Population structure in the *Neisseria*, and the biological significance of fuzzy species

Jukka Corander¹,*, Thomas R. Connor², Clíona A. O’Dwyer³, J. Simon Kroll³ and William P. Hanage⁴

¹Department of Mathematics and Statistics, University of Helsinki, PO Box 68, 00014 Helsinki, Finland
²The Wellcome Trust Sanger Institute, Wellcome Trust Genome Campus, Hinxton, Cambridge CB10 1SA, UK
³Molecular Infectious Diseases Group, Section of Paediatrics, Department of Medicine, Imperial College London, St Mary’s Hospital Campus, Norfolk Place, London W2 1PG, UK
⁴Center for Communicable Disease Dynamics, Department of Epidemiology, Harvard School of Public Health, Boston, MA 02115, USA

Phenotypic and genetic variation in bacteria can take bewilderingly complex forms even within a single genus. One of the most intriguing examples of this is the genus *Neisseria*, which comprises both pathogens and commensals colonizing a variety of body sites and host species, and causing a range of disease. Complex relatedness among both named species and previously identified lineages of *Neisseria* makes it challenging to study their evolution. Using the largest publicly available collection of bacterial sequence data in combination with a population genetic analysis and experiment, we probe the contribution of inter-species recombination to neisserial population structure, and specifically whether it is more common in some strains than others.

We identify hybrid groups of strains containing sequences typical of more than one species. These groups of strains, typical of a fuzzy species, appear to have experienced elevated rates of inter-species recombination estimated by population genetic analysis and further supported by transformation experiments. In particular, strains of the pathogen *Neisseria meningitidis* in the fuzzy species boundary appear to follow a different lifestyle, which may have considerable biological implications concerning distribution of novel resistance elements and meningococcal vaccine development. Despite the strong evidence for negligible geographical barriers to gene flow within the population, exchange of genetic material still shows directionality among named species in a non-uniform manner.

**Keywords:** fuzzy species; recombination; *Neisseria*

1. INTRODUCTION

The genus *Neisseria* consists of Gram-negative bacteria that colonize the mucous membranes of mammals. Two species, *Neisseria gonorrhoeae* and *Neisseria meningitidis* (the meningococcus), are important pathogens in humans, while other neisserial species have rarely been associated with symptomatic illness. *Neisseria lactamica* is the neisserial species that most commonly colonizes the nasopharynx of infants and young children [1,2], and in this population, meningococcal carriage is low, rising to high levels only in adolescence and young adult life [3].

Another feature of the *Neisseria* is their high rate of homologous recombination [4,5]. This has been known for some time, and has the important clinical consequence that it can spread genes encoding antibiotic resistance among lineages and even species [6]. It is also a crucial contributor to the rapid diversification of surface antigens, allowing immune evasion [7,8]. The result of this high rate of recombination is that within species phylogenies constructed from unlinked housekeeping loci are no more similar than random, recombination having scrambled phylogenetic signal at all but the most closely related strains [9]. Homologous recombination occurs more efficiently between closely related donor and recipient sequences, but it can and does also occur between more distantly related taxa, including species boundaries. Owing to the co-colonization of meningococcus and *N. lactamica* at the same body site, opportunities occur frequently in nature for the occasional transfer of DNA between them. Interspecific recombination means that it is not uncommon to find a strain securely identified as one species, but which contains a locus or loci with sequence typical of another species [10]. The impact of this admixture on species formation and identification has

*Author for correspondence (jukka.corander@helsinki.fi).

attracted considerable interest in recent years, and it has been proposed that it might explain historical difficulties with the taxonomy of the Neisseria (among others). The term ‘fuzzy species’ was coined to describe a situation in which, due to inter-species recombination, the fringes of a given named species cluster in a notional sequence space are not discrete and contain mosaics of genetic material from more than one named species [10].

Such mosaics were identified in Neisseria using data collected for multi-locus sequence typing (MLST) [11], which types pathogens using the DNA sequence at multiple unlinked housekeeping loci. The scheme developed for the meningococcus can also be used for related species, and trees built from concatenates of these clearly showed the existence of meningococci and N. lactamica strains containing multiple loci characteristic of more than one named species [10].

