Two distinct genotypes of prtF2, encoding a fibronectin binding protein, and the evolution of the gene family in Streptococcus pyogenes

V. Ramachandran  
*University of Wollongong,* vidiya@uow.edu.au

Jason D. McArthur  
*University of Wollongong,* jasonm@uow.edu.au

C. E. Behm  
*University of Wollongong*

C. Gutzeit  
*University of Wollongong*

M. Dowton  
*University of Wollongong*

*See next page for additional authors*

Publication Details

This article was originally published as Ramachandran, V et al, Two distinct genotypes of prtF2, encoding a fibronectin binding protein, and the evolution of the gene family in Streptococcus pyogenes, *Journal of Bacteriology,* 186(22), 2004, 7601-7609.
Two distinct genotypes of prtF2, encoding a fibronectin binding protein, and the evolution of the gene family in Streptococcus pyogenes

Abstract
The group A Streptococcus (GAS) is an important pathogen responsible for a wide range of human diseases. Fibronectin binding proteins (FBPs) play an important role in promoting GAS adherence and invasion of host cells. The gene prtF2 encodes a FBP and is contained in approximately 60% of GAS strains. In the present study we have examined 51 prtF2-positive GAS strains isolated from the Northern Territory of Australia and describe two genotypes of prtF2, which are mutually exclusive. Both genotypes have been previously identified in the literature as pfbp and fbaB. We show these genotypes map to the same chromosomal location within the highly recombinatorial fibronectin-collagen-T antigen (FCT) locus, indicating they have arisen from a common ancestor, and in this study these have been designated as the pfbp-type and fbaB-type. Phylogenetic analysis of 7 pfbp-types, 14 fbaB-types and 11 prtF2-negative GAS strains by pulsed-field gel electrophoresis (PFGE) produced 32 distinct PFGE patterns. Interpretation of evolution based on the PFGE dendrogram by parsimony suggests that the pfbp-type is of a recent origin compared to the fbaB type. Comparing multiple DNA sequences of pfbp and fbaB-types reveals a mosaic pattern for the amino terminal region of pfbp-types. The fbaB-type is generally conserved at the amino terminus but varies in the number of fibronectin binding repeats contained within the carboxy terminus. Our data also suggests a possible association of the pfbp-genotype with sof (84.2%) whilst the fbaB genotype was found in a majority of GAS strains negative for sof (90.6%) indicating these two prtF2 subtypes may be under different selective pressures.

Keywords
Streptococcus, pyogenes, fibronectin, CMMB

Disciplines
Life Sciences | Physical Sciences and Mathematics | Social and Behavioral Sciences

Publication Details
This article was originally published as Ramachandran, V et al, Two distinct genotypes of prtF2, encoding a fibronectin binding protein, and the evolution of the gene family in Streptococcus pyogenes, Journal of Bacteriology, 186(22), 2004, 7601-7609.

Authors

This journal article is available at Research Online: http://ro.uow.edu.au/scipapers/98
Two distinct genotypes of *prtF2*, encoding a fibronectin binding protein, and the evolution of the gene family in *Streptococcus pyogenes*

Running Title: Evolution of two distinct *prtF2* genotypes

V. Ramachandran 1, J. D. McArthur 2, C. E. Behm 1, C. Gutzeit 1, M. Dowton 1, P. K. Fagan 3,4, R. Towers 3,4, B. Currie 3,4, K. S. Sriprakash 2,3,4 and M. J. Walker 1*

1 School of Biological Sciences, University of Wollongong, NSW, Australia
2 Queensland Institute of Medical Research, Brisbane, Australia
3 Menzies School of Health Research, Darwin, Australia
4 Institute of Advanced Studies, Charles Darwin University, Darwin, Australia

* Corresponding author

Mailing address: School of Biological Sciences, University of Wollongong, NSW, Australia, 2522
Phone: 0061-2-42213438
Fax: 0061-2-42213400,
E-mail: mwalker@uow.edu.au
Abstract

