Group B streptococci causing neonatal infections in Barcelona are a stable clonal population: 18-year surveillance.

E. R. Martins¹, A. Andreu², P. Correia¹, T. Juncosa³, J. Bosch⁴, M. Ramirez¹* and J. Melo-Cristino¹ on behalf of the Microbiologist Group for the Study of Vertical Transmission Infections, from the Catalan Society for Clinical Microbiology and Infectious Diseases

¹Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal. ²Servei de Microbiología, Hospital Vall d’Hebron, Universitat Autònoma de Barcelona. ³Servei de Microbiología, Hospital Sant Joan de Deu, Esplugues de Llobregat. ⁴Servei de Microbiología, Hospital Clínic i Provincial, Barcelona.

Running title: *Streptococcus agalactiae* neonatal invasive infections

*Corresponding author:
Mario Ramirez
Instituto de Microbiologia
Faculdade Medicina Lisboa
Av. Prof. Egas Moniz
PT 1649-028 Lisboa
Portugal
Email: ramirez@fm.ul.pt
Tel: +351- 21 799 9460
Fax: +351- 21 799 9459
Abstract

We analyzed 212 group B streptococci (GBS) from invasive infections in newborns in the Barcelona area between 1992-2009, with the aim of documenting changes in the prevalence of serotypes, antimicrobial resistance and genetic lineages and evaluating their association with either early-onset (EOD) or late-onset disease (LOD). Serotypes III (n=118) and Ia (n=47) together accounted for nearly 78% of the isolates. All isolates carried an alpha or alpha-like protein gene, with specific associations between genes and serotypes such as, Ib/bca, II/bca, III/rib and V/alp3, reflecting the presence of particular genetic lineages. Macrolide resistance (14.2%) was significantly associated to serotype V. Pulsed-field gel electrophoresis (PFGE) clustering was an excellent predictor of serotype and antibiotic resistance. The combination of PFGE and multilocus sequence typing revealed a large number of genetically distinct lineages. Still, specific lineages were dominant in our collection, particularly the serotype III/ST17/rib that had enhanced potential to cause LOD. Serotype Ia was concentrated on a single PFGE cluster composed of two lineages: ST23/eps and ST24/bca. The ST24/bca sub-lineage of serotype Ia, that is infrequently found elsewhere, may be emerging as an important cause of neonatal invasive infections in the Mediterranean region. In spite of the introduction of prophylaxis, resulting in a pronounced decline of EOD, the study revealed a remarkably stable clonal structure of GBS causing neonatal infections in Barcelona in a period of 18 years.
Introduction

*Streptococcus agalactiae*, or group B streptococcus (GBS) is well established as a leading cause of neonatal sepsis and meningitis (25, 45). In neonates, early-onset disease (EOD) is defined as occurring within the first 7 days and late-onset disease (LOD) from day 8 to 90 (12). While vertical transmission is commonly accepted to be the cause of EOD (24, 37, 51), the source of bacterial strains causing LOD is less well understood (45). In 1996 guidelines for prevention of GBS neonatal infections by antimicrobial prophylaxis were published in the United States (12). Whereas a mixed risk-based and screening-based approach was initially suggested in the guidelines, shortly after the universal screening of pregnant women for GBS vaginal colonization at 35 to 37 weeks of gestation and the administration of intrapartum antimicrobial prophylaxis to carriers was proposed (11). As a consequence, the incidence of EOD where these guidelines were followed has fallen significantly over the past decade, yet this strategy is raising concern as to the widespread use of intrapartum antimicrobials that might delay, rather than prevent, GBS disease onset (10, 16). In the area of Barcelona, during the implementation of the intrapartum antimicrobial prophylaxis guidelines the incidence of EOD declined by 86% from 1.92 cases per 1000 live births in 1994 to 0.26 in 2001 (p<0.001) (2), remaining at low levels since then (varying from 0.47 in 2007 to 0.18 in 2009). In the same area the incidence of LOD increased from 0.11 in 1996 to 0.81 per 1000 live births in 2009 but, in spite of the difference between these values, these changes did not reflect a significant trend. Furthermore, the increase in antimicrobial use due to intrapartum antibiotic prophylaxis can lead to the emergence of resistant bacteria (46), a concern that has been strengthened by the recent description of GBS strains showing reduced susceptibility to beta-lactams (28).
These considerations are driving the search for alternative prevention strategies. Studies evaluating the potential impact of vaccines in the management of GBS disease suggest that vaccination may provide additional benefits over antimicrobial prophylaxis, especially due to the expected reduction in LOD (49). Vaccine formulations currently on trial are based on the GBS capsular polysaccharides however; they are not expected to provide optimal coverage in different regions due to geographical variations in serotype distribution. In order to overcome serotype-specificity, whole-genome based approaches have been directed towards identifying protein antigens, holding promise as components of globally effective vaccines (26, 33).

