Analysis of Invasiveness of Pneumococcal Serotypes and Clones Circulating in Portugal before Widespread Use of Conjugate Vaccines Reveals Heterogeneous Behavior of Clones Expressing the Same Serotype

Raquel Sá-Leão,1‡,‡* Francisco Pinto,2‡ Sandra Aguiar,3 Sónia Nunes,1 João A. Carriço,3 Nelson Frazão,1 Natacha Gonçalves-Sousa,1 José Melo-Cristino,3 Hermínia de Lencastre,1,4 and Mário Ramirez3

Instituto de Tecnologia Química e Biomédica, Universidade Nova de Lisboa, Oeiras, Portugal1; Centro de Química e Bioquímica, Departamento de Química e Bioquímica, Faculdade de Ciências da Universidade de Lisboa, Lisbon, Portugal2; Instituto de Microbiologia, Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Lisbon, Portugal3; and Laboratory of Microbiology, The Rockefeller University, New York, New York4

Received 30 August 2010/Returned for modification 17 November 2010/Accepted 19 January 2011

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 of isolates were also characterized by multilocus sequence typing. We found that the diversities of serotypes and genotypes of pneumococci responsible for invasive infections and carriage were identical and that most carried clones could also be detected as causes of invasive disease. Their ability to do so, however, varied 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, namely, 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.

The capsule of the pneumococcus 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 to 30% of the DNA content could be classified as an accessory genome (21). Thus, substantial genetic diversity that may impact 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 isolates (9, 11) and high among serotype 19A isolates (2), and substantial geographical and temporal variations have already been reported.

Recent studies 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 extended their analyses (by genotyping the isolates) in an attempt to establish the relative contributions 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 multivalent pneumococcal conjugate vaccines, such studies are particularly important to obtain information that will increase our understanding of—and maybe even help to anticipate—serotype and clonal replacement changes occurring in carriage and disease as a consequence of vaccination.

In this study, we 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; includes serotypes 4, 6B, 9V, 14, 18C, 19F, and 23F) (3, 27). Molecular typing was performed for all isolates, enabling us to explore the relative contributions of...
genotype versus capsular type to the invasive disease potential of pneumococcal clones. We found that a number of serotypes presented enhanced invasive disease potential, including several not included 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* isolates obtained in Portugal between 2001 and 2003 were analyzed in the study: one included invasive disease isolates, and the other included carriage isolates.

The 475 invasive disease isolates were obtained in 30 microbiology laboratories throughout Portugal between January 2001 and December 2003 and were 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 of up to 6 years old, 13 were from children of 6 to 13 years old, and 383 were from adults.

The 769 nasopharyngeal isolates were obtained from children of up to 6 years of age 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). During each period, one nasopharyngeal sample was obtained from each child.

Parts 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 Serum Institut, Copenhagen, Denmark) (32).

**PFGE and multilocus sequence typing (MLST).** All isolates were analyzed by pulsed-field gel electrophoresis (PFGE). Briefly, total DNA was extracted and digested with Smal, and the DNA fragments were resolved by PFGE as described previously (28). PFGE patterns were analyzed with BioNumerics software (version 4.3; Applied Maths, Ghent, Belgium). For each serotype, PFGE patterns were clustered by the unweighted-pair group method using average linkages (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 in the dendrogram (13, 33). The integration of two independent databases—one containing disease isolates and the other containing carriage isolates (generated in different laboratories)—was done by calibrating the band sizes of marker 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 according to the following criteria: (i) a minimum of two isolates from each PFGE cluster containing 10 or more isolates were selected, (ii) if such clusters included colonized and invasive disease isolates, at least one isolate from each group was selected, and (iii) a minimum of one isolate from each PFGE cluster containing five to nine isolates was selected. Isolates from PFGE clusters with fewer than five isolates were not selected systematically 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 http://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 disease potential of particular serotypes and clones was estimated using odds ratios (ORs) with 95% confidence intervals (95% CI), computed by 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 disease potential of a particular serotype or genotype were calculated using the numbers of invasive and carriage isolates in 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 disease or carriage). For all analyses, the resulting P values were corrected for multiple testing by controlling the false discovery rate (FDR) to ≤0.05 through the linear procedure of Benjamini and Hochberg (8).

