Microbiology (1998), 144, 3049–3060 Printed in Great Britain

A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease

Mark C. Enright and Brian G. Spratt

Author for correspondence: Brian G. Spratt. Tel: +44 1865 281301. Fax: +44 1865 310447. e-mail: brian.spratt@zoology.oxford.ac.uk

Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, Oxford OX1 3PS, UK

The population biology of *Streptococcus pneumoniae* is poorly understood. Most of the important issues could be addressed by the molecular characterization of large, well sampled populations from carriage and from the different manifestations of pneumococcal disease. The authors have therefore developed a pneumococcal multilocus sequence typing scheme and database by sequencing ~450 bp fragments of seven housekeeping loci from 295 isolates. The combination of alleles at the seven loci provided an allelic profile, or sequence type (ST), and the relatedness between isolates was obtained by constructing a dendrogram from the matrix of pairwise differences between STs. The typing scheme was validated using pneumococci of known genetic relatedness and could resolve >6 billion STs. Among 274 isolates from recent cases of invasive pneumococcal disease in eight countries, 143 STs were resolved, but 12 STs contained at least five isolates (range 5–21 isolates). The repeated recovery of indistinguishable isolates from invasive disease in different countries implies that these STs define strains with an increased capacity to cause invasive disease. The relationship between STs and serotypes suggested that, in the longer term, capsular genes have been distributed horizontally within the pneumococcal population, but in the short term, expansion of clones occurs with only occasional changes of serotype. The multilocus sequence typing scheme provides a powerful new approach to the characterization of pneumococci, since it provides molecular typing data that are electronically portable between laboratories, and which can be used to probe aspects of the population and evolutionary biology of these organisms. A Web site for the molecular characterization of pneumococci by MLST is available (http://mlst.zoo.ox.ac.uk).

Keywords: molecular typing, virulent clones, serotype stability, housekeeping genes, World Wide Web

INTRODUCTION

Streptococcus pneumoniae remains a major cause of morbidity and mortality worldwide, causing diseases which range in severity from otitis media and sinusitis, to pneumonia, septicaemia and meningitis (Feldman & Klugman, 1997). Pneumococci are resolved into >90 serotypes, based on the immunochemistry of their capsular polysaccharides, but isolates of many serotypes rarely cause disease and 16 serotypes cause approximately 90% of invasive disease worldwide (Feldman & Klugman, 1997; Scott *et al.*, 1996).

Some serotypes are particularly associated with disease in children or adults (Scott *et al.*, 1996), but there have been very few attempts to identify particular strains that are associated with disease. For example, we do not know whether most isolates of the 'virulent' serotypes have increased capacity to cause disease, or whether most isolates of these serotypes rarely cause disease, with the majority of disease being caused by a small number of virulent clones within each virulent serotype.

Abbreviations: I_A, index of association; MLEE, multilocus enzyme electrophoresis; MLST, multilocus sequence typing; ST, sequence type; UPGMA, unweighted pair group method with arithmetic means.

The sequences of each allele at the seven loci have been deposited in GenBank under accession numbers AJ232241–AJ232433.

Neither do we know whether different strains are associated with the different manifestations of pneumococcal disease. Similarly, although serotyping is universally used to characterize pneumococci, we do not know whether isolates of the same serotype are relatively uniform compared to the total pneumococcal population, or whether the diversity within each serotype approaches that within the whole population. We also know little about the extent of recombination in pneumococci, and consequently the population structure (Lomholt, 1995; Hall *et al.*, 1996), and we have only limited knowledge of the frequency of serotype exchange (Coffey *et al.*, 1998a), which is important in the context of the long-term efficacy of new conjugate vaccines that will protect against a limited number of serotypes (Kayhty & Eskola, 1996).

These basic questions could be addressed by the application of valid molecular typing methods to appropriate pneumococcal populations. Due to the large number of pneumococcal serotypes, the need to study isolates from carriage (which are the bulk of the population, since carriage is common but disease is rare), and the different types of disease, as well as antibiotic-susceptible and resistant isolates, an adequate understanding of the population requires the analysis of large numbers of isolates. With many molecular typing methods the results obtained in different studies are difficult to combine. This limitation has been removed by the development of a portable high-resolution molecular typing procedure, multilocus sequence typing (MLST; Maiden *et al.*, 1998), which is based on the principles of multilocus enzyme electrophoresis (MLEE; Selander *et al.*, 1986), but characterizes the alleles present at multiple housekeeping genes directly by nucleotide sequencing, rather than indirectly from the electrophoretic mobilities of their gene products. Like MLEE, MLST uses variation that accumulates slowly, and which is expected to be selectively neutral, and achieves very high resolution by analysing multiple loci. The most important advantage of MLST over MLEE, and the many typing methods that involve the comparisons of DNA fragments on gels, is the unambiguity and electronic portability of nucleotide sequence data (Maiden *et al.*, 1998). For any bacterial species, an expanding database of the sequences of ~ 450 bp regions of the chosen loci can be held on a World Wide Web site, with appropriate interrogation software, so that any laboratory that sequences the same gene fragments can compare their isolates with those in the database via the Internet. In this way the results obtained in different studies in the same laboratory, or in different laboratories, can be readily combined to provide a growing resource for understanding aspects of the global epidemiology of a bacterial species. In addition, the nucleotide sequences of multiple gene fragments, from hundreds of isolates, provides data that can be used to address aspects of the population and evolutionary biology of the species.

MLST was developed and validated using *Neisseria meningitidis* (Maiden *et al.*, 1998); here we describe a MLST scheme and database for *S. pneumoniae* and use the method to identify the major clones associated with serious invasive pneumococcal disease.

METHODS

Bacterial strains. The validation strains (Table 1) included two isolates of the major Spanish multiresistant serotype 23F clone (Muñoz *et al.*, 1991) and single isolates of serotype 19A and serotype 19F variants of this clone (Coffey *et al.*, 1996, 1998b), three isolates of the Spanish multiresistant serotype 6B clone (Soares *et al.*, 1993; Coffey *et al.*, 1996), and clusters of isolates that had been shown to be closely related by Hall *et al.* (1996). The 274 pneumococci from serious invasive disease, mostly meningitis and septicaemia in patients with no special risk factors, were recovered between 1990 and 1997 (excepting the strains from the Netherlands, which were from cases of meningitis between 1975 and 1992) from hospitals or public health laboratories in Australia (19 isolates), Canada (15 isolates), Denmark (55 isolates), Finland (6 isolates), the Netherlands (38 isolates), Sweden (10 isolates), Great Britain (120 isolates) and Uruguay (11 isolates), with no preselection for serotype, or antibiotic susceptibility or resistance. These isolates are therefore likely to be representative of those causing severe pneumococcal disease. The only exception was the collection from Denmark, which met all the above criteria, except that they were selected from the 12 serotypes most commonly causing invasive disease in children and adults. Reference strains of the penicillin-resistant Spanish serotype 9V clone (SP-665, Coffey *et al.*, 1991) and the multiresistant Spanish serotype 14 clone (SP-VH14; Coffey *et al.*, 1996) were included, as well as *Streptococcus mitis* strain 803.

Pneumococci were grown, and chromosomal DNA prepared, as described previously (Coffey *et al.*, 1996). The minimum inhibitory concentrations of antibiotics were determined by the E-test (AB Biodisk). Serotyping was kindly carried out by D. Griffiths at the Department of Microbiology, John Radcliffe Hospital, Oxford.

Selection of gene loci. The *ddl*, *spi* and *recP* sequences were available from GenBank and the *aroE*, *gdh*, *gki* and *xpt* sequences were kindly provided by Dr M. Burnham of SmithKline Beecham. Where possible, the sequences of the selected pneumococcal housekeeping genes were compared with their homologues in the genome sequence of *Streptococcus pyogenes* (dna1.chem.uoknor.edu/strep.html) to identify regions of maximal nucleotide sequence divergence. These regions were considered likely to be the most polymorphic within the pneumococcal population and, for each gene, primers were made that allowed an ~ 450 bp fragment from the selected region to be amplified by PCR.

