Molecular Population Genetic Evidence of Horizontal Spread of Two Alleles of the Pyrogenic Exotoxin C Gene (*speC*) among Pathogenic Clones of *Streptococcus pyogenes*

VIVEK KAPUR,¹ KIMBERLYN NELSON,² PATRICK M. SCHLIEVERT,³ ROBERT K. SELANDER,² AND JAMES M. MUSSER¹*

Section of Molecular Pathobiology, Department of Pathology, Baylor College of Medicine, Houston, Texas 77030¹; Institute of Molecular Evolutionary Genetics, Pennsylvania State University, University Park, Pennsylvania 16802²; and Department of Microbiology, University of Minnesota, Minneapolis, Minnesota 55455³

Received 31 March 1992/Accepted 4 June 1992

It has recently been demonstrated that the bacteriophage-borne gene (speC) encoding pyrogenic exotoxin C is harbored by phylogenetic lineages representing virtually the entire breadth of genomic differentiation present in the species *Streptococcus pyogenes* (J. M. Musser, A. R. Hauser, M. H. Kim, P. M. Schlievert, K. Nelson, and R. K. Selander, Proc. Natl. Acad. Sci. USA 88:2668-2672, 1991). To determine whether the *speC* genes occurring in association with divergent chromosomal genotypes (clones) are identical or represent a group of allelic variants, we sequenced *speC* from 23 *S. pyogenes* strains representing 15 clones identified by multilocus enzyme electrophoresis. Two alleles of *speC* are present in natural populations, and each allele occurs in clones that are well differentiated in overall chromosomal character; in one case, isolates of a single clone had different *speC* alleles. We interpret these patterns of toxin allele-clone distribution as evidence of occasional episodes of *speC* horizontal dissemination, presumably by bacteriophage-mediated gene transfer and recombination.

An increase in the number of unusually severe Streptococcus pyogenes invasive infections and acute rheumatic fever (ARF) episodes reported in recent years in the United States (1, 20, 25) and elsewhere (7) stimulated renewed interest in this pathogen. The organism expresses a heterogeneous array of extracellular molecules that are believed to be causally involved in the pathogenesis of various diseases. Among the putative virulence factors synthesized by S. pyogenes are streptokinase, hyaluronidase, and DNase and the extracellular antiphagocytic molecules M protein and hyaluronic acid. Strains of S. pyogenes also have the ability to produce one or more of three pyrogenic exotoxins (2). Pyrogenic exotoxin A (scarlet fever toxin) and exotoxin C are bacteriophage encoded and are variably present in isolates recovered from natural populations (16, 28, 29), whereas all isolates harbor sequences hybridizing with a specific probe for the pyrogenic exotoxin B gene (11). These toxins are members of a family of "superantigen" molecules that are also expressed by Staphylococcus aureus and other microbes (14). Most of them have similar amino acid sequences and functional activities.

Superantigen research has been an especially active area of investigation in recent years because these molecules stimulate lymphocytes in a novel fashion that apparently does not require antigen processing. In addition, there is much speculation (14) and some evidence (4, 10) that superantigens are involved in the pathogenesis of several autoimmune disorders, including ARF and rheumatic heart disease. It has recently been shown that streptococcal pyrogenic exotoxin (SPE) C is the most effective of the three streptococcal toxins in stimulating human T-cell mitogenesis (26).

Epidemiologic evidence has demonstrated that SPE A is preferentially associated with isolates recovered from patients with a recently described, unusually severe invasive disease termed streptococcal toxic shock-like syndrome (TSLS) (16). However, not all isolates cultured from patients with TSLS or other severe invasive diseases (SID) express SPE A. Many of these other strains produce pyrogenic exotoxin B, alone or in combination with exotoxin C (16), an observation suggesting that expression of the two toxins is involved in the pathogenesis of some cases of severe streptococcal diseases.