**RESULTS**

**Diversity of serotypes in disease and colonization.** In the total collection of isolates, 49 serotypes were identified, 32 of which were identified in both disease and carriage isolates (Fig. 1; see Table S1 in the supplemental material). Isolates that could not be typed with existing sera (nontypeable [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% of isolates), 1 (9.3%), 3 (9.1%), 4 (6.7%), 8 (5.7%), 19A (5.3%), and 7F (4.4%). Serotypes which were found exclusively in invasive disease isolates were 5 (n = 12), 12B (n = 10), 20 (n = 10), 9L (n = 6), 7C (n = 4), 10F (n = 4), and, with a single isolate each, 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% of isolates), 23F (9.0%), 6B (9.1%), 6A (8.2%), 14 (6.8%), and 19A (5.9%). Serotypes which were found exclusively in carriage isolates were 37 (n = 9), 21 (n = 6), 15F (n = 4), 42 (n = 4), and, with a single isolate each, 29 and 39.

The serotype diversities were indistinguishable for both collections: the SID at the serotype level for the invasive disease collection was 0.946 (95% CI, 0.939 to 0.953), and that for the carriage collection was 0.941 (95% CI, 0.936 to 0.946).

**Association of serotypes with carriage and disease.** When the relative proportions of individual serotypes being detected in carriage and disease were compared through estimation of ORs, 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 ORs of significantly more than 1 (Fig. 1; see Table S1 in the supplemental material). 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, but restricting the invasive disease isolates to those originating from children only (n = 79), serotypes 1 (OR, 135.78; 95% CI, 19.54 to 5,663.69), 7F (OR, 13.52; 95% CI, 2.24 to 94.04), and 18C (OR, 4.76; 95% CI, 1.44 to 13.93) were found to have a higher propensity to cause invasive disease, while serotype 19F (OR, 0.09; 95% CI, 0.00 to 0.55) was associated mostly 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 remainder 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 sequence types (STs) were found. The most common clones—PFGE clusters
and associated STs—are detailed in Table S1 in the supplemental material. The $W$ value between PFGE clusters and CC was high ($W_{\text{PFGE}\rightarrow\text{CC}} = 0.844; 95\% \text{ CI}, 0.774$ to $0.913$), indicating that at least five 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 identified congruent genetic lineages as defined by goeBURST. The mapping of PFGE and MLST data allowed comparison of our results with those of studies using only one of these techniques for typing of the isolates.

The diversities defined by PFGE clustering were indistinguishable for the disease and carriage collections: the SID for the invasive disease collection was $0.974 (95\% \text{ CI}, 0.969$ to $0.979$), and that for the carriage collection was $0.967 (95\% \text{ CI}, 0.963$ to $0.972$).

**Analysis of heterogeneity of invasive disease potential between clones expressing the same serotype.** To gain insights into the relative contributions of serotype and genotype to the invasive disease potential of a given isolate, the heterogeneity of invasive disease potential between lineages with the same serotype was investigated. In other words, for each serotype, the distributions of clones in invasive disease and carriage were 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 disease potential within serotypes was obtained for serotypes 3 and 14 (with ORs of significantly more than 1), serotypes 6A, 6B, 11A, 19F, 23F, and 34 (with ORs of significantly less than 1), and serotypes 19A, 22F, and NT (with ORs not significantly different from 1) (Fig. 1; see Table S1 in the supplemental material). 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 with children’s isolates only, heterogeneity of invasive disease potential was detected for serotypes 6A, 6B, 19A, 19F, and 23F, but the results were robust after FDR correction for serotypes 6A, 6B, and 19A only.

**Association of clones with invasiveness.** Table 1 and Fig. S1 in the supplemental material detail the significant ORs found among clones belonging to serotypes with heterogeneous invasive disease potential. A conservative approach was taken to estimate these ORs, 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, with the exception being serotype 11A) that was clearly associated per se with carriage or disease. In most cases, as expected, the individual clones mostly reflected the invasive capacities of the serotypes they expressed. The exceptions were serotypes 19A and 22F and the nontypeable isolates, which did not reach a significant OR when taken together but included clones that...
were significantly associated with invasive disease or colonization (Table 1; Fig. S1).

