CHAPTER 3

GBS MOLECULAR SEROTYPE IDENTIFICATION

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

1. Centre for Infectious Diseases and Microbiology, Insititute of Clinical Pathology and Medical Research, Westmead Hospital, Westmead, NSW 2145, Australia
2. Environmental Science and Research, Porirua, Wellington, New Zealand

This chapter has been published in *Journal of Clinical Microbiology* 2002; 40:216-226. (see Appendices section)

Statement of Joint Authorship

Kong, F. (candidate)
Did all the molecular serotype identification work, interpreted the data and wrote the manuscript.

Gowan, S. and Martin, D.
Provided and serotyped the New Zealand GBS strains.

James, G.
Provided all the needed molecular experiment equipment, reagents and software.

Gilbert G. L. (supervisor)
Supervised the overall project. Assisted in research design, in the analysis and interpretation of data, and made a significant contribution to the manuscript.

3.1. SUMMARY

Group B streptococcus (GBS; *Streptococcus agalactiae*) is the commonest cause of neonatal and obstetric sepsis and an increasingly important cause of septicaemia in the elderly and immunocompromised patients. Ongoing surveillance, to monitor GBS serotype distribution, will be needed to guide the development and use of GBS conjugate vaccines. Based on previously published sequences of the capsular polysaccharide synthesis (*cps*) gene clusters, we designed sequencing primers to further define partial *cps* gene clusters for eight of the nine GBS serotypes (Ia to VII). Subsequently, we designed and evaluated primers to identify serotypes Ia, Ib, III, IV, V and VI directly by PCR and all eight serotypes (Ia to VII) by sequence heterogeneity. 206 clinical GBS isolates were used to compare our molecular serotype (MS) identification method with conventional serotyping (CS). All clinical isolates were assigned an MS, compared with 188 of 206 (91.3%) using antisera. A small number of isolates (serosubtypes III-3 and III-4) showed different serotype specificities between PCR and sequencing, but PCR results correlated with CS. Overall agreement between MS and CS, for isolates for which results of both were available, was 100% (188/188). MS is a specific and practical alternative to conventional GBS serotyping and will facilitate epidemiological studies.

3.2. INTRODUCTION

Group B streptococcus (GBS; *Streptococcus agalactiae*) is the commonest cause of neonatal and obstetric sepsis and an increasingly important cause of septicaemia in the elderly and immunocompromised patients (Schuchat, 1999; Tyrrell *et al*., 2000). The incidence of neonatal GBS sepsis has been reduced in recent years by the use of intrapartum antibiotic prophylaxis (Schrag *et al*., 2000), but there are many problems with this approach (Levine *et al*., 1999). In future, vaccination is likely to be preferred and there has been considerable progress in development of conjugate polysaccharide GBS vaccines (Paoletti *et al*., 1999).

Before the introduction of vaccines, extensive epidemiological studies will be required to assess, not only the burden of disease, but also the distribution of GBS serotypes to determine the optimal formulation of vaccine antigens (Harrison *et al.*, 1998). Serotype distribution based on one geographic location or small numbers of patients may not be generally applicable (Hickman *et al.*, 1999; Lin *et al.*, 1998). Continued monitoring will be necessary to assess the suitability of combinations of GBS vaccine antigens for different target populations in different geographic locations (Harrison *et al.*, 1998; Hickman *et al.*, 1999).

Nine capsular polysaccharide GBS serotypes have been described (Harrison *et al.*, 1998; Hickman *et al.*, 1999). Various serotyping methods have been used, including immuno-precipitation (Wilkinson & Moody, 1969), enzyme immunoassay (Holm & Hakansson, 1988), coagglutination (Hakansson *et al.*, 1992), counter-immunoelectrophoresis, and capillary precipitation (Triscott & Davis, 1980), latex agglutination (Zuerlein *et al.*, 1991), fluorescence microscopy (Cropp *et al.*, 1974) and inhibition-ELISA (Arakere *et al.*, 1999). These methods are labour-intensive and require high-titer serotype-specific antisera, which are expensive and difficult to make and commercially available for only six serotypes, serotypes Ia to V (Arakere *et al.*, 1999). Molecular genotyping, such as pulsed-field gel electrophoresis (Rolland *et al.*, 1999), restriction endonuclease analysis (Nagano *et al.*, 1991) are useful for epidemiological studies but do not generally identify serotypes. Molecular serotype (MS) identification methods are theoretically attractive because of their potentially high discriminatory power and reproducibility (Sellin *et al.*, 2000). PCR-based assays have been used to detect and genotype GBS, but further development is needed to make them practicable for use in serotype identification (Ke *et al.*, 2000; Sellin *et al.*, 2000).

In this study, we used published sequences of the capsular polysaccharide synthesis (*cps*) gene clusters of GBS serotypes Ia and III (Chaffin *et al.*, 2000; Yamamoto *et al.*, 1999), and our own sequencing results, to analyse partial *cps* gene cluster sequences of eight serotypes (serotypes Ia to VII) and to develop a MS method.

Recently published sequences of *cps* gene clusters of serotypes Ib (Miyake *et al.*, submitted into GenBank, 2001 [GenBank accession number AB050723]), IV, V and VI (McKinnon *et al.,* submitted into GenBank, 2001 [GenBank accession numbers AF355776, AF349539, AF337958, respectively]) confirmed our sequence analysis. Subsequently, we used these sequences to design PCR primers specific for serotypes Ib, IV, V and VI.

