INFECTION AND IMMUNITY, Mar. 1995, p. 994–1003 Vol. 63, No. 3
0019-9567/95/$04.00+0

Genetic Diversity and Relationships among *Streptococcus pyogenes* Strains Expressing Serotype M1 Protein: Recent Intercontinental Spread of a Subclone Causing Episodes of Invasive Disease

JAMES M. MUSSER,[1]* VIVEK KAPUR,[1] JENNY SZETO,[2] XI PAN,[1]
DOUGLAS S. SWANSON,[1] AND DIANA R. MARTIN[2]

Section of Molecular Pathobiology, Department of Pathology, Baylor College of Medicine, Houston, Texas 77030,[1] and Communicable Disease Center, ESR: Health, Porirua, New Zealand[2]

Received 28 October 1994/Returned for modification 14 December 1994/Accepted 28 December 1994

Chromosomal diversity and relationships among 126 *Streptococcus pyogenes* strains expressing M1 protein from 13 countries on five continents were analyzed by multilocus enzyme electrophoresis and restriction fragment profiling by pulsed-field gel electrophoresis. All isolates were studied for the presence of the gene encoding streptococcal pyrogenic exotoxin A by PCR. Strain subsets were also examined by automated DNA sequencing for allelic polymorphism in genes encoding M protein (*emm*), streptococcal pyrogenic exotoxin A (*speA*), streptokinase (*ska*), pyrogenic exotoxin B (interleukin-1β convertase) (*speB*), and C5a peptidase (*scp*). Seven distinct *emm1* alleles that encode M proteins differing at one or more amino acids in the N-terminal variable region were identified. Although substantial levels of genetic diversity exist among M1-expressing organisms, most invasive disease episodes are caused by two subclones marked by distinctive multilocus enzyme electrophoretic profiles and pulsed-field gel electrophoresis restriction fragment length polymorphism (RFLP) types. One of these subclones (ET 1/RFLP pattern 1a) has the *speA* gene and was recovered worldwide. Identity of *speA*, *emm1*, *speB*, and *ska* alleles in virtually all isolates of ET 1/RFLP type 1a means that these organisms share a common ancestor and that global dispersion of this M1-expressing subclone has occurred very recently. The occurrence of the same *emm* and *ska* alleles in strains that are well differentiated in overall chromosomal character demonstrates that horizontal transfer and recombination play a fundamental role in diversifying natural populations of *S. pyogenes*.

Streptococcus pyogenes (group A streptococcus) causes a variety of important human diseases, including pharyngitis, pneumonia, acute rheumatic fever, poststreptococcal glomerulonephritis, cellulitis, sepsis, and meningitis (4, 53). Although several virulence factors have been identified, the primary focus of pathogenesis research for more than seven decades has been the M protein, an α-helical-coiled coil protein that projects outward from the bacterial cell wall (12, 25, 27). This molecule is considered to be a major *S. pyogenes* virulence determinant on the basis of the observation that some M proteins have been shown to be antiphagocytic and elicit type-specific immunity.

Early in the study of M protein, it was recognized that the molecule was polymorphic in natural populations of group A streptococci. Serological studies have identified more than 100 distinct M protein serotypes among isolates recovered from diseased humans and healthy carriers (12, 25, 27). The occurrence of extensive serological diversity forms the basis of a scheme widely used for classifying isolates, and until recently, essentially all epidemiologic and most virulence-related research has been conducted, and the results interpreted, in the context of the M protein serotype of strains.

In the past decade the molecular basis of M protein variation has been examined in detail (reviewed in references 12, 25, and 46). It is now known that there is a conserved leader sequence at the amino terminus which is followed by a short nonrepeat domain that is hypervariable and is the region against which type-specific antisera is directed. A central α-helical region consists of sets of repeats that are polymorphic in size and number among serotypes. The repeats are followed by a proline-glycine-rich domain that is cell wall associated and a membrane anchor region at the carboxy terminus. These latter two areas are well conserved at the nucleotide and amino acid levels.

Although there is a commonly held assumption that strains classified on the basis of the M protein serotype are genetically uniform or share close overall chromosomal affinity, the results of several recent studies have provided evidence that strains expressing the same M protein serotype are not genetically homogeneous. For example, Harbaugh et al. (14) analyzed a small group of serotype M1 strains and found nucleotide diversity (substitutions and a small insert) in a gene designated *emm1.1* that was associated with altered antibody recognition of opsonic epitopes. Examination (58) of seven serotype M5 strains identified one organism that had two distinct *emm*-like genes located in the Vir regulon (6), a chromosomal region with tandemly arranged genes coding for putative *S. pyogenes* virulence factors. Six other strains had only a single *emm* gene in this region (58). Two studies employing multilocus enzyme electrophoresis identified organisms with different allele profiles that shared the same M protein serologic reactivity pattern (23, 35), and strains of certain M types were shown to be heterogenous by restriction fragment length polymorphism (RFLP) analysis (51).

Inasmuch as there is accumulating evidence that *S. pyogenes* strains classified on the basis of the M protein serotype are genetically heterogeneous and considerable research has been devoted to formulating an M protein-based vaccine (1, 2, 13), it is important to investigate in detail the level of genomic

* Corresponding author. Mailing address: Department of Pathology, Baylor College of Medicine, One Baylor Plaza, Houston, TX 77030. Phone: (713) 798-3823. Fax: (713) 798-5838. Electronic mail address: jmusser@darwin.bcm.tmc.edu.

diversity present in organisms expressing the same M type recovered from different localities. Moreover, extensive human travel on a global scale provides an effective mechanism for dispersing strains with unusual virulence properties. Genetic diversity among isolates of an M type was studied in a sample of M1 strains for three reasons. First, M1 organisms are important causes of human invasive disease episodes. Second, in recent years, a striking increase in the frequency and severity of *S. pyogenes* infections has occurred on an intercontinental basis, and M1 is one of the two serotypes associated with this temporal disease variation (15, 36, 37, 45, 53, 54). M1 strains have been an especially important cause of contemporary invasive disease episodes in Europe (26). Third, because M1 strains are responsible for many severe infections, samples of organisms from widespread geographic sources are available for analysis.

To address the question of genetic diversity among M1 strains, multilocus enzyme electrophoresis (36, 48) was employed to estimate overall chromosomal relatedness among isolates, and each isolate was also characterized for its pulsed-field gel electrophoretic RFLP pattern (32, 51) and the presence of the gene (*speA*) encoding streptococcal pyrogenic exotoxin A (SPE A). Selected samples of isolates were examined for sequence variation in parts of structural genes for M protein, SPE A, pyrogenic exotoxin B (interleukin-1β convertase) (21), streptokinase (17, 18), and C5a peptidase (5).

MATERIALS AND METHODS

Bacterial strains. A sample of 126 strains of *S. pyogenes* (Table 1) cultured from patients in 13 countries on five continents was studied. The organisms were recovered from diseased individuals with a variety of streptococcal infections, including toxic-shock-like syndrome, other severe invasive disease episodes, scarlet fever, poststreptococcal glomerulonephritis, and acute rheumatic fever; many of the isolates were taken from individuals with pharyngitis. All isolates were recovered from individuals with no known direct epidemiologic association. Some of the organisms have been studied previously on a limited basis (23, 32, 35–37).

Electrophoresis of enzymes. Methods of protein extract preparation, electrophoresis of soluble enzymes, and selective enzyme staining have been described previously (48). The 12 enzymes assayed were α-naphthyl propionate esterase, leucylalanine peptidase, leucylglycylglycine peptidase, phosphoglucose isomerase, nucleoside phosphorylase, glyceraldehyde-3-phosphate dehydrogenase (NAD dependent), glyceraldehyde-3-phosphate dehydrogenase (NADP dependent), lactate dehydrogenase, leucine aminopeptidase, hydroxybutyrate dehydrogenase, β-glucuronidase, and adenylate kinase. Distinctive electromorphs (mobility variants) of each enzyme, numbered in order of decreasing rate of anodal migration, were equated with alleles at the corresponding structural gene locus. Isolates that lacked activity for a specific enzyme were assigned a null allelic state at the locus in question.

Each isolate was characterized by its combination of alleles at the 12 enzyme loci, and distinctive multilocus enzyme genotypes were designated as electrophoretic types (ETs). The ET designations are cognate with those used elsewhere (36). Bacteria previously characterized for RFLP patterns (32) were coded and analyzed for ETs in a blinded fashion.

Strain serotyping. All isolates expressing M and T proteins were serotyped by internationally standardized methods (31) in the laboratory of D. Martin, with antisera prepared against reference type strains at the New Zealand Communicable Disease Centre.

