Genomic Diversity between Strains of the Same Serotype and Multilocus Sequence Type among Pneumococcal Clinical Isolates

Nuno A. Silva,[1]† Jackie McCluskey,[1]† Johanna M. C. Jefferies,[1,2] Jason Hinds,[3] Andrew Smith,[4] Stuart C. Clarke,[2] Tim J. Mitchell,[1]* and Gavin K. Paterson[1]‡

Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom[1]; Scottish Meningococcus and Pneumococcus Reference Laboratory, Stobhill Hospital, Glasgow, G21 3UW, United Kingdom[2]; Department of Medical Microbiology, St. George's Hospital Medical School, Cranmer Terrace, London SW17 0RE, United Kingdom[3]; and Infection Research Group, Level 9, Glasgow Dental School, University of Glasgow, Glasgow G2 3JZ, United Kingdom[4]

The important human pathogen *Streptococcus pneumoniae* is known to be a genetically diverse species. We have used comparative genome hybridization (CGH) microarray analysis to investigate this diversity in a collection of clinical isolates including several capsule serotype 14 pneumococci, a dominant serotype among disease isolates. We have identified three new regions of diversity among pneumococcal isolates and, importantly, clearly demonstrate genetic differences between strains of the same multilocus sequence type (ST) and capsule serotype. CGH may therefore, under certain circumstances, prove to be a valuable tool to supplement current typing methods. Finally, we show that these clonal strains with the same serotype and ST behave differently in an animal model. Strains of the same ST and serotype therefore have important genetic and phenotypic differences.

Although found as a commensal of the upper respiratory tract, the gram-positive bacterium *Streptococcus pneumoniae* (the pneumococcus) remains an important cause of human morbidity and mortality worldwide. Pneumonia, sepsis, meningitis, and otitis media are the predominant diseases caused by the pneumococcus. Drawbacks with the current vaccines and the increase and spread of antibiotic resistance hamper prevention and treatment and highlight the need for improved understanding of pneumococcal biology (4, 30). Naturally transformable, pneumococcal genetic diversity and plasticity are evidenced by the presence of over 90 distinct capsular serotypes and the emergence of antibiotic resistance. Indeed, pneumococcal genetic diversity and genetic exchange with related organisms make it hard to clearly define the pneumococcus as a species (2, 12, 29). Along with this genetic diversity come important phenotypic differences with regards to the propensity of strains and serotypes to cause disease. For example, ~85% of disease is caused by only 20 different serotypes (18). In addition, certain multilocus sequence types (ST) are more associated with disease than others (6). Although, capsular serotype is recognized as a crucial contributing factor in these differences (6), other as yet uncharacterized genetic factors also contribute (20, 25, 26). The advent of genome sequencing and microarray technology has allowed this genetic diversity to be probed more fully, offering the potential to better understand pneumococcal strain and serotype differences. In addition to helping to understand carriage and disease processes, such data may also contribute to antimicrobial

and vaccine development through the identification of conserved targets found in all strains or serotypes. Furthermore, understanding the pneumococcal population structure may help predict and interpret its response to interventions such as antibiotics or vaccines, especially when these may be effective against only a subset of strains or serotypes. Previously, comparative genome hybridization (CGH) microarray analysis of 19 pneumococcal strains found that in any one strain 8 to 10% of genes were divergent from the reference sequenced strain TIGR4, with the pool of nonconserved genes represented approximately 20% of the TIGR4 genome (11). The majority of these genes were located in 25 clusters of TIGR4 genes (5). To further characterize pneumococcal genetic diversity, we have performed CGH using 13 test strains comprising commonly used laboratory strains and clinical isolates. We confirm previous data showing significant strain-to-strain genetic variation, and, importantly, we find that strains of a major disease causing clone with the same multilocus sequence type (ST124) and serotype (14) differ in gene content. Therefore, even strains that appear identical using high-resolution typing methods carry genetic differences that might impact their biology. CGH may therefore, under certain circumstances, prove to be a valuable tool to supplement current typing methods. Furthermore, we show that isolates with the same ST and serotype behave differently in an infection model, supporting the proposal that genetic differences within the same serotype ST clone may affect biology.

MATERIALS AND METHODS

Bacterial strains, culture conditions, and DNA extraction. The *S. pneumoniae* strains used for this study are described in Table 1. The strains were grown in brain heart infusion broth (BHI; Oxoid) and stored as frozen stocks at −80°C in BHI with 12% glycerol (vol/vol). For DNA preparations, cells were grown statically to mid-logarithmic phase (optical density at 600 nm of 0.6) in BHI at 37°C and harvested by centrifugation for 10 min at 5,000 rpm. DNA was isolated using QIAGEN 100/G genomic tips, following the manufacturer's instructions.

