Invasiveness of pneumococcal serotypes and clones circulating in Portugal before widespread use of conjugate vaccines revealing heterogeneous behavior of clones expressing the same serotype

Raquel Sá-Leão¹*, Francisco Pinto²*, Sandra Aguiar³, Sónia Nunes¹, João A. Carriço³, Nelson Frazão¹, Natacha Gonçalves-Sousa¹, José Melo-Cristino³, Hermínia de Lencastre¹,⁴, and Mário Ramirez³

¹Instituto de Tecnologia Química e Biológica, Universidade Nova de Lisboa, Oeiras, Portugal; ²Centro de Química e Bioquímica, Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade de Lisboa, Lisboa, Portugal; ³Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisboa, Portugal; ⁴Laboratory of Microbiology, The Rockefeller University, New York, NY, USA.

Corresponding author: Raquel Sá-Leão

Mailing address: Instituto de Tecnologia Química e Biológica, Rua da Quinta Grande, 6, Oeiras, 2780-156 Oeiras, Portugal. Phone: (351) 21 446 9872. Fax: (351) 21 4428766.
E-mail: rsaleao@itqb.unl.pt

These authors contributed equally to this work

Running title: Invasiveness of pneumococcal capsules and clones

Key words: Streptococcus pneumoniae, pneumococcus, carriage, colonization, invasive disease, MLST, PFGE, serotype, capsule
ABSTRACT

To estimate the invasive disease potential of serotypes and clones circulating in Portugal before extensive use of the seven-valent pneumococcal conjugate vaccine, we analyzed 475 invasive isolates recovered from children and adults and 769 carriage isolates recovered from children between 2001 and 2003. Isolates were serotyped and genotyped by pulsed-field gel electrophoresis and a selection was also characterized by multilocus sequence typing. We have found that the diversity of serotypes and genotypes of pneumococci responsible for invasive infections and carriage is identical and that most carried clones could also be detected as causes of invasive disease. Their ability to do so, however, can vary substantially. Serotypes 1, 3, 4, 5, 7F, 8, 9N, 9L, 12B, 14, 18C, and 20 were found to have an enhanced propensity to cause invasive disease while serotypes 6A, 6B, 11A, 15B/C, 16F, 19F, 23F, 34, 35F, and 37 were associated with carriage. In addition, significant differences in invasive disease potential between clones sharing the same serotype were found among several serotypes – 3, 6A, 6B, 11A, 14, 19A, 19F, 22F, 23F, 34, and NT. This heterogeneous behavior of the clones was found irrespective of the serotype’s overall invasive disease potential. Our results highlight the importance of the genetic background when analyzing the invasive disease potential of certain serotypes and provide an important baseline for its monitoring following conjugate vaccine use. Continuous surveillance should be maintained and current research should focus on uncovering the genetic determinants that contribute to the heterogeneity of invasive disease potential of clones sharing the same serotype.
INTRODUCTION

The capsule of pneumococci is considered its main virulence factor (23) and at least 93 capsular types (or serotypes) have been described (12). The different polysaccharide structures confer a range of properties to the bacterium leading to enhanced colonization capacity or hindering of phagocytosis (20). In addition, pneumococci possess various other virulence factors several of which are variable between strains. Some are prone to allelic variation, such as PspA (22), while others can be either present or absent, such as pili (4, 7).

Indeed, recent data from genome sequencing suggested that, for every pneumococcal strain, 20-30% of its DNA content could be classified as accessory genome (21). Thus, substantial genetic diversity that may impact in the overall colonization and virulence capacities of the bacterium may be expected between isolates. Furthermore, detailed molecular typing studies have shown that genetic diversity within serotypes can vary greatly, for instance being generally low among serotype 1 (9, 11) and high among serotype 19A (2), and substantial geographical and temporal variations were already reported.

