The Influence of Recombination on the Population Structure and Evolution of the Human Pathogen *Neisseria meningitidis*

Edward C. Holmes, Rachel Urwin, and Martin C. J. Maiden

Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford

The extent to which recombination disrupts the bifurcating treelike phylogeny and clonal structure imposed by binary fission on bacterial populations remains contentious. Here, we address this question with a study of nucleotide sequence data from 107 isolates of the human pathogen *Neisseria meningitidis.* Gene fragments from 12 housekeeping loci distributed around the meningococcal chromosome were analyzed, showing that (1) identical alleles are disseminated among genetically diverse isolates, with no evidence for linkage disequilibrium; (2) different loci give distinct and incongruent phylogenetic trees; and (3) allele sequences are incompatible with a bifurcating treelike phylogeny at all loci. These observations are consistent with the hypothesis that meningococcal populations comprise organisms assembled from a common gene pool, with alleles and allele fragments spreading independently, together with the occasional importation of genetic material from other species. Further, they support the view that recombination is an important genetic mechanism in the generation new meningococcal clones and alleles. Consequently, for anything other than the short-term evolution of this species, a bifurcating treelike phylogeny is not an appropriate model.

Downloaded from https://academic.oup.com/mbe/article/16/6/741/2925463 by guest on 16 August 2022

Introduction

Whereas eukaryotic organisms have evolved mechanisms of sexual reproduction in which extensive genetic recombination occurs as an integral part of propagation, the bacteria reproduce asexually by binary fission, with each haploid mother cell giving rise to two genetically identical daughter cells. In the absence of a sexual process, variation will occur solely by the accumulation of mutational events within lineages of bacteria that share a common ancestor. This imposes a bifurcating treelike phylogeny, with mutation generating its branches, and the resultant linkage disequilibrium of variant alleles in the population is characteristic of a ''clonal'' population structure, a predominant paradigm for bacteria (Selander and Levin 1980; Levin 1981).

Despite this, comparative nucleotide sequence analyses indicate that the parasexual processes of many bacteria (transformation, conjugation, and transduction) can have a major impact on their evolution (Maynard Smith et al. 1993; Maynard Smith 1995; Lawrence and Ochman 1997). Such processes usually result in the incorporation of a relatively small DNA fragment at a given location on the genome, a process that has been termed ''localized sex''(Maynard Smith, Dowson, and Spratt 1991). Although localized sex is apparently more limited in its effects than genuine sexual reproduction, certain aspects of bacterial biology increase its evolutionary importance. For example, while the potential for interbreeding may define species in multicellular organisms, bacteria share genetic information more widely by homologous recombination between DNA with up to ~30% sequence divergence, and by illegitimate recombination in which nonhomologous genetic material can be incorporated into the chromosome (Lorenz and Wackernagel 1994). Intragenic recombination of divergent DNA results in mosaic genes in which different gene segments exhibit different evolutionary histories and which may have novel biological properties (Feavers et al. 1992; Maiden 1993, 1998; Spratt et al. 1995).

The frequency of localized sex is an important factor in determining bacterial population structure, which ranges from fully clonal, where genetic exchange is effectively absent, to nonclonal or panmictic, where transfer of genetic material prevents the population from being made up of clones (Maynard Smith et al. 1993). However, different subpopulations of the same species may exhibit different population structures, and distinct structures may be simultaneously present in the same population but operate on different timescales. For example, the rapid spread of the descendants of a particularly successful bacterium may result in a short-term, or ''epidemic,'' clonal structure even in a population that is otherwise nonclonal (Maynard Smith et al. 1993).

Here, we examined the population structures and evolutionary patterns of the bacterial pathogen *Neisseria meningitidis,* which exchanges DNA by transformation and for which a range of population structures have been proposed (Maiden and Feavers 1995). In particular, we used phylogenetic techniques to ask whether the population structure of *N. meningitidis* is better represented as a bifurcating tree, as predicted under the clonal model, a network of interconnected nodes, as expected if recombination has occurred frequently, or as a star phylogeny in which lineages have arisen within a short time period, characteristic of recent positive selection or rapid population growth (Hudson 1990), as predicted under the epidemic clone model (Maynard Smith 1989).

Key words: *Neisseria meningitidis,* recombination, population structure, evolution, split decomposition.

Address for correspondence and reprints: Martin C. J. Maiden, Wellcome Trust Centre for the Epidemiology of Infectious Disease, Department of Zoology, University of Oxford, South Parks Road, Oxford OX1 3PS United Kingdom. E-mail: martin.maiden@zoo.ox.ac.uk.

Mol. Biol. Evol. 16(6):741–749. 1999
© 1999 by the Society for Molecular Biology and Evolution. ISSN: 0737-4038

Materials and Methods

Bacterial Isolates and Nucleotide Sequences

The data set comprised gene fragments of 12 housekeeping genes encoded by loci distributed around the chromosome from 107 isolates chosen to be representative of the known diversity of the meningococcus. The sequences, which totaled 5,659 bp, or approximate-

Table 1
Variability of the Loci Examined in the 107 Isolates

Locus	Size (bp)	Number of Variable Sites	Number of Alleles
abcZ	433	75	15
adk	465	17	10
aroE	490	166	19
fumC	465	38	19
gdh	501	28	16
mtg	497	76	16
pdhC	480	80	24
pgm	450	77	21
pilA	432	50	36
pip	416	26	19
ppk	579	77	23
serC	451	67	29

ly 0.25% of the *N. meningitidis* genome, were those collected during the development of a novel typing system for bacteria, multilocus sequence typing (MLST) (Maiden et al. 1998), supplemented with data from an additional locus, *fumC,* encoding fumarate hydratase (GenBank accession numbers AF086737–AF086755).