* Corresponding author. Mailing address: Division of Infection and Immunity, Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow G12 8QQ, United Kingdom. Phone: 44 141 330 3749. Fax: 44 141 330 3727. E-mail: t.mitchell@bio.gla.ac.uk.

† N.A.S. and J.M. contributed equally to this study.

‡ Present address: Biomedical Sciences, DSTL, Porton Down, Salisbury SP4 0JQ, United Kingdom.

TABLE 1. *S. pneumoniae* strains used in this study

Strain	Specimen	Serotype	ST	Yr of isolation	Location	Reference
TIGR4	Blood	4	205	NA[a]	Norway	1, 31
0100993	NA	3	180	NA	NA	32
R6	Laboratory strain derived from in vitro passage	2	128	1950s	NA	23
D39	NA	2	128	Deposited in NCTC in 1948	NA	3
P11	Blood	14	9	2003	Aberdeen, Scotland	17
N16	Blood	14	9	2000	Dundee, Scotland	17
P33	Eye	14	9	2001	Dumfries, Scotland	17
10	Blood	14	124	2000	Aberdeen, Scotland	17
48	Blood	14	124	2000	Glasgow, Scotland	17
50	Blood	14	124	2001	Glasgow, Scotland	17
P49	Blood	3	180	2002	Glasgow, Scotland	17
PMEN7	NA	19A	75	<1997	South Africa	28
PMEN13	NA	19A	41	<1997	South Africa	28
PMEN23	NA	6A	37	1994–2000	North Carolina	24

[a] NA, data not available or not known.

Microarray construction, DNA labeling, and hybridization. The pneumococcal microarrays employed in this study have been used previously (19). Briefly, microarrays were constructed by robotic spotting of PCR amplicons onto poly-L-lysine-coated glass microscope slides (MicroGrid II; BioRobotics, Huntingdon, United Kingdom) (15). Amplicons were designed to represent each of the annotated open reading frames (ORFs) present in *S. pneumoniae* strain TIGR4 (31) in addition to those unique to the other sequenced strain, R6. The process essentially designed multiple amplicons using Primer3 for all TIGR4 ORFs, as determined by automated analysis of BLASTN comparisons (16). A single amplicon was selected to represent each ORF based on its lack of similarity to other ORFs on the array using BLASTN analysis to ensure minimal cross-hybridization. DNA was fluorescently labeled and hybridized to the microarray chips as described previously (8). Essentially, fluorochrome Cy3 or Cy5 dCTP (Amersham Pharmacia Biotech) was incorporated into whole-genomic DNA by a randomly primed polymerization reaction using large fragment DNA polymerase I (Invitrogen). Whole-genomic DNA comparisons were carried out by competing DNA from the test strains listed in Table 1 with a standard reference DNA from TIGR4. Microarray slides were prehybridized for 20 min at 65°C in a buffer containing 3.5× SSC (1× SSC is 0.15 M NaCl plus 0.015 M sodium citrate), 0.1% sodium dodecyl sulfate (SDS), and 10 mg/ml bovine serum albumin. Labeled DNAs from TIGR4 and the test strains were mixed and purified using a Minielute kit (QIAGEN), after which 4× SSC and 0.3% SDS were added. After denaturation at 95°C for 2 min, the DNA mixture was applied to the microarray and hybridized during 18 h at 65°C. Before analysis, the slides were washed once in 1× SSC buffer with 0.06% SDS for 5 min at 65°C and twice in 0.06× SSC buffer at room temperature.

Comparative hybridizations were performed between a Cy5-labeled genomic DNA of the reference strain (TIGR4) in competition with Cy3-labeled genomic DNA. Reciprocal dye swap experiments were performed to minimize labeling artifacts.

Data generation and analysis. Hybridized slides were scanned using a ScanArray (Packard BioScience) according to the manufacturer's instructions, and the median pixel intensity values for each element on the array were quantified using Quantarray (Packard BioScience). The data were further analyzed using Gene-Spring 7.0 (Silicon Genetics). On the basis of preliminary PCR validations, we determined a cutoff for the normalized intensity ratio of >1.5 indicated the absence of a gene, unless the intensity was >1,500 in both channels (TIGR4 and test strain), in which case we decided that the gene is present. For intensities of <600 for both channels, the result was regarded as ambiguous and therefore was checked by PCR. Genes with a ratio of 1.4 gave poor agreement between the array and PCR results, and so the presence or absence of all such genes was determined by PCR.

PCR validation of microarray results. Validation of array results was performed by PCR using the primers employed to generate the microarray probe.

