Infection, Genetics and Evolution 54 (2017) 478–485

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Leptospira species molecular epidemiology in the genomic era

K. Caimi[a], S.A. Repetto[b], V. Varni[a], P. Ruybal[b,*]

[a] *Instituto Nacional de Tecnología Agropecuaria (INTA), Instituto de Biotecnología, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Buenos Aires, Argentina*

[b] *Universidad de Buenos Aires, Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET), Instituto de Investigaciones en Microbiología y Parasitología Médica (IMPaM), Facultad de Medicina, Buenos Aires, Argentina*

ARTICLE INFO

Keywords:
Leptospira spp.
MLST
Genomics
Evolution

ABSTRACT

Leptospirosis is a zoonotic disease which global burden is increasing often related to climatic change. Hundreds of whole genome sequences from worldwide isolates of *Leptospira* spp. are available nowadays, together with online tools that permit to assign MLST sequence types (STs) directly from raw sequence data. In this work we have applied R7L-MLST to near 500 genomes and strains collection globally distributed. All 10 pathogenic species as well as intermediate were typed using this MLST scheme. The correlation observed between STs and serogroups in our previous work, is still satisfied with this higher dataset sustaining the implementation of MLST to assist serological classification as a complementary approach. Bayesian phylogenetic analysis of concatenated sequences from R7-MLST loci allowed us to resolve taxonomic inconsistencies but also showed that events such as recombination, gene conversion or lateral gene transfer played an important role in the evolution of *Leptospira* genus. Whole genome sequencing allows us to contribute with suitable epidemiologic information useful to apply in the design of control strategies and also in diagnostic methods for this illness.

1. Introduction

Leptospirosis is a zoonotic disease caused by pathogenic species of the genus Leptospira. Transmission to humans occurs via direct contact with infected animals or via contaminated water with animal urine (Bharti et al., 2003). Annual worldwide case number was estimated at around 1 million with the majority of cases and death occurring in tropical regions (Costa et al., 2015). The highest disease burden was reported in tropical low and middle income countries, driven by climatic conditions, close human-animal contact, inadequate sewage disposal and water treatment (Levett, 2001). Epidemics in humans and animals are increasing and are often related to natural events like floods (Bandara et al., 2014).

The gold standards for laboratory diagnosis of leptospirosis are culture or a four-fold rise in antibody titre between admission and convalescent samples by the microscopic agglutination test (MAT) (Levett, 2001). Culture of *Leptospira* spp. is time consuming and MAT allows serogroup but no species identificaction. Even though MAT is the method of reference for *Leptospira* typing, the large number of serological variants characterized by 300 different antigenic types, produces a high diversity of strains making extremely difficult to track isolates through this serological approach.

Methods based on genomic DNA homology has been available and earlier phylogenetic analyses permits to classify the genus into 3 distinct lineages that include ten pathogenic species: *Leptospira interrogans*, *L. borgpetersenii*, *L. santarosai*, *L. noguchii*, *L. weilii*, *L. kirschneri*, *L. alexanderi*, *L. alstonii*, *L. kmetyi and L. mayottensis*; five intermediate species: *L. inadai*, *L. broomii*, *L. fainei*, *L. wolffii*, *L. licerasiae* and six non-infectious saprophytic species: *L. biflexa*, *L. wolbachii*, *L. vanthielii*, *L. terpstrae*, *L. meyeri*, *L. idonii*, *and L. yanagawae* (Lehmann et al., 2014). Whole genome sequencing (WGS) of some *Leptospira* species were primary released, including *L. interrogans*, *L. borgpetersenii*, *L. santarosai*, *L. licerasiae* and *L. biflexa* (Nascimento et al., 2004; Ren et al., 2003; Bulach et al., 2006; Chou et al., 2012; Ricaldi et al., 2012; Picardeau et al., 2008). Genomic comparisons indicate that while L. *biflexa* genome is relatively stable, the genomes of pathogenic species have undergone considerable insertion sequences mediated rearrangements revealing a high-level of plasticity of *Leptospira* genomes (Ricaldi et al., 2012). Recently, two different WGS *Leptospira* projects have been carried out, where the main goal was to obtain and compare genome information for all known *Leptospira* species and main serovars, providing together data of very fine-scale resolution of near 500 different *Leptospira* genomes (Fouts et al., 2016; Xu et al., 2016).

The availability of genome sequences allows the development of

* Corresponding author at: Instituto de Investigaciones en Microbiología y Parasitología Médica, Facultad de Medicina, Universidad de Buenos Aires, Paraguay 2155 Piso 13, 1121, Cdad. Autónoma de Buenos Aires, Argentina.
E-mail address: pruybal@gmail.com (P. Ruybal).

http://dx.doi.org/10.1016/j.meegid.2017.08.013
Received 15 May 2017; Received in revised form 26 July 2017; Accepted 13 August 2017
Available online 15 August 2017

different molecular typing techniques, being Multilocus Sequence Typing (MLST) the first sequence-based approach for strain resolution in many bacterial species (Maiden et al., 1998). MLST has already established promise in unraveling *Leptospira* strains typing and phylogeny, although in studies of limited species or strain panels or even in strains with restricted geographic prevalence (Ahmed et al., 2006; Thaipadungpanit et al., 2007; Romero et al., 2011; Boonsilp et al., 2013; Varni et al., 2014).

Currently, WGS has become widely feasible together with automated tools that permit to assign MLST sequence types (STs) directly from raw sequence data. Therefore, in this study we applied our reassessed 7 loci MLST (R7L-MLST) scheme (Varni et al., 2014) to an extended group of strains of worldwide distribution which whole genome sequences are now accessible, in the attempt to generate ready available typing profiles and a deeper phylogenetic and epidemiological scenario of *Leptospira* genus.

2. Materials and methods

2.1. Bacterial strains and genome sequences

A total of 436 *Leptospira* genomic sequences were downloaded from the NCBI database (https://www.ncbi.nlm.nih.gov). The sequences were generated by the "Leptospira Genomics and Human Health" genomic project from the John Craig Venter Institute (JCVI). These genomes corresponded to all known species of the genus including pathogenic, intermediate, saprophytic and undetermined pathogenicity *Leptospira* isolates and reference strains isolated worldwide. The collection of 116 strains characterized in our previous study was also included in this work (Varni et al., 2014) together with new sequences (11 isolates) added recently to the public MLST database (http://pubmlst.org/leptospira/) hosted at the Department of Zoology from the University of Oxford (Jolley and Maiden, 2010). Data associated with each strain used in this study are summarized in Supplementary Table 1.

2.2. Selection of genome sequences

The whole genome sequences used in this study were analyzed by the MLST 1.8 web application (http://cge.cbs.dtu.dk/services/MLST/) (Larsen et al., 2012), selecting *Leptospira* spp. #2 configuration which correspond to our *Leptospira* seven loci MLST (R7L-MLST) scheme database (http://pubmlst.or/leptospira/). Genomic sequences from intermediate species that failed to align the complete set of alleles were analyzed using a Protein Basic Local Alignment Search Tool (https://blast.ncbi.nlm.nih.gov/, blastp) in order to complete the scheme. Each allele sequence derived from MLST 1.8 hit in the genome, were included in our published sequences data (116 isolates) (Varni et al., 2014).

