A combined approach of VNTR and MLST analysis: improving molecular typing of Argentinean isolates of *Leptospira interrogans*

Karina Caimi[1], Vanina Varni[1], Yamil Melendez[1], Ariel Koval[3], Bibiana Brihuega[2], Paula Ruybal[1]/+

[1]Biotechnology Institute [2]Pathobiology Institute, National Institute of Agricultural Technology,
Castelar, Buenos Aires, Argentina [3]Biogénesis Bagó SA, Garín, Buenos Aires, Argentina

Leptospirosis is an emerging infectious disease that has been identified as both a human and animal health problem worldwide. Regular outbreaks associated with specific risk factors have been reported in Argentina. However, there are no available data concerning the genetic population level for this pathogen. Therefore, the aim of this work was to describe the genetic diversity of Leptospira interrogans *through the application of two molecular typing strategies: variable number of tandem repeats (VNTR) and multilocus sequence typing (MLST). For this purpose, seven reference strains and 18 non-epidemiologically related isolates from diverse hosts and Argentinean regions were analysed. Among them, nine genotypes and seven sequence types (STs), including three unreported STs, were described using VNTR and MLST, respectively. eBURST analysis demonstrated that ST37 was the most frequent and founder genotype of a clonal complex (CCs) containing STN1 and STN3, suggesting the importance of studying the serovars belonging to this CC in Argentina. The data from maximum parsimony analysis, which combined both techniques, achieved intra-serovar discrimination, surmounted microscopic agglutination test discrepancies and increased the discriminatory power of each technique applied separately. This study is the first to combine both strategies for* L. interrogans *typing to generate a more comprehensive molecular genotyping of isolates from Argentina in a global context.*

Key words: *Leptospira interrogans* - variable number tandem repeats - multilocus sequence typing

Leptospirosis is a systemic disease characterised by fever, renal and hepatic insufficiency, pulmonary manifestations and reproductive failure that affects humans, wild and domestic animals (primarily dogs), cattle and swine. In humans, leptospirosis is transmitted through direct or indirect contact with the urine of infected animals and can be lethal. Respiratory manifestations, involving severe pulmonary oedema and haemorrhages, are the primary cause of death in some outbreaks; 20% fatality rates have been reported (Levett 2001, Bharti et al. 2003, Adler & de la Peña Moctezuma 2010).

Leptospirosis is one of the most widely distributed zoonosis in the world (WHO 1999). The Intergovernmental Panel on Climate Change has suggested that the predicted increase in heavy rainfall in the XXI century may increase the risk of leptospirosis through the contamination of flood waters or run-off by rodent populations (Solomon et al. 2007). This disease is difficult to diagnose at the clinical laboratory level and therefore it is frequently not identified and consequently severely neglected. Leptospirosis is re-emerging globally and numerous outbreaks have occurred worldwide within the last decade. However, the true spread and increase of leptospirosis remains unknown because the quality and availability of diagnostic tests, testing facilities and surveillance systems are highly variable and frequently absent (Hartskeerl et al. 2011).

Outbreaks in humans and animals have regularly been reported in Argentina. Exposure could occur during floods and rural activities, with the majority of cases reported during the warm and rainy season. However, limited data on the current disease burden in Argentina are available due to limitations associated with national laboratory surveillance (i.e., passive case ascertainment, referral of case samples and information to a reference laboratory and representativeness of identified cases). The microscopic agglutination test (MAT) is the most widely used diagnostic tool for leptospirosis because it is specific for serogroups. However, this test cannot discriminate between vaccinated and infected hosts. Moreover, the MAT is laborious and some discrepancies have been reported. For instance, the panel of serovars used might not include all local circulating isolates (Vanasco et al. 2008).

The causal agent of leptospirosis is a spirochete of the genus *Leptospira*, which has the ability to parasitise a wide variety of wild and domesticated animals (Haake et al. 2004). Traditionally, two *Leptospira* species have been identified, i.e., *Leptospira interrogans* and *Leptospira biflexa*, as pathogenic and non-pathogenic leptospires, respectively. The serovar is the basic identifier, characterised on the basis of serological criteria. To date, nearly 300 serovars have been identified under the species *L. interrogans* alone, which are distributed among 25 different serogroups of antigenically similar serovars (Kmety & Dikken 1993). A classification system based on DNA-DNA hybridisation studies has been introduced, which now comprises 20 *Leptospira* species (Bezerra da Silva et al. 2011). The enormous inventory of serovars, based primarily on an ever-changing surface antigen repertoire, presents an artificial and unreliable scenario of strain diversity. It is therefore difficult to track strains

whose molecular identity changes according to the host and the environmental niches they inhabit and encounter. Molecular tools, including restriction endonuclease assays (Brown & Levett 1997), pulsed field gel electrophoresis (Herrmann et al. 1992), restriction fragment length polymorphism (Zuerner et al. 1993), arbitrarily primed polymerase chain reaction (PCR) (Perolat et al. 1994), variable number of tandem repeats (VNTR) analysis (Majed et al. 2005) and fluorescent amplified fragment length polymorphism (Vijayachari et al. 2004), have also been employed to sub-classify and catalogue leptospiral agents. However, these techniques suffer from certain disadvantages that include the requirement of a large quantity of pure and high-quality DNA, low discriminatory power, low reproducibility, ambiguous interpretation of data and problems associated with the transfer of data between different laboratories (Majed et al. 2005). In recent years, multilocus sequence typing (MLST) has emerged as another PCR-based technique, which makes use of automated DNA sequencers to assign and characterise the alleles present in different target genes. This method generates sequence data on a low to high-throughput scale, which is unambiguous and suitable for epidemiological and population studies (Maiden 2006). A number of MLST schemes have been

described to type *Leptospira* sp. (Ahmed et al. 2006, Leon et al. 2010, Thaipadungpanit et al. 2007). Thaipadungpanit et al. (2007) developed an MLST scheme to describe genetic variation, particularly amongst *L. interrogans* isolates from humans and rodents during the Thai outbreak in the early 2000s.

