Clonal analysis of non-typable *Haemophilus influenzae* by sodium dodecyl sulphate-polyacrylamide gel electrophoresis of whole cell polypeptides

K. D. BRUCE and T. H. PENNINGTON

Department of Medical Microbiology, University of Aberdeen Medical School, Foresterhill, Aberdeen AB9 2ZD

**Summary.** A collection of 223 strains of non-typable *Haemophilus influenzae* was assembled. Most strains were isolated from hospital in-patients in North East Scotland between 1984 and 1989. These isolates were examined by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) of whole cell polypeptides. Variability was assessed in terms of apparent molecular weight differences between the protein profiles. Isolates were grouped on the basis of Dice coefficients of similarity and assigned to clones. Of the 223 strains, 147 were unique; the remaining strains were assigned to multi-member clones of which 13 clones had two members, one clone had three members, three clones had four members, three clones had five members, two clones had six members and one clone had eight members. The longest time interval between isolation of clonally related strains virtually equalled the limits of the study. Isolates from a children's hospital were significantly less diverse than those from patients in adult wards.

**Introduction**

Evidence is accumulating that in many bacterial species the recombination rates of chromosomal genes are very low, indicating a clonal population structure. The evidence for this is largely circumstantial, but is strong. It rests on the observations that the same genotypes of organisms have been isolated worldwide and over long periods, that strong non-random associations of alleles over loci (linkage disequilibrium) occur, and, for *Escherichia coli*, of direct evidence of very low recombination rates in chemostat populations. In the short term, the evolution of these species is likely to be dominated by the stochastic extinction of clones and periodic selection in which neutral, or even slightly deleterious, alleles “hitch-hike” with favourable mutations selected by fitness.

*Haemophilus influenzae* is an important human pathogen. Capsulate strains are a frequent cause of meningitis and other invasive infections in children. Non-capsulate (non-typable) strains cause a variety of clinically important infections and are found in the respiratory tracts of up to 75% of healthy adults. This means that it is often technically difficult to establish with certainty the pathogenic role of strains isolated from sputum produced by patients with pneumonia or exacerbations of chronic bronchitis, although other evidence clearly points to non-typable *H. influenzae* playing an important aetiological role in these conditions. Population genetic work has largely centred on capsulate strains, particularly those of serotype b. In a recent very large study, nine electrophoretotypes accounted for 81% of 1975 serotype-b strains. Clonal assignments by whole cell polypeptide analysis have also shown that populations of type-b strains are dominated by a few clones.

In population genetic studies, non-typable *H. influenzae* have been compared with type-b isolates. For example, in the studies of Porras et al. and Musser et al. a total of 146 strains was analysed by multilocus enzyme electrophoresis; 135 electrophoretic types were detected indicating an extreme degree of genetic variability. It has been suggested that the rate of recombination among non-typable strains may be greater and their persistence time shorter than among the clones of capsulate strains.

In this study, the technique of analysis of whole cell polypeptide patterns generated by sodium dodecyl sulphate-polyacrylamide gel electrophoresis (SDS-PAGE) has been used to identify clones of non-capsulate strains of *H. influenzae*. This method was used because it is rapid and simple, and because it provides a large number of markers, a feature necessary for the analysis of the structure of a population. The aim of the study was to establish the temporal and geographical distribution of clones of non-typable *H. influenzae* in the North East of Scotland and to determine the number of clones circulating there. Particular attention was paid to the establishment of clonal survival times as no information on this is currently available. Very short clonal lifetimes
would almost certainly preclude the use of clonal analysis for epidemiological purposes and for constructing genetic frameworks for the interpretation of variation in phenotypic characters that may be causally related to virulence.

