Genetic Basis for the New Pneumococcal Serotype, 6C[∇]

In Ho Park,¹ Saeyoung Park,^{1,2} Susan K. Hollingshead,² and Moon H. Nahm^{1,2*}

*Departments of Pathology¹ and Microbiology,² University of Alabama at Birmingham,
845 19th Street South, BBRB 614, Birmingham, Alabama 35294*

Received 9 April 2007/Returned for modification 13 May 2007/Accepted 10 June 2007

We have recently reported a new pneumococcal serotype (6C), which is closely related to serotype 6A (I. H. Park et al., J. Clin. Microbiol. 45:1225–1233, 2007). To investigate the genetic basis for serotype 6C, we studied the capsule gene loci of 14 6C isolates from three different continents, including one isolated in Alabama 27 years ago. The *wciN* region of all 6C isolates has a 1,029-bp-long sequence that replaces the 1,222-bp-long sequence of the 6A *wciN* region. This recombination event has created a new 1,125-bp-long open reading frame which encodes a product that is also homologous to glycosyl transferases. Flanking this introduced gene is 300 bp upstream and 100 bp downstream with only about 90% homology with 6A and which is identical in all 6C isolates. Transfer of the *wciN* region converts 6A to 6C. Determination of the DNA sequence of the entire capsule gene locus of one 6C isolate showed that the 6C capsule gene locus is almost identical (>98% homologous) to that of 6A except for the *wciN* region. These findings indicate that the 6C capsule type originated more than 27 years ago by a single recombination event in a 6A locus in which 6A *wciN* was replaced by a gene of unknown origin.

Streptococcus pneumoniae is a major human pathogen that can cause various diseases, including pneumonia, otitis media, meningitis, and sepsis (9). Pneumococci produce at least 91 different types (serotypes) of the polysaccharide (PS) capsule (11, 19), which shields pneumococci from host phagocytes and is required for virulence (1). Antibodies to the capsular PS can abrogate the shielding effect of the capsule and thereby provide a host with capsule-type-specific protection (6). Consequently, currently available pneumococcal vaccines are designed to induce the production of antibodies to capsular PS.

Among the 91 different serotypes, certain serotypes are clinically important and have been extensively studied. For instance, 6A and 6B serotypes accounted for 4.6% and 10.6% of pneumococcal meningitis or sepsis in young children before the introduction of a pneumococcal conjugate vaccine (24). Extensive biochemical studies since their discovery in 1929 (7) have shown the 6A/6B PS to be isopolymers containing galactose-glucose-rhamnose-ribitol-phosphate, differing only in the rhamnose-ribitol linkage (12, 22). Genetic studies of the capsule loci from many isolates showed that the capsule loci of 6A and 6B are largely identical and that only one nucleotide in *wciP* (G584A; S195N) is consistently different between the capsule loci of the two serotypes (16). This suggests that the mutation at position 584 is responsible for the difference between 6A and 6B serotypes.

Notwithstanding such intensive studies, we have recently identified two subtypes in pneumococci that were serotyped as “6A” by classical quellung reactions (14, 19). Chemical studies of the capsular PS showed that the major subtype produces capsular PS matching the 6A PS structure described in the literature (galactose-glucose-rhamnose-ribitol-phosphate) and,

thus, the major subtype kept the serotype 6A designation (19). In contrast, the repeating unit of the minor subtype was found to be glucose-glucose-rhamnose-ribitol-phosphate, and the minor subtype was named serotype 6C (19). In the present study, we describe the genetic differences between 6A and 6C.

MATERIALS AND METHODS

Bacterial strains and culture. The pneumococcal strains used in this study are listed in Table 1. In addition to the 6C isolates from Brazil that were reported earlier (14), we identified additional 6C strains by retyping the preexisting pneumococcal isolates archived in our laboratory as the “6A” serotype. One isolate (BGO-2197) was obtained in 1979 from the nasopharynx of a healthy child in Birmingham, AL (S. K. Hollingshead, unpublished). The collection includes 6A isolates used for studies by Robinson (23) and Mavroidi (16). The TIGR4JS4 strain is a noncapsulated variant of the TIGR4 strain (26) and was produced by replacing the type 4 capsule gene locus with a Janus cassette (*kanA-rpsL*⁺) and backcrossing three times to wild-type TIGR4 (25, 27; Hollingshead, unpublished). TIGR6AX, TIGR6A4, TIGR6C4, and TIGR4JS4 are four variants of TIGR4 expressing, respectively, no capsule, 6A capsule, 6C capsule, or no capsule. These variants were produced as described below.

PCR and DNA sequencing. The PCR mixture had 10 to 30 ng of chromosomal DNA, 1 µl of each primer from a 100-pmol stock, 2 µl of 10 mM deoxynucleoside triphosphate, 5 µl of 10× buffer solution, 0.5 µl (2.5 U) of *Taq* polymerase (Takara Biomedical, Shiga, Japan), and 39.5 µl of sterile water (Sigma, St. Louis, MO). Chromosomal DNA was isolated with a Wizard genomic DNA purification kit (Promega, Madison, WI), according to the manufacturer's instructions. Thermal cycling conditions were customized for each primer pair. The size of the PCR products was determined by electrophoresis in a 1% to 1.5% agarose gel.