RFLP profiling. RFLP patterns were generated by pulsed-field gel electrophoresis of chromosomal DNA digested with the infrequently cutting restriction enzyme *Sfi*I by techniques described recently (32, 51). Briefly, DNA was obtained from strains grown overnight in 5 ml of Todd-Hewitt broth plus neopeptone. Five milliliters of EC lysis buffer (containing Tris, EDTA, Brij-58, deoxycholate, sarkosyl, lysozyme, and RNase), and ESP buffer (EDTA, sodium lauroylsarcosine, pronase) were used. The agarose plugs were washed in pulsed-field gel Tris-borate-EDTA buffer, and the gel electrophoresis conditions used for ramping were 30 to 90 s at 200 V for 23 h. This technique yields approximately five to seven well-separated DNA fragments in profiles that are less ambiguous than those produced with multicutting enzymes. Isolates were analyzed in a blinded fashion, without knowledge of the multilocus enzyme genotype.

Sequencing of genes encoding M protein (*emm*), SPE A (*speA*), interleukin-1β convertase (*speB*), streptokinase (*ska*), and C5a peptidase (*scp*). Strains selected to represent organisms of distinct ETs, RFLP types, and the breadth of geographic and clinical syndrome diversity were studied for sequence variation in five genes encoding putative virulence factors. The oligonucleotide primers and PCR conditions used to amplify target segments of the genes studied are listed in Table 2. For each reaction, PCR amplification of 0.5 μl of chromosomal DNA was performed in 30 μl of a mixture containing 50 mM KCl; 10 mM Tris-HCl (pH 8.3); 1.5 mM $MgCl_2$, 200 μM (each) dATP, dCTP, dGTP, and dTTP; 500 nM (each) primer; and 0.83 U of Ampli*Taq* DNA polymerase (Perkin-Elmer Cetus, Norwalk, Conn.). For each target gene segment, the thermocycler parameters used were 30 cycles of annealing at 50°C for 2.5 min and extension at 72°C for 2.5 min for *emm*, *scpA*, and *speA* and 30 cycles of annealing at 55°C for 2 min and extension at 72°C for 2.5 min for *speB* and *ska*. Each reaction was preceded by a single denaturation step at 94°C for 4 min and terminated by a single primer extension step at 72°C for 15 min. Each cycle had an initial denaturation step at 94°C for 1 min.

The unincorporated nucleotides and primers were separated from amplified DNA by filtration through Microcon 100 microconcentrators (Amicon, Inc., Beverly, Mass.). Sequencing reactions with the *Taq* DyeDeoxy terminator cycle sequencing kit (Applied Biosystems Inc., Foster City, Calif.) were performed with 7 μl of PCR-amplified DNA as the template and 3.2 pmol of either the forward or reverse primer. The unincorporated dye terminators and primers were separated from the extension products by spin column purification (Centri-Sep; Princeton Separations, Inc., Adelphia, N.J.). The products were dried in a vacuum centrifuge, resuspended in 4 μl of loading buffer (5:1 ratio of deionized formamide–50 mM EDTA [pH 8.0]), heat denatured for 2 min at 90°C, and immediately loaded on an acrylamide gel in an Applied Biosystems model 373A automated DNA sequencer. Both strands were sequenced. The data were assembled and edited with EDITSEQ, ALIGN, and SEQMAN programs (DNASTAR, Madison, Wis.), and the sequences were compared with published data.

The *speA* allele of each of 42 isolates was determined by automated DNA sequencing of PCR-amplified products. The entire *speA* gene was sequenced in approximately one-half of the sample; for the other organisms, a fragment of *speA* with a size of >300 bp that included a variable region differentiating *speA1*, *speA2*, *speA3*, and *speA4* (40) was characterized.

Variation in the *emm* gene was studied in 51 strains. Two regions of *emm* were sequenced, including nucleotides (nt) 79 to 378 (the numbering system used was described in reference 14) coding for the leader sequence and the N-terminal variable domain and nt 856 to 1155 specifying two of the C-repeat regions (C2 and C3) and part of the D-repeat region (12, 25, 46).

An ~350-bp region of *speB* (23) in seven strains representing four distinct ETs was sequenced. The area studied corresponds to a variable region of the mature cysteine protease that contains a linear B-cell epitope, as defined by reactivity of rabbit polyclonal sera and murine monoclonal antibodies (19). For *ska* (GenBank no. XO3399), an ~450-bp region corresponding to an unusually variable part of the gene (17, 18, 38, 42) was sequenced. The region of *scp* (GenBank no. JO5229) sequenced was a 405-bp fragment located between nt 2198 and 2604.

Nucleotide sequence accession number. The deduced N-terminal and C-terminal sequences of *S. pyogenes* serotype M1 *emm* alleles were deposited in GenBank under accession no. U20094 to U20104.

RESULTS

Genetic and genotypic diversity and estimates of phylogenetic relationships among multilocus enzyme genotypes. Of 126 isolates of M1 *S. pyogenes* analyzed, five structural genes assayed by multilocus enzyme electrophoresis were polymorphic and seven were monomorphic. Among the polymorphic loci, the number of alleles ranged from two to four.

Comparison of allele profiles over all loci identified six ETs (arbitrarily designated ET 1 and ET 1B to ET 1F), of which two (ET 1D and ET 1E) were represented by single isolates and four (ET 1, ET 1B, ET 1C, and ET 1F) were represented by multiple isolates (range, 2 to 115) (Table 3). For purposes of illustrating that strains expressing M1 protein do not represent a distinct subline of the species *S. pyogenes*, the multilocus enzyme genotypes of seven non-M1 strains representing the breadth of species diversity were combined with the M1 sample, and overall genetic relatedness was estimated by the average linkage method (48). As shown in Fig. 1, the cell lines expressing M1 protein represent a heterogeneous array of chromosomal lineages of *S. pyogenes*, rather than a single phylogenetic branch of the species. In overall chromosomal relatedness, several M1 strains were allied more closely with organisms expressing other serologic variants of M protein than with the other M1 strains.

