

Fibronectin-Binding Protein Gene Recombination and Horizontal Transfer between Group A and G Streptococci

Rebecca J. Towers,^{1,2,3} Daniel Gal,^{1,2} David McMillan,⁴ Kadaba S. Sriprakash,⁴
Bart J. Currie,^{1,2} Mark J. Walker,³ Gursharan S. Chhatwal,⁵
and Peter K. Fagan^{1,2,6*}

Menzies School of Health Research,¹ Cooperative Research Centre for Aboriginal Health,⁶ Institute of Advanced Studies,
Charles Darwin University,² Darwin, Northern Territory, Department of Biological Sciences, University of Wollongong,
Wollongong, New South Wales,³ and Queensland Institute of Medical Research, Brisbane, Queensland,⁴
Australia, and German National Research Centre for Biotechnology, Braunschweig, Germany⁵

Received 20 April 2004/Returned for modification 8 June 2004/Accepted 30 July 2004

We report evidence of interspecies gene transfer between the important virulence factor genes *sfbI* and *gfbA*. Because the identified group G streptococcus *gfbA* types possess DNA cassettes that can be identified in a number of group A streptococcus strains, it appears that homologous recombination is occurring between these species.

Group A streptococcus (GAS) is a human-specific pathogen primarily infecting the throat and skin. Group G streptococcus (GGS) is generally regarded as a less important organism, which cohabits with GAS at the same tissue sites. However, some GGS, specifically *Streptococcus dysgalactiae* subsp. *equisimilis*, are capable of causing serious and life-threatening diseases traditionally associated with GAS (1, 4, 9, 20, 28). These diseases include septicemia (10), endocarditis (25), septic arthritis (12), pneumonia, meningitis, pharyngitis, otitis media, acute poststreptococcal glomerulonephritis (22), and cellulitis (7). Although no definitive cases of GGS-associated acute rheumatic fever have been reported, antibodies raised against Australian GGS strains have recently been shown to cross-react with human heart muscle myosin, which has reported to indicate a potential to elicit an autoimmune response that may trigger acute rheumatic fever (13).

Identification of genes encoding C5a peptidase, M protein (*emm*) (5, 24, 26), streptococcal pyrogenic exotoxin G (SpeG) (23), streptolysin O, and streptokinase in GGS is consistent with the observation that these two bacteria have a common disease spectrum. Sequence analysis of these genes from various *S. dysgalactiae* subsp. *equisimilis* isolates suggested that horizontal genetic transfer between GAS and GGS may have contributed to the diversity of these genes. More recently, multilocus sequence typing revealed that horizontal acquisition may also have contributed to variations in selection-neutral housekeeping genes in GGS (16). The multilocus sequence typing data also suggests that the overall direction of genetic flow is predominantly from GAS to GGS. In order to test the generality of horizontal acquisition in contributing to diversity of virulence genes, we chose to investigate *sfbI*, encoding streptococcal fibronectin-binding protein I (SfbI; also known as protein F1) (14, 27).

SfbI is an important and well-characterized virulence factor in GAS that is involved in adherence to throat and skin cells (21, 27) and mediates GAS internalization in epithelial cells (15, 19). A detailed study of gene architecture of *sfbI* from different GAS strains revealed mosaic structures suggesting lateral genetic transfer between strains of GAS and recombination contributing to *sfbI* diversity (29). *sfbI* is situated in a chromosomal locus called the FCT region (named FCT for fibronectin, collagen binding, and T antigen) under the transcriptional regulation of *RofA*, which is also encoded in this region (3).

The *sfbI* homolog *gfbA* has been identified in GGS and is reportedly present in about 30% of GGS strains (17), although only one *gfbA* gene has been sequenced and characterized. The present study showed that approximately 10% of *S. dysgalactiae* subsp. *equisimilis* isolates from humans in the Northern Territory of Australia have the GGS *sfbI* homolog *gfbA* and that this gene is located adjacent to *rofA*, which is also the case in the GAS FCT locus. Comparison of derived *gfbA* sequences with the sequences of known *sfbI* genes revealed mosaicism, with sections of *gfbA* homologous to corresponding sections of different *sfbI* types.

