Infection, Genetics and Evolution xxx (2013) xxx–xxx

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

journal homepage: www.elsevier.com/locate/meegid

Reassessment of MLST schemes for *Leptospira* spp. typing worldwide

Vanina Varni [a], Paula Ruybal [a], Juan José Lauthier [b], Nicolás Tomasini [b], Bibiana Brihuega [c], Ariel Koval [d], Karina Caimi [a,*]

[a] *Biotechnology Institute, INTA Castelar, Buenos Aires, Argentina*
[b] *Molecular Epidemiology Unit, School of Health Sciences, National University of Salta, Salta, Argentina*
[c] *Pathobiology Institute, INTA Castelar, Buenos Aires, Argentina*
[d] *Biogenesis Bagó, Garín, Buenos Aires, Argentina*

ARTICLE INFO

Article history:
Available online xxxx

Keywords:
Leptospira spp.
Multilocus sequence typing
Reassessment

ABSTRACT

Leptospirosis is a neglected zoonosis of global importance. Several multilocus sequence typing (MLST) methods have been developed for *Leptospira* spp., the causative agent of leptospirosis.

In this study we reassessed the most commonly used MLST schemes in a set of worldwide isolates, in order to select the loci that achieve the maximum power of discrimination for typing *Leptospira* spp. Global eBURST algorithm was used to detect clonal complexes among STs and phylogenetic relationships among concatenated and individual sequences were inferred through maximum likelihood (ML) analysis. The evaluation of 12 loci combined to type a subset of strains rendered 57 different STs. Seven of these loci were selected into a final scheme upon studying the number of alleles and polymorphisms, the typing efficiency, the discriminatory power and the ratio dN/dS per nucleotide site for each locus. This new 7-locus scheme was applied to a wider collection of worldwide strains. The ML tree constructed from concatenated sequences of the 7 loci identified 6 major clusters corresponding to 6 *Leptospira* species. Global eBURST established 8 CCs, which showed that genotypes were clearly related by geographic origin and host. ST52 and ST47, represented mostly by Argentinian isolates, grouped the higher number of isolates. These isolates were serotyped as serogroups Pomona and Icterohaemorrhagiae, showing a unidirectional correlation in which the isolates with the same ST belong to the same serogroup.

In summary, this scheme combines the best loci from the most widely used MLST schemes for *Leptospira* spp. and supports worldwide strains classification. The Argentinian isolates exhibited congruence between allelic profile and serogroup, providing an alternative to serological methods.

Published by Elsevier B.V.

1. Introduction

Leptospirosis is probably the most widely distributed zoonosis in the world (World Health Organisation, 1999). This disease is transmitted through direct or indirect contact with the urine of infected animals (Bharti et al., 2003).The spectrum of human disease associated with leptospirosis is extremely wide, ranging from subclinic infections to severe pulmonary hemorrhagic syndrome (Segura et al., 2005). Fatality rates of 20% have been reported in different outbreaks and occurred particularly in humid and warm climates (Vanasco et al., 2008; Levett 2001; Bharti et al., 2003; Adler and de la Peña Moctezuma, 2010). The difficulty in the diagnosis of leptospirosis at the clinical laboratory level makes it a severely neglected disease, particularly in developing countries.

In Argentina, leptospirosis was first recognized in humans in 1915. However, there are limited data concerning the current disease burden. In a recent study, Icterohaemorrhagiae and Pomona have been identified as the major serogroups involved in human leptospirosis; moreover, the main risk factor in humans was the extended contact with floods that occurred in recent years in Argentina (Vanasco et al., 2008).

Leptospirosis is caused by infection with pathogenic spirochetes of the genus *Leptospira* spp. Traditionally, the serovar has been the basic serologic identifier characterized by 300 different antigenic types that produce a high diversity of strains. Usually, it is difficult to track isolates through the serological approach because molecular identity within a serological taxon varies according to the host and environmental niches they inhabit and encounter (Nalam et al., 2010). However, DNA composition analyses have identified 20 *Leptospira* species with nine major pathogenic species, *L. interrogans, L. borgpetersenii, L. santarosai, L. noguchii, L. weilii, L. kirschneri, L. alexanderi, L. alstonii* and*L. kmetyi* and six non-pathogenic species. Recently, an "intermediate" group of *Leptospira* species isolated from animals and humans has been identified, which

* Corresponding author. Address: Biotechnology Institute, INTA Castelar, Nicolas Repetto and Los Reseros s/n, B1686IGC Hurlingham, Buenos Aires, Argentina. Tel.: +54 11 4621 1447; fax: +54 11 4621 0199.
E-mail address: kcaimi@cnia.inta.gov.ar (K. Caimi).

1567-1348/$ - see front matter Published by Elsevier B.V.
http://dx.doi.org/10.1016/j.meegid.2013.08.002

produces mild clinical symptoms or no symptoms at all. The pathogenic implication of this group is yet unknown (Kmety and Dikken 1993; Boonsilp et al., 2013; Smythe et al. 2012).

Identification and typing play an important role in understanding disease epidemiology (Ahmed et al., 2011). In the last decade, the first molecular typing approaches based on genomic sequence applied to *Leptospira* spp. were the fluorescent amplified fragment length polymorphism (FAFLP) (Vijayachari et al., 2004), multilocus variable number of tandem repeats analysis (MLVA) (Majed et al., 2005; Slack et al., 2006) and multilocus sequence typing (MLST) (Ahmed et al., 2006; Thaipadungpanit et al., 2007; Leon et al., 2010). Among these, the disadvantage of FAFLP is that it requires high quality reagents and highly purified, concentrated genomic DNA. The MLVA methods described until now generally do not expand beyond *L. interrogans*. Therefore, MLST has become the method of choice according to its multiple benefits based on the generation of sequences from different genes on a high-throughput scale which is unambiguous and suitable for epidemiological and population studies (Maiden, 2006).

