

(19) 日本国特許庁(JP)

(12) 公表特許公報(A)

(11) 特許出願公表番号

特表2005-502381

(P2005-502381A)

(43) 公表日 平成17年1月27日(2005.1.27)

(51) Int.Cl. ⁷	F I	テーマコード (参考)
C 1 2 N 15/09	C 1 2 N 15/00	Z N A A 4 B O 2 4
C 1 2 Q 1/68	C 1 2 Q 1/68	A 4 B O 6 3
G O 1 N 33/53	G O 1 N 33/53	M
G O 1 N 33/569	G O 1 N 33/569	C
G O 1 N 37/00	G O 1 N 37/00	1 O 2
	審査請求 未請求 予備審査請求 有	(全 172 頁) 最終頁に続く

(21) 出願番号	特願2003-529988 (P2003-529988)	(71) 出願人	503334390 ウェスタン シドニー エリア ヘルス サービス
(86) (22) 出願日	平成14年9月18日 (2002. 9. 18)		
(85) 翻訳文提出日	平成16年3月19日 (2004. 3. 19)		
(86) 国際出願番号	PCT/AU2002/001281		オーストラリア国 2 1 4 5 エヌエスダ ブリュー, ウェストミード, ウェストミ ード ホスピタル
(87) 国際公開番号	W02003/025216	(74) 代理人	100102978 弁理士 清水 初志
(87) 国際公開日	平成15年3月27日 (2003. 3. 27)	(74) 代理人	100108774 弁理士 橋本 一憲
(31) 優先権主張番号	PR 7749	(74) 代理人	100128048 弁理士 新見 浩一
(32) 優先日	平成13年9月19日 (2001. 9. 19)		
(33) 優先権主張国	オーストラリア (AU)		

最終頁に続く

(54) 【発明の名称】 B群連鎖球菌の分子的分類法

(57) 【要約】

B群連鎖球菌の分子的分類方法およびこのような方法に有用なポリヌクレオチドを提供する。

【特許請求の範囲】**【請求項1】**

細菌のcpsD、cpsE、cpsF、cpsG、および/またはcpsI/M遺伝子内の1つまたは複数の領域のヌクレオチド配列を分析する工程を含み、該領域が型によって配列が異なる1つまたは複数のヌクレオチドを含む、B群連鎖球菌の分類方法。

【請求項2】

ヌクレオチド配列を、図1に記載の62位、78～86位、138位、139位、144位、198位、204位、211位、281位、240位、249位、300位、321位、419位、429位、437位、457位、466位、486位、602位、606位、627位、636位、645位、803位、971位、1026位、1044位、1173位、1194位、1251位、1278位、1413位、1495位、1500位、1501位、1512位、1518位、1527位、1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、1971位、2026位、2088位、2134位、2187位、および2196位に対応する1つまたは複数の位置について分析する、請求項1記載の方法。

10

【請求項3】

少なくとも1つの領域が、連鎖球菌のcpsE-cpsF-cspG遺伝子クラスターのcpsE遺伝子の3'末端の136塩基およびcpsG遺伝子の5'末端の218塩基で示される配列内にある、請求項1記載の方法。

【請求項4】

ヌクレオチド配列を、図1に記載の1413位、1495位、1500位、1501位、1512位、1518位、1527位、1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、1971位、2026位、2088位、2134位、2187位、および2196位に対応する1つまたは複数の位置について分析する、請求項3記載の方法。

20

【請求項5】

少なくとも1つの領域が、細菌のcpsI/M遺伝子内にある、請求項1～4のいずれか一項記載の方法。

【請求項6】

ヌクレオチド配列分析工程が、1つまたは複数の領域を配列決定する工程を含む、請求項1～5のいずれか一項記載の方法。

【請求項7】

ヌクレオチド配列分析工程が、細菌から得たポリヌクレオチドが1つまたは複数の領域を含むポリヌクレオチドプローブと選択的にハイブリッド形成するかどうかを決定する工程を含む、請求項1～5のいずれか一項記載の方法。

30

【請求項8】

細菌から得たポリヌクレオチドが1つまたは複数の領域に対応する複数のポリヌクレオチドプローブの1つまたは複数とハイブリッド形成するかどうかを決定する工程を含む、請求項7記載の方法。

【請求項9】

複数のポリヌクレオチドプローブがマイクロアレイとして存在する、請求項9記載の方法。

【請求項10】

ヌクレオチド配列分析工程が、1つまたは複数のプライマーを使用する増幅工程を含み、その少なくとも1つが型間で異なる配列と特異的にハイブリッド形成する、請求項1～5のいずれか一項記載の方法。

40

【請求項11】

ヌクレオチド配列分析工程が、プライマー対を使用する増幅工程を含み、その少なくとも1つが型間で異なる配列と特異的にハイブリッド形成する、請求項1～6のいずれか一項記載の方法。

【請求項12】

プライマーが、表2に記載のプライマーから選択される、請求項10または請求項11記載の方法。

50

【請求項13】

細菌のゲノム中における rib、alp2、または alp3 遺伝子から選択される1つまたは複数の表面タンパク質遺伝子の有無を決定する工程を含む、B群連鎖球菌の分類方法。

【請求項14】

表面タンパク質遺伝子の有無を決定する工程が、細菌から得たポリヌクレオチドが該表面タンパク質遺伝子の領域に対応するポリヌクレオチドプローブと選択的にハイブリッド形成するかどうかを決定する工程を含む、請求項13記載の方法。

【請求項15】

表面タンパク質遺伝子の有無を決定する工程が、該表面タンパク質遺伝子の領域を特異的に増幅する1つまたは複数のプライマーを使用した増幅工程を含む、請求項13記載の方法

10

【請求項16】

プライマーが、表6に記載のプライマーから選択される、請求項15記載の方法。

【請求項17】

細菌のゲノム中における rib、alp2、または alp3 遺伝子から選択される1つまたは複数の表面タンパク質遺伝子の有無を決定する工程をさらに含む、請求項1~12のいずれか一項記載の方法。

【請求項18】

細菌のゲノム中における IS861、IS1548、IS1381、ISSa4、および GBSi1 から選択される1つまたは複数の可動遺伝因子の有無を決定する工程を含む、B群連鎖球菌の分類方法。

20

【請求項19】

可動遺伝因子の有無を決定する工程が、細菌から得たポリヌクレオチドが該可動遺伝因子の領域に対応するポリヌクレオチドプローブに選択的にハイブリッド形成するかどうかを決定する工程を含む、請求項18記載の方法。

【請求項20】

可動遺伝因子の有無を決定する工程が、該可動遺伝因子の領域を特異的に増幅する1つまたは複数のプライマーを使用した増幅工程を含む、請求項18記載の方法。

【請求項21】

プライマーが、表10に記載のプライマーから選択される、請求項20記載の方法。

【請求項22】

細菌のゲノム中における IS861、IS1548、IS1381、ISSa4、および GBSi1 から選択される1つまたは複数の可動遺伝因子の有無を決定する工程をさらに含む、請求項13~17のいずれか一項記載の方法。

30

【請求項23】

本質的にB群連鎖球菌の cpsD-cpsE-cpsF-cpsG 遺伝子内の領域に対応する少なくとも10個の連続するヌクレオチドからなるポリヌクレオチドであって、B群連鎖球菌の血清型間で異なる1つまたは複数のヌクレオチドを含む、ポリヌクレオチド。

【請求項24】

B群連鎖球菌の血清型間で異なるヌクレオチドが、図1に記載の62位、78~86位、138位、139位、144位、198位、204位、211位、281位、240位、249位、300位、321位、419位、429位、437位、457位、466位、486位、602位、606位、627位、636位、645位、803位、971位、1026位、1044位、1173位、1194位、1251位、1278位、1413位、1495位、1500位、1501位、1512位、1518位、1527位、1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、1971位、2026位、2088位、2134位、2187位、および2196位の1つまたは複数に対応する、請求項23記載のポリヌクレオチド。

40

【請求項25】

本質的にB群連鎖球菌の cpsE-cpsF-cspG 遺伝子クラスターの cpsE の3'末端の136塩基対および cpsG の5'末端の218塩基対によって示される配列内の領域に対応する少なくとも10個の連続するヌクレオチドからなるポリヌクレオチドであって、B群連鎖球菌の型間で異なる1つまたは複数のヌクレオチドを含む、ポリヌクレオチド。

50

【請求項 26】

B群連鎖球菌の型間で異なるヌクレオチドが、図1に記載の1413位、1495位、1500位、1501位、1512位、1518位、1527位、1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、1971位、2026位、2088位、2134位、2187位、および2196位の1つまたは複数に対応する、請求項25記載のポリヌクレオチド。

【請求項 27】

本質的にB群連鎖球菌のcpsI/M遺伝子内の領域に対応する少なくとも10個の連続するヌクレオチドからなるポリヌクレオチドであって、連鎖球菌の血清型間で異なる1つまたは複数のヌクレオチドを含む、ポリヌクレオチド。

【請求項 28】

表2に記載のヌクレオチド配列から選択される、請求項27記載のポリヌクレオチド。

10

【請求項 29】

本質的にB群連鎖球菌のrib、alp2、またはalp3遺伝子内の領域に対応する少なくとも10個の連続するヌクレオチドからなるポリヌクレオチドであって、B群連鎖球菌の亜型間で異なる1つまたは複数のヌクレオチドを含む、ポリヌクレオチド。

【請求項 30】

表6に記載のヌクレオチド配列から選択される、請求項29記載のポリヌクレオチド。

【請求項 31】

B群連鎖球菌の血清型分類法および/または亜型分類法における、請求項23～30のいずれか一項記載のポリヌクレオチドの使用。

20

【請求項 32】

請求項23～30のいずれか一項記載の複数のポリヌクレオチドを含む組成物。

【請求項 33】

B群連鎖球菌の血清型分類法および/または亜型分類法における、請求項32記載の組成物の使用。

【請求項 34】

請求項23～30のいずれか一項記載の複数のポリヌクレオチドを含むマイクロアレイ。

【請求項 35】

B群連鎖球菌の血清型分類法および/または亜型分類法における、請求項34記載のマイクロアレイの使用。

30

【発明の詳細な説明】

【技術分野】

【0001】

発明の分野

本発明は、B群連鎖球菌の分子的分類方法およびこのような方法に有用なポリヌクレオチドに関する。

【背景技術】

【0002】

発明の背景

B群連鎖球菌 (GBS) -*Streptococcus agalactiae*-は、新生児および産科敗血症の最も一般的な原因であり、高齢の患者および免疫無防備状態の患者の敗血症の原因として重要性が増大している。新生児GBS敗血症の発生率は、分娩時抗生物質予防法の使用によって近年減少しているが、このアプローチに伴う問題は多数ある。今後、ワクチン接種が好ましいようであり、結合型 (conjugate) 多糖GBSワクチンの開発がかなりの進展を遂げている。

40

【0003】

結合型ワクチンの導入前に、疾病の負担だけでなく、ワクチン抗原の最適な処方を決定するために、GBS型 (莢膜多糖類遺伝子血清型、血清亜型; タンパク質抗原遺伝子の亜型; 可動遺伝子の亜型を含む) の分布を評価する、広範な疫学および他の関連研究が必要である。1つの地理的位置または少数の患者に基づいた型分布は、一般に適用不可能である。異なる地理的位置における異なる標的集団についてのGBSワクチン抗原の組み合わせ

50

の適合性を評価するために、継続的なモニタリングが必要である。

【0004】

9種の莢膜多糖類GBSの血清型が説明されている(Harrisonら、1998; Hickmanら、1999)。種々の血清型分類法(免疫沈降(Wilkinson and Moody、1969)、酵素免疫アッセイ(Holm and Hakansson、1988)、共凝集(Hakanssonら、1992)、対向免疫電気泳動法、および毛細管沈殿(Triscott and Davies、1979)、ラテックス凝集(Zuerleinら、1991)、蛍光顕微鏡法(Croppら、1974)、ならびに阻害ELISA(Arakereら、1999))などが使用されている。これらの方法は、骨の折れる作業であり、高価血清型特異的抗血清が必要であるが、これらは高価である上に作製が困難であり、たった6種の血清型(Ia~V)しか市販されていない(Arakereら、1999)。パルスフィールドゲル電気泳動(Rollandら、1999)、制限酵素分析(Naganoら、1991)などの分子的遺伝子型同定法は、疫学研究に有用であるが、一般に血清型は同定されない。したがって、GBS血清型同定のための信頼できる分子的方法が必要である。

10

【発明の開示】

【0005】

発明の概要

本発明者らは、異なる型(莢膜多糖類遺伝子血清型、血清亜型; タンパク質抗原遺伝子の亜型; 可動遺伝因子の亜型を含む)を識別するために使用することができるB群連鎖球菌ゲノム内の特異的な型間配列異種性領域を同定した。本発明者らは、これらの配列異種性を検出する分子的方法を使用して、B群連鎖球菌を正確に識別および分類できることを示す。

20

【0006】

したがって、第1の局面では、本発明は、細菌のcpsD、cpsE、cpsF、cpsG、cpsI/M遺伝子内の1つまたは複数の領域のヌクレオチド配列を分析する工程を含み、前記領域が型によって配列が異なる1つまたは複数のヌクレオチドを含む、B群連鎖球菌の分類方法を提供する。

【0007】

詳細には、ヌクレオチド配列を、図1に記載の62位、78~86位、138位、139位、144位、198位、204位、211位、281位、240位、249位、300位、321位、419位、429位、437位、457位、466位、486位、602位、606位、627位、636位、645位、803位、971位、1026位、1044位、1173位、1194位、1251位、1278位、1413位、1495位、1500位、1501位、1512位、1518位、1527位、1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、1971位、2026位、2088位、2134位、2187位、および2196位に対応する1つまたは複数の位置について分析することができる。

30

【0008】

好ましくは、少なくとも1つの領域は、B群連鎖球菌のcpsE-cpsF-cspG遺伝子クラスターのcpsE遺伝子の3'末端の136塩基およびcpsG遺伝子の5'末端の218塩基で示される配列内のものである。詳細には、ヌクレオチド配列を、図1に記載の1413位、1495位、1500位、1501位、1512位、1518位、1527位、1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、1971位、2026位、2088位、2134位、2187位、および2196位に対応する1つまたは複数の位置について分析することができる。

40

【0009】

1つの態様では、少なくとも1つの領域は、B群連鎖球菌のcpsI/M遺伝子内に存在する。

【0010】

本発明者らはまた、rib遺伝子、alp2遺伝子、またはalp3遺伝子を含む多数の表面タンパク質抗原遺伝子、および5個の可動遺伝因子を、GBSの分子的亜型分類に使用できることを示す。したがって、本発明はまた、細菌のゲノム中において、rib、alp2、またはalp3遺伝子から選択される1つまたは複数の表面タンパク質抗原遺伝子、ならびに/または、IS861、IS1548、IS1381、ISSa4、およびGBSi1から選択される1つまたは複数の可動遺伝因子の有無を決定する工程を含む、B群連鎖球菌の分類方法を提供する。好ましくは、このよう

50

な方法を、本発明の上記方法と組み合わせる。

【0011】

ヌクレオチド配列分析工程は、1つまたは複数の領域を配列決定する工程を含み得る。または、あるいはさらに、ヌクレオチド配列分析工程は、細菌から得たポリヌクレオチドが1つまたは複数の前記領域を含むポリヌクレオチドプローブ、好ましくは1つまたは複数の前記領域に対応する複数のポリヌクレオチドプローブの1つまたは複数と選択的にハイブリッド形成するかどうかを決定する工程を含み得る。

【0012】

好ましい態様では、分析手段として複数のプローブへのハイブリッド形成を使用する場合、複数のポリヌクレオチドプローブがマイクロアレイとして存在する。

10

【0013】

別の態様では、ヌクレオチド配列分析工程は、1つまたは複数のプライマーを使用し、少なくともその1つが型間で異なる配列と特異的にハイブリッド形成する増幅工程を含む。典型的には、少なくとも1つが型間で異なる配列と特異的にハイブリッド形成するプライマー対を使用する。好ましくは、プライマーは、表2および/または表6および/または表10に記載のプライマーから選択される。

【0014】

第2の局面では、本発明は、本質的にB群連鎖球菌のcpsD-cpsE-cpsF-cpsG遺伝子内の領域に対応する少なくとも10個の連続するヌクレオチドからなるポリヌクレオチドであって、GBSの型間で異なる1つまたは複数のヌクレオチドを含む、ポリヌクレオチドを提供する。

20

【0015】

好ましくは、GBS型間で異なるヌクレオチドは、1つまたは複数の図1に記載の62位、78~86位、138位、139位、144位、198位、204位、211位、281位、240位、249位、300位、321位、419位、429位、437位、457位、466位、486位、602位、606位、627位、636位、645位、803位、971位、1026位、1044位、1173位、1194位、1251位、1278位、1413位、1495位、1500位、1501位、1512位、1518位、1527位、1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、1971位、2026位、2088位、2134位、2187位、および2196位に対応する。

【0016】

本発明はまた、本質的にB群連鎖球菌のcpsE-cpsF-cspG遺伝子クラスターのcpsEの3'末端の136塩基対およびcpsGの5'末端の218塩基対によって示される配列内の領域に対応する少なくとも10個の連続するヌクレオチドからなるポリヌクレオチドであって、GBS型間で異なる1つまたは複数のヌクレオチドを含む、ポリヌクレオチドを提供する。

30

【0017】

好ましくは、B群連鎖球菌の型間で異なるヌクレオチドは、1つまたは複数の図1に記載の1413位、1495位、1500位、1501位、1512位、1518位、1527位、1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、1971位、2026位、2088位、2134位、2187位、および2196位に対応する。

【0018】

本発明はまた、本質的にB群連鎖球菌のcpsI/M遺伝子内の領域に対応する少なくとも10個の連続するヌクレオチドからなるポリヌクレオチドであって、B群連鎖球菌の型間で異なる1つまたは複数のヌクレオチドを含む、ポリヌクレオチドを提供する。

40

【0019】

好ましくは、ポリヌクレオチドは、表2に記載のヌクレオチド配列から選択される。

【0020】

本発明は、さらに、本質的にB群連鎖球菌のrib、alp2、またはalp3遺伝子内の領域に対応する少なくとも10個の連続するヌクレオチドからなるポリヌクレオチドであって、GBSタンパク質抗原の遺伝子型間で異なる1つまたは複数のヌクレオチドを含む、ポリヌクレオチドを提供する。

【0021】

50

好ましくは、ポリヌクレオチドは、表6に記載のヌクレオチド配列から選択される。

【0022】

本発明は、さらに、本質的にB群連鎖球菌のIS861、IS1548、IS1381、ISSa4、および/またはGBSi1内の領域に対応する少なくとも10個の連続するヌクレオチドからなるポリヌクレオチドであって、GBS可動遺伝因子の亜型間で異なる1つまたは複数のヌクレオチドを含む、ポリヌクレオチドを提供する。

【0023】

好ましくは、ポリヌクレオチドは、表10に記載のヌクレオチド配列から選択される。

【0024】

本発明のポリヌクレオチドを、B群連鎖球菌の血清型分類および/または亜型分類などの分類方法に使用することができる。 10

【0025】

第3の局面では、本発明は、複数の本発明の第2の局面のポリヌクレオチドを含む組成物を提供する。本組成物は、B群連鎖球菌の血清型分類および/または亜型分類などの分類方法に使用することができる。

【0026】

第4の局面では、本発明は、複数の本発明の第2の局面のポリヌクレオチドを含むマイクロアレイを提供する。本マイクロアレイは、B群連鎖球菌の血清型分類および/または亜型分類などの分類方法に使用することができる。

【0027】

発明の詳細な説明

他に定義しない限り、本明細書中で使用される全ての技術用語および科学用語は、当業者（例えば、細胞培養、分子遺伝学、核酸化学、ハイブリッド形成技術、および生化学）に一般的に理解されている意味を有する。分子的方法、遺伝子的方法、および生化学的方法（全般的に、Sambrookら、「Molecular Cloning: A Laboratory Manual」、第3版、2001、Cold Spring Harbor Laboratory Press、Cold Spring Harbor、N.Y.およびAusubelら、「Short Protocols in Molecular Biology」、1999、第4版、John Wiley & Sons, Inc.、およびタイトル「Current Protocols in Molecular Biology」の全版（参照として本明細書に組み入れられる）を参照のこと）ならびに化学的方法についての標準的技術を使用する。 20

【0028】

本発明の分子的分類法は、サンプル中、本発明者らが種々の型間で異なることを同定したB群連鎖球菌（GBS）のゲノム領域中の特定のポリヌクレオチド配列の存在を検出することに依存する。 30

【0029】

より詳細には、検出すべき特定のポリヌクレオチド配列は、GBSのcpsD、cpsE、cpsF、cpsG、cpsI、cpsM、rib、alp2、および/またはalp3遺伝子、ならびに可動遺伝因子IS861、IS1548、およびIS1381、ISSa4、およびGBSi1、好ましくは、cpsD、cpsE、cpsF、cpsG、および/またはcpsI/M遺伝子内のものである。

【0030】

上記遺伝子の範囲内の目的の領域は、その配列が2つまたはそれ以上の型間で変化する（すなわち、異種性）領域である。異種性は、異なる型の対応する領域間の挿入、欠失、および/または置換に起因し得る。rib、alp2、およびalp3の場合、異種性は、典型的には、全遺伝子の有無の形態をとる。同様に、因子IS861、IS1548、IS1381、ISSa4、およびGBSi1については、異種性は、典型的には、全配列の有無の形態をとる。 40

【0031】

特定の異種性領域には、以下の位置が含まれる。cpsD遺伝子-62位および78~86位；cpsD-cpsE遺伝子スペーサー-138位、139位、および144位；cpsE遺伝子-198位、204位、211位、281位、240位、249位、300位、321位、419位、429位、437位、457位、466位、486位、602位、606位、627位、636位、645位、803位、971位、1026位、1044位、1173位、1194位、12 50

51位、1278位、1413位、1495位、1500位、1501位、1512位、1518位、および1527位；cpsF遺伝子-1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、および1971位；ならびにcpsG遺伝子-2026位、2088位、2134位、2187位、および2196位（番号付けは図1に記載の番号付けに対応する）。

【0032】

特に好ましい目的の位置は、cpsE-cpsF-cpsGの790bp断片（およそcpsEの3'末端の136塩基およびcpsGの5'末端の218塩基からなる）内の位置、すなわち、図1に記載の1413位、1495位、1500位、1501位、1512位、1518位、1527位、1595位、1611位、1620位、1627位、1629位、1655位、1832位、1856位、1866位、1871位、1892位、1971位、2026位、2088位、2134位、2187位、および2196位である。

10

【0033】

別の異種性領域は、cpsDの62位、およびcpsDの78位から86位で見出される反復配列（TTACGGCGA）であるが、これはいくつかのGBS血清型におけるものであって、全てではない。

【0034】

特定の異種性領域には、図3の配列アラインメントに記載のcpsI/M遺伝子内の多数の位置も含まれる。

【0035】

これらの異種性領域を、配列決定、PCR、および標識プローブの結合を含む種々の手段を使用して分析することができる。

【0036】

血清型を同定するための配列決定の場合、異種性領域が存在する領域内またはその付近と特異的にハイブリッド形成するように配列決定プライマーを選択する。分子的血清型の割り当てに使用する配列決定プロセス時に実際の配列情報が得られるので、プライマーは特定の血清型に特異的である必要はない。したがって、プライマーは、全てのGBS血清型（少なくとも血清型IaからVII）または特定の血清型と特異的にハイブリッド形成することができる。

20

【0037】

好ましいプライマーは、図1に記載のcpsE-cpsF-cpsGの790bp領域内の異種性位置の100、50、または20個の連続するヌクレオチド内にアニーリングする。適切な配列決定プライマーの例を、表2に示す（cspES3、cspFA、cspFS、cspGA、およびcspGA1）。

30

【0038】

PCRおよび他の特異的ハイブリッド形成ベースの血清型分類法は、典型的には、2つまたはそれ以上の血清型間で異なるヌクレオチドを含むGBS血清型のゲノム領域に特異的に結合するヌクレオチドプライマー/プローブの使用を含む。したがって、プライマー/プローブは、このような領域の1つに相補的な配列を含み得る。異種性位置が互いに近接している場合（例えば、cpsEの198位、204位、211位、および218位）、2つまたはそれ以上の異種性位置を含むGBSゲノム領域に特異的にハイブリッド形成するプライマー/プローブを使用することが望ましい。したがって、例えば、cpsEのヌクレオチド195から220に相補的なプライマー/プローブをデザインすることができる。このようなプライマー/プローブは、特異性を改善し、偽陽性の可能性を減少させる可能性が高い。

40

【0039】

PCRベースの検出方法は、少なくとも1つが1つまたは複数であるが全てではない血清型中の目的の領域に特異的に結合するプライマー対の使用に依存し得る。両プライマーが結合しない限り、PCR産物は得られない。したがって、特異的PCR産物の有無を使用して、特定のGBS血清型を示す配列の存在を決定することができる。しかし、述べたように、目的の遺伝子（cpsD、cpsE、cpsF、cpsG、cpsIおよび/またはcpsM遺伝子など）中の異種性領域にたった1つのプライマーが対応する必要がある。他のプライマーは、前記遺伝子内の保存もしくは異種性領域、またはcpsH遺伝子などのGBSゲノムの別の部分内の領域さえも、血清型間で保存されているもしくは異種性のいずれにせよ、結合することができる。したがって、例えば、2172位から2210位を含むcpsG遺伝子領域に結合するプライマー（cpsGS

50

)と異種性のcpsH遺伝子領域 (IacpsHA1、IIIcpsHA) に結合するプライマーとの組み合わせを、血清型 (IaおよびIII) 識別の基本として使用することができる。

【0040】

さらに、異種性cpsI領域に結合するプライマーを、不変のcpsG領域に結合するプライマーと組み合わせることができる。そのようなプライマー対の例は、1517bpのPCR産物を生じ、且つGBS血清型VI特異的である、プライマー対VIcpsIAおよびcpsGS1である。

【0041】

または、GBSゲノムの保存領域に結合するが、血清型間で長さが異なる領域をはさむ、プライマーを使用することができる。この場合、GBS細菌が存在するがPCR産物のサイズが血清型間で異なる場合にPCR産物が常に得られる。

10

【0042】

さらに、一方または両方のプライマーの特異的結合と種々の長さのPCRプライマーとの組み合わせを、特定の分子的血清型の同定手段として使用することができる。

【0043】

cpsD、cpsE、cpsF、cpsG、cpsI、またはcpsM遺伝子をターゲティングする特異的プライマー/プローブの例には、以下のものが含まれる。

cpsDS GCA AAA GAA CAG ATG GAA CAA AGT GG
 cpsES CTT TTG GAG TCG TGG CTA TCT TG
 cpsEA1 GA/T/GA AAA AAG GAA AGT CGT GTC G/ATT G
 cpsES1 CTT GGA C/TTC CTC TGA AAA GGA TTG
 cpsEA2 AAA A/CGC TTG ATC AAC AGT TAA GCA GG
 cpsES2 GAT GGT/C GGA CCG GCT ATC TTT TCT C
 cpsEA3 CTT AAT TTG TTC TGC ATC TAC TCG C
 cpsES3 GTT AGA TGT TCA ATA TAT CAA TGA ATG GTC TAT TTG GTC AG
 cpsEFA CCT TTC AAA CCT TAC CTT TAC TTA GC
 cpsFS CAT CTG GTG CCG CTG TAG CAG TAC CAT T
 cpsFA GTC GAA AAC CTC TAT A/GT A AAC/T GGT CTT ACA A/GCC AAA
 TAA CTT ACC
 cpsGA AAG/C AGT TCA TAT CAT CAT ATG AGA G
 cpsGA1 CCG CCA/G TGT GTG ATA ACA ATC TCA GCT TC
 cpsGS ATG ATG ATA TGA ACT CTT ACA TGA AAG AAG CTG AGA TTG
 cpsGS1 GAA CTC TTA CAT GAA AGA AGC TGA GAT TGT TAT CAC AC
 IbcpsIA CTA TCA ATG AAT GAG TCT GTT GTA GGA CGG ATT GCA CG
 IbcpsIS GAT AAT AGT GGA GAA ATT TGT GAT AAT TTA TCT CAA AAA
 GAC G
 IbcpsIA1 CCT GAT TCA TTG CAG AAG TCT TTA CGA TGC GAT AGG TG
 IVcpsMA GGG TCA ATT GTA TCG TCG CTG TCA ACA AAA CCA ATC AAA TC
 VcpsMA CCC CCC ATA AGT ATA AAT AAT ATC CAA TCT TGC ATA GTC AG
 VIcpsIA GAA GCA AAG ATT CTA CAC AGT TCT CAA TCA CTA ACT CCG
 cpsIA GTA TAA CTT CTA TCA ATG GAT GAG TCT GTT GTA GTA CGG

20

30

40

プライマーの命名は、表2に記載のものに対応する。

【0044】

alp2、alp3、およびrib表面タンパク質抗原遺伝子に関して、異種性およびタンパク質抗原遺伝子の亜型を、B群連鎖球菌が遺伝子を含むかどうかのレベルでさらに評価する。本発明者らの結果は、GBSゲノム中に存在する表面タンパク質遺伝子の特定の組み合わせが血清型/血清亜型の指標であることを示す(表9を参照のこと)。したがって、本発明の方

50

法での使用に適切なプライマー/プローブは、特定の遺伝子に特異的なものである。したがって、alp2またはalp3またはribに特異的なプローブ/プライマーが好ましい。図4は、alp2に特異的またはalp3に特異的なプライマーをデザインするために使用したalp2およびalp3のアラインメントを示す。

【 0 0 4 5 】

alp2、alp3、およびrib遺伝子をターゲティングする特異的プライマー/プローブの例には、以下のものが含まれる。

bcaS1 GGT AAT CTT AAT ATT TTT GAA GAG TCA ATA GTT GCT GCA TCT
AC

bcaS2 CCAGGGA GTG CAG CGA CCT TAA ATA CAA GCA TC

balS GAT CCT CAA AAC CTC ATT GTA TTA AAT CCA TCA AGC TAT TC

balA CCA GTT AAG ACT TCA TCA CGA CTC CCA TCA C

bal23S1 CAG ACT GTT AAA GTG GAT GAA GAT ATT ACC TTT ACG G

bal23S2 CTT AAA GCT AAG TAT GAA AAT GAT ATC ATT GGA GCT CGT G

bal2S CTT CCG CCA GAT AAA ATT AAG

bal2A CTG TTG ACT TAT CTG GAT AGG TC

bal2A1 CGT GTT GTT CAA CAG TCC TAT GCT TAG CCT CTG GTG

bal2A2 GGT ATC TGG TTT ATG ACC ATT TTT CCA GTT ATA CG

bal3S GTT CTT CCG CTT AAG GAT AG

bal3A GAC CGT TTG GTC CTT ACC TTT TGG TTC GTT GCT ATC C

ribS2 GAAGTAATTCAG GAA GTG CTG TTA CGT TAA ACA CAA ATA TG

ribA1 GAA GGT TGT GTG AAA TAA TTG CCG CCT TGC CTA ATG

ribA2 AAT ACT AGC TGC ACC AAC AGT AGT CAA TTC AGA AGG

10

20

プライマーの命名は、表6に記載のものに対応する。

【 0 0 4 6 】

IS861、IS1548、IS1381、ISSa4、およびGBSi1に関して、異種性および亜型を、B群連鎖球菌が因子を含むかどうかのレベルでさらに評価する。因子数を評価することもできる。本発明者らの結果は、GBSゲノム中に存在する可動因子の特定の組み合わせが血清型/血清亜型の指標となることを示す(表12を参照のこと)。したがって、本発明の方法での使用に適切なプライマー/プローブは、特定の可動遺伝因子に特異的なものである。したがって、IS861、IS1548、IS1381、ISSa4、およびGBSi1に特異的なプローブ/プライマーが好ましい。

30

【 0 0 4 7 】

IS861、IS1548、IS1381、ISSa4、およびGBSi1をターゲティングする特異的プライマー/プローブの例には、以下のものが含まれる。

IS861S GAG AAA ACA AGA GGG AGA CCG AGT AAA ATG GGA CG
 IS861A1 CAC GAT TTC GCA GTT CTA AAT AAA TCC GAC GAT AGC C
 IS861A2 CAA ACT CCG TCA CAT CGG TAT AGC ACT TCT CAT AGG
 IS1548S CTA TTG ATG ATT GCG CAG TTG AAT TGG ATA GTC GTC
 IS1548S1 GTT TGG GAC AGG TAG CGG TTG AGG AGA AAA GTA ATG
 IS1548A1 CAT TAC TTT TCT CCT CAA CCG CTA CCT GTC CCA AAC
 IS1548A2 CCC AAT ACC ACG TAA CTT ATG CCA TTT G
 IS1548A3 CGT GTT ACG AGT CAT CCC AAT ACC ACG TAA CTT ATG CC
 IS1381S1 CTT ATG AAC AAA TTG CGG CTG ATT TTG GCA TTC ACG
 IS1381S2 GGC TCA GGC GAT TGT CAC AAG CCA AGG GAG
 IS1381A CTA AAA TCC TAG TTC ACG GTT GAT CAT TCC AGC
 ISSa4S CGT ATC TGT CAC TTA TTT CCC TGC GGG TGT CTC C
 ISSa4A1 GCC GAT GTC ACA ACA TAG TTC AGG ATA TAG CCA G
 ISSa4A2 CGT AAA GGA GTC CAA AGA TGA TAG CCT TTT TGA ACC
 GBSi1S1 CAT CTC GGA ACA ATA TGC TCG AAG CTT ACA AGC AAG TG
 GBSi1S2 GGG GTC ACT ATC GAG CAG ATG GAT GAC TAT CTT CAC
 GBSi1A1 AAT GGC TGT TTC GCA GGA GCG ATT GGG TCT GAA CC
 GBSi1A2 CCA GGG ACA TCA ATC TGT CTT GCG GAA CAG TAT CG

10

20

【0048】

好ましくは、プライマー/プローブは、少なくとも10、15、または20個のヌクレオチドを含む。典型的には、プライマー/プローブは、100、50、または30個未満のヌクレオチドからなる。プライマー/プローブは、一般に、デオキシヌクレオチドを含むポリヌクレオチドである。これらはまた、合成または修飾ヌクレオチドを含むポリヌクレオチドであり得る。多数の異なる型のオリゴヌクレオチド修飾法が当技術分野において公知である。これらには、メチルホスホネートおよびホスホロチオエート骨格、分子の3'および/または5'末端でのアクリジンまたはポリリジン鎖の付加が含まれる。本発明の目的のために、本明細書中に記載のポリヌクレオチドは、当技術分野において利用可能な任意の方法によって修飾してもよいと理解すべきである。プライマー/プローブを、放射性元素、蛍光分子、またはビオチンなどの任意の適切な検出可能な標識で標識することができる。

30

【0049】

1つの態様では、プライマー/プローブは、急速サイクル (rapid cycle) PCRで使用することができるように、高い融点 (>70)を有する。

【0050】

cpsD、cpsE、cpsF、cpsG、またはcpsI/M遺伝子内の1つまたは複数の領域を分析するために使用される複数のヌクレオチドを含む組成物はまた、cpsH遺伝子内の1つまたは複数の領域を分析するために使用することができるヌクレオチドをさらに含むことができる。実施例に適切なヌクレオチドを記載するが、当業者は、図3に記載の配列アラインメントに基づいて他の適切な配列をデザインすることができる。

40

【0051】

さらに、alp2、alp3、またはrib遺伝子内の1つまたは複数の領域を分析するために使用される複数のヌクレオチドを含む組成物はまた、C (bca) 遺伝子およびC (bac) 遺伝子 (C 遺伝子はbagとしても公知) 内の1つまたは複数の領域を分析するために使用することができるヌクレオチドをさらに含むことができる。

【0052】

種々の技術を使用して、目的の細菌のゲノム内の1つまたは複数の領域を分析することができる。典型的には、GBS細菌を含むと疑われる目的のサンプルを、サンプル中に存在する任意の微生物からゲノムDNAを得るための標準的技術を使用して処理する。多数のその

50

後の検出工程のために、ゲノムDNAを実質的に断片化する核酸調製技術を使用することが望ましい場合がある。サンプルは、細菌培養物または患者（典型的には、ヒト患者）由来の臨床サンプル由来であり得る。臨床サンプルは細菌培養物を産生するように培養され得る。しかし、培養工程を直接利用して臨床サンプルを試験することも可能である。

【0053】

次いで、ゲノムDNAを、配列決定、酵素的増幅、および/またはハイブリッド形成を含み得る1つまたは複数の分析工程に供する。これらの一般的なDNA分析技術は当技術分野において公知であり、例えば、Sambrookら、2001およびAusubelら、1999、前出で詳細に考察されている。

【0054】

血清型分類は、1つまたは複数の工程を含み得る。例えば、サンプル中にGBS血清型間で保存されているがいかなる他の生物でも見出されないヌクレオチド配列が存在するかどうかを決定する最初の工程を行うことが望ましい場合がある。任意（しかし唯一）のGBS細菌を検出するPCRプライマーの使用によってこれを行うことができる（例えば、プライマー対Sag59/Sag190および/またはDSF2/DSR1を使用-表2および3を参照のこと）。

【0055】

次いで、上記で考察したプローブ/プライマーに基づく配列決定、酵素的増幅、および/またはハイブリッド形成などの任意の適切な技術を使用して目的の領域中の1つまたは複数の異種性領域の存在を検出することによって、特異的GBS血清型の分子的血清型分類を行うことができる。

【0056】

特に好ましい検出技術は、急速サイクルPCR（Kongら、2000）などのPCRである。

【0057】

多工程血清型分類ストラテジー（アルゴリズム）の例を、図2に示す。しかし、種々の他のストラテジーが考えられ、当業者は本明細書中に記載の配列異種性情報を使用してデザインすることができる。特に、血清型分類手順は、cpsD、cpsE、cpsF、cpsG、および/またはcpsI/M遺伝子の1つまたは複数の領域の分析に基づいた少なくとも1つの分析工程を含むことが好ましい。この分析を、cpsH遺伝子内の1つまたは複数の領域の分析と選択的に組み合わせることができる。類似の技術を使用して、cpsH遺伝子領域を分析することができ、適切なプライマー配列および方法も、実施例に記載されている。

【0058】

実施例に記載のように、alp2、alp3、および/またはrib遺伝子の有無の分析を、C（bca遺伝子）、C（bac）遺伝子配列の有無の分析と選択的に組み合わせることができる。類似の技術を使用してこれらの領域を分析することができ、適切なプライマー配列およびPCR法もまた実施例に記載されている。

【0059】

さらに、alp2、alp3、および/またはrib遺伝子（選択的に、bcaおよびbac遺伝子）の有無の分析を、可動遺伝因子の有無の分析と組み合わせることができる。

【0060】

したがって、分類ストラテジーは、より多くの血清型分類および型分類情報を得るための種々の組み合わせにおけるcps遺伝子、表面タンパク質遺伝子、および/または可動遺伝因子の分析を含み得る。

【0061】

上記技術を使用したGBSゲノム配列の分析を溶液中で行い、その後ゲル電気泳動などの方法を使用した標準的分離を行うことができる。しかし、本発明の好ましい局面では、アレイを形成するためにプライマー/プローブを固体基質に固定する。

【0062】

ポリヌクレオチドプローブは、典型的には、固体基質の別々の領域上または領域中に固定される。基質は、基質内への固定のために多孔質であっても実質的に無孔であってもよく、この場合プローブは典型的には、基質表面に固定される。適切な固体基質の例には、板

10

20

30

40

50

ガラス（ホウケイ酸ガラスなど）、シリコンウェファ、マイカ、セラミックス、ならびにポリスチレンおよびポリメタクリレートを含むプラスチックなどの有機ポリマーが含まれる。広範に利用可能なニトロセルロースまたはナイロンメンブレンなどの半透膜を使用することも可能である。半透膜を、ガラスなどのより強固な固体表面に置くことができる。表面を、金、白金、または他の遷移金属などの金属層で選択的にコートすることができる。

【0063】

好ましくは、固体基質は、一般に、剛性または半剛性表面を有する材料である。好ましい態様では、少なくとも1つの基質表面は実質的に平面であるが、いくつかの態様では、例えば、隆起した領域または窪んだ溝を有する異なるポリマーのための合成領域を物理的に分離することが望ましい。固体基質が典型的には10000から40000 cm^{-2} の密度が得られる50から100 μm の個別の領域中のDNA配列の高密度適用に適していることも好ましい。

10

【0064】

固体基質は、都合よくセクションに分けられる。フォトエッチングなどの技術によって、または疎水性インク、例えばテフロンベースのインク（Cel-line、USA）の適用によって、これを行うことができる。それぞれの異なるプローブが配置される個別の位置は、任意の都合の良い形状（例えば、円形、長方形、楕円形、V字型など）を有することができる。

【0065】

ライブラリー配列は、共有的または非共有的手段によって基質へ結合しうる。ライブラリー配列を、ライブラリー配列が結合する分子層を介して基質に結合することができる。例えば、プローブをビオチンで標識し、基質をアビジンおよび/またはストレプトアビジンでコードすることができる。ビオチン化プローブ使用の都合の良い特徴は、固体基質へのカップリング効率を容易に決定することができるという点である。ポリヌクレオチドプローブがいくつかの固体基質にはあまり結合できないことがあるので、しばしば固体基質（ガラスの場合など）とプローブとの間に化学的境界面を提供する必要がある。したがって、基質表面を、例えば、疎水性を増減する化学物質でのコーティングまたはポリヌクレオチドプローブを共有結合させる化学物質でのコーティングによって調製することができる。いくつかの化学コーティングにより、疎水性を変化させ共有結合をさせることができる。固体基質上の疎水性を、シラン処理または当技術分野において公知の他の処理によって容易に増大することができる。適切な化学コーティングの例には、ポリリジンおよびポリ（エチレンジアミン）が含まれる。結合方法のさらなる詳細は、米国特許第6,248,521号に記載されている。分子への種々の官能基の導入による核酸の固定方法はまた、Bischoffら、1987（*Anal. Biochem.*、164、336～344およびKremskyら、1987（*Nucl. Acids Res.*、15、2891～2910）に記載されている。

20

30

【0066】

固定化核酸分子アレイの作製技術は、当技術分野において説明されている。有用な概説は、Schenら、1998、*TibTech*、16、301～306（これに記載の技術についての参考文献も記載されている）に記載されている。

【0067】

マイクロアレイ製造テクノロジーは、主に2つのカテゴリー-合成および送達に分類される。合成アプローチでは、マイクロアレイを、生化学的基礎単位からの核酸のインサイチュール合成による段階的様式で調製する。各合成ラウンドで、ヌクレオチドを付加して所望の長さに達するまで鎖を伸長させる。多数の先行技術の方法が、例えば、固体基質上の種々の個別の位置で配列を様々に並べかえるためのマスキング技術（フォトリソグラフィ）を使用した、インサイチュールでの一本鎖核酸分子ライブラリーの合成方法を記載している。米国特許第5,837,832号は、非常に大規模な組み込みテクノロジーに基づいたシリコン基質に固定化したDNAアレイの改良された作製方法を記載している。特に、米国特許第5,837,832号は、本発明の固定化DNAライブラリーの作製に使用することができる基質上の空間的に定義された位置に特定のプローブ組を合成するための「タイリング（tiling）」と呼

40

50

ばれるストラテジーを記載している。米国特許第5,837,832号はまた、使用することができる初期の技術についての参考文献も記載している。

【0068】

対照的に、送達テクノロジーは、チップ製造のためにあらかじめ調製された生化学物質の外因的配置を使用する。例えば、DNAを、ピン（機械的マイクロスポッティング）または圧電性デバイス（インクジェット）のいずれかを具備したロボットデバイスを使用して基質上に直接プリントすることもできる。機械的マイクロスポッティングでは、生化学的サンプルを、毛管現象によってスポッティングピンにロードし、ピンと固体基質との間の物理的接触によって、少量を固体表面に導入する。第1のスポッティングサイクル後、ピンを洗浄し、第2のサンプルをロードし、隣接するアドレスに配置させる。ロボット制御系および多重プリントヘッドにより、マイクロアレイ製造を自動化できる。インクジェットは、圧電付属品を具備した小型ノズルにポリヌクレオチドなどの生化学的サンプルをロードする工程を含み、電流を使用してジェットから基質へ正確な量の液体を発射する。第1のジェット工程後、ジェットを洗浄し、第2のサンプルをロードし、隣接するアドレスに配置させる。複数のジェットで一連のサイクルを繰り返すことにより、迅速なマイクロアレイ作製が可能である。

10

【0069】

1つの態様では、マイクロアレイは、約50個、好ましくは約100個または200個を超える異なる核酸プローブを含む高密度アレイである。このような高密度プローブは、 cm^2 あたり約50個、好ましくは約500個、より好ましくは約1000個、最も好ましくは約2000個を超える異なる核酸プローブのプローブ密度を含む。アレイは、さらに、ミスマッチコントロールプローブおよび/または基準プローブ（正のコントロールなど）を含み得る。

20

【0070】

本発明のマイクロアレイは、典型的には、複数の上記のプライマー/プローブを含む。プライマー/プローブは、任意の順序でアレイにグループ化することができる。しかし、特異的な型（莢膜多糖類遺伝子血清型、血清亜型；タンパク質抗原遺伝子の亜型；可動遺伝因子の亜型）または型（莢膜多糖類遺伝子血清型、血清亜型；タンパク質抗原遺伝子の亜型；可動遺伝因子の亜型）の群によってプライマー/プローブをグループ化することが望ましい場合がある。得られるパターンがコンピュータソフトウェアによるパターン認識に容易に影響を受けるように、このようなグループ化を行うことができる。

30

【0071】

アレイ中の因子は、たった1つのプローブ/プライマー型または多数の異なるプローブ/プライマーを含み得る。

【0072】

多数の技術を使用して、固定化プローブ/プライマーへのGBSゲノムDNAの結合を検出することができる。例えば、多数の型（莢膜多糖類遺伝子血清型、血清亜型；タンパク質抗原遺伝子の亜型；可動遺伝因子の亜型）に特異的な固定化プローブは、捕捉プローブとして機能することができる。アレイへのゲノムDNAの結合後、アレイを洗浄し、保存されたGBSゲノム領域と特異的にハイブリッド形成する1つまたは複数の標識検出プローブとインキュベートする。次いで、これらの検出プローブの結合を、標識の存在の検出によって決定することができる。例えば、標識は、蛍光標識であってよく、アレイを、電荷結合素子（CCD）カメラ下のX-Yリーダーに置くことができる。

40

【0073】

他の技術には、アレイとの接触前のゲノムDNAの標識が含まれる（例えば、ニックトランスレーションおよび標識dNTPを使用する）。次いで、ゲノムDNAの結合を、直接検出することができる。

【0074】

標識dNTPを使用した単一のPCR増幅工程を使用することも可能である。この態様では、ゲノムDNAフラグメントは、アレイ中の第1のプライマーに結合する。ポリメラーゼ、いくつかの標識dNTPを含むdNTP、および第2のプライマーの添加により、標識ヌクレオチドが組

50

み込まれたPCR産物が合成される。次いで、プレート上に捕捉された標識PCRフラグメントを検出することができる。

【0075】

多数の利用可能な検出技術は、標識は必要ないが、その代わりにリガンド結合時の質量の変化に依存する（例えば、表面プラズモン共鳴-SPR）。SPRの原理およびSPRの使用に必要な固体基質の型（例えば、BIACoreチップ）は、Ausubelら、1999、前出に記載されている。

【0076】

C.使用

上記のように、B群連鎖球菌（GBS）-*Streptococcus agalactiae*は、新生児および産科敗血症の最も一般的な原因であり、高齢の患者および免疫無防備状態の患者の敗血症の原因としての重要性が増大している。したがって、本発明の検出方法、プローブ/プライマー、およびマイクロアレイを、妊婦、高齢者、および/または免疫無防備状態の患者のGBS感染の診断で使用することができる。本明細書中に記載のPCRおよびマイクロアレイ技術は、従来の同定および血清型分類と比較して精度および感度が高く、妊婦の日常的な出生前スクリーニングおよび妊婦の感染診断で特に使用してもよい。これらの方法はまた、一般に十分な材料を得るために臨床サンプルを培養する必要がないので、より早く結果が得られる可能性が高い。さらに、ほとんどの研究所で専門的知識または試薬を必要としないで分子的技術を使用することができる。