The observation of relatively pervasive recombination between named species creates insecurity in current species concepts [12–14]. One response to this uncertainty is to focus on loci at which recombination is less frequently found, either because it is more rare or more difficult to detect. An approach using highly conserved ribosomal proteins, retrieved from multiple genomes of Neisseria and streptococci, has been developed [15]. While such strategies are likely to be of great value in the question of how to pragmatically use sequence data to define species, any limited set of loci can, by definition, tell us nothing about variation and recombination elsewhere in the genome.

Serogroup A meningococci are thought to undergo recombination at a lower rate than other members of the same species [16]. Such variation in recombination rate is a matter of interest because, as noted earlier, higher recombination rates are thought to be associated with a greater risk that the strains concerned acquire resistance or virulence determinants. Circumstantial evidence for such a link already exists for Streptococcus pneumoniae [17]. The extent to which meningococci vary in their recombination rate is poorly understood, but as noted may have important consequences for evolution. Strains that contain mosaics of sequence data typical of more than one species might be inferred to have a higher recombination rate, or to have performed so at some point in the past. The biological significance of these mosaic genotypes is not clear, and could be an epiphenomenon: in other words, the limited sample of genes studied is not predictive of the phenotype of the organism, and that these strains are simply rare examples of meningococci (or N. lactamica strains) that have undergone recombination at one of the loci used for MLST, and that this is of negligible importance for the study of the strain, the species, or the evolution of either. Even if there is no variation among strains in the rate of recombination, some will be expected to have more recombinant loci simply by chance. However, commensal neisserial strains have been suggested as live attenuated vehicles for vaccine strategies [18], and hence the degree and frequency of recombination between N. lactamica and the meningococcus is a matter of more than theoretical concern.

To fully assess the importance of mosaic strains requires both an experimental and computational approach. Although as noted earlier, in the meningococcus, the phylogeny of one locus is no better than chance at predicting that of another unlinked locus, clusters of related strains, and the degree of admixture (recombination) between them can be identified by a number of statistical approaches [19–21]. If any such clusters can be securely identified that show a greater frequency of admixture between species, they then can be examined to find whether this is associated with any other features, including difficulty in species definition. Furthermore, any such strains can be directly tested in the laboratory to estimate whether they are more permissive of transformation under experimental conditions.

In this paper, we use a combination of sequence data, statistical analysis and experiment to probe the contribution of recombination to neisserial population structure, and specifically apparent hybrid groups of strains containing sequence data typical of more than one species.

2. MATERIAL AND METHODS

2.1. Sources of sequence data

Among all human and animal pathogens and commensals that have been studied using a MLST scheme [11], the amount of publicly available data is most extensive for Neisseria. We used a complete set of the available sequence types (STs) for Neisseria from http://pubmlst.org/neisseria/ as of the 26 November 2010. The database contained 8619 STs with sequences available for the following seven genes: abeZ, adk, aroE, famC, gdh, pdhC and pgm. The database also contains information on which species each strain was assigned to in the laboratory submitting the sequence data. The input dataset comprised 8074 strains identified as meningococcus, 283 as N. lactamica, 193 as N. gonorrhoeae, four as N. cinerea, three as N. flava, two as N. flavescens, four as N. mucosa, four as N. mucosa, four as N. perflava, seven as N. polysaccharea, six as N. sicca and four as N. subflava. In addition, 31 strains had unresolved species status.

The strains used in the transformation experiments were obtained through the generosity of Prof. M. Maiden, University of Oxford, UK, Dr Heike Claus, Universität Würzburg, Germany and our own collection. Neisserial strains were routinely propagated at 37°C with 5 per cent CO2 on GC agar (Difco) supplemented with 1 per cent Vitox (Oxoid) or at 37°C, 180 r.p.m. in Mueller–Hinton broth (Oxoid) with 1 per cent Vitox.

Nalidixic acid-resistant derivatives of N. lactamica were selected by exposure of strains grown to the stationary phase in broth cultures to nalidixic acid at 10 μg ml⁻¹, followed by overnight growth on selection medium to isolate spontaneous gyrase mutants. The N. lactamica strains used were ST 595 and ST 3493.

2.2. Transformation experiments

Neisserial strains (1 × 10⁶ cfu ml⁻¹) were incubated with 1 μg ml⁻¹ chromosomal DNA isolated from a gyrase mutant in Mueller–Hinton broth supplemented with 1 per cent Vitox. Following 4 h incubation at 37°C
5 per cent CO<sub>2</sub>, the transformation mix was plated onto GC agar plates supplemented with 1 per cent Vitox containing 20 µg ml<sup>−1</sup> nalidixic acid. Transformation frequencies were determined as the number of antibiotic resistant colony-forming unit per millilitre recovered divided by the total colony-forming unit per millilitre counted on non-selective medium. For each meningococcal isolate, transformation frequencies were measured in triplicate against both derivatives of <i>N. lactamica</i> strains. Details of the meningococcal strains used in the experiments are given in the electronic supplementary material, table S1.

