Identification of Streptococcus agalactiae isolates from various phylogenetic lineages by matrix-assisted laser desorption ionization-time of flight mass spectrometry

Marie-Frédérique Lartigue, Geneviève Héry-Arnaud, Eve Haguenoer, Anne-Sophie Domelier, Pierre-Olivier Schmit, Nathalie van Der Mee-Marquet, Philippe Lanotte, Laurent Mereghetti, Markus Kostrzewa, Roland Quentin

JCM00175-09 - Revised Manuscript.

Identification of *Streptococcus agalactiae* isolates from various phylogenetic lineages by Matrix-Assisted Laser Desorption/Ionization -Time of Flight Mass Spectrometry

Marie-Frédérique Lartigue[1,2], Geneviève Héry-Arnaud[1], Eve Haguenauer[1], Anne-Sophie Domelier[1,2], Pierre-Olivier Schmit[3], Nathalie van der Mee-Marquet[1,2], Philippe Lanotte[1], Laurent Mereghetti[1], Markus Kostrzewa[4], and Roland Quentin[1, 2].

[1] Université François-Rabelais de Tours, UFR de Médecine, EA 3854 « Bactéries et risque materno-foetal », Institut Fédératif de Recherche 136 « Agents Transmissibles et Infectiologie », 37032 Tours Cedex, France

[2] Service de Bactériologie et Hygiène Hospitalière, Hôpital Trousseau, CHRU de Tours, 37044 Tours, France.

[3] Bruker Daltonique, 34 rue de l'Industrie, 67166 Wissembourg Cedex, France

[4] Bruker Daltonik GmbH, Permoser Str. 15, D-04318 Leipzig, Germany

Corresponding author. Mailing address :

Marie-Frédérique Lartigue

Service de Bactériologie-Hygiène

CHRU Trousseau

37044 Tours cedex 9

France.

Phone : 33 2 47 47 81 13

Fax : 33 2 47 47 85 30

E-mail : lartigue@med.univ-tours.fr

Abstract

Variations in proteins related to bacterial diversity may affect species identification performed using matrix-assisted laser desorption/ionization-time of flight mass spectrometry (MALDI-TOF MS). Using this method, we identified 110 *Streptococcus agalactiae* isolates characterized by serotyping and MLST. Serotype III and ST23 strains expressed the widest variations in molecular weights of putative "species-identifying" biomarker ions. Recognition of the diversity of MALDI patterns observed in strains that represent all major intraspecies lineages assists in the constitution of an optimal reference database.

35 *Streptococcus agalactiae* is the main cause of neonatal infections and has
36 emerged as an increasingly frequent pathogen in non-pregnant humans (15). Several
37 studies have identified diversity in *S. agalactiae* species, first by serotyping, and then
38 extensively by multilocus enzyme electrophoresis (MLEE) (12, 13). A large number of
39 different Sequence Types (STs) distributed over several major phylogenetic lineages or
40 clonal complexes (CC) have recently been identified by multilocus sequence typing
41 (MLST) (http://pubmlst.org/sagalactiae/). STs 17 and 19 account for the majority of
42 cases of *S. agalactiae* meningitis in infants (7). CCs 1, 12, 17, 19 and 23 are mostly
43 associated with infections in adults (7).

44 Matrix-assisted laser desorption/ionization-time of flight mass spectrometry
45 (MALDI-TOF MS) has emerged as a new technique for species identification. By
46 measuring the exact molecular masses of many peptides and small proteins, it is
47 possible to determine the species within a few minutes, whether the analysis is started
48 with whole cells, cell lysates, or crude bacterial extracts (4, 6, 9, 11). Nevertheless, wide
49 variations in protein expression have been reported, specifically according to the
50 distribution of strains in various phylogenetic lineages that make up many species. For
51 example, we have shown such variations in the expression of metabolic enzymes (12,
52 13), in catabolic functions (3), in the expression of surface-exposed bacterial proteins
53 involved in the adhesion and/or invasion of host cells (14). In addition, the diversity of
54 the rRNA gene region found in *S. agalactiae* species (2) may induce variations in the
55 expression of the ribosomal proteins detected by MALDI-TOF MS, a method based
56 mainly on the detection of ribosomal protein fractions of bacteria (1, 4, 11).