In addition to the capsular polysaccharides, multiple virulence factors have been recognized and extensively characterized in the past decades. These virulence factors may be unevenly distributed within a particular serotype and may contribute significantly to the invasive potential of a particular lineage, independently of its capsular polysaccharide. Molecular epidemiology has been used to discriminate genetic lineages in order to probe for associations between specific GBS genotypes and disease. Most of these studies using multilocus sequence typing (MLST), have identified a lineage with enhanced invasive capacity expressing serotype III and defined by sequence type 17 (ST17) (6, 15, 31). Moreover, in a study of carriage and invasive isolates from Portugal, we found that serotype Ia presented an enhanced invasive disease potential and that it was particularly associated with EOD (36). In that study, serotype Ia was associated mostly with a single pulsed-field gel electrophoresis (PFGE) cluster and with two sequence types (ST23 and ST24), again pointing to the possible existence of particular genetic lineages with enhanced invasive disease potential.

We undertook the analysis of GBS isolates responsible for invasive infections in newborns in the Barcelona area from 1992 to 2009 with the aim of documenting
changes over this 18-year period and of testing associations with EOD and LOD. To this end, we have characterized the isolates regarding serotype and antimicrobial resistance pattern and identified the genetic lineages present by PFGE profiling, MLST and surface protein gene profiling.
Materials and Methods

From 1994, 8 hospitals located in the Barcelona’s metropolitan area, monitored all EOD cases and from 1996 all the LOD cases. In these hospitals GBS prevention policies were progressively implemented from 1994 onwards. Invasive disease was defined as the presence of GBS in a normally sterile fluid (blood and CSF). During 1994 - 2009, a total of 351,950 live infants were born in the 8 hospitals and EOD was diagnosed in 243 cases (189 born in the 8 hospitals and 54 referred from other hospitals). During 1996-2009, in a total of 315,576 live births, LOD was diagnosed in 131 infants.

Bacterial isolates. We characterized the 207 isolates available from the 374 GBS cases identified between 1994 and 2009 in the 8 Barcelona-area hospitals. Additionally, we also included 5 isolates recovered from GBS invasive disease in the same centers between 1992 and 1993. Only the first isolate of each case was considered. A total of 212 GBS isolates were characterized, 123 from EOD and 89 from LOD.

Serotyping, antimicrobial susceptibility testing and macrolide-resistance phenotype. Capsular serotyping was carried out by a latex agglutination assay with a GBS serotyping kit (Essum, Umeå, Sweden) according to the manufacturer’s instructions.

All GBS isolates were tested for susceptibility to erythromycin, clindamycin, tetracycline, chloramphenicol, levofloxacin and penicillin using the disk diffusion method according to the Clinical and Laboratory Standards Institute (CLSI) guidelines (13). The macrolide resistance phenotype was determined according to a double-disk test as previously described (38).
**PCR determination of macrolide-resistance genotype.** Total bacterial DNA was isolated by treatment of the cells with mutanolysin and boiling. A multiplex PCR reaction was performed to detect the presence of the *erm*(B), *erm*(A) [*erm*(TR) subclass] and *mef* genes, as described elsewhere (17).

**Pulsed-field gel profiling and MLST.** Total bacterial DNA of the strains was isolated, digested with SmaI, and separated by PFGE as previously described (36). Whenever a complete digestion with SmaI was not achieved, the isoschizomer Cfr9I was used (48). PFGE patterns were compared by using Bionumerics software (Applied Maths, Sint-Martens-Latem, Belgium) to create dendrograms by the unweighted pair-group method with arithmetic averages (UPGMA). The Dice similarity coefficient was used with optimization and position tolerance settings of 1.0 and 1.5, respectively. PFGE-based clusters were defined as isolates with $\geq 80\%$ relatedness on the dendrogram (36). MLST was performed by sequencing seven housekeeping genes as described previously (27) and sequence type (ST) identification was done by using the *S. agalactiae* MLST database (http://pubmlst.org/sagalactiae) and analyzed using the entire database and goeBURST (20). Alleles and sequence types not previously described were deposited at the *S. agalactiae* MLST database. Analysis of DNA sequences was done by using the Bionumerics software.

**Surface protein gene profile.** Total bacterial DNA was isolated by treatment of the cells with mutanolysin and boiling. A multiplex PCR assay was performed for direct identification of GBS alpha-like-protein genes, as described elsewhere (14). This assay allowed the determination of the following GBS surface protein genes directly by the analysis of the amplicon size: the alpha-C protein gene (*bca*); the epsilon protein gene
(eps); and the rib, alp2/3, and alp4 genes. A previously described assay was performed to differentiate protein antigen genes alp2 and alp3 (34).