The aim of this work was to evaluate the genetic variability of *L. interrogans* in Argentina using a combination of VNTR and MLST analyses to improve the characterisation of Argentinean isolates of *L. interrogans* from different regions and diverse hosts (Majed et al. 2005, Thaipadungpanit et al. 2007).

MATERIALS AND METHODS

Strains - The bacterial strains used in this study belong to the collections of the Biogenesis-Bago Laboratories (Buenos Aires, Argentina) and the Leptospirosis Department of Pathobiology Institute [National Institute of Agricultural Technology (INTA), Castelar, Buenos Aires] (Table I) and correspond to non-epidemiologically related isolates. Both collections were previously characterised using partial 16S rRNA sequencing and the MAT (Merien et al. 1992, Levett 2001). The Royal Tropical Institute (KIT, The Netherlands) kindly provided the reference strains. The samples were cultured in Ellinghausen-Mc-Cullough-Johnson-Harris liquid medium until saturation.

TABLE I

Details of *Leptospira interrogans* strains and isolates analyzed in the present study

Serogroup	Serovar	Strain	Year	Location	Host
Icterohaemorrhagiae	Copenhageni	Fiocruz L1-130	1996	Brazil (ref. strain)	Human
Icterohaemorrhagiae	Icterohaemorrhagiae	IcteroI	1914	Japan (ref. strain)	Human
Icterohaemorrhagiae	Icterohaemorrhagiae	RGA	1915	Belgium (ref. strain)	Human
Icterohaemorrhagiae	Lai	Lai 56601	ND	China (ref. strain)	Human
Canicola	Canicola	Hond Utrecht IV	ND	Netherlands (ref. strain)	Dog
Canicola	Portlandvere	MY1039	ND	Jamaica (ref. strain)	Human
Pomona	Pomona	Pomona	ND	Australia (ref. strain)	Human
Pomona	ND	Marcos Juárez	1960	Marcos Juárez, Córdoba, Argentina	Bovine
Pomona	ND	Fulton	1976	Fulton, Buenos Aires, Argentina	Bovine
Pomona	ND	Longchamps	1977	Longchamps, Buenos Aires, Argentina	Human
Pomona	ND	Pujato	1981	Pujato, Santa Fe, Argentina	Bovine
Pomona	ND	Cañuelas I	1982	Cañuelas, Buenos Aires, Argentina	Porcine
Pomona	ND	San Alfredo I	1984	Santa Fe, Santa Fe, Argentina	Porcine
Pomona	ND	Marcos Paz	1982	Marcos Paz, Córdoba, Argentina	Porcine
Pomona	ND	Cañuelas II	1986	Cañuelas, Buenos Aires, Argentina	Porcine
Pomona[a]	ND	Corrientes 266	1985	Mercedes, Corrientes, Argentina	Bovine
Pomona	ND	Rojas	1985	Rojas, Buenos Aires, Argentina	Bovine
Pomona	ND	AK-RFB	2007	ND, Buenos Aires, Argentina	Bovine
Canicola	ND	Comadreja	2005	Saladillo, Buenos Aires, Argentina	Weasel
Sejroe	ND	3705 SENASA	ND	Garín, Buenos Aires, Argentina	Human
Sejroe[a]	ND	3705 LiW-04-W1	ND	Garin, Buenos Aires, Argentina	-
Sejroe[a]	ND	V3P	ND	ND, Córdoba, Argentina	ND
Icterohaemorrhagiae	ND	La Cava 1	1996	San Isidro, Buenos Aires, Argentina	Rat
Icterohaemorrhagiae	ND	Cañuelas III	1993	Cañuelas, Buenos Aires, Argentina	Porcine
Icterohaemorrhagiae	ND	E3	2005	ND, Buenos Aires, Argentina	Human

a: inconclusive microscopic agglutination test characterization; ND: not determined.

DNA extraction - A 500 μL aliquot of culture was centrifuged for 10 min at 10,000 rpm. The pellet was washed twice in phosphate buffered saline and resuspended in 200 μL of buffer containing 50 mM Tris-ethylenediamine tetraacetic acid (EDTA), 100 μM NaCl, 20 μL 10% sodium dodecyl sulfate and 10 μL 10 mg/mL proteinase K. The samples were incubated for 2 h at 37ºC, extracted in a (24:1) chloroform/isoamyl alcohol mixture and the DNA was precipitated from the aqueous phase using sodium acetate in isopropanol. The pellet was washed in 70% ethanol and resuspended in 30 μL of water.

VNTR - The PCR products for VNTR *loci* 4, 7, 10, 11, 19, 23 and 31 were assessed for all of the strains used in this study as previously described (Majed et al. 2005) with slight modifications. The primer sequences used in this study were the same as previously reported (Majed et al. 2005), except for VNTR11a and VNTR9b, which were redesigned to improve their amplification efficiency (VNTR11a-bis, 5'-CACAGGTCGGAATTT-GTCT-3' and VNTR9b-bis, 5'-TCGCTCTRCAGGTCG-GTGTT-3'). The PCR conditions consisted of an initial denaturation step at 94ºC for 5 min followed by 35 cycles of amplification with a denaturation step at 94ºC for 30 s, annealing at 55ºC for 30 s and extension at 72ºC for 1 min; a final extension step was performed at 70ºC for 10 min. The PCR products were analysed on 2% agarose gel stained with ethidium bromide and the molecular weights were estimated by comparison with a 100-bp DNA ladder (Promega, USA) using GelCompar software (Bionumerics, Applied Maths, Belgium).