The group A Streptococcus (GAS) is an important pathogen responsible for a wide range of human diseases. Fibronectin binding proteins (FBPs) play important role in promoting GAS adherence and invasion of host cells. The gene prtF2 encodes a FBP and is contained in approximately 60% of GAS strains. In the present study we have examined 51 prtF2-positive GAS strains isolated from the Northern Territory of Australia and describe two genotypes of prtF2, which are mutually exclusive. Both genotypes have been previously identified in the literature as pfbp and fbaB. We show these genotypes map to the same chromosomal location within the highly recombinatorial fibronectin-collagen-T antigen (FCT) locus, indicating they have arisen from a common ancestor, and in this study these have been designated as the pfbp-type and fbaB-type. Phylogenetic analysis of 7 pfbp-types, 14 fbaB-types and 11 prtF2-negative GAS strains by pulsed-field gel electrophoresis (PFGE) produced 32 distinct PFGE-patterns. Interpretation of evolution based on the PFGE dendrogram by parsimony suggests that the pfbp-type is of a recent origin compared to the fbaB-type. Comparing multiple DNA sequences of pfbp and fbaB-types reveals a mosaic pattern for the amino terminal region of pfbp-types. The fbaB-type is generally conserved at the amino terminus but varies in the number of fibronectin binding repeats contained within the carboxy terminus. Our data also suggests a possible association of the pfbp-genotype with sof (84.2%) whilst the fbaB-genotype was found in a majority of GAS strains negative for sof (90.6%) indicating these two prtF2 subtypes may be under different selective pressures.
Introduction

One adherence mechanism of *Streptococcus pyogenes* (group A streptococcus, GAS) to host cells and tissues is mediated by the interaction with the host ligand, fibronectin. Strains of GAS encode several proteins that have the capacity to bind fibronectin (9, 10, 16, 18, 19, 31, 32, 36, 37). This in itself strongly suggests that the FBP-fibronectin interaction may play an important role in the progression of GAS infection and disease. Whereas many different FBPs in GAS have been described, not all strains are genetically totipotent for each of these FBPs (12, 14, 24, 40). For example, genes encoding the FBPs *sfbI*, *sof* and *prtF2* are present in approximately 52%, 44% and 60% of GAS strains isolated from the Northern Territory (NT) of Australia, respectively (12).

GAS is a human specific pathogen, which can cause a wide range of diseases from benign mucosal and skin infections to life threatening diseases and sequelae such as acute post-streptococcal glomerulonephritis and rheumatic heart disease (11). Diversity in the repertoire of the genes encoding FBPs may have implications for GAS tissue tropism, persistence within the human host and the disease spectrum that these strains can cause. For instance, Neeman et al. (29) have shown an association between *sfbI*-positive GAS strains and persistence after antibiotic treatment. Likewise, an association between *prtF2* and GAS invasive diseases has been observed (12, 37). SfbI, SOF and PrtF2 are
distinct proteins and whilst sfbI and prtF2 are located in the same chromosomal location called the fibronectin-collagen-1 antigen (FCT) locus (5), the sof gene is situated outside this locus.

PrtF2 was originally described by Jaffe et al. (18). Subsequently, Rocha and Fischetti (32) described another FBP designated PFBP. PrtF2 and PFBP have very high sequence identity and possess similar domains. More recently, Terao et al. (37) identified FbaB, a FBP from M3 and M18 GAS serotypes. This protein and PFBP also share the same leader sequence and high sequence similarity in the C proximal region of the protein, which contains the fibronectin binding domains. These observations raise the important question of the evolutionary relationship between these FBP genes.

In order to address this question and further our understanding of the evolution of PrtF2, we have selected 51 prtF2-positive and 11 prtF2-negative genotypically distinct GAS strains. Here we report the characterization of the two distinct genotypes of PrtF2 using PCR and DNA sequence analysis of the strains by pulsed field gel electrophoresis (PFGE) to determine the evolutionary relationship of the prtf2 genotypes. The epidemiological and evolutionary implications of these data are discussed.

Materials and Methods
**Bacterial Strains**

Sixty-two GAS isolates belonging to distinct genotypes as judged by Vir-typing (17) or emm-sequence typing (1) were selected for this study. These strains were isolated from patients in the Northern Territory, Australia and have been described previously (12). In GAS, genomic diversity is predominantly contributed by recombination (13). Thus GAS isolates, from a defined geographical region such as the Northern Territory where the diversity of GAS strains and disease-burden is high, offer an opportunity to discern lineage of a single locus in relation to population structure.