Nucleotide sequencing and analysis. The sequences of each of the selected ~ 450 bp regions were obtained from a subset of 15 pneumococcal strains; those that distinguished a number of distinct alleles were selected for further evaluation, whereas those that were monomorphic, or nearly so, were discarded. For each of the selected genes, the sequences from all isolates were compared and allele numbers were assigned to each unique sequence. No weighting was given to the degree of sequence divergence between alleles since, in the absence of a knowledge of the proportion of allelic changes that are due to recombination rather than mutation, we cannot say that alleles differing at many sites are any more distantly related than those that differ at a single site. The alleles present at the seven loci define the allelic profile, or sequence type (ST), of that strain and the allelic profiles of the strains were entered

Table 1. Properties of the validation strains of *S. pneumoniae*

ET, electrophoretic type; GB, Great Britain; SP, Spain. The PN strains and their ET and PFGE patterns are from Hall *et al.* (1996). PFGE patterns with different numbers had <80% of DNA fragments in common.

Strain	Allele numbers							Serotype	Country	ET	PFGE pattern	
	aroE	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*				*ApaI*	*SmaI*
PN94-431	13	9	15	14	10	16	1	3	GB	1	3b	3c
PN93-1730	13	9	15	14	10	16	25	3	GB	1	3a	3a
PN94-261	13	9	15	14	10	16	25	3	GB	2	3a	3b
PN93-1160	7	15	2	10	6	1	22	3	GB	16	1b	1b
PN93-1639	7	15	2	10	6	1	22	3	GB	18	1b	1c
PN94-661	7	11	10	1	6	8	14	9V	GB	20	12e	12e
PN93-110	2	8	2	4	6	1	1	9N	GB	21	11a	11a
PN94-336	14	5	4	5	5	3	8	14	GB	4	14d	14d
PN94-153	1	5	4	1	5	3	8	14	GB	4	14c	14c
PN94-653	1	5	4	5	5	1	8	14	GB	3	14e	14a
PN93-872	1	5	4	5	5	1	8	14	GB	3	14a	14a
PN93-637	1	5	4	5	5	1	8	14	GB	3	14a	14a
SP-264	4	4	2	4	4	1	1	23F	SP	–	–	–
SP-577	4	4	2	4	4	1	1	23F	SP	–	–	–
SP-GM70	4	4	2	4	4	1	1	19A	SP	–	–	–
SP-GA71	4	4	2	4	4	1	1	19F	SP	–	–	–
SP-3026	5	6	1	2	6	3	4	6B	SP	–	–	–
SP-GM17	5	6	1	2	6	3	4	6B	SP	–	–	–
SP-681	5	6	1	2	6	3	4	6B	SP	–	–	–

into an Excel database. The data were loaded from Excel into the Statistica package (StatSoft), which was used to produce a matrix of pairwise differences in the allelic profiles, and to construct a dendrogram from the matrix by the unweighted pair group method with arithmetic means (UPGMA) method.

The primers used for amplification by the PCR (55 °C annealing temperature using Qiagen *Taq* polymerase) were: *aroE*-up, 5′ GCCTTTGAGGCGACAGC, and *aroE*-dn, 5′ TGCAGTTCAG/AAAACATA/TTTCTAA; *ddl*-up, 5′ TGCC/TCAAGTTCCTTATGTGG, and *ddl*-dn, 5′ CACTGGGTG/AAAACCA/TGGCAT; *gdh*-up, 5′ ATGGACAAACCAGCNAGC/TTT, and *gdh*-dn, 5′ GCTTGAGGTCCCATG/ACTNCC; *gki*-up, 5′ GGCATTGGAATGGGATCACC, and *gki*-dn, 5′ TCTCCCGCAGCTGACAC; *recP*-up, 5′ GCCAACTCAGGTCATCCAGG, and *rec*P-dn, 5′ TGCAACCGTAGCATTGTAAC; *spi*-up, 5′ TTATTCCTCCTGATTCTGTC, and *spi*-dn, 5′ GTGATTGGCCAGAAGCGGAA; *xpt*-up, 5′ TTATTAGAAGAGCGCATCCT, and *xpt*-dn, 5′ AGATCTGCCTCCTTAAATAC. The DNA fragments were purified using QIAquick (Qiagen) and were sequenced in each direction, using the primers used for amplification, on an Applied Biosystems Prism 377 automated sequencer with d-Rhodamine-labelled terminators (PE Applied Biosystems).

The index of association (I_A) was calculated as described by Maynard Smith *et al.* (1993), except that simulation was used to test the significance of the deviation of the observed variance in allelic mismatches from the variance expected under the null hypothesis of linkage equilibrium. The test of Sawyer (1989) was applied to the synonymous polymorphic sites within the alleles at each locus and the significance of any clustering of polymorphic sites was evaluated using 10000 resamplings of the data.

RESULTS

Sequencing of housekeeping gene fragments

Seven polymorphic gene fragments were chosen: *aroE* (shikimate dehydrogenase), *ddl* (D-alanine–D-alanine ligase), *gdh* (glucose-6-phosphate dehydrogenase), *gki* (glucose kinase), *recP* (transketolase), *spi* (signal peptidase I) and *xpt* (xanthine phosphoribosyltransferase). The seven gene fragments were sequenced from the 295 pneumococci and from *S. mitis*. The sequence diversity within the pneumococcal genes was low but was sufficient to distinguish 18–34 alleles (Table 2), even though in many cases alleles differed from each other at only one or a few nucleotide sites. The polymorphic sites

Table 2. Gene fragments used in MLST analysis

Gene	Fragment size (bp)	Polymorphic sites*	Alleles	Heterozygosity per locus
aroE	405	19 (11)	18	0·789
gdh	459	36 (27)	26	0·820
gki	483	51 (36)	30	0·862
recP	448	16 (13)	19	0·802
spi	472	25 (23)	23	0·820
xpt	486	35 (18)	34	0·810
ddl	441	40 (25)	26	0·852

* Number of synonymous polymorphic sites in parentheses.

(a)

```
                    1111111111111222222222222222233333333333334
                  11123567800011122234670012346777788889000134556788890
                  78237777357567147025215118097570369025810348343808179 4
allele 1          AGCTCGCCTCGTGAACTACGTTCATTCTACGTAACTGAGTTCGGCAGTCGGCTG
allele 2          .....A.G.....................................T..
allele 4          .....A.G.....................................A....A.CA
allele 5          .....G.....................................T.T..
allele 6          .....A.G.....................C.....G.....T..
allele 7          ....T.....A.G.....................................
allele 8          .....G.....................................
allele 9          .....G...T.....................................
allele 10         ....T.....A.G.....................................T..
allele 11         .....A.G.....................................
allele 12         ....T.....G.....................................
allele 13         ....T.....G.....................T.....
allele 14         .....T.....................................
allele 15         .....A.G.....................T.....
allele 16         ....T.....A.G.....T.....
allele 17         ...CT.....G.....T.....
allele 18         .....A.G....A.....T..
allele 19         ....T.....
S. mitis          G.T..AT....C.GG..GT..CGGCATAG..CGCTAAGAC.T.....C....CA
```