Three different MLST schemes have already been described for *Leptospira* spp. due to the significant genetic divergence of this genus, which hinders the development of a single scheme (Ahmed et al., 2006; Thaipadungpanit et al., 2007; Leon et al., 2010). These schemes are composed of 6 or 7 different loci, but initially only Ahmed's scheme allowed the typing of the major pathogenic *Leptospira* species, beyond *L. interrogans* and *L. kirschneri*. Ahmed and colleagues performed a comparison between the two major schemes (Ahmed's and Thaipadungpanit's schemes) concluding that the 7-locus scheme should be applied for *Leptospira* spp. typing (Ahmed et al., 2011). Recently the 7-locus scheme was modified by Boonsilp et al., (2013) that excluded *fad*D locus, included a new locus (*cai*B) and modified primers sequences in order to amplify all *Leptospira* species. This modification enabled the scheme to be applied to the seven pathogenic species of leptospires, similarly to what had been previously achieved by Ahmed's scheme. This new scheme was tested with a higher number of isolates, but most of them were primarily obtained from a single human outbreak occurred in Thailand, which could have biased the genetic background of the strain collection.

In this context, the aim of our study was to reassess the existing MLST schemes by finding a combination of loci that would reach the maximum power of discrimination between non-epidemiologically related strains from the pathogenic species of *Leptospira* spp.

2. Materials and methods

2.1. Bacterial isolates, DNA extraction, and species identification

The 91 isolates of *Leptospira* spp. used in this study were selected based on the available sequences of the 13 loci after combining the MLST schemes described by Ahmed and colleagues and by Thaipadungpanit and colleagues. Twenty-five Argentinian non-epidemiologically related isolates from different hosts and date of isolation were included and previously characterized using partial 16S rRNA sequencing and Microagglutination test (MAT) (Merien et al., 1992; Levett 2001; Caimi et al., 2012). Bacteria were cultured in Ellinghausen-McCullough-Johnson-Harris (EMJH) liquid medium until saturation, and DNA extraction was performed as previously described (Caimi et al., 2012). *Leptospira* spp. isolates and data from providers is shown in Supplementary Table 1.

2.2. MLST schemes, PCR conditions and sequencing

Twelve loci were tested in this study. They comprise NAD(P) transhydrogenase subunit alpha (*pnt*A), 2-oxoglutarate dehydrogenase decarboxylase component (*suc*A), ribokinase (*pfk*B), triosephosphateisomerase (*tpi*A), rod shape determining protein rodA (*mre*A), UDPN-acetyl glucosamine pyrophosphorylase (*glm*U), a putative long chain fatty acid-CoA ligase (*fad*D), adenylate kinase (*adk*), isocitrate dehydrogenase (*icd*A), outer membrane lipoprotein LipL32 (*lipL32*), 16S rRNA (*rrs*2), and the outer membrane lipoprotein LipL41 (*lipL41*). As many strains lacked preprotein translocase SecY (*sec*Y) sequence, this gene was excluded from the analysis. Moreover, Cerqueira et al., (2010) observed that *sec*Y presented only 16 alleles out of 38 described by the combination of*lig*B, *sec*Y,*rpo*B and *lipL41* loci, in contrast to *lipL41* which presented 28 alleles out of 38. On this basis it was decided to retain *lipL41* and not to include *sec*Y in this analysis.

The amplification reactions were performed in a final volume of 50 μL containing 0.5 pmol/μl of each primer, 0.8 mM of dNTPs, 2.5 mM of $MgCl_2$ and 1.25 U of Taq DNA polymerase (GoTaq, Promega, USA). The PCR cycling conditions consisted of an initial denaturation step at 95 °C for 5 min, followed by 35 amplification cycles (denaturation at 95 °C for 30 s, annealing at 52 °C for *pnt*A, *suc*A, *pfk*B, *tpi*A, *mre*A and *fad*D, 50 °C for *glm*U and 58 °C for *adk*, *icd*A, *lipL32*, *lipL41* and *rrs*2 for 30 s and extension at 72 °C for 50 s) and a final extension step at 72 °C for 7 min. The PCR products were analyzed on 1% agarose gels and purified using ethylene diamine tetra acetic acid disodium salt (EDTA)-ethanol precipitation.

Sequencing was performed in the Genotyping and Sequencing facility of the Biotechnology Institute (INTA). Both strands were sequenced on a Big Dye Terminator v3.1 kit from Applied Biosystems and analyzed on an ABI 3130XL genetic analyzer from the same supplier. Subsequently, the contigs were assembled using STADEN Package software (MRC-LMB, UK) (Staden, 1996). All the sequences for each locus of Argentinian isolates were deposited in GenBank database (see Table S2 for Accession numbers).

2.3. Sequence analysis

Sequence alignment was performed using MEGA version 5.1 (Tamura et al., 2011). Allelic profiles and sequence types (STs) were assigned by MLSTest software (http://mlstest.codeplex.com/). The number of alleles and polymorphisms, typing efficiency, and discriminatory power (Simpson's index) of each locus were also calculated through MLSTest software. START2 software was used to evaluate the ratio of non-synonymous (dN) to synonymous (dS) substitutions per nucleotide site (Nei and Gojobori method) (Jolley et al., 2001). Finally, we applied the goeBURST algorithm (goeburst.phyloviz.net/) to determine the relationships between STs (Francisco et al., 2009; Feil et al., 2004). Clonal complexes (CCs) were defined as STs linked by single locus variants (SLVs) criteria and named on the basis of the predicted founder ST. In addition, we analyzed the clonal complexes formed by linking STs through triple locus variants (TLVs) to expose further relations among STs.

