Capsulation in Distantly Related Strains of *Haemophilus influenzae* Type b: Genetic Drift and Gene Transfer at the Capsulation Locus

J. Simon Kroll* and E. Richard Moxon

Institute of Molecular Medicine and University Department of Paediatrics, John Radcliffe Hospital, Oxford OX3 9DU, United Kingdom

Received 24 July 1989/Accepted 30 November 1989

Among natural populations of capsulate *Haemophilus influenzae*, clones of strains with type b capsular polysaccharide are found in each of two widely separated phylogenetic divisions. The chromosomal capsulation locus found in strains from either division has a three-segment organization, with serotype-specific DNA nested between elements common to all serotypes, but pairwise comparison of the segments between the divisions suggests that they have distinct phylogenetic histories. Genes clustered in one of the non-serotype-specific segments appear to have diverged from an ancestral element, reflected in 12% nucleotide sequence divergence in one homologous pair. In contrast, genes conferring the capacity to produce type-specific polysaccharide exhibit no such divergence, and we speculate that these have been subject more recently to horizontal transfer within the bacterial population. Clinically important capsulate gram-negative bacteria share a common organization of their capsulation loci, arguing convergence on a successful arrangement of genes. In *H. influenzae* this appears to have allowed the occasional exchange of serotype-specific capsulation genes between strains, a event of potential clinical importance in this major bacterial pathogen.

*Haemophilus influenzae* is the commonest bacterial cause of meningitis and other life-threatening infections in young children in the United Kingdom and the United States (23), with a lifetime risk of invasive infection of about 1 in 400. More than 95% of these infections are caused by serotype b strains (25), and the type-specific capsular polysaccharide is a major determinant of virulence (18).

A recent study of the population diversity of capsule *H. influenzae* has shown that there are highly genetically divergent strains sharing serotype b (20, 21). Based on the electrophoretic mobilities of 17 metabolic enzymes, strains have been ascribed a multilocus electrotype, and cluster analysis of these electrotypes has revealed two primary phylogenetic divisions, I and II, separated by a genetic distance of 0.66, representing differences between electrotypes, on the average, at 10 of the 17 loci assayed (Fig. 1). Although more than 98% of type b strains have electrotypes in division I, there is a small cluster, recovered over many years from widely separated geographical sources, with electrotypes in division II. Notwithstanding the substantial genetic distance separating these strains from the majority of type b isolates, chemical and spectroscopic analyses of the respective capsular polysaccharides have shown them to be identical in composition and linkages (3, 4; W. Egan, personal communication). Here we report the characterization of the chromosomal capsulation locus (*cap*) in a type b strain from phylogenetic division II. Our results indicate that *cap* loci in type b strains from the two divisions are homologous but that different segments of *cap* have different phylogenetic histories. The results also suggest that the horizontal transfer of type-specific capsulation genes has played a part in the generation of diversity in this pathogen.

**MATERIALS AND METHODS**

Bacterial strains. (i) *H. influenzae*. RM926 is a type b isolate from a case of meningitis obtained from the National Collection of Type Cultures, Central Public Health Laboratory, United Kingdom (NCTC 8468). RM107 is a type a isolate from a case of respiratory infection obtained from the American Type Culture Collection, Rockville, Md. (ATCC 9006). The type b strains RM153 (Eagan), RM8055, and RM135 have been described previously (13, 14); capsule-deficient mutants of RM135, and the transformant RM684, are described in this work.

(ii) *Escherichia coli*. E. coli P2-392, a P2 lysogen obtained from Northumbria Biologicals Ltd., United Kingdom, was used to propagate bacteriophage lambda. Constructions in plasmid pUC13 were propagated in *E. coli* DH5α. Techniques used for the storage of strains, growth in liquid and on solid media, assessment of *H. influenzae* capsulation, and quantitation of type b capsular polysaccharide were as previously described (14).