MLST - Seven *loci*, including NAD(P) transhydrogenase subunit alpha (*pnt*A), 2-oxoglutarate dehydrogenase decarboxylase component (*suc*A), ribokinase (*pfk*B), triosephosphate isomerase (*tpi*A), rod shape-determining protein rodA (*mre*A), UDP-N-acetylglucosamine pyrophosphorylase (*glm*U) and a putative long-chain-fatty acid-CoA ligase (*fad*D), were analysed as previously described (Thaipadungpanit et al. 2007). The primers for the *fad*D *locus* were redesigned in the present study to increase their amplification efficiency (*fad*D-Fnew: 5'-ACGTGATCTCCCTTATGCCAAGCA-3', 5'-*fad*D-Rnew: ATCCAACCGACAGAAGTATGGCGT-3').

The amplification reactions were performed in a total volume of 50 μL containing 10 pmol of each primer, 100 mM dNTPs, 5 mM MgCl$_2$ and 0.05 U of DNA polymerase (GoTaq, Promega, USA). The PCR cycling consisted of an initial denaturation step at 94ºC for 5 min, followed by 30 amplification cycles (denaturation at 94ºC for 30 s, annealing at 52ºC for *pnt*A, *suc*A, *pfk*B, *tpi*A, *mre*A and *fad*D or 50ºC for *glm*U for 15 s and extension at 72ºC for 50 s) and a final extension step at 72ºC for 7 min. The PCR products were analysed on 1% agarose gels and purified using EDTA-ethanol precipitation. The sequencing was performed in the Genotyping and Sequencing facility at the Biotechnology Institute (INTA). The contigs were assembled using the STADEN Package software (MRC-LMB, UK). The sequence type (ST) was determined in each case according to the 7 MLST *loci* sequences published in the world database leptospira.mlst.net (hosted at the Imperial College, UK). Sequence Type Analysis and Recombination Tests software was used to evaluate allelic and profile frequen-

cies and to perform several MLST statistical analyses for recombination, such as the maximum chi-squared test (1,000 random trials), the index of association (1,000 random trials) and the ratio of non-synonymous (d_N) to synonymous (d_S) substitutions per nucleotide site (Nei and Gojobori method) (Jolley et al. 2001). The goeBURST algorithm (goeburst.phyloviz.net/), which uses the same clustering rules as eBURST but provides a global optimal solution, was used to determine the relationships between STs (Francisco et al. 2009, Feil et al. 2004). Clonal complexes (CCs) were defined as STs that are linked through single locus variants (SLVs) and named on the basis of the predicted founder ST, which is the ST associated with the most SLVs. Phylogenetic relationships among concatenated sequences of housekeeping alleles were inferred using the maximum-likelihood (ML) method and the Kimura-2-parameter nucleotide substitution model tested with 500 bootstrap replications as implemented in MEGA5 (Tamura et al. 2011). The *Leptospira kirschneri* serovar Grippotyphosa strain Moskva V (ST62) was considered as the outgroup.

MLST-VNTR combined analysis - Maximum-parsimony (MP) analysis was implemented in Tree Analysis Using New Technology, version 1.1 (Goloboff et al. 2008). The input data were encoded in a matrix comprising 25 taxa and 15 characters (7 MLST and 8 VNTR *loci*). Traditional search testing was performed with 100 replicates using tree bisection reconnection and 500 bootstrap replicates.

Calculation of the discriminatory power index - Simpson's index of diversity was calculated for each individual genotyping method and for the combination of methods as previously described (Hunter & Gaston 1988).

RESULTS

The national incidence and prevalence of leptospirosis in Argentina are unknown, particularly in animal health due to difficulties diagnosing the disease. To contribute to the knowledge concerning the genetic diversity of leptospirosis in Argentina, *L. interrogans* isolates were analysed using VNTR and MLST. We studied isolates primarily obtained from livestock and wild animals in Argentinean regions with the highest leptospirosis prevalence (Table I). A total of 18 Argentinean isolates were characterised. These isolates were previously serotyped using the MAT, which established 11 isolates belonging to the Pomona serogroup, three isolates belonging to the Icterohaemorrhagiae serogroup, three isolates belonging to the Sejroe serogroup and one isolate belonging to the Canicola serogroup. Among the 18 isolates identified, three presented inconclusive results using the MAT (Table I). The reference strains IcteroI, Fiocruz L1-130, RGA, Hond Utrecht, MY1039 and Pomona were also analysed to complete the VNTR profile and standardise the MLST technique. Furthermore, the MLST and VNTR profiles from the Lai 56601 strain were elucidated using genome sequence analysis.

L. interrogans VNTR - The application of the previously described VNTR typing method (Majed et al. 2005) resulted in grouping the isolates into nine different genotypes arbitrarily named "a" thru "i" (Table II). VNTR 4, 9, 10, 19 and 23 were the most informative and

VNTR 10 was the most discriminative. VNTR 7, 11 and 31 were less discriminative, particularly the latter, which allowed the identification of only two different alleles.

The isolates characterised as belonging to the Icterohaemorrhagiae serogroup were further divided into four genotypes. Genotype "a" grouped the isolate La Cava 1 together with the reference strains Ictero I, RGA and Fiocruz L1-130, while genotype "e" was unique among these isolates. Genotype "h" contained the reference strain Lai 56601, which is completely different from the other reference strains in the Icterohaemorrhagiae serogroup. Genotype "f" grouped the reference strain Hond Utrecht IV together with two isolates, E3 and Comadreja; the latter was serotyped as serogroup Canicola.

The clusters "b", "c" and "d" corresponded with the isolates characterised as serogroup Pomona, where genotype "b" was predominant, containing nine of the 11 identified isolates. The V3P isolate (serogroup Sejroe) had similar genotyping results. Genotype "i" grouped the Corrientes 266 isolate together with the MY1039 reference strain. The isolates belonging to the Sejroe serogroup were grouped into genotype "g".

L. interrogans MLST profiles - The seven *loci* (*pntA, sucA, pfkB, tpiA, mreA, glmU and fadD*) were analysed as previously described (Thaipadungpanit et al. 2007). Significant variation in the *fadD locus* amplification efficiency was detected and therefore a new set of primers was designed that allowed a 25-fold increase in the detection limit and generated a 700-bp PCR product that included the *fadD* MLST region (Fig. 1).