2.3. MLST scheme and sequence analysis

Our previous work proposed a seven loci MLST scheme (R7L-MLST) that includes locus: *adk*, *glm*U, *icd*A, *lip*L32, *lip*L41, *mre*A and *pnt*A (Varni et al., 2014). Sequence types (STs) designation, typing efficiency and discriminatory power estimation were performed by MLSTest software (http://mlstest.codeplex.com/) (Tomasini et al., 2013). All sequences are available in pubmlst database (http://pubmlst.or/leptospira/).

The ratio of non-synonymous (dN) to synonymous (dS) substitutions per nucleotide site (Nei and Gojobori method) was determined by DNAsp v.5 software (Librado and Rozas, 2009). Sequence alignment was performed using MEGA version 6 (Tamura et al., 2013). GoeBURST algorithm (http://www.phyloviz.net/goeburst/) was applied to determine the relationships between STs (Francisco et al., 2009; Feil et al., 2004). Clonal complexes (CCs) were defined as STs linked by triple locus variants (TLVs) criteria.

Phylogenetic relationships among concatenated sequences of loci were inferred by maximum likelihood (ML) tested with 500 bootstrap replications in MEGA 6 (Tamura et al., 2013). The selection of the nucleotide substitution model was performed through JModelTest 2 software (Darriba et al., 2012; Guindon and Gascuel, 2003). The tree figures were edited by FigTree v. 1.4.3 (http://tree.bio.ed.ac.uk/software/figtree/).

Congruence analysis between concatenated tree branches was performed by MLSTest software. Overall incongruence was evaluated by Incongruence Length Difference test implemented by BIO-Neighbor Joining method (ILD-BIONJ) with 100 permutations (Tomasini et al., 2013). Localized incongruence was evaluated by the number of topologically discordant gene trees respect to each branch in the concatenated tree. For this purpose, topological incongruence (TI) analysis was considered as high when > 40% of branches with *n*-1 loci were incompatible with the cluster in the concatenated tree (Tomasini et al., 2014).

The *phi* test for recombination was performed with SplitsTree v.4.12.6 (Bruen et al., 2006), and P values < 0.05 was considered as positive recombination. Standardized index of association (I_A^S) was calculated with Linkage Analysis v3.6 (Haubold and Hudson, 2000) with 100,000 iterations by Monte Carlo based on allelic profiles. Horizontal gene transfer (HGT) events were investigated by applying multiple recombination algorithms: Bootscan/Recscan (Martin et al., 2005), Chimaera (Posada and Crandall, 2001), GENECONV (Padidam et al., 1999), MaxChi (Maynard Smith, 1992), RDP (Martin and Rybicki, 2000), and SiScan (Gibbs et al., 2000), as implemented in the RDP4 package (Martin et al., 2015). The occurrence of a potential HGT event was accepted only if: both major and minor sequences parents identified, no possible misinterpretation of recombination was informed, at least three distinct validation methods were reported and were sustained by strong statistical support.

Bayesian phylogenetic trees were reconstructed with BEAST v1.8.1 software (Drummond et al., 2012). The model of evolution for each gene was determined using the jModelTest 2 program. The Bayesian analyses were performed using a Yule process of speciation, and a strict-clock model with the clock rate set to 1 as the tree priors, as well as other default parameters. We performed a Markov Chain Monte Carlo (MCMC) run of 200 (species tree) million generations, sampling every 20,000 generations. Posterior distributions for parameter estimates and likelihood scores to approximate convergence were visualized with the Tracer v1.6.0 program (Rambaut et al., 2014). Effective sample sizes (EES) values ≥ 200 confirmed that the analyses were adequately sampled. A maximum clade credibility (MCC) tree was chosen by TreeAnnotator v1.8.1 (Drummond et al., 2012) and visualized with the program FigTree. Distribution of species trees was evaluated by DensiTree 2.0 (Bouckaert, 2010).

All bioinformatic analysis was performed in a local server at Instituto de Investigaciones en Microbiología y Parasitología Médicas (IMPaM) which is part of National System of High Performance Computing (SNCAD) of Ministry of Science, Technology and Productive Innovation (MINCyT).

2.4. Statistical analysis

Both descriptive and inferential statistical analyses were performed using SPSS software package for Windows (version 21). Statistical associations among categorical variables were analyzed using the Fisher's test and presented as observed frequencies and proportions. The probability of finding the outcome of interest was calculated as odds ratio (OR). A P value lower than 0.05 was considered statistically significant for the inferential tests.

3. Results and discussion

The availability of new *Leptospira* whole genome sequences from different worldwide isolates makes possible the in silico analysis of MLST. After the analysis of 436 genome sequences for alleles determination, 411 were finally included in this work. Twenty-five genomes were excluded based on the lack of alignment in at least one of the alleles according to MLST 1.8 algorithm or blastp analysis. The remaining strains correspond to our previous collection of 116. Therefore, in this work we have applied the R7L-MLST scheme to a total of 527 strains.

Despite R7L-MLST scheme was arranged from two MLST schemes that were originally developed upon the premise of detection and typing just *Leptospira* pathogenic species, not all the pathogenic species were typed in those studies due to a very low rate of isolation at that moment. We have now included some strains belonging to those species, *L. alstoni* and *L. kmetyi*, taking advantage of their genomes availability. At the same time, some locus could be present also in intermediate species genomes enabling their inclusion in the analysis. Eight strains belonging to the five intermediate species (*L. licerasiae*, *L. wolffii*, *L. broomii*, *L. inadai* and *L. fanei*), showed a complete hit in the genome sequences for all loci. These results reinforce the recent similarities observed in the gene content between pathogenic and intermediate species (Fouts et al., 2016; Xu et al., 2016) and hence the pathogenic features of these species (Arzouni et al., 2002; Lehmann et al., 2014). Saprophytic strains were excluded from the analysis due to the low identity percentage obtained.

The final species distribution was as follows: 330 (62%) *L. interrogans*, 49 (9%) *L. borgpetersenii*, 46 (8%) *L. santarosai*, 40 (7%) *L. kirschneri*, 22 (4%) *L. weilii*, 19 (3%) *L. noguchi*, 5 (0.9%) *L. alstoni*, 3 (0.5%) *L. licerasiae*, 2 (0.3%) *L. inadai*, 2 (0.3%) *L. mayotensis*, 1 (0.1%) *L. alexanderi*, 1 (0.1%) *L. kmetyi*, 1 (0.1%) *L. fanei*, 1 (0.1%) *L. broomii*, 1 (0.1%) *L. wolffii* and 4 (0.7%) strains of undetermined species.

Our analysis identified a total of 271 sequence types (STs) among the 527 *Leptospira* spp. isolates. The number of alleles for each locus ranged from 67 at *lip*L32 to 111 at *adk* (Table 1). Absolute frequency analysis showed that of the 271 unique STs, 220 STs were represented by single isolates (81,18%), 50 STs were represented by two to 17 isolates (18,45%) and the ST-47 (0,37%) was the most frequent variant represented by 112 *Leptospira interrogans* isolates.

The most frequent variant among the collection (ST-47) was widely distributed throughout all continents and is present in four hosts (domestic dog, fox, rodent and human) and water but not in livestock. The second most frequent ST (ST-3, 17 isolates) was also widely distributed but absent in Africa and Oceania and was isolated from domestic dog, rodent, human, weasel and from livestock (bovine and porcine). The ST-52 (15 isolates) was absent in Asia and Europe and isolated from domestic dog, human, rodent, fox, livestock (bovine, porcine) and from marine mammals. All these STs were assigned to *L. interrogans* isolates which corresponds to the most representative species among the studied collection. Also, these three STs shared the Americas as continent of isolation and humans, rodents and domestic dogs as hosts (Supplementary Table 1), reinforcing with these results our previous observation with a minor set of strains (Varni et al., 2014).