For MLST, the data were presented as allelic variants, with each distinct allele sequence having a unique number. The number of alleles varied from 10 to 36, and the number of variable sites per gene fragment ranged from 17 to 166, depending on the locus (table 1). Allele numbers were used to define the sequence type (ST) of each isolate, the relatedness of which was represented by a dendrogram constructed by cluster analysis using the unweighted pair group method with arithmetic mean (UPGMA). Although this method grouped identical or very closely related isolates, which was useful for epidemiological typing (Maiden et al. 1998), it was not a reliable indicator of phylogenetic relationships among isolates.

Assessing the Extent of Recombination at the Genome Level

The clonal model of bacterial evolution implies that the genome forms a single linkage group and that all loci, whatever their genomic locations, will record the same phylogenetic history. Therefore, a simple means of assessing clonality would be to test the extent of congruence between trees reconstructed on each of the 12 loci, an approach previously used for *Borrelia* (Dykhuizen et al. 1993). However, this was not possible for the *N. meningitidis* data set presented here because of the large number of loci and isolates and the relatively short lengths of the sequences, all of which would reduce phylogenetic resolution and increase computing complexity. To overcome these problems, the congruence of trees was examined in a subset of the most diverse isolates.

Thirty strains, one from each of the major lineages observed in the MLST analysis (Maiden et al. 1998), were chosen as representative of the deep phylogenetic history of *N. meningitidis.* This subset contained isolates that were separated in the UPGMA cluster analysis by a linkage distance of >0.4. A maximum-likelihood (ML) phylogenetic tree was then constructed for each locus. However, because of the presence of many identical or very closely related alleles, trees were also reconstructed on four groups of three loci spliced together. Combining data in this way is valid, because clonality should lead to a single phylogenetic tree whatever data partition is used. These groups, which reflected the spatial ordering of the genes on the chromosome (i.e., genes were assembled in accordance with their gene order on a meningococcal chromosome map [Dempsey, Wallace, and Cannon 1995; Maiden et al. 1998]) were *aroE-mtg-pilA* (1,419 bp), *ppk-pgm-adk* (1,494 bp), *pip-abcZ-pdhC* (1,329 bp), and *gdh-fumC-serC* (1,417 bp).

All ML phylogenetic trees were reconstructed using the HKY85 model of DNA substitution with the value of the transition : transversion (Ts/Tv) ratio and the shape parameter (α) of a discrete approximation (with eight categories) to a gamma distribution of rate heterogeneity among sites estimated from the empirical data during tree reconstruction. The degree of incongruence between trees was assessed for each of the four groups of loci by comparing the likelihood of the ML tree for each group with those of the ML topologies obtained for the other groups and with 100 randomly generated trees of the same size. If the ML trees of each group are congruent, then all should have likelihoods higher than those of the random trees. To determine whether the incongruencies observed were due to low phylogenetic resolution or were strongly supported, as expected under recombination, 1,000 replicate bootstrap trees were reconstructed. Because of computational constraints, this analysis was undertaken with Neighbor-Joining bootstrap trees, but using the ML substitution model (including empirical Ts/Tv and α values). All of these analyses were undertaken with the 4.64 test version of PAUP*, kindly provided by David L. Swofford.

Assessing the Extent of Intragenic Recombination

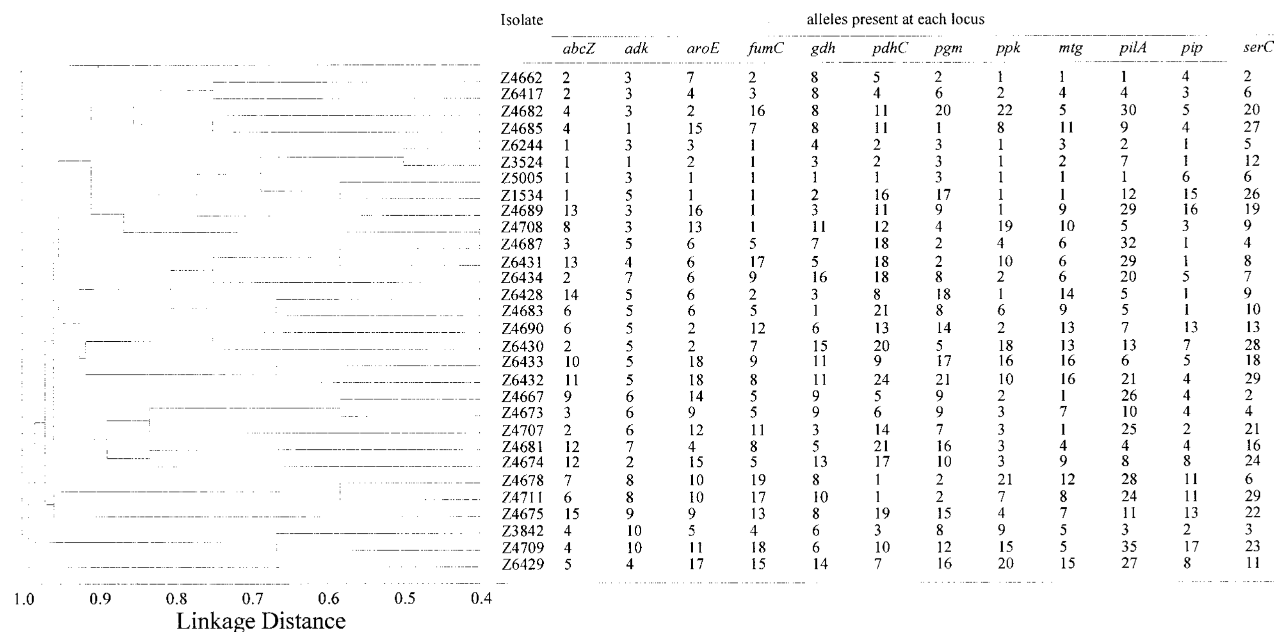
Because of the difficulties in confirming phylogenetic incongruence within genes caused by the low level of sequence variation at each locus, the method of split decomposition was used to assess the degree of treelike structure present in the alleles found for each locus in the complete set of 107 isolates (Bandelt and Dress 1992; Dopazo, Dress and von Haeseler 1993). This method, which depicts parallel edges between sequences if there is conflicting phylogenetic signal in the data, is a useful way of visualizing the extent of networked evolution among the sequences, itself indicative of recombination (Maynard Smith 1989). This analysis was undertaken with the SplitsTree program, version 2.4 (Huson 1998), with the input pairwise distances between sequences, on which the splits graph is constructed, estimated using hamming (uncorrected) distances, although similar results were found with other distance measures.

