Published in final edited form as:

Infect Genet Evol. 2007 December ; 7(6): 708–716.

Assessment of molecular typing methods to determine invasiveness and to differentiate clones of *Streptococcus pneumoniae*

Caroline A. Obert^{1,*}, Geli Gao², Jack Sublett², Elaine I. Tuomanen², and Carlos J. Orihuela³

¹ Hartwell Center for Bioinformatics and Biotechnology, St. Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105

² Department of Infectious Diseases, St. Jude Children's Research Hospital, 332 North Lauderdale, Memphis, TN 38105

³ Department of Microbiology and Immunology, University of Texas Health Science Center at San Antonio, 7703 Floyd Curl Drive, San Antonio, TX 78229

SUMMARY

In the United States *Streptococcus pneumoniae* is the leading cause of community acquired pneumonia and invasive bacterial disease. As antimicrobial resistance increases, it will become critical to determine if strains circulating in a population are likely to cause invasive pneumococcal disease (IPD). This is possible by comparison of an isolate's genotype to strains known to be invasive. In this work we compared pulse-field gel electrophoresis (PFGE) and multi-locus sequence typing (MLST), comparative genomic hybridization (CGH) and multi-invasive locus sequence typing (MILST) for their ability to distinguish between known IPD causing and carrier strains using phylogenetic analyses. In addition, we assess the ability of these techniques to determine true clones from highly related strains. The resulting trees suggest that despite similar overall topologies, the clearest picture of invasiveness and genetic relatedness can be viewed when typing methods are used collectively.

Keywords

Phylogenetics; Pneumococcus; MLST; PFGE; CGH; MILST; comparison

INTRODUCTION

In the United States, *Streptococcus pneumoniae* (the pneumococcus) is a leading cause of community-acquired pneumonia (CAP) and invasive bacterial disease (Musher 1992). Primarily a commensal, the pneumococcus colonizes the nasopharynx of 10% to 20% of health adults and 40% to 77% of healthy children (Givon-Lavi, Dagan et al. 1999; Harper 1999; Kellner and Ford-Jones 1999; Lopez, Cima et al. 1999). While in most instances colonization is asymptomatic, in the U.S. the pneumococcus is responsible for 100,000–135,000 cases of

*Address correspondence to: C. Obert, Ph.D., Hartwell Center, St. Jude Children's Research Hospital, 332 North Lauderdale St., Memphis, TN 38105-2794, caroline.obert@stjude.org, Tel: 901-495-3217, Fax: 901-495-2945.

Publisher's Disclaimer: This is a PDF file of an unedited manuscript that has been accepted for publication. As a service to our customers we are providing this early version of the manuscript. The manuscript will undergo copyediting, typesetting, and review of the resulting proof before it is published in its final citable form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

CAP, 60,000 cases of bacteremia and sepsis, and 3,300 cases of meningitis resulting in approximately 40,000 deaths per year (2005). At risk for infection are individuals at the extremes of age, individuals with underlying medical conditions, such as cardiovascular disease or sickle cell anemia, and the immunocompromised. Thus, surveillance studies that can predict whether a circulating clone is capable of causing invasive pneumococcal disease (IPD) or carries additional risks, such as antibiotic resistance, are critical.

The propensity of an isolate to cause invasive disease is dependent on its genome content. Initial attempts to distinguish between strains that cause IPD and those that are asymptomatic involved antibiotic susceptibility testing (Kellner, McGeer et al. 1998) and serotyping (Hausdorff, Bryant et al. 2000). More recently molecular epidemiology using pulsed-field gel electrophoresis (PFGE) and multilocus sequence typing (MLST) has demonstrated that within invasive serotypes, invasive and non-invasive clones exist (Robinson, Edwards et al. 2001; Sandgren, Sjostrom et al. 2004). Since pneumococci are capable of undergoing capsular switching (Coffey, Dowson et al. 1991) and the current conjugate vaccine will likely result in an increase in the amount of invasive disease caused by non-vaccine serotypes (Dagan, Givon-Lavi et al. 2002; Veenhoven, Bogaert et al. 2003), a better method of assessing invasive potential would involve determining an isolate's genetic relationship to other pneumococcal strains. To date, numerous groups have attempted to infer such relationships (Kornelisse, Sluijter et al. 1996; Swiatlo, Crain et al. 1996; Marchese, Ramirez et al. 1998; Song, Lee et al. 1999; Yoshida, Hirakata et al. 1999; Gherardi, Whitney et al. 2000; Ko, Reis et al. 2000; Llull, Lopez et al. 2000; Song, Yang et al. 2000; Zhou, Enright et al. 2000; Carlisle, Gratten et al. 2001; Arguedas, Dagan et al. 2003; Skoczynska and Hryniewicz 2003; Clarke, Lawrie et al. 2004; Clarke, Scott et al. 2004; Pletz, McGee et al. 2004; Medina, Greene et al. 2005; Miller, Obert et al. 2005; Richter, Heilmann et al. 2005; Tonnaer, Rijkers et al. 2005). Nevertheless, because pneumococci are recognized as one of the most recombinogenic bacteria (Feil, Smith et al. 2000), it is inherently difficult to infer invasive potential using a single method.

