J Mol Evol (1996) 43:631–640

JOURNAL OF MOLECULAR EVOLUTION

A Comparison of the Nucleotide Sequences of the *adk* and *recA* Genes of Pathogenic and Commensal *Neisseria* Species: Evidence for Extensive Interspecies Recombination Within *adk*

Edward Feil, Jiaji Zhou, John Maynard Smith, Brian G. Spratt

School of Biological Sciences, University of Sussex, Falmer, Brighton, BN1 9QG, U.K.

Received: 12 January 1996 / Accepted: 20 May 1996

Abstract. The sequences of the adenylate kinase gene (*adk*) and the RecA gene (*recA*) were determined from the same isolates of *Neisseria gonorrhoeae, N. meningitidis, N. lactamica, N. polysaccharea, N. cinerea, N. mucosa, N. pharyngis* var. *flava, N. flavescens,* and *N. animalis.* The patterns of sequence divergence observed at *adk* and *recA* were very different. Dendrograms constructed from the *recA* data using two different algorithms were statistically robust and were congruent with each other and with the relationships between the species previously proposed using other data. In contrast, the dendrograms derived from the *adk* data were noncongruent with each other, and with those from the *recA* data, and were statistically poorly supported. These results, along with the uniform distribution of pairwise sequence divergences between the species at *adk,* suggest there has been a history of interspecies recombination within the *adk* gene of the human *Neisseria* species which has obscured the phylogenetic relationships between the species. This view was supported by Sawyer's runs test, and the Index of Association (I_A) between codons, which provided significant evidence for interspecies recombination between the *adk* genes from the human *Neisseria* species, but no evidence of interspecies recombination between the *recA* sequences.

Key words: *Neisseria* — Interspecies recombination — Adenylate kinase — Index of Association — Linkage disequilibrium — Codon linkage

Correspondence to: B.G. Spratt

Introduction

The true *Neisseria* are members of the beta subclass of *Proteobacteria* and are Gram-negative diplococci. Several *Neisseria* species are opportunistic pathogens of animals, although most of the species which colonize humans are harmless commensals (Knapp 1988). There are two species which are important human pathogens: *N. gonorrhoeae* (the gonococcus) and *N. meningitidis* (the meningococcus) are the causative agents of gonorrhoea and meningococcal meningitis, respectively. *N. meningitidis* and the human commensal *Neisseria* species normally reside in the nasopharynx, although meningococci occasionally gain access to the blood and cerebrospinal fluid to cause septicemia and meningitis (Knapp 1988). *N. gonorrhoeae* is normally isolated from the urinogenital tract of humans, although the sexual activity of the host means that it is also commonly isolated from the rectum or nasopharynx.

Meningococci and gonococci are naturally transformable and this mechanism of genetic exchange has the potential to promote localized recombination between chromosomal segments (Maynard Smith et al. 1991; Maiden 1993; Spratt et al. 1995). The extent of recombination in nature is the subject of current debate and appears to vary widely between bacterial species (Istock et al. 1992; Maynard Smith et al. 1993; Boyd et al. 1994; Guttman and Dykhuizen 1994). In some species recombination appears to be very rare, such that their population biology is characterized by largely isolated clonal lineages. In these species phylogenetic relationships between isolates, whether inferred from the sequences of

different housekeeping genes or from isoenzyme studies, should be broadly congruent (Boyd et al. 1994). Similarly, the phylogenies of highly clonal bacterial species reconstructed from the sequences of housekeeping genes are unlikely to be distorted significantly by interspecies recombination (Dykhuizen and Green 1991). In other bacteria the frequency of recombination appears to be much greater, such that the alleles in the population are at, or close to, linkage equilibrium (Maynard Smith et al. 1993). In these nonclonal species, recombination renders any attempt to reconstruct the phylogenetic relationships between isolates from the sequences of housekeeping genes meaningless. Furthermore, the extent of recombination between closely related species may be much greater in certain bacterial genera than in others, and a history of repeated interspecies recombinational events may similarly compromise the validity of inferred species phylogenies.

The frequency of recombination appears to be high in meningococci and gonococci compared to that in most other bacteria. Thus alleles of housekeeping loci appear to be close to linkage equilibrium in gonococci (O'Rourke and Stevens 1993; Vázquez et al. 1993) and in serogroup B and C meningococci (Maynard Smith et al. 1993; Spratt et al. 1995), although sampling problems arising from the overrepresentation of disease-causing isolates can obscure the underlying nonclonal nature of the meningococcal population (Maynard Smith et al. 1993). Serogroup A meningococci have a very different epidemiology from serogroup B and C meningococci and appear to represent a clonal subpopulation (Olyhoek et al. 1987; Achtman 1994; Spratt et al. 1995). Evidence for intraspecies recombination is difficult to obtain for gonococci by nucleotide sequencing as housekeeping genes from this species are extremely uniform (Vázquez et al. 1993; E.F. and B.G.S., unpublished). However, sequencing of housekeeping genes from meningococci has supported the idea of frequent recombination as the phylogenies derived from different housekeeping genes are noncongruent (Zhou and Spratt 1992; Feil et al. 1995).

Interspecies recombination between gonococci and meningococci, and between the pathogenic and commensal species, has been reported in genes under strong directional selection (e.g., the penicillin-binding protein 2 gene of penicillin-resistant *N. gonorrhoeae* and *N. meningitidis*; Spratt et al. 1992) and under diversifying selection (e.g., in the *por* and *opa* genes that encode outer membrane proteins; Ward et al. 1992; Hobbs et al. 1994; Vázquez et al. 1995). However, interspecies recombination has been inferred from the sequences of genes under strong selection even in highly clonal species (Nelson and Selander 1994), presumably since very rare recombinational events may be observed if there is strong selection for the recombinant phenotype. Meningococci appear to differ from highly clonal species, as evidence for interspecies recombination with commensal *Neisseria* species has been found in most housekeeping genes that have been examined. Thus, interspecies recombination has been inferred from the sequences of the *adk*, *recA*, *glnA*, and *argF* genes from small samples of meningococci (Zhou and Spratt 1992; Spratt and Zhou 1994; Feil et al. 1995; J.Z. and B.G.S., unpublished).