Infection model. Strains were grown in BHI to mid-log phase, and glycerol stocks (12% [vol/vol]) were prepared and stored at −80°C. For the infection, a vial was rapidly thawed at 37°C and cells were collected by centrifugation at 13,000 rpm for 3 min and resuspended in phosphate-buffered saline to give a concentration of 10^8 CFU/ml. One hundred microliters of this bacterial suspension was administered by intraperitoneal injection of mice. Infected mice were

monitored regularly for clinical signs, and tail bleeds were taken from the superficial tail vein. C57/BL6 mice (female, 5 weeks old when infected) were purchased from Harlan Olac, Bicester, United Kingdom. Mice with bacterial counts below the detection limit (~100 CFU/ml) were ascribed a value just beneath that limit. All animal work was conducted with the appropriate local and Home Office approval and licensing.

RESULTS AND DISCUSSION

Regions of genetic diversity among clinical isolates. To further explore the genetic diversity of the important human pathogen *S. pneumoniae*, we employed comparative genome hybridization microarray analysis to examine gene content in a panel of clinical isolates and laboratory strains. Thirteen test strains were chosen to represent seven serotypes and eight multilocus sequence types (Table 1). These included six serotype 14 strains, currently a prominent serotype among disease isolates in the United Kingdom and elsewhere (6, 7, 13, 18). Importantly, strains of the same serotype and multilocus sequence type were included to investigate the level of diversity among these apparently closely related strains. The three reference strains R6, D39, and 0100093 were also included.

Twenty-five regions of diversity (RDs) totaling 248 genes were identified in which three or more contiguous genes (according to the TIGR4 annotation) were not conserved in all strains (Table 2). Importantly, there was strong agreement between the array results and the sequenced R6 genome sequence. Of the 248 RD genes identified, the array agreed with the result expected based on the R6 genome sequence for 242 of these genes (97.5%). In addition, the distribution of 207 genes from the other test strains was investigated by PCR. Of these 207, 197 (95.2%) showed agreement between the microarray and PCR. Together, these data confirm the microarray analysis to be a good predictor of the presence or absence of genes or probe sequences.

The RDs ranged considerably in size from 1.7 to 34.8 kb. Nineteen of the RDs identified here have previously reported by Bruckner et al. (5). We therefore confirm these previous findings showing these regions to be variable between strains and extend their data to a new collection of isolates. The limits of these RDs or clusters vary a little between the studies and likely reflect the use of different clinical strains or array probes.

TABLE 2. Regions of genetic diversity among clinical isolates

RD	Approximate size (kb)	SP no. of variable genes	Cluster[a] (SP no. of variable genes)
1	9	SP0067-0074	1 (SP0067-0072)
2	5.8	SP0109-0115	1* (SP0109-0113)
3	5.6	SP0163-0168	2 (SP0163-0171)
4	14.2	SP0346-0360	3 (SP0347-0360)
5	3.3	SP0378-0380	Identified but not as cluster
6	5.4	SP0394-0397	4* (SP0388-0399)
7	12.6	SP0460-0468	4 (SP460-470)
8	7.1	SP0473-0478	5* (SP0473-0478)
9	5.6	SP0531-0544	5 (SP0531-0544)
10	11	SP0643-0648	Identified but not as cluster
11	8.0	SP0664-0666	Identified but not as cluster
12	4.4	SP0692-0700	7* (SP691-698)
13	1.7	SP0888-0891	6 (SP887-891)
14	7.9	SP0949-0954	New
15	11.9	SP1050-1065	7 (SP1054-1064)
16	9.2	SP1129-1147	8 (SP1129-1146)
17	33.7	SP1315-1352	9 and 10* (SP1309-1352)
18	12.1	SP1433-1444	11* (SP1433-1437)
19	10.3	SP1612-1622	10 (SP1611-1622)
20	34.8	SP1756-1773	11 (SP1755-1773)
21	5.3	SP1793-1799	12 (SP1791-1794 and 1796-1798)
22	3.2	SP1828-1830	13* (SP2159-2166)
23	3.2	SP1911-1918	New
24	9.4	SP1948-1955	12* (SP1947-1958)
25	5.3	SP2159-2166	13* (SP2159-2166)

[a] Relates to work by Bruckner et al. (5). Clusters of strain-specific genes identified by Bruckner et al. but not found in this study are clusters 2*, 3*, 6*, 8*, and 9* (see Table 3 in reference 5).

In addition we identify three novel RDs not previously recognized: RD14, -22, and -23. These are described in detail here.

RD14. RD14 covers ~8 kb and six genes in TIGR4, SP0949 to -0954. SP0949 encodes a transposase (IS1515), and although predicted to be inactive in TIGR4, due to a frameshift mutation introducing a premature stop codon, this element may be responsible for the unequal strain distribution of this region. SP0950 encodes a predicted GNAT family acetyltransferase (GCN5-related *N*-acetyltransferase), as does SP0953, although they share only limited homology. SP0954 encodes the CeLA competence protein. SP0951 encodes a conserved hypothetical protein, which contains a putative TfoX N-terminal domain. Identified in *Haemophilus influenzae*, the TfoX/Sxy protein is essential for transformation in that species (33, 34). No data exist as yet for a role for SP0951 in pneumococcal transformation, and microarray analysis of the global response to competence-stimulating peptide in TIGR4 did not identify SP0951 as being a competence-stimulating peptide-responsive gene (21). The remaining gene in RD14, SP0952, is annotated as encoding an alanine dehydrogenase carrying an authentic frameshift resulting in a premature stop codon.

