Science 12 June 2009:
Vol. 324. no. 5933, pp. 1454 - 1457
DOI: 10.1126/science.1171908

REPORTS

Hyper-Recombination, Diversity, and Antibiotic Resistance in Pneumococcus

William Paul Hanage,[1,*] Christophe Fraser,[1] Jing Tang,[2] Thomas Richard Connor,[1] Jukka Corander[2]

***Streptococcus pneumoniae* is a pathogen of global importance that frequently transfers genetic material between strains and on occasion across species boundaries. In an analysis of 1930 pneumococcal genotypes from six housekeeping genes and 94 genotypes from related species, we identified mosaic genotypes representing admixture between populations and found that these were significantly associated with resistance to several classes of antibiotics. We hypothesize that these observations result from a history of hyper-recombination, which means that these strains are more likely to acquire both divergent genetic material and resistance determinants. This could have consequences for the reemergence of drug resistance after pneumococcal vaccination and also for our understanding of diversification and speciation in recombinogenic bacteria.**

[1] Department of Infectious Disease Epidemiology, Imperial College London, Norfolk Place, London W2 1PG, UK.
[2] Department of Mathematics, Åbo Akademi, FI-20500, Turku, Finland.

[*] To whom correspondence should be addressed. E-mail: w.hanage{at} imperial.ac.uk

Many bacteria undergo homologous recombination, in which short tracts of DNA in the recipient are replaced by the corresponding tract from a donor strain, resulting in a mosaic of DNA from different ancestors (*1*). Although this occurs mainly within species and declines markedly with increasing sequence divergence between donor and recipient (*2*), occasional gene transfers between species do occur. Such events have the potential to introduce new phenotypes, such as virulence or antibiotic resistance, into a new genetic background that may or may not be the same as the species of the donor strain (*3*–*7*) and may have considerable impacts on bacterial evolution and human health.

One group in which homologous recombination is frequent is the mitis group streptococci. This includes the major human pathogen *Streptococcus pneumoniae*,

the pneumococcus, which is responsible for at least 1 million deaths per year worldwide (8). The closely related species *S. oralis*, *S. mitis*, and *S. pseudopneumoniae* (among others) have a history of taxonomic confusion, which may be partly explained by genetic diversity within the mitis group (9, 10). Moreover, rare but important events have led to the acquisition of antibiotic resistance by pneumococcus as a result of the transfer of resistance determinants across species boundaries (4, 5). The high rates of recombination within the species have the potential to shuffle resistance determinants among pneumococcal genotypes. It is not known whether or not recombination, either at resistance loci or housekeeping genes, is equally likely for all members of the species or whether some strains are more likely to be involved in this process.

Although a vaccine is available for 7 of the more than 90 pneumococcal serotypes, this has not eliminated pneumococcal disease because the nonvaccine serotypes derive an ecological advantage from the removal of their competitors and have been increasing in carriage prevalence (11) and, concomitantly, in disease (12). Alongside this, we are observing the expansion of existing antibiotic-resistant clones with nonvaccine serotypes and the possible emergence of new ones (13, 14).

Multilocus sequence typing (MLST) (15) supplies genetic data to study recombination and population structure in the pneumococcal population. MLST characterizes an isolate by sequencing internal fragments of seven housekeeping genes. Together these define the sequence type (ST) of the isolate, which may readily be compared to others through the MLST database (16). This contains sequence data from many thousands of isolates reported by the global community of MLST users. It also contains associated epidemiological data including serotype and antibiotic resistance. We have previously published the sequences of MLST loci from multiple isolates of related species, allowing us to study recombination between them and the pneumococcus (9). The *ddl* locus used in MLST for pneumococcus is associated with a high frequency of interspecies gene transfer because of physical linkage with the penicillin binding protein (PBP) 2b locus (17), at which alleles containing DNA originating in other species lead to penicillin resistance. Because the linkage could bias any estimates of admixture, we have excluded this locus from the following analysis. Once the *ddl* locus is removed, the data set consists of 1930 distinct genotypes of *S. pneumoniae*, along with 40 identified as *S. mitis*, 39 *S. pseudopneumoniae*, and 15 *S. oralis* (9).

