Journal of Applied Microbiology

Journal of Applied Microbiology ISSN 1364-5072

ORIGINAL ARTICLE

Multilocus variable number tandem repeat analysis as a tool to discern genetic relationships among strains of *Yersinia enterocolitica* biovar 1A

P. Gulati[1], R.K. Varshney[2] and J.S. Virdi[1]

1 Microbial Pathogenicity Laboratory, Department of Microbiology, University of Delhi South Campus, New Delhi, India
2 Applied Genomics Laboratory, International Crops Research Institute for the Semi-Arid Tropics (ICRISAT), Patancheru, Andhra Pradesh, India

Keywords
epidemiology, India, MLVA, VNTR, *Yersinia enterocolitica*.

Correspondence
Jugsharan S. Virdi, Microbial Pathogenicity Laboratory, Department of Microbiology, University of Delhi South Campus, Benito Juarez Road, New Delhi – 110 021, India.
E-mail: virdi_dusc@rediffmail.com

2008/1428: received 19 August 2008, revised 10 December 2008 and accepted 20 January 2009

doi:10.1111/j.1365-2672.2009.04267.x

Abstract

Aims: To identify variable number tandem repeat (VNTR)-containing loci, and to use multilocus VNTR (MLVA) to discern genetic relationships among strains of *Yersinia enterocolitica* biovar 1A isolated from diverse sources.
Methods and Results: The whole genome sequence of *Y. enterocolitica* 8081 was analysed and eight VNTR loci with repeat sizes between 4 and 9 bp, and each containing more than four repeat copies were selected for MLVA typing of 88 strains of *Y. enterocolitica*. Of these, four loci were polymorphic and generated 26 MLVA genotypes among 81 strains of *Y. enterocolitica* biovar 1A. MLVA was found to be quite discriminatory (DI = 0·87). Cluster analysis and population modelling using minimum spanning tree (MST) clearly clustered *Y. enterocolitica* biovar 1A into two major groups.
Conclusions: The MLVA is easy to perform and can be used to discern clonal relationship among strains of *Y. enterocolitica*. Also the phylogenetic relationships obtained with MLVA genotypes were in good agreement with those established by other typing methods.
Significance and Impact of the Study: The MLVA method reported is relatively more discriminatory than the other genotyping methods and has the potential to be used as an epidemiological tool for the study of strains of *Y. enterocolitica* biovar 1A.

Introduction

Yersinia enterocolitica, an important food- and water-borne gastrointestinal agent is known to cause a variety of syndromes ranging from mild gastroenteritis to more invasive diseases like terminal ileitis, mesenteric lymphadenitis mimicking appendicitis and septicaemia (Robins-Browne 1997; Bottone 1999). Blood transfusion-related septicaemia due to *Y. enterocolitica* has been reported to have high mortality (Leclercq *et al.* 2005).

Yersinia enterocolitica is quite heterogeneous biochemically, serologically and pathogenically, and has been classified into six biovars (1A, 1B, 2, 3, 4 and 5) and more than 50 serotypes. As regards its pathogenicity, *Y. enterocolitica* can be broadly divided into three groups (Cornelis *et al.* 1987) viz. highly pathogenic strains (biovar 1B), strains with moderate pathogenicity (biovars 2–5) and the biovar 1A strains, which have generally been regarded as avirulent (Bottone 1999). Nevertheless, some biovar 1A strains produce symptoms indistinguishable from that produced by pathogenic biovars (Morris *et al.* 1991; Burnens *et al.* 1996). Biovar 1A strains have also been implicated in nosocomial (Ratnam *et al.* 1982) and food-borne outbreaks (Greenwood and Hooper 1990; Butt *et al.* 1991), and were isolated from extra-intestinal infections (Bissett *et al.* 1990). The biovar 1A strains are highly heterogeneous encompassing a plethora of serotypes and nonagglutinable (NAG) forms, and have been isolated from diverse sources (Tennant *et al.* 2003). However, analysis of biovar 1A strains by REP-PCR, ERIC-PCR

genomic fingerprinting (Sachdeva and Virdi 2004), ribotyping, PCR-ribotyping and *gyr*B gene polymorphism (Gulati and Virdi 2007) generally revealed limited genetic heterogeneity. Aside from clinical and wastewater O:6,30–6,31 (biovar 1A) isolates clustering into two separate groups (Sachdeva and Virdi 2004; Gulati and Virdi 2007), these studies failed to reveal any association between serotypes and the sources of isolation. It is, therefore, imperative to analyse biovar 1A by other methods to reveal such associations, if any.

The accumulating genomic data shows that variable number of tandem repeats (VNTR) are frequent in bacteria and loci containing these are polymorphic (van Belkum *et al.* 1997). The value of VNTRs as an epidemiological tool (van Belkum 1999) has been appreciated and multilocus variable number tandem repeat analysis (MLVA) has shown to be a promising method for genotyping several pathogenic bacteria (Lindstedt 2005; García-Yoldi *et al.* 2007; Kimura *et al.* 2008; Sperry *et al.* 2008). MLVA has been used to characterize several bacterial species that are genetically homogeneous and difficult to type by other means (Johansson *et al.* 2004; Lowell *et al.* 2005). Recently, Gierczynski *et al.* (2007) described a MLVA typing scheme for *Y. enterocolitica* subspecies *palearctica* bioserotype 4/O:3 using the draft genome of *Y. enterocolitica* Y11 (serotype O:3, biovar 4). Overall, the data showed that MLVA has high-discriminatory power but failed to establish the clonal nature of *Y. enterocolitica* O:3.

The main objective of the present study was to identify VNTR-containing loci using the whole genome sequence of *Y. enterocolitica* 8081, and to evaluate the ability of multilocus VNTR analysis (MLVA) to discern genetic relationship among strains of *Y. enterocolitica* biovar 1A.

Materials and methods

Bacterial strains

MLVA was performed with 81 strains of *Y. enterocolitica* biovar 1A isolated from diarrhoeic human stools (51 strains), wastewater (18 strains), pig throat (7 strains) and pork (5 strains). These included strains isolated in India (65 strains), Europe (15 strains) and the USA (one strain). Details of these strains namely serotype, source of isolation, country of origin and WHO reference laboratory accession numbers have been reported previously (Sachdeva and Virdi 2004). Reference strains *viz. Y. enterocolitica* 8081 (biovar 1B, serotype O:8); W22703, IP383 and IP26329 (biovar 2, serotype O:9); IP26249 (biovar 2, serotype O:5,27) and IP134 and IP26332 (biovar 4, serotype O:3) were also included in this study. *Y. enterocolitica* 8081 was provided by Mikael Skurnik (Department of Bacteriology and Immunology, Haartman Institute, University of Helsinki, Finland) while strain W22703 was obtained from G.R. Cornelis (Universite Catholique de Louvain, Belgium). All reference strains prefixed with IP were obtained from E. Carniel (*Yersinia* National Reference Laboratory and WHO collaborating Center, Institut Pasteur, Paris, France). All isolates were maintained on trypticase soy agar (TSA) slants at 4°C along with glycerol stocks (50% v/v) at −40°C. The strains were also maintained in lyophilized state. The purity of these isolates was checked periodically on *Yersinia* selective medium, i.e. cefsulodin-irgasan-novobiocin (CIN) agar.