The array data shows only three strains carry all genes in this RD: PMEN7, PMEN23, and P49. Six strains contain all genes with the exception of SP0949. RD14 is missing entirely in strains P11, P33, and N16, except for SP0954, which is present in P33 and N16. The remaining strain, PMEN7, lacks only SP0951 and SP0952.

None of the genes in this region were identified as virulence factors in the signature-tagged mutagenesis (STM) screen of TIGR4, and so their role in virulence is unclear as yet.

RD22. RD22 comprises three genes, the SP01828 to -1830, in a 3.2-kb region of the TIGR4 genome. All three genes were present in R6, D39, PMEN13, and PMEN23. In contrast, all were absent in N16, 10, 48, 50, 0100993, P49, and PMEN7. In the cases of P11 and P33, only SP1830 was found to be present. The genes are annotated in numerical order as coding for UDP-glucose 4-epimerase (*galE*), galactose-1-phosphate uridylyltransferase (*galT*), and phosphate transport system regulatory protein (*phoU*). The latter gene was identified in the TIGR4 STM screen, and when the mutant was analyzed further in competitive infections with the wild type, it was attenuated in models of pneumonia, bacteremia, and nasal colonization (14). Besides this, these pneumococcal genes are uncharacterized, but the STM data do show a potential for the selected distribution of these genes to influence the behavior of strains.

RD23. The eight SP1911 to SP1918 genes make up RD23 and cover ~3.2 kb of the TIGR4 genome. This region is fully present in 12 strains. One other strain, PMEN7, lacks only a single gene, SP1914. However, in strain 50, this entire region is missing. This region is poorly characterized, with four of the genes annotated as encoding hypothetical proteins; the significance of the absence of this region in strain 50 is therefore unclear. A function for one of these hypothetical proteins, coded for by SP1915, is suggested by the presence of a LytTr DNA-binding domain found in various bacterial transcriptional factors. The remaining genes are annotated as coding for a putative thioredoxin (SP1911), a cspC (cold shock protein)-related protein with an authentic point mutation resulting in a premature stop codon (SP1913), a PAP2 family protein (SP1916), a family of mainly phosphatase enzymes (PF-01569), and an ATP-binding protein (SP1918). None of the genes in this region was identified in the STM screen of TIGR4 (14).

Bruckner et al. (5) identified five clusters that we have not observed (Table 2). Presumably, this is due to the use of different strains, while the analysis of further strains would allow discovery of other RDs.

Diversity between strains of the same multilocus sequence type and serotype. Multilocus sequence typing is now widely used to type the pneumococcus and other pathogens, providing high-resolution discrimination of a large number of clones (9, 10, 33). Strains of the same ST are assumed to be clonal and to have descended from a recent common ancestor. Strains of the same ST can be of different capsular serotypes, showing that such strains are not necessarily identical despite being of the same ST. However, genetic differences in addition to the capsule locus have not been extensively characterized for strains of the same ST. The distribution of RD genes in this study provides a first example of the phenomenon of differences between strains of the same ST extending to noncapsular genes. The example involves RD17 and 23 and the three ST124 serotype 14 strains: 10, 48, and 50 (Fig. 1). In the case of RD17, it is present in its entirety in strain 10, but the first half of this region is absent in strains 48 and 50. RD23, which is a new region of diversity (see above), is absent in strain 50 but present in strains 10 and 48. Importantly, validation by PCR showed a strong agreement with the microarray results with 105/108 (97%) genes agreeing between the two methods (Fig. 1).