As mentioned above eight isolates corresponding to five pathogenic intermediate species were included in this study. Each of these species displayed a different and unique ST.

As shown in Table 1, the addition of 411 isolates to R7L-MLST scheme analysis showed a significant increase in the number of alleles and polymorphisms per locus which were under negative selective pressure (dN/dS < 1). The correlation between the increase in the number of isolates and alleles led to similar values of typing efficiency (TE, 0.367) and the discriminatory power (DP, 0.951) parameters when compared to those obtained in our previous work (Varni et al., 2014). This reinforces the high intra-species DP of R7L-MLST scheme since the 62.69% of the isolates belonged to *L. interrogans*. In other words, high values of DP were obtained for the six most represented species in the population: *L. weilii* (DP:1); *L. santarosai* (DP:0.994); *L. kirschneri* (DP:0.976); *L. borgpetersenii* (DP:0.932); *L. interrogans* (DP:0.878) and *L. noguchii* (DP:0.877).

Global eBURST analysis of the 271 unique STs defined 20 clonal complexes (CCs) when linked by up to triple locus variant level (TLV) and 74 singletons (Fig. 1, Supplementary Table 1). There was no co-existence of different species within each CC except for the CC-16 composed by strains *L. santarosai* CBC613 (ST-166) and *L. interrogans* ZV013 (ST-248). These two strains were highly related variants (Single Locus Variants, SLVs) and based on their 16S ribosomal DNA sequences analysis with Ribosomal Database Project (RDB, http://rdp.cme.msu.edu/), both strains showed high similarity just with the species *L. kirschneri*. Therefore, CC-16 also displayed STs groups from unique species.

A singleton represents an allelic profile that varied from N-1 STs in four to seven loci of the R7L-MLST scheme. Hence, given a specific population, the genetic diversity could be assessed by the number of singletons in the set of STs by the SiSt ratio (No. singletons/No. STs). Distribution of different species consistent with genetic relatedness between STs and the subsequent organization in CCs and singletons highlighted the high diversity within species *L. noguchii* and *L. santarosai* (Fig. 1). The SiSt ratio in both cases was 0.67 and 0.68, respectively. In contrast, singletons were less frequent in *L. interrogans* and *L. kirschneri* species (SiSt: 0.08 and 0.03, respectively) and a moderate SiSt ratio was described for *L. weilii* and *L. borgpetersenii* (0.36 and 0.35, respectively) (Table 2). Most variable species, *L. noguchii* and *L. santarosai*, included strains isolated from eight different host groups. *L. noguchii* was associated with North America (OR = 9,85; CI 95% = 3,82–25,36; $P < 0,01$) and *L. santarosai* with South America (OR = 2,75; CI 95% = 1,49–5,06; $P < 0,01$) and North America (OR = 6,47; CI 95% = 3,31–12,65; $P < 0,01$). Less variable species, *L. interrogans* and *L. kirschneri*, were isolated from 15 and 8 host groups, respectively. Furthermore, *L. interrogans* was associated with Asia (OR = 1,58; CI 95% = 1,08–2,31; $P < 0,01$) and South America (OR = 2,05; CI 95% = 1,37–3,07; $P < 0,01$) and L. *kirschneri* with Europe (OR = 4,28; CI 95% = 1,93–9,48; $P < 0,01$) and Africa (OR = 4,19; CI 95% = 1,76–9,97; $P < 0,01$). Finally, species with moderate SiSt ratio, *L. borgpetersenii* and *L. weilii*, included strains isolated from eight and three host groups, respectively. *L. borgpetersenii* was associated with Europe (OR = 3,11; CI 95% = 1,44–6,75; $P < 0,01$) and Africa (OR = 5,19; CI 95% = 2,37–11,35; $P < 0,01$) and *L. weilii* with Asia (OR = 8,43; CI 95% = 2,81–25,30; $P < 0,01$). These data suggest that species organization in CCs and singletons with

Table 1
Parameters analyzed in 527 isolates for the R7L-MLST scheme. ND: Not determined.

Parameters/loci	ST	*adk*	glmU	*icd*A	*lip*L32	*lip*L41	*mre*A	*pnt*A
No. of alleles	271	111	78	98	67	92	80	86
No. of polymorphisms	1668	227	243	233	294	236	202	233
Typing efficiency (TE)	0,366	0,485	0,321	0,421	0,228	0,390	0,396	0,369
Discriminatory power (DP)	0,951	0,876	0,827	0,861	0,768	0,874	0,912	0,832
dN/dS	ND	0,0530	0,0945	0,0245	0,1756	0,0657	0,0407	0,0245

