Determination of Accessory Gene Patterns Predicts the Same Relatedness among Strains of *Streptococcus pneumoniae* as Sequencing of Housekeeping Genes Does and Represents a Novel Approach in Molecular Epidemiology[▽]

Jessica Dagerhamn,[1,2] Christel Blomberg,[1,2] Sarah Browall,[1] Karin Sjöström,[1]
Eva Morfeldt,[1] and Birgitta Henriques-Normark[1,2]*

*Swedish Institute for Infectious Disease Control, Solna, Sweden,[1] and Department of Microbiology, Tumor Biology and
Cell Biology, Karolinska Institutet, Solna, Sweden[2]*

Relatedness between isolates of *Streptococcus pneumoniae* can be determined from sequences of multiple genes belonging to the core genome (multilocus sequence typing [MLST]), but these do not provide information on gene content that may affect the potential of isolates to cause invasive pneumococcal disease. Gene content data, obtained using microarrays, were gathered for 40 clinical isolates of 12 serotypes belonging to 30 multilocus sequence types. We found that sequence variations in housekeeping genes assessed by MLST correlated well with whole-genome microarray analyses identifying the presence/absence of accessory genes/ regions. However, isolates belonging to the same clonal complex, as determined by MLST, may not have identical gene contents, potentially affecting virulence. We found fewer intraclonal (same MLST sequence type) differences associated with pneumococcal serotypes of high invasive disease potential, i.e., serotypes rarely found among carriers compared to serotypes frequently found in carriage. Molecular typing of pneumococci based on the presence/absence of 25 genes localized to accessory regions shows the same relatedness among pneumococcal strains as MLST does. We conclude that molecular typing of pneumococci based on variation in the nucleotide sequences of parts of housekeeping genes (MLST) correlates with the presence/absence of genes in the accessory part of the genome. This covariation is likely due to the fact that both sequence variations and gene content variations are created primarily by recombination events in pneumococci.

Streptococcus pneumoniae is a devastating pathogen, killing 1 to 2 million people annually. It is also a common colonizer of the respiratory tract, and up to 60 to 70% of healthy children may harbor this organism in the nasopharynx (7). To assess the capability of pneumococci to cause disease and to study their spread, several studies have been conducted to compare invasive disease and carrier isolates, thereby estimating the abundances of the two groups (1, 2, 9). It has been suggested that disease potential is associated not only with serotype but also with clonal type, as determined with molecular methods such as multilocus sequence typing (MLST), comparing sequences of seven housekeeping genes belonging to the core genome and present in all isolates (9). At least 30% of the pneumococcal gene content is variable between strains (unpublished data). Most of these genes are present in regions that can be present or absent when different isolates are analyzed. These accessory regions may encode putative virulence proteins, such as adhesins, whereas others appear to contribute to the metabolism of pneumococci. It is not known how the extent of genetic similarity in individual isolates, as defined by MLST, relates to the extent of similarity in gene content. To study these issues, we performed microarray analyses of 40 pneumococcal isolates belonging to 12 different serotypes and 30 multilocus sequence types (STs).

MATERIALS AND METHODS

Pneumococcal isolates studied. Forty pneumococcal isolates of 12 different serotypes were used in this study. The following serotypes were included: 1 (n = 4), 3 (n = 3), 4 (n = 3), 6B (n = 3), 7F (n = 3), 9V (n = 5), 11A (n = 1), 14 (n = 8), 15B (n = 1), 19F (n = 6), 23F (n = 2), and 35B (n = 1). The laboratory strains R6 and TIGR4 were also used. The isolates belonged to 30 different STs, as assessed using MLST. The isolates were named according to their ST and serotype and were marked numerically when more than one isolate was present within one serotype and ST (Fig. 1).