Materials and methods

Bacteria

Isolates of *H. influenzae* were obtained from routine clinical specimens submitted to the diagnostic microbiology laboratories in Aberdeen from Aberdeen hospitals and from general practitioners in North East Scotland between 1984 and 1989. A total of 223 nontypable *H. influenzae* strains from the same number of individuals was examined. Of the 172 isolates from adults, one was from Portsoy hospital, two were from Aberdeen Maternity Hospital, 22 (including eight from specimens submitted by general practitioners) were from the City Hospital, Aberdeen, and the remaining 147 were from in-patients at Aberdeen Royal Infirmary. Of the 51 isolates from children, one was from the City Hospital, Aberdeen, and the remainder from wards of the Royal Aberdeen Children’s Hospital. Eighteen isolates of nontypable *H. influenzae* were received from Glasgow Royal Infirmary.

Sample preparation and SDS-PAGE

Sample preparation and SDS-PAGE were performed as described by Pennington and Freebairn. Each strain, cultured overnight on heated (chocolated) blood agar at 37°C in air + CO₂ 5%, was harvested with a sterile swab and transferred into 1 ml of sterile distilled water. The suspension was sonicated for 1 min with an ultrasound probe emitting 45 W at maximum power. Samples prepared in this manner were stored at −70°C.

SDS-PAGE of the samples was performed by the method of Laemmli. Sonicated suspensions were added to “boiling buffer” (SDS 2% w/v, 2-mercaptoethanol 5% w/v, glycerol 10% v/v, 0.05 mM Tris-HCl, pH 6-8, and bromophenol blue to colour) and were heated for 5 min in a boiling water bath. Polypeptides were electrophoresed in separating gels of acrylamide 11% by a constant 30 mA current until the blue marker was about 0.5 cm from the bottom of the gel. Gels were stained with Coomassie Brilliant Blue 0.25% w/v in methanol 45:5% and acetic acid 9%.

Analysis of SDS-PAGE profiles and definition of a clone

For initial comparison of the SDS-PAGE profiles, the approach of Murphy *et al.* was used. Tracks were compared in a region of the gel containing polypeptides of 34–42 Kda. This region contained two major outer-membrane proteins (OMPs) and some other polypeptides; band patterns in this region were particularly variable. Isolates were grouped according to similarities in this region and were electrophoresed again. The pattern of bands in the whole track was then compared. In most cases, about forty bands were visualised for each track. Those isolates showing a Dice coefficient greater than 95% were considered to be clonally related.

Dice coefficients (D%) were calculated from the formula:

\[
D\% = \frac{\text{Number of matching bands} \times 2}{\text{Total number of bands in both strains}} \times 100
\]

Enrichment for OMPs

A modification of the protocol of Murphy and Bartos for the enrichment of the OMP P2 of *H. influenzae* type b was used to fractionate cell extracts of non-typable strains. Bacteria were grown on one half of a chocolate agar plate overnight in air + CO₂ 5%, at 37°C, harvested and shaken in 30 μl of 1 M sodium acetate-0.001 M 2-mercaptoethanol (pH 4.0). To this, 270 μl of a solution of 3-(N,N-dimethylmyristylammonio)-propane sulphonate (Fluka) 5% in 0.5 M CaCl₂ was added to a final concentration of 20% v/v. After centrifugation for 10 min at 11 500 g at room temperature, the pellet was discarded. Ethanol was then added to the supernate to a concentration of 80% v/v. After centrifugation at 11 500 g for 20 min the protein pellet was suspended in water.

Estimation of the total number of clones in North East Scotland

The method of Fisher was used to estimate the total number of clones of *H. influenzae* in the host population studied. This method assumes that the abundance of each species (or, in this instance, clone) in a heterogeneous sample made of individuals from many different clones is distributed as \( \chi^2 \) (i.e., as a gamma or Eulerian distribution) and that the probability of N individuals of a given clone is given by the Poisson distribution, with a mean depending on the relative abundance of the clone and the sample size.

