JOURNAL OF CLINICAL MICROBIOLOGY, Apr. 1998, p. 902–907 Vol. 36, No. 4
0095-1137/98/$04.00+0

Group A Streptococcal Vir Types Are M-Protein Gene (*emm*) Sequence Type Specific

DON L. GARDINER,[1]* ALISON M. GOODFELLOW,[1] DIANA R. MARTIN,[2] AND KADABA S. SRIPRAKASH[1]

Menzies School of Health Research, Darwin, Australia,[1] and ESR, Communicable Disease Centre, Porirua, New Zealand[2]

Received 10 September 1997/Returned for modification 16 October 1997/Accepted 14 December 1997

The M-protein genes (*emm* genes) of 103 separate impetiginous *Streptococcus pyogenes* isolates were sequenced and the sequence types were compared to the types obtained by Vir typing. Vir typing is based on restriction fragment length polymorphism (RFLP) analysis of a 4- to 7-kb pathogenicity island encoding *emm* and other virulence genes. By using both *Hae*III and *Hin*fI to generate RFLP profiles, complete concordance between Vir type and *emm* sequence type was found. Comparison of the *emm* sequences with those in GenBank revealed new sequence types sharing less than 90% identity with known types. Diversity in the *emm* sequence was generated by corrected frameshift mutations, point mutations, and small in-frame mutations.

Infections caused by group A streptococci (GAS) lead to a spectrum of disease (4, 5) ranging from relatively benign conditions such as impetigo to more severe invasive diseases and serious nonsuppurative sequelae: acute poststreptococcal glomerulonephritis (APSGN) and acute rheumatic fever (ARF). Despite the recent reemergence of severe invasive diseases in North America and Europe, sequelae following streptococcal infection are now comparatively rare in much of the developed world. Nevertheless, both APSGN and ARF still cause significant morbidity and mortality within the Aboriginal population of northern Australia (6, 24, 36).

Because a significant proportion of isolates of GAS from the Northern Territory are nontypeable by serology, a molecular method called Vir typing (9) was developed. This method is based on the restriction fragment length polymorphisms (RFLPs) of a long PCR product of the Vir regulon of GAS encoding the M-protein family of genes and other virulence factors. More than 400 GAS isolates from skin sores of Australian Aboriginal children have been genotyped by this method, yielding 43 distinct RFLP patterns or Vir types (VTs) (11, 12). A further 193 isolates from four communities involved in a widespread outbreak of APSGN produced 17 distinct VTs (13). However, for a small minority of VTs, the *Hae*III restriction profiles were found to be uninformative, yielding only a single RFLP fragment in addition to the RFLP fragments which were common to the majority of VTs. In order to determine if these VTs could be subdivided, a second restriction enzyme, *Hin*fI, was used in this study.

Selected VTs were also subjected to sequence analysis of the hypervariable region of the *emm* gene in order to determine if the diversity observed in the VT patterns of these isolates was due to architectural diversity of the regulon or to variation in *emm*. Other studies have focused on the sequencing of reference M types (1, 2, 38, 39) which are collections of isolates predominantly from North America and Europe. Unique sequence types (STs) from M-nontypeable (MNT) isolates that are related to reference STs have also been described (18, 29–31). In this study, we examined isolates collected over the past 6 years from the skin of individuals in a small but widespread population from the Northern Territory of Australia. The nucleotide sequences of the hypervariable region of a number of *emm* genes divided the isolates into four categories: (i) isolates with 99 to 100% identity with reference types (group I), variants of reported reference types (group II), isolates with 100% identity with *emm* sequence types previously reported from among MNT isolates from this region (29, 31) (group III), and isolates with 70 to 90% identity with previously reported *emm* STs (group IV). For the isolates in this last group, there was sufficient divergence from previously reported *emm* STs for them to be considered new STs. Point mutations, corrected frameshift mutations, and short in-frame mutations accounted for the majority of the changes in the hypervariable regions of these new STs.

MATERIALS AND METHODS

Vir typing of *Streptococcus pyogenes* isolates. The 103 isolates of GAS selected for VT and *emm* sequence analysis were collected between 1990 and 1996 mainly from pyoderma lesions of children (10, 13). They were processed as described previously by the agarose microtiter tray DNA extraction procedure and long PCR (9, 11). Cycling conditions consisted of an initial denaturation step for 30 s at 95°C, followed by 30 cycles of 94°C for 15 s, 60°C for 60 s, and 68°C for 6 min. Five microliters of the PCR mixture was electrophoresed on a 1% agarose gel to determine the quality of the DNA that was amplified. Vir typing was conducted by digesting approximately 0.5 μg of PCR product (from 8 to 25 μl) with 2 U of *Hae*III or 2 U of *Hin*fI (Pharmacia).