### 2.3. Sequence analysis

We used BAPS software [21,22] to cluster the sequences into genetically distinct groups and to infer the amount of gene flow between the groups. The optimal clustering was obtained using 10 runs of the estimation algorithm with the prior upper bound of the number of clusters varying in the range (100, 300) over the 10 replicates. All estimation runs yielded highly congruent partitions of the ST data with either 50 or 51 clusters, indicating a strongly peaked posterior distribution in the neighbourhood of these partitions. The estimated posterior mode clustering had 50 clusters, and the admixture analysis was subsequently performed using these clusters with 100 Monte Carlo replicates for allele frequencies and by generating 100 reference genotypes to calculate <i>p</i>-values.

For reference cases, we used 10 iterations in estimation according to the guidelines of Corander & Marttinen [23]. An ST was considered significantly admixed if the <i>p</i>-value did not fall below the threshold of 5 per cent. Given the extremely large number of STs in the data, the population genetic analyses took approximately 5360 computational hours. For further details about the methods implemented in BAPS, see the original papers.

### 2.4. Phylogenetics

To examine the relatedness of hybrid clusters to each other and the entire sample, we constructed a maximum-likelihood tree using concatenates of MLST sequences for all STs in our dataset. The tree was produced using RAXML v. 7.0.4 [24], with a general time reversible model and gamma model of rate heterogeneity. We used the rapid hillclimbing mode (option –f), and the bootstrapping was calculated by drawing the bipartition information onto the tree, based on the results from 100 individual bootstrap replicates. The 100 bootstrap replicates were computed on a parallel environment and the whole analysis was repeated twice due to numerical problems in the first run, such that only results from the second run are included in the study. The RAXML analyses took in total approximately 14,544 computational hours. It must be noted that the details of these phylogenies are highly suspect owing to the impact of recombination, and hence taxon labels have been removed. We wish to further emphasize the fact that the trees reported are primarily intended for examining the molecular distances among the BAPS groups and to see which groups form well-resolved clusters of STs in the tree.

### 3. RESULTS

#### 3.1. Results of BAPS analysis and identification of significantly mixed groups

BAPS clustered the MLST data into 50 groups, of which 42 contained strains of a single species (<i>N. meningitidis</i>), whereas the remaining eight groups contained isolates of at least two species, in which case we label them as mixed (table 1). The observed species distribution of mixed <i>N. lactamica</i> and <i>N. meningitidis</i> groups (2, 31, 41, 46) was highly significant in comparison with the null hypothesis that all strains are equally likely to be erroneously identified at the species level, indicating that these clusters are indeed taxonomically problematic (for details of these tests, and for complete results of population genetic analyses, see the electronic supplementary material). In the following, we focus primarily on the earlier-mentioned groups containing strains significantly identified as more

<table>
<thead>
<tr>
<th>species/group</th>
<th>2</th>
<th>31</th>
<th>34</th>
<th>41</th>
<th>46</th>
<th>48</th>
<th>49</th>
<th>50</th>
</tr>
</thead>
<tbody>
<tr>
<td>Neisseria cinerea</td>
<td>1 (4.8%)</td>
<td>3 (6.3%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria flava</td>
<td>3 (14.3%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria flavescens</td>
<td>2 (9.5%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria gonorrhoeae</td>
<td>193 (99.5%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria lactamica</td>
<td>48 (88.9%)</td>
<td>225 (84.0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria meningitidis</td>
<td>6 (11.1%)</td>
<td>43 (16.0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria mucosa</td>
<td>1 (4.8%)</td>
<td>3 (37.5%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria perflava</td>
<td>4 (19.0%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria polysacchara</td>
<td>7 (14.6%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria sicca</td>
<td>3 (14.3%)</td>
<td>3 (37.5%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Neisseria subflava</td>
<td>3 (14.3%)</td>
<td>1 (12.5%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>unknown</td>
<td>3 (14.3%)</td>
<td>31 (64.6%)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Table 1. Frequencies of STs from different <i>Neisseria</i> species among mixed BAPS groups. Percentage of each species within a group is given within parentheses.

