2012 3rd International Conference on Biology, Environment and Chemistry
IPCBEE vol.46 (2012) © (2012) IACSIT Press, Singapore
DOI: 10.7763/IPCBEE. 2012. V46. 24

Characterization of *Shigella flexneri* Isolates by Tandem Repeat Analysis and Pulsed-Field Gel Electrophoresis

Ashraf Hosseini [1&2+], Shanta Dutta [3], Suman Nandy[3], Indira Ghosh [4] and P. S. Khandekar [5]

[1] Tehran University of Medical Science, School of Allied Medical Sciences, Tehran, Iran

[2] Institute of Bioinformatics and Biotechnology, University of Pune, Pune 411007, India

[3] National Institute of Cholera and Enteric Diseases P-33, Cit Road Scheme XmP.O.Box 177 Beliaghata Kolkata - 700010 .

[4] School of Information Technology, JNU, NewDelhi, 110067. 5. Department of Biotechnology, Sinhagad Engineering College, Sinhagad Institute of technology, Vadguon, Pune, India.

Abstract. Pulsed-field gel electrophoresis (PFGE) is highly discriminatory molecular typing tools for many bacterial pathogens and Multilocus variable-number tandem repeat (VNTR) analysis (MLVA) is prominent typing tool which has been developed for a variety of bacterial pathogens. Knowledge on molecular characterization and typing of *Shigella* strains of any geographical region is important for epidemiology, pathogenesis and treatment of shigellosis and as VNTRs appear to contain greater diversity, hence, analysis of these regions will provide greater discrimination among the strains of identical serotypes. In this report, we identified 8 repeat sequence loci in the genomic sequences of the *Shigella flexneri* 2a strains *301* and *2457T*. Seven tandem repeats were found to be variable after testing a set of 40 multidrug-resistant clinical isolates of *S.flexneri,* which were isolated from children with acute diarrhea and attending to the outpatient department of Dr. B. C. Roy Memorial Hospital for Children, Kolkata, India between 2000 and 2005. Seven VNTR loci were found to be the powerful markers for strain differentiation and phylogenetic studies of *S.flexneri*. This study would provide a new insight and impetus for further investigations on global epidemiology of *Shigella* isolates.

Keywords: MLVA, *Shigella flexneri*, PFGE, Molecular typing

1. Introduction

Recently, PCR-based methods have become increasingly popular for molecular typing efforts. (1). One of the most recent developments in molecular typing involves the analysis of VNTR sequences (2).VNTR loci are tandemly arranged nucleotide repeats in a single genetic block often vary in copy numbers, exhibit allelic variation in length and creating polymorphisms that can be detected after amplification using PCR or restriction enzymes. The VNTR-based assays are based on direct PCR amplification of specific loci. All the targets of VNTR assay are thus known, and are not "anonymous" bands as in PFGE or RAPD typing. (3). The availability of whole genome sequences and appropriate algorithms for searching repetitive sequences has led to the application of the VNTR typing approach, namely the development of multilocus VNTR analysis or MLVA method for typing of several pathogens, such as *Bacillus anthracis* (4), *Yersinia pestis* (5), *Francisella tularensis* (6). In this report, we described a VNTR analysis system that used eight marker loci to discriminate among different *S. flexneri* strains. This study aims at developing and evaluating the MLVA method particularly for *S. flexneri* (strains within a single serotype) which is most prevalent and endemic in developing countries.

2. Materials and Methods

[+] Corresponding author.
E-mail addresses: ash-hosseini@tums.ac.ir

2.1. Selection of VNTR and MLVA Loci

Selection of repeat loci was based on the size of the tandem repeat, number of repeat units and whether the sequence was conserved within the repeats. These potential VNTRs were selected from sequence genome of the reference strains *Shigella flexneri 2a str301* (NC_004851) and *Shigella flexneri 2a str2457T* (NC_004741), subsequently those were tested on 40 clinical isolates of *Shigella flexneri 2a.* Some of the microsatellite loci have been selected in the genome *of Shigella* by using simple sequence repeat results and 3 softwares as follows:

- Tandem repeats database accessible at http://minisatellites.u-psud.f
- Tandem Repeats Finder (TRF) software accessible at http://c3.biomath.mssm.edu/trf.html
- Strain comparison accessible at http://strain-comparison.igmors.u-psud.fr

Tools such as BLAST and Clustal W program were also used to determine the theoretical sizes of the amplicons and compare similarity regions and copy.