For three serotypes (6A, 6B, and 19A), more than one clone within each serotype showed a 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 cases 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 ORs that were associated with carriage (similar 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, which may be the reason that 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 ORs significantly different from 1, 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 frequencies of the various serotypes are different in colonization and infection led to the hypothesis that the capsular polysaccharide is the major factor in determining the invasive disease potential of a given isolate. This hypothesis was specifically addressed in several studies (9, 10, 19, 24, 29), and a few of those also tested if particular genetic lineages, irrespective of serotype, would also show increased invasive disease potential (9, 19, 28). As a result of these studies, some authors proposed that serotype alone determines invasiveness, while others found evidence that the clonal type is also important. In our study, we explored both hypotheses, based on

### TABLE 1. Invasive disease potential of clones (n ≥ 5) belonging to heterogeneous serotypes

<table>
<thead>
<tr>
<th>PFGE cluster</th>
<th>ST(s) (n)*</th>
<th>CCb</th>
<th>No. of isolates</th>
<th>Total no. of isolates</th>
<th>OR (95% CI)c</th>
<th>P valuecd</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>Invasive disease</td>
<td>Carriage</td>
<td></td>
<td></td>
</tr>
<tr>
<td>3-1</td>
<td>180 (5), 232 (2)</td>
<td>180</td>
<td>25</td>
<td>26</td>
<td>51</td>
<td>1.59 (0.87–2.90)</td>
</tr>
<tr>
<td>3-2</td>
<td>162, 1220, 1646</td>
<td>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</td>
<td>460, 1880</td>
<td>460</td>
<td>2</td>
<td>17</td>
<td>19</td>
<td>0.04 (0.00–0.13)</td>
</tr>
<tr>
<td>6A-2</td>
<td>1879 (2)</td>
<td>Singleton</td>
<td>0</td>
<td>12</td>
<td>12</td>
<td>0.00 (0.00–0.13)</td>
</tr>
<tr>
<td>6A-3</td>
<td>1878 (2)</td>
<td>315</td>
<td>0</td>
<td>14</td>
<td>14</td>
<td>0.00 (0.00–0.13)</td>
</tr>
<tr>
<td>6A-4</td>
<td>1152, 1647, 1876</td>
<td>473</td>
<td>3</td>
<td>5</td>
<td>8</td>
<td>0.97 (0.15–5.02)</td>
</tr>
<tr>
<td>6A-5</td>
<td>681</td>
<td>490</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>0.00 (0.00–1.12)</td>
</tr>
<tr>
<td>6A-7</td>
<td>395</td>
<td>395</td>
<td>0</td>
<td>2</td>
<td>2</td>
<td>0.00 (0.00–1.12)</td>
</tr>
<tr>
<td>6A-unique</td>
<td></td>
<td></td>
<td>5</td>
<td>0</td>
<td>5</td>
<td>Infinity (1.49–infinity)</td>
</tr>
<tr>
<td>6B-1</td>
<td>887 (2)</td>
<td>315</td>
<td>0</td>
<td>40</td>
<td>40</td>
<td>0.00 (0.00–0.13)</td>
</tr>
<tr>
<td>6B-2</td>
<td>176 (2)</td>
<td>176</td>
<td>0</td>
<td>19</td>
<td>19</td>
<td>4.08 (0.66–42.93)</td>
</tr>
<tr>
<td>6B-3</td>
<td>273</td>