Phylogenetic relationships among concatenated sequences of selected genes were inferred by maximum likelihood (ML) tested with 1000 bootstrap replications in MEGA5.1. The selection of the model of nucleotide substitution was performed through JModelTest software (Darriba et al., 2012; Guindon and Gascuel 2003) and the parameters were adjusted according to substitution model selected.

3. Results and discussion

3.1. Twelve loci MLST scheme

The application of the 12-locus scheme to a subset of 48 worldwide strains and 17 Argentinean isolates of *L. interrogans* and *L. kirschneri* rendered 57 different sequence types (STs). The

discriminatory power was of 0.85 and 1 for *L. interrogans* and *L. kirschneri* respectively. The number of alleles per locus ranged from 9 (*fad*D) to 28 (*adk*). Although the number of isolates tested was lower, these results agreed with those previously obtained by other authors (Ahmed et al., 2011; Boonsilp et al., 2013).

A maximum likelihood tree was constructed from the concatenated sequences of the 12 loci for the 65 isolates and was compared with ML trees obtained from concatenated sequences for each of the original schemes separately (Fig. S1). Although only *L. interrogans* and *L. kirschneri* were included, the phylogenetic analysis of the 12 loci supported the species assignments. Notably, the tree obtained from the 12 loci strongly sustained serogroup subdivision; it resolved some inconsistencies in the classification of strains observed in Ahmed's scheme, and showed phylogenetic relationships similar to those obtained with the Thaipadungpanit's scheme. The predominant serogroup within the *L. interrogans* cluster was Pomona followed by Icterohaemorrhagiae, corresponding to isolates primarily from Argentina.

Global eBURST analysis established 35 singletons and six clonal complexes (CC) by linking STs permissively through TLVs. Two preponderant CCs among the collection exhibited ST52 and ST47 as founder genotypes according to SLV criteria. Relations among STs described by CCs were consistent with ML analysis and showed that Argentinian isolates from different hosts were closely related (data not shown). Previous data support Icterohaemorrhagiae as the serogroup most frequently associated with human infections in Argentina, and Pomona mostly associated with pigs and cattle. These facts reinforce the role of animals as source of transmission of the disease to humans (Vanasco et al., 2008; Caimi et al., 2012).

3.2. Scheme standarization

The scheme optimization was performed estimating the genetic diversity (GD) for every possible combination of 2 to n−1 loci. When an increase in GD became unlikely after adding additional locus to the scheme, the minimum number of loci was reached. Starting from 12 loci and after analyzing 924 possible combinations, MLSTest software found four optimized 6-loci schemes which rendered a maximum of 57 STs. The combinations thrown by the software were: (a) *adk*, *icd*A, *lipL32*, *lipL41*, *mre*A, *glm*U; (b) *adk*, *icd*A, *lipL32, lipL41*, *rrs*2, *mre*A, *glm*U; (c) *adk*, *icd*A, *lipL32*, *lipL41*, *mre*A, *pnt*A and (d) *adk*, *icd*A, *lipL32*, *rrs*2, *mre*A, *pnt*A. Taken together, all these combinations comprise 5 loci from Ahmed's scheme and 3 from Thaipadungpanit's scheme. Based on this selection further analyses were performed. The different parameters analyzed (Table 1) included the number of alleles and the number of polymorphisms per locus, ranging from 10 (*lipL32*) to 28 (*adk*) and 18 (*rrs*2) to 106 (*adk*), respectively. Except for *rrs*2, the typing efficiency, defined as the number of different genotypes per polymorphic site, was similar between loci. The discriminatory power (DP) of the loci was also homogeneous, being *lipL41* the locus with the highest and *rrs*2 the locus with the lowest DP. The ratio dN/dS value for *rrs2* showed that this gene region is subjected to positive selection which could interfere in the subsequent phylogenetic signal studies. We therefore establish as the reassessed 7-loci scheme the combination of *adk, glm*U, *icd*A, *lipL32*,*lipL41*, *mre*A and *pnt*A genes. Although the software had thrown a 6-locus scheme, we decided to keep 7 loci in order to ensure the highest level of discrimination in larger isolate collections (Maiden, 2006).

3.3. Analysis of the reassessed 7 loci scheme

An extended collection of 116 isolates that belong to the major pathogenic species of *Leptospira* spp. was used to evaluate 7 selected loci (Table S1). This new scheme rendered a total of 96 different STs in the collection. These findings are similar to those obtained for the 12 loci scheme.

When the parameters studied for the loci selection were applied to the 7 loci, we obtained higher number of alleles and polymorphisms per locus. The typing efficiency was similar and the discriminatory power was higher than 0.837 for all loci. As expected, the dN/dS were in all cases less than 1 (Table 2). Even when the number of isolates analyzed in this study was smaller, the results were comparable to those obtained by Ahmed et al., (2011) and Boonsilp et al., (2013), which evidence the usefulness of the combination of loci selected above.