J. R. Soc. Interface (2012)
than one species, to address their relationship to admixture and mosaicism between meningococcus and *N. lactamica*.

The population genetic analysis identified four mixed groups of *N. lactamica* and *N. meningitidis* STs containing in total 282 *N. lactamica* and 92 *N. meningitidis* STs (table 1). In addition, there are three other mixed groups (34, 48, 49) with higher level of species heterogeneity and a single group (50) into which all *N. gonorrhoeae* STs were assigned. Frequencies of different serotypes of *N. meningitidis* within the four mixed BAPS groups and in total among *N. meningitidis*. Percentage of each serotype is given within parentheses.

<table>
<thead>
<tr>
<th>serotype/group</th>
<th>2</th>
<th>31</th>
<th>41</th>
<th>46</th>
<th>all groups</th>
</tr>
</thead>
<tbody>
<tr>
<td>29E</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>106 (1.4%)</td>
</tr>
<tr>
<td>A</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>147 (1.9%)</td>
</tr>
<tr>
<td>B</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>4043 (54.4%)</td>
</tr>
<tr>
<td>C</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>752 (9.8%)</td>
</tr>
<tr>
<td>H</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4</td>
<td>4 (&lt;0.01%)</td>
</tr>
<tr>
<td>K</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2</td>
<td>2 (&lt;0.01%)</td>
</tr>
<tr>
<td>NG</td>
<td>6</td>
<td>40</td>
<td>17</td>
<td>8</td>
<td>1574 (21.2%)</td>
</tr>
<tr>
<td>other</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9</td>
<td>9 (&lt;0.01%)</td>
</tr>
<tr>
<td>W-135</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>6</td>
<td>266 (3.6%)</td>
</tr>
<tr>
<td>X</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>117 (1.6%)</td>
</tr>
<tr>
<td>Y</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>3</td>
<td>371 (5.0%)</td>
</tr>
<tr>
<td>Z</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50</td>
<td>50 (&lt;0.01%)</td>
</tr>
</tbody>
</table>

Figure 1. BAPS estimates of admixture for the STs in the mixed groups. Each ST corresponds to a coloured vertical bar, where fractions of colours indicate the amount of variation attributed to a particular group. For STs with non-significant admixture (*p*-value greater than 5%), the bars are single-coloured. The dark blue colour refers to the sum of estimated admixture from any of the 42 groups harbouring solely *Neisseria meningitidis* STs.

Table 2. Frequencies of different serotypes of *Neisseria meningitidis* within four particular mixed BAPS groups and in total among *N. meningitidis*. Percentage of each serotype is given within parentheses.

To examine whether BAPS groups vary in the degree to which recombination has contributed to their composition, we used BAPS admixture analysis [21]. Given the large number of detected groups, it is not feasible to visualize them simultaneously with respect to the estimated level of admixture. Figure 1 shows the admixture estimates for the STs in the mixed groups such that the level of gene flow from purely *N. meningitidis* groups is also visible. The admixture results jointly for all groups are numerically presented in a table format in the electronic supplementary material. We here define a hybrid group as one that contains a greater proportion of variation characteristic of more than one cluster, than expected by chance (given the observed amount of admixture in the entire dataset). Since the recombination frequency is expected to be dependent on sequence similarity, it is necessary to take this into account when quantitatively comparing different groups. Figure 2 shows the distribution of levels of significant non-zero admixture between the following three sets of BAPS groups: (i) purely *N. meningitidis*, (ii) mixed *N. lactamica* and *N. meningitidis* and (iii) between sets in (i) and (ii).

3.2. Admixture analysis: identification of hybrid groups

3.2. Admixture analysis: identification of hybrid groups

To examine whether BAPS groups vary in the degree to which recombination has contributed to their composition, we used BAPS admixture analysis [21]. Given the large number of detected groups, it is not feasible to visualize them simultaneously with respect to the estimated level of admixture. Figure 1 shows the admixture estimates for the STs in the mixed groups such that the level of gene flow from purely *N. meningitidis* groups is also visible. The admixture results jointly for all groups are numerically presented in a table format in the electronic supplementary material. We here define a hybrid group as one that contains a greater proportion of variation characteristic of more than one cluster, than expected by chance (given the observed amount of admixture in the entire dataset). Since the recombination frequency is expected to be dependent on sequence similarity, it is necessary to take this into account when quantitatively comparing different groups. Figure 2 shows the distribution of levels of significant non-zero admixture between the following three sets of BAPS groups: (i) purely *N. meningitidis*, (ii) mixed *N. lactamica* and *N. meningitidis* and (iii) between sets in (i) and (ii).
The fraction of pairs of groups with an inferred average level of admixture equal to zero is 12.8 per cent, 0 per cent and 83 per cent for these three cases, respectively. The decreasing amount of admixture as a function of molecular distance is well visible in figure 2. However, the mixed *N. lactamica* and *N. meningitidis* groups show evidence of considerably higher level admixture at molecular distances equivalent to those found between the two species groups in general. Thus, those groups with evidence of taxonomic uncertainty are considerably more hybrid than non-mixed ones.

