Evaluation of multiple-locus variable-number tandem repeat analysis (MLVA) for genotyping of *Escherichia coli* isolated from Karaj River

Reza Ranjbar, Samaneh Shokouhi Mostafavi and Hamed Memariani

**ABSTRACT**

Most microbiological water quality regulations rely upon the detection of indicators of fecal pollution, such as coliform bacteria, or more specifically *Escherichia coli*. In order to further understand the source, fate, and implications for water quality regulation, environmental *E. coli* isolates should be assessed genetically to observe various levels of genotypic diversity. Multiple-locus variable-number tandem repeat (VNTR) analysis (MLVA) is a novel, simple and inexpensive polymerase chain reaction (PCR) based genotyping method which relies on the detection of different copy numbers inside each VNTR locus. In this study, we evaluated MLVA as a tool for the genotyping of *E. coli* strains of water samples collected from the Karaj River, Iran. Overall, high genetic diversity was observed among environmental *E. coli* isolates. We also proved the feasibility of MLVA as a complementary or even future replacement genotyping method for the average routine laboratory.

**Key words** | *Escherichia coli*, genotyping, MLVA, Karaj river, VNTR

**INTRODUCTION**

Fecal contamination of surface waters can not only impair the quality of water utilized for drinking and irrigation purposes but also pose a significant risk to public and environmental health by contributing pathogenic microorganisms (McLellan 2004; Mohapatra *et al.* 2007). Regulatory and public health agencies continue to face the challenges of controlling fecal contamination of surface water (Casarez *et al.* 2007). Many water quality regulations still rely on the detection of indicators of fecal pollution, such as coliforms, or more specifically, *Escherichia coli* (*E. coli*) or *Enterococcus* spp. (Casarez *et al.* 2007; Mohapatra *et al.* 2007).

*E. coli* is a prominent member of the bacterial microbiota in fecal material from humans and warm-blooded animals. This bacterium has an impressive ability to survive harsh conditions encountered in different environments which is attributable to its vast genetic diversity, as more genetic diversity usually improves adaptability and resistance to environmental changes (Ardakani & Ranjbar 2016; Kheiri *et al.* 2016). Detection of *E. coli* by routine microbiological methods provides no information as to the originating host source. However, identifying the source of fecal contamination is a high priority in order to better understand the potential health risk and to mitigate the source of pollution (McLellan 2004).

Several genotyping methods such as pulsed-field gel electrophoresis (PFGE) (Parveen *et al.* 2001; McLellan 2004; Casarez *et al.* 2007), ribotyping (Carson *et al.* 2001; Anderson *et al.* 2006), and enterobacterial repetitive intergenic consensus sequence-based polymerase chain reaction (ERIC-PCR) (Dombek *et al.* 2000; Casarez *et al.* 2007; Lyauté *et al.* 2010; Goto & Yan 2011) have been applied for elucidating the sources of fecal contamination. These source tracking techniques have different discriminatory powers to resolve differences between closely related bacterial strains. Of these methods, PFGE is considered as the
gold standard fingerprinting method for *E. coli*. However, PFGE has certain drawbacks including the need for highly trained staff, expensive equipments, and reduced comparability of results between different laboratories (McLellan 2004; Casarez et al. 2007; Saxena et al. 2015). In contrast to PFGE, multiple-locus variable number tandem repeat (VNTR) analysis (MLVA) is a simple and inexpensive PCR-based typing method which relies on the detection of different copy numbers inside each VNTR locus. It has also been successfully employed as an effective tool for investigating strains that are epidemiologically related or unrelated in specific outbreaks (Gorge et al. 2008; Christiansson et al. 2011; Lindstedt et al. 2013; Memariani et al. 2015).

