

# Natural genetic exchange between *Haemophilus* and *Neisseria*: Intergeneric transfer of chromosomal genes between major human pathogens

J. SIMON KROLL\*, KATHRYN E. WILKS, JAYNE L. FARRANT, AND PAUL R. LANGFORD

Molecular Infectious Diseases Group, Imperial College School of Medicine, Department of Paediatrics, St Mary's Hospital, Norfolk Place, London W2 1PG, United Kingdom.

Edited by Maurice S. Fox, Massachusetts Institute of Technology, Cambridge, MA, and approved August 17, 1998 (received for review July 2, 1998)

**ABSTRACT** Members of the bacterial families *Haemophilus* and *Neisseria*, important human pathogens that commonly colonize the nasopharynx, are naturally competent for DNA uptake from their environment. In each genus this process is discriminant in favor of its own and against foreign DNA through sequence specificity of DNA receptors. The *Haemophilus* DNA uptake apparatus binds a 29-bp oligonucleotide domain containing a highly conserved 9-bp core sequence, whereas the neisserial apparatus binds a 10-bp motif. Each motif ("uptake sequence", US) is highly over-represented in the chromosome of the corresponding genus, particularly concentrated with core sequences in inverted pairs forming gene terminators. Two *Haemophilus* core USs were unexpectedly found forming the terminator of *sodC* in *Neisseria meningitidis* (meningococcus), and sequence analysis strongly suggests that this virulence gene, located next to *IS1106*, arose through horizontal transfer from *Haemophilus*. By using USs as search strings in a computer-based analysis of genome sequence, it was established that while USs of the "wrong" genus do not occur commonly in *Neisseria* or *Haemophilus*, where they do they are highly likely to flag domains of chromosomal DNA that have been transferred from *Haemophilus*. Three independent domains of *Haemophilus*-like DNA were found in the meningococcal chromosome, associated respectively with the virulence gene *sodC*, the *bio* gene cluster, and an unidentified *orf*. This report identifies intergenerically transferred DNA and its source in bacteria, and further identifies transformation with heterologous chromosomal DNA as a way of establishing potentially important chromosomal mosaicism in these pathogenic bacteria.

*Haemophilus* and *Neisseria* spp. are among the commonest colonists of the human upper respiratory tract, from which site either species may occasionally invade to cause life-threatening infection in previously healthy people, most usually children. At any time up to 80% of healthy individuals carry *Haemophilus influenzae*, mainly noncapsulate strains, while non-*H. influenzae* species such as *Haemophilus parainfluenzae* are virtually ubiquitous in the nasopharyngeal flora (1). Commensal neisserial species such as *Neisseria lactamica* are widespread, while 5–40% of healthy individuals of different ages, the proportion at its greatest in late adolescence, are colonized with *Neisseria meningitidis* (meningococcus).

Many *Haemophilus* and neisserial species, including *H. influenzae/parainfluenzae* and *N. meningitidis/gonorrhoeae*, are naturally competent for transformation, with a capacity to take up fragments of DNA from their environment and integrate them into the chromosome. The process is not

indiscriminate: DNA containing a particular uptake sequence (US) is taken up preferentially. In *Haemophilus* this US is a 29-bp domain 5'-aAAGTGCGGTnRWWWWnnnnnnRW-WWWW (HmUS), flagged by a 9-bp core sequence 5'-AAGTGCGGT (2–4). This sequence, in which lowercase "a" represents an adenine occurring in more than 50% of HmUSs, "n" is any base, "R" is a purine, and "W" is adenine or thymine, is highly over-represented in the genome [expected occurrence of the 9-bp core sequence  $\approx$  8 copies/1.8 Mb double-stranded genome, actual occurrence in the completely sequenced *H. influenzae* strain Rd genome, 1465 copies (4, 5)]. This great excess reflects the role of HmUS as the ligand for surface DNA receptors expressed by competent bacterial cells (6). Its presence is both necessary and sufficient for facilitated uptake of DNA by this route, regardless of the origin of the rest of the DNA with which it is associated (7, 8). Beyond the numerical excess of 9-bp core HmUS sequences in the genome, an analysis of their distribution has revealed that it is skewed in two respects, with significant over-representation in intergenic regions, and in +/- inverted repeat configuration, separated by 6–8 bases (4, 9). Such a chromosomal location gives the resulting stem-loop structure the potential to play a part in transcriptional regulation or termination (10, 11), whereas such a separation in an inverted repeat allows the context in which the core of the HmUS lies on each strand most closely to approximate to the extended consensus [at the best, only a single "S" (C-G) for "W" (T-A) mismatch at position 25 on each strand (4)]. A quite different, 10-bp, sequence, 5'-GCCGTCTGAA (nUS), appears to fulfill the same functions in *Neisseria* (12, 13) and is similarly highly over-represented in the inverted repeat configuration. Any more extended consensus US domain in *Neisseria* awaits definition.