Phylogenetic analyses were performed through trees constructed from concatenated sequences of the reassessed scheme. Tree showed six clusters that matched species assignments. The *L. alexanderi* isolate grouped together with the *L. borgpetersenii* isolates. Clustering of *L. alexanderi* with other species had been observed by other authors. Ahmed et al., (2006) found that the only *L. alexanderi* isolate included in the analysis was genomically similar to *L. santarosai* isolates and clustered accordingly. In turn, Boonsilp and colleagues (2013) reported that *L. borgpetersenii* isolates were linked to the groups formed by the *L. santarosai* and *L. weilii* (Fig. 1). The major cluster belonged to *L. interrogans* isolates, which had a monophyletic origin with *L. kirschneri* and *L. noguchii* strains. Benjamin, 136/2/2 and Vleermius 90C strains did not cluster together with *L. interrogans* isolates. Coincidentally, they were distinct in that they belonged to Canicola serogroup. Also, two of them were related to *L. kirschneri. L. interrogans* An7705 strain did not group with any other species. It also had a unique allelic profile, the singleton ST10, that shared with *L. interrogans* strains only 3 (*glm*U, *mre*A and *pnt*A) out of 7 alleles. However, these alleles are present in two of the main *L. interrogans* STs: ST52 and ST3. This could provide some evidence about the similarities of this isolate with the others belonging to *L. interrogans*, despite its location in the dendrogram. Nevertheless, we cannot discard the possibility of a misclassification of the isolate in the database.

All Argentinian isolates were distributed within *L. interrogans* cluster, and were closely related through the STs that grouped the highest number of isolates (ST47, ST52 and ST3).

The individual contribution of loci to the phylogeny was also explored by constructing ML trees for each locus separately (Fig. S2). Most dendrograms supported the phylogenetic relatedness between species that had been obtained for the concatenated sequences. Remarkably, only the *glm*U and *pnt*A dendrograms

Table 1
Parameters analyzed in a subset of 65 isolates for the selection of the optimal number of loci. Numbers underlined indicate the values of the parameters for the discarded *rrs*2 locus. CC: clonal complex.

Parameters/loci	*adk*	*glm*U	*icd*A	*lipL32*	*lipL41*	*mre*A	*pnt*A	*rrs*2
N° of alleles	28	14	21	10	26	17	16	14
N° of polymorphisms	106	38	81	31	84	44	51	18
Typing efficiency (TE)	0.264	0.368	0.259	0.322	0.309	0.386	0.313	0.777
Discriminatory power (DP)	0.833	0.765	0.744	0.654	0.941	0.866	0.756	0.735
dN/dS	0.0469	0.0531	0.012	0.055	0.052	0.007	0.0082	1.3153
goeBURST (N° CC)	5	1	3	2	5	2	0	5

Please cite this article in press as: Varni, V., et al. Reassessment of MLST schemes for *Leptospira* spp. typing worldwide. Infect. Genet. Evol. (2013), http://dx.doi.org/10.1016/j.meegid.2013.08.002

Table 2
Parameters analyzed in 116 isolates for the 7-locus scheme. CC: Clonal Complex.

Parameters/loci	*adk*	*glm*U	*icd*A	*lipL32*	*lipL41*	*mre*A	*pnt*A
N° of alleles	54	31	45	33	49	38	35
N° of polymorphisms	135	138	147	51	109	136	157
Typing efficency (TE)	0.4	0.225	0.306	0.647	0.45	0.279	0.223
Discriminatory power (DP)	0.901	0.863	0.865	0.837	0.959	0.916	0.870
dN/dS	0.0415	0.0614	0.0177	0.0644	0.0522	0.0256	0.0141
goeBURST (N° CC)	4	2	7	4	5	4	1

