EPIDEMIOLOGY AND BACTERIAL TYPING

Comparison of Smal-defined genotypes of Campylobacter jejuni examined by KpnI: a population-based study

SOPHIE MICHAUD§, SUZANNE MENARD†, CHRISTIANE GAUDREAU‡ and ROBERT D. ARBEIT§

*Centre Hospitalier Universitaire de Sherbrooke and 1Direction de la Santé Publique de l’Estrie, Sherbrooke, 2Centre Hospitalier de l’Université de Montréal, Hôpital St-Luc, Montréal, Québec, Canada; ‡Research Service, VA Medical Center, and Department of Medicine, Boston University School of Medicine, Boston, MA, USA

Pulsed-field gel electrophoresis (PFGE) was used to analyse 147 isolates collected in two regions of Québec province (Estrie and Montréal) between March 1998 and Feb. 1999, to determine the utility of molecular strain typing for a population-based collection of Campylobacter jejuni and to compare directly the discriminatory power of Smal and KpnI restriction digests. With a combination of epidemiological criteria including space and time plus molecular strain typing, 49% of isolates from Estrie and 39% of isolates from Montréal were identified as belonging to a putative cluster. For 41% of the cases, sources were either missing or explicitly unknown; the remaining sources were subject to recall bias. Thus, the evaluation of sporadic cases of campylobacter enteritis by descriptive clinical investigation alone is neither sensitive nor reliable for identifying sources of infection. In the PFGE analysis, KpnI digests provided appreciably greater discriminatory power than Smal digests. When combining the PFGE analyses with basic epidemiological criteria, 30% of the putative Smal clusters were inconsistent with the epidemiological criteria compared with 17% of the KpnI clusters. Among the 98 isolates assigned to clusters by Smal, only 65% gave concordant results with KpnI. In contrast, among the 81 isolates assigned to clusters by KpnI, 92% gave concordant results with Smal. Finally, clusters that were epidemiologically related to ingestion of raw milk and specific water sources correlated better with the typing results based on KpnI than Smal. Thus, KpnI is the enzyme of choice for molecular epidemiology studies of C. jejuni. The combination of continuous epidemiological surveillance and molecular strain typing may be useful for identifying new sources and mechanisms of transmission for community-acquired C. jejuni infection and ultimately for developing new approaches to prevention.

Introduction

Campylobacter enteritis is the third most common notifiable infectious disease in Québec province. In 1998, the number of diarrhoeal illnesses in Québec attributed to campylobacter enteritis exceeded the combined total caused by Salmonella and Shigella species, Escherichia coli O157:H7 and Yersinia enterocolitica. Between 1990 and 1998, the annual reported incidence of campylobacter enteritis increased from 29.9 to 43.4 per 100 000 population. The reasons for this increase are not well defined; moreover, surveillance strategies vary among different regions within Québec province. In Estrie, all cases reported to the public health service are prospectively investigated by clinical nurses by means of a standardised questionnaire. Other regions, such as Montréal, do not routinely investigate reported cases.

Campylobacter jejuni is a commensal of the intestinal tract of a wide range of birds and mammals, including pets and domestic animals widely used for food production [1]. It is principally a food-borne pathogen, with well documented outbreaks related to ingestion of raw milk, untreated water and undercooked poultry [2]. However, most cases of campylobacter enteritis are deemed to be sporadic [3, 4].

Received 20 March 2000; revised manuscript received 22 Feb 2001; accepted 16 May 2001.
Corresponding author: Dr S. Michaud (e-mail: smichaud@courrier.usherb.ca).
The wide distribution of campylobacters makes it difficult to identify specific sources of enteritis by traditional epidemiological investigation. Molecular strain typing methods may help clarify the complex epidemiology of these infections. Pulsed-field gel electrophoresis (PFGE) of DNA digests has proved to be a highly reproducible and discriminatory typing technique for a range of bacterial pathogens [5]. Although there are several reports describing PFGE typing of C. jejuni [6–8], no clear consensus has emerged as to whether Smu1 or Kpn1 is the most effective restriction endonuclease for such studies [9]. Smu1 has been used most commonly, but some investigators suggest that Kpn1 digests provide greater discriminatory power [6, 10, 11].