### 3.3. Phylogenetic analysis

The phylogenetic tree in figure 3 shows the position of STs in groups 2, 31, 41 and 46 in the overall database.
It should be noted that the details of this tree must be considered suspect owing to recombination, and we make no particular claims regarding ordering of the branches. Groups 2, 31 and 41 are shown as closely related to each other, forming well-resolved clusters. In contrast, group 46 is not coherent, and its component STs are scattered around the tree.

Laboratory errors containing randomly selected alleles from the meningococcal and \textit{N. lactamica} populations would not be expected to cluster together in a phylogeny. Hence, the clustering of STs in groups 2, 31 and 41 strongly argues against the interpretation that these hybrid groups are laboratory artefacts, while group 46 may be. Finally, STs arising from laboratory errors should not be found more than once, and 46 STs in the hybrid groups have been recorded on more than one occasion, providing further evidence that these observations are secure.

3.4. Results of transformation experiments

The earlier-mentioned accounts suggest that strains in the mixed \textit{N. lactamica} and \textit{N. meningitidis} groups may be more likely to take up DNA than others. We therefore directly examined the degree of variation in transformation rates among strains from the mixed and non-mixed groups. The data are shown in the electronic supplementary material, figure S5. A linear-correlated random effects model was fitted to the data using SPSS 17 to account for the dependences within each triplicate measurement corresponding to the same combination of a meningococcal and \textit{N. lactamica} strain. Grouping with respect to mixed and non-mixed STs was used as a fixed effect in the model, and the difference between the two groups was significant at 5 per cent level (\(p\)-value 0.048). A single meningococcal isolate in the mixed group failed to produce any measurable transformation frequency against both \textit{N. lactamica} strains. Similarly, zero frequencies were also observed for all replicates of a single isolate in the non-mixed group. When these zero frequencies are excluded from the random effects analysis, the \(p\)-value for the difference equals 0.013.

4. DISCUSSION

It has been previously reported that the \textit{Neisseria} undergo recombination within and between species, to the extent that it confounds conventional phylogenetic analyses [9]. This has led to the ‘fuzzy species’ concept, first applied to the \textit{Neisseria}, which states that recombination between named species produces strains that contain a mosaic of DNA sequence characteristic of each, and which blurs the genotypic boundaries between species [10]. While the fuzzy species concept accurately reflects the observed distribution of genotypes in sequence space, the biological significance of the mosaic strains has remained uncertain. Are these strains phonetically as well as genetically plastic [12], and are they only transient, with little evolutionary future, because as previously stated ‘the fate of hybrid genotypes is important’ [25]? We have found groups of strains with evidence of significant past admixture between the named species \textit{N. meningitidis} and \textit{N. lactamica}, and moreover that these are also characterized by clear evidence of taxonomic confusion.

The identification of bacteria to the species level in clinical laboratories is an important part of patient care, determining appropriate therapies and treatment regimens. Despite this, occasional errors do occur. However, the ability of \textit{N. lactamica} to produce beta-galactosidase and acid from lactose is normally considered to distinguish it well from the meningococcus [26]. We have shown in this work that some groups of related \textit{N. lactamica} and meningococcus strains are more difficult to securely identify than this would suggest, given the mixture of species identifications within them. It is not possible to determine from the database whether the distribution of phenotypic characteristics in these groups is divergent from that found among strains in the main meningococcal and lactamica clusters. However, if we take species identification as a proxy for phenotype, then we can conclude that this is likely the case. Strikingly, we also find that these mixed groups contain sequence data found to be characteristic of both the meningococcus and \textit{N. lactamica}, providing an association (though not a causal link) between a history of recombination and phenotypic variation. Such phenotypic variation arising from genetic plasticity may be a contributing factor to difficulties with species classification. The relatively close relationships between the STs in these groups, as displayed by the tree shown in figure 3, argue that they are not laboratory artefacts produced by mixed cultures. Further evidence for this assertion comes from the preponderance of strains among those isolates identified as meningococcus for which no serogroup can be determined: greater than 90 per cent of isolates in groups 2, 31 and 41, in contrast with only 21 per cent of isolates in all meningococci (table 2). The observed accumulation of non-groupable strains is clearly significant (\(p\)-value less than 0.001) in these three groups (for details of tests, see the electronic supplementary material). The exception is group 46, which is scattered around the tree (figure 3) and contains a smaller proportion of non-groupable strains (\(p\)-value 0.075). Group 46, containing 28 STs, may indeed be composed of laboratory errors, but we are confident that this explanation can be rejected for groups 2, 31 and 41, which together contain 346 STs.