57 The aim of this study was therefore to determine whether variations in protein
58 expression related to the phylogenetic position of strains affect the results obtained by
59 MALDI-TOF MS when used to identify *S. agalactiae* isolates. One hundred and ten

strains were selected from an epidemiologically unrelated national collection (10, 13) isolated from the vagina, the anatomical site at which the genetic diversity of strains is the highest. As usually performed in medical laboratories, strains were plated on 5% sheep's blood agar (laboratoire bioMérieux, Marcy l'Etoile, France). Serotyping was performed with a Pastorex rapid latex agglutination test (Bio-Rad, Hercules, Calif.) and by a previously reported PCR serotype identification method (8). Forty-five isolates were from serotype III (40.9% of isolates), 17 from serotype II (15.5%), 16 from serotype V (14.5%), 15 from serotype Ia (13.6%), 7 from serotype Ib (6.4%), and 7 from serotype IV (6.4%). Three isolates were not typeable (2.7%). MLST analysis, carried out with the standard MLST scheme (7), identified 38 STs for the 110 strains (Fig. 1). The relationships between STs were defined by UPGMA analysis and represented as a tree generated from allelic profile data using Phylodendron (http://pubmlst.org/sagalactiae/) (Fig. 1). By adding together data from serotyping and MLST, 52 patterns were obtained that represent the wide diversity of the *S. agalactiae* population studied and the major STs and CCs (Fig. 1).

For MALDI-TOF-MS analysis, cell extracts were prepared using ten colonies for each preparation. Samples were prepared according to the microorganism profiling « ethanol/formic acid extraction » procedure as recommended by the manufacturer, with minor modifications as recently described (1). After drying and addition of a chemical matrix, the samples were analyzed by MALDI-TOF MS on a Bruker Ultraflex TOF/TOF III in positive linear mode (Bruker Daltonique, Wissembourg, France). The spectra were recorded as recently described (1). For each spectrum, 500 laser shots were collected and analyzed. For automated data analysis, raw spectra were processed using the MALDI Biotyper 1.1 software (Bruker Daltonique, France) with default settings. To identify bacteria, the peak lists generated were used for matches against the initial

MALDI Biotyper reference library (*S. agalactiae* ATCC 27956 THL and *S. agalactiae* DSM 2134) directly using the integrated pattern-matching algorithm of the software (1). In a typical analysis of *S. agalactiae* strains by MALDI-TOF MS, 70 prominent ion peaks were noted in the spectra in the region between 2,000 and 20,000 Da, the highest-intensity peaks being consistently in the range of 3,000 to 10,000 Da. The log(score) of the MALDI Biotyper pattern matching algorithm is calculated according to the log of the product of three factors: the matches of the unknown spectrum against the reference spectrum in the database (main spectrum), the reverse matches of the main spectrum with the unknown spectrum, and the correlation of relative intensities of the unknown spectrum and the main spectrum. The product has a maximum value of 1000, leading to a maximum log(score) of 3. Differences in the distribution of the strains in log (score) groups were tested by the Chi^2 test (STAT-ITCF software, Paris). A p value < 0.05 was considered to indicate statistical significance.

On this basis, the log (score) values obtained by MALDI-TOF MS correctly identified all 110 *S. agalactiae* isolates at the species level [log (score), ≥2.0]; 86 of the 110 (78%) being identified with excellent scores [log (score), ≥2.3] (Fig. 1, Table 1). Nevertheless, significant variations in the log (score) were observed according to serotypes and major STs. Indeed, excellent scores [log (score) ≥2.3] were obtained significantly less frequently (64%) for serotype III strains than for the other strains (73 to 100%) (Table 1) ($p = 0.0027$). Similarly, excellent scores [log (score), ≥2.3] were obtained less frequently for ST23 strains (43%) than for the other major STs (77 to 88%) (Table 1) (Chi^2 test, $p = 0.025$).