【0077】

本発明の分子的分類法はまた、GBS結合型ワクチンの導入前後にGBS単離物をモニターする必要がある疫学的研究および他の関連研究に有用な包括的株同定を補助することができる。

【0078】

本発明は、ここに、以下の実施例を参照してより詳細に記載されるが、これは例示のみを目的として本発明を限定しない。実施例は、図面を参照する。

【0079】

実施例

材料と方法

GBS基準株および臨床単離物

9個のGBS血清型のパネル（Ia~VIII）を、Lawrence Paoletti博士（Channing Laboratory、Boston USA）からご提供いただいた（基準パネル1）。Diana Martin博士（*Streptococcus* Reference Laboratory、ESR、Wellington、New Zealand）から血清型Ia~VIを含む9個の国際基準GBS型株の別のパネル（基準パネル2）を入手した（表1）。さらに、本発明者らは、血清型分類のためにニュージーランドの種々の研究所に参照された146種を含む205臨床例からの単離物およびシドニーの1つの診断研究所の10年間にわたる正常な無菌部位由来の59種の単離物を試験した。1つの培養物が混合されていることが続いて示され、206種の異なる単離物を試験した。従来の血清型分類（CS）は、*Streptococcus* Reference Laboratory、ESR、Wellington、New Zealandで行い、MSは、Centre for Infectious Diseases and Microbiology Laboratory Services、ICPMR、Sydney、Australiaで行った。

【0080】

GBS基準株の2つのパネルおよび63種の選択された臨床単離物を、MSで使用される適切な配列を同定するためのそれぞれの>2200塩基対（bp）の配列決定によってより詳細に研究した。次いで、これらおよび残りの臨床単離物を使用して、MS法を評価し、CSの結果と比較した。最初は一方の結果を考慮せずに、両方法による分類を行った。

【0081】

血液寒天プレート（5%ウマ血液を補ったColumbiaII寒天ベース）の継代培養によって細菌単離物を保存物から取り出し、37℃で一晩インキュベートした。

【0082】

侵襲性GBS臨床単離物

10

20

30

40

50

可動遺伝因子研究で使用した全194種の単離物を、191人の患者（女性107人、男性80人、4人は性別記録なし；3種の培養物はそれぞれ成長の異なる2つのGBS血清型を含んでいた）の血液（177）またはCSF（17）から回収した。108種の単離物は、1996年から2001年にCentre for Infectious Diseases and Microbiology Laboratory Services、ICPMR、Sydney、Australiaに培養のために提出された標本に由来し、83種の単離物は、1994年から2000年にニュージーランドの種々の診断研究所から血清型分類のためにInstitute of Environmental Science and Research（ESR）、Porirua、Wellington、New Zealandに寄託された。

【0083】

遺伝子型分布の分析のために、患者を以下のように年齢群に分類した：新生児、早期発症（0～6日）；新生児、後期発症（7日～3ヶ月）；幼児および小児（4ヶ月～14歳）；若年成人（15～45歳）；中年（46～60歳）；高齢者（>60歳）。 10

【0084】

これらの単離物は、主に上記の単離物のサブセットであるが、基準株を含み、非侵襲性単離物は排除した。

【0085】

従来 of 血清型分類（CS）

標準的な方法（Wilkinson and Moody、1969）を使用してCSを行った。簡単に述べれば、各単離物について酸加熱（56℃）抽出物を調製し、アガロース中の型特異的抗血清の免疫沈降によって血清型を決定した。生じる沈殿がコントロール株のものと同じのラインを形成した場合に、単離物を特定の血清型について陽性で見なした。ESRでウサギにおいて血清型Ia、Ib、Ic、II、III、IV、V、およびRタンパク質抗原に対する抗血清を調製して使用した。14種の選択した単離物（血清型I～Vに対する抗血清を使用して分類不可能な6種、CSとMSとで最初に結果が矛盾した6種、および混合培養物からの2種の個別の単離物を含む）を、全血清型に対する抗血清を使用して、Abbie Weisner氏およびAndroulla Efstratiou博士（Central Public Health Laboratory、Colindale、London、UK）が試験して下さった。 20

【0086】

分子的血清型同定（MS）；方法の開発

オリゴヌクレオチドプライマー

本研究で使用したオリゴヌクレオチドプライマー、その標的部位、および融点を、表2、6、および10に示す。その特異性および予想されるアンプリコンの長さを、表3、7、および11に示す。プライマーは、本発明者らの仕様に従って、Sigma-Aldrich（Castle Hill NSW、Australia）によって合成された。4つの以前に公開されたオリゴヌクレオチドプライマーおよび本発明者らによってデザインされた一連の新規のプライマーを使用して、目的の遺伝子、すなわち16S/23S rRNA遺伝子間スペーサー領域および部分的cps遺伝子クラスターを配列決定し、または、各GBS血清型の固有の配列を増幅した。6つの以前に公開されたオリゴヌクレオチドプライマーおよび本発明者らによってデザインされた一連の新規のプライマーを使用して、GBS表面タンパク質をコードする遺伝子の一部を配列決定し、および/またはこの遺伝子の特異的に増幅した。本発明者らはまた、5つの公知のGBS可動遺伝因子の一部を配列決定するためおよび/またはこれを特異的に増幅するための、一連のプライマーをデザインした。いくつかは、急速サイクルPCRで使用するために高融点（>70℃）でデザインした。 40

【0087】

DNA調製およびポリメラーゼ連鎖反応（PCR）

5つの個別のGBSコロニーまたは一連の培養物を、使い捨てループを使用してサンプリングし、2mlエッペンドルフチューブ中、1mlの消化緩衝液（10mM Tris-HCl（pH8.0）、0.45% Triton X-100、および0.45% Tween 20）に再懸濁した。GBS懸濁液を含むチューブを、100℃（ドライブロックヒーターまたは水浴）で10分間加熱し、その後氷上で反応停止させ、14,000rpmで2分間遠心分離して細胞残屑をペレット化した。抽出DNAを含む5μlの各上清 50

を、PCRのテンプレートとして使用した (Mawnら、1993)。

【0088】

PCRシステム (検出のみのためには25 μ L、検出および配列決定のためには50 μ L) を、以前に記載 (Kongら、1999) のように使用した。使用した変性、アニーリング、および伸長の温度および時間は、それぞれ96 で1秒間、55~72 (プライマーのT_m値による、または以前に記載) で1秒間、および74 で1~30秒間 (アンプリコンの長さによる) を35サイクルであった。

【0089】

10 μ LのPCR産物を、1.5%アガロースゲルでの電気泳動で分析し、0.5 μ g臭化エチジウムmL⁻¹で染色した。検出および/または血清型同定のために、紫外線透視によって示される予想される長さのPCRアンプリコンの存在を、陽性と認めた。配列決定のために、40 μ LのPCR産物を、ポリエチレングリコール沈殿法 (Ahmetら、1999) によってさらに精製した。

10

【0090】

配列決定

PCR産物を、標準的なプロトコールにしたがって、Applied Biosystems (ABI) Taq DyeDeoxyターミネーターサイクル配列決定キットを使用して配列決定した。対応する増幅プライマーまたは内部プライマーを、配列決定プライマーとして使用した。

【0091】

多配列アラインメントおよび配列比較

多配列分析プログラムグループのPileupおよびPrettyプログラムを使用して、多配列アラインメントを行った。比較プログラムグループのBestfitプログラムを使用して、配列を比較した。全てのプログラムは、WebANGIS、ANGIS (Australian National Genomic Information Service)、第3版から提供されている。

20

【0092】

表面タンパク質遺伝子プロフィールコード

表7に示す種々のプライマー対を使用した正のPCR結果にしたがって、各単離物にタンパク質遺伝子プロフィールコードを与えた。

【0093】

ヌクレオチド配列のアクセッション番号

記載の新規の配列データを、GenBankヌクレオチド配列データベースに提出し、以下のアクセッション番号が割り当てられた: AF291411~AF291419 (基準パネル1由来の血清型Ia~VIII基準株についての16S/23S rRNA遺伝子間スペーサー領域); AF332893~AF332917、AF363032~AF363060、AF367973、AF381030、およびAF381031 (基準株 (表) の2つのパネルの部分的cps遺伝子クラスターおよび選択された代表的臨床単離物); AF367974 (1つの単離物由来の挿入配列IS1381を含む部分的bac遺伝子配列); AF362685~AF362704 (全bac陽性単離物の部分的bac遺伝子配列); およびAF373214 (基準株Prague25/60 (Rタンパク質標準株) の部分的rib様遺伝子)。

30

【0094】

本明細書中で言及した以前に報告された配列データは、GenBankヌクレオチド配列データベース中で以下のアクセッション番号で認められた: AB023574 (16S rRNA遺伝子); U39765、L31412 (16S/23S rRNA遺伝子間スペーサー領域); X68427 (S.oralisの23S rRNA遺伝子); X72754 (cfb遺伝子); AB028896 (血清型Iaのcps遺伝子クラスター); AB050723 (血清型Ibの部分的cps遺伝子クラスター); AF163833 (血清型IIIのcps遺伝子クラスター); AF355776 (血清型IVのcps遺伝子クラスター); AF349539 (血清型Vのcps遺伝子クラスター); AF337958 (血清型VIのcps遺伝子クラスター); M97256 (bca遺伝子); X58470、X59771 (bac遺伝子); U58333 (rib遺伝子); AF208158 (alp2遺伝子)、AF291065~AF291072 (alp3遺伝子); AF064785 (IS1381); M22449 (IS861); Y14270 (IS1548); AF064785 (IS1381); AF165983 (ISSa4); およびAJ292930 (GBSi1)。

40

【0095】

統計分析および樹状図

50

統計分析にSSPSバージョン11ソフトウェアを使用した。SSPSバージョン11ソフトウェアのAverage Linkage (群間) およびHierarchical Cluster Analysisを使用して、樹状図を作成した。各マーカー-MS Ia、Ib、II、IV~VI、sst III-1~4; pgp「A」、「R」、「a」、「as」、「alp2」、「alp3」; bac亜群1、1a、2、3、3a、3b、3c、4、4b、5a、7、7a、8、9、9a、10、n1、n2; およびmge IS1381、IS861、IS1548、ISSa4、GBSII-の有無が、分析に含まれていた。遺伝子型を、それぞれ、分子的血清型分類 (MS) またはsst、pgp、およびmgeの個別の比較によって特徴付けた。

【0096】

実施例 1

GBSゲノムの特定領域中の血清型/血清亜型配列間および配列内異種性の研究、ならびに分子的血清型分類/血清亜型分類の適合性の評価 10

ポリメラーゼ連鎖反応

2つの例外を除いて、全てのGBS特異的プライマー対により、試験した全ての基準株および臨床単離物から予想されるサイズのアンプリコンが産生された (表3)。例外は、Sag59/Sag190およびCFBS/CFBAであった。両方ともcfb遺伝子をターゲティングするが、繰り返し試みたにもかかわらず1つの臨床単離物からアンプリコンを産生できなかった。本発明者らは、この単離物はcfb遺伝子を欠くか遺伝子が変異形態で存在すると予想した。cfb遺伝子をターゲティングするPCRにより全てのGBS単離物が同定されず (Hassanら、2000)、16S rRNA遺伝子に基づいた別のプライマー対DSF2/DSR1 (Ahmetら、1999) は完全に特異的ではないことが以前に示唆されている。したがって、本研究では、両プライマー対 (DSF2/DSR1およびSag59/Sag190) を使用して、全単離物がGBSであることを確認した。 20

【0097】

16S/23S rRNA遺伝子間スペーサー領域の配列異種性

基準パネル1由来の血清型Ia~VIIIの16S/23S rRNA遺伝子間スペーサー領域を配列決定した。多配列アラインメントにより、以下の2つの位置のみで血清型間で異なることが示された: 207 (血清型VIはTまたはC [T/C] であり、血清型VIIおよびVIIIはCであり、他はTである) および272 (血清型IIIはTであり、他はGである)。したがって、これらの領域は、MSには不適切である。

【0098】

cpsD-cpsE-cpsFの3'末端およびcpsGの5'末端の配列異種性 30

cpsD-cpsE-cpsFの3'末端およびcpsGの5'末端をターゲティングする一連のプライマーを使用して、9塩基反復配列の有無によって、基準株 (血清型IaからVII) の両パネルおよび63種の選択した臨床単離物から2226または2217bpを増幅および配列決定した。代表的な配列を、GenBankに寄託した。基準パネル株のGenBankアクセス番号については、表1を参照のこと。

【0099】

反復配列

cpsDの3'末端領域で、MS IaおよびIIのほとんどの単離物、MS IIIのいくつかの単離物、MS IV、V、およびVIIの全ての単離物で、9塩基反復配列 (TTA CGG CGA) が見出されたが、MS IbまたはVIの単離物では見出されなかった (表4)。この反復配列の有無を使用して、MS Ia、II、およびIIIをさらに亜型分類することができる (以下を参照のこと)。 40

【0100】

血清型内の異種性

一般に、血清型内異種性は低かった-血清型内異種性がより複雑なMS IIIを除いて、全ての血清型の少数の単離物で無作為なバリエーションが少し存在した。MS IIIを、22の位置 (62、139、144、204、300、321、429、437、457、486、602、636、971、1026、1194、1413、1501、1512、1518、1527、1629、および2134) での異種性および反復配列 (78~86) の有無に基づいて4つの配列亜型に分類することができた (表4)。

【0101】

60種のMS III単離物 (58種の臨床単離物および2種の基準株) のうち、血清型III-1 (30種 50

の単離物)およびIII-2(22種の単離物)が主であった。反復配列は、血清亜型III-1中には存在したが、III-2中には存在しなかった;7つの他の部位(139、144、204、300、321、636、および1629)で相違が存在した。

【0102】

反復配列を含み、3つの可変部位(139、144、および300)で血清亜型III-1と同一であり、且つ4つの部位(204、321、626、および1629)で血清亜型III-2と同一である血清亜型III-3に属する5つの単離物が存在した。血清亜型III-3は、7つの部位(486、1026、1413、1512、1518、1527、および2134)で血清亜型III-1およびIII-2と異なっていた。血清亜型III-3中のこれらの7つの部位は、MS Iaの対応する部位と同一であった。

【0103】

配列がMS IIの対応する配列とほぼ同一である血清亜型III-4単離物が3つ存在した。437位のみが例外であり、ヌクレオチドが血清亜型III-4中ではT(MS VIIのように)であり、MS II中ではCであった。(PCR(以下を参照のこと)に加えて)この相違を使用して、血清亜型III-4とMS IIとを区別することができる。2つの血清亜型III-4単離物は反復配列を含み、他の一つは含んでいなかった。少数の血清亜型III-4単離物のために、これらをさらに亜型分類するために反復配列を使用しなかった。

【0104】

血清型間の異種性

8つのMS間に56部位の異種性が存在した(表4)。MSのPCR/配列決定での使用に最も適切な部位は、領域の3'末端に最も近い23部位の群であった(表4、図1)。第1に、これらは基準株の2つのパネルおよびほとんどの臨床単離物(少数の血清亜型III-3およびIII-4単離物のみを除く、以下を参照のこと)で一致していた。第2に、これらは、単一反応での配列決定に都合の良い長さである790bp領域内に比較的集中していた。第3に、これらは、MS Ia~VIIの区別に十分な異種部位を含んでいた(いくつかの例外あり)。この790bp領域に基づくだけでは、血清亜型III-3をMS Iaと区別することも、血清亜型III-4をMS IIと区別することもできない。しかし、これらを、MS III特異的PCRによって同定することができる(以下を参照のこと)。

【0105】

血清型VIIIは、790bp領域をターゲティングするプライマー対を使用してもアンプリコンを形成しないが、GBSのPCR同定後の排除によって同定することができる。この研究では、(790bp領域を増幅するプライマー対に加えて)2226bp領域を増幅するプライマー対ではアンプリコンが産生されない1つのMS VIII単離物が同定された。この結果を、血清型VIII特異的抗血清の使用によって確認した。

【0106】

単一の単離物における混合血清型特異性

MS特異的PCRおよび配列全体(2226/2217bpセグメント内)に基づいて1つのMSとして11種の単離物を同定したが、その配列は同一のMSの単離物といくつかの部位で異なり、且つ別の部位特異的特徴を共有していた。これらは、5つの血清亜型III-3単離物および3つの血清亜型III-4を含んでいた(上記を参照のこと)。MS IIと同定された1つの血清型分類不可能な基準株(Prague25/60)は、領域の5'末端で5つの部位で他のMS II単離物と異なり、これらの部位の3つでMS IIIと同一であった。Prague25/60 MS III特異的PCRは、陰性であった。CS II、およびその配列全体に基づいてMS IIと同定された1つの臨床単離物は、血清型Ibに特徴的な領域の5'末端において9つの部位で塩基を有しており、MS Ib特異的PCRは陰性であった。最後に、1つのCS V基準株(Prague10/84)は、GenBankの対応する配列(AF349539)と同一の配列決定結果を有するが、両者は、研究した他のMS V株の配列と領域の5'末端において3つの部位で異なった。

【0107】

血清亜型III-3およびIII-4に関連する特異性を除いて、これらの混合血清型特異性は全て、2226/2217フラグメントの5'末端領域で生じた。このことは、MSの配列決定標的として790bpの3'末端の選択を支持した。この標的を使用して、MS III特異的PCRによって同定

10

20

30

40

50

することができる血清亜型III-3およびIII-4に属するMS IIIを除いて、全てのMSを正確に同定した（実施例2を参照のこと）。

【0108】

実施例2

cpsG-cpsH-cpsI/cpsMの3'末端をターゲティングするMS特異的PCRに基づいた分子的血清型同定（MS）

本発明者らの配列アラインメント結果により、cpsG-cpsH-cpsI/cpsMの3'末端中に意義深い配列異種性が存在することが示され（図3）、これにより、PCRによって直接的に血清型Ia、Ib、III、IV、V、およびVIを区別するための特異的プライマー対デザインの使用が適するようになった。考えられるさらなる将来的な要件（例えば、多重（multiplex）PCRの開発）を満たすためおよび/または配列分類法をさらに評価するために、各血清型のためのいくつかのプライマー対をデザインした（表2および3）。基準株の2つのパネルおよび特定の条件を使用して、全プライマー対により対応する血清型からのみDNAが増幅された。臨床単離物を試験した場合、2組のMS特異的プライマー対を使用して類似の結果が得られた。一般に、より小さなアンプリコンを生じるプライマーと共に、よりストリンジェントな条件（より低いプライマー濃度、より高いアニーリング温度）を使用することができる。MSについて選択された条件を、表3および図2に示す。

10

【0109】

以下のように、MSをPCRによって206種の臨床単離物中179種（86.9%）に割り当てた：MS Ia 40；MS Ib 35；MS III 58（以前に血清亜型III-3およびIII-4と同定されたものを含む）；MS IV 7；MS V 36；MS VI 3。

20

【0110】

実施例3

MSとCSとの間の血清型同定結果の比較

CSおよびMSが完了した後、結果を比較した。最初の結果は、15種の単離物で矛盾していたが、このうちの5種を除く全部（以下を参照のこと）は、再試験および/または誤記の訂正によって解決した。

【0111】

CSおよびMS/配列亜型分類の結果を、表5に示す。PCRおよび/または配列決定によって、CSによる206中188（91.3%）と比較して、MSを全ての単離物に割り当てた。MS IIおよびVII Iのための特異的PCRは開発されていないので、全MS II単離物を配列決定のみによって決定し、1つの推定MS VIII単離物を、排除によって決定した（実施例1を参照のこと）。他の全ての単離物について、血清亜型III-3およびIII-4ならびに上記の他の軽微な配列の相違（実施例1）を除いて、PCRおよび配列決定の結果は一致した。CSの結果は、PCRの結果と十分に相関した。

30

【0112】

CSおよびMSの最終結果は、両方法の結果が利用可能である188種全ての単離物（100%）で同一であった。CSによって血清型分類不可能な18種の臨床単離物を、以下のようにMSに割り当てた：Ia、2；Ib、5；II、1；血清亜型III-1、3；血清亜型III-2、1；V、5；およびVI、1。

40

【0113】

MS VIと同定された3種の臨床単離物の配列（2217bp）は、血清型VI基準株の配列およびGenBankの対応する配列（AF337958）と同一であった。

【0114】

混合培養物

4種の臨床単離物について、MS III特異的PCRで正の結果が得られたが、配列決定によってMS IIと暫定的に同定された。3つはCS IIIであり、1つはCS IIであり、血清型III抗血清と弱い交差反応が認められた。これらの単離物を、それぞれ12個の個々のコロニーの継代培養によってさらに研究した。全ての継代培養物を、MS III特異的PCRによって試験した。3種のCS III単離物の12個のコロニー継代培養物は全て、MS III特異的PCRによって陽性

50

であったので、単離物を血清亜型 III-4 と分類した（上記を参照のこと）。しかし、第4番目の単離物の12個のコロニー継代培養物のうちの11個は、MS III特異的PCRによって陰性であり、1個はMS III特異的PCRによって陽性であった。したがって、これは主にMS/CS IIの混合培養物と予想された。その後、その1個のMS III特異的PCR陽性コロニーは血清亜型 III-2 と同定され、さらなる臨床単離物として含まれた（全部で206）。

【0115】

実施例 4

PCRおよび配列決定によるGBSの血清型割り当てのアルゴリズム

GBS血清型同定の臨床的实施に上記のPCRおよび配列決定法がどのように使用されうるかということの例として、臨床用のアルゴリズムをデザインした。使用する全てのプライマー（内部配列決定プライマーを除く）に高い融点（ >70 ）を与え、急速サイクルPCRを使用した（図2）（プライマー配列については、表2を参照のこと）。

10

【0116】

実施例 5

タンパク質抗原遺伝子特異的亜型分類に適した *alp2*、*alp3*、および *rib* 遺伝子中の領域の同定

少数の例外を除き、全てのプライマー対により、正の結果が得られる単離物由来の予想される長さのアンプリコンが産生された（表7）。例外は、プライマー対 GBS1360S/GBS1937A および GBS1717S/GBS1937A（共に *bac* 遺伝子をターゲティングする）を使用したPCRによって陽性であるが、他の *bac* 遺伝子陽性単離物よりも有意に長いアンプリコンが産生される、1つの単離物を含んでいた。配列決定により、公開された配列（Tamuraら、2000）と比較して、アンプリコンは軽微なバリエーションを含む挿入配列 IS1381 を含むことが示された。プライマー IgAagGBS/RlgAagGBS および IgAS1/IgAA1（*bac* 遺伝子もターゲティングする）を使用して産生したアンプリコンは長さが異なり（Bernerら、1999）、さらなる亜型分類のために配列決定した（以下および表8を参照のこと）。

20

【0117】

アンプリコン配列決定の結果

デザインまたは改変された選択プライマー対の特異性を確認するために、基準株の2つのパネルおよび31種の無作為に選択した臨床単離物由来の、*bcaS1/bcaA*（*bca* 遺伝子の5'末端をターゲティングする）によって産生された23個のアンプリコンのうち10個、ならびに *ribS1/ribA3*（*rib* 遺伝子をターゲティングする）および GBS1360S/GBS1937A（*bac* 遺伝子をターゲティングする）によって産生された全てのものを配列決定した。

30

【0118】

プライマー *bcaS1/bcaA* の全10アンプリコンおよびプライマー *ribS1/ribA3* の13のうちの12は、GenBankの対応する遺伝子配列と同一であった（それぞれ、M97256は *bca* 遺伝子、およびU58333は *rib* 遺伝子）。1つのさらなる単離物（すなわち、基準パネル2中のPrague25/60（R抗血清の惹起に使用））により、低アニーリング温度（55）でのみプライマー対 *ribS1/ribA3* を使用してアンプリコンが産生されたが、*ribS2/ribA1* および *ribS2/ribA2* では産生されなかった。したがって、*rib* 遺伝子を含まないと予想されるが、アンプリコン配列は *rib* 遺伝子とかなりの相同性を示した（71.4%または66.6%、プライマー配列を含む場合または含まない場合）（図3）。この単離物は、試験した224種のうちのたった1種であり、PCRは *ribS2/ribA1* および *ribS2/ribA2* を使用して陰性であったが、*ribS1/ribA3* を使用して陽性であった。後者のプライマー対は、*rib* 遺伝子と完全に特異的ではないと考えられるので、配列決定のためだけに使用した。

40

【0119】

プライマー対 GBS1360S/GBS1937A（*bac* 遺伝子をターゲティングする）の10アンプリコンのうち4つは、GenBank中の対応する配列（X58470、X59771）と同一であった。残る6つの *bac* 遺伝子アンプリコンにおいて点変異（AからG、X59771の1441）が認められ、そのうちの1つは挿入配列 IS1381（上記およびAF367974参照）を含んでいた。

【0120】

50

プライマー対 *bcaS1/balA* (*alp2/alp3* 遺伝子をターゲティングする)、*bal23S1/bal2A2* (*alp2* 遺伝子をターゲティングする)、および *IgAagGBS/RIgAagGBS* (*bac* 遺伝子をターゲティングする) を使用して陽性の PCR 結果が得られる 224 種の全単離物由来のアンプリコンを配列決定した。

【0121】

プライマー対 *bcaS1/balA* を使用して、50 種の単離物からアンプリコンを産生した。9 個の配列は *alp2* 遺伝子の公開された配列 (AF208158) の対応する部分と同一であり、41 個は *alp3* 遺伝子 (AF291065) と同一であった。*bcaS1/balA* アンプリコンの配列中の *alp2* 遺伝子と *alp3* 遺伝子との間に 2 つの一致する異種部位が存在し (図 4)、*alp2* および *alp3* 遺伝子特異的 PCR に加えて、これらを区別するために使用することができる。プライマー対 *bal23S1/bal2A2* の 9 つ全てのアンプリコンは、GenBank の *alp2* 遺伝子配列 (AF208158) の対応する部分と同一であった。

10

【0122】

プライマー対 *IgAagGBS/RIgAagGBS* により、52 種の単離物中の *bac* 遺伝子が同定された。アンプリコン長および配列異種性に基づいてそれぞれ *bac* 遺伝子陽性単離物を 11 の群および 20 の亜群に分離することができる多数の配列バリエーションが存在した (表 8)。群は、B1 (20 種の単離物、2 亜群) および B4 (11 種の単離物、3 亜群) を除いて、少数 (1 から 5 種) の単離物を含んでいた。アンプリコン長の相違は概して、短い反復配列の有無に起因していた。

20

【0123】

表面タンパク質遺伝子特異的プライマー対の特異性のさらなる確認
プライマー特異性を確認するために、*bac* 遺伝子 PCR 用にデザインまたは改変したプライマー配列を使用した PCR の結果と以前に公開されたプライマーを使用した PCR とを比較し、100% 相関することを見出した。

30

【0124】

公開プライマー対 *bcaRUS/bcaRUA* (*bca* 遺伝子反復単位をターゲティングする) の以前に報告された非特異性を確認した。これらのプライマーを使用して、9 つ全ての *alp2* 遺伝子陽性 (*bcaS1/bcaA* 陰性) 単離物およびプライマー *bcaS1/bcaA*、*bcaS2/bcaA* (*bca* 遺伝子の 5' 末端をターゲティングする)、*bal23S1/bal2A2*、および *bal23S2/bal2A1* (*alp2* 遺伝子の 5' 末端をターゲティングする) を使用して PCR 陰性であった 53 種により、アンプリコンが産生された。配列決定により、*bca* 遺伝子および *alp2* 遺伝子は、*alp2* 遺伝子陽性株からアンプリコンが形成される *bcaRUS/bcaRUA* によってターゲティングされる領域中に有意な相同性を有することが示された。これらの偽陽性の結果は、*bca* 遺伝子反復単位に相同な領域 (*bca* 遺伝子反復単位様配列) を含む他の C 様タンパク質の存在に起因し得る。

30

【0125】

各遺伝子 (*rib*、*alp2*、および *alp3* 遺伝子) のためにデザインした 2 つまたはそれ以上のプライマー対を使用した PCR の結果が十分に相関し、各組の特異性を支持することも示した。上記のように、試験された 224 種の単離物のうちの 1 つに由来する非特異的アンプリコンを産生する *ribS1/ribA3* のみが例外であった。

40

【0126】

実施例 6

表面タンパク質抗原遺伝子プロフィールと *cps* 血清型/血清亜型との関係

表面タンパク質遺伝子プロフィール

各遺伝子 (*bca* 遺伝子反復単位または *bca* 遺伝子反復単位様領域を除く) について、PCR による GBS 表面タンパク質の同定および特徴づけのために 2 つのプライマー対を選択した。各単離物に、PCR の結果によって以下のタンパク質遺伝子プロフィールコードを与えた。

「A」: *bcaS1/bcaA* および *bcaS2/bcaA* によって増幅された *bca* 遺伝子の 5' 末端;

「a」または「as」: それぞれ複数または単一バンドのアンプリコンを有する、*bcaRUS/bcaRUA* によって増幅された *bca* 遺伝子反復単位または *bca* 遺伝子反復単位様領域;

「B」: GBS1360S/GBS1937A および *IgAagGBS/RIgAagGBS* によって増幅された *bac* 遺伝子 (配

50

列異種性に基づいた > 20の亜群) ;

「R」 : ribS2/ribA1およびribS2/ribA2によって増幅されたrib遺伝子 ;

「alp2」 : bal23S1/bal2A2およびbal23S2/bal2A1によって増幅されたalp2遺伝子 ; および

「alp3」 : bal23S1/bal3Aおよびbal23S2/bal3Aによって増幅されたalp3遺伝子 (表7)。

【0127】

4つの共通のプロフィールが224種の単離物中の203種(90.6%)を占めた: 「R」(62単離物)、「AaB」(51単離物)、「a」(49単離物)、および「alp3」(41単離物)(表4を参照のこと)。2種の単離物のみが、表面タンパク質遺伝子マーカーを含まなかった。bac遺伝子を含む単離物(「B」)は1つを除いて全て、反復単位を含むbca遺伝子も有し(「Aa」)、1種はrib遺伝子を有していた。全ての「alp2」単離物は、単一のbca反復単位様配列を含んでいた(「as」)。「A」、「R」、「alp2」、「alp3」は全て相互に排他的である。rib遺伝子を有する63種の単離物(「R」)のうちの62種およびalp3遺伝子を有する41種の単離物のうちの41種は、他のタンパク質抗原マーカーを有していなかった。

10

【0128】

表面タンパク質抗原遺伝子プロフィールとcps血清型/血清亜型との関係

cps分子的血清型(MS)を、実施例1から4に記載の方法にしたがって全単離物に割り当て、この結果は、抗血清を使用して分類不可能な224種の単離物のうちの19種を除いて従来の血清型分類(CS)の結果と相関した。表面タンパク質遺伝子プロフィールとcpsMSとの関係を、表9にまとめる。

【0129】

以下のものの間に強力な関係が確認または証明された: MS Iaとbca遺伝子反復単位またはbca遺伝子反復単位様配列(ほとんどがプロフィール「a」を有する); MS血清亜型III-1およびIII-2とrib遺伝子; MS血清亜型III-3とalp2遺伝子; MS Ibとbca/bac遺伝子; およびMS Vとalp3遺伝子。MS IIは、最も変化した表面タンパク質遺伝子プロフィールを示した。しかし、この関係は絶対的ではなく、cps血清型とタンパク質遺伝子プロフィールとの異なる組み合わせにより、bac遺伝子(「B」)亜群を考慮した場合に、31種の異なる血清変異型(serovariant)または51種が産生された。

20

【0130】

実施例7

表面タンパク質抗原とタンパク質遺伝子プロフィールとの関係

従来の血清型分類に基づいて、33種の単離物(CS Ia/c、Ib/c、IIc、IIb、IIIC、またはIIbに属する)は、C抗血清と反応した。これら全ての単離物の表面タンパク質遺伝子プロフィールは、bca遺伝子(「A」)またはbca遺伝子反復単位関連マーカー(「a」または「as」)を含んでいた: Aa、3; AaB、18; a、11; alp2as、1。29種の単離物はR抗血清と反応し、これらのうちの22種はrib遺伝子を含み、6種はalp3遺伝子を含んでいた。Rタンパク質抗血清を惹起するために使用した株(Prague25/60)は、推定rib様遺伝子を含んでいた(上記および図3を参照のこと)。

30

【0131】

実施例8

分子的亜型分類に適した可動遺伝因子の同定

GBS血清型中の5つの可動因子の存在をスクリーニングするための一連のPCRプライマーを開発した。

40

【0132】

プライマー対の特異性

全てのプライマー対により、いくつかの基準単離物および/またはいくつかの臨床単離物から予想される長さのアンプリコン(表11)が産生された(表12)。本発明者らのプライマー対の特異性を評価するために、プライマーIS1548S/IS1548A3およびISSa4S/ISSa4A2によって産生された全アンプリコン、ならびにIS861S/IS861A2(12単離物)、IS1381S1/IS1381A(24単離物)、およびGBSi1S1/GBSi1A2(11単離物)によって産生された基準単離物および臨床単離物両方から選択されたアンプリコンを配列決定した。

50

【 0 1 3 3 】

41種のIS1548および15種のISSa4アンプリコン配列の全てがGenBankの対応する配列（それぞれ、Y14270およびAF165983）と同一であった。12種のIS861アンプリコン配列のうちの5種は、GenBankの対応するIS861配列（M22449）と同一であった。他の7種は、公開された配列と732位が異なり（G A）、基準株Prague25/60は、M22449の576位および830位に2つのさらなる相違を有していた（それぞれG AおよびT A）。

【 0 1 3 4 】

以前に、本発明者らは、元の公開された配列（AF064785）と比較して以下のいくつかの相違を有する、臨床単離物のC 抗原遺伝子内に全長挿入配列IS1381（AF367974）を見出した：末端逆位反復配列は20塩基対（bp）ではなく15bpを含み、3bpが欠失し、（元のGenBank配列の）419位と429位との間の推定トランスポザーゼ偽遺伝子中で4つの個々の塩基対が異なる -

GGG ATC CGA TT

(AF064785) vs CAG A-- -GG TA (AF367974; 本発明者らの配列)

。12種の基準単離物および12種の選択された臨床単離物由来のプライマー対IS1381S1/IS1381Aの全てのアンプリコンは、互いに同一であり、且つGenBankの本発明者らのIS1381配列（AF367974）のアンプリコンと一致するが、上記のように、最初に報告されたIS1381配列（AF064785）と異なっていた。

【 0 1 3 5 】

4種のGBSi1陽性基準株および7種の選択した臨床単離物の全て由来のプライマー対GBSi1S1/GBSi1A2のアンプリコンを配列決定した。6種（3種の基準株のものを含む）は、GenBankの対応するGBSi1配列（AJ292930）と同一であった。4種の臨床単離物由来のアンプリコンは、3つの部位バリエーション（AJ292930配列の767位でのC T、846位でのA C、および923位でのT C）を示した。基準株Prague25/60は、これらの部位バリエーションのうち最初の2つのみを示した。

【 0 1 3 6 】

配列決定に加えて、各標的に対する2つまたはそれ以上のプライマー対のPCR結果を比較することによって本発明者らのプライマー対の特異性を評価した（表11）。全ての場合、同一の可動遺伝因子をターゲティングするPCRで試験した場合に同一の単離物の組は陽性の結果を示したので、プライマー対の特異性が確認された。

【 0 1 3 7 】

5つ全ての可動遺伝因子の特異的プライマー対を使用したPCRの結果それぞれ単離物の55%、18%、85%、7%、および19%でIS861、IS1548、IS1381、ISSa4、およびGBSi1が同定された。10種の単離物（4%）で可動因子は検出されなかった。前の実施例で試験した224種のGBS単離物におけるPCRによって同定された5つの可動因子の分布を、表12に示す。79種の単離物中でIS1381のみが検出され、1種の単離物中でGBSi1のみが検出された。46種の単離物は2つの異なる挿入配列（IS861およびIS1381、42単離物；IS1548およびIS1381、3単離物；ISSa4およびIS1381、1単離物）を含んでいた。44種の単離物は、3つの挿入配列（IS861、IS1548、およびIS1381、34；IS861、ISSa4、およびIS1381、10）を含み、1種の単離物は4つ全ての挿入配列を含んでいた。41種の単離物は、1つ（IS861、22；IS1381、1単離物）、2つ（IS861およびIS1381、11；ISSa4およびIS1381、3単離物）、または3つ（IS861、IS1548、およびIS1381、4単離物）の挿入配列と組み合わせてGBSi1を含んでいた。

【 0 1 3 8 】

5つ全ての可動遺伝因子に特異的プライマー対を使用した194種の侵襲性単離物のPCR結果異なる可動遺伝因子（mge）組み合わせ（単離物あたり0から4種）を含む単離物の数を、表13に示す。それぞれ単離物の87%、52%、17%、6%、および18%で、IS1381、IS861、IS1548、ISSa4、およびGBSi1が同定された。6種（3%）の単離物は、mgeを含んでいなかった。

【 0 1 3 9 】

実施例 9

cps血清型、血清型II、表面タンパク質遺伝子プロファイル、および可動遺伝因子の関係異なるcps血清型、血清型III型、および表面タンパク質遺伝子プロファイル中の5つの各可動遺伝因子の分布を、表12および13に示す。各血清型/血清型IIについて最も一致した所見は以下のとおりであった：

- 1) 血清型Ia-ほとんど (> 80%) が、C タンパク質と密接に関連するタンパク質を発現し、IS1381を含んでいた。
- 2) 血清型Ib-ほとんど (> 90%) が、C およびC タンパク質を発現し、IS861およびIS1381を含んでいた。
- 3) 血清型II-以下の2つの共通のパターンを示す：
 - a) > 50% が、C タンパク質 (およびしばしばC) を発現し、IS861、IS1381および時々他の可動因子、特にISSa4を含んでいた、または
 - b) > 25% が、Ribタンパク質を発現し、IS861、IS1381、およびGBSi1を含んでいた。
- 4) 血清型III-1：全てがRibタンパク質を発現し、IS861、IS1548、およびIS1381を含むがGBSi1を含まなかった。
- 5) 血清型III-2：全てがRibタンパク質を発現し、IS861およびGBSi1を含むがIS1548およびIS1381のいずれも含まなかった。
- 6) 血清型III-3：全てがC 様タンパク質2を発現し、可動遺伝因子を含まなかった。
- 7) 血清型III-4：種々のタンパク質を発現し、全てがGBSi1を含んでいた。
- 8) 血清型IV-ほとんどが、C タンパク質と密接に関連するタンパク質を発現し、IS1381を含んでいた。
- 9) 血清型V-ほとんどがC 様タンパク質3を発現し、IS1381を含んでいた。
- 10) GBSi1およびIS1548は血清型III (III-1、III-2、およびIII-4) において互いに排他的であるが、血清型IIではそうではない。
- 11) C 様タンパク質2を発現した単離物は全て、挿入配列を含まなかった。

10

20

【 0 1 4 0 】

MS/sst、pgp、およびmgeの主な関係

図5は、種々の遺伝子マーカー間の関係を示す。IS1381は、MS Ia、Ib、IV、V、およびVIのほぼ全ての単離物に存在したが、sstIII-2またはIII-3には存在しなかった。血清型IIまたはIIIでIS1548が排他的に見出され、GBSi1が最も共通して見出され；3種の単離物 (全てMS II) は、GBSi1およびIS1548の両方を含んでいた。全てのsst III-1およびIII-2ならびにほとんどのMS IIおよびIb単離物でIS861が見出されたが、他のMS単離物ではたった14%であった。ISSa4は、6%の単離物のみが存在し、その過半数はMS IIであり；これは1996年以前 (1994年) に得られた1つの侵襲性単離物中に存在していた。IS1381は、挿入配列を有さないクラスター8中の単離物 (pgp「alp2」) 以外のほとんどの単離物中で見出された。IS861は、pgp「AaB」 (クラスター3および4) を有するほとんどの遺伝子型およびpgp「R」 (クラスター6および7) を有する全ての遺伝子型で見出された。

30

【 0 1 4 1 】

MS/sst、pgp、bac型、およびmgeに基づく遺伝子型

MS/sst、pgp、bac型 (pgp「B」を含む単離物について)、およびmgeの種々の組み合わせの存在により、PCR/配列決定ベースの遺伝子型同定系が得られる。本研究における194種の侵襲性単離物により、bac型 (主にMS Ib) が含まれる場合に、pgpおよびmgeまたは56種の遺伝子型の分布に基づいて、7種の血清型、10種のMS/sst、41種の型が示された (図5)。

40

【 0 1 4 2 】

理論的GBSクローン集団構造

理論上は、13種の可能なGBS MS/sst (8種のMS-Ia、Ib、II、IV~VIII、4種のsst III 1~4およびcps遺伝子クラスター不在) および少なくとも10種のpgp (なし、「Aa」、「AaB」、「a」、「as」、「R」、「RB」、「alp2as」、「alp3」、または「alp4a」) が存在す

50

る。これまでに同定された22種のbac亜型が含まれる場合、31種までのpgpが存在する。5種のmgeが独立して無作為に分布して存在するまたは存在しない場合、 $13 \times 31 \times 2^5 = 12,896$ 種の分子マーカーの異なる組み合わせが存在する可能性がある。56種の異なる組み合わせしか見出されないという事実(図5)は、マーカーが無作為に分布しない、言い換えれば、これらの侵襲性オーストラレーシアGBS単離物はクローン集団構造を有することを示している。考えにくいことであるが、これらの単離物は非常に限られた数のGBS遺伝子型しか示さない可能性がある。

【0143】

オーストラレーシア侵襲性GBSの系統発生的関係

56種の遺伝子型は8種のクラスターを形成し、約16の遺伝的距離で分離されていた(あるいは、3つのクラスター群が約22.5の距離で分離していた)。pgpは、クラスター分離の主な決定因子であった(図5)。94%の単離物は、5種のMS(Ia、Ib、II、III、およびV)に属し、62%が5種(9%)の遺伝子型(Ia-1、Ib-1、III-1、III-2、V-1)に属し、92%が5種の最も大きいクラスター(1、2、4、6、および7)に属していた。クラスター群A(最も大きい)は、139種(72%)の単離物および48種(86%)の遺伝子型を含み、そのうちの45種が5種未満の単離物を含み、クラスター群Bは49種(25%)の単離物および5種(9%)の遺伝子型を含んでいた。

10

【0144】

各クラスターの主な特徴は、以下のとおりである：

クラスター1：「alp3」、IS1381(39種の単離物、4種のMS、11種の遺伝子型；主な遺伝子型V-1)。

20

クラスター2：「a」または「as」、IS1381(55種の単離物、4種のMS、12種の遺伝子型、主な遺伝子型Ia-1)。

クラスター3：「Aa」または「AaB」、MS II、IS1381、IS861(10種の単離物、6種の遺伝子型)。

クラスター4：「AaB」、IS1381、IS861(35種の単離物、2種のMS：VIまたはIb；18種の遺伝子型；主な遺伝子型Ib-1)。

クラスター5：「AaB」、IS861、GBSi1、遺伝子型III-4-1(1種の単離物)。

クラスター6：「R」、IS861、およびGBSi1(22種の単離物、3種のMS/遺伝子型；主な遺伝子型III-2)。

30

クラスター7：「R」、IS1381およびIS861(27種の単離物；2種のMS/遺伝子型；主な遺伝子型III-1)。

クラスター8：「alp2as」、ISなし(6種の単離物；3種のMS/遺伝子型；1つがGBSi1を含む)。

【0145】

系統発生研究により、SSPSによって推測された樹状図は非常に強固であることが示された。

【0146】

遺伝子型とGBS疾患パターンとの関係

侵襲性GBS疾患患者の異なる年齢群におけるMSおよび遺伝子型の分布を、表14に示す。1つを超える患者群で全ての共通のMSが示された。しかし、(全ての他の年齢群と比較すると)sstIII-2と遅発型新生児感染症との間($p = 0.0005$)およびMS Vと高齢者の感染症との間($p = 0.001$)に、非常に有意な関係が存在した。

40

【0147】

脳脊髄液標本由来の17種の単離物が存在し、そのうちの9種(53%)がMS III(それぞれ異なるクラスター中の3種の異なるsst/遺伝子型由来)であった。他の8種の単離物は、5種のMS、7種の遺伝子型、および4種のクラスターに分布していた。全年齢群で髄膜炎を発症していたが、他の全ての群での5%と比較して、遅発型新生児群では症例の23%であった。

【0148】

50

考察

GBSにおける莢膜産生は、我々の研究開始前に血清型 Ia および血清型 III について配列決定されていた莢膜多糖類合成 (cps) 遺伝子クラスターによって調節される。プロジェクトがほぼ終了して、血清型 Ib の対応する配列 (Miyake ら、2001 年 GenBank 提出、GenBank アクセション番号 AB050723) および血清型 IV、V、および VI に対応する配列 (McKinnon ら、2001 年 GenBank 提出、それぞれ GenBank アクセション番号: AF355776、AF349539、AF337958) が最近公開されたが、他の 3 種の血清型 (II、VII、および VIII) の対応する配列 (cps 遺伝子クラスター配列) は、以前に公開されていなかった。

【0149】

血清型 Ia および III の cps 遺伝子クラスターの配列は、cpsD-cpsE-cpsF の 3' 末端および cps G の 5' 末端で相同性が高かった。本発明者らは、この領域中の 2226/2217bp セグメントを増幅するための一連のプライマーをデザインし、VIII 以外の全ての血清型からアンプリコンが得られることを見出した。これにより、この領域で血清型 VIII が他の血清型と著しく異なるという以前の示唆が確認された。

【0150】

本発明者らは、8 種の血清型 (Ia から VII) の基準株を使用して、血清型の間で 50 種を超える異なる点を示した (図 1、表 4)。従来の方法によって血清型分類した 63 種の選択された臨床単離物を使用して、これらの血清型間の相違が一般に一貫し且つ特異的であり、特に 23 個の部位が領域の 3' 末端にクラスター化していることを見出した。本発明者らは、従来の方法によって得られた血清型に関する知識を使用せずに、これらの相違を使用して本 20 研究で回収した残りの臨床単離物に血清型を割り当てた。

【0151】

血清型 Ia、III、Ib、IV、V、および VI の cpsG-cpsH-cpsI/cpsM の 3' 末端の配列分析により、この領域の可変性が高く (図 3)、この領域は PCR による直接的血清型同定の適切な標的となることが示された。本発明者らは、MS Ia、Ib、III、IV、V、および VI のいくつかの MS 特異的プライマー対をデザインし、これを使用して 2 つの CS 基準パネルを試験した。206 種の臨床単離物の 86.9% の MS について、PCR のみによる選択したプライマー対を使用した。急速サイクル MS 特異的 PCR を使用して、1 作業日以内に結果を利用することができる。後に、この領域中の cps 遺伝子クラスター配列が血清型 II、VII、および VIII で利用可能である場合、全 MS に対してこの方法を拡大することが可能である。一方で、MS II および VII を、 30 cpsE-cpsF の 3' 末端-cpsG の 5' 末端の 790bp PCR アンプリコンの配列決定によって同定することができる (図 1、表 4)。790bp を増幅する全プライマーを使用した陽性の GBS 特異的 PCR 結果および陰性の PCR 結果により、排除によって MS VIII が同定された。

【0152】

後に、および現在いくつかの研究所において、全単離物の cpsE-cpsF の 3' 末端-cpsG の 5' 末端の 790bp PCR アンプリコンの配列決定は、たった 1 つの方法およびより少数のプライマーしか必要でないので、より都合がよい可能性がある。しかし、院内で配列決定が利用できない場合、納期がより長くなり、割合の低い血清型が間違っ て割り当てられることが考えられる (血清型 III-3 および III-4 がそれぞれ MS Ia および II として)。最初に MS III 特異的 PCR を使用してスクリーニングすることによって、これを回避することができる。7 40 90bp PCR アンプリコンの配列決定により、配列異種性に基づいて MS III を亜型分類することが可能である。

【0153】

以前の研究により、血清型 Ia、Ib、II、III、および V は、米国およびいくつかの国で正常な無菌部位から最も頻りに単離されるものであることが示されている。血清型 VI および VII は、日本の患者から単離された主な血清型であるが、他の地域では一般的ではない。本発明者らの単離物を選択したが、これらはおそらくオーストラレーシアで疾患を発症させるものの代表であり、Ia、Ib、II、III、および V は同定された最も一般的な血清型であったが、少数の血清型 IV、VI、および VIII が存在した。