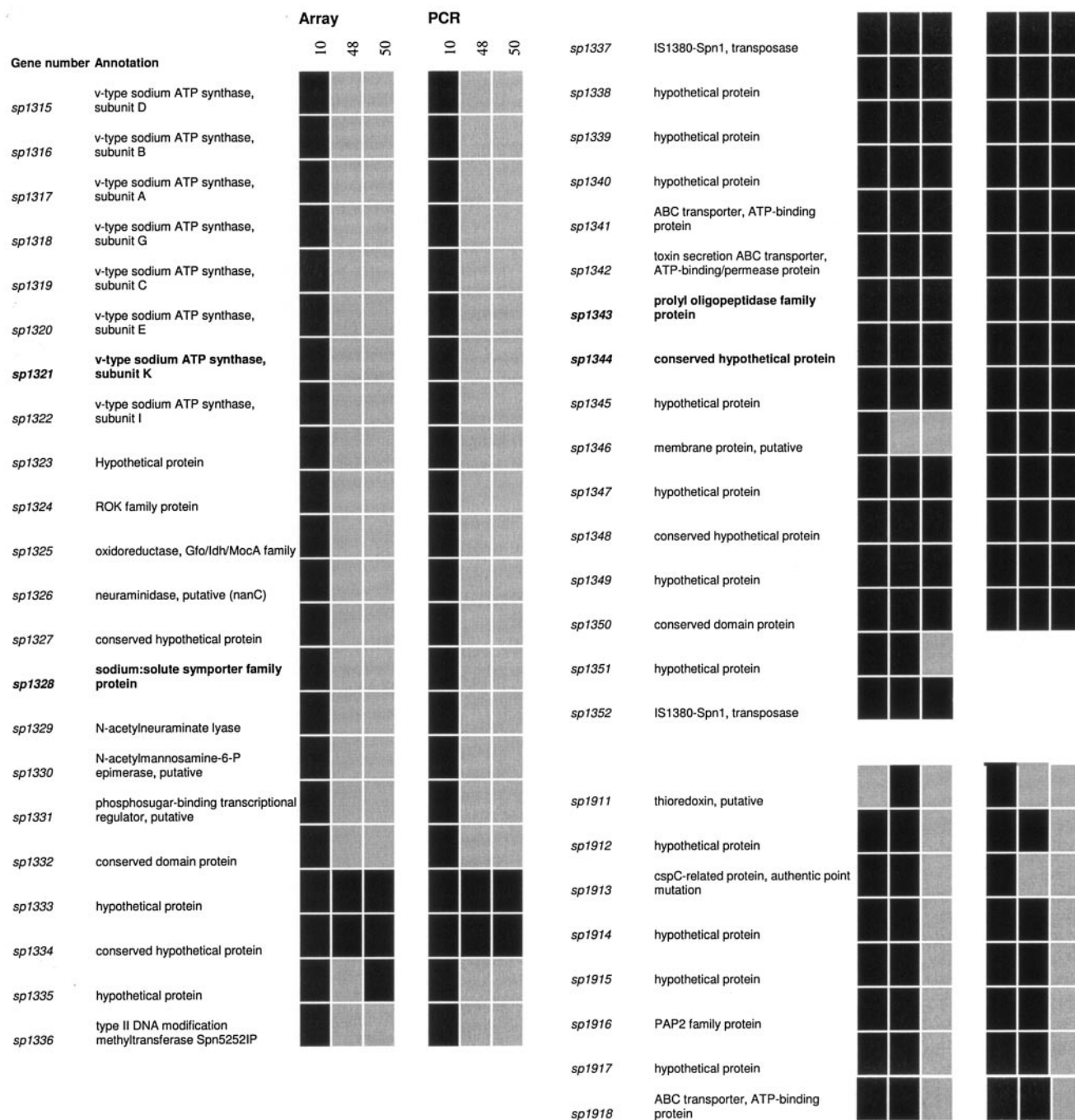The biological significance of these differences is uncertain,

FIG. 1. Genetic differences between pneumococcal strains of the same ST and serotype. Strain distribution of RD17 (SP1315 to -1352) and RD23 (SP1911 to 1918) genes as determined by microarray and PCR analysis. Black, positive; gray, negative. The gene number and annotation are taken from the TIGR4 genome at http://www.tigr.org. Genes and products highlighted in bold were identified as virulence factors in the TIGR4 STM screen (14).

however, as both RD17 and -23 are poorly characterized. Within these regions, the TIGR4 STM screen identified two genes (SP1321 and SP1328) with unequal strain distributions as pneumococcal virulence factors (14). There is therefore the potential for these genetic differences to effect phenotypic differences between these very similar strains.

Thus, strains of the same ST and serotype have genetic differences; although perhaps not a surprising finding, this is the first clear demonstration of this phenomenon. Microarray analysis may therefore be of utility in complementing current multilocus sequence typing and serotyping schemes by providing a higher resolution.