Recombination between gonococci and the other human *Neisseria* species appears to be much less common than recombination between meningococci and the latter species, and there are no known examples within housekeeping genes. This presumably reflects the ecological isolation of gonococci from meningococci and the human commensal *Neisseria* species (Vázquez et al. 1993).

In this work we have examined the consistency of phylogenetic trees of the human *Neisseria* species reconstructed using sequence data from two housekeeping genes, *recA* and *adk*, which encode the RecA recombinase and adenylate kinase (ADK), respectively. Our results show a noncongruence of trees and suggest that a history of frequent interspecies recombination, presumably mediated by genetic transformation, has distorted the phylogeny of species inferred from the *adk* sequence data. In contrast, the *recA* sequences examined here do not provide any evidence of interspecies recombination.

Materials and Methods

Bacterial Strains. The following strains were used: *N. gonorrhoeae* CH95, *N. meningitidis* S3446, *N. polysaccharea* NCTC11858, *N. lactamica* NCTC10617, *N. cinerea* LNP1646, *N. flavescens* LNP444, *N. mucosa* LNP405, *N. pharyngis* var. *flava* NCTC4590, and *N. animalis* NCTC10212. Strains designated NCTC were obtained from the National Collection of Type Cultures (Colindale, UK); LNP strains were kindly provided by Dr. J.-Y. Riou from the collection of the Laboratoire *Neisseria* de l'Institut Pasteur. The bacteria were grown on GC Base Agar (Difco) plus supplements at 37°C in an atmosphere of 95% air/5% CO_2 and chromosomal DNA was prepared as described (Bowler et al. 1994).

Generation of Sequence Data. The cloning and sequencing of the *adk* gene of *N. gonorrhoeae* and *N. meningitidis* has been described (Feil et al. 1995). The primers used previously to amplify the whole *adk* coding region from the pathogenic *Neisseria* species did not amplify the gene from several of the commensal *Neisseria* species. Degenerate primers based on sequences within the gonococcal *adk* gene (5′-GGCGCCGGCAAAGGNACNCAA/GGC-3′ nucleotides 28–50, and 5′-TTCGGCTTTCACNGCT/CTCNAC-3′, nucleotides 604–624) were therefore used to amply an internal fragment from all of the *Neisseria* species. The degenerate primers described previously (Zhou and Spratt 1992) were used to generate a 877-bp fragment of the *recA* coding region. Sequencing of a 522-bp region of *adk*, and a 822-bp region of *recA*, was carried out for all nine species on both strands, directly from the PCR products, using the λ-exonuclease method (Higuchi and Ochman 1989). The sequences of the *adk* and *recA* genes have been submitted to EMBL/GenBank under the access numbers L36471, L47158, U57708–U57713, U57901–U57910.

Phylogenetic Analysis. Sequences were aligned (no gaps were required) and phylogenetic reconstructions from sequence data were car-

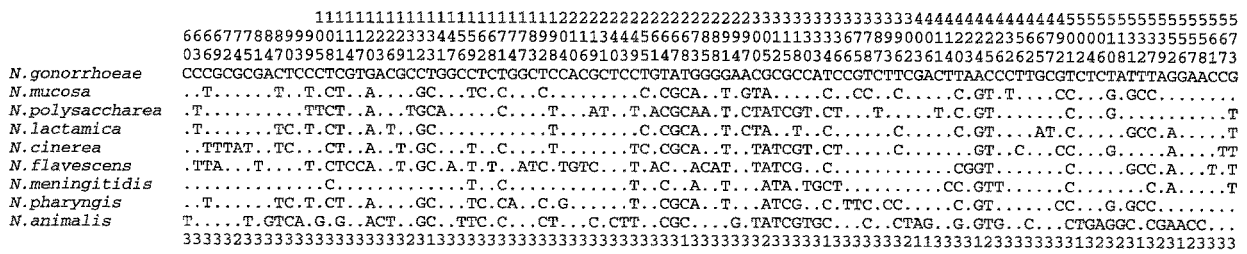

Fig. 1. Polymorphic sites within the *adk* gene of the *Neisseria* species. The *adk* sequence of *N. gonorrhoeae* is used as the master sequence and only those nucleotides in the *adk* genes of the other *Neisseria* species that differ from the *N. gonorrhoeae* sequence are shown. The position of each of these polymorphic sites is shown above in *vertical format,* and the position of the site within a codon is shown below. The *numbering* of the sequence is that of Feil et al. (1995).

ried out using the Molecular Evolutionary Genetics Analysis (MEGA) suite of programs (Kumar et al. 1993).

Statistical Analysis. Sequence data were subjected to two types of statistical analysis—Sawyer's runs test and the Index of Association between codons. Sawyer's runs test was performed following the method described previously (Sawyer 1989) using Multicomp (Reeves et al. 1994). The Index of Association (I_A) between codons is essentially the same test as that used previously for calculating the amount of linkage disequilibrium between electrophoretic allozymes (Brown et al. 1980; Maynard Smith et al. 1993), except that allozymes are equated with polymorphic codons, and the use of standard errors to estimate significance is replaced by simulation, as the published standard error estimation (Brown et al. 1980) is erroneous. Essentially, the I_A test looks at the distribution of the genetic differences between all pairs of strains. If there is an association between the alleles present at different loci in a given strain, then some pairs of strains will be very similar and other pairs will be very different. In other words, the variance of the between-strain differences will be higher than expected by chance. To apply the test to sequence data, the sequences are aligned and, for each polymorphic codon, the different codons are assigned a number—0, 1, 2, etc. Monomorphic codons are ignored. The observed variance (V_E) can then be calculated in basically the same way as for MLEE data with the electrophoretic parameters (in inverted commas) replaced with the sequence parameters (in parentheses).

Suppose there are r "loci" (polymorphic codons) within the "population" (set of n sequences) and P_{ij} is the frequency within the population of the ith "allele" (codon variant) at the jth locus. If k_{ij} = the total number of the ith allele at the jth locus (such that $P_{ij} = k_{ij}/n$) the probability that two individuals will be the same at the jth locus is given by:

$$\Sigma_i P_{ij} (k_{ij} - 1)/(n - 1) \quad (1)$$

The probability that two random individuals will be different at the jth locus is therefore given by:

$$pd = 1 - \Sigma_i P_{ij} (k_{ij} - 1)/(n - 1) \quad (2)$$

V_E can then be calculated from the following equation:

$$V_E = \Sigma pd (1 - pd) \quad (3)$$

and the Index of Association can then be calculated as:

$$I_A = V_O/V_E - 1 \quad (4)$$

If the data are in linkage equilibrium, the expected value of I_A is zero. An estimate of the statistical significance of I_A is made by calculating V_O from 1,000 random data sets. In each random set, the numbers of each kind of allele, at each locus, were the same as in the real data, but the alleles were assigned randomly to strains. If the data is in strong linkage disequilibrium (i.e., $P < 0.001$), then none of the values of V_O obtained from the random data sets should exceed that obtained with the real data.