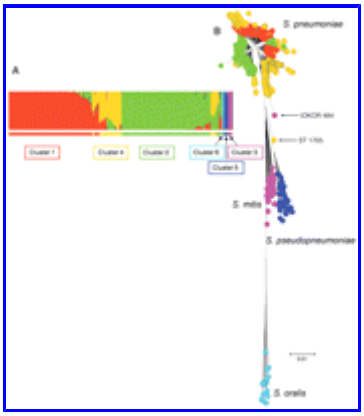
To identify populations and rates of admixture between them, we used the program Bayesian Analysis of Population Structure [BAPS (18–20)]. This program, freely available online (21), implements several models to identify clusters characterized by different allele frequencies within a population characterized by multilocus DNA sequences. Furthermore, cases of likely admixture, that is, isolates containing a DNA sequence characteristic of more than one population as a result of recombination, can be identified (22).

The results of the analysis of the streptococcal data set are summarized in Table 1 and Fig. 1 [and presented in more detail in (22)]. In total, six clusters were identified, three of which corresponded to the nonpneumococcal species (Table 1). The remaining three clusters (1, 2, and 4) represent subpopulations of the pneumococcus as defined within the BAPS analysis. Figure 1A shows the admixture graphic, in which each unique genotype is represented by a column, colored according to the proportion of sequence assigned to each cluster. For clusters 1 and 2, the vast majority of genotypes were characteristic of only one cluster. The reverse was true of cluster 4, which was mostly composed of mosaics. In Fig. 1B, which displays the clustering of these groups by using a phylogenetic tree, this is evident in

the scattering of cluster 4 genotypes around the pneumococcal cluster. Two anomalous genotypes were evident and are indicated in Fig. 1B: one assigned to cluster 4 arising from the branch leading to *S. pseudopneumoniae* and *S. mitis* (this is ST 1705 and is a pneumococcal strain containing multiple divergent alleles) and another highly divergent genotype assigned to cluster 3 at the end of a long branch arising from within the *S. pneumoniae* cluster. This strain [IOKOR 484 as described in (*23*)] was previously considered to be an example of *S. pseudopneumoniae* but in this analysis clustered with *S. mitis* strains. These illustrate the difficulty of assigning strains to species in such recombinant taxa.

View this table: [in this window] [in a new window]

Table 1. Association between named species and BAPS cluster. Blank fields indicate that no strains in the relevant cluster were identified as that species.

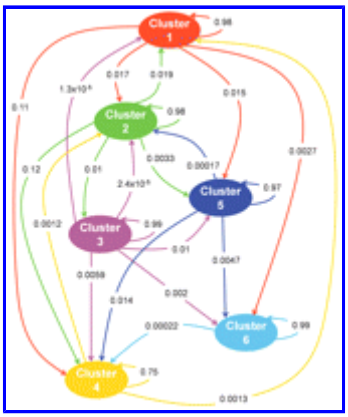


View larger version (19K): [in this window] [in a new window]

Fig. 1. (**A**) Admixture analysis of 2024 distinct streptococcal genotypes. Each column represents a single multilocus genotype, colored according to the proportion of genetic variation assigned to each cluster. The final cluster assignment is shown by the color of the line underneath. For example, a solid red column signifies a genotype for which 100% of the sequence was characteristic of group 1. Note that the variation characteristic of a cluster is not necessarily contiguous in the concatenated sequences used in the analysis. The relationships between the groups and named species are presented in Table 1. Group 4 strains are composed of a high proportion of mosaics. (**B**) Minimum evolution tree constructed by using MEGA4 (*28*) from concatenates of *S. pneumoniae* strains and related species as described in the text. The optimal tree is shown with distances computed by using the Kimura two-parameter method. Taxa are colored by BAPS group according to the same scheme used in (A). The relationships between BAPS clusters and named species are indicated and are presented in Table 1. The positions of the anomalous strains ST 1705 and IOKOR 484 are indicated.

We estimated the relative amounts of admixture between the clusters (*22*) (Fig. 2). This estimate shows cluster 4 to be a recipient of genetic information from the other clusters and other species. Hence for individual loci, we identified alleles that are divergent from typical pneumococcal alleles and are more similar to those found in

related species. Of 93 individual genotypes containing alleles that cluster with nonpneumococcal species (*22*), all but one were found in cluster 4.



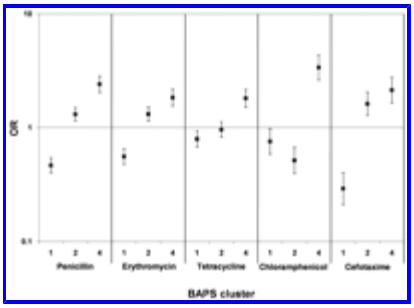
View larger version (30K):
[in this window]
[in a new window]

Fig. 2. Admixture between the clusters illustrated in Fig. 1. Arrows indicate the average fraction of sequence variation obtained from the source cluster by the strains assigned to the target cluster. Circular loops indicate the fraction of variation estimated as not arising from outside the cluster.