This report describes the application of PFGE molecular strain typing to a large collection of C. jejuni enteritis isolates from Québec. The goals of the study were to determine whether such typing would facilitate the identification of common source clusters of community-acquired infection, and to compare directly the discriminatory power of Smu1 and Kpn1 restriction digests.

**Materials and methods**

**Population**

All isolates of C. jejuni subsp. jejuni (hereafter C. jejuni) cultured at the Centre Universitaire de Santé de l’Éstrie and at the Centre Hospitalier de l’Université de Montréal, Hôpital St-Luc, were collected prospectively between 1 March 1998 and 28 Feb. 1999. Clinical information about individual cases of campylobacter infection in Estrie was obtained prospectively by the Direction de la Santé Publique (DSP) of l’Estrie. For patients from Montréal only information about age, sex, date of culture and history of foreign travel was available.

**PFGE**

C. jejuni isolates were grown on sheep blood 5% agar for 48 h at 37°C in a micro-aerobic atmosphere. Bacterial colonies were harvested, resuspended in 900 μl of cold saline and treated with 100 μl of formaldehyde on ice for 1 h to inactivate endogenous nucleases [12]; bacterial suspensions were then washed three times in 1000 μl of cold saline by centrifugation at 12 000 rpm and resuspended in 1000 μl of saline. The optical densities of the bacterial suspensions were then adjusted to 1.9 at 405 nm and 625–μl samples were gently mixed with 375 μl of InCert agarose 1.5% (FMC BioProducts, Rockland, ME, USA). The resulting mixture was poured into moulds and DNA was prepared as described previously [13]. DNA in agarose plugs was digested overnight at 25°C with Smu1 or at 37°C with Kpn1 and the resulting macrorestriction digests were electrophoresed in an agarose 1% gel in 0.5 × TBE buffer (45 mM Tris, 45 mM boric acid, 1 mM EDTA) at 200 V and 14°C with a contour-clamped homogeneous electric field apparatus (CHEF Mapper or CHEF DRII System, BioRad, Hercules, CA, USA). Pulsing was ramped from 1 to 40 s over 22 h for Smu1 digests and from 4 to 20 s over 23 h for Kpn1 digests. The gels were stained with ethidium bromide and photographed under ultraviolet light with a digital camera.

Each gel comprised 15 lanes and included Smu1 digests of *Staphylococcus aureus* NCTC 8325 in lanes 2, 8 and 14 as a reference standard and a Kpn1 digest of C. jejuni strain 153B-80 in lane 13 as a reproducibility control. Lanes 1 and 15 were left blank; the remaining lanes were used for study isolates.

**GelCompar II analysis**

The PFGE fingerprinting patterns were analysed with GelCompar II (version 1.01 for Windows; Applied Maths, Kortrijk, Belgium). Restriction fragments were identified visually and the PFGE patterns were normalised by interpolation to the nearest reference lane. The molecular sizes of the fragments detected for the study isolates were calculated based on the fragments of *S. aureus* NCTC 8325. Only fragments in the size range 80–674 kb were analysed; smaller fragments were not consistently resolved. Optimisation and a position tolerance of 1.0% were applied. Dice similarity coefficients were calculated based on pairwise comparison of the PFGE profiles of study isolates. The matrix of coefficients was used to generate dendrograms based on the unweighted pair group method using arithmetic averages (UPGMA).

**Criteria for defining clusters**

Three different sets of criteria were used to define clusters of related study isolates. (i) Isolates were considered to have closely related genotypes based on molecular typing if their PFGE profiles were related at ≥0.90, as determined by the GelCompar II analysis. (ii) Genotypically related isolates were considered to be clustered in space if they were cultured from patients whose infection was acquired in Québec province; infections acquired in a foreign country were excluded. (iii) Isolates that were genotypically and geographically related were considered to also be clustered in time if there were <3 months between sequential isolates.

