**Streptococcus pneumoniae** Serotype 6C: an Intra- and Interclonal Complex Comparison

J. C. Thomas,1,2 Y. Kong,3 V. Sabharwal,4 S. I. Pelton,4 and M. M. Pettigrew1*

Yale School of Public Health, Yale University School of Medicine, New Haven, Connecticut1; Department of Microbiology, University of Mississippi Medical Center, Jackson, Mississippi2; Department of Molecular Biophysics and Biochemistry and W. M. Keck Foundation Biotechnology Resource Laboratory, New Haven, Connecticut3; and Boston University School of Medicine and Public Health, Boston Medical Center, Boston, Massachusetts4

Received 13 April 2011/Accepted 25 April 2011

We report the annotated draft genome sequences of four serotype 6C *Streptococcus pneumoniae* isolates of differing genetic backgrounds. Serotype 6C isolates are increasing in prevalence and becoming progressively more resistant to antibiotics. As a result, these strains are likely to become more important in the near future.

**Streptococcus pneumoniae** is a leading cause of meningitis, pneumonia, and otitis media globally, in addition to asymptptomatically colonizing the upper respiratory tract of nearly half of all healthy children (2, 11, 16). Pneumococcal isolates express one of more than 90 distinct capsular polysaccharides (13). The introduction of a heptavalent pneumococcal conjugate vaccine (PCV7) has led to a decrease in invasive disease by isolates that express one of the targeted vaccine serotypes (14). However, during this period, nonvaccine serotypes have increased in prevalence (3, 9, 10), and novel serotypes such as 6C have recently been identified (13). The prevalence of serotype 6C isolates has increased in the United States over recent years (4, 12). Moreover, the proportion of serotype 6C isolates that are nonsusceptible to penicillin has also increased (8). While Gertz et al. (8) have determined several multilocus sequence types (STs) that belong to serotype 6C, this study presents the first genome sequences of serotype 6C isolates.

Four serotype 6C pneumococcal strains were isolated in Massachusetts between 2006 and 2007. Three of the isolates were obtained from nasopharyngeal swabs, while the fourth was acquired from a blood sample from a patient with meningitis. Multilocus sequence typing (7) identified the isolates as belonging to sequence types 1292 (isolate 07AR0125, from the meningitis patient), 1390 (isolate PT8114, and 1692 (isolates BR1064 and ND6012). Approximately 28% of serotype 6C isolates in the United States belong to ST1390 (4). The capsule type was determined by a Quellung reaction. DNA was extracted using Qiagen 100/G genomic tips and sequenced using the Solexa paired-end sequencing platform (Illumina, San Diego, CA). Fifty-base-pair reads were generated, resulting in an average coverage of >30-fold. bTrim (http://graphics.med.yale.edu/trim/readme) was used to trim paired-end sequences, based on their quality scores. Velvet was used to assemble sequences that passed trimming (17). Coding sequences were predicted using Glimmer3 (5). Genome sequences were automatically annotated using RAST (1), and RNA sequences were detected using tRNAscan-SE (15).

The draft genome sequences vary in size between 2.03 and 2.05 Mb, with a mean G+C content of 39.6%. The four genomes contain between 2,015 and 2,047 predicted protein-coding genes, a finding that is broadly consistent with those for previously sequenced pneumococcal isolates (6). The intraclonal complex isolates differed by 5 genes, while interclonal complex pairwise comparisons revealed a mean of 136 gene differences. The genomes contained between 34 and 39 tRNA genes, comprising all amino acids. Further analysis of these genomes is currently being performed to determine genetic factors unique to these serotype 6C isolates.

**Nucleotide sequence accession numbers.** Whole-genome shotgun projects for each of the four isolates have been deposited at DDBJ/EMBL/GenBank under accession numbers AFBY00000000, AFBZ00000000, AFCB00000000, and AFBZ00000000. The versions described in this paper are the first versions under the designations AFBY01000000, AFBZ01000000, AFCB01000000, and AFBZ01000000, respectively.

We thank the W. M. Keck Facility for their work in the sequencing of the pneumococcal isolates.

This work was supported by the National Institute of Allergy and Infectious Diseases (RO1 AI068043 to M.M.P.).

The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**REFERENCES**


7. Enright, M. C., and B. G. Spratt. 1998. A multilocus sequence typing scheme...