Construction of the lambda gene bank of RM926 and subcloning of *Haemophilus* DNA into the plasmid pUC13. Total cellular DNA was prepared from RM926 as previously described (11) and partially digested with restriction endonuclease *Sau*3AI. The resulting fragments were size fractionated by centrifugation through a sucrose density gradient, and the fractions containing 12- to 25-kilobase (kb) fragments were selected. These were ligated to *Bam*HI-digested arms of bacteriophage lambda EMBL4 (7), and the resulting concatamers were packaged to form infectious particles by using a kit purchased from Amersham International, United Kingdom. EcoR1 fragments from lambda clones of interest were subcloned by ligation into the polylinker cloning site in the plasmid vector pUC13. pJSK32, pJSK40, pJSK53, and pJSK56 are described in this work. Recombinants were distinguished from nonrecombinants by their failure to confer on DH5α the ability to generate a blue colony in the presence of 5-bromo-4-chloro-3-indolyl-β-D-galactopyranoside and isopropyl-β-D-thiogalactopyranoside.

DNA hybridization probes and analysis of cloned and genomic DNA. pUO38, containing *cap* DNA from an *H. influenzae* type b strain cloned in pBR322, has been described previously (14). Other DNA probes used in this work are shown in Fig. 2.
Methods used for deriving restriction endonuclease cleavage maps of cloned DNA and chromosomal loci, for bacterial transformation, and for Sanger dideoxy sequencing have been described previously (9, 13, 17, 19, 22).

RESULTS

Cloning and physical mapping of cap loci. We have delineated cap in the division I type b strain RM8055 (14, 15) as a structure with segmental organization, serotype-specific DNA lying nested between elements common to all serotypes (Fig. 2A). In this form, or more usually as a duplication (11), this physical map of cap is representative of more than 100 division I strains that we have examined, but comparison of EcoRI-digested DNA from RM8055 and the division II strain RM926 in a Southern blot probed with pUO38 showed obvious differences (Fig. 2B). Although the strains both showed strongly hybridizing fragments of 2.1 and 2.7 kb, another four fragments in each strain did not match for size or strength of hybridization. Southern blots of EcoRI-digested DNA from seven other division II strains showed the same pattern as that seen for RM926, except that in two cases the faintly hybridizing high-molecular-weight fragment was apparently even larger, >30 kb (data not shown).

A gene bank of RM926 DNA, λ 926, was made in lambda EMBL4, and clones hybridizing to pUO38 were selected for further analysis by plaque hybridization. Physical mapping of the DNA inserts identified overlaps and so allowed construction of a restriction endonuclease cleavage map of the corresponding chromosomal locus (Fig. 2C). This locus contains all of the fragments identified in the Southern blot of EcoRI-digested RM926 DNA except for the faintly hybridizing 20-kb segment. In studies to be reported elsewhere (J. S. Kroll, B. Loynds, S. R. M. Dobson, and E. R. Moxon, manuscript in preparation), we have found this to lie elsewhere in the chromosome and, though cross-hybridizing to pUO38, not to form part of the cap locus.

Fine-structure mapping was used to compare corresponding parts of the division I and division II cap loci in detail (Fig. 3). Over the 2.1-plus 2.7-kb EcoRI segment comprising the center of region 2 of cap in RM8055, 13 restriction sites have been mapped by using restriction endonucleases with hexanucleotide recognition sequences, and exactly the same pattern was found on mapping DNA cloned from the RM926 locus. In contrast, there was no obvious similarity between the map of cap region 1 in RM8055, containing nine sites within a 4-kb Scal-EcoRV fragment, and the corresponding part of the RM926 locus. Fewer sites have been mapped to cap region 3 in RM8055, but, whereas NcoI and ClaI sites could be aligned with corresponding sites in RM926, five further sites failed to match, suggesting a discordance of sequence between the divisions here as at the other end of the locus. This striking picture clearly differentiated DNA in the central region of cap from that flanking it.