Identification and selection of VNTR-containing loci

The published *Y. enterocolitica* 8081 genome sequence (Thomson *et al.* 2006) was accessed at the Sanger centre web server (http://www.sanger.ac.uk/Projects/Y_enterocolitica). Genomic sequence data was analysed for the presence of VNTR-containing loci using tandem repeats finder (Benson 1999; Denoeud and Vergnaud 2004) and microsatellite (SSRs) identification software (Jayashree *et al.* 2006). The results obtained from these were combined for detailed analysis of the occurrence of VNTRs in the genome of *Y. enterocolitica* 8081. After identification of tandem repeat containing loci, eight such loci each with short repeat unit length (4–9 bp) and repeated at least four times in tandem, were selected for multilocus-VNTR typing. To assess the functions associated with these loci, the recently annotated *Y. enterocolitica* genome was accessed (GenBank acc. no. AM286415) in addition to the homologies identified using Blastn and Blastx searches. Details of the VNTR loci selected for the present study along with the putative functions associated with these are described in Table 1.

DNA extraction and multilocus variable number tandem repeat analysis-PCR

Genomic DNA was extracted from cultures grown overnight at 28°C, using DNeasy tissue kit (Qiagen, Hilden, Germany) as per the manufacturer's instructions.

MLVA-PCR primers were designed using PrimerSelect software of the DNASTAR Package. For each of the VNTR loci selected, primers were designed for amplification and sequencing, such that amplicon of 200–300 bp containing the tandem repeat along with its flanking regions was obtained. Details of the PCR primers and amplification conditions are listed in Table 1. Optimized VNTR PCRs were performed for each of the selected loci in 25 μl reaction volume. The reaction mixture containing 50 ng of genomic DNA, 10 pmol of each primer,

Table 1 Characteristics of the VNTR loci, amplification conditions and the number of VNTR alleles obtained for *Y. enterocolitica* biovar 1A

VNTR locus	Location*	Function	Repeat sequence (5′–3′)	Primers used	Primer sequence	Annealing temperature (°C)	Amplicon size†	No. of alleles ‡	No. of repeats ‡		Nei's diversity index (D)
									Min.	Max.	
VNTR01	2203556–2203619	*Orf*528, GTP binding protein	CCAGCA	TR1 TR2	ACCGGCACTGAATCTGAAGATGC CTTATCTTCCATAATCTAGACCTC	55	–	7	4	10	0·73
VNTR02	1106385–1106522	Relaxase domain	CCAGCA	TR3 TR4	GCGTTTCCCTAGCTGTCTTGATTA GCAGCTTAACTACAACCCAGACCA	50	227	2	12	23	0·02
VNTR03	3036006–3036504	*San*A putative exporter protein	TGGACG	V5 V6	ATGAGCCGACTGGATGAGGATGA TGGAACACCGGTGGCAGAAGAA	54	260	11	2	11	0·79
VNTR04	2849331–2849385	*Orf*16 and *Orf*17 Hypothetical protein	TGCCACT	V7 V8	AGCTCCATTCCCCATTCTTGAT ATGCCTCTGATTCTATGCTTCTCC	51·5	332	1	–	7	–
VNTR05	3530219–35530274	Sugar transporter gene	ATCAACC	V9 V10	GGTTGAAGCCGCAGAGGTAAG AGTGGCTGGGCATAATAGTTG	52	355	9	1	46	0·54
VNTR06	943156–9432195	Putative sugar ABC transporter periplasmic protein	TATCGG	V11 V12	AGGCCAAAAGCTAAGGGGATAA CACTAAAATGGCACTAAGGCTGAT	52	224	1	–	4	–
VNTR07	2324500–2324543	Phage shock protein	ATCAAGGGA	Vi5 V16	TTGTTCAAACCCGCTCTCAC TGGCAATAACTTTAAACATCTCAT	54	264	1	–	4	–
VNTR08	1054682–1054707	Putative ABC transporter protein	TATT	V21 V22	CTTGGCATTTTCACTGCTCA CCATGCCTCTGATACGTGGC	53	257	3	0	4	0·10

*In *Y. enterocolitica* 8081 sequence (GenBank acc. no. AM286415).

†For *Y. enterocolitica* 8081.

‡Including data for *Y. enterocolitica* 8081 (reference strain).

200 μmol l^{-1} of dNTP mix, 1·5 mmol l^{-1} $MgCl_2$ and 2 U DNA polymerase (DyNAzyme^{TM}, Finnzymes, Finland) was used. PCR amplifications were performed in a PTC-100^{TM} Thermal Cycler with the following temperature profile: initial denaturation at 95°C for 10 min; followed by 35 cycles each of denaturation at 95°C for 45 s, annealing (50–55°C) for 1 min and extension at 72°C for 1 min, and a final extension at 72°C for 10 min. All steps in the PCR thermocyling programme except the annealing temperatures were identical for all the loci studied.

Ten micro litres of the PCR products obtained were size-fractionated by electrophoresis on 3% agarose gel in 1× TBE buffer at a constant voltage of 2·5 V cm^{-1} for 6 h and visualized after staining with ethidium bromide (0·5 μg ml^{-1}). In order to obtain better resolution, VNTR08 amplicons were separated on 15% PAGE. 50 or 100 bp DNA ladder (MBI Fermentas) was used as the size standard marker.

DNA sequencing of VNTR amplicons

Sequencing of the selected (representative) samples was performed, in order to verify that the differences observed in amplicon sizes were due to variability in the tandem repeat region rather than any other genetic event. For sequencing, the amplified DNA product from a single 50 μl reaction was purified using QIAquick PCR purification kit. The purified PCR products were sequenced using the ABI Prism 3730 automated genetic analyser by the dye termination method. Once the sequences were obtained and aligned, the number of the repeats was counted. The sequences have been submitted to GenBank under the accession numbers DQ976374–78.

Phylogenetic and statistical analysis of MLVA-PCR profiles

VNTR gel images were captured and stored as TIFF files using Gel Doc 2000 (Bio-Rad). These were analysed with Diversity Database software (version 2, Bio-Rad) and the number of repeats at each locus was deduced from the size of the amplicon, using the tandem repeat calculator tool of this database. A locus having different number of tandem repeats was designated as separate allele.