3.3. MATERIALS AND METHODS

3.3.1. GBS reference strains and clinical isolates.

A panel of nine GBS serotypes (Ia to VIII) was kindly provided by Dr Lawrence Paoletti, Channing Laboratory, Boston, USA (reference panel 1). Dr Diana Martin, Streptococcus Reference Laboratory, at Environmental Science and Research (ESR), Wellington, New Zealand, provided another panel of nine international reference GBS type-strains including serotypes Ia to VI (reference panel 2) (Table 3.1.). In addition, we tested isolates from 205 clinical cases including 146 which had been referred from various laboratories in New Zealand for serotyping and 59 isolated from normally sterile sites over a period of ten years in one diagnostic laboratory in Sydney. One culture was subsequently shown to be mixed, so 206 different isolates were examined. Conventional serotyping (CS) was performed at the Streptococcus Reference Laboratory, at ESR, Wellington, New Zealand, and MS at the Centre for Infectious Diseases and Microbiology (CIDM), Insititute of Clinical Pathology and Medical Research (ICPMR), Sydney, Australia.

The two panels of GBS reference strains and 63 selected clinical isolates were studied in more detail, by sequencing >2200 base pairs (bp) of each to identify appropriate sequences for use in MS. These and the remaining clinical isolates were then used to evaluate the MS method and compare results with those of CS. Typing by both methods was done initially without knowledge of results of the other.

Bacterial isolates were retrieved from storage by subculture on blood agar plates (Columbia II agar base supplemented with 5% horse blood) and incubated overnight at 37^{o}C.

3.3.2. Conventional serotyping (CS).

CS was performed using standard methodology (Wilkinson & Moody, 1969). Briefly, an acid-heated (56°C) extract was prepared for each isolate and the serotype determined by immuno-precipitation of type-specific antiserum in agarose. An isolate was considered positive for a particular serotype when the precipitation occurring formed a line of identity with that of the control strain. Antisera used were prepared at ESR in rabbits against serotypes Ia, Ib, II, III, IV and V. Fourteen selected isolates, including six that were nontypable using antisera against serotypes I-V, six that initially gave discrepant results between CS and MS and two separate isolates from a mixed culture, were kindly tested using antisera against all serotypes by Abbie Weisner and Dr Androulla Efstratiou at Central Public Health Laboratory, Colindale, London, UK.

3.3.3. Oligonucleotide primers.

Four previously published oligonucleotide primers, and a series of new primers designed by us were used to sequence the genes of interest, namely 16S-23S rRNA intergenic spacer region and partial *cps* gene cluster, or to amplify unique sequences of GBS *cps* clusters of each serotype. The sequences, target sites and melting temperatures (*T*m) of primers used in this study are shown in Table 3.2. Some were designed with high melting temperatures to be used in rapid cycle PCR (Ke *et al.*, 2000).

Table 3.1. Reference strainss used in GBS molecular serotype identification study.

Lab strain number	Source	Serotype	MS/subtype	GenBank numbers
Reference panel 1[a]				
090	Channing	Ia	Ia	AF332893
H36B	Channing	Ib	Ib	AF332903
18RS21	Channing	II	II	AF332905
M781	Channing	III	III-2[c]	AF332896
3139	Channing	IV	IV	AF332908
CJB 111	Channing	V	V	AF332910
SS1214	Channing	VI	VI	AF332901
7271	Channing	VII	VII	AF332913
JM9 130013	Channing	VIII	VIII	-
Reference panel 2[b]				
NZRM 908 (NCDC SS615)	ESR	Ia	Ia	AF332894
NZRM 909 (NCDC SS618)	ESR	Ib	Ib	AF332904
NZRM 910 (NCDC SS700)	ESR	Ia	Ia	AF332914
NZRM 911 (NCDC SS619)	ESR	II	II	AF332906
NZRM 912 (NCDC SS620)	ESR	III	III-3[c]	AF332897
NZRM 2217 (Prague 25/60)	ESR	Non-typable (R)	II	AF332907
NZRM 2832 (Prague 1/82)	ESR	IV	IV	AF332909
NZRM 2833 (Prague 10/84)	ESR	V	V	AF332911
NZRM 2834 (Prague 118754)	ESR	VI	VI	AF332902

Notes.

a. Reference panel 1: supplied by Dr Lawrence Paoletti, Channing Laboratory, Boston, USA.

b. Reference panel 2: supplied by Dr Diana Martin, ESR, Porirua, Wellington, New Zealand.

c. MS III subtypes based on sequence heterogeneity; see text for more detail.

3.3.4. DNA preparation and PCR.

Five individual GBS colonies or a sweep of culture were sampled using a disposable loop and resuspended in 200 µl of digestion buffer (10mM Tris-HCl [pH 8.0], 0.45% Triton X-100 and 0.45% Tween 20) in 2 ml Eppendorf tubes. The tubes containing GBS suspension were heated at 100^{o}C (dry block heater or water bath) for ten minutes then quenched on ice and centrifuged at 14,000 rpm (16,000x *g*) in an Eppendorf Centrifuge 5415C for two minutes to pellet the cell debris. 5 µl of each supernatant containing extracted DNA was used as template for PCR (Mawn *et al*., 1993).

PCR systems (25 µl for detection only, 50 µl for detection and sequencing) were used as previously described (Kong *et al*., 1999). The denaturation, annealing and elongation temperatures and times used were 96^{o}C for 1 second, 55-72^{o}C (according to the primer *T*m values or as previously described) for 1 second and 74^{o}C for 1 to 30 seconds (according to the length of amplicons), respectively, for 35 cycles.

10 µl of PCR products were analysed by electrophoresis on 1.5 % agarose gels, which were stained with 0.5 µg ethidium bromide ml^{-1}. For detection and serotype identification, the presence of PCR amplicons of expected length, shown by ultraviolet transillumination, were accepted as positive. For sequencing, 40 µl of PCR products were further purified by polyethylene glycol precipitation method (Ahmet *et al*., 1999).

3.3.5. Sequencing.

The PCR products were sequenced using Applied Biosystems (ABI) *Taq* DyeDeoxy terminator cycle-sequencing kits according to standard protocols. The corresponding amplification primers or inner primers were used as the sequencing primers.

Table 3.2. Primers used in GBS molecular serotype identification study.