Vir typing with *Hae*III results in very distinct RFLP profiles for the majority of isolates. Certain *Hae*III RFLP patterns were thought to possess a lower information content due to the presence of only a single RFLP fragment, in addition to the RFLP fragments which were common to the majority of VTs. Because it generates a considerable number of bands, *Hin*fI was then used to determine if the designation obtained by typing with *Hae*III could be further subdivided. Previous studies have shown that *Hin*fI digestion, in addition to *Hae*III digestion, of the Vir regulon is as discriminatory as multilocus enzyme electrophoresis with 20 alloenzymes in distinguishing strains of GAS (12).

***emm* sequence analysis.** The template DNA used to determine the hypervariable region of *emm* was either the PCR product from Vir typing as described above or an *emm*-specific amplification product (18, 28). PCR primers were removed by isopropanol precipitation before cycle sequencing with pF (28). To ensure that there was no sampling bias, the isolates of each VT used for sequencing were chosen, whenever possible, from different communities at different time points over the period of the study. Between 2 and 11 isolates of each VT were sequenced.

Sequences were examined by a suite of programs at the Australian National Genomic Information Service. Pairwise comparison of the nucleotide identities for the first 90 to 189 bases of the hypervariable region was conducted after the conserved leader sequence was excluded.

* Corresponding author. Mailing address: Menzies School of Health Research, Combined Health Building, Building 58, Royal Darwin Hospital, Rocklands Dr., Tiwi 0812, Northern Territory, Australia. Phone: 61-8-89228006. Fax: 61-8-89275187. E-mail don@menzies.su.edu.au.

Statistical analysis. The concordance of Vir typing and *emm* sequence typing was determined by the use of a contingency table. For example, two isolates had identical VTs (VT 4) and also had identical STs (NS27). What is the expected number of such identities? To illustrate the statistical analysis, note that 103 isolates had been both Vir typed and sequenced typed. Of these, the first occurrence of an observed concordance between VT 4 and *emm* NS27 serves to identify this pair. This reduces the observed numbers for testing the hypothesis by one, and we calculated the expected number of excess identities (over and above the first) as $1 \times 1/10^2 = 0.01$. For the 16 VTs with an apparently unique ST, there were 53 excess identities compared with 3.42 excess identities expected if the null hypothesis of no association between VT and *emm* ST is correct ($X_1 = 718.8$). For some VTs (e.g., VTs 3, 7, 17, and 29), several STs were identified. By a similar argument, as outlined above, there were 23 excess identities compared to the expected excess of 0.86 ($X_1 = 570.0$). When both of these groups were amalgamated, there were a total of 76 excess identities, compared to the expected excess of 4.38 ($X_1 = 1171.1$) ($P < 0.001$).

Nucleotide sequence accession numbers. Ten unique *emm* sequences (AF018176 to AF018185) were submitted to GenBank.

RESULTS

Improving the discriminatory power of Vir typing by *Hin*fI restriction. *Hae*III restriction of the PCR product derived from the amplification of the Vir regulon of GAS from impetiginous isolates of GAS generally gives rise to between four and eight fragments which vary in size from 200 to 4,000 bp (11). Of these fragments, up to three (1,400, 500, and 275 bp) may correspond to fragments common to the majority of isolates tested. These common fragments do not contribute to the discrimination of RFLP patterns without the presence of other higher-molecular-mass fragments. In the case of VTs 3, 7, and 17, only one other fragment not common to the majority of isolates was present. To improve the discrimination in such cases, another restriction enzyme was used. *Hin*fI could distinguish three subtypes of VT 3, four subtypes of VT 17, and two subtypes of VT 7 (Fig. 1). VTs which could be split by the use of *Hin*fI were designated by their primary *Hae*III VT type followed by a decimal point and their *Hin*fI RFLP type (namely, VT 3.1 and VT 3.2).

As expected, for VTs which had two or more fragments other than the common fragments, *Hin*fI restriction did not appear to provide any further discrimination except for VT 29, in which two *Hin*fI subtypes were found (VT 29.1 and VT 29.2). *Hin*fI has been found to generate high numbers of fragments (9, 12; this study) and is more useful in further discriminating the RFLP patterns obtained by *Hae*III digestion than as the primary restriction enzyme in Vir typing.

5′ *emm* sequence analysis. One hundred three isolates of GAS of 20 VTs obtained by *Hae*III digestion (not including subtypes) were analyzed for the DNA sequence corresponding to the hypervariable region of *emm*. Isolates from all the subtypes of VTs 3, 7, 17, and 29 obtained by *Hin*fI digestion were also included. Twenty-seven distinct N-terminal *emm* STs were found among the 20 VTs. Every VT obtained by *Hae*III digestion other than VTs 3, 7, 17, and 29 had a single corresponding unique ST, whereas every subtype of VTs 3, 7, 17, and 29 had a unique ST.

The concordance between VT and *emm* ST is presented in Table 1. There was a statistically significant concordance between VTs and STs ($P < 0.0001$); each VT represents at least one distinct *emm* ST, and in the case of VT 3, 7, 17, and 29, the VT represents more than one distinct *emm* ST. When VTs 3, 7, 17, and 29 are divided into their subtypes, there is an absolute concordance between the RFLP digests of the Vir regulon and the *emm* ST.