Clustering of the MLVA types was performed with BioNumerics software (ver. 4.61, Applied Maths, Kortrijk, Belgium) using categorical coefficient of similarity to estimate genetic differences and a dendrogram was constructed. The number of the repeat units for each locus was saved as the 'character type' data in BioNumerics software and then subjected to cluster analysis using minimum spanning tree (MST).

The Nei's diversity index (D value) was calculated for each locus as $1 - \Sigma$ (allele frequency)^2 (Weir 1990).

Comparison of MLVA with other genotyping methods

In order to compare the discriminatory ability of MLVA with other genotyping methods reported earlier, Simpson's diversity index (DI) and 95% confidence intervals were calculated (Hunter and Gaston 1988; Grundmann *et al.* 2001).

Results

Identification and characterization of tandem repeat containing loci

Yersinia enterocolitica 8081 genome was analysed for the presence of tandem repeats. In all, 701 (tandem repeat finder) and 237 (SSR identification software) loci that were 3–202 bp long were identified. The copy number of the repeats ranged from 2 to 28. Pentameric (5 bp) repeats were absent in the genome of *Y. enterocolitica* 8081. The reports summarizing the occurrence and distribution of these VNTRs are given in Supporting Information. Eight VNTR loci with repeat unit size between 4 and 9 bp, containing more than four copies each and 100% homology were selected for multilocus-VNTR-based typing of 88 (81 biovar 1A and 7 reference) strains of *Y. enterocolitica*. These loci were designated VNTR01-VNTR08. Blastn and Blastp searches revealed that these loci were generally located in the open reading frame of structural genes (Table 1). Eight separate PCR assays were developed to enumerate the number of repeats at each locus. The amplification was robust and yielded reproducible results for all the loci studied. The number of repeats at each locus was deduced from the size of the amplicon, using the tandem repeat calculator tool of Diversity Database. To verify that PCR amplicon size variations resulted from differences in repeat copy number, a randomly selected allele was sequenced for each locus using the ABI Prism 3730 automated genetic analyser. The number of repeat units counted after sequencing coincided with the copy number obtained from the amplicon size on gel using the tandem repeat calculator software.

Comparison of VNTR alleles among strains of *Y. enterocolitica* biovar 1A

Polymorphism was observed in four loci viz. VNTR01, VNTR03, VNTR05 and VNTR08 and the variations are summarized in Table 1.

Analysis of VNTR01 locus in *Y. enterocolitica* biovar 1A revealed seven different alleles with 4–10 repeats with allele

having eight copies being most predominant. For this locus, 17 isolates failed to yield PCR amplification and were regarded to possess a null allele. PCR amplification of VNTR02 yielded amplicon in just two strains viz. *Y. enterocolitica* 8081 (biovar 1B) and strain 54 (biovar 1A) with 23 and 12 repeats, respectively. For VNTR 03, 05 and 08 loci, PCR amplicons were obtained for all the isolates investigated. For VNTR04, VNTR06 and VNTR07 amplicon was observed in *Y. enterocolitica* 8081 only. As VNTR 02, 04, 06 and 07 loci were absent in most of the isolates, data from these were not included for cluster analysis.

Each isolate has been depicted as MLVA genotype (Fig. 1) or allele string representing number of repeats obtained at the four loci viz. VNTR01, VNTR03, VNTR05 and VNTR08. VNTR03 was the most diverse ($D = 0{\cdot}79$) and thus may have potential for epidemiological analyses. Eleven different alleles containing 2–10 repeats were obtained for this VNTR locus. The reference strain *Y. enterocolitica* 8081 contained an allele with 11 repeats. VNTR03 allele carrying five repeats was present predominantly in isolates belonging to 'clonal group B' as discerned in an earlier work by REP- and ERIC-PCR in the authors' laboratory (Sachdeva and Virdi 2004). Strains carrying VNTR03 with two repeats comprised mainly O:6,30–6,31 isolates from wastewaters. VNTR05 showed the largest range of repeat numbers (number of repeats ranged from 1 to 46) with nine distinct alleles. For VNTR05, all seven reference strains generated alleles distinct from that produced by biovar 1A strains. VNTR08 showed presence of only three alleles with 95% (77 of 81) of the isolates possessing allele that contained two repeats.

With five isolates, the VNTR03 PCR yielded two amplicons differing in size by four repeats ($n = 2$, $n = 6$) and one strain generated two amplicons varying in size by three repeats ($n = 2$, $n = 5$). Similarly for VNTR05, 17 strains produced two amplicons (15 and 27 repeats) while one strain produced two fragments containing 35 and 46 repeats. Double bands may appear either due to errors during PCR amplification or by the presence of multiple alleles of a given sequence. Erroneous amplification was, however, ruled out, as the bands were highly discrete and reproducible. Furthermore, sequencing of the amplicons showed that the only difference between the two copies was in the number of repeats.

The Nei's diversity index (D value) that reflects discriminatory potential of a marker was calculated for each of the VNTR locus. The D values for the VNTR loci in this study ranged from 0·02 (VNTR02) to 0·79 (VNTR03) (Table 1).

Reproducibility of MLVA profile

To determine the reproducibility and stability of the MLVA genotypes, five strains of *Y. enterocolitica* biovar 1A were subcultured serially in typticase soy broth and randomly typed using VNTRs. Indistinguishable MLVA genotypes were generated for VNTR01, VNTR03, VNTR05 and VNTR08 (data not shown) up to six subcultures.