Recent studies have looked at the distribution of pneumococcal serotypes in colonization and disease to estimate their invasive disease potential (9, 10, 19, 24, 29, 31). Some of these studies have extended their analysis (by genotyping the isolates) in an attempt to establish the relative contribution of serotype versus genotype to the disease potential of pneumococcal isolates (9, 19, 29). Although still a matter of debate, the serotype is generally accepted as the main predictor of invasive disease potential. In the era of multi-valent pneumococcal conjugate vaccines, such studies are particularly important to obtain information that will increase our understanding - and may even help to anticipate - serotype and clonal replacement changes occurring in carriage and disease as a consequence of vaccination.
In this study we have compared a large collection of invasive and colonizing isolates obtained between 2001 and 2003 in Portugal before massive use of the seven-valent pneumococcal conjugate vaccine (PCV7, including serotypes 4, 6B, 9V, 18C, 19F and 23F) (3, 27). Molecular typing was performed for all isolates enabling us to explore the relative contribution of genotype versus capsular type for the invasive disease potential of pneumococcal clones. We found that a number of serotypes presented enhanced invasive disease potential, including several not comprised in the highest valency conjugate vaccine available at the moment (PCV13, a 13-valent conjugate vaccine including serotypes 1, 3, 4, 5, 6A, 6B, 7F, 9V, 14, 18C, 19F, 19A and 23F). Moreover, we found within some serotypes a significant heterogeneity in invasive disease potential between clones, emphasizing the importance of the genetic context in which a serotype is expressed.

MATERIALS AND METHODS

Bacterial isolates. Two collections of S. pneumoniae obtained between 2001 and 2003 in Portugal were analyzed in the study: one included invasive disease isolates, the other included carriage isolates.

The 475 invasive isolates were obtained in 30 microbiology laboratories throughout Portugal between January 2001 and December 2003, and sent to a central laboratory for characterization. The clinical sources of the isolates were: blood (92.6%), cerebrospinal fluid (4.0%), pleural fluid (2.8%), and other normally sterile sites (0.6%). Seventy-nine isolates were from children aged up to six years old, 13 were from children >6 and ≤18 years, and 383 were from adults.
The 769 nasopharyngeal isolates were obtained from children up to six years old, attending day-care centers in the areas of Lisbon and Oeiras. Isolates were recovered during six sampling periods (January-February 2001, May-June 2001, November-December 2001, February-March 2002, November-December 2002, and February-March 2003). In each period one nasopharyngeal sample was obtained from each child.

Part of the collections analyzed in this study, i.e., the drug-resistant colonizing isolates (n=276), and the invasive isolates obtained in 2001 and 2002 (n=265), were initially characterized under the scope of other initiatives (25, 30).

**Serotyping.** Isolates were serotyped by the Quellung reaction using commercially available antisera (Statens Seruminstitut, Copenhagen, Denmark) (32).

**Pulsed-field gel electrophoresis (PFGE) and multilocus sequence typing (MLST).** All isolates were analyzed by PFGE. Briefly, total DNA was extracted, digested with Smal, and the DNA fragments were resolved by PFGE as described previously (28). PFGE patterns were analyzed with the Bionumerics software (version 4.5, Applied Maths, Gent, Belgium). For each serotype, PFGE patterns were clustered by UPGMA, and a dendrogram was generated from a similarity matrix calculated using the Dice similarity coefficient with an optimization of 1.0% and a tolerance of 1.5%. PFGE clusters were defined as isolates with a similarity of 80% or higher on the dendrogram (13, 33). The integration of two independent databases - one containing disease isolates, the other containing carriage isolates - generated in different laboratories was done by calibrating the band sizes of markers used in each separate database using the Bionumerics normalization procedure. PFGE clusters were named as a combination
of the serotype and arbitrary consecutive numbers; for instance, cluster 14-1 refers to cluster 1 of serotype 14.