To test the reproducibility of the method, five strains representative of the five main STs (ST1, ST10, ST17, ST19, and ST23) were grown in quintuplicate and analyzed. The mass spectrum was measured five times for each replicate of each of the

five representative strains of main STs (125 measurements). The log(score) calculated using the mass spectra data allowed determination of the coefficient of variation, CV=δ/μ in which δ is the standard deviation and μ is the average. The CV for intra runs (5 measurements for each replicate) varied from 0.006 to 0.025 and the highest CV for inter runs (5 different replicas) was 0.027. Moreover, reproducibility tests were also performed by cultivation of bacteria on different culture media (5% sheep blood agar, CPS ID 2®, chocolate PVX agar (bioMérieux), Granada® (Biolys, Taluyers, France), and Mueller Hinton agar). A strain of GBS ST17 was grown in triplicate on each medium. All the log(score) were above 2.3. The low values of the intra runs and inter runs CV and the low variations in the scores obtained using various culture media indicate that the reproducibility of the method was high. Thus the variations found between populations of various serotypes or STs could not be related to the procedure used for preparing the samples.

This study confirmed variations in expression of proteins according to the distribution of strains in serotypes and in various phylogenetic lineages that make up *S. agalactiae* species. Strains of serotype III and of the phylogenetic lineage ST23 expressed greater variations in the molecular weights of putative "species-identifying" biomarker ions than strains of other serotypes and STs. The ability to produce an intense peak in the protein pattern correlates with the ionization efficiency, combined with the protein quantity used. The variations observed for serotype III strains may be explained by the greater genetic diversity of this serotype population compared to the other serotype populations, as shown in Figure 1 and as previously described (7, 12, 13). As shown previously, CC23 (that mainly contains ST23 strains) is genetically a highly divergent clone in the species (Fig. 1) (5, 7, 16). This characteristic may explain why ST23 strains express differences in the nature or quantities of proteins when compared

to other major ST populations that are genetically more closely related. These variations did not affect the ability of MALDI-TOF MS combined with the MALDI Biotyper software method to identify *S. agalactiae* at a species level; log (score) values were ≥2.0 for all strains. Nevertheless, to optimize the ability of MALDI-TOF MS to identify *S. agalactiae* strains whatever their phylogenetic origin, we selected five strains to constitute a new reference database on the basis of the variations observed for each lineage. These strains were retained because they generated the best representative spectra for each major intraspecies lineage (ST1, ST10, ST17, ST19, and ST23). The log (score) values obtained by using the new reference database identified 109 of the 110 *S. agalactiae* isolates (>99%) with a higher log (score) (≥2.3) (Fig. 1, Table 1).

In conclusion, genetic diversity between bacterial species affects the diversity of MALDI patterns observed when using MALDI-TOF MS. Consequently, the most effective performance for identification of bacterial strains at the species level was obtained by using a reference database designed, not by selecting one or two strains randomly, but by choosing them from the major phylogenetic lineages that represent the species studied.

Conceived and designed the experiments: MFL, RQ. Performed the experiments: MFL, GH, EH, ASD, NVDM, PL, LM (serotyping and MLST), MFL, POS (MALDI-TOF MS). Analyzed the data: MFL, MK, RQ. Wrote the paper: MFL, RQ.

Acknowledgments. This study was supported by the Ministère de l'Enseignement Supérieur et de la Recherche and by the Centre Hospitalier Universitaire de Tours.