Various whole-genome molecular methods have been routinely used to classify pneumococcal isolates, including multilocus enzyme electrophoresis, restriction fragment length polymorphism analyses, and PFGE (McDougal, Facklam et al. 1992; Muller-Graf, Whatmore et al. 1999). Recently we used data obtained from comparative genomic hybridization (CGH) studies to sort isolates into IPD-causing and carrier cohorts (Obert, Sublett et al. 2006). These analyses identified pneumococcal regions of diversity whose presence was correlated with the ability to cause human disease. Although all of these methods provide a plethora of genomic information about the individual isolates, they are all limited by a lack of specific nucleotide information. Enright and Spratt (Enright and Spratt 1998) addressed this predicament with the establishment of MLST, in which the sequences of 7 housekeeping loci are compared with those in a database of isolate sequences. Work by Zhang *et al.* (Zhang, Jayarao et al. 2004) suggested a subtyping scheme for MLST (known as multi-virulence-locus sequence typing), in which loci correlated with virulence are similarly typed. We describe here a new sub-typing scheme that is a variant of multi-virulence-locus sequence typing, multi-invasive-locus sequence typing (MILST), in which loci whose presence has been correlated with the ability to cause IPD are sequenced.

In an effort to determine how well current typing methods distinguish between strains that may cause IPD from carrier strains, we have applied established molecular typing methods (PFGE and MLST) and the new methods (CGH and MILST) to a population of 20 serotype-14 isolates whose invasiveness had previously been determined (Obert, Sublett et al. 2006), and evaluated the phylogenetic robustness of the data provided by each method and the ability of each method to cluster strains by invasiveness. In addition, we report the utility of the above methods to discriminate between clones which differ in their ability to cause IPD, for epidemiologic surveillance purposes.

MATERIALS AND METHODS

Bacterial strains

Clinical isolates of *S. pneumoniae* were collected from children 2 to 95 months old, at The University of Texas Southwestern Medical Center at Dallas (Collin County, Texas) between February 1999 and January 2003 (Obert, Sublett et al. 2006). Isolates were grown on tryptic soy agar (Difco, Detroit, MI) supplemented with 4% defibrinated sheep blood or in defined semisynthetic casein liquid medium supplemented with 0.5% yeast extract (Lacks and Hotchkiss 1960). Of 20 serotype-14 isolates examined, 12 were obtained from the blood or cerebrospinal fluid of individuals with IPD and 8 were from nasopharyngeal swabs from healthy carriers. To identify genes that are significant to invasiveness and facilitate selection of loci for MLST, we included in our analysis genomic data from 15 invasive isolates of serotype 6A and 14 invasive isolates of serotype 6B that were part of a previous comparative genomic study of invasive and carrier pneumococcal strains (Obert, Sublett et al. 2006). Population genetic analyses of these isolates are provided in the supplemental data (Supplemental Table 1).

Antibiotic susceptibility testing

For each isolate, penicillin susceptibility was determined by disc diffusion using a 1 µg oxacillin disk. Only penicillin susceptibility is defined by CLSI for zones of inhibition (CLSI 2006). Zones smaller than 20mm are defined as non-susceptible. In addition, penicillin E-test gradient strips were placed on Mueller-Hinton agar plates supplemented with 5% defibrinated sheep blood to determine individual minimal inhibitory concentrations. Isolates were scored as susceptible, intermediate, or resistant according to the manufacturer's instruction. Each method was performed in duplicate. Discrepancies in antibiotic susceptibility testing methods (intermediate versus susceptible) were resolved by using the more conservative classification of susceptible since previous studies have indicated that discordant results can occur at breakpoints with E-tests (Macias, Mason et al. 1994).

PFGE

The PFGE procedure was adapted from the procedure described by Hermans et al (Hermans, Sluiter et al. 1995). Chromosomal restriction fragments that were obtained by cleavage with the restriction enzyme *SmaI* were embedded in agarose. The fragments then were separated on a 1% agarose gel, by use of PFGE performed for 24 hours at 14° C in 0.5X TBE, at 6 V/cm, with pulse times of 3 seconds to 30 seconds and at an angle of 120°. After electrophoresis, the gel was stained with 1 mg/mL of ethidium bromide for 30 min.

MLST

Genomic DNA preparation, PCR amplification, and nucleotide sequencing for MLST were performed as previously described (Miller, Obert et al. 2005).

MILST

Loci were selected on the basis of a number of specific criteria. Foremost, since this method is to be used to distinguish between clones that are related to IPD, the presence of the gene needed to be statistically significant to the overall invasive cohort (which was composed of isolates of serotype 6A, 6B, and 14), as determined by a one-tailed Fisher exact test (Obert, Sublett et al. 2006). Second, each statistically significant gene needed to be present in all strains identified as members of the invasive cohort, regardless of serotype. Third, to increase the likelihood of detecting allelic variation, which would distinguish between clones, the selected loci from the genome of the serotype-4 reference strain of *S. pneumoniae*, TIGR4, had to show ≤ 99% homology to those in the genome of strain R6 of *S. pneumoniae* (Hoskins, Albourn et

al. 2001). Other considerations, as described by Wertz *et al.* (Wertz, Valletta-Goldstone et al. 2003), were that genes needed to be >100 kb apart in the TIGR4 genome (Tettelin, Nelson et al. 2001), so that they would not be co-transcribed; that no known paralogs of the individual genes should be present; and that no one physiological process should be oversampled, to prevent artifacts of co-evolution. Primer sequences for the 6 loci selected from the invasive isolates are listed in Supplemental Table 2. PCR products were visualized by electrophoresis on a 1.0% wt/vol agarose gel, and the corresponding PCR products sequenced using forward and reverse primers. Sequences were contigged using Seqassem version 09/2004 (<http://www.gwdg.de/~dhepper/>), aligned with the Clustal W algorithm in Bioedit (Hall 1999) and subsequently queried against the NCBI non-redundant database, using the nucleotide-BLAST [blastn] and protein-BLAST [blastp] programs.