In a study (16) of the molecular population genetics of strains recovered from patients with TSLS and other SID, it was demonstrated that speC is associated with many distinct phylogenetic lineages of the species, some of which are highly divergent in overall chromosomal character. For example, several isolates differed at over half of the 12 loci assayed for electrophoretic variation, and some isolates diverged from one another at a genetic distance of 0.70, a level of differentiation that in many species approximates species-level distinction (23). However, it was not determined whether the speC sequences harbored by strains of different chromosomal genotypes are identical or whether they represent a family of allelic variants of this toxin gene. The purpose of the present study was to address this question by sequencing speC from strains which represent many evolutionary lineages of S. pyogenes and which were recovered from patients with defined clinical syndromes.

Our analysis has revealed that there are two distinct alleles of *speC* present in isolates from natural populations of *S. pyogenes*, including strains recovered from patients with TSLS, and in organisms expressing M types historically associated with ARF. Each allele is found in clones that are well differentiated in overall chromosomal character, and in one case, isolates of the same clone harbored different alleles. The most parsimonious hypothesis accounting for the observed toxin allele-clone distribution pattern postulates occasional episodes of *speC* horizontal transfer and

^{*} Corresponding author.

Clonal group (ET)	MGAS strain	Protein serotype ^a	Allele	Locality	Disease ^b
1	19	M1	speC1	Michigan	RP
1	307 ^c	M1	speC1	Texas	TSLS
2	276	T28	speC2	Colorado	SID
2	289	T28	speC2	Colorado	SID
2	332	M3	speC2	United States	Invasive
2	1251	M3	speC1	Unknown	Scarlet fever
3	270	NT	speC2	Minnesota	SID
8	280 ^c	Т9	speC2	Colorado	SID
9	323	T12/13	speC2	Idaho	Invasive
10	255 ^c	NE	speC2	California	SID
11	296	T9/5/27/44	speC2	Washington	Invasive
14	321	M4	speC1	Washington	TSLS
16	168	M66	speC2	Texas	Invasive
18	264	NE	speC2	California	SID
20	156	M18	speC1	Nebraska	TSLS
20	300	M18	speC1	Washington	Invasive
20	1585 ^c	M18	speC1	Great Lakes Naval Training Center, Ill.	ARF
20	1587 ^c	M18	speC1	Ohio	ARF
22	265	NE	speC2	California	Invasive
24	162	M22	speC1	Illinois	SID
28	254	NE	speC1	California	SID
28	258 ^c	NE	speC1	California	SID
32	317 ^c	NE	speC1	Idaho	Invasive

TABLE 1. Properties of 23 strains of S. pyogenes representing 15 clonal groups

^a NT, nontypeable; NE, not examined.

^b RP, recurrent pharyngitis; TSLS, toxic shock-like syndrome; SID, severe invasive disease; ARF, acute rheumatic fever associated.

^c Partial speC sequence data only.

recombination, presumably occurring by bacteriophagemediated transduction.

MATERIALS AND METHODS

Bacterial isolates. A sample of 20 S. pyogenes isolates representing 15 distinct clonal lineages, among which there is an estimated maximum genetic distance of 0.70, was drawn from a larger collection of strains previously characterized by multilocus enzyme electrophoresis (16). The sample includes one strain cultured from a child with recurrent pharyngitis, three strains recovered from patients with TSLS, seven strains obtained from individuals who had septicemia but for whom we lack sufficient clinical information to permit assignment to the TSLS or the SID category, and nine strains recovered from cases of SID. We also studied one strain (MGAS 1587) recovered in 1986 from the throat of a patient in Ohio in the course of an ARF outbreak that has been described elsewhere (6) and one strain (MGAS 1585) associated with an outbreak of ARF at the Great Lakes Naval Training Center in the 1940s (Table 1). This strain was used as the source of the speC gene originally sequenced by Goshorn and Schlievert (8).