Results

Distribution of Alleles Among Meningococci

A UPGMA dendrogram generated from the STs of the 30 more distantly related members of the strain collection is shown in fig. 1, combined with a table show-

Downloaded from https://academic.oup.com/mbe/article/16/6/741/2925463 by guest on 16 August 2022



Isolate	alleles present at each locus											
	abcZ	*adk*	*aroE*	*fumC*	*gdh*	*pdhC*	*pgm*	*ppk*	*mtg*	*pilA*	*pip*	*serC*
Z4662	2	3	7	2	8	5	2	1	1	1	4	2
Z6417	2	3	4	3	8	4	6	2	4	4	3	6
Z4682	4	3	2	16	8	11	20	22	5	30	5	20
Z4685	4	1	15	7	8	11	1	8	11	9	4	27
Z6244	1	3	3	1	4	2	3	1	3	2	1	5
Z3524	1	1	2	1	3	2	3	1	2	7	1	12
Z5005	1	3	1	1	1	1	3	1	1	1	6	6
Z1534	1	5	1	1	2	16	17	1	1	12	15	26
Z4689	13	3	16	1	3	11	9	1	9	29	16	19
Z4708	8	3	13	1	11	12	4	19	10	5	3	9
Z4687	3	5	6	5	7	18	2	4	6	32	1	4
Z6431	13	4	6	17	5	18	2	10	6	29	1	8
Z6434	2	7	6	9	16	18	8	2	6	20	5	7
Z6428	14	5	6	2	3	8	18	1	14	5	1	9
Z4683	6	5	6	5	1	21	8	6	9	5	1	10
Z4690	6	5	2	12	6	13	14	2	13	7	13	13
Z6430	2	5	2	7	15	20	5	18	13	13	7	28
Z6433	10	5	18	9	11	9	17	16	16	6	5	18
Z6432	11	5	18	8	11	24	21	10	16	21	4	29
Z4667	9	6	14	5	9	5	9	2	1	26	4	2
Z4673	3	6	9	5	9	6	9	3	7	10	4	4
Z4707	2	6	12	11	3	14	7	3	1	25	2	21
Z4681	12	7	4	8	5	21	16	3	4	4	4	16
Z4674	12	2	15	5	13	17	10	3	9	8	8	24
Z4678	7	8	10	19	8	1	2	21	12	28	11	6
Z4711	6	8	10	17	10	1	2	7	8	24	11	29
Z4675	15	9	9	13	8	19	15	4	7	11	13	22
Z3842	4	10	5	4	6	3	8	9	5	3	2	3
Z4709	4	10	11	18	6	10	12	15	5	35	17	23
Z6429	5	4	17	15	14	7	16	20	15	27	8	11

FIG. 1.—Alleles present at each locus and the relationships of isolates by UPGMA analysis. A UPGMA dendrogram, drawn on the basis of the alleles present at each of the 12 loci in the subset of the 30 most diverse meningococcal isolates, is shown, together with the alleles present at each locus for each isolate.

ing the alleles for each locus. The index of association (I_A) test for linkage disequilibrium (Maynard Smith et al. 1993) was performed on the 12 loci for both the 107 strains ($I_A = 4.57$) and the 30-strain subset ($I_A = 0.65$), showing evidence for linkage in the former but not the latter. Visual inspection revealed that for each locus, identical alleles were distributed among distantly related isolates. Examples included *abcZ* allele 2, which occurred in five isolates, *gdh* allele 8, present in six isolates, and *pgm* allele 2, found in five isolates at various locations on the dendrogram.

Degree of Congruence Among Loci

The ML phylogenetic trees for the four groups of concatenated *N. meningitidis* genes are presented in fig. 2. These trees are characterized by very different branching structures and by a distinct lack of congruence, such that isolates frequently changed positions among trees. For each group, only the ML tree had a likelihood better than those of 100 randomly generated trees of the same number of taxa; the ML topologies obtained for the other groups, tested on each data set, were no more likely than random assortments of the data (table 2). A similar pattern of widespread incongruence was found among trees constructed from each of the 12 loci separately (data not shown; these trees are available at http://mlst.zoo.ox.ac.uk/Meningococcus).

To further assess the effect of recombination on phylogenetic structure, we examined pairs of isolates found in each of the four phylogenies which were supported by ≥70% of bootstrap replications. If the 30 isolates had a single phylogenetic history at all loci, these strongly supported pairs of isolates should be found in all trees, provided there is sufficient resolution, and no isolate should be found in more than a single pair.

Among the four trees of three loci, 22 such isolate pairs were found (table 3). Only three of these pairings were found in more than a single tree, and only one of these, isolate pairing A22 : 71/94 (present in the *aroE-mtg-pilA* and *gdh-fumC-serC* trees), received ≥70% bootstrap support in more than one tree. In no case was a pairing found in all four data sets, and the *pip-abcZ-pdhC* data set had no pairing in common with any other group. Even if pairings that were split by a single sequence were included (denoted by "?" in table 3), there was little evidence of congruence. Furthermore, many of the strains formed different strongly supported pairings in each of the trees. For example, although the isolate pairing A22 : 71/94 was well supported in the *aroE-mtg-pilA* and *gdh-fumC-serC* trees, isolate 71/94 clustered with isolate BZ 198 in 100% of bootstrap replicates of the *pip-abcZ-pdhC* data set. In turn, isolate BZ 198 grouped with isolate DK 24 in 100% of the *aroE-mtg-pilA* bootstraps and with isolate BZ 147 in 92% of the *gdh-fumC-serC* trees. Similar observations were made for the trees constructed for the 12 loci separately. Of the 30 isolates analyzed, 22 were found in different pairings supported by >70% bootstrap support (15 sequences at >90% support) and sequences from two isolates, EG011 and NG 940, were found in six different pairings (data not shown; available at http://mlst.zoo.ox.ac.uk/Meningococcus). Together, these results revealed substantial variation in the phylogenetic signals obtained from different loci within populations of *N. meningitidis.*

