Sporadic isolates of *Escherichia coli* O157. H7 investigated by pyrolysis mass spectrometry

R. FREEMAN1*, P. R. SISSON1, D. R. JENKINS1, A. C. WARD2, N. F. LIGHTFOOT1 AND S. J. O’BRIEN3

1 Regional Public Health Laboratory, 2 Department of Microbiology and 3 Department of Epidemiology and Public Health, Medical School, Newcastle upon Tyne, NE4 6BE, UK

(Accepted 17 January 1995)

SUMMARY

Thirty-six encoded isolates of *Escherichia coli*, 32 of which were of serotype O157, were examined by pyrolysis mass spectrometry (PyMS). Thirty-one of the serotype O157 isolates possessed the flagellar antigen H7 and produced Verocytotoxin (VT). the other isolate serotyped as H45 and was non-toxigenic. Eighteen of the VT-producing *E. coli* (VTEC) isolates were from sporadic disease in residents of the Northern Region. Standard principal component (PC) and canonical variate (CV) analysis of the data distinguished only the four non-O157 isolates from the remainder which were indistinguishable by this approach. A similarity matrix based on differences between individual CV means distinguished a further ten isolates. The matrix correctly clustered 2 pairs of isolates from siblings and 4 isolates from an affected family. A further 5 clusters of 3 or more isolates and 6 pairs of isolates were defined. These groupings proved to be homogenous for toxin phenotype but occasionally entrained isolates of dissimilar phage type. However, in general, PyMS-derived clustering of apparently sporadic isolates accorded with geographical locations as determined by postcode. PyMS, which is a quick and high volume capacity phenotypic technique, may be a useful addition to existing methods in the investigation of the epidemiology of sporadic VTEC disease.

INTRODUCTION

Verocytotoxin (VT) producing strains of *Escherichia coli* (VTEC) are an important public health problem. Most VTEC belong to serogroup O157 and possess the flagellar antigen H7 [1, 2]. The O157. H7 serotype, which has increased markedly in frequency over the past decade in both North America and the UK, is genetically highly conserved and only distantly related to other VT-producing strains of other *E. coli* serotypes. Clonal descent of the serotype has been suggested [3].

Because of the propensity of *E. coli* O157. H7 to cause outbreaks of gastrointestinal disease, complicated by haemorrhagic colitis and the haemolytic-uraemic syndrome, the development of typing methods is a high priority but the

* Author for correspondence and reprints: Dr R. Freeman, Regional Public Health Laboratory, Newcastle General Hospital, Newcastle upon Tyne, NE4 6BE.
high degree of similarity between isolates has limited the options available. Plasmid analysis [4], phage typing [5], lipopolysaccharide (LPS) polymorphism [6] and VT genotyping [7] have all been used with some degree of success. Phage typing has defined over 80 types. A potentially highly discriminatory index is based on bacteriophage \( \lambda \)-restriction fragment length polymorphism (\( \lambda \)-RFLP) [8], but this method still awaits full evaluation by comparison with other techniques. Some of the current techniques are expensive and time-consuming, although, for instance, phage typing and VT genotyping can be both quick and relatively cheap. Careful studies on outbreaks have identified a bovine reservoir for \( E. coli \) O157:H7 [9]. However, there is also a need to address the commoner problem of sporadic cases in which the sources, mode of transmission and epidemiology are less well understood than those of outbreak disease.

Sporadic isolates accumulate over a period of time and, by definition, there is no apparent evidence of relatedness and no suspicion of a common source. A rapid and cheap, but highly discriminatory, technique, capable of dealing with large numbers of such apparently unrelated isolates might screen such collections and allow the subsequent targeted application of the DXA-based techniques to best effect, if previously unsuspected relationships were suggested.

Pyrolysis mass spectrometry (PyMS) is a highly discriminatory, rapid and, in consumable costs, relatively cheap method for inter-strain comparisons of microorganisms which has the additional advantage of being able to examine large numbers of isolates (up to 50) at a time. We have conducted a preliminary assessment of the suitability of PyMS for the examination of sporadic isolates of \( E. coli \) O157:H7.

**MATERIALS AND METHODS**

**E. coli isolates**

Thirty-six isolates of \( E. coli \) were studied. They comprised 28 consecutive isolates of serogroup O157 from patients in the Northern Region, obtained between June and September 1992, four O157 isolates kindly donated by the Laboratory for Enteric Pathogens (LEP). Central Public health Laboratory, Colindale, London and four isolates of \( E. coli \) shown not to be of serogroup O157. The isolates studied are listed in Table 1, together with brief epidemiological data and the final results of flagellar antigen typing and phage typing, which together with detection of VT genes, VT1 and VT2, were performed at LEP. PyMS examination was performed under a blinded code until the other results were available.