The atypical genotypes in cluster 4 reflect a history of recombination; hence, we examined the association between cluster 4 and antibiotic resistance. We used 3732 records deposited in the MLST database, which supplied the input data for the BAPS analysis and contains data on resistance to penicillin, erythromycin, tetracycline, chloramphenicol, and cefotaxime. We categorized an isolate as nonsusceptible if it was recorded as having a minimal inhibitory concentration (MIC) greater than the breakpoint for susceptibility according to British Society for Antimicrobial Chemotherapy or if it was recorded as resistant or of intermediate resistance [i.e., nonsusceptible in accordance with customary usage (*22*)].

Odds ratios (ORs) for the association of resistance with the clusters inferred by BAPS showed (Fig. 3) a significant association with cluster 4 and a negative relationship with cluster 1. Cluster 2 is intermediate in all cases except for chloramphenicol. This may be linked to the fact that cluster 2 contains a high proportion of strains with serotypes included in current conjugate vaccine formulations (Table 2). These serotypes have been previously known to be associated with resistance, and their removal after vaccination has led to a decline in resistance (*14*). In contrast, cluster 4 shows no association with these serotypes. The association between cluster 4 and nonsusceptibility is also robust to different criteria for associating genotypes with resistance (*22*).

Fig. 3. ORs showing the association between BAPS cluster and nonsusceptibility to five antibiotics. ORs were calculated relative to all other clusters as described in the text. Error bars indicate 95% confidence intervals.



View larger version (12K):
[in this window]
[in a new window]

View this table: [in this window] [in a new window]

Table 2. Associations between pneumococcal isolates with vaccine serotypes and BAPS clusters. The vaccine serotypes are those present in the seven valent pneumococcal vaccines (4, 6B, 9V, 14, 18C, 19F, and 23F), and all records present in the MLST database at the time of BAPS analysis were used to estimate ORs. Reestimation using records entered into the database since the initial analysis did not substantially alter the results (*22*).

The reasons cluster 4 associated with both resistance and mosaicism are unclear. Resistance to the different antibiotics may arise by the acquisition of additional loci (such as efflux pumps in the case of erythromycin resistance) as well as gene conversion–like processes (e.g., homologous recombination at PBP loci for β-lactam nonsusceptibility). Both involve the acquisition of DNA from other lineages. We have considered the possibility that this result could be a consequence of reverse causation: Resistance leads to strains, for some reason, being more likely to be grouped in cluster 4. However, any artifactual association between this cluster and antibiotic resistance seems very unlikely given that the BAPS classification and resistance phenotype are based on unlinked loci.

We hypothesize that cluster 4 is an amalgam of strains with a history of hyper-recombination, which leads to them being grouped together by BAPS because they share anomalous DNA sequences. By having a history of hyper-recombination, such strains are more likely to accept divergent DNA, both at housekeeping and resistance loci, and hence more likely to acquire resistance and housekeeping gene sequences from distantly related pneumococci or other species.

Within many bacterial populations and in particular those under strong antibiotic selective pressure, we can identify strains with an elevated mutation rate (*24*) usually through defects in the mismatch repair (MMR) system. In certain contexts [e.g., the cystic fibrotic lung (*25*)], it is thought that second-order selection on the resistant strains that arise in the mutator lineages leads to strains with a mutator phenotype being more common than predicted (*26*). However, the relationship between elevated recombination and antibiotic resistance is not well understood. Elevated mutation and/or recombination rates should carry a fitness cost, and as a result a

high rate of reversion to wild type is predicted, possibly by horizontal acquisition of wild-type MMR genes, as has been proposed for hypermutators in *Escherichia coli* (*27*).

Through analysis of MLST data collected for epidemiological purposes, we have identified pneumococcal strains showing evidence of past recombination, both with other pneumococci and with related species. These strains are significantly more likely to be resistant to all classes of antibiotics for which data are available. Because the resistance mechanisms involved include both homologous and illegitimate recombination, this implies a general tolerance for foreign DNA, suggesting a hyper-recombinant state. It is reasonable to suggest that this state could be important for adaptation to other environmental pressures beyond antibiotics. This demonstrates the importance of recombination in bacterial evolution over the long term and suggests that it may vary markedly within a species. The consequences for speciation and adaptation remain to be determined.