Within *Neisseria* and *Haemophilus*, US-mediated recombination is a plausible mechanism underlying the exchange of chromosomal gene segments that is now recognized to have occurred widely between related species. In the colonizing setting organisms are exposed to DNA bearing the appropriate US from lysed related, commensal, strains. In such circumstances natural competence, facilitating uptake of this cognate DNA, can explain the observed mosaicism of chromosomal genes, with phenotypic consequences ranging from the comparatively trivial (alteration in electrophoretic mobility, though not function, of central metabolic enzymes) to the profound (modification of structure of immunodominant antigens such as the meningococcal major porin) (14–22). Until now there has been no suggestion that a US-mediated mechanism might lead to natural movement of DNA between unrelated organisms. Here we identify the donor and recipient

The publication costs of this article were defrayed in part by page charge payment. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. §1734 solely to indicate this fact.

© 1998 by The National Academy of Sciences 0027-8424/98/9512381-5\$2.00/0 PNAS is available online at [www.pnas.org](http://www.pnas.org).

This paper was submitted directly (Track II) to the *Proceedings* office. Abbreviation: US, uptake sequence.

Data deposition: The sequence reported in this paper has been deposited in the GenBank database (accession no. AJ001313).

\*To whom reprint requests should be addressed. e-mail: [s.kroll@ic.ac.uk](mailto:s.kroll@ic.ac.uk).

of natural intergeneric chromosomal gene transfer events that have occurred between *Haemophilus* and *Neisseria*, with potentially far-reaching implications for understanding the evolution of virulence of these major pathogens.

**RESULTS AND DISCUSSION**

In experiments to investigate the role of periplasmic [Cu,Zn]-cofactored superoxide dismutase (SodC) in the virulence of *N. meningitidis* (23), a 3,572-bp chromosomal *Cla*I fragment containing *sodC* and flanking genes was cloned as pJSK205 from the virulent serogroup B strain MC58 and sequenced (Fig. 1; GenBank accession no. AJ001313). Upstream of *sodC*, on the same strand, lies an ORF deduced, by the high similarity of its translated product to the protein in *Escherichia coli*, to encode MutY, an adenine glycosylase involved in DNA mismatch repair (24). Downstream of *sodC*, from position 2313, lie elements from the meningococcal insertion sequence IS1106, including a partial copy of the DR2 element and the gene encoding the transposase (25). On the *sodC* strand, at positions 2,270–2,301 between the TAA stop codon (at position 2,250) and the start of IS1106 sequence (Fig. 1), a 32-nt segment of DNA has the potential to form a hairpin loop structure followed by a run of thymidine residues, characteristic of a rho-independent terminator. Surprisingly, its sequence is that typically found downstream of genes in *H. influenzae*, containing inverted repeats of the core HmUS separated by 8 nt such that on each strand the sequence makes the best possible fit (28/29) to the HmUS consensus (Fig. 2).

*sodC* encodes a determinant of virulence widely distributed in Gram-negative commensals and pathogens of mucosal surfaces (26–29) including a comprehensive range of *Haemophilus* species (30, 31). The immediate provenance of the gene sequence under present consideration is, however, indubitably meningococcal, *sodC* with the associated HmUS being present both in another example of the same serogroup B strain sequenced independently [in the TIGR meningococcal genome sequencing project (accessed at <http://www.tigr.org>)] and in the serogroup A strain Z2491 [in data released by the Sanger Centre meningococcal genome sequencing project (accessed at [http://www.sanger.ac.uk/Projects/N\\_meningitidis](http://www.sanger.ac.uk/Projects/N_meningitidis))].

SodC protein sequences from *N. meningitidis*, *H. influenzae* (GenBank accession no. M84012), and *E. coli* (GenBank accession no. X97766) were compared all-against-all by using the interactive system for sequence matching and data analysis, Data Analysis and Retrieval With Indexed Nucleotide/peptide sequences (DARWIN) (32, 33), accessed via electronic mail (cbrg@inf.ethz.ch). The meningococcal and *H. influenzae* protein sequences showed striking similarity (87% identity). The DARWIN program PHYLOTREE was used to put the differences between peptide sequences on a quantitative graphical basis. In the resulting unrooted dendrograms (Fig. 3), pairwise relationships are displayed according to the principle that the

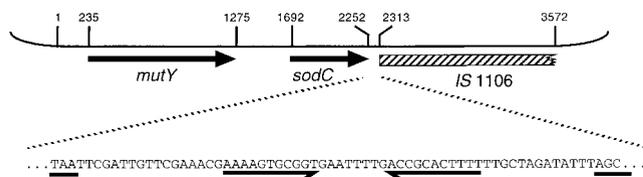


FIG. 1. Shown is the 3,572-bp *Cla*I DNA insert of pJSK205. *mutY* and *sodC* are indicated by arrows, and the hatched rectangle indicates the region with homology to IS1106. Coordinates mark start/stop codons and the beginning of IS1106 homology, and correspond to those in GenBank accession no. AJ001313. The expansion below shows the sequence in intergenic region 3' to *sodC*, with the inverted repeat highlighted. The *sodC* TAA stop codon and the first three bases of the region homologous to IS1106 are underlined.