Seven different STs were described among the 25 strains. The predominant ST was ST37 (serogroups Pomona and Canicola) followed by ST17 (serogroup Icterohaemorrhagiae), ST58 (V3P isolate), ST1 (Lai 56601 strain) (Table II) and three new STs identified in the San Alfredo I (STN1), 3705 LiW-04-W1 (STN2) and Longchamps (STN3) isolates. Moreover, STN1 and STN3 shared six alleles with ST37 and described new *pfkB* and *glmU loci* variants, respectively. STN2 consisted of a new allele combination between ST58 (*glmU, pntA, sucA, fadD* and *tpiA* alleles) and ST17 (*pfkB* and *mreA* alleles).

Clonal relationships among the STs - The allelic profiles among the 25 strains were compared using eBURST to detect CCs. This analysis indicated that 13 isolates were closely related, establishing a single CC within the collection studied. ST37 was the most frequent profile and common ancestor clone, generating two single-*locus* variants (SLVs): STN1 and STN3.

Recombination and selection at the MLST loci - Table II displays unique *locus* frequencies ranging from two-five alleles: four *glmU* alleles, two *pntA* alleles, three *sucA* alleles, two *fadD* alleles, four *tpiA* alleles, five *pfkB* alleles and four *mreA* alleles. The number of d_S and d_N were estimated, merging the new *glmU* and *pfkB* variants with the global allele database (leptospira.mlst.net). All genes had a d_N/d_S ratio of < 1, indicating that the gene regions included in the present analysis were subject to stabilising selection consistent with the general requirements for MLST *loci* (Maiden 2006). Significant linkage disequilibrium was observed in both local and global datasets. Moreover, intragenic recombination evaluated using

the maximum chi-squared test revealed putative recombination sites within two MLST *loci* (*sucA* and *pfkB*).

Phylogenetic relationships among the STs - The ML dendrogram, constructed from concatenated sequences, identified three major clusters, consistent with the previously established serogroups (Fig. 2). Although the Pomona, Icterohaemorrhagiae and Sejroe serogroups were discriminated, the Canicola serogroup clustered with Pomona because they shared the same MLST pattern.

Combined analysis of VNTR profiles and STs - According to the discriminatory power index, further discrimination of the genetic variants was obtained using VNTR (data not shown). However, both strategies displayed typing shortcomings, such as the inability to properly distinguish the Canicola serogroup from the Pomona serogroup (MLST) or the Icterohaemorrhagiae serogroup from the Canicola serogroup (VNTR). Consequently, the VNTR genotypes were combined with the STs to generate a new code: STVNTR (Table II), which described 13 different patterns that render an increased discriminatory power index. In addition, the combined patterns were subjected to MP analysis. As shown in Fig. 3, the MP cladogram described four clusters (cluster 1: Icterohaemorragiae; cluster 2: Sejroe; cluster 3: Canicola; cluster 4: Pomona) that corresponded with previously serotyped groups and enabled an accurate classification of the three inconsistently serotyped strains as Canicola (Corrientes 266), Sejroe (3705 LiW-04-W1) and Pomona (V3P).

DISCUSSION

In Argentina, rural occupations continue to be an important risk factor for leptospirosis. Previous data support Icterohaemorrhagiae as the most prevalent serogroup associated with human infections, while Pomona is associated with swine and bovine, Ballum with urban rodent transmission and the Sejroe serogroup with cattle (Vanasco et al. 2008). There are few studies concerning the prevalence of *Leptospira* in Argentina and many of these studies reported a number of cases with agglutination titres against multiple serogroups (Vanasco et al. 2008). Thus, the application of molecular techniques is necessary to both discern MAT discrepancies and describe the genetic variability of this pathogen in Argentina.

More reliable and robust methods have recently been applied in *L. interrogans* typing, such as VNTR and MLST (Majed et al. 2005, Ahmed et al. 2006, Leon et al. 2010, Thaipadungpanit et al. 2007). Both techniques are easy to perform and do not require large amounts of DNA. Whereas VNTR is the current method accepted worldwide, MLST promises a more straightforward characterisation of *L. interrogans* isolates because it is amenable to standardisation through available online databases, such as leptospira.mlst.net and allows access to current molecular epidemiology data from many laboratories (Adler & de la Peña Moctezuma 2010).

Although Majed et al. (2005) demonstrated the efficacy of VNTR to discriminate between *L. interrogans* serovars, VNTR 9 and 11 displayed little or no amplification products from the Argentinean isolates. Further analysis based on the *L. interrogans* Lai 56601 and Fiocruz L1-130 genomes revealed nucleotide polymorphisms within

TABLE II

Variable number of tandem repeats (VNTR) and multilocus sequence typing (MLST) *loci* described for Argentinean isolates and reference strains