Primer	Target gene name	*Tm* °C[a]	GenBank numbers	Sequence[b-d]
CFBS	*cfb*	56.7	X72754	**328**GAT GTA TCT ATC TGG AAC TCT AGT G**352**
Sag59[e]	*cfb*	77.4	X72754	**350**GTGGCTGGTGCATTGTTAT TTT CAC CAG CTG TAT TAG AAG TA**391**
Sag190[e]	*cfb*	76.8	X72754	**545**CATTAACCGGTTTTTCATAATCT GTT CCC TGA ACA TTA TCT TTG AT**500**
CFBA	*cfb*	63.2	X72754	**568**TTT TTC CAC GCT AGT AAT AGC CTC**545**
16SS	16S rRNA	69.3	AB023574	**1441**GCC GCC TAA GGT GGG ATA GAT G**1462**
23SA	23S rRNA	65.7	X68427	**70**CGT CGT TTG TCA CGT CCT TC**51**
DSF2[f]	16S rRNA	75.9	AB023574	**975**CATCCTTCTGACC GGC CTA GAG ATA GGC TTT CT**1007**
DSR1[f]	16S rRNA	81.5	AB023574	**1250**CGTCACCGG CTT GCG ACT CGT TGT ACC AA**1222**
cpsES3	*cpsE*	71.5	AB028896 (Ia), AF163833 (III)	**6410/6020**GTT AGA TGT TCA ATA TAT CAA TGA ATG GTC TAT TTG GTC AG**6450/6060**
cpsFS	*cpsF*	75.0	AB028896 (Ia), AF163833 (III)	**6777/6387**CAT CTG GTG CCG CTG TAG CAG TAC CAT T**6804/6414**
cpsFA	*cpsF*	73.2	AB028896 (Ia), AF163833 (III)	**6859/6469**GTC GAA AAC CTC TAT A/GT A AAC/T GGT CTT ACA A/GCC AAA TAA CTT ACC**6815/6425**
cpsGA	*cpsG*	54.7	AB028896 (Ia), AF163833 (III)	**7162/6772**AAG/C AGT TCA TAT CAT CAT ATG AGA G **7138/6748**
cpsGA1	*cpsG*	74.5	AB028896 (Ia), AF163833 (III)	**7199/6809**CCG CCA/G TGT GTG ATA ACA ATC TCA GCT TC**7171/6781**
IacpsHS1	*cpsH*	77.9	AB028896 (Ia)	**8463**GGC CTG CTG GGA TTA ATG AAT ATA GTT CCA GGT TTG C**8499**

cpsIA	*cpsI*	70.3	AB028896 (Ia), AF163833 (III)	**8816/8312**GTA TAA CTT CTA TCA ATG GAT GAG TCT GTT GTA GTA CGG**8778/8274**
IbcpsIS	*cpsI*	71.1	AB050723 (Ib)	**4116**GAT AAT AGT GGA GAA ATT TGT GAT AAT TTA TCT CAA AAA GAC G**4158**
IbcpsIA1	*cpsI*	78.6	AB050723 (Ib)	**4638**CCT GAT TCA TTG CAG AAG TCT TTA CGA TGC GAT AGG TG**4601**
IIIcpsHS	*cpsH*	72.1	AF163833 (III)	**7672**GAA TAC TAT TGG TCT GTA TGT TGG TTT TAT TAG CAT CGC**7710**
IVcpsHS1	*cpsH*	71.2	AF355776 (IV)	7887CCC AAG TAT AGT TAT GAA TAT TAG TTG GAT GGT TTT TGG7925
IVcpsMA	*cpsH*	80.7	AF355776 (IV)	**8265**GGG TCA ATT GTA TCG TCG CTG TCA ACA AAA CCA ATC AAA TC**8225**
VcpsHS2	*cpsH*	74.0	AF349539 (V)	**7871**CCC AGT GTG GTA ATG AAT ATT AGT TGG CTA GTT TTT GG**7908**
VcpsMA	*cpsM*	73.1	AF349539 (V)	**8244**CCC CCC ATA AGT ATA AAT AAT ATC CAA TCT TGC ATA GTC AG**8204**
VIcpsHS1	*cpsH*	77.2	AF337958 (VI)	**7767**CCT TAT TGG GCA AGG TAT AAG AGT TCC CTC CAG TGT G**7803**
VIcpsIA	*cpsI*	74.5	AF337958 (VI)	**8126**GAA GCA AAG ATT CTA CAC AGT TCT CAA TCA CTA ACT CCG**8088**

Notes.

a. The primer *T*m values were provided by the primer synthesiser (Sigma-Aldrich, Castle Hill, NSW, Australia)

b. Numbers represent the numbered base positions at which primer sequences start and finish (numbering start point "1" refer to the start points "1" of correspondent gene GenBank accession numbers).

c. Underlined sequences show bases added to modify previously published primers.

d. Letters behind "/" indicate alternative nucleotides in different serotypes.

e. From Ke *et al*., 2000.

f. From Ahmet *et al*., 1999.

3.3.6. Multiple sequence alignments.

Multiple sequence alignments were performed with *Pileup* and *Pretty* programs in Multiple Sequence Analysis program group. Both programs are provided in WebANGIS (Website: http://www1.angis.org.au/pbin/WebANGIS/wrapper.pl), ANGIS (Australian National Genomic Information Service).

3.3.7. Nucleotide sequence accession numbers.

The new sequence data reported in this chapter were deposited into the GenBank Nucleotide Sequence Databases with the following accession numbers: AF332893-AF332917, AF363032-AF363060, AF367973, AF381030 and AF381031 for partial *cps* gene clusters for two panels of reference strains (Table 3.1.) and selected representative clinical isolates. Previously reported sequence data used in this chapter have appeared in the GenBank Nucleotide Sequence Databases with the following accession numbers: AB023574 for the 16S rRNA gene, U39765, L31412 for the 16S-23S rRNA intergenic spacer regions, X68427 for the *S. oralis* 23S rRNA gene, X72754 for the *cfb* gene, AB028896 for the *cps* gene cluster for serotype Ia, AB050723 for the partial *cps* gene cluster for serotype Ib, AF163833 for the *cps* gene cluster for serotype III, AF355776 for the *cps* gene cluster for serotype IV, AF349539 for the *cps* gene cluster for serotype V, and AF337958 for the *cps* gene cluster for serotype VI.