The primers used to amplify the *wciN*, *wciO*, and *wciP* genes were described by Mavroidi et al. (16). Additional primers were designed by us using the DNA sequences of the 6A and 6B capsule gene loci in GenBank (accession numbers CR931638 and CR931639, respectively). The newly designed PCR primers are listed in Table 2.

The PCR products were purified with a Wizard PCR clean-up system (Promega), and the DNA sequences of the PCR products were determined by the UAB genomics core facility. DNA sequences were analyzed with Lasergene v. 5.1 software (DNASTAR, Madison, WI) and the Basic Local Alignment Search Tool (BLAST; (<http://www.ncbi.nlm.nih.gov/BLAST/>)). Potential operons and genes, promoters, and transcription terminators in the capsule gene locus were identified using fgenesB, BPROM, and FindTerm (Softberry Inc.), which are available at the website www.molquest.com. The sequences from the capsule gene locus were compared with the sequences previously reported (16).

* Corresponding author. Mailing address: University of Alabama at Birmingham, 845 19th Street South, BBRB 614, Birmingham, AL 35294. Phone: (205) 934-0163. Fax: (205) 975-2149. E-mail: nahm@uab.edu.

[∇] Published ahead of print on 18 June 2007.

TABLE 1. List of pneumococcus strains

Strain name	Serotype	Tissue location	Country of origin (yr[s] of isolation)	Source or reference
CHPA37	6C	Nasopharynx	United States (1999–2002)	17
CHPA388	6C	Nasopharynx	United States (1999–2002)	17
BGO-2197	6C	Nasopharynx	United States (1979)	Hollingshead, unpublished
MX-67 CMN	6C	Bronchus	Mexico (1996)	23
ACA-C21	6C	Nasopharynx	Canada (1995)	23
BZ17	6C	CSF ^a	Brazil (2003)	14
BZ39	6C	CSF	Brazil (2003)	14
BZ86	6C	CSF	Brazil (2003)	14
BZ650	6C	CSF	Brazil (2003)	14
ST260	6C	CSF	Brazil (2003)	This study
KK177	6C	Oropharynx	Korea (2005)	This study
CH66	6C	Nasopharynx	China (1997)	23
CH158	6C	Nasopharynx	China (1997)	23
CH199	6C	Nasopharynx	China (1998)	23
CHPA67	6A	Nasopharynx	United States (1999–2002)	17
CHPA78	6A	Nasopharynx	United States (1999–2002)	17
BZ652	6A	CSF	Brazil (2003)	14
KK58	6A	Oropharynx	Korea (2005)	This study
AAU-33	6A	Blood	United States (1998)	16
TIGR4JS4	Noncapsulated	Derived from TIGR4 ^b	Not applicable	27
TIGR6A4	6A	Derived from TIGR4JS4	Not applicable	This study
TIGR6AX	Noncapsulated	Derived from TIGR6A4	Not applicable	This study
TIGR6C4	6C	Derived from TIGR6A4	Not applicable	This study

^a CSF, cerebrospinal fluid.^b TIGR4 expresses serotype 4 and was originally isolated from blood (26).

Production of TIGR4 variants with 6A and 6C capsule gene loci. To investigate the role of the *wciN* gene in 6C capsule expression, we first inserted the 6A capsule locus into the genetic background of TIGR4 and then replaced the *wciN*_{6A} gene with *wciN*_{6C} using two different DNA cassettes (labeled cassette 1 and cassette 2 in Fig. 1) as described below. Cassette 1 has three parts: the target DNA and two flanking DNAs. The target DNA contains the kanamycin resistance (*kanA*) and streptomycin sensitivity (*rpsL*⁺) genes of the Janus cassette (25). The two flanking DNAs were obtained from either *wchA* or *wciO-P* genes from AAU-33 (a 6A strain) by PCR using the primer pairs described in Fig. 1 and Table 2. The three DNA fragments in cassette 1 were then linked together by digestion with the appropriate restriction enzyme, followed by ligation with T4 DNA ligase (New England BioLabs, Beverly, MA). The ligation product was then amplified by PCR using primers 5113 and 3102. Cassette 2, which was used to replace the antibiotic selection genes with the *wciN*_{6C} gene, was prepared by PCR of CHPA388 (a 6C strain) DNA using primers 5113 and 3102. The PCR products were purified with the Wizard PCR clean-up system (Promega). After confirming the DNA sequences, the PCR products were used for transformation.

The *wciN* replacement was performed as follows. TIGR4JS4, which has a Janus cassette (25) in place of TIGR4's original capsule gene locus (27), was transformed with genes from a 6A strain, AAU-33. A streptomycin-resistant but kanamycin-sensitive isolate was then selected, and DNA from it was used to backcross into TIGR4JS4. The backcross was repeated three times, each time testing for expression of 6A capsular PS. A final backcrossed transformant expressing 6A capsular PS was selected and named TIGR6A4. To remove the *wciN*_{6A} gene from TIGR6A4, it was transformed with cassette 1, and a kanamycin-resistant transformant was obtained. After the resulting transformant was backcrossed three times on the TIGR6A4 strain background, the final capsule-deficient kanamycin-resistant transformant was selected and named TIGR6AX. To insert the *wciN*_{6C} gene into TIGR6AX, it was then transformed with cassette 2, and a kanamycin-sensitive but streptomycin-resistant transformant was obtained. Backcrossing once more for three times on the TIGR6AX background resulted in a final 6C capsular PS transformant that was named TIGR6C.