After analysis of the dendrograms generated by combining the two collections, MLST was performed on representative isolates which were selected according to the following criteria: (i) a minimum of two isolates from each PFGE cluster containing 10 or more isolates; (ii) if such clusters included colonization and invasive disease isolates, at least one isolate from each group was selected; (iii) a minimum of one isolate from each PFGE cluster containing five to nine isolates. Isolates from PFGE clusters with less than five isolates were not systematically selected for characterization by MLST. Clonal complex (CC) assignment, was done by using the goeBURST program (17) and the complete *S. pneumoniae* MLST database available at spneumoniae.mlst.net. Clones were defined as groups of isolates clustered together by PFGE for which representative isolates were also characterized by MLST.

**Statistical analysis.** Simpson’s index of diversity (SID) was used to measure the population diversity (14) using the www.comparingpartitions.info website. Wallace coefficients (W) were used to compare the congruence between typing methods (14, 25a).

The invasive potential of particular serotypes and clones was estimated using odds ratios (OR) with 95% confidence intervals (CI95%) computed through the Fisher method implemented in the epitools package for the R language (5). OR significance was tested with the two-tailed Fisher exact test. The ORs for enhanced invasive potential of a particular serotype or genotype were calculated using the number of invasive and carriage isolates by reference to all other isolates from these two sources.
Heterogeneity of invasive disease potential within each serotype was evaluated through two-tailed Fisher exact tests applied to the contingency table of the PFGE clusters by source collection (invasive or carriage). For all the analyses, the resulting p-values were corrected for multiple testing by controlling the False Discovery Rate (FDR) under or equal to 0.05 through the linear procedure of Benjamini and Hochberg (8).

RESULTS

Diversity of serotypes in disease and colonization.

In the total collection 49 serotypes were identified, 32 of which were identified in both disease and carriage isolates (Figure 1, Table S1). Isolates that could not be typed with existing sera (non-typable, NT) were also found in both groups.

The 475 invasive disease isolates were distributed among 43 capsular types. The most common invasive serotypes – accounting for over half of the isolates - were 14 (12.8%), 1 (9.3%), 3 (9.1%), 4 (6.7%), 8 (5.7%), 19A (5.3%), and 7F (4.4%). Serotypes which were exclusively found in invasive disease were 5 (n=12), 12B (n=10), 20 (n=10), 9L (n=6), 7C (n=4), 10F (n=4), and with a single isolate of serotypes 13, 19C, 24B, 25F, and 35B.

The 769 carriage isolates were distributed among 38 capsular types. The most commonly carried serotypes – accounting for over half of the isolates - were 19F (12.1%), 23F (9.0%), 6B (9.1%), 6A (8.2%), 14 (6.8%), and 19A (5.9%). Serotypes which were exclusively found in carriage were 37 (n=9), 21 (n=6), 15F (n=4), 42 (n=4), and with a single isolate serotypes 29 and 39.
The serotype diversity was indistinguishable in both collections: the SID at the serotype level for the invasive disease collection was 0.946 (CI\textsubscript{95%}, 0.939 to 0.953) and for the carriage collection was 0.941 (CI\textsubscript{95%}, 0.936 to 0.946).

**Association of serotypes with carriage and disease.**

When the relative proportion of individual serotypes being detected in carriage and disease was compared through estimation of OR, different associations were revealed. In particular, serotypes 1, 3, 4, 5, 7F, 8, 9N, 9L, 12B, 14, 18C, and 20 were found to have an enhanced propensity to cause invasive disease as all had an OR significantly higher than 1 (Figure 1 and Table S1). In contrast, serotypes 6A, 6B, 11A, 15B/C, 16F, 19F, 23F, 34, 35F, and 37 were associated with carriage (i.e. had a low invasive disease potential).

When the same analysis was repeated restricting the invasive disease isolates to those originating from children only (n=79), serotypes 1 (OR, 135.78; CI\textsubscript{95%}, 19.54 to 5663.69), 7F (OR, 13.52; CI\textsubscript{95%}, 2.24 to 94.04), and 18C (OR, 4.76; CI\textsubscript{95%}, 1.44 to 13.93) were found to have a higher propensity to cause invasive disease while serotype 19F (OR, 0.09; CI\textsubscript{95%}, 0.00 to 0.55) was mostly associated with carriage.