<td>176</td>
<td>5</td>
<td>2</td>
<td>7</td>
<td>8.16 (0.91–83.17)</td>
</tr>
<tr>
<td>6B-unique</td>
<td></td>
<td></td>
<td>5</td>
<td>1</td>
<td>6</td>
<td>0.23 (0.08–0.54)</td>
</tr>
<tr>
<td>11A-1</td>
<td>62 (2), 408 (3)</td>
<td>53</td>
<td>6</td>
<td>41</td>
<td>47</td>
<td>2.70 (1.62–4.57)</td>
</tr>
<tr>
<td>14-1</td>
<td>156 (5)</td>
<td>156</td>
<td>44</td>
<td>28</td>
<td>72</td>
<td>0.77 (0.32–1.72)</td>
</tr>
<tr>
<td>14-2</td>
<td>15 (2), 409 (1)</td>
<td>15</td>
<td>10</td>
<td>21</td>
<td>31</td>
<td>0.20 (0.04–0.66)</td>
</tr>
<tr>
<td>19A-1</td>
<td>81 (3), 994</td>
<td>81, singleton</td>
<td>3</td>
<td>24</td>
<td>27</td>
<td>Infinity (1.92–infinity)</td>
</tr>
<tr>
<td>19A-3</td>
<td>230, 276</td>
<td>230</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>5.73 (1.08–55.68)</td>
</tr>
<tr>
<td>19A-4</td>
<td>193 (2), 416</td>
<td>193, 199</td>
<td>7</td>
<td>2</td>
<td>9</td>
<td>2.03 (0.44–10.30)</td>
</tr>
<tr>
<td>19A-5</td>
<td>1151 (2)</td>
<td>1151</td>
<td>5</td>
<td>4</td>
<td>9</td>
<td>0.00 (0.00–1.12)</td>
</tr>
<tr>
<td>19A-9</td>
<td>63 (2)</td>
<td>63</td>
<td>0</td>
<td>6</td>
<td>6</td>
<td>0.00 (0.00–1.12)</td>
</tr>
<tr>
<td>19F-1</td>
<td>177, 179 (2)</td>
<td>177</td>
<td>2</td>
<td>82</td>
<td>84</td>
<td>0.04 (0.00–0.13)</td>
</tr>
<tr>
<td>19F-2</td>
<td>1149 (2)</td>
<td>63</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></td>
<td></td>
<td>2</td>
<td>3</td>
<td>5</td>
<td>1.08 (0.09–9.46)</td>
</tr>
<tr>
<td>22F-1</td>
<td>443</td>
<td>Group 270</td>
<td>6</td>
<td>0</td>
<td>6</td>
<td>Infinity (1.92–infinity)</td>
</tr>
<tr>
<td>22F-2</td>
<td>433</td>
<td>433</td>
<td>1</td>
<td>8</td>
<td>9</td>
<td>0.00 (0.00–1.12)</td>
</tr>
<tr>
<td>23F-1</td>
<td>338, 732</td>
<td>176</td>
<td>12</td>
<td>40</td>
<td>52</td>
<td>0.07 (0.02–0.69)</td>
</tr>
<tr>
<td>23F-3</td>
<td>33 (2)</td>
<td>439</td>
<td>0</td>
<td>16</td>
<td>16</td>
<td>0.00 (0.00–0.13)</td>
</tr>
<tr>
<td>23F-4</td>
<td>81</td>
<td>81</td>
<td>2</td>
<td>5</td>
<td>7</td>
<td>0.65 (0.06–3.97)</td>
</tr>
<tr>
<td>23F-unique</td>
<td></td>
<td></td>
<td>1</td>
<td>5</td>
<td>6</td>
<td>0.32 (0.01–2.90)</td>
</tr>
<tr>
<td>34-1</td>
<td>2001 (2)</td>
<td>1439</td>
<td>0</td>
<td>12</td>
<td>12</td>
<td>0.00 (0.00–0.13)</td>
</tr>
<tr>
<td>34-2</td>
<td>1046</td>
<td>Group 238</td>
<td>0</td>
<td>5</td>
<td>5</td>
<td>0.00 (0.00–1.12)</td>
</tr>
<tr>
<td>NT-1</td>
<td>344 (2)</td>
<td>344</td>
<td>0</td>
<td>10</td>
<td>10</td>
<td>0.00 (0.00–1.12)</td>
</tr>
<tr>
<td>NT-2</td>
<td>1156</td>
<td>Group 157</td>
<td>0</td>
<td>7</td>
<td>7</td>
<td>0.00 (0.00–1.12)</td>
</tr>
<tr>
<td>NT-unique</td>
<td></td>
<td></td>
<td>7</td>
<td>3</td>
<td>10</td>
<td>3.81 (0.87–22.97)</td>
</tr>
</tbody>
</table>

a Number of isolates showing that particular sequence type, if different from one.