Results

Patterns of Divergence Between Species

All the species included in this investigation were isolated from humans, with the exception of *N. animalis,* which was isolated from a guinea pig nasopharynx (Barrett and Sneath 1994). This species was included on the assumption that it has remained ecologically isolated from the human *Neisseria* species, making it unlikely that it would have been able to recombine with them. The inclusion of this species therefore provides a negative control for interspecies recombination and an outgroup for the construction of dendrograms. Both DNA-DNA hybridization data (Guibourdenche et al. 1986) and the phenotypic analysis carried out by Barrett and Sneath (1994) have consistently placed *N. animalis* as distinct from the human species included in this study.

The polymorphic sites within the *adk* and *recA* genes of the nine *Neisseria* species are shown in Figs. 1 and 2, respectively. There were 241 polymorphic sites within the 822 bp of *recA* that were sequenced in all nine species, and 104 polymorphic sites within the 522 bp of *adk* sequenced in the same nine species. A visual inspection of these figures reveals no clear recombinational event between the species for either gene. Figure 3 shows the polymorphic residues amongst the derived amino acid sequences from *adk* and *recA*. It is clear from both genes that the most distinct species, as expected, is *N. animalis. N. gonorrhoeae, N. polysaccharea, N. lactamica,* and *N. meningitidis* share identical RecA amino acid sequences. Identical ADK amino acid sequences are shared by *N. mucosa, N. lactamica, N. meningitidis,* and *N. pharyngis* var. *flava.* (The amino acid sequences of ADK found in *N. polysaccharea, N. cinerea,* and *N. flavescens* each exhibit a single unique substitution.)

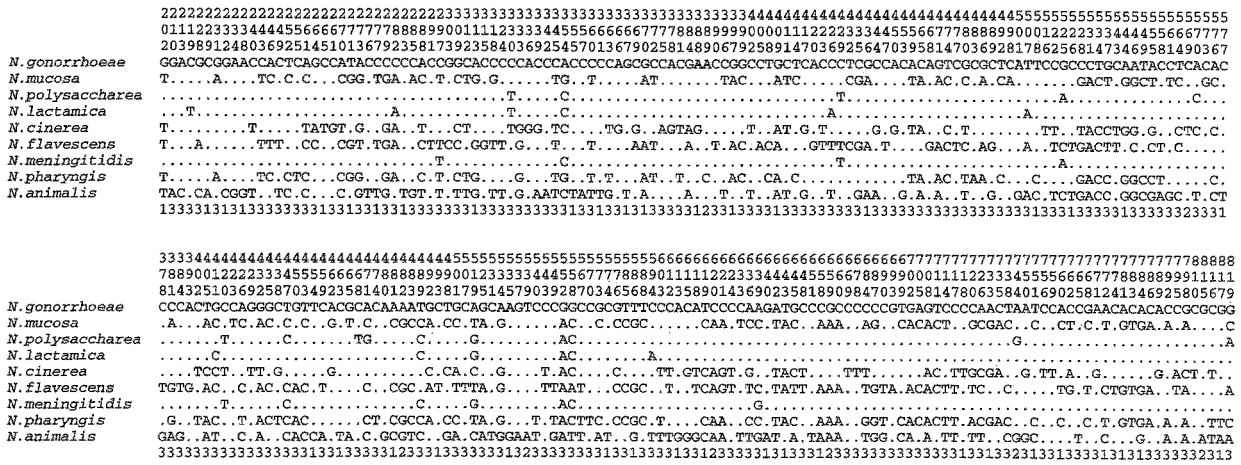

Fig. 2. Polymorphic sites within the *recA* gene of the *Neisseria* species. See legend to Fig. 1. The *numbering* of the sequence is that of Fyfe and Davies (1990).

A)

	111111111 249013346788 546225629306
N. gonorrhoeae	GPMGSVTTAEDE
N. mucosa	...DP..V....
N. polysaccharea	.L.DP..V....
N. lactamica	...DP..V....
N. cinerea	D..DP..V....
N. flavescens	..IDP..V....
N. meningitidis	...DP..V....
N. pharyngis	...DP..V....
N. animalis	...DPIAVDGGT

B)

	1111122222222222222222 222333334678889035680012333334678889 467234687190693288742398056783871455 6
N. gonorrhoeae	AMKQQENELIAVNVVVIDQKMFSTSEEVLRIININGA
N. mucosa	S.....DVV.I.IIIVE.R..A.QDD.IKVLYVE..
N. polysaccharea	.....
N. lactamica	.....
N. cinerea	S.....D.I.V.H.....Q...I.V.....
N. flavescens	S.....DVV.I.III.E.R..A.QDD.IK.LHVE..
N. meningitidis	.....
N. pharyngis	S.....DVV.ITII.VE.R..A.QDD.IKVLYVE.S
N. animalis	SINHKQD..VSI..IIVE.RVYAAQD.I.K..H.DDT

Fig. 3. Polymorphic amino acid residues in adenylate kinase and RecA protein. The sequence of *N. gonorrhoeae* is used as the master sequence. Only those residues that differ from the *N. gonorrhoeae* sequence are shown. The position of each of these polymorphic amino acid residues is shown above in vertical format. **A** Adenylate kinase. **B** RecA protein.

Tables 1 and 2 give the pairwise comparisons between all nine species, given as the number of nucleotide substitutions (upper right) and the percentage of nucleotide divergence (lower left) at both *adk* and *recA*. Figure 4 plots the distribution of these pairwise comparisons and shows the very striking differences between the *recA* and *adk* genes. The pairwise comparisons at *recA* fall into two distinct groups. There are six comparisons which are <1.9% and these represent all the comparisons between the four species, *N. gonorrhoeae, N. meningitidis, N. polysaccharea,* and *N. lactamica,* which are known to be closely related. Barring the comparison between *N. mucosa* and *N. pharyngis* var. *flava* (5.6%), all of the other 29 pairwise comparisons are >11%. *N. animalis* is the most diverged species with five out of the six pairwise comparisons ⩾18% involving this species.