3.4. RESULTS

3.4.1. PCR.

With two exceptions, all GBS-specific primer pairs produced amplicons of the expected size from all reference strains and clinical isolates tested (Table 3.2.). The exceptions were Sag59-Sag190 and CFBS-CFBA. Both target the *cfb* gene, but

failed to produce amplicons from one clinical isolate, despite repeated attempts. We assumed that this isolate either lacked the *cfb* gene or that the gene was present in a mutant form. It has been suggested previously that PCR targeting the *cfb* gene will not identify all GBS isolates (Hassan *et al*., 2000) and that another primer pair based on 16S rRNA gene, DSF2-DSR1 (Ahmet *et al*., 1999) was not entirely specific. Therefore, in this study, we used both primer pairs (DSF2-DSR1 and Sag59-Sag190) to confirm all the isolates were GBS.

3.4.2. Sequence heterogeneity at the 3'-end of *cpsD-cpsE-cpsF*-and the 5'-end of *cpsG*.

Using a series of primers targeting the 3'-end of *cpsD-cpsE-cpsF*-and the 5'-end of *cpsG*, we amplified and sequenced 2226- or 2217 bp – depending on the presence or absence of a nine-base repetitive sequence – from both panels of reference strains (serotypes Ia to VII) and 63 selected clinical isolates. Representative sequences were deposited into GenBank. See Table 3.1. for GenBank accession numbers of reference panel strains.

3.4.3. Repetitive sequence.

At the 3'-end region of *cpsD*, we found a nine-base repetitive sequence (TTA CGG CGA) in most isolates of MS Ia and II, some of MS III, all of MS IV, V, and VII, but none of the isolates of MS Ib or VI examined. (Table 3.3.). The presence or absence of this repetitive sequence can be used to further subtype MS Ia, II and III (see below).

Table 3.3. The heterogeneity of eight GBS serotypes in the regions of the 3'-end of *cpsD*- and the 5'-end of *cpsE*.

Sites[a]	**Ia**	**Ib**	**II/III-4**[d]	**III**	**IV**	**V**	**VI**	**VII**	**Specificity**
cpsD									
62	G	A	G[e]	A	A	A	A	G	Ia, II, VII
78-86 repetitive sequence - TTACGGCGA *cpsD-cpsE* spacer	- Ia-2[b]; + Ia-1[b]	-	- II-2[c, e]; + II-1[c]	- III-2[d]; + III-1[c], III-3[c]	+	+	-	+	See text
138	G	G	G	G	G	A[f]	G	G	V
139	G	G	G	A III-2; G III-1, III-3	G	G	G	G	III-2
144	T	T	T	G III-2; T III-1, III-3	T	T	T	T	III-2
cpsE									
198	A	C	A[e]	A	C	C[f]	A	A	Ib, IV, V
204	G	G	G	A III-2, III-3; G III-1	G	G	G	G	III-2, III-3
211	T	T	T	T	T	T	G	T	VI
218	C	C	C	C	C	C	T	C	VI
240	T	T	T	T	T	T	C	T	VI

249	T	C	T[e]	T	C	C[f]	T	T	Ib, IV, V
300	C	C	C	T III-2; C III-1, III-3	C	C	C	C	III-2
321	C	C	C	T III-1; C III-2, III-3	C	C	C	C	III-1
419	T	C	T[e]	T	T	T	T	T	Ib
429	A	T	A[e]	T	T	T	T	A	Ia, II, VII
437	C	C	C; T III-4	C	C	C	C	T	VII, III-4
457	T	A	C[e]	A	A	A	A	C	Ia, II, VII
466	G	G	G	G	A	G	G	A	IV
486	G	A	A	G III-3; A III-2, III-1	A	A	A	A	Ia, III-3
602	G	G	A[e]	G	G	G	G	A	II, VII
606	T	T	T	T	T	T	C	T	VI
627	T	C	C	C	C	C	C	C	Ia
636	C	T	T	C III-1; T III-2, III-3	T	T	T	T	Ia, III-1
645	C	T	C[e]	C	T	T	C	C	Ib, IV, V
803	A	A	A	A	A	A	T	A	VI

971	C	T	T	C	C	C	T	T	Ia, III, IV, V
1026	A	G	G	G III-2, III-1; A III-3	A	A	G	G	Ia, III-3, IV, V
1044	T	T	T	T	T	T	C	T	VI
1173	A	G	A	A	A	A	A	A	Ib
1194	C	C	C	A	A	C	A	C	III, IV, VI
1251	G	G	G	G	G	G	A	G	VI
1278	A	A	A	A	A	G	A	A	V

Notes.

a. Numbering start point "1" refers to the start point "1" of GenBank accession number AF332908 (for serotype IV reference strain 3139).

b. Repetitive sequence: Ia-1 present (+); Ia-2 absent (-) (see text).

c. Repetitive sequence: II-1 present (+); II-2 absent (-) (see text).

d. Repetitive sequence: III-1 and III-3 present (+); III-2 absent (-); III-4 variable (see text)

e. One CS II strain has mutations at the 9 sites (see text).

f. At positions 138, 198, and 249, one CS V reference strain (Prague 10/84) is identical with corresponding sequence in GenBank (GenBank accession number AF349539), the sequences are G, A and T, respectively; another CS V reference strain (CJB 111) and all the other sequenced CS V strains are identical, the sequences are A, C and C, respectively.