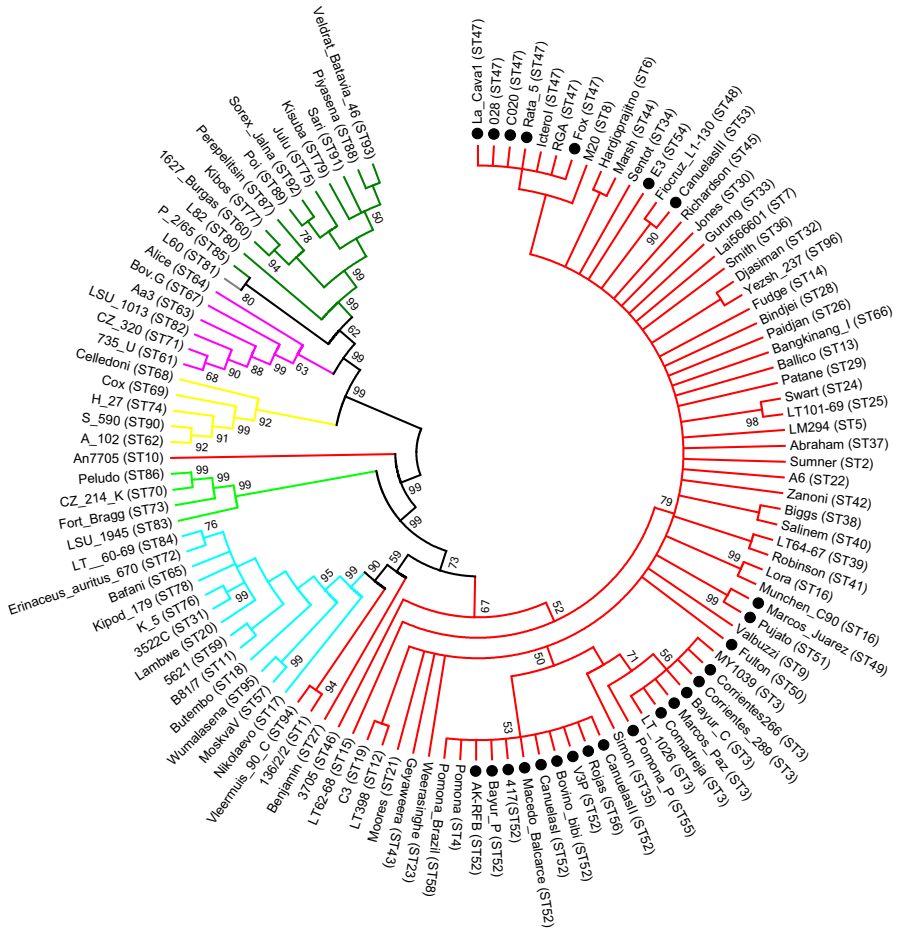

Fig. 1. Maximum-likelihood tree based on concatenated sequences of 7-locus MLST scheme for the 116 isolates. The dendrogram displays six major clusters corresponding to 6 of 7 *Leptospira* species analyzed in different colors: Red: *L. interrogans*; Light blue: *L. kirschneri*; Light green: *L. noguchi*; Yellow: *L. weilii*; Pink: *L. santarosai* and Green: *L. borgpetersenii*. The *L. alexanderi* isolate (Grey) was grouped together with the *L. borgpetersenii* cluster. Argentinian isolates are shown in black dots. STs are shown in parentheses. Phylogenetic relationships were inferred by maximum-likelihood, general time reversible model with gamma-distributed rate variation plus invariant sites, 1000 bootstrap replications. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

accurately defined seven different clusters as expected in this MLST analysis. In both dendrograms, the branch of *L. alexanderi* that had been included within *L. borgpetersenii* cluster for the concatenated sequences was now placed in a separated branch. *adk* and *icd*A displayed less reliable dendrograms regarding species distribution, by introducing some discrepancies into the phylogeny. However, the parameters studied pointed these loci as good candidates for typing because of their high variability and contribution for clonal complexes definition that strongly support their inclusion in the scheme. The trees for *lipL32* and l*ipL41* showed an unclustered species distribution, possibly because these are not housekeeping genes but they encode membrane lipoproteins that are less conserved and may provide a higher rate of variability. Furthermore, the Argentinian isolates were placed in separated branches within the *L. interrogans* cluster in the trees for each of these 2 loci. These results suggested that these loci could be useful for providing intra-specific discrimination.

Strain classification discrepancies that had been observed in the dendrogram of concatenated sequences were still unsolved in separate trees. This was the case of the isolate An7705 which was grouped in the *L. santarosai* cluster within the *adk*, *icd*A, *lipL32* and *lipL41* trees. This result was also obtained by Ahmed and colleagues after applying their 6-locus MLST scheme (Ahmed et al., 2006). In that case, the isolate had been classified as *L. santarosai*, and also clustered with this species in the phylogenetic tree. By contrast, the *glm*U, *mre*A and *pnt*A trees from this study pooled the isolate into the species *L. interrogans*, according to the results obtained by Thaipadungpanit et al., (2007) and Boonsilp et al., (2013). As mentioned above, the discrepancy around An7705 could have arisen from a misclassification of the isolate. Although it is