Comparison with known *emm* STs. Twelve of the *emm* sequences corresponded to known *emm* STs (group I, 99 to 100% identity) which have been characterized in North America and Europe (Table 2). However, in the case of the ST

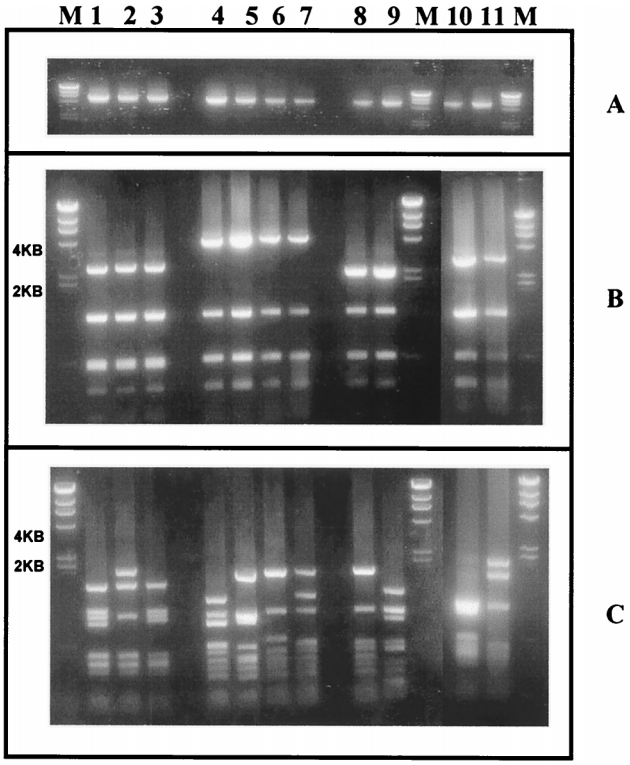

FIG. 1. Ethidium bromide-stained gel of RFLP profiles of VT 3 (lanes 1 to 3), VT 17 (lanes 4 to 7), VT 29 (lanes 8 and 9), and VT 7 (lanes 10 and 11). (A) Undigested PCR product. (B) PCR product digested with *Hae*III. (C) PCR product digested with *Hin*fI. Lane 1, VT 3.1; lane 2, VT 3.2; lane 3, VT 3.3; lane 4, VT 17.1; lane 5, VT 17.2; lane 6, VT 17.3; lane 7, VT 17.4; lane 8, VT 29.1; lane 9, VT 29.2; lane 10, VT 7.1; lane 11, VT 7.2; lane M, bacteriophage λ DNA digested with *Hin*dIII.

*emm*1.4 (25), the original strains used in this study came from either Australia or New Zealand and may represent a geographically local *emm* ST. Three other VTs (VTs 3.3, 10, and 40) had STs that had significant sequence homology to known reference STs and were considered variants of these known STs (group II). One of these variants, *emm*2.3 (accession no. AF018178), has been reported previously (30). In group III, five STs corresponded to previously characterized STs in northern Australia. Finally, seven STs (group IV) showed only 70 to 90% identity with any previously published *emm* sequences. These STs appear to be unique to this geographic location.

Variants of the characterized *emm* types. VT 3.3 represented by a variant of *emm*49, designated *emm*49.2 (accession no. AF018176), shares 88% identity with *emm*49 in the first 112 bases of the hypervariable region. VT 3.3 has three point mutations and one corrected frameshift region of 18 bases (Fig. 2). This is expected to change six amino acid residues in relation to the sequence of *emm*49 in this region. Excluding the frameshift mutation, the level of identity of these two sequences increases to 95%.

In VT 10, which has 90% identity with *emmL*2.1 over 184 bases, there are two point mutations and two insertions (insertions 15 and 9 bases) within the hypervariable region of the *emm* gene (Fig. 2). In the first insertion, the nucleotide sequence CCCTGY has been tandemly repeated three times, while in the second insertion the nucleotide sequence CAAAAT has been repeated twice. In *emm*2.1, only two repeats of CCCTGY were found, and interestingly, pyrimidine