A second feature of the ML trees was the great variability in branch lengths, with certain isolates highly divergent in some trees but not in others. For example, isolate 3906 was very divergent in the *aroE-mtg-pilA*

Downloaded from https://academic.oup.com/mbe/article/16/6/741/2925463 by guest on 16 August 2022

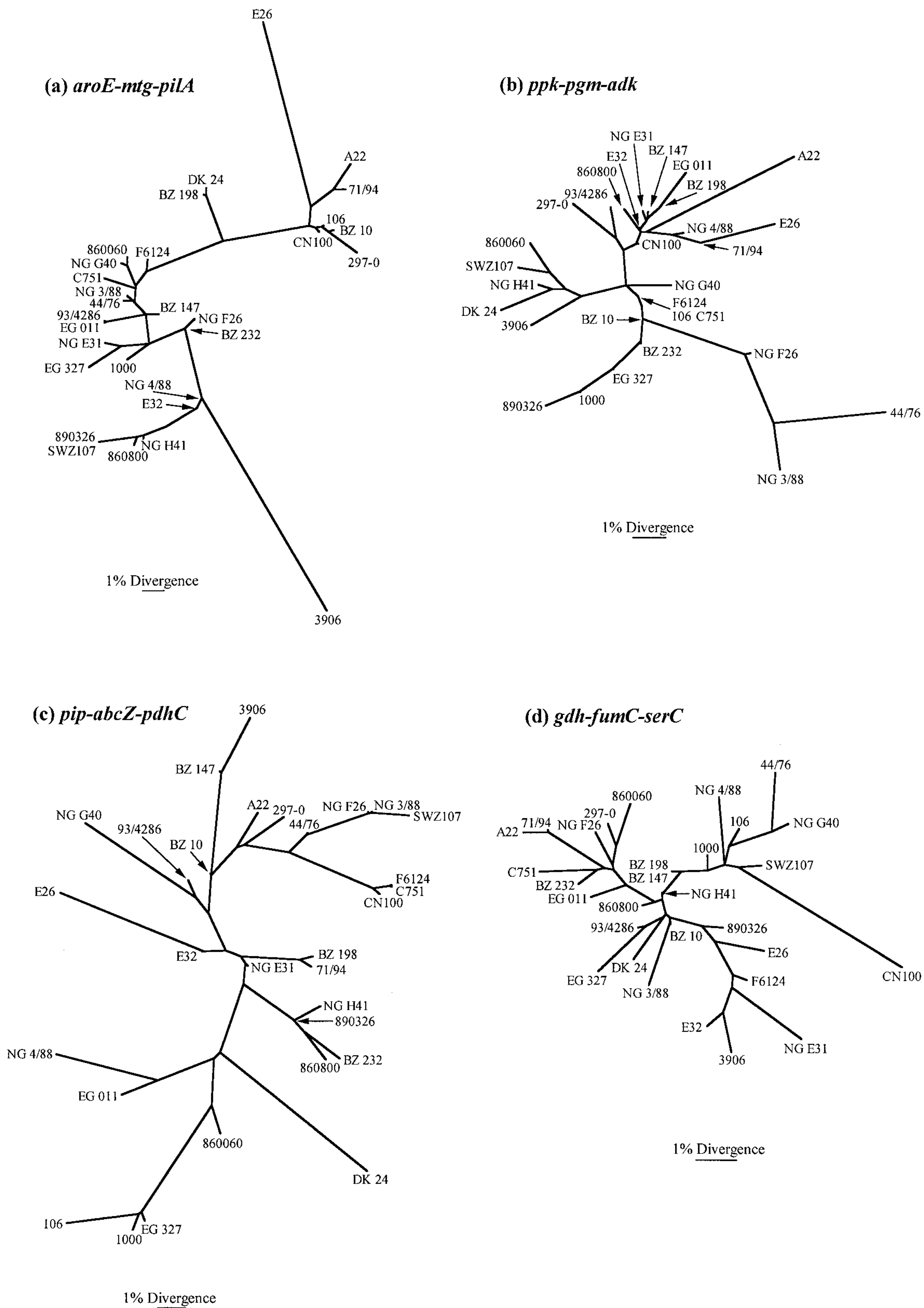


FIG. 2.—Maximum-likelihood trees for concatenated sets of loci from 30 strains of *N. meningitidis*. *a*, *aroE-mtg-pilA*. *b*, *ppk-pgm-adk*. *c*, *pip-abcZ-pdhC*. *d*, *gdh-fumC-serC*. All branch lengths are drawn to scale.

Downloaded from https://academic.oup.com/mbe/article/16/6/741/2925463 by guest on 16 August 2022

Table 2
Data Parameters of the Maximum-Likelihood (ML) Phylogenetic Analysis of 12 Gene Loci from *Neisseria meningitidis*

Group	Sequence Length (bp)	d[a]	Ts/Tv[b]	α[c]	*ln L*[d] of ML Tree	*ln L*[d] of ML Trees from other Groups (range)	*ln L* of 100 Random ML Trees (range)
aroE-mtg-pilA	1,419	0.081	2.176	0.122	−5,314.39207[e]	−7,288.30603 to −7,332.67235	−6,965.89465 to −7,521.51177
ppk-pgm-adk	1,494	0.035	3.154	0.047	−4,315.49046	−4,995.69087 to −5,105.34102	−4,922.33580 to −5,206.47962
pip-abcZ-pdhC	1,329	0.057	7.328	0.073	−4,359.87550[f]	−5,181.77561 to −5,375.76031	−5,147.84616 to −5,477.26468
gdh-fumC-serC	1,417	0.025	3.102	0.052	−3,877.21051	−4,299.25285 to −4,311.41617	−4,213.24706 to −4,348.88953

[a] Average pairwise genetic distance (under ML substitution model).
[b] Transition : transversion ratio.
[c] Alpha parameter of rate variation among sites.
[d] Log likelihood.
[e] 81 trees of equal likelihood were found.
[f] Nine trees of equal likelihood were found.

tree but in no other phylogeny, while isolate CN100, which was very divergent in the *gdh-fumC-serC* tree, showed only short branch lengths in the remaining trees.

