Towards a genotyping system for *Streptococcus agalactiae* (group B streptococcus): use of mobile genetic elements in Australasian invasive isolates

Fanrong Kong,[1] Diana Martin,[2] Gregory James[1] and Gwendolyn L. Gilbert[1]

Correspondence
Gwendolyn L. Gilbert
lyng@icpmr.wsahs.nsw.gov.au

[1]Centre for Infectious Diseases and Microbiology Laboratory Services, Institute of Clinical Pathology and Medical Research, Westmead Hospital, Darcy Rd, Westmead, New South Wales, 2145 Australia

[2]Institute of Environmental Science and Research (ESR), Porirua, Wellington, New Zealand

This study forms part of the development of an integrated genotyping system for *Streptococcus agalactiae* (group B streptococcus, GBS) that can be used to study the population genetics of the organism and the pathogenesis and epidemiology of GBS disease. In recent previous studies, two sets of markers, the capsular polysaccharide synthesis (*cps*) gene cluster and surface protein antigen genes, have been used to assign molecular serotypes (MS) and protein-gene profiles (PGP) to more than 200 isolates. In the present study, five mobile genetic elements (MGE) have been used as a third set of markers, to characterize further 194 invasive isolates, recovered from blood or cerebrospinal fluid (CSF). Of these, 97 % contained one or more of the five MGE, the distribution of which was related to MS and PGP, as illustrated by MS III, which is divisible into four serosubtypes with different combinations of the MGE (or none). Fifty-six different genotypes and eight genetic clusters were identified, each with different combinations of the three sets of molecular markers. Five predominant genotypes (Ia-1, Ib-1, III-1, III-2 and V-1) contained 62 % of the isolates and five of the eight genetic clusters contained 92 % of the isolates. The 17 CSF isolates were relatively widely distributed between 10 genotypes and across seven of the eight clusters. Further study is needed to determine whether these genotypes or clusters share common markers of increased virulence. In future, comparison of invasive with colonizing strains of GBS may elucidate the significance of these findings.

INTRODUCTION

Streptococcus agalactiae or group B streptococcus (GBS) is the commonest cause of neonatal sepsis and an increasingly important cause of infection in elderly patients with chronic medical conditions (Schuchat, 1998). However, it is also frequently carried in the normal faecal and/or vaginal flora. Integrated studies of strain virulence and the epidemiology and pathogenesis of infection are needed to identify strains that are most likely to cause disease (van Belkum *et al.*, 2001). Our aim is to develop a GBS genotyping system with enough discriminatory ability to be used for such studies.

Nine GBS serotypes are defined by differences in capsular polysaccharide antigens (Chaffin *et al.*, 2000), which are important virulence factors, encoded by the *cps* gene clusters. A family of variable surface protein antigens, including Rib, C alpha, C alpha-like and the IgA-binding protein, C beta, also contribute to virulence (Lachenauer *et al.*, 2000; Mawn *et al.*,

1993; Wastfelt *et al.*, 1996). Numerous mobile genetic elements (MGE), including the four insertion sequences (IS) and a group II intron selected for use in this study, have been identified in GBS (Tamura *et al.*, 2000; Rubens *et al.*, 1989; Granlund *et al.*, 1998, 2001; Spellerberg *et al.*, 2000; Franken *et al.*, 2001; Takahashi *et al.*, 2002; Tettelin *et al.*, 2002; Glaser *et al.*, 2002). Their presence and the number of copies can affect gene expression and virulence (Mahillon & Chandler, 1998; Mahillon *et al.*, 1999; Martinez-Abarca & Toro, 2000) and they may be involved in pathogenicity island-like structures in GBS (Glaser *et al.*, 2002; Tettelin *et al.*, 2002). They are also potentially useful epidemiological markers (Tamura *et al.*, 2000).

In our previous studies, we developed methods, based on PCR and sequencing, to identify serotypes and to define surface protein-gene profiles (PGP) using a panel of 224 GBS isolates and compared the results with those of conventional serotyping (Kong *et al.*, 2002a, b). Molecular and conventional serotyping results correlated well and, in addition, a molecular serotype (MS) was assigned to isolates that were non-typable using antisera (Kong *et al.*, 2002a, b). We

Abbreviations: CSF, cerebrospinal fluid; GBS, group B streptococcus; MGE, mobile genetic element; MS, molecular serotype; PGP, protein-gene profile; SST, serosubtype.

confirmed previously reported associations between capsular polysaccharide and surface protein antigens (Lachenauer *et al.*, 2000; Wastfelt *et al.*, 1996) and divided our 224 GBS isolates into more than 50 molecular types corresponding with different antigen combinations (and antigenic variants) (Kong *et al.*, 2002a, b).

In this study, we used five selected MGE as a third set of molecular markers, to increase the discriminatory ability of our genotyping system further (Kong *et al.*, 2002a, b). We then used this typing system to examine the population genetic structure and age-related disease distribution of genotypes among 194 invasive GBS isolates.

METHODS

Invasive GBS clinical isolates. All 194 isolates used in this study were recovered from the blood (177) or cerebrospinal fluid (CSF) (17) of 191 patients (107 female, 80 male, four sex unrecorded; three cultures each contained mixed growth of two GBS serotypes). One hundred and eight isolates were from specimens submitted for culture to the Centre for Infectious Diseases and Microbiology Laboratory Services, ICPMR, Sydney, Australia, during 1996–2001 and 83 were referred to the Institute of Environmental Science and Research, Porirua, Wellington, New Zealand, for serotyping from various diagnostic laboratories in New Zealand during 1994–2000.

Isolates and methods used to identify MS, serosubtypes (SST) and surface PGP have been described previously in detail (Kong *et al.*, 2002a, b). Patients were classified into age groups for analysis of genotype distribution as follows: neonatal, early onset (0–6 days); neonatal, late onset (7 days to 3 months); infant and child (4 months to 14 years); young adult (15–45 years); middle-aged (46–60 years); and elderly (> 60 years).

