Pathogenic *Neisseria meningitidis* isolates contain a polysaccharide capsule that is the main virulence determinant for this bacterium. Thirteen capsular polysaccharides have been described, and nuclear magnetic resonance spectroscopy has enabled determination of the structure of capsular polysaccharides responsible for serogroup specificity. Molecular mechanisms involved in *N. meningitidis* capsule biosynthesis have also been identified, and genes involved in this process and in cell surface translocation are clustered at a single chromosomal locus termed *cps*. The use of multiple names for some of the genes involved in capsule synthesis, combined with the need for rapid diagnosis of serogroups commonly associated with invasive meningococcal disease, prompted a requirement for a consistent approach to the nomenclature of capsule genes. In this report, a comprehensive description of all *N. meningitidis* serogroups is provided, along with a proposed nomenclature, which was presented at the 2012 XVIIIth International Pathogenic Neisseria Conference.

Thirteen *Neisseria meningitidis* serogroups have been described on the basis of serologic differences of the capsule; of these 13 serogroups, 6 (A, B, C, W, X, Y) cause invasive meningococcal disease. The polysaccharide capsule is a key virulence determinant, and for serogroups A, C, W, and Y, it forms the basis of polysaccharide conjugate vaccines. In one of the first reports distinguishing *N. meningitidis*, disease isolates were serologically classified into types I–IV on the basis of agglutination reactions with immune rabbit serum (*J*). In 1950, the subcommittee on *Neisseria* of the Nomenclature Committee of the International Association of Microbiologists recommended that types I and III be combined into serogroup A; type II become serogroup B; a type II subgroup, termed type II-α, become serogroup C; and type IV become serogroup D. After the report of a fourth serogroup, Z′ (later shown to be serogroup E), 3 new serogroups (X–Z) were identified by using double agar diffusion (*2, 3*). In 1981, three more serogroups (H, I, K) were proposed, and a fourth (serogroup L) was identified in 1983 (*4, 5*).

Nuclear magnetic resonance spectroscopy enabled determination of the structure of capsular polysaccharides responsible for serogroup specificity, and structures for 12 of the 13 serogroups (all but serogroup D) from *N. meningitidis* capsular polysaccharides have been reported (*6–15*). Molecular mechanisms of capsular polysaccharide synthesis have been elucidated; genes involved in polysaccharide biosynthesis and cell surface translocation are clustered at a single chromosomal locus termed *cps*. Genes within this locus are divided into 6 regions: A–D, D′, and E (*16*). Genes in region A encode enzymes for biosynthesis of the capsular polysaccharide, and genes in regions B and C are implicated in the translocation of the high molecular weight polysaccharides to the cell surface.

Complete nucleotide sequences of *cps* loci encoding serogroups A–C, W, and Y have been elucidated. Serogroup-specific capsule biosynthesis genes located in region A have been published for serogroup X, and...
nucleotide sequences for serogroups E, L, and Z have been submitted to GenBank (accession nos. AJ576117, AF112478, and AJ744766, respectively) (17–19). This study provides a comprehensive description of all \( N. \text{meningitidis} \) serogroups and presents proposed revisions to the nomenclature.

**Materials and Methods**

**Strain Selection**

Serogroup \( D, H, I, \) and \( K \) isolates were obtained from a collection maintained at the National Institute for Biological Standards and Controls, Potters Bar, UK; the isolates were originally from a collection (1980s) from the People’s Republic of China Committee for Culture Collection of Microorganisms. Serogroup \( E, W, X, \) and \( Y \) isolates were from the Bavarian \( N. \text{meningitidis} \) carriage collection (20). Three serogroup \( L \) isolates were analyzed with additional serogroup \( H, I, \) and \( K \) isolates obtained from Paula Kriz (Czech Republic), who had obtained the isolates from Fraser Ashton, who had acquired them from People’s Republic of China (Table 1). The sequenced genomes from serogroup \( A \) isolate Z2491, serogroup \( B \) isolate H44/76, and serogroup \( C \) isolates FAM18 and 053442 were used (Table 1) (21–24). Isolates were grown on Mueller-Hintonagar supplemented with 5% (vol/vol) defibrinated sterile horse blood for 15 h at 37°C in a 5% (vol/vol) \( \text{CO}_2 \) atmosphere. DNA was extracted by using the Wizard Genomic DNA Purification Kit (Promega, Southampton, UK) according to the manufacturer’s instructions.