**Typing concordance and statistics.** The Wallace (W) and adjusted Rand (AR) coefficients were calculated to determine the concordance between the different typing methods (9, 40, 47). AR provides a measurement of the overall concordance between the results of two methods, whereas W provides a directional measurement of clustering concordance between different typing methods, i.e., if the results of one typing method can predict the results of another method. The Simpson’s index of diversity (SID) was calculated to evaluate the diversity found among the isolates studied (9). Both these calculations were performed at the Comparing Partitions website, available at [www.comparingpartitions.info](http://www.comparingpartitions.info). The Fisher exact test was used to evaluate associations. Odds ratios (OR) with 95% Wald confidence intervals (CI_{95%}) (1), were calculated against all other serotypes or PFGE clusters and used to identify particular serotypes or PFGE clusters associated with certain characteristics, controlling for the false discovery rate (FDR) under or equal to 0.05 (3). The Spearman's non-parametric test was used to evaluate correlations (1).
Results

Capsular serotyping. The results of serotyping the 212 invasive GBS isolates from neonates are summarized in Table 1. Serotypes III (n=118) and Ia (n=47) were the most frequent among the population, together accounting for 77.8% of the isolates. The serotypes were also differently distributed in EOD and LOD (p=0.0067, Fisher’s exact test). Serotypes III and Ia were found in 44% and 26%, respectively of the cases of EOD and in 72% and 17%, respectively of the cases of LOD. In fact, the number of serotype III isolates found in LOD (n=64) more than doubled the sum of all other serotypes (n=25), and this was the only serotype that showed a significant association with disease presentation (OR=2.980, CI<sub>95%</sub> 1.581 to 5.734, indicating an association with LOD). Although there were more isolates recovered from the CSF in cases of LOD (n=25) than in EOD (n=21), similarly to what was described previously (36), we did not find an association between serotype and the biological product from which the isolate was recovered.

PFGE cluster analysis and MLST. All isolates were analyzed by PFGE and 43 different profiles were identified, grouped into 18 different PFGE clusters (≥3 isolates), of which the major five accounted for nearly 60% of the isolates (figure 1). The remaining isolates (n=34) were included in minor PFGE groups (≤2 isolates) or had unique profiles. The dendrogram depicting the relationship between these PFGE clusters is shown in figure 2. The SID for the classification of the isolates in PFGE clusters was 0.894 (CI<sub>95%</sub> 0.867 to 0.921), indicating that the collection analyzed is genetically very diverse. Each PFGE cluster was almost exclusively composed of isolates of the same serotype (W=0.955, CI<sub>95%</sub> 0.910 to 1.000), indicating a very good predictive power of the PFGE-based genotypes over serotype. The converse was not so,
and inspection of figure 2 and table 2 reveals that each serotype is subdivided into several PFGE clusters, as expected from the existence of several genetic lineages sharing the same serotype.

To further identify the genetic lineages associated with each PFGE clone, all isolates were characterized by MLST. Thirteen novel alleles and thirteen novel STs (ST469-ST471, and ST542-ST551) were identified among the isolates studied. The detailed characterization of the clusters depicted in figures 1 and 2 is summarized in Table 2. The SID for the classification of the isolates according to their MLST-based sequence types was 0.805 (CI95% 0.758 to 0.852), also indicative of a genetically diverse collection; however less diverse when compared to the classification in PFGE profiles. Overall, the isolates sharing the same ST also shared the same serotype (W=0.978, CI95% 0.961 to 0.997), indicating that MLST is a good predictor of serotype, similarly to what was found for the PFGE profiles.

In agreement with these observations, serotype III isolates representing ST17 were mostly found in PFGE clusters A33 and D15. Likewise, most serotype Ia (presenting mainly with ST23, but also a significant proportion with ST24) and V (ST1) isolates grouped into particular clusters, H38 and Q11, respectively.

Interestingly, most of the new STs were identified in serotype III isolates. Whereas ST469, ST542, ST543 and ST550 are single-locus variants (SLVs) of ST17, ST470 is a double-locus variant (DLV) of ST17 (exhibiting two novel alleles). On the other hand, ST471 and ST547 are SLVs of ST19, a sequence type previously associated with colonization (27, 31, 36). ST19 was found to be poorly represented in our collection, with the largest PFGE cluster where this ST was found being J9. This was expected since the collection analyzed included only isolates having caused neonatal invasive infections.
Six of the newly described STs (ST470, ST544, ST546, ST548, ST549, and ST551) were found in isolates included in minor PFGE clusters or showing unique profiles (Table 2). The combination of rare PFGE profiles and STs suggests that the GBS populations causing neonatal invasive disease may reflect to a limited extent the larger genetic diversity found in the overall GBS population.

When testing for an association between the largest PFGE clusters (\(n \geq 10\)) and the timing of disease presentation, both the major clusters of serotype III (A\textsubscript{53} and D\textsubscript{15}) presented significant OR for an association with LOD (OR=2.202, CI\textsubscript{95%} 1.121 to 4.376 and OR= 6.182, CI\textsubscript{95%} 1.598 to 35.220, respectively) but only the later was significant for FDR (p=0.018). Similarly to what was found for the serotypes, none of these largest PFGE clusters was associated with recovery from blood or CSF.