2270 2301  
aAAGTGC GG TnRWWWWnnnnnnRWWWW  
\*\*\*\*\*  
5' - AAAAGTGC GG TGAATTTTGACCGCAC TTTT -3'  
3' - TTTTCACGCCACTTAAACTGGCGTGAAAAA -5'  
\*\*\*\*\*  
WWWWWRnnnnnnWWWWWRnTGGCGTGAAa

FIG. 2. Shown is the sequence of both strands of domain of pJSK205 between coordinates 2,270 and 2,301 (the region of the inverted repeat 3' to *sodC*). Consensus HmUS is aligned with each strand [a, adenine in >50% HmUSs; n, any nucleotide; R, purine (A/G); W, weak (A/T)]. Matches are indicated by \*.

length of the branch path joining two peptide sequences is in proportion to their degree of homology, calculated as the PAM distance [percentage accepted mutations) separating two sequences (32, 33)]. Short PAM distances thus signify

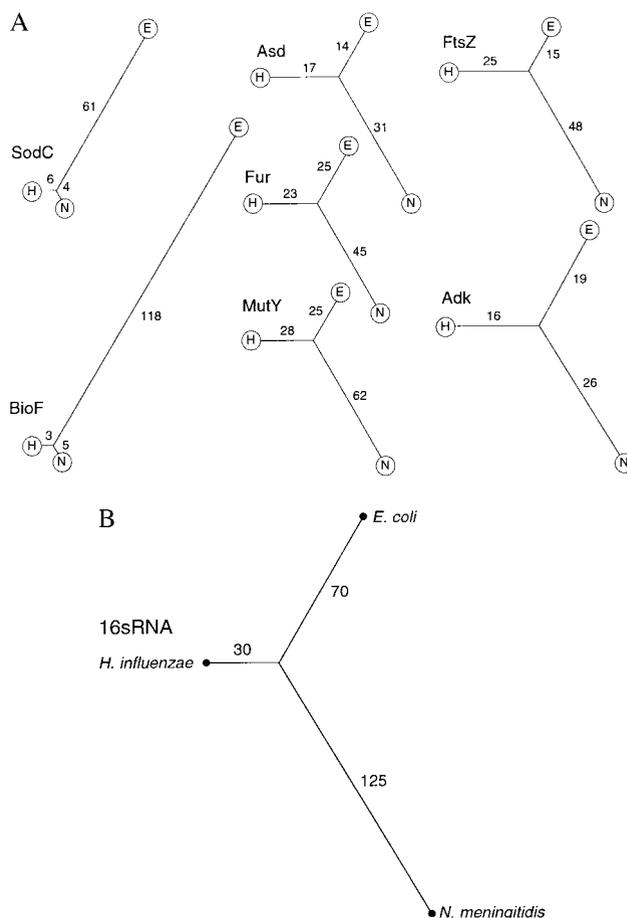


FIG. 3. (A) Phylograms based on protein sequence comparisons, computed with the DARWIN program PHYLOTREE of Gonnet *et al.* (32) accessed by e-mail (cbrg@inf.ethz.ch). In each case H represents *H. influenzae* (Hi), N = *N. meningitidis* (Nm), and E = *E. coli* (Ec). In the unrooted dendrograms, branch path length joining two peptide sequences (indicated) is in proportion to their degree of homology, calculated as the PAM (percentage accepted mutations) distance separating two sequences. Proteins (bacterial species, GenBank accession nos.) are represented as follows: Adk (Nm, L36469; Hi, X57315; Ec, X03038); Asd (Nm, Z14063; Hi, U32747; Ec, V00262); FtsZ (Nm, U43329; Hi, U32794; Ec, AE000119); Fur (Nm, L19777; Hi, U32704; Ec, X02589); MutY (Nm, this work; Hi, U32760; Ec, X52391); SodC (Nm, AJ001313; Hi, M84012; Ec, X97766); and BioF (Nm, Sanger sequence; Hi, U32830; Ec, AE000180). (B) Radial phylogram based on 16S rRNA sequences, constructed from output of the computer program SUBTREE of Maidak *et al.* (34) accessed at <http://rdp.life.uiuc.edu>. Branch lengths are in arbitrary units.