3.4.4. Intraserotype heterogeneity.

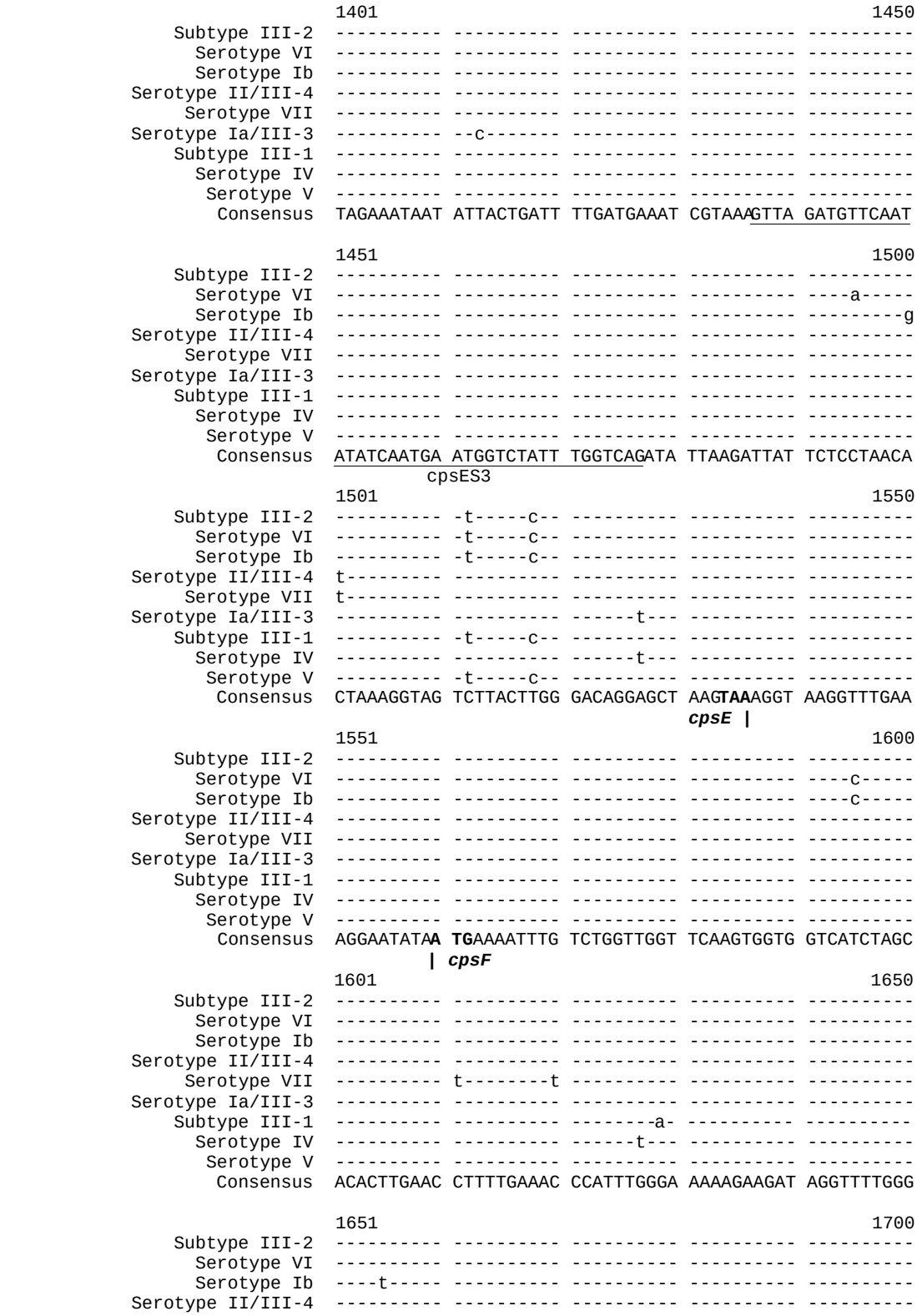
In general, intraserotype heterogeneity was low; there were minor random variations in a few isolates of all serotypes except MS III, in which the intraserotype heterogeneity was more complex. MS III could be divided into four sequence subtypes on the basis of heterogeneity at 22 positions (62, 139, 144, 204, 300, 321, 429, 437, 457, 486, 602, 636, 971, 1026, 1194, 1413, 1501, 1512, 1518, 1527, 1629 and 2134) and the presence or absence of the repetitive sequence (positions 78-86) (Figure 3.1. and Table 3.3.).

Among 60 MS III isolates (58 clinical isolates and two reference strains), subtypes III-1 (30 isolates) and III-2 (22 isolates) were predominant. The repetitive sequence was present in subtype III-1 but not III-2; there were differences at seven other sites (positions 139, 144, 204, 300, 321, 636 and 1629) (Table 3.3.).

There were five isolates belonging to subtype III-3, which contained the repetitive sequence and were identical with subtype III-1 at three variable sites (positions 139, 144 and 300) and with subtype III-2 at four (positions 204, 321, 626 and 1629). Subtype III-3 differed from both subtypes III-1 and III-2 at seven sites (positions 486, 1026, 1413, 1512, 1518, 1527 and 2134). These seven sites in subtype III-3 were identical with the corresponding sites of MS Ia.

There were three subtype III-4 isolates, whose sequences were nearly identical with the corresponding sequence of MS II. The only exception was at position 437, where the nucleotide was T in subtype III-4 (as in MS VII), and C in MS II. This difference can be used (in addition to PCR, see below) to differentiate subtype III-4 from MS II. Two subtype III-4 isolates contained the repetitive sequence, and the other did not. Because of the small number of subtype III-4 isolates, we did not use the repetitive sequence to subtype them further.