Nucleotide sequence accession numbers. The nucleotide sequences of the capsule gene locus of pneumococcal serotype 6C and the *wciN*_{6C} region of several pneumococcal serotype 6C isolates have been submitted to GenBank and assigned accession numbers EF538714 to EF538718.

RESULTS

Identification of additional 6C strains among our “6A” collections. To obtain a representative collection of 6C serotypes from various locations, we retested our preexisting collection of “6A” strains with our monoclonal antibodies (19) and identified nine additional 6C isolates from five countries on three different continents (Table 1). These isolates were obtained from spinal fluid, blood, and the nasopharynx samples, indicating that 6C can cause invasive pneumococcal infections as well as be isolated in asymptomatic carriage. One strain (BGO-2197) was isolated in 1979 in Birmingham, AL. This finding shows that the 6C serotype has been in existence for at least 27 years and is now found throughout the world.

The capsule gene loci of 6A and 6C differ in the region between the *wchA* and *wciO* genes. Because galactose of 6A PS is replaced with glucose in 6C PS (19), we hypothesized that the galactosyl transferase gene in the 6A serotype is replaced with a new glycosyl transferase gene in 6C. Since the *wciN* gene encodes galactosyl transferase, we used PCR to compare the sizes of *wciN* genes of 6A and 6C isolates. The *wciN* PCR products of all 6C isolates were about 1.8 kb long, whereas the *wciN* PCR products of all 6A isolates were about 2 kb long (Fig. 2). To distinguish between the two *wciN* genes from the 6A and 6C serotypes, we have named them *wciN*_{6A} and *wciN*_{6C}, respectively.

To further investigate *wciN*_{6C}, we determined the DNA sequences of the capsule gene region, including the *wchA*, *wciN*_{6C}, and *wciO* genes from five 6C strains (BZ17, BZ86, CHPA388, KK177, and ST-260). Since their sequences were almost identical, the actual DNA sequence is shown for only

TABLE 2. List of PCR primers

Primer direction and name	Primer site of accession no. CR931638	Description ^a	Sequence	Source or reference
Forward primers				
5101	6949–6966	In <i>wciN</i> , for INDEL detection	5'-ATTTGGTGTACTTCCTCC	16
5103	8146–8168	In <i>wciO</i> , for sequencing the 6C capsule gene	5'-AAACATGACATCAATTACA	This study
5106	5897–5916	In <i>wchA</i> , for <i>wciN</i> detection	5'-TACCATGCAGGGTGGAATGT	This study
5108	8350–8370	In <i>wciP</i> , for <i>wciP</i> allele detection	5'-ATGGTGAGAGATATTGTCAc	16
5112	Not applicable	In <i>kanA-rpsL</i> ⁺ with XbaI site	5'-CTAGTCTAGAGTTTGATTTTAAATGG	This study
5113	4870–4894	In <i>wze</i> , for fragment C	5'-GGGAAAAATAAAAAATAGGTCGGG	This study
5118	7613–7636	In <i>wciO</i> with BamHI site	5'-CGCGGATCCAGAAAACTATGTCGCCTGCTAAa	This study
5120	1–30	In <i>dexB</i> , for fragment A	5'-TGTCCAATGAAGAGCAAGACTTGACA GTag	27
5122	2187–2206	In <i>wzg</i> , for fragment B	5'-TTCGTCCATTACACCTTAG	This study
5123	8775–8794	In <i>wciP</i> , for fragment D	5'-ATCGTATATCTGGGGGTGTA	This study
5124	11274–11293	In <i>wzx</i> , for fragment E	5'-AATGATTGGGCGGATGTTT	This study
5125	13864–13883	In <i>rmlC</i> , for fragment F	5'-AGTGATTGATGCGAGTAAGG	This study
5140	9531–9551	In <i>wzy</i> , for <i>wzy</i> allele detection	5'-CCTAAAGTGGAGGGAATTTTCG	16
5141	11459–11478	In <i>wzx</i> , for <i>wzx</i> allele detection	5'-TTCGAATGGGAATTCAATGG	16
Reverse primers				
3101	7888–7905	In <i>wciO</i> , for INDEL and <i>wciN</i> detection	5'-CCATCCTTCGAGTATTGC	16
3103	9468–9487	In <i>wzy</i> , for Janus cassette and fragment C	5'-AACCCCTAACAAATATCAAAT	This study
3107	9226–9245	In <i>wciP</i> , for <i>wciP</i> allele detection	5'-AGCATGATGGTATATAAGCC	16
3112	Not applicable	In <i>kanA-rpsL</i> ⁺ with BamHI site	5'-CGCGGATCCGGGCCCCCTTTCCTTATGCTTTGG	This study
3113	6203–6224	In <i>wchA</i> with XbaI site	5'-CTAGTCTAGAAATAAAATTTCAATATCTTTCCAG	This study
3121	3676–3660	In <i>wzd</i> , for fragment A	5'-GATTGCGATTCACTACG	This study
3122	5380–5361	In <i>wchA</i> , for fragment B	5'-AACTCCCCAACCAACCTCATT	This study
3123	12978–12959	In <i>rmlA</i> , for fragment D	5'-AAAATCAAGGCAACGCTATC	This study
3124	14618–14600	In <i>rmlB</i> , for fragment E	5'-ACGGAGAGCTTGGGTTGTA	This study
3126	17611–17584	In <i>aliA</i> , for fragment F	5'-CAATAATGTCACGCCCGCAAGGGCAAGT	27
3143	10135–10115	In <i>wzy</i> , for <i>wzy</i> allele detection	5'-CCTCCCATATAACGAGTAGT	16
3144	12068–12049	In <i>wzx</i> , for <i>wzx</i> allele detection	5'-GCGAGCCAAATCGGTAAGTA	16