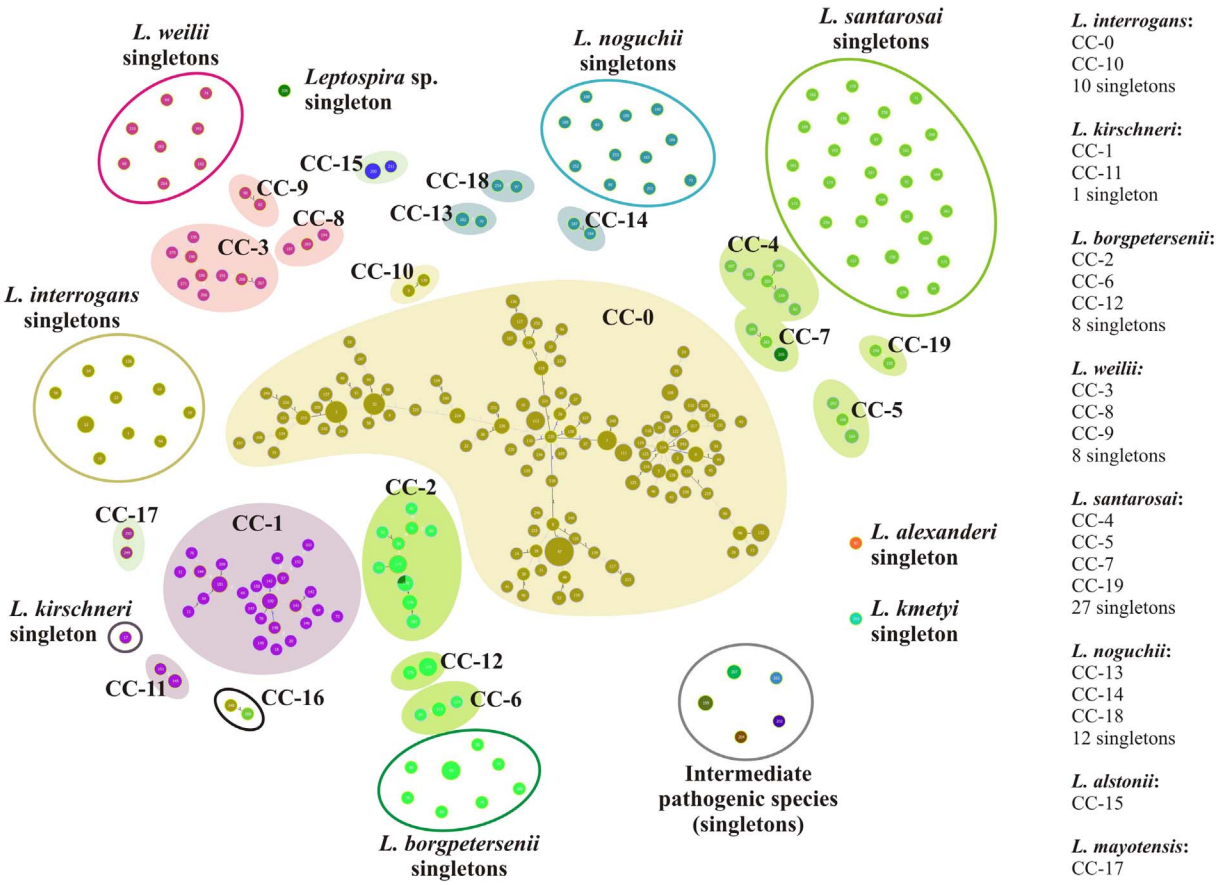

Fig. 1. *Leptospira* STs organization into clonal complexes (CC) and singletons. Each CC displayed groups of STs from unique species. CC-16 was composed by strains previously assigned as *L. santarosai* CBC613 (ST-166) and *L. interrogans* ZV013 (ST-248). Both 16S ribosomal DNA sequences were analyzed with Ribosomal Database Project (RDB, http://rdp.cme.msu.edu/) and reassigned to *L. kirschneri*.

Table 2
Leptospira species distribution according to the number of isolates, STs, CCs and singletons. SiSt ratio was calculated for each species. Number of isolates was specified according to the new taxonomic classification based on the present study results. ND: Not determined.

Species	No. isolates	No. STs	No. CCs	No. singletons	SiSt rate
L. alexanderi	1	1	0	1	1
L. alstonii	5	2	1	0	ND
L. borgpetersenii	49	23	3	8	0.35
L. interrogans	327	122	2	10	0.08
L. kirschneri	44	33	3	1	0.03
L. kmetyi	1	1	0	1	1
L. mayotensis	2	2	1	0	ND
L. noguchii	19	18	3	12	0.67
L. santarosai	45	40	4	27	0.68
L. weilii	22	22	3	8	0.36
L. licerasiae	3	1	0	1	1
L. wolffii	1	1	0	1	1
L. fainei	1	1	0	1	1
L. broomii	1	1	0	1	1
L. inadai	2	1	0	1	1
ND	4	2	0	1	1
Total	527	271	20	74	–

subsequent variations in genetic variability (SiSt values) was not associated neither to the geographic distribution or the host group dissemination of the isolates.

A taxonomic criterion considered to classify the genus *Leptospira* is the serovar categorization by cross agglutination absorption test (CAAT) or pulsed field gel electrophoresis (Kmety and Dikken, 1993; Herrmann et al., 1992). In addition, serogroup determination using Microagglutination Test (MAT) is the most widely employed method for serological characterization of isolates but lacks taxonomic value (Levett, 2001). Our previous work proposed a possible correlation between STs and serogroup organization with the goal of introducing a debate about the intention of moving forward to the molecular typing of the genus (Varni et al., 2014; Goarant, 2014). Previously, only four STs (ST-47, ST-3, ST-16 and ST-52) comprised two or more isolates associated to analyse this correlation. However, with the wider and more diverse population included in the present study, 51 allelic profiles (18.82%) comprised two or more isolates enabling an improved follow up of this correlation. Twenty STs (7.38%) were composed of two to seven isolates with unique serogroups while within 16 variants (5.9%) coexisted isolates with one serogroup population together with isolates with unresolved MAT characterization (Supplementary Table 2). Particularly, while ST-47, ST-3 and ST-52 significantly increased their isolate number, the serogroup distribution remained stable. Statistical analysis showed a significant association between ST-47 and serogroup icterohaemorrhagiae (OR = 85,58; CI 95% = 44,52–164,49; $P < 0,01$), ST-3 with serogroups canicola (OR = 31,49; CI 95% = 10,58–93,71; $P < 0,01$) and pomona (OR = 10,58; CI 95% = 3,61–30,93; $P < 0,01$) and ST-52 with serogroup pomona (OR = 67,65; CI 95% = 19,80–231,12; $P < 0,01$). Furthermore, six sub-CCs (CCs linked only up to SLVs) and six CCs (linked up to TLVs) displayed unique serogroups (Supplementary Fig. 1).

On the other hand, setting up as starting point of the analysis the serogroups in the population, there is no correlation between this variable and STs, continent of isolation and hosts group. This reinforces our previous observation that given an allelic profile population (or a set of STs in a CC) same correlation could be describe to assist