TABLE 1. Concordance between VT and ST

No. of emm STs and VT	ST	No. of Vir isolates typed	No. of isolates sequenced	No. of isolates with identical VTs and STs (no. of excess identities)	Expected no. of excess identities
One *emm* ST					
4	NS27	14	2	2 (1)	0.01
5	BSA29	4	2	2 (1)	0.01
6	BSB19	2	2	2 (1)	0.01
8	57	10	6	6 (5)	0.33
9	25	14	11	11 (10)	1.32
10	2.3	18	3	3 (2)	0.05
11	DRX4	11	2	2 (1)	0.01
12	BL18	11	2	2 (1)	0.01
16	1.4	15	6	6 (5)	0.33
18	55	15	6	6 (5)	0.33
19	74	9	3	3 (2)	0.05
23	NS1	11	4	4 (3)	0.12
26	PL1	10	7	7 (6)	0.47
27	33	5	3	3 (2)	0.05
39	80	9	5	5 (4)	0.21
40	19.2	7	5	5 (4)	0.21
More than one *emm* ST					
3.1	44/66		2	2 (1)	0.01
3.2	11		5	5 (4)	0.22
3.3	49.2		2	2 (1)	0.01
7.1	63		3	3 (2)	0.05
7.2	9		3	3 (2)	0.05
17.1	NS5		2	2 (1)	0.01
17.2	BL12		2	2 (1)	0.01
17.3	Donald		5	5 (4)	0.22
17.4	DRV8		2	2 (1)	0.01
29.1	52		5	5 (4)	0.22
29.2	PK14		3	3 (2)	0.05
Total			103	103 (76)	4.38

transitions were observed in both of these repeats. Due to the significant similarity of the majority of the nucleotide sequence to that of *emmL*2.1 and to prevent confusion with the *enn* gene called *emmL*2.2 (3), this ST has been designated *emm*2.3 (accession no. AF018178). *emm*2.3 also shares 92% identity with the recently described ST 2967 (2).

VT 40 corresponds closely to *emm*19.1 with 93% identity in the first 100 bases, with five point mutations and one deletion of 9 bases corresponding to one tandem repeat (Fig. 2). This sequence has been designated *emm*19.2 (accession no. AF018177).

New STs. VT 5, represented by STBSB29 (accession no. AF018177), resembled M6, but they showed 76% homology in the hypervariable region. Most of the changes seen were due to point mutations (45 in 150 bases), with two insertions and a corrected frameshift. Of the 45 point mutations, 26 were transversions rather than transitional changes. An excess of transversion point mutations was also noted for VT 11 (STDRX4; accession no. AF018181) and VT 17.2 (STBL12; accession no. AF018182) when the sequences of these *emm* STs were compared to those of their closest corresponding reference type (Table 2).

VT 6, represented by STBSB19 (accession no. AF018183), shows 84% homology to *emm*13, excluding a corrected frameshift region spanning 45 nucleotides. STBSB19 also has a 9-base deletion and two 3-base insertions. Another large frameshift region of 33 nucleotides was present in VT 17.4 (STDRV8; accession no. AF018180).

Two new types, STBL18 and STPK14, are part of the *emm*33 and *emm*70 family. The N-terminal *emm* gene sequence for M70 was first described as a local ST, type STBSB75, by Relf et al. (31) and was later identified by Beall and colleagues (1). The *emm*70 ST shares 83% identity with *emm*33 (Fig. 3A), an ST also found among our local isolates (VT 27). STBL18 (accession no. AF018184) from VT 12 has 92% identity with *emm*70 and 85% identity with *emm*33 (Fig. 3A). The similarity, which includes a contiguous stretch of 14 amino acids at the N-proximal region, between the translated sequence of STBL18 and *emm*70 is striking (Fig. 3B). A corrected frameshift spanning 9 bases is responsible for changes in 5 amino acid residues in the sub-N-terminal region.

STPK14 (accession no. AF018185) represented by VT 29.2 (Fig. 3A) also shares identity with *emm*70 (73%), *emm*33 (83%), and STBL18 (77%); however, the translated sequence of STPK14 differs from those of both M33 and M70 (Fig. 3B).

DISCUSSION

The discriminatory power of Vir typing with *Hae*III has been enhanced by the use of an additional restriction enzyme, *Hin*fI. The complete concordance between VTs and STs among these geographically related isolates indicates that each VT profile may represent at least one unique *emm* or *emm* ST among the strains tested.

These findings highlight the significant diversity of strains of GAS within the small, widespread Aboriginal communities of northern Australia because we have found 43 VTs circulating in the community (11, 13; this study) and a further 40 VTs in isolates from hospitalized patients (10). The observation that nearly one-third of Northern Territory isolates sequenced represent new STs differs considerably from a recent reports from the United States where a significant majority of the *emm* STs matched known STs (1, 2). Given the considerable diversity of *emm* STs found within the Aboriginal community, which has the highest prevalence of rheumatic heart disease reported in the world to date (6), ARF vaccines targeted to a few selected M types are unlikely to provide more than limited protection in these communities unless the same epitopes are shared by many members of an *emm* family.

This study supports recent conclusions by Beall and colleagues (2) that epidemiological typing is most meaningful when it is based upon a system reflecting M specificity. A study described in a recent report used an *emm* PCR-RFLP analysis, in which only the 1- to 2-kb *emm* gene is restricted with *Hae*III (35). Outbreak-related strains that were defined by serological M type were compared with the *emm* PCR-RFLP patterns of coexistent strains of the same serotype, and it was found that M5 could be split into five *Hae*III RFLP patterns, M76 could be split into six profiles, and R28 could be split into four distinct profiles. Unfortunately, sequencing of the *emm* gene was not done in that study, and in our experience, serotyping may group together isolates that are genetically distinct (12, 34). Nevertheless, *emm* RFLP analysis may be useful when examining small numbers of outbreak-related isolates in areas where strains of GAS are not endemic. In the endemic situation in which hundreds of isolates are examined, fragments in the range of 1 to 4 kb are helpful for discriminating between the different RFLP patterns obtained with *Hae*III and *Hin*fI. For these purposes, it is more convenient to analyze the RFLP profile of a 4- to 7-kb VT PCR product than a smaller *emm* PCR product.