Split Decomposition of Allele Sequences

Although the splits graphs generated from each of the 12 loci were unique, they all shared similar features, illustrated here with the results from *abcZ, pdhC,* and *aroE* (figs. 3–5); splits graphs for the other loci are available at http:/mlst.zoo.ox.ac.uk/Meningococcus. None of the loci gave a treelike structure, but instead comprised networks with long and short branches extending from them. In many cases, the splits graphs were sufficiently complex and the distances among alleles sufficiently great that the data sets had to be simplified by removing the alleles representing the longest branches to allow visualization of central networks (e.g., figs. 4 and 5). In most cases, the data were sufficiently complex that the ''fit'' of the splits graph of all alleles to the data set was poor, indicating the difficulty of resolving the relationships among alleles. In general, as alleles occupying the longest branches were successively removed, the fit parameter improved.

Mapping nucleotide differences onto the splits graphs provided further information, revealing that the networked structures comprised alleles which shared all of their polymorphisms in different combinations. The very small distances at this level made resolution difficult, such that inferences about phylogenetic structure could only be made with caution. The short branches usually, but not invariably, comprised alleles with one or two nucleotide changes not seen elsewhere in the data

Table 3
Incongruence Among Gene Trees of *Neisseria meningitidis*

Strain Pair	*aroE-mtg-pilA*	*ppk-pgm-adk*	*pip-abcZ-pdhC*	*gdh-fumC-serC*
DK 24 : BZ 198.	100%			*
93/4286 : EG011	100%			*
NG F26 : BZ 232	100%		*	?
860060 : NG G40	100%			*
A22 : 71/94.	93%		*	100%
44/76 : NG 3/88	93%	52%	*	*
3906 : NG 4/88	87%	?	*	
BZ 10 : 106	80%			
890326 : 860800	80%	*	*	
EG 327 : NG E31	79%			*
SWZ107 : 860060	*	83%		
1000 : 890326	*	76%	**	
F6124 : C751		?	100%	
BZ 147 : 3906	*		100%	*
BZ 198 : 71/94.	**		100%	**
NG 3/88 : NG F26.	**	?	91%	
NG H41 : 890326	*	*	91%	
EG 327 : 1000	?	?	70%	*
EG 327 : 93/4286.	**		*	96%
BZ 147 : BZ 198	*	50%	**	92%
44/76 : NG G40	**		?	81%
NG H41 : 860800	*		*	70%

NOTE.—All strain pairs supported by ≥70% of bootstrap replications in at least one group of loci are shown (left-hand column). The percentages of bootstrap support for these pairs in the other groups are shown in the remaining columns (blank cell, pairing not found). ?, pair split by a single sequence; *, one sequence of the pair forms another, i.e., incongruent, pair in ≥70% of bootstraps. **, both sequences of the pair form other, i.e., incongruent, pair in ≥70% of bootstraps.

Downloaded from https://academic.oup.com/mbe/article/16/6/741/2925463 by guest on 16 August 2022

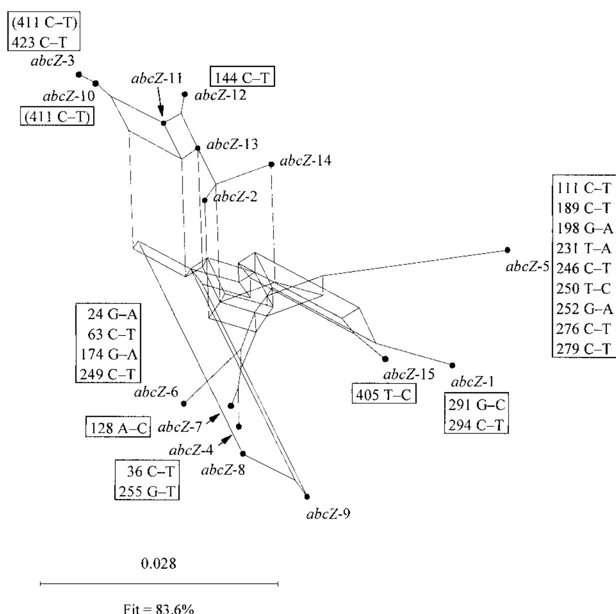


FIG. 3.—Annotated splits graph of the 15 alleles found at the *abcZ* locus for 107 strains. The nodes represent alleles and are labeled with MLST allele numbers (Maiden et al. 1998). The boxes adjacent to some nodes show the nucleotide changes unique to that allele in this data set. The arabic numbers indicate the base positions of the nucleotide changes: the first letter indicates the base present at that location in the majority of the alleles, and the letter after the dash indicates the base present in the allele in question. One nucleotide change which was unique to alleles *abcZ*-3 and *abcZ*-10, and consequently unique to a short branch, is shown in brackets. The scale bar represents uncorrected distances, and the fit parameter is shown.

set, while the longer branches represented alleles with numerous unique nucleotide changes. Several of the shorter branches contained no unique nucleotide changes (e.g., alleles *pdhC*-2, -8, -9, -10, and -24 in fig. 4 and *aroE*-3, - 4, and -15 in fig. 5) and while in the longer branches unique nucleotide changes were often localized within a small region of the gene in question (e.g., alleles *abcZ*-5, in fig. 3, and *aroE*-18, in fig. 5), the longest branches comprised alleles with many polymorphisms distributed throughout their sequences (e.g., *aroE*-13 in fig. 5). In some cases, longer branches formed networks themselves (e.g., alleles *pdhC*-1, -19, -20, -21, -22, and -23 in fig. 4). In a number of these examples, clear indications of mosaic structure could be seen among such alleles, some of which gave significant evidence of recombination as assessed by the maximum chi-squared test (e.g., alleles *pdhC*-21 and -22 are a mosaic of alleles -19 and -23, fig. 4; $P < 0.0001$ after 10,000 simulations of the data) (Maynard Smith 1992).