Oligonucleotide primers, DNA preparations and PCR. The target sites, sequences, melting temperatures and numbered base positions of the oligonucleotide primers used in this study are shown in Table 1. Expected amplicon lengths of different primer pairs can be calculated from the 5′-end positions of the corresponding primers. Primers were designed, specifically for this study, using published sequences of the relevant MGE. DNA extraction from GBS cultures and PCR were performed as described previously (Kong *et al.*, 2002a, b).

Sequencing. To evaluate the specificity of our primer pairs, we sequenced selected amplicons produced by primers IS1381S1/IS1381A (two isolates), IS861S/IS861A2 (10 isolates), IS1548S/IS1548A3 (21 isolates), ISSa4S/ISSa4A2 (nine isolates) and GBSi1S1/GBSi1A2 (three isolates). For sequencing, 40 µl aliquots of PCR products were further purified by the PEG precipitation method (Ahmet *et al.*, 1999). PCR products were sequenced using Applied Biosystems *Taq* DyeDeoxy terminator cycle-sequencing kits according to standard protocols. The corresponding amplification primers or inner primers were used as the sequencing primers.

Multiple sequence alignments and sequence comparison. Multiple sequence alignments were performed with the PILEUP and PRETTY programs in the Multiple Sequence Analysis program group. Sequences were compared using the BESTFIT program in the Comparison program group. All programs are provided in WebANGIS, version 3 (Australian National Genomic Information Service).

Nucleotide sequence accession numbers. The insertion sequence IS*1381*, sequenced in our previous study (Kong *et al.*, 2002b), has appeared in GenBank with accession number AF367974 (IS*1381* in partial C beta antigen gene). The five MGE previously sequenced by others have appeared in GenBank under accession numbers AF064785 (IS*1381*), M22449 (IS*861*), Y14270 (IS*1548*), AF165983 (ISSa*4*) and AJ292930 (GBSi1). Two recently released genome sequences (Glaser *et al.*, 2002; Tettelin *et al.*, 2002) have appeared in GenBank with the following accession numbers: NC_004116 (*S. agalactiae* 2603V/R,

Table 1. Oligonucleotide primers used in this study

Primer T_m values were provided by the manufacturer (Sigma-Aldrich). Numbers at the ends of the sequences represent the positions at which primer sequences start and finish (numbering start point 1 refers to the start point of the corresponding GenBank accession).

Primer	Target	T_m (°C)	Accession no.	Sequence
IS861S	IS*861*	77·4	M22449	[445]GAGAAAACAAGAGGGAGACCGAGTAAAATGGGACG[479]
IS861A1	IS*861*	77·3	M22449	[831]CACGATTTCGCAGTTCTAAATAAATCCGACGATAGCC[795]
IS861A2	IS*861*	76·1	M22449	[1020]CAAACTCCGTCACATCGGTATAGCACTTCTCATAGG[985]
IS1548S	IS*1548*	76·5	Y14270	[143]CTATTGATGATTGCGCAGTTGAATTGGATAGTCGTC[178]
IS1548S1	IS*1548*	77·0	Y14270	[539]GTTTGGGACAGGTAGCGGTTGAGGAGAAAAGTAATG[574]
IS1548A1	IS*1548*	77·0	Y14270	[574]CATTACTTTTCTCCTCAACCGCTACCTGTCCCAAAC[539]
IS1548A2	IS*1548*	70·3	Y14270	[915]CCCAATACCACGTAACTTATGCCATTTG[888]
IS1548A3	IS*1548*	78·0	Y14270	[930]CGTGTTACGAGTCATCCCAATACCACGTAACTTATGCC[893]
IS1381S1	IS*1381*	80·1	AF064785/AF367974	[272/818]CTTATGAACAAATTGCGGCTGATTTTGGCATTCACG[307/853]
IS1381S2	IS*1381*	81·7	AF064785/AF367974	[497/1040]GGCTCAGGCGATTGTCACAAGCCAAGGGAG[526/1069]
IS1381A	IS*1381*	73·1	AF064785/AF367974	[881/1424]CTAAAATCCTAGTTCACGGTTGATCATTCCAGC[849/1392]
ISSa4S	ISSa*4*	78·5	AF165983	[326]CGTATCTGTCACTTATTTCCCTGCGGGTGTCTCC[359]
ISSa4A1	ISSa*4*	75·2	AF165983	[639]GCCGATGTCACAACATAGTTCAGGATATAGCCAG[606]
ISSa4A2	ISSa*4*	74·5	AF165983	[780]CGTAAAGGAGTCCAAAGATGATAGCCTTTTTGAACC[745]
GBSi1S1	GBSi1	78·6	AJ292930	[721]CATCTCGGAACAATATGCTCGAAGCTTACAAGCAAGTG[758]
GBSi1S2	GBSi1	77·3	AJ292930	[789]GGGGTCACTATCGAGCAGATGGATGACTATCTTCAC[824]
GBSi1A1	GBSi1	83·9	AJ292930	[1058]AATGGCTGTTTCGCAGGAGCGATTGGGTCTGAACC[1024]
GBSi1A2	GBSi1	80·5	AJ292930	[1161]CCAGGGACATCAATCTGTCTTGCGGAACAGTATCG[1127]

complete genome), NC_004368 (*S. agalactiae* NEM316, complete genome).

Definitions of MS, SST and PGP. Our previous study (Kong *et al.*, 2002a) defined MS Ia, Ib and II–VIII (as for conventional serotypes) and four SST of MS III, based on heterogeneity within several regions of the *cps* gene cluster. A second study (Kong *et al.*, 2002b) defined PGP based on the presence of genes encoding Rib, C alpha, C alpha-like, C beta proteins as shown in Table 2.