Analysis of virulence of strains of the same ST and serotype. To test if the genetic differences between strains with the
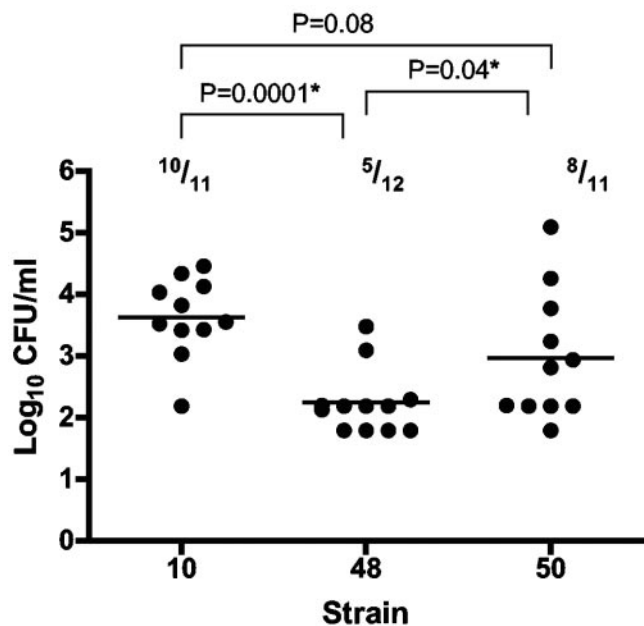
FIG. 2. Blood bacterial counts 6 h postinfection. Three pneumococcal strains of the same ST and serotype (ST124 serotype 14) were injected by the intraperitoneal route into C57/BL6 mice (dose of 10^7 CFU), and the blood bacterial viable count was taken at 6 h postinfection. Each point indicates the data from an individual mouse; the horizontal bar indicates the mean ($n = 11$ to 12 pooled from two experiments giving similar results). The proportions indicate the number of mice that had bacteremia above the detection limit ($\sim$log 2 CFU/ml). P values (Student's t test) relate to bacterial counts, with ≤ 0.05 considered significant (*).

same ST and serotype could impact biological significance, the three strains 10, 48, and 50 were tested for virulence in a mouse intraperitoneal infection model. Young (5 week old) female C57/BL6 mice were infected by the intraperitoneal route with 10^7 CFU, and survival and blood counts were monitored. All mice survived the infection, and none showed clinical signs ($n = 7$ to 8). However, at 6 h postinfection, a transient bacteremia was noted that was cleared by 24 h. Comparison of the blood bacterial counts at 6 h shows a significant difference between the strains (Fig. 2). The blood counts of mice infected with strain 48 were significantly lower than those of mice infected with strain 10 ($P < 0.0001$) or strain 50 ($P = 0.0441$). The mean bacterial count for strain 48 was approximately 24-fold lower than that for strain 10 and 5-fold lower than that for strain 50. In line with this, strain 10 had the lowest proportion of bacteremic animals, Fig. 2. Although there was a trend toward higher bacterial blood counts in mice infected with strain 10 compared to those infected with strain 50, the difference was not statistically significant ($P = 0.0882$). Therefore, these strains, despite being of the same serotype and ST, show differences in virulence in this mouse model. A causal relationship between genotypic differences and phenotype remains, however, to be confirmed empirically. A trend for differences in the behavior of strains of the same ST and serotype during mouse infections was recently shown, but not examined further, by Sandgren et al. (25). For example, two ST162 serotype 19F strains showed different propensities to cause pneumonia following intranasal infection. One strain

caused pneumonia in 80% of infected C57BL/6 and BALB/c mice, while the proportions for a second strain were 53% and 40%, respectively.

Concluding remarks. Although microarray analysis allows the whole genome to be interrogated easily, there are several caveats to be acknowledged. First, the true degree of population diversity is underestimated because test strain-specific genes are not included. How many genes do the test strains carry that are absent in TIGR4? Indeed, sequencing of genomic libraries from eight pneumococcal clinical isolates revealed a number of putative ORFs distinct from TIGR4, with many also unrelated to known streptococcal sequences (27). Importantly, a major shortcoming of investigations like this is the inability to determine if the absence of an array signal represents the absence of a particular gene or divergence in the probe sequences between strains. In addition, array analysis provides no details on gene location or number. For example, a gene may be present in multiple copies or in a different location compared to other strains, but this is overlooked in this analysis. Also, subtle but functionally significant differences will be missed such as promoter and coding sequence mutations that may alter the production and activity of gene products. Likewise, the bases for genetic differences are unclear: i.e., is presence or absence due to acquisition by one strain and not another or loss from one strain and not another? Finally, many of the genes identified here and in similar array analyses are annotated as encoding hypothetical or conserved hypothetical proteins with little or no data available on their function(s). Furthermore, even those with annotations lack functional confirmation. The potential significance of the absence or presence of these genes is therefore hard to interpret until they have been characterized more fully. However, acknowledging these drawbacks, we have employed CGH to identify large genomic regions of diversity between pneumococcal strains. We confirm the previous identification of several variable regions and identify three addition regions that are not conserved among strains. We provide a clear demonstration of genetic differences between strains of the same serotype and ST. In addition, we show differences in the virulence of these strains in a mouse infection model. Thus, even although strains may appear identical based on current typing methods, they may boast potentially important genetic and phenotypic differences. CGH may therefore be useful in providing higher-resolution typing. This will especially be valuable when particular virulence-associated genes and genotypes are identified: for example, as done recently for otitis media (22).