【0154】

13%までのGBS単離物は血清型分類することができず、本研究では、利用可能な抗血清を使用して、その割合は8.7% (18/206)であった。これは、型特異的抗原合成の減少；非莢膜形成期の変化；またはcps遺伝子クラスターの遺伝子の挿入もしくは変異に起因している可能性がある。本研究における1つの血清型分類不可能な株GBSは、cpsG遺伝子のT塩基が欠失しており、これによりcpsG遺伝子の読み取り枠が変化していた。

【0155】

本発明者らは、GBS表面タンパク質遺伝子を同定し、さらにこれらの単離物を特徴付けるPCRベースの方法も開発した。公開されたbac遺伝子配列を使用して、bac遺伝子特異的プライマーを改変し、全ての主要な表面タンパク質遺伝子をターゲティングする急速サイクルPCRに適切な高融点 (>70)を有する新規のプライマーをデザインした。

10

【0156】

以前に報告されたように、bca遺伝子反復単位 (bca遺伝子の3'末端) をターゲティングする公開されたPCRプライマー対は、bca遺伝子に完全に特異的なわけではなかった。本発明者らは、特異性を改善するために、bca遺伝子の5'末端をターゲティングする新規の2つのプライマー対をデザインした。しかし、これらのプライマーを使用して非常に少数の血清型1a株しか正の結果が得られないのに対して、bca遺伝子反復単位をターゲティングするプライマーを使用すると全てがPCR陽性であった。これらの結果は、bca遺伝子の5'末端をターゲティングするプローブは9種の血清型1a株の1種のみとしかハイブリッド形成しないが、タンデムリピート領域を含む巨大なbca遺伝子プローブは9種すべての株とハイブリッド形成するという以前の報告と一致した。

20

【0157】

rib、alp2、およびalp3遺伝子に特異的なPCRは、以前には記載されていなかった。本発明者らがデザインしたプライマー対は、主に遺伝子の5'末端をターゲティングし、遺伝子の異種性と関連遺伝子の配列との比較後に選択した。本発明者らは、同一の遺伝子をターゲティングする異なるPCRの結果の比較によってプライマーの特異性をチェックするための各遺伝子について2つまたはそれ以上のプライマー対をデザインした。タンパク質遺伝子プロフィール「alp2」および「alp3」を、alp2およびalp3遺伝子特異的PCRならびに/またはbcaS1/balAもしくはbcaS2/balAのアンプリコン中の2つの配列異種性部位に基づいて区別した。

【0158】

本発明者らのプライマーの特異性を確認するために、本発明者らは、これらを使用して2つの基準パネルおよび選択したGBS単離物を試験した。各遺伝子についてPCRによって産生された最も長いアンプリコンを配列決定して最大の配列情報を得て、確実に内部プライマーが株異種性部位に位置しないようにした。本発明者らの配列決定結果により、プライマーの特異性が確認された。各遺伝子の2つのプライマー対を、類似の結果と比較した。最後に、6種の遺伝子/領域特異的プライマー対 (bca遺伝子反復単位をターゲティングするものを含む) を使用して、224種全ての単離物のタンパク質抗原遺伝子プロフィールを定義した。

30

【0159】

本研究により、反復配列を含む表面タンパク質遺伝子ファミリー (rib、bca、alp2、およびalp3遺伝子) のたった1つのメンバーしか任意の1つの単離物に存在することができないことが示された。しかし、反復配列を含む表面タンパク質遺伝子ファミリーのメンバーではないbac遺伝子を含む全ての単離物もまた、bca遺伝子 (51/52) またはrib遺伝子 (1/52) のいずれかを含んでいた。

40

【0160】

bac遺伝子は、単離物の23%中に存在し、その類似の比率 (19~22%) が以前に報告されている。他と共通して、bac遺伝子中に可変性の小さな内部反復配列に起因するバリエーションが見出された。これらのbac遺伝子反復配列は不規則であった (bca-rib遺伝子ファミリーのものとは異なる) 。その役割は明らかではないが、疫学研究用の有用な分子マーカーである可能性がある。

50

【0161】

本発明者らのデータは、いくつかの血清型III単離物（MS血清亜型III-1およびIII-2）がrib遺伝子と密接に関連し、他の単離物（MS血清亜型III-3）がalp2遺伝子に関連することを示す。血清型Ibは、bcaおよびbac遺伝子と関連し、血清型Vはalp3と関連していた。しかし、この関係は絶対的ではないので、cps血清型-血清亜型/タンパク質遺伝子プロフィールの異なる組み合わせにより、疫学研究および結合型ワクチンの処方にも有用な多数の血清バリエーションが同定された。PCRのみに基づいて、224種の単離物を、bac遺伝子（B）群に基づいて31種の血清バリエーションに分類するか、亜群に基づいて51種に分類することができた。理論的には、さらなる血清バリエーションが存在する可能性が高い。

【0162】

本発明者らは、抗血清「c」および「R」タンパク質抗原が、ある特定のタンパク質遺伝子に完全に特異的なわけではないことを見出した。しかし、「c」抗血清との反応のほとんどが、C（bca遺伝子）および関連タンパク質抗原（少なくともalp2遺伝子を含む）をコードする遺伝子、ならびにRib（rib遺伝子）をコードする遺伝子を有する「R」、および関連タンパク質（少なくともalp3遺伝子および稀な推定rib様遺伝子を含む）に対する抗血清の存在を反映していた。

【0163】

GBSの異なる血清型の多数の可動因子の存在も調査した。以前にGBS中に4つの異なる挿入配列が同定されている。いくつかの血清型III単離物中のIS861の複数のコピーが、莢膜遺伝子発現の増大に関連していた。全ての血清亜型III-1およびIII-2ならびにほとんどの血清型IIおよびIb単離物中でIS861が見出されたが、その他はほとんど見出されなかった。全IS861含有単離物は、少なくとも1つのさらなる可動因子を含んでいた。

【0164】

IS1381の複数のコピーが高い比率でGBSおよび他のStreptococcus種（S. pneumoniaeを含む）で見出されており、これは、疫学研究のためのGBSの制限フラグメント長多型（RFLP）分析用のプローブとして使用されている（Tamuraら、2000）。本発明者らは、単離物全体の85%においてIS1381を見出した。これらは血清亜型III-1の全単離物中に存在していたが、血清亜型III-2またはIII-3には存在しなかった。24種の単離物由来のIS1381配列は互いに同一であるが、以前に記載された単離物（AF064785）由来のものはいくつかの部位が異なっていた。これらの相違の有意性は不明であるが、できるだけ多数の異なる株由来の配列を確認することの重要性が強調される。

【0165】

非溶血性GBS単離物中で、溶血素の産生に関与するABC輸送体の一部である遺伝子cyIBの挿入性不活化を引き起こす、ISSa4が最初に同定された。（主に溶血性の）GBS単離物のわずかな割合（4%）にISSa4が含まれ、その全てが1996年から単離されており、ISSa4はGBSに新規に獲得されたと推測された。本発明者らも単離物中のわずかな割合（7%）でISSa4を見出したが、1996年以前（44中4）および以降（162中11）で得られた臨床単離物と類似の割合で存在していた。

【0166】

IS1548は、いくつかのヒアルロニダーゼ陰性GBS血清型III単離物中で最初に発見され、これは、遺伝子hyIB（ヒアルロニダーゼ（重要なGBS毒性因子）の産生を担うクラスターの1つ）の挿入性不活化を引き起こす（Granlundら、1998）。IS1548のコピーもまた、これを含む単離物中のC5aペプチダーゼ遺伝子（毒性にも関連する）の下流で見出される。ほとんどのIS1548含有単離物は心内膜炎患者に由来し、ヒアルロニダーゼ産生の不活化および/またはC5aペプチダーゼに対するいくつかの効果によりGBS単離物が心臓弁に接着してそこで生存することができる可能性が示唆された。

【0167】

本発明者らは、表面（12中8）および正常に滅菌した（46中22）検体由来の回収物中の58種の血清型III単離物の52%を占める全血清亜型III-1単離物中にIS1548を見出した。後者は新生児（20中7）、成人（6中3）、および年齢が特定されていない被験体（20中12）（

10

20

30

40

50

データ示さず)に由来した。特定の臨床データが利用できなかったが、GBS心内膜炎は稀であり、存在するとしてもこれらの被験体にはわずしか存在しなかった可能性が高い。この挿入配列と特定の毒性因子および臨床的な症候群との関連を解明するためのさらなる研究が必要である。

【0168】

本発明者らは、224種の単離物全体の19%においてGBSi1(グループIIイントロン)を見出した;これは、IS861とよく関連し、血清型/血清亜型によって分布が異なっていた。IIおよびIII以外の血清型で見出されることは稀であった。IS1548も含んでいた4種を含む、50%をこえる血清型II単離物に存在していた。GBSi1は、IS1548が見出されない全ての血清型III-2およびIII-4単離物中に見出されたが、IS1548を含む血清型III-1単離物またはIS548を含まない血清型III-3単離物では見出されなかった。

10

【0169】

cps遺伝子クラスター内の相違に基づいたGBS血清型IIIの4つの血清型への下位区分は、本研究に記載の表面タンパク質遺伝子プロファイルの対応する相違および5種の可動因子の分布によって支持された。本発明者らの単離物はヒアルロニダーゼ活性について試験しなかったが、Ribタンパク質を発現し、IS1548、IS861、およびIS1381を含む本発明者らの血清型III-1は、Bohnsackら、2001に記載のヒアルロニダーゼ陰性型III-2に相当する可能性が高い。本発明者らの血清型III-2はまた、Ribタンパク質を発現し、IS861およびGBSi1を含み、おそらくBohnsackら、2001の型III-3に相当する。血清型III-3およびIII-4は、比較的少数の単離物で示された。前者(いくつかの血清型Ia単離物と共通)は、C様タンパク質2を発現し、可動因子を含まなかった(別の一般的でない所見)。後者は、血清型IIと密接に関連し、cps遺伝子クラスター区分、種々の表面タンパク質プロファイル、および可動因子と配列相同性を有する。

20

【0170】

まとめ

本発明者らの目的は、B群連鎖球菌(GBS)のための包括的遺伝子型同定システムの開発である。このようなシステムは、理想的には、再現性があり、客観的であり、研究所間で移動可能であり、他の分類法と同等の相補的なものであり、公知の毒性マーカーを組み込むことができなければならない。これらの基準に基づいて、本発明者らは、最初に、cps遺伝子クラスターに基づいた分子的血清分類(MS)法を開発した。これは、従来の血清分類法(CS)と比較して優れており、感度が高く、他で記載されているような血清型(sst)I IIのいくつかの型を同定することができた。本発明者らは、可変性表面タンパク質抗原をコードする遺伝子ファミリー(bca/rib/alp2/alp3/alp4)およびIgA結合タンパク質C(bac)に基づいた第2の分子的型分類法も開発し、これは、全ての単離物を分類することができず且つ時に誤った方向に導く従来のタンパク質血清型分類よりも、高感度且つ客観的である。本発明者らの方法は、可変性抗原遺伝子ファミリーのより多数のメンバーを同定し、多数のbac亜群を区別することもできる。第3の型分類法は、GBSで同定された4つの異なる挿入配列(IS)およびII型イントロンを含む5つの可動遺伝因子(mge)を使用する。この第3の方法の使用により、本発明者らの遺伝子型同定システムの識別能力が、さらに高められる。

30

40

【0171】

本発明者らは、次いで、集団遺伝子構造および194種の侵襲性GBS単離物の遺伝子型の年齢に関する疾患分布の試験に本発明者らの分類システムを使用した。

【0172】

主に、本発明者らの遺伝子型同定システムの実用的価値を証明し、クローン集団構造を確認し、異なる患者群の遺伝子型分布を決定するために、侵襲性GBS単離物を使用した。単離物は、GBS敗血症を有する全年齢の患者に由来した。約半分は、産科ユニットを有する総合病院の大規模な診断室での血液またはCSF由来の連続GBS単離物であった(すなわち、新生児以外の幼児由来の単離物は存在しなかった)。残りは、ニュージーランド全土からの血清分類のために参照された連続単離物であった。したがって、新生児早期を超える幼

50

児がおそらく過小評価されること以外は、全体的な年齢分布が、GBS疾患に罹患した集団の分布を表す。しかし、各年齢群内の遺伝子型の分布が示すべきである。

【0173】

194種のオーストラレーシア侵襲性GBS単離物のうち、本発明者らは56種の遺伝子型を同定し、そのうちの5種（Ia-1、Ib-1、III-1、III-2、およびV-1）が単離物の62%を占めていた。

【0174】

本発明者らの結果に由来する樹状図により、cps血清型とタンパク質遺伝子プロファイル（pgp）との関係が示された。本発明者らの結果はまた、特定の公知の毒性マーカー-C、C変異型、およびヒアルロニダーゼ産生（間接的）が異なるクローン系統と関連したことを示す。

10

【0175】

3つの遺伝子マーカー組に基づく本発明者らの遺伝子型同定システムは識別力が高い。抗原性組成物を含む有用な表現型データが得られるので、特に潜在的なGBSワクチンの使用に関するGBSの疫学的調査に有用である。推定高毒性遺伝子型と患者の特徴（年齢および/または基本的な危険因子）との関係、CSF単離物（または遺伝子型）と他の侵襲性またはコロニー化株との間に有意差があるかどうかについての研究が、本発明者らの遺伝子型同定によって容易になる。このシステムを使用して、侵襲性オーストラレーシアGBS単離物の間のクローン集団構造を証明した。このシステムを、コロニー化GBS単離物に應用して、毒性マーカーを同定しうる。

20

【0176】

したがって、本発明者らは、正確且つ再現性があり、PCR/配列決定法が利用可能な任意の研究所で実施することができ、重要なことには、維持が困難になる血清型特異的抗血清のパネルを必要としない、GBSの従来血清型分類の代替法を開発した。全ての単離物が血清型分類可能であり、比較的限られた790bp領域の配列決定によりMS IIIについてのさらなる血清型分類情報を得ることができる。タンパク質プロファイリング（またはタンパク質抗原型分類）および可動遺伝子の同定（または可動遺伝子の型分類）と共に血清型同定について本発明者らが述べた分子的方法により、さらなるGBSの系統発生学研究および疫学研究ならびにGBS結合型ワクチンの導入前後にGBS単離物をモニターするために必要な疫学研究および他の関連研究に有用な包括的株同定に潜在的に有用なマーカーが得られる。

30

【0177】

上記の各節に記載の本発明の種々の特徴および態様を、必要な変更を加えて他の節に適切に適用する。したがって、1つの節で特定された特徴を他の節で特定された特徴と適切に組み合わせることができる。

【0178】

上記明細書に記載の全ての刊行物は、参照として本明細書に組み入れられる。記載された本発明の方法およびシステムの種々の修正形態および変更形態は、本発明の範囲および本旨から逸脱することなく当業者に明らかである。本発明を特定の好ましい態様と併せて記載しているが、特許請求の範囲に記載の発明がこのような特定の態様に過度に制限されないと理解すべきである。実際、分子生物学分野または関連分野の当業者に容易に明らか本発明の実施のために記載の様式の種々の修正形態は、以下の特許請求の範囲の範囲内であることが意図される。

40

【0179】

参考文献

Ahmet, Z., P. Stanier, D. Harvey, and D. Holt. 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.

Arakere, G., A.E. Flores, P. Ferrieri, and C.E. Frasch. 1999. Inhibition enzyme-linked immunosorbent assay for serotyping of group B streptococcal isolates. *J. Clin. Microbiol.* **37**:2564-2567.

Bohnsack, J. F., S. Takahashi, S. R. Detrick, L. R. Pelinka, L. L. Hammitt, A. . Aly, A. A. Whiting, and E. E. Adderson. 2001. Phylogenetic Classification of Serotype III Group B Streptococci on the Basis of *hylB* Gene Analysis and DNA Sequences Specific to Restriction Digest Pattern Type III-3. *J. Infect. Dis.* **183**:1694-1697.

10

Cropp, C.B., R.A. Zimmerman, J. Jelinkova, A.H. Auernheimer, R.A. Bolin, and B.C. Wyrick. 1974. Serotyping of group B streptococci by slide agglutination fluorescence microscopy, and microimmunodiffusion. *J. Lab. Clin. Med.* **84**:594-603.

Granlund, M., L. Oberg, M. Sellin, and M. Norgren. 1998. Identification of a novel insertion element, *IS1548*, in group B streptococci, predominantly in strains causing endocarditis. *J. Infect. Dis.* **177**:967-976

20

Hakansson, S., L.G. Burman, J. Henrichsen, and S.E. Holm. 1992. Novel coagglutination method for serotyping group B streptococci. *J. Clin. Microbiol.* **30**:3268-3269.

Harrison, L.H., J.A. Elliott, D.M. Dwyer, J.P. Libonati, P. Ferrieri, L. Billmann, and A. Schuchat. 1998. Serotype distribution of invasive group B streptococcal isolates in Maryland: implications for vaccine formulation. Maryland Emerging Infections Program. *J. Infect. Dis.* **177**:998-1002.

30

Hassan, A.A., A. Abdulmawjood, A.O. Yildirim, K. Fink, C. Lammler, and R. Schlenstedt. 2000. Identification of streptococci isolated from various sources by determination of *cfb* gene and other CAMP-factor genes. *Can. J. Microbiol.* **46**:946-951.

Hickman, M.E., M.A. Rench, P. Ferrieri, and C.J. Baker. 1999. Changing epidemiology of group B streptococcal colonization. *Pediatrics.* **104**:203-209.

Holm, S.E., and S. Hakansson. 1988. A simple and sensitive enzyme immunoassay for determination of soluble type-specific polysaccharide from group B streptococci. *J. Immunol. Methods.* **106**:89-94.

Ke, D., C. Menard, F.J. Picard, M. Boissinot, M. Ouellette, P.H. Roy, and M.G. Bergeron. 2000. Development of conventional and real-time PCR assays for the rapid detection of group B streptococci. *Clin. Chem.* **46**:324-331.

Kong, F., X. Zhu, W. Wang, X. Zhou, S. Gordon, and G.L. Gilbert. 1999. Comparative analysis and serovar-specific identification of the multiple banded antigen genes of *Ureaplasma urealyticum* biovar one. *J. Clin. Microbiol.* **37**: 538-543. 10

Kong, F., S. Gordon, and G.L. Gilbert. 2000. Rapid-Cycle PCR for Detection and Typing of *Mycoplasma pneumoniae* in Clinical Specimens. *J. Clin. Microbiol.* **38**:4256-4259.

Maeland, J. A., O. G. Brakstad, L. Bevanger, and A. I. Kvam. 1997. *Streptococcus agalactiae* beta gene and gene product variations. *J. Med. Microbiol.* **46**:999-1005. 20

Maeland, J. A., O. G. Brakstad, L. Bevanger, and S. Krokstad. 2000. Distribution and expression of *bca*, the gene encoding the c alpha protein, by *Streptococcus agalactiae*. *J. Med. Microbiol.* **49**:193-198.

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

Nagano, Y., N. Nagano, S. Takahashi, K. Murono, K. Fujita, F. Taguchi, and Y. Okuwaki. 1991. Restriction endonuclease digest patterns of chromosomal DNA from group B beta-haemolytic streptococci. *J. Med. Microbiol.* **35**:297-303. 30

Rolland, K., C. Marois, V. Siquier, B. Cattier, and R. Quentin. 1999. Genetic features of *Streptococcus agalactiae* strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and *hylB* gene analysis. *J. Clin. Microbiol.* **37**:1892-1898.

Tamura, G. S., M. Herndon, J. Przekwas, C. E. Rubens, P. Ferrieri, and S. L. Hillier. 2000. Analysis of restriction fragment length polymorphisms of the insertion sequence *IS1381* in group B Streptococci. *J. Infect. Dis.* **181**:364-368.

Triscott, M.X., and G.H. Davis. 1979. A comparison of four methods for the serotyping of group B streptococci. *Aust. J. Exp. Biol. Med. Sci.* **57**:521-527. 40

Wilkinson, H.W., and M.D. Moody. 1969. Serological relationships of type I antigens of group B streptococci. *J. Bacteriol.* **97**:629-34.

Zuerlein, T.J., B. Christensen, and R.T. Hall. 1991. Latex agglutination detection of group-B streptococcal inoculum in urine. *Diagn. Microbiol. Infect. Dis.* **14**:191-194.

【 0 1 8 0 】

(表 1) 本研究で使用したGBS基準パネル

50

研究所株番号	供給源	血清型	MS/血清亜型	GenBank アクセッション番号	
基準パネル 1 ¹					
090	Channing	Ia	Ia	AF332893	
H36B	Channing	Ib	Ib	AF332903	
18RS21	Channing	II	II	AF332905	
M781	Channing	III	III-2 ³	AF332896	10
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		
基準パネル 2 ²					
NZRM 908 (NCDC SS615)	ESR	Ia	Ia	AF332894	
NZRM 909 (NCDC SS618)	ESR	Ib	Ib	AF332904	20
NZRM 910 (NCDC SS700)	ESR	Ic	Ia	AF332914	
NZRM 911 (NCDC SS619)	ESR	II	II	AF332906	
NZRM 912 (NCDC SS620)	ESR	III	III-3 ³	AF332897	
NZRM 2217 (Prague 25/60)	ESR	分類不可能 (R)	II	AF332907	30
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	

注釈

1. 基準パネル1: Lawrence Paoletti博士 (Channing Laboratory、Boston、USA) から提供 40
2. 基準パネル2: Diana Martin博士 (ESR、Porirua、Wellington、New Zealand) から提供されたNew Zealand Reference Medical Culture Collection株。
3. 配列異種性に基づいたMS III血清亜型 (詳細についてはテキストを参照のこと)。

【 0 1 8 1 】

(表2) 本研究で使用したオリゴヌクレオチドプライマー

プライマー	標的遺伝子	Tm °C ¹	GenBank	配列 ²⁻⁴
			アクセッション番号	
CFBS	<i>cfb</i>	56.7	X72754	328 GAT GTA TCT ATC TGG AAC TCT AGT G352
Sag59 ⁵	<i>cfb</i>	77.4	X72754	350 <u>G</u> TGGCTGGTGCAT <u>TTGTTAI</u> TTT CAC CAG CTG TAT TAG AAG TA 391
Sag190 ⁵	<i>cfb</i>	76.8	X72754	545 CATTAAACCGGTTTTTCATAA TAICI GTT CCC TGA ACA TTA TCT TTG AT 500
CFBA	<i>cfb</i>	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 G1462
23SA	23S rRNA	65.7	X68427	70 CGT CGT TTG TCA CGT CCT TC 51
DSF2 ⁶	16S rRNA	75.9	AB023574	975 <u>CATCCTTC</u> TGACC GGC CTA GAG ATA GGC TTT CT 1007
DSR1 ⁶	16S rRNA	81.5	AB023574	1250 <u>CGICACCGG</u> CTT GCG ACT CGT TGT ACC AA 1222
cpsDS	<i>cpsD</i>	69.1	AB028896 (Ia), AF163833 (III)	4892/4593 GCA AAA GAA CAG ATG GAA CAA AGT GG 5007/4618
cpsES	<i>cpsE</i>	65.7	AB028896 (Ia), AF163833 (III)	5300/4910 CTT TTG GAG TCG TGG CTA TCT TG 5322/4932
cpsEA1	<i>cpsE</i>	65.4	AB028896 (Ia), AF163833 (III)	5431/5041 GA/T/GA AAA AAG GAA AGT CGT GTC G/ATT G 5612/5017

cpsES1	cpsE	65.9	AB028896 (Ia), AF163833 (III)	5612/5222CTT GGA C/TTT CTC TGA AAA GGA TTG5635/5245
cpsEA2	cpsE	66.8	AB028896 (Ia), AF163833 (III)	5723/5333AAA A/CGC TTG ATC AAC AGT TAA GCA GG5698/5308
cpsES2	cpsE	70.2	AB028896 (Ia), AF163833 (III)	6012/5622GAT GGT/C GGA CCG GCT ATC TTTT TCT C6036/5646
cpsEA3	cpsE	63.7	AB028896 (Ia), AF163833 (III)	6116/5726CTT AAT TTG TTC TGC ATC TAC TCG C6092/5702
cpsES3	cpsE	71.5	AB028896 (Ia), AF163833 (III)	6410/6020GTT AGA TGT TCA ATA TAT CAA TGA ATG GTC TAT TTG GTC AG6450/6060
cpsEFA	cpsE/F スベ一サ一	62.1	AB028896 (Ia), AF163833 (III)	6526/6136CCT TTC AAA CCT TAC CTT TAC TTA GC6501/6111
cpsFS	cpsF	75.0	AB028896 (Ia), AF163833 (III)	6777/6387CAT CTG GTG CCG CTG TAG CAG TAC CAT T6804/6414
cpsFA	cpsF	73.2	AB028896 (Ia), AF163833 (III)	6859/6469GTC GAA AAC CTC TAT A/GT A AAC/T GGT CTT ACA A/GCC AAA TAA CTT ACC6819/6425
cpsGA	cpsG	54.7	AB028896 (Ia), AF163833 (III)	7162/6772AAG/C AGT TCA TAT CAT CAT ATG AGA G 7138/6748
cpsGA1	cpsG	74.5	AB028896 (Ia), AF163833 (III)	7199/6809CCG CCA/G TGT GTG ATA ACA ATC TCA GCT TC7171/6781
cpsGS	cpsG	72.24	AB028896 (Ia), AF163833 (III)	7145/6755ATG ATG ATA TGA ACT CTT ACA TGA AAG AAG CTG AGA TTG 7183/6793
cpsGS1	cpsG	71.62	AB028896 (Ia), AF163833 (III)	7155/6765GAA CTC TTA CAT GAA AGA AGC TGA GAT TGT TAT CAC AC 7192/6802

lacpsHS	cpsH	73.6	AB028896 (1a)	7698CAT TCT TTG TTT AAA AA/CT CCT GAT TTT GAT AGA ATT TTA GCA GC7741
lacpsHA	cpsH	75.2	AB028896 (1a)	7993GAA TAT TCA AAA AAT CCC ATT GCT CTT TGA GTA TGC ATA CC7953
lacpsHA1	cpsH	66.4	AB028896 (1a)	8271GTA AGT TAT CAA AAT ATA ACA TCA TTA CTA TTA CTA GTA GAA ACG G8226
lacpsHS1	cpsH	77.9	AB028896 (1a)	8463GGC CTG CTG GGA TTA ATG AAT ATA GTT CCA GGT TTG C8499
lacpsHA2	cpsH	58.5	AB028896 (1a)	8499GCA AAC CTG GAA CTA TAT TCA T8478
lbcpsHS0	cpsH	58.6	AB050723 (1b)	3013ATT GCT GCA TTC AAT TCA C3031
lbcpsHS	cpsH	81.9	AB050723 (1b)	3016GCT GCA TTC AAT TCA CTG GCA GTA GGG GTT GTG TCC3051
lbcpsHA	cpsH	67.7	AB050723 (1b)	3297GAT AGT TAA GGG TAT TAT AAG ATT TGA ATA TTC AAA GAA AGC3256
lbcpsHS1	cpsH	74.1	AB050723 (1b)	3546TTT GGT GAG CAT ATA TAA TAG AAT AAT CAA TTT GCG GTC G3585
lbcpsHS2	cpsH	73.7	AB050723 (1b)	3740CTG GCC TAT TTG GAC TAA TAA ATG TGA TTT TAG GTT TGT TTC3781
lbcpsHA01	cpsH	57.7	AB050723 (1b)	3781GAA ACA AAC CTA AAA TCA CAT TTA3758

IbcpsHA1	<i>cpsH</i>	78.5	AB050723 (Ib)	3894 GGC GCC ATC AAT ATC TTC AAG TGC AAA AAA TGA AAA TAG G3855
IbcpsIA	<i>cpsI</i>	78.2	AB050723 (Ib)	4086 CTA TCA ATG AAT GAG TCT GTT GTA GGA CGG ATT GCA CG 4049
IbcpsIS	<i>cpsI</i>	71.1	AB050723 (Ib)	4116 GAT AAT AGT GGA GAA ATT TGT GAT AAT TTA TCT CAA AAA GAC G4158
IbcpsIA1	<i>cpsI</i>	78.6	AB050723 (Ib)	4638 CCT GAT TCA TTG CAG AAG TCT TTA CGA TGC GAT AGG TG 4601
IIIVlcpsHS	<i>cpsH</i>	75.3	AF163833 (III), AF337958 (VI)	7275/7120 CAA GAG GAT ATA ACG TTT CAG CGA TTT ATT GCT GAG C7311/7156
IIlcpsHS	<i>cpsH</i>	72.1	AF163833 (III)	7672 GAA TAC TAT TGG TCT GTA TGT TGG TTT TAT TAG CAT CGC 7710
IIlcpsHA	<i>cpsH</i>	71.0	AF163833 (III)	7817 GTT ATA AGA AAA ACA AGCGGT GAT AAA TAA GAA AGT CAT ACC 7776
IVcpsHS	<i>cpsH</i>	74.1	AF355776 (IV)	7552 CCG TAC ATA CAA CTG TTC TTG TTA GCA TTT ACT TTT CTT TGC 7593
IVcpsHS1	<i>cpsH</i>	71.2	AF355776 (IV)	7887 CCC AAG TAT AGT TAT GAA TAT TAG TTG GAT GGT TTT TGG 7925
IVcpsHA	<i>cpsH</i>	77.3	AF355776 (IV)	7951 CAT CTA CAC CCC CAC AAA ATA TTT TCC CAA AAA CCA TC 7914
IVcpsHA1	<i>cpsH</i>	58.7	AF355776 (IV)	7958 TGT AAA TCA TCT ACA CCC CC 7939

IVcpsMA	cpsM	80.7	AF355776 (IV)	8265GGG TCA ATT GTA TCG TCG CTG TCA ACA AAA CCA ATC AAA TC8225
VcpsHS	cpsH	76.3	AF349539 (V)	6943GGG TTT AGG CGA GGG AAA CTC AGC TTA CAA AAT AGT G6979
VcpsHS1	cpsH	72.2	AF349539 (V)	7258CAA TTT TTA TAG GGA TGG ACA ATT TAT TCT GAG AAG TGA C7297
VcpsHA	cpsH	71.1	AF349539 (V)	7291TCT CAG AAT AAA TTG TCC ATC CCT ATA AAA ATT GAC ATA C7252
VcpsHS02	cpsH	59.0	AF349539 (V)	7616GAT GTT CTT TTA ACA GGT AGA TTA CAC7642
VcpsHA1	cpsH	66.8	AF349539 (V)	7658GTT GTA AAT GAG CAT AGT GTA ATC TAC CTG TTA AAA GAA C7619
VcpsHS2	cpsH	74.0	AF349539 (V)	7871CCC AGT GTG GTA ATG AAT ATT AGT TGG CTA GTT TTT GG7908
VcpsHA2	cpsH	58.6	AF349539 (V)	7945CTT TTT TAT AGG TTC GAT ACC ATC7922
VcpsMA	cpsM	73.1	AF349539 (V)	8244CCC CCC ATA AGT ATA AAT AAT ATC CAA TCT TGC ATA GTC AG8204
VlcpsHS	cpsH	76.7	AF337958 (VI)	7478CAC TAT TCC TAG TTT TTT GTG CAT ATT TGA CAG GGG CAA G7517
VlcpsHA	cpsH	76.7	AF337958 (VI)	7517CTT GCC CCT GTC AAA TAT GCA CAA AAA ACT AGG AAT AGT G7478

VlcpsHS1	cpsH	77.2	AF337958 (VI)	7767CCT TAT TGG GCA AGG TAT AAG AGT TCC CTC CAG TGT G7803	10
VlcpsHA1	cpsH	77.2	AF337958 (VI)	7804CCA CAC TGG AGG GAA CTC TTA TAC CTT GCC CAA TAA G7768	20
VlcpsIA	cpsI	74.5	AF337958 (VI)	8126GAA GCA AAG ATT CTA CAC AGT TCT CAA TCA CTA ACT CCG8088	30
cpsIA	cpsI	70.3	AB028896 (Ia), AF163833 (III)	8816/8312GTA TAA CTT CTA TCA ATG GAT GAG TCT GTT GTA GTA CCG8778/8274	40

注釈

1. プライマーの T_m 値は、プライマー合成機 (Sigma-Aldrich) によって得られる。
2. 数字は、プライマー配列が開始および終結する番号付けした塩基の位置を示す (開始点「1」の番号付けは、対応する遺伝子GenBankアクセッション番号の開始点「1」をいう)。
3. 下線を引いた配列は、以前に公開されたプライマーを改変するために付加した塩基を示す。
4. 「/」の後ろの文字は、異なる血清型における別のヌクレオチドを示す。
5. Keら、2000。

6. Ahmet ら、1999。

【 0 1 8 2 】

(表3) 異なるオリゴヌクレオチドプライマー対を使用したアンプリコンの特異性および予想される長さ

プライマー対*	特異性	アンプリコンの長さ (塩基対)	
Sag59/Sag190 ^a	GBS (<i>S. agalactiae</i>)	196	
CFBS/CFBA	GBS (<i>S. agalactiae</i>)	241	10
16SS/23SA	GBS (<i>S. agalactiae</i>)	433	
DSF2/DSR1 ^a	GBS (<i>S. agalactiae</i>)	276	
cpsDS/cpsEA1	血清型 Ia ~ VII	449/458	
cpsES/cpsEA2	血清型 Ia ~ VII	424	
cpsES1/cpsEA3	血清型 Ia ~ VII	505	
cpsES2/cpsEFA	血清型 Ia ~ VII	515	
cpsES3/cpsFA ^b	血清型 Ia ~ VII	450	
cpsFS/cpsGA1 ^b	血清型 Ia ~ VII	423	
cpsES3/cpsGA1 ^b	血清型 Ia ~ VII	790	20
cpsGS/cpsIA	血清型 Ia および III	1672/1558	
cpsGS1/cpsIA	血清型 Ia および III	1662/1548	
cpsGS/lacpsHA1	血清型 Ia	1127	
cpsGS1/lacpsHA1	血清型 Ia	1117	
lacpsHS/lacpsHA	血清型 Ia	296	
lacpsHS/lacpsHA1	血清型 Ia	574	
lacpsHS1/cpsIA ^c	血清型 Ia	354	
cpsGS/lbcpsHA1	血清型 Ib	1468	
cpsGS1/lbcpsHA1	血清型 Ib	1458	30
cpsGS/lbcpsIA	血清型 Ib	1660	
cpsGS1/lbcpsIA	血清型 Ib	1650	
lbcpsHS/lbcpsHA	血清型 Ib	282	
lbcpsHS1/lbcpsHA1	血清型 Ib	349	
lbcpsHS2/lbcpsIA	血清型 Ib	347	
lbcpsIS/lbcpsIA1 ^c	血清型 Ib	523	
cpsGS/IIlcpSHA	血清型 III	1063	
cpsGS1/IIlcpSHA	血清型 III	1053	
IIIVlcpHS/IIlcpSHA	血清型 III	543	40
IIlcpHS/cpsIA ^c	血清型 III	641	
cpsGS/IVcpSHA	血清型 IV	1372	
cpsGS1/IVcpSHA	血清型 IV	1362	
cpsGS/IVcpSMA	血清型 IV	1686	

cpsGS1/IVcpsMA	血清型 IV	1676	
IVcpsHS/IVcpsHA	血清型 IV	400	
IVcpsHS1/IVcpsMA ^c	血清型 IV	379	
cpsGS/VcpsHA1	血清型 V	1096	
cpsGS1/VcpsHA1	血清型 V	1086	
cpsGS/VcpsMA	血清型 V	1682	
CpsGS1/VcpsMA	血清型 V	1672	
VcpsHS/VcpsHA	血清型 V	349	10
VcpsHS1/VcpsHA1	血清型 V	401	
VcpsHS2/VcpsMA ^c	血清型 V	374	
IIIVlcpsHS1/VlcpsHA	血清型 VI	398	
cpsGS/VlcpsHA1	血清型 VI	1205	
cpsGS1/VlcpsHA1	血清型 VI	1195	
cpsGS/VlcpsIA	血清型 VI	1527	
cpsGS1/VlcpsIA	血清型 VI	1517	
VlcpsHS/VlcpsHA1 ^c	血清型 VI	327	
VlcpsHS1/VlcpsIA	血清型 VI	360	20

注釈

* プライマー配列については表2およびいくつかのプライマー部位については図1を参照のこと。分子的血清型同定のためのアルゴリズムで使用したプライマー-図2。a.GBSの同定のため、b.配列決定のため、c.MS特異的PCRのため。

【 0 1 8 3 】

(表4) cpsD-cpsE-cpsFの3'末端領域およびcpsGの5'末端領域中の8種のGBS血清型の異種性

部位	la	lb	III/III-4	III	IV	V	VI	VII	特異性
cpsD 遺伝子									
62	G	A	G ⁴	A	A	A	A	G	la, II, VII
78-86 反復配列	- la-2 ¹ ; + la-1 ¹	-	- II-2 ^{2,4} ; + II-1 ²	- III-2 ³ ; + III-1 ³ , III-3 ³	+	+	-	+	ラキストを参照のこと
- TTACGGCGA									
cpsD/cpsE 遺伝子 スペーサー									
138	G	G	G	G	G	A ⁵	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 遺伝子									
198	A	C	A ⁴	A	C	C ⁵	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 ⁴	T	C	C ⁵	T	T	T	Ib, IV, V
300	C	C	C	T III-2;	C	C	T III-2;	C	C	III-2
321	C	C	C	C III-1, III-3	C	C	C III-1;	C	C	III-1
419	T	C	T ⁴	T	T	T	T	T	T	Ib
429	A	T	A ⁴	T	T	T	T	A	A	Ia, II, VII
437	C	C	C;	C	C	C	C	T	T	VII, III-4
457	T	A	T III-4	A	A	A	A	C	C	Ia, II, VII
466	G	G	G	G	A	G	G	A	A	IV
486	G	A	A	G III-3;	A	A	A	A	A	Ia, III-3
602	G	G	A ⁴	A III-2, III-1	G	G	G	A	A	II, VII
606	T	T	T	G	T	T	T	C	T	VI
627	T	C	C	C	C	C	C	C	C	Ia
636	C	T	T	C III-1;	T	T	T	T	T	Ia, III-1
645	C	T	C ⁴	T III-2, III-3	T	T	T	C	C	Ib, IV, V
803	A	A	A	A	A	A	T	A	A	VI
971	C	T	T	C	C	C	T	T	T	Ia, III, IV, V
1026	A	G	G	G III-2, III-1;	A	A	A	G	G	Ia, III-3, IV, V
1044	T	T	T	A III-3	T	T	T	C	T	VI

10

20

30

1173	A	G	A	A	A	A	A	A	A	lb
1194	C	C	C	A	A	C	A	C	C	III, IV, VI
1251	G	G	G	G	G	G	A	G	G	VI
1278	A	A	A	A	A	A	A	A	A	V
1413	C	T	T	C III-3; T III-2, III-1	T	T	T	T	T	Ia, III-3
1495	C	C	C	C	C	C	A	C	C	VI
1500	A	G	A	A	A	A	A	A	A	lb
1501	C	C	T	C	C	C	C	T	T	II, VII
1512	C	T	C	T III-2, III-1; C III-3	C	T	T	C	C	Ia, II, III-3, IV, VII
1518	T	C	T	C III-2, III-1; T III-3	T	C	C	T	T	Ia, II, III-3, IV, VII
1527	T	A	A	T III-3; A III-2, III-1	T	A	A	A	A	Ia, III-3, IV
cpsF 遺伝子										
1595	T	C	T	T	T	T	T	T	T	lb, VI
1611	C	C	C	C	C	C	C	T	T	VII
1620	C	C	C	C	C	C	C	T	T	VII
1627	G	G	G	G	T	G	G	G	G	IV
1629	G	G	G	A III-1; G III-2, III-3	G	G	G	G	G	III-1
1655	C	T	C	C	C	C	C	C	C	lb
1832	C	C	C	C	T	C	C	C	C	IV

10

20

30

1856	T	T	T	T	T	T	T	T	T	lb	10
1866	G	G	G	G	G	G	G	G	A	VII	
1871	T	T	T	T	T	T	T	T	T	V	
1892	A	A	A	A	A	A	A	A	A	V	
1971	G	G	G	G	G	G	G	G	G	VI	
cpsG 遺伝子											
2026	G	G	G	G	G	G	G	G	G	lb	
2088	G	G	G	G	A	G	G	G	G	IV	
2134	T	T	T	C III-2, III-1; T III-3	T	T	T	T	T	III-2, III-1	
2187	C	C	C	C	C	C	C	C	C	VII	
2196	A	A	A	A	A	A	A	A	A	VII	
											20
											30

注釈

1. 反復配列：血清型 Ia-1あり (+)；血清型 Ia-2なし (-) (テキストを参照のこと)。
2. 反復配列：血清型 II-1あり (+)；血清型 II-2なし (-) (テキストを参照のこと)。
3. 反復配列：血清型 III-1および III-3あり (+)；血清型 III-2なし (-)；血清型 III-4可変 (テキストを参照のこと)。
4. 1つの CS II株は、9つの部位で変異を有する (テキストを参照のこと)。
5. 138位、198位、および249位で、1つの CS V基準株 (Prague10/84) が、GenBankの対応する配列 (GenBankアクセッション番号 AF349539) と同一であり、配列はそれぞれ G、A、および T である；別の CS V基準株 (CJB 111) および全ての他の配列決定した CS V株は同一で

40

50

あり、配列はそれぞれA、C、およびCである。

【0184】

(表5) 206種の臨床GBS単離物の従来の血清型分類(CS)と分子的血清型同定(MS)/亜型分類との結果比較

		MS/血清亜型										
CS		la	lb	II	III-1 ¹	III-2 ¹	III-3 ¹	III-4 ¹	IV	V	VI	VIII
la		38										
lb			30									
II				25								
III					27	20	4	3				
IV								7				
V									31			
VI											2	
VIII												1
NT ¹		2	5	1	3	1				5	1	
合計 (206) ²		40	35	26 ²	30	21 ²	4	3	7	36	3	1

10

20

注釈

1. MS III血清亜型の詳細については、テキストを参照のこと。

2. 1つの混合培養物は、2つの個別の単離物(1つが血清型II、1つが亜型III-2)として含まれていた。

【0185】

(表6) 本研究で使用したオリゴヌクレオチドプライマー

プライマー	標的遺伝子	Tm °C ¹	GenBank	配列 ^{2,3}
			アクセッション番号	
IgAagGBS ⁵	<i>bac</i>	73.8	X59771	2663 <u>GCGATTAAACAA</u> CAA ACT ATT TTT GAT A TTG ACA ATG CAA 2702
IgAS1 ⁴	<i>bac</i>	72.8	X59771	2765 GCT AAA TTT CAA AAA GGT CTA GAG ACA AAT ACG CCA G 2801
IgAA1 ⁴	<i>bac</i>	78.9	X59771	3157 CCC ATC TGG TAA CTT CGG TGC ATC TGG AAG C 3127
RigAagGBS ⁵	<i>bac</i>	76.3	X59771	3284 <u>CAGCCAACTCITTC</u> GTC GTT ACT TCC TTG AGA TGT AAC 3247
GBS1360S ⁶	<i>bac</i>	72.3	X59771	1325 <u>GIGAAATIGTAT</u> AAG GCT ATG AGT GAG AGC TTG GAG 1360
GBS1717S ⁴	<i>bac</i>	75.0	X59771	1685 ACA GTC ACA GCT AAA AGT GAT TCG AAG ACG ACG 1717
GBS1937A ⁶	<i>bac</i>	75.9	X59771	1976 <u>CCGIIIIAGAAICIII</u> CTG CTC TGG TGT TTT AGG AAC TTG 1937
BcaRUS ⁷	<i>bca</i> 反復単位	73.5	M97256	769 <u>GATAAATATGATCCAA</u> CAG GAG GGG AAA CAA CAG TAC 805
BcaRUA ⁷	<i>bca</i> 反復単位	77.2	M97256	1003 <u>CTGGIIIIIGGIGTCACAI</u> GAA CCG TTA CTT CTA CTG TAT CC 963

bcaS1 ⁴	bca/alp2/alp3	71.7	M97256 および AF291065	208/533GGT AAT CTT AAT ATT TTT GAA GAG TCA ATA GTT GCT GCA TCT AC251/576
bcaS2 ⁴	bca/alp2/alp3	78.0	M97256 および AF291065	256/581CCAGGGA GTG CAG CGA CCT TAA ATA CAA GCA TC288/613
bcaS ⁴	bca	58.9	M97256	370GTT TTA GAA CAA GGT TTT ACA GC392
baIS ⁴	alp2/alp3	73.8	AF291065	677GAT CCT CAA AAC CTC ATT GTA TTA AAT CCA TCA AGC TAT TC717
bcaA ⁴	bca	74.2	M97256	597CGTTCTAACTT CTT CAA TCT TAT CCC TCA AGG TTG TTG560
baIA ⁴	alp2/alp3	73.6	AF291065	978CCA GTT AAG ACT TCA TCA CGA CTC CCA TCA C948
baI23S1 ⁴	alp2/alp3	70.9	AF208158 および AF291065	1093/1373CAG ACT GTT AAA GTG GAT GAA GAT ATT ACC TTT ACG G1129 /1409
baI23S2 ⁴	alp2/alp3	72.9	AF208158 および AF291065	1174/1454CTT AAA GCT AAG TAT GAA AAT GAT ATC ATT GGA GCT CGT G1213/1493
baI2S ⁴	alp2	59.2	AF208158	1363GTT CTT CCG CCA GAT AAA ATT AAG1386
baI2A ⁴	alp2	58.3	AF208158	1576CTG TTG ACT TAT CTG GAT AGG TC1554
baI2A1 ⁴	alp2	78.3	AF208158	1426CGT GTT GTT CAA CAG TCC TAT GCT TAG CCT CTG GTG1391
baI2A2 ⁴	alp2	70.8	AF208158	1518GGT ATC TGG TTT ATG ACC ATT TTT CCA GTT ATA CG1484

10

20

30

40

bal3S ⁴	<i>alp3</i>	57.1	AF291065	1643GTT CTT CCG CTT AAG GAT AGC A1664	
bal3A ⁴	<i>alp3</i>	79.2	AF291065	1693GAC CGT TTG GTC CTT ACC TTT TGG TTC GTT GCT ATC C1657	
#ribS1 ⁴	<i>rib</i>	65.2	U58333	216TAC AGA TAC TGT GTT TGC AGC TGA AG241	
ribS2 ⁴	<i>rib</i>	73.0	U58333	238GAAGTAATTCAG GAA GTG CTG TTA CGT TAA ACA CAA ATA TG279	
ribA1 ⁴	<i>rib</i>	78.8	U58333	431GAA GGT TGT GTG AAA TAA TTG CCG CCT TGC CTA ATG396	
ribA2 ⁴	<i>rib</i>	72.6	U58333	462AAT ACT AGC TGC ACC AAC AGT AGT CAA TTC AGA AGG427	
#ribA3 ⁴	<i>rib</i>	61.3	U58333	570CAT CTA TTT TAT CTC TCA AAG CTG AAG554	

注釈

配列決定のみでの使用については、*rib*遺伝子に完全に特異的なわけではない。

1. プライマーのTm値は、プライマー合成機 (Sigma-Aldrich) によって得られる。
2. 数字は、プライマー配列が開始および終結する番号付けした塩基の位置を示す (開始点「1」の番号付けは、いくつかの配列に対しては2つ存在する、対応するGenBankアクセス番号の開始点「1」をいう)。

3. 下線を引いた配列は、以前に公開されたプライマーを改変するために付加した塩基を示す。

4. 本研究での使用のためにデザインしたプライマー。

5. Mawnら、1993。

6. Maelandら、1997。

7. Maelandら、2000。

【0186】

(表7) 異なるプライマー対を使用したアンプリコンの特異性および予想される長さ

プライマー対*	特異性	アンプリコンの長さ (塩基対)	タンパク質 プロフィール コード	
IgAagGBS/ RIgAagGBS	<i>bac</i>	532-838	B	10
IgAS1/IgAA1	<i>bac</i>	303-591	B	
GBS1360S/ GBS1937A	<i>bac</i>	652	B	
GBS1717S/ GBS1937A	<i>bac</i>	292	B	20
bcaS1/bcaA	<i>bca</i> の5'末端	390	A	
bcaS2/bcaA	<i>bca</i> の5'末端	342	A	
BcaRUS/bcaRUA	<i>bca</i> 反復単位/ <i>bca</i> 反復単位様領域	235	a/as	
bcaS1/balA	<i>alp2/alp3</i>	446	<i>alp2</i> または <i>alp3</i>	
bcaS2/balA	<i>alp2/alp3</i>	398	<i>alp2</i> または <i>alp3</i>	
balS/balA	<i>alp2/alp3</i>	302	<i>alp2</i> または <i>alp3</i>	30
bal23S1/bal2A1	<i>alp2</i>	334	<i>alp2</i>	
bal23S2/bal2A1	<i>alp2</i>	253	<i>alp2</i>	
bal23S1/bal2A2	<i>alp2</i>	426	<i>alp2</i>	
bal23S2/bal2A2	<i>alp2</i>	345	<i>alp2</i>	
bal23S1/bal3A	<i>alp3</i>	321	<i>alp3</i>	
bal23S2/bal3A	<i>alp3</i>	240	<i>alp3</i>	
#ribS1/ribA3	<i>rib/rib</i> 様	355	R/r	
ribS2/ribA1	<i>rib</i>	194	R	
ribS2/ribA2	<i>rib</i>	225	R	40
ribS2/ribA3	<i>rib</i>	333	R	