**Clonal diversity in disease and colonization.**

A total of 193 PFGE clusters were identified of which 81 comprised a single isolate and the remaining were clusters that had between 2 and 84 isolates. Forty-three PFGE clusters contained both carriage and invasive disease isolates and these included 56.4% of both carriage and invasive isolates. A total of 217 isolates were characterized by MLST and 115 different STs were found. The most common clones – PFGE clusters and associated sequence types (ST) - are detailed in Table S1. The W value between PFGE cluster and CC was high
(W_{PFGE-CC}=0.844, CI_{95\%} of 0.774-0.913) indicating that over five out of six pairs of isolates belonging to the same PFGE cluster were also classified together in the same CC. This reassured us that the PFGE clusters were identifying congruent genetic lineages as defined by goeBURST. The mapping of PFGE and MLST data allow the comparison of our results with studies using only one of these techniques for typing of the isolates.

The diversity defined by PFGE clustering was indistinguishable in both disease and carriage collections: the SID for the invasive disease collection was 0.974 (CI_{95\%}, 0.969 to 0.979) and for the carriage collection was 0.967 (CI_{95\%}, 0.963 to 0.972).

**Analysis of heterogeneity of invasive disease potential between clones expressing the same serotype.**

To gain insights into the relative contribution of serotype and genotype for the invasive disease potential of a given isolate, the heterogeneity of invasive potential between lineages with the same serotype was investigated. In other words, for each serotype the distribution of clones in invasive disease and carriage was compared. A serotype was considered homogeneous in its invasive disease potential if all clones expressing that serotype were distributed in an equal manner in carriage and invasive disease. Conversely, a serotype was considered heterogeneous in its invasive disease potential if at least one clone was associated with carriage or invasive infection in a different manner.

Evidence for heterogeneity of invasive potential within serotypes was obtained for serotypes 3 and 14 (both serotypes with OR significantly higher than 1), 6A, 6B, 11A, 19F, 23F, and 34 (all serotypes with OR significantly lower than 1), and serotypes 19A, 22F, and NT (all serotypes with OR not significantly different from 1) (Figure 1, Table S1). Although the
Fisher exact test suggested (p<0.05) that serotypes 16F, 18A, and 35F were heterogeneous, the results were not significant after FDR correction for multiple testing.

When the same analysis was repeated for children isolates only, heterogeneity of invasive disease potential was detected for serotypes 6A, 6B, 19A, 19F, and 23F, but the results were only robust after FDR correction for serotypes 6A, 6B, and 19A.

**Association of clones with invasiveness.**

Table 1 and Figure S1 detail the significant OR found among clones belonging to serotypes with heterogeneous invasive disease potential. A conservative approach was taken to estimate these OR as the entire collection of carriage and disease isolates (regardless of the serotype) was used as a reference. After FDR correction it was possible to identify at least one clone in all but one heterogeneous serotype (i.e. serotypes 3, 6A, 6B, 14, 19A, 19F, 22F, 23F, 34, and NT, the exception being 11A) that, per se, was clearly associated with carriage or disease. In most cases, as expected, the individual clones reflected mostly the invasive capacity of the serotype they expressed. The exceptions were serotypes 19A and 22F and the non-typable isolates that, although not reaching a significant OR when taken together, included clones that were significantly associated with invasive disease or colonization (Table 1 and Figure S1).

For three serotypes (6A, 6B and 19A) more than one clone within each serotype showed significant OR, even after FDR correction. While in serotype 6B both clones were associated with carriage, as was the serotype as a whole, in the case of serotypes 6A and 19A clones with opposing invasive disease potential could be identified. In serotype 6A, in contrast to the three clones with significant OR that were associated with carriage (similarly to the serotype as a whole), the five isolates not classified into any PFGE cluster (6A-unique), were collectively...
found to have an enhanced invasive disease capacity. In serotype 19A, while the major clone
was found to be associated with carriage, another clone presented an enhanced invasive
disease capacity, and this may be the reason why the overall serotype OR did not reach
significance.