Supporting Online Material

www.sciencemag.org/cgi/content/full/324/5933/1454/DC1

Materials and Methods

Fig. S1

Tables S1 to S3

References

References and Notes

1. E. J. Feil, B. G. Spratt, *Annu. Rev. Microbiol.* **55**, 561 (2001). [CrossRef] [ISI] [Medline]
2. C. Fraser, W. P. Hanage, B. G. Spratt, *Science* **315**, 476 (2007). [Abstract/Free Full Text]
3. R. Hakenbeck, *Electrophoresis* **19**, 597 (1998). [CrossRef] [ISI] [Medline]
4. C. G. Dowson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **87**, 5858 (1990). [Abstract/Free Full Text]
5. C. G. Dowson, T. J. Coffey, C. Kell, R. A. Whiley, *Mol. Microbiol.* **9**, 635 (1993). [ISI] [Medline]
6. C. G. Dowson *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8842 (1989). [Abstract/Free Full Text]
7. B. G. Spratt *et al*., *Proc. Natl. Acad. Sci. U.S.A.* **86**, 8988 (1989). [Abstract/Free Full Text]
8. K. Mulholland, *Lancet* **370**, 285 (2007). [CrossRef] [ISI] [Medline]
9. W. P. Hanage, C. Fraser, B. G. Spratt, *Philos. Trans. R. Soc. London Ser. B* **361**, 1917 (2006). [Abstract/Free Full Text]
10. C. J. Bishop *et al*., *BMC Biol.* **7**, 3 (2009). [CrossRef] [Medline]
11. S. S. Huang *et al*., *Pediatrics* **116**, e408 (2005). [Abstract/Free Full Text]
12. L. A. Hicks *et al*., *J. Infect. Dis.* **196**, 1346 (2007). [CrossRef] [ISI] [Medline]
13. A. B. Brueggemann, R. Pai, D. W. Crook, B. Beall, *PLoS Pathog.* **3**, e168 (2007).

[CrossRef] [Medline]

14. M. H. Kyaw *et al*., *N. Engl. J. Med.* **354**, 1455 (2006).[Abstract/Free Full Text]
15. M. C. Enright, B. G. Spratt, *Microbiology* **144**, 3049 (1998). [Abstract/Free Full Text]
16. http://spneumoniae.mlst.net/
17. M. C. Enright, B. G. Spratt, *Mol. Biol. Evol.* **16**, 1687 (1999).[Abstract]
18. J. Corander, P. Waldmann, P. Marttinen, M. J. Sillanpaa, *Bioinformatics* **20**, 2363 (2004). [Abstract/Free Full Text]
19. J. Corander, P. Waldmann, M. J. Sillanpaa, *Genetics* **163**, 367 (2003). [ISI] [Medline]
20. J. Corander, J. Tang, *Math. Biosci.* **205**, 19 (2007). [CrossRef] [ISI] [Medline]
21. http://www.abo.fi/fak/mnf/mate/jc/smack_software_eng.html
22. Materials and methods are available as supporting material on *Science* Online.
23. W. P. Hanage *et al*., *J. Bacteriol.* **187**, 6223 (2005). [Abstract/Free Full Text]
24. E. Denamur, I. Matic, *Mol. Microbiol.* **60**, 820 (2006). [CrossRef] [ISI] [Medline]
25. R. del Campo *et al*., *J. Clin. Microbiol.* **43**, 2207 (2005). [Abstract/Free Full Text]
26. M. C. Negri *et al*., *Antimicrob. Agents Chemother.* **46**, 528 (2002). [Abstract/Free Full Text]
27. E. Denamur *et al*., *Cell* **103**, 711 (2000). [CrossRef] [ISI] [Medline]
28. K. Tamura, J. Dudley, M. Nei, S. Kumar, *Mol. Biol. Evol.* **24**, 1596 (2007). [Abstract/Free Full Text]
29. We acknowledge the support of the Royal Society (W.P.H. and C.F.); the Computational Biology, Bioinformatics, and Biometry graduate school (J.T.); a UK Biotechnology and Biological Sciences Research Council studentship (to T.R.C.); and the Academy of Finland (J.C., grant no. 121301). We thank B. Spratt for helpful discussions.

Received for publication 5 February 2009. Accepted for publication 7 May 2009.

Science. ISSN 0036-8075 (print), 1095-9203 (online)