^a Fragments A through F refer to the corresponding fragments of the serotype 6C capsule gene locus used for capsule gene locus sequencing.

CHPA388 (Fig. 3A), and the sequences of the other isolates were deposited in GenBank (accession numbers EF538715 to EF538718). The sequence of the *wciN*_{6C} gene from CHPA388 was then compared with the 6A sequence of the corresponding region available in GenBank (accession no. CR931638). A summary of the comparison is shown in Fig. 3B.

Sequence comparison revealed clear differences in *wciN*_{6A} and *wciN*_{6C} genes. The 6C serotype has 1,029 bp in place of 1,222 bp in 6A (Fig. 3B and C). The two *wciN* genes are completely different, with a sequence homology of only about 50%. The decreased homology begins immediately after the termination of the *wchA* gene (position 1368) and ends 130 bases upstream of the beginning of *wciO* (6A-2631 and 6C-2398) (Fig. 3B and C). The different regions are indicated in Fig. 3B. When the DNA sequences flanking the replaced gene were compared between 6A and 6C, significantly more DNA polymorphisms were found in the two flanking regions than in the regions outside of the two flanking regions. For instance, the 300 bases upstream of the replaced gene have 25 different nucleotides, but the 150 bases located immediately upstream from the 300 bases have only one different base ($P < 0.001$ by Fisher's exact test) (Fig. 3B). Similarly, in the 3' direction, 20

bases differ in the proximal 110 bases (positions 2398 to 2508 for CHPA388), but only 1 base differs in the next 300 bases ($P < 0.001$ by Fisher's exact test) (Fig. 3B). These findings are not unique to this particular 6A sequence (accession no. CR931638), because similar results were obtained with the sequence of seven different 6A strains: AAU-33, D020-1B, HS3050, CHPA78, KK65, ST19, and ST558. These findings suggest that the two flanking regions were also introduced with *wciN*_{6C} into 6A to create 6C. The fact that all 6C isolates have the identical flanking region sequences is strong evidence that the major genetic replacement generating 6C took place only once and that the capsule gene locus present in all the 6C isolates must have originated from that initial replacement event.

With this gene replacement, *wciN*_{6C} has a new open reading frame (ORF) that is 1,125 bases long and encodes a peptide with 374 amino acids, which is designated WciN_{6C} (Fig. 3A). The termination codon of the new ORF is between the two potential start codons for *wciO*, which are located at positions 2497 and 2528 of the sequence of CHPA388 (Fig. 3A). When the sequence of the *wciN*_{6C} gene was compared with the sequences in the database, 110 bases (from 1627 to 1736 in 6C)

FIG. 1. *wciN* region exchange experiment diagram. In step A, the *wchA/wciN*_{6A}/*wciO-P* region of TIGR6A4 was replaced with cassette 1. Cassette 1 has three parts (central core and two flanking regions), and each part is about 1 kb long. The central core has antibiotic susceptibility genes *kanA* and *rpsL*⁺. The two flanking regions were made with *wchA* and *wciO-P* regions from the AAU-33 strain. In step B, cassette 1 in TIGR6AX is replaced with cassette 2. Cassette 2 has *wciN*_{6C}, *wchA*, and *wciO-P* regions from a 6C strain (CHPA388). TIGR6C4 shows the final product that is obtained after cassette 2 is inserted. *Xba*I and *Bam*HI sites in the PCR primers, which were introduced to simplify genetic manipulations, are shown.

of 6C demonstrated 81% homology to the 90 bases of the exopolysaccharide synthesis gene of *Streptococcus thermophilus* strain CNRZ1066 (3). Also, the translated sequence of the *wciN*_{6C} gene has 22% amino acid identity and 44% similarity to the translated sequence of the *capH* gene of *Staphylococcus aureus* (15). The *wciN*_{6C} gene product is a member of the WaaG family (20). Incidentally, the *waaG* gene product of *Escherichia coli* K-12 is an α -1,3-glucosyltransferase involved in lipopolysaccharide synthesis (10).