**Screening for genes encoding fibronectin binding proteins**

All GAS strains were screened for genes encoding FBPs namely prtF2, pfbp, fbaB, sfbl, sof, sfbX, FBP54 and fbaA. The prtF2, sfbl, sof and FBP54 status of these strains has been previously described (12). However, the prtF2 polymerase chain reaction (PCR) using primer sets situated within the fibronectin binding repeat domains described by Delvecchio et al. (12) does not differentiate between the two genotypes of prtF2 (pfbp, fbaB). Therefore, 2 sets of PCR reactions were designed and utilized in this study. The first amplification with primer sets VPrtf2-F and VPrtF2-R designed in the signal sequence and cell wall anchor region respectively distinguishes between the two genotypes of prtF2 as well as confirms the mutual exclusiveness of the 2 prtF2 genotypes, and the second PCR amplification with primer sets PFBP-F and ManR4 designed in the flanking region of the prtF2 open reading frame also distinguishes between the 2
genotypes genes as well as confirms the location of these genotypes in the chromosome. Primers SfbXF1 and SfbXR1 were used to screen for sfbX in all strains (19). Primers FbaA-F and FbaA-R were used to screen for fbaA in all strains (Table 1). PCR reactions was carried out in a 50 μl total volume containing 2 μl of DNA template extracted using the QIAGEN DNeasy Tissue Kit, 10 mM Tris-HCl (pH 8.3), 10 mM KCl, 2 mM MgCl2, 50 pmol of each primer, 200 μM of each dNTP and 1 U of Taq DNA.

DNA sequence analysis
Complete nucleotide sequences of 7 pfbp-type and 11 fbaB-type genes were determined. PCR products obtained with primers PFBP-F and ManR4 (Table 1) were used as template after the amplicons were purified using a QIAquick PCR Purification Kit (Qiagen, Australia). The nucleotide sequence of both strands of DNA was determined by primer walking. DNA sequence reactions were performed using the ABI Prism BigDyeTM cycle Sequencing Kit (Applied-Biosystems, Australia) and electrophoresed on an ABI prism 377 DNA sequencer (Perkin-Elmer, CA). Compilation and analysis of DNA sequence data was performed using Auto Assembler software (Perkin-Elmer). Nucleotide sequences of the 18 genes sequenced in this study were deposited into GenBank (NS101, AY612216; NS1140, AY612217; NS125, AY612218; NS178, AY612219; NS179, AY612220; NS192, AY612221; NS195, AY612222; NS210, AY612223; NS235, AY612224; NS240, AY612225; NS265, AY612226; NS436, AY612227; NS506, AY612228; NS53, AY612229; NS564, AY612230; NS581, AY612231; NS730;
Amino acid analysis was performed using programs accessed via the Australian National Genomic Information Services (ANGIS) (www.angis.org.au). Clustaw (38) was used to produce multiple sequence alignments of the pfbp and fbaB sequences determined in this study and previously determined sequences for these genes (the M3 GAS strain SSI-1, AB084272 [37]; the M5 GAS strain Manfredo, http://www.sanger.ac.uk; the M12 GAS strain A735, AF071083 [32]; the M18 GAS strain MGAS8382, AE009964 [33]; the M49 GAS strain 100076, U31980 [18]; the M49 GAS strain B737, AY049089 [5]).

PFGE and phylogenetic analysis
PFGE was carried out using the following modification of the method described by Chatellier et al. (7). Briefly, single colonies of each GAS strain were used to inoculate 2 ml of Todd-Hewitt broth (Difco) supplemented with 1% yeast and grown overnight at 37°C. Cells were harvested by centrifugation, washed twice with TSE buffer (10 mM Tris-Cl, 1.0 M NaCl, 50 mM EDTA [pH 8.0]) and resuspended in 200 μl TE buffer (pH 7.5). An equal volume of pre-warmed 1.5% low melt preparative grade agarose (Bio-Rad, CA) was mixed with the cell suspension, transferred into gel block molds, and allowed to solidify. The blocks were treated for 4 h at 37°C with 400 μl of lysis buffer (6 mM Tris-Cl [pH 7.6], 100 mM EDTA [pH 7.5], 1 M NaCl, 0.5% Brij 58, 0.2% deoxycholate, 0.5% sodium lauroyl sarcosine) containing freshly added lysozyme (1 mg/ml), mutanolysin (100 U/ml) and RNase (20 μg/ml). The lysis buffer was replaced with 300 μl of
deproteination solution (1 μg/ml proteinase K, 1% sodium lauroyl sarcosine and 500 mM EDTA [pH 8.5]) and the blocks were incubated overnight at 50°C. The blocks were prepared for restriction enzyme digestion by washing three times in TE buffer (pH 7.5) for 30 min each wash, the first wash containing 0.5 mM PMSF. Blocks were stored in 1 M EDTA (pH 8.5) at 4°C until use.