Most studies describing genetic diversity of environmental *E. coli* have focused on molecular methods such as REP-PCR, ERIC-PCR, and PFGE (Dombek et al. 2000; Parveen et al. 2001; McLellan 2004; Casarez et al. 2007; Lyaultey et al. 2010; Goto & Yan 2011). To the best of our knowledge, no studies have investigated the ability of MLVA for genotyping of *E. coli* strains in water samples. The Karaj River, one of the major rivers in northern Iran, has long been used for irrigation, fishing, and swimming. Some previous studies have showed that the levels of bacterial contamination have exceeded permissible limits (Torabian et al. 2011; Sakizadeh et al. 2015; Kheiri & Akhtari 2017). Therefore, the goal of this study was to examine the genotypic heterogeneity of *E. coli* in water samples collected from the Karaj River, Iran.

**METHODS**

**Water sample collection and *E. coli* isolation**

From July to September 2015, a total of 52 surface water samples were collected from the Karaj River with geographic coordinates of 35° 48′46.48″N; 51°0′43.05″E for enumeration of *E. coli*. Sampling bottles were immediately placed in a lightproof insulated box containing icepacks to ensure rapid cooling. Then, the samples were shipped to the laboratory for further analysis. For culture based tests, conventional methods according to Standard Methods for the Examination of Water and Wastewater 22nd edition were used (APHA/AWWA/WEF 2012). *E. coli* colonies from agar plates were picked and streaked for purity on EMB agar. Well-isolated colonies of purified *E. coli* were resuspended in trypticase soy broth with 20% glycerol and stored in −70°C for long-term storage.

**DNA extraction**

A pure culture of *E. coli* was plated on nutrient agar and incubated overnight at 37°C. A single colony was removed from the plate, suspended in 200μL of sterile deionized water, and boiled for 15 min. After centrifugation at 6,000 g for 8 min, the supernatant was transferred into a new tube for subsequent PCR analysis (Ranjbar & Memariani 2015).

**MLVA assay**

All of the *E. coli* isolates were typed by MLVA, using loci ms06, ms07, ms09, ms11, ms21, ms23, and ms32 (Gorge et al. 2008). The locus name, repeat size, primer sequences, and PCR annealing temperatures are shown in Table 1. For each locus, PCR was performed in 25 μL volume including 1X PCR buffer (50 mmol/L KCl, 10 mmol/L Tris, pH = 9), 2.5 mmol/L MgCl2, 0.2 mmol/L of each primer with 1 U of TaqDNA polymerase (CinnaGen Co., Iran), and 4 μL of the crude DNA extract. The PCR products were run on 1.5% (w/v) agarose gels, stained with ethidium bromide (Sigma-Aldrich, Steinheim, Germany), and visualized under ultraviolet transillumination. The number of repeats can be easily deduced from the PCR product sizes by manual reading. The product sizes were converted into repeat numbers based on formula as described previously (Ranjbar & Memariani 2015). A dendrogram of genetic relationships was also generated using the unweighted pair group method with arithmetic averages (UPGMA) based on allelic profiles. Moreover, a Hunter-Gaston discriminatory index (HGDI) was calculated for each VNTR locus as described previously (Ranjbar et al. 2016a).

**RESULTS**

The abundance of *E. coli* in surface waters was determined from 52 water samples over a 3-month period. Counts ranged from undetectable (detection limit, 1 CFU 100 ml−1)
to \(1 \times 10^5\) CFU 100 ml\(^{-1}\) (median, 80 CFU 100 ml\(^{-1}\)). Out of 52 samples, 45 showed detectable growth. Of these 45 samples, 27 were positive for \(E.\ coli\), with 71 \(E.\ coli\) isolates obtained and stored. The number of isolates genotyped per water sample ranged from one to five.

Overall, the 71 isolates were discriminated into 70 distinct genotypes (MLVA profiles). The genetic diversity based on HGDI for seven VNTR loci ranged from 0.347 to 0.835. VNTR locus 52 was identified to be the most polymorphic locus (HGDI = 0.835) while locus ms23 had the lowest diversity index (HGDI = 0.347). VNTR loci ms07, ms09, and ms52 had the highest number of different repeats (\(n = 9\)) whereas ms23 had the lowest number of different repeats (\(n = 5\)) (Table 2). Polymorphism of three VNTR loci in different \(E.\ coli\) strains is illustrated in Figure 1. The UPGMA dendrogram based on VNTR alleles with detailed information is shown in Figure 2. We have used a similarity cut-off of 70%, as mentioned previously in some similar studies (Casarez et al. 2007; Goto & Yan 2011). In our study, the \(E.\ coli\) isolates were grouped into five clusters.