MLVA-based cluster analysis

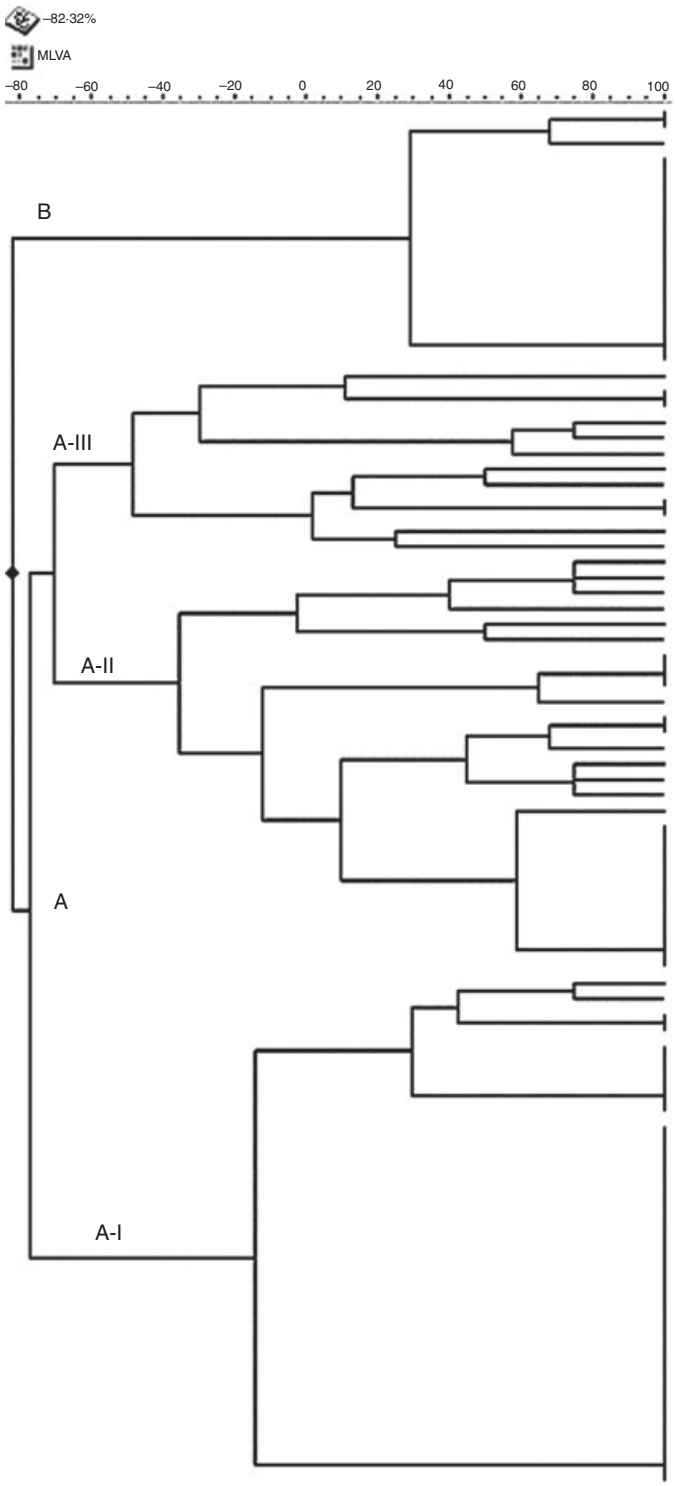
A total of 33 MLVA genotypes were observed among 88 strains of *Y. enterocolitica*. Twenty-six MLVA genotypes were discerned among 81 biovar 1A strains (Fig. 1) belonging to diverse sources, serotypes and geographical origin. Sixteen epidemiologically unrelated European biovar 1A strains generated 14 distinct MLVA genotypes and were more heterogeneous compared to Indian isolates.

The genetic relationship among 88 strains of *Y. enterocolitica* using polymorphism at four VNTR loci was examined by cluster analysis and a dendrogram was constructed (Fig. 1). MLVA clustered *Y. enterocolitica* biovar 1A strains clearly into two major groups – A and B. Group A included most (64 of 81) of the isolates of *Y. enterocolitica* biovar 1A, which was further divided into subgroups namely A-I, A-II and A-III. Subgroup A-I exclusively comprised of all the human serotype O:6,30–6,31 and O:6,30 isolates along with the three pig and one pork isolates. On the other hand, all wastewater serotype O:6,30–6,31 isolates generated identical MLVA genotype and clustered in subgroup A-II. All the reference strains (biovar 1B, 2 and 4) except one grouped in subgroup A-III. Genotypically group-B was found to be very homogeneous. All Indian isolates in this group were identical in their MLVA profiles.

Population modelling using the MST corroborated the two groups – A and B (Fig. 2) among strains of *Y. enterocolitica* biovar 1A. Each of the two groups was represented by a predominant MLVA genotype. The other MLVA genotypes radiated from these two dominant ancestral profiles. MST also revealed that in group-A, wastewater O:6,30–6,31 isolates represent the ancestral strains while the clinical O:6,30–6,31 isolates occupy a radial position. Three of the reference (biovar 2) isolates were related to group-A; the other biovar 2 reference strain along with the biovar 4 reference strains clustered together in a separate group. The single biovar 1B reference strain was closer to group-B (Fig. 2).

Comparison of MLVA with other genomic typing methods

The discriminatory ability of MLVA was compared with other genotyping methods by calculating simpson's diversity index (Table 2). The discriminatory ability of MLVA



