# High Levels of Recombination among *Streptococcus pneumoniae* Isolates from the Gambia

**E. S. Donkor,[a] C. J. Bishop,[b] M. Antonio,[c] B. Wren,[a] and W. P. Hanage[b]***

Department of Infectious and Tropical Diseases, London School of Hygiene and Tropical Medicine, London, United Kingdom[a]; Department of Infectious Disease Epidemiology, Imperial College London, London, United Kingdom[b]; and Bacterial Diseases Programme, Medical Research Council Laboratories (United Kingdom), Fajara, the Gambia[c]

* Present address: Department of Epidemiology, Harvard School of Public Health, Boston, Massachusetts, USA.

**ABSTRACT** **We carried out multilocus sequence typing (MLST) on 148 pneumococcal carriage isolates collected from children <24 months old in the Upper River Division, the Gambia. MLST revealed a diverse population. Seventy-six different sequence types (STs) were found, the most common of which were 802 and 919, associated with 23F and 6A serotypes, respectively. Comparison with the MLST database showed that only 11 of the STs found in the present sample had been reported outside Africa. Six STs showed evidence of capsular switching (172, 802, 847, 1730, 1736, and 1737). Serotype switches were confirmed by microarrays that detected capsule genes. Of isolates analyzed by using microarrays, 40/69 (58%) harbored the *tetM* resistance determinant. A statistical genetic analysis to detect recombination found that 49/144 (34%) isolates showed significant ($P < 0.05$) evidence of admixture, which is greater than that observed in similar samples from the United Kingdom (5%) and Finland (2%). We hypothesize that large amounts of admixture could reflect the high prevalence of multiple carriage in this region, leading to more opportunities for homologous recombination between strains. This could have consequences for the population response to conjugate vaccination.**

**IMPORTANCE** **The population structure of the pneumococcus in sub-Saharan Africa has barely been studied despite the high levels of morbidity and mortality due to pneumococcal disease in this region. We report the largest sample to date from carriage in sub-Saharan Africa to be typed by sequencing (multilocus sequence typing [MLST]) and microarray analysis. The results clearly show that the population is highly distinct and divergent from others studied in the United States and Europe. Moreover, in contrast with samples from developed countries, the population contains a high proportion of isolates showing a history of homologous recombination, which shuffles genetic information into new combinations and can generate drug-resistant and vaccine escape strains. This is likely to have important consequences for the evolutionary response of the pneumococcal population to conjugate vaccination targeting a subset of pneumococcal serotypes, which is under way in the Gambia and other countries.**

**Received** 3 March 2011 **Accepted** 19 May 2011 **Published** 21 June 2011

**Citation** Donkor ES, Bishop CJ, Antonio M, Wren B, Hanage WP. 2011. High levels of recombination among *Streptococcus pneumoniae* isolates from the Gambia. mBio 2(3): e00040-11. doi:10.1128/mBio.00040-11.

**Editor** Fernando Baquero, Ramón y Cajal University Hospital



Address correspondence to W. P. Hanage, whanage@hsph.harvard.edu.

The pneumococcus (*Streptococcus pneumoniae*) is a major pediatric pathogen, estimated to be responsible for mortality in excess of 800,000 per annum in children under 5 years and a particular threat in sub-Saharan Africa, where the burden of disease is concentrated (1). Despite this sobering toll, the vast majority of the pneumococcal population is found in asymptomatic nasopharyngeal carriage, which is responsible for almost all transmission and is a prerequisite for the development of invasive disease (2). It is known that pneumococcal serotypes vary in their propensity to cause invasive disease, with some (notably serotypes 1 and 5) being associated with a considerably higher invasive potential (3–7). Samples collected from invasive disease are therefore not representative of the carried population. Moreover, the pneumococcal population is diverse, and serotypes are frequently composed of multiple distinct lineages (8, 9).

To date, few data exist on the carriage population of pneumococci in sub-Saharan Africa, despite the great mortality in this region due to this pathogen. It is known that some serotypes that are comparatively rare in European or North American settings cause a greater proportion of disease in Africa (10), and the serotypes found in carriage are expected to differ accordingly. However, it is possible that the lineages making up those serotypes also found commonly in Europe and North America could be similar to those found in those settings or quite distinct. For example, it has been shown that serotype 1 is divided into two well-supported relatively distantly related clades, one of which is associated with North America and the United Kingdom and the other of which is associated with Africa and Europe (11). Similarly, sequence type (ST) 458 has been shown to be responsible for a high proportion of serotype 3 cases of invasive pneumococcal disease (IPD) in South Africa (12), whereas in other settings the distantly related ST 180 is by far the most common clone in this serotype.

Another feature of pneumococcal carriage in sub-Saharan Africa is the comparatively high rate of colonization with multiple serotypes or strains (13, 14). In combination with the very high carriage rate, it is expected that this could offer more opportunities for homologous recombination between pneumococcal strains. The pneumococcus is naturally competent, and recombination is known to be an important factor in its population genetics (9). Recombination can have major clinical consequences by spreading antibiotic resistance (15) and enabling serotype switching, in which the capsular biosynthetic locus is replaced with the corresponding locus from another serotype (16). Serotype switching has been reported to allow vaccine escape by some important resistant strains (17–19). Though it has been proposed that some pneumococcal strains are more likely than others to be involved in recombination (20), whether there is geographical variation in the genotypes showing a history of recombination is largely unknown. The degree to which recombination can contribute to the diversification of pneumococcal clones has recently been illuminated by the publication of 240 whole genomes of a single lineage, ST 81 (21). Within a global sample, 74% of the genome had been involved in recombination in at least one lineage. At present, it is not clear whether this is true for all pnemococcal lineages or indeed whether certain settings might lead to yet more recombination. However, it illustrates well the contribution made by homologous recombination to pneumococcal diversification.

Conjugate vaccines against a subset of pneumococcal serotypes are becoming available to some of those countries with high carriage rates that have borne the brunt of pneumococcal morbidity and mortality, among them the Gambia (where a seven-valent vaccine was introduced in August 2009). The population structure of pneumococci in high-carriage environments such as the Gambia has not been extensively studied to date. High rates of multiple carriage may feasibly have consequences for the extent of homologous recombination, of relevance to the spread of antibiotic resistance and vaccine escape, as described above. We present a study of carried pneumococci in children from the Gambia. To characterize this sample, to compare it with the carriage populations in other countries, and to detect recombination, we have applied multilocus sequence typing (MLST).