**PCR and Nucleotide Sequencing of the Capsule Locus**

The genes for the \( N. \text{meningitidis} \) \( cps \) locus are located between a gene encoding a putative inner membrane transport protein (NMC0044 FAM18 genome annotation) and a gene encoding the sodium/glutamate symport carrier protein, \( \text{gltS} \) (NMC0069). PCR reactions were performed for isolates 29013, \( a707, 29031, 29043, 29046, 3608, a275, a388, a162, WUE171, WUE172, \) and WUE173 (Table 1) by using the Expand Long Template PCR System (Roche Applied Science, Burgess Hill, UK) according to the manufacturer’s recommended protocol, with annealing at 55°C and extension at 68°C for 20 min. Initial reactions used 1 of the following primer pairs: \( \text{gltS} \) (NMC0069) to \( \text{tex} \) (NMC0059) (primers \( \text{gltS} \) 5′-CC-GACCAAGCCGTATTTGC + ATGATACTGAGGGC-GTGGTT-3′ and \( \text{tex} \) 5′-TGTCAGAGCCGTCATATCT + GCCCTGTCCAACAGTTGTG-3′) and \( \text{tex} \) to NMC0044 (primers \( \text{tex} \) 5′-CGGCCGCTGTCGTATCC + TTGTGCTGTGGTTAGCCGAATTCC-3′ and NMC0044 5′-CGGGCAACACGCCTAT + TATCGTTGGTG-CGCTGGTTAT-3′).

**Genome Sequencing**

Genomic data for serogroups \( D, H, I, K, \) and \( L \) were obtained by using the Illumina sequencing platform and de novo assembly (Illumina, San Diego, CA, USA), using the shuffle and Velvet Optimization scripts found within Velvet 1.1.03 (25) (Table 1). By using Velvet Optimiser version 2.1.7 (http://bioinformatics.net.au/software.velvetoptimiser.shtml), optimal k-mer lengths ranging from 41 to 65 were selected and contigs were generated. These contigs were deposited in the Bacterial Isolate Genome Sequence Database (BIGS\textsubscript{db}) along with the isolate provenance and name, after which contigs were scanned for genes within the \( cps \) locus and tagged and alleles were assigned (26). New alleles were manually checked for a correct start and stop codon and aligned with known alleles before assignment. Additional data (e.g., multilocus sequence type [MLST]), and PorA, PorB, and FetA designs were obtained.

**Annotation and Bioinformatic Methods**

Predicted proteins were clustered into homology groups by using TribeMCL algorithm with a cutoff of 1e-70 (http://doc.bioperl.org/bioperl-run/lib/Bio/Tools/Run/TribeMCL.html#General). The genes within the \( cps \) loci that encoded proteins with the same homology group were assigned the same name, exceptions being those genes found in Region A. BLASTp (http://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins) searches were done by using a reference sequence database with default settings.

Annotation was performed by using the genome viewer Artemis (27). The \( cps \) nucleotide sequence locus for each serogroup was stored in BIGS\textsubscript{db} and linked with the corresponding isolate record.

MEGA5 (http://megasoftware.net/) was used to calculate overall \( p \)-distances and \( G+C \) content and to obtain the number of polymorphisms observed within each gene. The \( cps \) loci were compared by using the Artemis Comparison Tool (www.sanger.ac.uk/resources/software/act/).

**Results and Discussion**

**General Organization**

All of the serogroups examined had comparable \( cps \) loci, and regions occurred in the following order D-A-C-E-D′-B, revealing a conserved gene synteny and a replication of the genomic rearrangements (Figure). Genes in regions B–D, D′ and E were conserved, whereas genes in region A were diverse; this finding is consistent with the distinct biochemical composition found within each serogroup (Table 2).