When the same analysis was repeated for isolates recovered from children only, three
serotype 6B clones and one 19A clone were identified as having OR significantly different
from one, but these were found to lose significance once the FDR correction was performed
(data not shown).

**DISCUSSION**

The capsular polysaccharide plays a major role in pneumococcal virulence. The longstanding
use of serotyping to discriminate pneumococcal strains and the realization that the frequency
of the various serotypes was different in colonization and infection, led to the hypothesis that
the capsular polysaccharide was the major factor in determining the invasive potential of a
given isolate. This hypothesis was specifically addressed in several studies (9, 10, 19, 24, 29,
31) and a few of those have also tested if particular genetic lineages, irrespective of serotype,
would also show increased invasive potential (9, 19, 29). As a result of these studies, some
authors propose that serotype alone determines invasiveness, while others found evidence that
the clonal type is also important. In our study we have explored both hypothesis based on the
comparison of the distribution of pneumococcal serotypes and clones among a large
collection of invasive and carriage isolates obtained in Portugal, between 2001 and 2003,
before massive use of PCV7.
While it is frequently assumed that only a subset of the serotypes and clones found in carriage is capable of causing invasive infections, over half of the isolates from each collection belonged to clones that were detected in both collections. In other words, most carried clones have the capacity to cause invasive disease and, when taken together, also account for the majority of disease isolates, indicating a close relationship between the two bacterial populations. Moreover, the diversity of both serotypes and genotypes among the isolates causing invasive infections was indistinguishable from that found in carriage. Still, when analyzed in more detail the capacity of individual serotypes and clones to cause invasive infections can vary substantially as discussed below.

Serotypes 1, 3, 4, 5, 7F, 8, 9N, 9L, 12B, 14, 18C, and 20 were commonly associated with invasive disease in general concordance with previous reports (9, 10, 19, 24, 29, 31). In this group most serotypes showed no evidence of significant heterogeneity – i.e. most serotypes were not comprised of clones with distinct invasive potential – except for serotypes 3 and 14 where individual clones with significant OR could be identified. For the remaining serotypes all genetic lineages had a comparable and elevated propensity to cause disease, not rejecting the hypothesis that, in these cases, serotype alone may be the determinant of invasive potential. In line with their high invasive disease potential, a recent study from Portugal found that in 2006-2008, following extensive use of PCV7, serotypes 1, 3 and 7F, together with serotype 19A discussed below, consolidated as the most frequent causes of invasive infections in children (1). However, these important serotypes are now included in the PCV13 formulation and its use is expected to lead to sharp declines of the number of invasive infections by these serotypes, as happened before with the PCV7 serotypes (3). Still, the group of homogeneously invasive serotypes included types 8, 9N, 9L, 12B, and 20 that are
not included in PCV13 raising the possibility that the prevalence of invasive infections by these serotypes in the post-vaccination era may increase.

On the other extreme, serotypes 6A, 6B, 11A, 15B/C, 16F, 19F, 23F, 34, 35F, and 37 were commonly associated with carriage, also in agreement with previous studies (9, 10, 19, 24, 29, 31). Within this group, heterogeneity of invasive potential between clones of the same serotype was detected for several serotypes - 6A, 6B, 11A, 19F, 23F, and 34 - although the low number of isolates often did not allow a clear discrimination of the properties of individual clones sharing the same serotype. Since most of the serotypes associated with carriage are not included in PCV13 (serotypes 11A, 15B/C, 34, 35F and 37) a declining frequency of PCV13 serotypes expected from vaccine use similarly to what was found for PCV7 (27), may afford an opportunity for increased colonization by these serotypes. Their low invasive disease potential may mean that we will not see their proportional increase in invasive disease; a desirable outcome of vaccination that only continuing surveillance will be able to confirm.