Strain no	Biovar	Serotype	Source	Country of origin	MLVA genotype	REP-PCR†	ERIC-PCR†
51	1A	O:10-34	Human	France	null-6-15,27-2	R17	E21
53	1A	O:10-34	Human	France	null-6-15,27-2	R19	E22
52	1A	O:10-34	Human	France	null-7-15,27-2	R18	E22
47	1A	O:10-34	Water	India	null-5-15,27-2	R2	E4
48	1A	O:10-34	Water	India	null-5-15,27-2	R2	E4
50	1A	O:10-34	Water	India	null-5-15,27-2	R2	E4
62	1A	O:7,8-8-8,19	Pork	India	null-5-15,27-2	R2	E4
63	1A	O:7,8-8-8,19	Pork	India	null-5-15,27-2	R2	E4
22	1A	O:6,30	Human	India	null-5-15,27-2	R2	E4
34	1A	O:6,30	Human	India	null-5-15,27-2	R2	E4
37	1A	O:6,30	Human	India	null-5-15,27-2	R2	E4
5	1A	O:6,30-6,31	Human	India	null-5-15,27-2	R2	E4
68	1A	ND	Swine	India	null-5-15,27-2	R2	E4
69	1A	ND	Swine	India	null-5-15,27-2	R2	E4
71	1A	ND	Swine	India	null-5-15,27-2	R2	E4
60	1A	NAG	Water	India	null-5-15,27-2	R2	E4
78	1A	NAG	Human	India	null-5-15,27-2	R2	E4
8081	1B	O:8	Human	USA	null-11-8-4	ND	ND
64	1A	O:7,8-8-8,19	Pork	India	4-7-27-4	R10	E11
65	1A	NAG	Pork	India	4-7-27-4	R10	E11
IP134	4	O:3	Human	Europe	6-9-46-1	ND	ND
IP26332	4	O:3	Human	Europe	6-9-45-1	ND	ND
IP383	2	O:9	Human	Europe	6-7-46-1	ND	ND
10	1A	O:6,30-6,31	Human	France	10-7-3-2	R1	E1
11	1A	O:6,30-6,31	Human	France	10-8-0-2	R1	E1
56	1A	O:41,42	Water	India	10-8-1-0	R7	E8
57	1A	O:41,43	Water	India	10-8-1-0	R8	E8
IP26329	2	O:9	Human	Europe	8-8-45-3	ND	ND
IP26249	2	O:5,27	Human	Europe	6-10-43-3	ND	ND
44	1A	O:6,30	NK	NK	4-6-1-2	R13	E17
54	1A	O:5	Human	NK	4-4-1-2	R14	E19
45	1A	O:6,30	NK	NK	4-10-1-2	R1	E18
41	1A	O:6,30	Human	France	7-5-1-2	R12	E15
38	1A	O:6,30	Human	France	6-4-6-2	R1	E15
W22703	2	O:9	Human	Europe	6-10-47-2	ND	ND
76	1A	NAG	Human	India	5-2-3-2	R1	E12
80	1A	NAG	Human	India	5-2-3-2	R1	E13
81	1A	NAG	Human	India	5-2-3-2	R11	E14
79	1A	NAG	Human	India	5-2-2-2	R1	E13
42	1A	O:6,30	Human	Germany	4-5-3-2	R1	E16
43	1A	O:6,30	Human	Germany	4-5-3-2	R2	E16
39	1A	O:6,30	Human	France	4-5-2-2	R15	E20
40	1A	O:6,30	Human	France	4-2-3-2	R16	E2
46	1A	O:10-34	Water	India	4-2,5-3-2	R5	E7
55	1A	O:15	Water	India	4-6-3-2	R2	E23
58	1A	O:6,31	Water	India	4-2-35,46-2	R7	E8
59	1A	NAG	Water	India	4-2-3-2	R8	E8
72	1A	ND	Swine	India	4-2-3-2	R1	E2
49	1A	O:10-34	Water	India	4-2-3-2	R6	E2
13	1A	O:6,30-6,31	Water	India	4-2-3-2	R1	E6
14	1A	O:6,30-6,31	Water	India	4-2-3-2	R1	E6
15	1A	O:6,30-6,31	Water	India	4-2-3-2	R1	E2
16	1A	O:6,30-6,31	Water	India	4-2-3-2	R1	E2
17	1A	O:6,30-6,31	Water	India	4-2-3-2	R1	E2
18	1A	O:6,30-6,31	Water	India	4-2-3-2	R1	E2
19	1A	O:6,30-6,31	Water	India	4-2-3-2	R1	E2
9	1A	O:6,30-6,31	Human	France	8-8-3-2	R1	E1
12	1A	O:6,30-6,31	Human	USA	6-8-3-2	R9	E1
73	1A	NAG	Human	India	8-2-3-2	R1	E2
74	1A	NAG	Human	India	8-2-3-2	R1	E2
75	1A	NAG	Human	India	8-2,6-3-2	R1	E1
77	1A	NAG	Human	India	8-2,6-3-2	R1	E1
61	1A	O:7,8-8-8,19	Pork	India	8-2,6-3-2	R1	E2
27	1A	O:6,30	Human	India	8-2,6-3-2	R1	E1
28	1A	O:6,30	Human	India	8-2,6-3-2	R1	E1
1	1A	O:6,30-6,31	Human	India	8-6-3-2	R1	E1
2	1A	O:6,30-6,31	Human	India	8-6-3-2	R1	E1
3	1A	O:6,30-6,31	Human	India	8-6-3-2	R1	E1
4	1A	O:6,30-6,31	Human	India	8-6-3-2	R3	E3
6	1A	O:6,30-6,31	Human	India	8-6-3-2	R4	E5
7	1A	O:6,30-6,31	Human	India	8-6-3-2	R1	E1
8	1A	O:6,30-6,31	Human	India	8-6-3-2	R1	E1
20	1A	O:6,30	Human	India	8-6-3-2	R1	E1
21	1A	O:6,30	Human	India	8-6-3-2	R1	E1
23	1A	O:6,30	Human	India	8-6-3-2	R1	E1
24	1A	O:6,30	Human	India	8-6-3-2	R1	E1
25	1A	O:6,30	Human	India	8-6-3-2	R1	E1
26	1A	O:6,30	Human	India	8-6-3-2	R1	E1
29	1A	O:6,30	Human	India	8-6-3-2	R1	E1
30	1A	O:6,30	Human	India	8-6-3-2	R1	E1
31	1A	O:6,30	Human	India	8-6-3-2	R1	E1
32	1A	O:6,30	Human	India	8-6-3-2	R1	E1
33	1A	O:6,30	Human	India	8-6-3-2	R1	E1
35	1A	O:6,30	Human	India	8-6-3-2	R1	E1
36	1A	O:6,30	Human	India	8-6-3-2	R1	E1
66	1A	ND	Swine	India	8-6-3-2	R1	E1
67	1A	ND	Swine	India	8-6-3-2	R1	E1
70	1A	ND	Swine	India	8-6-3-2	R1	E1

Figure 1 Dendrogram deduced from cluster analysis of 88 isolates of *Y. enterocolitica* using MLVA genotypes. The similarities between MLVA genotypes were performed by categorical coefficient and clustering was performed using the Ward's parameter. The MLVA genotype (allele string) represents the number of repeats obtained at the four loci depicted as VNTR01-03-05-08. In the MLVA genotype, for the isolates where two amplicons were obtained for a locus, these are separated by comma. 'Null' represents allele when no PCR amplification was observed. NAG, nonagglutinable; ND, not determined; NK, Not known, †See Sachdeva and Virdi (2004).

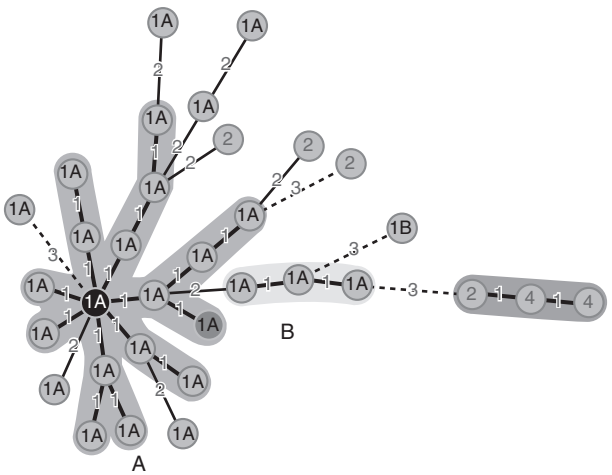


Figure 2 Population modelling by minimum spanning tree (MST) using MLVA data. Each circle in the tree represents a different MLVA type; letter in the circle denotes biovar of the isolate. The MLVA data separates the isolates into two distinct MST tree branches – A and B. Biovars 2 and 4 form a separate group. The thickness and the dotting of the lines indicate the distance between the circles. A thicker line denotes closer distance than a thin line, and a thin line denotes closer distance than a dotted line. The number on each line denotes the number of loci that differ between the MLVA types. The black circle indicates wastewater serotype O:6,30–6,31 isolates, while the grey circle indicates clinical serotype O:6,30–6,31 isolates.

Table 2 Discriminatory indices (DI) and 95% confidence interval of various genotyping techniques for *Y. enterocolitica* biovar 1A

Genotyping method	Simpson's index of diversity (DI)	Confidence interval (95% CI)	Reference
MLVA	0·87	0·85–0·90	This study
rep-PCR *	0·84	0·81–0·88	Sachdeva and Virdi (2004)
ERIC-PCR†	0·82	0·79–0·86	Sachdeva and Virdi (2004)
REP-PCR ‡	0·68	0·63–0·73	Sachdeva and Virdi (2004)
*gyr*B-RFLP	0·75	0·72–0·77	Gulati and Virdi (2007)
PCR-Ribotyping	0·68	0·66–0·70	Gulati and Virdi (2007)
Ribotyping	0·72	0·69–0·74	Gulati and Virdi (2007)

*Repetitive elements PCR.
†Enterobacterial repetitive intergenic consensus PCR.
‡Repetitive extragenic palindromic PCR.