Statistical analysis and dendrogram. SSPS version 11 software was used for statistic analysis. A dendrogram was formed using Average Linkage (between groups) and Hierarchical Cluster Analysis in SSPS version 11 software. The presence or absence of each marker, MS Ia, Ib, II and IV–VI, SST III-1–III-4; PGP 'A', 'a', 'as', 'R', 'alp2' and 'alp3'; *bac* subgroups 1, 1a, 2, 3, 3a, 3b, 3c, 4, 4b, 5a, 7, 7a, 8, 9, 9a, 10, N1 and N2; and MGE IS*1381*, IS*861*, IS*1548*, ISSa*4* and GBSil, was included in the analysis. The genotypes were each characterized by a distinct combination of the MS or SST, PGP and MGE.

RESULTS

Specificity of primer pairs

The sequencing results showed that 10 IS*861*, 21 IS*1548*, nine ISSa*4* and three GBSi1 amplicon sequences were either identical with their corresponding sequences in GenBank or had very minor mutations (one to three sites).

In our previous study, we found a full-length IS*1381* (AF367974) within the C beta antigen gene (*bac*) of one clinical isolate. Our sequence showed several differences compared with the original published sequence (AF064785): the terminal inverted repeats contained 17 rather than 20 bp and, between positions 419 and 429, there was a 3 bp deletion and four individual differences as follows: **GG**GA**TCC**GA**TT** (AF064785) vs **CA**GA --- **GG**T**A** (AF367974; our sequence). The sequence of the amplicons of primer pair IS1381S1/IS1381A, from two selected clinical isolates, were identical to that of our sequence (AF367974) and to the IS*1381* sequences in the GBS (2603V/R) genome (NC_004116), but different, as above, from the original reported IS*1381* sequence (AF064785).

In addition to sequencing, we evaluated the specificity of our

Table 2. Explanation of PGP codes

Code	Protein gene present
A	5′ end of *bca*
a	*bca* repetitive unit or *bca* repetitive unit-like sequence with multiple amplicons
as	*bca* repetitive unit or *bca* repetitive unit-like sequence with single amplicon
R	*rib*
alp2	*alp2*
alp3	*alp3*
B (> 20 subgroups)	*bac* (subgroups based on sequence heterogeneity)

primer pairs by comparing PCR results for two or more primer pairs for each target (Table 1). In all cases, the same sets of isolates produced amplicons of the expected length when tested with different primers targeting the same MGE, thus confirming the specificity of the primer pairs.

Distribution of MGE

The numbers of isolates containing different MGE combinations (from none to four per isolate) are shown in Table 3. IS*1381*, IS*861*, IS*1548*, ISSa*4* and GBSi1 were respectively identified in 87, 52, 17, 6 and 18 % of isolates. Six (3 %) isolates contained none of the five selected MGE.

Predominant relationships between MS/SST, PGP and MGE

Fig. 1 shows the relationships between the various genetic markers. IS*1381* was present in nearly all isolates of MS Ia, Ib, IV, V and VI, but in none of SST III-2 or III-3. IS*1548* and GBSil were found most commonly in serotypes II or III; three isolates (all MS II) contained both GBSi1 and IS*1548*. IS*861* was found in all SST III-1 and III-2 and most MS II and Ib isolates but in only 14 % of other MS isolates. ISSa*4* was present in only 6 % of isolates, more than half of which were MS II; it was present in one invasive isolate obtained before 1996 (in 1994). IS*1381* was found in most isolates except those in cluster 8, PGP 'alp2as', which had none of the four insertion sequences (Glaser *et al.*, 2002). IS*861* was found in most genotypes with PGP 'AaB' (clusters 3 and 4) and all genotypes with PGP 'R' (clusters 6 and 7).

Genotypes based on MS/SST, PGP, *bac* subtypes and MGE

MS/SST, PGP, *bac* subtype (for isolates with PGP 'B') and the presence of various combinations of MGE provide a PCR/sequencing-based genotyping system. The 194 invasive isolates in this study represented seven serotypes, ten MS/SST, 41 subtypes based on the distributions of PGP and MGE or 56 genotypes when *bac* subtypes (mainly in MS Ib) were included (Fig. 1).

Theoretical GBS genetic population structure

Theoretically, there are 13 possible GBS MS/SST (eight MS, Ia, Ib, II, IV–VIII; four SST, III-1–III-4; and *cps* gene cluster absent) and at least 10 PGP (none, 'Aa', 'AaB', 'a', 'as', 'R', 'RB', 'alp2as', 'alp3' or 'alp4a'). If the 22 *bac* subgroups identified so far are included, there are up to 31 PGP. If the five MGE were independently, randomly distributed and present or absent, there would be $13 \times 31 \times 2^5$ or 12 896 different possible combinations of molecular markers. The fact that only 56 different combinations were found (Fig. 1) demonstrates that markers are not distributed randomly or, in other words, these invasive Australasian GBS isolates have a clonal population structure. It is possible, but unlikely, that these isolates represent a very limited number of GBS genotypes.