Among serotypes with “neutral” (OR not significantly different from 1) invasive disease capacity, 19A, 22F, and NT stood out as serotypes with heterogeneity of invasive potential between the clones that comprised them. We have recently documented the epidemiology of serotype 19A in Portugal between 2001 and 2006 (1). In line with that study, we found that among the clones expressing this serotype we could find some proficient in colonization while others were more prone to cause invasive disease. Although serotype 22F as a whole did not present an OR significantly different from one, cluster 22F-1 (ST443) was characterized by an enhanced invasive disease potential. Although apparently contradictory, these results reflect the fact that the other large cluster in this serotype, cluster 22F-2,
comprises mostly carriage isolates although it does not reach a statistical significant OR. Serotype 19A, that was not included in PCV7, saw an extraordinary increase in frequency of invasive infections worldwide following PCV7 use (2), confirming the potential of these serotypes to emerge as important causes of invasive infections. Serotype 22F on the other hand, although still not notable in invasive infections in children in Portugal in the post-PCV7 period (1), it is not included in the PCV13 formulation raising the possibility that this vaccine could facilitate its emergence as an important cause of invasive infections.

Among the group of serotypes with a “neutral” invasive potential, serotypes 11A, 15B/C, 16F, 34, 35F, and 37 are also not included in PCV13. Given that their potential to cause invasive disease is low, they are not such a cause of concern as those previously discussed. However, the vacant nasopharyngeal niche expected to be left behind by the decline of PCV13 serotypes following its use, may increase their frequency as colonizers that can then be reflected in smaller but still significant increase in their overall frequency among invasive disease isolates.

The general concordance between our study and the previous ones reinforces the idea that serotype plays an important role in determining invasive disease potential. Nevertheless, for several serotypes we obtained evidence for differences in invasive disease potential between clones sharing the same serotype. This heterogeneity in clonal behavior was observed independently of the serotype’s overall invasive disease potential, indicating that for some serotypes other factors, in addition to the capsular polysaccharide, have a major contribution to the propensity of a given isolate to cause invasive disease. Why is it then that the clone’s genetic background seems to be important to determine its invasive disease potential in some serotypes but not others? One possibility is that in some cases the influence of the type of capsular polysaccharide in the isolate’s potential to cause invasive infection is so
overpowering that the influence of other factors is simply overwhelmed. Another possibility is that the genetic diversity of the serotypes that we found to be homogeneous in terms of invasive potential is limited and that therefore we are in fact evaluating only a few clones whose genetic background may also condition their invasive potential. Inspection of table S1 largely supports this later hypothesis. Further studies sampling a larger clonal diversity within each serotype are needed to address this issue.

Our study has some potential limitations. First, while disease isolates originated from all over the country, colonization isolates were geographically limited to the area of Lisbon and Oeiras. Previous studies did not find significant regional asymmetries in Portugal among disease isolates and thus those would also not be expected in carriage (3, 30). Secondly, isolates from carriage were obtained from children aged up to six years old while isolates from disease were obtained from patients of all ages. There is now ample evidence that children are the main reservoir of pneumococci in the community since in countries where widespread use of PCV7 has occurred, the incidence of invasive disease caused by vaccine-types has declined not only in the target group but also in all other age groups. This herd effect has been attributed to decreased transmission of carried pneumococci from young children to adults (26). In addition, while carriage rates are very high among young children they decline with age and remain low throughout life (16, 18). We thus opted for not restricting our analysis to young children. Still, for comparison purposes we have also performed analyses taking only into consideration the isolates recovered from children. As expected due to the low numbers of invasive disease isolates, fewer results reached statistical significance. However, it is important to note that all results were in agreement with those obtained when the entire collection was considered. Thirdly, isolates from carriers were all from day-care center attendees. While in some countries day-care attendance is relatively uncommon, in Portugal
close to 80% of children attend day-care regularly (five full days a week) and hence the patterns of carriage found should reflect the majority of the population of this age (6).