The *wciN* gene region is responsible for conversion from the 6A to 6C serotype. Although the above studies showed that the major difference is in the *wciN* region, it is possible that some other genetic differences outside of the capsule gene locus could be involved in the 6C expression. To show that only the

wciN region is involved, we determined whether the replacement of the *wciN*_{6A} region with the *wciN*_{6C} region could convert the 6A serotype to the 6C serotype using the “Janus” cassette (27). As shown in Fig. 1, we first produced TIGR6A by replacing the capsule locus of TIGR4 with the 6A capsule gene locus from strain AAU-33. We then removed the *wciN*_{6A} gene from TIGR6A by transforming it with cassette 1. The resulting strain, named TIGR6AX, was nonencapsulated and was found, via PCR, to have lost the *wciN*_{6A} gene. The *wciN*_{6C} region was then inserted into TIGR6AX using cassette 2, which contained the *wciN*_{6C} and *wciO* genes from CHPA388 as well as parts of *wchA* and *wciP* genes. The 3' flanking regions of CHPA388 and AAU-33 had only minimal differences: they differed by two amino acids between their *wciO* gene products

FIG. 2. Electrophoretic patterns of the PCR products of the *wciN* region of 6A and 6C isolates. The primers used were 5106 and 3101, which are located in *wchA* and *wciO*, respectively. M, DNA ladder. Standard markers with 2,000 and 1,650 bp are indicated on the left. Lanes 1 to 13 contain PCR products of 6C isolates CHPA37 (lane 1), CHPA388 (lane 2), BG2197 (lane 3), BZ17 (lane 4), BZ39 (lane 5), BZ86 (lane 6), BZ650 (lane 7), KK177 (lane 8), CH66 (lane 9), CH158 (lane 10), CH199 (lane 11), MX-67 (lane 12), and ACA-C21 (lane 13). Lanes 14 to 18 contain PCR products of 6A isolates CHPA67 (lane 14), CHPA78 (lane 15), BZ652 (lane 16), KK58 (lane 17), and AAU-33 (lane 18).

and by one amino acid between the parts of the *wciP* gene products involved in the genetic exchange. After the insertion of cassette 2, the resulting strain was named TIGR6C, and it was found to express serotype 6C. In addition, TIGR6C was

found to have the *wciN_{6C}* gene at the expected location when the region was sequenced. This confirmed that the *wciN_{6C}* gene and the 1-kb regions surrounding *wciN_{6C}* are sufficient for the serotype conversion. Also, although 6A and 6C have few

FIG. 3. A. The nucleotide sequence of the *wciN_{6C}* ORF is shown along with the nucleotide sequences of the 3' end of *wchA* and the 5' end of *wciO*. The derived amino acid sequence of the *wciN_{6C}* ORF is shown below the nucleotide sequence. Also shown are putative termination sites of *wchA* and *wciN_{6C}* as well as putative initiation sites of *wciN_{6C}* and *wciO*. *wciO* has two potential initiation sites. B. DNA sequences of *wciN_{6A}* and *wciN_{6C}* regions (shown in boldface) of a 6A strain (GenBank accession no. CR931638) and a 6C strain (CHPA388). The sequence of the nonhomologous mid-region of *wciN* (about 900 to 1,110 bases) is not shown. Sites of PCR primers (5106, 3101, 5118, and 3113) are shown. Also shown are potential termination sites of *wchA* and *wciN_{6C}* and potential initiation sites of *wciN_{6C}* and *wciO*. C. Genetic map of the capsule loci surrounding *wciN* of 6A and 6C isolates. The map shows *wchA* (hatched), *wciN* (horizontal bars or black), *wciO* (checkered), and *wciP* (wavy lines) genes. The 6A locus has two unexpressed DNA fragments (indicated by arrows) upstream (95 bases long) or downstream (312 bases long) of *wciN_{6A}*. An alternative initiation site for *wciO* is located 32 bases upstream of the initiation site shown (position 2721 for 6A). For 6C isolates, the native DNA (1,222 bases; indicated by horizontal bars) in the *wciN_{6A}* locus is replaced with new DNA (1,029 bases; black). The replacement creates a new ORF (named *wciN_{6C}*) that has 1,125 bases. Nucleotide position 1 in this figure corresponds to nucleotide position 4902 of the 6A capsule genome sequence (GenBank accession no. CR931638).

amino acid differences in the *wciO* and *wciP* gene products, the *wciN* gene is most likely responsible for the serotype conversion from 6A to 6C.

The sequences of the capsule gene loci of the 6A and 6C serotypes differ only slightly in regions other than the *wciN* gene. To determine if the 6A and 6C capsule gene loci differ only in the *wciN* region, we determined the sequence of the entire capsule locus of a 6C isolate, CHPA388, by PCR amplifying the entire capsule gene locus between *dexB* and *aliA* in six overlapping DNA fragments (fragments A to E) using primers shown in Table 2. Figure 4 shows the genetic map of the sequence of the capsule gene locus. The 6C capsule gene locus is about 17 kb long and contained 14 ORFs (Fig. 4) involved in capsular PS synthesis. The 6C ORFs are all on the same strand and correspond exactly to those found in the 6A capsule locus with the exception of the *wciN* region. *wciP* of 6C has G at position 584, as *wciP* of 6A does. The ORFs of 6C begin with *cpsA* at the 5' end and terminate with *rmlD* at the 3' end. Both 6A and 6C capsule loci have four potential transcription start sites and a putative rho-independent transcription terminator site. Also, there are "transposase-like" sequences (Fig. 4) at both ends of the 6C capsule gene locus, as are commonly found for many pneumococcal capsule cassette

(2). The nucleotide sequence of the entire locus has been deposited in GenBank (accession number EF538714).