(b)

```
                    1111111111111222222222222333333333344444444
                  26666790012367789990112234566780015688991236778888
                  23312694702472748902843959738143769547140710851505 6
allele 1          GTACGGTAGATGGGCCCCGCAGCCCTTCGGGTGGCTTCGCCCCGTCCTTGC
allele 3          .....G.....................................
allele 4          .....GC.....................C.....C.....T
allele 5          .....GC.....................C.....T
allele 6          .....A.....A..C.....C.....T
allele 7          A.....
allele 8          A.....GC.....
allele 9          .....T.....
allele 10         .....A.....CA.....T
allele 11         .....G.....C.A..A.....T
allele 12         .....A..CA.....T
allele 13         .....GC.....C.....
allele 14         .....A.C.....C.....T
allele 15         .....GC...T.T..GA..TCC..A.C...C.....
allele 16         .C.....A.....C.....
allele 17         .....C.....T
allele 18         A...A...GC.....
allele 19         .....G.....A..C.....C.....T
allele 20         .....GC.....C.....
allele 21         .C.....C.....T.....
allele 22         .....G.....T...A..C.....
allele 23         .....C.....
allele 24         .....T.....
allele 25         .....C.....T
allele 26         .....G.....T
allele 27         .....G.....C.A...A.....
allele 28         .....T
allele 29         .....G.....T.....
allele 30         ..G.....
allele 31         ....A.....T
allele 32         .....GC.....C.....C..A.CT
allele 33         .....A..C.AT.....C.....T
allele 34         A.....GC.A.....
allele 35         .....GC.....C.....CT....T
S. mitis          ...A.AC.TG.A.....T.ATT..CT..AC...CC.AT.GTAC.T.C.T
```

Fig. 1. The variable sites in each of the unique sequences (alleles) of the *aroE* (a) and *xpt* (b) fragments from the 295 pneumococci are shown. The sites are numbered above in vertical format. Allele 3 of *aroE*, and allele 2 of *xpt*, were unique to *S. mitis* and are shown at the bottom of each figure.

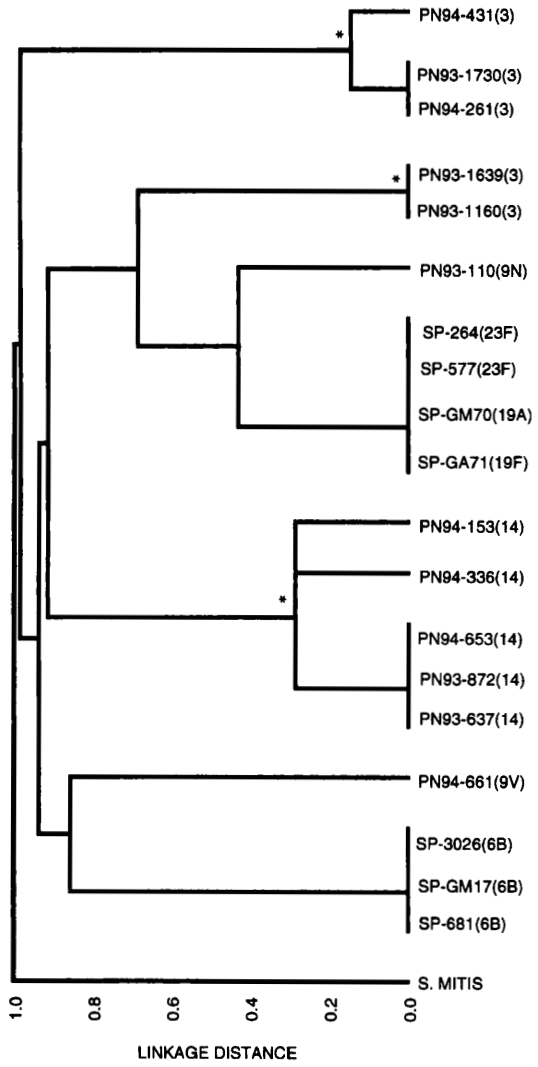

Fig. 2. Dendrogram of the genetic relatedness among the validation strains. The linkage distance is shown at the bottom. The PN isolates emanating from the nodes marked with asterisks were previously shown to be related by MLEE (identical alleles at $\geqslant$8/10 loci), and to give highly related PFGE patterns (Table 1; Hall *et al.*, 1996). SP-264, SP-577, SP-GM70 and SP-GA71 are previously characterized members of the multiresistant serotype 23F clone (or serotype variants of this clone); SP-3026, SP-GM17 and SP-681 are previously characterized members of the Spanish serotype 6B clone.

within the alleles at two typical loci, *aroE* and *xpt*, are shown in Fig. 1. The sequences of the alleles at all seven loci, and the properties of the pneumococci, are available from the MLST Web site (http://mlst.zoo.ox.ac.uk).

For the 295 pneumococci, the mean number of alleles per locus was 25·1, providing the potential to distinguish $>6 \times 10^9$ different genotypes. The frequency of the most common allele in the population ranged from 0·24 (*gki*) to 0·38 (*xpt*) and, from the product of the allele frequencies, isolates with the most common allele at each locus would be expected to be present by chance in the population at a frequency of 0·00038. The ST that had the most common alleles at each locus was not present in the dataset and the allelic profile in the dataset most similar to this ST differed at 3/7 loci. As the expected frequencies of all other STs are lower, and often very much lower, than that of the ST with the most common alleles at each locus, it is unlikely that unrelated isolates will by chance have the same ST.

Validation of the MLST scheme

Isolates that had previously been shown to be closely related using highly discriminatory typing methods (validation strains) were used to check the congruence of MLST with the other methods. One group of isolates included two members of the Spanish multiply antibiotic-resistant serotype 23F clone (Muñoz *et al.*, 1991), and single members of the serotype 19A and 19F variants of this clone. These four isolates have been shown to be indistinguishable by PCR with repetitive element primers (REP-PCR) and very similar, or identical, by MLEE (Coffey *et al.*, 1996, 1998a, 1998b). Using MLST they had identical sequences at all seven loci. Similarly,

three reference isolates of the Spanish multiresistant serotype 6B (Coffey *et al.*, 1996), were identical at all seven loci. Isolates characterized by Hall *et al.* (1996) (Table 1) were also analysed and the relationships between these strains by MLST are shown in Fig. 2. The groupings obtained with MLST were highly congruent with those obtained previously using MLEE and pulsed-field gel electrophoresis (PFGE), except that the serotype 9N (PN93/110) and 9V (PN94/661) strains did not appear to be closely related by MLST (different alleles at 6/7 loci), although they differed at only 2/10 loci by MLEE (Hall *et al.*, 1996). These two strains were, however, not considered to be highly related by PFGE (Hall *et al.*, 1996).

Relatedness of the pneumococcal isolates

Fig. 3 shows a dendrogram constructed from the matrix of pairwise allelic differences between the STs of the 274 isolates from invasive disease. The isolates were resolved into 143 STs, 34 of which contained more than one isolate. The properties of the isolates within each ST are shown in Table 3. The 12 STs that included at least five invasive isolates (range 5–21) were defined as virulent clones (labelled 1–12 in Fig. 3) and are described by a reference strain for the clone, with the predominant clonal serotype in superscript (e.g. the $M7^{6B}$ clone). Each of these clones included isolates from both meningitis and septicaemia.

The most prevalent virulent clone, $M99^{14}$, was widely distributed within the UK and was also found in Australia. A second serotype 14 clone ($M4^{14}$) was recovered throughout northern Europe and from Australia and Canada. These two serotype 14 virulent clones were not closely related as they had different alleles at 6/7 loci. The second most prevalent clone, $M80^{7F}$, was found in the UK, Denmark, Finland, The Netherlands and Uruguay.

Isolates of all but one of the 12 virulent clones were susceptible to penicillin, although all the isolates of the $M99^{14}$ clone from the UK, but only some of those from Australia, were resistant to erythromycin. Five of the 10 single-locus variants of the $M99^{14}$ clone were also resistant to erythromycin. The $M99^{14}$ clone was identical by MLST to the erythromycin-resistant serotype 14 isolates from the study of Hall *et al.* (1996) included in the validation strains (e.g. strains PN93/872, PN93/637 and PN94/653). The $M99^{14}$ clone therefore corresponds to the erythromycin-resistant serotype 14 clone recently associated with invasive disease in the UK (Hall *et al.*, 1996). Isolates of the $M40^{9V}$ virulent clone were penicillin-resistant and this clone is discussed further below.