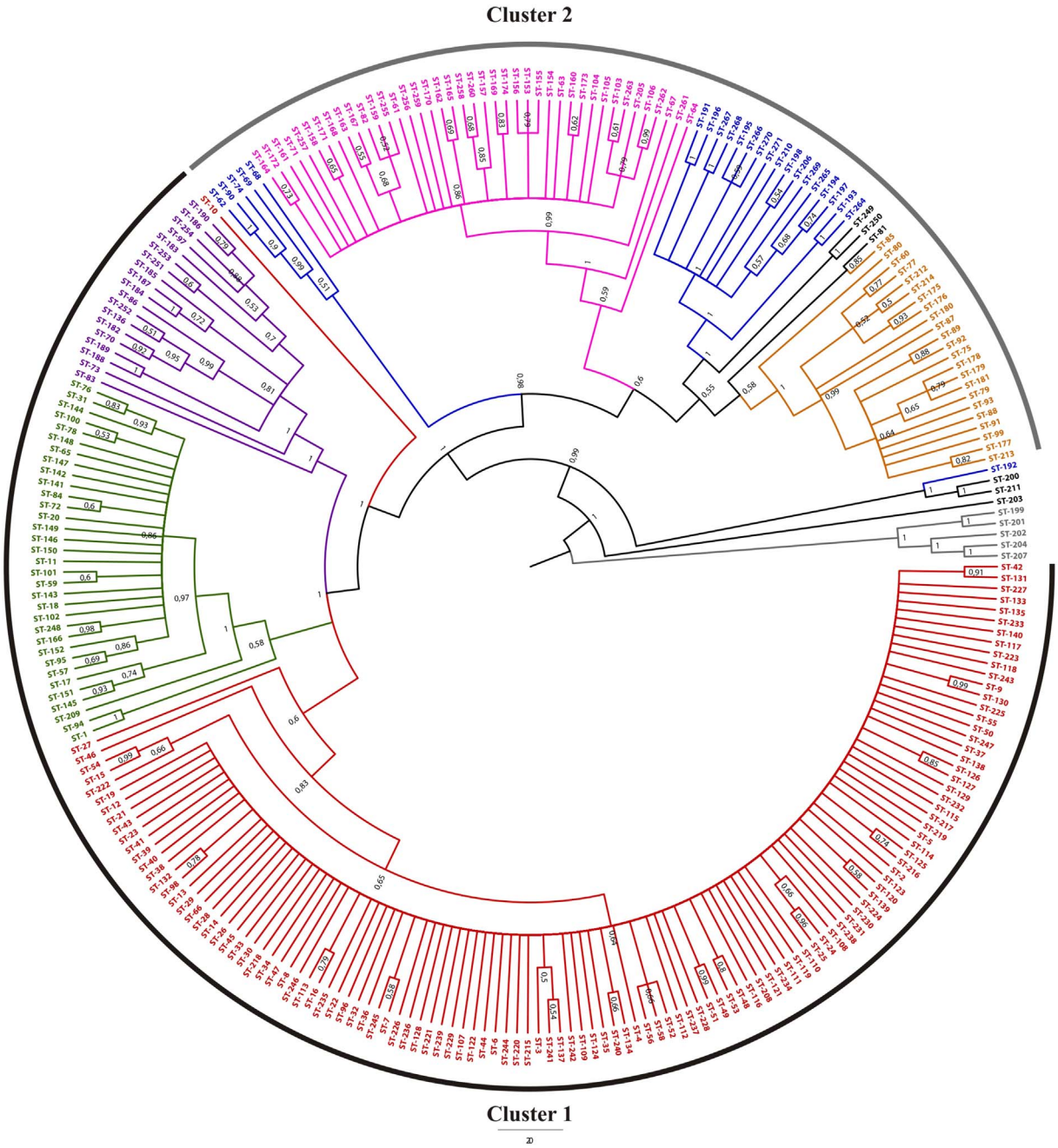

Fig. 2. Dendrogram topology based on concatenated sequences of R7L MLST scheme for the 271 STs. Phylogenetic relationships were inferred by maximum-likelihood method, Tamura-Nei model with gamma distribution and invariant sites and 500 bootstrap replications. The dendrogram topology displays two major clusters. Cluster 1 included *L. interrogans* (red), *L. kirschneri* (green) and *L. noguchii* (purple) and Cluster 2 *L. borgpetersenii* (orange), *L. santarosai* (pink) and *L. weilii* (blue). Both *L. alexanderi* (ST-81) and *L. mayotensis* (ST-249 and ST-250) were also present in cluster 2. Intermediate species (grey) were grouped as a root cluster (ST-199, ST-201, ST-202, ST-204, ST-207) while L. *kmetyi* and *L alstoni* strains showed to be ancestor species of clusters 1 and 2. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

serological classification but the opposite was not satisfied because the set of isolates from a given serogroup displayed a variety of STs.

Phylogenetic analysis was performed based on concatenated sequences constructed from each unique ST generated from the of the R7L-MLST scheme. As showed in Fig. 2, the six most represented *Leptospira* species were distributed into two major clusters as follows: *L. interrogans*, *L. kirschneri*, and *L. noguchii* (Cluster 1) and *L. borgpetersenii*, *L. santarosai*, and *L. weilii* (Cluster 2). Both clusters further subdivided in six subclusters that matched species assignments with few exceptions. Species misclassification of ST-166 and ST-248 that was previously observed by goeBURST analysis was confirmed in current phylogenetic analysis since both STs were grouped within *L. kirschneri* subcluster.

The ST-136 (only present in *L. interrogans* strain HAI1536) was a singleton in the goeBURST analysis avoiding any possible misclassification. Nonetheless, this *L. interrogans* strain was grouped within *L. noguchii* subcluster, classification that was confirmed by and 16S ribosomal DNA sequences analysis with RDB.

Four strains, Sh9 (ST-178), P2653 (ST-206), Fiocruz LV4135 and Fiocruz LV3954 (both ST-205), were listed as *Leptospira* spp. (undetermined species). The strain Sh9 described the same ST as three *L. borgpetersenii* strains (TE0159, 56142, 56214). Strain P2653 was grouped with *L. weilii* and both Fiocruz LV4135 and LV3954 within *L. santarosai* subcluster. All these species designations were confirmed by 16S ribosomal DNA sequences analysis with RDB.

Two other taxonomic inconsistencies were found in the dendrogram topology. The strains P2/65 (ST-85) and ICFT (ST-192) which are listed as *L. borgpetersenii* and *L. weilii*, respectively, did not match the species/cluster correlation. Unfortunately, the 16S ribosomal DNA sequences analysis could not be performed due to the lack of sequences both from the strain or the RDB. Then, the taxonomy at the species level of these two strains should be reviewed.

The advent of the genomic era made possible the access to large amount of data improving dataset size to test new hypotheses. Concatenated tree analysis is a common procedure in bacterial MLST studies. In our study, the phylogenetic analysis of concatenated sequences from R7-MLST loci allowed us to resolve taxonomic inconsistencies but also, led us to uncover common events in evolution of prokaryotes.

Even when taxonomy conflicts were resolved (Supplementary Table 3) maximum-likelihood tree showed that only three species form monophyletic subclusters (*L. kirschneri*, *L. noguchii* and *L. santarosai*) and low bootstrap support values were obtained in terminal branches (high level of polytomy or multifurcation) despite high number of polymorphic sites within each locus (Table 1). This could indicate a low genetic structure of the population based on high incongruence among gene trees. Major causes of gene trees incongruence are genetic drift, gene duplication (GD), or horizontal gene transfer (HGT) (Jeffroy et al., 2006). Recently, HGT and GD were described as responsible for significant gene gain throughout *Leptospira* genus, particularly for pathogenic species where both events allowed the acquisition of new virulence and host adaptation genes (Xu et al., 2016). Even though positive selection genes are not included in MLST schemes, these genetic arrangements comprise any kind of genes such as outer membrane or leptospiral immunoglobulin-like protein genes (Ralph and McClelland, 1994; Haake et al., 2004; McBride et al., 2009), that as lipoproteins in our scheme, could account for the observed trees incongruence.

Despite the concordance between branches in individual loci observed in our previous work, high levels of localized incongruence for each branch in the concatenated tree built from this new dataset was observed (88% of branches with *n*-1 gene fragments topologically incompatible with the cluster in the concatenated tree). Higher number of incongruent loci was observed in deeper nodes of the concatenated tree. By contrast, at the tips of the branches much greater congruence between individual loci and the concatenated trees was observed (disappearing tree phenomenon) (Thiergart et al., 2014). On the other hand, bootstrap support values of deeper branches were high (Fig. 2). As proposed by Tomasini et al. (2013), observed incongruence could be due to either a random distribution of contradictory phylogenetic signals among loci (random homoplasy) or concentrated inconsistencies in certain loci (incongruent loci). Overall incongruence evaluated by Incongruence Length Difference (ILD) test implemented by BIO-Neighbor Joining method (ILD-BIONJ) showed that the observed concatenated inconsistencies were due to incongruent loci but not to a stochastic distribution of contradictory phylogenetic signals ($P < 0{,}01$, 100 permutations).

Contradictory phylogenetic signals exposed by incongruence analysis were confirmed by frequent recombination in the whole population (271 STs) as well as in the six most representative pathogenic species included in this study (Table 3). Several approaches were considered in order to analyse the genetic flow within the collection. First, ST overlapping between different species was absent but eleven loci (*adk*-15, *adk*-29, *glm*U-1, *icd*A-5, *lip*L32-3, *lip*L32-35, *lip*L41-20, *lip*L41-30, *mre*A-1, *mre*A-37 and *pnt*A-2) coexisted in two species. Second, as shown in Table 3, *P* value determined by the *phi* test for the whole 271 STs and for each of the six species populations were all < 0.001. Third, 12 potential HGT events were detected by at least three distinct methods of the RDP4 package affecting six out of seven genes of the R7-MLST scheme (Table 4). In this incongruence/recombination scenario the STs I_A^S values showed a significant tendency to a non-random association of MLST alleles which could be due to a recent divergence of the genus and therefore linkage disequilibrium has not yet been reached. Based on these results a bayesian phylogenetic approach was carried out to estimate the species tree given the multi-individual multilocus sequence dataset (Heled and Drummond, 2010). As shown in Fig. 3, high values of posterior probabilities were obtained in the maximum clade credibility (MCC) tree (burn in 2×10^7) which showed the same topology as maximum-likelihood (ML) dendrogram for intermediate species. On the other hand, pathogenic species diverge into two main clusters. Cluster 1 correlates with ML topology including *L. interrogans*, *L. kirschneri* and *L. noguchii* species. However, differing from ML tree, *L. alstonii* and *L. kmetyi* species showed to be closely related to ML cluster 2 species (*L. weilii*, *L. santarosai*, *L. mayotensis*, *L. alexanderii* and *L. borgpetersenii*). Consensus trees density plots displayed phylogenetic signal conflicts among this cluster (Supplementary Fig. 2).