Species	Serogroup	Serovar	Strain	MLST								VNTR									
				glmU	pntA	sucA	fadD	tpiA	pfkB	mreA	ST	4	7	9	10	11	19	23	31	VNTR	STVNTR
L. interrogans	Icterohaemorrhagiae	Copenhageni	Fiocruz L1-130	1	1	2	1	2	10	4	17	2	1	13	7	1	2	0	3	a	A
L. interrogans	Icterohaemorrhagiae	Icterohaemorrhagiae	Icterol	1	1	2	1	2	10	4	17	2	1	13	7	1	2	0	3	a	A
L. interrogans	Icterohaemorrhagiae	Icterohaemorrhagiae	RGA	1	1	2	1	2	10	4	17	2	1	13	7	1	2	0	3	a	A
L. interrogans	Icterohaemorrhagiae	Lai	Lai 56601	1	1	1	1	1	1	1	1	5	3	4	2	4	6	5	4	h	K
L. interrogans	Canicola	Canicola	Hond Utrecht IV^a	3	3	3	2	3	4	5	37	1	10	2	3	2	10	2	3	f	H
L. interrogans	Canicola	Portlandvere	MY1039^a	3	3	3	2	3	4	5	37	1	10	2	3	2	10	2	4	i	L
L. interrogans	Pomona	Pomona	Pomona^a	3	3	3	2	3	4	5	37	2	0	6	14	2	8	1	3	d	M
L. interrogans	Pomona	ND	Marcos Juárez	3	3	3	2	3	4	5	37	4	1	6	10	2	8	2	3	b	B
L. interrogans	Pomona	ND	Fulton	3	3	3	2	3	4	5	37	4	1	6	10	1	8	2	3	c	C
L. interrogans	Pomona	ND	Longchamps	New	3	3	2	3	4	5	STN3	4	1	6	10	2	8	2	3	b	D
L. interrogans	Pomona	ND	Pujato	3	3	3	2	3	4	5	37	4	1	6	10	2	8	2	3	b	B
L. interrogans	Pomona	ND	Cañuelas I	3	3	3	2	3	4	5	37	4	1	6	10	2	8	2	3	b	B
L. interrogans	Pomona	ND	San Alfredo I	3	3	3	2	3	New	5	SNT1	2	0	6	14	2	8	1	3	d	E
L. interrogans	Pomona	ND	Marcos Paz	3	3	3	2	3	4	5	37	4	1	6	10	2	8	2	3	b	B
L. interrogans	Pomona	ND	Cañuelas II	3	3	3	2	3	4	5	37	4	1	6	10	2	8	2	3	b	B
L. interrogans	Pomona	ND	Rojas	3	3	3	2	3	4	5	37	4	1	6	10	1	8	2	3	c	C
L. interrogans	Pomona	ND	AK-RFB	3	3	3	2	3	4	5	37	4	1	6	10	2	8	2	3	b	B
L. interrogans	Pomona^b	ND	V3P	3	3	3	2	3	4	5	37	4	1	6	10	2	8	2	3	b	B
L. interrogans	Canicola^b	ND	Corrientes 266	3	3	3	2	3	4	5	37	1	10	2	3	2	10	2	4	i	L
L. interrogans	Canicola	ND	Comadreja	3	3	3	2	3	4	5	37	1	10	2	3	2	10	2	3	f	H
L. interrogans	Sejroe	ND	3705 SENASA	9	2	1	1	4	13	9	58	3	1	3	11	2	17	7	3	g	I
L. interrogans	Sejroe^b	ND	3705 LiW-04-W1	9	2	1	1	4	10	4	STN2	3	1	3	11	2	17	7	3	g	J
L. interrogans	Icterohaemorrhagiae	ND	La Cava 1	1	1	2	1	2	10	4	17	2	1	13	7	1	2	0	3	a	A
L. interrogans	Icterohaemorrhagiae	ND	Cañuelas III	1	1	2	1	2	10	4	17	2	2	13	7	2	2	0	3	e	F
L. interrogans	Icterohaemorrhagiae	ND	E3	1	1	2	1	2	10	4	17	1	10	2	3	2	10	2	3	f	G

a: MLST and VNTRs profiles obtain from bibliography; *b*: strains reclassification by molecular typing; *fadD*: long-chain-fatty acid-CoA ligase; *glmU*: UDP-N-acetylglucosamine pyrophosphorylase; *mreA*: rod shape-determining protein rodA; *pntA*: NAD(P) transhydrogenase subunit alpha; *sucA*: 2-oxoglutarate dehydrogenase decarboxylase component; *tpiA*: triosephosphate isomerase; ND: not determined; *pfkB*: ribokinase. Sequence type (ST) number was determinate based on the global database (leptospira.mlst.net). STN1, STN2 and STN3 correspond to new STs due to new *loci* variants (New) or new combination of preexistent *loci* variants. VNTR genotypes were encoded with letters "a" to "i". ST and VNTR genotypes were merged in a new code named STVNTR.

the VNTR11a and VNTR9a primer sequences. Therefore, this study designed alternative primers for these two *loci* to obtain better amplification results.

The VNTR analysis rendered nine genotypes among the isolates from the Argentinean collections. The Icterohaemorrhagiae serogroup was discriminated into four genotypes. Among these, genotype "f" contained two isolates belonging to different serogroups: Icterohaemorrhagiae and Canicola. As Pavan et al. (2011) previously showed, the Comadreja and Corrientes 266 isolates exhibited VNTR profiles consistent with the serogroup Canicola serovar Canicola (genotype Hond Utrecht IV) and serogroup Canicola serovar Portlandvere (genotype MY 1039), respectively. The three genotypes described serogroup Pomona within this collection and "b" was the predominant genotype.

The addition of VNTR11 allowed confirmation of the isolates classified by Pavan et al. (2011), particularly Corrientes 266, which had previously been misclassified using MAT. Moreover, the current results were consistent with the data published by Pavan, where both the Icterohaemorrhagiae and Pomona serogroups were further discriminated upon the inclusion of this *locus*.

The advantages of the VNTR strategy lie in the comparison of PCR products on agarose gels according to the size of the repeats, ranging between 15-50 bp; the amplified products indicate a high conservation of repeat units and flanking regions among *L. interrogans* serovars (Majed et al. 2005). Nevertheless, the absence of complete unit repeats, insertions and/or deletions could result in amplified products that are not compatible with variations in the copy number of a full-length repeat.

The emergence of MLST as a reference typing technique in molecular microbiology resulted from the improved knowledge of bacterial evolution and population biology. The increasing availability and decreasing cost of high-throughput nucleotide sequence determination and developments in information technology, specifically the development of the Internet, avoids the expensive and potentially hazardous exchange of isolates (Maiden 2006).

Curated databases of nucleotide sequence data enable direct comparisons of bacterial isolates and allow microbiologists a rapid and accurate method for isolate characterisation. The development of web-accessible MLST databases offers scalability and flexibility, enabling participating laboratories to maintain ownership and control of their data while benefiting from centralised assignments of allele sequences and profiles, ensuring data integrity and consistency (Jolley et al. 2004).