159 **REFERENCES**

160

161 1. **Barbuddhe, S.B., Maier, T., Schwarz, G., Kostrzewa, M., Hof, H., Domann,**
162 **E., Chakraborty, T., and T. Hain**. 2008. Rapid Identification and Typing of
163 *Listeria* species using Matrix Assisted Laser Desorption Ionization-Time of
164 Flight Mass Spectrometry. Appl. Environ. Microbiol. **74**:5402-5407.
165 2. **Chatellier, S., Huet, H., Kenzi, S., Rosenau, A., Geslin, P., and R. Quentin.**
166 1996. Genetic diversity of rRNA operons of unrelated Streptococcus agalactiae
167 strains isolated from cerebrospinal fluid of neonates suffering from meningitis.
168 J. Clin. Microbiol. **34**:2741-2747.
169 3. **Domelier, A.S., van der Mee-Marquet, N., Grandet, A., Mereghetti, L.,**
170 **Rosenau, A., and R. Quentin.** 2006. Loss of catabolic function in
171 *Streptococcus agalactiae* strains and its association with neonatal meningitis. J.
172 Clin. Microbiol. **44**:3245-3250.
173 4. **Fenselau, C., and P.A. Demirev.** 2001. Characterization of intact
174 microorganisms by MALDI mass spectrometry. Mass. Spectrom. Rev. **20:**157-
175 171.
176 5. **Héry-Arnaud, G., Bruant, G., Lanotte, P., Brun, S., Picard, B., Rosenau, A.,**
177 **van der Mee-Marquet, N., Rainard, P., Quentin, R., and L. Mereghetti.**
178 2007. Mobile genetic elements provide evidence for a bovine origin of clonal
179 complex 17 of *Streptococcus agalactiae*. Appl. Environ. Microbiol. **73**:4668-
180 4672.
181 6. **Holland, R.D., Wilkes, J.G., Rafii, F., Sutherland, J.B., Persons, C.C.,**
182 **Voorhees, K.J., and J.O.J. Lay.** 1996. Rapid identification of intact whole
183 bacteria based on spectral patterns using matrix-assisted laser

desorption/ionization with time-of-flight mass spectrometry. Rapid. Commun. Mass. Spectrom. **10:**1227-1232.

7. **Jones, N., Bohnsack, J.F., Takahashi, S., Oliver, K.A., Chan, M.S., Kunst, F., Glaser, P., Rusniok, C., Crook, D.W., Harding, R.M., Bisharat, N., and B.G. Spratt.** 2003. Multilocus sequence typing system for group B *Streptococcus*. J. Clin. Microbiol. **41:**2530-2536.
8. **Kong, F., Gowan, S., Martin, D., James, G., and G.L. Gilbert.** 2002. Serotype identification of group B streptococci by PCR and sequencing. J. Clin. Microbiol. **40:** 216-226.
9. **Krishnamurthy, T., Ross, P.L., and U. Rajamani.** 1996. Detection of pathogenic and non-pathogenic bacteria by matrix-assisted laser desorption/ionization time-of-flight mass spectrometry. Rapid. Commun. Mass. Spectrom. **10**:883-888.
10. **Loulergue, J., Couhé, C., Grasmick, C., Laudat., P., and R. Quentin.** 2004. Sensibilité aux antibiotiques des souches de streptocoque du groupe B de portage vaginal isolées en France, 2003. B.E.H. **18**:69-70.
11. **Mellmann, A., Cloud, J., Maier, T., Keckevoet, U., Ramminger, I., Iwen, P., Dunn, J., Hall, G., Wilson, D., Lasala, P., Kostrzewa, M., and D. Harmsen.** 2008. Evaluation of matrix-assisted laser desorption ionization-time-of-flight mass spectrometry in comparison to 16S rRNA gene sequencing for species identification of nonfermenting bacteria. J. Clin. Microbiol. **46**:1946-1954.
12. **Musser, J.M., Mattingly, S.J., Quentin, R., Goudeau, A., and R.K. Selander**. 1989. Identification of a high-virulence clone of type III *Streptococcus agalactiae* (group B *Streptococcus*) causing invasive neonatal disease. Proc. Natl. Acad. Sci. U. S. A. **86:** 4731-4735.

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

14. **Rosenau, A., Martins, K., Amor, S., Gannier, F., Lanotte, P., van der Mee-Marquet, N., Mereghetti, L., and R. Quentin.** 2007. Evaluation of the ability of *Streptococcus agalactiae* strains isolated from genital and neonatal specimens to bind to human fibrinogen and correlation with characteristics of the *fbsA* and *fbs*B genes. Infect. Immun. **75**:1310-1317.