Genomic DNA preparation. Genomic DNA was prepared from isolates grown on brain heart infusion agar plates. In general, cells were scraped from one plate, suspended in 800 μ l of 10 mM Tris-50 mM EDTA (pH 8.0) (TE), heated at 65°C for 15 min, washed, resuspended in 500 μ l of TE containing 5 μ g of mutanolysin, and incubated at 37°C for 2 h. The cells were lysed by adding 100 μ l of 10% sodium dodecyl sulfate and heating at 65°C for 20 min. After centrifugation for 10 min, the supernatant was transferred to a clean tube and incubated overnight at 37°C with 100 μ g of RNase and 50 μ g of proteinase K. The DNA was then extracted with phenol-chloroform, precipitated with ethanol, and suspended in 100 μ l of TE. Genetic relationships among clones. Methods of estimating genetic relationships among *S. pyogenes* clones by multilocus enzyme electrophoresis have been described previously (16). Each isolate was characterized by its combination of alleles at 12 enzyme loci (16). The electrophoretic type (ET) designations are cognate with those used previously (16).

Sequencing of speC. The speC gene was amplified by polymerase chain reaction (PCR) with the following synthetic oligonucleotide primers: 5' primer (SPEC-1), 5'-AT GAAAAAGATTAACATCATCAAAAT-3'; and 3' primer (SPEC-X), 5'-TTCAAGATAAATATCGAAATGAC-3'. Single-stranded DNA templates were produced by lambda exonuclease digestion following PCR amplification (12), and the resulting 702-bp fragments were sequenced in both directions with the aid of four additional internal sequencing primers with Sequenase version 2.0 (U.S. Biochemicals, Cleveland, Ohio). The sequences were assembled and edited with SEQMAN and SEQMANED programs (Dnastar, Madison, Wis.). Seven of the strains we studied were not sequenced for the entire 702-bp fragment described above; for these strains, we sequenced a fragment of speC of at least 300 bp that included the variable region differentiating speC1 and speC2.

Nucleotide sequence accession numbers. The DNA sequence data for *speC1* and *speC2* are available from Gen-Bank under accession numbers M97156 and M97157, respectively.

RESULTS AND DISCUSSION

Phylogenetic relationships among multilocus genotypes. The chromosomal relationships among the multilocus enzyme genotypes of 19 of the strains were presented earlier (16). The multilocus metabolic enzyme allele profile of strain MGAS 19 was identical to that reported previously for isolates of ET 1, the profile of MGAS 1251 was that of ET 2,



FIG. 1. Dendrogram showing estimates of genetic relationships of 15 ETs of 23 isolates of *S. pyogenes* harboring *speC*, based on allele profiles at 12 enzyme loci. The dendrogram was generated from a matrix of genetic distances between pairs of ETs by the average-linkage method (21). The ET designations are cognate with those employed in earlier studies (15, 16). There are six major clusters identified by the letters A through F. M and T protein serotypes and *speC* alleles of isolates of each ET are indicated. Strains are designated by the ET numbers given in Table 1.

and the allele profiles of MGAS 1585 and MGAS 1587 were those of ET 20.

Estimates of the overall genetic relationships among the isolates are summarized in the dendrogram in Fig. 1. At a genetic distance of 0.35, there were six branches, designated A through F. Clusters A and B were separated from lineages C and D at a genetic distance of about 0.45. Cluster E diverges from clusters A through D at a genetic distance of 0.50, which means that ETs in these two divisions of the dendrogram differ, on average, at half of the 12 enzyme loci assayed. Lineage F (ET 32) was even more divergent, and was separated from phylogenetic lines A through E at a genetic distance of 0.70. It is noteworthy that estimates of genetic relatedness based on multilocus enzyme electrophoresis of Haemophilus spp. (19), Legionella spp. (22), and other bacteria (23) have shown that divergence of clones at a genetic distance of 0.55 frequently means that the lineages will show less than about 70% nucleotide sequence relatedness by DNA hybridization, a value now widely accepted as the criterion of species limits in many groups of bacteria (3, 27). Hence, many of these chromosomal lines are strongly differentiated from one another and have not recently shared a common ancestor.