**Pyrolysis mass spectrometry**

All organisms were checked for purity and then subcultured in duplicate, overnight on nutrient agar (Lab M) at 37°C. Using a flamed straight wire, triplicate smears were then made onto Ni-Fe pyrolysis foils (Horizon Instruments Ltd., Heathfield, Sussex) from well-separated colonies from each of the duplicate subcultures. Foils were then inserted into pyrolysis tubes (Horizon Instruments) and the assembled tubes were then heated at 80°C for 5 min in a hot air oven.

All 216 sample tubes (two subcultures per isolate, each sampled in triplicate) were then processed as a single batch on a Horizon Instruments PYMS 200X pyrolysis mass spectrometer. The time required for the analysis of each individual
Table 1. Isolate designations, phage type, toxin status, date of isolation and postcode of affected patient for 28 isolates of Escherichia coli, serogroup 0157 from the Northern Region (NR numbers), 4 isolates from the Laboratory for Enteric Pathogens (LEP numbers) and 4 isolates of E. coli not of serogroup 0157 (EC numbers), examined by pyrolysis mass spectrometry. Unless otherwise indicated serogroup 0157 isolates had the flagellar antigen H7

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Phage type</th>
<th>Toxin status</th>
<th>Date</th>
<th>Post code</th>
</tr>
</thead>
<tbody>
<tr>
<td>XR1</td>
<td>49</td>
<td>VT2</td>
<td>22/6</td>
<td>HD6</td>
</tr>
<tr>
<td>XR2*</td>
<td>RDNC</td>
<td>Neg</td>
<td>29/6</td>
<td>NE12</td>
</tr>
<tr>
<td>XR3</td>
<td>2</td>
<td>VT2</td>
<td>4/7</td>
<td>NE32</td>
</tr>
<tr>
<td>XR4</td>
<td>2</td>
<td>VT2</td>
<td>5/7</td>
<td>NE33</td>
</tr>
<tr>
<td>XR5</td>
<td>2</td>
<td>VT2</td>
<td>6/7</td>
<td>NE33</td>
</tr>
<tr>
<td>XR6</td>
<td>32</td>
<td>VT2</td>
<td>22/5</td>
<td>CA7</td>
</tr>
<tr>
<td>XR7</td>
<td>32</td>
<td>VT2</td>
<td>31/5</td>
<td>CA10</td>
</tr>
<tr>
<td>XR8</td>
<td>2</td>
<td>VT2</td>
<td>8/7</td>
<td>NE32</td>
</tr>
<tr>
<td>XR9</td>
<td>2</td>
<td>VT2</td>
<td>10/7</td>
<td>DL12</td>
</tr>
<tr>
<td>XR10</td>
<td>34</td>
<td>VT2</td>
<td>12/7</td>
<td>SR6</td>
</tr>
<tr>
<td>XR11</td>
<td>2</td>
<td>VT2</td>
<td>16/7</td>
<td>DH9</td>
</tr>
<tr>
<td>XR12</td>
<td>2</td>
<td>VT2</td>
<td>23/6</td>
<td>DH9</td>
</tr>
<tr>
<td>XR13</td>
<td>2</td>
<td>VT2</td>
<td>20/6</td>
<td>DH7</td>
</tr>
<tr>
<td>XR14</td>
<td>2</td>
<td>VT2</td>
<td>3/8</td>
<td>CA11</td>
</tr>
<tr>
<td>XR15</td>
<td>2</td>
<td>VT2</td>
<td>31/7</td>
<td>CA6</td>
</tr>
<tr>
<td>XR16</td>
<td>49</td>
<td>VT2</td>
<td>3/8</td>
<td>DL16</td>
</tr>
<tr>
<td>XR17</td>
<td>RDNC</td>
<td>VT2</td>
<td>1/8</td>
<td>DL16</td>
</tr>
<tr>
<td>XR18</td>
<td>1</td>
<td>VT1/2</td>
<td>6/8</td>
<td>CA10</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Isolate</th>
<th>Phage type</th>
<th>Toxin status</th>
<th>Date</th>
<th>Post code</th>
</tr>
</thead>
<tbody>
<tr>
<td>NR19</td>
<td>2</td>
<td>VT2</td>
<td>19/8</td>
<td>NE63</td>
</tr>
<tr>
<td>NR20*</td>
<td>RDNC</td>
<td>VT2</td>
<td>26/9</td>
<td>NE63</td>
</tr>
<tr>
<td>NR21</td>
<td>49</td>
<td>VT2</td>
<td>23/8</td>
<td>DL15</td>
</tr>
<tr>
<td>NR22</td>
<td>49</td>
<td>VT2</td>
<td>22/8</td>
<td>DL17</td>
</tr>
<tr>
<td>NR23</td>
<td>2</td>
<td>VT2</td>
<td>Northampton</td>
<td></td>
</tr>
<tr>
<td>NR24+</td>
<td>4</td>
<td>VT1/2</td>
<td>30/8</td>
<td>CA3</td>
</tr>
<tr>
<td>NR25†</td>
<td>4</td>
<td>VT2</td>
<td>30/8</td>
<td>CA3</td>
</tr>
<tr>
<td>NR26‡</td>
<td>1</td>
<td>VT2</td>
<td>29/8</td>
<td>LA14</td>
</tr>
<tr>
<td>NR27‡</td>
<td>1</td>
<td>VT1/2</td>
<td>7/9</td>
<td>LA14</td>
</tr>
<tr>
<td>NR28‡</td>
<td>1</td>
<td>VT2</td>
<td>7/9</td>
<td>LA14</td>
</tr>
<tr>
<td>NR29‡</td>
<td>1</td>
<td>VT2</td>
<td>Not serogroup 0157</td>
<td></td>
</tr>
<tr>
<td>NR30‡</td>
<td>1</td>
<td>VT1/2</td>
<td>Not serogroup 0157</td>
<td></td>
</tr>
<tr>
<td>NR31‡</td>
<td>1</td>
<td>VT1/2</td>
<td>Not serogroup 0157</td>
<td></td>
</tr>
</tbody>
</table>