The Vir regulon of GAS shows structural as well as sequence heterogeneity among isolates (15, 27). In this study, isolates

TABLE 2. Classification of 5′ *emm* ST found among 27 VTs and Vir subtypes

Group and VT	Closest match (accession no.)	% Identity (no. of bases compared)[a]	New accession no.
I			
VT 3.1	*emm*44/61 (U11964)	100 (168)	
VT 3.2	*emm*11 (U11938)	99 (150)	
VT 7.1	*emm*63 (U11982)	100 (189)	
VT 7.2	*emm*9 (U12002)	100 (147)	
VT 8	*emm*57 (X60959)	100 (138)	
VT 9	*emm*25 (U11952)	100 (138)	
VT 16	*emm*1.4 (U20098)	99 (162)	
VT 18	*emm*55 (U11973)	100 (135)	
VT 19	*emm*74 (U11994)	99 (153)	
VT 27	*emm*33 (U11942)	100 (144)	
VT 29.1	*emm*52 (L27098)	100 (132)	
VT 39	*emm*80 (L27097)	100 (135)	
II			
VT 3.3 *emm*49.2	*emm*49 (M31789)	88 (111)	AF018176
VT 10 *emm*2.3	*emm*2.1 (X56398)	90 (184)	AF018178
VT 40 *emm*19.2	*emm*19.1 (U39838)	93 (90)	AF018177
III			
VT 4	ST NS27 (L27094)	100 (150)	
VT 17.1	ST NS5 (L27093)	100 (150)	
VT 17.3	ST Donald (L05017)	100 (129)	
VT 23	ST NS1 (L05022)	100 (144)	
VT 26	ST PL1 (L28822)	100 (120)	
IV			
VT 5 (BSA29)	*emm*6 (M11338)	76 (159)	AF018179
VT 6 (BSB19)	*emm*13 (AF025950)	84 (100)	AF018183
VT 11 (DRX4)	*emm*52 (L27098)	86 (153)	AF018181
VT 12 (BL18)	*emm*70-*emm*33 (L27095-U11942)	92/85 (144)	AF018184
VT 17.2 (BL12)	*emm*71 (L46652)	71 (150)	AF018182
VT 17.4 (DRV8)	*emm*PT2110 (U11957)	75 (132)	AF018180
VT 29.2 (PK14)	*emm*70-*emm*33 (L27095-U11942)	73/83 (102)	AF018185

[a] Number of bases corresponding to the hypervariable region of the mature protein (i.e., conserved leader sequences were not used in the comparison).

were collected from impetiginous lesions, and their Vir regulons did not show significant structural heterogeneity, as evidenced by the remarkable similarities in the sizes of the initial PCR products when compared to the 4- to 7-kb PCR product found when isolates of GAS from the skin and throats of subjects in the northern hemisphere were used (9). Thus, the heterogeneity demonstrated by Vir typing of impetiginous isolates of GAS is due mainly to variations in the *emm* gene rather than diversity in the architecture of the regulon. Since the same ST has not yet been found in different VTs (this study), it is reasonable to hypothesize that *emm* may be evolving faster than other *emm*-like genes in our region. This may lead to the restricted diversity of *enn* in comparison to that of *emm*, as was observed previously (22, 40).

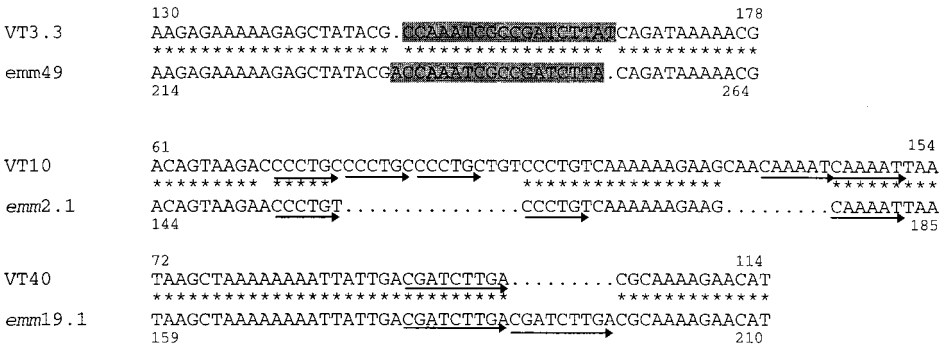

FIG. 2. Alignment of nucleotide sequences of VT 3.3 (accession no. AF018176) and *emm*49 (accession no. M31789), VT 10 (accession no. AF018178) and *emm*L2.1 (accession no. X56398), and VT40 (accession no. AF018177) and *emm*19.1 (accession no. U39838) in the region of in-frame and frameshift insertions or deletions. Asterisks represent identity of the corresponding nucleotide, and dots represent missing nucleotides. Highlighted regions in VT 3.3 and *emm*49 represent the position of the frameshift mutation, while the arrows in VT 10, VT 40, *emm*2.1, and *emm*19.1 represent tandem repeats. The numbers attached to the sequences correspond to the position in the nucleotide sequence in GenBank.