Table 3
phi recombination test and linkage disequilibrium analysis. I_A^S: standardized index of association.

Population	No. STs	Recombination	Linkage desequilibrium	
		phi	I_A^S	*P* value
Whole collection	271	$P < 0.001$	0.2857	< 0.001
L. interrogans	125	$P < 0.001$	0.1073	< 0.001
L. kirschneri	31	$P < 0.001$	0.2063	< 0.001
L. borgpetersenii	23	$P < 0.001$	0.2204	< 0.001
L. santarosai	40	$P < 0.001$	0.1016	< 0.001
L. noguchii	18	$P < 0.001$	0.1807	< 0.001
L. weilli	22	$P < 0.001$	0.2082	< 0.001

As genomic big data analysis is exhibiting, concatenation of individual genes or even complete genomes analysis from prokaryotes collapse from bootstrap analysis and/or congruence perspectives. The assumption that different genetic regions share a common phylogeny neglects processes such as recombination, gene duplication and lateral gene transfer, which as seen in this study and before, played an important role in the evolution of *Leptospira* genus (Xu et al., 2016).

From the epidemiological point of view R7-MLST scheme represents a great support for the correct taxonomic classification, species identification and tracking, and follow up of isolated strains or clinical samples (manuscript in preparation), within a given host or geographic region population. Furthermore, the correlation observed between STs and serogroups in our previous work, is still satisfied with this higher dataset sustaining the implementation of MLST to assist serological classification as a complementary approach.

In sumary, this work allowed us to deepen the understanding of the genetic variability of the genus *Leptospira*, thus contributing with suitable epidemiologic information useful to apply in the design of control strategies and also in diagnostic methods for this illness.

Supplementary data to this article can be found online at http://dx.doi.org/10.1016/j.meegid.2017.08.013.

Acknowledgements

We thank Dr. Julia Sabio y García for their support, encouragement and critical reading of this manuscript. We also thank Dr. Keith Jolley from Oxford University for his valuable help in the *Leptospira* database implementation, Dario Carballido for his technical and bioinformatics support, Verónica Lia for Bayesian phylogenetic guidance and Nicolas Tomasini for congruence analysis assistance. We wanted to thank Dr. Mette Voldby Larsen and Dr. Vanessa Jurtz from the Technical University of Denmark.

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 postdoctoral fellowship from the CONICET. K.C. and P.R. are CONICET fellows.

Table 4
Horizontal gene transfer events analysis in R7-MLST scheme. N: number of STs involved in HGT event. PPST: potential parent ST. M: major potential parent. m: minor potential parent. A: number of different methods supporting the HGT event. *P*: highest *P* value obtained after multiple-comparison correction supported for at least three algorithms.

HGT event	Species	STs involved in HGT event	N	PPST		A	Breakpoint		Genes(s) involved in HGT event	*P*
				M	m		Start	End		
1	*L. santarosai*	64	1	169	27	6	1450	2375	*lip*L32–*lip*L41	1510×10^{-61}
2	*L. weilli*	69, 62, 90, 74	4	206	27	6	1418	2375	*lip*L31–*lip*L41	2222×10^{-25}
3	*L. interrogans*	46	1	16	160	5	1870	2367	*lip*L41	5621×10^{-48}
4	*L. interrogans*	10	1	209	160	6	880	2331	*icd*A–*lip*L31–*lip*L41	4804×10^{-47}
5	*L. weilli*	68	1	193	27	6	1411	2375	*lip*L31–*lip*L41	2969×10^{-44}
6	*L. weilli*	68	1	264	19	6	29	425	*adk*	3297×10^{-41}
7	*L. borgpetersenii*	85	1	154	87	5	2367	3327	*mre*A–*pnt*A	5883×10^{-24}
8	*L. noguchii*	83	1	70	27	6	950	2370	*lip*L31–*lip*L41	3811×10^{-17}
9	*L. alexanderi*	81	1	193	155	6	865	2372	*icd*A–*lip*L31–*lip*L41	8833×10^{-17}
10	*L. borgpetersenii*L. *santarosai* (ST67)	60, 67, 75, 77, 79, 80, 87, 88, 89, 91, 92, 93, 99, 175, 176, 177, 178, 179, 180, 181, 212, 213, 214	23	162	54	3	1758	1905	*lip*L41	1137×10^{-11}
11	*L. noguchii*	73, 83	2	97	16	5	39	426	*adk*	1037×10^{-06}
12	*L. weilli*	74	1	17	81	4	1916	2138	*lip*L41	3765×10^{-04}

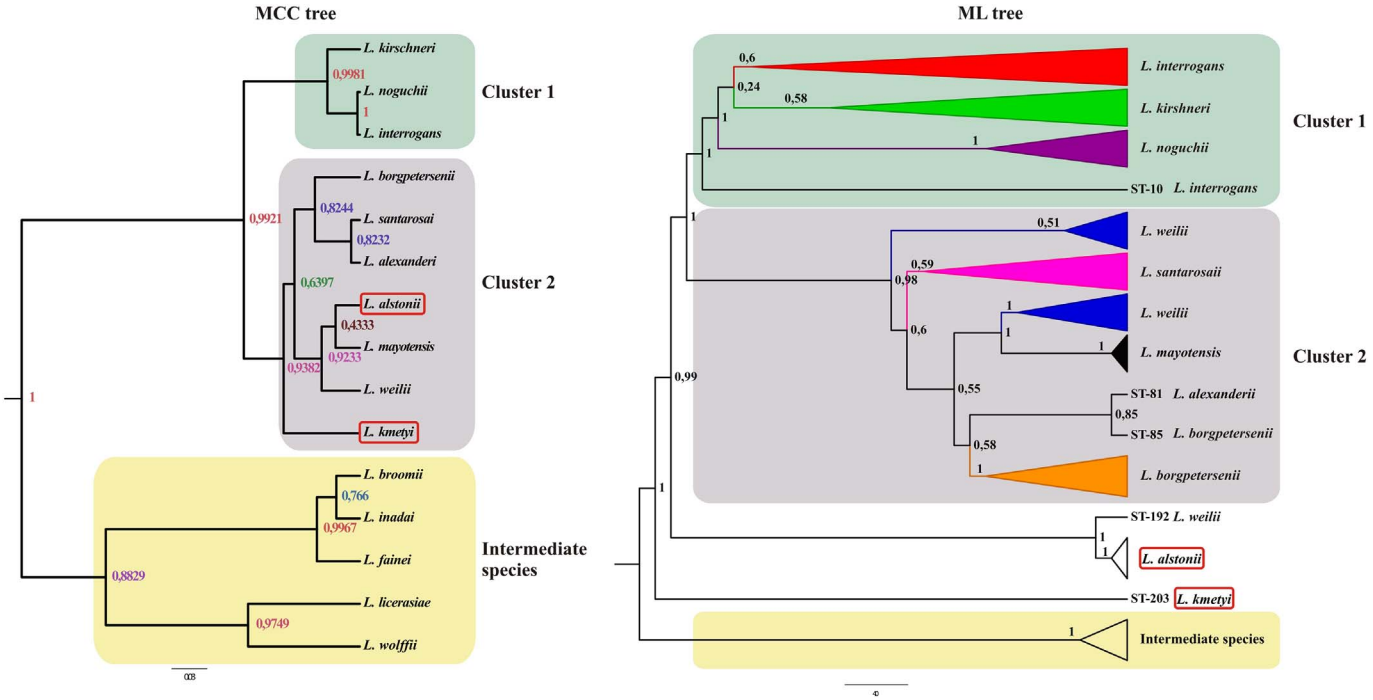

Fig. 3. Maximum-likelihood (ML) and Bayesian (MCC) phylogenetic analysis. Intermediate species were resolved from pathogenic species. Different topologies were obtained by each approach for pathogenic species.