close similarity of sequence; long distances denote correspondingly divergent sequence. This treatment demonstrates that the meningococcal and *Haemophilus* SodC sequences are much more similar to, and by inference closely related to, each other than either is to the *E. coli* sequence. *adk*, *asd*, *ftsZ*, and *fur*, chosen as examples of genes encoding central metabolic functions that have been identified in the same three bacterial species, were subjected to the same comparative analysis. In contrast to the SodC result, there was no obvious skew in the quantitatively displayed relationships (Fig. 3). In each case the *Haemophilus* and *E. coli* protein sequences were somewhat more closely similar to each other than either was to meningococcal sequence, reflecting the result of 16S rRNA sequence comparison (ref. 34; Fig. 3). Taken together, these data indicate that the meningococcal *sodC* gene is anomalous, with characteristics suggesting it has arisen through horizontal transfer from *Haemophilus*.

Identification of the meningococcal homologue of *mutY* upstream of *sodC* provided an opportunity to assess the chromosomal extent of this unexpected sequence similarity between the meningococcal and *Haemophilus* loci. All-against-all comparison of the translated protein sequence with MutY from *E. coli* (GenBank accession no. X52391) and *H. influenzae* (GenBank accession no. U32760) gave the expected, unskewed, result, quite unlike the SodC relationship (Fig. 3). The GC-content of *mutY* (55.8%) was typical of a neisserial gene (meningococcal GC-content 51%) in contrast to 43.5% in the case of *sodC* (more *Haemophilus* like; GC-content, 38.2%), and toward the 3' end of the *mutY* coding sequence three copies of the nUS are found. *MutY* has sequence characteristics typical of a neisserial gene: it does not have the anomalous *Haemophilus*-like character of *sodC*.

*N. gonorrhoeae*, unlike *N. meningitidis*, appears on experimental grounds to lack *sodC* (23). The expanding Oklahoma database of contigs from the genome of *N. gonorrhoeae* strain FA1090 (accessed over the Internet at <http://www.genome.ou.edu>) was interrogated with meningococcal *sodC* and *mutY* sequences. As anticipated, there was no match to *sodC*, but a gene almost identical to meningococcal *mutY* (94% DNA identity; typical for neisserial genes present in both species) was present. Downstream of gonococcal *mutY*, a sequence identical to the meningococcal sequence in pJSK205 continued for a further 328 bp, leading up to a point of abrupt divergence corresponding to position 1,610 in the meningococcal sequence, 81 bp upstream of *sodC*. Beyond this point on the gonococcal contig, instead of *sodC*, there is a converging gene with extremely high similarity to the meningococcal gene *pspA* (N. J. Simpson and B. G. Spratt, GenBank accession no. AF030941, 1997).

These data together strongly suggest that *Neisseria* has acquired *sodC* through the horizontal transfer of at least 0.7 kb of DNA and that the likely source was an *Haemophilus* species, one of the common commensal organisms sharing the same mucosal ecological niche. The juxtaposition of *sodC* with *IS1106* sequence in the independent meningococcal genomes examined, and the absence of both in the gonococcal chromosome, suggests the possibility that the transposable element may have played a part in the acquisition of this DNA or, alternatively, that it facilitated a subsequent deletion amongst those apparently involved, alongside rearrangements, in the evolution of gonococcus from its neisserial ancestor.

Availability of sequence data from the near complete Sanger meningococcal sequencing project provided the chance to assess a minimum frequency for the 9-bp core HmUS in a meningococcal genome and look for further possible examples of DNA transfer. The complete *E. coli* genome sequence was selected as a control sequence for the same analysis. Searching both strands of all meningococcal contigs published up to March 6, 1998 (2.174 Mb of the  $\approx$ 2.3-Mb genome), revealed 17 different 9/9 matches, and a further 4 which at that point

contained a single ambiguous nucleotide. In 2 Mb of double-stranded DNA the expected result would be around 16 examples ( $2 \times 2,000,000/4^9$ ), so it is reasonable to conclude that even when the whole genome sequence is completed there will be no significant excess representation of this sequence. A similar search of the complete 4.6-Mb *E. coli* genome yielded 9 matches. On examining the chromosomal context of each of these core HmUSs in the meningococcal genome, another example of the full 29-bp HmUS was found. No such extended HmUS matches occur in the substantially larger *E. coli* genome. The likelihood of this sequence occurring even once by chance is approximately  $1$  in  $4 \times 10^9$  (1 in a 1,000 in 2 Mb of double-stranded sequence) and inspection of the new meningococcal locus identified in this way strongly suggests that this example has not arisen through chance, but rather also through an *Haemophilus*-to-*Neisseria* DNA transfer event.