An eBURST analysis of all isolates in the MLST database was performed using the default values of seven loci per isolate and six as the minimal number of identical loci for group definition (4). We named eBURST groups by their predicted founders as of April 2007. Six different eBURST groups were represented by more than one isolate included in the study. These were the ST306 clonal complex (ST306 and ST228 of serotype 1), the ST124 clonal complex (ST124 and ST307 of serotype 14), the ST191 clonal complex (ST191 and ST1838 of serotype 7F), the ST176 clonal complex (ST176 and ST138 of serotype 6B), the ST180 clonal complex (ST180 and ST1826 of serotype 3), and the ST162 clonal complex (ST162, ST156, and ST838, of serotypes 9V, 14, and 19F).

Serotyping. All pneumococcal isolates were serotyped at the Swedish Institute for Infectious Disease Control by gel diffusion as described previously (7).

MLST. The MLST method was adapted from the procedure described by Enright and Spratt (3) and was performed as previously described (12).

Microarray analysis. Comparative genomic hybridizations were carried out using a reference design as previously described (11). In brief, the microarray for *S. pneumoniae* consisted of 2,797 50-mer oligonucleotides (MWG Biotech, Edensberg, Germany) spotted in triplicate in random order on glass slides (Qarray arrayer [Genetix, Boston, MA] and MWG epoxy slides [MWG Biotech, Edensberg, Germany]). Oligonucleotides were based on predicted open reading frames of the two fully sequenced strains R6 and TIGR4. For each of the 40 test

* Corresponding author. Mailing address: Swedish Institute for Infectious Disease Control/MTC, Karolinska Institutet, SE-171 82 Solna, Sweden. Phone: 46 8 457 24 13. Fax: 46 8 30 25 66. E-mail: Birgitta .Henriques@smi.ki.se.
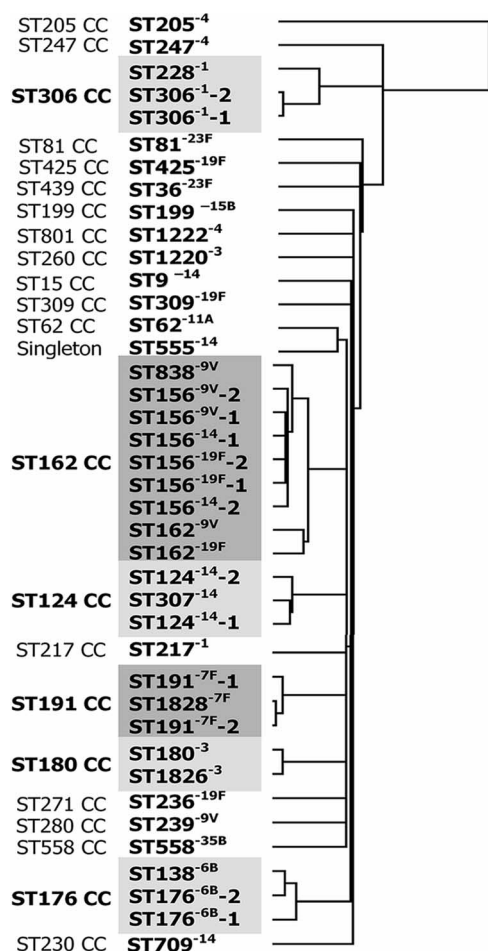
FIG. 1. Hierarchical clustering of microarray results for all 40 pneumococcal isolates tested. Clustering was based on the filtered normalized values for all oligonucleotides obtained from the microarray experiments. The left column indicates clonal complexes (CCs) determined by eBURST. Gray boxes highlight the six clonal complexes that include several STs.

isolates, we performed three or four replicate experiments, including a dye swap experiment with each isolate. The analysis was carried out as previously described (11), and the arrays were scanned using Genepix Pro 6.0 software (Axon Instruments). The $\log_2$ fluorescence ratios were normalized to the median by using the R Project for Statistical Computing (http://www.r-project.org/). For statistical analysis, we used a Bayesian linear model (13) and the Holm multiple testing correction to adjust individual P values. This method was used to compare the isolates to the two reference strains as well as to each other in a pairwise manner. Genes were considered absent if they had a P value of <0.01 with an M value of <−1. Clustering was based on the filtered normalized values of all oligonucleotides obtained from the microarray experiments and was done with an uncentered correlation and average linkage, using Gene Cluster v 2.11. The visualization of clustering was done using Treeview v 1.60. The statistical analysis of the microarray data was performed under rather stringent conditions, meaning that there was a greater likelihood for categorizing a gene as present when it was in fact absent than the opposite. Also, some genes that appeared to be absent may still have been present, but with many differences in sequence over the oligonucleotide being used to define the gene. All isolates were compared to all others with regard to the number of genes differing between the isolates, with a total of 780 comparisons. The comparisons were divided into four groups based on the same or different clonal complexes and the same or different serotypes. We used the Welch two-sample t test to test differences between the groups. Based on the results from the microarray analysis, 25 genes situated in accessory