Discussion

The increasing availability of data sets comprising nucleotide sequences of multiple loci from extensive collections of bacterial isolates requires the development of appropriate analytical methods. This study has applied a variety of such techniques to 605,513 bp of nucleotide sequence data from a total of 107 isolates to explore the phylogenetic and population structure of *N. meningitidis*. The phylogenetic trees of the 30 most diverse isolates, whether reconstructed for each locus or for groups of three loci combined, exhibited more differences than similarities. The extent of this conflicting phylogenetic signal was such that a strain which has a position in one gene tree, strongly supported by bootstrap analysis, could occupy a very different but equally well supported position in another. Such large-scale discrepancies among the genes can only be explained by frequent recombination events. Additional evidence came from the variation in branch lengths, which was strongly suggestive of recombination events involving strains not included in the data set, perhaps including the importation of genetic variability from other *Neisseria* species. It is unlikely that differences in base composition caused the observed incongruence, given that GC content only varies from 54.2% to 56.3% among the data sets, and the low genetic distances among sequences (table 2) would lessen the effect of multiple substitution.

Genetic exchange during the evolution of *N. meningitidis* also provided the most reasonable explanation for the results of the split decomposition analyses, which demonstrated that for each locus, the nucleotide sequences of alleles were inconsistent with a branching treelike phylogeny. Three types of recombination event could account for the features of the splits graphs obtained. Networks could be generated by extensive recombination over a prolonged timescale, with consequent distribution of the polymorphisms present in the population among alleles. Longer branches are consistent with the importation of divergent genes, or large fragments of them, from other organisms, most probably other *Neisseria* species, and shorter branches are consistent with the simultaneous movement of adjacent nucleotide changes as a result of intragenic recombination events. While some of the short branches could be explained by the accumulation of de novo mutations, it is also possible that they were generated by the importation of nucleotide changes from other alleles of the meningococcal or *Neisseria* population that were not present in the isolate collection analyzed here. In any case, the lack of extensive treelike structures in the splits graphs for the alleles at *any* locus suggests that mutation is not a major mechanism for allele evolution in this species or that these structures do not persist.

Evidence for genetic exchange was also provided by more quantitative measures of recombination. When the whole set of 107 strains was examined, multiple representatives of hyperinvasive lineages (Maiden et al. 1998) provided the data set with a high I_A (4.57), implying that the population is clonal and therefore essentially treelike. However, when the deeper phylogeny was considered by eliminating multiple members of hyperinvasive lineages, no linkage disequilibrium was observed ($I_A = 0.65$). Such a change in I_A is characteristic of an ''epidemic'' population structure and is consistent with analysis of multilocus enzyme electrophoresis data sets for meningococci (Maynard Smith et al. 1993; Maynard Smith 1995), which is probably due, at least in part, to an overrepresentation of meningococcal isolates ob-

Downloaded from https://academic.oup.com/mbe/article/16/6/741/2925463 by guest on 16 August 2022