A measure of the diversity of population, \( \alpha \), was initially estimated from Fisher’s nomogram; its value was then calculated with the formula

\[
S = \log_e(1 + N/\alpha)
\]

where N is the total number of individual specimens and S the number of clones observed within that sample of specimens. An estimate of the number of clones of the organism circulating within the geographical area under study was also calculated from the formula. The difference between \( \alpha \) values was analysed by the \( \chi^2 \) test.
CLONAL ANALYSIS OF H. INFLUENZAE

Results

Number of clones

In fig. 1, SDS-PAGE profiles of strains were selected to show the range of variation of whole cell polypeptides. Polypeptides of 34–42 Kda appeared to be especially variable. Enrichment of whole cell polypeptide extracts for OMPs showed that the major protein in this region was probably a member of this class and was probably an analogue of the type b-P2 OMP (results not shown). Fig. 1 shows 28 mobility variants of this protein. Variability was not confined to these proteins; mobility differences were apparent in polypeptides of many mol. wts.

Strains were assigned to clones after estimation of Dice coefficients of similarity of polypeptide profiles. Of the initial sample of 223 isolates from Aberdeen, 147 (66%) were designated as being unique, 26 (12%) were assigned to clones with two members, 3 (1%) to clones with three members, 12 (5%) to clones with four members, 15 (7%) to clones with five members and 12 (5%) to clones with six members. A single clone with eight members (4%) was also identified. Table I shows the relationships between clonal type and clinical source of isolates. With the exception of the group from sputum samples from adults, the numbers in each section are relatively small, making interpretation of the effect of type of specimen difficult. However, apart from the predominance of isolates from children in clones 11 and 13, no relationships between clonal type and specimen source were evident.

Temporal relatedness of members of clones and place of isolation

The dates of isolation of clone members are shown in fig. 2. The longest interval between the isolation of clonally related strains was 55 months, an interval which virtually equalled the temporal limits of the study period. Members of one other clone were isolated more than 4 years apart. In three clones, more than 3 years separated the isolation of members, in seven clones, more than 2 years, and in six clones more than 1 year. The members of five clones were isolated close together in time.

Geographical distribution of clones

The population of North East Scotland is distributed unevenly, being concentrated in towns, villages and rural areas along the coast (fig. 3a); 41% of the population lives in the city of Aberdeen. Of the 72 strains of H. influenzae belonging to multi-member clones, 35 came from patients with home addresses in Aberdeen. Another 35 came from other places in North East Scotland, one came from a patient in Aberdeen who lived in Merseyside, England, and one was isolated in Glasgow; both these strains belonged to clones with a predominantly North East membership. Figs. 3b–d show the geographical distribution of clones with more than three members. In 18 of the 21 multi-member clones, members were isolated from patients with home addresses many miles apart; no geographical clustering of clonally related strains was

Fig. 1. Whole-cell polypeptide profile (Coomassie Brilliant Blue stain) of 28 non-typable strains of H. influenzae. M, s are shown on the left (Kda); 34–42-Kda polypeptides are indicated on the right.
Table I. Clonal type and clinical source of isolates

<table>
<thead>
<tr>
<th>Source of isolates</th>
<th>Number of isolates allocated to multi-member clone no.</th>
<th>Number of unique clones</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21 22 23</td>
<td></td>
</tr>
<tr>
<td>Sputum Adults</td>
<td>5 1 2 2 2 1 4 2 4 1 3 2 1 1 1 2 4 4 2 1 3 2 4 99</td>
<td>2</td>
</tr>
<tr>
<td>Children</td>
<td>1 1</td>
<td>6</td>
</tr>
<tr>
<td>Tracheal aspirate</td>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Adults</td>
<td></td>
<td>1 0</td>
</tr>
<tr>
<td>Children</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Throat swab Adults</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Children</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Eye swab Adults</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Children</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>(Peri) nasal swab</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Adults</td>
<td></td>
<td>3</td>
</tr>
<tr>
<td>Children</td>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Nasopharyngeal catheter tip Adults</td>
<td>1 4 1</td>
<td>3</td>
</tr>
<tr>
<td>Children</td>
<td></td>
<td>17</td>
</tr>
<tr>
<td>Other Adults</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Children</td>
<td>1</td>
<td>1</td>
</tr>
<tr>
<td>Total</td>
<td>5 2 2 2 2 2 6 2 4 2 8 3 4 2 2 6 4 2 2 5 2 5 147</td>
<td></td>
</tr>
</tbody>
</table>