Relationship between serotypes and STs

In 26 of the 34 STs that contained multiple invasive isolates there was a perfect congruence between ST and serotype; in the other eight STs a single isolate had a serotype different from the other isolate(s). Similarly, isolates of closely related STs (clonal complexes) almost invariably had the same serotype. The letters A–O in Fig. 3 show the nodes from which the descendent STs shared the same serotype. In all cases where isolates in the same, or closely related, STs differed in serotype, the serotypes were rechecked and shown to be correct.

Identity of serotype did not, however, imply close genetic relatedness. Thirteen serotypes were represented by at least five invasive isolates and, for 10 of these serotypes, there were isolates of the same serotype that were only distantly related in genotype (Table 4). Invasive isolates of serotypes 7F and 12F were exceptions and, in each case, they were restricted to a single clonal complex that included isolates of only these serotypes. Serotype 9V isolates were also restricted to a cluster of closely related STs, but the 23 isolates within this clonal complex included five isolates of serotype 19F and three of serotype 14 (plus one non-typable isolate).

Evidence of recombination

There was evidence of a history of intraspecific recombination within the *gki* and *spi* genes fragments as the test of Sawyer (1989) showed a highly non-random distribution of synonymous polymorphic sites ($P < 0{\cdot}0005$). There was no significant clustering of synonymous polymorphic sites within the other gene fragments, but the number of such sites was probably too low to apply this test to *aroE*, *recP* and *xpt* (Table 2). There was a possible example of an interspecific recombinational event in *gdh* (Fig. 4) as the mean divergence between allele 20 and the other pneumococcal *gdh* alleles was 4·58%, which is much higher than the mean diversity within the other pneumococcal alleles (0·95%), and similar to the mean divergence between these alleles and that of *S. mitis* (5·16%).

The level of linkage disequilibrium between alleles was assessed using the I_A (Maynard Smith *et al.*, 1993). An I_A of 0·478 was obtained when one isolate of each of the 143 STs was included in the analysis. Using randomized datasets, this value was shown to be significantly greater than the I_A of zero expected for a population at linkage equilibrium ($P < 0{\cdot}01$). The experimental dataset is a highly biased sample of the total pneumococcal population as it is dominated by the hypervirulent clonal complexes which have risen to high frequency in our sample of isolates from invasive disease. To reduce this sampling bias, the dendrogram was truncated at a genetic distance of 0·2, so that only one isolate of each hypervirulent clonal complex was included in the analysis (Maynard Smith *et al.*, 1993). After truncation, the number of lineages was reduced to 95, and an I_A of 0·267 was obtained when the analysis was performed using a single isolate from each of these 95 lineages. This value was not significantly greater than zero ($P > 0{\cdot}1$), implying no departure form the null hypothesis of linkage equilibrium between alleles.

Characterization of penicillin-resistant isolates

The pneumococci from invasive disease were from countries that have a low incidence of antibiotic

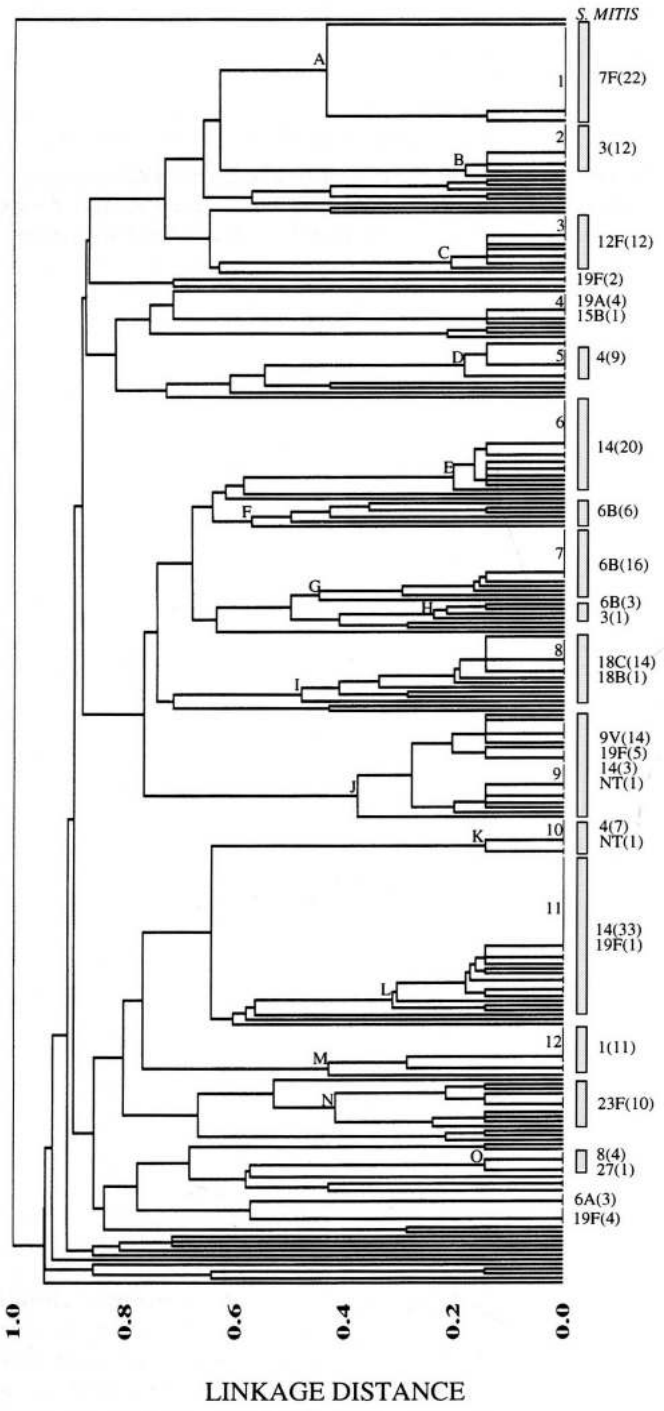

Fig. 3. Dendrogram of the genetic relatedness among the 274 invasive isolates of *S. pneumoniae*. The linkage distance is shown at the bottom. The letters A–O identify nodes from which all descendent STs (shown as shaded rectangles) share a common, or predominant, serotype and which are defined as clonal complexes. The serotypes of the descendent STs and the numbers of isolates of each serotype (in parentheses) are shown. The serotypes and numbers of isolates are also shown for the other STs that include multiple isolates. The numbers 1–12 identify STs that include at least five isolates and which are defined as virulent clones. NT, non-typable.

resistance. However, 17 of the invasive isolates (6%) had minimal inhibitory concentrations (MICs) of $\geqslant 0{\cdot}5$ μg benzylpenicillin ml^{-1}. All of these could be assigned by MLST to the previously characterized high-level penicillin-resistant clones by comparison with reference strains. All six isolates of the $M40^{9V}$ virulent clone were penicillin-resistant and were identical by MLST to SP665, a reference strain of the globally distributed Spanish penicillin-resistant serotype 9V clone (Coffey *et al.*, 1991; Gasc *et al.*, 1997; Hermans *et al.*, 1997); six additional resistant isolates differed at only a single locus and were also assigned to this clone. M225 from Australia was identical by MLST to GM17, a reference strain of the multiresistant Spanish serotype 6B clone (Coffey *et al.*, 1996), and M41 from the UK was identical to SP264, a reference strain of the multiresistant Spanish serotype 23F clone (Coffey *et al.*, 1996). A further resistant isolate (M170 from Denmark) was also assigned to the latter clone as it differed at only a single locus. All these penicillin-resistant isolates had the same serotypes as the reference isolates of the resistant clones, except that one of the members of the Spanish serotype 9V ($M40^{9V}$) clone, and one single-locus variant of this clone, were serotype 14. One other single-locus variant could not be serotyped.