## RESULTS

**Serotypes.** One hundred forty-eight pneumococcal isolates were collected from nasopharyngeal carriage in children under 2 years of age in the Upper River Division, the Gambia. The results of MLST and serotyping are shown in Table 1, together with the results of subsequent microarray analysis. The most common serotype was 6A (50/148). Microarray analysis was used to confirm the serotype of 69 isolates for which independent evidence of the observed ST/serotype combination was not available. In cases where multiple serotypes were detected by microarray analysis, the ST was assigned a serotype as described in Materials and Methods. It was not possible to determine the serotypes for three isolates (PNC 179, PNC 31, and PNC 203). It is possible that some of those isolates originally classified as 6A were the recently described serotype 6C. Microarray analysis of DNA from 17 of these isolates (representing 14 different STs) confirmed them all as consistent with 6A (Table 1). No instances of 6C were found. After 6A, the next most common serotypes were 19F (18/148), 19A (14/148), 23F (11/148), and 14 (8/148), which pooled together make up 68% of the data set. Notably, these are all relatively common serotypes in samples from unvaccinated subjects in the United States and Europe (for examples, see references 22 to 24).

**MLST.** The STs for all samples are presented along with the serotype data in Table 1. A total of 76 different STs were found, the most common of which were 802 and 919, which were associated mainly with 23F and 6A serotypes, respectively. Only 11 STs in the sample were present in the MLST database from non-African sources at the time of writing. Six STs (172, 802, 847, 1730, 1736, and 1737) were found expressing more than one serotype, indicating a history of serotype switching. STs 802 and 847 expressed both 19A and 23F, STs 1730 and ST 1736 expressed both 6A and 6B, ST 172 expressed 23F and 19F, and ST 1737 expressed 6A and 24A capsules.

**Multiple carriage.** Of the 69 samples selected for microarray analysis, 13 (19%) were found to contain evidence for colonization by more than one serotype, reflecting the high rate of multiple carriage in this setting. This was despite selection of a single colony for serotyping and DNA extraction. The observed ST was in almost all cases concordant with the majority serotype according to microarray analysis (Table 1). In cases where it was not, the sample was not included in further analyses.

**eBURST analysis.** STs were grouped into clonal complexes (CCs) of closely related strains in a "population snapshot" of the entire sample (Fig. 1) using eBURST. The largest clonal complex found in the data contained ST 912. Table 1 shows that almost half the 6A isolates in the sample are in the clonal complex descending from ST 912 (24/50, 48%). Most STs in the sample are singletons, or STs with no close relatives in the sample. This reflects a highly diverse population and contrasts with carriage samples from settings in the United States and Europe.

As noted above, the specific STs present are also different and divergent from those collected in other settings. Even among serotypes such as 19F, 23F, 19A, and 6A, which are common in both this sample and samples from the United States and Europe, the vast majority of STs shown in Table 1 have not been found in the United States or Europe despite intensive sampling in these regions. The divergence is shown in Fig. 2, in which the sample is compared with the contents of the entire MLST database (as of 14 April 2010). The STs found in the samples described here are shown surrounded by a pink identifier. The majority are not associated with the major clonal complexes, which are easily identified. The 6A clonal complex found in this work descending from ST 912 (visible on the edge of the diagram at one o'clock) is not found to be associated with any larger clonal complex or group of STs. ST 63 is found in association with a serotype 14 capsule, which is consistent with other studies of the carried pneumococcal population in the Gambia (8). However, ST 63, in samples from the United Kingdom and Europe, is typically associated with a 15A capsule (for example, see reference 25).

**Antibiotic resistance.** The microarray that was used to ascertain serotypes is also able to detect the genes associated with some antibiotic resistance determinants. The presence or absence of the *aphA3*, *cat*, *ermB*, *ermC*, *mefA*, *sat4*, *tetK*, *tetL*, *tetM*, and *tetO* loci is shown in Table 1. Of the 69 isolates that were selected for further testing to confirm serotype, 41 (59%) were found to harbor one of the above loci. The most common determinant was *tetM*, which was found in all but one of these, in five cases along with *cat* and in one with *tetL*. A single example of an isolate carrying *cat* and no other resistance loci was detected.