**Results**

**Clinical epidemiology**

Among 111 cases (72 men and 39 women) reported to the DSP of Estrie, complete clinical information was available for 94 (85%) patients. Clinical information was available for only a few patients from Montréal and therefore is not described here.
The mean age of the cases was 25 years (range 10 months–85 years), with the highest proportion in the age groups 20–24 years and 45–64 years (17% in each category). Cases occurred year round, with a peak incidence between July and October. The majority of patients presented with diarrhoea, abdominal pain and fever (92%, 86% and 60% respectively); 25% of patients had bloody stools. Although most patients (73%) had symptoms for <2 weeks, 6% reported symptoms for ≥6 weeks. Four percent of patients were asymptomatic; all were children screened in the context of an international adoption. Overall, 47% of patients were treated with antibiotics and 12% were hospitalised for >24 h.

The most frequently identified suspected sources of infection were chicken (13%), raw milk (7%), a foreign trip (9%), a meal outside the home (10%), beef (5%), contaminated water (5%) and miscellaneous (10%). However, the source was either missing or unknown for 41% of patients and the validity of even the suspected sources is questionable because they were not confirmed by either microbiological analysis or by a case-control study.

**Characterisation of PFGE analyses**

Of the 154 isolates available, 153 were from human stools and one from a blood culture. More than 95% of the isolates were typable by PFGE; seven isolates were excluded from the analysis because PFGE patterns could not be obtained after multiple attempts with SmaI, KpnI or with both enzymes (two, three and two isolates, respectively). Thus, 147 isolates (88 from Estrie and 59 from Montréal) were analysed by both SmaI and KpnI. SmaI macrorestriction digests comprised 4–9 fragments in the size range analysed (674–80 kb), compared with 5–12 fragments with KpnI.

The *S. aureus* NCTC 8325 replicates were related at the level of 0.95 across 20 gels (n = 59 lanes) for SmaI digests and 0.94 among 26 gels (n = 75 lanes) for KpnI digests. Replicate profiles of *C. jejuni* strain 153B-80 were related at 0.94 for SmaI digests (n = 16) and 0.96 for KpnI digests (n = 23). Thus, technical reproducibility was excellent. As suitable sets of unequivocally epidemiologically related isolates were not available, biological reproducibility could not be evaluated [5].

Among the 147 isolates resolved by SmaI, the overall genetic diversity was 0.37, including 49 (33%) isolates with unique genotypes plus 27 clusters; in contrast, KpnI gave an overall diversity of 0.33 including 66 (45%) isolates with unique genotypes plus 29 clusters.

**Molecular epidemiology**

Table 1 summarises the effect of different epidemiological criteria on the distribution of the size of clusters of *C. jejuni* isolates identified by PFGE analysis of SmaI and KpnI digests. Molecular typing with SmaI alone defined 27 clusters comprising 98 isolates, compared with 29 clusters representing 81 isolates with KpnI. The addition of space criteria excluded nine isolates from the clusters defined by SmaI typing, compared with two isolates among those clustered by KpnI. The application of both space and time criteria excluded 27 (28%) of the isolates and 8 (30%) of the clusters putatively identified by SmaI typing, but only 14 (17%) isolates and 5 (17%) of the clusters defined by KpnI typing.

The study also examined the concordance between the typing results obtained with each of the endonucleases (Table 2). As previously, isolates were designated ‘genotypically related’ if the similarity coefficient

<table>
<thead>
<tr>
<th></th>
<th>Number of clusters identified by</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Typing alone</td>
</tr>
<tr>
<td>Number of related isolates in cluster</td>
<td>2</td>
</tr>
<tr>
<td>Number of clusters with SmaI</td>
<td>3</td>
</tr>
<tr>
<td>Number of clusters with KpnI</td>
<td>4</td>
</tr>
<tr>
<td>Number of clusters</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>18</td>
</tr>
<tr>
<td></td>
<td>2</td>
</tr>
<tr>
<td>Number of clusters with KpnI</td>
<td>3</td>
</tr>
<tr>
<td></td>
<td>4</td>
</tr>
<tr>
<td>Number of clusters</td>
<td>5</td>
</tr>
<tr>
<td>Total</td>
<td>6</td>
</tr>
<tr>
<td></td>
<td>7</td>
</tr>
<tr>
<td>Number of clusters</td>
<td>29</td>
</tr>
</tbody>
</table>
Table 2. Disagreement between clustering of isolates based on PFGE profiles of \( \text{KpnI} \) and \( \text{SmaI} \) digests