Gene sequences of exotoxin C. Failure to identify the published *speC* sequence by comparative analysis of this gene from 16 isolates representing 15 distinct clonal lineages suggested that the sequence encoding the signal peptide region reported previously (8) is incorrect. This notion was confirmed by reanalysis of the *speC* gene from strain T18P



FIG. 2. Nucleotide and deduced amino acid sequences of SPE C. (A) Sequencing gel depicting nucleotide sequence encoding the putative signal peptide of SPE C. The arrows at the top and bottom indicate the base numbers relative to the start codon. The arrow at position 60 indicates the presence of a single T residue followed by an A residue at position 61, representing a difference from the sequence previously reported (8). Similarly, the arrow at position 76 indicates an A residue followed by three additional A residues, which also represents a difference from the previously reported sequence. (B) Alignment of previously reported (I) and corrected (II) nucleotide sequences of speC. The reading frame of the corrected sequence shifts at position 61 and reverts back to the reported sequence at position 77. (C) Alignment of previously reported (I) and corrected (II) deduced amino acid sequences of SPE C. The shift in reading frame of the sequence at position 61 results in changes in six amino acid residues from positions 21 through 26, as indicated by the colons.

(MGAS 1585), which was used in the earlier study by Goshorn and Schlievert (8). Compared with the published sequence, the sequences of all strains examined lacked a T at position 61 and had an A at position 76 (Fig. 2) (numbering is cognate with that in reference 8). As a consequence of these revisions, a stretch of six amino acids extending from position 21 through position 26 in the putative signal sequence of SPE C changes from Tyr-Phe-Thr-Tyr-His-Gln to Ile-Ser-Pro-Ile-Ile-Lys (Fig. 2). The presence of an additional A residue at position 76 allows the reading frame to revert to the previously reported sequence (8). We refer to this sequence as allele *speC1*, and it is otherwise identical to the previously published sequence.

We also identified a second allele, designated speC2, which differs from speC1 by nucleotide changes at positions



FIG. 3. Alleles of SPE C. The arrows at the top and the bottom indicate the nucleotide positions with respect to the start codon and are cognate with the previously reported numbering system (8). The allele *speC2* differs from *speC1* at positions 438 and 453, as indicated by the arrows. The G-to-A transitions at both positions represent synonymous changes in codons for lysine residues.

438 and 453, both of which are synonymous (silent) A-to-G transitions in the third position of codons for lysine (Fig. 3).

Distribution of exotoxin C alleles among clonal lineages. The distribution of two speC alleles among the streptococcal clonal lineages is shown in Fig. 1. Both are widely distributed in the range of chromosomal diversity in our sample. For example, speC1 occurred in association with ET 1, ET 2, ET 14, ET 20, ET 24, ET 28, and ET 32. Hence, there is no special speC allele-genotype cluster relationship, a circumstance that contrasts with that previously reported for alleles of the gene (speA) encoding SPE A (17). With the single exception of isolates of ET 2, all isolates assigned to individual chromosomal genotypes harbored the same speC allele. Isolates of ET 2 were associated with both speC1 (1 isolate) and speC2 (3 isolates), and intriguingly, all three isolates with speC2 were recovered from contemporary invasive disease episodes, whereas strain MGAS 1251, with speC1, was cultured in the 1920s. Strain MGAS 1251 was characterized on two separate occasions for multilocus enzyme genotype and speC allele.

Evidence for SPE C gene dissemination by horizontal gene transfer. Although some of the speC allele-clone associations presented in Fig. 1 can be explained on the basis of chromosomal divergence occurring after the differentiation of speC1 and speC2, two lines of evidence suggest the idea that toxin genes are sometimes disseminated in natural populations of streptococci by episodes of horizontal gene transfer and additive recombination, presumably by bacteriophage-mediated transduction. First, speC1 and speC2 each occur in phylogenetic lineages that are in many cases highly divergent in chromosomal character, as indexed by multilocus enzyme electrophoresis. For example, MGAS 317, assigned to ET 32, harbors speC1, and the same toxin allele occurs in association with six other multilocus enzyme genotypes that diverge from lineage F at a genetic distance of 0.70, a level of differentiation that in many genera approximates species-level distinction (23). Similarly, MGAS 321 (ET 14) harbors the speC1 allele, yet cluster analysis indicates that ET 14 is nested among seven other ETs of lineage

B, all of which carry the speC2 allele. A second line of evidence is the occurrence of both speC1 and speC2 among isolates assigned to the same clonal lineage (ET 2). It is most unlikely that the toxin allele-clone associations revealed by our analysis are the result of multiple episodes of evolutionary convergence involving mutation and natural selection, because the only two base pair changes differentiating speC1 and speC2 are synonymous substitutions, and are, therefore, unlikely to affect fitness.