CGH: Microarray-based genome analysis

CGH was performed using *S. pneumoniae* version 2 genomic microarrays obtained from the Pathogen Functional Genomics Resource Center at The Institute for Genomic Research as previously described (Obert, Sublett et al. 2006). Microarrays consisted of amplicons representing 2131 unique open reading frames (ORFs) from strain TIGR4 (Tettelin, Nelson et al. 2001), 164 ORFs from strain R6 (Hoskins, Alborn et al. 2001), and 399 ORFs from strain G54 (Dopazo, Mendoza et al. 2001). Gene presence or absence was determined by comparing hybridization signal strengths between Cy dye-labeled genomic DNA from the clinical isolates and TIGR4. Genes with a normalized log ratio of less than -1.5 were designated genes as absent; ≥ -1.5 were considered present. Each chip was normalized using a “within chip” method, with Spotfire DecisionSite for Functional Genomics software (version 7.2; Spotfire, Somerville, MA), to avoid intensity-specific bias.

Phylogenetic analysis

Phylogenetic relationships among the 20 pneumococcal strains were inferred by using the neighbor-joining algorithm with PAUP* software version 4.0b10 (Swofford 1997). The data obtained from the PFGE and CGH analyses were extrapolated into character presence or absence matrices, whereas the phylogenetic trees constructed from MLST and MILST data used concatenated nucleotide sequences, which were trimmed to the +1 frame. Phylogenetic relationships for MLST and MILST data were additionally inferred using Bayesian analysis, where a model of evolution was selected using algorithms provided by the MrModeltest (Nylander 2004) package. Although the topologies of the Bayesian trees were highly correlated to that found with the neighbor-joining algorithm (data not shown), only distance trees were used for analyses, for more direct comparisons. Statistical support of the branch points for all constructed trees was estimated by performing 5000 bootstrap replicates in PAUP*. The trees were rooted using data from the TIGR4 strain (Tettelin, Nelson et al. 2001). The discrimination index (D.I.) values were calculated as previously described (Hunter and Gaston 1988; Zhang, Jayarao et al. 2004).

RESULTS

Technical considerations

To avoid any potential differences resulting from varying gel analysis software, the gel rendered from PFGE was manually analyzed using Adobe Photoshop software. Only the region which was visually clearest (the center of the gel) was used to determine the presence or absence of bands. Using this software, lines were drawn across the 21 lanes and those isolates with intersecting bands were classified as having a positive band, thus band tolerance settings were not used. Additionally, although prior studies, including (Gamboa, Camou et al. 2002; Gertz, McEllistrem et al. 2003; Serrano, Melo-Cristino et al. 2005), have generally used the Dice similarity coefficient and UPGMA algorithm to analyze PFGE, after initial analysis with these

methods, we opted to use the neighbor-joining algorithm instead. Because the goal of this manuscript was to assess four different typing methods using phylogenetic topologies, we felt that it was essential to employ the same inference method in each case. Since PAUP software allows users to input not only sequence data but also presence/absence data under the neighbor joining algorithm, this algorithm was chosen for use with all four methods. UPGMA was not chosen to be used with all four methods, because while it is easier to calculate, it is generally viewed as a more inferior algorithm when compared to neighbor joining. Finally, allele numbers are missing in strains IPD6, IPD12x, NP7 and NP8 because those genes are absent from those strains. Their absence, however, does not violate the criteria set forth in this manuscript for selecting candidate genes, since these four strains were all previously determined to be part of a carrier cohort instead of an invasive one (Obert, Sublett et al. 2006). We assigned a VST (virulence sequence type) to these strains, despite the gene absence because we felt that the strain inclusion may yield important information about the spread of pneumococcal disease at a later time. The importance of examining pneumococcal strains with variations in allelic pattern (i.e. truncation) was previously discussed by Diggle and Clarke (Diggle and Clarke 2005).

Comparison of typing and sub-typing methodologies

Twenty pneumococcal isolates were subjected to PFGE, MLST, MILST, and CGH to assess the ability to distinguish invasiveness among strains and the phylogenetic robustness of each method. Although a larger sample size would provide an unquestionable amount of additional detail, the current size was used to facilitate direct and clear comparisons of clones between methods. Because nucleotide sequence information is not provided by PFGE or CGH, the restriction digest pattern (Fig. 1) and the hybridization ratios (Obert, Sublett et al. 2006) obtained by these methods for the 20 strains were extrapolated into presence or absence matrices. The phylogenetic relationships among strains were inferred from these data and from the concatenated nucleotide sequence data from MLST and MILST using the neighbor-joining algorithm. Because pneumococci are highly recombinogenic, deep evolutionary relationships can not be inferred by trees. Therefore, for these analyses only more recent relationships (outer clades) are examined and discussed.

The overall topologies of the resulting phylogenetic trees (Fig. 2a–d) indicate the existence of 2 main cohorts. Cohort 1 (bolded in Fig. 2) is composed of 9 clinical strains known to have caused IPD (IPD1–5, IPD7, and IPD9–11) and 3 strains from asymptomatic carriers (NP1, NP5, and NP6). Strains in this large paraphyletic clade were previously described as being part of an overall invasive cohort (Obert, Sublett et al. 2006). Cohort 2 (not bolded in Fig. 2) is composed of 3 IPD-causing strains isolated from patients diagnosed with meningitis (IPD6, IPD8, and IPD12X) and 5 carrier strains (NP2–4, NP7, and NP8) and was previously described as part of an overall carrier cohort. One notable difference in the topologies is the placement of IPD4, which clusters with cohort 2 when PFGE, MLST and MILST methods are applied (Figure 2a–c) but is grouped with cohort1 when CGH is utilized (Figure 2d).