Figure 3.1. Molecular serotype identification based on the sequence heterogeneity of the 790 bp fragment (positions 1437-2226).



```
     Serotype VII  ---------- ---------- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- ---------- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
        Consensus  TAACCTTTGA TAAAGAAGAT GCTAGGAGTA TTCTAAGAGA AGAGATTGTA

                   1701                                                1750
    Subtype III-2  ---------- ---------- ---------- ---------- ----------
      Serotype VI  ---------- ---------- ---------- ---------- ----------
      Serotype Ib  ---------- ---------- ---------- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ---------- ---------- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- ---------- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
        Consensus  TATCATTGCT TCTTTCCAAC AAACCGTAAT GTCAAAAACT TGGTAAAAAA

                   1751                                                1800
    Subtype III-2  ---------- ---------- ---------- ---------- ----------
      Serotype VI  ---------- ---------- ---------- ---------- ----------
      Serotype Ib  ---------- ---------- ---------- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ---------- ---------- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- ---------- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
        Consensus  TACTATTCTA GCTTTTAAGG TCCTTAGAAA AGAAAGACCA GATGTTATCA

                   1801                                                1850
    Subtype III-2  ---------- ---------- ---------- ---------- ----------
      Serotype VI  ---------- ---------- ---------- ---------- ----------
      Serotype Ib  ---------- ---------- ---------- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ---------- ---------- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- -t-------- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
        Consensus  TATCATCTGG TGCCGCTGTA GCAGTACCAT TCTTTTATAT TGGTAAGTTA
                                    cpsFS
                   1851                                                1900
    Subtype III-2  ---------- ---------- ---------- ---------- ----------
      Serotype VI  ---------- ---------- ---------- ---------- ----------
      Serotype Ib  -----c---- ---------- ---------- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ----------  -----a---- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- ---------- ----------
       Serotype V  ---------- ---------- c--------- ---------- -g--------
        Consensus  TTTGGTTGTA AGACCGTTTA TATAGAGGTT TTCGACAGGA TAGATAAACC
                                   cpsFA
                   1901                                                1950
    Subtype III-2  ---------- ---------- ---------- ---------- ----------
      Serotype VI  ---------- ---------- ---------- ---------- ----------
      Serotype Ib  ---------- ---------- ---------- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ---------- ---------- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- ---------- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
        Consensus  AACTTTGACA GGAAAATTAG TGTATCCTGT AACAGATAAA TTTATTGTTC

                   1951                                                2000
    Subtype III-2  ---------- ---------- ---------- ---------- ----------
      Serotype VI  ---------- ---------- a--------- ---------- ----------
```

```
      Serotype Ib  ---------- ---------- ---------- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ---------- ---------- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- ---------- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
         Consensus  AGTGGGAAGA AATGAAAAAA GTTTATCCTA AGGCAATTAA TTTAGGAGGA

                    2001                                                2050
    Subtype III-2  ---------- ---------- ---------- ---------- ----------
      Serotype VI  ---------- ---------- ---------- ---------- ----------
      Serotype Ib  ---------- ---------- -----a---- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ---------- ---------- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- ---------- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
         Consensus  ATTTTTTAAT GATTTTTGTC ACAGTGGGGA CACATGAACA GCAGTTCAAC
                       cpsF | cpsG
                    2051                                                2100
    Subtype III-2  ---------- ---------- ---------- ---------- ----------
      Serotype VI  ---------- ---------- ---------- ---------- ----------
      Serotype Ib  ---------- ---------- ---------- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ---------- ---------- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- -------a-- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
         Consensus  CGTCTTATTA AAGAAGTTGA TAGATTAAAA GGGACAGGTG CTATTGATCA

                    2101                                                2150
    Subtype III-2  ---------- ---------- ---------- ---c------ ----------
      Serotype VI  ---------- ---------- ---------- ---------- ----------
      Serotype Ib  ---------- ---------- ---------- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ---------- ---------- ---------- ---------- ----------
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---c------ ----------
      Serotype IV  ---------- ---------- ---------- ---------- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
         Consensus  AGAAGTGTTC ATTCAAACGG GTTACTCAGA CTTTGAACCT CAGAATTGTC

                    2151                                                2200
    Subtype III-2  ---------- ---------- ---------- ---------- ----------
      Serotype VI  ---------- ---------- ---------- ---------- ----------
      Serotype Ib  ---------- ---------- ---------- ---------- ----------
Serotype II/III-4  ---------- ---------- ---------- ---------- ----------
     Serotype VII  ---------- ---------- ---------- ------g--- -----g----
Serotype Ia/III-3  ---------- ---------- ---------- ---------- ----------
    Subtype III-1  ---------- ---------- ---------- ---------- ----------
      Serotype IV  ---------- ---------- ---------- ---------- ----------
       Serotype V  ---------- ---------- ---------- ---------- ----------
         Consensus  AGTGGTCAAA ATTTCTCTCA TATGATGATA TGAACTCTTA CATGAAAGAA
                                          cpsGA
                   2201                  2226
      Subtype III-2  ---------- ---------- ------
        Serotype VI  ---------- ---------- ------
        Serotype Ib  ---------- ---------- c-----
  Serotype II/III-4  ---------- ---------- ------
       Serotype VII  ---------- ---------- ------
  Serotype Ia/III-3  ---------- ---------- ------
      Subtype III-1  ---------- ---------- ------
        Serotype IV  ---------- ---------- ------
         Serotype V  ---------- ---------- ------
           Consensus  GCTGAGATTG TTATCACACA TGGCGG
                         cpsGA1       *
```

Notes.

- Underlining parts show the relevant PCR and sequencing primers.
- Lowercase letters show the sequence polymorphism sites.
- Boldface parts show the gene start and stop codons.

3.4.5. Interserotype heterogeneity.

There were 56 sites of heterogeneity between the eight MS. The most suitable sites, for use in PCR and sequencing for MS, were a group of 23 sites nearest to the 3'-end of the region (Figure 3.1.), which were contained in the regions corresponding to the amplicons of primer pair of cpsES3-cpsGA1. Firstly, they were consistent across two panels of reference strains and most clinical isolates (the only exceptions were the small number of subtypes III-3 and III-4 isolates, see below). Secondly, they were relatively concentrated within a 790 bp region (positions 1437-2226), which is a convenient length for sequencing in a single reaction. Thirdly, they contained enough heterogeneity sites to allow differentiation, with few exceptions, of MS Ia-VII. Based only on this 790 bp region, subtype III-3 cannot be distinguished from MS Ia, nor subtype III-4 from MS II. However, they can be identified by MS III-specific PCR (see below).