The possibility that these patients had acquired clonally related organisms in hospital was investigated by studying their location by ward and date of sampling. It was concluded that only four strains could have been transmitted nosocomially. In only one clone, which had two members, were organisms isolated from patients with home addresses in different parts of North East Scotland. The other strains were related by time and place of isolation in hospital, having been isolated from patients in the same ward, less than a month apart, and also by town of residence of their hosts.

Diversity of clones

Table II shows that isolates from the Children's Hospital were significantly less diverse than those from patients in adult wards (p < 0.001).

Estimation of the number of clones of non-typable H. influenzae circulating within the population of North East Scotland

The following assumptions were made:

1. the human population from which samples were collected was 500,000 (the population served by the acute hospitals in Aberdeen);
2. the carriage rate in this population was 70%;18
3. individuals carried only one clonal type at a time;19
4. the isolates studied were a representative sample of strains carried by the population in North East Scotland.

A calculation of Fisher's distribution by the formula gave an estimate of 2280 clones circulating in North East Scotland during the study period.

Discussion

This study has shown that non-typable strains of H. influenzae show marked strain to strain variability in their SDS-PAGE whole cell polypeptide profiles. These results confirm and extend those of Paterson et al.9 who used this technique to examine a small number of Scottish strains, and complement those of Loeb and Smith20 who found unique SDS-PAGE mobility patterns of OMPS for each of 18 strains. Multilocus enzyme electrophoretotyping has also shown that non-typable strains of H. influenzae are very variable. Musser et al.11 examined 65 strains and
found that all had different electrophoretotypes. Porras et al.\textsuperscript{10, 21} examined fewer enzymes than Musser et al. and showed that some strains had identical electrophoretotypes; 73 different electrophoretotypes occurred in a population of 82 strains.

Our demonstration that the SDS-PAGE polypeptide profiles of a significant number of strains unrelated in time or place of isolation were closely similar provides further support for the hypothesis that non-typable strains, like capsulate strains,\textsuperscript{8} have a clonal population structure. Earlier evidence for this hypothesis has already been provided by the finding of specific enzyme electrophoretotypes at a greater frequency than predicted by chance\textsuperscript{21} and the observation that certain independently isolated strains have identical SDS-PAGE polypeptide profiles.\textsuperscript{9} Additional evidence has been provided by Weinberg et al.\textsuperscript{22} who showed that non-typable strains isolated from the blood of children with lower respiratory tract infection in Pakistan had a clonal population structure. Nine clonal groups (defined by multilocus enzyme electrophoretotype, OMP profile and biotype) were found among 34 isolates, with five of these clonal types accounting for 82% of the strains.

The degree of concordance of population structure type revealed by enzyme electrophoretotyping and whole cell polypeptide electrophoretotyping was not entirely unexpected in the light of reported correlations between the polymorphisms revealed by multilocus enzyme electrophoresis and SDS-PAGE analysis of OMPs of \textit{E. coli}\textsuperscript{23} and \textit{Neisseria meningitidis}.\textsuperscript{24} That such correlations occur, however, is surprising. Multilocus enzyme electrophoresis detects charge differences between native proteins, and the genetic and evolutionary significance of these polymorphisms have been interpreted almost exclusively in terms of single amino acid substitutions.\textsuperscript{22} In SDS-PAGE, charge differences are masked because large amounts of SDS bind to the proteins during electrophoresis and migration rates are proportional to log mol. wt.