REFERENCES

1. **Aaberge, I. S., J. Eng, G. Lermark, and M. Lovik.** 1995. Virulence of *Streptococcus pneumoniae* in mice: a standardized method for preparation and frozen storage of the experimental bacterial inoculum. Microb. Pathog. **18:** 141–152.
2. **Arbique, J. C., C. Poyart, P. Trieu-Cuot, G. Quesne, M. D. G. S. Carvalho, A. G. Steigerwalt, R. E. Morey, D. Jackson, R. J. Davidson, and R. R. Facklam.** 2004. Accuracy of phenotypic and genotypic testing for identifica-

tion of *Streptococcus pneumoniae* and description of *Streptococcus pseudopneumoniae* sp. nov. J. Clin. Microbiol. **42:**4686–4696.

3. Avery, O. T., C. M. MacLeod, and M. McCarty. 1979. Studies on the chemical nature of the substance inducing transformation of pneumococcal types. Inductions of transformation by a desoxyribonucleic acid fraction isolated from pneumococcus type III. J. Exp. Med. **149:**297–326.

4. Bogaert, D., P. W. Hermans, P. V. Adrian, H. C. Rumke, and R. de Groot. 2004. Pneumococcal vaccines: an update on current strategies. Vaccine **22:**2209–2220.

5. Bruckner, R., M. Nuhn, P. Reichmann, B. Weber, and R. Hakenbeck. 2004. Mosaic genes and mosaic chromosomes-genomic variation in *Streptococcus pneumoniae*. Int. J. Med. Microbiol. **294:**157–168.

6. 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.

7. Denham, B. C., and S. C. Clarke. 2005. Serotype incidence and antibiotic susceptibility of *Streptococcus pneumoniae* causing invasive disease in Scotland, 1999–2002. J. Med. Microbiol. **54:**327–331.

8. Dorrell, N., J. A. Mangan, K. G. Laing, J. Hinds, D. Linton, H. Al-Ghusein, B. G. Barrell, J. Parkhill, N. G. Stoker, A. V. Karlyshev, P. D. Butcher, and B. W. Wren. 2001. Whole genome comparison of *Campylobacter jejuni* human isolates using a low-cost microarray reveals extensive genetic diversity. Genome Res. **11:**1706–1715.

9. 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.

10. Enright, M. C., and B. G. Spratt. 1999. Multilocus sequence typing. Trends Microbiol. **7:**482–487.

11. Hakenbeck, R., N. Balmelle, B. Weber, C. Gardès, 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.

12. Hanage, W. P., T. Kaijalainen, E. Herva, A. Saukkoriipi, R. Syrjänen, and B. G. Spratt. 2005. Using multilocus sequence data to define the pneumococcus. J. Bacteriol. **187:**6223–6230.

13. Hanage, W. P., T. H. Kaijalainen, R. K. Syrjänen, K. Auranen, M. Leinonen, P. H. Mäkelä, and B. G. Spratt. 2005. Invasiveness of serotypes and clones of *Streptococcus pneumoniae* among children in Finland. Infect. Immun. **73:**431–435.

14. Hava, D. L., and A. Camilli. 2002. Large-scale identification of serotype 4 *Streptococcus pneumoniae* virulence factors. Mol. Microbiol. **45:**1389–1406.

15. Hinds, J., K. G. Laing, J. A. Mangan, and P. D. Butcher. 2000. Glass slide microarray for bacterial genomes, p. 83–99. *In* B. W. Wren and N. Dorell (ed.), Methods in microbiology: functional microbial genomics. Academic Press, London, United Kingdom.

16. Hinds, J., A. A. Witney, and J. K. Vass. 2000. Microarray design for bacterial genomes, p. 67–82. *In* B. W. Wren and N. Dorell (ed.), Methods in microbiology: functional microbial genomics. Academic Press, London, United Kingdom.

17. Jefferies, J. M. C., A. Smith, S. C. Clarke, C. Dowson, and T. J. Mitchell. 2004. Genetic analysis of diverse disease-causing pneumococci indicates high levels of diversity within serotypes and capsule switching. J. Clin. Microbiol. **42:**5681–5688.

18. Kalin, M. 1998. Pneumococcal serotypes and their clinical relevance. Thorax **53:**159–162.

19. McCluskey, J., J. Hinds, S. Husain, A. Witney, and T. J. Mitchell. 2004. A two-component system that controls the expression of pneumococcal surface antigen A (PsaA) and regulates virulence and resistance to oxidative stress in *Streptococcus pneumoniae*. Mol. Microbiol. **51:**1661–1675.