Functional comparison of cap DNA in type b strains from the two phylogenetic divisions. The discordance of the endo-
nuclease cleavage maps of cap region 1 between phylogenetic divisions could reflect a fundamental difference in the loci or extensive divergence from a common ancestral origin. This was explored by DNA-DNA hybridization with cloned fragments from region 1 in the two strains (Fig. 4). pJKS32 and pJKS40, containing respectively the 5.5- and 4.1-kb EcoRI cap fragments cloned from λ 926.11, were digested with various combinations of restriction endonucleases to allow unambiguous identification of different small segments. Southern blots of these fragments were probed with labeled pSKH1 and pJSK4, respectively. Despite the map disparity, 12 of 13 DNA segments in this part of cap in RM926 hybridized strongly to DNA cloned from the corresponding region in the division I locus. The exception, a 0.56-kb ScaI-BglII fragment (fragment 9, Fig. 4), did not hybridize to any other part of the chromosome of a division I type b strain, indicating that the division II cap locus contains between 0.56 kb (ScaI-BglII segment) and 1.9 kb (BglII-EcoRV segment) of extra DNA in this region that is not essential for capsulation in division I strains. These data suggested that cap region 1 in the two divisions might differ both by sequence divergence and the presence, in the division II locus, of a DNA insertion. When the maps of the left-hand ends of cap were compared after removal of 1.1 kb (centered on the 0.56-kb ScaI-BglII segment) from the RM926 5.5-kb EcoRI fragment, SpII, EcoRI, and EcoRV sites could be visually aligned (Fig. 3). Nucleotide sequences from each cap locus immediately next to the EcoRI and EcoRV sites in question were found to be identical (data not shown), supporting the hypothesis that the loci indeed differed in part by the presence or absence of an insertion between these sites.

The possibility of functional equivalence of DNA from region 1 of the division I and division II type b cap loci was examined by the construction of chimeric transformants. RM132 is a mutant of the division I type b strain RM135 (14), which is capsule deficient as the result of a point mutation close to the EcoRV site in the 4.4-kb EcoRI segment. The mutation is located in DNA with type a and b serotype specificity, in a gene tentatively associated with the incorporation of ribitol into the capsule polysaccharide (10, 27). Transformation with pJKS32 yielded type b capsulate progeny like RM5684, in which the 4.4-kb EcoRI fragment had been replaced by a 5.5-kb piece (Fig. 5). However, restriction mapping of several independent transformants (data not presented) showed that the ends of the integrated DNA differed in each case, demonstrating that different combina-

![FIG. 3. Detailed physical maps of the cap loci of RM8055 and RM926. EcoRI sites are shown as long vertical lines dividing each locus into fragments of the sizes indicated in kilobases, and those defining the type b-specific region are used to align the maps. Shorter lines indicate the cleavage sites of other restriction endonucleases: B, BamHI; Bg, BglII; C, CiaI; V, EcoRV; M, MluI; Nc, NcoI; Nr, NrdI; Ns, NsiI; P, PstI; Pvu, PvuII; Sca, ScaI; Sp, SphI; Xb, XbaI; X, XhoI. Regions of overall map similarity are shown by orientation of the labels of restriction sites toward each other, whereas overall dissimilarity is shown by the opposite orientation. Dotted lines join restriction sites deduced to be equivalent, as described in the text.](http://jb.asm.org/)

![FIG. 4. Cross-hybridization of DNA in region 1. The linear map shows the contiguous EcoRI (E) fragments of 4.1 and 5.5 kb from the RM926 cap locus, divided into numbered segments by restriction endonucleases labeled as in Fig. 3. The upper panels show agarose gels containing the electrophoretically separated products of multiple endonuclease digests of the EcoRI fragments cloned in pUC13, stained with ethidium bromide, and visualized with UV light. Combinations of enzymes were used for the digests in each track to allow unambiguous identification of one or more of the numbered segments, as indicated above the figures. Where two or three fragments are seen, they are identified by their numerical labels in the same order vertically above the track. A dash above a lane signifies a fragment containing vector DNA as well as the insert. The size ranges of the fragments are shown in kilobases beside each gel. The lower panels show autoradiographs of Southern blots of the gels probed with [α-32P]dCTP-labeled pJSK4 (segments 1 through 4) and pSKH1 (segments 5 through 13). Segment 9 (white arrow) is the only fragment not hybridizing to the probe (black arrow).](http://jb.asm.org/)
tions of division I and division II gene products were capable of subserving the functions of region 1. The identity of DNA in region 2 of the division I and division II b strains was tested similarly. RM2103 (kindly provided by S. Ely) is an engineered mutant of RM135 made capsule deficient as the result of a 0.9-kb deletion in the 2.1-kb EcoRI segment of region 2. The capsulate phenotype was restored by transformation of this strain with the 2.1-kb EcoRI fragment cloned from RM926, with reconstitution of an intact 2.1-kb segment.