regions were selected to form a present/absent matrix. Additionally, some of these data were confirmed using PCR (data not shown).

Microarray data accession number. Microarray data have been submitted to Array Express under accession number E-TABM-320.

RESULTS

STs determined with MLST correlate well with whole-genome microarray analysis. Clustering of microarray data from the 40 pneumococcal isolates showed that isolates belonging to the same clonal complex, as determined by the eBURST analysis, grouped together, suggesting that they were closely related. In contrast, isolates not belonging to the same eBURST group did not cluster together and were less genetically related (Fig. 1) (http://www.mlst.net). Isolates of the same serotype but belonging to different clonal complexes were, on the other hand, not more similar to each other than to isolates of unrelated STs and of different serotypes. For some serotypes, i.e., 6B, 7F, and 14, isolates of the same ST were genetically more correlated with a single-locus variant than with other isolates of the same ST. This finding can be explained by solitary recombination events. In other cases, such as for the type 1 isolates of ST306 or ST228 or the ST156 isolates, isolates of the same ST were more genetically related to each other than to isolates belonging to single-locus variants or double-locus variants of that ST.

ST is a good predictor of genetic content, in contrast to serotype. All isolates were compared to all other isolates in a pairwise manner in order to obtain an approximate number of genes that differed between any two isolates (Table 1). Significance was set to P values of <0.01. Genes coding for type 2

TABLE 1. Differences among isolates belonging to the same or related STs (clonal complexes), as assessed by pairwise comparisons, based on microarray results for all genes except those coding for the type 2 and type 4 capsular genes

| Comparison	Serotype	No. of different genes	No. of base pair changes in MLST alleles that differed between two STs
Comparisons within the same ST			
ST306	1	0	
ST156	9V/14/19F	1–17	
ST162	9V/19F	5	
ST191	7F	0	
ST124	14	8	
ST176	6B	39	
Comparisons between related STs (same clonal complex)			
ST306-ST228	1	40–44	spi, 3; gki, 10
ST156-ST162	9V/14/19F	30–53	ddl, 2
ST156-ST838	9V/14/19F	10–29	ddl, 37
ST162-ST838	9V/19F	34–35	ddl, 35
ST191-ST1838	7F	0	spi, 2
ST124-ST307	14	5–11	xpt, 4
ST176-ST138	6B	4–44	recP, 2; gdh, 5
ST180-ST1823	3	4	xpt, 1
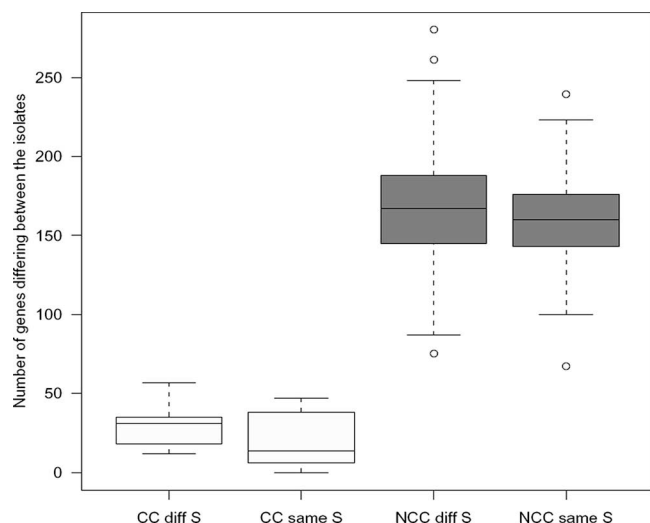
FIG. 2. Correlation of genetic content, as estimated by microarray analysis, with MLST pattern (CC) and serotype. A summary of the pairwise comparisons between all isolates tested is shown. The white boxes represent the numbers of genes differing between isolates belonging to the same clonal complex (CC) and having the same or different serotypes (S). The gray boxes represent numbers of genes differing between isolates not belonging to the same clonal complex (NCC) and having the same or different serotypes. Genes encoding the serotype 2 and 4 capsules on the array were excluded from the analysis.