The 38 GGS isolates used in this study were collected between 1995 and 2003 from remote Aboriginal communities in the tropical Top End of the Northern Territory, Australia, or as clinical samples from the Royal Darwin Hospital, Darwin, Northern Territory, Australia. In Aboriginal communities of the Top End, streptococcal infections are endemic (11), and the rates of poststreptococcal sequelae are among the highest in the world (6, 18). These GGS isolates are of special importance, because it has been hypothesized that pharyngeal carriage of GGS may trigger rheumatic fever in this population (13). All bacteria selected for use in this study were large colony-forming beta-hemolytic streptococci that reacted with group G-specific antisera (Oxoid, Basingstoke, England). They were identified as *S. dysgalactiae* subsp. *equisimilis* using commercially available gram-positive identification test cards in a VITEK system

* Corresponding author. Mailing address: Division of Infectious Disease, Menzies School of Health Research, P.O. Box 41096, Casuarina, Northern Territory 0811, Australia. Phone: 61-8-89227793. Fax: 61-8-89275187. E-mail: peterf@menzies.edu.au.

TABLE 1. Primer sequences used to detect *gfbA* and *rofA* in GGS^a

Primer	Sequence
Forward primers	
<i>sfbl</i> -F1	5'-GTCTTTCTTGACAATAACGTGGTAAGCTC-3'
<i>rofA</i> -F	5'-GCCAATAACTGAGGTAGC-3'
Reverse primers	
<i>sfbl</i> -R3	5'-GTATCTTCAACAATGGTCACTGTTTCACTG-3'
<i>sfbl</i> -R6	5'-GCAGCATAGGCTACTTGACCAAAAC-3'
<i>rofA</i> -R	5'-GGTTTGTCTTTTAGGT-3'

^a The primer locations within the genes and PCR strategy were described previously by Towers et al. (29).

(BioMerieux, Durham, N.C.). PCR screening of GGS isolates for *gfbA* and *rofA* and DNA sequence analysis were performed as described previously (29) using the same primer sets (Table 1) and PCR parameters.

The entire *gfbA* gene was present in 4 of the 38 GGS isolates designated 59G, NS3402, NS3404, and NS3404 (Fig. 1). Size variation in the length of *gfbA* (Fig. 1A) was attributed to either differences in the numbers of fibronectin-binding repeats or recombination within the proline-rich domain. All four GGS *gfbA* genes identified in this study were adjacent to the gene encoding the global regulator RofA (Fig. 1B). In fact, the presence of *rofA* was detected in all of the 38 GGS strains tested (results not shown). RofA is a global regulator of GAS (8) found in the FCT region, a locus which exhibits high rates of recombination (3). Therefore, it is possible that an analogous FCT region under the control of RofA is present in all human *S. dysgalactiae* subsp. *equisimilis* strains and that horizontal transfer from GAS could incorporate genes into this recombinatorial hot spot. Further investigation of the downstream genes is necessary to elucidate whether this region of the GGS chromosome is truly homologous with the GAS FCT region. However, finding this regulator adjacent to *gfbA* does suggest an analogous regulatory mechanism influencing bacterial persistence and virulence factor expression in *S. dysgalactiae* subsp. *equisimilis*. Furthermore, other virulence determi-

nant genes that are influenced by RofA in GAS have been detected in human GGS isolates (2, 24, 30).

Partial sequence from the variable 5' end of *gfbA* (Fig. 2A) and the published *gfbA* sequence (accession number U31115) (17) were compared with previously reported *sfbl* sequences from GAS. Overall, the architectural arrangements of predicted GfbA proteins are the same as that of SfbI (Fig. 2). From the N to C terminus, the proteins include a signal sequence, an aromatic amino acid-rich domain (Aro domain), a proline-rich domain comprising a proline-rich region (Pro region) and several proline-rich repeats, an upstream fibronectin-binding domain, a fibronectin-binding repeat region, and an anchor. The C-proximal half beginning with the proline-rich repeats is highly conserved and differs primarily in the number of repeats, but the N-proximal Aro and Pro domains are highly variable. In GAS, 15 Aro domain sequence types (Aro types) and 2 Pro region sequence types (ProA and ProB) were identified.

The five *gfbA* sequences were aligned against the previously described GAS *sfbl* sequences. The alignment revealed four *gfbA* types (ST35 to ST38) with three novel Aro types in GGS (Aro16 to Aro18). Similar to *sfbl* in GAS, the *gfbA* gene in GGS exhibits a mosaic structure within the Aro domain. To better illustrate this structure, the Aro domain has been divided into four main sections (Fig. 2A). These sections represent distinct DNA cassettes with local homology between Aro types, which lie between highly conserved "junction" sequences. For instance, in GGS 1750, the Aro domain sections 1, 2, 3, and 4 may have evolved from GAS Aro5, Aro6, Aro6, and Aro5, respectively. Likewise, sections 1 to 4 in GGS 59G appear to have been derived from GAS Aro1, Aro2 or Aro4, Aro2, and Aro3, respectively (Fig. 2B).