2.2. Source of Clinical Isolates of *S. flexneri*

Stool samples were collected from children suffering from acute dysentery and attending the outpatient department of a Govt. Referral Children Hospital in Kolkata, India during 2000 and 2005 and were processed in the microbiology division of National institute of Cholera and Enteric diseases, Kolkata for *Shigella* isolates using standard microbiological culture technique (WHO1987). A total no of 40 isolates of *S.flexneri 2a* were selected for this study and to test the potential VNTR loci. ***Shigella* Strains, Growth Conditions and DNA Purification**

A total of 40 *S.flexneri 2a* isolates from different areas of the Kolkata city were selected for MLVA typing and study of different antibiotic resistance patterns inclooding: ampicillin, amoxicilin, tetracycline, ciprofloxacin, chloramphenicol, gentamicin, norfloxacin, amikacin, nalidixic Acid, sulfamethoxazole, cefotaxine and furazolidone. *Shigella* isolates were incubated overnight at 37°C with shaking at 250 rpm using Luria broth media. Isolation of genomic DNA has been performed by the method of Chen and Kuo (1993). Extracted DNA has been checked by running 3 µl in an agarose gel.

2.4. PCR Primers

Primer sets were designed by using Primer3 software (http://frodo.wi.mit.edu/cgi-bin/primer3/primer3_www.cgi), GCG package (http://202.41.70.101:8009/mgr.shtml) and SGD primer software (http://www.yeastgenome.org). The primers targeted conserved sequences on either side of the VNTR locus. VNTR1 is present in 5 *shigella* strains available in GenBank including: *S.flexneri 2a str.301*, *S.flexneri 2a str.2457T, S.sonnei, S.boydii, S.flexneri* and *S.dysenteriae.* VNTRs number 2, 4 and 5 are present in *Shigella* and *E.coli* strains available in Gene Bank. VNTRs number 3 and 6 are present only in both reference strains of *S.flexneri 2a.*

2.5. PCR Amplification and Electrophoresis

PCR amplification of the VNTR loci was performed using Gene Amp 9700 thermal cycler (Applied Biosystem, USA) in a total volume of 25 µL. Each 25µl PCR mixture contained 2.5 µl of 10x *Taq* buffer (with 15 mM $MgCl_2$), 0.5 mM extra $MgCl_2$, 25 pM each of forward and reverse primer (Bangalore Genei), 2.5 mM each dNTPs, 1U of the *Taq* DNA polymerase (Bangalore Genei), and 5 µl of DNA template and double distilled water (ddH2O) making up the volume to 25 µL. The PCR temperature cycle comprising of an initial denaturation at 96°C for 5 minutes was followed by 40 cycles of a three step cycle protocol: 96°C for 30 seconds, 58°C for 30 seconds and 72°C for 60 seconds and a final extension of 72°C for 7 minutes.

2.6. DNA Sequencing

DNA sequencing has been done as per the standard protocols (Yi Wei Tang et al., 1997) and recommendations of the ABI 3730 sequencer team using Big Dye terminator chemistry.

2.7. PFGE

A total of 21 *S.flexneri* 2a isolates were subject to subtyping by standard PFGE method. The isolates were sub cultured on appropriate growth media and incubated at 37ºC overnight or longer for further procedure.

PFGE using Xba1 restriction enzyme (Bangalore Genei, Bangalore, India) for digestion of DNA from *S.flexneri 2a* isolates was performed following standard protocol (Ramesh Gautom, 1997). Digested DNA was electrophoresed on a 1% agarose gel using CHEF-DR III (Bio-Rad, Hercules, CA) with switch times of 5-35s at 6.0 V/cm for 24 h at 14ºC along with λ ladder PFGE molecular weight marker (New England Biolabs, Beverly, MA). After electrophoresis the gels were stained in ethidium bromide (10mg/ ml) followed by destained by washing in distilled water. Macro restriction fragment patterns were analyzed using FP quest software (BIO-RAD, US).