and type 4 capsules were excluded from the comparisons. The numbers of genes differing between the different isolates are summarized in Fig. 2. By pairwise comparisons, the isolates were grouped into the following groups: group A, clonal complex with the same serotype (24 comparisons); group B, clonal complex with different serotypes (25 comparisons); group C, not a clonal complex but having the same serotype (49 comparisons); and group D, not a clonal complex and having different serotypes (780 comparisons). Differences found between the different groups correlated well with whether the isolates belonged to the same clonal complex but not to whether the isolates were of the same serotype. Furthermore, isolates of the same serotype but of unrelated STs were not more genetically related to each other than to isolates of other capsular types, and a statistically significant difference ($P < 0.0001$) was found between the two groups with different STs (group C [same serotype] or group D [different serotypes]) and the two groups with similar STs (group A [same serotype] or group B [different serotypes]).

The presence or absence of genes in accessory regions predicts genetic relatedness between pneumococcal isolates. MLST is an expensive and time-consuming method in which parts of seven housekeeping genes are sequenced. To investigate whether the absence/presence of genes in the accessory part of the genome could function as a predictor of relatedness, we chose a set of 25 genes from our microarray analysis (Table 2). These genes were all present in different accessory regions consisting of three or more genes that were present or absent in the isolates and were evenly scattered around the genome (unpublished data). Three of the genes chosen were absent in TIGR4, 9 were absent in R6, and the remaining 13 were present in both TIGR4 and R6. Their presence varied

among the 40 clinical isolates studied. The genes were tested using a BLASTn approach of the refseq genomic database at the NCBI website (www.ncbi.nlm.nih.gov) to ensure that there were no other similar genes within the sequenced pneumococcal genomes. SP_0505 showed a high similarity to SP_0507, another gene located within the same accessory region. All 24 other genes had no significant similarity to other pneumococcal genes. In Table 3, the presence/absence of the 25 genes among the 40 clinical isolates is indicated, and the total pattern is described by assignment of a profile number. We found a strong correlation between these profiles and the MLST data at the level of clonal complexes. Thus, profiles of isolates belonging to the same clonal complex did not differ in more than 2 of the 25 genes. Isolates that belonged to different clonal complexes based on MLST differed in at least 4 of the 25 genes.

Serotypes with high invasive disease potential, i.e., those rarely causing carriage, show few intraclonal differences. Next, we studied intraclonal differences among the six clonal complexes representing more than one isolate in this study. We found no genetic differences between the isolates within ST306 and ST191, of the highly invasive (1, 2, 9) serotypes 1 and 7F, respectively (Table 1). However, approximately 40 genes differed between ST306 isolates and the double-locus variant isolate ST228^{-1}. All isolates belonging to the ST124 clonal

TABLE 2. Accessory genes selected to estimate genetic relatedness between pneumococcal isolates[a]