As for sequence types, ST17 was the only genetic lineage defined by MLST that was overrepresented in LOD (OR=4.972, CI\textsubscript{95%} 2.659 to 9.488). The fact that from the two major clusters of serotype III/ST17 defined by PFGE (A\textsubscript{53} and D\textsubscript{15}) only the later is significantly associated to LOD suggests that PFGE could be discriminating between two ST17 lineages otherwise indistinguishable by their serotype, ST and surface protein, but that could have other significant differences. This hypothesis remains to be further studied.

**Antimicrobial susceptibility testing.** All isolates were susceptible to penicillin and to levofloxacin. The overall rate of erythromycin resistance was 14.2\% (\(n=30\)). Among erythromycin-resistant isolates, all displayed the constitutive MLS\textsubscript{B} phenotype (cMLS\textsubscript{B}), defined by cross-resistance to all macrolides, lincosamides and to streptogramin B. None of the erythromycin-resistant isolates carried the mef gene, the \textit{erm}(B) gene was present in 76.7\% (\(n=23\)) and the \textit{erm}(A) [\textit{erm}(TR) subclass] gene in
23.3% (n=7) of the isolates. Resistance to tetracycline was found in 89.2% (n=189) of the isolates and to chloramphenicol in 2.8% (n=6).

While resistance to tetracycline and chloramphenicol was not clustered in particular serotypes, the same was not true for resistance to erythromycin (p<0.001, Fisher’s exact test) (table 2). An analysis of individual serotypes revealed that serotypes Ia [n=2/47 (4.3%)] and III [n=11/118 (9.3%)] presented less erythromycin resistance than expected (OR=0.219, CI_{95%} 0.024 to 0.926 and OR=0.408, CI_{95%} 0.165 to 0.962, respectively) while serotypes II [n=5/11 (45.5%)] and V [n=9/14 (64.3%)] presented more erythromycin resistant isolates than expected (OR=5.788, CI_{95%} 1.297 to 24.706 and OR=14.806, CI_{95%} 4.024 to 61.874, respectively), but only serotype V was significant for FDR (p<0.001).

An analysis of the largest PFGE clusters (n≥10) revealed three clusters with significant associations with erythromycin resistance, clusters H_{38} (OR=0, CI_{95%} 0 to 0.528) and A_{53} (OR=0.087, CI_{95%} 0.002 to 0.552) with reduced resistance and cluster Q_{11} (OR=21.092, CI_{95%} 4.645 to 131.992) with higher resistance, all significant for FDR (p=0.009 for both H_{38} and A_{53} and p<0.001 for Q_{11}). These results were not surprising. PFGE cluster Q_{11} includes the majority of serotype V isolates, all presenting with ST1, a sequence type that was also significantly associated to erythromycin resistance (OR=12.589, CI_{95%} 3.303 to 53.599). On the other hand, PFGE cluster H_{38} includes the majority of serotype ST23/serotype Ia isolates, and cluster A_{53} represents the ST17/serotype III and related lineages. Both clusters were associated with erythromycin susceptibility, a result that had already been suggested by the serotype analysis but that had failed to reach statistical significance. However the MLST analysis revealed that both ST23 and ST17 were significantly associated to erythromycin susceptibility (OR=...
0, CI$_{95\%}$ 0 to 0.658 and OR=0.040, CI$_{95\%}$ 0.001 to 0.256, respectively), supporting the indication given by the PFGE cluster analysis.

Even though ST19 was poorly represented in our collection, in agreement with being associated with colonization (31), it was also overrepresented among erythromycin resistant isolates (OR=6.885, CI$_{95\%}$ 2.082 to 22.557). This association was previously reported in Portugal (19).

**Surface protein gene profiling.** All isolates gave positive results for the presence of only one surface protein gene, with the exception of one isolate that we failed to amplify any of the genes tested. The surface protein gene rib was the most prevalent, followed by the eps, bca and alp3 genes, showing variable distributions across serotypes (Table 3). No alp2 or alp4 genes were found among the isolates. There was a very high correspondence between serotype and surface protein genes (AR=0.789, CI$_{95\%}$ 0.713 to 0.867). This was reflected in significant ORs of association of most serotypes (n≥10) with particular surface protein genes: serotype Ia and eps, Ib and II and bca, III and rib, V and alp3, all significant for FDR (p<0.001) (Table 3).

