An improved multiple-locus variable number of tandem repeats analysis for *Leptospira interrogans* serovar Australis: a comparison with fluorescent amplified fragment length polymorphism analysis and its use to redefine the molecular epidemiology of this serovar in Queensland, Australia

Andrew Slack, Meegan Symonds, Michael Dohnt and Lee Smythe

Correspondence

Andrew Slack
andrew_slack@health.qld.gov.au

WHO/FAO/OIE Collaborating Centre for Reference and Research on Leptospirosis, Western Pacific Region, Centre for Public Health Sciences, Queensland Health Scientific Services, Brisbane, Australia

In this study, an improved multiple-locus variable number of tandem repeats analysis (MLVA) method based upon a previously published method is described. Improvements to the method included redesigned primers and PCR conditions, combined with pooled capillary electrophoresis using multicolored dyes. Allele sizes were converted into an allele string, and each unique allele string was assigned a numerical MLVA type (MVT). The improved MLVA method was then applied to 96 previously characterized *Leptospira interrogans* serovar Australis isolates from human and animal sources. The improved MLVA was found to have between six and 13 alleles at each locus, compared with three to eight in the original. The mean Hunter–Gaston diversity index (HGDI) for the improved MLVA method was 0.654, compared with 0.599 in the original; this increase in diversity was largely due to changes in the analysis of the variable number of tandem repeat (VNTR) data. When the improved MLVA method was compared with the fluorescent amplified fragment length polymorphism (FAFLP) method, there was a high level of concordance between the profiles; however, the MLVA method produced an additional four unique profiles amongst the subset of 30 isolates tested. Given that the improved MLVA method was found to be superior to the original MLVA method, it was subsequently used to redefine the molecular epidemiology of *L. interrogans* serovar Australis in Queensland, Australia. Using cluster analysis, the authors were able to demonstrate clonal links amongst rodent isolates, rodent and human isolates, and rodent and canine isolates. These results highlight the role of rodents in the disease, and also the potential role of MLVA in defining the molecular epidemiology of *L. interrogans*.

Received 13 June 2006

Accepted 27 July 2006

INTRODUCTION

Leptospira interrogans is one of the causative agents of the disease leptospirosis. Leptospirosis is considered an emerging zoonotic disease with a worldwide distribution [Bharti *et al.*, 2003; World Health Organization (WHO), 1999]. Molecular investigations into the epidemiology of *Leptospira* infections are important to determine whether an outbreak caused by a clonal source has occurred, and which

species of animal is the vector of the disease (Slack *et al.*, 2005).

Various molecular techniques have been described for studying the molecular epidemiology of *Leptospira*, including PFGE, and randomly amplified polymorphic DNA (RAPD) and fluorescent amplified fragment length polymorphism (FAFLP) analysis. All of these methods suffer from significant drawbacks, including low discriminatory power, poor reproducibility, and the high level of technical skill required to perform the method. As an alternative to these methods we have developed a multiple locus variable number of tandem repeats analysis (MLVA) based upon agarose gel electrophoresis. Whilst this method fulfils the aim of producing a technically simple method that is reproducible and suitable for any laboratory with the bare minimum of equipment, it does have several drawbacks,

Abbreviations: FAFLP, fluorescent amplified fragment length polymorphism; FAM, 6-carboxyfluorescein; HEX, 6-hexachlorofluorescein; HGDI, Hunter–Gaston diversity index; I_a , index of association; MLVA, multiple-locus variable number of tandem repeats analysis; MLST, multilocus sequence typing; MST, minimum spanning tree; MVT, MLVA type; NED, 2,7,8-benzo-5-fluoro-2,4,7-trichloro-5-carboxyfluorescein; SLV, single-locus variant; VNTR, variable number of tandem repeat.

Table 1. Primers and PCR conditions used in the *L. interrogans* MLVA method

Locus	Forward primer	Reverse primer	PCR conditions	
			Number of cycles	Annealing temperature (°C)
V27	HEX-AAGTTCGTCGGGTGAGC	TGATTTCTTTCGGTGCC	25	50
V29	NED-TGGGTGCCGGGTGT	ATGCCACATCTCATCCATTAC	25	50
V30	FAM-ATAGGTTTCGGCGTTAGTA	TTTAGATGTTTCGCTTTGG	25	54
V36	HEX-TGGCGTCGAAGACAAA	ACTCTACCAGGAGATTATCAAA	33	52
V50	NED-CTTGTTGGATCACAATACGAACTATA	GGTAAGGGACAAAGTAAGTGAAGC	25	58

including a loss of resolution through the use of agarose gel electrophoresis and the rounding of repeats to whole numbers to simplify analysis (Slack *et al.*, 2005).