Inasmuch as our data are consistent with episodes of bacteriophage-mediated horizontal transfer and recombination of *speC* in mediating toxin gene dissemination, it is noteworthy that several investigators (5, 9, 13, 18) have provided evidence that lysogenic conversion to SPE C production can occur in the laboratory. Although we did not investigate strains for the presence of other bacteriophageencoded genes, a previous study (9) of *speC*-positive organisms showed that the *speC* gene harbored by phage CS112 was linked to a common phage-derived DNA fragment in several strains. Additional studies will be required to determine exactly how many distinct phages harbor *speC*.

Concluding comment. Gene sequencing and other molecular population genetic approaches have provided evidence (16, 20) that the genes (*speA* and *speC*) encoding bacterial superantigens SPE A and SPE C and the gene (*emm12*) encoding protein serotype M12 (24) have been laterally transferred among streptococcal clonal lineages. It is possible that horizontal gene transfer plays an important role in the molecular origin of new and perhaps unusually virulent cell lineages of this pathogen.

ACKNOWLEDGMENTS

This study was supported by American Heart Association grantin-aid 92-6640, National Institutes of Health Biomedical Research Support grant RR-05425, and Public Health Service grants AI-33119 to J. M. Musser and AI-24332 to R. K. Selander.

We thank Uma Shah for technical assistance and Susan Fogg for assistance in preparation of the manuscript.

REFERENCES

- 1. Bisno, A. L. 1991. Group A streptococcal infections and acute rheumatic fever. N. Engl. J. Med. 325:783-793.
- Bohach, G. A., D. J. Fast, R. D. Nelson, and P. M. Schlievert. 1990. Staphylococcal and streptococcal pyrogenic toxins involved in toxic shock syndrome and related illnesses. Crit. Rev. Microbiol. 17:251-272.
- Brenner, D. J. 1984. Family I. Enterobacteriaceae Rahn 1937, Nom. fam. cons. Opin. 15, Jud. Comm. 1958, 73; Ewing, Farmer, and Brenner 1980, 674; Judicial Commission 1981, 104, p. 408-516. In N. R. Krieg and J. G. Holt (ed.), Bergey's manual of systematic bacteriology, vol. 1. Williams & Wilkins, Baltimore.
- 4. Choi, Y., J. A. Lafferty, J. R. Clements, J. K. Todd, E. W. Gelfand, J. Kappler, P. Marrack, and B. L. Kotzin. 1990. Selective expansion of T cells expressing V β 2 in toxic shock syndrome. J. Exp. Med. 172:981–984.
- Colon-Whitt, A., R. S. Whitt, and R. M. Cole. 1979. Production of an erythrogenic toxin (streptococcal pyrogenic exotoxin) by a nonlysogenized group A streptococcus, p. 64–65. *In* M. T. Parker (ed.), Pathogenic streptococci. Reedbooks Ltd., Chertset, England.
- Congeni, B., C. Rizzo, J. Congeni, and V. V. Sreenivasan. 1987. Outbreak of acute rheumatic fever in northeast Ohio. J. Pediatr. 111:176–179.
- Gaworzewska, E., and G. Colman. 1988. Changes in the pattern of infection caused by *Streptococcus pyogenes*. Epidemiol. Infect. 100:257-269.
- 8. Goshorn, S. C., and P. M. Schlievert. 1988. Nucleotide sequence

of streptococcal pyrogenic exotoxin type C. Infect. Immun. 56:2518-2520.