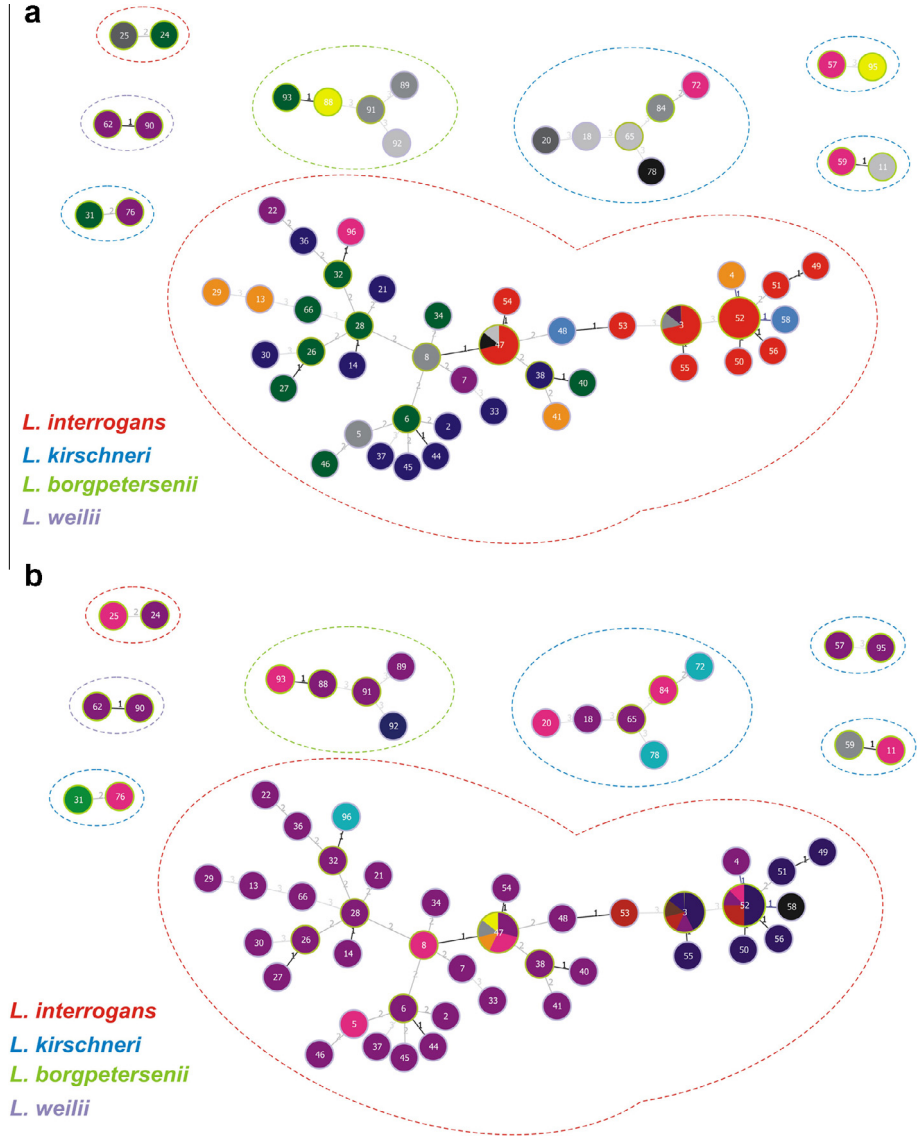

Fig. 2. *Leptospira* spp. STs relationship using goeBURST. Clonal complexes (CCs) were built based on STs linkage by TLV criteria. Singletons were excluded. The size of each circle is proportional to the number of isolates in each ST. The colored dotted lines correspond to the different *Leptospira* species in each CC follow: Red: *L. interrogans*; Light-blue: *L. kirschneri*; Green: *L. borgpetersenii* and Purple: *L. weilii*. A: Country distribution among CCs. Red: Argentina; Green: Indonesia; Blue: Malaysia; Sky-blue: Brazil; Orange: Australia; Purple: China; Violet (ST3): USA; Pink: Russia; Yellow: Sri Lanka; Light grey (ST11): Bulgaria, Light grey (ST18 and ST65): Zaire; Light grey (ST47): Belgium; Light grey (ST92): Czeckoslovakia; Grey (ST3 and ST84): Jamaica; Grey (ST5): Roumanica; Grey (ST8): Denmark; Dark grey (ST20): Kenya; Dark grey (ST25): The Phillipines; Black (ST47): Japan; Black (ST78): Israel. Host or source distribution among CCs. Blue: Bovine; Purple: humans; Brown: Weasel; Pink: Rat; Dark orange: Porcine; Light orange: Dog; Grey: Opossum; Yellow: Water. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.).

unlikely as this same isolate was classified as belonging to *L. interrogans* in other studies different from MLST schemes (Salaun et al., 2006). Probably, the underlying reason may be based on a genetic mosaicism of this particular isolate.

The usefulness of housekeeping *vs*. non-housekeeping genes in a MLST scheme has been already discussed by Pérez Losada et al., (2013). The authors stated that the proteins encoded by housekeeping genes are more reliable markers since they have a lower evolution rate, providing a robust phylogenetic signal which can allow assessment of population structure and also keep the relationship between strains. Although this observation is widely accepted, they also showed that in studies performed by Cooper and Feil (2006) on *Staphylococcus aureus* the inclusion of rapidly evolving genes did not hamper the evolutionary studies, while more standard MLST genes provided poorest phylogenetic resolution. From these results the authors proposed that loci selection, at least at the intra-species level, should be based on nucleotide diversity rather than gene function. Moreover, Maiden suggested that if higher resolution is required the inclusion of fast-evolving genes might be better than adding more MLST loci (Maiden, 2006). For all these reasons we decided to keep *lipL32* and *lipL41* in the proposed scheme.

Global eBURST analysis showed that the species are confined within the different CCs and that there is no coexistence of different species within the same CC (Fig. 2A and B). These findings reinforce the phylogenetic results regarding the distribution of the isolates into the different species. These results strongly support the robustness of the reassessed 7 loci scheme. Global eBURST established 35 singletons and 8 CCs linked by TLVs. The major CC grouped 40 of 96 STs and included 59 isolates. This CC contained ST47 and ST52, which were the founder genotypes of two different CCs in the previous 12-locus analysis. This analysis also showed connections between ST47, ST52 and ST3 involving the largest number of isolates and describing subgroups within this CC. The

ST52 grouped 8 isolates only from Argentina (Fig. 2A). These isolates were obtained from bovine (4 isolates), porcine (2 isolates), rat (1 isolate) and human (1 isolate) cases (Fig. 2B). A TLV from the ST52, the ST3, grouped 7 isolates from different sources including 5 obtained from cattle, pigs and a weasel in Argentina and 2 from USA and Jamaica, this last one being the reference strain MY 1039. Furthermore, both ST52 and ST3 are closely linked to others that also represent bovine isolates from Argentina. ST47 grouped 7 isolates, 5 from Argentina which include isolates from fox, rats, dog and water (Fig. 2A and B). The remaining ones were from Belgium and Japan and corresponded to the RGA and Ictero I reference strains from humans respectively. Unlike the previous, ST47 was linked to other STs that gathered isolates from diverse countries such as Malaysia, Brazil and Denmark, and come primarily from humans and rats. This finding is consistent with the previous reports that stated the role of rats as important carrier of leptospirosis and its source of leptospiral infection in human (Fig. 2A and B) (Li et al., 2013).