### DISCUSSION

Recent studies have shown that there is differential survival and even growth of some \(E.\ coli\) strains from animal or

<table>
<thead>
<tr>
<th>LOCUS</th>
<th>PRIMER SEQUENCE (5’ TO 3’)</th>
<th>REPEAT SIZES AT EACH LOCUS (BP)</th>
<th>ANNEALING TEMPERATURE (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ms06</td>
<td>F-AAA CGG GAG AGC CGG TTA TT R-TGG TGG TAC AAC GGC TCC TG</td>
<td>39</td>
<td>55°C</td>
</tr>
<tr>
<td>ms07</td>
<td>F-GTC AGT TCG CCC ACA CAC AG R-CGG TAT ACG GAG CAG GCT AGT</td>
<td>39</td>
<td>55°C</td>
</tr>
<tr>
<td>ms09</td>
<td>F-GTG CCA TCG GGC AAA ATT AG R-CGG TGA AGT ACG GAG CAG GCT AGT</td>
<td>179</td>
<td>55°C</td>
</tr>
<tr>
<td>ms11</td>
<td>F-GAA ACA GGC CCA GGC TAC AC R-CTG GCG GTG TAT AGT GGT AT</td>
<td>96</td>
<td>55°C</td>
</tr>
<tr>
<td>ms21</td>
<td>F-GCT GAT GGC GAA GGA GAA GA R-GGG AGT ATG CGG TCA AAA GC</td>
<td>141</td>
<td>55°C</td>
</tr>
<tr>
<td>ms23</td>
<td>F-GCT CCG CTG ATT GAC TCC TT R-CGG TTG CTC GAC CAC TAA CA</td>
<td>375</td>
<td>55°C</td>
</tr>
<tr>
<td>ms32</td>
<td>F-GAG ATT GCC GAA GTG TTG C R-AAC TGG CGG CGT TTA TCA AG</td>
<td>101</td>
<td>55°C</td>
</tr>
</tbody>
</table>

### Table 2 | Characteristics of VNTR loci used for MLVA genotyping

<table>
<thead>
<tr>
<th>VNTR locus</th>
<th>Primer sequence (5’ to 3’)</th>
<th>Repeat sizes at each locus (bp)</th>
<th>Annealing temperature (°C)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ms06</td>
<td>F-AAA CGG GAG AGC CGG TTA TT R-TGG TGG TAC AAC GGC TCC TG</td>
<td>39</td>
<td>55°C</td>
</tr>
<tr>
<td>ms07</td>
<td>F-GTC AGT TCG CCC ACA CAC AG R-CGG TAT ACG GAG CAG GCT AGT</td>
<td>39</td>
<td>55°C</td>
</tr>
<tr>
<td>ms09</td>
<td>F-GTG CCA TCG GGC AAA ATT AG R-CGG TGA AGT ACG GAG CAG GCT AGT</td>
<td>179</td>
<td>55°C</td>
</tr>
<tr>
<td>ms11</td>
<td>F-GAA ACA GGC CCA GGC TAC AC R-CTG GCG GTG TAT AGT GGT AT</td>
<td>96</td>
<td>55°C</td>
</tr>
<tr>
<td>ms21</td>
<td>F-GCT GAT GGC GAA GGA GAA GA R-GGG AGT ATG CGG TCA AAA GC</td>
<td>141</td>
<td>55°C</td>
</tr>
<tr>
<td>ms23</td>
<td>F-GCT CCG CTG ATT GAC TCC TT R-CGG TTG CTC GAC CAC TAA CA</td>
<td>375</td>
<td>55°C</td>
</tr>
<tr>
<td>ms32</td>
<td>F-GAG ATT GCC GAA GTG TTG C R-AAC TGG CGG CGT TTA TCA AG</td>
<td>101</td>
<td>55°C</td>
</tr>
</tbody>
</table>

