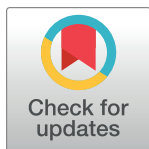


RESEARCH ARTICLE

Phylogenomics of *Leptospira santarosai*, a prevalent pathogenic species in the AmericasDiana Chinchilla¹✉, Cecilia Nieves²✉, Ricardo Gutiérrez¹, Vallier Sordoillet³, Frédéric J. Veyrier², Mathieu Picardeau³*✉

1 Centro Nacional de Referencia de Bacteriología, Instituto Costarricense de Investigación y Enseñanza en Nutrición y Salud (INCIENSA), La Unión, Cartago, Costa Rica, **2** Bacterial Symbionts Evolution, Centre Armand-Frappier Santé Biotechnologie, Institut National de la Recherche Scientifique, Université du Québec, Laval, Québec, Canada, **3** Institut Pasteur, Université Paris Cité, Biology of Spirochetes Unit, Paris, France

✉ These authors contributed equally to this work.

* mathieu.picardeau@pasteur.fr

Abstract

Background

Leptospirosis is a complex zoonotic disease mostly caused by a group of eight pathogenic species (*L. interrogans*, *L. borgpetersenii*, *L. kirschneri*, *L. mayottensis*, *L. noguchii*, *L. santarosai*, *L. weillii*, *L. alexanderi*), with a wide spectrum of animal reservoirs and patient outcomes. *Leptospira interrogans* is considered as the leading causative agent of leptospirosis worldwide and it is the most studied species. However, the genomic features and phylogeography of other *Leptospira* pathogenic species remain to be determined.

Methodology/principal findings

Here we investigated the genome diversity of the main pathogenic *Leptospira* species based on a collection of 914 genomes from strains isolated around the world. Genome analyses revealed species-specific genome size and GC content, and an open pangenome in the pathogenic species, except for *L. mayottensis*. Taking advantage of a new set of genomes of *L. santarosai* strains isolated from patients in Costa Rica, we took a closer look at this species. *L. santarosai* strains are largely distributed in America, including the Caribbean islands, with over 96% of the available genomes originating from this continent. Phylogenetic analysis showed high genetic diversity within *L. santarosai*, and the clonal groups identified by cgMLST were strongly associated with geographical areas. Serotype identification based on serogrouping and/or analysis of the *O*-antigen biosynthesis gene loci further confirmed the great diversity of strains within the species.

Conclusions/significance

In conclusion, we report a comprehensive genome analysis of pathogenic *Leptospira* species with a focus on *L. santarosai*. Our study sheds new light onto the genomic diversity, evolutionary history, and epidemiology of leptospirosis in America and globally. Our findings also expand our knowledge of the genes driving *O*-antigen diversity. In addition, our work

OPEN ACCESS

Citation: Chinchilla D, Nieves C, Gutiérrez R, Sordoillet V, Veyrier FJ, Picardeau M (2023) Phylogenomics of *Leptospira santarosai*, a prevalent pathogenic species in the Americas. PLoS Negl Trop Dis 17(11): e0011733. <https://doi.org/10.1371/journal.pntd.0011733>

Editor: Joseph M. Vinetz, Yale University School of Medicine, UNITED STATES

Received: August 15, 2023

Accepted: October 17, 2023

Published: November 2, 2023

Copyright: © 2023 Chinchilla et al. This is an open access article distributed under the terms of the [Creative Commons Attribution License](https://creativecommons.org/licenses/by/4.0/), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Data Availability Statement: All underlying data supporting the findings are freely accessible via the following links: [urldefense.com/v3/___/www.ncbi.nlm.nih.gov/bioproject/PRJNA967664/___!!JFdNOqOXpB6UZWO!oy47RazM7x709WJddyTH2NSE3_oY47RazM7x709WJddyTH2NSE3_Lt45Ccz050dvsmPtuYj22klBOsl_RsgyF5tFuWkgO-WZa43gemZb7jfkGa71LMoQ\\$](https://urldefense.com/v3/___/www.ncbi.nlm.nih.gov/bioproject/PRJNA967664/___!!JFdNOqOXpB6UZWO!oy47RazM7x709WJddyTH2NSE3_oY47RazM7x709WJddyTH2NSE3_Lt45Ccz050dvsmPtuYj22klBOsl_RsgyF5tFuWkgO-WZa43gemZb7jfkGa71LMoQ$) and at bigsd.bpasteur.fr/cgi-bin/bigsd/bigsd.pl?db=pubmlst_leptospira_isolates&page=query&project_list=21&submit=1

Funding: This research was supported by Santé Publique France through the french National Reference Center of Leptospirosis, the Institut Pasteur through grant PTR 30-2017 (MP and FJV) and by National Institutes of Health grant P01 AI 168148 (MP and FJV). CN received a Ph.D. studentship Calmette & Yersin from the Institut Pasteur International Network. FJV received a Junior 1 and Junior 2 research scholar salary award from the Fonds de Recherche du Québec—Santé. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing interests: The authors have declared that no competing interests exist.

provides a framework for understanding the virulence and spread of *L. santarosai* and for improving its surveillance in both humans and animals.