**Table 11.** Diversity of \textit{H. influenzae} isolates from different sources

<table>
<thead>
<tr>
<th>Source</th>
<th>Number of strains (N)</th>
<th>Number of clones (S)</th>
<th>Index of diversity (s)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Children's wards</td>
<td>48</td>
<td>41</td>
<td>133</td>
</tr>
<tr>
<td>Adult wards</td>
<td>175</td>
<td>137</td>
<td>291</td>
</tr>
</tbody>
</table>

Difference between children's and adult wards statistically significant, \( p = 0.001 \) (\( \chi^2 \) test).
Polypeptides differing only by a single amino-acid substitution would be expected to co-migrate and the differences seen in our SDS-PAGE gels almost certainly reflect much greater sequence differences. The degree of concordance between different electrophoretic systems is also in sharp contrast with the results obtained with eukaryotic systems. Thus Leigh-Brown and Langley26 found that the genetic variability of proteins from Drosophila visualised in two dimensional PAGE was much lower (average heterozygosity 4%) than the corresponding value obtained for enzymes by starch gel electrophoresis (average heterozygosity 14%). Similar results have been found with human proteins. Neel27 has speculated that the much lower (< 50%) variation shown by two dimensional PAGE may be due to mutability differences related to protein class, rather than being caused by more stringent biological selection, or, trivially, having a technical explanation.

A failure to find identical enzyme electrophoretotypes in small groups of H. influenzae strains from Sweden and the USA has led to the tentative suggestion that the persistence time of clones of non-typable strains may be less than in the capsule strains.10 These workers have also suggested that the high genetic diversity of these strains provides evidence which favours a higher recombination rate. Some workers have claimed that the capsular polysaccharide of H. influenzae is a barrier to exogenous DNA,28 but others have demonstrated that, although the majority of strains are non-transformable, a significant proportion can undergo transformation.29 Other factors which may explain the apparently longer clone persistence time and low frequency of recombination of capsule strains are the low carriage rates and limited duration of carriage of these strains.30

Long term studies will be needed to establish the persistence time of clones. In the present study, members of one clone were isolated 55 months apart, virtually at the temporal extremes of the study period. Four clones had members isolated more than three years apart, and the majority of multi-member clones (19 out of 23) had members isolated more than a year apart. Our demonstration that many clones were maintained by the frequent generation of new variants may explain the apparently longer clone persistence time and low frequency of recombination between clones, or whether variation is being transmitted over quite long distances in the community.

Population biologists have made little use of Fisher's logarithmic distribution since its original application to the study of species diversity in populations of Lepidoptera and Hemiptera. However, Zei and co-workers11 have shown that it offers a useful approximation to the Karlin-McGregor distribution,12 which generates an exact distribution of the number of selectively neutral alleles under mutation and drift in a finite population for any number of alleles. Zei and co-workers successfully used Fisher's distribution to show that surnames, whose transmission is haploid and unisexual, behave like neutral alleles.33 Several assumptions had to be made when using Fisher's distribution to calculate the clonal diversity of populations and the total number of clones circulating during the study period, and the results of these calculations must be regarded as tentative rather than definitive. Nevertheless, we believe that the results are useful in that they indicate areas which require further study. Clones found did not appear to be related to the type of specimen from which H. influenzae was isolated. However, the finding of a statistically significant difference in strain diversity between isolates from the paediatric and the adult wards was unexpected. Possible explanations are that children preferentially carry a sub-set of clones or that transmission of infection occurred within the Children's Hospital. Work is in progress to test these hypotheses. The establishment of persistence times and geographical distribution of clones also requires further investigation, and studies are in progress which aim to establish whether the extremely polymorphic population of non-typable H. influenzae strains is made up of a large number of clones with similar population persistence times, with little recombination between clones, or whether variation is being maintained by the frequent generation of new variants by intrachromosomal re-arrangements or transformation.

This work was supported by Grant K/MRS/50/C1029 from the Chief Scientist's Office, Scottish Home and Health Department. K.D.B. was supported by Aberdeen University Faculty of Medicine Endowment Funds.

References