References

Ahmed, N., Devi, S.M., Valverde, Mde.L., Vijayachari, P., Machang'u, 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.

Arzouni, J.P., Parola, P., La Scola, B., Postic, D., Brouqui, P., Raoult, D., 2002. Human infection caused by *Leptospira fainei*. Emerg. Infect. Dis. 8, 865–868.

Bandara, M., Ananda, M., Wickramage, K., Berger, E., Agampodi, S., 2014. Globalization of leptospirosis through travel and migration. Glob. Health 10, 61. http://dx.doi.org/10.1186/s12992-014-0061-0.

Bharti, A.R., Nally, J.E., Ricaldi, J.N., Matthias, M.A., Diaz, M.M., Lovett, M.A., Levett, P.N., Gilman, R.H., Willig, M.R., Gotuzzo, E., Vinetz, J.M., 2003. Peru-United States Leptospirosis Consortium. Leptospirosis: a zoonotic disease of global importance. Lancet Infect. Dis. 3 (12), 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. http://dx.doi.org/10.1371/journal.pntd.0001954.

Bouckaert, R.R., 2010. DensiTree: making sense of sets of phylogenetic trees. Bioinformatics 26, 1372–1373. http://dx.doi.org/10.1093/bioinformatics/btq110.

Bruen, T.C., Philippe, H., Bryant, D., 2006. A simple and robust statistical test for detecting the presence of recombination. Genetics 172, 2665–2681.

Bulach, D.M., Zuerner, R.L., Wilson, P., Seemann, T., McGrath, A., Cullen, P.A., Davis, J., Johnson, M., Kuczek, E., Alt, D.P., Peterson-Burch, B., Coppel, R.L., Rood, J.I., Davies, J.K., Adler, B., 2006. Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. Proc. Natl. Acad. Sci. U. S. A. 103, 14560–14565.

Chou, L.F., Chen, Y.T., Lu, C.W., Ko, Y.C., Tang, C.Y., Pan, M.J., Tian, Y.C., Chiu, C.H., Hung, C.C., Yang, C.W., 2012. Sequence of *Leptospira santarosai* serovar Shermani genome and prediction of virulence-associated genes. Gene 511, 364–370. http://dx.doi.org/10.1016/j.gene.2012.09.074.

Costa, F., Hagan, J.E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M.S., Stein, C., Abela-Ridder, B., Ko, A.I., 2015. Global morbidity and mortality of leptospirosis: a systematic review. PLoS Negl. Trop. Dis. 9, e0003898. http://dx.doi.org/10.1371/journal.pntd.0003898.

Darriba, D., Taboada, G.L., Doallo, R., Posada, D., 2012. jModelTest 2: more models, new heuristics and parallel computing. Nat. Methods 9, 772. http://dx.doi.org/10.1038/nmeth.2109.

Drummond, A.J., Suchard, M.A., Xie, D., Rambaut, A., 2012. Bayesian phylogenetics with BEAUti and the BEAST 1.7. Mol. Biol. Evol. 29, 1969–1973. http://dx.doi.org/10.1093/molbev/mss075.

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 related bacterial genotypes from multilocus sequence typing data. J. Bacteriol. 186, 1518–1530.

Fouts, D.E., Matthias, M.A., Adhikarla, H., Adler, B., Amorim-Santos, L., Berg, D.E.,

Bulach, D., Buschiazzo, A., Chang, Y.F., Galloway, R.L., Haake, D.A., Haft, D.H., Hartskeerl, R., Ko, A.I., Levett, P.N., Matsunaga, J., Mechaly, A.E., Monk, J.M., Nascimento, A.L., Nelson, K.E., Palsson, B., Peacock, S.J., Picardeau, M., Ricaldi, J.N., Thaipandungpanit, J., Wunder Jr., E.A., Yang, X.F., Zhang, J.J., Vinetz, J.M., 2016. What makes a bacterial species pathogenic?: comparative genomic analysis of the genus *Leptospira*. PLoS Negl. Trop. Dis. 10, e0004403. http://dx.doi.org/10.1371/journal. pntd.0004403.

Francisco, A.P., Bugalho, M., Ramirez, M., Carriço, J.A., 2009. Global optimal eBURST analysis of multilocus typing data using a graphic matroid approach. BMC Bioinf. 10, 152. http://dx.doi.org/10.1186/1471-2105-10-152.

Gibbs, M.J., Armstrong, J.S., Gibbs, A.J., 2000. Sister-scanning: a Monte Carlo procedure for assessing signals in recombinant sequences. Bioinformatics 16, 573–582.

Goarant, C., 2014. Leptospirosis: time to move to molecular epidemiology comments on "reassessment of MLST schemes for *Leptospira* spp. typing worldwide" by Varni and colleagues. Infect. Genet. Evol. 21, 484–485. http://dx.doi.org/10.1016/j.meegid.2013.10.018.

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

Haake, D.A., Suchard, M.A., Kelley, M.M., Dundoo, M., Alt, D.A., Zuerner, R.A., 2004. Molecular evolution and mosaicism of leptospiral outer membrane proteins involves horizontal DNA transfer. J. Bacteriol. 186, 2818–2828.

Haubold, B., Hudson, R.R., 2000. LIAN 3.0: detecting linkage disequilibrium in multilocus data. Linkage analysis. Bioinformatics 16, 847–848.

Heled, J., Drummond, A.J., 2010. Bayesian inference of species trees from multilocus data. Mol. Biol. Evol. 27, 570–580. http://dx.doi.org/10.1093/molbe v/msp274.

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.

Jeffroy, O., Brinkmann, H., Delsuc, F., Philippe, H., 2006. Phylogenomics: the beginning of incongruence? Trends Genet. 22 (4), 225–231.

Jolley, K.A., Maiden, M.C., 2010. BIGSdb: scalable analysis of bacterial genome variation at the population level. BMC Bioinf. 11, 595. http://dx.doi.org/10.1186/1471-2105-11-595.

Kmety, E., Dikken, H., 1993. Classification of the Species *Leptospira interrogans* and History of Its Serovars. University Press, Groningen (104 pp.).

Larsen, M.V., Cosentino, S., Rasmussen, S., Friis, C., Hasman, H., Marvig, R.L., Jelsbak, L., Sicheritz-Pontén, T., Ussery, D.W., Aarestrup, F.M., Lund, O., 2012. Multilocus sequence typing of total genome sequenced bacteria. J. Clin. Micobiol. 50, 1355–1361. http://dx.doi.org/10.1128/JCM.06094-11.