3.4. Application of allelic profile for serogroup inference

MAT allows the classification of *Leptospira* spp. isolates into serogroups, but these lack a genetic correlation as in the case of serovar that is determined by cross agglutination absorption test (CAAT) or pulsed field gel electrophoresis (Kmety and Dikken, 1993; Herrmann et al., 1992). Even when serovar is considered of taxonomic relevance, it was observed that it is a poor indicator of genetic relatedness (Boonsilp et al., 2013). Moreover, Argentinian isolates are frequently characterized up to serogroup level by MAT, but serovar is not determined. Based on this, we examined a possible correlation between allelic profiles and serogroups. We found that only 4 of 96 STs (ST3, ST16, ST47 and ST52) were represented by more than one isolate. These STs were also identified as founder genotypes of CCs. These isolates were mainly Argentinian and belonged to *L. interrogans*, turning them into ideal candidates for MLST vs. serogroup analysis. ST3 included 7 isolates of serogroups Pomona (4/7) and Canicola (3/7). ST47 grouped 7 isolates that were serotyped as Icterohaemorrhagiae (Fig. S3). Eight isolates described by ST52 were classified as serogroup Pomona, except for one that was classified as Sejroe. Finally, 2 isolates belonging to ST16 were serotyped as Australis. In order to evaluate if the isolates belonging to the serogroups mentioned above consistently referred to the same allelic profiles, we reviewed the entire collection. We found that the relationship observed between ST and serogroup was not met inversely, since the isolates of each serogroup were mostly described by many different allelic profiles. Furthermore, the analysis was extended to all the STs that appeared in CCs. We found that the SLVs from ST52 represented isolates classified as serogroup Pomona. In turn, ST47 with its SLVs, were all from Icterohaemorrhagiae serogroup. ST3 and ST55 are SLVs, and their respective isolates were serotyped as Pomona or Canicola.

These results implied that the collection of isolates exhibited congruence between allelic profile and serogroup, since a given profile (or set of allelic profiles in a clonal complex) belonged to at most, two serogroups. However, the opposite was not satisfied because the set of isolates from a given serogroup displayed a variety of STs.

Data previously reported by our group, revealed associations between genotype and serogroup in Argentinian isolates using a different MLST scheme plus VNTRs typing (Caimi et al., 2012). Furthermore, Resch et al., (2007) proposed the usefulness of genotypes from phylogenetic trees constructed with Random Amplified polymorphic DNA banding patterns in predicting not only serogroup but also serovar of an unknown *Leptospira* isolate.

Although the findings of the current study should be supported by increasing the number of isolates typed, the possibility of inferring serogroup by a genetic technique may provide an alternative to serological methods. Depending on the capabilities of each laboratory or the research objective, serological methods are not always available due to their complexity and the need to maintain live strains. The application of the allelic profile method would be useful to resolve discrepancies generated by MAT, but also in cases where the samples are not determined serologically and a quick approach to serogroup is required.

4. Conclusions

This study is an attempt to unify the existing MLST schemes, taking the advantages of each gene set by selecting the minimum number of loci and maximizing the information for an accurate molecular typing of this pathogen.

According to the results obtained in the phylogenetic studies, goeBURST algorithm and the exhaustive comparative analysis of each locus, we propose a 7 loci MLST scheme that includes the loci: *adk*, *glm*U, *icd*A, *lipL32*, *lipL41*, *mre*A and *pnt*A. This combined scheme was both successful in resolving species in a global strain collection, but also providing a higher level of intra-species discrimination. Furthermore, the correlation observed between STs and serogroups supports the implementation of MLST as a complementary approach to contribute to the classification of leptospira strains in a global scale.

Although sequences are already available in GenBank, the next step will be the implementation of this combined scheme in a public MLST database according to the requirements for MLST schemes.

Finally, the agreement in the utilization of a single scheme would allow researchers to increase the number of univocally typed strains therefore facilitating the global comparison, which constitutes the main advantage of MLST technique.

Acknowledgements

We thank Dr Rudy Hartskeerl and Dr Janjira Thaipadungpanit for kindly provide the sequences used in this study. We also want to thank Dr Patricio Diosque, Dr Andrea Gioffré and Dr Julia Sabio y García for their support, encouragement and critical reading of this manuscript, and Valeria Rocha, for her valuable technical help. This work was benefited from grants from the Fondo para la Investigación Científica y Tecnológica (FONCYT) (PICT 0117) and from the Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET) (PIP 11220090100666). V.V. holds a doctoral fellowship from the CONICET. K.C. and P.R. are CONICET fellows.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at http://dx.doi.org/10.1016/j.meegid.2013.08.002.

References

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

Ahmed, N., Devi, S.M., Valverde, M.D.L.Á., Vijayachari, P., Machang, R.S., Ellis, W.A., Hartskeerl, R.A., 2006. Multilocus sequence typing method for identification and genotypic classification of pathogenic *Leptospira* species. Ann. Clin. Microbiol. Antimicrob. 5, 28.

Ahmed, A., Thaipadungpanit, J., Boonsilp, S., Wuthiekanun, V., Nalam, K., Spratt, B., Aanensen, D., Smythe, L.D., Ahmed, N., Feil, E.J., Hartskeerl, R.A., Peacock, S.J., 2011. Comparison of 2 Multilocus Sequence Based Genotyping Schemes for *Leptospira* Species. PLoS Negl. Trop. Dis. 5, e1374.

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.

Boonsilp, S., Thaipadungpanit, J., Amornchai, P., Wuthiekanun, V., Bailey, M.S., Holden, M.T., Zhang, C., Jiang, X., Koizumi, N., Taylor, K., Galloway, R., Hoffmaster, A.R., Craig, S., Smythe, L.D., Hartskeerl, R.A., Day, N.P., Chantratita, N., Feil, E.J., Aanensen, D.M., Spratt, B.G., Peacock, S.J., 2013. A Single multilocus sequence typing (MLST) scheme for seven pathogenic leptospira species. PLoS Negl. Trop. Dis. 7, e1954.