TIGR4 locus	R6 locus	Description
No locus	spr0106	Transporter truncation
	spr1403	Hypothetical protein
	spr1619	ABC transporter, membrane-spanning permease, sugar transporter
SP_0069		Choline binding protein I
SP_0109		Bacteriocin putative
SP_0171		ROK family protein
SP_0308	spr0280	Phosphotransferase system sugar-specific EII component
SP_0394	spr0356	Mannitol PTS EII
SP_0467		Putative sortase
SP_0476	spr0423	Phosphotransferase system sugar-specific EII component
SP_0505	spr0448	Type I site-specific DNase chain S
SP_0695	spr0608	HesA/MoeB/ThiF family protein
SP_1052	spr0953	Hypothetical protein
SP_1131		Putative transcriptional regulator
SP_1222	spr1102	Uncharacterized restriction enzyme, interrupted N
SP_1344	spr1205	Conserved hypothetical protein
SP_1434	spr1289	ABC transporter, ATP binding/permease protein
SP_1620		Putative PTS system nitrogen regulatory component IIA
SP_1640		Hypothetical protein
SP_1761		Hypothetical protein
SP_1796		ABC transporter, substrate binding protein
SP_1830	spr1649	Phosphate transport system, putative regulatory protein PhoU
SP_1933	spr1749	Hypothetical protein
SP_1949	spr1766	Hypothetical protein
SP_2159	spr1965	Hypothetical protein

[a] The presence or absence of these genes is suggested to present a potential novel method for studying molecular epidemiology.

TABLE 3. Presence (+) or absence (−) of 25 accessory genes among 40 pneumococcal isolates[b,d]

Isolate	Serotype	ST	Gene locus/clonal complex	SP_0069T4	SP_0109T4	SP_0171T4	SP_0308	SP_0394	SP_0467T4	SP_0476	SP_0505	SP_0695	SP_1052	SP_1131T4	SP_1222	SP_1344	SP_1434	SP_1620T4	SP_1640	SP_1761T4	SP_1796T4	SP_1830	SP_1933	SP_1949	SP_2159	spr0106R6	spr1403R6	spr1619R6	Profile no.
ST306-1-1[a]	1	306	**306**	−	−	+	+	−	−	+	−	+	−	+	−	+	+	−	+	+	−	−	−	−	+	−	+	+	**1**
ST306-1-2[a]	1	306	**306**	−	−	+	+	−	−	+	−	+	−	+	−	+	+	−	+	+	−	−	−	−	+	−	+	+	**1**
ST228-1	1	228	**306**	−	−	+	+	−	−	+	−	+	−	+	−	+	+	−	+	+	−	−	−	−	+	−	+	+	**2**
ST217-1[a]	1	217	217	−	+	−	−	+	−	−	+	−	+	−	−	−	−	−	−	−	−	−	+	−	−	−	−	+	13
ST1826-3	3	1826	**180**	−	+	−	+	−	−	−	+	−	+	−	−	−	+	+	+	−	+	−	−	−	−	+	−	−	**6**
ST180-3[a]	3	180	**180**	−	+	−	+	−	−	−	+	−	+	−	−	−	+	+	+	−	+	−	−	−	−	+	−	−	**6**
ST1220-3	3	1220	260	+	−	−	−	+	−	−	−	+	+	−	−	−	+	−	−	−	−	−	−	−	+	−	−	−	28
ST1222-4	4	1222	801	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	7
ST247-4	4	247	247	+	−	−	−	+	−	−	−	+	+	−	+	−	+	−	−	+	−	−	−	−	+	−	−	−	11
ST205-4[a]	4	205	205	−	−	−	−	+	−	−	−	−	−	−	+	−	−	−	−	−	−	−	−	+	−	−	−	−	21
ST176-6B-1	6B	176	**176**	−	−	+	+	+	+	−	+	+	−	−	−	+	+	+	+	+	+	−	−	−	+	−	−	−	**17**