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

c Data shown in bold are OR estimates whose confidence interval does not include 1.

d P values that were significant after FDR correction are marked by asterisks.

e For the purpose of this analysis, all isolates that were not grouped with any other in a PFGE cluster were analyzed together.
comparison of the distributions 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, taken together, also account for the majority of disease isolates, indicating a close relationship between the two bacterial populations. Moreover, the diversities of both serotypes and genotypes among the isolates causing invasive infections were indistinguishable from those found in carriage. Still, analyzed in more detail, the capacities 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 disease potentials, except for serotypes 3 and 14, for which individual clones with significant ORs could be identified. For the remaining serotypes, all genetic lineages had comparable and elevated propensities to cause disease, not rejecting the hypothesis that, in these cases, serotype alone may be the determinant of invasive disease potential. In line with their high invasive disease potential, a recent study from Portugal found that in 2006 to 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 by these serotypes, as happened with the PCV7 serotypes (3). Still, analyzed in more detail, the capacities 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 disease potentials, except for serotypes 3 and 14, for which individual clones with significant ORs could be identified. For the remaining serotypes, all genetic lineages had comparable and elevated propensities to cause disease, not rejecting the hypothesis that, in these cases, serotype alone may be the determinant of invasive disease potential. In line with their high invasive disease potential, a recent study from Portugal found that in 2006 to 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).

Among the group of serotypes with a “neutral” invasive disease potential, serotypes 6C, 10A, 18A, 21, 23A, 23B, 24F, 31, and 33F are also not included in PCV13. Given that their potential to cause invasive disease is low, they are not as much a cause for concern as those previously discussed. However, the vacant nasopharyngeal niche expected to be left by the decline of PCV13 serotypes following its use may increase their frequency as colonizers, which may then be reflected in a smaller but still significant increase in their overall frequency among invasive disease isolates.

The general concordance between our study and 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, factors other than 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 in determining the invasive disease potential for some serotypes but not others? One possibility is that, in some cases, the influence of the type of capsular polysaccharide on 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 disease potential is limited and that we were in fact evaluating only a few clones whose genetic background may also condition their invasive disease potential. Inspection of Table S1 in the supplemental material largely supports the latter hypothesis. Further studies sampling a larger degree of 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 limited geographically to the area of Lisbon and Oeiras. Previous studies did not find significant regional asym-
metrics among disease isolates in Portugal, and thus those would also not be expected in carriage (3, 30). Second, isolates from carriage were obtained from children aged up to 6 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 not to restrict our analysis to young children. Still, for comparative purposes, we also performed analyses taking into consideration only the isolates recovered from children. As expected due to the small number of invasive disease isolates, fewer results reached statistical significance. However, it is important that all results were in agreement with those obtained when the entire collection was considered. Third, 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. Second, 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. Third, detailed molecular data were available for all isolates, allowing us to investigate which genotypes were associated significantly with carriage or disease. Fourth, by introducing a statistical correction for multiple testing, our results should contain fewer false-positive associations than those in previous studies characterizing serotype- or genotype-dependent 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 disease potential between clones sharing the same serotype was found for 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 disease potential. With the massive introduction of conjugate vaccines, it will be important to monitor the evolution of nonvaccine 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 a high propensity to cause disease but also on heterogeneous serotypes and their clones, since these clearly show the capacity to increase in frequency in invasive infections following the expected decline in PCV13 serotypes accompanying the widespread use of this vaccine. 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 disease potential within some serotypes, as they may point to important and clone-dependent virulence factors.

ACKNOWLEDGMENTS

The members of the Portuguese Surveillance Group for the Study of Respiratory Pathogens are gratefully thanked for their valuable collaboration in this study.

S. Aguiar, S. Nunes, and N. Frazão were supported by grants SFRH/BD/27518/2006, SFRH/BD/4396/2007, and SFRH/BD/30103/2006, respectively, from Fundação para a Ciência e Tecnologia, Portugal. Partial support for this work was provided by EURIS (contract OKL2-CT-2000-01020) and PREVIS (LSHM-CT-2003-503413) from the European Community, by Fundação para a Ciência e Tecnologia (PTDC/SAU-ESA/6488/2006, PTDC/SAU-ESA/65048/2006, PICT/83065/2007, and PTDC/BIA-BEC/098298/2008), and by Fundação Calouste Gulbenkian.

REFERENCES


20. Hall/CRC, Boca Raton, FL.