The new HmUS lies in a locus tentatively identified by its sequence similarity to entries in GenBank as *bioF*, encoding 7-keto-8-aminopelargonic acid synthetase, an enzyme in the biotin synthesis pathway (35). The meningococcal locus has strikingly similar sequence (near 90% identity) to a corresponding locus in the *H. influenzae* strain Rd genome over a continuous 2.5-kb span encompassing two of the putative *Haemophilus bio* genes and a further unidentified ORF (*HI1552*) (Fig. 4). In this case, in contrast to the example of *sodC*, the same high degree of identity, complete with HmUS, was found to DNA in the Oklahoma database of gonococcal sequence. The putative neisserial *bioF* homologue has an unusually low GC-content for the genus (44.3%) whereas the whole 2.5-kb domain is flanked by stretches of extreme AT-richness (70% and 80%, respectively). Because it is scarcely credible that both of these sequencing projects might have been contaminated with the same *Haemophilus* clone, the conclusion must be that both neisserial genomes contain a putative *bio* locus anomalously similar to that in *Haemophilus*. An all-against-all comparison of translated *bioF* protein sequence from *N. meningitidis*, *H. influenzae* (GenBank accession no. U32830) and *E. coli* (GenBank accession no. AE000180) showed the same highly skewed pattern seen with SodC (Fig. 3). The conclusion seems inescapable that at this locus too the neisserial chromosome (in this case both *N. meningitidis* and *N. gonorrhoeae*) contains DNA bearing the distinctive signature of originating from *Haemophilus*. This finding supports the hypothesis that transfer may have occurred early in neisserial history, before the divergence of *N.*

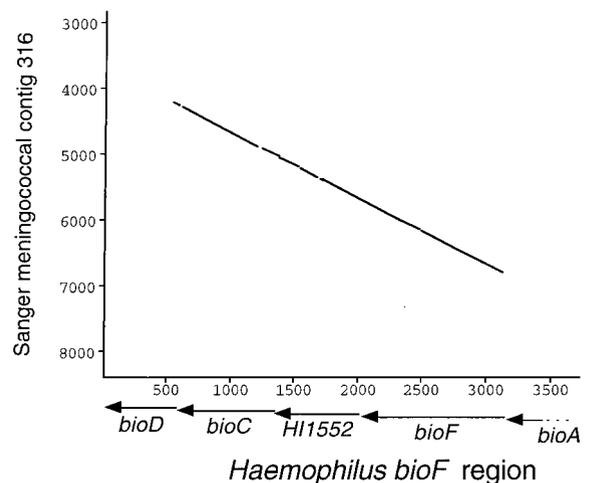


FIG. 4. Graphical comparison of DNA sequences at the *bio* locus of *H. influenzae* (x-axis) and putative *bio* locus of *N. meningitidis* from Sanger contig 316 (y-axis). Points plotted represent minimum 80% sequence identity over a 30-nt moving window. Positions of known genes shown in relation to the *Haemophilus* sequence.

*meningitidis* and *N. gonorrhoeae*. Interestingly, the short peptide domain encoded by the HmUS would not appear to be crucial for BioF activity, as it is not present in the corresponding *E. coli* sequence (Fig. 5). The provocative conclusion is that its presence in the *Haemophilus* and neisserial genomes reflects an importance beyond its coding capacity as a signal for DNA uptake or recombination.

Although extended HmUS sequences in the meningococcal genome appear to identify DNA that originated in *Haemophilus*, for *Haemophilus*-to-*Neisseria* DNA transfer to occur through an US-mediated process there needs to be an nUS sequence in the locus involved. The 10-bp nUS was accordingly used to search both strands of the *H. influenzae* Rd genome sequence. Four matches were found (the expected number to occur by chance in two strands of a 1.8-Mb genome) and the regions so identified were used to search the neisserial databases as before. Three regions found no significant match, but the fourth region, containing an nUS at position 1,120,756 in the *Haemophilus* genome, lay within the ORF of *HII055*, encoding an unidentified 514-aa protein, which was an extremely close match (93% DNA sequence identity) to a corresponding neisserial ORF, found as in the case of *bioF* in both meningococcus and gonococcus. The GC-content of *HII055* is 35.9%, which is as expected for an *Haemophilus* gene, but anomalously low for a neisserial gene. Again there was an extended domain of close similarity in the *Haemophilus* and neisserial sequences, in this case 86% DNA sequence identity over 3.4 kb, separated by about a kilobase from a further domain of high similarity. In the meningococcal chromosome, at one end of the domain and extending out of it into a region no longer closely similar to Rd, was the gene *lldA*. This encodes a protein verified experimentally to be L-lactate dehydrogenase and previously recognized to lie in the meningococcal chromosome next to DNA with similarity to *HII053*, which is two genes along from *HII055* in the *H. influenzae* chromosome (36). The presence of this 3.4-kb domain in each neisserial species, containing in the case of *N. meningitidis* part of a previously identified gene, again confirms that the match was not the trivial result of contamination of a neisserial library with a *Haemophilus* clone, but rather that the *Neisseria* and *Haemophilus* genomes truly contain very highly similar DNA at this location. The low GC-content of the *Haemophilus* gene argues an origin (or long-standing residence) in that genus, whereas the presence of the nUS indicates a basis for uptake of DNA from this locus by *Neisseria*. Intriguingly, the region in the neisserial sequence contains both two perfect (10/10) nUSs and one near-perfect (26/29) extended HmUS (gAAGcGCGGTCAAAAATCACAAGTTTTc; lowercase letters represents divergence from HmUS consensus); the (T-to-C) mutation in the core (underlined) HmUS sequence is not seen in the corresponding position in the *H. influenzae* chromosome.