ST176-6B-2	6B	176	**176**	−	−	+	+	+	+	−	+	−	−	−	−	+	+	+	+	+	+	−	−	−	+	−	−	−	**18**
ST138-6B	6B	138	**176**	−	−	+	+	+	+	−	+	−	−	−	−	+	+	+	+	+	+	−	−	−	+	−	−	−	**18**
ST191-7F-1[a]	7F	191	**191**	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	+	**8**
ST191-7F-2[a]	7F	191	**191**	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	+	**8**
ST1828-7F	7F	1838	**191**	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	+	−	−	−	−	+	**8**
ST156-9V-1[a]	9V	156	**162**	−	+	−	−	−	+	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	+	+	−	−	**3**
ST156-9V-2[a]	9V	156	**162**	−	+	−	−	−	+	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	+	+	−	−	**3**
ST156-14-1[a]	14	156	**162**	−	+	−	−	−	+	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	+	+	−	−	**3**
ST156-14-2[a]	14	156	**162**	−	+	−	−	−	+	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	+	+	−	−	**3**
ST156-19F-1[a]	19F	156	**162**	−	+	−	−	−	+	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	+	+	−	−	**3**
ST156-19F-2[a]	19F	156	**162**	−	−	−	−	−	+	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	+	+	−	−	**4**
ST838-9V	9V	838	**162**	−	+	−	−	−	+	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	+	+	−	−	**3**
ST162-9V	9V	162	**162**	−	−	−	−	−	+	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	+	+	−	−	**5**
ST162-19F	19F	162	**162**	−	−	−	−	−	+	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	+	+	−	−	**5**
ST239-9V	9V	239	280	−	−	−	−	−	−	+	+	−	−	−	−	−	+	+	+	−	+	−	−	−	−	+	−	−	12
ST62-11A	11A	62	62	+	−	−	−	−	−	−	−	+	+	−	−	−	−	−	−	+	−	−	+	−	+	−	−	+	20
ST555-14	14	555	NA	+	−	−	−	−	−	−	−	−	−	−	+	−	+	−	−	−	−	−	+	−	−	+	−	−	19
ST9-14[a]	14	9	15	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	25
ST709-14	14	709	230	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	10
ST124-14-1[a]	14	124	**124**	−	+	+	−	+	+	−	+	−	+	−	−	+	+	−	−	+	+	−	−	−	+	+	+	−	**14**
ST124-14-2[a]	14	124	**124**	−	+	+	−	+	+	−	+	−	+	−	−	+	+	−	−	+	+	−	−	−	+	+	+	−	**15**
ST307-14	14	307	**124**	−	+	+	−	+	+	−	+	−	+	−	−	+	+	−	−	+	+	−	−	−	+	+	+	−	**15**
ST199-15B[a]	15B	199	199	−	−	−	−	−[c]	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	16
ST236-19F[a]	19F	236	271	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	−	9
ST309-19F	19F	309	309	−	−	−	−	+	−	−	−	−	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	22
ST425-19F	19F	425	425	−	−	+	−	+	−	+	−	−	−	−	−	−	+	−	−	+	−	−	+	+	+	−	+	+	23
ST36-23F	23F	36	439	−	−	−	−	+	−	−	−	+	−	−	−	−	−	−	−	−	−	−	+	−	−	−	−	−	26
ST81-23F[a]	23F	81	81	−	−	+	−	+	−	+	+	−	−	−	−	−	+	−	−	−	−	−	+	−	+	−	−	−	27
ST558-35B	35B	558	558	−	−	−	−	+	−	−	+	+	−	−	−	−	+	+	−	−	−	−	+	−	+	−	−	−	24