(DI = 0·87) was highest among all the techniques used to genotype *Y. enterocolitica* biovar 1A strains (Sachdeva and Virdi 2004; Gulati and Virdi 2007).

Discussion

The majority of the tandem repeats identified in the genome of *Y. enterocolitica* 8081 was between 4 and 40 bp and amenable to analysis by PCR and agarose gel electrophoresis. Several reports show that simple tandem repeats are quite diverse (Keim *et al.* 2000; Farlow *et al.* 2001) as slip strand repair mutations occur more frequently on short repeats (van Belkum *et al.* 1997). Thus, in the present study, eight VNTR loci with repeat unit size between 4 and 9 bp were selected for multilocus-VNTR-based typing of 88 (81 biovar 1A and seven reference) strains of *Y. enterocolitica*. The absence of pentameric repeats was consistent with a previous report suggesting that such repeats were not ubiquitous and occurred in relatively limited number of bacterial genomes (van Belkum *et al.* 1999).

The VNTR01 locus showed the existence of seven alleles containing 4–10 repeats. This locus has also been described in *Y. enterocolitica* Y56 (biovar 4, serotype 3) containing 15 tandem copies (De Benito *et al.* 2004). Analysis of this locus in a collection of 55 clinical isolates of *Y. enterocolitica* (belonging to serotype O:3 biovar 4, except one) showed existence of eight different alleles with 4–14 repeats (De Benito *et al.* 2004). The allele with 13 copies was most common among these isolates. Such variations may be attributed to different strains (biovar 1A *vs* biovar 4) used in the two studies. *In silico* analysis showed that VNTR01 locus containing ten repeats was present in the *orf528* locus of *Y. enterocolitica* 8081 encoding a GTP-binding protein whose function is not known. But PCR amplification of VNTR01 was not seen experimentally in this strain. Analysis of this locus in the genome of *Y. enterocolitica* 8081 revealed that it carried a deletion in the *sat* (*s*treptogramin *a*cetyl *t*ransferase) gene, a region containing the binding site for one (TR2) of the primers. Comparison of sequences around *orf528* in *Y. enterocolitica* strains Y56 and 8081 revealed several differences in gene order, suggesting that this locus may represent a hotspot for gene rearrangement in this species.

For the loci (VNTR03 and VNTR05), for which PCR yielded two amplicons differing in size, sequencing showed that the only difference between the two copies was in the number of the repeats. This implied duplication of such loci in these strains as reported for other organisms like *Haemophilus influenzae* (van Belkum *et al.* 1997) and *Bordetella pertussis* (Schouls *et al.* 2004).

For VNTR04, VNTR06 and VNTR07, except for the positive control (*Y. enterocolitica* 8081, biovar 1B), amplicon was not observed in any other strain. *In silico* analysis of VNTR04 locus in *Y. enterocolitica* 8081 genome sequence showed that it was located within the *orf*16 and *orf*17 which are part of the right arm of the

high-pathogenicity island (HPI) (Carniel *et al.* 1996). BLASTN searches revealed that this locus was also present in another strain namely, *Y. enterocolitica* WA314 (biovar 1B) (GenBank acc. no. AJ132945). The comparison of the locus in *Y. enterocolitica* 8081 and WA314 showed that the right arm of HPI was highly conserved in biovar 1B strains with variations in the number of repeat units. HPI is known to be present only in highly pathogenic biovar 1B strains of *Y. enterocolitica* and absent in other biovars (1A and 2–5) of this organism (Carniel 1999). Thus, VNTR04 may be exploited for typing exclusively *Y. enterocolitica* biovar 1B strains. The VNTR07 was located within the *psp* (*p*hage *s*hock *p*rotein) locus in the genome of strain 8081. As *psp* locus has been reported to be present only in biovar 1B strains (Darwin and Miller 2001; Darwin 2005), VNTR07 may also be used for typing highly pathogenic biovar 1B strains. Interestingly, as both VNTR04 and VNTR07 are present within the virulence-related loci these may result in variations in the pathogenicity as reported for *H. influenzae* (Moxon *et al.* 1995; Peak *et al.* 1996).

Based on the number of the distinct MLVA genotypes generated, the European biovar 1A strains appeared more heterogeneous compared to the Indian isolates. High degree of genetic heterogeneity of the European biovar 1A has also been reported earlier using PFGE (Najdenski *et al.* 1994) and FAFLP (Fearnley *et al.* 2005; Kuehni-Boghenbor *et al.* 2006). PFGE revealed 14 different pulsotypes among a collection of sixteen biovar 1A (serotype O:5) strains (Najdenski *et al.* 1994) while FAFLP identified 18 types among 22 European biovar 1A strains (Fearnley *et al.* 2005). All the Indian biovar 1A strains belonging to serotype O:6,30 or O:6,30,31 were epidemiologically related as these were isolated from diarrhoeic patients from North India only. These strains were also investigated earlier using different typing techniques namely, REP-PCR, ERIC-PCR (Sachdeva and Virdi 2004), *rrn* and *gyr*B loci-based genotyping (Gulati and Virdi 2007) and the results showed that these were clonally related. This clearly indicated that inability to distinguish these strains by MLVA was not due to limitations of VNTR typing.

The discriminatory ability of MLVA described in this study was less than MLVA based on the draft genome of *Y. enterocolitica* Y11 that identified 45 genotypes among 64 isolates of *Y. enterocolitica* biovar 4, serotype O:3 (Gierczynski *et al.* 2007). The MLVA used by this group could not establish the clonal nature of these strains as suggested by other genotyping techniques (Najdenski *et al.* 1994; Fearnley *et al.* 2005; Howard *et al.* 2006). In the present study, however, the clonal relationships established amongst *Y. enterocolitica* biovar 1A strains with MLVA were in good agreement with those established on the basis of other genotyping methods (Sachdeva and Virdi 2004; Gulati and Virdi 2007). Fluorescent amplified fragment length polymorphism (FAFLP) likewise clustered biovar 1A strains into two groups. In one group, biovar 1A strains clustered with biovars 2, 3 and 4; in the second group biovar 1A strains were related to the single biovar 1B isolate (Fearnley *et al.* 2005). Genotyping of *Y. enterocolitica* biovar 1A strains using different genotypic approaches revealed that strains with different serotypes (O antigen types) produced identical genotypes and were closely related genetically. Also in some cases the same O antigen type was shared by strains that were different genotypically. These observations indicated O antigen switching or/and lateral transfer of genes for O antigen in strains of *Y. enterocolitica* as suggested recently by multilocus sequence typing (Kotetishvili *et al.* 2005).