Three of the GGS isolates contain Aro18 *gfbA* genes. Two of the three isolates are clonal (NS3402 and NS3404) as determined by random amplification of polymorphic DNA and pulsed-field gel electrophoretic analysis. NS3572 is not clonal, although the pulsed-field gel electrophoresis fingerprint indicates a limited degree of shared loci (results not shown). This Aro type (Aro18) has evolved from mosaic

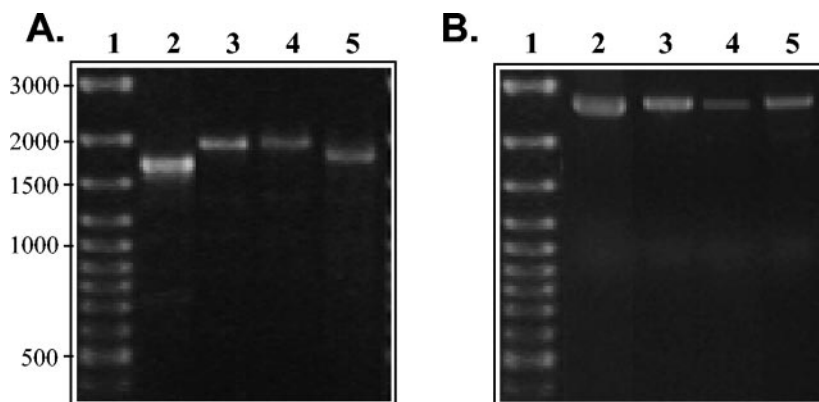
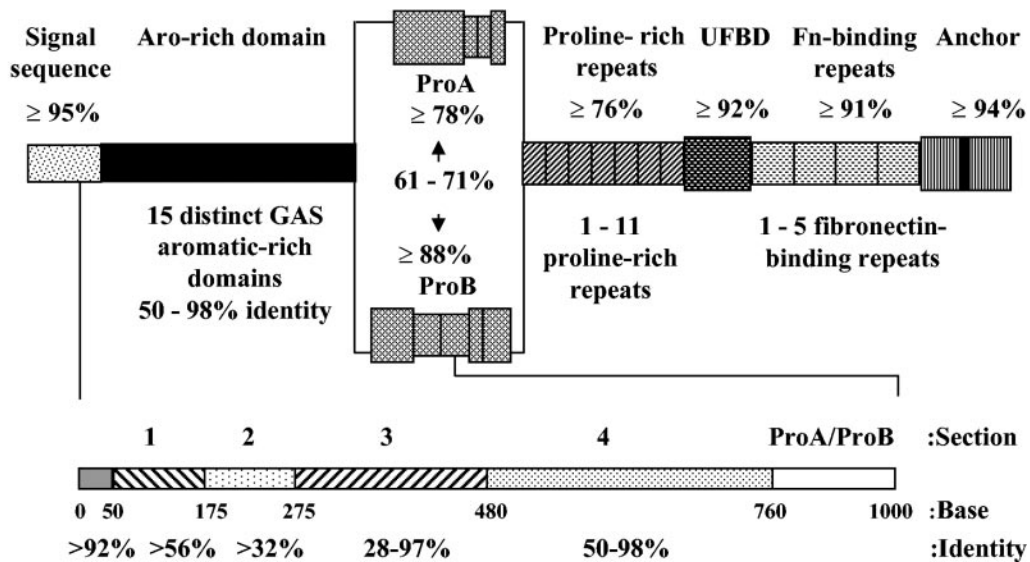


FIG. 1. Ethidium bromide-stained 1% agarose gels showing the results of PCR screening for *gfbA* and *rofA* in *S. dysgalactiae* subsp. *equisimilis* isolates. (A) Screening for full-length *gfbA*. Variations in the amplified product are due to differences in the numbers of fibronectin-binding and proline-rich repeats and the type of proline-rich region present. (B) Screening for the presence of *rofA* adjacent to *gfbA*. Lanes 1 to 5 contain GeneRuler 100-bp ladder (MBI Fermentas), GGS 59G, GGS NS3402, NS3404, and NS3404, respectively.

A.



B.