RFLP patterns. The analysis identified 16 distinct RFLP

TABLE 1. Properties of 126 isolates of *S. pyogenes* expressing M1 protein

ET	RFLP type[a]	MGAS no.	New Zealand no.	Locality or strain	Disease or source of isolate[b]	Genotype[c]	
						speA allele	Other allele(s)
1	1a	166	950026	Minnesota	TSLS	*speA2*	*emm1.0 speB2 scp1*
	1a	247	930785	California	SID	*speA* pos	*emm1.0*
	1a	250	930787	California	SID	*speA2*	
	1a	251	930788	California	SID	*speA2*	
	1a	256	930791	California	SID	*speA2*	
	1a	257	930792	California	SID	*speA* pos	
	1a	285	930796	Colorado	SID	*speA2*	*speB2*
	1a	287	930797	Colorado	SID	*speA2*	
	1a	290	930798	Colorado	SID	*speA* pos	
	1a	294	930799	Washington	Invasive	*speA* pos	
	1a	297	930800	Washington	TSLS	*speA2*	
	1a	298	930801	Unknown	Invasive	*speA2*	
	1a	301	930802	Utah	TSLS	*speA* pos	*emm1.0*
	1a	304	930803	Unknown	Invasive	*speA* pos	
	1a	307	930804	Texas	TSLS	*speA* pos	*emm1.0*
	1a	309	930805	Oregon	TSLS	*speA* pos	
	1a	310	930806	Unknown	Invasive	*speA2*	
	1a	312	930807	New York	TSLS	*speA2*	*emm1.0*
	1a	313	930808	Wyoming	TSLS	*speA* pos	
	1a	314	930809	Texas	Invasive	*speA2*	*emm1.0*
	1a	316	930810	Unknown	TSLS	*speA* pos	
	1a	336	930812	California	Invasive	*speA2*	
	1a	337	930813	New York	Invasive	*speA* pos	
	1a	340	930814	Oregon	TSLS	*speA2*	
	1a	532	930815	Sweden	TSLS	*speA2*	
	1a	533	930816	Sweden	TSLS	*speA2*	
	1a	534	930817	Sweden	TSLS	*speA2*	
	1a	535	930818	Sweden	TSLS	*speA2*	*emm1.0*
	1a	536	930819	Sweden	TSLS	*speA* pos	
	1a	575	930823	Canada	SF	*speA* neg	*emm1.3*
	1a	1272	930827	Denmark	Sepsis	*speA2*	*emm1.0*
	1a	1283	930828	Denmark	Sepsis	*speA2*	
	1a	1301	930829	Germany	SF	*speA* pos	*emm1.0*
	1a	1333	930830	Germany	SF	*speA* pos	
	1a	1400	930832	Germany	SF	*speA* neg	
	1a	1430	930833	Sweden	SF	*speA2*	
	1a	1508	940356	Czechoslovakia	SF	*speA* pos	
	1a	1593	940357	United States	Pharyngitis	*speA* pos	
	1a	1650	930839	United States	ARF	*speA2*	*emm1.0*
	1a	1666	930841	United Kingdom	Sepsis	*speA2*	*emm1.5*
	1a	1667	930842	United Kingdom	Sepsis	*speA2*	*emm1.0*
	1a	1668	930843	United Kingdom	Sepsis	*speA2*	*emm1.0*
	1a	1669	930844	United Kingdom	Sepsis	*speA2*	
	1a	1670	930845	United Kingdom	Sepsis	*speA* pos	
	1a	1700	940363	United States	Pharyngitis	*speA* pos	
	1a	1735	940364	United States	Pharnygitis	*speA* pos	
	1a	1922	930847	Netherlands	Sepsis	*speA* pos	
	1a	1930	930848	Netherlands	Sepsis	*speA* pos	
	1a	1960	940367	United States	Pharyngitis	*speA* pos	*emm1.0*
	1a	2077	841657	New Zealand	ARF	*speA2*	
	1a	2079	850921	New Zealand	ARF	*speA2*	*emm1.0*
	1a	2080	890385	New Zealand	Sepsis	*speA2*	*emm1.0*
	1a	2081	890637	New Zealand	Cellulitis	*speA2*	*emm1.0*
	1a	2082	900012	New Zealand	Sepsis	*speA2*	
	1a	2092	840221	New Zealand	Throat	*speA2*	
	1a	2093	851532	New Zealand	ARF	*speA2*	
	1a	2094	880711	New Zealand	Sepsis	*speA2*	
	1a	2115	850852	New Zealand	ARF	*speA* pos	
	1a	2121	920197	New Zealand	Meningitis	*speA* pos	
	1a	2122	920297	New Zealand	Tonsillitis	*speA* pos	
	1a	2124	920478	New Zealand	TSLS	*speA* pos	
	1a	2126	920727	India	ARF	*speA2*	*emm1.0*
	1a	2128	920739	New Zealand	Cellulitis	*speA* pos	
	1a	2129	920743	New Zealand	Sepsis	*speA* pos	
	1a	2131	930522	New Zealand	Endometritis	*speA* pos	
	1a	2132	930534	New Zealand	ARDS	*speA* pos	
	1a	2133	930653	New Zealand	Sepsis	*speA* pos	

Continued on following page

TABLE 1—*Continued*

ET	RFLP type[a]	MGAS no.	New Zealand no.	Locality or strain	Disease or source of isolate[b]	Genotype[c]	
						speA allele	Other allele(s)
	1a	2134	930707	New Zealand	ARF	*speA* pos	
	1a	2136	930726	Ethiopia	Tonsillitis	*speA* pos	
	1a	2137	930727	Ethiopia	Tonsillitis	*speA* pos	
	1a	2138	930728	Ethiopia	Tonsillitis	*speA* pos	
	1a	2139	930730	Ethiopia	Tonsillitis	*speA* pos	*emm1.0*
	1a	2141	930751	New Zealand	SF	*speA* pos	
	1a	2142	930906	Ethiopia	Tonsillitis	*speA* pos	
	1a	2203	940368	United States	Pharyngitis	*speA* pos	
	1a	2219	820905	Australia	SF	*speA* pos	*emm1.0*
	1a	2221	880388	Australia	SF	*speA* pos	*emm1.0*
	1a	2224	880709	New Zealand	AGN	*speA* pos	
	1a	2225	890590	New Zealand	Meningitis	*speA* pos	
	1a	2286	930369	United States	Pharyngitis	*speA* pos	*emm1.0*
	1a	2345	940370	United States	Pharyngitis	*speA* pos	
	1a	2350	940371	United States	Pharyngitis	*speA* pos	*emm1.0*
	1k	249	930786	California	SID	*speA* neg	*emm1.3*
	1k	252	930789	California	SID	*speA* neg	
	1k	253	930790	California	SID	*speA* neg	
	1k	267	930793	Minnesota	SID	*speA* neg	
	1k	279	930794	Colorado	SID	*speA* neg	*emm1.3*
	1k	284	930795	Colorado	SID	*speA* neg	
	1k	326	930811	Pennsylvania	TSLS	*speA* neg	*emm1.3 speB2 scp1*
	1k	570	930820	Canada	SF	*speA* neg	*emm1.3*
	1k	571	930821	Canada	SF	*speA* neg	
	1k	572	930822	Canada	SF	*speA* neg	
	1k	1264	930825	Denmark	Sepsis	*speA* neg	*emm1.0*
	1k	1377	930831	Germany	SF	*speA* neg	
	1k	1503	930834	Germany	SF	*speA* pos	*emm1.0*
	1k	1642	940360	United States	ARF	*speA* neg	*emm1.3*
	1k	1632	940358	United States	ARF	*speA* neg	*emm1.3*
	1k	1633	940359	United States	ARF	*speA* neg	*emm1.3*
	1k	1653	940361	United States	ARF	*speA* neg	*emm1.3*
	1c	2116	86J28a	New Zealand	Tonsillitis	*speA* pos	*emm1.0*
	1c	2119	900332	United States	Tonsillitis	*speA* pos	
	1c	2218	850067	New Zealand	Tonsillitis	*speA* pos	
	1e	2223	791451	New Zealand	LRTI	*speA* neg	
	1e	2108	791575	New Zealand		*speA* neg	*emm1.6*
	1e	1265	930826	Denmark	Sepsis	*speA* neg	*emm1.0*
	1l	1253	930824	119/6		*speA1*	*emm1.0 speB2 scp1*
	1l	1600	930836	SF130		*speA1*	
	1l	1829	930846	AP1		*speA1*	
	1l	2217	2031R	Prague 40/58		*speA1*	*emm1.0*
	1m	1626	930837	United States	ARF	*speA2*	*emm1.0*
	1m	1627	930838	United States	ARF	*speA2*	*emm1.0*
	1p	2130	930059	New Zealand	Cellulitis	*speA* pos	*emm1.0*
	1q	1781	940365	United States	Pharyngitis	*speA* neg	
	1q	1796	940366	United States	Pharyngitis	*speA* neg	
	1s	2143	930907	Ethiopia	Tonsillitis	*speA2*	
1B	1f	2110	820053	New Zealand	Eczema	*speA* neg	*emm1.4 speB101 scp2 ska2*
	1d	2112	830096	Papua New Guinea	Wound	*speA* neg	
1C	1o	2120	920057	Australia	Impetigo	*speA* neg	*emm1.4 speB102 scp1 ska1*
	1o	2125	910616	Australia	Impetigo	*speA* neg	*emm1.4*
1D	1a	2127	920728	India	ARF	*speA2*	*emm1.0*

Continued on following page

TABLE 1—*Continued*

ET	RFLP type[a]	MGAS no.	New Zealand no.	Locality or strain	Disease or source of isolate[b]	Genotype[c]	
						speA allele	Other allele(s)
1E	1k	2140	930731	Ethiopia	Throat	*speA* neg	
1F	1b	2109	791632	New Zealand	Tonsillitis	*speA* neg	*emm1.1* + 21 bp *speB100 scp1 ska1*
	1j	2123	910367	Australia	Scabies	*speA1*	*emm1.1*
	1b	2144	115031	New Zealand	Tonsillitis	*speA* neg	*emm1.1* + 21 bp
	1h	2216	870376	New Zealand	Tonsillitis	*speA* neg	*emm1.1* + 21 bp
	1b[d]	2226	820157	New Zealand	Tonsillitis	*speA* neg	

[a] The 1 designation for RFLP type refers to M protein serotype, as described in references 32 and 51.
[b] TSLS, toxic-shock-like syndrome; SID, severe invasive disease; invasive, invasive infection (from normally sterile body site); SF, scarlet fever; ARF, acute rheumatic fever; ARDS, adult respiratory distress syndrome; AGN, acute glomerulonephritis; LRTI, lower respiratory tract infection.
[c] *speA* pos, *speA* positive by PCR but not characterized to the allele level by DNA sequencing; *speA1*, allele *speA1*; *speA2*, allele *speA2*; *speA* neg, *speA* negative by PCR; + 21 bp, included a 21-bp in-frame insertion.
[d] Similar to but slightly different from RFLP pattern b.

patterns that differed from one another by at least one band (Fig. 2). Most isolates (83 [66%]) were assigned to a single RFLP type, designated 1a, and 17 isolates (13%) were RFLP type 1k. The remaining 14 RFLP types had four or fewer isolates (Table 1).