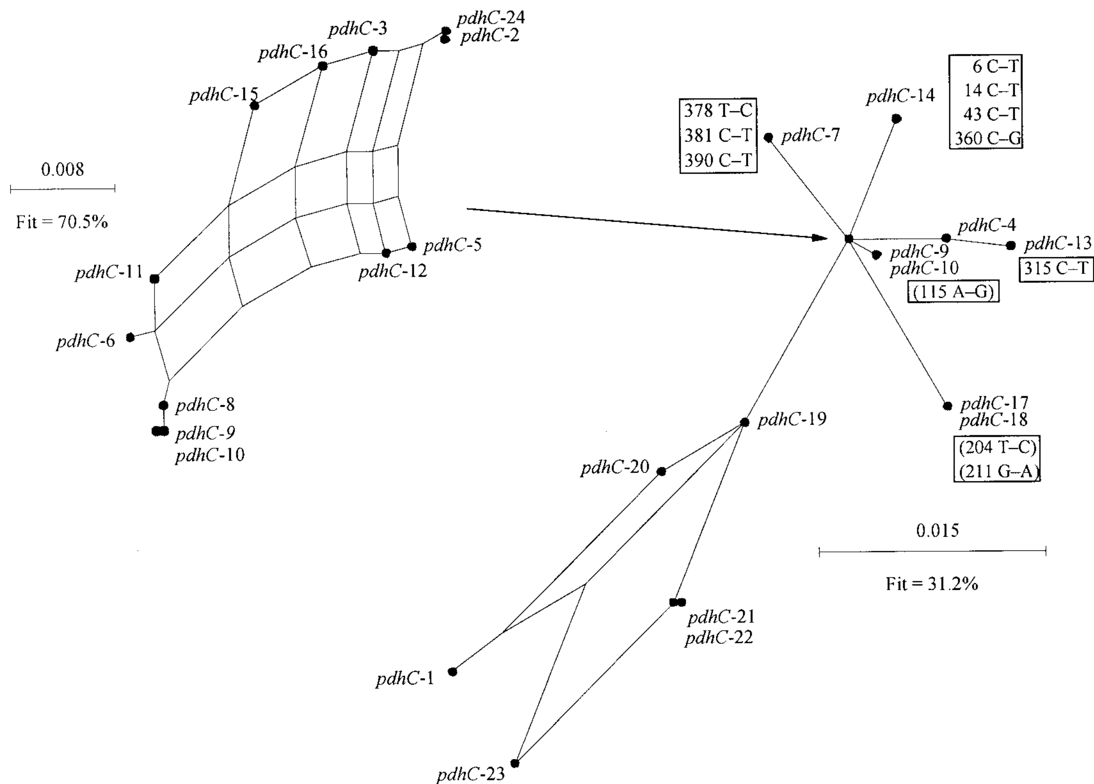


FIG. 4.—Annotated splits graphs of the 24 alleles found at the meningococcal *pdhC* locus for 107 strains, drawn following the same conventions used in figure 3. In this case, resolution of the central network is shown with a smaller splits graph; its location within the larger splits graph is indicated by the arrow. In addition to the unique nucleotide changes, one nucleotide change shared by alleles *pdhC*-9 and *pdhC*-10 and two shared by alleles *pdhC*-17 and *pdhC*-18 are shown. In this data set, the latter three nucleotide changes are unique to the short branch on which they occur.

tained from invasive disease in strain collections (Maiden and Feavers 1995).

Together, these analyses confirm that there is no single bifurcating tree which can link all strains of *N. meningitidis* and reveal that, as recombination occurs at a high rate both within and among loci, a network is often a more biologically accurate representation of relationships among strains. In these circumstances, a more informative way to reconstruct the evolutionary history of bacteria would be with a method, such as split decomposition, which allows them be to connected without assuming an evolutionary process a priori. It is essential to appreciate that the phylogenetic tree is an hypothesis of evolutionary history, not a truism.

Ideally, it would be possible to estimate an absolute rate of recombination, or, more precisely, the probability that any nucleotide has changed by recombination or by mutation. Unfortunately, such measures are difficult to obtain, given that data samples are often nonrandom and too small for the relevant statistics to work with accuracy (Hudson 1987), and the occurrence of multiple substitution, codon bias, and natural selection can distort any signal recovered (Maynard Smith and Smith 1998). Furthermore, other key population parameters, such as population size and mutation rate, may be unknown, or recombination may have involved some sequences not in the data analyzed. As these limitations will affect many sequence data sets, we suggest that determining how recombination influences population structure represents a profitable way to understand this important evolutionary process.

In summary, recombination within meningococcal populations is sufficient to disrupt a branching treelike phylogeny, which is clearly not an appropriate model for the long-term evolution of *N. meningitidis.* This is consistent with the concept of a global gene pool for *Neisseria* species (Maiden, Malorny, and Achtman 1996). Even if the short branches in the splits graphs represent de novo point mutations, it seems that in the absence of a strong selective advantage, such alleles do not generally escape from the network. Ultimately, while all variation must have been generated by mutation, the data presented here suggest that this slow process is not the primary route by which new allelic variants arise in the meningococcus.

Acknowledgments

E.C.H. is a Royal Society University Research Fellow and M.C.J.M. is a Wellcome Senior Research Fellow. We are grateful for financial support provided by the Wellcome Trust and the Royal Society. We thank Dr. Ed Feil for assistance with the I_A analyses.

Downloaded from https://academic.oup.com/mbe/article/16/6/741/2925463 by guest on 16 August 2022

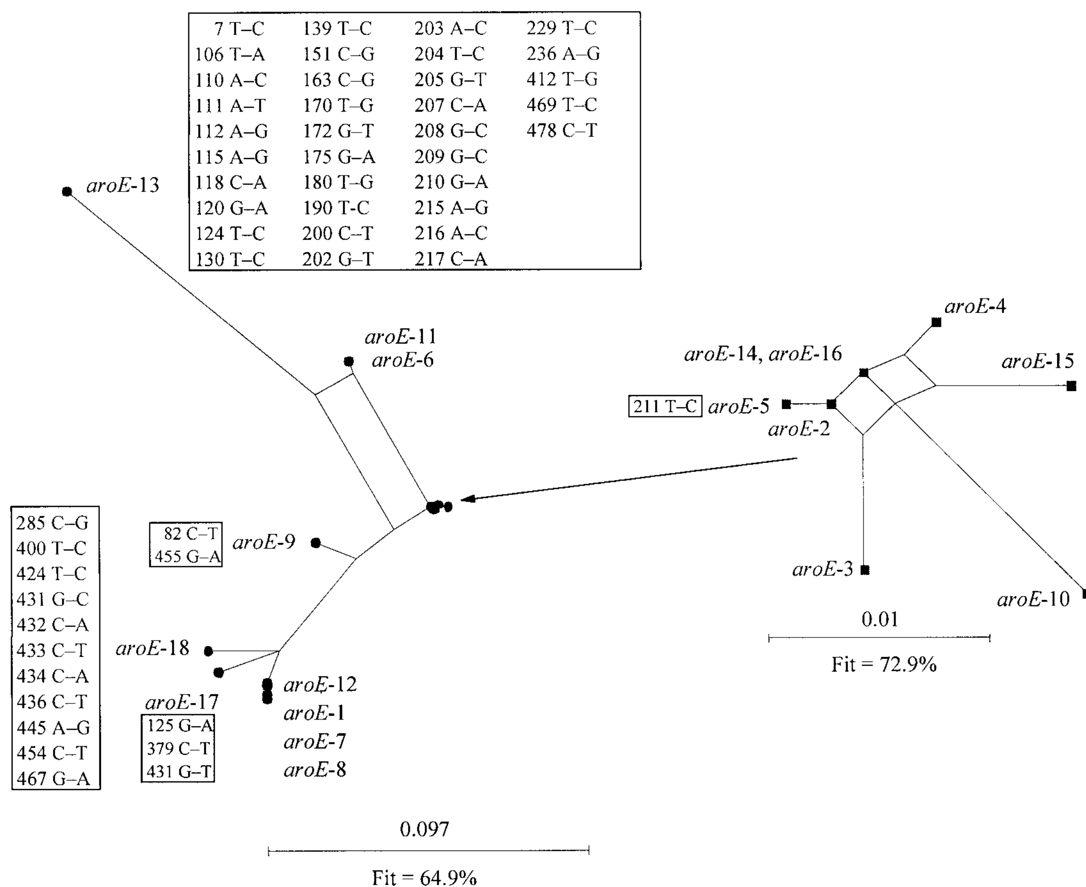


FIG. 5.—Annotated splits graphs of the 18 alleles found at the meningococcal *aroE* locus for 107 strains, drawn with the same conventions used in figures 3 and 4.