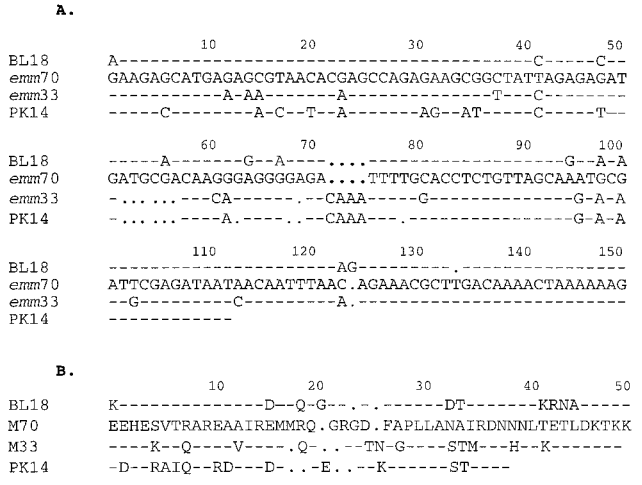

FIG. 3. (A) Aligned nucleotide sequences of STBL18, *emm*70 (STBSB75), *emm*33, and STPK14. The *emm*33 sequence was published by Whatmore et al. (38), while the *emm*70 sequence was first described from the Northern Territory but was identified more recently (1, 31). Hyphens indicate identical nucleotide sequence, and dots represent missing nucleotides. (B) Aligned translated protein sequences of STBL18, *emm*70 (STBL25), *emm*33, and STPK14. Hyphens indicate identical amino acid sequence, and dots represent missing amino acids.

By RFLP analysis, each fragment does not represent an independent locus, because the creation or elimination of a single restriction site will alter two restriction fragments (37). Thus, single base changes could theoretically significantly alter the RFLP profile of any individual VT without altering the 5′ *emm* sequence. However, the 5′ *emm* sequence from every VT sequenced differed significantly. The lack of transitional forms of any ST, i.e., different VTs, with either a single or a few base pair changes, even among STs with the same serological profiles, can be explained if a recent report by Gupta et al. (14) is correct. The host immune response will structure the populations of infectious pathogens into stable collections of independently transmitted strains with nonoverlapping repertoires of dominant polymorphic determinants, despite the effects of recombination. Since the M protein is the principal immunodominant protein of GAS (7, 23), the structures of strains of GAS will be based on the M protein. Previous reports (25) have ascribed a clonal population structure to GAS on the basis of the observation that specific M types are almost exclusively associated with specific multilocus enzyme electrophoretic types. However, other multilocus enzyme electrophoresis data (16) indicate that while there is strong linkage disequilibrium between M types and electrophoretic types, there is no significant linkage disequilibrium between any of the alleles of the housekeeping genes used to produce the electrophoretic profiles, indicating that recombinations are not uncommon and that the apparently "clonal" population structure (25) may be directly related to the host immune response and is not a function of the organism per se.

Antigenic variation due to corrected frameshift mutations (29–31), point mutations (17, 33), and small insertions and deletions (17) have frequently been observed within the hypervariable region of *emm*. Antigenic variation in the streptococcal M protein may also be due to the deletion of repeat blocks (8, 19). Corrected frameshift mutations can result in drastic changes in translated sequences, without significant changes in DNA homology. Thus, single frameshifts may not be reflected by changes in VT. Consequently, very closely related sequence types which show frameshifts may have the same VT but may have different serological M types. Corrected frameshift mutations were first described in the M52/M53/M80 family of isolates from the Northern Territory of Australia (29–31) and in an M5 family isolate, STPL1 (18). In this study we have extended those original observations and found corrected frameshift mutations in *emm*49, *emm*6, *emm*13, *emm*33-*emm*70, and PT2110-family STs. Generally, these STs had only a few point mutations. In contrast, STs that had numerous point mutations compared to the number of point mutations in their closest reference M type rarely exhibited frameshift mutations. Small insertions and/or deletions were noted in STBL18, STPK14, STBSB29, STBL12, and STBSB19, as were previously noted for M1 (17). The short deletions were associated with short repeat sequences within the hypervariable region.