Two serotype 19F isolates from the UK (M97 and M109; MIC of 1 μg penicillin ml^{-1}) were identical to each other, but were only distantly related to the reference strains of the known penicillin-resistant clones. These isolates have recently been shown by MLST to be members of a multiresistant serotype 19F clone that is prevalent in Taiwan (Z.-Y. Shi, M. Enright and B. G. Spratt, unpublished).

DISCUSSION

Several molecular typing procedures, including REP-PCR (Versalovic *et al.*, 1993; Coffey *et al.*, 1998b), BOX-PCR (van Belkum *et al.*, 1996), MLEE (Muñoz *et al.*, 1991; McDougal *et al.*, 1992; Sibold *et al.*, 1992) and PFGE (Soares *et al.*, 1993; Hall *et al.*, 1996), are currently used to characterize pneumococci. Most studies have analysed penicillin-resistant isolates and each of the above techniques can identify closely related isolates within these populations. One disadvantage of these techniques is the difficulty of comparing results between laboratories and PFGE has emerged as the most suitable technique for comparing results as, with careful control over experimental conditions, very similar fragment patterns can be obtained for the same strains in different laboratories (Tenover *et al.*, 1995). However, the comparison of fragment sizes on gels is not ideal, and the unambiguity and electronic portability of DNA sequence data that provide the rationale for MLST have very considerable advantages for molecular typing procedures (Maiden *et al.*, 1998).

MLST using seven loci provided a high-resolution portable molecular typing procedure for pneumococci. In our study of meningococci we validated MLST by using isolates that had previously been characterized by

Table 3. Properties of the isolates within the 143 STs among the 274 invasive pneumococci

Clonal complex*	Reference strain/ST†	*N*	Serotype	Countries‡	Resistance to§:		Allele nos						
					Pen	Ery	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*
A	M1	1	7F	S	S	S	9	12	2	1	6	9	17
A	M80†	20	7F	GB/D/F/N/U	S	S	8	9	2	1	6	1	17
A	M139	1	7F	D	S	S	8	9	2	1	10	1	17
B	M195†	8	3	GB/C/D/N	S	S	7	15	2	10	6	1	22
B	M147	2	3	D	S	S	7	15	2	10	10	1	22
B	M148	1	3	D	S	S	7	15	2	10	7	1	22
B	M211	1	3	GB	S	S	7	15	27	10	6	1	22
–	M8	1	9N	S	S	S	2	8	2	4	6	1	1
–	M18	1	–	GB	–	–	11	8	2	4	6	1	1
–	M94	1	22	GB	S	S	2	13	2	4	6	1	1
–	M41	1	23F	GB	R	S	4	4	2	4	4	1	1
–	M170	1	23F	D	R	S	4	4	2	4	5	1	1
–	M212	1	16F	GB	S	S	1	5	2	20	1	1	1
–	M75	1	22F	GB	S	S	10	16	16	1	6	17	6
–	M283	1	20	U	S	S	10	16	32	1	6	28	31
C	M150†	6	12F	GB/C/D/U	S	S	10	20	14	1	6	1	29
C	M151	1	12F	D	S	S	10	20	14	1	5	1	29
C	M153	1	12F	D	S	S	10	20	14	1	9	1	29
C	M154	2	12F	GB/D	S	S	10	20	14	1	10	1	29
C	M176	1	12F	GB	S	S	10	20	14	1	23	1	29
C	M191	1	12F	GB	S	S	10	20	14	4	6	1	29
–	M259	1	6A	N	S	S	10	8	30	5	6	1	9
–	M97	2	19F	GB	R	R	15	16	19	15	6	20	26
–	M106	1	6	GB	S	S	15	17	4	16	6	19	17
–	M9	1	18C	S	S	S	5	13	11	4	15	12	19
–	M52†	5	19A/15B(1)	GB/N	S	S	8	13	14	4	17	4	14
–	M130	1	14	D	S	S	8	13	14	4	1	4	14
–	M292	1	19F	C	S	S	8	13	14	10	9	4	14
–	M242	1	19A	N	S	S	2	24	2	4	9	4	31
–	M255	1	19F	N	S	S	2	27	2	4	9	4	31
–	M276	1	19F	N	S	S	2	27	2	4	9	32	31
D	M112	2	4	GB	S	S	16	13	4	5	6	10	18
D	M126†	5	4	D/N	S	S	16	13	4	5	6	10	14
D	M125	2	4	D	S	S	16	13	20	5	6	10	14
–	M235	1	6B	A	S	S	10	13	2	1	6	19	14
–	M243	1	2	N	S	S	2	13	4	1	6	6	14
–	M303	1	35B	C	S	S	8	13	4	8	6	22	34
–	M132	1	14	D	I	S	12	19	2	17	6	22	14
E	M4†	11	14	A/GB/C/F/N/S	S	S	7	5	1	8	14	11	14
E	M10	2	14	GB/S	S	S	7	5	1	8	14	11	5
E	M119	1	14	F	S	S	7	5	1	8	1	11	14
E	M133	2	14	D/N	S	S	7	5	1	8	6	11	14
E	M184	1	14	GB	S	S	7	5	1	8	9	11	14
E	M267	1	14	N	S	S	7	5	1	8	22	11	14
E	M278	1	14	N	S	S	7	5	1	8	10	11	14
E	M223	1	14	A	S	S	7	5	1	8	14	14	14
–	M286	1	6B	U	S	S	7	2	1	10	27	4	14
–	M230	1	19A	A	S	S	7	23	1	1	14	28	31
F	M54	1	6B	GB	S	S	7	6	1	2	6	1	23
F	M55	1	6B	GB	S	S	7	6	1	2	6	15	14
F	M294	1	6B	C	S	S	7	6	1	1	6	15	14
F	M256	1	6	N	S	S	7	10	1	1	6	31	14