Our study also has some strengths. First, it was confined in time as all data were obtained during three consecutive years. Secondly, a large collection of invasive disease and carriage isolates was put together enabling us to determine the invasive disease potential of serotypes that had not been accessed before. Thirdly, detailed molecular data was available for all isolates allowing us to investigate which genotypes were significantly associated with carriage or disease. Fourthly, by introducing a statistical correction for multiple testing, our results should contain less false positive associations, as compared with previous studies characterizing serotype or genotype invasive disease potential.

In summary, we have described the invasive disease potential of pneumococcal clones circulating in Portugal before widespread use of PCV7. Heterogeneity of invasive potential between clones sharing the same serotype was found in various serotypes, regardless of the serotype’s overall potential to cause invasive disease, suggesting that the genetic background has an important contribution to a given isolate’s invasive potential. With the massive introduction of conjugate vaccines it will be important to monitor the evolution of non-vaccine serotypes as dramatic changes in colonization and disease patterns are known to occur (3, 15, 26, 27). This close surveillance should focus not only on serotypes with high propensity to cause disease but also on heterogeneous serotypes and its clones, since these clearly show the capacity to increase in frequency in invasive infections following the expected decline in PCV13 serotypes accompanying its widespread use. This surveillance should be maintained among carried and infection causing pneumococci and should rely not only on serotyping but also on the identification of genetic lineages. The current data can
form the basis for future work aiming to elucidate the determinants contributing to heterogeneity in invasive potential within some serotypes as they may point to important and clone dependent virulence factors.

ACKNOWLEDGEMENTS
Members of the Portuguese Surveillance Group for the Study of Respiratory Pathogens are gratefully thanked for their valuable collaboration in this study. S.I. Aguiar, S. Nunes, and N. Frazão were supported by grants SFRH/BD/27518/2006, SFRH/BD/40706/2007, and SFRH/BD/30103/2006, respectively from Fundação para a Ciência e Tecnologia, Portugal. Partial support for this work was provided by EURIS (contract QKL2-CT-2000-01020) and PREVIS (LSHM-CT-2003-503413) from the European Community, Fundação para a Ciência e Tecnologia (PDTC/SAU-ESA/64888/2006, PTDC/SAU-ESA/65048/2006, PIC/IC/83065/2007, and PTDC/BIA-BEC/098289/2008), and Fundação Calouste Gulbenkian.