Serotype VIII does not form amplicons with primer pairs targeting the 790 bp region, but can be identified by exclusion after PCR identification of GBS. In this study, one MS VIII isolate was identified, for which none of the primer pairs that amplify the 2226 bp region (in addition to those that amplify the 790 bp region) produced amplicons.

3.4.5. Mixed serotype-specificities in single isolates.

Eleven isolates were identified as one MS on the basis of the MS-specific PCR and overall sequence (within the 2226 or 2217 bp segment) but their sequences differed at some sites from isolates of the same MS and shared site-specific characteristics of

another. They included five subtype III-3 isolates and three subtype III-4 (see above). One non-serotypable reference strain (Prague 25/60), which was identified as MS II, differed from other MS II isolates at five sites at the 5'-end of the region, and was identical with MS III at three of these sites. Prague 25/60 MS III-specific PCR was negative. One clinical isolate identified as CS II, and MS II on the basis of its overall sequence, had bases at nine sites at the 5'-end of the region, that were characteristic of serotype Ib; MS Ib-specific PCR was negative. Finally, one CS V reference strain (Prague 10/84) had the same sequencing result as the corresponding sequence in GenBank (AF349539), but both were different, at three sites at the 5'-end of the region, from sequences of the other MS V strains that we studied.

All of these mixed-serotype specificities, except for those associated with subtypes III-3 and III-4, occurred at the 5'-end region of the 2226 or 2217 bp fragment. This supported our selection of the 3'-end 790 bp as the sequencing target for MS. Using this target, all MS were correctly identified except for MS III belonging to subtypes III-3 and III-4, which can be identified by MS III-specific PCR (see below).

3.4.6. MS identification based on MS-specific PCR targeting the 3'-end of *cpsG-cpsH-cpsI/cpsM*.

Our sequence alignment results showed that there was significant sequence heterogeneity in the 3'-end of *cpsG-cpsH-cpsI/cpsM* (data not shown), which makes it appropriate for use in the design of specific primer pairs for differentiation of serotypes Ia, Ib, III, IV, V, and VI directly by PCR (Table 3.2.). Using two panels of reference strains and the specified conditions, all MS-specific primer pairs amplified DNA only from the corresponding serotypes. When clinical isolates were tested, a MS was assigned, by PCR only, to 179 of 206 (86.9%) clinical isolates as follows: MS Ia 40; MS Ib 35; MS III 58; MS IV 7; MS V 36; MS VI 3.

3.4.7. Comparison of serotype identification results between MS and CS.

After CS and MS had been completed, the results were compared. Initial results were discrepant for 15 isolates, all were resolved by retesting or correction of clerical errors.

Table 3.4. Comparison of the results of conventional serotyping (CS) and molecular serotype identification (MS)/subtyping of 206 clinical GBS isolates.

	MS/subtype										
CS	**Ia**	**Ib**	**II**	**III-1[a]**	**III-2[a]**	**III-3[a]**	**III-4[a]**	**IV**	**V**	**VI**	**VIII**
Ia	38										
Ib		30									
II			25								
III				27	20	4	3				
IV								7			
V									31		
VI										2	
VIII											1
NT[b]	2	5	1	3	1				5	1	
Total (206)[c]	40	35	26[c]	30	21[c]	4	3	7	36	3	1

Notes.

a. For details of MS III sequence subtypes see text.

b. NT=nontypable.

c. One mixed culture was included as two separate isolates (one serotype II, one subtype III-2).

The CS and MS and sequence subtyping results are shown in Table 3.4. A MS was assigned to all isolates by PCR and/or sequencing, compared with 188 of 206 (91.3%) by CS. Specific PCR has not yet been developed for MS II and VIII, so all MS II isolates were determined by sequencing only and one MS VIII isolate by exclusion (see above). For all other isolates, the results of PCR and sequencing were consistent, except for subtypes III-3 and III-4 and other minor sequence differences described above. CS results correlated well with PCR results.

Final CS and MS results were the same for all 188 isolates (100%) for which results for both methods were available. Eighteen clinical isolates that were non-serotypable by CS, were assigned MS as follows: Ia, two; Ib, five; II, one; subtype III-1, three; subtype III-2, one; V, five; and VI, one.

3.4.8. Mixed culture.

Four clinical isolates gave positive results with MS III-specific PCR, but were provisionally identified as MS II by sequencing. Three were CS III and one CS II, with a weak cross-reaction with serotype III antiserum. These isolates were studied further by subculturing 12 individual colonies of each. All subcultures were tested by MS III-specific PCR. All 12 colony subcultures of the three CS III isolates were positive by MS III-specific PCR and the isolates were therefore classified as subtype III-4 (see above). However, 11 of 12 colony subcultures of the fourth isolate were negative by MS III-specific PCR; and one was positive by MS III-specific PCR. It was therefore assumed that this was a mixed culture, predominantly of MS/CS II. The CS of these two isolates were confirmed by retesting each individually. The one MS III-specific PCR positive and CS III colony was subsequently identified as subtype III-2 and included as an additional clinical isolate (total 206 in all).

3.4.9. Algorithm for serotype assignment of GBS by PCR and sequencing (see Table 3.2. for primer sequences).

Table 3.5. Algorithm for GBS molecular serotype (MS) identification by PCR and sequencing.

Amplification primer pairs*	PCR product size (base pairs)	Interpretation
GBS identification primer pairs		
Sag59-Sag190	196	GBS (*S. agalactiae*)
DSF2-DSR1	276	GBS (*S. agalactiae*)
GBS MS identification by MS-specific PCR		
IacpsHS1-cpsIA	354	serotype Ia
IbcpsIS-IbcpsIA1	523	serotype Ib
IIIcpsHS-cpsIA	641	serotype III
IVcpsHS1-IVcpsMA	379	serotype IV
VcpsHS2-VcpsMA	374	serotype V
VIcpsHS1-VIcpsIA	360	serotype VI
GBS MS identification by sequencing		
cpsES3-cpsGA1	790	refer to Figure 3.1. to identify
cpsES3-cpsFA	450	serotypes according to the
cpsFS-cpsGA1	423	sequence heterogeneity

Note.