- 9. Goshorn, S. C., and P. M. Schlievert. 1989. Bacteriophage association of streptococcal pyrogenic exotoxin type C. J. Bacteriol. 171:3068-3073.
- Gray, E. D., L. W. Wannamaker, E. M. Ayoub, A. El Kholy, and Z. H. Abdin. 1981. Cellular immune responses to extracellular streptococcal products in rheumatic heart disease. J. Clin. Invest. 68:665–671.
- 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.
- 12. Higuchi, R. G., and H. Ochman. 1989. Production of singlestranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res. 17:5865.
- Johnson, L. P., P. M. Schlievert, and D. W. Watson. 1980. Transfer of group A streptococcal pyrogenic exotoxin production to nontoxigenic strains by lysogenic conversion. Infect. Immun. 28:254-257.
- 14. Marrack, P., and J. Kappler. 1990. The staphylococcal enterotoxins and their relatives. Science 248:705-711.
- 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.
- 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.
- 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.
- Nida, S. K., and J. J. Ferretti. 1982. Phage influence on the synthesis of extracellular toxins in group A streptococci. Infect. Immun. 36:745-750.
- Quentin, R., A. Goudeau, R. J. Wallace, Jr., A. L. Smith, R. K. Selander, and J. M. Musser. 1990. Urogenital, maternal and neonatal isolates of *Haemophilus influenzae*: identification of unusually virulent serologically non-typable clone families and

evidence for a new *Haemophilus* species. J. Gen. Microbiol. 136:1203-1209.

- Schwartz, B., R. R. Facklam, and R. F. Breiman. 1990. Changing epidemiology of group A streptococcal infection in the USA. Lancet 336:1167–1171.
- 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.
- Selander, R. K., R. M. McKinney, T. S. Whittam, W. F. Bibb, D. J. Brenner, F. S. Nolte, and P. E. Pattison. 1985. Genetic structure of populations of *Legionella pneumophila*. J. Bacteriol. 163:1021–1037.
- Selander, R. K., and J. M. Musser. 1990. Population genetics of bacterial pathogenesis, p. 11-36. *In* B. H. Iglewski and V. L. Clark (ed.), Molecular basis of bacterial pathogenesis. Academic Press, Inc., San Diego, Calif.
- 24. Simpson, W. J., J. M. Musser, and P. P. Cleary. 1992. Evidence consistent with horizontal transfer of the gene (*emm12*) encoding serotype M12 protein between group A and group G pathogenic streptococci. Infect. Immun. 60:1890–1893.
- Stevens, D. L., M. H. Tanner, J. Winship, R. Swarts, K. Ries, P. M. Schlievert, and E. Kaplan. 1989. Severe group A streptococcal infections associated with a toxic shock-like syndrome and scarlet fever toxin A. N. Engl. J. Med. 321:1–7.
- 26. Tomai, M. A., P. M. Schlievert, and M. Kotb. 1992. Distinct T-cell receptor V β gene usage by human T lymphocytes stimulated with the streptococcal pyrogenic exotoxins and pep M5 protein. Infect. Immun. 60:701–705.
- Wayne, L. G., D. J. Brenner, R. R. Colwell, P. A. D. Grimont, O. Kandler, M. I. Krichevsky, L. H. Moore, W. E. C. Moore, R. G. E. Murray, E. Stackebrandt, M. P. Starr, and H. G. Trüper. 1987. Report of the ad hoc committee on reconciliation of approaches to bacterial systematics. Int. J. Syst. Bacteriol. 37:463-464.
- Yu, C.-E., and J. J. Ferretti. 1989. Molecular epidemiologic analysis of the type A streptococcal exotoxin (erythrogenic toxin) gene (*speA*) in clinical *Streptococcus pyogenes* strains. Infect. Immun. 57:3715–3719.
- 29. Yu, C.-E., and J. J. Ferretti. 1991. Frequency of the erythrogenic toxin B and C genes (*speB* and *speC*) among clinical isolates of group A streptococci. Infect. Immun. 59:211-215.