To improve resolution and accuracy in bacterial variable number of tandem repeat (VNTR) methods, capillary electrophoresis has become the preferred methodology, due to the availability of multiple fluorescent labels, and to greater accuracy and reproducibility (Lindstedt *et al.*, 2004a, b; Lista *et al.*, 2006). Whilst capillary electrophoresis is the preferred methodology, the expense of purchasing a DNA sequencer may prevent some laboratories from changing from agarose gel electrophoresis; however, this can be compensated by the other critical improvements to the method, such as the optimized PCR amplification reactions and reaction conditions.

To further harness the power of capillary electrophoresis, we utilized pooled PCR products, and used three fluorescent dyes and a commercially available labelled DNA marker, which allowed the sizing of fragments of up to 1000 bp. The analysis of the MLVA data was also improved by using an allele designation system proven in several published MLVA articles (Lindstedt *et al.*, 2004a; Whatmore *et al.*, 2006) combined with the assignment of an MLVA type number similar to that used in multilocus sequence typing (MLST) experiments. To ascertain the effects of these changes, we applied the improved MLVA method to the *L. interrogans* serovar Australis strains that had been previously characterized by the original MLVA method, and present in this study a redefined epidemiological model of this serovar in Queensland, Australia. Furthermore, in this research, we established the relative value of the MLVA method as an epidemiological tool by comparing it with a previously published FAFLP method, using a subset of the *L. interrogans* serovar Australis isolates.

METHODS

***L. interrogans* serovar Australis strains.** A total of 96 isolates of *L. interrogans* serovar Australis were obtained from the culture collection housed at the WHO/FAO/OIE Collaborating Centre for Reference and Research on Leptospirosis. Fifty-seven isolates were from human sources, whilst 39 were from non-human sources, such

as rodents, dogs and native Australian animals. All isolates had been previously identified using standard serological methods, including the cross-agglutination absorption test (CAAT). All isolates had been previously characterized by MLVA (Slack *et al.*, 2005).

DNA extraction. *Leptospira* DNA was extracted by the following method: Ellinghausen McCullough Johnson Harris (EMJH) broth (500 µl) containing actively growing leptospires was centrifuged in a microcentrifuge tube at 12 000 *g* for 4 min. The supernatant was removed and the pellet resuspended in 200 µl PBS. This suspension was then extracted using the High Pure PCR Template Preparation kit (Roche) as per the manufacturer's instructions.

MLVA analysis. The VNTR loci used in this method have been previously described (Slack *et al.*, 2005); however, the primers were redesigned to prevent the spurious PCR products that were evident in the previous version of the method. The primers were synthesized by Invitrogen or Applied Biosystems, and the 5' end of the forward primers was labelled with the fluorophores 6-carboxyfluorescein (FAM), 2,7,8-benzo-5-fluoro-2,4,7-trichloro-5-carboxyfluorescein (NED) or 6-hexachlorofluorescein (HEX) (Table 1). PCR amplification was performed in 25 µl final volumes. All reactions contained 1 × PCR buffer, 2 mM MgCl₂, 200 µM dNTPs, 12.5 pmol forward and reverse primer (Table 1), 1 U AmpliTaq Gold (Applied Biosystems), 2.0 µl template DNA, and double-distilled water (ddH₂O) to make up the final volume. The PCR reactions were run on a PE9700 thermal cycler (Applied Biosystems) using the following conditions: 95 °C for 10 min, varying cycles of 94 °C for 30 s, annealing at varying temperatures (refer to Table 1) for 30 s, and extension at 72 °C for 1 min. The high-stringency conditions combined with the use of minimal thermal cycling prevented the formation of spurious PCR products that could interfere with the subsequent fragment sizing. Five microlitres of the PCR products from loci V27, V29 and V30 were pooled in a microcentrifuge tube, as were the products from V36 and V50. Both pooled PCR mixes were further diluted with 100 µl ddH₂O, and 3 µl of each pooled PCR product was combined with 21.5 µl HiDi formamide (Applied Biosystems) and 0.5 µl X-rhodamine MapMarker 1000 XL (BioVenture). The samples were denatured at 95 °C for 5 min and cooled rapidly to 4 °C before being loaded onto the ABI 310 capillary sequencer (Applied Biosystems). Electrophoresis was performed using a 47 cm capillary filled with performance-optimized polymer 4 (Applied Biosystems) at 60 °C for 45 min with a running voltage of 15 kV, and an injection time of 10 s at an injection voltage of 15 kV. Each VNTR locus could be identified by colour and assigned a size by the GENESCAN software (Applied Biosystems). This size was then converted into an allele designation, as shown in Table 2, which in turn formed the allele string for the five loci. The allele string was