Caimi, K., Varni, V., Meléndez, Y., Koval, A., Brihuega, B., Ruybal, P., 2012. A combined approach of VNTR and MLST analysis: improving molecular typing of Argentinean isolates of *Leptospira interrogans*. Mem. Inst. Oswaldo Cruz 107, 644–651.

Cerqueira, G.M., McBride, A.J., Hartskeerl, R.A., Ahmed, N., Dellagostin, O.A., Eslabão, M.R., Nascimento, A.L., 2010. Bioinformatics describes novel loci for high resolution discrimination of *Leptospira* isolates. PLoS One 5, e15335.

Cooper, J.E., Feil, E.J., 2006. The phylogeny of *Staphylococcus aureus*- which genes make the best intra-species markers? Microbiology 152, 1297–1305.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. JModelTest 2: more models, new heuristics and parallel computing. Nat Methods 9, 772.

Francisco, A.P., Bugalho, M., Ramirez, M., Carrico, J.A., 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroidapproach. BMC Bioinformatics 10, 152.

Feil, E.J., Li, B.C., Aanensen, D.M., Hanage, W.P., Spratt, B.G., 2004. EBURST: inferring patterns of evolutionary descent among clusters of re lated bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 186, 1518–1530.

Guindon, S., Gascuel, O., 2003. A simple, fast and accurate method to estimate large phylogenies by maximum-likelihood. Syst. Biol. 52, 696–704.

Herrmann, J.L., Bellenger, E., Perolat, P., Baranton, G., Saint Girons, I., 1992. Pulsed-field gel electrophoresis of NotI digests of leptospiral DNA: a new rapid method of serovar identification. J. Clin. Microbiol. 30, 1696–1702.

Jolley, K.A., Feil, E.J., Chan, M.S., Maiden, M.C., 2001. Sequence type analysis and recombinational tests (START). Bioinformatics 17, 1230–1231.

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, P.N., 2001. Leptospirosis. Clin. Microbiol. Rev 14, 296–326.

Li, S., Wang, D., Zhang, C., Wei, X., Tian, K., Li, X., Nie, Y., Liu, Y., Yao, G., Zhou, J., Tang, G., Jiang, X., Yan, J., 2013. Source tracking of human leptospirosis: serotyping and genotyping of Leptospira isolated from rodents in the epidemic area of Guizhou province China. BMC Microbiol. 13, 75.

Maiden, M.C., 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.

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, S.M., Francalacci, P., Baig, M., Sechi, L.A., Hartskeerl, R.A., Ahmed, N., 2010. Genetic affinities within a large global collection of pathogenic leptospira: implicationsfor strain identification and molecular epidemiology. PLoS One 5, e12637.

Pérez- Losada, M., Cabezas, P., Castro-Nallar, E., Crandall, K.A., 2013. Pathogen typing in the genomics era: MLST and the future of molecular typing. Infect. Gen. Evol. 16, 38–53.

Resch, G., Awad-Masalmeh, M., Bakoss, P., Jarekova, J., 2007. Utility of phylogenetic studies in the identification of Leptospira strains. Epidemiol. Infect. 135, 1266–1273.

Salaün, L., Mérien, F., Gurianova, S., Baranton, G., Picardeau, M., 2006. Application of multilocus variable-number tandem-repeat análisis for molecular typing of the agent of leptospirosis. J. Clin. Microbiol. 44, 3954–3962.

Segura, E.R., Ganoza, C.A., Campos, K., Ricaldi, J.N., Torres, S., Céspedes, M.J., Matthias, M.A., Swancutt, M.A., López Liñán, R., Gotuzzo, E., Guerra, H., Gilman, R.H., Vinetz, J.M., 2005. Clinical spectrum of pulmonary involvement in leptospirosis in a region of endemicity, with quantification of leptospiral burden. Clin. Infect. Dis. 40, 343–351.

Slack, A., Symonds, M., Dohnt, M., Smythe, L., 2006. 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 J. Med. Microbiol. 55, 1549–1557.

Smythe, L., Adler, B., Hartskeerl, R.H., Galloway, R.L., Turenne, C.Y., Levett, P.N.International Committee on Systematics of Prokaryotes Subcommittee on the Taxonomy of Leptospiraceae, 2012. Classification of genomospecies 1, 3, 4 and 5 as *Leptospira alstonii* sp nov., *Leptospira vanthielii* sp. nov., *Leptospira terpstrae* sp. nov., and *Leptospira yanagawae* sp. nov., respectively. Int. J. Syst. Evol. Microbiol. 63, 1859–1862, the.

Staden, R., 1996. The staden sequence analysis package. Mol. Biotechnol. 5, 233–241.

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, L.D., Petkanchanapong, W., Limpaiboon, R., Apiwatanaporn, A., Slack, A.T., Suputtamongkol, Y., White, N.J., Feil, E.J., Day, N.P., Peacock, S.J., 2007. A dominant clone of *Leptospira interrogans* associated with an outbreak of human leptospirosis in Thailand. PLoS Negl. Trop. Dis. 1, e56.

Vanasco, N.B., Schmeling, M.F., Lottersberg, J., Costa, F., Ko, A.I., Tarabla, H.D., 2008. Clinical characteristics and risk factors of humans leptospirosis in Argentina (1999–2005). Acta Trop. 107, 255–258.

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.

World Health Organization, 1999. Leptospirosis worldwide, 1999. Wkly. Epidemiol. Rec. 74, 237–244.

Please cite this article in press as: Varni, V., et al. Reassessment of MLST schemes for *Leptospira* spp. typing worldwide. Infect. Genet. Evol. (2013), http://dx.doi.org/10.1016/j.meegid.2013.08.002