Lehmann, J.S., Matthias, M.A., Vinetz, J.M., Fouts, D.E., 2014. Leptospiral pathogenomics. Pathogens. 3, 280–308. http://dx.doi.org/10.3390/pathogens3020280.

Levett, P.N., 2001. Leptospirosis. Clin. Microbiol. Rev. 14, 296–326. http://dx.doi.org/10.1128/CMR.14.2.296-326.

Librado, P., Rozas, J., 2009. DnaSP v5: a software for comprehensive analysis of DNA polymorphism data. Bioinformatics 25, 1451–1452. http://dx.doi.org/10.1093/bioinformatics/btp187.

Maiden, M.C., Bygraves, J.A., Feil, E., Morelli, G., Russel, J.E., Urwin, R., Zhang, Q., Zhou, J., Zurth, K., Caugant, D.A., Feavers, I.M., Achtman, M., Spratt, B.G., 1998. Multilocus sequence typing: a portable approach to the identification of clones within populations of pathogenic microorganisms. Proc. Natl. Acad. Sci. U. S. A. 95, 3140–3145.

Martin, D., Rybicki, E., 2000. RDP: detection of recombination amongst aligned sequences. Bioinformatics 16, 562–563.

Martin, D.P., Posada, D., Crandall, K.A., Williamson, C., 2005. A modified bootscan algorithm for automated identification of recombinant sequences and recombination breakpoints. AIDS Res. Hum. Retrovir. 21, 98–102.

Martin, D.P., Murrell, B., Golden, M., Khoosal, A., Muhire, B., 2015. RDP4: detection and analysis of recombination patterns in virus genomes. Virus Evol. 1, vev003. http://dx.doi.org/10.1093/ve/vev003.

Maynard Smith, J., 1992. Analyzing the mosaic structure of genes. J. Mol. Evol. 34, 126–129.

McBride, A.J.A., Cerqueira, G.M., Suchard, M.A., Moreira, A.N., Zuerner, R.L., Reis, M.G., Haake, D.A., Ko, A.I., Dellagostin, O.A., 2009. Genetic diversity of the Leptospiral immunoglobulin-like (Lig) genes in pathogenic *Leptospira* spp. Infect. Genet. Evol. 9, 196–205. http://dx.doi.org/10.1016/j.meegid.2008.10.012.

Nascimento, A.L., Verjovski-Almeida, S., Van Sluys, M.A., Monteiro-Vitorello, C.B., Camargo, L.E., Digiampietri, L.A., Harstkeerl, R.A., Ho, P.L., Marques, M.V., Oliveira, M.C., Setubal, J.C., Haake, D.A., Martins, E.A., 2004. Genome features of *Leptospira interrogans* serovar Copenhageni. Braz. J. Med. Biol. Res. 37, 459–477.

Padidam, M., Sawyer, S., Fauquet, C.M., 1999. Possible emergence of new geminiviruses by frequent recombination. Virology 265, 218–225.

Picardeau, M., Bulach, D.M., Bouchier, C., Zuerner, R.L., Zidane, N., Wilson, P.J., Creno, S., Kuczek, E.S., Bommezzadri, S., Davis, J.C., McGrath, A., Johnson, M.J., Boursaux-Eude, C., Seemann, T., Rouy, Z., Coppel, R.L., Rood, J.I., Lajus, A., Davies, J.K., Médigue, C., Adler, B., 2008. Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. PLoS One 3, e1607. http://dx.doi.org/10.1371/journal.pone.0001607.

Posada, D., Crandall, K.A., 2001. Evaluation of methods for detecting recombination from DNA sequences: computer simulations. Proc. Natl. Acad. Sci. U. S. A. 98, 13757–13762.

Ralph, D., McClelland, M., 1994. Phylogenetic evidence for horizontal transfer of an intervening sequence between species in a spirochete genus. J. Bacteriol. 176, 5982–5987.

Rambaut, A., Suchard, M.A., Xie, D., Drummond, A.J., 2014. Tracer v1.6. Available from. http://beast.bio.ed.ac.uk/Tracer.

Ren, S.X., Fu, G., Jiang, X.G., Zeng, R., Miao, Y.G., Xu, H., Zhang, Y.X., Xiong, H., Lu, G., Lu, L.F., Jiang, H.Q., Jia, J., Tu, Y.F., Jiang, J.X., Gu, W.Y., Zhang, Y.Q., Cai, Z., Sheng, H.H., Yin, H.F., Zhang, Y., Zhu, G.F., Wan, M., Huang, H.L., Qian, Z., Wang, S.Y., Ma, W., Yao, Z.J., Shen, Y., Qiang, B.Q., Xia, Q.C., Guo, X.K., Danchin, A., Saint, Girons I., Somerville, R.L., Wen, Y.M., Shi, M.H., Chen, Z., Xu, J.G., Zhao, G.P., 2003. Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. Nature 422, 888–893.

Ricaldi, J.N., Fouts, D.E., Selengut, J.D., Harkins, D.M., Patra, K.P., Moreno, A., Lehmann, J.S., Purushe, J., Sanka, R., Torres, M., Webster, N.J., Vinetz, J.M., Matthias, M.A., 2012. Whole genome analysis of *Leptospira licerasiae* provides insight into leptospiral evolution and pathogenicity. PLoS Negl. Trop. Dis. 6, e1853. http://dx.doi.org/10.1371/journal.pntd.0001853.

Romero, E.C., Blanco, R.M., Galloway, R.L., 2011. Analysis of multilocus sequence typing for identification of *Leptospira* isolates in Brazil. J. Clin. Microbiol. 49, 3940–3942. http://dx.doi.org/10.1128/JCM.01119-11.

Tamura, K., Stecher, G., Peterson, D., Filipski, A., Kumar, S., 2013. MEGA6: molecular evolutionary genetics analysis version 6.0. Mol. Biol. Evol. 30, 2725–2729. http://dx.doi.org/10.1093/molbev/mst197.

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.

Thiergart, T., Landan, G., Martin, W.F., 2014. Concatenated alignments and the case of the disappearing tree. BMC Evol. Biol. 14, 266. http://dx.doi.org/10.1186/s12862-014-0266-0.

Tomasini, N., Lauthier, J.J., Llewellyn, M.S., Diosque, P., 2013. MLSTest: novel software for multi-locus sequence data analysis in eukaryotic organisms. Infect. Genet. Evol. 20, 188–196. http://dx.doi.org/10.1016/j.meegid.2013.08.029.

Tomasini, N., Lauthier, J.J., Ayala, F.J., Tibayrenc, M., Diosque, P., 2014. How often do they have sex? A comparative analysis of the population structure of seven eukaryotic microbial pathogens. PLoS One 9, e103131. http://dx.doi.org/10.1371/journal.pone.0103131.

Varni, V., Ruybal, P., Lauthier, J.J., Tomasini, N., Brihuega, B., Koval, A., Caimi, K., 2014. Reassessment of MLST schemes for *Leptospira* spp. typing worldwide. Infect. Genet. Evol. 22, 216–222. http://dx.doi.org/10.1016/j.meegid.2013.08.002.

Xu, Y., Zhu, Y., Wang, Y., Chang, Y.F., Zhang, Y., Jiang, X., Zhuang, X., Zhu, Y., Zhang, J., Zeng, L., Yang, M., Li, S., Wang, S., Ye, Q., Xin, X., Zhao, G., Zheng, H., Guo, X., Wang, J., 2016. Whole genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic *Leptospira*. Sci Rep 6, 20020. http://dx.doi.org/10.1038/srep20020.