注釈

* プライマー配列については表6を参照のこと。

配列決定のみでの使用については、*rib*遺伝子に完全に特異的なわけではない(詳細については、テキストを参照のこと)。

【0187】

(表8) アンプリコン長(プライマーIgAagGBS/RIgAagGBSを使用)および配列異種性に基
づいた*bac*遺伝子(C タンパク質遺伝子)の遺伝子群および亜群

群または亜群	N=	アンプリコン 長	GenBank アクセッション番号	主な群と比較した (c.f.)異なる 部位の数	分子的血清型 / 血清亜型	
B1	19	532	X58470		17 = lb; 2 = II	
B1a	1	532	AF362686	1 (c.f. B1)	lb	
B2	3	550	AF362687		lb, II, III-4	
B3	2	586	AF362688		2=lb	10
B3a	1	586	AF362689	4 (c.f. B3)	V	
B3b	1	586	AF362690	21 (c.f. B3)	VI	
B3c	1	586	AF362691	24 (c.f. B3)	lb	
B4	8	604	AF362692		4 = lb; 4 = II	
B4a	1	604	AF362693	1 (c.f. B4)	II	
B4b	2	604	AF362694	2 (c.f. B4)	2 = lb	
B5	2	622	AF362695		la, VI	
B5a	1	622	AF362696	2 (c.f. B5)	la	
B6	1	640	AF362697		lb	20
B7	1	658	AF362698		lb	
B7a	1	658	AF362699	34 (c.f. B7)	VI	
B8	1	712	AF362700		lb	
B9	2	748	AF362701		2 = II	
B9a	1	748	AF362702	13 (c.f. B9)	lb	
B10	2	820	AF362703		2 = lb	
B11	1	838	AF362704		lb	

注釈

* タンパク質抗原との血清型 / 血清亜型の関係の詳細については表9を参照のこと。

【 0 1 8 8 】

(表9) GBSタンパク質遺伝子プロフィールと莢膜多糖類 (cps) 分子的血清型 / 血清亜型との関係

血清型 / 血清亜型*	N=	なし	Aa	AaB	R	alp 3	a	as	alp2as	RB	R a
la	43	-	-	2	-	-	35	3	3	-	-
lb	37	-	1	35	-	1	-	-	-	-	-
II	29	-	3	10	8	2	5	-	-	-	1
III-1	30	-	-	-	30	-	-	-	-	-	-
III-2	22	-	-	-	22	-	-	-	-	-	-
III-3	5	-	-	-	-	-	-	-	5	-	-
III-4	3	-	-	1	-	1	-	-	1	-	-
IV	9	-	-	-	1	-	8	-	-	-	-
V	38	1	-	-	1	35	-	-	-	1	-
VI	5	-	1	3	-	-	1	-	-	-	-
VII	1	-	-	-	-	1	-	-	-	-	-
VIII	2	1	-	-	-	1	-	-	-	-	-
合計	224	2	5	51	62	41	49	3	9	1	1

10

20

注釈

* cps血清亜型の説明についてはテキストを参照し、タンパク質抗原遺伝子プロフィールコードの説明については表7を参照のこと。

【 0 1 8 9 】

(表 1 0) 本研究で使用したオリゴヌクレオチドプライマー

プライマー	標的	Tm °C ¹	GenBank アクセッション番号	配列 ²	
IS861S	IS861	77.4	M22449	445 GAG AAA ACA AGA GGG AGA CCG AGT AAA ATG GGA CG479	
IS861A1	IS861	77.3	M22449	831 CAC GAT TTC GCA GTT CTA AAT AAA TCC GAC GAT AGC C795	10
IS861A2	IS861	76.1	M22449	1020 CAA ACT CCG TCA CAT CGG TAT AGC ACT TCT CAT AGG985	
IS1548S	IS1548	76.5	Y14270	143 CTA TTG ATG ATT GCG CAG TTG AAT TGG ATA GTC GTC178	
IS1548S1	IS1548	77.0	Y14270	539 GTT TGG GAC AGG TAG CGG TTG AGG AGA AAA GTA ATG574	20
IS1548A1	IS1548	77.0	Y14270	574 CAT TAC TTT TCT CCT CAA CCG CTA CCT GTC CCA AAC539	
IS1548A2	IS1548	70.3	Y14270	915 CCC AAT ACC ACG TAA CTT ATG CCA TTT G888	
IS1548A3	IS1548	78.0	Y14270	930 CGT GTT ACG AGT CAT CCC AAT ACC ACG TAA CTT ATG CC893	30
IS1381S1	IS1381	80.1	AF064785/ AF367974	272/818 CTT ATG AAC AAA TTG CGG CTG ATT TTG GCA TTC ACG 307/853	
IS1381S2	IS1381	81.7	AF064785/ AF367974	497/1040 GGC TCA GGC GAT TGT CAC AAG CCA AGG GAG526/1069	40

IS1381A	IS1381	73.1	AF064785/ AF367974	881/1424 CTA AAA TCC TAG TTC ACG GTT GAT CAT TCC AGC849/1392	
ISSa4S	ISSa4	78.5	AF165983	326 CGT ATC TGT CAC TTA TTT CCC TGC GGG TGT CTC C359	
ISSa4A1	ISSa4	75.2	AF165983	639 GCC GAT GTC ACA ACA TAG TTC AGG ATA TAG CCA G606	10
ISSa4A2	ISSa4	74.5	AF165983	780 CGT AAA GGA GTC CAA AGA TGA TAG CCT TTT TGA ACC745	
GBSi1S1	GBSi1	78.6	AJ292930	721 CAT CTC GGA ACA ATA TGC TCG AAG CTT ACA AGC AAG TG758	
GBSi1S2	GBSi1	77.3	AJ292930	789 GGG GTC ACT ATC GAG CAG ATG GAT GAC TAT CTT CAC824	20
GBSi1A1	GBSi1	83.9	AJ292930	1058 AAT GGC TGT TTC GCA GGA GCG ATT GGG TCT GAA CC1024	
GBSi1A2	GBSi1	80.5	AJ292930	1161 CCA GGG ACA TCA ATC TGT CTT GCG GAA CAG TAT CG1127	30

注釈

1. プライマーのT_m値は、プライマー合成機 (Sigma-Aldrich) によって得られた。
2. 数字は、プライマー配列が開始および終結する番号付けした塩基の位置を示す (開始点「1」の番号付けは、対応する遺伝子GenBankアクセッション番号の開始点「1」をいう)

【 0 1 9 0 】

(表11) 異なるオリゴヌクレオチドプライマー対を使用したアンプリコンの特異性および予想される長さ

プライマー対*	特異性	アンプリコン長 (塩基対)	
IS861S/IS861A1	IS861	387	
IS861S/IS861A2	IS861	576	
IS1548S/IS1548A1	IS1548	432	
IS1548S/IS1548A2	IS1548	773	
IS1548S/IS1548A3	IS1548	788	10
IS1548S1/IS1548A2	IS1548	377	
IS1548S1/IS1548A3	IS1548	392	
IS1381S1/IS1381A	IS1381	610/607#	
IS1381S2/IS1381A	IS1381	385	
ISSa4S/ISSa4A1	ISSa4	314	
ISSa4S/ISSa4A2	ISSa4	455	20
GBSi1S1/GBSi1A1	GBSi1	338	
GBSi1S1/GBSi1A2	GBSi1	441	
GBSi1S2/GBSi1A1	GBSi1	270	
GBSi1S2/GBSi1A2	GBSi1	373	

注釈

* プライマー配列については表10を参照のこと。

本発明者らの配列決定結果 (GenBankアクセッション番号AF367974) は、Tamuraら、2000 30
0によって以前に記載された結果 (GenBankアクセッション番号AF064785) よりも3bp短か
った。

【0191】

(表12) 可動遺伝因子と莢膜多糖類血清型、血清型III亜型と表面タンパク質遺伝子プ
ロファイルとの関係

血清型 / 血清亜型	タンパク質 遺伝子 プロフィール	N=	IS861	IS1548	IS1381	ISSa 4	GBSi1	可動因子数	
la	AaB	2	2	-	2	-	-	-	
la	alp2as	3	-	-	-	-	-	3	
la	a	35	3	1	35	1	-	-	
la	as	3	-	-	3	-	-	-	
小計		43	5	1	40	1	-	3	10
lb	Aa	1	-	-	-	-	-	1	
lb	AaB	35	30	-	35	1	-	-	
lb	alp3	1	-	-	1	-	-	-	
小計		37	30	-	36	1	-	1	
II	Aa	3	3	1	3	2	1	-	
II	AaB	10	10	5	10	5	1	-	
II	alp3	2	1	1	2	-	-	-	
II	R	8	8	-	8	-	8	-	
II	Ra	1	1	-	-	-	1	-	20
II	a	5	2	2	5	3	5	-	
小計		29	25	9	28	10	16	-	
III-1	R	30	30	30	30	1	-	-	
III-2	R	22	22	-	-	-	22	-	
III-3	alp2as	5	-	-	-	-	-	5	
III-4	AaB	1	1	-	1	-	1	-	
III-4	alp2as	1	-	-	-	-	1	-	
III-4	alp3	1	-	-	1	-	1	-	
小計		60	53	30	32	1	25	5	30
IV	R	1	1	-	1	-	1	-	
IV	a	8	2	-	8	-	-	-	
小計		9	3	-	9	-	1	-	
V	alp3	35	3	1	35	1	1	-	
V	R	1	1	-	1	1	-	-	
V	RB	1	1	-	1	-	-	-	
V	なし	1	-	-	-	-	-	1	
小計		38	5	1	37	1	1	2	

VI	Aa	1	-	-	1	-	-	-
	AaB	3	3	-	3	-	-	-
	a	1	-	-	1	-	-	-
小計		5	3	-	5	-	-	-
VII	alp3	1	-	-	1	-	-	-
VIII	alp3	1	-	-	1	-	-	-
	none	1	-	-	1	-	-	-
小計		2	-	-	2	-	-	-
合計		224	124	41 (18)	190	15 (7)	43 (19)	10 (4)
			(55)		(85)			

10

注釈

A : bca遺伝子 (C タンパク質) の5' 末端 ;

a : bca遺伝子反復単位またはbca遺伝子反復単位様配列 (多バンドアンプリコン) ;

as : bca遺伝子反復単位またはbca遺伝子反復単位様配列 (単一バンドアンプリコン) ;

B : C /IgA結合タンパク質 (bac) 遺伝子 ;

R : Ribタンパク質 (rib) 遺伝子 ;

alp2 : C 様タンパク質2 (alp2) 遺伝子 ;

alp3 : C 様タンパク質3 (alp3) 遺伝子 ;

r : 推定Rib様タンパク質遺伝子。

20

【 0 1 9 2 】

(表 1 3) 194種の侵襲性GBS単離物の可動遺伝因子の分布

存在する可動遺伝因子						
合計 N =	IS1381	IS861	IS1548	ISSa4	GBSi1	なし
6	-	-	-	-	-	6
78	78	-	-	-	-	-
2	-	-	-	-	2	-
37	37	37	-	-	-	-
1	1	-	1	-	-	-
3	3	-	-	3	-	-
29	29	29	29	-	-	-
6	6	6	-	6	-	-
8	8	8	-	-	8	-
18	-	18	-	-	18	-
1	1	-	-	-	1	-
1	1	-	1	-	1	-
2	2	2	2	-	2	-
2	2	-	-	2	2	-
合計	168 (87%)	100 (52%)	33 (17%)	11 (6%)	34 (18%)	6 (3%)
	(n=194)					

30

40

注釈

データは、種々のmge組み合わせを含む単離物数である。

【 0 1 9 3 】

(表 1 4) GBS遺伝子型と侵襲性疾患年齢との関係

50

血清型 遺伝子型	年齢群/疾患 ¹						合計
	0-6d	7-3m	4m-14yr	15-45 yr	46-60 yr	>60 yr	
Ia-1	14	4+1	1	7	3	6	35+1 (19%)
Ia-(2-8)	4	2	-	1	-	3	10
Ia 合計	18 (34%)	6+1 (21%)	1 (10%)	8 (28%)	3 (18%)	9 (17%)	45+1 (24%)
<i>Ib-1</i>	2	1+1	-	3	2	5+1	13+2
<i>Ib-(2-16)</i>	3	4+2	-	3	1	5	16+2
Ib 合計	5 (9.4%)	5+3 (24%)	-	6 (21%)	3	10+1	29+4 (17%)
II	8 (15%)	1 (3%)	-	4+1 (17%)	1	4 (7%)	18+1 (10%)
III-1	6+1 (13%)	4 (12%)	1+1 (20%)	1+1 (7%)	6+1 (41%)	4	22+4 (13%)
III-2	5 (9%)	5+4 (39%) ³	1 (10%)	2	-	-	13+4 (9%)
III-(3-4)	1+1	1	-	1	1	1	5+1
III 合計	12+2 (26%)	10+4 (41%)	2+1 (30%)	4+1 (17%)	7+1 (44%)	5 (9%)	40+9 (25%)
IV 合計	3	-	-	-	-	4	7 (4%)
V-1	3	3	2	4	2	13+1	27+1 (14%)
V-(2-7)	1	1	-	1	-	4	7
V 合計	4 (8%)	4 (12%)	2 (20%)	5 (17%)	2 (11%)	17+1 (33%)⁴	34+1 (18%)
VI 合計	1	-	-	-	+1	3	4+1 (3%)
合計	51+2=53	26+8=34	5+2=7	27+1=29	16+2=18	52+2=54	177+17=194

10

20

30

注釈

1. 「+」の後の番号は、CSF単離物をさす；その他は全て血液に由来する。

2. 5症例は4ヶ月から1歳であり、1症例は3歳であった。

3. 他の全ての群と比較した遅発型感染症における sst III-2 : $p = 0.0005$ 、オッズ比 (OR) 6.8 ; 95% 信頼区間 (CI) 2.4 ~ 19.4。

他の全年齢群と比較した高齢者の MS-V : $p = 0.001$ 、OR 0.28、95% CI 0.13 ~ 0.59)。

【図面の簡単な説明】

【0194】

40

【図1】 cpsD-cpsE-cpsFの3'末端およびcpsGの5'末端の配列異種性に基づいた分子的血清型同定を示す図である (関連プライマーを示す)。

【図2】 PCRおよび配列決定によるGBS分子的血清型 (MS) 同定のアルゴリズムを示す図である。

【図3】 血清型 Ia、Ib、II、III、IV、V、およびVIについての cpsG-cpsH-cpsI/Mの遺伝子配列の多配列アラインメントを示す図である (開始コドンおよび終止コドンを太字で強調する)。

【図4】 識別するために使用した alp2 (AF208158、上のライン) と alp3 (AF291065、下のライン) との間の配列異種性の2つの部位 (*) を示す図である (関連プライマーを示す)。

50

【 図 1 - 7 】

血清型 V	-----	-----	-----	-----
血清型 Ia-2	TATTAGTCCA	CAATCAAAAG	ACAGGGCTAA	TGTTTAAGTT AGACGATGAT
コンセンサス	-----	-----	-----	-----
1201	-----	-----	-----	1250
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	CCTAGATTTA	CTAAAATAGG	AAAATTTATT	CGAAAAACAA GCATAGATGA
コンセンサス	-----	-----	-----	-----
1251	-----	-----	-----	1300
血清型 III-2	a-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	GTTGCCTCAA	TCTATATAGT	TTTTAAAAGG	TGATATGAST TTAGTAGGAA
コンセンサス	-----	-----	-----	-----
1301	-----	-----	-----	1350
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	CACGCCCTCC	CACAGTTGAT	GAATATGAAA	AGTATATTC AACGCAGAG
コンセンサス	-----	-----	-----	-----
1351	-----	-----	-----	1400
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----

【 図 1 - 8 】

血清型 IV	-----	-----	-----	-----
血清型 V	CGACGCCTTA	GTTTTAAGCC	AGGAATCACT	GGTTTTGPGC AAATATCTGG
血清型 Ia-2	-----	-----	-----	-----
コンセンサス	-----	-----	-----	-----
1401	-----	-----	-----	1450
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	TAGAAATAT	ATTACTGATT	TTGATGAAT	CCTAAAGTTA GATGTTCAT
コンセンサス	-----	-----	-----	-----
1451	-----	-----	-----	1500
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	ATATCAATGA	ATGGTCTATT	TGGTCAGATA	TTAAGATTAT TCTCTAAGA
コンセンサス	-----	-----	-----	-----
cpsE3	-----	-----	-----	-----
1501	-----	-----	-----	1550
血清型 III-2	-----	-t-----	-c-----	-----
血清型 VI	-----	-t-----	-c-----	-----
血清型 Ib	-----	-t-----	-c-----	-----
血清型 II/III-4	t-----	-t-----	-c-----	-----
血清型 VII	t-----	-----	-----	-----
血清型 III-3	-----	-----	-t-----	-----
血清型 Ia-1	-----	-----	-t-----	-----
血清型 III-1	-----	-t-----	-c-----	-----
血清型 IV	-----	-t-----	-c-----	-----
血清型 V	-----	-t-----	-c-----	-----
血清型 Ia-2	CTAAAGGTAG	TCTTACTTGG	GACAGGAGCT	AAGTAAAGTT AAGGTTTGAA
コンセンサス	-----	-----	-----	-----
cpsE cpsEFA	-----	-----	-----	-----
1551	-----	-----	-----	1600
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----

【 図 1 - 9 】

血清型 V	-----	-----	-----	-----
血清型 Ia-2	AGGAATATAA	TGAAAATTTG	TCTGTGTTGT	TCAAGTGGTG GTCACTTAGC
コンセンサス	-----	-----	-----	-----
cpsF	-----	-----	-----	-----
1601	-----	-----	-----	1650
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	t-----	t-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	ACACTTGAAC	CTTTTGAAC	CCATTTGGGA	AAAAGAGAT AGSTTTTGGG
コンセンサス	-----	-----	-----	-----
1651	-----	-----	-----	1700
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	TAACCTTGA	TAAAGAGAT	GCTAGGACTA	TCTAAGAGA AGAGATTGTA
コンセンサス	-----	-----	-----	-----
1701	-----	-----	-----	1750
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	TATCATTTGCT	TCTTTCCAAC	AAACCGTAT	GTCAAAACT TGGTAAAAA
コンセンサス	-----	-----	-----	-----
1751	-----	-----	-----	1800
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----

【 図 1 - 10 】

血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	TACTATTCTA	GCTTTTAAAG	TCCTTAGAAA	AGAAAGACCA GATGTTATCA
コンセンサス	-----	-----	-----	-----
1801	-----	-----	-----	1850
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	TATCATCTGG	TGCCCGTGA	GCAGTACCAT	TCTTTTATAT TGCTAAGTTA
コンセンサス	-----	-----	-----	-----
cpsF3	-----	-----	-----	-----
1851	-----	-----	-----	1900
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	TTTGGTTGTA	AGACCGTTA	TATAGAGTT	TTCCAGGGA TAGATAAAC
コンセンサス	-----	-----	-----	-----
cpsFA	-----	-----	-----	-----
1901	-----	-----	-----	1950
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	-----	-----	-----
血清型 Ib	-----	-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	AACTTTGACA	GGAAATAG	TGTATCCTGT	AACAGATAAA TTTATGTTTC
コンセンサス	-----	-----	-----	-----
1951	-----	-----	-----	2000
血清型 III-2	-----	-----	-----	-----
血清型 VI	-----	a-----	-----	-----
血清型 Ib	-----	a-----	-----	-----
血清型 II/III-4	-----	-----	-----	-----
血清型 VII	-----	-----	-----	-----
血清型 III-3	-----	-----	-----	-----
血清型 Ia-1	-----	-----	-----	-----
血清型 III-1	-----	-----	-----	-----
血清型 IV	-----	-----	-----	-----
血清型 V	-----	-----	-----	-----
血清型 Ia-2	-----	-----	-----	-----
コンセンサス	-----	-----	-----	-----

【 図 1 - 1 1 】

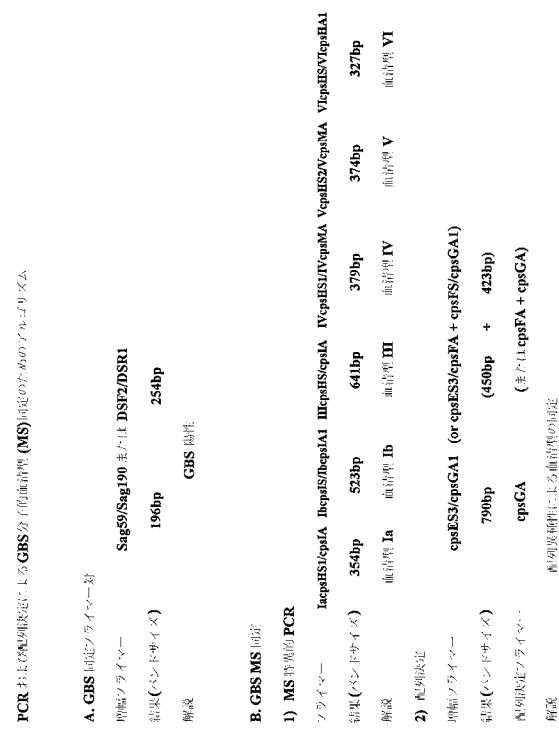
GenBank accession numbers and sequence alignments for various blood types (IV, V, Ia-2, III-2, etc.) across different cps genes (cpsF, cpsG, cpsG8).

【 図 1 - 1 2 】

GenBank accession numbers and sequence alignments for various blood types (III-2, VI, Ib, II/III-4, etc.) across different cps genes (cpsG1, cpsG2, cpsG3).

注釈
番号付(開始点)1 目1 GenBank アクセッション番号 AF332908 (血清型 IV 基準株 3139)
の開始点1 目を示す
血清型 Ia-1: 090 株 (GenBank アクセッション番号 AF332893);
血清型 Ia-2: NZRM 908 株 (NCDC SS615) (GenBank アクセッション番号 AF332894);
血清型 Ib: H36B 株 (GenBank アクセッション番号 AF332903);
血清型 II/III-4: 18RS21 株 (GenBank アクセッション番号 AF332905);
血清型 III-1: SG99/056 株 (GenBank アクセッション番号 AF332899);
血清型 III-2: M781 株 (GenBank アクセッション番号 AF332896);
血清型 III-3: NZRM 912 株 (NCDC SS620) (GenBank アクセッション番号 AF332897);
III-4 (血清 III-4): SG96/220 株 (GenBank アクセッション番号 AF363036);
血清型 IV: 3139 株 (GenBank アクセッション番号 AF332908);
血清型 V: CJB 111 株 (GenBank アクセッション番号 AF332910);
血清型 VI: SS1214 株 (GenBank アクセッション番号 AF332901);
血清型 VII: 7271 株 (GenBank アクセッション番号 AF332913)

【 図 2 】



【 図 3 - 1 】

Multiple sequence alignment of cpsG-cpsH-cpsH-cpsH regions for blood types Ia, Ib, II, III, IV, V, and VI, showing conserved regions and variations.

【 図 3 - 2 】

血清型 IV	GGAAAANAAC	TATTGTGGTT	CCTAGACAAG	AAACAGTTGG	AGAGCATGTG	300
血清型 V	GGAAAANAAC	TATTGTGGTT	CCTAGACAAG	AAACAGTTGG	AGAGCATGTG	
血清型 Ia	GGAAAANAAC	TATTGTGGTT	CCTAGACAAG	AAACAGTTGG	AGAGCATGTG	
血清型 Ib	GGAAAANAAC	TATTGTGGTT	CCTAGACAAG	AAACAGTTGG	AGAGCATGTG	
血清型 III	GGAAATPACC	AGTTGTGGTT	CCTAGACAAG	AAACAGTTGG	TGACATATTC	
血清型 VI	GGAAATPACC	AGTTGTGGTT	CCTAGACAAG	AAACAGTTGG	TGACATATTC	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	AATAATCATC	AGTGGGATT	TGTTAAATAAG	GTAATAACAA	TGTATAATTT	350
血清型 V	AATAATCATC	AGTGGGACTT	TGTTAAATAAG	GTAATAACAA	TGTATAATTT	
血清型 Ia	AATAATCATC	AGTGGGACTT	TGTTAAATAAG	GTAATAACAA	TGTATAATTT	
血清型 Ib	AATAATCATC	AGTGGGACTT	TGTTAAATAAG	GTAATAACAA	TGTATAATTT	
血清型 III	AAATGATCATC	AAATACAAAT	TTTAAATAAA	ATTGCCACCC	TGTATCCOCT	
血清型 VI	AAATGATCATC	AAATACAAAT	TTTAAATAAA	ATTGCCACCC	TGTATCCOCT	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	TGATATCCCT	CTAGATATTC	AAAGGTATCA	AAATGTATTC	TATGAGGGGA	400
血清型 V	TGATATCCCT	CTAGATATTC	AAAGGTATCA	AAATGTATTC	TATGAGGGGA	
血清型 Ia	AGATATATAT	TGAAATATCA	GTGAATATCA	GAATATATAT	AAGGAAAAAA	
血清型 Ib	AGATATATAT	TGAAATATCA	GTGAATATCA	GAATATATAT	AAGGAAAAAA	
血清型 III	GCCTTGGATT	GAGATATGAT	ATGGACTTCC	GGAAAGGTT	..GAAARAGGA	
血清型 VI	GCCTTGGATT	GAGATATGAT	ATGGACTTCC	GGAAAGGTT	..GAAARAGGA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	CGATGAATCC	TCCGTTTTTA	GAACTATACA	GAATATATTT	TATTT.....	450
血清型 V	CGATGAATCC	TCCGTTTTTA	GAACTATACA	GAATATATTT	TATTT.....	
血清型 Ia	ATATATCTAC	TAGTAATATC	ATATATCTAC	AACTATATTT	TTGTTCCCT	
血清型 Ib	ATATATCTAC	TAGTAATATC	ATATATCTAC	AACTATATTT	TTGTTCCCT	
血清型 III	ATATATCTAC	TAGTAATATC	ATATATCTAC	AACTATATTT	TTGTTCCCT	
血清型 VI	ATATATCTAC	TAGTAATATC	ATATATCTAC	AACTATATTT	TTGTTCCCT	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	500
血清型 V	
血清型 Ia	TTCAAAATTC	AACTATATTT	AACTATATTT	AACTATATTT	TTTGTGGGA	
血清型 Ib	TTCAAAATTC	AACTATATTT	AACTATATTT	AACTATATTT	TTTGTGGGA	
血清型 III	
血清型 VI	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	GTGTGTGATG	GAAA.....	ATCAATAAAA	AACCTTTTAT	TTTATATTC	550
血清型 V	GTGTGTGATG	GAAA.....	ATCAATAAAA	AACCTTTTAT	TTTATATTC	
血清型 Ia	GAARAAATTT	GAATATATTC	TTCATATCCA	AACTATATTT	TATGAGGGG	
血清型 Ib	GAARAAATTT	GAATATATTC	TTCATATCCA	AACTATATTT	TATGAGGGG	
血清型 III	GAATATATTC	GAATATATTC	GAATATATTC	GAATATATTC	TATGAGGGG	
血清型 VI	GAATATATTC	GAATATATTC	GAATATATTC	GAATATATTC	TATGAGGGG	
コンセンサス	*****	*****	*****	*****	*****	

【 図 3 - 4 】

血清型 IV	AGGATCCGGT	ATTTTGGGAG	TTCATATAAA	TCAAAATTTT	GTGAAATTAG	851
血清型 V	AGGATCCGGT	ATTTTGGGAG	TTCATATAAA	TCAAAATTTT	GTGAAATTAG	
血清型 Ia	ATTTTATATTT	AAATGGGTAA	TCCATATTTT	TTTAAAGGGA	ACATATATTT	
血清型 Ib	ATTTTATATTT	AAATGGGTAA	TCCATATTTT	TTTAAAGGGA	ACATATATTT	
血清型 III	GGAAAGGGTT	AACTATATTT	TTTGGTTTTC	TATACAGGTT	TATATTT...	
血清型 VI	GGAAAGGGTT	AACTATATTT	TTTGGTTTTC	TATACAGGTT	TATATTT...	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	ATTTAATAGA	AATTAATATC	ATCAAAATTT	ATAGGGATTC	ACAAATTTAT	901
血清型 V	ATTTAATAGA	AATTAATATC	ATCAAAATTT	ATAGGGATTC	ACAAATTTAT	
血清型 Ia	ATTTTATATTT	AAATGGGTAA	TCCATATTTT	TTTAAAGGGA	ACATATATTT	
血清型 Ib	ATTTTATATTT	AAATGGGTAA	TCCATATTTT	TTTAAAGGGA	ACATATATTT	
血清型 III	
血清型 VI	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	CTGAGAATGC	ACT.....TAGG	1000
血清型 V	CTGAGAATGC	ACT.....TAGG	
血清型 Ia	ATTGGGTCAG	ACTGGATAAA	TGGTATGATC	ACTCAAAGAC	CAATGGGATC	
血清型 Ib	ATTGGGTCAG	ACTGGATAAA	TGGTATGATC	ACTCAAAGAC	CAATGGGATC	
血清型 III	TTTTCATATG	ATGGGATTTT	CTCAATATTC	AACTATATTT	TTGCGGATTA	
血清型 VI	TTTTCATATG	ATGGGATTTT	CTCAATATTC	AACTATATTT	TTGCGGATTA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	TTTGGTCAAT	CCTAATCTTA	TTCATATATTT	TTTGGCAGTA	ACTGTTTTTT	1050
血清型 V	TTTGGTCAAT	CCTAATCTTA	TTCATATATTT	TTTGGCAGTA	ACTGTTTTTT	
血清型 Ia	TTTGGATAT	TCAAAATTTT	TATGATATTT	AACTATATTT	TTGCGGATTA	
血清型 Ib	TTTGGATAT	TCAAAATTTT	TATGATATTT	AACTATATTT	TTGCGGATTA	
血清型 III	TTTGGATAT	TCAAAATTTT	TATGATATTT	AACTATATTT	TTGCGGATTA	
血清型 VI	TTTGGATAT	TCAAAATTTT	TATGATATTT	AACTATATTT	TTGCGGATTA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	TATATATATC	ACTTTTATAT	AGAAACATTA	GAT..TATATA	CTATTGCTTT	1051
血清型 V	TATATATATC	ACTTTTATAT	AGAAACATTA	GAT..TATATA	CTATTGCTTT	
血清型 Ia	TATATATATC	ACTTTTATAT	AGAAACATTA	GAT..TATATA	CTATTGCTTT	
血清型 Ib	TATATATATC	ACTTTTATAT	AGAAACATTA	GAT..TATATA	CTATTGCTTT	
血清型 III	TAT.....	TTTTTTGGAA	ACCCCAAGAA	CATATGGAAA	AAATTTTAA	
血清型 VI	TAT.....	TTTTTTGGAA	ACCCCAAGAA	CATATGGAAA	AAATTTTAA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	TATTTTAACT	CTAATTTACT	TTTGTATACA	GTATATTTAT	TCAGAACCTG	1101
血清型 V	TATTTTAACT	CTAATTTACT	TTTGTATACA	GTATATTTAT	TCAGAACCTG	
血清型 Ia	TTTGTATACA	TATTTTAACT	TTTGTATACA	GTATATTTAT	TCAGAACCTG	
血清型 Ib	TTTGTATACA	TATTTTAACT	TTTGTATACA	GTATATTTAT	TCAGAACCTG	
血清型 III	CTCAATATTC	TTTGGGATTA	TTTGGGATTA	TTTGGGATTA	TTTGGGATTA	
血清型 VI	CTCAATATTC	TTTGGGATTA	TTTGGGATTA	TTTGGGATTA	TTTGGGATTA	
コンセンサス	*****	*****	*****	*****	*****	

【 図 3 - 3 】

血清型 IV	ANATTTTTTA	GTATATTTTT	TATATTTACT	AGTTTATAGA	GAGGGAACT	600
血清型 V	ANATTTTTTA	GTATATTTTT	TATATTTACT	AGTTTATAGA	GAGGGAACT	
血清型 Ia	AAATTTTGGT	CTAATTTTAT	TTTAAAGGCA	AATGCACTT	TGTATATTT	
血清型 Ib	AAATTTTGGT	CTAATTTTAT	TTTAAAGGCA	AATGCACTT	TGTATATTT	
血清型 III	CTTTTGGGTA	CTAATTTTAT	TACCAACCCA	ATGATATCAG	TTTTTAAATTA	
血清型 VI	CTTTTGGGTA	CTAATTTTAT	TACCAACCCA	ATGATATCAG	TTTTTAAATTA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	CAACTTACAA	AATAGTATGC	TTTGTTCGAA	TCTTCTGTGG	TGGAAATAAA	650
血清型 V	CAACTTACAA	AATAGTATGC	TTTGTTCGAA	TCTTCTGTGG	TGGAAATAAA	
血清型 Ia	TAGCATTAAT	AGTTTACTTT	ATTTTATAGA	GTATATAGAA	AAAATATGAA	
血清型 Ib	TAGCATTAAT	AGTTTACTTT	ATTTTATAGA	GTATATAGAA	AAAATATGAA	
血清型 III	TTTAAATAAA	ATTGCCACCC	TGTATCCOCT	TGTATCCOCT	TTTAAATAAA	
血清型 VI	TTTAAATAAA	ATTGCCACCC	TGTATCCOCT	TGTATCCOCT	TTTAAATAAA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	TTTTTA....	..TTAGATAG	CCTTATTTTT	GAAGAGAA	AACCTGTTAT	700
血清型 V	TTTTTA....	..TTAGATAG	CCTTATTTTT	GAAGAGAA	AACCTGTTAT	
血清型 Ia	TTTTTAAATA	TGGCTGAAAT	TTTTTCAAT	GAATTTTAA	TGATTTATTT	
血清型 Ib	TTTTTAAATA	TGGCTGAAAT	TTTTTCAAT	GAATTTTAA	TGATTTATTT	
血清型 III	TCTATAGCA	ATCTTCAAT	ACTATTTCTG	CTTTGGTAT	TTATTTATTT	
血清型 VI	TCTATAGCA	ATCTTCAAT	ACTATTTCTG	CTTTGGTAT	TTATTTATTT	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	CATCTTTTAA	TTATTTAATG	CGACCAATTT	GAATTTAATC	TTTGTTCATA	750
血清型 V	CATCTTTTAA	TTATTTAATG	CGACCAATTT	GAATTTAATC	TTTGTTCATA	
血清型 Ia	AACCTGTTAT	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 Ib	AACCTGTTAT	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 III	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 VI	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	AGGTTACTTT	TATATTTAA	800
血清型 V	AGGTTACTTT	TATATTTAA	
血清型 Ia	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 Ib	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 III	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 VI	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	TTTCTAGCAT	TAAAGGATAT	CTCTCTAATA	AAAGCTTTCT	CATATTAAT	850
血清型 V	TTTCTAGCAT	TAAAGGATAT	CTCTCTAATA	AAAGCTTTCT	CATATTAAT	
血清型 Ia	AAACGGTGGT	ATAAGATATC	AACTTTGGAG	TATGATATTA	TATTAAGAGC	
血清型 Ib	AAACGGTGGT	ATAAGATATC	AACTTTGGAG	TATGATATTA	TATTAAGAGC	
血清型 III	TATATATATC	TATGATATTA	TATGATATTA	TATGATATTA	TATGATATTA	
血清型 VI	TATATATATC	TATGATATTA	TATGATATTA	TATGATATTA	TATGATATTA	
コンセンサス	*****	*****	*****	*****	*****	

【 図 3 - 5 】

血清型 IV	GAATTTAAT	AGTACTCTTA	TTTATACCTA	TATATATATC	TACAAAGAT	1151
血清型 V	GAATTTAAT	AGTACTCTTA	TTTATACCTA	TATATATATC	TACAAAGAT	
血清型 Ia	CTGGAATATG	AGTACTCTTA	TTTATACCTA	TATATATATC	TACAAAGAT	
血清型 Ib	CTGGAATATG	AGTACTCTTA	TTTATACCTA	TATATATATC	TACAAAGAT	
血清型 III	TGCTTGTAT	GTGGTTTTTA	TTAGCATGCG	TCTTTTATGA	TATATATATC	
血清型 VI	TGCTTGTAT	GTGGTTTTTA	TTAGCATGCG	TCTTTTATGA	TATATATATC	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	AACCTGATTA	GGAAATTTT	TATGATATTT	GTCCGATACA	TACAAAGAT	1201
血清型 V	AACCTGATTA	GGAAATTTT	TATGATATTT	GTCCGATACA	TACAAAGAT	
血清型 Ia	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 Ib	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 III	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 VI	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	CTTGTATGCA	TTTACTTTTC	TTTGGTCTAC	TATTTTTTTC	AACCTGATTA	1255
血清型 V	CTTGTATGCA	TTTACTTTTC	TTTGGTCTAC	TATTTTTTTC	AACCTGATTA	
血清型 Ia	TCTATATGTA	ATAGATATGA	TGTATATTT	TGATATTTT	CTCAATATTT	
血清型 Ib	TCTATATGTA	ATAGATATGA	TGTATATTT	TGATATTTT	CTCAATATTT	
血清型 III	AACTATATTT	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 VI	AACTATATTT	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	CTTGTATGCA	TTTACTTTTC	TTTGGTCTAC	TATTTTTTTC	AACCTGATTA	1300
血清型 V	CTTGTATGCA	TTTACTTTTC	TTTGGTCTAC	TATTTTTTTC	AACCTGATTA	
血清型 Ia	TCTATATGTA	ATAGATATGA	TGTATATTT	TGATATTTT	CTCAATATTT	
血清型 Ib	TCTATATGTA	ATAGATATGA	TGTATATTT	TGATATTTT	CTCAATATTT	
血清型 III	AACTATATTT	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
血清型 VI	AACTATATTT	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	TTTAAAGGCA	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	TTTGTCAAAA	ATTATATAGC	CTTTTGACAG	GTAG.....	1350
血清型 V	TTTGTCAAAA	ATTATATAGC	CTTTTGACAG	GTAG.....	
血清型 Ia	TTTGTCAAAA	ATTATATAGC	CTTTTGACAG	GTAG.....	
血清型 Ib	TTTGTCAAAA	ATTATATAGC	CTTTTGACAG	GTAG.....	
血清型 III	TTTGTCAAAA	ATTATATAGC	CTTTTGACAG	GTAG.....	
血清型 VI	TTTGTCAAAA	ATTATATAGC	CTTTTGACAG	GTAG.....	
コンセンサス	*****	*****	*****	*****	*****	
血清型 IV	..GTTAAACT	ATGCTCATTT	ACAGCTGTGA	GAGGGTTTAA	CCTTTTGGG	1400
血清型 V	..GTTAAACT	ATGCTCATTT	ACAGCTGTGA	GAGGGTTTAA	CCTTTTGGG	
血清型 Ia	GTATATATAT	ATACAGTATA	CTCTCTAATA	TATATATATC	TATTTTGGG	

【 図 3 - 6 】

1451		1500	
血清型 IV	ATTTGATAAT AGCTACCTCA TGTATTGAG TATGTATGGT GTAGTACTTA		
血清型 V	ATTTGATAAT AGCTACCTCA TGTATTGAG TATGTATGGT GTAGTACTTA		
血清型 Ia	ATTCACAGTA TATAGGCTAT TTCTACAAA GTGGCTGCTT GGAATTAATG		
血清型 Ib	ATTCACAGTA TATAGGCTAT TTCTACAAA GTGGCTGCTT GGAATTAATG		
血清型 III	ATTCACAGTA TATAGGCTAT TTCTACAAA GTGGCTGCTT GGAATTAATG		
血清型 VI	ATTCACAGTA TATAGGCTAT TTCTACAAA GTGGCTGCTT GGAATTAATG		
コンセンサス	***-----*-----*-----*-----*-----*		
1501		1550	
血清型 IV	CCGATTTTTG TATGATAATC TATTATATCT ATAGTAAAA AGTCAATGTA		
血清型 V	CCGATTTTTG TATGATAATC TATTATATCT ATAGTAAAA AGTCAATGTA		
血清型 Ia	AATATAGTTC CAGGTTTGCT TTTAAT TTT TACTAATAT GTTAGGAAAG		
血清型 Ib	AATATAGTTC CAGGTTTGCT TTTAAT TTT TACTAATAT GTTAGGAAAG		
血清型 III	AATATAGTTC CAGGTTTGCT TTTAAT TTT TACTAATAT GTTAGGAAAG		
血清型 VI	AATATAGTTC CAGGTTTGCT TTTAAT TTT TACTAATAT GTTAGGAAAG		
コンセンサス	***-----*-----*-----*-----*-----*		
1551		1600	
血清型 IV	GTTGAGCTCC AGATACCTTT GTTTA.....		
血清型 V	AFTGAACTTC AACCTACTCT ATTTA.....TA		
血清型 Ia	CTAAGACTC ASCTTTTAT TATGAGTAG TAGGACACT TATACCTTA		
血清型 Ib	CTAAGACTC AGATTTCTAT TATGAGTAG TAGGCTGCT CATACCTTA		
血清型 III	TATATAACT TACAGATTAT TTT.....T TATACCTTA		
血清型 VI	CAGGATACT TACAGATTAT TTT.....T TGGATTGTA		
コンセンサス	***-----*-----*-----*-----*-----*		
1601		1650	
血清型 IV	AGCTCTTAG TATTATTTAC AGAGAGTTTT TACCAGTA TGTATAGAA		
血清型 V	AGCTCTTAG TATTATTTAC TGAAGTTTT TATCCAGTG TGTATAGAA		
血清型 Ia	TTTCTATTT TTSCACTTGA AGATCTTGC GGAGCTAAT GGCTATTTG		
血清型 Ib	TTTCTATTT TTSCACTTGA AGATCTTGC GGAGCTAAT GGCTATTTG		
血清型 III	TTTCTATTT AGCTCTTGA GATTTTGA AATAGAGAT TTCTATGTC		
血清型 VI	TTTCTATTT AGCTCTTGA GATTTTGA AATAGAGAT TTCTATGTC		
コンセンサス	***-----*-----*-----*-----*-----*		
1651		1700	
血清型 IV	TATTAGTTGG APTGTTTTG GGAARAAT TTCTGGGGT CIAGATGAT		
血清型 V	TATTAGTTGG APTGTTTTG GGAARAAT TTCTGGGGT CIAGATGAT		
血清型 Ia	TTTTATTTT ACAGTTTTG GATTTTGA AATAGAGAT TTTTATGTC		
血清型 Ib	TTTTATTTT ACAGTTTTG GATTTTGA AATAGAGAT TTTTATGTC		
血清型 III	ATTTATTTT TCACTTTTG GATTTTGA AATAGAGAT TTTTATGTC		
血清型 VI	ATTTATTTT ACTACATTG GATTTTGA GAGAG. GSA.		
コンセンサス	***-----*-----*-----*-----*-----*		
1701		1750	
血清型 IV	TACAC.....GAGGTT CACTTGGAG GCAATFAAA ATTAGTGTG		
血清型 V	TACAC.....AGGAAI TACI...ATT GTGATAATA TATGACATP		
血清型 Ia	AACTTAAAG GTGGAAAGT TAAAGGAAA AGCAATACT GTTCTATCA		
血清型 Ib	AACTTAAAG GTGGAAAGT TAAAGGAAA AGCAATACT GTTCTATCA		
血清型 IIIAAAT GAAAGAAA GTAACATCA		
血清型 VIATGAT AAAAAACTA GTTAGTGTG		
コンセンサス	***-----*-----*-----*-----*-----*		

【 図 3 - 8 】

2051		2100	
血清型 IV	AGAGTAGACA AGTAGACTGG GTCAGTCTA ATACAAJAA AATTACTCT		
血清型 V	AGCTTTTAA AAGTTTAAAT CTTTGAAGA TTCTAGAGT TTTCCTAGCA		
血清型 Ia	TCACAGAGA TGCATGATA GCAGAGTAG ATTTTGA... TATTTCGAT		
血清型 Ib	TCACAGAGA TGCATGATA GCAGAGTAG ATTTTGA... TATTTCGAT		
血清型 III	TCACAGAGA TGCATGATA GCAGAGTAG ATTTTGA... TATTTCGAT		
血清型 VI	TACCGAGAA CCGAGATATA TCAGAAATTT ATTTGCA... AGTTTCGAT		
コンセンサス	***-----*-----*-----*-----*-----*		
2101		2150	
血清型 IV	TACTGTCTA ACTTATATA TAAAGGACT GAATACATA ATGCTCTTA		
血清型 V	AGGAGATTA GTGTGAGT GAATACAGA TAAATATG TGCCTCTTA		
血清型 Ia	GAGAGAGAT ATGAAAGAA GAAAGAGCA AACTTTTATA AGCTCTTAA		
血清型 Ib	GAGAGAGAT ATGAAAGAA GAAAGAGCA AACTTTTATA AGCTCTTAA		
血清型 III	GAGAGAGAT ATGAAAGAA GAAAGAGCA AACTTTTATA AGCTCTTAA		
血清型 VI	GA...TTTT ATAAAGAAA AAAAAAGAA GTTACTATA GAGTTTICA		
コンセンサス	***-----*-----*-----*-----*-----*		
2151		2200	
血清型 IV	TAAACAAGT TTCCATACG AATTCTGAG TACAATFAG ATTTTATGTT		
血清型 V	AGGACATCAC TTTTAAAAT CAATATCTC TAAATATGAC AAAAGTGA		
血清型 Ia	AAACAAATAC TCTTAAAAG APTTTTTC AGGCAATGA GTGGAAATA		
血清型 Ib	AAACAAATAC TCTTAAAAG APTTTTTC AGGCAATGA GTGGAAATA		
血清型 III	GATTAATAC TCTTAAAAG APTTTTTC AGGCAATGA GTGGAAATA		
血清型 VI	AAACAATAG TCCICAAAG APTTTTTC AGGCAATGA GTGGAAATG		
コンセンサス	***-----*-----*-----*-----*-----*		
2201		2250	
血清型 IV	CAGTCTGGA GGGTTGTTA TCTAGAGAT TACCTTAAA AATAAAATC		
血清型 V	TACTCTCT TAAAGACA TGTGTAGAG TACACTTA TTTATGTA		
血清型 Ia	TCTTCTAC AAACTATAT AAAAAAGTA TAAATGGA APTGAGGTT		
血清型 Ib	TCTTCTAC AAACTATAT AAAAAAGTA TAAATGGA APTGAGGTT		
血清型 III	TCTTCTAC AAACTATAT AAAAAAGTA TAAATGGA APTGAGGTT		
血清型 VI	TCTTCTAC AAACTATAT AAAAAAGTA TAAATGGA APTGAGGTT		
コンセンサス	***-----*-----*-----*-----*-----*		
2251		2300	
血清型 IV	CCTGAGAAA AAAAT...A TGAAGTACA CAGTTTAT TTGATCTAT		
血清型 V	AACAGAGGC TTAAGA...A TAAAGATTT ATCAAAA GA...TGA		
血清型 Ia	GATGAGACT TAAAATG TGAGGATTA CTTTTAAT GTAACCTCT		
血清型 Ib	GATGAGACT TAAAATG TGAGGATTA CTTTTAAT GTAACCTCT		
血清型 III	GATGAGACT TAAAATG TGAGGATTA CTTTTAAT GTAACCTCT		
血清型 VI	GATGAGACT TAAAATG TGAGGATTA CTTTTAAT GTAACCTCT		
コンセンサス	***-----*-----*-----*-----*-----*		
2301		2350	
血清型 IV	AAAAATGCT AATAAGTTT TATTATAG CCAACCTTT TATAATTACT		
血清型 V	ATACAAAT ATCCAGAAA TATTTTAT CAAAAGAT TATTAACA...		
血清型 Ia	AGTCAAGG CACCTATG TCGAGAGAC GACTCTCC TTTATPACT		
血清型 Ib	ATGCAAGG CACCTATG TCGAGAGAC GACTCTCC TTTATPACT		
血清型 III	ATGCAAGG CACCTATG TCGAGAGAC GACTCTCC TTTATPACT		
血清型 VI	AAATAAGAA CACCTATG TCGAGAGAC TAGAAGATCA CTTATACTT		
コンセンサス	***-----*-----*-----*-----*-----*		