## CONCLUSIONS

Although natural interspecies exchange of DNA is recognized to have occurred within both *Neisseria* and *Haemophilus*, intergeneric transfer has until now only been mooted as a

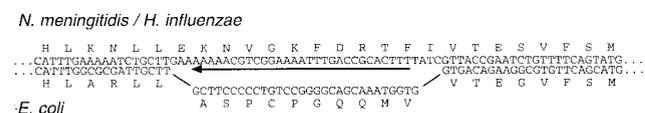


FIG. 5. DNA sequence alignment of *H. influenzae*/*N. meningitidis* *bioF* with *E. coli* *bioF* in the region of the HmUS (underlined; arrow indicates consensus sequence on noncoding strand). Coding strands and their translations are shown: *Haemophilus*/*Neisseria* (identical to each other) above, *E. coli* below. Separation of the sequences represents divergence where *Haemophilus*/*Neisseria* contains the HmUS.

hypothetical possibility, though one with obvious major implications for the evolution of virulence of either species. Only the most tenuous evidence has until now been found for such a process. Thompson *et al.* (37) found a highly degenerate version of the *Haemophilus* insertion element *IS1016* in meningococcus, eliciting speculation that there might have been exchange of DNA between the genera. Read and Farley (38) reported that a copy of a 21-bp extragenic element present in multiple copies in the *Haemophilus* chromosome was also found in relation to a meningococcal gene, and commented on the possibility that this might represent horizontal gene exchange between these distantly related bacteria. Nizet *et al.* (39) have described an unusually virulent noncapsulate clinical isolate of *H. influenzae* that has acquired cryptic characteristics from an unknown source by transformation. Martin *et al.* (40) have reported the presence in *Haemophilus* of a pathogenicity island of tryptophanase genes with enterobacteriaceal sequence characteristics flanked by HmUSs and have suggested that this may have arisen by horizontal transfer from an unidentified donor. Other pathogenicity islands are widely described in other bacterial species (41), though the origin of the DNA in question is generally obscure. In the present work, however, the donor and recipient species of natural intergeneric chromosomal gene transfer events have been identified. Following an appreciation of the anomalous, *Haemophilus*-like nature of meningococcal *sodC*, a prospective search for species-specific DNA uptake motifs in the “wrong” genus (HmUS in *Neisseria*, and nUS in *Haemophilus*) has successfully identified three unrelated chromosomal loci at which transfer of between 0.7 and 3.4 kb DNA appears to have occurred, anomalies in GC-content in the respective chromosomes strongly suggesting that in these cases the events were in the *Haemophilus*-to-*Neisseria* direction. The search strategy used, seeking perfect consensus USs, is a highly parsimonious one, necessarily adopted because of limitations in our computing power. The pressure to maintain the consensus US in either species relates to its species-specific importance in DNA uptake and is relaxed outside this context, so that for example HmUSs might be expected to degrade more quickly in the neisserial genome than in *Haemophilus*. The 26/29 match to the HmUS sequence seen in relation to the neisserial homologue of *HII055* is an example of such a degraded motif. Many more examples are likely to emerge with the eventual completion of the meningococcal genome, when it will be interesting to establish the true extent of chromosomal mosaicism with sufficient computing power to search for US sequences with mismatches as well as to compare the whole genomes without a preliminary US screen.

The mere presence of HmUSs flagging more extensive *Haemophilus*-like DNA in *Neisseria* does not inform on the mechanism by which DNA transfer took place. The best circumstantial evidence supporting transformation comes from the example associated with *HII055*, where the presence of the nUS in the *Haemophilus* chromosome provides a molecular basis for facilitated uptake of this DNA into *Neisseria*.