The lowest GC content was in regions A and C, indicative of genes in these regions resulting from horizontal recombination (Table 2; Table 3, Appendix, wwwnc.cdc.gov/EID/article/19/4/11-1799-T3.htm); regions B, D, and E displayed GC contents of 50%, 52%, and 52%, respectively.
respectively, similar to those found in the Neisseria genomes (21,22,33). Bacterial genomic GC content varies considerably among species but remains uniform within a bacterial genome, such that genes acquired through horizontal genetic exchange will have a GC content different from the overall GC content found within the genome.

Nomenclature

Reports identifying genes and proteins involved in N. meningitidis capsule biosynthesis, combined with the increasing availability of bacterial genomes, have necessitated a more unified approach to the nomenclature of genes within N. meningitidis cps locus (Table 3, Appendix). The capsule locus from each serogroup has been uploaded to the BIGSdb platform, enabling sequences from each gene within the N. meningitidis cps locus to be indexed and multiple cps loci to be typed. However, during the process, it was found that the nomenclature of some genes within this locus posed a problem. For instance, genes within region B were thought to encode lipidation enzymes and, thus, have been known as lipA and lipB. However, within the annotated genomes from N. meningitidis FAM18, MC58, Z2491, and 05342 and from N. lactamica 020–06, two other distinct lipA and lipB genes have been described encoding a lipoic acid synthetase and a lipoate-protein ligase protein, respectively. Furthermore, there are, in some instances, multiple names for the same gene.

Continued surveillance of meningococcal disease combined with the use of multiple names for some genes made it apparent that a consistent approach to the nomenclature of capsule genes was needed. To meet that need, we propose a comprehensive description of all N. meningitidis serogroups and a nomenclature for the cps locus, which was presented at the 2012 XVIIIth International Pathogenic Neisseria Conference (Table 3, Appendix).

We propose that the capsule biosynthesis genes within region A should be termed cs (for capsule synthesis) followed by a letter representing the serogroup and by a capital letter defining each gene according to the Demerec system of genetic nomenclature (34). For example, serogroup A capsule biosynthesis genes would be termed cssA–D (cs for capsule synthesis and a for serogroup A) (Table 3, Appendix). Under this proposal, the sialic acid capsule biosynthesis genes would be termed cssA–C (cs for capsule synthesis and s for sialic acid capsule).

Serogroups B, C, W, and Y are commonly associated with invasive meningococcal disease, and rapid diagnosis of the serogroup is key in monitoring the epidemiology of the disease and in developing prevention strategies. Thus, it is necessary to have nomenclature that identifies the serogroup simply and quickly. The polysialyltransferase genes belonging to serogroups B and C share >70% sequence identity and should be termed csb and csc, respectively. The equivalent gene in serogroups W and Y is also a sialyltransferase, but it also has a glycosyltransferase function and is distinct from the serogroups B and C genes. These distinctions should be reflected in the nomenclature: we propose that the gene for serogroups W and Y be termed csy and csc, respectively. The O-acetyltransferases should be termed cssE for serogroup C and cssF for serogroups W and Y; the nomenclature for ctrG, which has been shown to have a role in surface
translocation of sialic capsules, should be retained (30). The remaining serogroups would also follow this scheme (Table 3, Appendix). Genes within region C that encode the capsule transport genes have been termed ctrA–D (after capsule transport), and this nomenclature should be preserved (Table 3, Appendix) (32). Genes within region B that are involved in capsule translocation should be designated ctrE and ctrF (31). Capsule genes are accessible through the PubMLST database (http://pubmlst.org/neisseria). In addition to the common gene name, these loci are allocated a value-free nomenclature consistent with the FAM18 genome annotation but using the prefix NEIS instead of NMC (Table 3, Appendix).

The W-135 and 29E serogroup designations originated at the Walter Reed Army Institute of Research as a result of a paper published by Evans et al. (35). We propose to rename these W and E because the numbers are historic and supply no useful information.