### Table 2 | Diversity indices, number of alleles, and typeability for each VNTR locus

<table>
<thead>
<tr>
<th>VNTR locus</th>
<th>Hunter-Gatson diversity of index</th>
<th>95% Confidence interval (CI)</th>
<th>No. of alleles</th>
<th>Typeability(^a) (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>ms06</td>
<td>0.744</td>
<td>0.709-0.780</td>
<td>6</td>
<td>100</td>
</tr>
<tr>
<td>ms07</td>
<td>0.830</td>
<td>0.791-0.868</td>
<td>9</td>
<td>94.4</td>
</tr>
<tr>
<td>ms09</td>
<td>0.787</td>
<td>0.727-0.846</td>
<td>9</td>
<td>92.9</td>
</tr>
<tr>
<td>ms11</td>
<td>0.667</td>
<td>0.576-0.758</td>
<td>8</td>
<td>100</td>
</tr>
<tr>
<td>ms21</td>
<td>0.589</td>
<td>0.469-0.708</td>
<td>7</td>
<td>95.8</td>
</tr>
<tr>
<td>ms23</td>
<td>0.347</td>
<td>0.211-0.484</td>
<td>5</td>
<td>92.9</td>
</tr>
<tr>
<td>ms32</td>
<td>0.835</td>
<td>0.799-0.871</td>
<td>9</td>
<td>100</td>
</tr>
</tbody>
</table>

\(^a\)The ability of each VNTR locus to type the isolates was measured as follows: Number of isolates amplified in each VNTR locus/71.

![Figure 1](https://image.pollardをしている。2055591/issue18010160.pdf) Polymorphism of 3 VNTR loci in different E. coli isolates. This image illustrates how the number of repeats can be directly deduced by manual reading.
Figure 2 | UPGMA analysis of E. coli isolates based on VNTR profiles.
human faeces in gut habitats compared to secondary habitats such as soils, sediments, and waters (Casarez et al. 2007; Goto & Yan 2011). In order to further understand the source, fate, and implications for water quality regulation, environmental E. coli isolates should be assessed genetically to observe various levels of genotypic diversity (McLellan 2004; Byappanahalli et al. 2006; Ishii et al. 2006; Lyautey et al. 2010; Goto & Yan 2011). Although E. coli is one of the most useful fecal indicators, the genetic heterogeneity of E. coli has received little attention because current water quality regulations are based primarily on enumeration of fecal bacteria (Byappanahalli et al. 2006; Goto & Yan 2011).

In this study, we successfully applied the MLVA method to analyze E. coli strains isolated from surface waters. MLVA is a rapid and simple method with low cost, good reproducibility and high discriminatory power for monitoring outbreaks and clonal spread of bacterial isolates. It has been successfully applied to investigate the clonal relationship and epidemiology of clinical E. coli isolates (Gorge et al. 2008; Memariani et al. 2015). In spite of this, most source tracking studies have used genotyping methods other than MLVA for characterization of environmental E. coli strains (Dombek et al. 2000; Casarez et al. 2007; Lyautey et al. 2010; Goto & Yan 2011). For instance, Casarez et al. used PFGE and ERIC-PCR for the genotyping of 555 E. coli isolates recovered from natural waters in Texas, USA. Their E. coli isolates showed 461 PFGE genotypes. However, only 75 genotypes (16%) were represented by clusters of more than one isolate each. They also found that the discriminatory power of PFGE is far superior to ERIC-PCR (Casarez et al. 2007). In another study conducted by McLellan et al., E. coli isolates from sources of fecal pollution, gulls, and sewages were characterized by using REP-PCR and compared with E. coli isolates from urban rivers and beach waters in Milwaukee, USA. They showed that there was less diversity in E. coli isolates obtained from river and beach sites than with the isolates recovered from human and non-human sources (McLellan 2004). In a similar study from Canada, diversity of E. coli strains recovered from the surface waters of eastern Ontario was assessed using REP-PCR. They observed 7,325 distinct genotypes among 21,307 E. coli isolates (Lyautey et al. 2010). Goto and Yan also used REP-PCR for the genotyping of E. coli isolates in tropical watersheds in Hawaii and observed high overall genotypic diversity (35.5% unique genotypes) (Goto & Yan 2011). In general, most of these surveys have showed that there is a high amount of diversity within the environmental E. coli population; hence, accurate strain representation is needed to reflect what might be expected in surface waters contaminated with fecal pollution. The present study showed that the genetic diversity of E. coli from the Karaj River waters was very large, which is almost in accordance with results of some previous studies (Casarez et al. 2007; Lyautey et al. 2010).