The virulence of either *Neisseria* or *Haemophilus* may have been, or may in the future be, modified by US-mediated intergeneric exchange. That the three examples here appear to have been in the *Haemophilus*-to-*Neisseria* direction may simply reflect the ubiquity of *Haemophilus* species, and inferentially of *Haemophilus* DNA, in the meningococcal environment. The alarming possibility of the reverse process is suggested by the phenomenon of Brazilian purpuric fever.

Although both *N. meningitidis* and *H. influenzae* may invade the bloodstream from the site of colonization in the nasopharynx to cause life-threatening infection, the clinical features of the diseases they cause are very different. *N. meningitidis* is one of the most important causes all over the world of community-acquired septicemia and meningitis. The hallmarks of me-

ningococcal infection at its fulminant worst are the result of widespread endothelial damage, with capillary leak, hemorrhage, shock, multi-organ failure, and a mortality near 50%. In contrast *H. influenzae* only very rarely causes such a syndrome of septic shock, but capsule serotype b strains in particular ( $\approx 2\%$  of *Haemophilus* colonists) are an important cause of meningitis in children.

An outbreak of infection with many clinical features of meningococcal sepsis, with 70% mortality, occurred in children at several locations in the São Paulo province of Brazil just over a decade ago (42). Investigations established that the condition termed Brazilian purpuric fever, rather than being caused by meningococcus, was the result of septicaemia with a clone of *H. influenzae* biogroup *aegyptius* that had inexplicably acquired a highly virulent "meningococcal" phenotype. The organism had never before been known to cause anything more serious than conjunctivitis. The prospect of a major epidemic in a crowded developing world setting provoked considerable concern at the time of its description, although this has largely dissipated with the nonemergence of such a catastrophe. However the possibility that US-mediated transformation of commensal *Haemophilus* strains with neisserial genes conferred a clone with the phenotype of rapidly fatal invasive infection with septic shock must be recognized, and complacency would seem entirely inappropriate.

We are grateful to Dr Michael Coulthart (Laboratory Centre for Disease Control, Ottawa, Canada) for his help with DNA sequencing. We gratefully acknowledge the practice of sequencing centers who release data into the public domain in timely fashion, without whom this work would not have been possible. Meningococcal sequence data were produced by (i) the *Neisseria meningitidis* Sequencing Group at the Sanger Centre and can be obtained from [http://www.sanger.ac.uk/Projects/N\\_meningitidis](http://www.sanger.ac.uk/Projects/N_meningitidis), and (ii) the Institute for Genomic Research ([ftp://ftp.tigr.org/pub/data/n\\_meningitidis](ftp://ftp.tigr.org/pub/data/n_meningitidis)). Gonococcal sequence data came from the Gonococcal Genome Sequencing Project (supported by U.S. Public Health Service/National Institutes of Health Grant AI38399) at <ftp://ftp.genome.ou.edu/pub/gono>. This work is supported by the Meningitis Research Foundation.