【 図 3 - 7 】

1751		1800	
血清型 IV	TTCCACAGT ATATATTGG AAACAATTT TATAGCTTG CTTGATTTA		
血清型 V	TTCCACAGT ATATATTGG AAACAATTT TATAGCTTG CTTGATTTA		
血清型 Ia	TTATACCTAT ATACACTCA GAAGCAATC TTAAGAAATC TTGCAATCC		
血清型 Ib	TTATACCTAT ATACACTCA GAAGCAATC TTAAGAAATC TTGCAATCC		
血清型 III	TTATACCTAT ATACACTCA GAAGCAATC TTAAGAAATC TTGCAATCC		
血清型 VI	TTGTTCCAGT TTAATATTGG GAGTATGTA TTGAGACTC TTGATAATC		
コンセンサス	***-----*-----*-----*-----*-----*		
cps1/H		1850	
血清型 IV	ATTAGAAAAC AAACATATA GAATTTGGA ATTATTCTG TTAATGATG		
血清型 V	AGATTATTGT TGGTTTGGG GAATTCCTT ACCAGATAA TTAAGAAAT		
血清型 Ia	GTACTACAC AGACTCATCC ATTGATAGA GTTACTCTA TTGATGATG		
血清型 Ib	GTCTACAC AGACTCATC ATTGATAGA GTTACTCTA TTGATGATG		
血清型 III	GTACTACAC AGACTCATC ATTGATAGA GTTACTCTA TTGATGATG		
血清型 VI	TTGCTTCAAC AAACATACC AGAATAGA ATTATTATA TTAGATGATG		
コンセンサス	***-----*-----*-----*-----*-----*		
1851		1900	
血清型 IV	ATCACAGAT GGTAGTAAAG AGTTACTGA GGATAGA AATCAAGG		
血清型 V	ATTAFAAA...CTGGG AGAATCTC CCGATTAG AATCAATGA		
血清型 Ia	ATCCACTAT AATAGTGGG AAATTTGGA TAAATATCT CAGAAGTA		
血清型 Ib	ATCCACTAT AATAGTGGG AAATTTGGA TAAATATCT CAGAAGTA		
血清型 III	ATCCACTAT AATAGTGGG AAATTTGGA TAAATATCT CAGAAGTA		
血清型 VI	ATCACAGAT AAAAGTATC ATATTTTGA TAAATTTTA AAAAGGATA		
コンセンサス	***-----*-----*-----*-----*-----*		
1901		1950	
血清型 IV	AAAGAATTA GACATTTAC AAACAATG GAGCAATC AAGCGAAG		
血清型 V	ATGGAATGAG CATAATTAT AGTTAGTA AAATGTTTT ATGAGAAAG		
血清型 Ia	ATCCACTAT TGTATTTAT AAAAATAA GAGGGCTCC TCGCAAGG		
血清型 Ib	ATCCACTAT TGTATTTAT AAAAATAA GAGGGCTCC TCGCAAGG		
血清型 III	ATCCACTAT TGTATTTAT AAAAATAA GAGGGCTCC TCGCAAGG		
血清型 VI	GTCCGTAAA AGTATCAT AAATCAATG GAGGCTCC ATCACAGA		
コンセンサス	***-----*-----*-----*-----*-----*		
1951		2000	
血清型 IV	AAATTAGTA TTTTATCT TACAGAGAT TTGTTGTT TTCTCAGC		
血清型 V	CAATCTA...TCTT TATGTTCTG ACTAGCAG ACTAGCAG		
血清型 Ia	AACTTAGCT TAGTAACT CACAGGAAA TCAACACT TTGCAATG		
血清型 Ib	AACTTAGCT TAGTAACT CACAGGAAA TCAACACT TTGCAATG		
血清型 III	AACTTAGCT TAGTAACT CACAGGAAA TCAACACT TTGCAATG		
血清型 VI	AAATGGGAC TTGAGTGGC AAGAGTGA TTTAATCT TTGATGAT		
コンセンサス	***-----*-----*-----*-----*-----*		
2001		2050	
血清型 IV	CGAGATACA AITGACCTA AATGTATGA AACGTTACTA AATATATG		
血清型 V	ATFGATATT AITATACT TTGGGGGTT CACTAGAT ACTGATGG		
血清型 Ia	TGATGATTT GTAGCAGCA ATATGATTA AATATGTA AAAATTTAA		
血清型 Ib	TGATGATTT GTAGCAGCA ATATGATTA AATATGTA AAAATTTAA		
血清型 III	TGATGATTT GTAGCAGCA ATATGATTA AATATGTA AAAATTTAA		
血清型 VI	CGATGATTT GTCCGCTTA ATATGATTA AATATGCTA AATATTTGT		
コンセンサス	***-----*-----*-----*-----*-----*		

【 図 3 - 9 】

2351		2400	
血清型 IV	ACTACAGAAA AAATAGTACA AACAATCTC CATATAGTAG CTACATATG		
血清型 V	...GTAAGTT TGAATGCTG ACTAGTGA CTTATTTAT ACCTATPAC		
血清型 Ia	ATCCACTAT AAAACTTCC GCA...TGAAT CAGAAATCA ACAAAACTC		
血清型 Ib	ATCCACTAT AAAACTTCC GCA...TGAAT CAGAAATCA ACAAAACTC		
血清型 III	ATCCACTAT AAAACTTCC GCA...TGAAT CAGAAATCA ACAAAACTC		
血清型 VI	ATCCACTAT AAAACTTCC GCA...TGAAT CAGAAATCA ACAAAACTC		
コンセンサス	***-----*-----*-----*-----*-----*		
2401		2450	
血清型 IV	ATCATATCC ATATCTGAC TCAAGTTAT TATATGCA AGATTTTA		
血清型 V	ATCATATCC ATATCTGAC TCAAGTTAT TATATGCA AGATTTTA		
血清型 Ia	ATGATATTT ATACAAATC TTAATGAGT AATGATTTG TTAGATGCA		
血清型 Ib	ATGATATTT ATACAAATC TTAATGAGT AATGATTTG TTAGATGCA		
血清型 III	ATGATATTT ATACAAATC TTAATGAGT AATGATTTG TTAGATGCA		
血清型 VI	ATGATATTT ATGATATTT TTAATGAGT TCAACAGAT ACTCCAGC		
コンセンサス	***-----*-----*-----*-----*-----*		
2451		2500	
血清型 IV	TGGATTGAA GAAGTTGCT TTCAAGATT ATTGTTGCA TATCTGATG		
血清型 V	AGTAGGCTC ATATGATTT TTTTATTTG ATATGATCT TATGAAATG		
血清型 Ia	AATTGGCTA TTAGTTGAA GCGAATTT TAAGAGAAA GATAAATGT		
血清型 Ib	AATTGGCTA TTAGTTGAA GCGAATTT TAAGAGAAA GATAAATGT		
血清型 III	AATTGGCTA TTAGTTGAA GCGAATTT TAAGAGAAA GATAAATGT		
血清型 VI	AATTGGCTA TTAGTTGAA GCGAATTT TAAGAGAAA GATAAATGT		
コンセンサス	***-----*-----*-----*-----*-----*		
2501		2550	
血清型 IV	TAGCTAATA AATTGTATA AATAAAGAT ATAGAAAA CGAAGAATA		
血清型 V	TCTAAGSTT TCTAAGSTT AAGAAATAG.....		
血清型 Ia	CTCCGAAAA TGTGTAAT AGTAGAAT ATTCAGTA AATCAAGT		
血清型 Ib	CTCCGAAAA TGTGTAAT AGTAGAAT ATTCAGTA AATCAAGT		
血清型 III	CTCCGAAAA TGTGTAAT AGTAGAAT ATTCAGTA AATCAAGT		
血清型 VI	TAAAGAAA TGTGTAAT AGGAAATA GCTGATGAA AITTACTTT		
コンセンサス	***-----*-----*-----*-----*-----*		
2551		2600	
血清型 IV	AGTAAA.....		
血清型 V		
血清型 Ia	ACACAGGAG ATTTTTTCA AAGCAATTA ATCATACCC TCTATAAAG		
血清型 Ib	ACACAGGAG ATTTTTTCA AAGCAATTA ATCATACCC TCTATAAAG		
血清型 III	ACACAGGAG ATTTTTTCA AAGCAATTA ATCATACCC TCTATAAAG		
血清型 VI	ACAGGATAT AAATTTGCG AAGATATTA ATCATATCA ATAGGATAG		
コンセンサス	***-----*-----*-----*-----*-----*		
2601		2650	
血清型 IV		
血清型 V		
血清型 Ia	CGGTAAATA CTACACTA AAGGATTA TAACTTTTA TATAGAAA		
血清型 Ib	CGGTAAATA CTACACTA AAGGATTA TAACTTTTA TATAGAAA		
血清型 III	CGGTAAATA CTACACTA AAGGATTA TAACTTTTA TATAGAAA		
血清型 VI	CAATAAGTT CTACACTA AAACATCTC GTAGCTATA TTAGTAAA		
コンセンサス	***-----*-----*-----*-----*-----*		

【 図 3 - 10 】

血清型 IV	2651	2700
血清型 V		
血清型 Ia	TGTTACCTA AGCTATATG TATGGCAT	AGAAGTTC AACAAGTAC
血清型 Ib	TGTTACCTA TCCTATATG TATGGCAT	AGAAGTTC AACAAGTAC
血清型 III	TGTTACCTA AGCTATATG TATGGCAT	AGAAGTTC AACAAGTAC
血清型 VI	TATTTTCGT ACCTATATG AAAGATGAT	AATAAATTC AAAAGCAAT
コンセンサス		
血清型 IV	2701	2728
血清型 V		
血清型 Ia	TGGAGAAAT GGGAAAGAG	ATTATATA
血清型 Ib	A	
血清型 III	G	
血清型 VI	A	
コンセンサス		

注釈

- 血清型 Ia: GenBank アクセッション番号 AB028896;
- 血清型 Ib: GenBank アクセッション番号 AB050723;
- 血清型 III: GenBank アクセッション番号 AF163833;
- 血清型 IV: GenBank アクセッション番号 AF355776;
- 血清型 V: GenBank アクセッション番号 AF349539;
- 血清型 VI: GenBank アクセッション番号 AF337958.

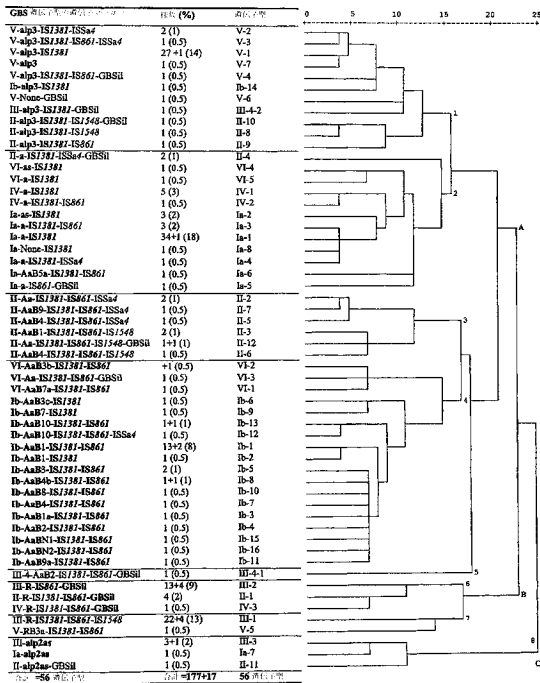
【 図 4 】

識別に使用した *alp2* (AF208158, 上のライン) と *alp3* (AF291065, 下のライン)

との間の配列異種性の2つの部位 (*) (固線フラインクを示す)

251	AAGTAATCTTAATATTTTGAAGATCAATAGTCTGTCATCACAAT	300
531	AAGTAATCTTAATATTTTGAAGATCAATAGTCTGTCATCACAAT	580
301	CCAGGAGTGCAGGACCTTAAATACAAGCATCACTAAAATATACAAA	350
581	CCAGGAGTGCAGGACCTTAAATACAAGCATCACTAAAATATACAAA	630
351	CGCAACCGTACATAGATTTATATGATTAAGAAATGGATTTGATTCAC	400
631	CGCAACCGTACATAGATTTATATGATTAAGAAATGGATTTGATTCAC	680
401	CTCAAAACCTCAITGTATTAATCCATCAAGCTATTGAGCAATATATAT	450
681	CTCAAAACCTCAITGTATTAATCCATCAAGCTATTGAGCAATATATAT	730
451	ATCAACAAGTCTTAATATATATGATTAATCCAGGATGAATACACAAC	500
731	ATCAACAAGTCTTAATATATATGATTAATCCAGGATGAATACACAAC	780
501	TGTTACCTAATATTTTGAAGATCAATAGTCTGTCATCACAAT	550
781	TGTTACCTAATATTTTGAAGATCAATAGTCTGTCATCACAAT	830
551	ATAAAAAGCTGATGGACAAATGATATAGTTAGTGAATTTAACTATA	600
831	ATAAAAAGCTGATGGACAAATGATATAGTTAGTGAATTTAACTATA	880
601	TATGATTCACAGCTTTAAGAAATAGGATAGATGAATTAATAATGCG	650
881	TATGATTCACAGCTTTAAGAAATAGGATAGATGAATTAATAATGCG	930
651	AAATGATCCTAAGTGGAGTGTGGAGTCTGATGAATCTTAATGAT	700
931	AAATGATCCTAAGTGGAGTGTGGAGTCTGATGAATCTTAATGAT	980

【 図 5 】



【国際公開パンフレット】

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau(43) International Publication Date
27 March 2003 (27.03.2003)

PCT

(10) International Publication Number
WO 03/025216 A1

- (51) International Patent Classification: C12Q 1/68, 1/46 (74) Agent: F.B. RICE & CO; 139 Rathdowne Street, Carlton South, Victoria 3053 (AU).
- (21) International Application Number: PCT/AU02/01281 (81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, OM, PI, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW.
- (22) International Filing Date: 18 September 2002 (18.09.2002) (84) Designated States (regional): ARIPO patent (GI, GM, KI, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, U, UM), European patent (AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, NI, SN, TD, TG).
- (25) Filing Language: English
- (26) Publication Language: English
- (30) Priority Data: PR 7749 19 September 2001 (19.09.2001) AU
- (71) Applicant (for all designated States except US): WESTERN SYDNEY AREA HEALTH SERVICE [AU/AU]; Westmead Hospital, Westmead, New South Wales 2145 (AU).
- (72) Inventors; and
(75) Inventors/Applicants (for US only): FANRONG, Kong [AU/AU]; Villa 23/93 Bridge Road, Westmead, New South Wales 2145 (AU), GILBERT, Gwendolyn [AU/AU]; 27 Kooyong Road, Riverview, New South Wales 2066 (AU).
- Published:
with International search report
- For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 03/025216 A1

(54) Title: MOLECULAR TYPING OF GROUP B STREPTOCOCCI

(57) Abstract: Molecular methods are provided for typing group B streptococci, as well as polynucleotides useful in such methods.

MOLECULAR TYPING OF GROUP B STREPTOCOCCIField of the invention

5 The present invention relates to molecular methods of typing group B streptococci, as well as polynucleotides useful in such methods.

Background to the invention

10 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. The incidence of neonatal GBS sepsis has been reduced in recent years by the use of intrapartum antibiotic prophylaxis, but there are many problems with this approach. In future, vaccination is likely to be preferred and there has been considerable progress in development of conjugate polysaccharide GBS
15 vaccines.

Before the introduction of conjugate vaccines, extensive epidemiological and other related studies will be required to assess, not only the burden of disease, but also the distribution of GBS types (including capsular polysaccharide gene serotypes, serosubtypes; protein antigen gene subtypes; mobile genetic
20 element subtypes) to determine the optimal formulation of vaccine antigens. Type distribution based on one geographic location or small numbers of patients may not be generally applicable. Continued monitoring will be necessary to assess the suitability of combinations of GBS vaccine antigens for different target populations in different geographic locations.

25 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 and Moody, 1969), enzyme immunoassay (Holm and Hakansson, 1988), coagglutination (Hakansson et al., 1992), counter-immunoelectrophoresis, and capillary precipitation (Triscott and
30 Davies, 1979), 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-titered serotype-specific antisera, which are expensive and difficult to make and commercially available for only six serotypes - Ia to V (Arakere et al., 1999). Molecular genotyping methods, such as pulsed-field gel electrophoresis (Rolland et al., 1999), restriction endonuclease analysis
35 (Nagano et al., 1991) are useful for epidemiological studies but do not generally identify serotypes. Consequently, there is a need for a reliable molecular method for GBS serotype identification.

WO 03/025216

PCT/AU02/01281

2

Summary of the invention

We have identified specific regions within the genome of group B streptococci of inter-type sequence heterogeneity that can be used to distinguish
5 different types (including capsular polysaccharide gene serotypes and serosubtypes; protein antigen gene subtypes; and mobile genetic element subtypes). We have shown that molecular methods that detect these sequence heterogeneities can be used to accurately distinguish and type group B streptococci.

Accordingly in a first aspect the present invention provides a method of typing a group B streptococcal bacterium which method comprises analysing the nucleotide sequence of one or more regions within the *cpsD*, *cpsE*, *cpsF*, *cpsG*, *cpsI/M* genes of said bacterium, said region(s) comprising one or more nucleotides whose sequence varies between types.

In particular, the nucleotide sequence may be analysed for one or more positions corresponding to positions 62, 78-86, 138, 139, 144, 198, 204, 211, 281, 240, 249, 300, 321, 419, 429, 437, 457, 466, 486, 602, 606, 627, 636, 645, 803, 971, 1026, 1044, 1173, 1194, 1251, 1278, 1413, 1495, 1500, 1501, 1512, 1518, 1527, 1595, 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892, 1971, 2026, 2088, 2134, 2187 and 2196 as shown in Figure 1.

10 Preferably at least one region is within a sequence delineated by the 3' 136 bases of the *cpsE* gene and the 5' 218 bases of the *cpsG* gene of the *cpsE-cpsF-cpsG* gene cluster of said group B streptococcal bacterium. In particular, the nucleotide sequence may be analysed for one or more positions corresponding to positions 1413, 1495, 1500, 1501, 1512, 1518, 1527, 1595,
15 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892, 1971, 2026, 2088, 2134, 2187 and 2196 as shown in Figure 1.

In one embodiment, at least one region is within the *cpsI/M* genes of said group B streptococcal bacterium.

We have also shown that a number of surface protein antigen genes,
20 including *rib*, *alp2* or *alp3* genes, and five mobile genetic elements may be used to molecular subtype GBS. Accordingly, the present invention also provides a method of typing a group B streptococcal bacterium which method comprises determining the presence or absence in the genome of said bacterium of one or more surface protein antigen genes selected from a *rib*, *alp2* or *alp3* gene, and/or
25 one or more mobile genetic elements selected from IS861, IS1548, IS1381,

WO 03/025216

PCT/AU02/01281

3

ISSa4 and GBSi1. Preferably, such as method is combined with the above methods of the invention.

The nucleotide sequence analysis step may comprise sequencing said one or more regions. Alternatively, or in addition, the nucleotide sequence analysis step may comprises determining whether a polynucleotide obtained from said bacterium selectively hybridises to a polynucleotide probe comprising one or more of the said regions, preferably to one or more of a plurality of polynucleotide probes corresponding to one or more of the said regions.

In a preferred embodiment, where hybridisation to a plurality of probes is used as a means of analysis, the plurality of polynucleotide probes are present as a microarray.

In another embodiment, the nucleotide sequence analysis step comprises an amplification step using one or more primers, at least one of which hybridise specifically to a sequence which differs between types. Typically, primer pairs are used, at least one of which hybridise specifically to a sequence which differs between types. Preferably, said primers are selected from the primers shown in Table 2 and/or Table 6 and/or Table 10.

In a second aspect, the present invention provides a polynucleotide consisting essentially of at least 10 contiguous nucleotides corresponding to a region within a *cpsD-cpsE-cpsF-cpsG* gene of a group B streptococcal bacterium, said polynucleotide comprising one or more nucleotides which differ between GBS types.

Preferably the nucleotides which differ between GBS types correspond to one or more of positions 62, 78-86, 138, 139, 144, 198, 204, 211, 281, 240, 249, 300, 321, 419, 429, 437, 457, 466, 486, 602, 606, 627, 636, 645, 803, 971, 1026, 1044, 1173, 1194, 1251, 1278, 1413, 1495, 1500, 1501, 1512, 1518, 1527, 1595, 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892, 1971, 2026, 2088, 2134, 2187 and 2196 as shown in Figure 1.

The present invention also provides a polynucleotide consisting essentially of at least 10 contiguous nucleotides corresponding to a region within a sequence delineated by the 3' 136 base pairs of *cpsE* and the 5' 218 base pairs of *cpsG* of the *cpsE-cpsF-cspG* gene cluster of a group B streptococcal bacterium, said polynucleotide comprising one or more nucleotides which differ between GBS types.

Preferably the nucleotides which differ between group B streptococcal types correspond to one or more of positions 1413, 1495, 1500, 1501, 1512, 1518, 1527, 1595, 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892, 1971, 2026, 2088, 2134, 2187 and 2196 as shown in Figure 1.

WO 03/025216

PCT/AU02/01281

4

The present invention also provides a polynucleotide consisting essentially of at least 10 contiguous nucleotides corresponding to a region within a *cpsI/M* gene of a group B streptococcal bacterium, said polynucleotide comprising one or more nucleotides which differ between group B streptococcal types.

5 Preferably the polynucleotide is selected from the nucleotide sequences shown in Table 2.

The present invention further provides a polynucleotide consisting essentially of at least 10 contiguous nucleotides corresponding to a region within a *rib*, *alp2* or *alp3* gene of a group B streptococcal bacterium, said polynucleotide
10 comprising one or more nucleotides which differ between GBS protein antigen gene subtypes.

Preferably the polynucleotide is selected from the nucleotide sequences shown in Table 6.

The present invention further provides a polynucleotide consisting essentially of at least 10 contiguous nucleotides corresponding to a region within
15 *IS861*, *IS1548*, *IS1381*, *ISSa4* and/or *GBSi1* of a group B streptococcal bacterium, said polynucleotide comprising one or more nucleotides which differ between GBS mobile genetic element subtypes.

Preferably the polynucleotide is selected from the nucleotide sequences
20 shown in Table 10.

The polynucleotides of the invention may be used in a method of typing, such as serotyping and/or subtyping, a group B streptococcal bacterium.

In a third aspect the present invention provides a composition comprising a plurality of polynucleotides of the second aspect of the invention. The
25 composition may be used in a method of typing, such as serotyping and/or subtyping, a group B streptococcal bacterium.

In a fourth aspect the present invention provides a microarray comprising a plurality of polynucleotides according to the second aspect of the invention. The
30 microarray may be used in a method of typing, such as serotyping and/or subtyping, a group B streptococcal bacterium.

Detailed description of the invention

Unless defined otherwise, all technical and scientific terms used herein have the same meaning as commonly understood by one of ordinary skill in the
35 art (e.g., in cell culture, molecular genetics, nucleic acid chemistry, hybridization techniques and biochemistry). Standard techniques are used for molecular, genetic and biochemical methods (see generally, Sambrook *et al.*, Molecular

WO 03/025216

PCT/AU02/01281

5

Cloning: A Laboratory Manual, 3rd ed. (2001) Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y. and Ausubel *et al.*, Short Protocols in Molecular Biology (1999) 4th Ed, John Wiley & Sons, Inc. - and the full version entitled Current Protocols in Molecular Biology, which are incorporated herein by reference) and chemical methods.

The molecular typing methods of the present invention rely on detecting the presence in sample of specific polynucleotide sequences in regions of the genome of group B streptococci (GBS) that we have identified as varying between different types.

More specifically, the specific polynucleotide sequences that are to be detected lie within *cpsD*, *cpsE*, *cpsF*, *cpsG*, *cpsI*, *cpsM*, *rib*, *alp2* and/or *alp3* genes of GBS as well as mobile genetic elements *IS861*, *IS1548* and *IS1381*, *ISSa4* and *GBSI1*, preferably the *cpsD*, *cpsE*, *cpsF*, *cpsG* and/or *cpsI/M* genes.

Regions of interest within those genes mentioned are regions whose sequence varies between two or more types, i.e. are heterogenous. Heterogeneity may be due to insertions, deletions and/or substitutions between corresponding regions in different types. In the case of *rib*, *alp2* and *alp3*, heterogeneity typically takes the form of the presence or absence of the entire gene. Similarly for elements *IS861*, *IS1548*, *IS1381*, *ISSa4* and *GBSI1* heterogeneity typically takes the form of the presence or absence of the entire sequence.

Specific regions of heterogeneity include the following positions within *cpsD* gene- 62 and 78-86; *cpsD-cpsE* gene spacer - 138, 139 and 144; *cpsE* gene - 198, 204, 211, 281, 240, 249, 300, 321, 419, 429, 437, 457, 466, 486, 602, 606, 627, 636, 645, 803, 971, 1026, 1044, 1173, 1194, 1251, 1278, 1413, 1495, 1500, 1501, 1512, 1518 and 1527; *cpsF* gene - 1595, 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892 and 1971; and *cpsG* gene - 2026, 2088, 2134, 2187 and 2196 (numbering corresponds to numbering shown in Figure 1).

Particularly preferred positions of interest are those that lie within a 790 bp fragment of *cpsE-cpsF-cpsG* (which consists of approximately the 3' 136 bases of *cpsE* to the 5' 218 bases of *cpsG*), namely positions 1413, 1495, 1500, 1501, 1512, 1518, 1527, 1595, 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892, 1971, 2026, 2088, 2134, 2187 and 2196 as shown in Figure 1.

Another region of heterogeneity is position 62 of *cpsD* and a repetitive sequence (TTACGGCGA) found at positions 78 to 86 of *cpsD* in some but not all GBS serotypes.

WO 03/025216

PCT/AU02/01281

6

Specific regions of heterogeneity also include a number of positions within the *cpsI/M* gene as shown in the sequence alignment depicted in Figure 3.

These regions of heterogeneity may be analysed using a variety of means including sequencing, PCR and binding of labelled probes.

5 In the case of sequencing to identify serotype, the sequencing primers are selected such that they hybridise specifically to a region within or near to a region within which a region of heterogeneity is present. The primers need not be specific to particular serotypes since the actual sequence information obtained during the sequencing process which is used to assign molecular serotype. Thus
10 the primers may hybridise specifically to all GBS serotypes (at least serotypes Ia to VII), or to specific serotypes.

Preferred primers anneal within 100, 50 or 20 contiguous nucleotides of a heterogeneous position within the 790 bp region of *cpsE-cpsF-cpsG* shown in Figure 1. Examples of suitable sequencing primers are shown in Table 2
15 (*cpsES3*, *cpsFA*, *cpsFS*, *cpsGA* and *cpsGA1*).

PCR and other specific hybridisation- based serotyping methods will typically involve the use of nucleotide primers/probes which bind specifically to a region of the genome of a GBS serotype which includes a nucleotide which varies between two or more serotypes. Thus the primers/probes may comprise a
20 sequence which is complementary to one of such regions. Where positions of heterogeneity are close together (e.g. positions 198, 204, 211 and 218 of *cpsE*), it may be desirable to use a primer/probe which hybridises specifically to a region of the GBS genome that comprises two or more positions of heterogeneity. Thus for example, a primer/probe may be designed that is complementary to
25 nucleotides 195 to 220 of *cpsE*. Such primers/probes are likely to have improved specificity and reduce the likelihood of false positives.

PCR-based methods of detection may rely upon the use of primer pairs, at least one of which binds specifically to a region of interest in one or more, but not all, serotypes. Unless both primers bind, no PCR product will be obtained.
30 Consequently, the presence or absence of a specific PCR product may be used to determine the presence of a sequence indicative of specific GBS serotypes. However, as mentioned, only one primer need correspond to a region of heterogeneity in the genes of interest (such as the *cpsD*, *cpsE*, *cpsF*, *cpsG*, *cpsI* and/or *cpsM* genes). The other primer may bind to a conserved or heterogeneous region within said gene or even a region within another part of the GBS genome,
35 such as the *cpsH* gene, whether said region is conserved or heterogeneous between serotypes. Thus, for example, a combination of a primer (*cpsGS*) which binds to a region of the *cpsG* gene including positions 2172 to 2210, and a primer

WO 03/025216

PCT/AU02/01281

7

which binds to a region of *cpsH* gene which is heterogeneous (IacpsHA1, IIIcpsHA), may be used as the basis of distinguishing serotypes (Ia and III).

Further, a primer which binds to a region of *cpsI* which is heterogeneous may be combined with a primer which binds to a region of *cpsG* which is constant. An example of such as primer pair is primer pair VIcpsIA, and cpsGS1, which give rise to a PCR product of 1517 bp and GBS serotype VI specific.

Alternatively, primers that bind to conserved regions of the GBS genome but which flank a region whose length varies between serotypes may be used. In this case, a PCR product will always be obtained when GBS bacteria are present but the size of the PCR product varies between serotypes.

Furthermore, a combination of specific binding of one or both primers and variations in the length of PCR primer may be used as a means of identifying particular molecular serotypes.

Examples of specific primers/probes which target the *cpsD*, *cpsE*, *cpsF*, *cpsG*, *cpsI* or *cpsM* genes include the following:

cpsDS GCA AAA GAA CAG ATG GAA CAA AGT GG
 cpsES CTT TTG GAG TCG TGG CTA TCT TG
 cpsEA1 GA/T/GA AAA AAG GAA AGT CGT GTC G/ATT G
 20 cpsES1 CTT GGA C/TTC CTC TGA AAA GGA TTG
 cpsEA2 AAA A/CGC TTG ATC AAC AGT TAA GCA GG
 cpsES2 GAT GGT/C GGA CCG GCT ATC TTT TCT C
 cpsEA3 CTT AAT TTG TTC TGC ATC TAC TCG C
 cpsES3 GTT AGA TGT TCA ATA TAT CAA TGA ATG GTC TAT TTG GTC AG
 25 cpsEFA CCT TTC AAA CCT TAC CTT TAC TTA GC
 cpsFS CAT CTG GTG CCG CTG TAG CAG TAC CAT T
 cpsFA GTC GAA AAC CTC TAT A/GT A AAC/T GGT CTT ACA A/GCC AAA
 TAA CTT ACC
 cpsGA AAG/C AGT TCA TAT CAT CAT ATG AGA G
 30 cpsGA1 CCG CCA/G TGT GTG ATA ACA ATC TCA GCT TC
 cpsGS ATG ATG ATA TGA ACT CTT ACA TGA AAG AAG CTG AGA TTG
 cpsGS1 GAA CTC TTA CAT GAA AGA AGC TGA GAT TGT TAT CAC AC
 IbcpsIA CTA TCA ATG AAT GAG TCT GTT GTA GGA CGG ATT GCA CG
 IbcpsIS GAT AAT AGT GGA GAA ATT TGT GAT AAT TTA TCT CAA AAA
 35 GAC G
 IbcpsIA1 CCT GAT TCA TTG CAG AAG TCT TTA CGA TGC GAT AGG TG
 IVcpsMA GGG TCA ATT GTA TCG TCG CTG TCA ACA AAA CCA ATC AAA TC
 VcpsMA CCC CCC ATA AGT ATA AAT AAT ATC CAA TCT TGC ATA GTC AG

WO 03/025216

PCT/AU02/01281

8

Vlcp1A GAA GCA AAG ATT CTA CAC AGT TCT CAA TCA CTA ACT CCG
 cps1A GTA TAA CTT CTA TCA ATG GAT GAG TCT GTT GTA GTA CCG

The primer designations correspond to those given in Table 2.

- 5 In relation to the *alp2*, *alp3* and *rib* surface protein antigen genes, heterogeneity and protein antigen gene subtype is assessed more at the level of whether a group B streptococcal bacterium contains the gene or not. Our results show that the specific combination of surface proteins genes present in a GBS genome is indicative of serotype/serosubtypes (see Table 9). Consequently, primers/probes suitable for use in the methods of the present invention are those that are specific for the particular genes. Thus probes/primers that are specific for *alp2* or *alp3* or *rib* are preferred. Figure 4 shows an alignment of *alp2* and *alp3* that was used to design primers specific for *alp2* or specific for *alp3*.

- 15 Examples of specific primers/probes which target the *alp2*, *alp3* and *rib* genes include the following:

- bcaS1 GGT AAT CTT AAT ATT TTT GAA GAG TCA ATA GTT GCT GCA TCT
 AC
 bcaS2 CCAGGGA GTG CAG CGA CCT TAA ATA CAA GCA TC
 20 balS GAT CCT CAA AAC CTC ATT GTA TTA AAT CCA TCA AGC TAT TC
 balA CCA GTT AAG ACT TCA TCA CGA CTC CCA TCA C
 bal23S1 CAG ACT GTT AAA GTG GAT GAA GAT ATT ACC TTT ACG G
 bal23S2 CTT AAA GCT AAG TAT GAA AAT GAT ATC ATT GGA GCT CGT G
 bal2S CTT CCG CCA GAT AAA ATT AAG
 25 bal2A CTG TTG ACT TAT CTG GAT AGG TC
 bal2A1 CGT GTT GTT CAA CAG TCC TAT GCT TAG CCT CTG GTG
 bal2A2 GGT ATC TGG TTT ATG ACC ATT TTT CCA GTT ATA CG
 bal3S GTT CTT CCG CTT AAG GAT AG
 bal3A GAC CGT TTG GTC CTT ACC TTT TGG TTC GTT GCT ATC C
 30 ribS2 GAAGTAATTCAG GAA GTG CTG TTA CGT TAA ACA CAA ATA TG
 ribA1 GAA GGT TGT GTG AAA TAA TTG CCG CCT TGC CTA ATG
 ribA2 AAT ACT AGC TGC ACC AAC AGT AGT CAA TTC AGA AGG

The primer designations correspond to those given in Table 6.

- 35 In relation to the IS861, IS1548, IS1381, ISSa4 and GBSi1, heterogeneity and subtype is assessed more at the level of whether a group B streptococcal bacterium contains the element or not. The number of elements may also be assessed. Our results show that the specific combination of mobile elements present in a GBS genome is indicative of serotype/serosubtype (see Table 12).

WO 03/025216

PCT/AU02/01281

9

Consequently, primers/probes suitable for use in the methods of the present invention are those that are specific for the particular mobile genetic elements. Thus probes/primers that are specific for IS861, IS1548, IS1381, ISSa4 and GBSi1 are preferred.

- 5 Examples of specific primers/probes which target IS861, IS1548, IS1381, ISSa4 and GBSi1 include the following:

	IS861S	GAG AAA ACA AGA GGG AGA CCG AGT AAA ATG GGA CG
	IS861A1	CAC GAT TTC GCA GTT CTA AAT AAA TCC GAC GAT AGC C
10	IS861A2	CAA ACT CCG TCA CAT CGG TAT AGC ACT TCT CAT AGG
	IS1548S	CTA TTG ATG ATT GCG CAG TTG AAT TGG ATA GTC GTC
	IS1548S1	GTT TGG GAC AGG TAG CGG TTG AGG AGA AAA GTA ATG
	IS1548A1	CAT TAC TTT TCT CCT CAA CCG CTA CCT GTC CCA AAC
	IS1548A2	CCC AAT ACC ACG TAA CTT ATG CCA TTT G
15	IS1548A3	CGT GTT ACG AGT CAT CCC AAT ACC ACG TAA CTT ATG CC
	IS1381S1	CTT ATG AAC AAA TTG CGG CTG ATT TTG GCA TTC ACG
	IS1381S2	GGC TCA GGC GAT TGT CAC AAG CCA AGG GAG
	IS1381A	CTA AAA TCC TAG TTC ACG GTT GAT CAT TCC AGC
	ISSa4S	CGT ATC TGT CAC TTA TTT CCC TGC GGG TGT CTC C
20	ISSa4A1	GCC GAT GTC ACA ACA TAG TTC AGG ATA TAG CCA G
	ISSa4A2	CGT AAA GGA GTC CAA AGA TGA TAG CCT TTT TGA ACC
	GBSi1S1	CAT CTC GGA ACA ATA TGC TCG AAG CTT ACA AGC AAG TG
	GBSi1S2	GGG GTC ACT ATC GAG CAG ATG GAT GAC TAT CTT CAC
	GBSi1A1	AAT GGC TGT TTC GCA GGA GCG ATT GGG TCT GAA CC
25	GBSi1A2	CCA GGG ACA TCA ATC TGT CTT GCG GAA CAG TAT CG

- Preferably, the primers/probes comprise at least 10, 15 or 20 nucleotides. Typically, primers/probes consist of fewer than 100, 50 or 30 nucleotides. Primers/probes are generally polynucleotides comprising deoxynucleotides. They
- 30 may also be polynucleotides which include within them synthetic or modified nucleotides. A number of different types of modification to oligonucleotides are known in the art. These include methylphosphonate and phosphorothioate backbones, addition of acridine or polylysine chains at the 3' and/or 5' ends of the molecule. For the purposes of the present invention, it is to be understood that
- 35 the polynucleotides described herein may be modified by any method available in the art. Primers/probes may be labelled with any suitable detectable label such as radioactive atoms, fluorescent molecules or biotin.

In one embodiment, primers/probes have a high melting temperature of >70°C so that they may be used in rapid cycle PCR.

Compositions comprising a plurality of nucleotides that are used to analyse one or more regions within the *cpsD*, *cpsE*, *cpsF*, *cpsG*, or *cpsM* genes may also further comprise nucleotides that may be used to analyse one or more regions within the *cpsH* gene. Suitable nucleotides are described in the Examples, although a person skilled in the art could design other suitable sequences based on the sequence alignment shown in Figure 3.

Further, compositions comprising a plurality of nucleotides that are used to analyse one or more regions within *alp2*, *alp3* or *rib* genes may also further comprise nucleotides that may be used to analyse one or more regions within the C alpha (*bca*) and C beta (*bac*) genes (C beta gene also known as *bag*).

A variety of techniques may be used to analyse one or more regions within the genome of a bacterium of interest. Typically, a sample of interest, which is suspected of containing GBS bacteria is treated, using standard techniques to obtain genomic DNA from any microorganisms present in the sample. It may be desirable for a number of subsequent detection steps to use nucleic acid preparation techniques that result in substantial fragmentation of the genomic DNA. The sample may be from a bacterial culture or a clinical sample from a patient, typically a human patient. Clinical samples may be cultured to produce a bacterial culture. However, it is also possible to test clinical samples directly with a culturing step.

The genomic DNA is then subjected to one or more analysis steps which may include sequencing, enzymatic amplification and/or hybridisation. These general techniques of DNA analysis are known in the art and are discussed in detail in, for example, Sambrook et al. 2001 and Ausubel et al. 1999 *supra*.

Serotyping may involve a one or more steps. For example, it may be desirable to carry out an initial step of determining whether there are nucleotide sequences present in the sample which are conserved between GBS serotypes but not found in any other organism. This may be achieved by using PCR primers that detect any (but only) GBS bacteria (e.g. using primer pairs Sag59/Sag190 and/or DSF2/DSR1 - see Tables 2 and 3).

Molecular serotyping for specific GBS serotypes can then be performed by detecting the presence of one or more regions of heterogeneity in the regions of interest using any suitable technique such as sequencing, enzymatic amplification and/or hybridisation based on the probes/primers discussed above.

A particularly preferred detection technique is PCR, such as rapid cycle PCR (Kong et al., 2000).

An example of a multi-step serotyping strategy (algorithm) is shown in Figure 2. However, a variety of other strategies are envisaged and can be designed by the skilled person using the sequence heterogeneity information presented herein. In particular, it is preferred that the serotyping procedure
5 comprise at least one analysis step based on analysing one or regions of the *cpsD*, *cpsE*, *cpsF*, *cpsG* and/or *cps/M* genes. This analysis may optionally be combined with an analysis of one or more regions within the *cpsH* gene. Similar techniques may be used to analyse the *cpsH* gene regions and suitable primer sequences and methods are also described in the Examples.

10 Analysis of the presence or absence of the *alp2*, *alp3* and/or *rib* genes may optionally be combined with an analysis of the presence or absence of C alpha (*bca* gene), C beta (*bac*) gene sequences as is described in the Examples. Similar techniques may be used to analyse these regions and suitable primer sequences and PCR methods are also described in the Examples.

15 Furthermore, analysis of the presence or absence of the *alp2*, *alp3* and/or *rib* genes (and optionally the *bca* and *bac* genes) may be combined with an analysis of the presence or absence of mobile genetic elements.

20 Thus a typing strategy may involve an analysis of *cps* genes, surface protein genes and/or mobile genetic elements in various combinations to provide more serosubtyping and subtyping information.

Analysis of GBS genomic sequences using the above techniques may take place in solution followed by standard resolution using methods such as gel electrophoresis. However in a preferred aspect of the invention, the primers/probes are immobilised onto a solid substrate to form arrays.

25 The polynucleotide probes are typically immobilised onto or in discrete regions of a solid substrate. The substrate may be porous to allow immobilisation within the substrate or substantially non-porous, in which case the probes are typically immobilised on the surface of the substrate. Examples of suitable solid substrates include flat glass (such as borosilicate glass), silicon wafers, mica, ceramics and organic polymers such as plastics, including polystyrene and
30 polymethacrylate. It may also be possible to use semi-permeable membranes such as nitrocellulose or nylon membranes, which are widely available. The semi-permeable membranes may be mounted on a more robust solid surface such as glass. The surfaces may optionally be coated with a layer of metal, such as gold,
35 platinum or other transition metal.

Preferably, the solid substrate is generally a material having a rigid or semi-rigid surface. In preferred embodiments, at least one surface of the substrate will be substantially flat, although in some embodiments it may be

desirable to physically separate synthesis regions for different polymers with, for example, raised regions or etched trenches. It is also preferred that the solid substrate is suitable for the high density application of DNA sequences in discrete areas of typically from 50 to 100 μm , giving a density of 10000 to 40000 cm^{-2} .

5 The solid substrate is conveniently divided up into sections. This may be achieved by techniques such as photoetching, or by the application of hydrophobic inks, for example teflon-based inks (Cel-line, USA). Discrete positions, in which each different probes are located may have any convenient shape, e.g., circular, rectangular, elliptical, wedge-shaped, etc.

10 Attachment of the library sequences to the substrate may be by covalent or non-covalent means. The library sequences may be attached to the substrate via a layer of molecules to which the library sequences bind. For example, the probes may be labelled with biotin and the substrate coated with avidin and/or streptavidin. A convenient feature of using biotinylated probes is that the efficiency of coupling to the solid substrate can be determined easily. Since the polynucleotide probes may bind only poorly to some solid substrates, it is often necessary to provide a chemical interface between the solid substrate (such as in the case of glass) and the probes. Thus, the surface of the substrate may be prepared by, for example, coating with a chemical that increases or decreases the hydrophobicity or coating with a chemical that allows covalent linkage of the polynucleotide probes. Some chemical coatings may both alter the hydrophobicity and allow covalent linkage. Hydrophobicity on a solid substrate may readily be increased by silane treatment or other treatments known in the art. Examples of suitable chemical coatings include polylysine and poly(ethyleneimine). Further details of methods for the attachment of are provided in US Patent No. 6,248,521. Methods for immobilizing nucleic acids by introduction of various functional groups to the molecules are also described in Bischoff *et al.*, 1987 (Anal. Biochem., 164:336-344) and Kremsky *et al.*, 1987 (Nucl. Acids Res. 15:2891-2910).

30 Techniques for producing immobilised arrays of nucleic acid molecules have been described in the art. A useful review is provided in Schena *et al.*, 1998, TibTech 16: 301-306, which also gives references for the techniques described therein.

35 Microarray-manufacturing technologies fall into two main categories—synthesis and delivery. In the synthesis approaches, microarrays are prepared in a stepwise fashion by the *in situ* synthesis of nucleic acids from biochemical building blocks. With each round of synthesis, nucleotides are added to growing chains until the desired length is achieved. A number of prior art methods describe

WO 03/025216

PCT/AU02/01281

13

how to synthesise single-stranded nucleic acid molecule libraries *in situ*, using for example masking techniques (photolithography) to build up various permutations of sequences at the various discrete positions on the solid substrate. U.S. Patent No. 5,837,832 describes an improved method for producing DNA arrays immobilised to
5 silicon substrates based on very large scale integration technology. In particular, U.S. Patent No. 5,837,832 describes a strategy called "filling" to synthesize specific sets of probes at spatially-defined locations on a substrate which may be used to produced the immobilised DNA libraries of the present invention. U.S. Patent No. 5,837,832 also provides references for earlier techniques that may also be used.

10 The delivery technologies, by contrast, use the exogenous deposition of preprepared biochemical substances for chip fabrication. For example, DNA may also be printed directly onto the substrate using for example robotic devices equipped with either pins (mechanical microspotting) or piezo electric devices (ink jetting). In mechanical microspotting, a biochemical sample is loaded into a
15 spotting pin by capillary action, and a small volume is transferred to a solid surface by physical contact between the pin and the solid substrate. After the first spotting cycle, the pin is washed and a second sample is loaded and deposited to an adjacent address. Robotic control systems and multiplexed printheads allow automated microarray fabrication. Ink jetting involves loading a biochemical
20 sample, such as a polynucleotide into a miniature nozzle equipped with a piezoelectric fitting and an electrical current is used to expel a precise amount of liquid from the jet onto the substrate. After the first jetting step, the jet is washed and a second sample is loaded and deposited to an adjacent address. A repeated series of cycles with multiple jets enables rapid microarray production.

25 In one embodiment, the microarray is a high density array, comprising greater than about 50, preferably greater than about 100 or 200 different nucleic acid probes. Such high density probes comprise a probe density of greater than about 50, preferably greater than about 500, more preferably greater than about 1,000, most preferably greater than about 2,000 different nucleic acid probes per
30 cm². The array may further comprise mismatch control probes and/or reference probes (such as positive controls).

Microarrays of the invention will typically comprise a plurality of primers/probes as described above. The primers/probes may be grouped on the array in any order. However, it may be desirable to group primers/probes
35 according to types (capsular polysaccharide gene serotypes, serosubtypes; protein antigen gene subtypes; mobile genetic elements subtypes), or groups of types (capsular polysaccharide gene serotypes, serosubtypes; protein antigen gene subtypes; mobile genetic elements subtypes) for which they are specific.

Such grouping may be arranged such that the resulting patterns are easily susceptible to pattern recognition by computer software.

Elements in an array may contain only one type of probe/primer or a number of different probes/primers.

5 Detection of binding of GBS genomic DNA to immobilised probes/primers may be performed using a number of techniques. For example, the immobilised probes which are specific to a number of types (capsular polysaccharide gene serotypes, serosubtypes; protein antigen gene subtypes; mobile genetic elements subtypes), may function as capture probes. Following binding of the genomic
10 DNA to the array, the array is washed and incubated with one or more labelled detection probes which hybridise specifically to regions of the GBS genome which are conserved. The binding of these detection probes may then be determined by detecting the presence of the label. For example, the label may be a fluorescent label and the array may be placed in an X-Y reader under a
15 charge-coupled device (CCD) camera.

Other techniques include labelling the genomic DNA prior to contact with the array (using nick-translation and labelled dNTPs for example). Binding of the genomic DNA can then be detected directly.

20 It is also possible to employ a single PCR amplification step using labelled dNTPs. In this embodiment, the genomic DNA fragment binds to a first primer present in the array. The addition of polymerase, dNTPs, including some labelled dNTPs and a second primer results in synthesis of a PCR product incorporating labelled nucleotides. The labelled PCR fragment captured on the plate may then be detected.

25 A number of available detection techniques do not require labels but instead rely on changes in mass upon ligand binding (e.g. surface plasmon resonance- SPR). The principles of SPR and the types of solid substrates required for use in SPR (e.g. BIACore chips) are described in Ausubel *et al.*, 1999, *supra*.