**Region A Capsule Biosynthesis Genes**

**Serogroup A**

The serogroup A capsule is composed of repeating units of O-acetylated (α1→6)-linked N-acetyl-d-mannosamine-1-phosphate (13). The capsule biosynthesis region is 4,365 bp long and contains 4 genes: csaA–D (also known as mynA–D or sacA–D) (18, 28). The first gene, csaA, encodes the UDP-N-acetyl-d-glucosamine (UDP-GlcNAc) 2-epimerase, which converts UDP-GlcNAc into UDP-N-acetyl-d-mannosamine (UDP-ManNAc); csaB is the polymerase linking ManNAc-phosphate monomers together with csaC, encoding an O-acetyltransferase, which transfers acetyl groups to ManNAc (18, 28). The fourth gene, csaD, is predicted to be involved in either capsule transport or in cross-linking of the capsule to the meningococcal cell surface. The serogroup A cps did not contain insertion sequences, and genes within region A contained some of the lowest GC content values among all serogroups (Table 3, Appendix).

**Serogroups B, C, W, and Y**

Capsular polysaccharides from serogroups B, C, W, and Y are composed of sialic acid derivatives; serogroups B and C express (α2→8)- and (α2→9)-linked sialic acid homopolymers, and alternating sequences of d-galactose or d-glucose and sialic acid are expressed by serogroups W and Y (6, 7). Region A is 5,313 bp long in serogroup B, 6,690 bp long in serogroup C, and averages ≈7,581 bp in serogroups W and Y.

All 4 serogroups contain the conserved cssA–C genes for cytidine-5′-monophosphate-N-acetylmuraminduronic acid synthesis; other designations for cssA–C include siaA–C, synA–C, and neuA–C (17). These are followed by csb, csc, csw, and csy genes (siaD, siaDBCWY, or synD–G), which are the capsule polymerases and determine the functional and nucleotide specificity for the 4 serogroups.
Serogroups C, W, and Y contain O-acetyltransferase genes termed cssE (ootC) and cssF (ootWY) (29). The 2 sequenced serogroup C and serogroup Y isolates harbored functional cssE/cssF genes with an IS1301 transposase adjacent to cssF in serogroup Y. Serogroup W cssF was interrupted by the insertion sequence IS1301. Another gene, ctrG, found in all 4 serogroups, encoded a protein essential in enabling the correct expression of sialic acid polysaccharides (30). The orientation of this gene differed between serogroups such that it was in the same direction as the css genes in serogroups B and C and the opposite in serogroups W and Y (Figure).

Serogroups W and Y contained an additional gene in region A, galU, encoding UTP-glucose-1-phosphate uridyltransferase, which was also found in region A of serogroup H. BLASTp searches of the derived amino acid sequence of GalU from serogroups W, Y, and H revealed that this protein shared an average of 49% sequence identity with GalU proteins from Streptococcus pneumoniae isolates. GalU has an essential function in capsule formation among S. pneumoniae isolates, catalyzing the reversible formation of UDP-Glc (uridine diphosphate glucose) and inorganic pyrophosphate from UTP (uridine 3-phosphate) and glucose 1-phosphate, and a role in virulence has been recognized for GalU in several bacterial species (36). Additional galU genes were found adjacent to the gene argH encoding argininosuccinate lyase in the sequenced genomes from isolates α275, 29031, H-ASH/87, and H-ZANE/83, and although the genes were conserved, they were not identical to the galU genes found within the cps locus. The genomes belonging to N. meningitidis Z2491, MC58, FAM18, and 053442; N. lactamica 020–06; and N. gonorrhoeae FA1090 also contained galU genes adjacent to argH. These were highly conserved (p-distance = 0.015) but were more distantly related to those found within the cps locus (p-distance = 0.180), indicating a different origin for cps-associated galU genes.