Prior to digestion with restriction enzyme, 2-3 mm slices were aseptically cut from the blocks, rinsed for 10 min with sterile distilled water and equilibrated for 30 min in SmaI restriction buffer. The slices were then incubated overnight at 25°C in fresh restriction buffer containing 20 U of SmaI restriction enzyme (Roche). The digested DNA was resolved by PFGE using a CHEF-DRTM electrophoresis cell (BioRad, Australia) employing the following parameters: 0.5 X Tris-borate-EDTA running buffer, 6 V/cm, 2 to 40 s linearly ramped switch times, at 10°C for 23 h. λ DNA concatemers (New England Biolabs, Beverly, MA) were included as a molecular size standard. The gel was then stained with ethidium bromide (1μg/ml) for 30 min and visualized under UV illumination using a GelDoc 1000 image analysis station (BioRad, Australia).

PFGE restriction fragment patterns were visually assessed following the criteria of Tenover et al. (35), and analyzed using GelCompar software (Applied Maths, Kortrijk, Belgium version 4.2). Genetic similarity was compared by clustering methods (unweighted pair group method with arithmetic means) using the Dice coefficient. A tolerance in the band positions of 0.2% was applied for comparison
of fingerprint profiles. MacClade v. 3.08 (27) was used to assign genotypes to ancestral nodes of the PFGE dendogram.

**Results**

*pfbp* and *fbaB* are two distinct mutually exclusive genotypes of *prtF2*

Analysis of GAS chromosomal DNA from 51 *prtF2* GAS strains chosen for this study (Table 2) using the PCR primers VPrtF2-F and VPrtF2-R, reveals amplicons of two size classes; one of approximately 1.9 kb and the other of approximately 3.3 kb (data not shown). None of these 51 strains yielded both amplicon sizes suggesting that these are mutually exclusive. The *prtF2* gene is known to reside within the FCT locus of GAS (5). PCR using primers PFBP-F and ManR4, the known chromosomal position of the *prtF2* open reading frame yielded amplicons of approximately 2.6 kb and 4.0 kb (data not shown), indicating the chromosomal position of *prtF2* is within the FCT locus in the 51-*prtF2* GAS strains examined, adjacent to the flanking *Spy0136* open reading frame (Fig. 1). To confirm that these amplicons are indeed located within the FCT locus, we have sequenced upstream of the *prtF2* open reading frame to confirm the presence of the *msmR* gene (data not shown) and the junction region between *prtF2* and the downstream *Spy0136* gene from a representative group of 18 *prtF2*-positive strains. The generated nucleotide sequence data was aligned with selected FCT junction sequences deposited within GenBank and other databases. The *prtF2*-positive nucleotide sequences are highly conserved across
this junction sequence (>93%) whereas the \textit{prtF2}-negative nucleotide sequences (M1 and M6) only share significant homology from nucleotide 207 of the junction sequence (data not shown).

To further characterize these amplicons, we selected 7 of the larger amplicons (NS179, NS192, NS210, NS240, NS436, NS564 and NS730) and 11 of the shorter amplicons (NS1140, NS101, NS125, NS178, NS195, NS235, NS265, NS506, NS53, NS581 and NS803) and subjected these to complete DNA sequence analysis. The \textit{prtF2} open reading frame encoded within these two amplicon sizes share high nucleotide sequence identity within the N-terminal signal sequence (>98%), part of the upstream fibronectin-binding domain (>98%) and the C-terminal anchor region (>94%). A moderate level of identity (>91%) within individual fibronectin binding repeats was also found; differences in the number of fibronectin binding repeats within this domain occurred (Fig. 2). The central domain exhibited no significant homology between the 2 amplicons. The larger amplicons displayed a high degree of similarity to the published \textit{pfbp} gene sequence (32) and the smaller to the published \textit{fbaB} gene (37).

The central domain of the 11 \textit{fbaB}-like amplicons sequenced show a high degree of similarity (>99%; Fig. 2), however this domain does not share significant similarity with other sequences deposited in GenBank (data not shown). The fibronectin binding repeat domains, as defined by Jaffe et al., (18) of the 11 \textit{fbaB}-like amplicons varied and contained either 2 (NS265), 3 (NS53, NS101, NS125,
NS178, NS195, NS235, NS506, NS581, NS803) or 4 (NS1140) fibronectin binding repeats. By contrast, 3 intact fibronectin binding repeats were found in all 8 *pfbp*-like amplicons sequenced (Fig. 2). The *pfbp*-like central domains displayed >87.1% similarity and contain distinct DNA cassettes flanked by highly conserved junction sequences, indicating that horizontal gene transfer may generate the gene mosaic pattern found within this region (Fig. 3). The nucleotide alignment of the *pfbp*-like sequences can be found at GenBank accession number ALIGN_000736; ftp://ftp.ebi.ac.uk/pub/database/embl/align/ALIGN_000736.dat (data not shown).