LITERATURE CITED

BANDELT, H. J., and A. W. DRESS. 1992. Split decomposition: a new and useful approach to phylogenetic analysis of distance data. Mol. Phylogenet. Evol. **1**:242–252.

DEMPSEY, J. A., A. B. WALLACE, and J. G. CANNON. 1995. The physical map of the chromosome of a serogroup A strain of *Neisseria meningitidis* shows complex rearrangements relative to the chromosome of the two mapped strains of the closely related species *N. gonorrhoeae.* J. Bacteriol. **177**:6390–6400.

DOPAZO, J., A. DRESS, and A. VON HAESELER. 1993. Split decomposition: a technique to analyze viral evolution. Proc. Natl. Acad. Sci. USA **90**:10320–10324.

DYKHUIZEN, D. E., D. S. POLIN, J. J. DUNN, B. WILSKE, V. PREAC MURSIC, R. J. DATTWYLER, and B. J. LUFT. 1993. *Borrelia burgdorferi* is clonal: implications for taxonomy and vaccine development. Proc. Natl. Acad. Sci. USA **90**: 10163–10167.

FEAVERS, I. M., A. B. HEATH, J. A. BYGRAVES, and M. C. MAIDEN. 1992. Role of horizontal genetic exchange in the antigenic variation of the class 1 outer membrane protein of *Neisseria meningitidis.* Mol. Microbiol. **6**:489–495.

HUDSON, R. R. 1987. Estimating the recombination parameter of a finite population model without selection. Genet. Res. **50**:245–250.

———. 1990. Gene genealogies and the coalescent process. Pp. 1–44 *in* D. FUTUYMA and J. ANTONOVICS, eds. Oxford surveys in evolutionary biology. Oxford University Press, Oxford.

HUSON, D. H. 1998. SplitsTree: a program for analysing and visualising evolutionary data. Bioinformatics **14**:68–73.

LAWRENCE, J. G., and H. OCHMAN. 1997. Amelioration of bacterial genomes: rates of change and exchange. J. Mol. Evol. **44**:383–397.

LEVIN, B. R. 1981. Periodic selection, infectious gene exchange and the genetic structure of *E. coli* populations. Genetics **99**:1–23.

LORENZ, M. G., and W. WACKERNAGEL. 1994. Bacterial gene transfer by natural genetic transformation in the environment. Microbiol. Rev. **58**:563–602.

MAIDEN, M. C. J. 1993. Population genetics of a transformable bacterium: the influence of horizontal genetical exchange on the biology of *Neisseria meningitidis.* FEMS Microbiol. Lett. **112**:243–250.

———. 1998. Horizontal genetic exchange, evolution, and spread of antibiotic resistance in bacteria. Clin. Infect. Dis. **27**:S12–S20.

MAIDEN, M. C. J., J. A. BYGRAVES, E. FEIL et al. (12 coauthors). 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. USA **95**: 3140–3145.

MAIDEN, M. C. J., and I. M. FEAVERS. 1995. Population genetics and global epidemiology of the human pathogen *Neisseria meningitidis.* Pp. 269–293 *in* S. BAUMBERG, J. P. W. YOUNG, E. M. H. WELLINGTON, and J. R. SAUNDERS, eds. Population genetics of bacteria. Cambridge University Press, Cambridge, England.

Downloaded from https://academic.oup.com/mbe/article/16/6/741/2925463 by guest on 16 August 2022

MAIDEN, M. C. J., B. MALORNY, and M. ACHTMAN. 1996. A global gene pool in the neisseriae. Mol. Microbiol. **21**: 1297–1298.

MAYNARD SMITH, J. 1989. Trees, bundles or nets. Trends Ecol. Evol. **4**:302–304.

———. 1992. Analysing the mosaic structure of genes. J. Mol. Evol. **34**:126–129.

———. 1995. Do bacteria have population genetics? Pp. 1–12 *in* S. BAUMBERG, J. P. W. YOUNG, E. M. H. WELLINGTON, and J. R. SAUNDERS, eds. Population genetics of bacteria. Cambridge University Press, Cambridge, England.

MAYNARD SMITH, J., C. G. DOWSON, and B. G. SPRATT. 1991. Localized sex in bacteria. Nature **349**:29–31.

MAYNARD SMITH, J., and N. H. SMITH. 1998. Detecting recombination from gene trees. Mol. Biol. Evol. **15**:590–599.

MAYNARD SMITH, J., N. H. SMITH, M. O'ROURKE, and B. G. SPRATT. 1993. How clonal are bacteria? Proc. Natl. Acad. Sci. USA **90**:4384–4388.

SELANDER, R. K., and B. R. LEVIN. 1980. Genetic diversity and structure in *Escherichia coli* populations. Science **210**: 545–547.

SPRATT, B. G., N. H. SMITH, J. ZHOU, M. O'ROURKE, and E. FEIL. 1995. The population genetics of the pathogenic *Neisseria.* Pp. 143–160 *in* S. BAUMBERG, J. P. W. YOUNG, E. M. H. WELLINGTON, and J. R. SAUNDERS, eds. Population genetics of bacteria. Cambridge University Press, Cambridge, England.

HOWARD OCHMAN, reviewing editor

Accepted February 15, 1999

Downloaded from https://academic.oup.com/mbe/article/16/6/741/2925463 by guest on 16 August 2022