Although all 4 methods, in general, cluster the isolates in the two cohort manner, the main difference between the results obtained by each method is the level of discrimination between the strains. Using PFGE, the pneumococcal isolates were placed into 11 distinct groups (D.I. = 0.93), whereas these same isolates were resolved into 6 groups (D.I. = 0.70) by MLST, 9 groups (D.I. = 0.77) by MILST, and 16 groups (D.I. = 0.97) by CGH (Fig. 1a–d). The D.I. values indicate that CGH is the most discriminating method for distinguishing between clones, followed by PFGE, MILST, and MLST. However, it should be noted that MILST and MLST provide more robust trees than does PFGE in that deeper evolutionary relationships are resolved, though again not as much as those seen with CGH and with the caveat that deep ancestral relationships are not necessarily accurate due to recombination in the pneumococcus.

DISCUSSION

The utility of certain molecular typing methods is well established (Tenover, Arbeit et al. 1997; Maiden, Bygraves et al. 1998; Savor, Pfaller et al. 1998; Foxman and Riley 2001; Cooper and Feil 2004). However, comparisons between new techniques (CGH and MLST) and existing methods (PFGE and MLST) have not previously been evaluated. PFGE has the advantage of being an inexpensive and highly reproducible technique for which all strains can be typed regardless of prior knowledge of the genome. It is most effective in examining short-term or local epidemiological studies (Maiden, Bygraves et al. 1998). However, since comparing PFGE results obtained by different laboratories is difficult and obtaining specific information about genes is not possible, PFGE is less suitable for long-term studies. By comparison, MLST (although a more expensive technique than PFGE) provides unambiguous results that are electronically portable and therefore suitable for global surveillance studies via the internet (Maiden, Bygraves et al. 1998; Cooper and Feil 2004). A major disadvantage of MLST, however, is that phylogenetic relationships and resolution of clones can be masked by the use of slowly evolving housekeeping genes. Zhang *et al.* therefore suggested, in their study of *Listeria monocytogenes*, that more rapidly evolving virulence loci should be used to resolve deeper relationships and provide a higher level of strain discrimination (Zhang, Jayarao et al. 2004). We tested that theory in the present study by assessing 6 loci in *S. pneumoniae* that are associated with invasive disease. Although we found greater discrimination between clones by using this technique and it cost no more than MLST, a major disadvantage of MLST is the lack of a worldwide database that would allow comparison of isolates. This drawback could be easily resolved, however, by allowing the inclusion of invasive loci data in the MLST database as a sub-typing method. Thus MLST would not replace MLST but rather expand the genes examined and enhance our current understanding of genetic relationships among strains that would seem to be identical but may have unique individual attributes, such as resistance patterns or virulence expression. The combination of MLST and MLST data for the strains used in this study increased the robustness of the phylogenetic tree (D.I.= 0.78, data not shown). Although the addition of MLST data to MLST did yield a higher D.I. value, it is not a perfect solution to discrimination amongst clones. The MLST topology indicated that strains classified as ST13 or ST656 by MLST were clones, even though PFGE and CGH data indicated otherwise. This contradiction is clearly the result of limited variability in the genes selected for MLST analyses. However, this does not mean that we should not consider the inclusion of additional genes as a sub-typing scheme for MLST. Although genes associated with invasiveness were selected to establish this variant of MLST, the addition of any new gene would add to the discrimination between strains given that the gene showed marked variability. Since we were looking to distinguish between isolates that may cause IPD from carriers, these genes were specifically chosen. It is highly probable that either the inclusion of more genes or the selection of new genes would resolve this particular conundrum.

Finally, it is evident from our study that CGH provides the most detailed phylogeny of the methods examined, in that deeper relationships are resolved and greater clade discrimination is apparent. However, this method has several limitations. Foremost, researchers are restricted to available microarray chips, which at present are very costly. Second, as with PFGE, sharing data between different laboratories will be restricted to the specific microarray chip used. For example, although arrays available from Affymetrix GeneChip® can clearly provide information about the presence or absence of specific genes, chips currently available from TIGR allow assessment of genomic content only by comparison with a known genome. Therefore, only information provided by that chip can be assessed. Third, although CGH clearly provides a plethora of additional information about specific clones that is not available from PFGE and MLST, at this current time it is not a feasible typing method for use on thousands of strains due to cost and time limitations. As such it currently, would need to be used as a sub-typing method to supply additional data on clonal groups until such time that improvements

to the technology alleviates those constraints in a similar manner to what has occurred with nucleotide sequencing.

Epidemiological Relationship of Pneumococcal Isolates

As mentioned in Silva *et al* (Silva, McCluskey et al. 2006), if CGH data, like MLST data, were to be combined with MLST data, and used as sub-typing schemes to differentiate between clones, an unquestionably more complete picture of genetic structure among isolates could be obtained that would be useful in long-term epidemiological studies. In the case of the 20 serotype-14 isolates used in the current study, our combined data indicate that most of the isolates are ST13 (Table 1), a single-locus variant (SLV) of an England 14-9 clone known to be susceptible to penicillin. However, 9 of these 11 ST13 isolates were resistant to penicillin (Table 1). Although this would seem to indicate that there has merely been an emergence of clones with a different susceptibility pattern (which might be attributed to differing antibiotic pressures in varying geographic regions), further examination of the topologies indicate that other factors may be present. If the 11 isolates resulted from the spread of a single strain, the topology would reveal one large monophyletic clade. However, the PFGE and CGH derived topologies show that the isolates have, in fact, undergone further differentiation most likely resulting from the transfer and/or recombination of additional genomic regions (Fig. 2a and d). These differences, which could also potentially result in a more variable antibiogram than just the change in penicillin susceptibility from the England 14-9 clone, would not have been observed were it not for the use of PFGE and/or CGH as a sub-typing scheme. In addition should such differences be found, CGH (and not PFGE) can also provide researchers with specific regions/genes to examine.