Recently, the MLST method for *L. interrogans* typing was incorporated into the web-accessible database mlst.net (Thaipadungpanit et al. 2007). This dataset includes 262 reference strains and STs primarily obtained from isolates from Asia, Australia, Europe, Brazil and Jamaica, to name a few.

Despite the scattered spatial and local geographic distribution of the analysed isolates, seven STs among 25 strains were described using the method according to Thaipadungpanit. The eBURST analysis identified a single CC composed of the founder and most frequent genotype ST37, as well as two SLVs (STN1 and STN3) comprising only isolates from the Pomona serogroup. STN1 and STN3 were generated from the new *pfk*B and *glm*U variants, respectively (Table II).

The global analysis of the predominant ST37 revealed that this profile was widely distributed in relation to the host range (30% bovine, 25% human, 15% porcine, 5% dog, 5% opossum and 5% weasel), serogroups (50% Pomona, 20% Canicola, 10% Pyrogenes and 5% Grippotyphosa) and geographic distribution (50% Argentina, 15% Laos, 10% Thailand, 5% Australia, 5% Brazil, 5% Jamaica, 5% Netherlands and 5% Sri Lanka). In addition, this ST grouped isolates from the main cattle production region in Argentina and it has been present since 1960, which could indicate that these isolates belong to a biologically successful clone that might have survival advantages in detrimental environments.

Previous data support that whilst mosaic patterns of genes encoding OMPs and Lig proteins have been reported, mosaicism does not affect all leptospiral outer membrane proteins (McBride et al. 2009, Hartskeerl et al. 2011). However, it was previously suggested that recombination events occur within the housekeeping genes, particularly within the *loci* selected for MLST analyses (Nalam et al. 2010). In the present study, the global recombination
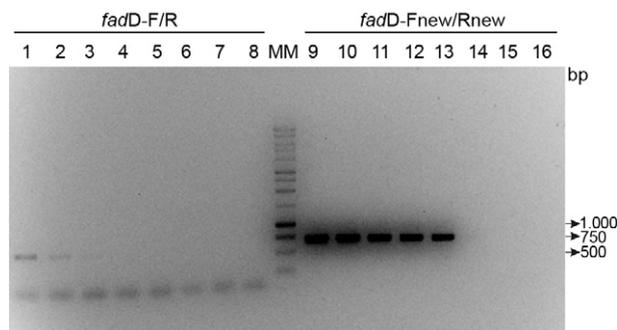
Fig. 1: former and new long-chain-fatty acid-CoA ligase (*fad*D) primers sensitivity. Lines 1-7, 9-15: 100 ng, 50 ng, 25ng, 10 ng, 1 ng, 1 pg, 1 fg; 8, 16: negative controls; MM: molecular marker, 1Kb DNA ladder (Promega, USA).
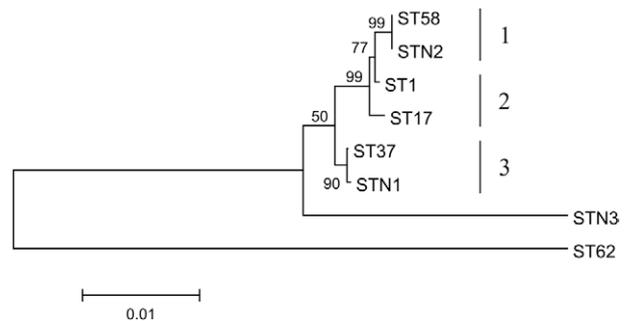
Fig. 2: phylogenetic relationships between sequence types (STs) based on maximum likelihood (ML) method. Phylogenetic relationships between concatenated sequences of housekeeping alleles were inferred using ML method and Kimura-2-parameter nucleotide substitution model tested with 500 bootstrap replications as implemented in MEGA5. *Leptospira kirschneri* serovar Grippotyphosa strain Moskva V (ST62) was considered as outgroup. Clusters described were identified as follows: 1: serogroup Sejroe; 2: serogroup Icterohaemorragiae; 3: serogroups Canicola and Pomona.

analysis for seven MLST *loci* displayed statistically significant putative recombination sites at the *suc*A and *pfk*B genes that could generate new alleles and therefore new haplotypes, such as STN1. Nevertheless, the d_N/d_S ratio revealed that all *loci* were subject to stabilising selection. Because recombination events seemed to be rare within the *L. interrogans* population, only a small proportion of recombinant alleles will drift to an observable frequency, while the less frequent events will be lost from the population. This idea is consistent with the low number of ST variants described in the present study. Moreover, alleles from different *loci* in the population are at linkage disequilibrium, which, along with a low recombination rate, leads to the diversification of lineages through a bifurcating tree-like process (Feil & Spratt 2001).

Regarding the host range, the analysed collection demonstrated a bias, where the bovine isolates exhibited more variability in the VNTR analysis than in the MLST, in which a single ST was described (ST37) (Tables I, II). However, the swine isolates revealed greater variability for both techniques (Tables I, II). In addition, STVNTR "B" was the most frequent pattern in our collection (Table II), which was present in bovine and porcine isolates from the main productive region in Argentina despite the year of isolation. These data were consistent with the wide distribution of ST37 as previously discussed.

The phylogenetic relationships within the STs described for the Argentinean collection were consistent with the serogroup classification, particularly the Comadreja and E3 isolates, which shared the VNTR "f", ST37 (serogroup Canicola) and ST17 (serogroup Icterohaemor-

rhagiae), respectively. Both the VNTR analysis and the MLST characterisation confirmed the misclassification of Corrientes 266 using MAT and the grouping of the V3P isolates into serogroups Canicola and Pomona, respectively. Consistent with the calculated discrimination index, the VNTR analysis was able to differentiate the Pomona and Canicola serogroups, as well as the genetic variants within serogroup Pomona.