Table 2. Allele designation for *L. interrogans* serovar Australis

Allele designation	Size (bp) of:				
	V27	V29	V30	V36	V50
1	317	415	583	746	491
2	363	744	722	709	583
3	367	791	676	532	480
4	414	651	532	593	796
5	644	464	630	796	537
6	818	840	475	737	396
7	—	697	576	758	345
8	—	701	555	643	632
9	—	774	—	695	465
10	—	607	—	500	531
11	—	—	—	658	486
12	—	—	—	—	445
13	—	—	—	—	404

constructed in the following order: V27-V29-V30-V36-V50. Each unique allele string was given a unique MLVA type (MVT) number, using a process similar to that used in MLST experiments (Fig. 1).

FAFLP analysis. FAFLP analysis was performed as described elsewhere, using the AFLP Microbial Fingerprinting kit (Applied Biosystems) (Vijayachari *et al.*, 2004). One microlitre of the six selective PCR products generated by the kit was mixed with 23 µl HiDi formamide and 1 µl GeneFlo-625 (CHIMERx), and denatured at 95 °C for 5 min. The products were then loaded onto the ABI 310 instrument, and injected into a 47 cm capillary filled with performance-optimized polymer 4 at 15 kV for 12 s. The fragments were separated at 13 kV for 35 min. The resulting electropherograms were manipulated using the Genotyper version 2.5 software (Applied Biosystems), and the combined allele sizes were exported into an Excel spreadsheet. An Excel macro described elsewhere (Rinehart, 2004) was used to convert the alleles into a binary sequence suitable for analysis, using Bionumerics software (Applied Maths). Each unique FAFLP binary pattern was assigned a letter code (e.g. AA, BB, CC) to allow easier referencing of the data.

Phylogenetic and statistical data analysis. Phylogenetic analysis was performed using Bionumerics version 4.0 (Applied Maths). Dendrograms were constructed using the categorical MLVA combined with the Ward algorithm. Population modelling was performed using the minimum spanning tree (MST) algorithm with the highest number of single-locus variants (SLVs) as the priority rule and the creation of hypothetical types was disabled. The Hunter–Gaston diversity index (HGDI) was calculated as previously described (Hunter & Gaston, 1988), using the VNTR diversity and confidence extractor software (V-DICE) available at the Health Protection Agency bioinformatics tools website (<http://www.hpa-bioinfotools.org.uk/cgi-bin/DICI/DICI.pl>). The degree of linkage disequilibrium was determined using the index of association (I_a) using the software available at the MLST website (<http://www.mlst.net>). I_a is calculated by comparing the observed variance (V_o) in the distribution of allelic mismatches in all pair-wise comparisons of the allelic profiles with that of the expected variance (V_e) in a freely recombining population minus one [$I_a = (V_o/V_e) - 1$]. Significant linkage disequilibrium is established if the observed variance in the MLVA allele profiles is greater than the maximum variance observed with 1000 randomized allele profiles ($P < 0.001$) (Smith *et al.*, 1993).