[*Continued overleaf*

Table 3 *(cont.)*

Clonal complex*	Reference strain/ST†	*N*	Serotype	Countries‡	Resistance to§:		Allele nos						
					Pen	Ery	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*
F	M226	1	6B	A	S	S	7	22	1	2	5	1	14
F	M225	1	6B	A	R	R	5	6	1	2	6	3	4
G	M7†	11	6B	GB/D/S	S	S	7	5	8	5	10	6	14
G	M159	1	6B	D	S	S	7	5	8	5	5	6	14
G	M181	1	6B	GB	S	R	7	9	8	5	10	6	14
G	M227	1	6B	A	S	S	7	5	8	18	10	6	14
G	M82	1	6B	GB	S	S	7	5	8	5	5	1	14
G	M298	1	6B	C	S	S	7	5	8	10	6	6	27
H	M49	1	6B	GB	S	S	7	13	8	6	10	6	14
H	M50	1	3	GB	–	–	1	13	8	6	10	6	14
H	M158	1	6B	D	S	S	7	13	8	6	6	6	14
H	M228	1	6B	A	S	S	7	13	4	6	10	6	14
–	M172	1	23F	D	I	S	7	13	1	6	5	6	8
–	M210	1	19A	D	S	S	7	13	8	6	25	6	8
–	M209	1	–	GB	–	–	7	21	26	1	10	25	14
I	M42	1	18C	GB	S	S	7	2	1	13	10	1	21
I	M43†	6	18C/18B(1)	GB/N	S	S	7	2	1	1	10	1	21
I	M161	2	18C	D	S	S	7	2	1	5	10	1	21
I	M202	1	18C	GB	S	S	7	2	25	1	10	1	21
I	M203	1	18C	GB	S	S	7	2	1	1	24	1	21
I	M200	1	18C	GB	S	S	7	2	24	1	10	24	21
I	M163	1	18C	D	S	S	7	2	1	10	6	1	6
I	M164	1	18C	D	S	S	7	2	1	1	6	1	30
I	M208	1	18C	GB	S	S	7	2	4	19	10	10	21
–	M155	1	6B	D	S	S	7	2	4	2	22	20	28
–	M249	1	6	N	S	S	7	25	4	2	15	1	28
J	M14	1	19F	GB	S	S	7	11	10	6	6	8	14
J	M78	1	9V	GB	S	S	7	11	10	5	6	8	14
J	M189	4	19F/14(1)	GB/C	S/R(1)	S/R(1)	7	11	10	1	6	8	14
J	M198	1	9V	GB	S	S	7	11	10	12	6	8	14
J	M142	1	9V	D	S	S	7	11	10	1	1	8	14
J	M180	3	9V/19F(1)	GB	S	S	7	11	10	1	10	8	14
J	M40†	6	9V/14(1)/NT(1)	GB/C/D/U	R	S/R(2)	7	11	10	1	6	8	1
J	M140	2	9V	D	R	S	7	11	10	1	1	8	1
J	M141	1	9V	D	R	S	7	11	10	1	20	8	1
J	M143	1	9V	D	R	S	7	11	10	1	21	8	1
J	M288	1	14	U	R	S	7	11	10	1	6	34	1
J	M91	1	9V	GB	S	S	7	11	10	1	5	18	1
K	M2†	6	4/NT(1)	A/C/D/S	S	S	10	5	4	5	13	10	18
K	M76	2	4	GB	S	S	10	5	17	5	13	10	18
L	M99†	21	14/19F(1)	A/GB	S	R/S(4)	1	5	4	5	5	1	8
L	M12	2	14	GB/N	S	S/R	1	5	4	5	5	3	8
L	M224	1	14	A	S	S	1	5	4	5	5	27	8
L	M60	1	14	GB	S	S	1	5	4	5	5	1	24
L	M233	1	14	A	S	R	1	5	4	5	5	1	14
L	M79	2	14	A/GB	S	S	1	5	4	1	5	1	8
L	M88	2	14	GB	S	R	1	5	4	5	17	1	8
L	M183	1	14	GB	S	R	1	5	4	5	10	1	8
L	M302	1	14	C	S	S	1	5	4	12	5	27	8
L	M77	1	14	GB	–	–	1	5	10	5	5	3	8
L	M273	1	14	N	S	S	1	5	31	5	5	3	8
–	M201	1	19F	GB	S	S	1	10	4	1	9	3	8

Table 3 *(cont.)*

Clonal complex*	Reference strain/ST†	*N*	Serotype	Countries‡	Resistance to§:		Allele nos						
					Pen	Ery	*aroE*	*gdh*	*gki*	*recP*	*spi*	*xpt*	*ddl*
–	M305	1	19F	C	S	S	18	5	4	1	6	35	8
–	M245	1	6A	N	S	S	1	5	7	12	26	1	14
M	M22†	8	1	GB/D	S	S	12	5	13	5	17	4	20
M	M121	2	1	D	S	S	12	8	1	5	17	4	20
M	M123	1	1	D	S	S	17	8	13	5	17	4	28
–	M16	1	23A	GB	S	S	8	8	9	9	6	4	6
N	M92	1	23F	GB	S	S	1	8	4	1	5	4	6
N	M93	1	23F	GB	S	S	1	8	9	1	5	4	6
N	M175	1	23F	GB	S	S	1	8	4	1	23	4	6
N	M197	3	23F	GB/C	S	S	1	8	4	1	1	4	6
N	M171	1	23F	D	S	S	1	8	6	2	6	4	6
N	M173	1	23F	D	S	S	1	8	6	2	5	4	6
N	M174	1	23F	D	S	S	1	8	6	2	22	4	6
N	M251	1	23F	N	S	S	1	8	1	2	6	4	6
–	M160	1	18C	D	S	S	1	5	9	1	14	14	6
–	M246	1	8	N	S	S	1	5	9	1	9	14	6
–	M263	1	18	N	S	S	19	5	9	1	9	14	6
–	M19	1	15A	GB	S	S	2	5	12	10	6	10	5
–	M274	1	19	N	S	S	2	5	12	10	6	10	32
O	M21	3	8/27(1)	GB/N	S	S	2	5	1	11	16	3	14
O	M261	2	8	N	S	S	2	5	1	11	6	3	14
–	M131	1	14	D	S	S	2	5	21	12	10	21	14
–	M196	2	33F	GB/N	S	S	2	5	23	18	10	3	1
–	M253	1	18	N	S	S	2	5	29	18	9	3	18
–	M107	3	6A	GB	S	S	2	7	4	10	10	1	27
–	M165	4	19F	D	S	S	2	14	4	12	1	1	14
–	M252	1	3	N	S	S	1	26	28	11	13	30	14
–	M290	1	3	U	S	S	1	26	28	11	6	1	14
–	M23	1	22F	GB	S	S	1	1	4	1	18	13	18
–	M95	1	14	GB	S	S	2	1	18	1	19	19	14
–	M120	1	1	D	S	S	10	18	4	1	7	19	9
–	M177	1	19F	GB	S	S	18	2	22	1	9	23	14
–	M282	1	5	U	S	S	16	12	9	1	15	33	33
–	M247	1	14	N	S	S	6	25	1	5	27	29	5
–	M146	1	3	D	S	S	13	9	15	14	4	16	1
–	M217	1	21	GB	S	S	8	10	2	16	1	26	1
–	M236	1	6B	A	S	S	8	13	2	16	1	26	1
–	M280	1	19F	U	S	S	7	10	19	16	14	7	1
–	M166	1	19F	D	S	S	5	5	7	7	8	5	4

* The clonal complexes correspond to those shown in Fig. 3.

† STs are named after the reference isolate for each ST. The STs are listed in the same order as in Fig. 3. The 12 STs that include at least five invasive isolates are marked with daggers.

‡ A, Australia; GB, Great Britain; C, Canada; D, Denmark; F, Finland; N, Netherlands, S, Sweden; U, Uruguay.

§ Isolates with MICs of $\geqslant 0{\cdot}5$ μg benzylpenicillin ml^{-1}, or $\geqslant 4$ μg erythromycin ml^{-1}, were scored as resistant (R); otherwise they were scored as susceptible (S). Two isolates had low level resistance to penicillin (MIC of $0{\cdot}25$ μg ml^{-1}) and were scored I.

–, Not determined. Where there were isolates in a ST with more than one serotype, or with different antibiotic susceptibilities, the predominant phenotype is shown first and the numbers of isolates of the minority phenotype are shown in parentheses.

MLEE (Maiden *et al.*, 1998). This is clearly not necessary, and here we validated the method using a small number of isolates that had already been typed by MLEE, but the majority of isolates were from recent episodes of serious invasive pneumococcal disease and had not previously been characterized. The validity of

Table 4. Serotype distribution of invasive pneumococcal isolates

Serotype	No.	Clusters*
14	61	8 (2)
6B	28	7 (3)
7F	22	1
19F	20	11 (2)
3	16	4 (1)
4	16	2 (2)
18C	16	3 (1)
9V	14	1
23F	13	3 (1)
1	12	2 (1)
12F	12	1
6A	5	3 (1)
8	5	2 (1)

* Numbers of clusters of isolates, or individual isolates, of each serotype that are separated by a genetic distance of $>0{\cdot}5$. The numbers of major clusters of each serotype are shown in parentheses. Serotypes that are represented by fewer than five invasive isolates are not included.