In contrast to the *recA* data, where only a single pairwise comparison (2.8% of all comparisons) fell between 5% and 7.9% sequence divergence, 67% of all pairwise comparisons at *adk* fell within this range. Eight out of the nine comparisons of *adk* above 8% correspond to comparisons with *N. animalis,* the one exception being the 9% divergence found between *N. flavescens* and *N. gonorrhoeae.* At the other extreme, the three pairwise comparisons at *adk* under 4.9% correspond to *N. mucosa* compared with *N. pharyngis* var. *flava* and *N. lactamica* (2.7% and 3.8% diverged, respectively) and *N. gonorrhoeae* compared with *N. meningitidis* (4.4% diverged).

The ratio of synonymous to nonsynonymous substitutions (d_S/d_N) was calculated for *adk* and *recA* by the method of Nei and Gojobori (1986) using the two most divergent sequences (*N. animalis* and *N. flavescens* for *adk* and *N. animalis* and *N. cinerea* for *recA*). The d_S/d_N ratio was twice as high for *adk* (18.8:1) as for *recA* (9.5:1).

Phylogenetic Analysis of adk *and* recA *from the* Neisseria *Species*

The MEGA suite of programs (Kumar et al. 1993) was used to construct phylogenetic trees from both the *recA* and *adk* sequences. The trees given in Fig. 5 were constructed using the neighbor-joining method (Saitou and Nei 1987) and the UPGMA algorithm, both with the Jukes/Cantor correction for synonymous changes (Jukes and Cantor 1969). The dendrograms for *recA* are highly consistent with each other and clearly show the grouping of the four species *N. gonorrhoeae, N. lactamica, N. polysaccharea,* and *N. meningitidis.* This grouping is statistically very significant as it was supported by 100% of 500 bootstrap replicates, irrespective of which algorithm

Table 1. Pairwise comparisons between the *adk* genes of the *Neisseria* species[a]

	NGON	NMUC	NPOL	NLAC	NCIN	NFLA	NMEN	NPHA	NANI
NGON	—	35	39	34	41	47	23	40	58
NMUC	6.7%	—	33	20	28	38	32	14	48
NPOL	7.5%	6.3%	—	28	28	38	29	33	53
NLAC	6.5%	3.8%	5.4%	—	28	36	31	27	50
NCIN	7.6%	5.4%	5.4%	5.4%	—	41	29	27	46
NFLA	9.0%	7.3%	7.3%	6.9%	7.9%	—	37	36	64
NMEN	4.4%	6.1%	5.6%	5.9%	5.6%	7.1%	—	32	50
NPHA	7.7%	2.7%	6.3%	5.2%	5.2%	6.9%	6.1%	—	45
NANI	11.1%	9.2%	10.2%	9.6%	8.8%	12.3%	9.6%	8.6%	—

[a] NGON, *N. gonorrhoeae;* NMUC, *N. mucosa;* NPOL, *N. polysaccharea;* NLAC, *N. lactamica;* NCIN, *N. cinerea;* NFLA, *N. flavescens;* NMEN, *N. meningitidis;* NPHA, *N. pharyngis* var *flava;* NANI, *N. animalis.* Upper right, number of nucleotide substitutions; lower left, % nucleotide divergence

Table 2. Pairwise comparisons between the *recA* genes of the *Neisseria* species[a]

	NGON	NMUC	NPOL	NLAC	NCIN	NFLA	NMEN	NPHA	NANI
NGON	—	119	15	12	103	136	11	124	152
NMUC	14.5%	—	122	119	132	93	117	46	126
NPOL	1.8%	14.9%	—	15	105	139	8	127	150
NLAC	1.5%	14.5%	1.8%	—	102	137	13	124	151
NCIN	12.5%	16.1%	12.8%	12.4%	—	148	105	128	155
NFLA	16.6%	11.3%	16.9%	16.7%	18.0%	—	135	95	137
NMEN	1.3%	14.2%	1.0%	1.6%	12.8%	16.4%	—	122	148
NPHA	15.1%	5.6%	15.5%	15.1%	15.6%	11.6%	14.8%	—	128
NANI	18.5%	15.3%	18.2%	18.4%	18.9%	16.7%	18.0%	15.6%	—

[a] NGON, *N. gonorrhoeae;* NMUC, *N. mucosa;* NPOL, *N. polysaccharea;* NLAC, *N. lactamica;* NCIN, *N. cinerea;* NFLA, *N. flavescens;* NMEN, *N. meningitidis;* NPHA, *N. pharyngis* var *flava;* NANI, *N. animalis.* Upper right, number of nucleotide substitutions; lower left, % nucleotide divergence

was used to construct the tree. Furthermore, in both dendrograms *N. polysaccharea* appears to be most closely related to *N. meningitidis,* whereas *N. gonorrhoeae* is closer to *N. lactamica.* However, these groupings are not so statistically significant as there is little nucleotide variation amongst these four sequences. *N. cinerea* is the next most closely related species to this group at *recA,* a node supported by 100% of bootstrap trees using the neighbor-joining method and 99% using the UPGMA algorithm. The grouping of *N. mucosa* with *N. pharyngis* var. *flava* was also strongly supported (100% of bootstrap trees irrespective of which algorithm is used). *N. flavescens* is also significantly grouped with these two species irrespective of which algorithm is used. The dendrograms for these nine species at *recA* are therefore statistically very robust, which suggests that they reflect a true phylogeny.

In contrast, the trees derived from the *adk* sequences using the neighbor-joining method and the UPGMA algorithm are inconsistent with each other and with the trees derived from the recA sequences. There are two significant groupings, *N. mucosa* with *N. pharyngis* var. *flava* (which is supported by 86% of bootstrap trees using the neighbor-joining method and 92% using the UPGMA algorithm), and *N. gonorrhoeae* with *N. meningitidis* (which is supported by 78% of bootstrap trees using the UPGMA algorithm and 96% of trees using the neighbor-joining method). The other nodes varied according to which algorithm was used but all exhibited very low bootstrap scores. For example, using the neighbor-joining method, *N. flavescens* appeared to group with *N. polysaccharea* and the pathogenic species, but this node has a very low bootstrap score (31%), and is not supported by the tree calculated using the UPGMA method or the trees derived from *recA.* Similarly, *N. lactamica* groups with *N. pharyngis* var. *flava* and *N. mucosa* using the UPGMA method applied to the *adk* data, whereas it clusters with the pathogenic species using the *recA* data.