<table>
<thead>
<tr>
<th>Enzyme digest used to define clusters</th>
<th>Number of isolates assigned to clusters (^a)</th>
<th>Number (%) of isolates with concordant (^a) results for both enzymes</th>
<th>Number (%) of isolates with discordant (^b) results for the two enzymes</th>
</tr>
</thead>
<tbody>
<tr>
<td>( \text{SmaI} )</td>
<td>98</td>
<td>64 (65%)</td>
<td>34 (35%)</td>
</tr>
<tr>
<td>( \text{KpnI} )</td>
<td>81</td>
<td>75 (92%)</td>
<td>6 (8%)</td>
</tr>
</tbody>
</table>

\(^a\)Clusters were defined as isolates whose restriction digests had an SC \( \geq 90\% \).

\(^b\)Concordant isolates were isolates that were assigned to the same cluster by the first enzyme and also had an SC \( \geq 90\% \) with the second enzyme.

\(^c\)Discordant isolates were isolates that were assigned to the same cluster by the first enzyme, but whose restriction digests had an SC \( < 80\% \) with the second enzyme and were thereby assigned to either a different cluster or to no cluster.

An SC of their PFGE profiles was \( \geq 90\% \). Typing results for the two enzymes were considered discordant if isolates whose PFGE profiles had a similarity coefficient of 0.90 with one endonuclease (i.e., were considered clustered) had a similarity coefficient of 0.80 when analysed with the other enzyme. Among the 98 isolates assigned to clusters by \( \text{SmaI} \), only 64 (65\%) gave concordant results with \( \text{KpnI} \). Among the 34 isolates with discordant results, half were assigned to other clusters and half represented unique \( \text{KpnI} \) genotypes within the study population. In contrast, among the 81 isolates assigned to clusters by \( \text{KpnI} \), 75 (92\%) gave concordant results with \( \text{SmaI} \) (p < 0.0001, Fisher’s two-tailed exact test).

Fig. 1 illustrates an example of isolates with discordant typing results. Analysis of \( \text{SmaI} \) digests of the 147 study isolates defined a subset of 49 isolates that were related at the 0.73 level (Fig. 1a) and included a single large cluster of 18 isolates related at the 0.90 level. \( \text{KpnI} \) patterns of the same 49 isolates demonstrated considerably more diversity, with overall relatedness of 0.43 (Fig. 1b). Of note, the 18 isolates that were closely related by \( \text{SmaI} \) digests were distributed across the entire \( \text{KpnI} \) dendrogram, with six of the isolates no longer identified as closely related to any other isolate within this subgroup.

Discriminatory power was further assessed with 16 study isolates that were acquired outside Québec and, therefore, considered to be epidemiologically unrelated. The \( \text{SmaI} \) profiles of six of these isolates were closely related to isolates from Québec, compared with only one with \( \text{KpnI} \). The overall matrix of SC indicated that pairwise comparison of each of these 16 isolates with each of the isolates from Québec yielded a mean SC of 0.482 (95\% CI: 0.475–0.488) for \( \text{SmaI} \) digests compared with a mean SC of 0.392 (95\% CI: 0.387–0.398) for \( \text{KpnI} \) digests (p < 0.0001, Wilcoxon rank sum test).

Correlation between clinical and molecular epidemiology data

Clinical investigation in Estrie identified three cases of campylobacter enteritis that were possibly related to the ingestion of contaminated municipal water, although the source was never definitively established. \( \text{KpnI} \) typing confirmed that isolates from these three cases had closely related genotypes and identified two additional such isolates cultured within the same week. In contrast, \( \text{SmaI} \) typing included only one of these two additional cases within the cluster, and assigned the remaining one to a different cluster that proved to be inconsistent with time and space criteria.

Between July and October, six cases in Estrie were attributed to the ingestion of raw milk. The cases appeared epidemiologically unrelated, as they were all acquired on different farms. Two of these isolates were identified as belonging to a cluster by \( \text{SmaI} \) digests, but had an SC of 0.80 with \( \text{KpnI} \). Two other isolates belonged to a large cluster of 18 isolates identified by \( \text{SmaI} \) but were classified within two different clusters by \( \text{KpnI} \).