Surface protein genes were differently distributed across PFGE clusters (Figure 2), correlating to the proportion of serotypes within the clusters. A main exception was serotype Ia isolates that grouped in PFGE cluster H$_{38}$ regardless of presenting with surface protein gene eps or bca. In this cluster, an absolute association was found between ST and the surface protein gene, with all ST23 isolates carrying the eps gene and all ST24 isolates exclusively the bca gene, in support of our hypothesis that they constitute sub-lineages. In addition, we have also identified two isolates, one representing ST223 and the other the newly identified ST545, which are SLVs of ST23 but not of ST24, and both harboring the eps gene, suggesting diversification of ST23.
while retaining the characteristic surface protein gene.
Discussion

This study comprises a considerable number of invasive GBS isolates collected over an 18-year period in the Barcelona region. In spite of the overall large number of isolates, the number of yearly infections was low, preventing a detailed evaluation of the temporal changes of serotypes or PFGE clusters. Still, serotype III was present in all years and serotype Ia was absent in only two years of the study period. The remaining serotypes were represented by fewer isolates than the number of years studied, but there was an overall correlation between the number of isolates and the number of years in which they were found (r= 0.964, p=0.0028, Spearman’s test), indicating that no significant changes in the serotypes causing neonatal infections occurred in Barcelona in the 18 years studied. A similar analysis by PFGE cluster is complicated by their larger number, still each of the five largest clusters (n>10) was found in at least seven years and the two largest (A53 and H38) were found in all but two of the study years.

Taken together these data reveal a remarkably stable clonal structure of the GBS causing neonatal infections in Barcelona in a period of 18 years. This occurred in spite of the major epidemiological changes in GBS neonatal infections due to the introduction of prophylaxis in 1994 that resulted in a pronounced decline in EOD (2). We found substantial diversity among the GBS isolates causing neonatal invasive disease, not only in terms of capsular polysaccharides, but also in genetic lineages defined by both PFGE and MLST (figure 2 and Table 2). However, most isolates belonged to two serotypes and to a few STs and major PFGE clusters. The serotype distribution found in the population was similar to that described in some European countries, where capsular types Ia and III prevail among isolates causing neonatal invasive infections (5, 36, 41). In contrast to some US studies, where the prevalence of serotype V goes up to 30% (42), and to its recent increase reported in Scandinavia (4,
in Barcelona serotype V was much less frequently found, similarly to most studies across Europe.

Classification by PFGE and MLST defined groups of isolates frequently sharing the same serotype and surface protein gene, indicating that both techniques are identifying groups of closely related isolates. A strong correlation between the genes encoding surface proteins and the serotype was also found (Table 3), with most serotypes associated primarily with a single surface protein gene. The exception was serotype Ia that was associated with two proteins, although only one reached significance, with approximately a quarter of the isolates carrying the bca gene and the remaining three quarters carrying the eps gene. While our data is consistent with that described in Europe, contrasting data can be found in some studies from the U.S. that report the absence of the bca surface protein gene in all serotype Ia isolates causing neonatal invasive infections (32). We believe that this is a defining characteristic of a sub-lineage of serotype Ia, as discussed below. Also in contrast to the data reported here, another study from the U.S. found the bca surface protein gene as the most prevalent in serotype V isolates (50). Considering that most studies describe a strong association of serotype V with the alp3 gene (21, 29, 43), it is possible that the higher prevalence of serotype V in the U.S. when compared to Europe results from the expansion of a sub-lineage not found in Europe. This despite most serotype V isolates in either continent sharing the same sequence type (ST1).

As more collections are analyzed by their complement of surface protein genes, a broader understanding of their relationship with serotype may be obtained. However, our data suggests that it may be naïve to expect an absolute correlation with serotype, with additional typing methods, such as PFGE or MLST being potentially useful in
identifying distinct genetic lineages within each serotype that may also differ in their surface proteins.

Of note in this context was the high prevalence of ST24 found among serotype Ia isolates that concentrated mostly in PFGE cluster H38 together with the more widely disseminated ST23. ST24 is a DLV of ST23 that has rarely been found among large collections of GBS isolates characterized by MLST, with the exception of Portugal where in 2007 a significant prevalence of this ST was first reported (36). Also in 2007, a study from Italy showed most serotype Ia isolates grouping together in the same PFGE cluster, presenting either the bca or eps surface protein genes as described here, and the one representative isolate of this cluster characterized by MLST harbored the bca gene and was ST24 (21). Later, ST24 was found in 3 out of 52 serotype Ia invasive isolates from neonates in the US, and reported as a rare invasive clone (8). This suggests that among serotype Ia there are two different sub-lineages not distinguishable by PFGE, but discriminated by MLST and surface protein gene profiling. In agreement with this hypothesis is the absolute association between ST and surface protein gene in our collection, with all ST23 isolates carrying the eps gene and all ST24 isolates exclusively the bca gene. Previous results from our laboratory had provided strong support for the circulation of both sub-lineages in Portugal (35). Taken together these observations suggest that the presence of a particular alpha or alpha-like surface protein gene is rather a clonal property than a feature of the serotype. Interestingly, we have also found an additional ST24/bca isolate presenting with serotype III that may be the result of capsular switching (34).