Phylogenetic analysis using MST further confirmed grouping of the biovar 1A strains into distinct clusters irrespective of their geographical origin suggesting the wide spread dissemination of both the groups. Also MST seem to suggest that clinical serotype O:6,30–6,31 isolates originated from ancestral wastewater serotype O:6,30–6,31 isolates by host adaptation and genetic change.

In conclusion, four polymorphic VNTR-containing loci were used to develop MLVA typing for strains of *Y. enterocolitica* biovar 1A. Twenty-six MLVA genotypes were identified among 81 such strains and the method was found to have good discriminatory (DI = 0·87) ability. The MLVA clustered *Y. enterocolitica* biovar 1A strains into two distinct groups and exhibited congruence with REP-PCR, ERIC-PCR, *rrn* and *gyr*B genotyping. The MLVA scheme presented here could be further improved with the use of fluorescently labelled primers and multiplexing of targets in order to increase both its practicality and discriminatory power (Lindstedt *et al.* 2005). Also more VNTR loci need to be characterized and tested for selection of an optimal set of VNTRs for MLVA typing. In addition to improving the methodology *per se*, comparisons against other molecular typing methods such as MLST, PFGE and AFLP would ultimately determine the utility of the MLVA as a molecular epidemiology tool for *Y. enterocolitica*. Moreover, further analyses of a larger number of isolates from various geographic areas is required to unravel the potential of MLVA typing for epidemiological investigations across all biovars of *Y. enterocolitica*.

Acknowledgements

This work was supported by a research grant to J.S.V. from the Department of Biotechnology, Government of India, New Delhi, and Senior Research Fellowship to P.G. from Indian Council of Medical Research, New Delhi, India. We are grateful to E. Carniel, J. Heesemann,

M. Skurnik and G.R. Cornelis for providing reference strains of *Y. enterocolitica*.

References

van Belkum, A. (1999) The role of short sequence repeats in epidemiologic typing. *Curr Opin Microbiol* **2**, 306–311.

van Belkum, A., Scherer, S., van Leeuwen, W., Willemse, D., van Alphen, L. and Verbrugh, H. (1997) Variable number of tandem repeats in clinical strains of *Haemophilus influenzae*. *Infect Immun* **65**, 5017–5027.

Benson, G. (1999) Tandem repeats finder: a program to analyse DNA sequences. *Nucleic Acids Res* **27**, 573–580.

Bissett, M.L., Powers, C., Abbott, S.M. and Janda, J.M. (1990) Epidemiologic investigations of *Yersinia enterocolitica* and related species: sources, frequency and serogroup distribution. *J Clin Microbiol* **28**, 910–912.

Bottone, E.J. (1999) *Yersinia enterocolitica*: overview and epidemiologic correlates. *Microb Infect* **1**, 323–333.

Burnens, A.P., Frey, A. and Nicolet, J. (1996) Association between clinical presentation, biogroups and virulence attributes of *Yersinia enterocolitica* strains in human diarrhoeal disease. *Epidemiol Infect* **116**, 27–34.

Butt, H.L., Gordon, D.L., Lee-Archer, I., Moritz, A. and Merrell, W.H. (1991) Relationship between clinical and milk isolates of *Yersinia enterocolitica*. *Pathology* **23**, 153–157.

Carniel, E. (1999) The *Yersinia* high-pathogenicity island. *Int Microbiol* **2**, 161–167.

Carniel, E., Guilvout, I. and Prentice, M. (1996) Characterization of a large chromosomal ''high-pathogenicity island'' in biotype 1B *Yersinia enterocolitica*. *J Bacteriol* **178**, 6743–6751.

Cornelis, G., Laroche, Y., Balligand, G., Sory, M.P. and Wauters, G. (1987) *Yersinia enterocolitica*, a primary model for bacterial invasiveness. *Rev Infect Dis* **9**, 64–87.

Darwin, A.J. (2005) The phage-shock-protein response. *Mol Microbiol* **57**, 621–628.

Darwin, A.J. and Miller, V.L. (2001) The *psp* locus of *Yersinia enterocolitica* is required for virulence and for growth *in vitro* when the Ysc type III secretion system is produced. *Mol Microbiol* **39**, 429–444.

De Benito, I., Cano, M.E., Aguero, J. and Garcia-Lobo, J.M. (2004) A polymorphic tandem repeat potentially useful for typing in the chromosome of *Yersinia enterocolitica*. *Microbiology* **150**, 199–204.

Denoeud, F. and Vergnaud, G. (2004) Identification of polymorphic tandem repeats by direct comparison of genome sequence from different bacterial strains: a web-based resource. *BMC Bioinformatics* **5**, 4.

Farlow, J., Smith, K.L., Wong, J., Abrams, M., Lytle, M. and Keim, P. (2001) *Francisella tularensis* strain typing using multiple-locus variable-number tandem repeat analysis. *J Clin Microbiol* **39**, 3186–3192.

Fearnley, C., On, S.L., Kokotovic, B., Manning, G., Cheasty, T. and Newel, D.G. (2005) Application of fluorescent amplified fragment length polymorphism for comparison of human and animal isolates of *Yersinia enterocolitica*. *Appl Environ Microbiol* **71**, 4960–4965.

García-Yoldi, D., Le Fleche, P., De Miguel, M.J., Muñoz, P.M., Blasco, J.M., Cvetnic, Z., Marín, C.M., Vergnaud, G. *et al.* (2007) Comparison of multiple-locus variable-number tandem-repeat analysis with other PCR-based methods for typing *Brucella suis* isolates. *J Clin Microbiol* **45**, 4070–4072.

Gierczynski, R., Golubov, A., Neubauer, H., Pham, J.N. and Rakin, A. (2007) Development of a multiple-locus variable-number tandem repeat analysis (MLVA) for *Yersinia enterocolitica palearctica* and its application to bioserogroup 4/O3 subtyping. *J Clin Microbiol* **45**, 2508–2515.

Greenwood, M.H. and Hooper, W.L. (1990) Excretion of *Yersinia* spp. associated with consumption of pasteurized milk. *Epidemiol Infect* **104**, 345–350.

Grundmann, H., Hori, S. and Tanner, G. (2001) Determining confidence intervals when measuring genetic diversity and the discriminatory abilities of typing methods for microorganisms. *J Clin Microbiol* **39**, 4190–4192.

Gulati, P.S. and Virdi, J.S. (2007) The *rrn* locus and *gyr*B genotyping confirm the existence of two clonal groups in strains of *Yersinia enterocolitica* subspecies *palearctica* biovar 1A. *Res Microbiol* **158**, 236–243.