[a] Internationally recognized clones, as determined by the Pneumococcal Molecular Epidemiology Network (www.sph.emory.edu/PMEN/pmen_ww_spread_clones.html) as of April 2006.

[b] T4, genes found in TIGR4 but not in R6; R6, genes found in R6 but not in TIGR4.

[c] Negative according to DNA detection but positive with PCR and Southern blotting.

[d] Data for isolates belonging to the same clonal complex are shown in bold.

complex were genetically highly related. However, ST124^{-14}-1 was more genetically related to ST307^{-14} than to ST124^{-14}-2. Four to 44 genes differed between the isolates belonging to the ST176 clonal complex, and within the ST180 clonal complex, 4 genes differed between the two type 3 isolates. These genes were not present in any accessory region and were scattered around the genome. One to 17 genes differed between the six ST156 isolates, compared to 10 to 53 genes between isolates of different STs within the ST162 clonal complex.

The number of base pairs that differed in the seven alleles responsible for the difference between STs was more than one in most cases, suggesting recombinational events as the main mechanism creating allelic diversity (Table 1). Recombination events are most likely also the cause of differences in gene content among different pneumococcal isolates. Our microarray data suggest that different isolates of clones that cause mainly invasive disease and are rarely found among carriers, i.e., ST306 and ST191, of types 1 and 7F, respectively, are more genetically related than isolates belonging to clones common among carriers, such as the ST176 clonal complex, suggesting that fewer recombination events are associated with the former group of isolates not found in carriage and hence less exposed to cocolonization opportunities.

DISCUSSION

In this study, we compared the genetic contents of 40 clinical pneumococcal isolates of different serotypes and STs and related the data obtained to clonal complexes determined by MLST. Few microarray studies of these organisms have been performed previously and have usually included a limited number of serotypes and clones (6, 10). Hence, to our knowledge, this is the first study including a wide variety of serotypes and clones with different invasive disease potentials. Here we show both inter- and intraclonal differences affecting the interpretation of MLST analysis. We could firmly conclude that sequence differences within housekeeping genes, not serotype, may predict the extent of gene content differences in pneumococcal isolates.

We found that the microarray data obtained correlated well with the eBURST groups identified using MLST. However, serotype did not predict genetic relatedness unless the isolates belonged to the same clonal complex. This is in agreement with previous studies (3) investigating MLST and serotype diversity and is important knowledge, especially since the most used method to study epidemiology is serotyping. Also, the available vaccines today target the capsular polysaccharide, and there are several discussions ongoing regarding the threat of immune escape through serotype switching, meaning that clones identified using molecular techniques such as MLST may emerge with nonvaccine capsular types.

We also observed gene content differences within clones, which is in agreement with a previous study by Silva et al. (10). In that study, a limited number of serotypes, mainly type 14, were characterized using microarray analysis, and intraclonal variations (the same ST and serotype) were found to affect outcomes in animal models. Interestingly, we found that intraclonal differences were more rare among serotypes and clones commonly causing invasive disease and rarely causing carriage (types 1 and 7F) than among clones common among carriers

(types 6B and 9V) (1, 2, 9). However, the type 1 isolates used in this study were from Sweden, and infections with these clones have not led to fatal outcomes (12). In contrast, a different clonal complex of ST217 dominated invasive serotype 1-mediated disease in sub-Saharan Africa, leading to high lethality (14). Hence, we included a Swedish serotype 1 isolate of ST217 in the study. In comparing isolates of ST306 and ST217, we found that as many as 166 to 173 genes differed, in contrast to no differences between the two ST306 isolates. Thus, unrelated clones of serotype 1 were not more genetically similar to one another than to isolates of carrier serotypes of unrelated STs. This was also true for other serotypes investigated in this study.

Our microarray data show that the pattern of presence/absence of accessory genes representing at least 30% of the pneumococcal genome (corresponding to around 600 genes) (unpublished data) predicts strain relatedness in the same way that sequencing of parts of housekeeping genes of the core genome does. It has been estimated that approximately 90% of the allelic variations seen in the seven housekeeping genes used for MLST of pneumococci can be attributed to homologous recombination (5). Most likely, the absence/presence of accessory genes and regions is also due to horizontal gene transfer events and recombination. It is therefore not surprising that both MLST and gene content patterns obtained from microarrays predict the same relatedness among strains. In comparing the presence or absence of a subset of 25 genes belonging to accessory regions distributed around the genome among the 40 isolates, we still found a strong positive correlation with the MLST data. Therefore, a multiplex PCR of a subset of accessory genes might provide a cheaper and simpler way of identifying genetic relatedness among pneumococcal isolates. With increasing knowledge of the actual roles played by different accessory genes/regions, this typing method may also yield additional information on disease potential and transmissibility (11). However, further studies using an even larger collection of isolates and extended microarrays including additional pneumococcal accessory genes (8) are needed to fully evaluate this novel approach of studying relatedness among pneumococci.