Multilocus enzyme types and RFLP patterns. There were 18 different combinations of ET and RFLP patterns identified in the sample. Among the 115 isolates of ET 1, nine RFLP patterns were found, although most (100 [87%]) were either RFLP type 1a or 1k. For the five non-ET-1 lineages, ET 1F was the most heterogeneous in RFLP pattern, with four types identified among the five strains.

***speA* alleles and subclone distribution.** All 126 isolates were tested by either PCR or Southern hybridization for presence of the gene (*speA*) encoding SPE A. In several studies (36, 37, 54) this gene has been statistically associated with strains recovered from patients with streptococcal toxic-shock-like syndrome. The *speA* gene was found in 94 of the isolates. Virtually all (81 of 83) strains of ET 1/RFLP 1a had this gene, but in contrast, 16 of 17 ET 1/RFLP 1k organisms lacked *speA*. The gene was also identified in strains assigned to six other ET-RFLP classes, but only two of the non-ET-1 strains had *speA* (Table 1).

Four alleles of *speA*, designated *speA1* to *speA4*, have been identified on the basis of nucleotide sequence characterization of strains from diverse localities and time periods (40). In order to determine the distribution of *speA* alleles among the six M1 ETs and the 17 ET/RFLP types, 40 strains of ET 1 and the two non-ET 1 *speA*-positive organisms were characterized by automated DNA sequencing. All isolates of ET 1/RFLP 1a, including strains from throughout the United States and eight other countries, had the *speA2* allele. In addition, with the exception of strains of ET 1/RFLP 1l, which were unusual in having the *speA1* allele, all other ET 1 organisms had the *speA2* variant. (On the basis of previously described organism pedigrees [37], it is likely that Musser group A streptococcus no. [MGAS] 1253 is a derivative of MGAS 1600.) The two non-ET-1 strains (MGAS 2123 and MGAS 2127) had *speA1* and *speA2*, respectively (Table 1).

***emm* gene sequences.** M1 serotype expression by strains representing very divergent lines of *S. pyogenes* could be explained at the molecular level in several ways. One hypothesis postulates episodes of horizontal transfer and recombination of all or part of an *emm1* gene (convergence by recombination). Alternatively, the occurrence of M protein of the same serotype in distantly related *S. pyogenes* strains may be due to mutational convergence in amino acid sequence. A third possibility is retention from a common ancestor. If recombinational convergence has occurred recently, the *emm1* genes in the distantly related lineages would be expected to be identical or nearly so. Under the hypothesis of retention from a common ancestor or mutational convergence, it is expected that differentiation of the *emm1* genes over time has occurred, resulting in several *emm1* alleles, and the topology (overall configuration) of the *emm* gene segment trees should be closely similar or identical to the topology of the trees on the basis of multilocus enzyme electrophoresis.

To differentiate among the hypotheses, a portion of the *emm* gene was characterized in 51 strains with distinct ETs and RFLP types, from different localities, and which were isolated at different times (Table 1). First, a region encoding the N-terminal variable domain of M protein was sequenced in 51 strains. The gene segment characterized includes codons 27 through 147 (GenBank sequence Z21845 [14]). Seven *emm* alleles were identified in the sample, including four that have not been previously described (Fig. 3). Twenty-four strains of

TABLE 2. Oligonucleotide primers (5′→3′) used for *S. pyogenes* gene amplification

Gene	Forward primer	Reverse primer	Product size (bp)
emm	GGGAATTCTATTSGCTTAGAAAATTAA	GCAAGTTCTTCAGCTTGTTT	Variable
speB	GTTGTCAGTGCAACTAACCGT	ATCTGTGTCTGATGGATATGCTT	1,484[a]
ska	AACCTTGCCGACCCAACCTGT	GTGAACAGTTTCAAGTGACTGCGAT	513
scpA	CAATCCCCAAAAAACCATCACC	CATACATCGTTGCTGCTGAAGC	416
speA	ATGGAAAACAATAAAAAAGTATTG	TTACTTGGTGTTAGGTAGCTTC	708

[a] Additional internal oligonucleotide primers were used for DNA sequencing.

TABLE 3. Allele profiles at five polymorphic loci in *S. pyogenes* serotype M1 strains

ET	No. of isolates	Enzyme locus[a]				
		EST	PEP	LGG	NSP	G3P
1	115	8	8	7	5	5
1D	1	8	7.5	7	5	5
1B	2	5	7	12	5	5
1C	2	8	5	7	5	5
1E	1	8	7	8	3	10
1F	5	8	5	8	5	3

[a] EST, esterase; PEP, leucylalanine peptidase; LGG, leucylglycylglycine peptidase; NSP, nucleoside phosphorylase; G3P, glyceraldehyde-3-phosphate dehydrogenase (NADP dependent).

ET 1/RFLP type 1a were studied for *emm* variation in this region; 22 had allele *emm1.0*, and 1 strain each had allele *emm1.3* (MGAS 575) and *emm1.5* (MGAS 1666). Compared with the *emm1.0* sequence, the aligned sequences of the open reading frames of alleles *emm1.3* and *emm1.5* have single nucleotide changes resulting in amino acid substitutions at residues designated 20 (N↔S) and 27 (D↔G), respectively (Fig. 3).

The *emm1.0* allele was identified in ET 1/RFLP type 1a organisms from diverse localities, including the United States, Denmark, Sweden, United Kingdom, Germany, Ethiopia, India, New Zealand, and Australia (Table 1). Strains of ET 1/RFLP type 1k had either the *emm1.3* allele ($n = 6$) or *emm1.0* ($n = 2$). The *emm1.0* allele was also identified in other ET 1 members, including organisms assigned to RFLP types 1c, 1e, 1l, 1m, and 1p. One isolate (MGAS 2108) of ET 1/RFLP type 1e had allele *emm1.6*, which differs from allele *emm1.0* by one nucleotide change resulting in an amino acid replacement at residue 42 (H↔Y) (Fig. 3).

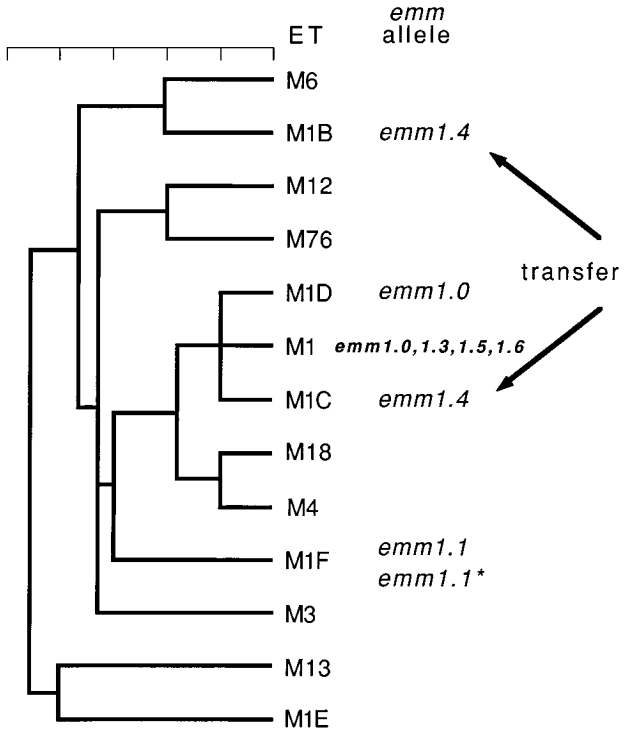

FIG. 1. Dendrogram showing estimates of overall chromosomal relatedness among six ETs expressing serotype M1 antigen and seven ETs of other M types selected to represent the range of species diversity based on multilocus enzyme electrophoresis. Each ET was characterized by its combination of alleles at 12 loci, and the dendrogram was generated from a matrix of genetic distances between pairs of ETs by the average linkage method (48). M1 strains are designated by the ETs given in Table 1. The asterisk beside *emm1.1* denotes the *emm1.1* allele plus a 21-bp insert, as described in the text. "Transfer" denotes a probable episode of horizontal transfer and recombination involving the *emm1.4* allele between two ETs that are divergent in overall chromosomal character.