3. Results

3.1. Evaluation of Tandem Repeats Polymorphism

The Tandem Repeat Finder program (TRF) identified 84 repeat motifs within the genome of *S.flexneri 2a str 301* and *2457T* and Tandem Repeat comparison program identified 24 tandem repeats with length difference between *S.flexneri 2a str. 2457T* and *S.flexneri 2a str. 301* >= 5 bp and <= 5000 bp. Initially, 8 potential VNTR loci with short lengths of repeat units (~15 bp) were selected from above mentioned lists (Table.1). Two of the 8 existed in multiple copies with various repeat numbers in *S.flexneri*. Figure 1 shows the VNTR3 locus as an example of the polymorphic patterns detected by agarose gel electrophoresis. *S.flexneri* isolates showed four drug resistance profiles. Pattern 1 indicates susceptibility to gentamicin, cefotaxine and amikacin and resistance to other antibiotics.Preliminary testing against the strains identified 7 loci out of 8 that were polymorphic between different strains. Only one MLVA locus (MLVA2) failed to amplify any DNA. Seven loci including VNTR1-6 and MLVA1 were found to contain variable repeat copy numbers within 2 strains of *S.flexneri 2a* and then retested using those 40 odd strains. The copy number of repeat per locus varied between 1.5 to 15.0. The copy number of repeat per locus is from 5.3 to 13.1 in VNTR1, 1.5 to 3.2 in VNTR2, 7.2 to 11.2 in VNTR3, 5.5 to 15 in VNTR4, 2.7 to 4.6 in VNTR5, 5.0 to 11.5 in VNTR6 and 1 to 2 in MLVA2. The genomic characteristics of the 7 VNTR loci are listed in Table.2. The lengths of repeat units for the 7 repeat loci ranged from 6 to 53 bp, one of the 7 loci were multiples of 53 bp. 6 loci of the 8 loci were located in coding region of annotated genes. The most hyper-variable locus was observed in VNTR1 and VNTR3. The observed copy number (calculated using formula mentioned above) of repetitions for each locus at 6 loci for 40 *S.flexneri* strains is listed in Table.2.

3.2. Cluster Analysis

After sequencing during verification process, the allele sizes were converted into repeat numbers of the seven loci and entered into CLUSTAL W (1.81) Multiple Sequence Alignments and GeneBee Molecular Biology Server for cluster analysis and Phylogenetic Tree Prediction. The dendrogram produced by CLUSTAL W that comprised 40 isolates were classified into two main clusters based by similarity matrices. Discriminatory capacity (http://insilico.ehu.es/mini_tools/discriminatory_power/index.php) of seven VNTR combinations for 40 isolates is high with a *D* value of 0.981. There is significant differences between distributing of drug resistance pattern and year of sample collection with profile types of VNTR combination by $\chi 2$ test ($P<0.05$). Lower levels of correlation were observed between sex, age, fever and abdominal pain with VNTR clusters. There is no correlation between VNTR clusters with diarrhea type, local address and vomiting (Fig.2). The dendrogram produced by the Gene Bee Molecular Biology Server indicates 69.9 percent homology among 21 study strains. Most Distance Matrix has been observed in isolates No.15 and 33 respectively (0.384 and 0.325). Discriminatory capacity of seven VNTR combinations for 21 isolates is high with a *D* value of 0.9667

4. Discussions

Variable copy numbers of the repeats for each locus within the study strains suggested that the loci selected were highly polymorphic. Our study also showed some relationship between MLVA profiles with drug resistance profiles and the year of sample collection but dendogram showed no bias towards geographical clustering and clinical features of strains. all 3 isolates of one sub cluster (no 24, 25 and 27) and 2 out of 3 isolates of 2 sub clusters (No 19, 16, 26 and 31) are drug resistance pattern 3. Thus, different sets of VNTR loci may be useful for phylogenetic investigation of isolates evolving over different time

scales. The method has potential application in further understanding of molecular epidemiology of *Shigella*. This method was standardized in laboratory of National Institute of Cholera and Enteric Disease (NICED), Kolkata, India, and has repeated in laboratory of Institute of Bioinformatics and Biotechnology (IBB), Pune, India with different thermal cyclers employed to achieve the same results. This level of standardization at laboratory would allow the transfer of the method into another laboratory more effectively than that of a method that was operator or equipment specific.