In addition, the global eBURST analysis revealed 12 CCs spanning 51 STs from diverse serogroups (Fig. 4). The predominant CC comprised 10 different STs, where ST37 was the founder genotype. Interestingly, two of the five SLVs (from the founder genotype) were derived from the Argentinean collection. The ML phylogenetic study was also performed using the collection containing 51 STs (data not shown). The topology of this new dendrogram was consistent with the dendrogram shown in Fig. 2 and the CCs from the eBURST analysis.

Recently, Nalam et al. (2010) compared the FA-FLP, MLST and VNTR techniques, demonstrating that MLST is a more efficient approach, obtaining species level discrimination. In this study, the MLST analysis and VNTR profiles were combined using a simplified matrix with an MP approach. The resulting cladogram reflected contributions from both techniques, displaying a more accurate description of the genetic divergence among Argentinean *L. interrogans*.

More isolates and *loci* are needed to obtain an improved typing method for *L. interrogans* characterisation in Argentina. Hence, the isolation of *Leptospira* must be encouraged. Meanwhile, as the present study suggests, the combined VNTR and MLST analysis seems to be a better approach for the molecular typing
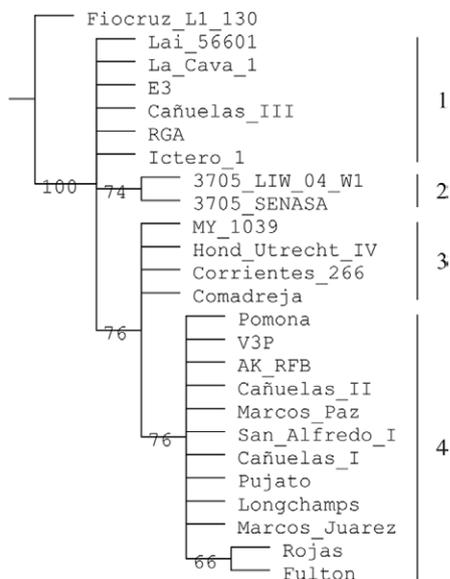

Fig. 3: cladogram obtained by maximum-parsimony (MP) analysis from sequence type variable number of tandem repeats (STVNTR) matrix data. MP analysis was implemented in Tree Analysis Using New Technology, version 1.1 (Goloboff et al. 2008). Input data was encoded in a matrix composed by 25 taxa and 15 characters (7 multilocus sequence typing and 8 VNTR *loci*). Traditional search testing was performed with 100 replicates by tree bisection reconnection and a 500 replicates bootstrap. Clusters described were identified as follows: 1: serogroup Icterohaemorragiae; 2: serogroup Sejroe; 3: serogroup Canicola; 4: serogroup Pomona.
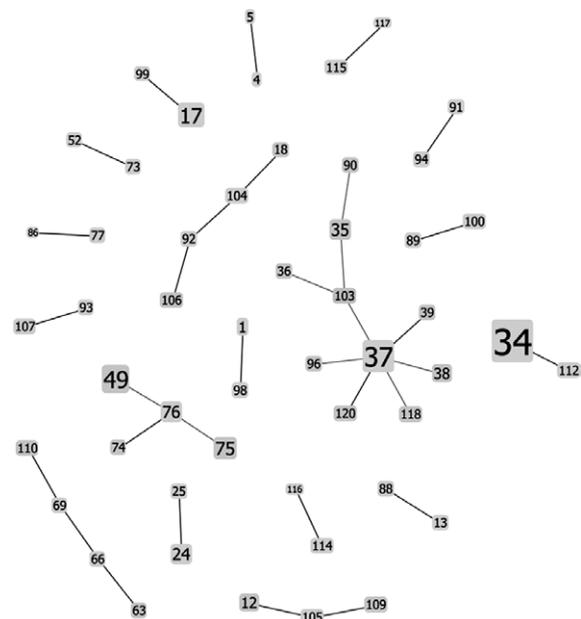

Fig. 4: eBURST diagram of relationships between 51 *Leptospira* spp sequence types (STs). Representation of the 18 clonal complexes (CC) for *Leptospira* spp with ST37 as the determined founder for the largest CC formed by 10 STs. The size of each square is proportional to the number of isolates with that particular ST in a logarithmic scale. STs assigned to the same CC are linked by straight lines.

of local *Leptospira* isolates to obtain greater insight into the evolutionary biology and epidemiology of this important pathogen in Argentina.

REFERENCES

Adler B, de la Peña Moctezuma A 2010. *Leptospira* and leptospirosis. *Vet Microbiol 140*: 287-296.

Ahmed N, Devi SM, Valverde M de L, Vijayachari P, Machang'u RS, Ellis WA, Hartskeerl RA 2006. Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. *Ann Clin Microbiol Antimicrob 5*: 28.

Bezerra da Silva J, Carvalho E, Hartskeerl RA, Ho PL 2011. Evaluation of the use of selective PCR amplification of LPS biosynthesis genes for molecular typing of *Leptospira* at the serovar level. *Curr Microbiol 62*: 518-524.

Bharti A, Nally J, Ricaldi J, Matthias M, Diaz M, Lovett M, Levett P, Gilman R, Willig M, Gotuzzo E, Vinetz J 2003. Leptospirosis: a zoonotic disease of global importance. *Lancet Infect Dis 3*: 757-771.

Brown PD, Levett PN 1997. Differentiation of *Leptospira* species and serovars by PCR restriction endonuclease analysis, arbitrarily primed PCR and low-stringency PCR. *J Med Microbiol 46*: 173-181.

Feil EJ, Li BC, Aanensen DM, Hanage WP, Spratt BG 2004. eBURST: inferring patterns of evolutionary descent among clusters of related bacterial genotypes from multilocus sequence typing data. *J Bacteriol 186*: 1518-1530.

Feil EJ, Spratt BG 2001. Recombination and the population structures of bacterial pathogens. *Annu Rev Microbiol 55*: 561-590.