Aro2 ProA2: GAS strain BL16 (<i>sfbl</i>-ST8) (GAS example)					
	Section 1	Section 2	Section 3	Section 4	Pro-type
Aro 12	81.7%	47.5%	53.9%	57.1%	B1
Aro 17	63.5%	76.2%	95.9%	61.6%	A1
Aro 13	71.2%	45.5%	54.3%	96%	A2
Aro16 ProB1: GGS strain 1750 (<i>sfbl</i>-ST35) (<i>gfbA</i>)					
	Section 1	Section 2	Section 3	Section 4	Pro-type
Aro 5	89.4%	82.2%	77.1%	92.3%	B1, B2
Aro 6	86.5%	92.1%	94.3%	73.5%	B1, B2, B3
Aro17 ProA1: GGS strain 59G (<i>sfbl</i>-ST36)					
	Section 1	Section 2	Section 3	Section 4	Pro-type
Aro 1	93.4%	55.4%	66%	66.2%	A1
Aro 2	63.5%	76.2%	95.9%	61.6%	A1, A2
Aro 3	77.9%	49.5%	59.6%	92.5%	A1
Aro 4	73.1%	76.2%	32.7%	87.5%	A1
Aro18 ProB2: GGS strain NS3402, NS3404 (<i>sfbl</i>-ST37) & NS3572 (<i>sfbl</i>-ST38)					
	Section 1	Section 2	Section 3	Section 4	Pro-type
Aro 4	90%	56%	29.7%	54.9%	A1
Aro 6	70.7%	44%	91%	61.6%	B1, B2 , B3
Aro 8	70.6%	50.5%	71%	91%	A2

FIG. 2. Variation seen in the *sfbl/gfbA* gene family in GAS and GGS. (A) Schematic showing the range of percentage DNA identity for each domain (above) and for each section of the aromatic amino acid-rich domain (below). Gaps introduced to aid alignment have been treated as single base changes. The diagram is not drawn to scale. The lower schematic defines DNA cassettes or sections used to illustrate the gene mosaic structure of *sfbl/gfbA*. UFBD, upstream fibronectin-binding domain; Fn-binding repeats, fibronectin-binding repeats. (B) Analysis demonstrating similarity between DNA cassettes found in *sfbl* and *gfbA*. Sections 1 to 4 correspond to the region encoding the aromatic amino acid-rich domain (described in the text). Section 5 corresponds to the proline-rich region type (ProA/ProB). Sections with the greatest percent identity to the corresponding *sfbl* Aro types in pairwise comparison are shown in bold type. GAS strain BL16 is an example used to demonstrate the analogous mosaic arrangement in GAS which has been previously described (29). Aro16 ProB1 was determined by analyzing the previously characterized *gfbA* from GGS strain 1750 (16).

arrangements. Aro domain sections 1, 3, and 4 have evolved from GAS Aro4, a novel Aro type, Aro6, and Aro8, respectively. Comparing NS3572 to NS3404, there are five base changes in the Aro-rich domain, only two of which are silent. Further differences are apparent in the proline-rich repeats with NS3404 having three nonidentical repeats. In contrast, NS3572 contains a single proline-rich repeat, which is identical to those seen in some GAS strains. These differences can be explained by accumulations of point mutations and duplication or deletion events within the genome. These differences can also be explained, however, by ongoing interspecies lateral transfer rearranging mosaic cassettes. It is not possible from these data to determine whether recombinational diversification is occurring in GAS prior to horizontal transfer or whether it is ongoing within GGS. A majority of the mosaic cassettes have been identified in GAS isolates which cohabit with *S. dysgalactiae* subsp. *equisimilis*. When the GGS aromatic-rich domain gene sequences are aligned against the known GAS aromatic-rich domain gene sequences using pairwise comparisons, the different GGS-derived Aro types group among GAS isolates. In fact, there is very little similarity between the GGS sequences (54.3 to 67.3% in pairwise analysis). This observation, in light of the fact that the GGS sequences show less similarity between themselves than with the homologous GAS sequences, supports the hypothesis that it is unlikely that the differences seen are the result of intragenomic or intergenomic rearrangement with other GGS.