Three isolates were classified as ST124, a major serotype-14 clone predominantly associated with bacteremia. MLST-derived phylogeny suggests that these three clones, obtained from an individual with meningitis (IPD4) and 2 healthy asymptomatic carriers (NP2 and NP4) although related, they are distinct enough from each other that NP2 and NP4 are placed into a paraphyletic clade, with IPD4 related to them in the larger polyphyletic clade. The strains are therefore not clones which would reside within the monophyletic clade, as suggested by MLST analysis. Further examination of the CGH-derived phylogeny indicates that the isolates had major genomic differences resulting in their placement within completely different clades. In particular 2 regions of DNA, Region of Diversity (RD) 5 (~5kb), and the pathogenicity island RD8a (~20kb) (Obert, Sublett et al. 2006), were present in IPD4 and absent in NP2 and NP4 (Table 2). Previously, we have shown that the presence of the pathogenicity island RD8a is associated with the ability to cause human IPD (Obert, Sublett et al. 2006). Thus, CGH was able to discriminate between an invasive isolate and 2 presumably non-invasive isolates by the presence/absence of these RDs whereas MLST and PFGE were not. Moreover, this is not the first report of isolates within the same group of clones (as defined by MLST and PFGE) differentiated by the presence of an RD and distinct virulence properties. Silva et al. described 3 pneumococcal isolates of the same serotype and MLST differentiated by the presence of RD8a and RD23 (Silva, McCluskey et al. 2006). These strains had differences in virulence when C57/BL6 mice were challenged intraperitoneally; a finding that emphasizes the necessity to discriminate clones beyond the details provided by PFGE and MLST and which underscores the role of RDs in pneumococcal virulence.

Smaller differences in the presence/absence of various individual genes (not RDs) also resulted in discrepancies between the typing methods: An ST656 clone (IPD8) was isolated from a patient with meningitis, whereas the other (NP3) was from an asymptomatic carrier. CGH and PFGE analyses both separated these isolates into paraphyletic clades, whereas MLST and MILST clustered them into a monophyletic clade. ST782, a SLV of the Sweden 15A-25 clone generally associated with bacteremia, was isolated from a patient with meningitis (IPD6) and

from 2 asymptomatic carriers (NP7 and NP8). CGH and MILST allowed assessment of distinguishing features that classified the strains into a paraphyletic clade, whereas MLST and PFGE grouped these isolates into a single monophyletic clade.

Currently, a plethora of typing methods have been established and applied to numerous bacterial strains in an attempt to distinguish one strain over another. While each method has its own unique advantages, numerous caveats are equally found. PFGE and MLST have become the standard methods by which *S. pneumoniae* strains are identified. However, its high level of recombination necessitates improved and/or additional typing methods to determine, for epidemiological purposes, if established clones have acquired new genes that modify virulence. The results of this study suggest that PFGE and MLST, although relatively inexpensive and widely used, are not sufficiently sensitive to allow identification of discrete changes in genomic content, such as the acquisition of pathogenicity islands, and thus may classify strains with dramatically different genomic contents as clones. In contrast, lesser used and more expensive typing methods CGH and MILST improves sensitivity by distinguishing between such isolates with such changes. In conclusion, because of the merits and weaknesses of any given molecular typing system, the only way to get a true picture of population structure is to combine analyses. Given the level of discrimination between typing schemes we suggest that MILST and in particular CGH be incorporated as sub-typing schemes to enhance current PFGE and MLST data. Such information would not only be applicable to clones, but also to more diverse populations, since combining typing methods diminishes the effect of high recombination rates in bacterial species such as *S. pneumoniae* and *Neisseria meningitidis*.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank Dr. Barry G. Hall and Dr. William Hanage for their insights on the phylogenetic analyses and this manuscript. We also thank Drs. Robert Fleischmann and Scott Peterson at the Pathogen Functional Genomics Resource Center at The Institute for Genomic Research for providing the pneumococcal microarrays necessary for this work. This work was supported by the NIH R01 award AI27913 and the American Lebanese Syrian Associated Charities (ALSAC).

References

- Streptococcus pneumoniae Disease. 2005. Retrieved July 13, 2006, 2006, from http://www.cdc.gov/ncidod/dbmd/diseaseinfo/streppneum_t.htm
- Arguedas A, Dagan R, et al. Microbiology of otitis media in Costa Rican children, 1999 through 2001. *Pediatr Infect Dis J* 2003;22(12):1063–1068. [PubMed: 14688566]
- Carlisle JB, Gratten M, et al. Molecular epidemiology of multiple drug resistant type 6B Streptococcus pneumoniae in the Northern Territory and Queensland, Australia. *Epidemiol Infect* 2001;126(1):25–29. [PubMed: 11293678]
- Clarke SC, Lawrie D, et al. Genetic relatedness of antibiotic-resistant pneumococci isolated during case clusters. *J Med Microbiol* 2004;53(Pt 11):1097–9. [PubMed: 15496386]
- Clarke SC, Scott KJ, et al. Erythromycin resistance in invasive serotype 14 pneumococci is highly related to clonal type. *J Med Microbiol* 2004;53(Pt 11):1101–1103. [PubMed: 15496387]
- CLSI. Approved Standard M100-S16. Wayne, PA: Clinical Laboratory Standards Institute; 2006. Performance standards for Antimicrobial Susceptibility Testing.
- Coffey TJ, Dowson CG, et al. Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of Streptococcus pneumoniae. *Mol Microbiol* 1991;5(9):2255–60. [PubMed: 1766389]
- Cooper JE, Feil EJ. Multilocus sequence typing--what is resolved? *Trends Microbiol* 2004;12(8):373–377. [PubMed: 15276613]