20. Mizrachi Nebenzahl, Y., N. Porat, S. Lifshitz, S. Novick, A. Levi, E. Ling, O. Liron, S. Mordechai, R. K. Sahu, and R. Dagan. 2004. Virulence of *Streptococcus pneumoniae* may be determined independently of capsular polysaccharide. FEMS Microbiol. Lett. **233:**147–152.

21. Peterson, S. N., C. K. Sung, R. Cline, B. V. Desai, E. C. Snesrud, P. Luo, J. Walling, H. Li, M. Mintz, G. Tsegaye, P. C. Burr, Y. Do, S. Ahn, J. Gilbert, R. D. Fleischmann, and D. A. Morrison. 2004. Identification of competence pheromone responsive genes in *Streptococcus pneumoniae* by use of DNA microarrays. Mol. Microbiol. **51:**1051–1070.

22. Pettigrew, M. M., and K. P. Fennie. 2005. Genomic subtraction followed by dot blot screening of *Streptococcus pneumoniae* clinical and carriage isolates identifies genetic differences associated with strains that cause otitis media. Infect. Immun. **73:**2805–2811.

23. Ravin, A. W. 1959. Reciprocal capsular transformations of pneumococci. J. Bacteriol. **77:**296–309.

24. Richter, S. S., K. P. Heilmann, S. L. Coffman, H. K. Huynh, A. B. Brueggemann, M. A. Pfaller, and G. V. Doern. 2002. The molecular epidemiology of penicillin-resistant *Streptococcus pneumoniae* in the United States, 1994–2000. Clin. Infect. Dis. **34:**330–339.

25. Sandgren, A., B. Albiger, C. J. Orihuela, E. Tuomanen, S. Normark, and B. Henriques-Normark. 2005. Virulence in mice of pneumococcal clonal types with known invasive disease potential in humans. J. Infect. Dis. **192:**791–800.

26. 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.

27. Shen, K., J. Gladitz, P. Antalis, B. Dice, B. Janto, R. Keefe, J. Hayes, A. Ahmed, R. Dopico, N. Ehrlich, J. Jocz, L. Kropp, S. Yu, L. Nistico, D. P. Greenberg, K. Barbadora, R. A. Preston, J. C. Post, G. D. Ehrlich, and F. Z. Hu. 2006. Characterization, distribution, and expression of novel genes among eight clinical isolates of *Streptococcus pneumoniae*. Infect. Immun. **74:**321–330.

28. Smith, A. M., and K. P. Klugman. 1997. Three predominant clones identified within penicillin-resistant South African isolates of *Streptococcus pneumoniae*. Microb. Drug Resist. **3:**385–389.

29. Suzuki, N., M. Seki, Y. Nakano, Y. Kiyoura, M. Maeno, and Y. Yamashita. 2005. Discrimination of *Streptococcus pneumoniae* from viridans group streptococci by genomic subtractive hybridization. J. Clin. Microbiol. **43:**4528–4534.

30. Tan, T. Q. 2003. Antibiotic resistant infections due to *Streptococcus pneumoniae*: impact on therapeutic options and clinical outcome. Curr. Opin. Infect. Dis. **16:**271–277.

31. Tettelin, H., K. E. Nelson, I. T. Paulsen, J. A. Eisen, T. D. Read, S. Peterson, J. Heidelberg, R. T. DeBoy, D. H. Haft, R. J. Dodson, A. S. Durkin, M. Gwinn, J. F. Kolonay, W. C. Nelson, J. D. Peterson, L. A. Umayam, O. White, S. L. Salzberg, M. R. Lewis, D. Radune, E. Holtzapple, H. Khouri, A. M. Wolf, T. R. Utterback, C. L. Hansen, L. A. McDonald, T. V. Feldblyum, S. Angiuoli, T. Dickinson, E. K. Hickey, I. E. Holt, B. J. Loftus, F. Yang, H. O. Smith, J. C. Venter, B. A. Dougherty, D. A. Morrison, S. K. Hollingshead, and C. M. Fraser. 2001. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. Science **293:**498–506.

32. Throup, J. P., K. K. Koretke, A. P. Bryant, K. A. Ingraham, A. F. Chalker, Y. Ge, A. Marra, N. G. Wallis, J. R. Brown, D. J. Holmes, M. Rosenberg, and M. K. Burnham. 2000. A genomic analysis of two-component signal transduction in *Streptococcus pneumoniae*. Mol. Microbiol. **35:**566–576.

33. Williams, P. M., L. A. Bannister, and R. J. Redfield. 1994. The *Haemophilus influenzae sxy-1* mutation is in a newly identified gene essential for competence. J. Bacteriol. **176:**6789–6794.

34. Zulty, J. J., and G. J. Barcak. 1995. Identification of a DNA transformation gene required for com101A+ expression and supertransformer phenotype in *Haemophilus influenzae*. Proc. Natl. Acad. Sci. USA **92:**3616–3620.