Statistical Analysis of the adk *and* recA *Sequences from the* Neisseria *Species*

Two statistical tests were used to examine the recombinational histories of *recA* and *adk*—Sawyer's runs test (Sawyer 1989) and the Index of Association (I_A), which is normally used for the analysis of linkage disequilibrium in isoenzyme data (Maynard Smith et al. 1993).

Sawyer's runs test was performed on the total data sets for all nine species for both *recA* and *adk.* Tests were carried out using all silent polymorphic sites, and only

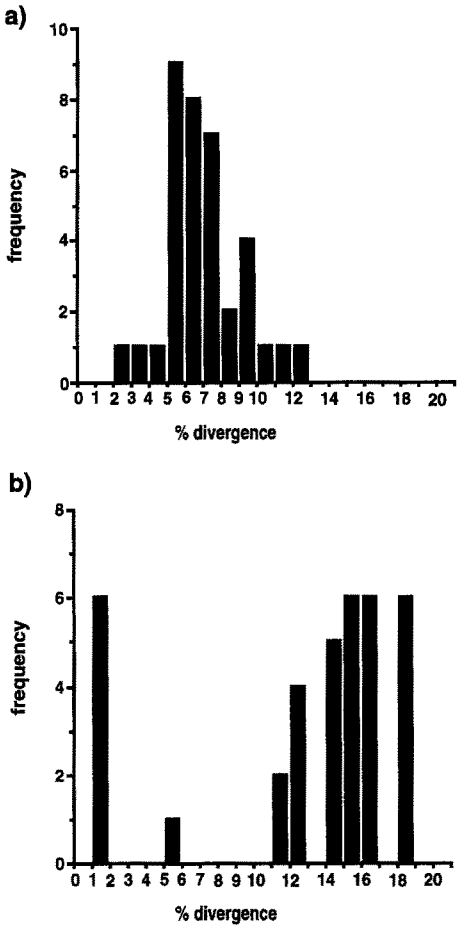

Fig. 4. Distribution of pairwise nucleotide sequence comparisons between the *adk* (**A**) and *recA* (**B**) genes of the *Neisseria* species.

informative silent sites. The results for the *recA* data were not significant, irrespective of whether all silent polymorphic sites or only informative polymorphic sites were included in the analysis. However, the same test on the *adk* data was highly significant when only informative silent sites were included in the analysis [$P = 0.006$, based on the sums of squares of condensed fragments (SSCF); Sawyer 1989]. Although not evident from a visual inspection of the data, this result argues that the sequences contain a much higher number of "runs" of identical polymorphic sites than would be expected by chance, which suggests a history of frequent recombination between the *adk* genes of these species. When the noninformative polymorphic sites were included in the analysis, the P value of the *adk* data set rose to 0.04. The apparent homogenization of the *adk* sequences, as revealed by phylogenetic analysis and an examination of the pairwise distances, argues that recombination has primarily occurred inclusively to the data set, i.e., between the species analyzed. Noninformative polymorphisms (i.e., those unique to a single sequence), which probably represent point mutations, or possibly rare recombinational replacements from species exclusive to the data set, will therefore weaken the significance of the data by interrupting runs.

If the process which has been influencing the divergence of a sequence most profoundly is stochastic mutation, rather than recombination, then one should expect very high levels of linkage disequilibrium between polymorphic codons. However, if sequences have encountered an evolutionary history most profoundly influenced by recombination, then codons will be more randomly assorted between the sequences and the level of linkage between codons will decrease. The I_A test was carried out on the *recA* sequences and, unsurprisingly, this revealed highly significant levels of linkage (0/1,000 simulations resulted in a V_O greater than observed with the real data). The same was also found with the *adk* data when taken as a whole.

When the *adk* sequence from *N. animalis* was removed from the analysis, I_A decreased from 3.47 (a greater variance was observed in 0/1,000 simulations) to 1.389 (where 65/1,000 simulations resulted in a greater variance than the real data), implying relatively low levels of linkage disequilibrium between codons for the human *Neisseria* species. Conversely, the *recA* data was still in highly significant linkage disequilibrium, even after the exclusion of the sequence from *N. animalis.* This result supports the suggestion of a high rate of interspecific recombination between the human *Neisseria* species at *adk* but not *recA*. Furthermore, this test provides circumstantial evidence that the nonhuman commensal, *N. animalis,* which is assumed to have been ecologically isolated, has, at least in relatively recent times, been recombining less frequently at *adk* with the human *Neisseria* species.

The analysis was taken further by dividing the 96 polymorphic codons within the complete *adk* data set into four regions, each of 24 polymorphic codons. I_A was then calculated for each of the four regions in turn. The *N. animalis* sequence was then removed from the analysis and the remaining 79 polymorphic codons were divided into three regions of 20 polymorphic codons, and one of 19 polymorphic codons. I_A was then calculated for all of these regions. There are a total of 197 polymorphic codons within the *recA* data set. V_E increases with r and so, in order to make an appropriate comparison with the *adk* data set, I_A was calculated from four regions of the *recA* data, each of 20 polymorphic codons, rather than by dividing the entire data set of *recA* into four regions, as was done for *adk.*

Figure 6 shows the I_A values for the four regions of the *adk* and *recA* sequence data sets (with and without the sequence from *N. animalis*). The I_A values for the *adk* data (without *N. animalis*) show a decline moving from the promoter-proximal to the promoter-distal part of the *adk* coding region, reflecting lower levels of linkage between codons. The final 20 polymorphic codons of the *adk* sequence (which represent a total of 54 codons) give an I_A value of −0.04, which suggests that there is no linkage between the polymorphic codons in this region,