This study, together with previous data from Portugal and Italy (21, 36) suggests that a particular sub-lineage of serotype Ia may be disseminated in the Mediterranean region.

Although we could not find a correlation between each of the surface protein defined
sub-lineages and isolate source (blood or CSF) or the timing of disease presentation (EOD or LOD), the ST24 sub-lineage may have other properties that could explain its success.

In spite of the recent description of penicillin non-susceptibility among GBS (28) and the intensive use of β-lactams in prophylaxis in Barcelona since 1994, all isolates were fully susceptible to penicillin. The macrolides can also be used in chemoprophylaxis and in contrast to penicillin, a significant proportion of erythromycin resistance was found (14.2%) in line with previous results from a multicentre study in Spain (13.7%) (22). The phenotypes and genotypes of macrolide resistant isolates from Barcelona were also similar to those previously identified in Spain (23). Similarly to other studies in different geographic regions, we found an association between macrolide resistance, serotype V (7, 53), and ST1 (44). Still, only 30% of erythromycin resistance was found in serotype V, while the remaining was dispersed in all other serotypes except serotype IV, a situation observed in a few previous studies (18, 30, 52). Resistance was also found to be dispersed in many genetic lineages as defined by PFGE and MLST, in agreement with the methylase genes being carried in mobile genetic elements. However, clonal expansion is also important in macrolide resistance; since PFGE cluster Q11, expressing serotype V and ST1 was found to be significantly associated with resistance. On the other hand, cluster A53 representing the highly virulent ST17 and associated lineages, was less resistant than expected, reflecting the general observation that isolates representing ST17 are rarely macrolide resistant. The reasons why this highly virulent and successful lineage is seldom resistant is unknown.

The characterization of the population of GBS causing invasive infections in neonates in the Barcelona region revealed the existence of a large number of genetically distinct lineages that were present over a significant time-span. The stability and dominance of a
few lineages that are responsible for the majority of infections in spite of continuous antibiotic and immune selective pressures, suggest that they are extremely well adapted to their particular niche. Although most of these lineages are widely disseminated worldwide, we have also identified seemingly regionally successful clones, raising the possibility of an ongoing selection and expansion of specific virulent GBS clones. Continuous surveillance will shed further light on these processes and will determine if these clones will further expand beyond their current geographical boundaries.
Acknowledgments

This work was partly supported by a grant from Fundação Calouste Gulbenkian, by Fundação para a Ciência e a Tecnologia (POCI/SAU-ESP/57646/2004) and an unrestricted grant from Glaxo Smithkline Portugal. E.R.M. was supported by a grant from Fundação para a Ciência e a Tecnologia (SFRH/BD/41761/2007). The Microbiologist Group for the Study of Vertical Transmission Infections: M Gimenez, Hospital Germans Trias i Pujol, Badalona, M Sierra, Hospital Barcelona-SCIAS, Barcelona, J Lite, Hospital Mutua Terrassa, Terrassa, I Sanfeliu, Coorporació Parc Taulí, Sabadell, F Sanchez, Hospital de Sant Pau, Barcelona, E Dopico, Laboratori Clinic l’Hospitalet, Hospitalet, C Guardià, Laboratori Clinic Barcelones Nord i Valles Oriental, Badalona.
Bibliography


Figure Legends

Figure 1. Representative SmaI macrorestriction profiles of each PFGE cluster.
Capital letters above the lanes correspond to cluster designations. The lane designated N shows an incomplete digestion with SmaI. The lane designated N* indicates the same strain in N but now digested with Cfr9I (see text). λ, Lambda Ladder PFG Marker (New England Biolabs, Beverly, MA).

Figure 2. Dendrogram of the PFGE profiles of 212 GBS isolates and distribution of the surface protein genes. The dendrogram was constructed using the unweighted pair group method with arithmetic mean (UPGMA) method. Dice coefficients (percentages) are indicated in the scale above the dendrogram. Each cluster (defined as a group of three or more isolates with a Dice coefficient ≥80%) is represented by a triangle proportional to the number of isolates included in the cluster. Clusters are designated by capital letters and a subscript number indicating the number of isolates included in the cluster. The surface protein genes found in each particular cluster are indicated. If more than one gene was found in strains of the same PFGE cluster, the number of isolates carrying each gene is indicated in parenthesis.
Table 1. Serotype distribution among invasive GBS isolates causing EOD and LOD