Given a) the two mutually exclusive amplicons from the 51 *prtF2*-positive strains were amplified using the same primers sets, b) the mapping of each of these open reading frames to the same position at the end of the FCT region, and c) the downstream sequences across the junction of the 18 open-reading frames analyzed and *Spy0136* are virtually identical, we suggest that *pfbp* and *fbaB* represent two distinct genotypes of *prtF2*. These 2 genotypes will be designated as the *pfbp*-type and the *fbaB*-type henceforth in this study.

**The *pfbp* lineage is more recent than the *fbaB* lineage**

PFGE analysis of thirty two strains comprising of 7 *pfbp*-positive, 14 *fbaB*-positive and 11 *prtF2*-negative GAS strains revealed that they belong to 32 PFGE patterns, consistent with the strains being genetically distinct. Genetic similarity observed by band based clustering in the PFGE profiles, using 0.2% tolerance
was conducted (Fig. 4). The strains were very diverse (<45% similarity) when
diversity was measured by following the criteria of Tenover et al. (35). The
isolates subjected to PFGE belonged to diverse emm sequence types (ST) of the
with the exception of strains NS1122 and NS931, which were of emmST65.
However, despite having the same emmST and a prtF2-negative genotype,
NS1122 and NS931 were found to contain distinct Vir types and genetically
unrelated PFGE fingerprint profiles indicating these strains are genetically
distinct.

We examined the evolution of different prtF2 genotypes (pfbp-type, fbaB-type,
and prtF2-negative) using MacClade (27) and the PFGE dendrogram (Fig. 4).
This program plots the contemporary genotypes onto the tips of the tree, inferring
the evolution of different genotypes by minimizing the total number of changes
over the tree (Fig. 5). When the gain or loss of a genotype was equally weighted
the pattern of evolution shown in Fig. 5A was recovered. This inferred that the
fbaB genotype was the older genotype and that pfbp genotype or the prtF2-
negative genotype evolved from a fbaB-positive progenitor. Strains NS50.1 and
NS235, which are part of a branch predominantly occupied by prtF2-negative
strains, may have horizontally acquired the fbaB gene; inter- and intra-species
horizontal acquisition is common in GAS (5, 20, 21, 34). The analysis also
indicates that fbaB-type can be lost from the genome, as exemplified in NS1045,
and in the ancestral NS1122/NS25 strain and the ancestral NS1096/NS1210
strain. However, in our limited data set, loss of the pfbp-type was not observed.
The \textit{pfbp} genotype may have been acquired in a \textit{fbaB}-background; as exemplified by NS179, NS730, NS436, NS192, and in the ancestor of strains NS564, NS240 and NS210. In this analysis gene gains and losses were equally weighted, and a similar pattern emerges when losses were weighted lower than gains (i.e. when gains were penalized relative to losses).

As an alternative evolutionary scenario to examine the effect of positive selection pressure on the \textit{pfbp} or \textit{fbaB} genotype, a new Maclade analysis of the PFGE dendogram was performed with gene gains weighted higher than losses (2:1 or higher). The resulting evolution pattern is shown in Fig. 5B. Here, the \textit{prtF2}\textit{-negative} genotype was present originally, with the \textit{fbaB} genotype generally evolving first, followed by the \textit{pfbp} genotype evolving in a \textit{fbaB}-type genetic background. Potential horizontal gene acquisition of \textit{fbaB}-type was observed in strains NS50.1, NS235, NS836, NS101 and NS1033, which are part of branches dominated by \textit{prtF2}-negative strains. Another difference between the two evolutionary patterns is the equivocal branch giving rise to strains NS265 and NS195 possessing the \textit{fbaB} genotype and strains NS179 and NS730 possessing the \textit{pfbp} genotype. This could be inferred as NS179 and NS730 evolving from a \textit{fbaB} ancestral genotype or vice versa.

We consider that the analysis shown in Fig. 5A (which is most commonly used) represents the most reasonable working hypothesis for the evolution of different \textit{prtF2} genotypes. Nevertheless, as Fig. 5B reveals, a small change in
the weight given to gene gains and losses impacts on the analysis, such that Fig. 5A should be considered a starting point only. Further, these analyses assume that the topology shown in the 0.2% tolerance PFGE dendrogram (Fig. 4) is correct. Although some branches may be misplaced by this analysis, the above evolutionary scenario represents a reasonable working hypothesis to direct subsequent investigations. Overall the above analysis suggests that the pfbp-type is of more recent origin and arises only in a fbaB-type genetic background.