When the 6C sequence was compared with the capsule locus of a 6A strain (GenBank accession no. CR931638), we found that, except for the *wciN* region described above, the capsule gene locus of 6C was very homologous (~98%) to that of 6A. Also, homology was significantly low (about 78%) for about 60 bp in the middle of the *cpsA* ORF, and the "transposase-like" sequences found at either end of the capsule gene loci were different between the 6A and 6C capsule gene loci. The 6C capsule gene locus did not have the INDEL that is present upstream of *wciO* in some 6A or 6B capsule loci (16). Despite these differences, the most prominent differences between the 6A and 6C capsule loci were found in the *wciN* region.

DISCUSSION

Following our initial discovery of 6C among "6A" isolates using monoclonal antibodies (14, 19), we now show the genetic basis for the new serotype, 6C. The capsule gene locus of 6C is very similar to the 6A locus except for *wciN*: 6A strains have *wciN_{6A}* but 6C strains have *wciN_{6C}*, which is dramatically different in its size (about 200 bases shorter) as well as its se-

FIG. 4. Capsule locus of a 6A strain (GenBank accession no. CR931638) and a 6C strain (CHPA388). All ORFs involved in capsule synthesis are shown as horizontal arrows, and their direction indicates the transcriptional orientation. For both 6A and 6C loci, the putative transcription initiation sites (bent arrows) and putative termination sites (vertical lines with a solid circle) were identified using fgenesB, BPRM, and FindTerm (Softberry Inc.), available at www.molquest.com. “Transposase” sequences (black boxes, labeled “tnp”) are found at either end of the capsule gene locus.

quence (only about 50% DNA homology) compared to *wciN*_{6A}. The size differences have been confirmed with additional 6C isolates by PCR (unpublished information). Chemical studies of purified 6C PS found that repeating unit is identical to 6A PS except for glucose in place of the galactose (19). Consistent with this, *WciN*_{6A} belongs to glycosyl transferase family 8, which includes many galactosyl transferases (4), but *WciN*_{6C} belongs to glycosyl transferase family 1 and resembles RfaG (WwaG), which is a glucosyl transferase involved in *E. coli* lipopolysaccharide synthesis (8). Furthermore, a 6A strain can be converted to a 6C strain by replacing the *wciN*_{6A} region with the *wciN*_{6C} region. Also involved in this genetic exchange are *wciO* and parts of the *wciP* and *wchA* genes. When the amino acid sequences of these coincidentally involved genes were compared between 6A and 6C strains, differences were minimal: two amino acids for *wciO*, one amino acid for the 5' end of *wciP*, and nine amino acids for *wchA*. These findings strongly support that *wciN*_{6A} encodes a galactosyl transferase and *wciN*_{6C} encodes a glucosyl transferase. Thus, we conclude that the *wciN*_{6C} gene is the basis for 6C serotype.

Capsule production involves many enzymes, and an alteration in the repeating unit of the capsular PS must still be recognized by other enzymes, such as the flippase and polymerase, which are involved in transporting and linking the altered repeating units. It is unlikely, however, that new genes other than *wciN*_{6C} are involved in producing 6C PS compared to 6A. We found that homologous recombination between the two *wciN* regions is sufficient to alter 6A serotype to 6C. Also, when the changes in the repeating unit are small, the other enzymes can accommodate the new repeating unit without any changes. 6A and 6B serotypes have different rhamnosyl transferases and produce PS with different rhamnose-ribitol linkages, but other genes can accommodate both repeating units. Similarly, glucose and galactose have very small chemical differences—a different orientation of the hydroxyl group of the fourth carbon molecule—and this minor chem-

ical difference is often accommodated. For instance, 9L PS has a galactose molecule and 9N PS has a glucose molecule. Their capsule loci resemble each other except for one gene, *wcjA*, which encodes a galactosyl transferase for 9L and a glucosyl transferase for 9N (2).

In the case of the 9L and 9N serotypes, their *wcjAs* are very similar and must have originated from a common precursor gene. In contrast, the *wciN*_{6A} and *wciN*_{6C} genes are very different. Perhaps the *wciN*_{6C} gene is from an organism other than pneumococci and was inserted into the 6A capsule gene locus. Horizontal gene transfers between *S. pneumoniae* and another bacterial species have been demonstrated with antibiotic resistance genes (5, 18) and have been suggested to occur for capsule gene loci because they have low G+C content (2). The “flanking regions” are known to be critical for homologous recombination in pneumococci (21). Indeed, an examination of *wciN*_{6C} shows evidence for the two flanking regions that may have participated in the homologous recombination.