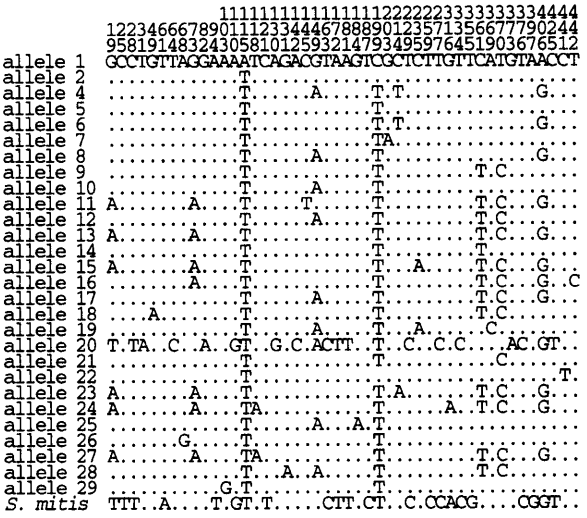

Fig. 4. Variable sites in the *gdh* gene. The variable sites in each of the alleles of the 295 pneumococci and *S. mitis* are shown. Allele 3 was found only in *S. mitis* and is shown at the bottom of the figure.

the pneumococcal MLST procedure was very strongly supported by the fact that almost all isolates within single STs, or clusters of closely related STs, shared the same serotype. The minor exceptions, where isolates of different serotypes were found within the same ST, could be explained either by a lack of resolving power of MLST or by the occurrence of a serotype change.

The expected frequency of occurrence of any ST by chance can be estimated from the products of the frequencies of their alleles in the population of 274 invasive isolates. Isolates of the eight STs that contained different serotypes would be expected by chance at frequencies of between $3{\cdot}5 \times 10^{-5}$ and $1{\cdot}6 \times 10^{-6}$ and it is therefore highly unlikely that a lack of resolving power has resulted in the grouping of unrelated isolates in the same ST. A much more plausible explanation is the occurrence of a change of serotype within these lineages. Serotype changes mediated by recombinational exchanges at the capsular biosynthetic locus are well documented from studies of penicillin-resistant pneumococci (Coffey *et al.*, 1991, 1998a; Hermans *et al.*, 1997; Nesin *et al.*, 1998).

MLST was developed for the identification of clones and clonal complexes within populations of bacteria (or other organisms); from this study, and the previous study of meningococci, it appears very well suited to this purpose. Twelve major clusters of pneumococcal isolates with identical STs were identified and these presumably represent recently emerged virulent lineages that have become over-represented within pneumococcal populations recovered from serious invasive disease (Maynard Smith *et al.*, 1993). It seems likely that most of these virulent clones are widely distributed as, even with the relatively small numbers of isolates from outside Europe, seven of the 12 virulent clones were found in at least two continents, and the $M2^{4}$, $M4^{14}$, $M150^{12F}$ and $M40^{9V}$ clones were found on three continents (Table 3).

Although MLST allows the identification of pneumococcal clones and clonal complexes, and probably provides reliable information about the genetic relatedness of isolates that differ at less than four of the seven loci, the relationships between isolates on deeper branches of the dendrogram should be treated with extreme caution. As lineages diverge they will relatively rapidly accumulate at least one nucleotide change in each of the seven gene fragments. Thus, the relationships between the major lineages are highly unreliable as they will usually involve comparisons of isolates that differ at most or all loci. In any case, it is unwise to assume there is any phylogenetic information in the deeper branches of dendrograms unless it is known that recombination is rare, and Fig. 3 should be viewed only as a convenient visual representation of the clusters of isolates with identical, or very similar, allelic profiles.

Although isolates with the same ST and closely related STs usually had the same serotype, isolates of most serotypes were distributed across the dendrogram. The most likely explanation is that in the long term the capsular genes that specify serotype spread horizontally within the population, but in the short term most isolates of invasive pneumococcal clones maintain their ancestral serotype as they expand and begin to diversify. Isolates of serotype 19F were the most widely distributed across the dendrogram, and this may reflect the fact that this serotype is commonly carried in the nasopharynx (Feldman & Klugman, 1997), providing opportunities for the horizontal transfer of its capsular genes to co-

infecting lineages of other serotypes. The genetic heterogeneity of serotype 19F isolates was also noted in a recent study of pneumococci from Finland (Takala *et al.*, 1996)

Serotypes 7F, 9V and 12F were atypical as in each case all invasive isolates appeared to be closely related and are presumably descended from a recent common ancestor. The limited genetic diversity among serotype 7F and 9V isolates was noted in recent studies of pneumococci from Finland (Takala *et al.*, 1996) and France (Gasc *et al.*, 1997), respectively. The Finnish study also found that all serotype 14 isolates were closely related, but in our larger study of isolates from worldwide sources we found two major clusters of serotype 14 lineages, and eight other serotype 14 isolates were in six different STs or clonal complexes.

The congruence between serotype and genetic relatedness was slightly unexpected as pneumococci are naturally transformable, and the population has been proposed to be weakly clonal (Lomholt, 1995; Hall *et al.*, 1996), implying that recombinational exchanges are relatively frequent. Invasive isolates are not an ideal sample for estimating the extent of recombination in a species like *S. pneumoniae*, where the bulk of the population is found in healthy carriers, as any lineages that have an increased capacity to cause disease become highly over-represented in this type of sample (Maynard Smith *et al.*, 1993). However, the linkage equilibrium between alleles that was observed when the distorting effect of this type of epidemic population structure (Maynard Smith *et al.*, 1993) was minimized (by analysing only those lineages present when the dendrogram was truncated at a linkage distance of 0·2) supports the view (Lomholt, 1995; Hall *et al.*, 1996) that recombination in the longer term may shuffle the alleles in the pneumococcal population. A much better estimate of the significance of recombination would be obtained by an analysis of linkage disequilibrium within an MLST dataset from an unbiased sample of the pneumococcal population (e.g. isolates from nasopharyngeal carriage in different countries).

In bacterial species there is often a poor congruence between genetic relatedness inferred by MLEE and serology, as recombinational exchanges that alter serotype may be strongly selected by the host immune system (Selander & Musser, 1990). The congruence observed here may reflect the fact that the clonal complexes defined by MLST are much younger than those defined by MLEE, since any nucleotide difference at a locus produces a new ST, whereas in MLEE about 26 nucleotide changes are required to alter the electrophoretic mobility of an enzyme (Boyd *et al.*, 1994) and produce a new electrophoretic type. Furthermore, the impact of recombination may be less in invasive isolates than in the bulk of the population which exists within the nasopharynges of healthy carriers, and there may be less congruence between genetic identity and serotype in the latter population, since isolates of different lineages are more likely to meet each other.

The pneumococcal MLST scheme and the MLST database provide the start of a new and powerful approach to the study of the global molecular epidemiology of pneumococcal disease. The study reported here was strongly biased to isolates from Great Britain but further studies of invasive isolates from other countries are in progress. These ongoing studies, and studies of isolates from different manifestations of pneumococcal disease, and from carriage, should provide answers to some of the basic questions about the population biology of this major pathogen. Other laboratories can easily compare their isolates with those in our growing pneumococcal MLST database by submitting the sequences from the seven housekeeping gene fragments to our World Wide Web site for the MLST of pneumococci.

MLST can also be used to characterize unusual isolates (e.g. optochin-resistant, non-serotypable strains), that may be either atypical pneumococci or viridans group streptococci, as in the latter case they should, like the *S. mitis* isolate included in this study, have highly diverged alleles at each locus. MLST also assigned all of the penicillin-resistant invasive isolates to the known resistant clones, and the ability to characterize penicillin-resistant isolates, by comparing their allelic profiles via the Internet with those in our database of resistant isolates, provides an unambiguous approach that has substantial advantages over the current methods for the characterization of these strains.