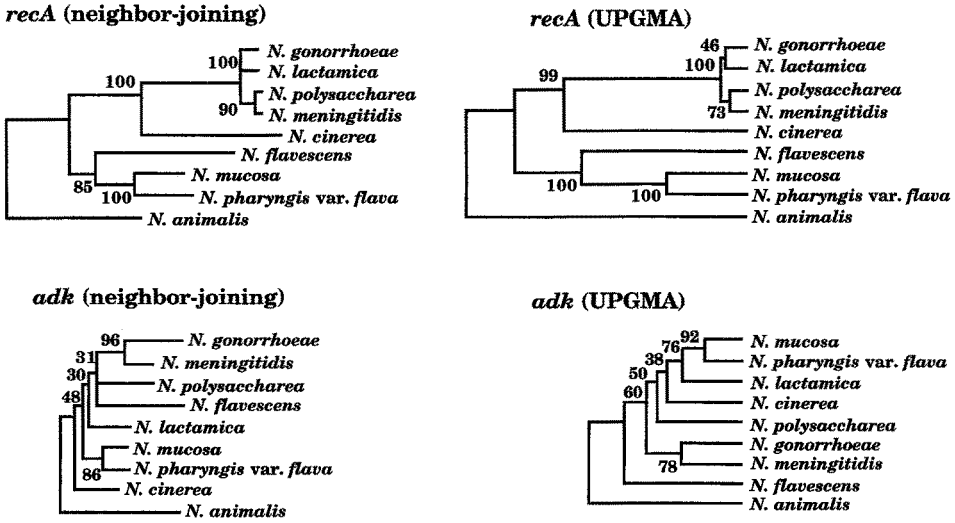

Fig. 5. Dendrograms constructed from the *adk* and *recA* sequences using the neighbor-joining and UPGMA methods. The values at the nodes are the percentages of the 500 bootstrap resamplings in which the species to the right of the node were separated from the other species.

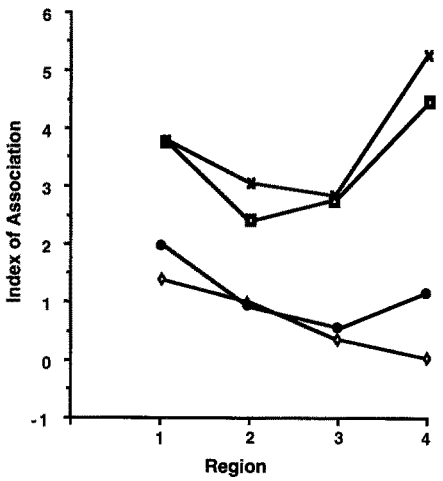

Fig. 6. Linkage between polymorphic codons within the *adk* and *recA* genes. The values of I_A were determined for four regions of the *adk* and *recA* genes from the *Neisseria* species. ●, *adk* genes from all species; ◇, *adk* genes without *N. animalis;* □, *recA* genes from all species; ✖, *recA* genes without *N. animalis.*

i.e., they are randomly assorted. However, with the inclusion of the sequence from *N. animalis,* the level of linkage is raised, to give an I_A value of 1.05. The difference in linkage disequilibrium is reflected in the simulation results for this region at 43/1,000 trials gave a greater variance than the real data when the *N. animalis* sequence was included compared to 484/1,000 trials when it was omitted. These results therefore argue that while the polymorphic codons in this region of *adk* are randomly assorted between the species isolated from human hosts, the polymorphic codons within the *adk* sequence from *N. animalis* are linked within this species. The inclusion of the data from *N. animalis* also increased the level of linkage within the first region of the *adk* gene (as well as the gene as a whole), but did not increase linkage within the second and third regions of the gene.

The *adk* data thus show evidence of recombination between the human *Neisseria* species, particularly within the downstream region of the gene. Conversely, the data for *recA* provides evidence for very high levels of linkage disequilibrium, within each of the four sample regions of the gene, as well as in the gene as a whole, irrespective of whether *N. animalis* is included in the analysis or not (Fig. 6).

Discussion

This paper provides evidence which suggests that the patterns of sequence divergence between the same *Neisseria* species are very different at *recA* and *adk,* despite the fact that both of these genes encode housekeeping proteins, and would therefore be expected to be subject to similar selection pressures. The sequence divergence between the nine species at *recA* showed no great anomalies. The dendrograms produced using the *recA* data appeared to be robust and did not change significantly when different algorithms were used. The phylogeny was also highly consistent with that obtained from previous studies. For example, the grouping of the pathogenic species *N. gonorrhoeae* and *N. meningitidis* with *N. lactamica* and *N. polysaccharea* is congruent with the results from DNA-DNA hybridization experiments (Guibourdenche et al. 1986) and sequence analysis of the *arg*F and *pen*A genes (Zhou and Spratt 1992; Bowler et al. 1994). This grouping is also supported by Barrett and Sneath's (1994) phenotypic analysis, which placed all these species within "Area A" of the dendrogram. *N. cinerea* was also within Area A and was the next most closely related species using the *recA* data. The grouping of *N. mucosa* with *N. pharyngis* var. *flava* is also supported by Barrett and Sneath's study (both are placed in Area B characterized by pharyngeal strains). *N. flave-*

scens and *N. animalis* are both placed in the more heterogeneous Area C.

The *adk* data showed several anomalies. Firstly, most of the pairwise distances between the *adk* sequences fall within a narrow range. Secondly, the dendrograms derived from the *adk* data using two different algorithms were grossly noncongruent with each other and with the dendrograms derived from the *recA* data. The low bootstrap values for the nodes using the *adk* data indicate that a consistent phylogeny could not be obtained for these nine species. Finally, although the d_S/d_N ratios suggest that *adk* is evolving more slowly than *recA,* the nucleotide sequence divergence between the two pathogenic *Neisseria* species at *adk* was over three times that observed for *recA,* and over twice that found in any other housekeeping gene that has been studied.