We have also examined differences in transformation rate between strains directly by experiment. While relatively few transformation experiments were performed, we nevertheless found a significant difference, among transformable strains, between the mean transformation efficiencies of members of mixed BAPS groups and those that were not. While this observation would in isolation be difficult to interpret, together with the results of the computational analyses presented here, it provides further support for the existence of marked variation in recombination rate among \textit{Neisseria} lineages.

The reasons for the variation in interspecific recombination that we observe among BAPS groups are obscure. Both mechanistic and ecological factors are plausible contributors. \textit{Neisseria lactamica} is known to typically colonize younger children than \textit{N. meningitidis} [2]. A sub-population of \textit{N. lactamica} that preferentially colonized older hosts would be expected to encounter meningococci more frequently, just as similar
opportunities for interspecific recombination would be offered by a meningococcal sub-population preferentially colonizing younger children. Other plausible factors include features of the mismatch repair system [27], or restriction modification systems [28,29]. Since the lineages with extended interspecific recombination have survived long enough to produce a cluster of related strains, any selective impact cannot be too deleterious. On the other hand, these lineages are also rare, which possibly suggests either some selective cost or that they have arisen only very recently. One possible scenario is that these strains have derived a short-term selective advantage by lowering the barriers to the incorporation of DNA from the other species, and as a result have acquired loci from that other species that allow them to colonize a new region of niche space. However, this must remain speculation in the absence of more whole genome data, and an improved understanding of the relevant ecological niche structure.

It is evident that extensive recombination has shaped the evolution of Neisseria. However, even if the gene flow within the population appears to have very, if indeed any, limited geographical barriers, it still exhibits directionality among named species in a non-uniform manner. Exchange of particular genes (aroE, ghaA) between N. lactamica and N. meningitidis strains has been reported previously [30]. However, the large-scale population study conducted here reveals separate evolutionary lineages within the N. meningitidis population, such that the level of recombination between them is smaller than within the lineages, even when controlling for the molecular distance. There is also a clear and consistent directionality in recombination between gonococci and other named species. Consistent with previous observations, the gonococci in the database contain no sequence characteristic of other clusters [10,31]. However, there is clear evidence of transfer from gonococci to mixed groups, established by examining sequence variation in individual loci (electronic supplementary material).

5. CONCLUSION

Bacteria can fall into a bewilderingly complicated range of population structures. The principles underlying each however can be abstractly understood in terms of the generation and shuffling of molecular variation. We have shown here the synergy between genetics, statistics and laboratory experiments in helping to elucidate some features of neisserial population structure. The methods we discuss and apply may be complemented by others both already available and in development, and the increasing numbers of whole genomes will enormously aid in this enterprise.

This publication made use of the Neisseria multi-locus sequence typing website (http://pubmlst.org/neisseria/) developed by Keith Jolley and sited at the University of Oxford. The development of this site has been funded by the Wellcome Trust and European Union. C.A.O’D. and J.S.K. were supported by generous grants to J.S.K. from Caroline Conran, from the Livanos Trust and funding from the Health Protection Agency, UK. We wish to thank Martin Maiden for providing the N. lactamica strains. The authors would like to thank the anonymous reviewers for their comments, which led to considerable improvements in the article. This work benefited significantly from discussions with participants of the 2nd Permafrost Workshop. The work of J.C. was supported by ERC grant no. 239784 and a grant from Sigrid Juselius Foundation. W.P.H. was supported by award number U54GM088558 from the National Institute of General Medical Sciences. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institute of General Medical Sciences or the National Institutes of Health.

REFERENCES