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

16. **Van der Mee-Marquet N, Fourny L, Arnault L, Domelier AS, Salloum M, Lartigue MF, Quentin R.** 2008. Molecular characterization of human-colonizing *Streptococcus agalactiae* strains isolated from throat, skin, anal margin, and genital body sites. J. Clin. Microbiol. 46:2906-2911

Figure Legend

Figure 1. MLST UPGMA tree of 110 *S. agalactiae* strains used to test MALDI-TOF MS for identification of *S. agalactiae* strains and log (score) obtained for each strain analyzed.

[a] The MALDI-TOF MS analysis calculates the log (score) that allows classification of the results of identification in four categories: i) a log (score) ≥2.3 indicates a highly probable level of species identification, ii) a log (score) ≤2.299 and ≥2.0 indicates a highly probable level of genus identification and probable species identification, iii) a log (score) ≤1.999 and ≥1.7 indicates a probable genus identification and iv) a log (score) ≤1.699 indicates non-reliable identification. Whatever the phylogenetic origin of the strain tested, no log (score) ≤1.999 was found.

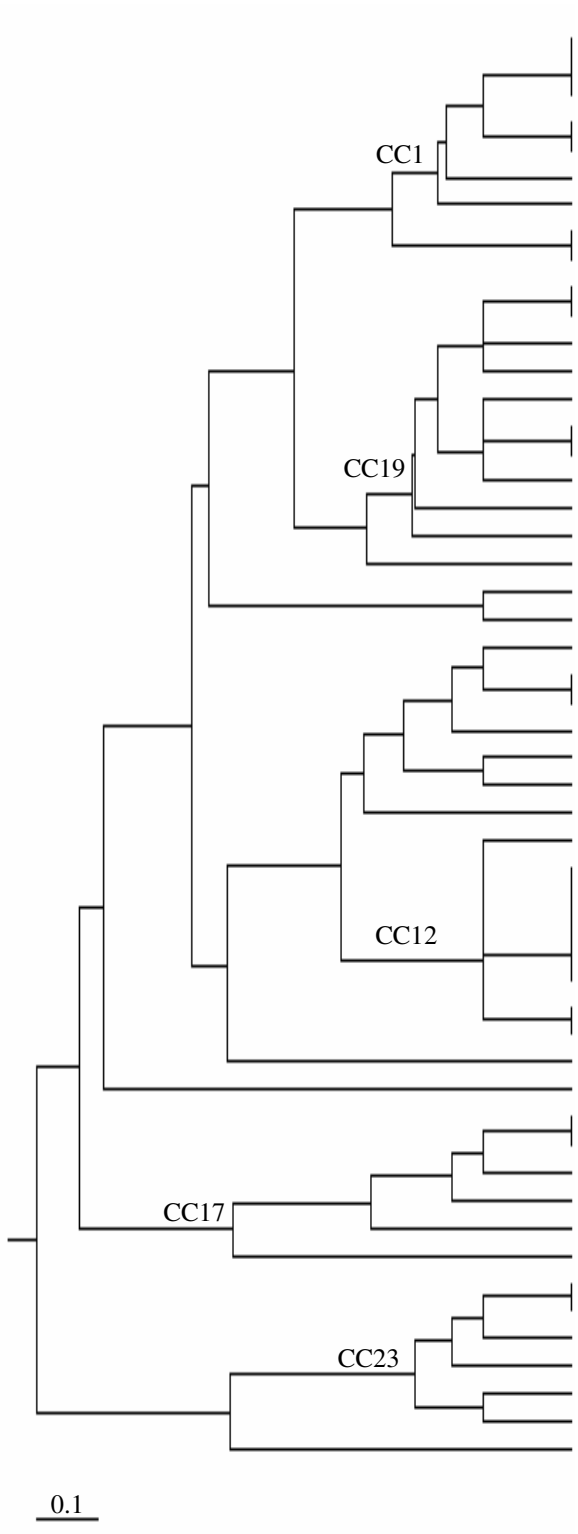
[b] Initial MALDI Biotyper reference library consisted of main spectra of *S. agalactiae* ATCC 27956 THL and *S. agalactiae* DSM 2134.