The source of *wciN*_{6C} is not yet known. *wciN*_{6C} resembles no pneumococcal genes, including the capsule gene locus genes of the 90 different, non-6C serotypes (2). About 100 bases of the *wciN*_{6C} gene are similar (81% homology) to the *epsG* gene of *S. thermophilus*, a gene involved in the synthesis of exopolysaccharide by *S. thermophilus*. Although the organization of exopolysaccharide synthesis gene loci is similar to pneumococcal capsule gene loci, the homology is short, suggesting that *S. thermophilus* is not the direct source for *wciN*_{6C}. The protein sequence of *WciN*_{6C} resembles the *waaG* (*rfaG*) gene product of the *E. coli* K-12 strain, and some pneumococcal genes apparently did come from gram-negative organisms, such as *Haemophilus* (13). Thus, it is possible that the *wciN*_{6C} gene could have come from a gram-negative species as well. Nevertheless, other oral streptococci, such as *Streptococcus salivarius*, *Streptococcus mitis*, and *Streptococcus oralis*, are the leading candidates, since they coexist in the oral cavity with pneumococci and many antibiotic resistance genes came from *S. oralis*.

When the *wciN*_{6C} region was examined for a 27-year-old

isolate and multiple recent isolates from different continents, flanking region sequences to *wciN*_{6C} were found to be identical. Also, the 6C serotype has only one (or two) capsule gene locus profile (data not shown) (14), whereas the 6A and 6B serotypes have diverse capsule gene locus profiles (16). These findings suggest that a single bacterium producing 6A PS captured *wciN*_{6C} and that all the 6C isolates found throughout the world have the capsule gene locus that originated in this 6C founder bacterium. The lack of diversity in the 6C capsule gene loci suggests that the capture of *wciN*_{6C} might have occurred recently compared to the history of 6A or 6B capsule gene loci. This singular origin of the 6C capsule gene locus may make 6C a good model for studying bacterial genetic evolution in response to the host immune system. Also, pneumococcal vaccines may not protect us against 6C as well as against 6A and the vaccination may increase the prevalence of 6C. If this happens, it would make 6C an even more interesting model for studying bacterial genetic evolution.

Because of the medical importance, pneumococcal serotypes have been extensively investigated using serological and biochemical as well as genetic tools. Thus, it has been commonly believed that additional pneumococcal serotypes, if found, would be discovered among nontypeable pneumococci. We now show that even a well-characterized serotype such as serotype "6A" harbors a distinct serotype with a relatively large genetic change that has eluded our intensive investigations. Thus, we should be aware that new serotypes can exist among the serotypeable pneumococci, as well as among the nontypeable.

ACKNOWLEDGMENTS

We thank Janet Yother, David Briles, William Benjamin, Jr., and Susan Michalek for helpful discussions and comments on the manuscript. We are also grateful to Kyung Hyo Kim and M. Catherine McEllistrem for providing us with pneumococcal isolates. We are indebted to the genomics core facility at UAB for DNA sequencing.

The work was supported by NIH funding AI-031473 and AI-30021 to M.H.N.