30

C. Uses

35 As discussed above, 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. Thus, the detection methods, probes/primer and microarrays of the invention may be used in the diagnosis of GBS infections in pregnant women, elderly and/or immunocompromised patients. The PCR and

WO 03/025216

PCT/AU02/01281

15

microarray techniques described herein may be of particular use in routine antenatal screening of pregnant women as well as in diagnosing infections in pregnant women given the increased accuracy and sensitivity compared to conventional identification and serotyping. These methods are also likely to give
5 faster results since it will not generally be necessary to culture clinical samples to obtain enough material. Further, the molecular techniques can be used in most laboratories without the need for specialist expertise or reagents.

The molecular typing methods of the invention may also assist in comprehensive strain identification that will be useful for epidemiological and
10 other related studies that will be needed to monitor GBS isolates before and after introduction of GBS conjugate vaccines.

The present invention will now be described in more detail with reference to the following examples, which are illustrative only and non-limiting. The
15 examples refer to Figures:

Detailed description of the Figures.

Figure 1. Molecular serotype identification based on the sequence heterogeneity of the 3'-end of *cpsD-cpsE-cpsF*-and the 5'-end of *cpsG* (relevant primers are shown).
20

Figure 2. Algorithm for GBS molecular serotype (MS) identification by PCR and sequencing.
25

Figure 3. Multiple sequence alignments of the gene sequences of *cpsG-cpsH-cpsI/M* for serotypes Ia, Ib, II, III, IV, V and VI (start and stop codons are highlighted in bold).

30 Figure 4. Two sites (*) of sequence heterogeneity between *alp2* (AF208158, upper lines) and *alp3* (AF291065, lower lines) used to distinguish them (relevant primers are shown).

Figure 5. Genetic relationship of 194 invasive Australasia GBS strains (or 56
35 genotypes).

Notes for column headed "Genetic Markers of GBS genotypes":

Protein antigen gene profile codes are:

WO 03/025216

PCT/AU02/01281

16

"A": 5' end of *bca* positive;
 "a" or "as": *bca* repetitive unit or *bca* repetitive unit-like region positive,
 with multiple or single band amplicons, respectively;
 "B": *bac* positive;
 5 "R": *rib* positive;
 "alp2": *alp2* positive;
 "alp3": *alp3* positive;
 "None": isolate contains none of the above protein genes.

The molecular markers in bold type show the common features in each cluster.

10

Notes for column headed "No. of strains":

After "+" are the numbers of CSF isolates, the others are blood isolates.

Notes for column headed "Genotypes":

15 Each genotype was characterized by a distinct combination of the *cps*
 genes, protein gene profiles and mobile genetic elements. The predominant
 genotype in each serotype were named as the number "1" genotype of that
 serotype.

Notes for the dendrogram:

20 At about distance 16, the 56 genotypes could be separated into 8 clusters
 (1-8); at about distance 22.5 the 56 genotypes could be separated into 3 cluster
 groups (A, B, C).

EXAMPLES

25 MATERIALS AND METHODS

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
 30 Diana Martin, Streptococcus Reference Laboratory, at ESR, Wellington, New
 Zealand, provided another panel of nine international reference GBS type-strains
 including serotypes Ia to VI (reference panel 2) (Table 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
 35 sterile sites over a period of 10 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

WO 03/025216

PCT/AU02/01281

17

Reference Laboratory, at ESR, Wellington, New Zealand, and MS at the Centre for Infectious Diseases and Microbiology Laboratory Services, ICPMR, Sydney, Australia.

5 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.

10 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°C.

Invasive GBS clinical isolates

15 All 194 isolates used in the study of mobile genetic elements were recovered from the blood (177) or CSF (17) of 191 patients (107 female, 80 male, four sex unrecorded; three cultures each contained mixed growth of two GBS serotypes). 108 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 Institute of Environmental Science and Research (ESR), Porirua, Wellington, New Zealand for serotyping, from various diagnostic laboratories in New Zealand, during 1994-2000.

20 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-14 years); young adult (15-45 years); middle-aged (46-60 years); elderly (>60 years).

25 These isolates are mainly a subset of the isolates described above but with reference strains and non-invasive isolates excluded.

30 Conventional serotyping (CS).

CS was performed using standard methodology (Wilkinson and 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, Ic, II, III, IV, V and the R protein antigen. Fourteen selected isolates, including six that

WO 03/025216

PCT/AU02/01281

18

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.

Molecular serotype identification (MS); development of method.

Oligonucleotide primers.

The oligonucleotide primers used in this study, their target sites and melting temperatures are shown in Tables 2, 6 and 10. Their specificities and expected lengths of amplicons are shown in Tables 3, 7 and 11. The primers were synthesised according to our specifications by Sigma-Aldrich (Castle Hill NSW, Australia). 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 individual GBS serotypes. Six previously published oligonucleotide primers and a series of new primers designed by us were used to sequence parts of and/or to specifically amplify genes encoding GBS surface proteins. We also designed a series of primers to sequence parts of and/or to specifically amplify five known GBS mobile genetic elements. Some were designed with high melting temperatures (>70 °C) to be used in rapid cycle PCR.

DNA preparation and polymerase chain reaction (PCR).

Five individual GBS colonies or a sweep of culture were sampled using a disposable loop and resuspended in 1 ml 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°C (dry block heater or water bath) for 10 minutes then quenched on ice and centrifuged for 2 minutes at 14,000 rpm 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 (25µL 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°C for 1 second, 55-72°C (according to the primer T_m values or as previously described) for 1 second and 74°C for 1 to 30 seconds (according to the length of amplicons), respectively, for 35 cycles.

WO 03/025216

PCT/AU02/01281

19

10 μ L of PCR products were analysed by electrophoresis on 1.5 % agarose gels, which were stained with 0.5 μ g ethidium bromide mL⁻¹. For detection and/or serotype identification, the presence of PCR amplicons of expected length, shown by ultraviolet transillumination, were accepted as
5 positive. For sequencing, 40 μ L volumes of PCR products were further purified by polyethylene glycol precipitation method (Ahmet et al., 1999).

Sequencing.

The PCR products were sequenced using Applied Biosystems (ABI) *Taq*
10 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 Pileup and Pretty
15 programs in Multiple Sequence Analysis program group. Sequences were compared using Bestfit program in Comparison program group. All programs are provided in WebANGIS, ANGIS (Australian National Genomic Information Service), 3rd version.

Surface protein gene profile codes

Each isolate was given a protein gene profile code according to positive
20 PCR results using various primer pairs, as shown in Table 7.

Nucleotide sequence accession numbers.

The new sequence data described have been submitted to the GenBank
Nucleotide Sequence Databases and allocated the following accession numbers:
AF291411-AF291419 (16S/23S rRNA intergenic spacer regions for serotypes Ia
to VIII reference strains from reference panel 1); AF332893-AF332917,
30 AF363032-AF363060, AF367973, AF381030 and AF381031 (partial *cps* gene
clusters for two panels of reference strains (Table) and selected representative
clinical isolates); AF367974 (partial *bac* gene sequence, with an insertion
sequence *IS1381* from one isolate), AF362685-AF362704 (partial *bac* gene
sequences for all *bac*-positive isolates) and AF373214 (partial *rib*-like gene for
35 reference strain Prague 25/60, an R protein standard strain).

Previously reported sequence data referred to herein have appeared in the
GenBank Nucleotide Sequence Databases with the following accession numbers:
AB023574 (16S rRNA gene); U39765, L31412 (16S/23S rRNA intergenic spacer

WO 03/025216

PCT/AU02/01281

20

regions; X68427 (*S. oralis* 23S rRNA gene); X72754 (*cfb* gene); AB028896 (*cps* gene cluster for serotype Ia); AB050723 (partial *cps* gene cluster for serotype Ib); AF163833 (*cps* gene cluster for serotype III); AF355776 (*cps* gene cluster for serotype IV); AF349539 (*cps* gene cluster for serotype V); AF337958 (*cps* gene cluster for serotype VI); M97256 (*bca* gene); X58470, X59771 (*bac* gene); U58333 (*nib* gene); AF208158 (*alp2* gene), AF291065-AF291072 (*alp3* gene); AF064785 (IS1381); M22449 (IS861); Y14270 (IS1548); AF064785 (IS1381); AF165983 (ISSa4); and AJ292930 (GBSi1).

10 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, IV-VI, sst III-1-4; pgp "A", "R", "a", "as", "alp2", alp3"; *bac* subgroups 1, 1a, 2, 3, 3a, 3b, 3c, 4, 4b, 5a, 7, 7a, 8, 9, 9a, 10, n1, n2; and *mge* IS1381, IS861, IS1548, ISSa4, GBSii - were included in the analysis. The genotypes were each characterized by a distinct combination of the molecular serotyping (MS) or sst, pgp and *mge*.

20 Example 1 - Study of inter- and intra-serotype/serosubtype sequence heterogeneity in specific regions of the GBS genome and assessment of suitability for molecular serotyping/serosubtyping.

Polymerase chain reaction.

25 With two exceptions, all GBS-specific primer pairs produced amplicons of the expected size from all reference strains and clinical isolates tested (Table 3). 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.

35 Sequence heterogeneity of 16S/23S rRNA intergenic spacer regions.

The 16S/23S rRNA intergenic spacer regions were sequenced for the serotypes Ia to VIII from reference panel 1. Multiple sequence alignment showed

WO 03/025216

PCT/AU02/01281

21

differences between serotypes at only two positions: 207 (serotype V is T or C [T/C], serotypes VII and VIII are C, others are T) and 272 (serotype III is T, others G). These regions are therefore unsuitable for MS.

5 **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 1 for GenBank accession numbers of reference panel strains.

Repetitive sequence.

15 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 4). The presence or absence of this repetitive sequence can be used to further subtype MS Ia, II and III (see below).

Intra-serotype heterogeneity.

In general, intra-serotype heterogeneity was low - there were minor random variations in a few isolates of all serotypes except MS III, in which the intra-serotype 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 (at 78-86) (Table 4).

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

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

WO 03/025216

PCT/AU02/01281

22

There were three serosubtype 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 serosubtype III-4 (as in MS VII), and C in MS II. This difference can be used (in addition to PCR, see below) to differentiate serosubtype III-4 from MS II. Two serosubtype III-4 isolates contained the repetitive sequence, and the other did not. Because of the small number of serosubtype III-4 isolates, we did not use the repetitive sequence to subtype them further.

10 *Inter-serotype heterogeneity.*

There were 56 sites of heterogeneity between the eight MS (Table 4). The most suitable sites, for use in PCR/sequencing for MS, were a group of 23 sites nearest to the 3'-end of the region (Table 4, Figure 1). Firstly, they were consistent across two panels of reference strains and most clinical isolates (the only exceptions were the small number of serosubtypes III-3 and III-4 isolates, see below). Secondly, they were relatively concentrated within a 790 bp region, 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, serosubtype III-3 cannot be distinguished from MS Ia, nor serosubtype 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. This result was confirmed by the use of serotype VIII-specific antiserum.

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 2226/2217 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 serosubtype III-3 isolates and three serosubtype 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

WO 03/025216

PCT/AU02/01281

23

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 serosubtypes III-3 and III-4, occurred at the 5'-end region of the 2226/2217 fragment. This supported our selection of the 790 bp 3'-end as the sequencing target for MS. Using this target, all MS were correctly identified except for MS III belonging to serosubtypes III-3 and III-4, which can be identified by MS III-specific PCR (see Example 2).

Example 2 - Molecular serotype identification (MS) based on MS-specific PCR targeting the 3'-end of *cpsG-cpsH-cps I/cpsM*.

Our sequence alignment results showed that there was significant sequence heterogeneity in the 3'-end of *cpsG-cpsH-cps I/cpsM* (Figure 3), 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. To fulfil possible additional future requirements - for example, development of multiplex PCR and/or to allow further evaluation of the sequence typing method, we designed several primer pairs for each serotype (Tables 2 & 3). Using two panels of reference strains and the specified conditions, all primer pairs amplified DNA only from the corresponding serotypes. When clinical isolates were tested, similar results were obtained with two sets of MS-specific primer pairs. In general, more stringent conditions (lower primer concentration, higher annealing temperatures) could be used with primers generating smaller amplicons. Those selected for MS are shown in Table 3 and Figure 2.

A MS was assigned, by PCR, to 179 of 206 (86.9%) clinical isolates as follows: MS Ia 40; MS Ib 35; MS III 58 (including those previously identified as serosubtypes III-3 and III-4); MS IV 7; MS V 36; MS VI 3.

Example 3 - 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 but five of which (see below) were resolved by retesting and/or correction of clerical errors.

The CS and MS/sequence subtyping results are shown in Table 5. 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 presumptive MS VIII isolate was decided by exclusion (see Example 1). For all other isolates, the results of PCR and sequencing were consistent, except for serosubtypes III-3 and III-4 and other minor sequence differences described above (Example 1). 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; serosubtype III-1, three; serosubtype III-2, one; V, five; and VI, one.

Sequences (2217 bp) of three clinical isolates that we identified as MS VI, were identical with those for serotype VI reference strains and the corresponding sequence in GenBank (AF337958).

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 serosubtype 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 one MS III-specific PCR positive colony was subsequently identified as serosubtype III-2 and included as an additional clinical isolate (total 206 in all).

35

Example 4 - Algorithm for serotype assignment of GBS by PCR and sequencing

As an example of how the PCR and sequencing methods described above may be used clinically to perform GBS serotype identification, we designed an algorithm for clinical use. All the primers (except the inner sequencing primers) used were given high melting temperature (>70 °C), so rapid cycle PCR could be used (Figure 2) (see Table 2 for primer sequences).

Example 5 - Identification of regions in the *alp2*, *alp3* and *rib* genes suitable for protein antigen gene specific subtyping

Polymerase chain reactions.

With few exceptions, all primer pairs produced amplicons of predicted length from isolates giving positive results (Table 7). The exceptions included one isolate that was positive by PCR using primer pairs GBS1360S/GBS1937A and GBS1717S/GBS1937A (which both target *bac* gene) but produced amplicons significantly longer than those of other *bac* gene-positive isolates. Sequencing showed that the amplicon contained the insertion sequence IS1381 with minor variations compared with the published sequences (Tamura et al., 2000). The amplicons produced using primers IgAagGBS/RlgAagGBS and IgAS1/IgAA1 (also targeting *bac* gene) varied in length (Berner et al., 1999) and were sequenced for further subtyping (see below and Table 8).

Amplicon sequencing results.

To confirm the specificity of selected primer pairs that we had designed or modified, we sequenced 10 of 23 amplicons produced by *bcaS1/bcaA* (targeting the 5'-end of *bca* gene) and all of those produced by *ribS1/ribA3* (targeting *rib* gene) and GBS1360S/GBS1937A (targeting *bac* gene), from the two panels of reference strains and 31 randomly selected clinical isolates.

All 10 amplicons of primers *bcaS1/bcaA* and 12 of 13 of primers *ribS1/ribA3* were identical with the corresponding gene sequences in GenBank (M97256, *bca* gene and U58333, *rib* gene, respectively). One additional isolate, namely Prague 25/60 in reference panel 2 (which is used to raise R antiserum), produced an amplicon with primer pair *ribS1/ribA3* only at a lower annealing temperature (55 °C) but not with *ribS2/ribA1* and *ribS2/ribA2*. It was therefore assumed not to contain *rib* gene, although the amplicon sequence showed considerable homology with *rib* gene (71.4% or 66.6% according to whether or not the primer sequences were included) (Figure 3). This isolate was the only

one, of 224 tested, for which PCRs were negative using ribS2/ribA1 and ribS2/ribA2 but positive using ribS1/ribA3. The latter primer pair is assumed to be not entirely specific for *rib* gene and was therefore used only for sequencing.

Four of 10 amplicons of primer pair GBS1360S/GBS1937A (targeting *bac* gene) were identical with the corresponding sequence in GenBank (X58470, X59771). A single point mutation (A to G, 1441 of X59771) was found in the remaining six *bac* gene amplicons, including the one which contained the insertion sequence IS1381 (see above and AF367974).

Amplicons from all of the 224 isolates that gave positive PCR results using primer pairs bcaS1/balA (targeting *alp2/alp3* genes), bal23S1/bal2A2 (targeting *alp2* gene) and IgAagGBS/RlgAagGBS (targeting *bac* gene) were sequenced.

Fifty isolates produced amplicons using primer pair bcaS1/balA. The sequences of nine were identical with the corresponding portions of the published sequence of *alp2* gene (AF208158) and 41 with that of *alp3* gene (AF291065). There are two consistent heterogeneity sites between *alp2* and *alp3* genes in the sequences of bcaS1/balA amplicons (Figure 4), which can be used to distinguish them, in addition to *alp2* and *alp3* gene -specific PCR. All nine amplicons of primer pair bal23S1/bal2A2 were identical with the corresponding portion of the *alp2* gene sequence in GenBank (AF208158).

The primer pair IgAagGBS/RlgAagGBS identified *bac* gene in 52 isolates. There was considerable sequence variation, which allowed separation of *bac* gene -positive isolates into 11 groups and 20 subgroups based on amplicon length and sequence heterogeneity, respectively (Table 8). The groups contained small numbers (one to five) of isolates except for B1 (20 isolates, 2 subgroups) and B4 (11 isolates, 3 subgroups). The differences in amplicon length was generally caused by the presence or absence of short repetitive sequences.

Further confirmation of specificity of surface protein gene-specific primer pairs.

To confirm primer specificity, we compared the results of PCR using the primer sequences we had designed or modified for *bac* gene PCR, with those of PCR using previously published primers and found 100% correlation.

The previously reported non-specificity of the published primer pair bcaRUS/bcaRUA (targeting the *bca* gene repetitive unit) was confirmed. Using these primers, all nine *alp2* gene positive (bcaS1/bcaA negative) isolates and 53 which were PCR negative using the primers bcaS1/bcaA, bcaS2/bcaA (targeting the 5'-end of *bca* gene), bal23S1/bal2A2 and bal23S2/bal2A1 (targeting the 5'-end of *alp2* gene) produced amplicons. Our sequencing showed that *bca* gene

WO 03/025216

PCT/AU02/01281

27

and *alp2* gene have significant homology in the regions targeted by *bcaRUS/bcaRUA* allowing amplicon formation from *alp2* gene -positive strains. These false positive results could be due to the presence of other C alpha-like proteins, containing regions homologous with the *bca* gene repetitive unit (*bca* gene repetitive unit-like sequence).

We also showed that the results of PCR using two or more primer pairs that we had designed for individual genes (*rib*, *alp2*, and *alp3* genes) correlated well, supporting the specificity of each set. The only exception, as mentioned above, was *ribS1/ribA3*, which produced a non-specific amplicon from one of 224 isolates tested.

Example 6 - The relationship between surface protein antigen gene profiles and cps serotypes/serosubtypes.

Surface protein gene profiles.

For each gene (except *bca* gene repetitive unit or *bca* gene repetitive unit-like region), we selected two primer pairs to identify and characterise GBS surface protein by PCR. Each isolate was given a protein gene profile code according to PCR results as follows:

"A": 5'end of *bca* gene amplified by *bcaS1/bcaA* and *bcaS2/bcaA*;

"a" or "as": *bca* gene repetitive unit or *bca* gene repetitive unit-like region amplified by *bcaRUS/bcaRUA*, with multiple or single band amplicons, respectively;

"B": *bac* gene amplified by GBS1360S/GBS1937A and IgAagGBS/RlgAagGBS (>20 subgroups based on sequence heterogeneity).

"R": *rib* gene amplified by *ribS2/ribA1* and *ribS2/ribA2*;

"alp2": *alp2* gene amplified by *bal23S1/bal2A2* and *bal23S2/bal2A1* and

"alp3": *alp3* gene amplified by *bal23S1/bal3A* and *bal23S2/bal3A* (Table 7).

Four common profiles accounted for 203 of 224 (90.6%) isolates: "R" (62 isolates), "AaB" (51 isolates), "a" (49 isolates) and "alp3" (41 isolates) (see Table 4). Only two isolates contained no surface protein gene markers. All but one isolate with the *bac* gene ("B") also had *bca* gene, with its repetitive unit ("Aa"); one had *rib* gene. All "alp2" isolates contained single *bca* repetitive unit-like sequences ("as"). "A", "R", "alp2" and "alp3" were all mutually exclusive. 62 of 63 isolates with *rib* gene ("R") and 41 of 41 isolates with *alp3* gene had no other protein antigen markers.

The relationship between surface protein antigen gene profiles and cps serotypes/serosubtypes.

5 A *cps* molecular serotype (MS) was assigned to all isolates in accordance with the methods described in Examples 1 to 4 and the results correlated with conventional serotyping (CS) results except for 19 of 224 isolates that were nontypable using antisera. The relationship between surface protein gene profiles and *cps* MS are summarised in Table 9.

10 The following strong associations were confirmed or demonstrated between: MS Ia and *bca* gene repetitive unit or *bca* gene repetitive unit-like sequence (most with profile "a"); MS serosubtypes III-1 and III-2 and *rib* gene; MS serosubtype III-3 and *alp2* gene; MS Ib and *bca/bac* genes and MS V and *alp3* gene. MS II showed the most varied surface protein gene profiles. However, the relationships were not absolute and different combinations of *cps* serotypes and protein gene profiles produced 31 different serovariants or 51 when *bac* gene ("B") subgroups were considered.

Example 7 - The relationship between surface protein antigens and protein gene profiles.

20 Based on conventional serotyping, 33 isolates (belonging to CS Ia/c, Ib/c, IIc, IIb, IIIc or IIId) reacted with the C antiserum. The surface protein gene profiles of all these isolates contained *bca* gene ("A") or *bca* gene repetitive unit-related markers ("a" or "as"): Aa, 3; AaB, 18; a, 11; *alp2as*, 1. Twenty nine isolates reacted with the R antiserum and, of these, 22 contained *rib* gene and six, *alp3* gene. The strain used to raise the R protein antiserum (Prague 25/60) contained a presumed *rib*-like gene (see above and Figure 3).

Example 8 - Identification of mobile genetic elements suitable for molecular subtyping

30 We developed a series of PCR primers to screen for the presence of five mobile elements in GBS serotypes.

Specificity of primers pairs.

35 All the primer pairs produced amplicons of the expected lengths (Table 11) from some reference and/or some clinical isolates (Table 12). To evaluate the specificity of our primer pairs, we sequenced all amplicons produced by primers IS1548S/IS1548A3 and ISSa4S/ISSa4A2, and amplicons, selected from both

WO 03/025216

PCT/AU02/01281

29

reference and clinical isolates, produced by IS861S/IS861A2 (12 isolates), IS1381S1/IS1381A (24 isolates) and GBSi1S1/GBSi1A2 (11 isolates).

All 41 IS1548 and 15 ISSa4 amplicon sequences were identical with the corresponding sequences in GenBank (Y14270 and AF165983, respectively).

5 Five of 12 IS861 amplicon sequences were identical with the corresponding IS861 sequence in GenBank (M22449). The other seven differed, at position 732, from the published sequence (G to A) and the reference strain Prague 25/60 had two additional differences - G to A and T to A - at positions 576 and 830 of M22449, respectively.

10 Previously, we found a full-length insertion sequence IS1381 (AF367974) within C beta antigen gene of a clinical isolate, with several differences compared with the original published sequence (AF064785): the terminal inverted repeats contained 15, rather than 20 base pairs (bp); there was a three bp deletion and four individual bp differences in the putative transposase pseudogene between
15 positions 419 to 429 (of the original GenBank sequence) - GGG ATC CGA TT (AF064785) vs CAG A-- -GG TA (AF367974; our sequence). All amplicons of primer pair IS1381S1/IS1381A from 12 reference and 12 selected clinical isolates were identical with each other and with that of our IS1381 sequence in GenBank (AF367974) but different, as above, from the original reported IS1381 sequence
20 (AF064785).

The amplicons of primer pair GBSi1S1/GBSi1A2 from all four GBSi1-positive reference strains and seven selected clinical isolates were sequenced. Six (including those of three reference strains) were identical with the corresponding GBSi1 sequence in GenBank (AJ292930). Amplicons from four
25 clinical isolates showed three site-variations (C to T at position 767, A to C at position 846 and T to C at position 923 of AJ292930 sequence). The reference strain Prague 25/60 showed only the first two of these site-variations.

In addition to sequencing, we evaluated the specificity of our primer pairs by comparing PCR results for two or more primer pairs for each target (Table 11).

30 In all cases, the same sets of isolates gave positive results when tested with PCR targeting the same mobile genetic elements, thus confirming the specificity of the primer pairs.

PCR results using specific primer pairs for all five mobile genetic elements.

35 IS861, IS1548, IS1381, ISSa4 and GBSi1 were identified in 55%, 18%, 85%, 7% and 19% of isolates, respectively. None of the mobile elements was detected in 10 (4%) isolates. The distributions of the five mobile elements identified by PCR in the 224 GBS isolates tested in the previous examples are shown in Table 12. IS1381

WO 03/025216

PCT/AU02/01281

30

was detected alone in 79 isolates and GBSi1 alone in one. Forty-six isolates contained two different insertion sequences (IS861 and IS1381, 42 isolates ; IS1548 and IS1381, three isolates; ISSa4 and IS1381, one isolate). Forty-four isolates contained three (IS861, IS1548 and IS1381 34; IS861, ISSa4 and IS1381, 10) and one contained all four insertion sequences. Forty-one isolates contained GBSi1 in combination with one (IS861, 22; IS1381, one isolate) two (IS861 and IS1381, 11; ISSa4 and IS1381, three isolates) or three (IS861, IS1548 and IS1381, four isolates) insertion sequences.

10 **PCR results for the 194 invasive isolates using specific primer pairs for all five mobile genetic elements - .**

The numbers of isolates containing different mobile genetic elements (mge) combinations (from none to four per isolate) are shown in Table 13. IS1381, IS861, IS1548, ISSa4 and GBSi1 were identified in 87%, 52%, 17%, 6% and 18% of isolates, respectively. Six (3%) isolates contained no mge.

Example 9 - The relationships between *cps* serotypes, serosubtypes, surface protein gene profiles and mobile genetic elements.

The distribution of each of the five mobile genetic elements in different *cps* serotypes, serotype III subtypes and surface protein gene profiles are shown in Tables 12 and 13. The most consistent findings for each sero/serosubtype were:

- 1) Serotype Ia - most (>80%) expressed proteins that closely related with C alpha protein and contained IS1381
- 2) Serotype Ib - most (>90%) expressed C alpha and C beta proteins and contained IS861 and IS1381
- 3) Serotype II - exhibited two common patterns:
 - a) >50% expressed C alpha protein (and often C beta) and contained IS861, IS1381 and sometimes other mobile elements, especially ISSa4 or
 - b) >25% expressed Rib protein and contained IS861, IS1381 and GBSi1
- 4) Serosubtype III-1 - all expressed Rib protein and contained IS861, IS1548 and IS1381 but not GBSi1.
- 5) Serosubtype III-2 - all expressed Rib protein and contained IS861 and GBSi1 but neither IS1548 nor IS1381.
- 6) Serosubtype III-3 - all expressed C alpha-like protein 2 and contained no mobile genetic elements.
- 7) Serosubtype III-4 - expressed various proteins; all contained GBSi1.

WO 03/025216

PCT/AU02/01281

31

- 8) Serotype IV - most expressed proteins that closely related with C alpha protein and contained IS1381
- 9) Serotype V - most expressed C alpha-like protein 3 contained IS1381
- 10) GBSi1 and IS1548 were mutually exclusive in serotype III (III-1, III-2 and III-4) but not in serotype II.
- 11) All isolates that expressed C alpha-like protein 2 contained no insertion sequences.

Predominant relationships between MS/sst, pgp and mge.

- Figure 5 shows the relationships between the various genetic markers. IS1381 was present in nearly all isolates of MS Ia, Ib, IV, V and VI, but in none of sst III-2 or III-3. IS1548 was found exclusively, and GBSi1 most commonly, in serotypes II or III; three isolates (all MS II) contained both GBSi1 and IS1548. IS861 was found in all sst III-1 and III-2 and most MS II and Ib isolates but only in 14% of other MS isolates. ISSa4 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 (1994). IS1381 was found in most isolates except those in cluster 8, pgp "alp2", which had no insertion sequences. IS861 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 (Figure 5).

Theoretical GBS clonal population structure.

- Theoretically there are 13 possible GBS MS/sst (eight MS - Ia, Ib, II, IV-VIII, four sst III 1-4 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 $13 \times 31 \times 2^5 = 12,896$ different possible combinations of molecular markers. The fact that only 56 different combinations were found (Figure 5), demonstrates that markers are not randomly distributed or, in other words, these invasive Australasian GBS isolates have a clonal population structure. It

WO 03/025216

PCT/AU02/01281

32

is possible, but unlikely, that these isolates represent a very limited number of GBS genotypes.

The phylogenetic relationship of Australasian invasive GBS.

5 The 56 genotypes formed eight clusters, separated at a genetic distance of about ~16 (or three cluster groups separated at a distance of ~22.5). The *gpg* was the main determinant of cluster separation (Figure 5). 94% 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:

15 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).

20 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 GBSii (22 isolates, three MS/genotypes; predominant genotype III-2).

25 Cluster 7: "R", IS1381 and IS861 (27 isolates; two MS/genotypes; predominant genotype III-1).

Cluster 8: "alp2as", no IS (six isolates; three MS/genotypes; one contained GBSi1).

The phylogenetic study showed that the dendrogram inferred by SPPS was very robust.

30

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 14. 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 ($p=0.001$).

35

WO 03/025216

PCT/AU02/01281

33

There were 17 isolates from cerebrospinal fluid 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

Capsule production in GBS is controlled by capsular polysaccharide synthesis (*cps*) gene cluster, which had been sequenced for serotype Ia and serotype III before we began our study. Corresponding sequences for serotype Ib (Miyake *et al.*, 2001 submitted into GenBank, GenBank accession number: AB050723), and for serotypes IV, V, and VI (McKinnon *et al.*, 2001 submitted into GenBank, GenBank accession numbers: AF355776, AF349539, AF337958, respectively) were released recently 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 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*. We designed a series of primers to amplify a 2226/2217 bp segment in this region and found that amplicons were obtained from all serotypes except VIII. This confirmed a previous suggestion that serotype VIII is significantly different from other serotypes in this region.

Using eight serotype (Ia to VII) reference strains, we showed more than 50 heterogeneity points between serotypes (Figure 1, Table 4). 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. 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 (Figure 3), making this region a suitable target for direct serotype identification by PCR. We designed several pairs of MS-specific primers for MS Ia, Ib, III, IV, V and VI and used them to test two CS reference panels. Selected primer pairs were used for MS, by PCR alone, of 86.9% of our 206 clinical isolates. 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

WO 03/025216

PCT/AU02/01281

34

this region are available for serotypes II, VII and VIII. Meanwhile, 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* (Figure 1, Table 4). A positive GBS-specific PCR and negative PCR results with all the primers that amplify the 790 bp, identified MS 5 VIII, by exclusion.

In future, and in some laboratories currently, sequencing of the 790 bp PCR amplicons of the 3'-end of *cpsE-cpsF*-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 10 and a small proportion of serotypes would be wrongly assigned (serosubtypes 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 15 most frequently isolated from normally sterile sites, in the United States and several countries. Serotypes VI and VIII are the predominant serotypes isolated from patients in Japan, 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 most common serotypes identified, 20 although there were small numbers of serotypes IV, VI and, VIII.

Up to 13 % of GBS isolates are non-serotypable and in our study the proportion was 8.7% (18/206) using the antisera available. This may be due to decreased type-specific-antigen synthesis; non-encapsulated phase variation; or insertion or mutation in genes of *cps* gene clusters. One non-serotypable strain 25 GBS in our study had a T base deletion in *cpsG* gene, which caused a change in the *cpsG* gene reading frame.

We have also developed PCR-based methods to identify GBS surface protein genes and further characterise these isolates. Using the published *bac* gene sequence, we modified *bac* gene-specific primers and designed new 30 primers, with high melting temperatures (>70 °C) suitable for rapid cycle PCR targeting all major surface protein genes.

As previously reported, a published PCR primer pair targeting the *bca* gene repetitive unit (at the 3'-end of *bca* gene), was not entirely specific for *bca* gene. We designed two new primer pairs targeting the 5'-end of *bca* gene, to 35 improve the specificity. However, very few serotype Ia strains gave positive results using these primers whereas all were PCR positive using primers targeting the *bca* gene repetitive unit. These results were consistent with a previous report, that a probe targeting the 5'-end of *bca* gene hybridized with only

WO 03/025216

PCT/AU02/01281

35

one of nine serotype Ia strains, but a large *bca* gene probe, including the tandem repeat region, hybridized with all nine strains.

5 PCR specific for *rib*, *alp2* and *alp3* genes has not been described previously. The primer pairs we designed mainly targeted the 5'-ends of the gene and were chosen after comparing the gene heterogeneity with related gene sequences. We designed two or more primer pairs for each gene to check primer specificity by comparison of results of different PCR targeting the same genes. Protein gene profiles "alp2" and "alp3" were distinguished on the basis of the *alp2* and *alp3* gene -specific PCR and/or two sequence heterogeneity sites in the amplicons of *bcaS1/balA*, or *bcaS2/balA*.

10 To confirm the specificity of our primers, we used them to examine two reference panels and selected GBS isolates. The longest amplicons produced by PCR for each gene were sequenced, to provide maximal sequence information and ensure that the inner primers were not located at strain heterogeneity sites. Our sequencing results confirmed the specificity of the primers. Two pairs of primers for each gene were compared, with similar results. Finally, six gene/region specific primer pairs (including the one targeting the *bca* gene repetitive unit) were used to define protein antigen gene profiles for all 224 isolates.

20 The study showed that only one member of the surface protein gene family containing repetitive sequences - *rib*, *bca*, *alp2*, and *alp3* genes-could be present in any single isolate. However, all isolates containing *bac* gene, which is not a member of the surface protein gene family containing repetitive sequences, also contained either *bca* gene (51/52) or *rib* gene (1/52).

25 *Bac* gene was present in 23% of isolates, a similar proportion to that (19-22%) previously reported. In common with others, we found variations in the *bac* gene due to variable small internal repetitive sequences. These *bac* gene repetitive sequences were irregular (unlike those of the *bca-rib* gene family). Their role is not clear, but they are potentially useful molecular markers for epidemiological studies.

30 Our data show that some serotype III isolates (our MS serosubtypes III-1 and III-2) were closely associated with *rib* gene, and others (our MS serosubtype III-3) with *alp2* gene. Serotype Ib was associated with *bca* and *bac* genes and serotype V with *alp3* gene. However, as the relationship was not absolute, different combinations of *cps* serotypes-serosubtypes/protein gene profiles identified many serovariants, which will be useful in epidemiological studies and in formulation of conjugate vaccines. Based on PCR only, we were able to divide

our 224 isolates into 31 serovariants based on *bac* gene (B) groups or 51, based on subgroups. Theoretically, there are likely to be additional serovariants.

We found that the antisera to "c" and "R" protein antigens were not entirely specific for any particular protein genes. However, reaction with "c" antiserum mostly reflected the presence of genes encoding C alpha (*bca* gene) and related protein antigens (at least including *alp2* gene) and the antiserum to "R" with those encoding Rib (*rib* gene) and related proteins (at least including *alp3* gene, and the rare presumed *rib*-like gene).

We have also investigated the presence of a number of mobile element in different serotypes of GBS. Four different insertion sequences have been identified previously in GBS. Multiple copies of *IS861* in some serotype III isolates were associated with increased capsule gene expression. We found *IS861* in all serosubtypes III-1 and III-2 and most serotype II and Ib isolates but few others. All *IS861*-containing isolates contained at least one additional mobile element.

Multiple copies of *IS1381* have been found in a high proportion GBS and other *Streptococcus* species, including *S. pneumoniae* and used as probes for restriction fragment length polymorphism (RFLP) analysis of GBS for epidemiological studies (Tamura et al., 2000). We found *IS1381* in 85% of isolates overall. They were present in all isolates of serosubtype III-1 but none of serosubtypes III-2 or III-3. Our *IS1381* sequences, from 24 isolates, were identical with each other, but differed at several sites, from that previously described (AF064785). The significance of these differences is unknown, but it emphasizes the importance of confirming sequences from as many different strains as possible.

ISSa4 was first identified in a nonhemolytic GBS isolate, in which it caused insertional inactivation of the gene *cyfB*, which is part of an ABC transporter involved in production of hemolysin. Only a small proportion of (mainly hemolytic) GBS isolates (4%) contained *ISSa4*, all of which had been isolated since 1996 and it was postulated that *ISSa4* had been newly acquired by GBS. We also found *ISSa4* in only a small proportion of isolates (7%) but it was present in similar proportions of clinical isolates obtained before (4 of 44) and during or after (11 of 162) 1996.

IS1548 was first discovered in some hyaluronidase-negative GBS serotype III isolates, in which it caused insertional inactivation of the gene *hyfB* (one of a cluster responsible for production of hyaluronidase, an important GBS virulence factor) (Granlund et al., 1998). A copy of *IS1548* is also found downstream of the *C5a* peptidase gene (also associated with virulence), in

WO 03/025216

PCT/AU02/01281

37

isolates that contain it. Most *IS1548*-containing isolates were from patients with endocarditis and it was postulated that inactivation of hyaluronidase production and/or some effect on C5a peptidase may allow GBS isolates to adhere to and survive on heart valves.

5 We found *IS1548* in all serosubtype III-1 isolates, which represented 52% of 58 serotype III isolates in our collection, from superficial (eight of 12) and normally sterile (22 of 46) specimens. The latter were from neonates (seven of 20), adults (three of six) and subjects of unspecified age (12 of 20) (data not shown). Although specific clinical data were unavailable, GBS endocarditis is
10 uncommon and likely to have been present in few, if any, of these subjects. Further study is required to elucidate the association with this insertion sequence with specific virulence factors and clinical syndromes.

We found *GBSi1*, a group II intron, in 19% of our 224 isolates overall; it was commonly associated with *IS861*, and the distribution varied with
15 serotype/serosubtype. It was rarely found in serotypes other than II and III. It was present in more than 50% of serotype II isolates, including four, which also contained *IS1548*. It was found in all serosubtypes III-2 and III-4 isolates, in which *IS1548* was not found, but in no serosubtype III-1 isolates which did contain *IS1548* or serosubtype III-3 isolates which did not.

20 Our subdivision of GBS serotype III into four serosubtypes, based on differences within the *cps* gene cluster was supported by corresponding differences in surface protein gene profiles and distribution of the five mobile elements described in this study. Although we did not test our isolates for hyaluronidase activity, it is likely that our serosubtype III-1, which expresses Rib
25 protein and contains *IS1548*, *IS861* and *IS1381*, corresponds with the hyaluronidase negative subtype III-2, described by Bohnsack et al., 2001. Our serosubtype III-2 also expresses Rib protein and contains *IS861* and *GBSi1* and probably corresponds with subtype III-3 of Bohnsack et al., 2001. Serosubtypes III-3 and III-4 were represented by relatively few isolates. The former (in common
30 with some serotype Ia isolates) expressed the C alpha-like protein 2 and contained no mobile elements (an otherwise uncommon finding). The latter is closely related to serotype II, with which it shares sequence homology in a section of the *cps* gene cluster and various surface protein profiles and mobile elements.

35

Summary

Our aim has been to develop a comprehensive genotyping system for group B streptococcus (GBS). Such a system should ideally be reproducible, objective and transportable between laboratories, comparable with and complementary to other typing methods and able to incorporate known virulence markers. Based on these criteria, we first developed a molecular serotyping (MS) method based on the *cps* gene cluster. It compared favourably with, but was more sensitive than, conventional serotyping (CS) and allowed us to identify several subtypes of serotype (sst) III, as described by others. We have also developed a second molecular subtyping method based on the family of genes encoding variable surface protein antigens (*bca/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. Our methods also can identify more members of the family of variable antigen genes and distinguish numerous *bac* subgroups. A third subtyping method uses five mobile genetic elements (mge) including four different insertion sequences (IS) and a type II intron, which have been identified in GBS. The use of this third method further enhances the discriminatory ability of our genotyping system.

We then used our typing system to examine the population genetic structure and age-related disease distribution of genotypes among 194 invasive GBS isolates.

We used mainly invasive GBS isolates to demonstrate the practical value of our genotyping system, confirm their clonal population structure and determine the distribution of genotypes in different patient groups. The isolates originated from patients of all ages with GBS sepsis. About half were consecutive GBS isolates from blood or CSF, at a large diagnostic laboratory in a general adult hospital, with an obstetric unit (i.e there were no isolates from children other than neonates). The rest were consecutive isolates referred for serotyping from all over New Zealand. Thus the overall age distribution is representative of that in the population affected by GBS disease, except that children beyond the early neonatal period are probably under-represented. However, the distribution of genotypes within each age-group should be representative.

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.

The phylogenetic tree derived from our results showed relationships between *cps* serotype and protein gene profiles (pgp). Our results also show that certain known virulence markers – C beta, C alpha variants and hyaluronidase production (indirectly) – were associated with distinct clonal lineages.

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. Study of the relationships between putative high-virulence genotypes and patient characteristics (age and/or underlying risk factors), and whether there are significant differences between CSF isolates (or genotypes) and other invasive or colonising strains, will be facilitated by our genotyping system. Using this system, we have demonstrated a clonal population structure among invasive Australasian GBS isolates. This system will be applied to colonising GBS isolates, to identify markers of virulence.

Thus, 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/sequencing and, importantly, does not require panels of serotype-specific antisera that are increasingly difficult to maintain. All isolates are serotypable and sequencing of a relatively limited 790 bp region can provide additional serosubtyping information for MS III. The molecular methods we have described for serotype identification, together with the protein profiling (or protein antigen subtyping) and identification of mobile genetic elements (or mobile genetic elements subtyping) provide potentially useful markers for further phylogenetic and epidemiological studies of GBS as well as comprehensive strain identification that will be useful for epidemiological and other related studies that will be needed to monitor GBS isolates before and after introduction of GBS conjugate vaccines.

The various features and embodiments of the present, referred to in individual sections above apply, as appropriate, to other sections, mutatis mutandis. Consequently features specified in one section may be combined with features specified in other sections, as appropriate.

All publications mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described methods and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the described modes for carrying out the invention which are readily apparent to those skilled in molecular biology or related fields are intended to be within the scope of the following claims.

REFERENCES

- 5 **Ahmet, Z., P. Stanier, D. Harvey, and D. Holt.** 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.
- Arakere, G., A.E. Flores, P. Ferrieri, and C.E. Frasch.** 1999. Inhibition enzyme-linked immunosorbent assay for serotyping of group B streptococcal isolates. *J. Clin. Microbiol.* **37**:2564-2567.
- 10 **Bohnsack, J. F., S. Takahashi, S. R. Detrick, L. R. Pelinka, L. L. Hammitt, A. . Aly, A. A. Whiting, and E. E. Adderson.** 2001. Phylogenetic Classification of Serotype III Group B Streptococci on the Basis of *hylB* Gene Analysis and DNA Sequences Specific to Restriction Digest Pattern Type III-3. *J. Infect. Dis.* **183**:1694-1697.
- 15 **Cropp, C.B., R.A. Zimmerman, J. Jelinkova, A.H. Auernheimer, R.A. Bolin, and B.C. Wyrick.** 1974. Serotyping of group B streptococci by slide agglutination fluorescence microscopy, and microimmunodiffusion. *J. Lab. Clin. Med.* **84**:594-603.
- 20 **Granlund, M., L. Oberg, M. Sellin, and M. Norgren.** 1998. Identification of a novel insertion element, *IS1548*, in group B streptococci, predominantly in strains causing endocarditis. *J. Infect. Dis.* **177**:967-976
- Hakansson, S., L.G. Burman, J. Henrichsen, and S.E. Holm.** 1992. Novel coagglutination method for serotyping group B streptococci. *J. Clin. Microbiol.* **30**:3268-3269.
- 25 **Harrison, L.H., J.A. Elliott, D.M. Dwyer, J.P. Libonati, P. Ferrieri, L. Billmann, and A. Schuchat.** 1998. Serotype distribution of invasive group B streptococcal isolates in Maryland: implications for vaccine formulation. Maryland Emerging Infections Program. *J. Infect. Dis.* **177**:998-1002.
- 30 **Hassan, A.A., A. Abdulmawjood, A.O. Yildirim, K. Fink, C. Lammler, and R. Schlenstedt.** 2000. Identification of streptococci isolated from various sources by determination of *cfb* gene and other CAMP-factor genes. *Can. J. Microbiol.* **46**:946-951.
- Hickman, M.E., M.A. Rench, P. Ferrieri, and C.J. Baker.** 1999. Changing epidemiology of group B streptococcal colonization. *Pediatrics.* **104**:203-209.

- Holm, S.E., and S. Hakansson. 1988. A simple and sensitive enzyme immunoassay for determination of soluble type-specific polysaccharide from group B streptococci. *J. Immunol. Methods.* **106**:89-94.
- 5 Ké, D., C. Menard, F.J. Picard, M. Boissinot, M. Ouellette, P.H. Roy, and M.G. Bergeron. 2000. Development of conventional and real-time PCR assays for the rapid detection of group B streptococci. *Clin. Chem.* **46**:324-331.
- 10 Kong, F., X. Zhu, W. Wang, X. Zhou, S. Gordon, and G.L. Gilbert. 1999. Comparative analysis and serovar-specific identification of the multiple banded antigen genes of *Ureaplasma urealyticum* biovar one. *J. Clin. Microbiol.* **37**: 538-543.
- Kong, F., S. Gordon, and G.L. Gilbert. 2000. Rapid-Cycle PCR for Detection and Typing of *Mycoplasma pneumoniae* in Clinical Specimens. *J. Clin. Microbiol.* **38**:4256-4259.
- 15 Maeland, J. A., O. G. Brakstad, L. Bevanger, and A. I. Kvam. 1997. *Streptococcus agalactiae* beta gene and gene product variations. *J. Med. Microbiol.* **46**:999-1005.
- Maeland, J. A., O. G. Brakstad, L. Bevanger, and S. Krokstad. 2000. Distribution and expression of bca, the gene encoding the c alpha protein, by *Streptococcus agalactiae*. *J. Med. Microbiol.* **49**:193-198.
- 20 Mawn, J.A., A.J. Simpson, and S.R. Heard. 1993. Detection of the C protein gene among group B streptococci using PCR. *J. Clin. Pathol.* **46**:633-636.
- Nagano, Y., N. Nagano, S. Takahashi, K. Murono, K. Fujita, F. Taguchi, and Y. Okuwaki. 1991. Restriction endonuclease digest patterns of chromosomal DNA from group B beta-haemolytic streptococci. *J. Med. Microbiol.* **35**:297-303.
- 25 Rolland, K., C. Marois, V. Siquier, B. Cattier, and R. Quentin. 1999. Genetic features of *Streptococcus agalactiae* strains causing severe neonatal infections, as revealed by pulsed-field gel electrophoresis and hylB gene analysis. *J. Clin. Microbiol.* **37**:1892-1898.
- Tamura, G. S., M. Herndon, J. Przekwas, C. E. Rubens, P. Ferrieri, and S. L. Hillier. 2000. Analysis of restriction fragment length polymorphisms of the insertion sequence IS1381 in group B Streptococci. *J. Infect. Dis.* **181**:364-368.
- 30 Triscott, M.X., and G.H. Davis. 1979. A comparison of four methods for the serotyping of group B streptococci. *Aust. J. Exp. Biol. Med. Sci.* **57**:521-527.

WO 03/025216

PCT/AU02/01281

42

Wilkinson, H.W., and M.D. Moody. 1969. Serological relationships of type I antigens of group B streptococci. *J. Bacteriol.* **97**:629-34.

Zuerlein, T.J., B. Christensen, and R.T. Hall. 1991. Latex agglutination detection of group-B streptococcal inoculum in urine. *Diagn. Microbiol. Infect. Dis.* **14**:191-194.

WO 03/025216

PCT/AU02/01281

43

Table 1. GBS reference panels used in this study.

Lab strain number	Source	Serotype	MS/ serosubtype	GenBank accession numbers
Reference panel 1¹				
090	Channing	Ia	Ia	AF332893
H36B	Channing	Ib	Ib	AF332903
18RS21	Channing	II	II	AF332905
M781	Channing	III	III-2 ³	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 2²				
NZRM 908 (NCDC SS615)	ESR	Ia	Ia	AF332894
NZRM 909 (NCDC SS618)	ESR	Ib	Ib	AF332904
NZRM 910 (NCDC SS700)	ESR	Ic	Ia	AF332914
NZRM 911 (NCDC SS619)	ESR	II	II	AF332906
NZRM 912 (NCDC SS620)	ESR	III	III-3 ³	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

WO 03/025216

PCT/AU02/01281

44

Notes.