* This organism was flagellar type H45, non-toxigenic and fermented sorbitol.
† Isolates from siblings.
‡ Isolates from probable point source outbreak.
VT2 = produces Verocytotoxin 2; VT1/2 = produces Verocytotoxins 1 and 2.

sample was approximately 2 min and pyrolysis was at a Curie point of 530 °C. Products were ionised by collision with a beam of low energy electrons (25 eV), and the ions were separated in a quadropole mass spectrometer which scanned the pyrolysate 160 times at intervals of 0.35 s from the start of pyrolysis.

Integrated ion counts at mass intervals of 51–200 were recorded on floppy disk, together with the pyrolysis sequence number and total ion count for each individual sample. The triplicate foils from each subculture were labelled as a separate group.

Data analysis

Principal component and canonical variate analysis

Spectral data from all 36 isolates were initially subjected to principal component (PC) canonical variate (CV) multivariate analysis, using the GENSTAT 5 program (Numerical Algorithm Group, Oxford). For this analysis, spectra were normalized by an iterative technique to eliminate variations due solely to differences in the amount of sample pyrolysed. Replicate spectra of the same isolate, including data from both subcultures were then labelled as distinct groups and analysed for between-group to within-group variations in each mass ion. The mass ion peaks were then ranked in order of discrimination and the most discriminatory mass ions were then used in the PCCV analysis.
The results of this initial analysis were presented as an ordination diagram of PCCV 1 v. PCCV 2, each individual datapoint representing the mean of 6 samples (duplicate subcultures, each sampled in triplicate) for each labelled group (that is, isolate). Circles indicating the 95% confidence zones around each datapoint were superimposed and any groups clearly differentiated from the remainder were identified as outliers. Outlying groups were edited from the dataset and the analysis was repeated until no further outlying groups could be identified. That is, the 95% confidence zones for all the remaining isolates overlapped. These remaining isolates were then regarded as indistinguishable by this standard form of analysis which has been previously described [10].

**Direct comparisons of individual PCCV means**

Data from those isolates unable to be further distinguished by standard PCCV analysis were then used in a series of analyses in which the principal component canonical variate mean (PCCV mean) of the data from each isolate was directly compared with the PCCV mean of the data from each other isolate. The differences between the two PCCV means for each such comparison were then used to construct a similarity matrix, the order of which was determined by the Unweighed Pair Group Method with Averages (UPGMA) analysis. Provided that the results obtained in the comparisons of duplicate subcultures conform to a normal distribution, these values approximate to those of $\chi^2$ distribution with one degree of freedom. Hence, values of 3.84, 6.63 and 10.84 were taken to indicate, respectively 95%, 99% and 99.9% confidence limits [11].