<table>
<thead>
<tr>
<th>Serotype</th>
<th>Blood</th>
<th>CSF</th>
<th>Total (%)</th>
<th>Blood</th>
<th>CSF</th>
<th>Total (%)</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td>Ia</td>
<td>27</td>
<td>5</td>
<td>32 (26)</td>
<td>11</td>
<td>4</td>
<td>15 (17)</td>
<td>47</td>
</tr>
<tr>
<td>Ib</td>
<td>8</td>
<td>0</td>
<td>8 (7)</td>
<td>4</td>
<td>1</td>
<td>5 (6)</td>
<td>13</td>
</tr>
<tr>
<td>II</td>
<td>9</td>
<td>1</td>
<td>10 (8)</td>
<td>1</td>
<td>0</td>
<td>1 (1)</td>
<td>11</td>
</tr>
<tr>
<td>III</td>
<td>41</td>
<td>13</td>
<td>54 (44)</td>
<td>46</td>
<td>18</td>
<td>64 (72)</td>
<td>118</td>
</tr>
<tr>
<td>IV</td>
<td>2</td>
<td>0</td>
<td>2 (2)</td>
<td>0</td>
<td>1</td>
<td>1 (1)</td>
<td>3</td>
</tr>
<tr>
<td>V</td>
<td>11</td>
<td>1</td>
<td>12 (10)</td>
<td>2</td>
<td>0</td>
<td>2 (2)</td>
<td>14</td>
</tr>
<tr>
<td>NT&lt;sup&gt;a&lt;/sup&gt;</td>
<td>4</td>
<td>1</td>
<td>5 (4)</td>
<td>0</td>
<td>1</td>
<td>1 (1)</td>
<td>6</td>
</tr>
<tr>
<td>Total</td>
<td>102</td>
<td>21</td>
<td>123 (100)</td>
<td>64</td>
<td>25</td>
<td>89 (100)</td>
<td>212</td>
</tr>
</tbody>
</table>

<sup>a</sup>NT, nontypeable
Table 2. Properties of the genetic lineages found among the 212 invasive isolates

<table>
<thead>
<tr>
<th>PFGE cluster⁴</th>
<th>Serotype</th>
<th>EOD/LOD</th>
<th>STs in PFGE cluster with the same serotype (n)⁵</th>
<th>No. of isolates not susceptible to⁶</th>
<th>Susceptible isolates (n)⁷</th>
</tr>
</thead>
<tbody>
<tr>
<td>A₃³</td>
<td>III</td>
<td>23/30</td>
<td>[ST17 (48), ST147 (1), ST148 (1), ST542 (1), ST543 (1), ST550 (1)]</td>
<td>1 0 51 2</td>
<td></td>
</tr>
<tr>
<td>B₂</td>
<td>III</td>
<td>2/1</td>
<td>[ST17 (2), ST180 (1)]</td>
<td>0 0 3 0</td>
<td></td>
</tr>
<tr>
<td>C₃</td>
<td>III</td>
<td>3/2</td>
<td>[ST17 (4), ST148 (1)]</td>
<td>0 0 5 0</td>
<td></td>
</tr>
<tr>
<td>D₁₃</td>
<td>III</td>
<td>3/12</td>
<td>ST17 (15)</td>
<td>0 0 15 0</td>
<td></td>
</tr>
<tr>
<td>E₄</td>
<td>III</td>
<td>2/4</td>
<td>[ST17 (4), ST469 (2)]</td>
<td>0 1 6 0</td>
<td></td>
</tr>
<tr>
<td>F₂</td>
<td>III</td>
<td>1/3</td>
<td>ST17 (4)</td>
<td>0 0 4 0</td>
<td></td>
</tr>
<tr>
<td>G₃</td>
<td>la</td>
<td>2/0</td>
<td>ST23 (2)</td>
<td>0 0 2 0</td>
<td></td>
</tr>
<tr>
<td>III</td>
<td>Ia</td>
<td>1/0</td>
<td>ST19 (1)</td>
<td>1 1 1 0</td>
<td></td>
</tr>
<tr>
<td>H₁₈</td>
<td>la</td>
<td>22/14</td>
<td>[ST23 (25), ST24 (9), ST223 (1), ST545 (1)]</td>
<td>0 1 34 3</td>
<td></td>
</tr>
<tr>
<td>I₃</td>
<td>III</td>
<td>0/2</td>
<td>ST17 (1), ST27 (1)</td>
<td>0 0 1 1</td>
<td></td>
</tr>
<tr>
<td>J₉</td>
<td>III</td>
<td>6/2</td>
<td>[ST19 (5), ST456 (1), ST471 (1), ST547 (1)]*</td>
<td>3⁷ 0 6 1</td>
<td></td>
</tr>
<tr>
<td>K₅</td>
<td>III</td>
<td>1/0</td>
<td>ST28 (1)*</td>
<td>0 0 1 0</td>
<td></td>
</tr>
<tr>
<td>L₈</td>
<td>II</td>
<td>3/0</td>
<td>[ST2 (1), ST9 (1), ST12 (1)]*</td>
<td>1 0 0 2</td>
<td></td>
</tr>
<tr>
<td>Ia</td>
<td>2/0</td>
<td>ST7 (2)*</td>
<td>2⁶ 0 1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>lb</td>
<td>0/1</td>
<td>ST1 (1)*</td>
<td>0 0 1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>M₄</td>
<td>la</td>
<td>3/1</td>
<td>[ST8 (1), ST9 (2), ST10 (1)]</td>
<td>0 0 2 2</td>
<td></td>
</tr>
<tr>
<td>N₄</td>
<td>III</td>
<td>2/1</td>
<td>[ST19 (2), ST27 (1)]*</td>
<td>0 0 2 0</td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>1/0</td>
<td>ST19 (1)*</td>
<td>0 0 2 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>O₃</td>
<td>lb</td>
<td>2/1</td>
<td>ST12 (3)</td>
<td>0 0 3 0</td>
<td></td>
</tr>
<tr>
<td>P₃</td>
<td>II</td>
<td>3/0</td>
<td>ST22 (3)</td>
<td>3⁷ 0 3 0</td>
<td></td>
</tr>
<tr>
<td>Q₅</td>
<td>V</td>
<td>9/2</td>
<td>ST1 (11)</td>
<td>8 0 11 0</td>
<td></td>
</tr>
<tr>
<td>R₃</td>
<td>la</td>
<td>2/0</td>
<td>ST23 (2)</td>
<td>0 0 2 0</td>
<td></td>
</tr>
<tr>
<td>NT²</td>
<td>1/0</td>
<td>ST19 (1)</td>
<td>0 0 0 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Other₅₄</td>
<td>III</td>
<td>11/4</td>
<td>[ST17 (4), ST19 (6), ST24 (1), ST106 (2), ST470 (1), ST546 (1), ST547 (1)]</td>
<td>6 3 14 1</td>
<td></td>
</tr>
<tr>
<td>lb</td>
<td>3/2</td>
<td>ST2 (1), ST9 (1), ST10 (1), ST12 (1), ST548 (1)</td>
<td>2 0 3 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>II</td>
<td>3/1</td>
<td>ST1 (1), ST12 (1), ST22 (1), ST544 (1)</td>
<td>1⁷ 0 3 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>2/1</td>
<td>ST2 (1), ST196 (1), ST549 (1)</td>
<td>0 0 2 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>NT²</td>
<td>3/0</td>
<td>ST19 (1), ST27 (1), ST130 (1)</td>
<td>1⁷ 0 1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Ia</td>
<td>1/1</td>
<td>ST7 (1), ST196 (1)</td>
<td>0 0 1 1</td>
<td></td>
<td></td>
</tr>
<tr>
<td>V</td>
<td>2/0</td>
<td>ST26 (1), ST551 (1)</td>
<td>1⁷ 0 1 0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td>123/89</td>
<td>30 6 189 19</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