**TABLE 1** MLST serotype, microarray, and admixture results

| Strain name | ST | *aroE* | *gdh* | *gki* | *recP* | *spi* | *xpt* | *ddl* | Serotype[a] | Comment[b] | Resistance locus or loci[c] | Significantly admixed[d] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PNC 64 | 458 | 2 | 32 | 9 | 47 | 6 | 21 | 17 | 3 | | ND | No |
| PNC 14 | 923 | 13 | 5 | 1 | 1 | 15 | 1 | 18 | 13 | Serotype verified by microarray | − | No |
| PNC 54 | 2180 | 50 | 9 | 123 | 10 | 15 | 167 | 74 | 13 | Serotype verified by microarray | *tetM* | No |
| PNC 74 | 63 | 2 | 5 | 36 | 12 | 17 | 21 | 14 | 14 | | ND | No |
| PNC 23 | 63 | 2 | 5 | 36 | 12 | 17 | 21 | 14 | 14 | | ND | No |
| PNC 92 | 63 | 2 | 5 | 36 | 12 | 17 | 21 | 14 | 14 | Serotype verified by microarray | *tetM* | No |
| PNC 31 | 908 | 2 | 5 | 36 | 12 | 2 | 21 | 123 | 14 | Serotype verified by microarray | *tetM* | No |
| PNC 285 | 915 | 6 | 60 | 4 | 5 | 27 | 123 | 6 | 14 | Serotype verified by microarray | − | No |
| PNC 311 | 915 | 6 | 60 | 4 | 5 | 27 | 123 | 6 | 14 | | ND | No |
| PNC 327 | 915 | 6 | 60 | 4 | 5 | 27 | 123 | 6 | 14 | | ND | No |
| PNC 347 | 915 | 6 | 60 | 4 | 5 | 27 | 123 | 6 | 14 | | ND | No |
| PNC 55 | 2169 | 7 | 5 | 4 | 1 | 1 | 20 | 11 | 21 | Serotype verified by microarray | − | No |
| PNC 332 | 1746 | 10 | 13 | 4 | 16 | 15 | 1 | 6 | 21 | | ND | No |
| PNC 294 | 1745 | 8 | 9 | 52 | 5 | 10 | 122 | 11 | 28 | | ND | Yes |
| PNC 283 | 1778 | 2 | 5 | 54 | 38 | 27 | 88 | 6 | 34 | Serotype verified by microarray | − | − |
| PNC 9 | 1788 | 5 | 5 | 4 | 50 | 6 | 1 | 18 | 34 | | ND | − |
| PNC 66 | 2159 | 5 | 15 | 4 | 16 | 6 | 1 | 6 | 38 | Serotype verified by microarray | *tetM* | No |
| PNC 69 | 2159 | 5 | 15 | 4 | 16 | 6 | 1 | 6 | 38 | Serotype verified by microarray | *tetM* | No |
| PNC 189 | 2159 | 5 | 15 | 4 | 16 | 6 | 1 | 6 | 38 | Serotype verified by microarray | *tetM* | No |
| PNC 89 | 393 | 10 | 43 | 41 | 18 | 13 | 49 | 6 | 38 | Serotype verified by microarray | − | No |
| PNC 237 | 909 | 2 | 42 | 2 | 1 | 6 | 19 | 20 | 10F | Serotype verified by microarray | − | No |
| PNC 236 | 909 | 2 | 42 | 2 | 1 | 6 | 19 | 20 | 10F | | ND | No |
| PNC 287 | 989 | 12 | 5 | 89 | 8 | 6 | 112 | 14 | 12F | | ND | Yes |
| PNC 338 | 989 | 12 | 5 | 89 | 8 | 6 | 112 | 14 | 12F | | ND | Yes |
| PNC 73 | 989 | 12 | 5 | 89 | 8 | 6 | 112 | 14 | 12F[e] | 12F (88%) + 15C (12%) | *cat* + *tetM* | Yes |
| PNC 81 | 916 | 7 | 5 | 4 | 4 | 15 | 20 | 14 | 15A | | ND | No |
| PNC 96 | 917 | 7 | 5 | 4 | 4 | 15 | 20 | 5 | 15A | | ND | No |
| PNC 204 | 1738 | 7 | 5 | 1 | 4 | 15 | 20 | 5 | 15A | | ND | No |
| PNC 38 | 1727 | 60 | 82 | 4 | 4 | 6 | 1 | 18 | 15B | Serotype verified by microarray | *cat* + *tetM* | No |
| PNC 208 | 1748 | 12 | 111 | 9 | 84 | 6 | 83 | 74 | 16F | Serotype verified by microarray | − | Yes |
| PNC 244 | 927 | 23 | 12 | 4 | 1 | 43 | 12 | 74 | 16F | Serotype verified by microarray | − | Yes |
| PNC 250 | 924 | 13 | 8 | 69 | 1 | 6 | 1 | 6 | 17F | Serotype verified by microarray | − | No |
| PNC 18 | 924 | 13 | 8 | 69 | 1 | 6 | 1 | 6 | 17F | | ND | No |
| PNC 19 | 1233 | 10 | 11 | 34 | 16 | 15 | 1 | 145 | 18C | Serotype verified by microarray | *tetM* | No |
| PNC 30 | 1233 | 10 | 11 | 34 | 16 | 15 | 1 | 145 | 18C | Serotype verified by microarray | *tetM* | No |
| PNC 325 | 2167 | 7 | 5 | 1 | 2 | 6 | 112 | 6 | 19A | Serotype verified by microarray | − | No |
| PNC 345 | 2168 | 7 | 5 | 4 | 1 | 6 | 112 | 14 | 19A | | ND | No |
| PNC 336 | 847 | 7 | 11 | 4 | 1 | 6 | 112 | 14 | 19A | Serotype verified by microarray | − | No |
| PNC 339 | 847 | 7 | 11 | 4 | 1 | 6 | 112 | 14 | 19A | | ND | No |
| PNC 1 | 847 | 7 | 11 | 4 | 1 | 6 | 112 | 14 | 19A | | ND | No |
| PNC 29 | 847 | 7 | 11 | 4 | 1 | 6 | 112 | 14 | 19A | | ND | No |
| PNC 39 | 847 | 7 | 11 | 4 | 1 | 6 | 112 | 14 | 19A | | ND | No |
| PNC 211 | 847 | 7 | 11 | 4 | 1 | 6 | 112 | 14 | 19A | | ND | No |
| PNC 320 | 847 | 7 | 11 | 4 | 1 | 6 | 112 | 14 | 19A | | ND | No |
| PNC 292 | 921 | 7 | 78 | 4 | 1 | 6 | 112 | 14 | 19A | Serotype verified by microarray | − | No |
| PNC 302 | 921 | 7 | 78 | 4 | 1 | 6 | 112 | 14 | 19A | Serotype verified by microarray | − | No |
| PNC 329 | 1787 | 8 | 13 | 7 | 16 | 36 | 173 | 6 | 19A | Serotype verified by microarray | − | − |
| PNC 111 | 802 | 10 | 13 | 53 | 1 | 72 | 38 | 31 | 19A | | ND | Yes |
| PNC 45 | 802 | 10 | 13 | 53 | 1 | 72 | 38 | 31 | 19A | | ND | Yes |
| PNC 301 | 1786 | 2 | 5 | 4 | 10 | 17 | 1 | 198 | 19F | Serotype verified by microarray | *tetM* | No |
| PNC 227 | 2170 | 7 | 5 | 4 | 4 | 6 | 20 | 14 | 19F | Serotype verified by microarray | − | No |
| PNC 112 | 172 | 7 | 13 | 8 | 6 | 25 | 6 | 8 | 19F | Serotype verified by microarray | *tetM* | Yes |
| PNC 219 | 925 | 15 | 16 | 19 | 15 | 6 | 20 | 19 | 19F | | ND | No |
| PNC 42 | 925 | 15 | 16 | 19 | 15 | 6 | 20 | 19 | 19F | | ND | No |
| PNC 61 | 925 | 15 | 16 | 19 | 15 | 6 | 20 | 19 | 19F | | ND | No |
| PNC 34 | 925 | 15 | 16 | 19 | 15 | 6 | 20 | 19 | 19F | Serotype verified by microarray | *tetM* | No |
| PNC 334 | 926 | 15 | 16 | 19 | 15 | 6 | 20 | 14 | 19F | | ND | No |
| PNC 221 | 926 | 15 | 16 | 19 | 15 | 6 | 20 | 14 | 19F | | ND | No |
| PNC 272 | 928 | 53 | 76 | 8 | 15 | 6 | 20 | 19 | 19F | Serotype verified by microarray | *tetM* | No |
| PNC 276 | 928 | 53 | 76 | 8 | 15 | 6 | 20 | 19 | 19F | Serotype verified by microarray | *tetM* | No |
| PNC 331 | 929 | 53 | 76 | 19 | 15 | 6 | 20 | 19 | 19F | | ND | No |
| PNC 41 | 929 | 53 | 76 | 19 | 15 | 6 | 20 | 19 | 19F | | ND | No |
| PNC 303 | 929 | 53 | 76 | 19 | 15 | 6 | 20 | 19 | 19F | | ND | No |
| PNC 333 | 929 | 53 | 76 | 19 | 15 | 6 | 20 | 19 | 19F | | ND | No |
| PNC 21 | 929 | 53 | 76 | 19 | 15 | 6 | 20 | 19 | 19F | Serotype verified by microarray | *tetM* | No |