In earlier report MLVA method has been exploited as a routine typing method for surveillance and outbreak investigations of diseases caused by *S.sonnei* and found to be better than PFGE (7). In this study since a limited number of *S.flexneri* 2a isolates from limited area have been tested by both methods, it was obvious that MLVA typing could exhibit a higher discriminatory power than PFGE. Though tested in a limited collection of isolates, still high discriminatory power of seven VNTR combinations in our study, suggests that these 7 loci are sufficient for the sub-typing of *S.flexneri*. This is supported by similar findings reported in the recently study on the *S.sonnei* by Shiu-Yun Liang et al (March 2007). Their **S**tudy indicated that MLVA is a powerful typing tool to distinguish isolates for outbreak investigation and that it exhibited a good discrimination of the 22 PFGE-indistinguishable isolates(8).

VNTRs are rapidly evolving genomic elements that have been used successfully for the molecular typing of other pathogens such as *Bacillus anthracis, Yersinia pestis,* and *Mycobacterium tuberculosis* (5). VNTR loci have been developed for *S.flexneri* can provide useful information for epidemiological investigation of *S.flexneri.* However, study on a greater number of strains from a variety of geographical regions will be needed to determine which loci need to be included in a standard MLVA typing scheme for disease surveillance and outbreak investigation.

To provide a continuous and robust coverage of the *S.flexneri* more strains from a variety of geographical regions must be typed. It is hoped that this data will constitute an easy-to-use high-resolution classification resource which will then help address medical and epidemiological issues regarding *S.flexneri*.

5. Conclusion

The VNTR loci chosen for this assay showed highly polymorphic variations in *S.flexneri 2a*. This method provides a new insight for further epidemiological investigations of infections of *S.flexneri* and other *Shigella* serotypes. We believe that MLVA will be an invaluable tool for investigations of population dynamics, maintenance, and molecular epidemiology studies of *S.flexneri*.

6. Acknowledgements

Help and support of staff of Microbiology laboratory of NICED, Director and faculty members and staff of Bioinformatics Centre and IBB, University of Pune is gratefully acknowledged.

7. References

[1] Williams, J. G., A. R, Kubelik., K. J, Livak., J. A, Rafalski., and S. V, Tingey. , 1990. DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. Nucleic Acids Res. 18:6531-6535.

[2] Van Belkum, A., S, Scherer, L. van Alphen, and H, Verbrugh. ,1998. Short-sequence DNA repeats in prokaryotic genomes. Microbiol. Mol. Biol. Rev. 62:275-293. [Abstract/Free Full Text].

[3] Bjŕn-Arne, Lindstedt. , 2005. Multiple-locus variable number tandem repeats analysis for genetic fingerprinting of pathogenic Bacteria *Electrophoresis* 2005, *26,* 2567–2582.

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

[5] Klevytska, A. M., Price, L. B., Schupp, J. M., Worsham, P. L., Wong, J. & Keim, P. , 2001. Identification and characterization of variable-number tandem repeats in the *Yersinia pestis* genome. *J Clin Microbiol* 39, 3179–3185.

[6] Farlow, J., Postic, D., Smith, K. L., Jay, Z., Baranton, G. & Keim, P. , 2002. Strain typing of *Borrelia burgdorferi*, *Borrelia afzelii*, and *Borrelia garinii* by using multiple-locus variable-number tandem repeat analysis. *J Clin*

Microbiol 40, 4612–4618.

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

[8] Shiu-Yun, Liang., Haruo, Watanabe., Jun, Terajima., Chun-Chin, Li., Jui-Cheng, Liao., Sheng, Kai. Tung., and Chien-Shun, Chiou. (November 2007). Multilocus Variable-Number Tandem-Repeat Analysis for Molecular Typing of *Shigella sonnei.* J Clin Microbiol.; 45(11): 3574–3580.