One explanation for these anomalies is that the expected patterns of sequence divergence between the *Neisseria* species have been obscured by a history of frequent recombination at *adk* but not at *recA.* Frequent recombination would act both to homogenize the *adk* sequences, resulting in the unimodal distribution of pairwise sequence comparisons and the unusually large divergence between the pathogenic *Neisseria* species, and would engender severe phylogenetic distortion. The assumption of a history of interspecies recombination at *adk* but not at *recA* is supported by statistical analysis of the sequences. Sawyer's runs test was significant for the *adk* data set, but not for the *recA* data set. The I_A values also revealed much lower levels of linkage between codons within the *adk* data set than in the *recA* data set (particularly within the promoter-distal region of the gene). Interestingly, inter- and intraspecies recombination at *adk* has recently been proposed to have produced the minor electrophoretic variants of ADK found in some *N. meningitidis* strains (Feil et al. 1995). The reason for the higher rate of interspecies recombination at *adk* compared to *recA* is unclear. The *adk* gene is not under diversifying selection, and is unlikely to have been influenced by hitchhiking at a neighboring locus under diversifying selection, as the average diversity among the meningococcal *adk* gene (excepting the rare strains expressing electrophoretic variants) was only 1.1% (Feil et al. 1995), and no polymorphisms have been found in the *adk* genes of ten gonococci that have been examined (E.F. and B.G.S., unpublished).

There is also evidence from the I_A values that the ecologically isolated *N. animalis* has been recombining to a lesser degree with the human species as the inclusion of the data from *N. animalis* appeared to increase the level of linkage in the first and final region of *adk* but not in the middle two regions. These differences may reflect effectively random influences resulting from the distribution of polymorphisms within the *adk* sequences or differences in the selective constraint along the length of the coding region. However, the results from the I_A test were not highly significant regarding the absence, or reduced level, of recombination between *N. animalis* and the human species, and Sawyer's runs test was unable to provide any support, so this evidence is not decisive.

The *adk* gene is more highly conserved between the *Neisseria* species than *recA.* For example, the highest level of nucleotide divergence between the nine species at *adk* is 12.3%, but there are 28 pairwise comparisons at *recA* (78%) which exceed this value. The reduced variation at *adk* may be due in part to the homogenizing effect of recombination, but also to stronger selection against nonsynonymous substitutions as the d_S/d_N ratio at *adk* was twice that at *recA.* The constraints against nonsynonymous substitutions and the homogenizing effect of recombination may be causally linked, as it is well documented from work on *Neisseria* sp. (Frosch and Meyer 1992), and from work on *Bacillus* sp. (Roberts and Cohan 1993), that the efficiency of transformation in the laboratory decreases as the degree of sequence divergence between the parental molecules increases. A gene which is under unusually high stabilizing selection will be more homogenous than one under less constraint, making it likely to encounter intragenic recombinational replacements at a higher frequency. Recombination will in turn act to homogenize the sequences, tempering the divergence between them, and so making subsequent recombination even more likely. This process will obscure the divergence of species through stochastic mutation and will render meaningless the use of the gene to determine the phylogenetic relationships between species. It is possible that such a process may be affecting the evolution of *adk* between the eight species isolated from humans, but not *N. animalis,* because it is ecologically isolated. Presumably, the relatively relaxed selection at *recA* allowed the sequences to obtain the threshold level of divergence necessary to escape from this process, thus allowing divergence to saturation. However, variation in the rates of evolution of different genes is likely to be only one factor in the observed differences in the extent of interspecies recombination in *Neisseria* genes. Variations in the absolute rates of recombination within genes (for example, due to the location of uptake sequences; Graves et al. 1982) and in the strength of selection for or against the recombinants are also likely to contribute.

The surprising finding of a high level of interspecies recombination within *adk* but not *recA* has important implications for the microbial species concept. Dykhuizen and Green (1991) have suggested a method by which microbial species can be defined, and the premise of this method rests on the biological species concept. They suggest that if recombination is frequent within a single species then the phylogenies of different strains within the species will be discordant because of the effects of recombination. However, the gene pool within a species should still represent a distinct phylogenetic cluster when compared to other species if recombination is very rare

or absent between species. Thus the phylogenies of single genes from individuals of different species should not be significantly different. This definition therefore categorizes species as recombinationally isolated entities, although recombination within a species may be frequent.

Regarding this definition, Dykhuizen and Green (1991) wrote: "a possible difficulty is seen with the various *Neisseria* species," noting that within this genus, recombination can occur between relatively distantly related individuals. The present work is significant because it implies that recombination can have a profound effect on the divergence of a particular locus even when it is not believed to be under strong diversifying selection pressures, and because it suggests that recombination rates within such genes can vary widely within a single genome. Cohan (1995) has argued that because bacterial recombination is a much less cohesive force than eukaryotic sex (as it occurs less frequently and has more localized effects), sexual isolation is not necessarily a prerequisite for permanent sequence divergence between bacteria. Sequence data has been generated for both *adk* and *recA* within populations of both *N. gonorrhoeae* and *N. meningitidis* (Zhou and Spratt 1992; Feil et al. 1995; E.F. and B.G.S., unpublished), and this has revealed that the level of diversity within these named species is much less than the degree of divergence between the named species revealed by the present work. It therefore seems likely that the sequences reported here represent distinct sequence clusters despite the fact that recombination between the species may be very frequent at *adk*. Although the *Neisseria* species investigated in this work do not correspond to distinct species under Dykhuizen and Green's definition (at least at *adk*), they are therefore likely to correspond to distinct species according to the definition suggested by Cohan (1995)—that species represent permanent sequence clusters. However, more detailed studies of other housekeeping genes are required to establish the extent to which interspecies recombination distorts the phylogeny of the human *Neisseria* species.

Acknowledgments. This work was supported by the Wellcome Trust. B.G.S. is a Wellcome Trust Principal Research Fellow. E.F. is a Wellcome Trust Prize Fellow.

References

Achtman M (1994) Clonal spread of serogroup A meningococci: a paradigm for the analysis of microevolution in bacteria. Mol Microbiol 11:15–22

Barrett SJ, Sneath PHA (1994) A numerical phenotypic taxonomic study of the genus *Neisseria*. Microbiol 140:2867–2891

Bowler LD, Zhang Q-Y, Riou J-Y, Spratt BG (1994) Inter-species recombination between the *penA* genes of *Neisseria meningitidis* and commensal *Neisseria* species during the emergence of penicillin resistance in *N. meningitidis:* natural events and laboratory simulation. J Bacteriol 176:333–337

Boyd EF, Nelson K, Wang F-S, Whittam TS, Selander RK (1994) Molecular genetic basis of allelic polymorphism in malate-dehydrogenase (*mdh*) in natural-populations of *Escherichia coli* and *Salmonella enterica*. Proc Natl Acad Sci USA 91:1280–1284