The serogroup D isolate was found to contain serogroup C capsule biosynthesis genes (Figure and online Technical Appendix Figure 1, panel A, wwwnc.cdc.gov/EID/article/19/4/11-1799-Techapp1.pdf). The cps locus from the prototype serogroup D isolates deposited by Gordon and Murray in 1917 and Branham in 1928 also contained serogroup C-specific capsule genes; however, neither isolate gave precipitins with antiserum to serogroup C, suggesting that these isolates were unencapsulated (C. Frasch. pers. comm.). The presence of internal stop codons in the ctrA and ctrE genes is consistent with this and further confirms that serogroup D does not exist.

**Serogroup E**

The serogroup E capsule consists of alternating α-galactosamine and 2-keto-3-deoxyoctulosonate (KDO) residues (9). Region A is 9,613 bp long, including IS1301, and contains 7 genes, cseA–G (formerly 29eA–H); gene cseE encodes a putative 3-deoxy-phospho-octulosonate synthase, cseF encodes a putative 3-deoxy-α-manno-octulosonate cytidylyltransferase, and cseG encodes a α-arabinose 5-phosphate isomerase protein. The deduced amino
Serogroups H and Z

The biochemical structures of serogroups H and Z contain monosaccharide glycerol-3-phosphate repeat units and share similarities with teichoic acid polymers. Region A varied from 6,182 bp in serogroup H to 7,198 bp in serogroup Z, with the latter containing the insertion element IS1301. Four genes were present in both serogroups (Figure and online Technical Appendix Figure 3, panel A). BLASTp searches of the deduced amino acid sequences of the first 2 genes, termed cshA/cszA and cshB/cszB, shared 91% sequence identity with each other and also shared 76% and 63% sequence identity with the capsule biosynthesis genes cps2B and cps2C belonging to Actinobacillus pleuropneumoniae serovars 2, 3, 6, 7, 9, 11, and 13 (38). These genes are predicted to encode a glycerol-3-phosphate cytidylyltransferase, and a hypothetical protein containing a LicD domain for a role in phosphorylcholine incorporation into teichoic acid polymers has been suggested (38). The third genes in region A, termed cshC and cszC, each shared 65% sequence identity with a putative teichoic thiose genes (cps2D and cps9D, respectively) belonging to A. pleuropneumoniae serovars 2, 7, 9, and 13. Last, cshD and cszD were 75% homologous to A. pleuropneumoniae capsule synthesis genes (cps7E).

Serogroups I and K

Different structural compositions have been described for these serogroups, with serogroup I consisting of O-acetylated alternating N-acetyl-guluronic acid and N-acetylmannosaminuronic acid units and serogroup K composed of O-acetylated disaccharide repeat units containing N-acetylmannosaminuronic acid (10,11). Both serogroups had almost identical capsule biosynthesis genes, csiA–E or cskA–E; region A was ≈9,026 bp long (Figure and online Technical Appendix Figure 3, panel B). MLST analysis showed that serogroup I and K isolates investigated in this study did not possess the same sequence types or PorA and FetA variable regions (Table 1). Antiserum is not commercially available to verify these serogroups; however, the immunochemical difference between serogroup I and K polysaccharides may reflect differences in the original methods used to purify them (10,11). In addition, 2 nonsynonymous substitutions were observed in the capsule biosynthesis genes csi/C/cskC and csiD/cskD, for which the deduced amino acid sequence encoded putative glycosyltransferases belonging to families 1 and 2, respectively. The capsule polymerases belonging to serogroups W and Y (csw and csy) are closely related; a single amino acid substitution in the N-terminal glycosyltransferase domain of the capsule polymerase produces the glucose or galactose substrate specificity for the enzyme (39) (online Technical Appendix). It is therefore possible that the nonsynonymous changes observed in the glycosyltransferases, csi/C/cskC and csiD/cskD, may produce the serogroup I and K capsules. Further investigation of these serogroups is required.

The derived amino acid sequences from csiA, B, and E/cskA, B, and E shared >70% sequence identity with capsule biosynthesis proteins belonging to Mannheimia haemolytica serotype A1. The capsule of M. haemolytica serotype
A1 is composed of a disaccharide repeat of N-acetylmannosaminuronic acid linked with N-acetylmannosamine.