Howard, S.L., Gaunt, M.W., Hinds, J., Witney, A.A., Stabler, R. and Wren, B.W. (2006) Application of comparative phylogenomics to study the evolution of *Yersinia enterocolitica* and to identify genetic differences relating to pathogenicity. *J Bacteriol* **188**, 3645–3653.

Hunter, P.R. and Gaston, M.A. (1988) Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol* **26**, 2465–2466.

Jayashree, B., Punna, R., Prasad, P., Bantte, K., Hash, C.T., Chandra, S., Hoisington, D.A. and Varshney, R.K. (2006) A database of simple sequence repeats from cereal and legume expressed sequence tags mined in silico: survey and evaluation. *In Silico Biol* **6**, 607–620.

Johansson, A., Farlow, J., Larsson, P., Dukerich, M., Chambers, E., Bystrom, M., Fox, J., Chu, M. *et al.* (2004) Worldwide genetic relationships among *Francisella tularensis* isolates determined by multiple-locus variable-number tandem repeat analysis. *J Bacteriol* **186**, 5808–5818.

Keim, P., Price, L.B., Klevytska, A.M., Smith, K.L., Schupp, J.M., Okinaka, R., Jackson, P. J. and Hugh-Jones, M.E. (2000) Multiple-locus variable-number tandem repeat analysis reveals genetic relationships within *Bacillus anthracis*. *J Bacteriol* **182**, 2928–2936.

Kimura, B., Sekine, Y., Takahashi, H., Tanaka, Y., Obata, H., Kai, A., Morozumi, S. and Fujii, T. (2008) Multiple-locus variable-number of tandem-repeats analysis distinguishes *Vibrio parahaemolyticus* pandemic O3:K6 strains. *J Microbiol Methods* **72**, 313–320.

Kotetishvili, M., Kreger, A., Wauters, G., Morris, J.G. Jr, Sulakvelidze, A. and Stine, O.C. (2005) Multilocus

sequence typing for studying genetic relationships among *Yersinia* species. *J Clin Microbiol* **43**, 2674–2684.

Kuehni-Boghenbor, K., On, S.L., Kokotovic, B., Baumgartner, A., Wassenaar, T.M., Wittwer, M., Bissig-Choisat, B. and Frey, J. (2006) Genotyping of human and porcine *Yersinia enterocolitica*, *Yersinia intermedia*, and *Yersinia bercovieri* strains from Switzerland by amplified fragment length polymorphism analysis. *Appl Environ Microbiol* **72**, 4061–4066.

Leclercq, A., Martin, L., Vergnes, M.L., Ounnoughene, N., Laran, J.F., Giraud, P. and Carniel, E. (2005) Fatal *Yersinia enterocolitica* biotype 4 serovar O:3 sepsis after red blood cell transfusion. *Transfusion* **45**, 814–818.

Lindstedt, B.A. (2005) Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic bacteria. *Electrophoresis* **26**, 2567–2582.

Lowell, J.L., Wagner, D.M., Atshabar, B., Antolin, M.F., Vogler, A.J., Keim, P., Chu, M.C. and Gage, K.L. (2005) Identifying sources of human exposure to plague. *J Clin Microbiol* **43**, 650–656.

Morris, J.G. . Jr, Prado, V., Ferreccio, C., Robins-Browne, R.M., Bordun, A-M., Cayazzo, M., Kay, B.A. and Levine, M.M. (1991) *Yersinia enterocolitica* isolated from two cohorts of young children in Santiago, Chile: incidence of and lack of correlation between illness and proposed virulence factors. *J Clin Microbiol* **29**, 2784–2788.

Moxon, E.R., Rainey, P.B., Nowak, M.A. and Lenski, R.E. (1995) Adaptive evolution of highly mutable loci in pathogenic bacteria. *Curr Biol* **4**, 24–33.

Najdenski, H., Iteman, I. and Carniel, E. (1994) Efficient subtyping of pathogenic *Yersinia enterocolitica* strains by pulsed-field gel electrophoresis. *J Clin Microbiol* **32**, 2913–2920.

Peak, I.R., Jennings, M.P., Hood, D.W., Bisercic, M. and Moxon, E.R. (1996) Tetrameric repeat units associated with virulence factor phase variation in *Haemophilus* also occur in *Neisseria* spp. and *Moraxella catarrhalis*. *FEMS Microbiol Lett* **137**, 109–114.

Ratnam, S., Mercer, E., Picco, B., Parsons, S. and Butler, R. (1982) A nosocomial outbreak of diarrheal disease due to *Yersinia enterocolitica* serotype O:5, biotype 1A. *J Infect Dis* **145**, 242–247.

Robins-Browne, R.M. (1997) *Yersinia enterocolitica*. In *Food Microbiology: Fundamentals and Frontiers* ed. Doyle, M.P., Beuchat, L.R. and Montiville, T.J. pp. 192–215 Washington, DC: American Society for Microbiology.

Sachdeva, P. and Virdi, J.S. (2004) Repetitive elements sequence (REP/ERIC)-PCR based genotyping of clinical and environmental strains of *Yersinia enterocolitica* biotype 1A reveal existence of limited number of clonal groups. *FEMS Microbiol Lett* **240**, 193–201.

Schouls, L.M., van der Heide, H.G., Vauterin, L., Vauterin, P. and Mooi, F.R. (2004) Multiple-locus variable-number tandem repeat analysis of Dutch *Bordetella pertussis* strains reveals rapid genetic changes with clonal expansion during the late 1990s. *J Bacteriol* **186**, 5496–5505.

Sperry, K.E., Kathariou, S., Edwards, J.S. and Wolf, L.A. (2008) Multiple-locus variable-number tandem-repeat analysis as a tool for subtyping *Listeria monocytogenes* strains. *J Clin Microbiol* **46**, 1435–1450.

Tennant, S.M., Grant, T.H. and Robins-Browne, R.M. (2003) Pathogenicity of *Yersinia enterocolitica* biotype 1A. *FEMS Immunol Med Microbiol* **38**, 127–137.

Thomson, N.R., Howard, S., Wren, B.W., Holden, M.T., Crossman, L., Challis, G.L., Churcher, C., Mungall, K. *et al.* (2006) The complete genome sequence and comparative genome analysis of the high pathogenicity *Yersinia enterocolitica* strain 8081. *PLoS Genet* **2**, e206.

Weir, B.S. (1990) *Genetic Data Analysis: Methods for Discrete Population Genetic Data Analysis*. Sunderland, MA: Sinauer Associates Inc.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Table S1 VNTR-containing loci identified in the genome of *Y. enterocolitica* 8081 using tandem repeat finder.

Table S2 VNTR-containing loci identified in the genome of *Y. enterocolitica* 8081 using SSR identification software.

Please note: Wiley-Blackwell are not responsible for the content or functionality of any supplementary materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the article.