1. Reference panel 1: supplied by Dr Lawrence Paoletti, Channing Laboratory, Boston, USA.
2. Reference panel 2: New Zealand Reference Medical Culture Collection strains supplied by Dr Diana Martin, ESR, Porirua, Wellington, New Zealand.
3. MS III serosubtypes based on sequence heterogeneity; see text for more detail

Table 2. Oligonucleotide primers used in this study.

Primer	Target gene	T _m °C ¹	GenBank accession numbers	Sequence ^{2,4}
CFBS	<i>cfb</i>	56.7	X72754	328GAT GTA TCT ATC TGG AAC TCT AGT G352
Sag59 ⁵	<i>cfb</i>	77.4	X72754	350GTGGCTGGTGCATTGTTAI TTT CAC CAG CTG TAT TAG AAG TA381
Sag190 ⁵	<i>cfb</i>	76.8	X72754	545CAITTAACCGGTTTTTCATAATCI GTT CCC TGA ACA TTA TCT TTG AT500
CFBA	<i>cfb</i>	63.2	X72754	568TTT TTC CAC GCT AGT AAT AGC CTC545
16SS	16S rRNA	69.3	AB023574	1441GCC GCC TAA GGT GGG ATA GAT G1462
23SA	23S rRNA	65.7	X68427	70CGT CGT TTG TCA CGT CCT TC51
DSF2 ⁶	16S rRNA	75.9	AB023574	975CAITCCTTCCTGACC GGC CTA GAG ATA GGC TTT CT1007
DSR1 ⁶	16S rRNA	81.5	AB023574	1250CGTCACCCGG CTT GCG ACT CGT TGT ACC AA1222
<i>cpsDS</i>	<i>cpsD</i>	69.1	AB028896 (Ia), AF163833 (III)	4892/4593GCA AAA GAA CAG ATG GAA CAA AGT GG5007/4618
<i>cpsES</i>	<i>cpsE</i>	65.7	AB028896 (Ia), AF163833 (III)	5300/4910CTT TTG GAG TCG TGG CTA TCT TG5322/4932
<i>cpsEA1</i>	<i>cpsE</i>	65.4	AB028896 (Ia), AF163833 (III)	5431/5041GAT/GA AAA AAG GAA AGT CGT GTC GIATT G5612/5017

cpsES1	cpsE	65.9	AB028896 (Ia), AF163833 (III)	5612/5222CTT GGA C/TTT CTC TGA AAA GGA TTG5635/5245
cpsEA2	cpsE	66.8	AB028896 (Ia), AF163833 (III)	5723/5333AAAA A/CGC TTG ATC AAC AGT TAA GCA GG5698/5308
cpsES2	cpsE	70.2	AB028896 (Ia), AF163833 (III)	6012/5622GAT GGT/C GGA CCG GCT ATC TTT TCT C8036/5646
cpsEA3	cpsE	63.7	AB028896 (Ia), AF163833 (III)	6116/5726C/TT AAT TTG TTC TGC ATC TAC TCG C8092/5702
cpsES3	cpsE	71.5	AB028896 (Ia), AF163833 (III)	6410/6020GTT AGA TGT TCA ATA TAT CAA TGA ATG GTC TAT TTG GTC AG6450/6060
cpsEFA	cpsE/F spacer	62.1	AB028896 (Ia), AF163833 (III)	6526/6136CCT TTC AAA CCT TAC CTT TAC TTA GC6501/6111
cpsFS	cpsF	75.0	AB028896 (Ia), AF163833 (III)	6777/6387CAT CTG GTG CCG CTG TAG CAG TAC CAT T8804/6414
cpsFA	cpsF	73.2	AB028896 (Ia), AF163833 (III)	6859/6469GTC GAA AAC CTC TAT A/GT A AAC/T GGT CTT ACA A/GCC AAA TAA CTT ACC6819/6425
cpsGA	cpsG	54.7	AB028896 (Ia), AF163833 (III)	7162/6772AAG/C AGT TCA TAT CAT CAT ATG AGA G 7138/6748
cpsGA1	cpsG	74.5	AB028896 (Ia), AF163833 (III)	7199/6809CCG CCAG TGT GTG ATA ACA ATC TCA GCT TC7171/6781
cpsGS	cpsG	72.24	AB028896 (Ia), AF163833 (III)	7145/6755ATG ATG ATA TGA ACT CTT ACA TGA AAG AAG CTG AGA TTG 7183/6793
cpsGS1	cpsG	71.62	AB028896 (Ia), AF163833 (III)	7155/6765GAA CTC TTA CAT GAA AGA AGC TGA GAT TGT TAT CAC AC 7192/6802

WO 03/025216

47

PCT/AU02/01281

lacpsHS	cpsH	73.6	AB028696 (le)	7698CAT TCT TTG TTT AAA AA/CT CCT GAT TTT GAT AGA ATT TTA GCA GC7741
lacpsHA	cpsH	75.2	AB028696 (le)	7993GAA TAT TCA AAA AAT CCC ATT GCT CTT TGA GTA TGC ATA CC7953
lacpsHA1	cpsH	66.4	AB028696 (le)	8271GTA AGT TAT CAA AAT ATA ACA TCA TTA CTA TTA CTA GTA GAA ACG G8226
lacpsHS1	cpsH	77.9	AB028696 (le)	8463GGC CTG CTG GGA TTA ATG AAT ATA GTT CCA GGT TTG C8499
lacpsHA2	cpsH	58.5	AB028696 (le)	8499GCA AAC CTG GAA CTA TAT TCA T8478
lbcpsHS0	cpsH	58.6	AB050723 (lb)	3013ATT GCT GCA TTC AAT TCA C3031
lbcpsHS	cpsH	81.9	AB050723 (lb)	3016GCT GCA TTC AAT TCA CTG GCA GTA GGG GTT GTG TCC3051
lbcpsHA	cpsH	67.7	AB050723 (lb)	3297GAT AGT TAA GGG TAT TAT AAG ATT TGA ATA TTC AAA GAA AGC3256
lbcpsHS1	cpsH	74.1	AB050723 (lb)	3546TTT GGT GAG CAT ATA TAA TAG AAT AAT CAA TTT GCG GTC G3585
lbcpsHS2	cpsH	73.7	AB050723 (lb)	3740CTG GCC TAT TTG GAC TAA TAA ATG TGA TTT TAG GTT TGT TTC3781
lbcpsHA01	cpsH	57.7	AB050723 (lb)	3781GAA ACA AAC CTA AAA TCA CAT TTA3758

WO 03/025216

48

PCT/AU02/01281

IbcpsHA1	<i>cpsH</i>	78.5	AB050723 (Ib)	3894GGC GCC ATC AAT ATC TTC AAG TGC AAA AAA TGA AAA TAG G3885
IbcpsIA	<i>cpsI</i>	78.2	AB050723 (Ib)	4086CTA TCA ATG AAT GAG TCT GTT GTA GGA CGG ATT GCA CG4049
IbcpsIS	<i>cpsI</i>	71.1	AB050723 (Ib)	4116GAT AAT AGT GGA GAA ATT TGT GAT AAT TTA TCT CAA AAA GAC G4188
IbcpsIA1	<i>cpsI</i>	78.6	AB050723 (Ib)	4638CCT GAT TCA TTG CAG AAG TCT TTA CGA TGC GAT AGG TG4601
IIIVcpsHS	<i>cpsH</i>	75.3	AF163833 (III), AF337958 (VI)	72787120CAA GAG GAT ATA ACG TTT CAG CGA TTT ATT GCT GAG C73117156
IIcpsHS	<i>cpsH</i>	72.1	AF163833 (III)	7672GAA TAC TAT TGG TCT GTA TGT TGG TTT TAT TAG CAT CGC7710
IIcpsHA	<i>cpsH</i>	71.0	AF163833 (III)	7817GTT ATA AGA AAA ACA AGCGGT GAT AAA TAA GAA AGT CAT ACCT7776
IVcpsHS	<i>cpsH</i>	74.1	AF355776 (IV)	7552CCG TAC ATA CAA CTG TTC TTG TTA GCA TTT ACT TTT CTT TGC7593
IVcpsHS1	<i>cpsH</i>	71.2	AF355776 (IV)	7887CCC AAG TAT AGT TAT GAA TAT TAG TTG GAT GGT TTT TGG7925
IVcpsHA	<i>cpsH</i>	77.3	AF355776 (IV)	7951CAT CTA CAC CCC CAC AAA ATA TTT TCC CAA AAA CCA TC7914
IVcpsHA1	<i>cpsH</i>	58.7	AF355776 (IV)	7958TGT AAA TCA TCT ACA CCC CC 7939

WO 03/025216

PCT/AU02/01281

49

VcpsMA	cpsM	80.7	AF355776 (V)	8265GGG TCA ATT GTA TCG TCG CTG TCA ACA AAA CCA ATC AAA TC8225
VcpsHS	cpsH	76.3	AF349539 (V)	6943GGG TTT AGG CGA GGG AAA CTC AGC TTA CAA AAT AGT G6979
VcpsHS1	cpsH	72.2	AF349539 (V)	7258CAA TTT TTA TAG GGA TGG ACA ATT TAT TCT GAG AAG TGA C7297
VcpsHA	cpsH	71.1	AF349539 (V)	7291TCT CAG AAT AAA TTG TCC ATC CCT ATA AAA ATT GAC ATA C7252
VcpsHS02	cpsH	59.0	AF349539 (V)	7616GAT GTT CTT TTA ACA GGT AGA TTA CAC7642
VcpsHA1	cpsH	66.8	AF349539 (V)	7658GTT GTA AAT GAG CAT AGT GTA ATC TAC CTG TTA AAA GAA C7619
VcpsHS2	cpsH	74.0	AF349539 (V)	7871CCC AGT GTG GTA ATG AAT ATT AGT TGG CTA GTT TTT GG7908
VcpsHA2	cpsH	58.6	AF349539 (V)	7945CTT TTT TAT AGG TTC GAT ACC ATC7922
VcpsMA	cpsM	73.1	AF349539 (V)	8244CCC CCC ATA AGT ATA AAT AAT ATC CAA TCT TGC ATA GTC AG8204
VlcpshS	cpsH	76.7	AF337958 (VI)	7478CAC TAT TCC TAG TTT TTT GTG CAT ATT TGA CAG GGG CAA G7517
VlcpshA	cpsH	76.7	AF337958 (VI)	7517CTT GCC CCT GTC AAA TAT GCA CAA AAA ACT AGG AAT AGT G7478

WO 03/025216

PCT/AU02/01281

50

VlcpsHS1	<i>cpsH</i>	77.2	AF337958 (VI)	<u>7767</u> CCT TAT TGG GCA AGG TAT AAG AGT TCC CTC CAG TCT G7803
VlcpsHA1	<i>cpsH</i>	77.2	AF337958 (VI)	<u>7804</u> CCA CAC TGG AGG GAA CTC TTA TAC CTT GCC CAA TAA G7768
Vlcps/A	<i>cpsI</i>	74.5	AF337958 (VI)	<u>8126</u> GAA GCA AAG ATT CTA CAC AGT TCT CAA TCA CTA ACT CCG8088
cps/A	<i>cpsI</i>	70.3	AB028896 (Ia), AF163833 (III)	<u>8816</u> 8312GTA TAA CTT CTA TCA ATG GAT GAG TCT GTT GTA GTA CGG8778274

Notes.

1. The primer T_m values are provided by the primer synthesiser (Sigma-Aldrich).
2. 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).
3. Underlined sequences show bases added to modify previously published primers.
4. Letters behind "I" indicate alternative nucleotides in different serotypes.
5. Ke *et al.*, 2000.
6. Ahmet *et al.*, 1999

50

Table 3. Specificity and expected lengths of amplicons of using different oligonucleotide primer pairs.

Primer pairs*	Specificity	Length of amplicons (base pairs)
Sag59/Sag190 ^a	GBS (<i>S. agalactiae</i>)	196
CFBS/CFBA	GBS (<i>S. agalactiae</i>)	241
16SS/23SA	GBS (<i>S. agalactiae</i>)	433
DSF2/DSR1 ^a	GBS (<i>S. agalactiae</i>)	276
cpsDS/cpsEA1	serotypes Ia to VII	449/458
cpsES/cpsEA2	serotypes Ia to VII	424
cpsES1/cpsEA3	serotypes Ia to VII	505
cpsES2/cpsEFA	serotypes Ia to VII	515
cpsES3/cpsFA ^b	serotypes Ia to VII	450
cpsFS/cpsGA1 ^b	serotypes Ia to VII	423
cpsES3/cpsGA1 ^b	serotypes Ia to VII	790
cpsGS/cpsIA	serotypes Ia and III	1672/1558
cpsGS1/cpsIA	serotypes Ia and III	1662/1548
cpsGS/lacpsHA1	serotype Ia	1127
cpsGS1/lacpsHA1	serotype Ia	1117
lacpsHS/lacpsHA	serotype Ia	296
lacpsHS/lacpsHA1	serotype Ia	574
lacpsHS1/cpsIA ^c	serotype Ia	354
cpsGS/lbcpsHA1	serotype Ib	1468
cpsGS1/lbcpsHA1	serotype Ib	1458
cpsGS/lbcpsIA	serotype Ib	1660
cpsGS1/lbcpsIA	serotype Ib	1650
lbcpsHS/lbcpsHA	serotype Ib	282
lbcpsHS1/lbcpsHA1	serotype Ib	349
lbcpsHS2/lbcpsIA	serotype Ib	347
lbcpsIS/lbcpsIA1 ^c	serotype Ib	523
cpsGS/IIIcpsHA	serotype III	1063
cpsGS1/IIIcpsHA	serotype III	1053
IIIVcpsHS/IIIcpsHA	serotype III	543
IIIcpsHS/cpsIA ^c	serotype III	641
cpsGS/IVcpsHA	serotype IV	1372
cpsGS1/IVcpsHA	serotype IV	1362
cpsGS/IVcpsMA	serotype IV	1686

WO 03/025216

PCT/AU02/01281

52

cpsGS1/IVcpsMA	serotype IV	1676
IVcpsHS/IVcpsHA	serotype IV	400
IVcpsHS1/IVcpsMA ^c	serotype IV	379
cpsGS/VcpsHA1	serotype V	1096
cpsGS1/VcpsHA1	serotype V	1086
cpsGS/VcpsMA	serotype V	1682
CpsGS1/VcpsMA	serotype V	1672
VcpsHS/VcpsHA	serotype V	349
VcpsHS1/VcpsHA1	serotype V	401
VcpsHS2/VcpsMA ^c	serotype V	374
IIIVcpsHS1/IVcpsHA	serotype VI	398
cpsGS/IVcpsHA1	serotype VI	1205
cpsGS1/IVcpsHA1	serotype VI	1195
cpsGS/IVcpsIA	serotype VI	1527
cpsGS1/IVcpsIA	serotype VI	1517
IVcpsHS/IVcpsHA1 ^c	serotype VI	327
IVcpsHS1/IVcpsIA	serotype VI	360

Notes.

*See Table 2 for primer sequences and Figure 1 for some primer sites.

Primers used in Algorithm for molecular serotype identification-Figure 2

a. to identify GBS, b. for sequencing, c. for MS-specific PCR

Table 4. The heterogeneity of 8 GBS serotypes in the regions of the 3'-end of *cpsD-cpsE-cpsF*-and the 5'-end of *cpsG*.

Sites	Ia	Ib	II/III-4	III	IV	V	VI	VII	Specificity
<i>cpsD</i> gene									
62	G	A	G ⁴	A	A	A	A	G	Ia, II, VII
78-86	-Ia-2 ¹ ; +Ia-1 ¹	-	-II-2 ^{2,4} ; +II-1 ²	-III-3 ³ ; +III-1 ³ ; III-3 ³	+	+	-	+	See text
repetitive sequence -TTACGGCGA									
<i>cpsD/cpsE</i> genes									
spacer									
138	G	G	G	G	G	A ⁵	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
<i>cpsE</i> gene									
198	A	C	A ⁴	A	C	C ⁵	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

1173	A	G	A	A	A	A	A	A	A	lb
1194	C	C	C	A	A	C	A	C	A	III, IV, VI
1251	G	G	G	G	G	A	G	A	G	VI
1278	A	A	A	A	A	G	A	A	V	
1413	C	T	T	C	III-3;	T	T	T	T	Ia, III-3
					T III-2, III-1					
1495	C	C	C	C	C	C	A	C	VI	
1500	A	G	A	A	A	A	A	A	lb	
1501	C	C	T	C	C	C	C	T	II, VII	
1512	C	T	C	T	III-2, III-1;	C	T	T	C	Ia, II, III-3, IV, VII
					C III-3					
1518	T	C	T	T	C	C	T	Ia, II, III-3, IV, VII		
					C III-2, III-1;					
					T III-3					
1527	T	A	A	T	III-3;	T	A	A	Ia, III-3, IV	
					A III-2, III-1					
cpsF gene										
1595	T	C	T	T	T	C	T	C	T	lb, VI
1611	C	C	C	C	C	C	C	T	VII	
1620	C	C	C	C	C	C	T	VII		
1627	G	G	G	G	T	G	G	IV		
1629	G	G	G	G	A	III-1;	G	G	III-1	
					G III-2, III-3					
1655	C	T	C	C	C	C	C	C	lb	
1832	C	C	C	C	T	C	C	C	IV	

WO 03/025216

PCT/AU02/01281

57

Table 5. Comparison of the results of conventional serotyping (CS) and molecular serotype identification (MS)/subtyping of 206 clinical GBS isolates.

CS	MS/serosubtype										
	la	lb	II	III-1 ¹	III-2 ¹	III-3 ¹	III-4 ¹	IV	V	VI	VIII
la	38										
lb		30									
II			25								
III				27	20	4	3				
IV								7			
V									31		
VI										2	
VIII											1
NT ¹	2	5	1	3	1				5	1	
Total (206) ²	40	35	26 ²	30	21 ²	4	3	7	36	3	1

Notes.

1. For details of MS III serosubtypes see text.
2. One mixed culture was included as two separate isolates (one serotype II, one subtype III-2).

Table 6. Oligonucleotide primers used in this study.

Primer	Target gene	T _m °C ¹	GenBank Accession numbers	Sequence ^{2,3}
IgAagGBS ⁵	<i>bac</i>	73.8	X59771	<u>2663GGGATTAAACAA</u> CAA ACT ATT TTT GAT A TTG ACA ATG CAA2702
IgAS1 ⁴	<i>bac</i>	72.8	X59771	2765GGCT AAA TTT CAA AAA GGT CTA GAG ACA AAT ACG CCA G2801
IgAA1 ⁴	<i>bac</i>	78.9	X59771	3157CCC ATC TGG TAA CTT CGG TGC ATC TGG AAG C3127
RigAagGBS ⁵	<i>bac</i>	76.3	X59771	<u>3284CAGCCAACTCITTC</u> GTC GTT ACT TCC TTG AGA TGT AAC3247
GBS1360S ⁶	<i>bac</i>	72.3	X59771	<u>1325GTGAAATTGTAT</u> AAG GCT ATG AGT GAG AGC TTG GAG1360
GBS1717S ⁴	<i>bac</i>	75.0	X59771	1885ACA GTC ACA GCT AAA AGT GAT TCG AAG ACG ACG1717
GBS1937A ⁶	<i>bac</i>	75.9	X59771	<u>1976CCGIIITAGAATCIII</u> CTG CTC TGG TGT TTT AGG AAC TTG1937
BcaRUS ⁷	<i>bca</i> repetitive unit	73.5	M97256	<u>769GATAAATATGATCCAA</u> CAG GAG GGG AAA CAA CAG TAC805
BcaRUA ⁷	<i>bca</i> repetitive unit	77.2	M97256	<u>1003CTGGIIITGGIGTCAGAI</u> GAA CCG TTA CTT CTA CTG TAT CC963

WO 03/025216

29

PCT/AU02/01281

bcas1 ⁴	bcas/alp2/alp3	71.7	M97256 and AF291065	208/533GGT AAT CTT AAT ATT TTT GAA GAG TCA ATA GTT GCT GCA TCT AC 251/576
bcas2 ⁴	bcas/alp2/alp3	78.0	M97256 and AF291065	258/581CCAGGGA GTG CAG CGA CCT TAA ATA CAA GCA TC 288/613
bcas ⁴	bca	58.9	M97256	370GTT TTA GAA CAA GGT TTT ACA GC392
balS ⁴	alp2/alp3	73.8	AF291065	677GAT CCT CAA AAC CTC ATT GTA TTA AAT CCA TCA AGC TAT TC 717
bcasA ⁴	bca	74.2	M97256	597CGTCTAACTT CTT CAA TCT TAT CCC TCA AGG TTG TTG 560
balA ⁴	alp2/alp3	73.6	AF291065	978CCA GTT AAG ACT TCA TCA CGA CTC CCA TCA C 948
bal23S1 ⁴	alp2/alp3	70.9	AF208158 and AF291065	1093/1373CAG ACT GTT AAA GTG GAT GAA GAT ATT ACC TTT ACG G 1129/1409
bal23S2 ⁴	alp2/alp3	72.9	AF208158 and AF291065	1174/1454CTT AAA GCT AAG TAT GAA AAT GAT ATC ATT GGA GCT CGT G 1213/1493
bal2S ⁴	alp2	59.2	AF208158	1363GTT CTT CCG CCA GAT AAA ATT AAG1386
bal2A ⁴	alp2	58.3	AF208158	1576CTG TTG ACT TAT CTG GAT AGG TC1554
bal2A1 ⁴	alp2	78.3	AF208158	1426CGT GTT GTT CAA CAG TCC TAT GCT TAG CCT CTG GTG 1391
bal2A2 ⁴	alp2	70.8	AF208158	1518GGT ATC TGG TTT ATG ACC ATT TTT CCA GTT ATA CG 1484

WO 03/025216

PCT/AU02/01281

60

bal3S ⁴	<i>alp3</i>	57.1	AF291065	1643 GTT CTT CCG CTT AAG GAT AGC A 1664
bal3A ⁴	<i>alp3</i>	79.2	AF291065	1693 GAC CGT TTG GTC CTT ACC TTT TGG TTC GTT GCT ATC C 1657
#ribS1 ⁴	<i>rib</i>	65.2	U58333	216 TAC AGA TAC TGT GTT TGC AGC TGA AG 241
ribS2 ⁴	<i>rib</i>	73.0	U58333	238 GAAGTAATTTCCAG GAA GTG CTG TTA CGT TAA ACA CAA ATA TG 279
ribA1 ⁴	<i>rib</i>	78.8	U58333	431 GAA GGT TGT GTG AAA TAA TTG CCG CCT TGC CTA ATG 396
ribA2 ⁴	<i>rib</i>	72.6	U58333	462 AAT ACT AGC TGC ACC AAC AGT CAA TTC AGA AGG 427
#ribA3 ⁴	<i>rib</i>	61.3	U58333	570 CAT CTA TTT TAT CTC TCA AAG CTG AAG 554

Notes.

#For sequencing use only, not entirely specific for rib gene.

1. The primer Tm values are provided by the primer synthesiser (Sigma-Aldrich).

2. Numbers represent the numbered base positions at which primer sequences start and finish (numbering start point "1" refer to the start point "1" of corresponding GenBank accession number, of which there are two for some sequences).

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

4. Primers designed by us for this study.

5. Mawn *et al.*, 1993.

6. Maeland *et al.*, 1997.

7. Maeland *et al.*, 2000.

WO 03/025216

PCT/AU02/01281

61

Table 7. Specificity and expected lengths of amplicons of using different primer pairs.

Primer pairs*	Specificity	Length of amplicons (base pairs)	Protein profile code
IgAagGBS/ RIgAagGBS	<i>bac</i>	532-838	B
IgAS1/IgAA1	<i>bac</i>	303-591	B
GBS1360S/ GBS1937A	<i>bac</i>	652	B
GBS1717S/ GBS1937A	<i>bac</i>	292	B
bcaS1/bcaA	5'-end of <i>bca</i>	390	A
bcaS2/bcaA	5'-end of <i>bca</i>	342	A
BcaRUS/bcaRUA	<i>bca</i> repetitive unit/ <i>bca</i> repetitive unit-like region	235	a/as
bcaS1/balA	<i>alp2/alp3</i>	446	alp2 or alp3
bcaS2/balA	<i>alp2/alp3</i>	398	alp2 or alp3
balS/balA	<i>alp2/alp3</i>	302	alp2 or alp3
bal23S1/bal2A1	<i>alp2</i>	334	alp2
bal23S2/bal2A1	<i>alp2</i>	253	alp2
bal23S1/bal2A2	<i>alp2</i>	426	alp2
bal23S2/bal2A2	<i>alp2</i>	345	alp2
bal23S1/bal3A	<i>alp3</i>	321	alp3
bal23S2/bal3A	<i>alp3</i>	240	alp3
#ribS1/ribA3	<i>rib/rib-like</i>	355	R/r
ribS2/ribA1	<i>rib</i>	194	R
ribS2/ribA2	<i>rib</i>	225	R
ribS2/ribA3	<i>rib</i>	333	R

Notes.

*See Table 6 for primer sequences.

#For sequencing use only, not entirely specific for *rib* gene (see text for more detail).

WO 03/025216

PCT/AU02/01281

62

Table 8. Genetic groups and subgroups of *bac* gene (C beta protein gene) based on amplicon length (using primers IgAagGBS/RlgAagGBS) and sequence heterogeneity.

Group or Subgroup	N=	Amplicon length	GenBank accession numbers	No. of different sites compared with (c.f.) main group	Molecular serotype/serosubtypes
B1	19	532	X58470		17 = Ib; 2 = II
B1a	1	532	AF362686	1 (c.f. B1)	Ib
B2	3	550	AF362687		Ib, II, III-4
B3	2	586	AF362688		2=Ib
B3a	1	586	AF362689	4 (c.f. B3)	V
B3b	1	586	AF362690	21 (c.f. B3)	VI
B3c	1	586	AF362691	24 (c.f. B3)	Ib
B4	8	604	AF362692		4 = Ib; 4 = II
B4a	1	604	AF362693	1 (c.f. B4)	II
B4b	2	604	AF362694	2 (c.f. B4)	2 = Ib
B5	2	622	AF362695		Ia, VI
B5a	1	622	AF362696	2 (c.f. B5)	Ia
B6	1	640	AF362697		Ib
B7	1	658	AF362698		Ib
B7a	1	658	AF362699	34 (c.f. B7)	VI
B8	1	712	AF362700		Ib
B9	2	748	AF362701		2 = II
B9a	1	748	AF362702	13 (c.f. B9)	Ib
B10	2	820	AF362703		2 = Ib
B11	1	838	AF362704		Ib

Note.

*See Table 9 for further details of serotype/serosubtype relationships with protein antigens.

WO 03/025216

PCT/AU02/01281

63

Table 9. The relationship between GBS protein gene profiles and capsular polysaccharide (cps) molecular serotypes/serosubtypes.

Serotype/ serosubtype *	N=	None	Aa	AaB	R	alp 3	a	as	alp2as	RB	R a
Ia	43	-	-	2	-	-	35	3	3	-	-
Ib	37	-	1	35	-	1	-	-	-	-	-
II	29	-	3	10	8	2	5	-	-	-	1
III-1	30	-	-	-	30	-	-	-	-	-	-
III-2	22	-	-	-	22	-	-	-	-	-	-
III-3	5	-	-	-	-	-	-	-	5	-	-
III-4	3	-	-	1	-	1	-	-	1	-	-
IV	9	-	-	-	1	-	8	-	-	-	-
V	38	1	-	-	1	35	-	-	-	1	-
VI	5	-	1	3	-	-	1	-	-	-	-
VII	1	-	-	-	-	1	-	-	-	-	-
VIII	2	1	-	-	-	1	-	-	-	-	-
Total	224	2	5	51	62	41	49	3	9	1	1

Note.

*See text for explanation of cps serosubtypes and Table 7 for explanation of protein antigen gene profile codes.

Table 10. Oligonucleotide primers used in this study.

Primer	Target	Tm °C ¹	GenBank accession numbers	Sequence ²
IS861S	IS861	77.4	M22449	445 GAG AAA ACA AGA GGG AGA CCG AGT AAA ATG GGA CG479
IS861A1	IS861	77.3	M22449	831 CAC GAT TTC GCA GTT CTA AAT AAA TCC GAC GAT AGC C795
IS861A2	IS861	76.1	M22449	1020 CAA ACT CCG TCA CAT CGG TAT AGC ACT TCT CAT AGG985
IS1548S	IS1548	76.5	Y14270	143 CTA TTG ATG ATT GCG CAG TTG AAT TGG ATA GTC GTC178
IS1548S1	IS1548	77.0	Y14270	539 GTT TGG GAC AGG TAG CGG TTG AGG AGA AAA GTA ATG574
IS1548A1	IS1548	77.0	Y14270	574 CAT TAC TTT TCT CCT CAA CCG CTA CCT GTC CCA AAC539
IS1548A2	IS1548	70.3	Y14270	915 CCC AAT ACC ACG TAA CTT ATG CCA TTT G888
IS1548A3	IS1548	78.0	Y14270	930 CGT GTT ACG AGT CAT CCC AAT ACC ACG TAA CTT ATG CC893
IS1381S1	IS1381	80.1	AF064785/ AF367974	272/818 CTT ATG AAC AAA TTG CGG CTG ATT TTG GCA TTC ACG 307/853
IS1381S2	IS1381	81.7	AF064785/ AF367974	497/1040 GGC TCA GGC GAT TGT CAC AAG CCA AGG GAG526/1069

WO 03/025216

PCT/AU02/01281

65

IS1381A	IS1381	73.1	AF064785/ AF367974	881/1424 CTA AAA TCC TAG TTC ACG GTT GAT CAT TCC AGC 849/1392
ISSa4S	ISSa4	78.5	AF165983	326 CGT ATC TGT CAC TTA TTT CCC TGC GGG TGT CTC C359
ISSa4A1	ISSa4	75.2	AF165983	639 GCC GAT GTC ACA ACA TAG TTC AGG ATA TAG CCA G606
ISSa4A2	ISSa4	74.5	AF165983	780 CGT AAA GGA GTC CAA AGA TGA TAG CCT TTT TGA ACC745
GBSi1S1	GBSi1	78.6	AJ292930	721 CAT CTC GGA ACA ATA TGC TCG AAG CTT ACA AGC AAG TG758
GBSi1S2	GBSi1	77.3	AJ292930	789 GGG GTC ACT ATC GAG CAG ATG GAT GAC TAT CTT CAC824
GBSi1A1	GBSi1	83.9	AJ292930	1058 AAT GGC TGT TTC GCA GGA GCG ATT GGG TCT GAA CC1024
GBSi1A2	GBSi1	80.5	AJ292930	1161 CCA GGG ACA TCA ATC TGT CTT GCG GAA CAG TAT CG1127

Notes.

1. The primer T_m values were provided by the primer synthesiser (Sigma-Aldrich).
2. Numbers represent the numbered base positions at which primer sequences start and finish (numbering start point "1" refers to the start point "1" of corresponding gene GenBank accession number).

WO 03/025216

PCT/AU02/01281

66

Table 11. Specificity and expected lengths of amplicons of using different oligonucleotide primer pairs.

Primer pairs*	Specificity	Length of amplicons (base pairs)
IS861S/IS861A1	IS861	387
IS861S/IS861A2	IS861	576
IS1548S/IS1548A1	IS1548	432
IS1548S/IS1548A2	IS1548	773
IS1548S/IS1548A3	IS1548	788
IS1548S1/IS1548A2	IS1548	377
IS1548S1/IS1548A3	IS1548	392
IS1381S1/IS1381A	IS1381	610/607#
IS1381S2/IS1381A	IS1381	385
ISSa4S/ISSa4A1	ISSa4	314
ISSa4S/ISSa4A2	ISSa4	455
GBSi1S1/GBSi1A1	GBSi1	338
GBSi1S1/GBSi1A2	GBSi1	441
GBSi1S2/GBSi1A1	GBSi1	270
GBSi1S2/GBSi1A2	GBSi1	373

Notes.

*See table 10 for primer sequences.

Our sequencing result (GenBank accession number: AF367974) was 3 bp shorter than that previously described by Tamura et al., 2000 (GenBank accession number: AF064785).

WO 03/025216

PCT/AU02/01281

67

Table 12. Relationship between mobile genetic elements and capsular polysaccharide serotypes, serotype III subtypes and surface protein gene profiles.

Serotype/ serosubtype	Protein gene profile	N=	IS861	IS1548	IS1381	ISSa 4	GBS11	No mobile element
la	AaB	2	2	-	2	-	-	-
la	alp2as	3	-	-	-	-	-	3
la	a	35	3	1	35	1	-	-
la	as	3	-	-	3	-	-	-
subtotal		43	5	1	40	1	-	3
lb	Aa	1	-	-	-	-	-	1
lb	AaB	35	30	-	35	1	-	-
lb	alp3	1	-	-	1	-	-	-
subtotal		37	30	-	36	1	-	1
II	Aa	3	3	1	3	2	1	-
II	AaB	10	10	5	10	5	1	-
II	alp3	2	1	1	2	-	-	-
II	R	8	8	-	8	-	8	-
II	Ra	1	1	-	-	-	1	-
II	a	5	2	2	5	3	5	-
subtotal		29	25	9	28	10	16	-
III-1	R	30	30	30	30	1	-	-
III-2	R	22	22	-	-	-	22	-
III-3	alp2as	5	-	-	-	-	-	5
III-4	AaB	1	1	-	1	-	1	-
III-4	alp2as	1	-	-	-	-	1	-
III-4	alp3	1	-	-	1	-	1	-
subtotal		60	53	30	32	1	25	5
IV	R	1	1	-	1	-	1	-
IV	a	8	2	-	8	-	-	-
subtotal		9	3	-	9	-	1	-
V	alp3	35	3	1	35	1	1	-
V	R	1	1	-	1	1	-	-
V	RB	1	1	-	1	-	-	-
V	none	1	-	-	-	-	-	1
subtotal		38	5	1	37	1	1	2

WO 03/025216

PCT/AU02/01281

68

VI	Aa	1	-	-	1	-	-	-
	AaB	3	3	-	3	-	-	-
	a	1	-	-	1	-	-	-
subtotal		5	3	-	5	-	-	-
VII	alp3	1	-	-	1	-	-	-
VIII	alp3	1	-	-	1	-	-	-
	none	1	-	-	1	-	-	-
subtotal		2	-	-	2	-	-	-
Total		224	124	41 (18)	190	15 (7)	43 (19)	10 (4)
			(55)		(85)			

Note.

A: 5'-end of *bca* gene (C alpha protein);a: *bca* gene repetitive unit or *bca* gene repetitive unit-like sequence (multiple band amplicon);as: *bca* gene repetitive unit or *bca* gene repetitive unit-like sequence (single band amplicon);B: C beta/IgA binding protein (*bac*) gene.R: Rib protein (*rib*) gene;alp2: C alpha-like protein 2 (*alp2*) gene;alp3: C alpha-like protein 3 (*alp3*) gene;

r: assumed Rib-like protein gene.

WO 03/025216

PCT/AU02/01281

69

Table 13. Distribution of mobile genetic elements among 194 invasive GBS isolates.

Total N =	Mobile genetic elements present					None
	IS1381	IS861	IS1548	ISSa4	GBSi1	
6	—	—	—	—	—	6
78	78	—	—	—	—	—
2	—	—	—	—	2	—
37	37	37	—	—	—	—
1	1	—	1	—	—	—
3	3	—	—	3	—	—
29	29	29	29	—	—	—
6	6	6	—	6	—	—
8	8	8	—	—	8	—
18	—	18	—	—	18	—
1	1	—	—	—	1	—
1	1	—	1	—	1	—
2	2	2	2	—	2	—
2	2	—	—	2	2	—
Total	168 (87%)	100 (52%)	33 (17%)	11 (6%)	34 (18%)	6 (3%)
(n=194)						

Note.

Data are numbers of isolates containing various combinations of mge

WO 03/025216

PCT/AU02/01281

70

Table 14 Relationship between GBS genotypes and invasive disease age.

Serotype Genotype	Age-group/disease ¹						Total
	0-6d	7-3m	4m-14yr	15-45 yr	46-60 yr	>60 yr	
Ia-1	14	4+1	1	7	3	6	35+1 (19%)
Ia-(2-8)	4	2	-	1	-	3	10
Ia total	18 (34%)	6+1 (21%)	1 (10%)	8 (28%)	3 (18%)	9 (17%)	45+1 (24%)
Ib-1	2	1+1	-	3	2	5+1	13+2
Ib-(2-16)	3	4+2	-	3	1	5	16+2
Ib total	5 (9.4%)	5+3 (24%)	-	6 (21%)	3	10+1	29+4 (17%)
II	8 (15%)	1 (3%)	-	4+1 (17%)	1	4 (7%)	18+1 (10%)
III-1	6+1 (13%)	4 (12%)	1+1 (20%)	1+1 (7%)	6+1 (41%)	4	22+4 (13%)
III-2	5 (9%)	5+4 (39%) ³	1 (10%)	2	-	-	13+4 (9%)
III-(3-4)	1+1	1	-	1	1	1	5+1
III total	12+2 (26%)	10+4 (41%)	2+1 (30%)	4+1 (17%)	7+1 (44%)	5 (9%)	40+9 (25%)
IV total	3	-	-	-	-	4	7 (4%)
V-1	3	3	2	4	2	13+1	27+1 (14%)
V-(2-7)	1	1	-	1	-	4	7
V total	4 (8%)	4 (12%)	2 (20%)	5 (17%)	2 (11%)	17+1 (33%)⁴	34+1 (18%)
VI total	1	-	-	-	+1	3	4+1 (3%)
TOTAL	51+2=53	26+8=34	5+2=7	27+1=29	16+2=18	52+2=54	177+17=194

Notes:

1. Numbers after "+" refer to CSF isolates; all others are from blood.
 2. Five aged 4m-1yr and one case was aged 3 yr.
 3. Sst III-2 in late onset infection compared with all other groups: p=0.0005, odds ratio (OR) 6.8; 95% confidence interval (CI) 2.4-19.4.
- MS-V in elderly compared with all other age-groups: p=0.001, OR 0.28; 95% CI 0.13-0.59).

CLAIMS

1. A method of typing a group B streptococcal bacterium which method comprises analysing the nucleotide sequence of one or more regions within the *cpsD*, *cpsE*, *cpsF*, *cpsG* and/or *cpsI/M* genes of said bacterium, said region(s) comprising one or more nucleotides whose sequence varies between types.
2. A method according to claim 1 wherein the nucleotide sequence is analysed for one or more positions corresponding to positions 62, 78-86, 138, 139, 144, 198, 204, 211, 281, 240, 249, 300, 321, 419, 429, 437, 457, 466, 486, 602, 606, 627, 636, 645, 803, 971, 1026, 1044, 1173, 1194, 1251, 1278, 1413, 1495, 1500, 1501, 1512, 1518, 1527, 1595, 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892, 1971, 2026, 2088, 2134, 2187 and 2196 as shown in Figure 1.
3. A method according to claim 1 wherein at least one region is within a sequence delineated by the 3' 136 bases of the *cpsE* gene and the 5' 218 bases of the *cpsG* gene of the *cpsE-cpsF-cspG* gene cluster of said streptococcal bacterium.
4. A method according to claim 3 wherein the nucleotide sequence is analysed for one or more positions corresponding to positions 1413, 1495, 1500, 1501, 1512, 1518, 1527, 1595, 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892, 1971, 2026, 2088, 2134, 2187 and 2196 as shown in Figure 1.
5. A method according to any one of claims 1 to 4 wherein at least one region is within the *cpsI/M* genes of said bacterium.
6. A method according to any one of claims 1 to 5 wherein the nucleotide sequence analysis step comprises sequencing said one or more regions.
7. A method according to any one of claims 1 to 5 wherein the nucleotide sequence analysis step comprises determining whether a polynucleotide obtained from said bacterium selectively hybridises to a polynucleotide probe comprising one or more of the said regions.
8. A method according to claim 7 which comprises determining whether the polynucleotide obtained from said bacterium hybridises to one or more of a plurality of polynucleotide probes corresponding to one or more of the said regions.

9. A method according to claim 9 wherein the plurality of polynucleotide probes are present as a microarray.
10. A method according to any one of claims 1 to 5 wherein the nucleotide sequence analysis step comprises an amplification step using one or more primers, at least one of which hybridises specifically to a sequence which differs between types.
11. A method according to any one of claims 1 to 6 wherein the nucleotide sequence analysis step comprises an amplification step using primer pairs, at least one of which hybridise specifically to a sequence which differs between types.
12. A method according to claim 10 or claim 11 wherein said primers are selected from the primers shown in Table 2.
13. A method of typing a group B streptococcal bacterium which method comprises determining the presence or absence in the genome of said bacterium of one or more surface protein genes selected from *rib*, *alp2* or *alp3* genes.
14. A method according to claim 13 wherein determining the presence or absence of said surface protein genes comprises determining whether a polynucleotide obtained from said bacterium selectively hybridises to a polynucleotide probe corresponding to a region of said surface protein genes.
15. A method according to any one of claim 13 wherein determining the presence or absence of said surface protein genes comprises an amplification step using one or more primers which amplify specifically a region of said surface protein genes.
16. A method according to claim 15 wherein said primers are selected from the primers shown in Table 6.
17. A method according to any one of claims 1 to 12 which further comprises determining the presence or absence of in the genome of said bacterium of one or more surface protein genes selected from *rib*, *alp2* or *alp3* genes.

18. A method of typing a group B streptococcal bacterium which method comprises determining the presence or absence in the genome of said bacterium of one or more mobile genetic elements selected from *IS861*, *IS1548*, *IS1381*, *ISSa4* and *GBSi1*.

19. A method according to claim 18 wherein determining the presence or absence of said mobile genetic elements comprises determining whether a polynucleotide obtained from said bacterium selectively hybridises to a polynucleotide probe corresponding to a region of said mobile genetic elements.

20. A method according to any one of claim 18 wherein determining the presence or absence of said mobile genetic elements comprises an amplification step using one or more primers which amplify specifically a region of said mobile genetic elements.

21. A method according to claim 20 wherein said primers are selected from the primers shown in Table 10.

22. A method according to any one of claims 13 to 17 which further comprises determining the presence or absence in the genome of said bacterium of one or more mobile genetic elements selected from *IS861*, *IS1548*, *IS1381*, *ISSa4* and *GBSi1*.

23. A polynucleotide consisting essentially of at least 10 contiguous nucleotides corresponding to a region within a *cpsD-cpsE-cpsF-cpsG* gene of a group B streptococcal bacterium, said polynucleotide comprising one or more nucleotides which differ between group B streptococcal serotypes.

24. A polynucleotide according to claim 23 wherein said nucleotides which differ between group B streptococcal serotypes correspond to one or more of positions 62, 78-86, 138, 139, 144, 198, 204, 211, 281, 240, 249, 300, 321, 419, 429, 437, 457, 466, 486, 602, 606, 627, 636, 645, 803, 971, 1026, 1044, 1173, 1194, 1251, 1278, 1413, 1495, 1500, 1501, 1512, 1518, 1527, 1595, 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892, 1971, 2026, 2088, 2134, 2187 and 2196 as shown in Figure 1.

25. A polynucleotide consisting essentially of at least 10 contiguous nucleotides corresponding to a region within a sequence delineated by the 3' 136 base pairs of

WO 03/025216

PCT/AU02/01281

74

cpsE and the 5' 218 base pairs of *cpsG* of the *cpsE-cpsF-cspG* gene cluster of a group B streptococcal bacterium, said polynucleotide comprising one or more nucleotides which differ between group B streptococcal types.

26. A polynucleotide according to claim 25 wherein said nucleotides which differ between group B streptococcal types correspond to one or more of positions 1413, 1495, 1500, 1501, 1512, 1518, 1527, 1595, 1611, 1620, 1627, 1629, 1655, 1832, 1856, 1866, 1871, 1892, 1971, 2026, 2088, 2134, 2187 and 2196 as shown in Figure 1.

27. A polynucleotide consisting essentially of at least 10 contiguous nucleotides corresponding to a region within a *cpsM* gene of a group B streptococcal bacterium, said polynucleotide comprising one or more nucleotides which differ between streptococcal serotypes.

28. A polynucleotide according to claim 27 wherein the polynucleotide is selected from the nucleotide sequences shown in Table 2.

29. A polynucleotide consisting essentially of at least 10 contiguous nucleotides corresponding to a region within a *rib*, *alp2* or *alp3* gene of a group B streptococcal bacterium, said polynucleotide comprising one or more nucleotides which differ between group B streptococcal subtypes.

30. A polynucleotide according to claim 29 wherein the polynucleotide is selected from the nucleotide sequences shown in Table 6.

31. Use of a polynucleotide according to any one of claims 23 to 30 in a method of serotyping and/or subtyping a group B streptococcal bacterium.

32. A composition comprising a plurality of polynucleotides according to any one of claims 23 to 30.

33. Use of a composition according to claim 32 in a method of serotyping and/or subtyping a group B streptococcal bacterium.

34. A microarray comprising a plurality of polynucleotides according to any one of claims 23 to 30.

WO 03/025216

PCT/AU02/01281

75

35. Use of a microarray according to claim 34 in a method of serotyping and/or subtyping a group B streptococcal bacterium.

Figure 1. Multiple sequence alignments of the regions of the 3' end of cpsD-cpsE-cpsF-and the 5' end of cpsG for reference strains of serotypes Ia to VII.