Francisco AP, Bugalho M, Ramirez M, Carrico JA 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. *BMC Bioinformatics 10*: 152.

Goloboff PA, Farris JS, Nixon KC 2008. TNT, a free program for phylogenetic analysis. *Cladistics 24*: 774-786.

Haake DA, Suchard MA, Kelley MM, Dundoo M, Alt DP, Zuerner RL 2004. Molecular evolution and mosaicism of leptospiral outer membrane proteins involves horizontal DNA transfer. *J Bacteriol 186*: 2818-2828.

Hartskeerl RA, Collares-Pereira M, Ellis WA 2011. Emergence, control and re-emerging leptospirosis: dynamics of infection in the changing world. *Clin Microbiol Infect 17*: 494-501.

Herrmann JL, Bellenger E, Perolat P, Baranton G, Girons IS 1992. Pulsed field gel electrophoresis of NotI digests of leptospiral DNA: a new rapid method of serovars identification. *J Clin Microbiol 30*: 1696-1702.

Hunter PR, Gaston MA 1988. Numerical index of the discriminatory ability of typing systems: an application of Simpson's index of diversity. *J Clin Microbiol 26*: 2465-2466.

Jolley KA, Feil EJ, Chan MS, Maiden MC 2001. Sequence type analysis and recombinational tests (START). *Bioinformatics 17*: 1230-1231.

Jolley KA, Chan MS, Maiden MC 2004. mlstdbNet - distributed multilocus sequence typing (MLST) databases. *BMC Bioinformatics 5*: 86.

Kmety E, Dikken H 1993. *Classification of the species Leptospira interrogans and history of its serovars*, University Press, Groningen, 104 pp.

Leon A, Pronost S, Fortier G, Andre-Fontaine G, Leclercq R 2010. Multilocus sequence analysis for typing *Leptospira interrogans* and *Leptospira kirschneri. J Clin Microbiol 48*: 581-585.

Levett PN 2001. Leptospirosis. *Clin Microbiol Rev 14*: 296-326.

Levett PN, Morey RE, Galloway RL, Steigerwalt AG 2006: *Leptospira broomii* sp. nov. isolated from humans with leptospirosis. *Int J Syst Evol Microbiol 56*: 671-673.

Maiden MC 2006. Multilocus sequence typing of bacteria. *Annu Rev Microbiol 60*: 561-588.

Majed Z, Bellenger E, Postic D, Pourcel C, Baranton G, Picardeau M 2005. Identification of variable-number tandem-repeat *loci* in *Leptospira interrogans sensu stricto. J Clin Microbiol 43*: 539-545.

McBride AJ, Cerqueira GM, Suchard MA, Moreira AN, Zuerner RL, Reis MG, Haake DA, Ko AI, Dellagostin OA 2009. Genetic diversity of the leptospiral immunoglobulin-like (Lig) genes in pathogenic *Leptospira* spp. *Infect Genet Evol 9*: 196-205.

Merien F, Amouriaux P, Perolat P, Baranton G, Saint Girons I 1992. Polymerase chain reaction for detection of *Leptospira* spp in clinical samples. *J Clin Microbiol 30*: 2219-2224.

Nalam K, Ahmed A, Devi SM, Francalacci P, Baig M, Sechi LA, Hartskeerl RA, Ahmed N 2010. Genetic affinities within a large global collection of pathogenic *Leptospira*: implications for strain identification and molecular epidemiology. *PLoS ONE 5*: e12637.

Pavan ME, Cairó F, Pettinari MJ, Samartino L, Brihuega B 2011. Genotyping of *Leptospira interrogans* strains from Argentina by multiple-*locus* variable-number tandem repeat analysis (MLVA). *Comp Immunol Microbiol Infect Dis 34*: 135-141.

Perolat P, Merien F, Ellis WA, Baranton G 1994. Characterization of *Leptospira* isolates from serovars hardjo by ribotyping, arbitrarily primed PCR and mapped restriction site polymorphisms. *J Clin Microbiol 32*: 1949-1957.

Solomon S, Qin D, Manning M, Marquis M, Averyt K, Tignor M, Miller HL 2007. Contribution of Working Group I to the Fourth Assessment Report of the Intergovernmental Panel on Climate Change 2007. Available from: ipcc.ch/publications_and_data/ar4/wg1/en/contents.html.

Tamura K, Peterson D, Peterson N, Stecher G, Nei M, Kumar S 2011. MEGA5: Molecular Evolutionary Genetics Analysis using Maximum Likelihood, Evolutionary Distance and Maximum Parsimony Methods. *Mol Biol Evol 24*: 1596-1599.

Thaipadungpanit J, Wuthiekanun V, Chierakul W, Smythe LD, Petkanchanapong W, Limpaiboon R, Apiwatanaporn A, Slack AT, Suputtamongkol Y, White NJ, Feil EJ, Day NPJ, Peacock SJ 2007. A Dominant clone of *Leptospira interrogans* associated with an outbreak of human leptospirosis in Thailand. *PLoS Negl Trop Dis 1*: 1-6.

Vanasco NB, Schmeling MF, Lottersberg J, Costa F, Ko AI, Tarabla HD 2008. Clinical characteristics and risk factors of humans leptospirosis in Argentina (1999-2005). *Acta Tropica 107*: 255-258.

Vijayachari P, Ahmed N, Sugunan AP, Ghousunnisa S, Rao KR, Hasnain SE, Sehgal SC 2004. Use of fluorescent amplified fragment length polymorphism for molecular epidemiology of leptospirosis in India. *J Clin Microbiol 42*: 3575-3580.

WHO - World Health Organization 1999. Leptospirosis worldwide. *Wkly Epidemiol Rec 74*: 237-244.

Zuerner RL, Herrmann JL, Girons IS 1993. Comparison of genetic maps for two *Leptospira interrogans* serovars provides evidence for two chromosomes and intra species heterogeneity. *J Bacteriol 175*: 5445-5451.