REFERENCES


FIGURE LEGENDS

FIGURE 1. Homogeneous and heterogeneous distribution of clones in carriage and invasive infections of the various serotypes. Only serotypes found in at least five isolates are represented. The number of isolates expressing each serotype is indicated. Isolates recovered from asymptomatic carriers are indicated by white bars and isolates recovered from invasive infections by black bars. Serotypes included in PCV13 are indicated by asterisks. Serotypes comprised of clones showing different invasive disease potentials are indicated on the left. Serotypes comprised of clones showing similar invasive disease potentials are indicated on the right.
<table>
<thead>
<tr>
<th>Genotype</th>
<th>No. of isolates from</th>
<th>Total no. of isolates</th>
<th>OR (95% CI)</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>ST (n*)</td>
<td>Invasive disease</td>
<td>Carriage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-1 180 (5), 232 (2)</td>
<td>25</td>
<td>26</td>
<td>51</td>
<td>1.59 (0.87-2.90)</td>
</tr>
<tr>
<td>3-2 162, 1220, 1646, 156, 260, singleton</td>
<td>13</td>
<td>1</td>
<td>14</td>
<td>21.57 (3.22-915.55)</td>
</tr>
<tr>
<td>6A-1 460, 1880</td>
<td>2</td>
<td>17</td>
<td>19</td>
<td>0.19 (0.02-0.79)</td>
</tr>
<tr>
<td>6A-2 1879 (2) singleton</td>
<td>0</td>
<td>12</td>
<td>12</td>
<td>0.00 (0.00-0.08)</td>
</tr>
<tr>
<td>6A-3 1878 (2)</td>
<td>0</td>
<td>14</td>
<td>14</td>
<td>0.00 (0.00-0.04)</td>
</tr>
<tr>
<td>6A-4 1152, 1647, 1876</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>0.97 (0.15-5.02)</td>
</tr>
<tr>
<td>6A-5 681</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>0.00 (0.00-1.12)</td>
</tr>
<tr>
<td>6A-7 395</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0.00 (0.00-8.62)</td>
</tr>
<tr>
<td>6A-unique</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6B-1 887 (2)</td>
<td>0</td>
<td>40</td>
<td>40</td>
<td>0.00 (0.00-0.15)</td>
</tr>
<tr>
<td>6B-2 176 (2)</td>
<td>0</td>
<td>19</td>
<td>19</td>
<td>0.00 (0.00-0.24)</td>
</tr>
<tr>
<td>6B-3 273</td>
<td>0</td>
<td>5</td>
<td>7</td>
<td>0.00 (0.00-0.48)</td>
</tr>
<tr>
<td>6B-unique</td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>8.16 (0.91-386.17)</td>
</tr>
<tr>
<td>11A-1 62 (2), 408 (3)</td>
<td>3</td>
<td>0</td>
<td>3</td>
<td>0.23 (0.09-0.54)</td>
</tr>
<tr>
<td>11A-2 156 (5)</td>
<td>15</td>
<td>28</td>
<td>22</td>
<td>2.70 (1.62-4.57)</td>
</tr>
<tr>
<td>19A-1 81 (3), 994</td>
<td>3</td>
<td>24</td>
<td>27</td>
<td>0.20 (0.04-0.66)</td>
</tr>
<tr>
<td>19A-2 115 (2), 416</td>
<td>7</td>
<td>0</td>
<td>7</td>
<td>5.73 (1.08-30.68)</td>
</tr>
<tr>
<td>19A-3 115 (2)</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td>2.35 (0.44-13.36)</td>
</tr>
<tr>
<td>19A-4 63 (2)</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>0.00 (0.00-1.27)</td>
</tr>
<tr>
<td>19F-1 177, 179 (2)</td>
<td>2</td>
<td>82</td>
<td>84</td>
<td>0.04 (0.00-0.13)</td>
</tr>
<tr>
<td>19F-2 1149 (2)</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>0.00 (0.00-1.12)</td>
</tr>
<tr>
<td>19F-unique</td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>1.08 (0.09-9.46)</td>
</tr>
<tr>
<td>22F-1 443</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>Inf (1.92-Inf)</td>
</tr>
<tr>
<td>22F-2 433</td>
<td>1</td>
<td>8</td>
<td>9</td>
<td>0.00 (0.00-1.51)</td>
</tr>
<tr>
<td>23F-1 338, 732</td>
<td>12</td>
<td>40</td>
<td>52</td>
<td>0.47 (0.22-0.93)</td>
</tr>
<tr>
<td>23F-2 33 (2)</td>
<td>0</td>
<td>16</td>
<td>16</td>
<td>0.00 (0.00-0.41)</td>
</tr>
<tr>
<td>23F-4 81</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>0.65 (0.06-3.97)</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
<td>---</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

No. of isolates showing that particular sequence type (ST) if different from one.

The predicted founder of the clonal complex, defined by goeBURST, is indicated. When the predicted founder could not be identified with certainty, the arbitrary group number is indicated.

Highlighted in bold are OR estimates whose confidence interval does not include one. Inf, infinite.

Values in bold are indicated when the confidence interval does not include one.

Significant p-values after FDR correction are marked by asterisks.

For the purpose of this analysis all isolates that were not grouped with any other in a PFGE cluster were analyzed together.
No. of isolates

Invasive Carriage

Heterogeneous invasive disease potential

Homogeneous invasive disease potential

OR > 1 OR < 1 OR > 1 OR < 1

* * * * * * * * * * ***

Heterogeneous invasive disease potential

Homogeneous invasive disease potential