RESULTS AND DISCUSSION

Improvements to the published MLVA method

The improved MLVA method incorporated several notable changes from the original MLVA method (Slack *et al.*, 2005). The first change was the redesign of the primers used to amplify the VNTR loci; the original primers for V27–V36 produced spurious products due to partial homology of the primer to the flanking regions of the target sequence. The primers used in this method were redesigned to prevent mispriming and thus produce a single band for each locus. In addition, the V8 locus used in the previous MLVA method (Slack *et al.*, 2005) was not used in this study, as under capillary electrophoresis it produced two alleles in certain strains (A. T. Slack and others, unpublished data). Although the changes to the PCR primers and conditions increased the complexity of the method and would make multiplexing the PCR difficult, the priority of this study was to ensure that the correct PCR product could be identified during capillary electrophoresis, with no interference from background or spurious amplification products. To utilize the full potential of the ABI 310 instrument, we pooled the PCR products labelled with three spectrally separate fluorophores (FAM, HEX and NED) and employed filter set D to sort and size the VNTR products. This resulted in clear and interpretable electropherograms for analysis. MapMarker 1000 XL contains 23 single-stranded DNA bands from 50 to 1000 bp, and the use of this marker was critical to the study, given that there is no GENESCAN marker available to size fragments > 500 bp.

The final improvement to the method was the conversion of VNTR allele sizes into a numerical string that would ultimately be suitable for both the end-user and the analysis software. The method used was adapted from a previously published method (Lindstedt *et al.*, 2004a). Briefly, the fragment sizes were assigned an allele designation as shown in Table 2; this designation was converted into an allele string in the following order: V27–V29–V30–V36–V50. To further simplify this allele string for both the end-user and also for discussion, we assigned an MVT number for every unique allele string present in the dataset, e.g. the allele string 1-1-1-1-1 was designated MVT1. This reduction of the allele string to an MVT brings the method into line with other common typing methods, such as PFGE and MLST, and allows easier interpretation of the result by non-specialist end-users of the data.

Comparison of MLVA data from the improved method with that from the original method

When the improved MLVA method was applied to the previously characterized *L. interrogans* serovar Australis isolates, each VNTR locus was found to have between six (V27) and 13 (V50) different alleles (Table 3). This is an increase in the number of alleles found in the VNTR loci when compared with the original method, which had between three (V27) and eight (V50) alleles. The differences in the number of alleles found in the modified method could

Fig. 1. (a, b) Clustering analysis of *L. interrogans* serovar Australis using improved MLVA data. *R. sordidus*, *Rattus sordidus*; *R. leucopus*, *Rattus leucopus*; *R. fuscipes*, *Rattus fuscipes*; *R. lorif*, *Rattus lorif*.

be attributed to the change in the data analysis: every unique size was given a unique allele designation, whereas in the original method the size was converted to a repeat unit and was further rounded to a whole number.

In addition to the changes in the number of alleles present in the VNTR loci, there was also a change in the diversity index for each locus. The diversity index is usually modelled on the mathematical equations of Nei (Weir, 1990), Simpson

Table 3. Comparison of data between the original (Slack *et al.*, 2005) and modified MLVA methods

Locus	Original MLVA method				Modified MLVA method			
	HDGI (<i>n</i> =96)*	Confidence interval†	Number of different repeats present at each locus	Max (pi)‡	HDGI (<i>n</i> =96)*	Confidence interval†	Number of different repeats present at each locus	Max (pi)‡
V27	0.098	0.060–0.135	3	0.949	0.594	0.548–0.640	6	0.604
V29	0.817	0.799–0.835	8	0.323	0.794	0.769–0.819	10	0.375
V30	0.763	0.745–0.780	7	0.333	0.689	0.663–0.715	8	0.396
V36	0.642	0.617–0.667	5	0.475	0.633	0.595–0.671	11	0.542
V50	0.674	0.648–0.699	6	0.424	0.563	0.510–0.616	13	0.646

*HGDI (Hunter & Gaston, 1988) as calculated by the V-DICE software (<http://www.hpa-bioinfotools.org.uk/cgi-bin/DICI/DICI.pl>).

†Confidence interval: the precision of the HGDI, expressed as the 95 % upper and lower boundaries.

‡Max (pi): the proportion of samples (expressed as a decimal) that possess the most frequently found repeat at the VNTR locus.