REFERENCES

1. **Brueggemann, A. B., D. T. Griffiths, E. Meats, T. Peto, D. W. Crook, and B. G. Spratt.** 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.
2. **Brueggemann, A. B., T. E. Peto, D. W. Crook, J. C. Butler, K. G. Kristinsson, and B. G. Spratt.** 2004. Temporal and geographic stability of the serogroup-specific invasive disease potential of Streptococcus pneumoniae in children. J. Infect. Dis. **190:**1203–1211.
3. **Enright, M. C., and B. G. Spratt.** 1998. A multilocus sequence typing scheme

for Streptococcus pneumoniae: identification of clones associated with serious invasive disease. Microbiology **144:**3049–3060.

4. **Feil, E. J., B. C. Li, D. M. Aanensen, W. P. Hanage, and B. G. Spratt.** 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. **186:**1518–1530.

5. **Feil, E. J., J. M. Smith, M. C. Enright, and B. G. Spratt.** 2000. Estimating recombinational parameters in Streptococcus pneumoniae from multilocus sequence typing data. Genetics **154:**1439–1450.

6. **Hakenbeck, R., N. Balmelle, B. Weber, C. Gardes, W. Keck, and A. de Saizieu.** 2001. Mosaic genes and mosaic chromosomes: intra- and interspecies genomic variation of *Streptococcus pneumoniae*. Infect. Immun. **69:**2477–2486.

7. **Henriques Normark, B., B. Christensson, A. Sandgren, B. Noreen, S. Sylvan, L. G. Burman, and B. Olsson-Liljequist.** 2003. Clonal analysis of Streptococcus pneumoniae nonsusceptible to penicillin at day-care centers with index cases, in a region with low incidence of resistance: emergence of an invasive type 35B clone among carriers. Microb. Drug Resist. **9:**337–344.

8. **Hiller, N. L., B. Janto, J. S. Hogg, R. Boissy, S. Yu, E. Powell, R. Keefe, N. E. Ehrlich, K. Shen, J. Hayes, K. Barbadora, W. Klimke, D. Dernovoy, T. Tatusova, J. Parkhill, S. D. Bentley, J. C. Post, G. D. Ehrlich, and F. Z. Hu.** 2007. Comparative genomic analyses of seventeen *Streptococcus pneumoniae* strains: insights into the pneumococcal supragenome. J. Bacteriol. **189:**8186–8195.

9. **Sandgren, A., K. Sjostrom, B. Olsson-Liljequist, B. Christensson, A. Samuelsson, G. Kronvall, and B. Henriques Normark.** 2004. Effect of clonal and serotype-specific properties on the invasive capacity of Streptococcus pneumoniae. J. Infect. Dis. **189:**785–796.

10. **Silva, N. A., J. McCluskey, J. M. Jefferies, J. Hinds, A. Smith, S. C. Clarke, T. J. Mitchell, and G. K. Paterson.** 2006. Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. Infect. Immun. **74:**3513–3518.

11. **Sjostrom, K., C. Blomberg, J. Fernebro, J. Dagerhamn, E. Morfeldt, M. A. Barocchi, S. Browall, M. Moschioni, M. Andersson, F. Henriques, B. Albiger, R. Rappuoli, S. Normark, and B. H. Normark.** 2007. Clonal success of piliated nonsusceptible pneumococci. Proc. Natl. Acad. Sci. USA **104:**12907–12912.

12. **Sjostrom, K., C. Spindler, A. Ortqvist, M. Kalin, A. Sandgren, S. Kuhlmann-Berenzon, and B. Henriques-Normark.** 2006. Clonal and capsular types decide whether pneumococci will act as a primary or opportunistic pathogen. Clin. Infect. Dis. **42:**451–459.

13. **Smyth, G. K.** 2004. Linear models and empirical Bayes methods for assessing differential expression in microarray experiments. Stat. Appl. Genet. Mol. Biol. **3:**Article3.

14. **Yaro, S., M. Lourd, Y. Traore, B. M. Njanpop-Lafourcade, A. Sawadogo, L. Sangare, A. Hien, M. S. Ouedraogo, O. Sanou, I. Parent du Chatelet, J. L. Koeck, and B. D. Gessner.** 2006. Epidemiological and molecular characteristics of a highly lethal pneumococcal meningitis epidemic in Burkina Faso. Clin. Infect. Dis. **43:**693–700.