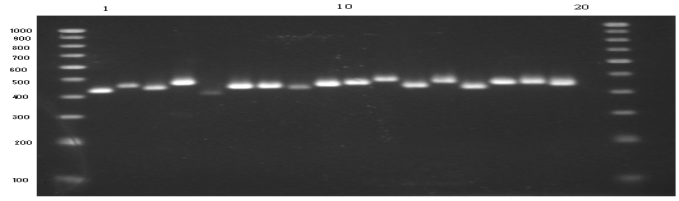
Table 1: The characteristics of selected loci and primers designed

VNTRs/ MLVAs	motif	Location	Position in Sh,f 301	L PRIMER 5'-3' R PRIMER 5'-3'	No of Repeat in Sh,f 301
VNTR1	CCAGCC	membrane-spanning protein of hydrogenase 3	2816426	GCCATCAACCCGACTAACAT GTTCCGATATCTGGCTGCTC	8.8
VNTR2	AGATAAAGATCGACC	putative transport protein	2880524	CGCCCATAAAGAAGAAGACG GTTGCTGGATCGTTTTGGTT	3.2
VNTR3	TTAATGATT	cytidine/deoxycytidine deaminase (SI 50)	3106561	CTGTATCCCGCGATGAGACT CCTGTTTGTGTCTGTTGAGCA	12.2
VNTR4	TGCAGG	uroporphyrinogen III methylase a late step of protoheme IX synthesis	3996030	TGGCAAACAGCACTACCATC AGAACGTCTCCACCTGGGTA	9.2
VNTR5	TTCAGCCAG	putative transcriptional regulator LYSR-type	1683226	TTCCCTCACCTGAGAGCATT CATTTACAAACCCGC GAAAT	3
VNTR6	ACCTGC	uroporphyrinogen III methylase a late step of protoheme IX synthesis	3776761	AAGCTGTACCGCGTCATCA ATCATCGGGCCAACCACTAT	5.5
MLVA1	ACGGGCACTGCGTTTATCCGG ATGATGTCGCTGTTGCTGACA GTGGTAATACG	IS629	251699, 52338, 314962, 315601	AGTGCATCCAGCACGAATGT CTGAACGTCAGCGTCTGAAA	2.1-3

VNTR= variable number tandem repeat MLVA = Multiple Loci VNTR Analysis

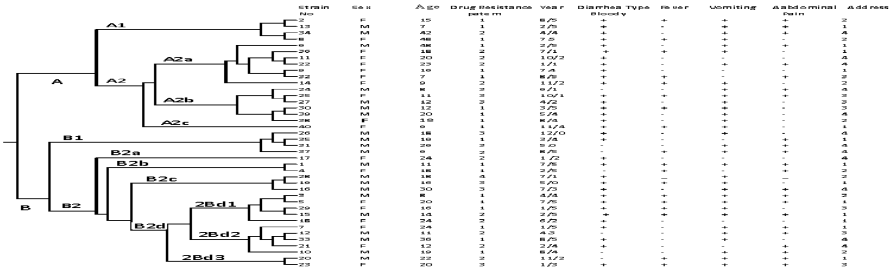
Table 2: Experimentally observed numbers of repetitions at 6 VNTR loci of 40 isolates of *S. flexneri*