The distribution of fbaB-type, pfbp-type and other GAS genes encoding fibronectin binding proteins

GAS may possess genes encoding several FBPs, including fbaB, pfbp, sfbl, sof, sfbX, fbaA and FBP54. We have investigated the distribution of sfbl, sof, sfbX, fbaA and fbp54 in the 32 fbaB-type and 19 pfbp-type GAS strains examined in this study. The fbaA and FBP54 genes were found in the majority of GAS strains possessing both pfbp and fbaB genotype examined. Thirteen of 19 (68.4%) pfbp-type GAS strains possessed sfbl, sof, and sfbX. In contrast, only 2 of 32 (6.3%) fbaB-type GAS strains possessed sfbl, sof, and sfbX, while in 21 of 32 (65.6%) fbaB-type GAS strains these genes were not detected (Table 2). GAS strains have been defined as belonging to either class I or class II M types based on epitopes present within the conserved C repeat region of M proteins (2, 3). Class II GAS strains are usually sof-positive while class I GAS strains are generally sof-negative (2, 4, 15). Examination of the sof status of the 51 prtF2-positive GAS strains examined in this study reveals 16/19 (84.2%) of the pfbp-positive strains
possessed sof. By contrast, only 4/32 (13.3%) fbaB-positive strains contained sof suggesting that the pfbp-type is linked to class II GAS and the fbaB-type is linked to class I GAS.

Discussion

Strains causing GAS infections among the Aboriginal communities of northern Australia have demonstrated high diversity and turnover rates (6, 28). There is also little evidence of the emergence of a dominant clone, which has been a common cause of GAS invasive infections elsewhere in the world (8). Similarly, all strains used in this study were found to be genetically distinct as indicated by emm typing, Vir typing and PFGE profiles.

The gene encoding the FBP PrtF2 is situated in the highly recombinatorial FCT region in GAS (5). PrtF2 facilitates the binding of host fibronectin enabling GAS to adhere/internalize host epithelial cells (18, 25). Epidemiological evidence suggests that the presence of this gene may confer upon the pathogen a greater propensity to cause invasive diseases (12). In the present study we have examined 51 prtF2-positive GAS strains isolated from patients in the northern Australia and describe two genotypes of prtF2, which are mutually exclusive. Both genotypes map to the same chromosomal location within the FCT locus, between genes encoding the potential gene regulator msmA and a hypothetical protein Spy0136, and have been designated as the pfbp-type and fbaB-type.
Other researchers have shown expression of PrtF2 to be influenced by the global negative regulator nra, located within the FCT region (30). Recent evidence from Kreikemeyer et al. (25), and the fact nra is not always found in prtF2 positive strains (5), suggests prtF2 expression is controlled by additional regulatory elements. Immediately upstream of prtF2 is a putative transcriptional regulator gene that has been termed msmR. MsmR belongs to the AraC family of regulators (5). Further the GAS genes controlled by this putative regulator have not yet been identified. In all strains studied to date, msmR is present upstream of prtF2. This may suggest msmR is affecting prtF2 expression directly or is regulating other genes that are requirements for PrtF2 function.

The major difference between the pfbp gene sequence (32) and the original prtF2 gene sequence (18) is the presence of 2 single base pair sequence changes in prtF2. These changes result in a frameshift mutation resulting in the loss of the additional 105 amino acids at the N terminus of PFBP, which was not identified in PrtF2. However, the upstream prtF2 gene sequence (GenBank accession number U31980) contains genetic information encoding for part of the 105 amino acid sequence. None of the 18 prtF2 genes sequenced in this study contains a frameshift mutation similar to that reported in Jaffe et al. (18) suggesting that this mutation is infrequent. Clearly however, the reported gene sequences of prtF2 (18), pfbp (32) and fbaB (37) are closely related. We therefore suggest that pfbp and fbaB represent two distinct genotypes of prtF2.
Multiple alignment analysis of the \textit{pfbp}-type DNA sequences has revealed the presence of a mosaic structure within the \textit{pfbp}-type central domain. This mosaic structure is not present in \textit{fbaB}-type sequences suggesting that only the \textit{pfbp}-type gene sequences are capable of the intergenomic exchange required to produce such arrangements. Several surface exposed GAS proteins have been shown to display a mosaic gene structure. These include \textit{emm} (41), \textit{ska} (22) and \textit{sfbI} (39). It is believed that this type of genetic recombination plays a fundamental role in providing a mechanism for evading host immune responses. This is highlighted by the serotype specific protection displayed by anti-M-protein antibodies (26). Such horizontal gene transfer may also produce mosaic gene structures that generate functional diversity of surface exposed proteins. For instance, various M or M-like proteins have been found to bind fibrinogen, Fc-domains of various human immunoglobulins, plasminogen and/or proteins involved in the complement cascade (23).