(Simpson, 1949) or Hunter and Gaston (Hunter & Gaston, 1988), and is based on the probability of two unrelated strains being characterized as the same type. These indexes may be used to compare typing methods and select the most discriminatory system (Hunter & Gaston, 1988). For this study we used the HGDI, as calculated by the V-DICE software. With the exception of V27, all other loci showed a drop in the HGDI when compared with the results of the original MLVA method. However, this drop in the HGDI at the individual locus must be tempered by the change in the mean diversity across the five loci, given that the allele string is analysed rather than individual loci. The modified method had a mean diversity of 0.654 and the original had a mean diversity of 0.599. The reason for positive change in mean diversity is that the V27 locus had an increase in diversity from 0.098 in the original method to 0.549 in the modified method (Table 3).

The *I*_a for the modified MLVA data was 0.827 (*P* < 0.001); this result shows that there was significant linkage disequilibrium, which implies a low rate of recombination between the alleles (Smith *et al.*, 1993).

Comparison of the improved MLVA method with FAFLP analysis

FAFLP analysis was performed on 30 MLVA-typed *L. interrogans* serovar Australis isolates, and clustering analysis was performed using Bionumerics for comparison with the modified MLVA (Fig. 2). Sixteen FAFLP profiles (AA–PP) were present amongst the 30 isolates. FAFLP type PP was the most prevalent type in the dataset (9/30 profiles, 30 %). The results from the comparison show that the FAFLP results were comparable with those obtained by MLVA. To highlight the concordance of the two methods, we have linked the two cluster analyses using a series of connecting solid and dashed lines and roman numeral designations (i–v). MVT clones such as MVT1 (FAFLP type PP), MVT32 (LL) MVT50 (NN), MVT5 (OO) and MVT6 (MM) were grouped together with the respective isolates by FAFLP. The majority of the remaining MVTs had unique FAFLP profiles

(AA–KK), with the exception of the following two isolates: QHR581B (MVT3) belonged to profile PP by FAFLP (the same FAFLP profile as MVT1) and LT1342 (MVT28) was found to have profile OO (the same as MVT5) in the FAFLP analysis (Fig. 2). Overall, the two methods were found to be comparable with each other; however, the MLVA method produced a greater number of unique patterns than the FAFLP method. A larger comparison study using other serovars may need to be conducted to establish which of the two methods has the highest resolution for the typing of *L. interrogans* strains. MLVA and FAFLP required equivalent levels of technical skill and equipment to perform the method; however, the data analysis for the MLVA method was easier to perform, given that it involves only five alleles compared with the 150–200 alleles produced by the FAFLP method. A further advantage of the MLVA method over FAFLP analysis is that the MLVA method could be applied in a non-culture-based situation, i.e. it could be possible to type an organism directly from a *Leptospira* PCR-positive DNA extract.

Redefining the molecular epidemiology of *L. interrogans* serovar Australis in Queensland, Australia, using the improved MLVA method

As the improved MLVA method was found to be superior to the original MLVA method, we used the data generated from the improved method to redefine the molecular epidemiology of *L. interrogans* serovar Australis in Queensland, Australia. Fifty MVTs were found in the 96 isolates tested in this study (Fig. 1a, b). When clustering analysis was applied to the isolates, it revealed eight clades (A–G), each containing isolates from both human and non-human sources (Fig. 1a, b). Clades A, B and C contained mostly the isolates from the Innisfail region (36/49, 73 %), and clades D, E, F and G contained mostly isolates from the Tully region (30/46, 65 %). Clade B was the largest group, containing 28 isolates, the majority of which were from the Innisfail area (24/28 isolates, 86 %). In the Innisfail region, MVT1 was the dominant clone (20/39, 51 %) (Fig. 3), and in the Tully region, MVT28 (8/32, 25 %) was the dominant

Fig. 2. Clustering analysis of FAFLP data in comparison with that of the improved MLVA method.

clone (Fig. 4). Similarities in the clustering analysis were present between the improved and the original MLVA method, such as the geographical specificity of certain MVTs to either the Innisfail or the Tully areas. The two dominant

clones, MVT1 and MVT28, showed remarkable geographical specificity for the Innisfail and Tully regions (Figs 2 and 3). The improved MLVA method allowed the examination of the epidemiological links between the animal and human

Fig. 3. Distribution of MVTs in the Innisfail region.

Fig. 4. Distribution of MVTs in the Tully region.

isolates of *L. interrogans* serovar Australis. Using the clustering analysis we could demonstrate strain clonality between rodent isolates, human and rodent isolates, and also rodent and canine isolates (Fig. 1a, b).