*Continued on following page*

**TABLE 1**—*Continued*

| Strain name | ST | *aroE* | *gdh* | *gki* | *recP* | *spi* | *xpt* | *ddl* | Serotype[a] | Comment[b] | Resistance locus or loci[c] | Significantly admixed[d] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PNC 35 | 929 | 53 | 76 | 19 | 15 | 6 | 20 | 19 | 19F | Serotype verified by microarray | *tetM* | No |
| PNC 5 | 929 | 53 | 76 | 19 | 15 | 6 | 20 | 19 | 19F | Serotype verified by microarray | *tetM* | No |
| PNC 228 | 1777 | 5 | 5 | 8 | 88 | 9 | 1 | 31 | 20[e] | 20 (94%) + 19F (5%) + 9L (1%) | *tetM* (19F) | − |
| PNC 32 | 1746 | 10 | 13 | 4 | 16 | 15 | 1 | 6 | 21[e] | 21 (93%) + 19A (7%) | *tetL* + *tetM* | No |
| PNC 53 | 910 | 5 | 5 | 6 | 5 | 9 | 17 | 19 | 22A | Serotype verified by microarray | − | Yes |
| PNC 312 | 802 | 10 | 13 | 53 | 1 | 72 | 38 | 31 | 23A | | ND | Yes |
| PNC 214 | 2160 | 6 | 5 | 5 | 5 | 27 | 3 | 5 | 23F | | ND | No |
| PNC 258 | 2160 | 6 | 5 | 5 | 5 | 27 | 3 | 5 | 23F | | ND | No |
| PNC 179 | 847 | 7 | 11 | 4 | 1 | 6 | 112 | 14 | 23F | | ND | No |
| PNC 107 | 172 | 7 | 13 | 8 | 6 | 25 | 6 | 8 | 23F | | ND | Yes |
| PNC 234 | 2174 | 7 | 16 | 8 | 8 | 6 | 142 | 14 | 23F | | ND | No |
| PNC 233 | 802 | 10 | 13 | 53 | 1 | 72 | 38 | 31 | 23F | | ND | Yes |
| PNC 243 | 802 | 10 | 13 | 53 | 1 | 72 | 38 | 31 | 23F | | ND | Yes |
| PNC 255 | 802 | 10 | 13 | 53 | 1 | 72 | 38 | 31 | 23F | | ND | Yes |
| PNC 44 | 802 | 10 | 13 | 53 | 1 | 72 | 38 | 31 | 23F | Serotype verified by microarray | *tetM* | Yes |
| PNC 47 | 802 | 10 | 13 | 53 | 1 | 72 | 38 | 31 | 23F | Serotype verified by microarray | *tetM* | Yes |
| PNC 65 | 802 | 10 | 13 | 53 | 1 | 72 | 38 | 31 | 23F[e] | 23F (74%) + 38 (26%) | *tetM* | Yes |
| PNC 46 | 1737 | 6 | 57 | 34 | 28 | 6 | 1 | 9 | 24A | Serotype verified by microarray | *tetM* | Yes |
| PNC 75 | 1747 | 10 | 16 | 54 | 10 | 7 | 1 | 31 | 35A | Serotype verified by microarray | − | Yes |
| PNC 254 | 1740 | 7 | 13 | 4 | 5 | 6 | 122 | 18 | 35B | | ND | No |
| PNC 43 | 2157 | 1 | 8 | 36 | 5 | 9 | 60 | 14 | 6A | Serotype verified by microarray | *tetM* | No |
| PNC 217 | 1730 | 2 | 5 | 4 | 10 | 17 | 1 | 8 | 6A | | ND | No |
| PNC 307 | 1730 | 2 | 5 | 4 | 10 | 17 | 1 | 8 | 6A | | ND | No |
| PNC 289 | 2161 | 6 | 15 | 34 | 28 | 6 | 1 | 9 | 6A | | ND | No |
| PNC 71 | 911 | 6 | 57 | 34 | 58 | 6 | 1 | 9 | 6A | | ND | Yes |
| PNC 80 | 912 | 6 | 57 | 34 | 28 | 7 | 1 | 9 | 6A | | ND | Yes |
| PNC 82 | 912 | 6 | 57 | 34 | 28 | 7 | 1 | 9 | 6A | | ND | Yes |
| PNC 94 | 912 | 6 | 57 | 34 | 28 | 7 | 1 | 9 | 6A | | ND | Yes |
| PNC 212 | 912 | 6 | 57 | 34 | 28 | 7 | 1 | 9 | 6A | | ND | Yes |
| PNC 215 | 912 | 6 | 57 | 34 | 28 | 7 | 1 | 9 | 6A | | ND | Yes |
| PNC 220 | 912 | 6 | 57 | 34 | 28 | 7 | 1 | 9 | 6A | Serotype verified by microarray | *tetM* | Yes |
| PNC 253 | 913 | 6 | 57 | 83 | 28 | 7 | 19 | 9 | 6A | | ND | Yes |
| PNC 256 | 914 | 6 | 57 | 83 | 28 | 7 | 1 | 9 | 6A | | ND | No |
| PNC 279 | 914 | 6 | 57 | 83 | 28 | 7 | 1 | 9 | 6A | Serotype verified by microarray | *tetM* | No |
| PNC 11 | 1736 | 6 | 57 | 34 | 1 | 7 | 1 | 74 | 6A | | ND | Yes |
| PNC 218 | 1736 | 6 | 57 | 34 | 1 | 7 | 1 | 74 | 6A | | ND | Yes |
| PNC 242 | 1736 | 6 | 57 | 34 | 1 | 7 | 1 | 74 | 6A | | ND | Yes |
| PNC 180 | 1736 | 6 | 57 | 34 | 1 | 7 | 1 | 74 | 6A | | ND | Yes |
| PNC 223 | 1737 | 6 | 57 | 34 | 28 | 6 | 1 | 9 | 6A | Serotype verified by microarray | *cat* + *tetM* | Yes |
| PNC 6 | 1737 | 6 | 57 | 34 | 28 | 6 | 1 | 9 | 6A | | ND | Yes |
| PNC 37 | 1737 | 6 | 57 | 34 | 28 | 6 | 1 | 9 | 6A | | ND | Yes |
| PNC 67 | 1737 | 6 | 57 | 34 | 28 | 6 | 1 | 9 | 6A | | ND | Yes |
| PNC 298 | 2162 | 6 | 57 | 34 | 1 | 7 | 1 | 8 | 6A | | ND | Yes |
| PNC 308 | 2163 | 6 | 57 | 34 | 28 | 6 | 1 | 19 | 6A | | ND | Yes |
| PNC 309 | 2164 | 6 | 57 | 34 | 28 | 7 | 1 | 18 | 6A | Serotype verified by microarray | *cat* + *tetM* | Yes |
| PNC 310 | 2165 | 6 | 57 | 34 | 28 | 7 | 1 | 8 | 6A | | ND | Yes |
| PNC 323 | 2166 | 6 | 57 | 34 | 28 | 7 | 1 | 14 | 6A | | ND | Yes |
| PNC 2 | 2171 | 7 | 8 | 1 | 2 | 15 | 60 | 14 | 6A | Serotype verified by microarray | *tetM* | No |
| PNC 10 | 2173 | 7 | 13 | 1 | 5 | 42 | 60 | 31 | 6A | | ND | No |
| PNC 12 | 2173 | 7 | 13 | 1 | 5 | 42 | 60 | 31 | 6A | | ND | No |
| PNC 206 | 2173 | 7 | 13 | 1 | 5 | 42 | 60 | 31 | 6A | | ND | No |
| PNC 28 | 919 | 7 | 25 | 29 | 2 | 27 | 122 | 122 | 6A | Serotype verified by microarray | − | No |
| PNC 27 | 919 | 7 | 25 | 29 | 2 | 27 | 122 | 122 | 6A | Serotype verified by microarray | − | No |
| PNC 24 | 919 | 7 | 25 | 29 | 2 | 27 | 122 | 122 | 6A | Serotype verified by microarray | − | No |
| PNC 239 | 919 | 7 | 25 | 29 | 2 | 27 | 122 | 122 | 6A | | ND | No |
| PNC 240 | 919 | 7 | 25 | 29 | 2 | 27 | 122 | 122 | 6A | | ND | No |
| PNC 319 | 919 | 7 | 25 | 29 | 2 | 27 | 122 | 122 | 6A | | ND | No |
| PNC 328 | 919 | 7 | 25 | 29 | 2 | 27 | 122 | 122 | 6A | | ND | No |
| PNC 4 | 919 | 7 | 25 | 29 | 2 | 27 | 122 | 122 | 6A | | ND | No |
| PNC 321 | 919 | 7 | 25 | 29 | 2 | 27 | 122 | 122 | 6A | | ND | No |
| PNC 198 | 920 | 7 | 25 | 29 | 2 | 27 | 112 | 122 | 6A | | ND | No |
| PNC 281 | 1741 | 7 | 25 | 29 | 2 | 112 | 172 | 122 | 6A | Serotype verified by microarray | − | Yes |
| PNC 286 | 1785 | 7 | 25 | 29 | 1 | 36 | 122 | 197 | 6A | Serotype verified by microarray | − | No |
| PNC 306 | 2175 | 7 | 25 | 29 | 1 | 36 | 122 | 18 | 6A | Serotype verified by microarray | − | No |