Serial No	Strain No	Number of repetitions at the following locus					
		VNTR 1	VNTR 2	VNTR 3	VNTR 4	VNTR 5	VNTR 6
1	6458	7.7	2.5	8.2	10	-	7.5
2	6536	7.7	2.5	11	11	4.6	8.5
3	5856	12.2	1.5	-	8.2	2.7	8.5
4	6281	6.7	2.5	8.2	10	4.6	7.5
5	6430	9.1	1.5	11.2	9	3.8	7.5
6	6279	7.7	1.5	7.2	8.2	3.8	7.5
7	6270	9.1	1.5	9.2	8.2	3.8	7.5
8	6321	9.1	1.5	9.2	8.2	3.8	7.5
9	6023	9.1	1.5	8.2	8.2	3.8	7.5
10	6037	6.7	1.5	10.2	7.6	3.8	7.5
11	5003	9.1	2.5	10.2	7.2	2.7	8.5
12	5223	9.1	1.5	-	7.2	2.7	6.6
13	6297	7.7	2.5	11.2	9	2.7	8.5
14	5014	11.1	1.5	9.2	7.6	2.7	7.5
15	5654	5.3	1.5	9.2	5.5	2.7	5.9
16	4189	13.1	2.5	8.2	8.3	2.7	8.5
17	4678	7.7	2.5	10	8.3	2.7	7.5
18	4979	6.7	3.2	10	9	3.8	7.5
19	2544	12.1	3.2	9.2	11	3.8	8.5
20	5062	10.1	2.5	8.2	5.5	3.8	7.5
21	5322	9.1	1.5	9.2	6.5	3.8	7.5
22	4317	9.1	1.5	12	-	2.7	8.5
23	4021	11.1	2.5	7.1	6.5	3.8	-
24	3020	10.1	2.5	11.2	6.5	2.7	6.6
25	3651	6.7	2.5	9.2	6.5	3.8	6.6
26	2594	7.7	-	8.2	9	3.8	7.5
27	3726	7.7	3.2	8.2	6.5	3.8	5.9
28	3158	7.7	3.2	10	6.5	2.7	5.9
29	6256	9.1	2.5	10	9	3.8	7.5
30	6280	7.7	2.5	9.2	9	3.8	7.5
31	6556	6.7	3.2	8.2	-	3.8	6.6
32	6436	9.1	3.2	8.2	9	4.6	8.5
33	6494	9.1	3.2	9.2	8.5-15	2.7	6.6-11.5
34	5837	6.7	3.2	10	9	2.7	8.5
35	5722	7.7	3.2	-	8.5	3.8	7.5
36	4508	5.3	3.2	7.2	9	2.7	7.5
37	6534	7.7	-	10.8	8.5	3.8	8.5
38	6066	10.1	2.5	9.2	10.4	3.8	8.5
39	5922	7.7	2.5	9.2	10.4	2.7	8.5
40	6152	6.7	2.5	9.2	11.3	2.7	8.5



Lanes 2 to 18: Amplificons from *Sh.flexneri* 2a isolate 1-2, 4-11 and 13-19 respectively. Lane 19: amplificon from *Sh.dysenteriae*. lane 1 and 20: DNA ladders with 100-bp mol. size marker.

Fig. 1: Polymorphic patterns of VNTR3 locus



The screening set comprised 40 strains of *Sh.flexneri 2a*. The second column indicates sex of the patient, the third column indicates age of the patient, the fourth column indicates drug resistance pattern, the fifth column indicates year of sample collection, the sixth to ninth columns indicate clinical features and the tenth column indicates local addresses in Kolkatta City which are geographical regions separated by some geographical pattern including river, lake and different life level of the patients.

Fig. 2: Dendrogram produced by use of clustalw program of 7 combined loci

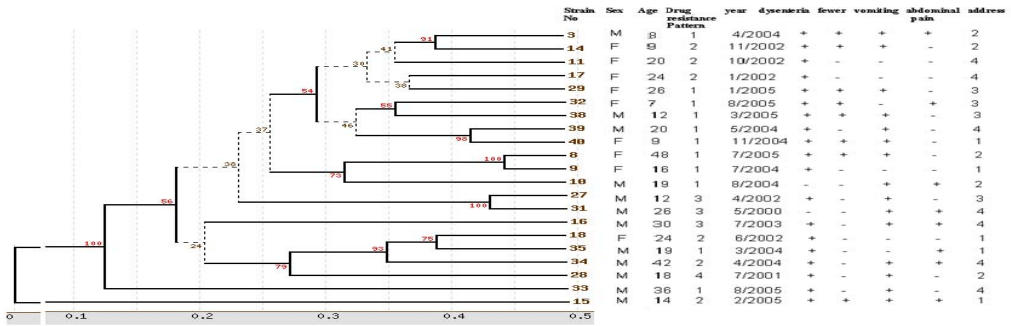


Fig. 3: Dendrogram produced by use of GeneBee Services of 6 combined loci for 20 clinical isolates of *S. flexneri* 2a