When population modelling was performed using the MST algorithm (Fig. 5), all MVTs, with the exception of MVT21, 33, 39, 41 and 45, clustered together.

MVT1 was considered to be the progenitor clone, given that it had the highest number of SLV isolates within the dataset. The population modelling using MST paralleled the geographical proximity of the isolates, with the majority of MVTs belonging to a single, albeit quite diverse, complex (Fig. 5).

In conclusion, in this study we have described an improved MLVA method for *L. interrogans* that utilizes novel primers

Fig. 5. MST population modelling of *L. interrogans* serovar Australis based upon MLVA data. The numbers shown are the MVT numbers.

and optimized PCR conditions for VNTR amplification, and multiple fluorescent dye capillary electrophoresis to sort and size the PCR products. This improved method was found to be superior to the original method, and produced a greater number of unique profiles when compared with FAFLP analysis. Additionally, we were able to use the improved MLVA method to redefine the molecular epidemiology of *L. interrogans* serovar Australis, and through clustering analysis we were able to demonstrate strain clonality amongst rodent, canine and human isolates.

ACKNOWLEDGEMENTS

The authors wish to thank Mr Shane Byrne for critically reviewing this manuscript.

REFERENCES

- Bharti, A. R., Nally, J. E., Ricaldi, J. N. & 8 other authors (2003). *Leptospirosis*: a zoonotic disease of global importance. *Lancet Infect Dis* **3**, 757–771.
- Hunter, P. R. & 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.
- Lindstedt, B. A., Vardund, T., Aas, L. & Kapperud, G. (2004a). Multiple-locus variable-number tandem-repeats analysis of *Salmonella enterica* subsp. *enterica* serovar Typhimurium using PCR multiplexing and multicolor capillary electrophoresis. *J Microbiol Methods* **59**, 163–172.
- Lindstedt, B. A., Vardund, T. & Kapperud, G. (2004b). Multiple-locus variable-number tandem-repeats analysis of *Escherichia coli* O157 using PCR multiplexing and multi-colored capillary electrophoresis. *J Microbiol Methods* **58**, 213–222.
- Lista, F., Faggioni, G., Valjevac, S. & 13 other authors (2006). Genotyping of *Bacillus anthracis* strains based on automated capillary 25-loci multiple locus variable-number tandem repeats analysis. *BMC Microbiol* **6**, 33.
- Rinehart, T. A. (2004). AFLP analysis using GeneMapper software and an Excel macro that aligns and converts output to binary. *Biotechniques* **37**, 186–188.
- Simpson, E. H. (1949). Measurement of diversity. *Nature* **163**, 688.
- Slack, A. T., Dohnt, M. F., Symonds, M. L. & Smythe, L. D. (2005). Development of a multiple-locus variable number of tandem repeat analysis (MLVA) for *Leptospira interrogans* and its application to *Leptospira interrogans* serovar Australis isolates from Far North Queensland, Australia. *Ann Clin Microbiol Antimicrob* **4**, 10.
- Smith, J. M., Smith, N. H., O'Rourke, M. & Spratt, B. G. (1993). How clonal are bacteria? *Proc Natl Acad Sci U S A* **90**, 4384–4388.
- Vijayachari, P., Ahmed, N., Sugunan, A. P., Ghousunnissa, S., Rao, K. R., Hasnain, S. E. & Sehgal, S. C. (2004). Use of fluorescent amplified fragment length polymorphism for molecular epidemiology of leptospirosis in India. *J Clin Microbiol* **42**, 3575–3580.
- Weir, B. S. (1990). *Genetic Data Analysis: Methods for Discrete Population Genetic Data Analysis*. Sunderland, MA: Sinauer Associates.
- Whatmore, A. M., Shankster, S. J., Perrett, L. L., Murphy, T. J., Brew, S. D., Thirlwall, R. E., Cutler, S. J. & MacMillan, A. P. (2006). Identification and characterization of variable-number tandem-repeat markers for typing of *Brucella* spp. *J Clin Microbiol* **44**, 1982–1993.
- WHO (1999). *Leptospirosis* worldwide, 1999. *Wkly Epidemiol Rec* **74**, 237–242.