Concordance between VT and *emm* ST is a significant observation because it gives insight into the overall diversity of the virulence locus and allows discrimination of different *emm* STs. Vir typing has clear-cut advantages over randomly amplified polymorphic DNA analysis (9, 12) and M typing with oligonucleotide probes (20, 21, 26, 32). It has recently been proposed that routine *emm* sequencing may be used to designate the M type (1, 2). This is not feasible for outbreaks or for laboratories that do not have the resources to run an automated sequencer. Vir typing has been shown to be a rapid method of sorting large numbers of diverse GAS into distinct genotypes (9–12) and is a method applicable to areas where GAS are endemic and where the majority of isolates are MNT by serotyping.

ACKNOWLEDGMENTS

We thank the district medical officers and nursing staff of the Rural Health Service and the Aboriginal health workers of the six communities for help in obtaining the streptococcal isolates. We thank John Mathews and Wendy Relf for review of our paper, Jenny Powers for critical comments on the statistical methods used, and Elizabeth Wilson for laboratory assistance.

This work was supported by grants from the National Health and Medical Research Council of Australia and the National Heart Foundation.

REFERENCES

1. **Beall, B., R. R. Facklam, and T. Thompson.** 1996. Sequencing emm-specific PCR products for routine and accurate typing of group A streptococci. J. Clin. Microbiol. **34:**953–958.
2. **Beall, B., R. Facklam, T. Hoenes, and B. Schwartz.** 1997. Survey of emm gene sequences and T antigen types from systemic *Streptococcus pyogenes* infection isolates collected in San Francisco, California; Atlanta, Georgia; and Connecticut in 1994 and 1995. J. Clin. Microbiol. **35:**1231–1235.
3. **Bessen, D. E., and V. A. Fischetti.** 1992. Nucleotide sequences of two adjacent M or M-like protein genes of group A streptococci: different RNA transcript levels and identification of a unique immunoglobulin A-binding protein. Infect. Immun. **60:**124–135.
4. **Bisno, A. L.** 1991. Group A streptococcal infections and acute rheumatic fever. N. Engl. J. Med. **325:**783–793.
5. **Breiman, R. F., J. P. Davies, R. R. Facklam, et al.** 1993. Defining the group A streptococcal toxic shock syndrome: rationale and consensus definition. The Working Group on Severe Streptococcal Disease. JAMA **269:**390–391.
6. **Carapetis, J. R., D. R. Wolff, and B. J. Currie.** 1996. Acute rheumatic fever and rheumatic heart disease in the Top End of Australia's Northern Territory. Med. J. Aust. **164:**146–149.
7. **Fischetti, V. A.** 1989. Streptococcal M protein: molecular design and biological behavior. Clin. Microbiol. Rev. **2:**285–314.
8. **Fischetti, V. A., M. Jarymowycz, K. F. Jones, and J. R. Scott.** 1986. Streptococcal M protein size mutants occur at high frequency within a single strain. J. Exp. Med. **164:**971–980.
9. **Gardiner, D., J. Hartas, B. Currie, J. D. Mathews, D. J. Kemp, and K. S. Sriprakash.** 1995. Vir typing: a long-PCR typing method for group A streptococci. PCR Methods Appl. **4:**288–293.
10. **Gardiner, D., J. Hartas, M. Hibble, A. Goodfellow, B. Currie, and K. S. Sriprakash.** 1997. Molecular epidemiology of group A streptococcal infec-