Based on the MLVA genotype diversity of water isolates, there is little evidence that a small number of environmentally-adapted E. coli strains represent dominant populations in this study. We anticipated finding broadly diverse VNTR profiles for our environmental E. coli isolates, which would reflect the diffuse nature of the bacterial contamination in surface waters. However, an accurate estimation of diversity would require determining the best compromise given the sample size. With a small sample size, the data should be interpreted with caution and should not be extrapolated to a larger population, or be used to make conclusions (Casarez et al. 2007; Lyautey et al. 2010; Goto & Yan 2011). Because of this, we think that larger sample sizes from different locations in the region would be more favorable to obtain a clearer separation of clonal groups. In our study, 71 E. coli isolates were discriminated into 70 distinct genotypes, indicating a wide diversity of E. coli population in the Karaj River. Another factor which affects the genetic diversity of E. coli in the environment is urban and agricultural land uses. Some previous studies have shown that the water affected by agricultural land use may contain higher E. coli genotypic diversity than that of urban land use (Larned et al. 2004; Traister & Anisfeld 2006; Goto & Yan 2011). The Karaj surface waters includes both urban and agricultural land uses within a large geographic area, providing a model system for studying the effects of such land uses on the genetic diversity of E. coli strains isolated from the same watersheds (Kheiri et al. 2016; Ranjbar et al. 2017). In addition, municipal wastewater discharge and water sport activities such as swimming may contribute to pollution of ground waters (Torabian et al. 2011; Sakizadeh...
et al. 2015). In a very recent study, Kheiri and Akhtari used ERIC-PCR to determine the genetic diversity of E. coli isolates in the Karaj River. They found considerable diversity among E. coli isolates (39 unique profiles among 50 isolates), which is almost in agreement with the results of our study (Kheiri & Akhtari 2017).

In conclusion, our study showed high genotypic diversity of environmental E. coli strains isolated from the Karaj River. In other words, there was no evidence that a given group of E. coli isolates with distinct origin may represent a dominant population (Ranjbar et al. 2015; Ranjbar et al. 2016). As a method, MLVA is high-throughput, inexpensive, easy to perform, rapid, and reliable. However, in order to further evaluate this new approach, future studies should carefully compare MLVA with other genotyping techniques such as PFGE, ERIC-PCR, and REP-PCR.

ACKNOWLEDGEMENTS

We would like to thank the ‘Clinical Research Development Center of Baqiyatallah Hospital’ for their kind cooperation. This study was financially supported in part by the ‘Clinical Research Development Center of Baqiyatallah Hospital’.

REFERENCES


Ardakani, M. A. & Ranjbar, R. 2016 Molecular typing of uropathogenic E. coli strains by the ERIC-PCR method. Electron Physician 8 (4), 2291–2296.


Mclellan, S. L. 2004 Genetic diversity of Escherichia coli isolated from urban rivers and beach water. Applied and Environmental Microbiology 70 (8), 4658–4665.


First received 20 November 2016; accepted in revised form 10 May 2017. Available online 5 June 2017