The relationship of DNA immediately flanking cap in the two divisions was assessed by probe hybridization. The 1.45- and 0.7-kb EcoRI fragments lying to the left of cap in RM926 (cloned from λ 926.11 in pJSK53 and pJSK56, respectively) failed to hybridize to DNA in the 6.8- or 10.7-kb EcoRI fragments adjacent to cap in division I strains, but located EcoRI fragments (again of 1.45 and 0.7 kb) elsewhere in the division I chromosome (data not presented). DNA cloned from the right-hand ends of the two cap loci similarly failed to cross-hybridize. This suggests that the chromosomal positions of cap genes in the division I and II type b strains are different.

Sequence analysis of a gene in region 1 of cap in a type b strain from each phylogenetic division. By using DNA from strain RM153 (another division I type b strain) the 2.1-kb PvuII segment lying (in RM8055) within the 6.8-kb EcoRI fragment was cloned into the bacteriophage M13, and the nucleotide sequences of both strands were determined (13). The NcoI-EcoRI segment of the 4.1-kb EcoRI fragment of cap in RM926 (Fig. 3) was subcloned into pUC13, and part of its nucleotide sequence was similarly determined. All endonuclease restriction sites were crossed in the process of sequencing. Comparison of the sequences showed each to contain an open reading frame of 795 nucleotides, the gene in the RM153 DNA being that previously identified immediately 5' to bexA (13). The sequences of these genes (bexB) are shown aligned in Fig. 6. There are 95 of 795 nucleotide mismatches between the genes, indicating 88% sequence identity. However, 76 of the 95 mismatches are translationally silent, and the deduced peptide sequences differ only 14 of 265 positions (95% identity). The amino acid substitutions that are found are functionally conservative (12), and analysis of the amino-acid sequences by computer for hydrophobicity with the program of Kyte and Doolittle (16) with a 15-residue window shows each to have an identical profile, characteristic of an integral membrane protein.

DISCUSSION

The population of encapsulated *H. influenzae* shows a genetic diversity comparable to that of *E. coli* (24) or *Salmonella typhimurium* (1), as reflected in the genetic distance of 0.66 that separates strains into two phylogenetic divisions (20) (Fig. 1). Despite this, strains of the same serotype are found segregating to both divisions. The production of the same capsular polysaccharide even by completely unrelated organisms is well recognized; *E. coli* K1 and *Neisseria meningitidis* group B provide examples. In these organisms no cross-hybridization of capsulation genes has been detected (5), although the question remains open whether the loci have evolved independently or the genes involved have diverged extensively from a common ancestral sequence. In the study reported here, however, we have shown that in the case of highly disparate *Haemophilus* strains, the evidence points to a combination of genetic drift and the horizontal transfer of cap genes as the basis for serotype identity.

The overall organization of cap in RM926, and in all seven other division II type b strains so far analyzed, is similar to that in a single-copy division I strain such as RM8055 (14). In RM926 and RM8055, the cap loci each contain about 18 kb of DNA, although cap in RM926 carries an extra 1.1 kb of unknown function. However, the chromosomal location of cap in the two strains appears to differ; probes cloned from DNA flanking cap in the division II strain hybridize away from cap in the division I chromosome.

In division I strains cap contains three regions, with a 'cassette' of serotype-specific DNA (region 2) sandwiched between DNA common to all serotypes (regions 1 and 3) (6, 15) (Fig. 2). The successful complementation of engineered mutations in regions 1 and 2 of division I strains with genes from equivalent parts of the RM926 cap locus suggests the same functional organization in division II strains. This is supported by the observation that the cap locus in the type a strain RM107 (Fig. 1) appears very similar to that in RM926 but bears type a-specific DNA in place of the central type b-specific region (A. Dhir, Ph.D. thesis, Oxford University, Oxford, United Kingdom, 1989). However, pairwise comparison of the separate regions of the type b cap locus between divisions suggests that different segments have different phylogenetic histories.