ACKNOWLEDGEMENTS

This work was supported by a grant from the Wellcome Trust. B.G.S. is a Wellcome Trust Principal Research Fellow. We are grateful to P. Wilkinson, P. Pickerill and M. Daniels for assistance; T. C. Boswell, D. W. Crook and the Oxford pneumococcal surveillance group, C. G. Dowson, A. Efstratiou, S. H. Gillespie, L. M. C. Hall, H. Janzen, R. M. Hargreaves, J. Henrichsen, P. W. Hermans, M. Hortal, D. Low, H. Lomholt, A. F. Maggs, H. McKenzie, J. C. Paton, J. Paul and E. Tarkka for strains; D. Griffiths for serotyping; Martin Maiden for stimulating discussions; and M. Burnham of SmithKline Beecham for pneumococcal sequences.

REFERENCES

van Belkum, A., Sluijter, M., de Groot, R., Verbrugh, H. & Hermans, P. W. (1996). Novel BOX repeat PCR assay for high-resolution typing of *Streptococcus pneumoniae* strains. *J Clin Microbiol* **34**, 1176–1179.

Boyd, E. F., Nelson, K., Wang, F. S., Whittam, T. S. & Selander, R. K. (1994). Molecular genetic basis of allelic polymorphism in malate dehydrogenase (MDH) in natural populations of *Escherichia coli* and *Salmonella enterica*. *Proc Natl Acad Sci USA* **91**, 1280–1284.

Coffey, T. J., Dowson, C. G., Daniels, M., Zhou, J., Martin, C., Spratt, B. G. & Musser, J. M. (1991). Horizontal gene transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes in natural populations of *Streptococcus pneumoniae*. *Mol Microbiol* **5**, 2255–2260.

Coffey, T. J., Berrón, S., Daniels, M., Garcia-Leoni, E., Cercenado, E., Bouza, E., Fenoll, A. & Spratt, B. G. (1996). Multiply antibiotic-

resistant *Streptococcus pneumoniae* recovered from Spanish hospitals (1988–1994): novel major clones of serotypes 14, 19F and 15F. *Microbiology* **142**, 2747–2757.

Coffey, T. J., Enright, M. C., Daniels, M., Morona, J. K., Morona, R., Hryniewicz, W., Paton, J. C. & Spratt, B. G. (1998a). Recombinational exchanges at the capsular polysaccharide biosynthetic locus lead to frequent serotype changes among natural isolates of *Streptococcus pneumoniae*. *Mol Microbiol* **27**, 73–83.

Coffey, T. J., Enright, M. C., Daniels, M., Wilkinson, P., Berrón, S., Fenoll, A. & Spratt, B. G. (1998b). Serotype 19A variants of the Spanish serotype 23F multiresistant clone of *Streptococcus pneumoniae*. *Microb Drug Resist* **4**, 51–55.

Feldman, C. & Klugman, K. P. (1997). Pneumococcal infections. *Curr Opin Infect Dis* **10**, 109–115.

Gasc, A. M., Giammarinaro, P., Ton-Hoang, B., Geslin, P., van der Giezen, M. & Sicard, M. (1997). Structural organization of the *Streptococcus pneumoniae* chromosome and relatedness of penicillin-sensitive and -resistant strains in type 9V. *Microb Drug Resist* **3**, 65–72.

Hall, L. M. C., Whiley, R. A., Duke, B., George, R. C. & Efstratiou, A. (1996). Genetic relatedness within and between serotypes of *Streptococcus pneumoniae* from the United Kingdom – analysis of multilocus enzyme electrophoresis, pulsed-field gel electrophoresis, and antimicrobial resistance patterns. *J Clin Microbiol* **34**, 853–859.

Hermans, P. W., Sluijter, M., Dejsirilert, S., Lemmens, N., Elzenaar, K., van Veen, A., Goessens, W. H. & de Groot, R. (1997). Molecular epidemiology of drug-resistant pneumococci: toward an international approach. *Microb Drug Resist* **3**, 243–251.

Kayhty, H. & Eskola, J. (1996). New vaccines for the prevention of pneumococcal infections. *Emerging Infect Dis* **2**, 289–298.

Lomholt, H. (1995). Evidence of recombination and an antigenically diverse immunoglobulin A1 protease among strains of *Streptococcus pneumoniae*. *Infect Immun* **63**, 4238–4243.

McDougal, L. K., Facklam, R., Reeves, M., Hunter, S., Swenson, J. M., Hill, B. C. & Tenover, F. C. (1992). Analysis of multiply antimicrobial-resistant isolates of *Streptococcus pneumoniae* from the United States. *Antimicrob Agents Chemother* **36**, 2176–2184.

Maiden, M. C. J., Bygraves, J. A., Feil, E. & 10 other authors (1998). Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci USA* **95**, 3140–3145.

Maynard Smith, J., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci USA* **90**, 4384–4388.

Muñoz, R., Coffey, T. C., Daniels, M. & 8 other authors (1991). Intercontinental spread of a multiresistant clone of serotype 23F *Streptococcus pneumoniae*. *J Infect Dis* **164**, 302–306.

Nesin, M., Ramirez, M. & Tomasz, A. (1998). Capsular transformation of a multidrug-resistant *Streptococcus pneumoniae in vivo*. *J Infect Dis* **177**, 707–713.

Sawyer, S. A. (1989). Statistical tests for detecting gene conversion. *Mol Biol Evol* **2**, 539–556.

Scott, J. A. G., Hall, A. J., Dagan, R. & 12 other authors (1996). Serogroup-specific epidemiology of *Streptococcus pneumoniae* – associations with age, sex, and geography in 7,000 episodes of invasive disease. *Clin Infect Dis* **22**, 973–981.

Selander, R. K. & Musser, J. M. (1990). Population genetics of bacterial pathogenesis. In *Molecular Basis of Bacterial Infections*, pp. 11–36. Edited by B. H. Iglewski & V. L. Clark. San Diego: Academic Press.

Selander, R. K., Caugant, D. A., Ochman, H., Musser, J. M., Gilmour, M. N. & Whittam, T. S. (1986). Methods of multilocus enzyme electrophoresis for bacterial population genetics and systematics. *Appl Environ Microbiol* **51**, 873–884.

Sibold, C., Wang, J., Henrichsen, J. & Hakenbeck, R. (1992). Genetic relationships of penicillin-susceptible and -resistant *Streptococcus pneumoniae* strains isolated on different continents. *Infect Immun* **60**, 4119–4126.

Soares, S., Kristinsson, K. G., Musser, J. M. & Tomasz, A. (1993). Evidence for the introduction of a multiresistant clone of serotype 6B *Streptococcus pneumoniae* from Spain to Iceland in the late 1980's. *J Infect Dis* **168**, 158–163.

Takala, A. K., Vuopiovarkila, J., Tarkka, E., Leinonen, M. & Musser, J. M. (1996). Subtyping of common pediatric pneumococcal serotypes from invasive disease and pharyngeal carriage in Finland. *J Infect Dis* **173**, 128–135.

Tenover, F. C., Arbeit, R. D., Goering, R. V., Mickelsen, P. A., Murray, B. E., Persing, D. H. & Swaminathan, B. (1995). Interpreting chromosomal DNA restriction patterns produced by pulsed-field gel electrophoresis: criteria for bacterial strain typing. *J Clin Microbiol* **33**, 2233–2239.

Versalovic, J., Kapur, V., Mason, E. O., Shah, V., Keouth, T., Lupski, J. R. & Musser, J. M. (1993). Penicillin-resistant *Streptococcus pneumoniae* strains recovered in Houston: identification and molecular characterization of multiple clones. *J Infect Dis* **167**, 850–856.

Received 15 May 1998; revised 12 July 1998; accepted 22 July 1998.