In striking contrast to the very closely related *emm* variants identified in ET 1, with a single exception (MGAS 2127, which is virtually identical to ET 1 in metabolic enzyme allele profile), the *emm1* alleles found in non-ET-1 organisms were well differentiated from alleles *emm1.0*, *emm1.3*, *emm1.5*, and *emm1.6* (Fig. 1). For example, two strains of ET 1C and one strain of ET 1B had allele *emm1.4*, which differs from *emm1.0* in the region studied at 33 nucleotides, resulting in 24 amino acid changes. Similarly, one ET 1F isolate (MGAS 2123) had allele *emm1.1*, which differs from *emm1.0* in a region of 18 nucleotides which produce 14 amino acid substitutions. Interestingly, the three other ET 1F isolates studied for *emm1* structure had an allele that was identical to *emm1.1* but included a 21-bp in-frame insertion, resulting in seven additional amino acids in the deduced mature protein. Heterogeneity in RFLP pattern was also detected among isolates of ET 1F.

A 300-bp (nt 856 to 1155) segment of *emm* which encodes two of the C-repeat regions (C2 and C3) and part of the D repeat region was then sequenced in a subset of 32 of the 51 strains examined for N-terminal variable domain variation. All 24 ET 1 organisms, regardless of RFLP type, were identical in the region characterized, and the sequences identified were the same as that recently reported for *emm1.2* (GenBank sequence Z21845) (14). Among the other eight strains studied, four distinct sequences were identified. Comparison of the relatedness of the gene sequences in the C-repeat region and those in the N-terminal variable region revealed that the overall topologies of the trees were similar (data not shown), which means that the two regions have evolved in parallel with one another but at different rates.

Nucleotide variation at other loci (*scp*, *ska*, and *speB*). In order to gain additional insight into the molecular variation and evolutionary genetics of M1-expressing *S. pyogenes* clones, regions of genes encoding streptokinase (*ska*), C5a peptidase (*scp*), and interleukin-1β convertase (*speB*) were characterized in representative isolates of ET 1 (MGAS 166), ET 1F (MGAS 2109), ET 1B (MGAS 2110), and ET 1C (MGAS 2120). In the 300-bp region of *scp* studied, a total of three sites were polymorphic and two alleles were identified in the four isolates. MGAS 166, MGAS 2109, and MGAS 2120 were identical and differed from the *scp* variant in strain MGAS 2110 at 3 nucleotides (Fig. 4).

Two *ska* variants were identified in the sample of four isolates. MGAS 166 (ET 1), MGAS 2109 (ET 1F), and MGAS 2120 (ET 1C) were identical but differed from the *ska* allele in MGAS 2110 (ET 1B) at 1 nucleotide (position 913), resulting in an amino acid replacement (AAA↔CAA, Lys↔Gln) at residue 305 in the deduced sequence.

In contrast to *ska* and *scp*, a distinct *speB* variant was identified in four isolates. Five nucleotide sites were polymorphic with three having synonymous base changes and two having nonsynonymous base changes.

Geographic distribution of clones and subclones. Individual ETs may be widely distributed. Isolates of clone ET 1 and subclone ET 1/RFLP type 1a were cultured from patients in the United States (10 states) and in 12 additional countries on five continents. Non-ET-1 bacteria were far less abundant and were confined to considerably more-narrow geographic ranges.

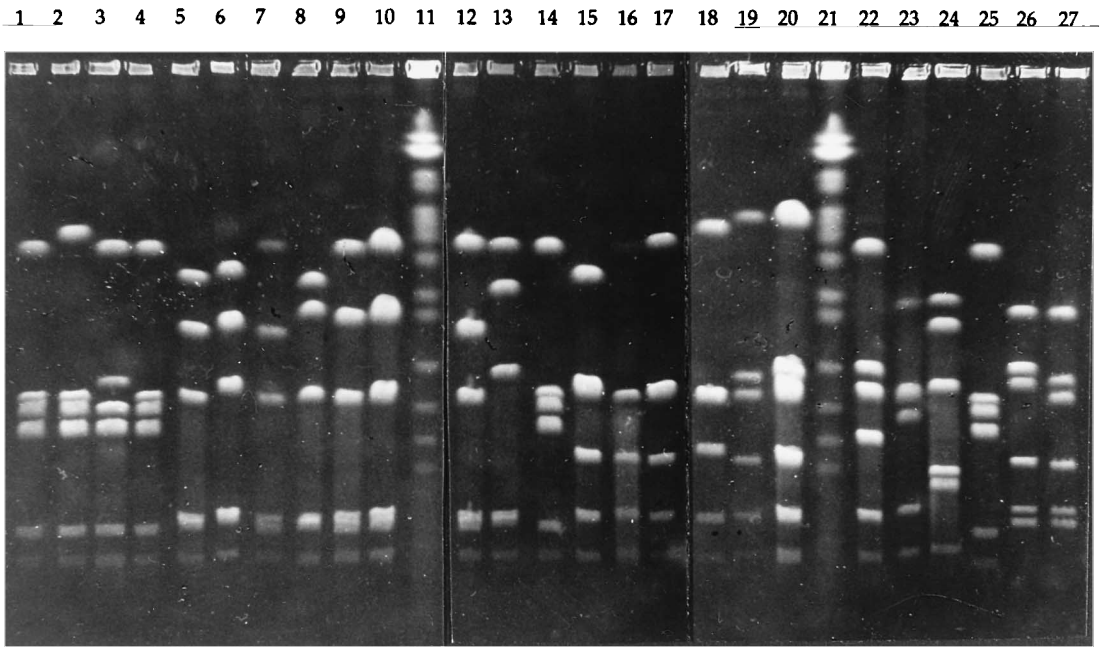

FIG. 2. RFLP patterns of representative *S. pyogenes* serotype M1. Lanes: 1, type 1a; 2, type 1s; 3, type 1p; 4, type 1a; 5, type 1b* (asterisk denotes minor variant); 6, type 1b; 7, type 1e; 8, type 1h; 9, type 1c; 10, type 1c; 11, *Saccharomyces cerevisiae* standard; 12, type 1e; 13, type 1o; 14, type 1a; 15, type 1t; 16, type 1k; 17, type 1k; 18, type 1m; 19, type 1l; 20, type 1l; 21, *Saccharomyces cerevisiae* standard; 22, type 1j; 23, type 1d; 24, type 1f; 25, type 1a; 26 and 27, non-serotype M1 strains.

For example, isolates of ET 1C were found only in Australia, and ET 1B organisms were identified only in New Zealand and Papua New Guinea but were found to differ in RFLP pattern (Table 1).

DISCUSSION

Strains expressing serotype M1 protein are genetically heterogeneous. For more than 70 years, the primary basis for strain classification and interpretation of epidemiologic and pathogenesis information of *S. pyogenes* has been M protein serotyping. The generally accepted idea that strains classified in this fashion are genetically uniform, share close overall chromosomal affinity, and behave in a biologically similar fashion has evolved. However, our data effectively rule out the notion that M1 strains represent a single phylogenetic lineage or group of closely related cell lines of *S. pyogenes*. The data also suggest the occurrence of additional non-ET-1 strains expressing M1 antigen in localities not yet sampled to an appreciable extent, such as Africa, Asia, and South America.

Temporal variation in disease frequency and severity is associated with alterations in M1 organisms. In recent years, serious infections caused by M1 strains have increased in frequency on an intercontinental scale. A critical observation, based on data presented here, is that, clearly, a distinct ET 1 subclone (ET 1/RFLP 1a/*speA* positive), rather than a heterogeneous array of sublines, has increased in frequency. Because virtually all isolates of ET 1/RFLP type 1a have the same alleles of *speA*, *speB*, *ska*, and *emm1*, we infer that the organisms have very recently spread globally and insufficient time has elapsed to accumulate synonymous mutations or other types of nucleotide changes. It is also important that RFLP (32, 41, 50) and similar analysis (39) of strain samples from Scandinavia and New Zealand have identified limited diversity among contemporary M1 isolates cultured from invasive disease episodes. Taken together, these observations imply an increase in the frequency of progeny of a more fit cell line, perhaps because of selection acting on a bacterium representing ET 1/RFLP 1a or its precursor. The observation (36, 40) that contemporary strains of M1 express a variant form of the bacteriophage-encoded (16) SPE A compared with organisms cultured from patients earlier in this century suggests that evolution of a mutant SPE A molecule has been an important factor in the contemporary resurgence of disease. It is also noteworthy that Stevens et al. (53) found that strains of M1 expressing SPE A were infrequently recovered in the United States in the 1970s and early 1980s prior to severe disease resurgence. If M1 strains circulating in that period lacked the *speA* gene, our data suggest that most of the organisms would

	10	20	30	40	50	60	70	80	90	100
emm1.0	TVLGAGFANQ	TEVKANGDGN	PREVIEDLAA	NNPAIQNIRL	RHENKDLKA-	------RLEN	AMEVAGRDFK	RAEELEKAKQ	ALEDQRKDLE	TKLKELQQDY
emm1.1	.......T..	......N..R	S.D.T.EI..	..TTV.....	.N...N...-	------....	..N.....	....	....	....
emm1.2	.......T..	......N..R	S.D.T.EI..	..TTV.....	.N...N...K	NEDLEA....	..N.....	....	....	....
emm1.3	....	.........S	....	....	.........-	------....	....	....	....	....
emm1.4	..V.....	......ENVG	..D.VKE.VE	KD.VL..K..	.S..QK..E-	------S...	.RD.....	....	...A.....	....
emm1.5	....	....	......G...	....	.........-	------....	....	....	....	....
emm1.6	....	....	....	....	.Y.......-	------....	....	....	....	....