Discussion

PFGE is established as a highly reproducible and discriminatory molecular typing method for a wide range of clinically and epidemiologically significant bacterial pathogens [5]. However, the best strategy for applying PFGE to strain typing of \( \text{C. jejuni} \) has been controversial. The majority of recent reports have relied on \( \text{SmaI} \) [7, 8, 14, 15] and considered the results obtained with additional enzymes only for selected isolates [6, 11]. Some laboratories have reported that \( \text{KpnI} \) digests discriminate among isolates with indistinguishable or closely related \( \text{SmaI} \) profiles [6, 10, 11] and the use of more than one restriction endonuclease has been recommended for analysing such isolates [16, 17]. This study is the first to provide a quantitative analysis of a large set of \( \text{C. jejuni} \) isolates examined with both \( \text{KpnI} \) and \( \text{SmaI} \). The results demonstrate that \( \text{KpnI} \) digests provide results that are more discriminatory than \( \text{SmaI} \) and also more consistent with available epidemiological observations.

In assessing discriminatory power, it was observed that epidemiologically unrelated isolates which had indis-
Fig. 1. Dendrograms representing relatedness among PFGE profiles of (a) Smal and (b) KpnI digests of 49 isolates of C. jejuni. The analysis of the Smal digests of the 147 study isolates indicated that these 49 isolates were related at the 0.73 level and included a single large cluster of 18 isolates (●) related at the 0.90 level. Analysis of the KpnI patterns of the same 49 isolates demonstrated considerably more diversity, with overall relatedness of 0.43 (■). Moreover, the 18 isolates previously in a single cluster based by the Smal digests were distributed across the entire dendrogram, with 6 isolates no longer identified as closely related (0.90 level) to any other isolate in that subset ● clusters of isolates related at the 0.90 level, ■ the root of the dendrogram.
tistinguishable genotypes by SmuI digests could fre-
quently be differentiated by KpnI digests. This may
reflect the consistently greater number of restriction
fragments generated by KpnI or, alternatively, that
SmuI sites are preferentially found in C. jejuni within
loci with restricted genetic variation (e.g., ribosomal
operons). As lineages of pathogenic bacteria, including
campylobacters, can be distributed worldwide, there is
no a priori imperative that all geographically unrelated
isolates represent distinct genotypes. However, in prac-
tice, for a molecular typing method to be robust and
useful it must be able to differentiate >95% of
epidemiologically unrelated isolates [5]; only KpnI
digests met this criterion. Another useful quantitative
indication of the discriminatory power of a typing
system is the total genetic diversity resolved across a
collection of isolates and KpnI resolved appreciably
more genetic diversity among the isolates studied.

This study used two approaches to evaluate the validity
of the molecular typing results. First, the PFGE
analyses were combined with basic epidemiological
criteria, i.e., relatedness in time and space. Among 147
isolates, 98 (67%) were assigned to 27 clusters based
on SmuI profiles, and 81 (55%) isolates were assigned
to 29 clusters based on KpnI digests. However, 8
(30%) of the putative SmuI clusters were inconsistent
with the epidemiological criteria compared with 5
(17%) of the KpnI clusters. Overall, the total number
of clustered isolates decreased to 28% and 17%,
respectively, when epidemiological criteria were com-
bined with the results of strain typing.

This study also specifically examined isolates that had
closely related restriction profiles with one endonu-
cl ease and re-analysed them with the other endonu-
cl ease. More than one-third of isolates that appeared
closely related on the basis of SmuI digests were
unrelated when analysed by KpnI digests. Moreover,
the 18 isolates assigned to a single large cluster based
on SmuI digests were distributed across the entire
dendrogram defined by KpnI, with six of the isolates
no longer identified as closely related to any other
isolate within this subgroup. Finally, clusters that were
epidemiologically related to ingestion of raw milk and
specific water sources correlated better with the typing
results based on KpnI than SmuI. Taken together, these
observations strongly support the conclusion that KpnI
is the endonuclease of choice for molecular epidemio-
logical studies of C. jejuni.