*Continued on following page*

**TABLE 1**—*Continued*

| Strain name | ST | *aroE* | *gdh* | *gki* | *recP* | *spi* | *xpt* | *ddl* | Serotype[a] | Comment[b] | Resistance locus or loci[c] | Significantly admixed[d] |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| PNC 282 | 1742 | 7 | 57 | 83 | 28 | 7 | 1 | 8 | 6A | Serotype verified by microarray | *cat* + *tetM* | Yes |
| PNC 225 | 1750 | 75 | 6 | 1 | 2 | 6 | 1 | 14 | 6A | | ND | No |
| PNC 245 | 912 | 6 | 57 | 34 | 28 | 7 | 1 | 9 | 6A[e] | 6A (87%) + 19F (13%) | *tetM* (6A) | Yes |
| PNC 249 | 912 | 6 | 57 | 34 | 28 | 7 | 1 | 9 | 6A[e] | 6A (84%) + 19F (16%) | *tetM* (6A) + *cat* (19F) | Yes |
| PNC 7 | 2172 | 7 | 8 | 54 | 6 | 15 | 60 | 18 | 6A[e] | 6A (79%) + 19F (21%) | *tetM* | Yes |
| PNC 291 | 920 | 7 | 25 | 29 | 2 | 27 | 112 | 122 | 6A[e] | 6A (93%) + NT3b (7%) | *cat* (NT3b) | No |
| PNC 203 | 1730 | 2 | 5 | 4 | 10 | 17 | 1 | 8 | 6B | Serotype verified by microarray | *tetM* | No |
| PNC 25 | 273 | 5 | 6 | 1 | 2 | 6 | 1 | 14 | 6B | | ND | No |
| PNC 251 | 1736 | 6 | 57 | 34 | 1 | 7 | 1 | 74 | 6B | | ND | Yes |
| PNC 190 | 2178 | 10 | 20 | 14 | 1 | 6 | 20 | 29 | 7F | Serotype verified by microarray | *tetM* | Yes |
| PNC 36 | 1735 | 6 | 5 | 2 | 5 | 36 | 1 | 5 | 9L[e] | 9L (95%) + 6A (5%) | *tetM* (6A) | No |
| PNC 324 | 1871 | 1 | 16 | 4 | 10 | 7 | 1 | 31 | 9V | Serotype verified by microarray | *tetM* | No |
| PNC 50 | 1871 | 1 | 16 | 4 | 10 | 7 | 1 | 31 | 9V | Serotype verified by microarray | *tetM* | No |
| PNC 230 | 2179 | 15 | 17 | 4 | 4 | 6 | 1 | 17 | 9V[e] | 9V (86%) + 6A (14%) | − | No |
| PNC 57 | 1728 | 8 | 77 | 84 | 1 | 10 | 171 | 72 | NT3a | Serotype verified by microarray | − | Yes |
| PNC 40 | 922 | 8 | 5 | 15 | 10 | 2 | 1 | 59 | NT3b | Serotype verified by microarray | − | No |
| PNC 13 | 918 | 7 | 13 | 8 | 6 | 9 | 6 | 14 | U | 46 (66%) + 21 (34%) | − | Yes |
| PNC 290 | 2176 | 8 | 5 | 2 | 16 | 1 | 26 | 1 | U | 46 (59%) + 21 (41%) | − | No |
| PNC 60 | 2177 | 8 | 5 | 54 | 10 | 7 | 1 | 168 | U | 22A (56%) + 11D (44%) | *tetM* | Yes |

[a] Serotype was determined as described in Materials and Methods.
[b] Indicates whether microarray was used to confirm serotype. Where signal from more than one capsule was found, the proportions are identified.
[c] The presence of resistance loci and (if relevant) which serotype they were associated with. −, no resistance loci were identified. ND, not done.
[d] Whether the ST was previously identified (20) as showing significant admixture. −, data not available.
[e] Isolates which were found to be mixed as described in Comment column. U, serotype was unclassifiable.

**Recombination.** To quantify the contribution made by homologous recombination to the carried population in the Gambia, the STs were compared with the previously published results of an analysis of the whole pneumococcal MLST database using the BAPS package (20). This analysis identifies STs with evidence of admixture—meaning in this case recombination. The results are shown in Table 1. Four STs (1777, 1778, 1787, and 1788) were not present in the MLST database at the time of the original BAPS analysis and have been excluded. Forty-nine of 144 (34%) isolates in the Gambian sample showed significant ($P < 0.05$) evidence of admixture.