FIG. 3. Alignment of deduced N-terminal amino acid sequences of seven alleles of *emm* from *S. pyogenes* serotype M1 isolates. The region shown represents amino acids 27 through 110 of the sequence deposited in GenBank under accession number X07860. Amino acid residues that are identical to those encoded by *emm1.0* are represented by periods, and gaps are inserted to maximize amino acid identity among the variants.

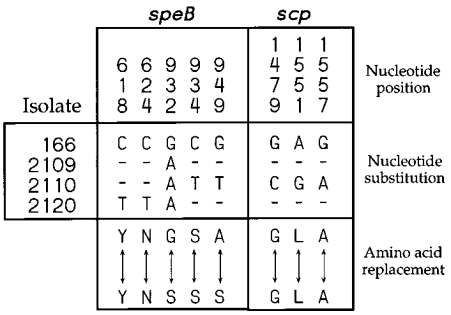

FIG. 4. Nucleotide sequence diversity in *speB* and *scp* of four clones of *S. pyogenes* serotype M1. A region of ~350 bp of *speB* that corresponds to a variable region of the mature cysteine protease which contains a linear B-cell epitope, as defined by reactivity of rabbit polyclonal sera and murine monoclonal antibodies, was sequenced (19). The region of *scp* sequenced is a 405-bp fragment located between nt 2198 and 2604. The polymorphic nucleotide positions of *speB* and *scp* are labelled in vertical format at the top of the figure, and the MGAS isolate numbers are listed on the left. The amino acid replacements (if any) resulting from the polymorphic nucleotides are shown at the bottom of the figure.

have expressed an M variant other than M1.0, such as M1.3. Harbaugh et al. (14) have shown that the 13 amino acid substitutions found in M1.1 relative to M1.0 and M1.2 result in an N-terminal variable domain that has altered opsonic epitopes in this region. In addition, evidence has recently been presented that human opsonic antibody is not necessarily type specific but rather may be strain specific (6a). It is therefore important to determine if the other variant M alleles identified in this study encode M protein with altered opsonic properties. Although three of the *emm* alleles (*emm1.3*, *emm1.5*, and *emm1.6*) would express M proteins differing by only single amino acids in the N-terminal variable domain, there is ample precedent that single amino acid changes in virulence factors can profoundly affect the relative virulence of a microbial pathogen in natural populations (8, 9, 24, 30, 33, 34, 55).

At a more defined local level, a study of 665 strains of *S. pyogenes* serotype M1 recovered from patients in New Zealand between 1980 and 1991 found that two distinct waves of invasive disease episodes were caused by strains of RFLP pattern 1a. Evidence was presented that isolates of RFLP pattern 1b were supplanted by the RFLP 1a type early in the 1980s. The multilocus enzyme electrophoretic data show that strains of RFLP types 1b and 1a represent highly divergent streptococcal phylogenetic lines, which means that they display a battery of distinct epitopes to the host. This idea is supported by the demonstration of considerable predicted amino acid diversity in the N-terminal variable region of M protein. Isolates of these two lineages also have distinct variants of SPE B, which differ at several amino acids, including position 307 (Gly↔Ser). This observation is important because it has very recently been shown that anti-cysteine protease antibodies confer protective immunity against invasive disease in a mouse model (20), and preliminary evidence indicates that a linear B-cell epitope occurs in this region of the enzyme (19). These data suggest that the ET 1/RFLP type 1a subclone may have had a selective advantage when introduced into an immunologically naive population with little or no previous exposure to the repertoire of toxins and antigens expressed by this organism. Hence, the abrupt transition from ET 1F/RFLP type 1b to the distantly related ET 1/RFLP type 1a organisms probably represents a phenomenon analogous to the antigenic shift that has been well described for influenza epidemics rather than a simple antigenic drift.

Confirmation that slipped-strand mispairing has participated in evolution of the *emm* locus. *S. pyogenes* isolates expressing the antigenic profile M1T8, which were first described in 1972 in association with an outbreak of acute streptococcal nephritis in Trinidad (43), were found to have a chromosomal lineage very different from that of commonly recovered M1 strains. One M1T8 isolate recovered in Trinidad (designated 71-155) and one M1, T-nontypeable organism (designated 76-088) cultured from an Egyptian patient (10) were recently shown (14) to have the same 21-bp insert that we found in three isolates from New Zealand which also express the T8 antigen, examined in this study. On the basis of the strong homology with an adjacent stretch of nucleotides, and the relatively small size of the insert, it was speculated (14) that the 21 bp arose because of slipped-strand mispairing (28) in the course of replication. Because the present analysis demonstrated that isolates of the same ET, which probably have shared lineal descent from a common ancestor, can have the identical *emm* sequence with or without the 21-bp insert, our data are fully consistent with the slipped-strand mispairing hypothesis.

Implications of chromosomal diversity among *S. pyogenes* strains expressing M1 protein for epidemiologic and pathogenesis research. The demonstration that M1 strains are genetically very heterogeneous has several implications for *S. pyogenes* pathogenesis and epidemiologic research. One common theme which has emerged from study of the genetic structure of natural populations of pathogenic bacteria is that although most species, including *S. pyogenes*, are polymorphic in multilocus enzyme genotype, a relatively small number of cell lineages marked by distinct ETs cause most of the serious disease episodes (49). This observation means that the interlineage variance in biologic behavior, including relative pathogenicity, is very large. The phenomenon is readily illustrated in natural populations of *S. pyogenes* by the observation that two ETs (ET 1 and ET 2) are responsible for well over one-half of all cases of streptococcal toxic shock syndrome in the United States and Europe (36). By demonstrating that virtually all serious invasive infections caused by M1 strains are due to organisms of only 2 of 17 ET-RFLP types expressing M1 serotype (1 of which is strongly associated with SPE A production), this theme is extended to the subclone level. In this regard, it is also noteworthy that in two mouse models of invasive streptococcal disease, isolates of ET 1 that express SPE A (subclone ET 1/RFLP 1a) have been shown to be more virulent than isolates of ET 1 that do not produce this toxin (subclone ET 1/RFLP 1k). For example, the 50% lethal dose of an ET 1/RFLP 1a strain was more than an order of magnitude lower than that of an ET 1/RFLP 1k organism (37).

Implicit in the application of M typing, and other schemes of bacterial strain typing, is the notion that distinctive organisms or cell lines are being typed or identified, rather than merely individual traits. In clonally structured populations such as *Escherichia coli* and *Legionella pneumophila*, knowledge of the state of one character provides information on the probable states of other traits (49). Because many *S. pyogenes* strains of the same M type are clonally related (36), this typing system has provided important insights into the epidemiology of streptococcal infections, mainly by delineating nonrandom associations of M type with certain disease syndromes, most notably poststreptococcal glomerulonephritis and acute rheumatic fever. However, when the same M protein serotype trait occurs in association with distinct chromosomal genotypes, as we have found for the M1 organisms, it is likely that associations among serotypes and other characters will be far less strong. In essence, the serotyping scheme becomes merely a convenient

method for pigeonholing strains and, hence, loses much of its utility. This means that when M serotyping is practiced, it should be conducted with the understanding that substantial levels of genetic diversity can exist among isolates assigned to the same M serotype and that the interclonal variances in traits such as virulence factors, relative virulence, and disease associations are likely to be very large. Although only M1 strains were examined in this study, it will be shown elsewhere that clonal diversity among strains classified on the basis of M protein serotype is common (unpublished data).