In this study, we used software that utilised both the PFGE dendrogram and \textit{prtF2} genotype status to infer the evolution of the different genotypes by minimising the total number of changes over the tree. Using this methodology it is suggested that the smaller more conserved \textit{fbaB} genotype is of a more ancient origin than \textit{pfbp}-type. This pattern of inheritance was similar when gains or losses were equally weighted or when gains were weighted higher than losses. As \textit{pfbp}-type only appears in a \textit{fbaB}-type background we hypothesise that \textit{pfbp}-
type arose from an insertional event within \textit{fbaB}-type and the genetic rearrangement seen in the central domain of this gene is a subsequent process. This may also suggest, the acquisition of the \textit{pfbp}-type gene may have some selectable advantage over \textit{fbaB}-type. One possible advantage may be the result of extra functions gained by the protein through the additional \textit{pfbp}-type central domain sequence. Additionally or alternatively, the mosaic structure of the \textit{pfbp}-type central domain may allow for immune evasion to occur.

Interestingly all \textit{pfbp}-type genes sequenced contained 3 repeat regions in the FBRD while \textit{fbaB}-type gene sequences had 2 to 4 repeats. Jaffe et al. (18) localised the UFBD for PrtF2 between amino acids 679 and 783. Residues in the PrtF2 UFBD critical for fibronectin binding were mapped to positions 679 to 717. Homology of the FbaB UFBD region only begins from amino acid 740 of PrtF2. It is not known if this truncated UFBD of FbaB will alter fibronectin binding. Possibly, 3 fibronectin binding repeats are a functional constraint for an intact UFBD in Pfbp. Variations in the number of repeats in the FBRD of FbaB may be due loss of this constraint when the UFBD is truncated.

Thirteen of 19 (68.4\%) \textit{pfbp}-type GAS strains investigated in this study, possessed \textit{sfbI}, \textit{sof}, and \textit{sfbX}. In contrast, only 2 of 32 (6.3\%) \textit{fbaB}-type GAS strains possessed \textit{sfbI}, \textit{sof}, and \textit{sfbX}. Interestingly, 21 of 32 (65.6\%) \textit{fbaB}-type GAS strains lacked these 3 genes. The biological implications of the \textit{pfbp}-type generally being associated with other FBP genes (\textit{sfbI}, \textit{sof} and \textit{sfbX}) and of the
fbAB-type with fewer FBP genes are not known. Previous study has shown linkage association between genes sfbl and sof (14). In this study we report a linkage association exists between pfbp and sof. Sixteen of 19 (84.2%) pfbp-type strains possessed sof. By contrast, only 4 of 32 (13.3%) fbAB-type strains contained sof. Class I GAS strains are generally sof-negative and have been associated with acute rheumatic fever, unlike class II GAS strains which are generally sof-positive and usually associated with skin tropic infections (2, 4, 15). Therefore, our data suggest that pfbp-type maybe linked to class II GAS strains and the fbAB-type linked to class I.

GAS is a highly specific yet extremely versatile pathogen of humans. An increasing number of studies are now revealing the extent of the genetic diversity displayed by this species and are implicating horizontal gene transfer and recombination as major mechanisms influencing the generation of this diversity. Future work, leading to a better understanding of the evolution of GAS and in particular the emergence of highly virulent strains is warranted and may provide new strategies for predicting GAS disease trends and pathogenic mechanisms.

Acknowledgements

This work was supported by the National Health and Medical Research Council (NHMRC) of Australia. We thank Kent Wu for his assistance with the analysis of the pulsed field gel electrophoresis data using the GelCompar software.
References


   Enright, R. Goldstein, D. W. Hood, A. Kalia, C. E. Moore, J. Zhou, and
   B. G. Spratt. 2001. Recombination within natural populations of
   pathogenic bacteria: short-term empirical estimates and long-term

   of fibronectin binding proteins (SfbI and SfbII) of Streptococcus pyogenes
   38:389-392.

   regulons of group A streptococci parallels opacity factor phenotype and M

   protein, is an adhesin of the group A streptococcus Streptococcus

   locus-specific DNA typing method (Vir typing) for Streptococcus pyogenes.

   Protein F2, a novel fibronectin-binding protein from Streptococcus


expression profile, binding characteristics, and impact on eukaryotic cell


Phylogeny and Character Evolution (v. 3.08). Sinauer, Sunderland.

fever: a chink in the chain that links the heart to the throat? The Lancet
Infect. Dis. 4:240-245.