**RESULTS**

**Principal component and canonical variate analysis**

On the first PCCV analysis the ordination diagram (not shown) showed isolates EC1, EC2, EC3 and EC4 as clear outliers. On re-analysis following the removal of these data from the database, no further outliers were detected, the 95% confidence zones of all the remaining datapoints overlapping.

**Direct comparisons of individual PCCV means**

The resulting similarity matrix, ordered by UPGMA analysis, of the 'one to one' comparisons of the PCCV means of those 32 isolates unable to be further distinguished in the ordination diagram is seen in Fig. 1. This shows the distribution of those comparisons in which the value obtained was < 3.84, that is, a value within 95% confidence limit.

Isolates NR1, NR2, NR6, NR7, NR10, NR15, NR18, LEP1, LEP3 and LEP4 showed differences from all other isolates, and each other, which exceeded the 95% confidence limit. In the case of NR1, NR2 and LEP4 these differences also exceeded the 99% confidence limit (6.63). Isolate NR1 was from a traveller from without the Northern Region and isolate NR2 was of serotype O157:H45 and non-toxigenic. Isolates LEP1, LEP3 and LEP4 were from the Midlands. Isolates NR6, NR7, NR15 and NR18 were 4 of only 5 sporadic isolates of *E. coli* O157:H7 from the geographically distinct Cumbrian division of the Northern Region. Isolate NR10 was the only sporadic isolate of *E. coli* O157:H7 of 13 from the
Pyrolysis of sporadic E. coli O157. H7

Fig. 1. A similarity matrix compiled from the results of individual comparisons of the principal component canonical variate means of pyrolysis mass spectra of 32 isolates of Escherichia coli O157. The matrix was ordered by UPGMA analysis. Solid squares denote values of 3.84 or less.

Northumbrian/Durham division of the Northern Region to be shown to have no relationship with any other sporadic isolate at the 95% level. Thus, of 18 sporadic isolates of E. coli O157. H7 from residents of the Northern Region, 4 of 5 isolates from the Cumbrian division were distinct from all other isolates at the 95% level in contrast to only 1 of 13 from the Northumbrian/Durham division ($\chi^2 = 6.2$; $P < 0.01$).

Those isolates epidemiologically highly likely to belong to single strains (isolates NR19/NR20, from siblings and isolates NR23/NR24, also from siblings) were indistinguishable at the 95% confidence limit. Three of the four isolates from an affected family (NR25/NR26/NR28), were also appropriately clustered at the 95% level. In the case of the fourth isolate from this family (NR27), the difference between the PCCV means when compared with NR28 was within the 95% confidence limit but the differences between the PCCV means of NR27 and NR25 and NR26 were 9.25 ($> 99\% < 99.9\%$) and 5.9 ($> 95\% < 99\%$), respectively.
When the sibling isolate criterion of relatedness, that is, differences between individually paired CV means being within the 95% confidence limit, was applied to the rest of the matrix containing data from sporadic, apparently unrelated isolates, 5 clusters of 3 or more isolates were found.

In the first, isolates NR4, NR5 and NR9 were related. All three isolates were of phage type 2 and produced VT2. Isolates NR4 and NR5 were obtained within 24 h of each other from apparently unrelated patients within the same postal district. Similarly, isolates NR5 and NR8, which were clustered with NR13 were isolated within 48 h of each other from patients in adjoining postal districts. The cluster containing isolates NR3, NR16 and NR20 included two isolates of phage type 2, producing VT2 (NR3 and NR20) and one (NR16) of phage type 49. All three were from patients in separate and unrelated postal districts. Isolates NR16 and NR17, which clustered with NR20, were both VT2-producing and were isolated within 48 h of each other from patients in the same postal district. NR16 is of phage type 49 whereas NR17 was phage non-typable. Finally, a cluster of 4 isolates comprised NR3, NR11, NR12 and NR16. Isolates NR11 and NR12 were obtained over a 23 day interval, but from patients in the same postal district.

Six pairs of isolates also showed differences which fell within the 95% confidence limit. These were NR4/NR5 (same phage and toxin type; same postal district), NR17/NR22 (one phage untypable and one phage type 49 but both same toxin type; adjoining postal districts), NR21/NR22 (same phage and toxin type; adjoining postal districts), NR12/NR19, NR22/NR26 and NR13/NR14. In these latter three pairs no epidemiological link was evident. Finally, strain LEP2 was seen to differ by less than the 95% confidence limit from isolates NR5 and NR8.