Horizontal transfer of genes conferring adaptive traits. The data presented here add to the general concept in bacterial evolutionary genetics and pathobiology that horizontal transfer and recombination of genes encoding or mediating traits thought to confer adaptive advantage are important mechanisms used by pathogenic microbes to diversify populations and enhance survival. Convincing molecular data that recombination events have been major forces responsible for enhancing diversity in the *Salmonella enterica* flagellin gene (29), *Salmonella* O antigen (44), *Neisseria meningitidis* class 1 outer membrane protein gene (11), penicillin-binding protein genes in *Neisseria* sp. and *Streptococcus* sp. (7, 52), and genes in several other pathogenic bacteria (47) have been presented. For *S. pyogenes*, there is evidence of intragenic or assortive recombination occurring for genes encoding exotoxin A (40), exotoxin B (interleukin-1β convertase) (23), exotoxin C (22), Emm (3, 57, 58), Enn (56), and streptokinase (18). Lateral transfer of virulence-related and other genes clearly is a common theme in *S. pyogenes* and warrants further study at the molecular level.

ACKNOWLEDGMENTS

We thank Wezenet Tewordos (Karolinska Institute, Stockholm, Sweden), Shipra Dutta (Lady Hardinge Medical College, New Delhi, India), and Edward Kaplan (University of Minnesota, Minneapolis) for providing some of the strains.

This study was supported by American Heart Association grant-in-aid 92-006640, by National Institutes of Health grant AI-33119, and by funds supplied by the Health Research Council of New Zealand. J.M.M. is an Established Investigator of the American Heart Association.

REFERENCES

1. **Beachey, E. H., and M. Kotb.** 1992. Prospects for group A streptococcal vaccine, p. 115–136. *In* M. Hook and L. M. Switalski (ed.), Microbial adhesion and invasion. Springler-Verlag, New York.
2. **Bessen, D., and V. A. Fischetti.** 1988. Influence of intranasal immunization with synthetic peptides corresponding to conserved epitopes of M protein on mucosal colonization by group A streptococci. Infect. Immun. **56:**2666–2672.
3. **Bessen, D. E., and S. K. Hollingshead.** 1994. Allelic polymorphism of *emm* loci provides evidence for horizontal gene spread in group A streptococci. Proc. Natl. Acad. Sci. USA **91:**3280–3284.
4. **Bisno, A. L.** 1991. Group A streptococcal infections and acute rheumatic fever. N. Engl. J. Med. **325:**783–793.
5. **Chen, C. C., and P. P. Cleary.** 1990. Complete nucleotide sequence of the streptococcal C5a peptidase gene of *Streptococcus pyogenes*. J. Biol. Chem. **265:**3161–3167.
6. **Cleary, P. P., D. LaPenta, D. Heath, E. J. Haanes, and C. Chen.** 1991. A virulence regulon in *Streptococcus pyogenes*, p. 147–151. *In* G. M. Dunny, P. P. Cleary, and L. L. McKay (ed.), Genetics and molecular biology of streptococci, lactococci, and enterococci. American Society for Microbiology, Washington, D.C.

6a. **de Malmanche, S. A., and D. R. Martin.** Protective immunity to the group A *Streptococcus* may be only strain specific. Med. Microbiol. Immunol., in press.

7. **Dowson, C. G., A. Hutchinson, N. Woodford, A. P. Johnson, R. C. George, and B. G. Spratt.** 1990. Penicillin-resistant viridans streptococci have obtained altered penicillin-binding protein genes from penicillin-resistant strains of *Streptococcus pneumoniae*. Proc. Natl. Acad. Sci. USA **87:**5858–5862.
8. **Duim, B., L. van Alphen, P. Eijk, H. M. Jansen, and J. Dankert.** 1994. Antigenic drift of non-encapsulated *Haemophilus influenzae* major outer membrane protein P2 in patients with chronic bronchitis is caused by point mutations. Mol. Microbiol. **11:**1181–1189.
9. **Ehata, T., M. Omata, W. L. Chuang, O. Yokosuka, Y. Ito, K. Hosoda, and M. Ohto.** 1993. Mutations in core nucleotide sequence of hepatitis B virus correlate with fulminant and severe hepatitis. J. Clin. Invest. **91:**1206–1213.
10. **El Kholy, A., N. I. Guirguis, L. W. Wannamaker, and R. M. Krause.** 1975. Newly recognized distinctive M protein associated with certain M type 1 group A streptococci. Infect. Immun. **11:**551–555.
11. **Feavers, I. M., A. B. Heath, J. A. Bygraves, and M. C. J. Maiden.** 1992. Role of horizontal gene exchange in the antigenic variation of the class I outer membrane protein of *Neisseria meningitidis*. Mol. Microbiol. **6:**489–495.
12. **Fischetti, V. A.** 1989. Streptococcal M protein: molecular design and biological behavior. Clin. Microbiol. Rev. **2:**285–314.
13. **Fischetti, V. A., W. M. Hodges, and D. E. Hruby.** 1989. Protection against streptococcal pharyngeal colonization with a vaccinia: M protein recombinant. Science **244:**1487–1490.
14. **Harbaugh, M. P., A. Podbielski, S. Hugl, and P. P. Cleary.** 1993. Nucleotide substitutions and small-scale insertion produce size and antigenic variation in group A streptococcal M1 protein. Mol. Microbiol. **8:**981–991.
15. **Hauser, A. R., D. L. Stevens, E. L. Kaplan, and P. M. Schlievert.** 1991. Molecular analysis of pyrogenic exotoxins from *Streptococcus pyogenes* isolates associated with toxic shock-like syndrome. J. Clin. Microbiol. **29:**1562–1567.
16. **Johnson, L. P., and P. M. Schlievert.** 1984. Group A streptococcal phage T12 carries the structural gene for pyrogenic exotoxin type A. Mol. Gen. Genet. **194:**52–56.
17. **Johnston, K. H., J. E. Chaiban, and R. C. Wheeler.** 1992. Analysis of the variable domain of the streptokinase gene associated with post-streptococcal glomerulonephritis. Zentralbl. Mikrobiol. **22**(Suppl.)**:**339–342.
18. **Kapur, V., S. Kanjilal, M. R. Hamrick, L.-L. Li, T. S. Whittam, S. A. Sawyer, and J. M. Musser.** Molecular population genetic analysis of the streptokinase gene of *Streptococcus pyogenes*: mosaic alleles generated by recombination. Mol. Microbiol., in press.
19. **Kapur, V., L.-L. Li, and J. M. Musser.** 1994. Identification of linear B-cell epitopes in a conserved *Streptococcus pyogenes* extracellular cysteine protease (interleukin-1β convertase), abstr. B31, p. 62. *In* Abstracts of the 34th Interscience Conference on Antimicrobial Agents and Chemotherapy. American Society for Microbiology, Washington, D.C.
20. **Kapur, V., J. T. Maffei, R. S. Greer, L.-L. Li, G. J. Adams, and J. M. Musser.** 1994. Vaccination with streptococcal extracellular cysteine protease (interleukin-1β convertase) protects mice against challenge with heterologous group A streptococci. Microb. Pathog. **16:**443–450.
21. **Kapur, V., M. W. Majesky, L.-L. Li, R. A. Black, and J. M. Musser.** 1993. Cleavage of interleukin 1β (IL-1β) precursor to produce active IL-1β by a conserved extracellular cysteine protease from *Streptococcus pyogenes*. Proc. Natl. Acad. Sci. USA **90:**7676–7680.
22. **Kapur, V., K. Nelson, P. M. Schlievert, R. K. Selander, and J. M. Musser.** 1992. Molecular population genetic evidence of horizontal spread of two alleles of the pyrogenic exotoxin C gene (*speC*) among pathogenic clones of *Streptococcus pyogenes*. Infect. Immun. **60:**3513–3517.
23. **Kapur, V., S. Topouzis, M. W. Majesky, L.-L. Li, M. R. Hamrick, R. J. Hamill, J. M. Patti, and J. M. Musser.** 1993. A conserved *Streptococcus pyogenes* extracellular cysteine protease cleaves human fibronectin and degrades vitronectin. Microb. Pathog. **15:**327–346.
24. **Kawaoka, Y., and R. G. Webster.** 1988. Molecular mechanism of acquisition of virulence in influenza virus in nature. Microb. Pathog. **5:**311–318.
25. **Kehoe, M. A.** 1994. Cell-wall associated proteins in Gram-positive bacteria. New Compr. Biochem. **27:**217–261.
26. **Kohler, W.** 1990. Streptococcal toxic shock syndrome. Zentralbl. Bakteriol. **272:**257–264.
27. **Lancefield, R. C.** 1962. Current knowledge of the type specific M antigens of group A streptococci. J. Immunol. **89:**307–313.
28. **Levinson, G., and G. A. Gutman.** 1987. Slipped-strand mispairing: a major mechanism for DNA sequence evolution. Mol. Biol. Evol. **4:**203–221.
29. **Li, J., K. Nelson, A. C. McWhorter, T. S. Whittam, and R. K. Selander.** 1994. Recombinational basis of serovar diversity in *Salmonella enterica*. Proc. Natl. Acad. Sci. USA **91:**2552–2556.
30. **Liang, T. J., K. Hasegawa, N. Rimon, J. R. Wands, and E. Ben-Porath.** 1991. A hepatitis B virus mutation associated with an epidemic of fulminant hepatitis. N. Engl. J. Med. **324:**1705–1709.
31. **Martin, D. R., G. E. Meekin, and L. A. Finch.** 1983. Streptococci—are they different in New Zealand? N. Z. Med. J. **96:**298–301.
32. **Martin, D. R., and L. A. Single.** 1993. Molecular epidemiology of group A streptococcus M type 1 infections. J. Infect. Dis. **167:**1112–1117.
33. **McGuinness, B. T., N. Clarke, P. R. Lambden, A. K. Barlow, J. T. Poolman, D. M. Jones, and J. E. Heckels.** 1991. Point mutation in meningococcal *porA* gene associated with increased endemic disease. Lancet **337:**514–517.
34. **McGuinness, B. T., P. R. Lambden, and J. E. Heckels.** 1993. Class 1 outer membrane protein of *Neisseria meningitidis*: epitope analysis of the antigenic diversity between strains, implications for subtype definition and molecular epidemiology. Mol. Microbiol. **7:**505–514.
35. **Musser, J. M., B. M. Gray, P. M. Schlievert, and M. E. Pichichero.** 1992. *Streptococcus pyogenes* pharyngitis: characterization of strains by multilocus