tion in the Northern Territory of Australia. Adv. Exp. Med. Biol. **418:**317–321.
11. **Gardiner, D. L., and K. S. Sriprakash.** 1996. Molecular epidemiology of impetiginous group A streptococci in aboriginal communities of Northern Australia. J. Clin. Microbiol. **34:**1448–1452.
12. **Gardiner, D. L.** 1996. Ph.D. thesis. Molecular epidemiology of *Streptococcus pyogenes*. University of Sydney, Sydney, Australia.
13. **Goodfellow, A. M., and D. L. Gardiner.** 1997. Searching for APSGN-associated *Streptococcus pyogenes* in communities with endemic streptococcal skin infections. Adv. Exp. Med. Biol. **418:**103–108.
14. **Gupta, S., M. C. Maiden, I. M. Feavers, S. Nee, R. M. May, and R. M. Anderson.** 1996. The maintainence of strain structure in populations of recombining infectious agents. Nature Med. **2:**437–442.
15. **Haanes, E. J., D. G. Heath, and P. P. Cleary.** 1992. Architecture of the *vir* regulons of group A streptococci parallels opacity factor phenotype and M protein class. J. Bacteriol. **174:**4967–4976.
16. **Haase, A. M., A. Melder, J. D. Mathews, D. J. Kemp, and M. Adams.** 1994. Clonal diversity of *Streptococcus pyogenes* within some M types reveled by multilocus enzyme electrophoresis. Epidemiol. Infect. **113:**455–462.
17. **Harbaugh, M. P., A. Podbielski, S. Hugl, and P. P. Cleary.** 1993. Nucleotide substitution and small scale insertion produce size and antigenic variation in group A streptococcal M1 protein. Mol. Microbiol. **8:**981–991.
18. **Hartas, J., A. M. Goodfellow, B. Currie, and K. S. Sriprakash.** 1995. Characteristics of group A streptococcal isolates from tropical Australia with high prevalence of rheumatic fever: probing for signature sequences to identify members of the family of serotype 5. Microb. Pathog. **18:**345–354.
19. **Hollingshead, S. K., V. A. Fischetti, and J. R. Scott.** 1987. Size variation in group A streptococcal M protein is generated by homologous recombination between intragenic repeats. Mol. Gen. Genet. **207:**196–203.
20. **Kaufhold, A., A. Podbielski, D. R. Johnson, E. L. Kaplan, and R. Lutticken.** 1992. M protein gene typing of *Streptococcus pyogenes* by nonradioactively labeled oligonucleotide probes. J. Clin. Microbiol. **30:**2391–2397.
21. **Kaufhold, A., A. Podbielski, G. Baumgarten, M. Blokpoel, J. Top, and L. Schouls.** 1994. Rapid typing of group A streptococci by the use of DNA amplification and non-radioactive allele-specific oligonucleotide probes. FEMS Microbiol. Lett. **119:**19–25.
22. **Kehoe, M. A., V. Kapur, A. M. Whatmore, and J. M. Musser.** 1996. Horizontal gene transfer among group A streptococci: implications for pathogenesis and epidemiology. Trends Microbiol. **4:**436–443.
23. **Lancefield, R. C.** 1962. Current knowledge of type-specific M antigens of group A streptococci. J. Immunol. **89:**301–313.
24. **Martin, D. R., and K. S. Sriprakash.** 1996. Epidemiology of group A streptococcal disease in Australia and New Zealand. Recent Adv. Microbiol. **4:**1–40.
25. **Musser, J. M., V. Kapur, J. Szeto, X. Pan, D. S. Swanson, and D. R. Martin.** 1995. 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. **63:**994–1003.
26. **Podbielski, A., O. Kuhnemund, and R. Lutticken.** 1990. Identification of group A type 1 streptococcal M protein gene by non-radioactive oligonucleotide detection method. Med. Microbiol. Immunol. **179:**255–262.
27. **Podbielski, A.** 1993. Three different types of organisation of the vir regulon in group A streptococci. Mol. Gen. Genet. **237:**287–300.
28. **Relf, W. A., and K. S. Sriprakash.** 1990. Limited repertoire of the C-terminal region of the M protein in *Streptococcus pyogenes*. FEMS Microbiol. Lett. **71:**345–350.
29. **Relf, W. A., D. R. Martin, and K. S. Sriprakash.** 1992. Identification of sequence types among M-nontypeable group A streptococci. J. Clin. Microbiol. **30:**3190–3194.
30. **Relf, W. A.** 1993. Ph.D. thesis. Molecular diversity of the M protein gene of *Streptococcus pyogenes*. University of Sydney, Sydney, Australia.
31. **Relf, W. A., D. R. Martin, and K. S. Sriprakash.** 1994. Antigenic diversity within a family of M proteins from group A streptococci: evidence for the role of frameshift and compensatory frameshift mutations. Gene **144:**25–30.
32. **Saunders, N. A., G. Hallas, E. T. Gaworzewska, L. Metherell, A. Efstratiou, J. V. Hookey, and R. C. George.** 1997. PCR–enzyme-linked immunosorbent assay and sequencing as an alternative to serology for M-antigen typing of *Streptococcus pyogenes*. J. Clin. Microbiol. **35:**2689–2691.
33. **Scott, J. R.** 1990. The M protein of group A streptococcus: evolution and regulation, p. 117–203. *In* B. Iglewski (ed.), The bacteria, vol. XI. Academic Press, Inc., New York, N.Y.
34. **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.
35. **Stanley, J., M. Desai, J. Xerry, A. Tanna, A. Efstratiou, and R. George.** 1996. High resolution genotyping elucidates the epidemiology of group A streptococcal outbreaks. J. Infect. Dis. **174:**500–506.
36. **Streeton, C. L., J. N. Hanna, R. D. Messer, and A. Merianos.** 1995. An epidemic of acute poststreptococcal glomerulonephritis among aboriginal children. Paediatr. Child Health **31:**245–248.
37. **Tenover, F. C., R. D. Arbeit, R. V. Goering, P. A. Mickelsen, B. E. Murray, D. H. Persing, and B. Swaminathan.** 1995. Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. J. Clin. Microbiol. **33:**2233–2239.
38. **Whatmore, A. M., V. Kapur, D. J. Sullivan, J. M. Musser, and M. A. Kehoe.** 1994. Noncongruent relationships between variations in *emm* gene sequences and population genetic structure of group A streptococci. Mol. Microbiol. **14:**613–619.
39. **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.
40. **Whatmore, A. M., V. Kapur, J. M. Musser, and M. A. Kehoe.** 1995. 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. **15:**1039–1048.