[c] New reference library consisted of the 5 newly created mass spectra for *S. agalactiae* strains representative of each major intraspecies lineage (ST1, ST10, ST17, ST19, and ST23).

Table 1. Identification scores produced by the pattern-matching algorithm against the initial MALDI Biotyper reference library (main spectra of *S. agalactiae* ATCC 27956 THL and *S. agalactiae* DSM 2134) and against a new reference library (mass spectra of five *S. agalactiae* strains selected to represent each major intraspecies lineage ST1, ST10, ST17, ST19, and ST2) for 110 *S. agalactiae* strains, according to serotypes and major intraspecies STs.

	Log(score)			
	Initial reference library		New reference library	
S. agalactiae strains (N)	≥2.3 N (%)	<2.3-≥2.0 N (%)	≥2.3 N (%)	<2.3-≥2.0 N (%)
All (110)	86 (78)	24 (22)	109 (99)	1 (1)
Serotype Ia (15)	11 (73)	4 (27)	15 (100)	0
Serotype Ib (7)	7 (100)	0	7 (100)	0
Serotype II (17)	15 (88)	2 (12)	17 (100)	0
Serotype III (45)	29 (64)	16 (36)	44 (98)	1 (2)
Serotype IV (7)	7 (100)	0	7 (100)	0
Serotype V (16)	15 (94)	1 (6)	16 (100)	0
ST1 (16)	13 (81)	3 (19)	16 (100)	0
ST10 (8)	7 (88)	1 (12)	8 (100)	0
ST17 (15)	12 (80)	3 (20)	15 (100)	0
ST19 (13)	10 (77)	3 (23)	13 (100)	0
ST23 (7)	3 (43)	4 (57)	7 (100)	0

MLST UPGMA tree



ST	Serotype	Log(score)[a]			
		Initial reference library[b]		New reference library[c]	
		≥2.3	<2.3-≥2.0	≥2.3	<2.3-≥2.0
	(N of strains)	N of strains	N of strains	N of strains	N of strains
ST1	Ia (1)	1		1	
ST1	III (3)	1	2	3	
ST1	V (12)	11	1	12	
ST2	II (2)	2		2	
ST2	IV (1)	1		1	
ST49	II (1)		1	1	
ST394	III (1)	1		1	
ST196	IV (5)	5		5	
ST196	nt (1)		1	1	
ST19	III (10)	7	3	10	
ST19	II (3)	3		3	
ST408	nt (1)	1		1	
ST395	III (1)		1	1	
ST27	III (2)	1	1	2	
ST28	II (4)	4		4	
ST28	III (1)	1		1	
ST392	III (1)	1		1	
ST393	III (1)	1		1	
ST107	III (1)		1	1	
ST106	III (1)		1		1
ST314	Ia (1)	1		1	
ST226	nt (1)	1		1	
ST6	III (1)		1	1	
ST7	Ia (2)	1	1	2	
ST7	IV (1)	1		1	
ST41	V (1)	1		1	
ST84	Ib (1)	1		1	
ST399	Ia (1)		1	1	
ST402	II (1)	1		1	
ST8	Ib (2)	2		2	
ST10	Ia (1)	1		1	
ST10	Ib (2)	2		2	
ST10	II (3)	2	1	3	
ST10	III (1)	1		1	
ST10	V (1)	1		1	
ST12	Ib (2)	2		2	
ST12	II (1)	1		1	
ST401	II (1)	1		1	
ST26	V (1)	1		1	
ST17	III (14)	11	3	14	
ST17	V (1)	1		1	
ST201	III (1)	1		1	
ST315	Ia (1)	1		1	
ST405	III (1)	1		1	
ST22	II (1)	1		1	
ST23	Ia (5)	3	2	5	
ST23	III (2)		2	2	
ST199	Ia (1)	1		1	
ST366	III (1)		1	1	
ST221	Ia (2)	2		2	
ST403	III (1)	1		1	
ST404	III (1)	1		1	
Total	(110)	86	24	109	1