enzyme genotype, M and T protein serotype, and pyrogenic exotoxin gene probing. J. Clin. Microbiol. **30:**600–603.
36. **Musser, J. M., A. R. Hauser, M. H. Kim, P. M. Schlievert, K. Nelson, and R. K. Selander.** 1991. *Streptococcus pyogenes* causing toxic-shock-like syndrome and other invasive diseases: clonal diversity and pyrogenic exotoxin expression. Proc. Natl. Acad. Sci. USA **88:**2668–2672.
37. **Musser, J. M., V. Kapur, S. Kanjilal, U. Shah, D. M. Musher, N. L. Barg, K. H. Johnston, P. M. Schlievert, J. Henrichsen, D. Gerlach, R. M. Rakita, A. Tanna, B. D. Cookson, and J. C. Huang.** 1993. Geographic and temporal distribution and molecular characterization of two highly pathogenic clones of *Streptococcus pyogenes* expressing allelic variants of pyrogenic exotoxin A (scarlet fever toxin). J. Infect. Dis. **167:**337–346.
38. **Musser, J. M., V. Kapur, J. E. Peters, C. W. Hendrix, D. Drehner, G. D. Gackstetter, D. R. Skalka, P. L. Fort, J. T. Maffei, L.-L. Li, and G. P. Melcher.** 1994. Real-time molecular epidemiologic analysis of an outbreak of *Streptococcus pyogenes* invasive disease in US Air Force trainees. Arch. Pathol. Lab. Med. **118:**128–133.
39. **Mylvaganam, H., B. Bjorvatn, T. Hofstad, R. Hjetland, E. A. Hoiby, and S. E. Holm.** 1994. Small-fragment restriction endonuclease analysis in epidemiological mapping of group A streptococci. J. Med. Microbiol. **40:**256–260.
40. **Nelson, K., P. M. Schlievert, R. K. Selander, and J. M. Musser.** 1991. Characterization and clonal distribution of four alleles of the *speA* gene encoding pyrogenic exotoxin A (scarlet fever toxin) in *Streptococcus pyogenes*. J. Exp. Med. **174:**1271–1274.
41. **Norgren, M., A. Norrby, and S. E. Holm.** 1992. Genetic diversity in T1M1 group A streptococci in relation to clinical outcome of infection. J. Infect. Dis. **166:**1014–1020.
42. **Okhuni, H., Y. Todome, H. Suzuki, M. Mizuse, N. Kotani, K. Horiuchi, N. Shikama, A. Tsugita, and K. H. Johnston.** 1992. Immunochemical studies and complete amino acid sequence of the streptokinase from *Streptococcus pyogenes* (group A) M type 12 strain. Infect. Immun. **60:**278–283.
43. **Potter, E. V., M. Svartman, E. G. Burt, J. F. Finklea, T. Poon-King, and D. P. Earle.** 1972. Relationship of acute rheumatic fever to acute glomerulonephritis in Trinidad. J. Infect. Dis. **125:**619–625.
44. **Reeves, P. R.** 1993. Evolution of *Salmonella* O antigen variation by interspecific gene transfer on a large scale. Trends Genet. **9:**17–22.
45. **Schwartz, B., R. R. Facklam, and R. F. Breiman.** 1990. Changing epidemiology of group A streptococcal infection in the USA. Lancet **336:**1167–1171.
46. **Scott, J. R.** 1990. The M protein of group A *Streptococcus*: evolution and regulation, p. 177–203. *In* B. H. Iglewski and V. L. Clark (ed.), Molecular basis of bacterial pathogenesis. Academic Press, New York.
47. **Selander, R. K., J. Li, E. F. Boyd, F.-S. Wang, and K. Nelson.** 1994. DNA sequence analysis of the genetic structure of populations of *Salmonella enterica* and *Escherichia coli*, p. 1–36. *In* F. G. Priest, A. Ramos-Cormenzana, and R. Tindall (ed.), Bacterial systematics and diversity. Plenum, New York.
48. **Selander, R. K., D. A. Caugant, H. Ochman, J. M. Musser, M. N. Gilmour, and T. S. Whittam.** 1986. Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. Appl. Environ. Microbiol. **51:**873–884.
49. **Selander, R. K., and J. M. Musser.** 1990. Population genetics of microbial pathogenesis, p. 11–38. *In* B. H. Iglewski and V. L. Clark (ed.), Molecular basis of bacterial pathogenesis. Academic Press, New York.
50. **Seppälä, H., J. Vuopio-Varkila, M. Osterblad, M. Jahkola, M. Rummukainen, S. E. Holm, and P. Huovinen.** 1994. Evaluation of methods for epidemiologic typing of group A streptococci. J. Infect. Dis. **169:**519–525.
51. **Single, L. A., and D. R. Martin.** 1992. Clonal differences within M-types of the group A streptococcus revealed by pulsed field gel electrophoresis. FEMS Microbiol. Lett. **91:**85–90.
52. **Spratt, B. G., L. D. Bowler, Q. Y. Zhang, J. Zhou, and J. Maynard Smith.** 1992. Role of interspecies transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. J. Mol. Evol. **34:**115–125.
53. **Stevens, D. L.** 1992. Invasive group A streptococcus infections. Clin. Infect. Dis. **14:**2–13.
54. **Talkington, D. F., B. Schwartz, C. M. Black, J. K. Todd, J. Elliott, R. F. Breiman, and R. R. Facklam.** 1993. Association of phenotypic and genotypic characteristics of invasive *Streptococcus pyogenes* isolates with clinical components of streptococcal toxic shock syndrome. Infect. Immun. **61:**3369–3374.
55. **Tucker, P. C., E. G. Strauss, R. J. Kuhn, J. H. Strauss, and D. E. Griffin.** 1993. Viral determinants of age-dependent virulence of Sindbis virus for mice. J. Virol. **67:**4605–4610.
56. **Whatmore, A. M., V. Kapur, J. M. Musser, and M. A. Kehoe.** Molecular population genetic analysis of the *enn* subdivision of group A streptococcal *emm*-like genes: horizontal transfer and restricted variation among *enn* genes. Mol. Microbiol., in press.
57. **Whatmore, A. M., V. Kapur, D. J. Sullivan, J. M. Musser, and M. A. Kehoe.** 1994. Non-congruent relationships between variation in *emm* gene sequences and the population genetic structure of group A streptococci. Mol. Microbiol. **14:**619–631.
58. **Whatmore, A. M., and M. A. Kehoe.** 1994. Horizontal gene transfer in the evolution of group A streptococcal *emm*-like genes: gene mosaics and variation in Vir regulons. Mol. Microbiol. **11:**363–374.