Table 3. Distribution and combinations of MGE among 194 invasive GBS isolates

Isolates (n)	IS1381	IS861	IS1548	ISSa4	GBSi1	None
6	−	−	−	−	−	+
78	+	−	−	−	−	−
2	−	−	−	−	+	−
37	+	+	−	−	−	−
1	+	−	+	−	−	−
3	+	−	−	+	−	−
29	+	+	+	−	−	−
6	+	+	−	+	−	−
8	+	+	−	−	+	−
18	−	+	−	−	+	−
1	+	−	−	−	+	−
1	+	−	+	−	+	−
2	+	+	+	−	+	−
2	+	−	−	+	+	−
Total (n = 194)	168 (87 %)	100 (52 %)	33 (17 %)	11 (6 %)	34 (18 %)	6 (3 %)

Phylogenetic relationship of Australasian invasive GBS

The 56 genotypes formed eight clusters, if they were separated at a genetic distance of about 16 (or three cluster groups separated at a distance of ∼22·5). The PGP was the main determinant of cluster separation (Fig. 1). Ninety-four per cent of isolates belonged to five MS (Ia, Ib, II, III and V), 62 % belonged to five (9 %) genotypes (Ia-1, Ib-1, III-1, III-2, V-1) and 92 % belonged to the five largest clusters (1, 2, 4, 6 and 7). Cluster group A, the largest, contained 139 (72 %) isolates and 48 (86 %) genotypes, 45 of which contained fewer than five isolates, whereas cluster group B contained 49 (25 %) isolates and five (9 %) genotypes.

The main characteristics of each cluster were as follows. Cluster 1: 'alp3', IS1381 (39 isolates, four MS, 11 genotypes; predominant genotype V-1). Cluster 2: 'a' or 'as', IS1381 (55 isolates, four MS, 12 genotypes, predominant genotype Ia-1). Cluster 3: 'Aa' or 'AaB', MS II, IS1381, IS 861 (10 isolates, six genotypes). Cluster 4: 'AaB', IS1381, IS861 (35 isolates, two MS: VI or Ib; 18 genotypes; predominant genotype Ib-1). Cluster 5: 'AaB', IS861, GBSi1, genotype III-4-1 (one isolate). Cluster 6: 'R', IS861 and GBSil (22 isolates, three MS/genotypes; predominant genotype III-2). Cluster 7: 'R', IS1381 and IS861 (27 isolates; two MS/genotypes; predominant genotype III-1). Cluster 8: 'alp2as', none of the four selected IS (six isolates; three MS/genotypes; one contained GBSi1).

The relationship between genotypes and GBS disease patterns

The distribution of MS and genotypes in different age groups of patients with invasive GBS disease is shown in Table 4. All common MS were represented in more than one patient group. However, there were highly significant associations (when compared with all other age-groups) between SST III-2 and late-onset neonatal infection ($P = 0·0005$) and MS V and infection in the elderly ($P = 0·001$).

There were 17 isolates from CSF specimens, nine (53 %) of which were MS III (from three different SST/genotypes, each in a different cluster). The other eight isolates were distributed among five MS, seven genotypes and four clusters. Meningitis occurred in all age groups but comprised 23 % of cases in the late-onset neonatal group compared with 5 % in all other groups.

DISCUSSION

Our aim is to develop a comprehensive genotyping system to study the epidemiology and pathogenesis of GBS infection. It should be reproducible, objective and transportable between laboratories, comparable with and complementary to other typing methods and able to incorporate known virulence markers (Hauge et al., 1996; van Belkum et al., 2001). Based on these criteria, we first developed a molecular serotyping method based on the cps gene cluster. It compared favourably with, but was more sensitive than, conventional serotyping (Kong et al., 2002a) and allowed us to identify several subtypes of serotype III, as described by others (Takahashi et al., 2002). A second subtyping method based on the family of genes encoding variable surface protein antigens (bca/a/as/rib/alp2/alp3/alp4) and the IgA-binding protein C beta (bac) is more sensitive and objective than conventional protein serotyping, which cannot type all isolates and is sometimes misleading (Lachenauer et al., 2000). Our methods can also identify more members of the family of variable antigen genes and distinguish numerous bac subgroups (Kong et al., 2002b).

To extend our typing system, we selected five MGE that have been identified in GBS (Tamura et al., 2000; Rubens et al., 1989; Granlund et al., 1998, 2001; Spellerberg et al., 2000),