- Dagan R, Givon-Lavi N, et al. Reduction of nasopharyngeal carriage of *Streptococcus pneumoniae* after administration of a 9-valent pneumococcal conjugate vaccine to toddlers attending day care centers. *J Infect Dis* 2002;185(7):927–36. [PubMed: 11920317]
- Diggle MA, Clarke SC. Truncated xpt gene present in invasive *Streptococcus pneumoniae* may have implications for MLST schemes. *J Med Microbiol* 2005;54(10):909–912. [PubMed: 16157542]
- Dopazo J, Mendoza A, et al. Annotated draft genomic sequence from a *Streptococcus pneumoniae* type 19F clinical isolate. *Microb Drug Resist* 2001;7(2):99–125. [PubMed: 11442348]
- Enright MC, Spratt BG. A multilocus sequence typing scheme for *Streptococcus pneumoniae*: identification of clones associated with serious invasive disease. *Microbiology* 1998;144(Pt 11):3049–3060. [PubMed: 9846740]
- Feil EJ, Smith JM, et al. Estimating recombinational parameters in *Streptococcus pneumoniae* from multilocus sequence typing data. *Genetics* 2000;154(4):1439–1450. [PubMed: 10747043]
- Foxman B, Riley L. Molecular epidemiology: focus on infection. *Am J Epidemiol* 2001;153(12):1135–41. [PubMed: 11415945]
- Gamboa L, Camou T, et al. Dissemination of *Streptococcus pneumoniae* Clone Colombia5–19 in Latin America. *J Clin Microbiol* 2002;40(11):3942–3950. [PubMed: 12409356]
- Gertz RE Jr, McEllistrem MC, et al. Clonal Distribution of Invasive Pneumococcal Isolates from Children and Selected Adults in the United States Prior to 7-Valent Conjugate Vaccine Introduction. *J Clin Microbiol* 2003;41(9):4194–4216. [PubMed: 12958247]
- Gherardi G, Whitney CG, et al. Major related sets of antibiotic-resistant Pneumococci in the United States as determined by pulsed-field gel electrophoresis and pbp1a-pbp2b-pbp2x-dhf restriction profiles. *J Infect Dis* 2000;181(1):216–29. [PubMed: 10608770]
- Givon-Lavi N, Dagan R, et al. Marked differences in pneumococcal carriage and resistance patterns between day care centers located within a small area. *Clin Infect Dis* 1999;29(5):1274–80. [PubMed: 10524975]
- Hall T. BioEdit: a user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT “. *Nucleic Acids Symposium Series* 1999;41:95–98.
- Harper MB. Nasopharyngeal colonization with pathogens causing otitis media: how does this information help us? *Pediatr Infect Dis J* 1999;18(12):1120–4. [PubMed: 10608643]
- Hausdorff WP, Bryant J, et al. Which pneumococcal serogroups cause the most invasive disease: implications for conjugate vaccine formulation and use, part I. *Clin Infect Dis* 2000;30(1):100–121. [PubMed: 10619740]
- Hermans PW, Sluijter M, et al. Comparative study of five different DNA fingerprint techniques for molecular typing of *Streptococcus pneumoniae* strains. *J Clin Microbiol* 1995;33(6):1606–12. [PubMed: 7650196]
- Hoskins J, Alborn WE Jr, et al. Genome of the bacterium *Streptococcus pneumoniae* strain R6. *J Bacteriol* 2001;183(19):5709–5717. [PubMed: 11544234]
- Hunter PR, Gaston MA. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* 1988;26(11):2465–2466. [PubMed: 3069867]
- Kellner JD, Ford-Jones EL. *Streptococcus pneumoniae* carriage in children attending 59 Canadian child care centers. Toronto Child Care Centre Study Group. *Arch Pediatr Adolesc Med* 1999;153(5):495–502. [PubMed: 10323630]
- Kellner JD, McGeer A, et al. The use of *Streptococcus pneumoniae* nasopharyngeal isolates from healthy children to predict features of invasive disease. *Pediatr Infect Dis J* 1998;17(4):279–286. [PubMed: 9576381]
- Ko AI, Reis JN, et al. Clonally related penicillin-nonsusceptible *Streptococcus pneumoniae* serotype 14 from cases of meningitis in Salvador, Brazil. *Clin Infect Dis* 2000;30(1):78–86. [PubMed: 10619737]
- Kornelis RF, Sluijter M, et al. Molecular epidemiological characteristics of pneumococcal meningitis in children. *Eur J Clin Microbiol Infect Dis* 1996;15(8):635–638. [PubMed: 8894570]
- Lacks S, Hotchkiss RD. A study of the genetic material determining an enzyme in *Pneumococcus*. *Biochim Biophys Acta* 1960;39:508–518. [PubMed: 14413322]
- Llull D, Lopez R, et al. Clonal origin of the type 37 *Streptococcus pneumoniae*. *Microb Drug Resist* 2000;6(4):269–275. [PubMed: 11272254]