- Kilian, M. (1976) *J. Gen. Microbiol.* **93**, 9–62.
- Danner, D. B., Deich, R. A., Sisco, K. L. & Smith, H. O. (1980) *Gene* **11**, 311–318.
- Goodgal, S. H. & Mitchell, M. (1984) *J. Bacteriol.* **157**, 785–788.
- Smith, H. O., Tomb, J. F., Dougherty, B. A., Fleischmann, R. D. & Venter, J. C. (1995) *Science* **269**, 538–540.
- Fleischmann, R. D., Adams, M. D., White, O., Clayton, R. A., Kirkness, E. F., Kerlavage, A. R., Bult, C. J., Tomb, J. F., Dougherty, B. A., Merrick, J. M., *et al.* (1995) *Science* **269**, 496–512.
- Deich, R. A. & Smith, H. O. (1980) *Mol. Gen. Genet.* **177**, 369–374.
- Sisco, K. L. & Smith, H. O. (1979) *Proc. Natl. Acad. Sci. USA* **76**, 972–976.
- Danner, D. B., Smith, H. O. & Narang, S. A. (1982) *Proc. Natl. Acad. Sci. USA* **79**, 2393–2397.
- Goodgal, S. H. & Mitchell, M. A. (1990) *J. Bacteriol.* **172**, 5924–5928.
- Tomb, J. F., El-Hajj, H. & Smith, H. O. (1991) *Gene* **104**, 1–10.
- Kroll, J. S., Loynds, B. M. & Langford, P. R. (1992) *Gene* **114**, 151–152.
- Mathis, L. S. & Scocca, J. J. (1982) *J. Gen. Microbiol.* **128**, 1159–1161.
- Elkins, C., Thomas, C. E., Seifert, H. S. & Sparling, P. F. (1991) *J. Bacteriol.* **173**, 3911–3913.
- Bowler, L. D., Zhang, Q. Y., Riou, J. Y. & Spratt, B. G. (1994) *J. Bacteriol.* **176**, 333–337.
- Feil, E., Carpenter, G. & Spratt, B. G. (1995) *Proc. Natl. Acad. Sci. USA* **92**, 10535–10539.
- Feil, E., Zhou, J., Maynard-Smith, J. & Spratt, B. G. (1996) *J. Mol. Evol.* **43**, 631–640.
- Kroll, J. S. & Moxon, E. R. (1990) *J. Bacteriol.* **172**, 1374–1379.
- Kroll, J. S., Moxon, E. R. & Loynds, B. M. (1994) *J. Infect. Dis.* **169**, 676–679.
- Lujan, R., Zhang, Q. Y., Saez-Nieto, J. A., Jones, D. M. & Spratt, B. G. (1991) *Antimicrob. Agents Chemother.* **35**, 300–304.
- Spratt, B. G., Bowler, L. D., Zhang, Q. Y., Zhou, J. & Smith, J. M. (1992) *J. Mol. Evol.* **34**, 115–125.
- Vazquez, J. A., Berron, S., O'Rourke, M., Carpenter, G., Feil, E., Smith, N. H. & Spratt, B. G. (1995) *Mol. Microbiol.* **15**, 1001–1007.
- Zhou, J., Bowler, L. D. & Spratt, B. G. (1997) *Mol. Microbiol.* **23**, 799–812.
- Wilks, K. E., Dunn, K. L. R., Farrant, J. L., Reddin, K. M., Gorrington, A. R., Langford, P. R. & Kroll, J. S. (1998) *Infect. Immun.* **66**, 213–217.
- Michaels, M. L., Pham, L., Nghiem, Y., Cruz, C. & Miller, J. H. (1990) *Nucleic Acids Res.* **18**, 3841–3845.
- Knight, A. I., Ni, H., Cartwright, K. A. V. & McFadden, J. J. (1992) *Mol. Microbiol.* **6**, 1565–1573.
- Kroll, J. S., Langford, P. R., Wilks, K. E. & Keil, A. D. (1995) *Microbiology* **141**, 2271–2279.
- Imlay, K. R. C. & Imlay, J. A. (1996) *J. Bacteriol.* **178**, 4807–4813.
- Langford, P. R., Loynds, B. M. & Kroll, J. S. (1996) *Infect. Immun.* **64**, 5035–5041.
- Farrant, J. L., Sansone, A., Canvin, J. R., Pallen, M. J., Langford, P. R., Wallis, T. S., Dougan, G. & Kroll, J. S. (1997) *Mol. Microbiol.* **25**, 785–796.
- Kroll, J. S., Langford, P. R. & Loynds, B. M. (1991) *J. Bacteriol.* **173**, 7449–7457.
- Langford, P. R., Loynds, B. M. & Kroll, J. S. (1992) *J. Gen. Microbiol.* **138**, 517–522.
- Gonnet, G. H., Cohen, M. A. & Benner, S. A. (1992) *Science* **256**, 1443–1445.
- Gonnet, G. H. (1994) in *Computational Methods in Genome Research*, ed. Suhai, S. (Plenum, New York), pp. 153–161.
- Maidak, B. L., Olsen, G. J., Larsen, N., Overbeek, R., McCaughey, M. J. & Woese, C. R. (1997) *Nucleic Acids Res.* **25**, 109–111.
- Otsuka, A. J., Buoncristiani, M. R., Howard, P. K., Flamm, J. & Johnson, O. (1988) *J. Biol. Chem.* **263**, 19577–19585.
- Erwin, A. L. & Gotschlich, E. C. (1996) *J. Bacteriol.* **178**, 4807–4813.
- Thompson, S. A., Wang, L. L. & Sparling, P. F. (1993) *Mol. Microbiol.* **9**, 85–96.
- Read, T. D. & Farley, M. M. (1997) *Mol. Microbiol.* **23**, 627–628.
- Nizet, V., Colina, K. F., Almquist, J. R., Rubens, C. E. & Smith, A. L. (1996) *J. Infect. Dis.* **173**, 180–186.
- Martin, K., Morlin, G., Smith, A., Nordyke, A., Eisenstark, A. & Golomb, M. (1998) *J. Bacteriol.* **180**, 107–118.
- Hacker, J., Blum-Oehler, G., Muhldorfer, I. & Tschape, H. (1997) *Mol. Microbiol.* **23**, 1089–1097.
- Brenner, D. J., Mayer, L. W., Carlone, G. M., Harrison, L. H., Bibb, W. F., Brandileone, M. C., Sottnek, F. O., Irino, K., Reeves, M. W., Swenson, J. M., *et al.* (1988) *J. Clin. Microbiol.* **26**, 1524–1534.