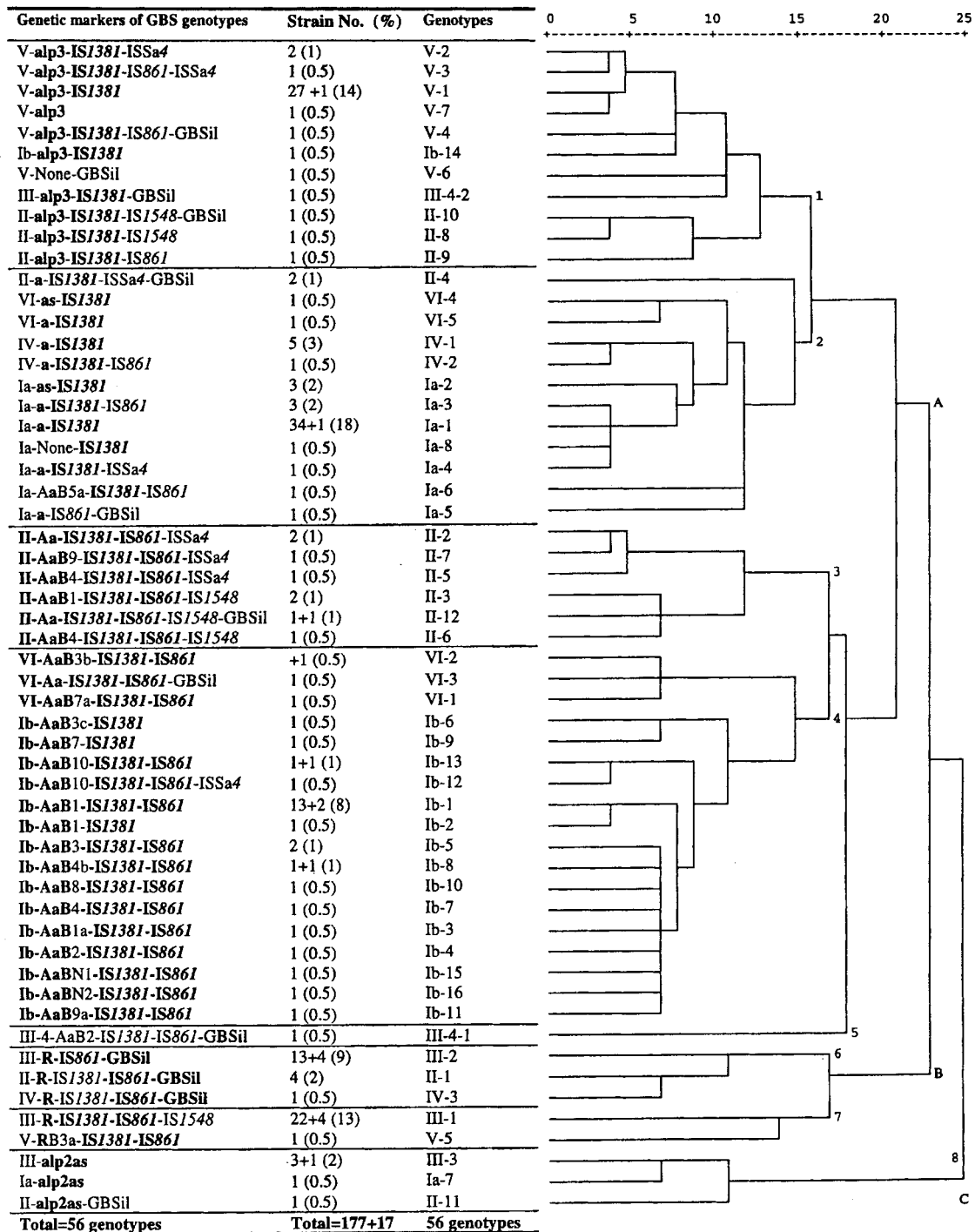| Genetic markers of GBS genotypes	Strain No. (%)	Genotypes
V-**alp3**-IS*1381*-ISSa*4*	2 (1)	V-2
V-**alp3**-IS*1381*-IS*861*-ISSa*4*	1 (0.5)	V-3
V-**alp3**-IS*1381*	27 +1 (14)	V-1
V-**alp3**	1 (0.5)	V-7
V-**alp3**-IS*1381*-IS*861*-GBSil	1 (0.5)	V-4
Ib-**alp3**-IS*1381*	1 (0.5)	Ib-14
V-None-GBSil	1 (0.5)	V-6
III-**alp3**-IS*1381*-GBSil	1 (0.5)	III-4-2
II-**alp3**-IS*1381*-IS*1548*-GBSil	1 (0.5)	II-10
II-**alp3**-IS*1381*-IS*1548*	1 (0.5)	II-8
II-**alp3**-IS*1381*-IS*861*	1 (0.5)	II-9
II-**a**-IS*1381*-ISSa4-GBSil	2 (1)	II-4
VI-**as**-IS*1381*	1 (0.5)	VI-4
VI-**a**-IS*1381*	1 (0.5)	VI-5
IV-**a**-IS*1381*	5 (3)	IV-1
IV-**a**-IS*1381*-IS*861*	1 (0.5)	IV-2
Ia-**as**-IS*1381*	3 (2)	Ia-2
Ia-**a**-IS*1381*-IS*861*	3 (2)	Ia-3
Ia-**a**-IS*1381*	34+1 (18)	Ia-1
Ia-None-IS*1381*	1 (0.5)	Ia-8
Ia-**a**-IS*1381*-ISSa*4*	1 (0.5)	Ia-4
Ia-AaB5a-IS*1381*-IS*861*	1 (0.5)	Ia-6
Ia-**a**-IS*861*-GBSil	1 (0.5)	Ia-5
II-**Aa**-IS*1381*-IS*861*-ISSa*4*	2 (1)	II-2
II-**AaB9**-IS*1381*-IS*861*-ISSa*4*	1 (0.5)	II-7
II-**AaB4**-IS*1381*-IS*861*-ISSa*4*	1 (0.5)	II-5
II-**AaB1**-IS*1381*-IS*861*-IS*1548*	2 (1)	II-3
II-**Aa**-IS*1381*-IS*861*-IS*1548*-GBSil	1+1 (1)	II-12
II-**AaB4**-IS*1381*-IS*861*-IS*1548*	1 (0.5)	II-6
VI-**AaB3b**-IS*1381*-IS*861*	+1 (0.5)	VI-2
VI-**Aa**-IS*1381*-IS*861*-GBSil	1 (0.5)	VI-3
VI-**AaB7a**-IS*1381*-IS*861*	1 (0.5)	VI-1
Ib-**AaB3c**-IS*1381*	1 (0.5)	Ib-6
Ib-**AaB7**-IS*1381*	1 (0.5)	Ib-9
Ib-**AaB10**-IS*1381*-IS*861*	1+1 (1)	Ib-13
Ib-**AaB10**-IS*1381*-IS*861*-ISSa*4*	1 (0.5)	Ib-12
Ib-**AaB1**-IS*1381*-IS*861*	13+2 (8)	Ib-1
Ib-**AaB1**-IS*1381*	1 (0.5)	Ib-2
Ib-**AaB3**-IS*1381*-IS*861*	2 (1)	Ib-5
Ib-**AaB4b**-IS*1381*-IS*861*	1+1 (1)	Ib-8
Ib-**AaB8**-IS*1381*-IS*861*	1 (0.5)	Ib-10
Ib-**AaB4**-IS*1381*-IS*861*	1 (0.5)	Ib-7
Ib-**AaB1a**-IS*1381*-IS*861*	1 (0.5)	Ib-3
Ib-**AaB2**-IS*1381*-IS*861*	1 (0.5)	Ib-4
Ib-**AaBN1**-IS*1381*-IS*861*	1 (0.5)	Ib-15
Ib-**AaBN2**-IS*1381*-IS*861*	1 (0.5)	Ib-16
Ib-**AaB9a**-IS*1381*-IS*861*	1 (0.5)	Ib-11
III-4-AaB2-IS*1381*-IS*861*-GBSil	1 (0.5)	III-4-1
III-**R**-IS*861*-GBSil	13+4 (9)	III-2
II-**R**-IS*1381*-IS*861*-GBSil	4 (2)	II-1
IV-**R**-IS*1381*-IS*861*-GBSil	1 (0.5)	IV-3
III-**R**-IS*1381*-IS*861*-IS*1548*	22+4 (13)	III-1
V-**RB3a**-IS*1381*-IS*861*	1 (0.5)	V-5
III-**alp2as**	3+1 (2)	III-3
Ia-**alp2as**	1 (0.5)	Ia-7
II-**alp2as**-GBSil	1 (0.5)	II-11
Total=56 genotypes	Total=177+17	56 genotypes