⁴PFGE clusters are identified as indicated in figure 1. Clusters are designated by capital letters and a subscript number indicating the number of isolates included in the cluster.

⁵Brackets indicate STs that were grouped into the same PFGE cluster and belonged to the same clonal complex by goeBURST and expressed the same serotype. Asterisks indicate STs or groups of STs that were found in the same PFGE cluster and belonged to the same clonal complex by goeBURST but expressed different serotypes.

⁶ERY, erythromycin; CLI, clindamycin; CHL, chloramphenicol; TET, tetracycline

⁷All isolates were simultaneously resistant to erythromycin and clindamycin, presenting with the cMLSB phenotype associated to the erm(B) gene, except when indicated.

⁸Isolates susceptible to all antimicrobials tested (erythromycin, clindamycin, chloramphenicol and tetracycline)

⁹One isolate carried the erm(A) gene

¹⁰NT, nontypeable

¹¹Two isolates carried the erm(A) gene
Table 3. Distribution of genes encoding surface proteins across serotypes

<table>
<thead>
<tr>
<th>Serotype</th>
<th>Surface protein gene&lt;sup&gt;a&lt;/sup&gt;</th>
<th>Total</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>bca</td>
<td>eps</td>
</tr>
<tr>
<td>Ia</td>
<td>12</td>
<td>35</td>
</tr>
<tr>
<td>Ib</td>
<td>11</td>
<td>1</td>
</tr>
<tr>
<td>II</td>
<td>7</td>
<td>1</td>
</tr>
<tr>
<td>III</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>IV</td>
<td>1</td>
<td>2</td>
</tr>
<tr>
<td>V</td>
<td>2</td>
<td></td>
</tr>
<tr>
<td>NT&lt;sup&gt;b&lt;/sup&gt;</td>
<td>1</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>34</td>
<td>41</td>
</tr>
</tbody>
</table>

<sup>a</sup>Highlighted in bold whenever a significant correlation between surface protein gene and serotype was found (see text)
<sup>b</sup>NT, nontypeable
<sup>c</sup>one isolate failed to amplify any of the surface proteins tested