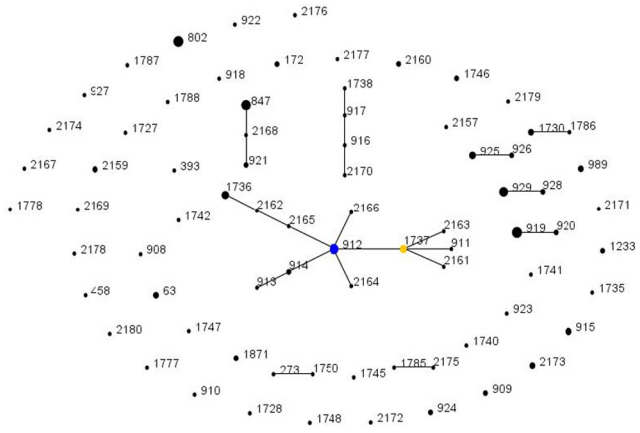


**FIG 1** Population snapshot of the Gambian carriage sample. Each ST is represented by a point, the size of which is determined by the number of isolates with that ST in the combined data set. STs differing at a single genetic locus are linked by a straight line. A clonal complex (CC) is a group of STs sharing 6 of 7 alleles with at least one other member of the group. The putative ancestral ST of each CC is shown in blue, with subgroup ancestors shown in yellow. Those STs that cannot be linked to any other in the sample are termed singletons and appear as unlinked points. For more information see http://spneumoniae.mlst.net/eburst/.

## DISCUSSION

Despite several recent publications studying the pneumococcal population in sub-Saharan Africa (8, 13, 14, 26, 27), this region remains undersampled in general in relation to the amounts of invasive disease that occur there (1). In this paper, we present a sample from children under 2 years of age from the Gambia. All isolates have been characterized using MLST, which allows us to examine relationships between strains in the data set and also relationships with the MLST database, which contains isolates from many other countries. This is to date the largest carriage sample from sub-Saharan Africa to have been typed by MLST and hence reflects one of the best insights into the population structure of this important pathogen in the region where it exacts its greatest toll.

One of the most striking findings of this work is the difference between the STs found in the samples here and the more intensively sampled locations in the United States and Europe. This can be illustrated by comparison with the MLST database, which records MLST and epidemiological data from researchers in multiple labs and continents (http://spneumoniae.mlst.net/). Because submission of MLST data is voluntary, the database is far from an ideal sample of pneumococcal global diversity, but it nevertheless allows us to ask whether an ST or lineage is sufficiently common to have been sampled. The vast majority of records reflect the majority of pneumococcal research from developed countries, with 5,102 of 7,278 records for which location data are available at the time of writing being from Europe or North America. The diverse nature of the sample set that we describe here, in contrast with the large clonal complexes that make up the majority of the MLST database, is evident in Fig. 2. Also evident is the high frequency of 6A in this sample, in the largest clonal complex and several unre-

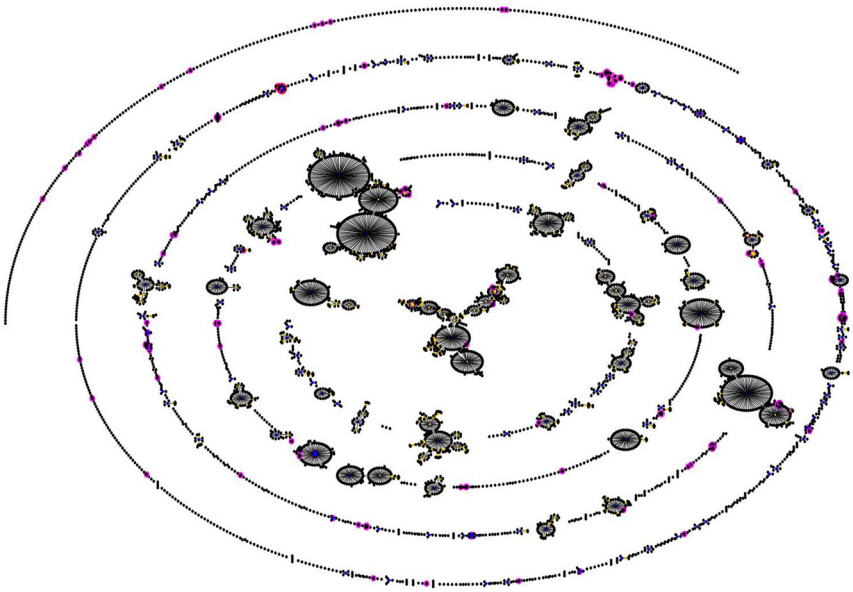

**FIG 2** Comparison of the STs shown in Figure 1 with the whole MLST data set (as of 20 April 2010). Clonal complexes are defined as described above. STs found in this work are shown surrounded by pink.

lated (singleton) lineages (Fig. 1). This observation suggests that 6A may have been common for some time, in order to have generated such diversity.

We have also found evidence of a high prevalence of resistance to at least some antibiotics. It is not possible to determine the precise prevalence of the resistance determinants detected by the microarray, because the isolates tested were chosen not for this purpose but in order to verify their serotype and as a result are a biased sample of the carriage population. However, it is reasonable to conclude that *tetM* prevalence is very high. If we assume that resistance is predominantly a clonal property, and that all isolates with STs which were found to be *tetM*$^+$ also contained this locus, then we can estimate that at least 68/148 of the Gambian carriage samples were resistant to tetracycline. This takes no account of other resistance determinants, nor of the possibility that other STs, closely related to those found with the *tetM* locus, are also resistant. With this in mind, it should be noted that ST 912, the most common ST in the data set and the inferred ancestor of the only clonal complex within it (Fig. 1), was found to carry *tetM*. It may be counted as surprising that the combination of *ermB* with *tetM*, frequent in other samples, was not noted here. At present, we cannot be sure whether this was due to the absence of the *ermB* resistance locus or the presence of a highly divergent *ermB* locus that could not be detected by our methods. The nature of the region harboring *tetM* resistance in these strains is currently being investigated.

From the data presented here, we cannot conclude whether the prevalence and wide phylogenetic distribution of these resistance determinants are associated with recombination, but we have good evidence that recombination has made a considerable contribution to the Gambian pneumococcal population. This is particularly evident if we compare previously published carriage data sets from Finland (6) and Oxford, United Kingdom (23). In the Oxford data set, 14/264 (5%) isolates were associated with an inferred history of admixture, while this was true for just 4/217 (2%) isolates in the Finnish sample. This should be compared with 49/144 (34%) isolates in this sample. That recombination in sub-Saharan Africa might be more common than elsewhere is perhaps not surprising given the high rates of carriage and the relatively high frequency with which multiple serotypes are carried in this setting, as directly if inadvertently detected in 13 of 69 (19%) samples subjected to microarray analysis. In contrast, in a recent study using the microarray in combination with other methods to examine multiple carriage among Swiss outpatients, only 7.9% of those carrying pneumococci showed evidence of more than one serotype being present (28). However, we emphasize that the samples and methods used in that study were not identical to those that we employed. In those cases of multiple carriage detected in our data set, a clear MLST profile was determined despite the clear presence of DNA from more than one serotype. While it might be expected that sequencing would detect mixed cultures as generating mixed sequence, the ability of sequencing to detect the presence of minority DNA has not been systematically studied, and this may suggest that mixed DNA samples in other data sets are more common than is immediately apparent from MLST analysis.