It is also notable that some isolates showed a broader relatedness than other isolates. Of the 18 sporadic isolates of \( \textit{E. coli} \) 0157. H7 from residents of the Northern Region, isolates NR5 and NR16 were indistinguishable at the 95% level from, respectively, 6 and 5 other isolates and formed parts of, respectively, 4 and 3 clusters or pairs at this level of difference.

**DISCUSSION**

In this study, the 4 isolates of \( \textit{E. coli} \) of serogroups other than O157 were shown to differ substantially from the remaining 32 isolates of \( \textit{E. coli} \) O157 by standard PyMS analysis. Further analysis of the remaining, predominantly sporadic and apparently unrelated, isolates, showed that the differences between the PCCV means of some isolates fell within the 95% confidence limit, suggesting apparent relationships between them.

The differences between the PCCV means in two pairs of isolates from siblings (and, therefore, likely to be isolates of the same strain) fell within the 95% confidence limit. This lends support to the potential validity of the other relationships suggested. Further support was lent by the clustering of 3 of 4 isolates from a single affected family. The fourth isolate showed a difference within the 95% limit in relation to 1 of the other 3.

In 4 of 5 clusters containing 3 or more apparently sporadic isolates, at least 2 isolates in each cluster were obtained from patients in the same or immediately adjoining postal districts. Additionally, in 3 of 6 pairs of isolates, both isolates in
the pair had been taken from patients in the same or immediately adjoining postal
district. The clustered isolates, whether in groups of three or more or simply in pairs,
were usually of the same toxin and phage types within each cluster. However,
there were exceptions and isolates of differing phage types (although of the same
toxin phenotype) were entrained in two of the clusters.

The high degree of overall similarity of the isolates examined by direct ‘one to
one’ comparisons of PCCV means is emphasized by the finding that of the 32
isolates only 10 showed differences beyond the 95% confidence limit from all other
isolates and all the isolates were indistinguishable at the 99.9% level. These
findings suggest that the relationships suggested in the similarity matrix reflect
very small differences between otherwise highly similar organisms.

PyMS of whole cells is a method which is capable, within a single batch of
organisms, of high levels of discrimination [10], but which is nonetheless,
phenotypic. However, comparison of the PyMS findings with the results of the
other tests on our studied collection show that many of the previously described
phenotypic variables in E. coli O157 strains are unlikely to be the basis of the
PyMS-based distinctions and groups found. Thus, all the isolates involved in the
clusters in the similarity matrix were of serotype O157.H7 and the toxin pheno-
type was homogenous within each cluster. Previous studies have suggested that
only one phenotype of the lipopolysaccharide can be expected in isolates which are
uniformly E. coli O157.H7 toxin-producers [6]. However, the inclusion of more
than one phage type in a cluster of isolates suggests that the characters used in the
analysis may be phenotypic rather than genotypic, as does the small differences
seen between NR5 and NR8 and the geographically distant LEP2. The larger
differences between NR27 and the other three isolates from an affected family
(NR25, NR26 and NR28) are also probably best explained as phenotypic.

PyMS analyses have occasionally resulted in discrimination approaching that
obtained by DNA-based methods on the same populations of organisms [13, 14]
but it is unlikely that the results obtained here reflect genotypic differences. In a
recently reported study it was found that the λ-RFLP patterns of sporadic
isolates were typically distinct, in contrast to their high level of similarity in
outbreak strains [8]. In contrast, here, the occasional clustering of isolates of the
same toxin phenotype but of discordant phage types strongly suggests that the
observed relatedness is based on phenotypic characteristics. It is interesting that
these discordant clusters entrained phage types 2 and 49; the commonest types
encountered in the UK [5]. It will be important to address this issue in order
correctly to assess the significance of our findings. PyMS studies on geographically
distinct sporadic isolates of individual phage types as well as further studies on
isolates of diverse phage types will be necessary. Since the presumptively
phenotypic differences detected by PyMS are unlikely to be any of those
previously described, it will also be very important to combine any such further
PyMS studies with other methods, such as phage typing, VT typing, plasmid
analysis and RFLP analysis in order to maximize the discrimination to be
achieved.

Whatever the basis of the differences found in this study of E. coli O157.H7, the
PyMS method has apparently shown relationships between these sporadic isolates
which accord to some extent with their geographical sources. If these groupings
are based upon small but significant phenotypic differences, they might represent clusters of genotypically distinct strains occupying specific ecological niches and expressing common phenotypic characteristics. PyMS analysis of much larger numbers of apparently sporadic human isolates of *Escherichia coli* O157:H7 might link presumptive phenotypes with geographical locations over prolonged intervals and provide an insight into the epidemiology of sporadic disease associated with this organism.

REFERENCES