Much of the difference between the maps of cap region 1 in the two divisions can be explained by assuming that there has been divergence from an ancestral sequence. There is strong cross-hybridization of equivalent fragments of DNA between the divisions, and sequenced genes are clearly homologous, although showing 12% nucleotide sequence disparity. (This disparity is not found in strains from the same phylogenetic division; in separate experiments not reported here, stretches of homologous sequence from different strains in division I were compared, and sequence differences were found to be very uncommon.) A divergence from sequence identity of this magnitude is comparable to that recorded between homologous genes in *E. coli* and *S. typhimurium* (J. M. Musser, personal communication). Restriction mapping within region 1 suggests that the whole region departs from identity to about the same extent between the divisions. In the 4-kb Scal-EcoRV segment of region 1 in RM8055, the nine mapped restriction sites represent a sampling of 54 nucleotides. In the corresponding part of cap from RM926, the situation is complicated by the addition of nonhomologous DNA. Assuming that a DNA insertion in the division II locus accounts for the disparity in
size of the corresponding regions, this leaves nine sites (excluding the Scal and BglII sites flanking the 0.56-kb nonhomologous segment) in putatively homologous DNA, of which three can now be aligned with sites in the division 1 locus (Fig. 3). If the most parsimonious assumption is made, that the creation or loss of a particular hexanucleotide recognition sequence in each case has occurred as the result of a single nucleotide change, this means that of a total of 90 nucleotides \((9 \times 6) + (9 - 3) \times 6\) sampled over the merged span, a maximum of 78 nucleotides \((3 \times 6) + (12 \times 3)\) are unchanged, representing about 13% sequence disparity, as found by sequencing. The situation is completely different when region 2 is considered. No difference was found on mapping 13 sites in each locus, a difference from the situation obtaining in region 1 that would be expected to arise by chance with a probability of only approximately \(5 \times 10^{-9}\) (Fisher exact test). Nucleotide sequencing of a portion of region 2 has, as expected, that sequence divergence must be very rare (no mismatches in 230 base pairs compared between the divisions). Identity of region 2 DNA between the divisions is most simply explained by assuming that a cassette of type b-specific DNA has been transferred horizontally between strains, sufficiently recently in comparison to the creation of the cap locus regions 1 and 3 to have avoided significant sequence variation. This interpretation of our findings is strengthened by the observation of equivalent map differences at the cap loci of type a strains segregated to the two divisions. In exactly the same way as type b, such strains have the same (now type a-specific) DNA sandwiched between flanks with physical maps very similar to those found here for type b (15; Dhir, Ph.D. thesis). The alternative explanation for constancy of the region 2 map—that the entire nucleotide sequence is critical for its function and catastrophically sensitive to translationally silent mutations—is without precedent for so large an element and seems very unlikely.

Our observations on the genetic basis of capsulation can be generalized beyond Haemophilus species, for it is becoming clear that polysaccharide capsule production proceeds by the agency of a locus with this common theme in widely differing circumstances. All three of the major gram-negative pathogens causing meningitis in infants and children—H. influenzae type b, E. coli K1 (2, 26), and N. meningitidis group B (8)—have capsulation loci organized in the same segmental fashion, arguing that we are seeing genetic convergence on a highly successful arrangement. In Haemophilus species, one consequence of having a capsulation locus with a cassette structure appears to have been to allow occasional exchange of serotype-specific capsulation genes between strains, a event of possible importance in the evolution of pathogenicity.
ACKNOWLEDGMENTS

We gratefully acknowledge the expert technical help provided by Barbara Loynds and Lisa Brophy and helpful discussions with J. M. Musser and A. Dhir.

This work was funded in part by a program grant from the Medical Research Council, and generous support from the National Meningitis Trust. J.S.K. is a Lister Institute Research Fellow.

LITERATURE CITED