In order for strains to undergo recombination, they must cocolonize the same host, and the high rate of multiple carriage is expected to provide more opportunities for this to occur. As noted below, the STs found in this paper (and others) are divergent from those that have been recorded to date. This may be partially a consequence of the high rate of recombination, which will generate novel combinations of the alleles at the MLST loci. As noted by Croucher et al. (21), recombination in the pneumococcus is capable of generating considerable adaptation over a time scale of a few decades, and within high-carriage settings, this could be enhanced by more opportunities for homologous recombination.

Examining the MLST database, we note that the most common ST in this sample (ST 919) has been recorded only in the Gambia, as is also the case for ST 912 (the predicted ancestor of the largest clonal complex shown in Fig. 1 and divergent from STs in the

MLST database as shown in Fig. 2). Similarly striking is the observation that the common 23F clone ST 802 has been recorded only once in the United States or Europe. At the time of writing, 26 isolates of ST 802 are present in the MLST database. Of these, 13 are African in origin, being from other studies in the Gambia or Niger (http://spneumoniae.mlst.net). However, it is particularly striking that 10 additional records (at the time of writing) of ST 802 in carriage come from studies in the Mae La refugee camp in Thailand (http://spneumoniae.mlst.net). A direct epidemiological link between these settings is hard to envisage. That a clone should be common in such distant locations suggests the intriguing possibility that some pneumococcal lineages are adapted to settings other than those that have been extensively studied to date. Certain host genotypes predispose to invasive pneumococcal disease (29), but no association is known between host genotype and carriage. Figure 2 shows that the STs found in this work are comparatively divergent from the rest of the MLST database, and other MLST studies on strains from the Gambia have also found a large proportion of STs that had not previously been recorded (8); the vast majority of pneumococcal morbidity and mortality falls in locations such as sub-Saharan Africa. If the pneumococcal population found in these regions is distinct, it is possible that it might also differ in terms of invasiveness. Whether or not this is the case can only be resolved by improved sampling of carriage and disease from those locations where most pneumococcal disease occurs.

## MATERIALS AND METHODS

**Strains.** Nasopharyngeal swab samples were collected from children under 2 years of age in the Upper River Division, the Gambia. While the children themselves were unimmunized, prior to delivery their mothers had received a 9-valent pneumococcal conjugate vaccine (30). Pneumococci were identified on gentamicin (5 mg/ml) sheep blood agar by their morphological characteristics and optochin sensitivity. Serotyping was performed on site, with capsular and factor typing sera (Statens Serum Institut) using the latex agglutination technique (26). Isolates with equivocal serotyping results were rechecked by Quellung reaction. A single colony was picked, and DNA was extracted for MLST as previously described (31). The Medical Research Council (MRC) microbiology laboratory is enrolled in the external quality assurance program of the United Kingdom National External Quality Assessment Scheme (http://www.ukneqas.org/uk). The study was approved by the Joint Gambian Government and MRC Ethics Committee and the Ethics Committees of the London School of Hygiene and Tropical Medicine.

**MLST.** Sequence types (STs) of isolates were determined by MLST as previously described (31). Sequences of each of the seven gene fragments used in the pneumococcal MLST scheme were obtained on both DNA strands with an ABI 3700 DNA analyzer. The sequences were aligned and trimmed to defined start and end positions using MEGA version 4 (32). Allele and ST assignments were made using the MLST website (http://spneumoniae.mlst.net). All alleles not already present in the pneumococcal MLST database were verified by resequencing the gene fragment on both strands.

**Microarray analysis to confirm serotype.** In all cases where the reported serotype was discordant with that expected either from identical or closely related STs in the sample or the MLST database, the isolate was reserotyped. In 20 cases this was not possible because of loss of the original culture. In these cases, the DNA used for MLST was applied to a microarray designed to allow the unique combinations of genes determining capsular type to be identified (28, 33). This revealed a high rate of error in the original serotyping results, and as a result, microarray analysis was used to confirm the serotype of all MLST types for which there was no independent evidence for that serotype/ST combination (i.e., STs found only once, and with no close relatives either in the data set here or in the database). Where the microarray identified DNA from more than one capsular type (the result of a mixed culture), the isolate was recorded as the majority serotype if it was responsible for >70% of the signal. In all cases where the majority serotype was responsible for <70% of the signal, the isolate was recorded as unclassified. In addition to loci for serotype assignment, the microarray also contained the *aphA3*, *cat*, *ermB*, *ermC*, *mefA*, *sat4*, *tetK*, *tetL*, *tetM*, and *tetO* genes, allowing the presence and transfer of these antibiotic resistance genes to be assessed.

**eBURST and detection of recombination.** The population structure of the samples was analyzed using eBURST (34). This program groups related STs into clonal complexes (CCs), identifies the probable ancestor of each CC as the ST with the largest number of minor differences, and outputs a graphical representation of these relationships.

To assess the impact of recombination on the population, the STs found in this work were compared with the results of a previously published analysis of the entire MLST database using the program Bayesian Analysis of Population Structure (BAPS) (35–37). This program divides a data set into populations on the basis of allele frequencies and identifies significantly ($P < 0.05$) admixed genotypes that contain sequence characteristic of more than one population.

## ACKNOWLEDGMENTS

We acknowledge the study participants and their parents and the field team and laboratory support group in the Gambia, in particular Richard Adegbola and Kawsu Sankareh. We gratefully acknowledge the BUGs microarray facility, St. George's Hospital, for provision of the *S. pneumoniae* serotyping microarray and Brian Spratt for comments on the manuscript.

This work was funded by the award of a Royal Society University Research Fellowship to W.P.H., by a Ghanaian Education Trust Fund scholarship to E.S.D., and by the Medical Research Council UK (M.A.) and the Wellcome Trust (C.J.B.).

Regarding potential conflicts of interest, W.P.H. has acted as an advisor to GlaxoSmithKline. The other authors have no conflicts of interest to report.