Regions B and C
The genes ctrE and ctrF (formerly lipA and lipB) are required for surface expression of a properly anchored capsule polymer. The ABC (ATP binding cassette) transport system is characterized by the hydrophobic outer and inner membrane proteins CtrA and CtrB, respectively; the integral inner membrane–associated protein, CtrC; and the ATP binding protein, CtrD, and homologous genes are found among other group II capsule-expressing bacteria (40) (see online Technical Appendix).

Conclusions
We compared nucleotide sequence data from complete cps loci from all described N. meningitidis serogroups, revealing that the genetic organization is similar for all loci and that region A contains the capsule-specific biosynthesis genes. Distinct capsule operons corresponding to serogroups A, B, C, E, H, I, K, L, W, X, Y, and Z have been described herein, with the serogroup D capsule described as being an uncleavable serogroup C variant. Nucleotide sequence data for each capsule locus have been deposited in BIGSdb, with genes from each region defined and organized into schemes (26). However, it became apparent that a consistent approach to the nomenclature of capsule genes was required. In 2012, the approach detailed in this paper was approved at the XVIIIth International Pathogenic Neisseria Conference in Würzburg, Germany.

Horizontal genetic transfer of cps genes in region A was evident in serogroups H, I, K, and Z isolates, indicating acquisition of genes from external sources, including the bacterial species A. pleuropneumoniae and M. haemolytica. Combined with the low GC content observed in regions A and C, these data are consistent with acquisition of capsular genetic material from other species. Serogroup determination has been unresolved in some isolates. These isolates may contain capsule genes that are not expressed and will not be detectable by using conventional seroglutination techniques, or they may be serogroup E, H, I, K, L, X, or Z, which are not routinely searched for and for which commercial antisera is not available. This study provides additional tools to detect all capsule loci and may ultimately permit determination of the distribution of all serogroups among N. meningitidis populations and detection of cni isolates.

Acknowledgments
We thank Paula Kriz, Dominique Caugant, and Robert Mulhall for their kind donation of isolates.

M.C.J.M. is a Wellcome Trust Senior Research Fellow.

Dr Harrison is a postdoctoral research assistant in the Department of Zoology, University of Oxford. Her research interests include the molecular epidemiology and population structure of Neisseria meningitidis.

References

Address for correspondence: Odile B. Harrison, University of Oxford, Department of Zoology, South Parks Rd, Oxford OX1 3PS, UK; email: odile.harrison@zoo.ox.ac.uk
Description and Nomenclature of *Neisseria meningitidis* Capsule Locus

Technical Appendix

A
Figure 1. Panel A depicts a comparison of *Neisseria meningitidis* capsule loci from serogroups C, C (previously defined as serogroup D), B, W, and Y. Panel B displays a comparison of serogroups W and Y. The red and blue bands represent forward and reverse matches, respectively, with the intensity of the color proportional to the percentage sequence identity such that red indicates a good match (typically ranging from 90% to 100% sequence identity) and white indicates a low-scoring match. The average G+C content (%) across the bases is displayed at the top. Analyses were done by using the Artemis comparison tool (27).
Figure 2. The figure compares *Neisseria meningitidis* serogroups A, X, E, L, and C. The red and blue bands represent forward and reverse matches respectively with the intensity of the color proportional to the percentage sequence identity such that red indicates a good match (typically ranging from 90% to 100% sequence identity) and white indicates a low-scoring match. The average G+C content (%) across the bases is displayed at the top. Analyses were done by using the Artemis comparison tool (27).
Figure 3. Panel A depicts *Neisseria meningitidis* serogroups H and Z. Panel B depicts serogroups I and K. The red and blue bands represent forward and reverse matches, respectively, with the intensity of the color proportional to the percentage sequence identity such that red indicates a good match (typically ranging from 90% to 100% sequence identity) and white indicates a low-scoring match. The average G+C content (%) across the bases is displayed at the top. Analyses were done by using the Artemis comparison tool (27).