- Lopez B, Cima MD, et al. Epidemiological study of *Streptococcus pneumoniae* carriers in healthy primary-school children. *Eur J Clin Microbiol Infect Dis* 1999;18(11):771–6. [PubMed: 10614950]
- Macias EA, Mason EO Jr, et al. Comparison of E test with standard broth microdilution for determining antibiotic susceptibilities of penicillin-resistant strains of *Streptococcus pneumoniae*. *J Clin Microbiol* 1994;32(2):430–2. [PubMed: 8150953]
- Maiden MC, Bygraves JA, et al. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. *Proc Natl Acad Sci U S A* 1998;95(6):3140–3145. [PubMed: 9501229]
- Marchese A, Ramirez M, et al. Molecular epidemiology of penicillin-resistant *Streptococcus pneumoniae* isolates recovered in Italy from 1993 to 1996. *J Clin Microbiol* 1998;36(10):2944–2949. [PubMed: 9738048]
- McDougal LK, Facklam R, et al. Analysis of multiply antimicrobial-resistant isolates of *Streptococcus pneumoniae* from the United States. *Antimicrob Agents Chemother* 1992;36(10):2176–84. [PubMed: 1444297]
- Medina MJ, Greene CM, et al. Novel antibiotic-resistant pneumococcal strains recovered from the upper respiratory tracts of HIV-infected adults and their children in Kisumu, Kenya. *Microb Drug Resist* 2005;11(1):9–17. [PubMed: 15770088]
- Miller ML, Obert CA, et al. Cephalosporin-resistant pneumococci and sickle cell disease. *Emerg Infect Dis* 2005;11(8):1192–1196. [PubMed: 16102306]
- Muller-Graf CD, Whatmore AM, et al. Population biology of *Streptococcus pneumoniae* isolated from oropharyngeal carriage and invasive disease. *Microbiology* 1999;145(Pt 11):3283–93. [PubMed: 10589738]
- Musher DM. Infections caused by *Streptococcus pneumoniae*: clinical spectrum, pathogenesis, immunity, and treatment. *Clin Infect Dis* 1992;14(4):801–7. [PubMed: 1576274]
- Nylander, J. MrModeltest: Program distributed by the author. Evolutionary Biology Centre; Uppsala University: 2004.
- Obert C, Sublett J, et al. Identification of a candidate *Streptococcus pneumoniae* core genome and regions of diversity correlated with invasive pneumococcal disease. *Infection and Immunity* 2006;74(8):4766–4777. [PubMed: 16861665]
- Pletz MW, McGee L, et al. Levofloxacin-resistant invasive *Streptococcus pneumoniae* in the United States: evidence for clonal spread and the impact of conjugate pneumococcal vaccine. *Antimicrob Agents Chemother* 2004;48(9):3491–3497. [PubMed: 15328116]
- Richter SS, Heilmann KP, et al. The molecular epidemiology of *Streptococcus pneumoniae* with quinolone resistance mutations. *Clin Infect Dis* 2005;40(2):225–35. [PubMed: 15655739]
- Robinson DA, Edwards KM, et al. Clones of *Streptococcus pneumoniae* isolated from nasopharyngeal carriage and invasive disease in young children in central Tennessee. *J Infect Dis* 2001;183(10):1501–7. [PubMed: 11319686]
- Sandgren A, Sjostrom K, et al. Effect of clonal and serotype-specific properties on the invasive capacity of *Streptococcus pneumoniae*. *J Infect Dis* 2004;189(5):785–96. [PubMed: 14976594]
- Savor C, Pfaller MA, et al. Comparison of genomic methods for differentiating strains of *Enterococcus faecium*: assessment using clinical epidemiologic data. *J Clin Microbiol* 1998;36(11):3327–3331. [PubMed: 9774587]
- Serrano I, Melo-Cristino J, et al. Characterization of the Genetic Lineages Responsible for Pneumococcal Invasive Disease in Portugal. *J Clin Microbiol* 2005;43(4):1706–1715. [PubMed: 15814989]
- Silva NA, McCluskey J, et al. Genomic diversity between strains of the same serotype and multilocus sequence type among pneumococcal clinical isolates. *Infect Immun* 2006;74(6):3513–8. [PubMed: 16714583]
- Skoczynska A, Hryniewicz W. Genetic relatedness, antibiotic susceptibility, and serotype distribution of *Streptococcus pneumoniae* responsible for meningitis in Poland, 1997–2001. *Microb Drug Resist* 2003;9(2):175–182. [PubMed: 12820803]
- Song JH, Lee NY, et al. Spread of drug-resistant *Streptococcus pneumoniae* in Asian countries: Asian Network for Surveillance of Resistant Pathogens (ANSORP) Study. *Clin Infect Dis* 1999;28(6):1206–1211. [PubMed: 10451154]

- Song JH, Yang JW, et al. Molecular characterization of multidrug-resistant *Streptococcus pneumoniae* isolates in Korea. The Asian Network for Surveillance of Resistant Pathogens (ANSORP) Study Group. *J Clin Microbiol* 2000;38(4):1641–1644. [PubMed: 10747158]
- Swiatlo E, Crain MJ, et al. DNA polymorphisms and variant penicillin-binding proteins as evidence that relatively penicillin-resistant pneumococci in western Canada are clonally related. *J Infect Dis* 1996;174(4):884–888. [PubMed: 8843237]
- Swofford, D. PAUP*: Phylogenetic Analysis Using Parsimony (* and other methods). Sunderland, MA: Sinauer Associates; 1997.
- Tenover FC, Arbeit RD, et al. How to select and interpret molecular strain typing methods for epidemiological studies of bacterial infections: a review for healthcare epidemiologists. Molecular Typing Working Group of the Society for Healthcare Epidemiology of America. *Infect Control Hosp Epidemiol* 1997;18(6):426–439. [PubMed: 9181401]
- Tettelin H, Nelson KE, et al. Complete genome sequence of a virulent isolate of *Streptococcus pneumoniae*. *Science* 2001;293(5529):498–506. [PubMed: 11463916]
- Tonnaer EL, Rijkers GT, et al. Genetic relatedness between pneumococcal populations originating from the nasopharynx, adenoid, and tympanic cavity of children with otitis media. *J Clin Microbiol* 2005;43(7):3140–3144. [PubMed: 16000426]
- Veenhoven R, Bogaert D, et al. Effect of conjugate pneumococcal vaccine followed by polysaccharide pneumococcal vaccine on recurrent acute otitis media: a randomised study. *Lancet* 2003;361(9376):2189–95. [PubMed: 12842372]
- Wertz JE, Valletta-Goldstone CM, et al. A molecular phylogeny of enteric bacteria and implications for a bacterial species concept. *J Evol Biol* 2003;16(6):1236–1248. [PubMed: 14640415]
- Yoshida R, Hirakata Y, et al. Genetic analysis of serotype 23F *Streptococcus pneumoniae* isolates from several countries by penicillin-binding protein gene fingerprinting and pulsed-field gel electrophoresis. *Chemotherapy* 1999;45(3):158–165. [PubMed: 10224337]
- Zhang W, Jayarao BM, et al. Multi-virulence-locus sequence typing of *Listeria monocytogenes*. *Appl Environ Microbiol* 2004;70(2):913–920. [PubMed: 14766571]
- Zhou J, Enright MC, et al. Identification of the major Spanish clones of penicillin-resistant pneumococci via the Internet using multilocus sequence typing. *J Clin Microbiol* 2000;38(3):977–986. [PubMed: 10698983]