This study is the first demonstration that ongoing lateral transfer leading to genetic diversity may extend to a streptococcal virulence gene with limited strain distribution and that there is ongoing mosaic rearrangement within the *sfbl/gfbA* gene family in GAS or GGS. This horizontal gene transfer and gene rearrangement are of particular significance in light of the fact that the regulatory genes involved in the control of these pathogenicity factors (i.e., *rofA*) have also been identified in *S. dysgalactiae* subsp. *equisimilis*. We believe that taken together, these results indicate that selective pressure may result in the emergence of virulent human pathogenic GGS strains with the potential to cause diseases classically associated with GAS.

Nucleotide sequence accession numbers. All newly determined nucleotide sequences were deposited in GenBank The nucleotide sequences and accession numbers follow: GGS isolate 59G (ST36), AJ605760; GGS isolate NS3402 (ST37), AJ605743; GGS isolate NS3404 (ST37), AJ605744; and GGS isolate NS3572 (ST38), AJ605745. The alignment of the GGS *gfbA* variable regions with published GAS *sfbl* Aro types may be accessed through the EMBL nucleotide database (Align_000635).

This work was supported in part by the Cooperative Research Centre for Aboriginal Health (Australia), the National Health and Medical Research Council of Australia, and the BMBF Network "Pathogenomik" (Germany). P.K.F. is the recipient of an Alexander von Humboldt Foundation Fellowship.

We thank M. McDonald and P. Dasari for their contributions to the MSHR Northern Territory Streptococcal Collection, the source of the strains mentioned in this paper.

REFERENCES

- Barnham, M. 1983. The gut as a source of the haemolytic streptococci causing infection in surgery of the intestinal and biliary tracts. *J. Infect.* **6**:129-139.
- Beckert, S., B. Kreikemeyer, and A. Podbielski. 2001. Group A streptococcal *rofA* gene is involved in the control of several virulence factor genes and eukaryotic cell attachment and internalization. *Infect. Immun.* **69**:534-537.
- Bessen, D. E., and A. Kalia. 2002. Genomic localization of a T serotype locus to a recombinatorial zone encoding extracellular matrix-binding proteins in *Streptococcus pyogenes*. *Infect. Immun.* **70**:1159-1167.
- Christensen, K. K., P. Christensen, L. Flamholz, and T. Ripa. 1974. Frequencies of streptococci of groups A, B, C, D, and G in urethra and cervix swab specimens from patients with suspected gonococcal infection. *Acta Pathol. Microbiol. Scand. Sect. B* **82**:470-474.
- Collins, C. M., A. Kimura, and A. L. Bisno. 1991. Group G streptococcal M protein exhibits structural features analogous to those of class I M protein of group A streptococci. *Infect. Immun.* **60**:3689-3696.
- Currie, B. 1993. Medicine in tropical Australia. *Med. J. Aust.* **158**:609, 612-615.
- Efstratiou, A. 1997. Pyogenic streptococci of Lancefield groups C and G as pathogens in man. *Soc. Appl. Bacteriol. Symp. Ser.* **26**:72S-79S.
- Fogg, G. C., C. M. Gibson, and M. G. Caparon. 1994. The identification of *rofA*, a positive-acting regulatory component of *prtF* expression: use of an m gamma delta-based shuttle mutagenesis strategy in *Streptococcus pyogenes*. *Mol. Microbiol.* **11**:671-684.
- Förner, C. B., and P. D. Ellner. 1979. Distribution of hemolytic streptococci in respiratory specimens. *J. Clin. Microbiol.* **10**:69-71.
- Galloway, A., I. Noel, A. Efstratiou, E. Saint, and D. R. White. 1994. An outbreak of group C streptococcal infection in a maternity unit. *J. Hosp. Infect.* **28**:31-37.
- Gardiner, D. L., and K. S. Sriprakash. 1996. Molecular epidemiology of impetiginous group A streptococcal infections in aboriginal communities of northern Australia. *J. Clin. Microbiol.* **34**:1448-1452.
- Gaunt, P. N., and D. V. Seal. 1987. Group G streptococcal infections. *J. Infect.* **15**:5-20.
- Haidan, A., S. R. Talay, M. Rohde, K. S. Sriprakash, B. J. Currie, and G. S. Chhatwal. 2000. Pharyngeal carriage of group C and group G streptococci and acute rheumatic fever in an Aboriginal population. *Lancet* **356**:1167-1169.
- Hanski, E., and M. Caparon. 1992. Protein F, a fibronectin-binding protein, is an adhesin of the group A streptococcus *Streptococcus pyogenes*. *Proc. Natl. Acad. Sci. USA* **89**:6172-6176.
- Jadoun, J., V. Ozeri, E. Burstein, E. Skutelsky, E. Hanski, and S. Sela. 1998. Protein F1 is required for efficient entry of *Streptococcus pyogenes* into epithelial cells. *J. Infect. Dis.* **178**:147-158.
- Kalia, A., M. C. Enright, B. G. Spratt, and D. E. Bessen. 2001. Directional gene movement from human-pathogenic to commensal-like streptococci. *Infect. Immun.* **69**:4858-4869.
- Kline, J. B., S. Xu, A. L. Bisno, and C. M. Collins. 1996. Identification of a fibronectin-binding protein (GfbA) in pathogenic group G streptococci. *Infect. Immun.* **64**:2122-2129.
- MacDonald, K. T., and A. C. Walker. 1989. Rheumatic heart disease in aboriginal children in the Northern Territory. *Med. J. Aust.* **150**:503-505.
- Molinari, G., S. R. Talay, P. Valentini-Weigand, M. Rohde, and G. S. Chhatwal. 1997. The fibronectin-binding protein of *Streptococcus pyogenes*, SfbI, is involved in the internalization of group A streptococci by epithelial cells. *Infect. Immun.* **65**:1357-1363.
- Nunthapisud, P., S. Sililertpanrana, K. Tatiyakavee, and S. Chumdermpadetsuk. 1990. Occurrence of beta-hemolytic streptococcus group G in school children and sick children. *Southeast Asian J. Trop. Med. Public Health* **21**:215-218.
- Okada, N., A. P. Pentland, P. Falk, and M. G. Caparon. 1994. M protein and protein F act as important determinants of cell-specific tropism of *Streptococcus pyogenes* in skin tissue. *J. Clin. Investig.* **94**:965-977.
- Reid, H. F., D. C. Bassett, E. Gaworzewska, G. Colman, and T. Poon-King. 1990. Streptococcal serotypes newly associated with epidemic post-streptococcal acute glomerulonephritis. *J. Med. Microbiol.* **32**:111-114.
- Sachse, S., P. Seidel, D. Gerlach, E. Gunther, J. Rodel, E. Straube, and K. H. Schmidt. 2002. Superantigen-like gene(s) in human pathogenic *Streptococcus dysgalactiae*, subsp. *equisimilis*: genomic localisation of the gene encoding streptococcal pyrogenic exotoxin G (*speG^{abs}*). *FEMS Immunol. Med. Microbiol.* **34**:159-167.
- Schnitzler, N., A. Podbielski, G. Baumgarten, M. Mignon, and A. Kaufhold. 1995. M or M-like protein gene polymorphisms in human group G streptococci. *J. Clin. Microbiol.* **33**:356-363.
- Smyth, E. G., A. P. Pallett, and R. N. Davidson. 1988. Group G streptococcal endocarditis: two case reports, a review of the literature and recommendations for treatment. *J. Infect.* **16**:169-176.
- Sriprakash, K. S., and J. Hartas. 1996. Lateral genetic transfers between group A and G streptococci for M-like genes are ongoing. *Microb. Pathog.* **20**:275-285.