	1				50
Serotype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serotype III-3	-----	-----	-----	-----	-----
Serotype Ia-1	-----	-----	-----	-----	-----
Serotype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serotype Ia-2	-----	-----	-----	-----	-----
Consensus	GCAAAAGAAC	AGATGGAACA	AAGTGGTTCA	AAGTTCCTAG	GTATTATCTT
		cpsDS			
	51				100
Serotype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-g-----	-----	-----	-----
Serotype VII	-----	-g-----	-----	-----	-----
Serotype III-3	-----	-----	-----	-----	-----
Serotype Ia-1	-----	-g-----	-----	-----	-----
Serotype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serotype Ia-2	-----	-g-----	-----	-----	-----
Consensus	TAATAAAGTT	AATGAATCTG	TTGCTACTTA	CGCGGATTAC	GCGGATTATG
	101				150
Serotype III-2	-----	-----	-----	-----	-a-g-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serotype III-3	-----	-----	-----	-----	-----
Serotype Ia-1	-----	-----	-----	-----	-----
Serotype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serotype Ia-2	-----	-----	-----	-----	-----
Consensus	GAAATTACGG	AAAAAGGGAT	AGAAAAAGGA	AGTAAAGGGC	TCTTGTATTG
			cpsD		
	151				200
Serotype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serotype III-3	-----	-----	-----	-----	-----
Serotype Ia-1	-----	-----	-----	-----	-----
Serotype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----

WO 03/025216

PCT/AU02/01281

2/25

```

Serotype V -----M--
Serosubtype Ia-2 -----
Consensus AAAGAAAAG AAAATATACA AAAGATTATT ATAGCGATGA TTCAAACAGT
                | cpsE
201                                     250
Serosubtype III-2 ---a-----
Serotype VI -----g-----t-----c-----
Serotype Ib -----c-----
Serotype II/III-4 -----
Serotype VII -----
Serosubtype III-3 -----a-----
Serosubtype Ia-1 -----
Serosubtype III-1 -----
Serotype IV -----c-----
Serotype V -----f-----
Serosubtype Ia-2 -----
Consensus TGTGGTTTAT TTTTCTGCAA GTTTGACATT AHCATTAATT ACTCCCAATT
251                                     300
Serosubtype III-2 -----t-----
Serotype VI -----t-----
Serotype Ib -----
Serotype II/III-4 -----
Serotype VII -----
Serosubtype III-3 -----
Serosubtype Ia-1 -----
Serosubtype III-1 -----
Serotype IV -----
Serotype V -----
Serosubtype Ia-2 -----
Consensus TTAAAAGCAA TAAGATTTA TTGTTTGTC TATTGATACA TTATATTGTC
301                                     350
Serosubtype III-2 -----
Serotype VI -----
Serotype Ib -----
Serotype II/III-4 -----
Serotype VII -----
Serosubtype III-3 -----
Serosubtype Ia-1 -----
Serosubtype III-1 -----t-----
Serotype IV -----
Serotype V -----
Serosubtype Ia-2 -----
Consensus TTTTATCTTT CTGATTTTTA CAGAGACTTT TCGAGTCGTG GCTATCTTGA
                cpsES
351                                     400
Serosubtype III-2 -----
Serotype VI -----
Serotype Ib -----
Serotype II/III-4 -----
Serotype VII -----
Serosubtype III-3 -----
Serosubtype Ia-1 -----
Serosubtype III-1 -----

```

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

3/25

Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	AGAGTTTAAA	ATGGTATTGA	AATACAGCTT	TTACTATATT	TTCATATCAA
	401				450
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----c-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----a-----	-----t-----	-----
Serotype VII	-----	-----	-----a-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----g-----	-----
Serosubtype Ia-1	-----	-----	-----	-----g-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	GTTTATTATT	TTTTATTTT	AAAAACTCTT	TTACAACGAC	ACGACTTTCC
	451				500
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----c-----	-----	-----	-----
Serotype VII	-----	-----c-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----g-----	-----
Serosubtype Ia-1	-----	-----t-----	-----	-----g-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----a-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	TTTTTACTT	TTATTGCTAT	GAATTCGATT	TTATTATATC	TATTGAATTC
	501				550
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	ATTTTTAAAA	TATTATCGAA	AATATTCTTA	CGCTAAGTTT	TCACGAGATA
	551				600
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

4/25

Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	CCAAAGTTGT	TTTGATAACG	AATAAGGATT	CTTTATCAAA	AATGACCTTT
	601				650
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	AGGAATAAAT	ACGACCATAA	TTATATCGCT	GTCTGTATCT	TGGACTCCTC
	651				700
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	TGAAAAGGAT	TCTTATGATT	TGAACATAA	CTCGTTAAGG	ATAATAAACA
	701				750
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	AAGATGCTCT	TACTTCAGAG	TTAACCTGCT	TAACGTGTTG	TCAAGCTTTT
	751				800
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

5/25

Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	ATTACATAC	CCATTGAATT	ATTGGTAAA	TACCAATAC	AAGATATTAT
	901				850
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	TAATGACATT	GAAGCAATGG	GAGTGATTGT	CAATGTTAT	GTAGAGGCAC
	851				900
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	TTAGCTTTGA	TAATATAGGA	GAAAAGCGAA	TCCAACCTTT	TGAGGATAT
	901				950
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	AGTGTATTATA	CATATTCTAT	GAAATTCAT	AAATATAGTC	ACCTTATAGC
	951				1000
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

6/25

```

Serotype IV -----
Serotype V -----
Serosubtype Ia-2 -----
Consensus AAAACGATTT TTGGATATCA CGGGTGCTAT TATAGGTTTG CTCATATGTC
1001 1050
Serosubtype III-2 -----
Serotype VI -----
Serotype Ib -----
Serotype II/III-4 -----
Serotype VII -----
Serosubtype III-3 -----
Serosubtype Ia-1 -----
Serosubtype III-1 -----
Serotype IV -----
Serotype V -----
Serosubtype Ia-2 -----
Consensus GCATTGTGGC AATTTTCTA GTTCGCGAAA TCAGAAAAGA TGGTGGACCG
1051 1100
Serosubtype III-2 -----
Serotype VI -----
Serotype Ib -----
Serotype II/III-4 -----
Serotype VII -----
Serosubtype III-3 -----
Serosubtype Ia-1 -----
Serosubtype III-1 -----
Serotype IV -----
Serotype V -----
Serosubtype Ia-2 -----
Consensus GCTATCTTTT CTCAAAATAG AGTAGTTCGT AATGGTAGGA TTTTAGATT
cpsES2 1150
Serosubtype III-2 -----
Serotype VI -----
Serotype Ib -----
Serotype II/III-4 -----
Serotype VII -----
Serosubtype III-3 -----
Serosubtype Ia-1 -----
Serosubtype III-1 -----
Serotype IV -----
Serotype V -----
Serosubtype Ia-2 -----
Consensus CTATAAATTC AGATCAATGC GAGTAGATGC AGACAAAATT AAGAAAGATT
cpsEA3 1151 1200
Serosubtype III-2 -----
Serotype VI -----
Serotype Ib -----
Serotype II/III-4 -----
Serotype VII -----
Serosubtype III-3 -----
Serosubtype Ia-1 -----
Serosubtype III-1 -----
Serotype IV -----

```

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

7/25

Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	TATTAGTTCA	CAATCAAATG	ACAGGGCTAA	TGTTTAAGTT	AGACGATGAT
	1201				1250
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	CCTAGAATTA	CTAAAATAGG	AAAATTTATT	CGAAAAACAA	GCATAGATGA
	1251				1300
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	GTTGCCTCAA	TTCTATPATG	TTTTAAAGG	TGATATGAGT	TTAGTAGGAA
	1301				1350
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	CACGCCCTCC	CACAGTTGAT	GAATATGAAA	AGTATAATTC	AACGCAGAAG
	1351				1400
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

8/25

Serotype IV	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----
Consensus	CGAGCCTTA	GTTTAAAGCC	AGGAATCACT	GGTTTGTGGC
	1401			1450
Serosubtype III-2	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----
Consensus	TAGAAATAAT	ATTACTGATT	TTGATGAANT	CCTAAAGTGA
	1451			1500
Serosubtype III-2	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----
Consensus	ATATCAATGA	ATGGTCTATT	TGGTCAGATA	TTAAGNTTAT
	1501			1550
Serosubtype III-2	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----
Consensus	CTAAAGGTAG	TCTTACTTGG	GACAGGACCT	AACTAAAGGT
	1551			1600
Serosubtype III-2	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

9/25

Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	AGGAATATAA	TGAAAATTG	TCTGGTGGT	TCAAGTGGT	GTCACTAGC
			cpsF		
	1601				1650
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	t-----t	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	t-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	ACACTTGAAC	CITTTGAAAC	CCATTTGGGA	AAAAGAGAT	AGGTTTTGGG
	1651				1700
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	t-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	TAACTTTGA	TAAACAAGAT	GCTAGGAGTA	TTCTAAGAGA	AGAGATTGTA
	1701				1750
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	TATCATTGCT	TCTTCCAAC	AAACCGTAAT	GTCAAAAAC	TGCTAAAAA
	1751				1800
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

10/25

Serosubtype III-1	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----
Consensus	TACTATTCTA	GCTTTTAAGG	TCCTTAGAAA	AGAAAGACCA GATGTTATCA
	1801			1850
Serosubtype III-2	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----
Consensus	TATCATCTGG	TGCCGCTGTA	GCAGTACCAT	TCTTTTATAT TGGTAACTTA
	1851	cpsFS		1900
Serosubtype III-2	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----
Consensus	TTTGTTGTA	AGACCGTTA	TATAGAGGT	TTCGACAGGA TAGATAAACC
	1901	cpsFA		1950
Serosubtype III-2	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----
Consensus	AACTTTGACA	GGAAAATTAG	TGTATCCCTGT	AACAGATAAA TTPATTGTTC
	1951			2000
Serosubtype III-2	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

11/25

Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	AGTGGGAAGA	AATGAAAAAA	GTTTATCCTA	AGGCAATTAA	TTTAGGAGGA
	2001				2050
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----a-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	ATTTTTTAAT	GATTTTGTG	ACAGTGGGA	CACATGAACA	GCAGTTCAAC
	cpsF	I	cpsG		
	2051				2100
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----a-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	CGTCTTATTA	AAGAAGTTGA	TAGATFAAAA	GGGACAGGTG	CTATTGATCA
	2101				2150
Serosubtype III-2	-----	-----	-----	-----c-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----c-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	AGAAGTGTTT	ATTCAAAACGG	GTTACTCAGA	CTTTGAACCT	CAGAATTGTC
				cpsG8	
	2151				2200
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----g-----	-----g-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

12/25

Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	<u>AGTGGTCAAA</u>	<u>ATTTCTCTCA</u>	<u>TATGATGATA</u>	<u>TGAACTCTTA</u>	<u>CATGAAAGAA</u>
		cpsGA1		cpsGA2	
	2201		2226		
Serosubtype III-2	-----	-----	-----	-----	-----
Serotype VI	-----	-----	-----	-----	-----
Serotype Ib	-----	-----	c-----	-----	-----
Serotype II/III-4	-----	-----	-----	-----	-----
Serotype VII	-----	-----	-----	-----	-----
Serosubtype III-3	-----	-----	-----	-----	-----
Serosubtype Ia-1	-----	-----	-----	-----	-----
Serosubtype III-1	-----	-----	-----	-----	-----
Serotype IV	-----	-----	-----	-----	-----
Serotype V	-----	-----	-----	-----	-----
Serosubtype Ia-2	-----	-----	-----	-----	-----
Consensus	<u>GCTGAGATTG</u>	<u>TTATCACACA</u>	<u>TGGCGG</u>		
		cpsGA3			

Notes.

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

Serosubtype Ia-1: strain 090, GenBank accession number AF332893;
 Serosubtype Ia-2: strain NZRM 908(NCDC SS615), GenBank accession number AF332894;
 Serotype Ib: strain H36B, GenBank accession number AF332903;
 Serotype II/III-4: strain 18RS21, GenBank accession number AF332905;
 Serosubtype III-1: strain SG99/056, GenBank accession number AF332899;
 Serosubtype III-2: strain M781, GenBank accession number AF332896;
 Serosubtype III-3: strain NZRM 912 (NCDC SS620), GenBank accession number AF332897;
 III-4 (Subtype III-4): strain SG96/220, GenBank accession number AF363036;
 Serotype IV: strain 3139, GenBank accession number AF332908;
 Serotype V: strain CJB 111, GenBank accession number AF332910;
 Serotype VI: strain SS1214, GenBank accession number AF332901;
 Serotype VII: strain 7271, GenBank accession number AF332913.

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

13/25

Figure 2. Algorithm for GBS molecular serotype (MS) identification by PCR and sequencing.

A. GBS identification primer pairs.	
Amplification primers	Sag59/Sag190 or DSF2/DSR1
Result (band size)	196bp 254bp
Interpretation	GBS positive
B. GBS MS identification.	
1) MS-specific PCR.	
Primers	IacpHSI/cpsIA IbcpsS/IbcpsIA1 IIIcpsHS/cpsIA IVcpsHSI/IVcpsMA VcpsHSZ/VcpsMA VIcpsHSV/VcpHAI
Result (band size)	354bp 523bp 641bp 379bp 374bp 327bp
Interpretation	serotype Ia serotype Ib serotype III serotype IV serotype V serotype VI
2) Sequencing.	
Amplification primers	cpsES3/cpsGA1 (or cpsES3/cpsFA + cpsFS/cpsGA1)
Result (band size)	790bp (450bp + 423bp)
Sequencing primers	cpsGA (or cpsFA + cpsCA)
Interpretation	identification serotypes according to the sequence heterogeneity

SUBSTITUTE SHEET (RULE 26)

Figure 3. Multiple sequence alignments of the gene sequences of the cpsG-cpsH-cpsI/M for serotypes Ia, Ib, II, III, IV, V and VI (start and stop codons were highlighted).

```

1                               50
Serotype IV ATGATTTTTG TCACAGTGGG GACACATGAA CAGCAGTTCA ACCGCTTTAT
Serotype V  ATGATTTTTG TCACAGTGGG GACACATGAA CAGCAGTTCA ACCGCTTTAT
Serotype Ia ATGATTTTTG TCACAGTGGG GACACATGAA CAGCAGTTCA ACCGCTTTAT
Serotype Ib ATGATTTTTG TCACAGTAGG GACACATGAA CAGCAGTTCA ACCGCTTTAT
Serotype III ATGATTTTTG TCACAGTGGG GACACATGAA CAGCAGTTCA ACCGCTTTAT
Serotype VI ATGATTTTTG TCACAGTGGG GACACATGAA CAGCAGTTCA ACCGCTTTAT
Consensus *****_*** *****_*** *****_*** *****_***

cpsG
51                               100
Serotype IV TAAAGAAGTT GATAGATTAA AAGGGACAGA TGCTATTGAT CAAGAAGTGT
Serotype V  TAAAGAAGTT GATAGATTAA AAGGGACAGG TGCTATTGAT CAAGAAGTGT
Serotype Ia TAAAGAAGTT GATAGATTAA AAGGGACAGG TGCTATTGAT CAAGAAGTGT
Serotype Ib TAAAGAAGTT GATAGATTAA AAGGGACAGG TGCTATTGAT CAAGAAGTGT
Serotype III TAAAGAAGTT GATAGATTAA AAGGGACAGG TGCTATTGAT CAAGAAGTGT
Serotype VI TAAAGAAGTT GATAGATTAA AAGGGACAGG TGCTATTGAT CAAGAAGTGT
Consensus *****_*** *****_*** *****_*** *****_***

101                               150
Serotype IV TCATTCAAAC GGGTTACTCA GACTTTGAAC CTCAGAATTG TCAGTGGTCA
Serotype V  TCATTCAAAC GGGTTACTCA GACTTTGAAC CTCAGAATTG TCAGTGGTCA
Serotype Ia TCATTCAAAC GGGTTACTCA GACTTTGAAC CTCAGAATTG TCAGTGGTCA
Serotype Ib TCATTCAAAC GGGTTACTCA GACTTTGAAC CTCAGAATTG TCAGTGGTCA
Serotype III TCATTCAAAC GGGTTACTCA GACTTTGAAC CTCAGAATTG TCAGTGGTCA
Serotype VI TCATTCAAAC GGGTTACTCA GACTTTGAAC CTCAGAATTG TCAGTGGTCA
Consensus *****_*** *****_*** *****_*** *****_***

151                               200
Serotype IV AAATTCTCTC CATATGATGA TATGAACCTC TACATGAARG ARGCTGAGAT
Serotype V  AAATTCTCTC CATATGATGA TATGAACCTC TACATGAARG ARGCTGAGAT
Serotype Ia AAATTCTCTC CATATGATGA TATGAACCTC TACATGAARG ARGCTGAGAT
Serotype Ib AAATTCTCTC CATATGATGA TATGAACCTC TACATGAARG ARGCTGAGAT
Serotype III AAATTCTCTC CATATGATGA TATGAACCTC TACATGAARG ARGCTGAGAT
Serotype VI AAATTCTCTC CATATGATGA TATGAACCTC TACATGAARG ARGCTGAGAT
Consensus *****_*** *****_*** *****_*** *****_***

201                               250
Serotype IV TGTTATCACA CATGGCGGTC CAGCGACGTT TATGAATGCA GTTTCTAAAG
Serotype V  TGTTATCACA CATGGCGGTC CAGCGACGTT TATGAATGCA GTTTCTAAAG
Serotype Ia TGTTATCACA CATGGCGGTC CAGCGACGTT TATGAATGCA GTTTCTAAAG
Serotype Ib TGTTATCACA CACGGCGGTC CAGCACGTTT TATGAATGCA GTTTCTAAAG
Serotype III TGTTATCACA CATGGCGGTC CAGCGACGTT TATGTCAGTT ATTTCTTTAG
Serotype VI TGTTATCACA CATGGCGGTC CAGCGACGTT TATGTCAGTT ATTTCTTTAG
Consensus *****_***_*****_***_*****_***_*****_***_*****_***_

```

WO 03/025216

PCT/AU02/01281

15/25

```

251                                     300
Serotype IV  GGAAAAAAC TATTGTGGTT CCTAGACAAG AACAGTTTGG AGAGCAATGTG
Serotype V   GGAAAAAAC TATTGTGGTT CCTAGACAAG AACAGTTTGG AGAGCAATGTG
Serotype Ia  GGAAAAAAC TATTGTGGTT CCTAGACAAG AACAGTTTGG AGAGCAATGTG
Serotype Ib  GGAAAAAAC TATTGTGGTT CCTAGACAAG AACAGTTTGG AGAGCAATGTG
Serotype III GGAAATTACC AGTTGTTGTT CCTAGAGAAA AGCAGTTTGG TGAACATATC
Serotype VI  GGAAATTACC AGTTGTTGTT CCCAGGAGAA AGCAGTTTGG TGAACATATC
Consensus   *-*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-*

301                                     350
Serotype IV  AATAAICATC AGGIGGACIT TGTTAATAAG GTAAAAACAA TGTATAATTT
Serotype V   AATAAICATC AGGIGGACIT TGTTAATAAG GTAAAAACAA TGTATAATTT
Serotype Ia  AATAAICATC AGGIGGACIT TTTGAAAGAG TTATCTPTGA AAATTGAATTT
Serotype Ib  AATAAICATC AGGIGGACIT TTTGAAAGAG TTATCTPTGA AAATTGAATTT
Serotype III AATGATCATC AATAACAATT TTTAAAAAAA ATTGCCACCC TGTATCCCTT
Serotype VI  AATGATCATC AATAACAATT TTTAAAAATTG ATTGCCACCC TGTATCCCTT
Consensus   *-*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-*

351                                     400
Serotype IV  TGATATCCTT GTAGATATTC AAAGCTTACA AATGTAGTC TATGAGGGGA
Serotype V   TGATATCCTT GTAGATATTC AAAGCTTACA AATGTAGTC TATGAGGGGA
Serotype Ia  AGATTATATT TTGAATATCA GTGAATTAGA GAATATPATP AAGGAARAAA
Serotype Ib  AGATTATATT TTGAATATCA GTGAATTAGA GAATATPATP AAGGAARAAA
Serotype III GCCTTGGATT GAAGATGTAG ATGGACTTGC GGAAGCCTT. .GAAARAGGA
Serotype VI  GCCTTGGATT GAAGATGTAG ATGGACTTGC GGAAGCCTT. .GAAARAGGA
Consensus   -*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-*

401                                     450
Serotype IV  CGATGAATCG TCCGTTTTTA GAACTAACA GAAGTAATTT TATT.....
Serotype V   TGATGAATCG TCCGTTTTTA GAACTAACA GTAGTAATTT TATT.....
Serotype Ia  ATATATCTAC TAGTAAAGTA ATATCACAAA ACAATGATTT TTGTTCCTCT
Serotype Ib  ATATATCTAC TAGTAAAGTA ATATCACAAA ACAATGATTT TTGTTCCTCT
Serotype III ATATATCTAC AGAAAAATAT CAGGGAATA ATGATATGTT TTGT.....
Serotype VI  ATATATCTAC AGAAAAATAT CAGGGAATA ATGATATGTT TTGT.....
Consensus   -*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-*

451                                     500
Serotype IV  ..... GAAGAA TTTAAGGTAA TATTAAGGA
Serotype V   ..... GAAGAA TTTAAGGTAA TATTAAGGA
Serotype Ia  TTCAAAAATG AACATTTTCA TAACTATTTG AATAAATATA TTTTGTGGGA
Serotype Ib  TTCAAAAATG AAC..TTTCT TAACTATTTG AATAAATATA TTTTGTGGGA
Serotype III ..... CATTA AATTAGAAAA AATTATAGGT
Serotype VI  ..... CATTA AATTAGAAAA AATTATAGGT
Consensus   ----- *-*-*-*-* *-*-*-*-* *-*-*-*-*

501                                     550
Serotype IV  GTTGTGCGAT GAAA..... ATCAATAAAA AACTCTTTAT TTTATATTCG
Serotype V   GTTGTGCGAT GAAA..... ATCAATAAAA AACTCTTTAT TTTATATTCG
Serotype Ia  GAAAAAATTT GAAATTAACA TATCAATCCA AAGTATTTGT TAATAGGAGG
Serotype Ib  GAAAAAATTT GAAATTAAC. TATCAATCCA AAGTATTTGT TAATAGGAGG
Serotype III GAAATATGAG GAAAT....A TCTAGATTTA GATTATTCCT TATTTATATG
Serotype VI  GAAATATGAG GAAAT....A TCTAGATTTA GATTATTCCT TATTTATATG
Consensus   *-*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-* *-*-*-*-*

```

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

16/25

```

551                                     600
Serotype IV AATAATTTTAA GTTAATTTTT TTAATCACT AGCTTTAGGA GAGGGGAAC
Serotype V AATAATTTTAA GTTAATTTTT TTAATCACT GGGTTTAGGC GAGGGGAAC
Serotype Ia AATTTTCGCT TTAACCCCTAT TTTCAAAGCC AATGCAACTT TTGTTACTTT
Serotype Ib AATTTTCGCT TTAACCCCTAT TTTCAAAGCC AATGCAACTT TTGTTACTTT
Serotype III TCTTTGGGTA CTTATTTTAG TACCAAACCA ATGGTATCAG TTTTAAATTA
Serotype VI TCTTTGGGTA CTTATTTTAG TACCAAACCA ATGGTATCAG TTTTAAATTA
Consensus ---*---*---*---*---*---*---*---*---*---*---*---*---*---*---*
                                     cpsH
601                                     650
Serotype IV CAACCTACAA AATAGTGATG TTTGTTGCAA TCTTCTCTTG TGGAAATAAA
Serotype V CAGCCTACAA AATAGTGATG TTAGTTGCAA TTTTACTGTG TGGAAATAAA
Serotype Ia TAGCATTAAAT AGTTTTACTT ATTGTAGTA GTTAAAGAA AAAAATGAAA
Serotype Ib TAGCATTAAAT AGTTTTACTT ATTGTAGTA GTTAAAGAA AAAAATGAAA
Serotype III TTACCATTAT AGTCTATA TTACTTTGGA AGAGTGAGTT TAGAAT...A
Serotype VI TTACCATTAT AGTCTATA TTACTTTGGA AGAGTGAGTT TAGAAT...A
Consensus ---*---*---*---*---*---*---*---*---*---*---*---*---*---*---*

651                                     700
Serotype IV TTTTAA... ..TTAGATAG CCTTTATTTT GAAAGAAGAA AACTCGTAT
Serotype V TTTTAA... ..TTAGATAG CCTTTATTTT GAAAGAAGAA AACTCGTAT
Serotype Ia TTTTAAATA TGGCTGAAT TTTTTCATTT GTATTTATA TCATTTATTT
Serotype Ib TTTTAAATA TGGCTGAAT TTTTTCATTT GTATTTATA TCATTTATTT
Serotype III TCTATAAGCA ATTCTCAAT ACTATTTCTG CTTTGGTAT TTATTTATTT
Serotype VI TCTATAAGCA ATTCTCAAT ACTATTTCTG CTTTGGTAT TTATTTATTT
Consensus *---*---*---*---*---*---*---*---*---*---*---*---*---*---*---*

701                                     750
Serotype IV CACTTTTTTA TTATTTATG CGACCATTTT GAATTTATTC TTGTTTATA
Serotype V CACTTTTTTA TTATTTATG CGACCATTTT GAATTTATTC TTGTTTATA
Serotype Ia AACTTCATA TTGCTACAT CTTTGTTTAA AACTCCTGAT TTGATAGAA
Serotype Ib AGTATCATA GTATTAARTT CGTTATTTAG AACTCCGAA TTTCATAGAG
Serotype III ATTTGCATA CTCATTAGAG GTACTCAAGA GGATATAACG TTTCAGCGAT
Serotype VI ATTTGCATA CTCATTAGAG GTACTCAAGA GGATATAACG TTTCAGCGAT
Consensus ---*---*---*---*---*---*---*---*---*---*---*---*---*---*---*

751                                     800
Serotype IV AGGTTACTTT TATATTAA... ..C TTTAATTTTT
Serotype V AGGTTACTTT TATATTAA... ..C TTTAATTTTT
Serotype Ia TTTTAGCAGC TTTTAACTCG TTGATTATCG GTATAGTATC AGTGGGTTTG
Serotype Ib TCATTGCTGC ATTCAATTCA CTGGCAGTAG GGGTTGTTCT GTTATTTATTT
Serotype III TTATTGCTGA GCTATTAAAA CTAATTAGTA CAGGATATGC TTTATTTTTT
Serotype VI TTATTGCTGA GCTATTAAAA CTAATTAGTA CAGGATATGC TTTATTTTTT
Consensus ---*---*---*---*---*---*---*---*---*---*---*---*---*---*---*

801                                     850
Serotype IV TTTCTAGCAT TAAAGGATAT CTCCTAAAAA AARGCTTCT CTATAATAAT
Serotype V TTTCTAGCAT TAAAGGATAT CTCCTAAAAA AARGCTTCT CTATAATAAT
Serotype Ia AAACGGTGGT ATAAGGATAC AACTTTGGAG TTACATAAAA TATTAAGAAG
Serotype Ib TRCCATTACT ATAAGGATAC TAATTTGAA TTACATAAAA TGCTAAATC
Serotype III TATAATTAAT ATAGAAAGC TGATTTTAAT AGTTCAGTTG TAAGGAATGT
Serotype VI TATAATTAAT ATAGAAAGC TGATTTTAAT AGTTCAGTTG TAAGGAATGT
Consensus ---*---*---*---*---*---*---*---*---*---*---*---*---*---*---*

```


WO 03/025216

PCT/AU02/01281

18/25

	1151		1200
Serotype IV	GATATTATAT	AGTACTCTTA	TTTATACTTA
Serotype V	GGTATTATAT	CGTAATTTTA	TTTATTGTAC
Serotype Ia	CTGGAAATAGT	AGCTATPATTG	GCGCAGATGT
Serotype Ib	CTGGTATTAT	AGTTGTGCTA	CTACAGGTTA
Serotype III	TGGCTGTAT	GTTGGTTTTA	TTAGCATCGC
Serotype VI	TAATTTGTAT	GATAATTTTA	TTAGGTTATT
Consensus	-----*	-----*	-----*
	1201		1250
Serotype IV	AACCTGATAA	GGAAAATTTT	TATGATAGTT
Serotype V	AGCTTAAATA	AAAGASTATT	TATGAAATTA
Serotype Ia	GTTCGCGTAA	AGAGCAAAAC	GATAAGATTT
Serotype Ib	ATTCTAATAA	AAAGCAAAAC	GATAAGATTT
Serotype III	TTAAAACATA	ATTGCAAAAT	GACCAAGAAA
Serotype VI	ATAAATTTA	ACCTAAAAT	TACTAAAAAA
Consensus	-----*	-----*	-----*
	1251		1300
Serotype IV	CTTGTAGCA	TTTACTTTTC	TTTGTCTTAC
Serotype V	TTTATTAGTA	TTTACTTTTT	TGATCTTAC
Serotype Ia	TCTACTAGTA	ATAGTAATGA	TGTATATATP
Serotype Ib	ACTATATATA	CTATAGTGA	TATTACGTTT
Serotype III	GACTTTCTTA	TTTATCACCG	CTTGTTTTTC
Serotype VI	AGGATATA	TTATTATGG	TATGTTTTTC
Consensus	-----*	-----*	-----*
	1301		1350
Serotype IV	TTGTTCAAAA	ATTAGATAGC	CTTTGACAG
Serotype V	TTGTTCAAAA	ATTAGATGTT	CTTTAACAG
Serotype Ia	ATAATAGTAT	AATCAATTG	CGATCCGGGA
Serotype Ib	ATAATAGTAT	AATCAATTG	CGATCCGGGA
Serotype III	TTAAAAAAT	AATTAATGAC	AGAAACCAAA
Serotype VI	TCAATTATAT	AATACACTAT	AGATTTCAAA
Consensus	-----*	-----*	-----*
	1351		1400
Serotype IV	..GTTAAACT	ATGCTCATT	ACAGCITGTA
Serotype V	..ATTACACT	ATGCTCATT	ACAACITGTA
Serotype Ia	GTATATAAAG	ATACAGTAAA	CATCGTATA
Serotype Ib	TTGTACAAGG	ATACCGTACA	CTCAGTAAAT
Serotype III	GTTTACAAG	AACTATATAT	TGAAGTTCTA
Serotype VI	GCTTATTACG	AAAGTATAAG	AGCGATTTTA
Consensus	-----*	-----*	-----*
	1401		1450
Serotype IV	AAATAGTTTT	AAGGAG....A
Serotype V	AAATAGTTTT	AAGGA....A
Serotype Ia	AGAGGAGTT	AAAGGTTAT	GGTTAAATAG
Serotype Ib	AAAAGGTGTA	AAAGNATTCT	GGTTAAATAG
Serotype III	ACAGGGTATA	AGGA...TTC	CATCAAGTGA
Serotype VI	GCAAGGTATA	AGAG...TTC	CCTCCAGTGT
Consensus	-----*	-----*	-----*

SUBSTITUTE SHEET (RULE 26)

WO 03/025216

PCT/AU02/01281

20/25

```

1751                               1800
Serotype IV TTGTACCAGT ATATAATTCG AAACAMATTT TAATAGCTTG CGTTGATTCA
Serotype V TTGCTCTGAT ATGGCAGGAG GTAAGGAAGG AAAATGATAC CTAAGATTAT
Serotype Ia TTATACCTAT ATACAACCTCA GAAGCATACC TTAAGAAATG TGTGCAATCC
Serotype Ib TTATACCTAT ATACAACCTCG GAAGCATATC TTAAGAAATG CGTGCATCC
Serotype III TTATACCTAT ATACAACCTCA GAAGCATACC TTAAGAAATG TGTGCAATCC
Serotype VI TTGTTCCAGT TTATAATTCG GAGTTAGTGA TTGAGAAGTG TGTAGAATCT
Consensus **-----*-----*-----*-----*-----*-----*-----*

```

cpsI/M

```

1801                               1850
Serotype IV ATTAGAAAAC AAACAMTAA GAAATTTGAA AYTATTCTTG TTAATGATGG
Serotype V ACATTAATTGT TGGTTTGGAG GAAATCCCTT AGCAGATPAT TTAAGAAAT
Serotype Ia GTACTACAAC AGACTCATCC ATTGATAGAA GTTAACTATA TTGATGATGG
Serotype Ib GTCCTACAAC AGACTCATCC ATTGATAGAA GTTAACTATA TTGATGATGG
Serotype III GTACTACAAC AGACTCATCC ATTGATAGAA GTTAACTATA TTGATGATGG
Serotype VI TTGCTTCAAC AAACMTACCC AGAARTAGAA ATTTTATTA TAGATGATGG
Consensus -----*-----*-----*-----*-----*-----*-----*

```

```

1851                               1900
Serotype IV ATCAACAGAT GGTAGTAAAG AGTTAIGTGA GGAGATANGA AAATCAGATG
Serotype V ATATAAAAA. . . CTGGGAG AGACCAATGT CCGGATATG AAATATTGTA
Serotype Ia ATCCACTGAT AATAGTGGAG AAATTTGTGA TAATTTTCTC CAAGAAGATA
Serotype Ib ATCCACTGAT AATAGTGGAG AAATTTGTGA TAATTTTCTC CAAGAAGATA
Serotype III ATCCACTGAT AATAGTGGAG AAATTTGTGA TAATTTTCTC CAAGAAGATA
Serotype VI ATCTACAGAT AAAAGTAGTC ATATTTGTAA TAATTTTTTA AAAGGGATA
Consensus ***-----*-----*-----*-----*-----*-----*

```

```

1901                               1950
Serotype IV AAAGAAITAA GACATTTTAC AAACCAATG GAGGCAATC AAGCGCAAGG
Serotype V ATGGAATGAG CATAATTATG ATGTTAGTAA AAATGTTTTT ATGAGAGAGG
Serotype Ia ATCCGATACT TGTATTTTCA AAAAAAAAAATG GAGGGGTCTC TTCGGCAAGG
Serotype Ib ATCCGATACT TGTATTTTCA AAAAAAAAAATG GAGGGGTCTC TTCGGCAAGG
Serotype III ATCCGATACT TGTATTTTCA AAAAAAAAAATG GAGGGGTCTC TTCGGCAAGG
Serotype VI GTCCGCTAAA AGTCTATCAT AAATCAATG GAGGTGCATC ATCAGCAAGA
Consensus ---*-----*-----*-----*-----*-----*-----*

```

```

1951                               2000
Serotype IV AATTTAGGTA TTTTATACTC TACAGGAGAT TTGATTGGTT TTGTGACAG
Serotype V CATATACTAA GAAGAATTT. . . . . TGCT TATGTTCTG ACIATGCAAG
Serotype Ia AACCTAGGTC TAGATAAATC CACAGGAGAA TTCATAACAT TTGTGGATAG
Serotype Ib AACCTAGGTC TTGATAAATC CACAGGAGAA TTCATAACAT TTGTGGATAG
Serotype III AACCTAGGTC TAGATAAATC CACAGGAGAA TTCATAACAT TTGTGGATAG
Serotype VI AATGTGGGAC TTGAGATGGC AGAAGGTGAA TTTTAACTT TTGTAGATAG
Consensus -*-----*-----*-----*-----*-----*-----*

```

```

2001                               2050
Serotype IV CGACGATACA ATTGACCCTA AAATGTATGA AACGTTACTA AATATATATG
Serotype V ATTGGATATT ATTTTACTTT ATGGGGGGTT CTATCTAGAT ACTGATGTGG
Serotype Ia TGATGATTTT GTAGCACCGA ATATGATTTA AATAATGTTA AAAAATTTAA
Serotype Ib TGATGATTTT GTAGCACCGA ATATGATTTA AATAATGTTA AAAAATTTAA
Serotype III TGATGATTTT GTAGCACCGA ATATGATTTA AATAATGTTA AAAAATTTAA
Serotype VI CGATGATTTT GTGCACCTAA ATATGATTTA AATTATGCTG AATAATTTGT
Consensus -----*-----*-----*-----*-----*-----*

```


WO 03/025216

PCT/AU02/01281

23/25

```

                2651                                2700
Serotype IV  .....
Serotype V   .....
Serotype Ia  TGTTCACCTA AACTATATGT TATGGCATAT AGAAGATTCA AAACAAGTAGC
Serotype Ib  TGTTCACCCA TCTTGTATAT AAAATTATAT GACAGATTTC AAAAACAGTA
Serotype III TGTTCACCTA AACTATATGT TATGGCATAT AGAAGATTTC AAAAACAGTA
Serotype VI  TATTTCCGGT ACGTATATAT AAAGATGTAT AATAAATTTC AAAAGCAATA
Consensus   -----
                2701                                2728
Serotype IV  .....
Serotype V   .....
Serotype Ia  TGGAGAAATT GGGAAAGAGA AATTATAAA
Serotype Ib  A.....
Serotype III G.....
Serotype VI  A.....
Consensus   -----

```

Notes.

Serotype Ia: GenBank accession number AB028896;
 Serotype Ib: GenBank accession number AB050723;
 Serotype III: GenBank accession number AF163833;
 Serotype IV: GenBank accession number AF355776;
 Serotype V: GenBank accession number AF349539;
 Serotype VI: GenBank accession number AF337958.

Figure 4. Two sites (*) of sequence heterogeneity between *alp2* (AF208158, upper lines) and *alp3* (AF291065, lower lines) used to distinguish them (relevant primers are shown).

```

251 AAGGTAAATCCTAATATTTTTGAAAGATCAATAGTTGCTGCATCTACAATT 300
|||||
531 AAGGTAAATCCTAATATTTTTGAAAGATCAATAGTTGCTGCATCTACAATT 580
      bcaS1
301 CCRGGGAGTGCAGCGACCTTAAATACARGCATCACTAAAATATACAAAA 350
|||||
581 CCRGGGAGTGCAGCGACCTTAAATACARGCATCACTAAAATATACAAAA 630
      bcaS2
351 CGGAAACGCTTACATAGATTTATATGATGTAAGGAATGGATTGATGATC 400
|||||*|||||
631 CGGAAATGCTTACATAGATTTATATGATGTAAGGAATGGATTGATCGATC 680
      bcaS
401 CTCAAAACCTCATTTGATTAATCCATCAAGCTATTTCAGCAATATATAT 450
|||||
681 CTCAAAACCTCATTTGATTAATCCATCAAGCTATTTCAGCAATATATAT 730
      ba1S
451 ATCAAAACAGGTGCTAAATATATAGTAATCCGAGTGAATTAACAACAAC 500
|||||
731 ATCAAAACAGGTGCTAAATATATAGTAATCCGAGTGAATTAACAACAAC 780
      ba1A
501 TGGTTCAGCAACTATTACTTTTAAATACTTGATGAACTGGAAATCCAC 550
|||||
781 TGGTTCAGCAACTATTACTTTTAAATACTTGATGAACTGGAAATCCAC 830
      ba1A
551 ATAAAAAGCTGATGGACAATTCATATAGTTAGTCTGAATTTAACTATA 600
|||||
831 ATAAAAAGCTGATGGACAATTCATATAGTTAGTCTGAATTTAACTATA 880
      ba1A
601 TATGATTCACAGCTTTAAGAAATAGGATAGATGAAGTAATAAATAATGC 650
|||||
881 TATGATTCACAGCTTTAAGAAATAGGATAGATGAAGTAATAAATAATGC 930
      ba1A
651 AAATGATCCTAAGTGGAGTGAATGGAGTCTGATGAAGTCTTAACTGGAT 700
|||||
931 AAATGATCCTAAGTGGAGTGAATGGAGTCTGATGAAGTCTTAACTGGAT 980
      ba1A

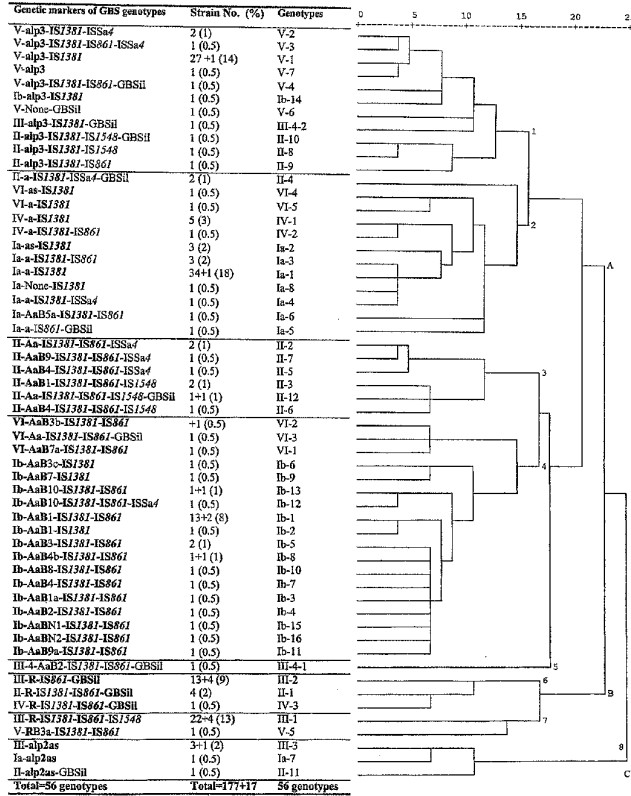
```

WO 03/025216

25/25

PCT/AU02/01281

Figure 5



【 国際調査報告 】

INTERNATIONAL SEARCH REPORT		International application No. PCT/AU02/01281
A. CLASSIFICATION OF SUBJECT MATTER		
Int. Cl. ⁷ : C12Q 1/68, 1:46		
According to International Patent Classification (IPC) or to both national classification and IPC		
B. FIELDS SEARCHED		
Minimum documentation searched (classification system followed by classification symbols) SEE ELECTRONIC DATABASE		
Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched SEE ELECTRONIC DATABASE		
Electronic data base consulted during the international search (name of data base and, where practicable, search terms used) WPIDS: CA: MEDLINE. KEYWORDS: Streptococcus, Streptococcus agalactiae, cps, capsular polysaccharide, capsular antigen, sero?, rib, alp2, alp3, surface antigen, surface protein, mobile genetic element, transposon.		
C. DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
P, X	Kong et al. "Serotype Identification of Group B Streptococci by PCR and Sequencing" Journal of Clinical Microbiology (2002) Vol 40 (1): pages 216-226 (see the whole document)	1-12, 23-28 31-35
P, X	Kong et al. "Molecular Profiles of Group B Streptococcal Surface protein Antigen Genes: Relationship to Molecular Serotypes" Journal of Clinical Microbiology (2002) Vol 40 (2): pages 620-626 (see the whole document)	13-17, 29-35
X	Lachenaier et al. "Mosaicism in the Alpha-Like Protein Genes of Group B Streptococci" Proc Natl Acad Sci USA. (2000) Vol 97 (17): pages 9630-9635 (see the whole document) EMBL DATABASE AF208158, ACCESSION NUMBER AF208158 23 August 2000	13-16, 29
<input checked="" type="checkbox"/> Further documents are listed in the continuation of Box C <input type="checkbox"/> See patent family annex		
* Special categories of cited documents:		
"A" document defining the general state of the art which is not considered to be of particular relevance	"T" later document published after the international filing date or priority date and not in conflict with the application, but cited to understand the principle or theory underlying the invention	
"E" earlier application or patent but published on or after the international filing date	"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone	
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)	"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art	
"O" document referring to an oral disclosure, use, exhibition or other means	"&" document member of the same patent family	
"P" document published prior to the international filing date but later than the priority date claimed		
Date of the actual completion of the international search 28 November 2002	Date of mailing of the international search report 09 DEC 2002	
Name and mailing address of the ISA/AU AUSTRALIAN PATENT OFFICE PO BOX 200, WODEN ACT 2606, AUSTRALIA E-mail address: pct@ipaustralia.gov.au Facsimile No. (02) 6285 3929	Authorized officer TERRY MOORE Telephone No : (02) 6283 2632	

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU02/01281

C (Continuation). DOCUMENTS CONSIDERED TO BE RELEVANT		
Category*	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
X	Tamura et al. "Analysis of Restriction Fragment Length Polymorphisms of the Insertion Sequence ISI381 in Group B Streptococci" The Journal of Infectious Diseases (2000) Vol 181: pages 364-368 (see the whole document)	18-20
X	Yamamoto S et al. "Molecular characterization of type-specific capsular polysaccharide biosynthesis genes of Streptococcus agalactiae type Ia" Journal of Bacteriology (1999) Vol 181: pages: 5176-5184 EMBL DATABASE ENTRY AB028896.2, ACCESSION NUMBER AB028896 16 July 1999	23
X	Miyake K et al. "CpsJ of Streptococcus agalactiae type Ib shows beta-1,3-galactosyltransferase activity" Unpublished. EMBL DATABASE ENTRY AB050723.1, ACCESSION NUMBER AB050723 5 February 2001	27

INTERNATIONAL SEARCH REPORT

International application No.
PCT/AU02/01281

Box I	Observations where certain claims were found unsearchable (Continuation of item 2 of first sheet)
This international search report has not been established in respect of certain claims under Article 17(2)(a) for the following reasons:	
1.	<input type="checkbox"/> Claims Nos : because they relate to subject matter not required to be searched by this Authority, namely:
2.	<input type="checkbox"/> Claims Nos : because they relate to parts of the international application that do not comply with the prescribed requirements to such an extent that no meaningful international search can be carried out, specifically:
3.	<input type="checkbox"/> Claims Nos : because they are dependent claims and are not drafted in accordance with the second and third sentences of Rule 6.4(a)
Box II	Observations where unity of invention is lacking (Continuation of item 3 of first sheet)
This International Searching Authority found multiple inventions in this international application, as follows:	
Invention I: Claims 1 to 12, 23 to 28 completely and claims 17, 22, and 31 to 35 partially.	
Invention II: Claims 13 to 16, 29 and 30 completely and claims 17, 22, and 31 to 35 partially.	
Invention III: Claims 18 to 21 completely and claim 22 partially. (see supplemental sheet)	
1.	<input type="checkbox"/> As all required additional search fees were timely paid by the applicant, this international search report covers all searchable claims
2.	<input checked="" type="checkbox"/> As all searchable claims could be searched without effort justifying an additional fee, this Authority did not invite payment of any additional fee.
3.	<input type="checkbox"/> As only some of the required additional search fees were timely paid by the applicant, this international search report covers only those claims for which fees were paid, specifically claims Nos.:
4.	<input type="checkbox"/> No required additional search fees were timely paid by the applicant. Consequently, this international search report is restricted to the invention first mentioned in the claims; it is covered by claims Nos.:
Remark on Protest <input type="checkbox"/> The additional search fees were accompanied by the applicant's protest. <input type="checkbox"/> No protest accompanied the payment of additional search fees.	

INTERNATIONAL SEARCH REPORT

International application No.

PCT/AU02/01281

Supplemental Box

(To be used when the space in any of Boxes I to VIII is not sufficient)

Continuation of Box No: II

The international application does not comply with the requirements of unity of invention because it does not relate to one invention or to a group of inventions so linked as to form a single general inventive concept. In coming to this conclusion the International Searching Authority has found that there are different inventions as follows:

Invention I relates to primers, compositions and methods for typing group B streptococcus based on sequence analysis of capsular proteins *cpsD*, *cpsE*, *cpsF*, *cpsG* and/or *cpsI/M*.

Invention II relates to primers, compositions and methods for typing group B streptococcus based on determining the presence or absence of one or more of the surface protein genes, *rib*, *alp2* and *alp3*.

Invention III relates to primers and methods for typing group B streptococcus based on determining the presence or absence of one or more mobile genetic elements selected from *IS861*, *IS1548*, *IS1381*, *ISSa4* and *GBSi1*.

The technical feature common to all three inventions is considered to be methods of typing group B streptococcus using nucleic acid primers. However this feature is obvious in view of the disclosure of Lachenauer et al. which prefigures methods of typing based on mosaicism in the *alp2* and *alp3* genes in GBS strains V and VIII. Therefore methods of typing group B streptococcus can not be regarded as a unifying technical feature.

フロントページの続き

(51) Int.Cl.⁷ F I テーマコード(参考)
C 1 2 N 15/00 F

(81) 指定国 AP(GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZM, ZW), EA(AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), EP(AT, BE, BG, CH, CY, CZ, DE, DK, EE, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, SK, TR), OA(BF, BJ, CF, CG, CI, CM, GA, GN, GQ, GW, ML, MR, NE, SN, TD, TG), AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EC, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, N O, NZ, OM, PH, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TN, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZM, ZW

(特許庁注：以下のものは登録商標)

テフロン

(72) 発明者 ファンロング コング

オーストラリア国 ニュー サウス ウェールズ州 ウェストミード ブリッジ ロード ピラ
23 / 93

(72) 発明者 ギルバート グウェンドリン

オーストラリア国 ニュー サウス ウェールズ州 リバービュー コーヨン ロード 27

Fターム(参考) 4B024 AA13 CA01 CA11 HA12

4B063 QA01 QA18 QQ06 QQ42 QQ52 QR08 QR42 QR55 QR62 QR82

QS25 QS34 QS36 QX02

专利名称(译)	B群連鎖球菌の分子的分类法		
公开(公告)号	JP2005502381A	公开(公告)日	2005-01-27
申请号	JP2003529988	申请日	2002-09-18
申请(专利权)人(译)	西悉尼地区卫生服务		
[标]发明人	ファンロングコング ギルバートグウェンドリン		
发明人	ファンロング コング ギルバート グウェンドリン		
IPC分类号	G01N33/53 C12N15/09 C12Q1/68 C12Q1/689 G01N33/569 G01N37/00		
CPC分类号	C12Q1/689		
FI分类号	C12N15/00.ZNA.A C12Q1/68.A G01N33/53.M G01N33/569.C G01N37/00.102 C12N15/00.F		
F-TERM分类号	4B024/AA13 4B024/CA01 4B024/CA11 4B024/HA12 4B063/QA01 4B063/QA18 4B063/QQ06 4B063/QQ42 4B063/QQ52 4B063/QR08 4B063/QR42 4B063/QR55 4B063/QR62 4B063/QR82 4B063/QS25 4B063/QS34 4B063/QS36 4B063/QX02		
代理人(译)	清水初衷		
优先权	2001PR7749 2001-09-19 AU		
其他公开文献	JP2005502381A5		
外部链接	Espacenet		

摘要(译)

提供了用于B组链球菌的分子分类的方法和用于这种方法的多核苷酸。

(34) J P 2005- 502381 A

研究菌株番号	供給源	血清型	MS/血清型	GenBank アクセッション番号
基準パネル 1¹				
O90	Channing	Ia	Ia	AF332893
H36B	Channing	Ib	Ib	AF332903
18RS21	Channing	II	II	AF332905
M731	Channing	III	III-2 ³	AF332895
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	
基準パネル 2²				
NZRM 908 (NCDC SS615)	ESR	Ia	Ia	AF332894
NZRM 909 (NCDC SS618)	ESR	Ib	Ib	AF332904
NZRM 910 (NCDC SS700)	ESR	Ic	Ia	AF332914
NZRM 911 (NCDC SS619)	ESR	II	II	AF332906
NZRM 912 (NCDC SS620)	ESR	III	III-3 ³	AF332897
NZRM 2217 (Prague 25/60)	ESR	分類不可能 (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

注釈

1. 基準パネル1: Lawrence Paoletti 博士 (Channing Laboratory, Boston, USA)