FIG. 1.
PFGE of restriction digests of nasopharyngeal strains and invasive strains. Isolates were digested with *Sma*I restriction enzyme.

FIG. 2. Phylogenetic relatedness of 20 serotype-14 strains as defined by (a) PFGE, (b) MLST, (c) MILST, and (d) CGH. Branch lengths are provided beneath each tree. Bolded strains were previously classified as members of an invasive cohort (Obert, Sublett et al. 2006). Bootstrap values >80% are presented.

Table 1

Genotypes of the clinical Serotype-14 isolates

Isolate	Diagnosis	Site	Date isolated	MILST allele numbers					VST ^D		MLST allele numbers					ST ^E		PEN ^F		CLD ^J	
				pepS	dinP	valS	tdk	str	rok	aroE	gdh	gki	xpt	recP	spi	ddl	CLE ^G	MIC ^H			
	Bacteremia	Blood	Oct '99	1	1	1	1	1	1	1	1	1	5	4	5	5	27	8	R ^J	3	NS ^M
	Bacteremia	Blood	Oct '99	1	1	1	1	1	1	1	1	1	5	4	5	5	27	8	R ^J	4	NS
	Meningitis	CSF ^B	Jun '99	1	2	1	1	1	1	2	1	5	4	5	5	5	27	8	S ^K	0.023	S
	Meningitis	CSF	Oct '00	4	2	1	1	3	5	3	7	5	1	8	14	11	14	124	S	0.016	S
	Meningitis	CSF	Feb '00	1	1	1	1	1	1	1	1	5	4	5	5	5	27	8	R ^J	3	NS
	Meningitis	CSF	Jan '00	5	3	5	3	N/	4	4	2	5	36	12	17	21	5	782	I	0.125	S
	Bacteremia	Blood	Apr '01	1	1	1	1	1	1	1	1	5	4	5	5	5	27	8	S	0.023	S
	Meningitis	CSF	May '01	4	2	2	1	3	6	5	41	5	1	8	14	11	14	656	S	0.016	S
	Meningitis	CSF	Dec '01	1	1	1	1	1	1	1	1	5	4	5	5	5	27	8	R	3	NS
	Meningitis	CSF	Mar '02	1	1	1	1	1	1	1	1	5	4	5	5	5	27	8	R	3	NS
	Bacteremia	Blood	Jan '03	1	1	1	1	1	1	1	1	5	4	5	5	5	27	8	R	4	NS
	Meningitis	CSF	Feb '99	3	N/A	4	1	2	3	6	7	11	10	1	6	8	1	156	R	2	NS
	Healthy	NP	N/A ^C	1	1	1	1	1	1	1	1	5	4	5	5	5	27	8	R	3	NS
	Healthy	NP	Oct '00	4	2	2	1	3	5	7	7	5	1	8	14	11	14	124	S	0.016	S
	Healthy	NP	Dec '00	4	2	2	1	3	6	5	41	5	1	8	14	11	14	656	S	0.032	S
	Healthy	NP	Dec '00	4	2	2	1	3	6	5	7	5	1	8	14	11	14	124	S	0.023	S
	Healthy	NP	Jan '02	1	1	1	1	1	1	1	1	5	4	5	5	5	27	8	R	2	NS
	Healthy	NP	Jan '02	1	1	1	1	1	1	1	1	5	4	5	5	5	27	8	R	3	NS
	Healthy	NP	Feb '02	5	3	5	1	N/	4	8	2	5	36	12	17	21	5	782	I ^L	0.19	S
	Healthy	NP	Jun '02	5	3	5	1	N/	4	8	2	5	36	12	17	21	5	782	I	0.125	S

Infect Genet Evol. Author manuscript; available

Infect Genet Evol. Author manuscript; available in PMC 2007 December 10.

nasopharynx;

cerebrospinal fluid;

not available;

, virulence sequence type;

sequence type;

penicillin;

classification by E-test;

minimum inhibitory concentration;

classification by disc diffusion;

resistant;

susceptible;

L_I , intermediate susceptibility;

M_{NS} , non-susceptible.

Table 2
RDs present in clone ST-124 isolates as determined by CGH

Genes (TIGR Annotation)	IPD4	NP2	NP4
<i>Region of Diversity 5</i>			
SP0691	+	—	—
SP0692	+	—	—
SP0694	+	—	—
SP0695	+	—	—
SP0696	+	—	—
SP0697	+	—	—
SP0698	+	—	—
SP0700	+	—	—
<i>Region of Diversity 8a</i>			
SP1315	+	—	—
SP1316	+	—	—
SP1317	+	—	—
SP1318	+	—	+
SP1319	—	—	—
SP1320	+	—	—
SP1321	+	—	—
SP1322	+	—	—
SP1323	+	+	—
SP1324	+	—	—
SP1325	+	—	+
SP1326	+	—	—
SP1327	+	—	—
SP1328	+	—	—
SP1329	+	—	—
SP1330	+	—	—
SP1331	+	—	—
SP1332	—	—	—