*See Table 3.2. for primer sequences.

In order to make GBS serotype identification using our PCR and sequencing method practicable, we designed an algorithm for clinical use (Table 3.5.). All the primers (except the inner sequencing primers) used were given high melting temperature (>70 ^{o}C), so rapid cycle PCR could be used.

3.5. DISCUSSION

Capsule production in GBS is controlled by capsular polysaccharide synthesis (*cps*) gene cluster (Chaffin *et al*., 2000), which had been sequenced for serotype Ia (Yamamoto *et al*., 1999) and serotype III (Chaffin *et al*., 2000) before we began our study. Differences between the two serotypes and serotype-specific definition regions of *cpsH* (capsular polysaccharide repeating unit polymerase gene, serotype definition gene) had been identified (Chaffin *et al*., 2000; Yamamoto *et al*., 1999). Corresponding sequences for serotype Ib (Miyake *et al.*, submitted into GenBank, 2001 [GenBank accession number: AB050723]), and for serotypes IV, V, and VI (McKinnon *et al.*, submitted into GenBank, 2001 [GenBank accession numbers: AF355776, AF349539, AF337958, respectively]) were released when the project was nearly finished but those for the other three serotypes (II, VII and VIII), the sequences of *cps* gene clusters, have not been published previously.

The published sequences of *cps* gene clusters for serotypes Ia and III showed considerable homology at the 3'-end of *cpsD-cpsE-cpsF*-and the 5'-end of *cpsG* (Chaffin *et al*., 2000; Yamamoto *et al*., 1999) and we hypothesised that these regions would also be relatively conserved in other serotypes. Our study and the recently released sequences for *cps* gene clusters of serotypes Ib, IV, V and VI have supported this hypothesis. We designed a series of primers to amplify a 2226 or 2217 bp segment in this region and found that amplicons were obtained from all serotypes except VIII (Chaffin *et al*., 2000; Yamamoto *et al*., 1999). This confirmed a previous suggestion that serotype VIII is significantly different from other serotypes in this region (Cieslewicz *et al*., 2001).

Using eight serotype (Ia to VII) reference strains, we showed more than 50 heterogeneity points between serotypes (Figure 3.1.). Using 63 selected clinical isolates that had been serotyped by conventional methods, we found that these inter-serotype differences were generally consistent and specific, especially the 23 sites clustered at the 3'-end of the regions (corresponding to the regions amplified by the primer pair cpsES3-cpsGA1). We used these differences to assign serotypes to the remaining clinical isolates collected in this study, without knowledge of the serotype obtained by conventional methods.

Sequence analysis of the 3'-end of *cpsG-cpsH-cpsI*/*cpsM* for serotypes Ia, III, Ib, IV, V and VI showed that this region is highly variable, and *cpsH* also contains serotype-specific definition sites (Chaffin ***et al***., 2000), making this region a suitable target for direct serotype identification by PCR. We designed MS-specific primers for MS Ia, Ib, III, IV, V and VI and confirmed their specificity, initially, using two CS reference panels. When used to test 206 clinical isolates, MS-specific PCR alone correctly identified 86.9%. Using rapid-cycle MS-specific PCR, results are available within one working day. In future, it will be possible to extend this method to all MS, when *cps* gene cluster sequences in this region are available for serotypes II, VII and VIII.

MS II and VII can be identified by sequencing the 790 bp PCR amplicons of the 3'-end of *cpsE-cpsF*-the 5'-end of *cpsG*. A positive GBS-specific PCR and negative PCR results with all the primers that amplify the 790 bp, identified MS VIII by exclusion. In future, and in some laboratories currently, sequencing of the 790 bp PCR amplicons of the 3'-end of *cpsE-cpsF*-and the 5'-end of *cpsG* for all isolates may be more convenient, as only one method and fewer primers are needed. However, if sequencing is not available in-house, the turn-around time is longer and a small proportion of serotypes would be wrongly assigned (subtypes III-3 and III-4 as MS Ia and II, respectively). This could be avoided by screening with MS III-specific PCR first. Sequencing the 790 bp PCR amplicon, allows MS III to be subtyped on the basis of the sequence heterogeneity.

Previous studies have shown that serotypes Ia, Ib, II, III and V are those most frequently isolated from normally sterile sites, in the United States and several countries (Hickman *et al*., 1999; Kalliola *et al*., 1999). Serotypes VI and VIII are the predominant serotypes isolated from patients in Japan (Lachenauer *et al*., 1999), but are uncommon elsewhere. Although our isolates were selected, they were probably representative of those causing disease in Australasia; Ia, Ib, II, III and V were the commonest serotypes identified, although there were small numbers of serotypes IV, VI and VIII.

Up to 13 % of GBS isolates are non-serotypable (Tyrrell *et al*., 2000); in our study the proportion was 8.7% (18/206) using antisera against all nine serotypes. Failure to react with anisera may be due to decreased type-specific-antigen synthesis (Palacios *et al*., 1997); non-encapsulated phase variation; or insertion or mutation in genes of *cps* gene clusters (Cieslewicz *et al*., 2001; Sellin *et al*., 2000). One non-serotypable strain GBS in our study had a T base deletion in *cpsG* gene, which caused a change in the *cpsG* reading frame.

In summary, we have developed an alternative to conventional serotyping for GBS, which is accurate and reproducible, can be performed by any laboratory with access to PCR and sequencing and, importantly, does not require panels of serotype-specific antisera that are increasingly difficult to maintain. All isolates are serotypeable and sequencing of a relatively limited 790 bp region can provide additional subtyping information for MS III. In future we will combine this with further PCR typing to identify members of the family of variable surface proteins that are important virulence factors for GBS and known GBS mobile genetic elements (to be reported separately). These extended typing methods will provide comprehensive strain identification that will be useful for epidemiological studies that will be needed to monitor GBS isolates before and after introduction of GBS conjugate vaccines.