Brown AHD, Feldman MW, Nevo E (1980) Multilocus structure of natural populations of *Hordeum spontanium*. Genetics 96:523–536

Cohan FM (1995) Does recombination constrain neutral divergence among bacterial taxa. Evolution 49:164–175

Dykhuizen DE, Green L (1991) Recombination in *Escherichia coli* and the definition of biological species. J Bacteriol 173:7257–7268

Feil E, Carpenter G, Spratt BG (1995) Electrophoretic variation in adenylate kinase of *Neisseria meningitidis* is due to inter- and intra-species recombination. Proc Natl Acad Sci USA 92:10535–10539

Frosch M, Meyer TF (1992) Transformation-mediated exchange of virulence determinants by co-cultivation of pathogenic *Neisseria*. FEMS Microbiol Lett 100:345–350

Fyfe JAM, Davies JK (1990) Nucleotide sequence and expression in *Escherichia coli* of the *recA* gene of *Neisseria gonorrhoeae*. Gene 93:151–156

Graves JF, Biswas GD, Sparling PF (1982) Sequence-specific DNA uptake in transformation of *Neisseria gonorrhoeae*. J Bacteriol 152: 1071–1077

Guibourdenche M, Popoff MY, Riou JY (1986) Deoxyribonucleic acid relatedness among *Neisseria gonorrhoeae, N. meningitidis, N. lactamica, N. cinerea* and "*Neisseria polysaccharea.*" Ann Inst Pasteur Microbiol 137B:177–185

Guttman DS, Dykhuizen DE (1994) Clonal divergence in *Escherichia coli* as a result of recombination, not mutation. Science 266:1380–1383

Higuchi RG, Ochman H (1989) Production of single-stranded DNA templates by exonuclease digestion following the polymerase chain reaction. Nucleic Acids Res 17:5865

Hobbs MM, Seiler A, Achtman M, Cannon JG (1994) Microevolution within a clonal population of pathogenic bacteria—recombination, gene duplication and horizontal genetic exchange in the *opa* gene family of *Neisseria meningitidis*. Mol Microbiol 12:171–180

Istock CA, Duncan KE, Ferguson N, Zhou X (1992) Sexuality in a natural population of bacteria—*Bacillus subtilis* challenges the clonal paradigm. Mol Ecol 1:95–103

Knapp JS (1988) Historical perspectives and identification of *Neisseria* and related species. Clin Microbiol Rev 1:415–431

Kumar S, Tamura K, Nei M (1993) MEGA: Molecular Evolutionary Genetics Analysis software for microcomputers. Comput Appl Biosci 10:189–191

Jukes TH, Cantor CR (1969) Evolution of protein molecules. In: Munro HN (ed) Mammalian protein metabolism. Academic Press, New York, pp 21–132

Maiden MCJ (1993) Population genetics of a transformable bacterium: the influence of horizontal genetic exchange on the biology of *Neisseria meningitidis*. FEMS Microbiol Lett 112:243–250

Maynard Smith J, Dowson CG, Spratt BG (1991) Localised sex in bacteria. Nature 349:29–31

Maynard Smith J, Smith NH, O'Rourke M, Spratt BG (1993) How clonal are bacteria? Proc Natl Acad Sci USA 90:4384–4388

Nei M, Gojobori T (1986) Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol Biol Evol 3:418–426

Nelson K, Selander RK (1994) Intergeneric transfer and recombination of the 6-phosphogluconate gene (*gnd*) in enteric bacteria. Proc Natl Acad Sci USA 91:10227–10231

Olyhoek T, Crowe BA, Achtman M (1987) Clonal population structure of *Neisseria meningitidis* serogroup A isolated from epidemics and pandemics between 1915 and 1983. Rev Infect Dis 9:665–682

O'Rourke M, Stevens E (1993) Genetic structure of *Neisseria gonorrhoeae* populations: a non-clonal pathogen. J Gen Microbiol 139: 2603–2611

Reeves PR, Farnell L, Lan R (1994) Multicomp: a program for preparing sequence data for phylogenetic analysis. Comput Appl Biosci 10:281–284

Roberts MS, Cohan FM (1993) The effect of DNA sequence divergence on sexual isolation in *Bacillus*. Genetics 134:402–408

Saitou N, Nei M (1987) The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol Biol Evol 4:406–425

Sawyer S (1989) Statistical tests for detecting gene conversion. Mol Biol Evol 6:526–538

Spratt BG, Zhou J (1994) Recombination in nature between chromosomal genes of pathogenic and commensal *Neisseria* species. In: Conde-Glez CJ, Morse S, Rice P, Sparling F, Calderon E (eds) Pathobiology and immunobiology of *Neisseriaceae*. INSP, Cuernavaca, Mexico, pp 575–581

Spratt BG, Bowler LD, Zhang Q-Y, Zhou J, Maynard Smith J (1992) Role of inter-species transfer of chromosomal genes in the evolution of penicillin resistance in pathogenic and commensal *Neisseria* species. J Mol Evol 34:115–125

Spratt BG, Smith NH, Zhou J, O'Rourke M, Feil E (1995) The population genetics of the pathogenic *Neisseria*. In: Baumberg S, Young JPW, Wellington EMH, Saunders JR (eds) The population genetics of bacteria. SGM Symposium No. 52, Cambridge University Press, Cambridge, pp 143–160

Vázquez JA, de la Fuente L, Berrón S, O'Rourke M, Smith NH, Zhou J, Spratt BG (1993) Ecological separation and genetic isolation of *Neisseria gonorrhoeae* and *Neisseria meningitidis*. Curr Biol 3: 567–572

Vázquez JA, Berrón S, O'Rourke M, Carpenter G, Feil E, Smith NH, Spratt BG (1995) Interspecies recombination in nature: a meningococcus that has acquired a gonococcal PIB porin. Mol Microbiol 15:1001–1007

Ward MJ, Lambden PR, Heckels JE (1992) Sequence analysis and relationships between meningococcal class 3 serotype proteins and other porins from pathogenic and non-pathogenic *Neisseria* species. FEMS Microbiol Lett 94:283–290

Zhou J, Spratt BG (1992) Sequence diversity within the *argF*, *fbp* and *recA* genes of natural isolates of *Neisseria meningitidis:* interspecies recombination within the *argF* gene. Mol Microbiol 6:2135–2146