human beings. J Exp Med **1958**; 108:299–309.
41. Cleary PP, LaPenta D, Vessela R, Lam H, Cue D. A globally disseminated M1 subclone of group A streptococci differs from other subclones by 70 kilobases of prophage DNA and capacity for high-frequency intracellular invasion. Infect Immun **1998**; 66:5592–7.
42. Sriskandan S, Unnikrishnan M, Krausz T, Cohen J. Molecular analysis of the role of streptococcal pyrogenic exotoxin A (SPEA) in invasive soft-tissue infection resulting from *Streptococcus pyogenes*. Mol Microbiol **1999**; 33:778–90.
43. Unnikrishnan M, Cohen J, Sriskandan S. Complementation of a *speA* negative *Streptococcus pyogenes* with *speA:* effects on virulence and production of streptococcal pyrogenic exotoxin A. Microb Pathog **2001**; 31:109–14.
44. Sriskandan S, Unnikrishnan M, Krausz T, et al. Enhanced susceptibility to superantigen-associated streptococcal sepsis in human leukocyte antigen–DQ transgenic mice. J Infect Dis **2001**; 184:166–73.
45. Llewelyn M, Sriskandan S, Peakman M, et al. HLA class II polymorphisms determine responses to bacterial superantigens. J Immunol **2004**; 172:1719–26.
46. Brinkmann V, Reichard U, Goosmann C, et al. Neutrophil extracellular traps kill bacteria. Science **2004**; 303:1532–5.
47. Kimoto H, Fujii Y, Yokota Y, Taketo A. Molecular characterization of NADase-streptolysin O operon of hemolytic streptococci. Biochim Biophys Acta **2005**; 1681:134–49.
48. Liu M, Tolstorukov M, Zhurkin V, Garges S, Adhya S. A mutant spacer sequence between −35 and −10 elements makes the P_{lac} promoter hyperactive and cAMP receptor protein-independent. Proc Natl Acad Sci USA **2004**; 101:6911–6.
49. Borst DW, Betley MJ. Mutations in the promoter spacer region and early transcribed region increase expression of staphylococcal enterotoxin A. Infect Immun **1993**; 61:5421–5.
50. Kansal RG, McGeer A, Low DE, Norrby-Teglund A, Kotb M. Inverse relationship between disease severity and expression of the streptococcal cysteine protease, SpeB, among clonal M1T1 isolates recovered from invasive group A streptococcal infection cases. Infect Immun **2000**; 68:6362–9.