Fig. 1. Genetic relationship of 194 Australasian invasive GBS strains (or 56 genotypes). At a distance about of 16, the 56 genotypes could be separated into eight clusters (1–8); at a distance of about 22·5, the 56 genotypes could be separated into three cluster groups (A, B, C). Genetic markers of GBS genotypes: see text for explanation of *cps* SST and the five MGE. PGP codes are explained in Table 2; 'None', isolate contains none of the protein genes. Molecular markers in bold show the common features in each cluster. Strain No. (%): numbers after '+' are the numbers of CSF isolates, the others are blood isolates. Genotypes: each genotype was characterized by a distinct combination of the *cps* genes, PGP and MGE. The predominant genotype in each serotype was named the number '1' genotype of that serotype.

Table 4. Relationship between genotypes of invasive GBS isolates and patient age groups

Genotype	0–6 days	7 days–3 months	4 months–14 years*	15–45 years	46–60 years	>60 years	Total
Ia total	18	6 + 1†	1	8	3	9	45 + 1
Ia-1	14	4 + 1	1	7	3	6	35 + 1
Ia-2–Ia-8	4	2	–	1	–	3	10
Ib total	5	5 + 3	–	6	3	10 + 1	29 + 4
Ib-1	2	1 + 1	–	3	2	5 + 1	13 + 2
Ib-2–Ib-16	3	4 + 2	–	3	1	5	16 + 2
II total	8	1	–	4 + 1	1	4	18 + 1
III total	12 + 2	10 + 4	2 + 1	4 + 1	7 + 1	5	40 + 9
III-1	6 + 1	4	1 + 1	1 + 1	6 + 1	4	22 + 4
III-2	5	5 + 4‡	1	2	–	–	13 + 4
III-3–III-4	1 + 1	1	–	1	1	1	5 + 1
IV total	3	–	–	–	–	4	7
V total	4	4	2	5	2	17 + 1§	34 + 1
V-1	3	3	2	4	2	13 + 1	27 + 1
V-2–V-7	1	1	–	1	–	4	7
VI total	1	–	–	–	+ 1	3	4 + 1
All genotypes	51 + 2 = 53	26 + 8 = 34	5 + 1 = 6	27 + 2 = 29	16 + 2 = 18	52 + 2 = 54	177 + 17 = 194

*Five cases aged 4 months to 1 year and one case aged 3 years.

†Numbers after '+' refer to CSF isolates; all others are from blood.

‡SST III-2 in late-onset infection (9/34 = 26 %) compared with all other patient age groups (8/160 = 5 %): $P = 0.0005$, odds ratio (OR) 6·8; 95 % confidence interval (CI) 2·4–19.4.

§MS V in elderly (18/54 = 33 %) compared with all other age groups (17/140 = 12 %): $P = 0.001$, OR 3·6; 95 % CI 1·7–7·7).

including four IS and a type II intron, and developed PCR methods to detect their presence in GBS isolates. This significantly increased the level of discrimination of our GBS genotyping system. Several new MGE have been found recently in GBS (Franken *et al.*, 2001; Takahashi *et al.*, 2002; Tettelin *et al.*, 2002; Glaser *et al.*, 2002), some of which, after evaluation, could be added to the genotyping system to increase its discriminatory ability further. It could also be increased, for epidemiological studies, by combining it with other methods, such as multilocus enzyme electrophoresis (MLEE) (Musser *et al.*, 1989), multilocus sequence typing (Enright & Spratt, 1999), random amplification of polymorphic DNA (RAPD) (Chatellier *et al.*, 1996; Limansky *et al.*, 1998; Martinez *et al.*, 2000) or restriction length fragment polymorphism (RFLP) (Tamura *et al.*, 2000). In future, genome microarray-based assays have the potential to provide the greatest discriminatory ability. Combining genotyping and identification of antigens and virulence markers with clinical data would allow better definition of potentially virulent genotypes.

We used only invasive GBS isolates in this study to demonstrate the practical value of our genotyping system and to determine the distribution of genotypes in different patient age groups. Among our 194 Australasian invasive GBS isolates we identified 56 genotypes, of which five (Ia-1, Ib-1, III-1, III-2 and V-1) accounted for 62 % of isolates.

Whether this indicates that these genotypes are more virulent or just more prevalent among GBS isolates in general is not clear. Previous studies, mostly confined to isolates from neonates and pregnant carriers, have suggested that there is less diversity among invasive than colonizing strains (Musser *et al.*, 1989; Quentin *et al.*, 1995). Another study confirmed associations between genotypes and virulence factors (hyaluronidase, IgA-binding protein, C alpha and variants of C beta proteins) but found no differences between invasive neonatal and colonizing genital isolates (Hauge *et al.*, 1996). Because genotypes associated with invasive GBS disease vary in different age groups and, potentially, in different geographical areas and at different times, valid comparisons between invasive and colonizing strains will require selection of both, contemporaneously, from similar populations.