27. Talay, S. R., P. Valentin-Weigand, P. G. Jerlstrom, K. N. Timmis, and G. S. Chhatwal. 1992. Fibronectin-binding protein of *Streptococcus pyogenes*: sequence of the binding domain involved in adherence of streptococci to epithelial cells. *Infect. Immun.* **60**:3837–3844.
28. Teare, E. L., R. D. Smithson, A. Efstratiou, W. R. Devenish, and N. D. Noah. 1989. An outbreak of puerperal fever caused by group C streptococci. *J. Hosp. Infect.* **13**:337–347.
29. Towers, R. J., P. K. Fagan, S. R. Talay, B. J. Currie, K. S. Sriprakash, M. J. Walker, and G. S. Chhatwal. 2003. Evolution of *sfbI* encoding streptococcal fibronectin-binding protein I: horizontal genetic transfer and gene mosaic structure. *J. Clin. Microbiol.* **41**:5398–5406.
30. Woo, P. C., J. L. Teng, S. K. Lau, P. N. Lum, K. L. Leung, K. L. Wong, K. W. Li, K. C. Lam, and K. Y. Yuen. 2003. Analysis of a viridans group strain reveals a case of bacteremia due to Lancefield group G alpha-hemolytic *Streptococcus dysgalactiae* subsp. *equisimilis* in a patient with pyomyositis and reactive arthritis. *J. Clin. Microbiol.* **41**:613–618.