## REFERENCES

1. **O'Brien KL, et al.** 2009. Burden of disease caused by *Streptococcus pneumoniae* in children younger than 5 years: global estimates. Lancet **374**: 893–902.
2. **Crook DW, Brueggemann AB, Sleeman KL, Peto TEA.** 2004. Pneumococcal carriage, p. 136–148. *In* Tuomanen EI, Mitchell TJ, Morrison DA, Spratt BG (ed.), The pneumococcus. ASM Press, Washington, DC.
3. **Adegbola RA, et al.** 2006. Serotype and antimicrobial susceptibility patterns of isolates of *Streptococcus pneumoniae* causing invasive disease in The Gambia 1996-2003. Trop. Med. Int. Health **11**:1128–1135.
4. **Brueggemann AB, et al.** 2003. Clonal relationships between invasive and carriage *Streptococcus pneumoniae* and serotype- and clone-specific differences in invasive disease potential. J. Infect. Dis. **187**:1424–1432.
5. **Brueggemann AB, et al.** 2004. Temporal and geographic stability of the serogroup-specific invasive disease potential of *Streptococcus pneumoniae* in children. J. Infect. Dis. **190**:1203–1211.
6. **Hanage WP, et al.** 2005. Invasiveness of serotypes and clones of *Streptococcus pneumoniae* among children in Finland. Infect. Immun. **73**: 431–435.
7. **Antonio M, et al.** 2008. Seasonality and outbreak of a predominant *Streptococcus pneumoniae* serotype 1 clone from The Gambia: expansion of ST217 hypervirulent clonal complex in West Africa. BMC Microbiol. **8**:198.
8. **Antonio M, et al.** 2008. Molecular epidemiology of pneumococci obtained from Gambian children aged 2-29 months with invasive pneumococcal disease during a trial of a 9-valent pneumococcal conjugate vaccine. BMC Infect. Dis. **8**:81.
9. **Spratt BG, Hanage WP, Brueggemann AB.** 2004. Evolutionary and population biology of *Streptococcus pneumoniae*, p. 119–135. *In* Tuomanen EI, Mitchell TJ, Morrison DA, Spratt BG (ed.), The pneumococcus. ASM Press, Washington, DC.
10. **Obaro S.** 2001. Differences in invasive pneumococcal serotypes. Lancet **357**:1800–1801.

11. **Brueggemann AB, Spratt BG.** 2003. Geographic distribution and clonal diversity of *Streptococcus pneumoniae* serotype 1 isolates. J. Clin. Microbiol. **41**:4966–4970.
12. **Mothibeli KM, et al.** 2009. An unusual pneumococcal sequence type is the predominant cause of serotype 3 invasive disease in South Africa. J. Clin. Microbiol. **48**:184–191.
13. **Donkor ES, et al.** 2010. Invasive disease and paediatric carriage of *Streptococcus pneumoniae* in Ghana. Scand. J. Infect. Dis. **42**:254–259.
14. **Hill PC, et al.** 2008. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian infants: a longitudinal study. Clin. Infect. Dis. **46**:807–814.
15. **Hakenbeck R.** 1998. Mosaic genes and their role in penicillin-resistant *Streptococcus pneumoniae*. Electrophoresis **19**:597–601.
16. **Coffey TJ, et al.** 1998. Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. Mol. Microbiol. **27**:73–83.
17. **Brueggemann AB, Pai R, Crook DW, Beall B.** 2007. Vaccine escape recombinants emerge after pneumococcal vaccination in the United States. PLoS Pathog. **3**:e168.
18. **Moore MR, et al.** 2008. Population snapshot of emergent *Streptococcus pneumoniae* serotype 19A in the United States, 2005. J. Infect. Dis. **197**: 1016–1027.
19. **Pelton SI, et al.** 2007. Emergence of 19A as virulent and multidrug resistant pneumococcus in Massachusetts following universal immunization of infants with pneumococcal conjugate vaccine. Pediatr. Infect. Dis. J. **26**:468–472.
20. **Hanage WP, Fraser C, Tang J, Connor TR, Corander J.** 2009. Hyperrecombination, diversity, and antibiotic resistance in pneumococcus. Science **324**:1454–1457.
21. **Croucher NJ, et al.** 2011. Rapid pneumococcal evolution in response to clinical interventions. Science **331**:430–434.
22. **Hanage WP, et al.** 2004. Ability of pneumococcal serotypes and clones to cause acute otitis media: implications for the prevention of otitis media by conjugate vaccines. Infect. Immun. **72**:76–81.
23. **Meats E, et al.** 2003. Stability of serotypes during nasopharyngeal carriage of *Streptococcus pneumoniae*. J. Clin. Microbiol. **41**:386–392.
24. **Marchisio P, et al.** 2002. Nasopharyngeal carriage of *Streptococcus pneumoniae* in healthy children: implications for the use of heptavalent pneumococcal conjugate vaccine. Emerg. Infect. Dis. **8**:479–484.
25. **Hanage WP, et al.** 2007. Diversity and antibiotic resistance among nonvaccine serotypes of *Streptococcus pneumoniae* carriage isolates in the post-heptavalent conjugate vaccine era. J. Infect. Dis. **195**:347–352.
26. **Hill PC, et al.** 2006. Nasopharyngeal carriage of *Streptococcus pneumoniae* in Gambian villagers. Clin. Infect. Dis. **43**:673–679.
27. **Hill PC, et al.** 2010. Transmission of *Streptococcus pneumoniae* in rural Gambian villages: a longitudinal study. Clin. Infect. Dis. **50**:1468–1476.
28. **Brugger SD, Frey P, Aebi S, Hinds J, Muhlemann K.** 2010. Multiple colonization with *S. pneumoniae* before and after introduction of the seven-valent conjugated pneumococcal polysaccharide vaccine. PLoS One **5**:e11638.
29. **Brouwer MC, et al.** 2009. Host genetic susceptibility to pneumococcal and meningococcal disease: a systematic review and meta-analysis. Lancet Infect. Dis. **9**:31–44.
30. **Obaro SK, et al.** 2004. Serotype-specific pneumococcal antibodies in breast milk of Gambian women immunized with a pneumococcal polysaccharide vaccine during pregnancy. Pediatr. Infect. Dis. J. **23**: 1023–1029.
31. **Enright MC, Spratt BG.** 1998. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. Microbiology **144**:3049–3060.
32. **Tamura K, Dudley J, Nei M, Kumar S.** 2007. MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. Mol. Biol. Evol. **24**:1596–1599.
33. **Hinds J, et al.** 2010. Molecular serotyping of *Streptococcus pneumoniae*: a microarray-based tool with enhanced utility. Proceedings of the 7th International Symposium on Pneumococci and Pneumococcal Diseases, Tel Aviv, Israel.
34. **Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186**: 1518–1530.
35. **Corander J, Tang J.** 2007. Bayesian analysis of population structure based on linked molecular information. Math. Biosci. **205**:19–31.
36. **Corander J, Waldmann P, Marttinen P, Sillanpää MJ.** 2004. BAPS 2: enhanced possibilities for the analysis of genetic population structure. Bioinformatics **20**:2363–2369.
37. **Corander J, Waldmann P, Sillanpää MJ.** 2003. Bayesian analysis of genetic differentiation between populations. Genetics **163**:367–374.