The phylogenetic tree derived from our results showed both similarities to, and differences from, that derived by Hauge *et al.* (1996). Between them, the two studies examined most known GBS genotypic and phenotypic markers (more than 30 together). Both showed relationships between *cps* serotype and PGP and demonstrated a similar level of discrimination: 56 (this study) and 58 (Hauge *et al.*, 1996) genotypes among 194 and 91 isolates, respectively. Both showed that certain known virulence markers, C beta, C alpha variants and hyaluronidase production (indirectly in our study; see below), were associated with distinct clonal lineages. The

main difference between the two studies was in the source of isolates. Ours included a high proportion from adults with invasive disease, in which serotype V was common, whereas the other study examined invasive isolates from infants and genital isolates, with a greater proportion of serotype III.

Our 17 CSF isolates were genetically heterogeneous, although more than half of them belonged to MS III. They were distributed among all age groups but the highest proportion of cases of meningitis was among cases of late-onset neonatal infection. Others have found that CSF isolates from neonates are much more likely than vaginal isolates to be serotype III (77 vs 27 %) and were confined to fewer genotypes, based on RAPD and MLEE typing, than genital tract isolates (Chatellier *et al.*, 1996). The relative heterogeneity of our isolates is probably related to the wide age range of patients from whom they were isolated.

GBS serotype III has been subdivided previously into two or more phylogenetic subgroups or subtypes on the basis of differences among *cps* gene clusters (Sellin *et al.*, 1995), RFLP plus *hylB*, *infB* genes, MGE (Takahashi *et al.*, 2002) or rRNA genes (Chatellier *et al.*, 1996). Our subdivision of serotype III into four SST, based on differences within the *cps* gene cluster (Kong *et al.*, 2002a), was supported by corresponding differences in surface PGP (Kong *et al.*, 2002b) and distribution of MGE. It is likely that our SST III-1, III-2, III-3 and III-4 correspond with subtypes III-2, III-3, III-1, III-4, respectively, described by Takahashi *et al.* (2002), based on similar MGE content and their relative predominance. SST III-1 and III-2 are the predominant genotypes of MS III. Our study, and that of Takahashi *et al.* (2002), indicate that SST III-2 (subtype III-3) is the most virulent. SST III-3 and III-4 genotypes are apparently related to MS Ia and MS II, respectively. They share sequence similarity in sections of the corresponding *cps* clusters (Kong *et al.*, 2002a) and genotype Ia-7, MS II genotypes in cluster 3 and genotype II-10 are closely related to genotypes III-3, III-4-1 and III-4-2, respectively (Fig. 1) (Kong *et al.*, 2002b), suggesting that they have common ancestors. These results suggest that SST III-3 and III-4 have arisen by recombination between serotypes, as occurs quite commonly in *Streptococcus pneumoniae* (Jiang *et al.*, 2001). The serotype III strain NEM316, of which the genome sequence has been published recently, belongs to SST III-3 (Glaser *et al.*, 2002). The other recently published GBS genome was of a serotype V strain, 2603 V/R, related to the relatively uncommon genotype V-5 (Tettelin *et al.*, 2002).

Our genotyping system, based on three sets of genetic markers, is highly discriminatory. Because it provides useful phenotypic data, including antigenic composition, it will be useful for epidemiological surveillance of GBS, especially in relation to potential GBS vaccine use. Further investigation of the distribution and multiplicity of MGE (including newly described ones) and their associations with virulence markers is needed to determine their role, if any, in the pathogenesis of GBS disease. The relationships between putative high-virulence genotypes (Musser *et al.*, 1989) and

patient characteristics (age and/or underlying risk factors) and whether there are significant differences between CSF isolates (or genotypes) and other invasive or colonizing strains require further study, which will be facilitated by our genotyping system. Using this system, we have demonstrated a clonal population structure among invasive Australasian GBS isolates. Comparison of the genotypes of invasive neonatal isolates with those of colonizing vaginal GBS isolates is in progress, in order to identify potential markers of virulence.

ACKNOWLEDGEMENTS

REFERENCES

Ahmet, Z., Stanier, P., Harvey, D. & Holt, D. (1999). New PCR primers for the sensitive detection and specific identification of group B beta-hemolytic streptococci in cerebrospinal fluid. *Mol Cell Probes* **13**, 349–357.

Chaffin, D. O., Beres, S. B., Yim, H. H. & Rubens, C. E. (2000). The serotype of type Ia and III group B streptococci is determined by the polymerase gene within the polycistronic capsule operon. *J Bacteriol* **182**, 4466–4477.

Chatellier, S., Huet, H., Kenzi, S., Rosenau, A., Geslin, P. & Quentin, R. (1996). Genetic diversity of rRNA operons of unrelated *Streptococcus agalactiae* strains isolated from cerebrospinal fluid of neonates suffering from meningitis. *J Clin Microbiol* **34**, 2741–2747.

Enright, M. C. & Spratt, B. G. (1999). Multilocus sequence typing. *Trends Microbiol* **7**, 482–487.

Franken, C., Haase, G., Brandt, C., Weber-Heynemann, J., Martin, S., Lammler, C., Podbielski, A., Lutticken, R. & Spellerberg, B. (2001). Horizontal gene transfer and host specificity of beta-haemolytic streptococci: the role of a putative composite transposon containing *scpB* and *lmb*. *Mol Microbiol* **41**, 925–935.

Glaser, P., Rusniok, C., Buchrieser, C. & 9 other authors (2002). Genome sequence of *Streptococcus agalactiae*, a pathogen causing invasive neonatal disease. *Mol Microbiol* **45**, 1499–1513.