The combination of KpnI typing and epidemiological
(space and time) criteria assigned putative clusters for
49% of isolates from Estrie and 39% of isolates from
Montréal. This suggests that common source outbreaks
of C. jejuni may be substantially more frequent than
has been suspected on the basis of conventional epi-
demiological data alone. However, validation of these
clusters as instances of recent transmission or common
source outbreaks requires additional epidemiological
evidence (e.g., case-control study) or molecular typing
of a source isolate. The available epidemiological data
were insufficient to confirm or refute these putative
clusters, as in 41% of the cases, sources were either
missing or explicitly unknown and even those sources
proposed were subject to recall bias. Further insights
into the epidemiology of sporadic cases of campylo-
bacter enteritis will require prospective application of
both clinical and molecular investigations.

This study has several notable limitations. A set of
isolates representing an unambiguous outbreak was not
available; consequently, the biological reproducibility
of C. jejuni molecular typing by SmuI and KpnI
digests could not be formally evaluated. Because of the
unreliability of the suspected sources, the study could
not demonstrate unambiguously that the greater dis-
criminatory power of KpnI yielded more appropriate
epidemiological data. A case-control study combining
both clinical and molecular epidemiology data is
needed to resolve this issue.

Most of the observed clusters represented two or three
isolates. However, this estimate is likely to be arti-
ficially low. First, the study analysed only 147 isolates
of the 3500 C. jejuni isolates cultured annually in
Québec province. Furthermore, many people with
enteritis do not seek medical attention and, even
among those, only some will have a stool specimen
cultured for enteric pathogens. A recent study suggests
that among eight cases of campylobacter enteritis in the
community only one is included in national surveil-
lance data [4]. Thus, the small clusters observed
probably represent only a fraction of the epidemiolo-
gically related cases. The time criterion clearly
impacted on both the number and the size of the
clusters identified. A more stringent requirement, i.e.,
requiring only 2 or 3 weeks between two sequential
genetically related isolates, would have resulted in
fewer and smaller clusters, but would not have changed
the overall conclusions.

To facilitate studying this relatively large collection of
isolates, a computer-based analysis of scanned gel
images was applied, rather than visual comparison
based on the Tenover criteria [18]. Several issues that
proved critical to success with this approach must be
emphasised. In each gel, only 13 of 15 available lanes
were used, leaving the outermost lanes empty to mini-
mise distortion artifacts. The study included (a) a
molecular size standard in every fifth lane, so that
when normalising the patterns no isolate was more than
two lanes from a standard, and (b) an additional in-
dependent standard in every gel so that technical
reproducibility could be quantified across the entire
series. It is essential that this reproducibility standard
is distinct from the molecular size standard; it is helpful
to use an isolate of the same species being analysed.
The reproducibility patterns were invaluable in validat-
ing the user-adjustable parameters present in the ana-
lytical software; these parameters have a strong influence on the quantitative results and should be noted explicitly.

In conclusion, PFGE analysis of KpnI digests of C. jejuni is more discriminatory and epidemiologically informative than analysis of SmalI digests. Computer analysis of the patterns greatly facilitates the identification of related genotypes among large sets of isolates. As campylobacters are very widely distributed in nature, a combination of approaches will probably be required to identify sources of infection accurately. We propose the use of molecular typing to define clusters of genotypically related isolates, followed quickly by clinical investigation and case-control studies to identify specific common exposures and, finally, selective culturing of food, animal and environmental specimens and molecular typing of the isolates obtained to confirm specific sources. Continuous prospective surveillance studies with this strategy are needed to define the mechanisms of community-acquired C. jejuni infection and ultimately to develop new, effective approaches to prevention.

This work was sponsored by the Medical Research Service of the Department of Veterans Affairs (RDA) and by the Centre de Recherche Medecale de l’Universite de Sherbrooke (S.M.). We thank Boris Ehrenstein for invaluable assistance with the computer analysis, Janet Senott from the Laboratory Institute of the Massachusetts Department of Public Health for confirming the identification of some C. jejuni isolates, David Carroll for his support and useful suggestions and Alain Lajeunesse for his technical expertise in preparing the figure.

References

3. Blaser MJ. Epidemiologic and clinical features of Campylo-

8. Shi ZY, Liu PY, Lau YL, Lin YH, Hu BS, Tsai HN. Comparison of polymerase chain reaction and pulsed-field gel electrophoresis for the epidemiological typing of Campylo-