Prevalence of internalisation-associated gene, prtF1, among persisting
group-A streptococcus strains isolated from asymptomatic carriers.

Characterization of nra, a global negative regulator gene in group A

sequence of the serum opacity factor of group A streptococci: identification

32. Rocha, C. L., and V. A. Fischetti. 1999. Identification and
characterization of a novel fibronectin-binding protein on the surface of


**FIGURE LEGENDS**

**Figure 1.** Arrangement of the FCT region. Accession numbers are shown in parentheses: M1, strain SF370 (AE006482/3); M6, strain D471 (U01312, L10919,
AY049087); M12, strains A735 and A374 (AF447492, AY049088, AF071083);
M49, strains CS101/B737 (U49397, AY049089) and 100076 (U31980); M3, strain
MGAS315 (AE014138); M5, strain Manfredo
(http://www.sanger.ac.uk/Projects/S_pyogenes); and M18, strain MGAS8382
(AE009963/4). Open reading frame (ORF) designations (above) are as reported
in the M1 genome. The region is demarcated by the two highly conserved ORFs,
Spy0123 and Spy0136 (black arrows). With the exception of sfbl, which is used
instead of prtF1, all other designations are as reported in Podbielski et al. (30)
and Bessen and Kalia (5). Dotted lines represent gaps introduced to aid
alignment. ORFs with no significant homology are left unshaded. Chevron within
nra18 indicates a stop codon caused by a single point mutation. Similarly, the
single ORF in M49 spanning eftLSL.A and eftLSL.B is caused by a single point
mutation resulting in the substitution of a stop codon (TAA) with a Glu (Q) codon
(CAA).

**Figure 2.** Comparison of pfbp-type and fbaB-type variants. Values indicate the
range of identities observed in pairwise comparisons of DNA sequences both
within and between variants. The two mature proteins differ significantly in their
central domain but are highly similar over all other domains. The central domain
is highly conserved in the fbaB-type but considerable variation is seen in the
pfbp-type central domain. The pfbp-type variants always possess three non-
identical fibronectin-binding repeats in the FBRD while the fbaB-type variants
have variable numbers of repeats due to duplication or deletion of repeat 2
Overall homology of the FBRD of *fbaB*-type could not be determined due to the resultant gaps in the alignment, however the individual repeats are highly homologous both within and between variants.

**Figure 3.** Parts of the *pfbp*-type encoding regions of the amino-terminus have been horizontally transferred between different *pfbp*-type genes producing mosaic alleles. Throughout the sequence highly conserved junction regions were identified (shaded, <93.1% identity in pairwise comparisons) flanking distinct sections or cassettes, which appear to have been exchanged between alleles resulting in a mosaic structure. Regions with identity to the *pfbp*-type gene of the M12 strain A735 (*prtF2.12*) have been underlined to illustrate the mobility of individual DNA cassettes.

**Figure 4.** Dendrogram generated by the GelCompar software showing the genetic relationship between 21 *S. pyogenes* isolates possessing *prtF2* genotypes and 11 isolates not possessing the gene. The dendogram was constructed with the use of UPGMA cluster analysis of the PFGE patterns obtained after macro restriction with the *Sma*I enzyme. A tolerance in the band positions of 0.2% was applied for comparison of fingerprint profiles. PFGE fingerprint patterns are depicted next to the corresponding branches of the dendogram. The *prtF2* genotypes and *emm*ST are indicated after the strain number. P, *pfbp*-type; F, *fbaB*-type; N, absence of *prtF2*. The scale indicates percent similarity.
Figure 5. The evolution of *prtF2* genotypes, inferred by parsimony using MacClade v.3.08 (27) based on the 0.2% PFGE dendogram. Evolution pattern recovered when (A) gain or loss of a genotype were equally weighted (B) gains were weighted higher than losses of a genotype. *pfbp*-type (solid black); *fbaB*-type (stippled); equivocal (vertical stripe); neither genotype (unshaded).
Fig 1
no significant homology

>98.1%  >87.1%  >98.8%  >91.3%  >96.4%  >96.9%  >95.8%

signal sequence  central domain  UFBD  FBRD  carboxy-terminus

>98.1%  >98.3%  >98.3%  >96.1%

>99.4%  >98.3%  >99.3%  >94.5%

>98.7%  >91.3%  >94.6%  >95.3%

fbaB-type

pfbp-type

>99.7%  >99.4%  >98.3%  >92.4%  >93.7%  >95.3%  >96.4%

>93.7%  >95.3%  >94.5%  >96.4%

no significant homology

Fig 2
Fig 3
Fig 4
Fig 5