REFERENCES

1. Avery, O. T., and R. Dubos. 1931. The protective action of a specific enzyme against type III pneumococcus infection in mice. *J. Exp. Med.* **54**:73–89.
2. Bentley, S. D., D. M. Aanensen, A. Mavroidi, D. Saunders, E. Rabinowitsch, M. Collins, K. Donohoe, D. Harris, L. Murphy, M. A. Quail, G. Samuel, I. C. Skovsted, M. S. Kalltoft, B. Barrell, P. R. Reeves, J. Parkhill, and B. G. Spratt. 2006. Genetic analysis of the capsular biosynthetic locus from all 90 pneumococcal serotypes. *PLoS Genet.* **2**:e31.
3. Bolotin, A., B. Quinquis, P. Renault, A. Sorokin, S. D. Ehrlich, S. Kulakauskas, A. Lapidus, E. Goltsman, M. Mazur, G. D. Pusch, M. Fonstein, R. Overbeek, N. Kyprides, B. Purnelle, D. Prozzi, K. Ngui, D. Masuy, F. Hancy, S. Burteau, M. Boutry, J. Delcour, A. Goffeau, and P. Hols. 2004. Complete sequence and comparative genome analysis of the dairy bacterium *Streptococcus thermophilus*. *Nat. Biotechnol.* **22**:1554–1558.
4. Campbell, J. A., G. J. Davies, V. Bulone, and B. Henrissat. 1997. A classification of nucleotide-diphospho-sugar glycosyltransferases based on amino acid sequence similarities. *Biochem. J.* **326**:929–939.
5. Coffey, T. J., C. G. Dowson, M. Daniels, J. Zhou, C. Martin, B. G. Spratt, and J. M. Musser. 1991. Horizontal transfer of multiple penicillin-binding protein genes, and capsular biosynthetic genes, in natural populations of *Streptococcus pneumoniae*. *Mol. Microbiol.* **5**:2255–2260.
6. Cole, R. 1913. Treatment of pneumonia by means of specific serums. *JAMA* **61**:663–666.
7. Cooper, G., M. Edwards, and C. Rosenstein. 1929. The separation of types among the pneumococci hitherto called group IV and the development of therapeutic antisera for these types. *J. Exp. Med.* **49**:461–474.
8. Coutinho, P. M., E. Deleury, G. J. Davies, and B. Henrissat. 2003. An evolving hierarchical family classification for glycosyltransferases. *J. Mol. Biol.* **328**:307–317.
9. Fedson, D. S. 1988. Pneumococcal vaccine, p. 271–299. In S. A. Plotkin and E. A. Mortimer, Jr. (ed.), *Vaccines*. W.B. Saunders Co., Philadelphia, PA.
10. Heinrichs, D. E., J. A. Yethon, and C. Whitfield. 1998. Molecular basis for structural diversity in the core regions of the lipopolysaccharides of *Escherichia coli* and *Salmonella enterica*. *Mol. Microbiol.* **30**:221–232.
11. Heinrichs, J. 1995. Six newly recognized types of *Streptococcus pneumoniae*. *J. Clin. Microbiol.* **33**:2759–2762.
12. Kenne, L., B. Lindberg, and J. K. Madden. 1979. Structural studies of the capsular antigen from *Streptococcus pneumoniae* type 26. *Carbohydr. Res.* **73**:175–182.
13. Lacks, S. A. 2004. Transformation, p. 89–115. In E. I. Tuomanen, T. J. Mitchell, D. A. Morrison, and B. G. Spratt (ed.), *The pneumococcus*. ASM Press, Washington, DC.
14. Lin, J., M. S. Kalltoft, A. P. Brandao, G. Echaniz-Aviles, M. C. Brandileone, S. K. Hollingshead, W. H. Benjamin, Jr., and M. H. Nahm. 2006. Validation of a multiplex pneumococcal serotyping assay with clinical samples. *J. Clin. Microbiol.* **44**:383–388.
15. Lin, W. S., T. Cunneen, and C. Y. Lee. 1994. Sequence analysis and molecular characterization of genes required for the biosynthesis of type 1 capsular polysaccharide in *Staphylococcus aureus*. *J. Bacteriol.* **176**:7005–7016.
16. Mavroidi, A., D. Godoy, D. M. Aanensen, D. A. Robinson, S. K. Hollingshead, and B. G. Spratt. 2004. Evolutionary genetics of the capsular locus of serogroup 6 pneumococci. *J. Bacteriol.* **186**:8181–8192.
17. McEllistrem, M. C., J. M. Adams, K. Patel, A. B. Mendelsohn, S. L. Kaplan, J. S. Bradley, G. E. Schutze, K. S. Kim, E. O. Mason, and E. R. Wald. 2005. Acute otitis media due to penicillin-nonsusceptible *Streptococcus pneumoniae* before and after the introduction of the pneumococcal conjugate vaccine. *Clin. Infect. Dis.* **40**:1738–1744.
18. Muller-Graf, C. D., A. M. Whatmore, S. J. King, K. Trzcinski, A. P. Pickerill, N. Doherty, J. Paul, D. Griffiths, D. Crook, and C. G. Dowson. 1999. Population biology of *Streptococcus pneumoniae* isolated from oropharyngeal carriage and invasive disease. *Microbiology* **145**:3283–3293.
19. Park, I. H., D. G. Pritchard, R. Cartee, A. Brandao, M. C. C. Brandileone, and M. H. Nahm. 2007. Discovery of a new capsular serotype (6C) within serogroup 6 of *Streptococcus pneumoniae*. *J. Clin. Microbiol.* **45**:1225–1233.
20. Parker, C. T., E. Pradel, and C. A. Schnaitman. 1992. Identification and sequences of the lipopolysaccharide core biosynthetic genes *rfaQ*, *rfaP*, and *rfaG* of *Escherichia coli* K-12. *J. Bacteriol.* **174**:930–934.
21. Prudhomme, M., V. Libante, and J. P. Claverys. 2002. Homologous recombination at the border: insertion-deletions and the trapping of foreign DNA in *Streptococcus pneumoniae*. *Proc. Natl. Acad. Sci. USA* **99**:2100–2105.
22. Rebers, P. A., and M. Heidelberger. 1961. The specific polysaccharide of type VI pneumococcus. II. The repeating unit. *J. Am. Chem. Soc.* **83**:3056–3059.
23. Robinson, D. A., D. E. Briles, M. J. Crain, and S. K. Hollingshead. 2002. Evolution and virulence of serogroup 6 pneumococci on a global scale. *J. Bacteriol.* **184**:6367–6375.
24. Robinson, K. A., W. Baughman, G. Rothrock, N. L. Barrett, M. Pass, C. Lexau, B. Damaske, K. Stefonek, B. Barnes, J. Patterson, E. R. Zell, A. Schuchat, and C. G. Whitney. 2001. Epidemiology of invasive *Streptococcus pneumoniae* infections in the United States, 1995–1998: opportunities for prevention in the conjugate vaccine era. *JAMA* **285**:1729–1735.
25. Sung, C. K., H. Li, J. P. Claverys, and D. A. Morrison. 2001. An *rpsL* cassette, Janus, for gene replacement through negative selection in *Streptococcus pneumoniae*. *Appl. Environ. Microbiol.* **67**:5190–5196.
26. 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.
27. Trzcinski, K., C. M. Thompson, and M. Lipsitch. 2003. Construction of otherwise isogenic serotype 6B, 7F, 14, and 19F capsular variants of *Streptococcus pneumoniae* strain TIGR4. *Appl. Environ. Microbiol.* **69**:7364–7370.