Granlund, M., Oberg, L., Sellin, M. & Norgren, M. (1998). Identification of a novel insertion element, IS*1548*, in group B streptococci, predominantly in strains causing endocarditis. *J Infect Dis* **177**, 967–976.

Granlund, M., Michel, F. & Norgren, M. (2001). Mutually exclusive distribution of IS*1548* and GBSi1, an active group II intron identified in human isolates of group B streptococci. *J Bacteriol* **183**, 2560–2569.

Hauge, M., Jespersgaard, C., Poulsen, K. & Kilian, M. (1996). Population structure of *Streptococcus agalactiae* reveals an association between specific evolutionary lineages and putative virulence factors but not disease. *Infect Immun* **64**, 919–925.

Jiang, S. M., Wang, L. & Reeves, P. R. (2001). Molecular characterization of *Streptococcus pneumoniae* type 4, 6B, 8, and 18C capsular polysaccharide gene clusters. *Infect Immun* **69**, 1244–1255.

Kong, F., Gowan, S., Martin, D., James, G. & Gilbert, G. L. (2002a). Serotype identification of group B streptococci by PCR and sequencing. *J Clin Microbiol* **40**, 216–226.

Kong, F., Gowan, S., Martin, D., James, G. & Gilbert, G. L. (2002b). Molecular profiles of group B streptococcal surface protein antigen genes: relationship to molecular serotypes. *J Clin Microbiol* **40**, 620–626.

Lachenauer, C. S., Creti, R., Michel, J. L. & Madoff, L. C. (2000). Mosaicism in the alpha-like protein genes of group B streptococci. *Proc Natl Acad Sci U S A* **97**, 9630–9635.

Limansky, A. S., Sutich, E. G., Guardati, M. C., Toresani, I. E. & Viale, A. M. (1998). Genomic diversity among *Streptococcus agalactiae* isolates detected by a degenerate oligonucleotide-primed amplification assay. *J Infect Dis* **177**, 1308–1313.

Mahillon, J. & Chandler, M. (1998). Insertion sequences. *Microbiol Mol Biol Rev* **62**, 725–774.

Mahillon, J., Leonard, C. & Chandler, M. (1999). IS elements as constituents of bacterial genomes. *Res Microbiol* **150**, 675–687.

Martinez, G., Harel, J., Higgins, R., Lacouture, S., Daignault, D. & Gottschalk, M. (2000). Characterization of *Streptococcus agalactiae* isolates of bovine and human origin by randomly amplified polymorphic DNA analysis. *J Clin Microbiol* **38**, 71–78.

Martinez-Abarca, F. & Toro, N. (2000). Group II introns in the bacterial world. *Mol Microbiol* **38**, 917–926.

Mawn, J. A., Simpson, A. J. & Heard, S. R. (1993). Detection of the C protein gene among group B streptococci using PCR. *J Clin Pathol* **46**, 633–636.

Musser, J. M., Mattingly, S. J., Quentin, R., Goudeau, A. & Selander, R. K. (1989). Identification of a high-virulence clone of type III *Streptococcus agalactiae* (group B streptococcus) causing invasive neonatal disease. *Proc Natl Acad Sci U S A* **86**, 4731–4735.

Quentin, R., Huet, H., Wang, F. S., Geslin, P., Goudeau, A. & Selander, R. K. (1995). Characterization of *Streptococcus agalactiae* strains by multilocus enzyme genotype and serotype: identification of multiple virulent clone families that cause invasive neonatal disease. *J Clin Microbiol* **33**, 2576–2581.

Rubens, C. E., Heggen, L. M. & Kuypers, J. M. (1989). IS*861*, a group B streptococcal insertion sequence related to IS*150* and IS*3* of *Escherichia coli*. *J Bacteriol* **171**, 5531–5535.

Schuchat, A. (1998). Epidemiology of group B streptococcal disease in the United States: shifting paradigms. *Clin Microbiol Rev* **11**, 497–513.

Sellin, M., Hakansson, S. & Norgren, M. (1995). Phase-shift of polysaccharide capsule expression in group B streptococci, type III. *Microb Pathog* **18**, 401–415.

Spellerberg, B., Martin, S., Franken, C., Berner, R. & Lutticken, R. (2000). Identification of a novel insertion sequence element in *Streptococcus agalactiae*. *Gene* **241**, 51–56.

Takahashi, S., Detrick, S., Whiting, A. A., Blaschke-Bonkowksy, A. J., Aoyagi, Y., Adderson, E. E. & Bohnsack, J. F. (2002). Correlation of phylogenetic lineages of group B streptococci, identified by analysis of restriction-digestion patterns of genomic DNA, with *infB* alleles and mobile genetic elements. *J Infect Dis* **186**, 1034–1038.

Tamura, G. S., Herndon, M., Przekwas, J., Rubens, C. E., Ferrieri, P. & Hillier, S. L. (2000). Analysis of restriction fragment length polymorphisms of the insertion sequence IS*1381* in group B streptococci. *J Infect Dis* **181**, 364–368.

Tettelin, H., Masignani, V., Cieslewicz, M. J. & 40 other authors (2002). Complete genome sequence and comparative genomic analysis of an emerging human pathogen, serotype V *Streptococcus agalactiae*. *Proc Natl Acad Sci U S A* **99**, 12391–12396.

van Belkum, A., Struelens, M., de Visser, A., Verbrugh, H. & Tibayrenc, M. (2001). Role of genomic typing in taxonomy, evolutionary genetics, and microbial epidemiology. *Clin Microbiol Rev* **14**, 547–560.

Wastfelt, M., Stalhammar-Carlemalm, M., Delisse, A. M., Cabezon, T. & Lindahl, G. (1996). Identification of a family of streptococcal surface proteins with extremely repetitive structure. *J Biol Chem* **271**, 18892–18897.