Author summary

Leptospirosis is an emerging zoonosis caused by pathogenic species of a highly heterogeneous genus. Most studies have focused on *Leptospira interrogans* that is responsible for the majority of human infection cases worldwide. On the contrary, our knowledge is very limited for other pathogenic species, including *L. santarosai*, which may represent a public health problem in both humans and animals in the American continent. Our comparative genomic analyses of the pathogenic species revealed that *L. santarosai* is characterized by an open pangenome state with high genetic and serovar diversity. This first study of *L. santarosai* isolates not only contributes to the global understanding of genomics and evolution within the *Leptospira* pathogenic species but also provides the groundwork for better surveillance of this pathogen.

Introduction

Leptospira is a highly heterogeneous bacterial genus divided into pathogenic and saprophytic species and then further divided into more than 300 serovars, which are defined according to structural heterogeneity of the lipopolysaccharide (LPS) O-antigen. Nowadays, strain identification is mainly based on genome analysis, and core genome multilocus sequence typing (cgMLST) [1] enables identification of the species and below. Recent studies have also shown that whole-genome sequences can be used for predicting *Leptospira* serotypes on the basis of the *rfb* locus which contains the genes for the O-antigen biosynthesis [2,3]. This approach offers a promising alternative to the conventional serotyping method, which is laborious, time-consuming, expensive and requires a high level of expertise.

Over the past decade, the number of *Leptospira* species described has rapidly extended from 22 in 2014 to 69 in 2022 [4], largely due to the use of improved protocols for culture isolation from the environment [5,6] and the generalization of next generation sequencing [7]. Among the genus *Leptospira*, eight species (*L. interrogans*, *L. kirschneri*, *L. noguchii*, *L. santarosai*, *L. mayottensis*, *L. borgpetersenii*, *L. alexanderi* and *L. weilii*), which diverged after a specific node of evolution, constitute the most virulent group of pathogenic species [8]. These *Leptospira* species are the causative agents of leptospirosis in both human and animals, leading to a high disease burden in tropical countries [9] and major economic losses in the livestock sector [10].

Our previous analysis of the distribution of pathogenic *Leptospira* species showed that *L. interrogans* is the most frequently encountered and globally distributed species [1]. This cosmopolitan species is also by far the most studied in terms of virulence, and molecular epidemiology, among other aspects. On the contrary, to date, very little is known about the geographical distribution, reservoirs, genomic features and virulence factors of pathogenic species other than *L. interrogans*. In the same analysis, we showed that some pathogenic species were geographically restricted [1]. Thus, only limited reports have described the existence of *L. santarosai* outside the American continent. *L. santarosai*, named after Carlos A. Santa Rosa, a Brazilian veterinary microbiologist who pioneered the study of leptospirosis in Brazil, was first described in 1987 [11]. *L. santarosai* is predominant in many countries from Central and South America.

In the present study, we first performed an analysis of the pangenome in pathogenic *Leptospira* species and then took a closer look at the genetic diversity of *L. santarosai* including a set of strains recently isolated from patients in Costa Rica, which is an endemic country for leptospirosis [12].

Phylogenomics analysis of *L. santarosai* genomes will enable to better understand the genetic diversity and genome features of this pathogenic species which is prevalent in most countries of the American continent.

Material and methods

Ethics statement

According to the decree number 40556-s of the General Health Law of Costa Rica, epidemiological studies that incorporate the review of clinical records do not require the approval of an ethics-scientific committee. Additionally, no written informed consent from patients was required, as the study was conducted as part of the routine diagnosis at the Centro Nacional de Referencia de Bacteriología of the Instituto Costarricense de Investigación y Enseñanza en Nutrición y Salud (INCIENSA). No additional clinical specimens were collected for the purpose of the study. Human samples were anonymized, and collection of the samples was conducted according to the Declaration of Helsinki.

Strains

Isolates sequenced in this study (n = 153) were obtained from the collections of the French National Reference Center for Leptospirosis (Institut Pasteur, Paris, France), Laboratorio de Genética Molecular (Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela), Institut Pasteur of Alger (Algiers, Algeria), Institute of Veterinary Bacteriology (University of Bern, Switzerland), Molecular Epidemiology and Public Health Laboratory (School of Veterinary Sciences, Massey University, New Zealand), Instituto de Higiene (Facultad de Medicina, Universidad de la República, Montevideo, Uruguay), Universidade Federal Fluminense (Rio de Janeiro, Brazil), Faculty of Veterinary Medicine (University of Zagreb, Croatia), National Collaborating Centre for Reference and Research on Leptospirosis (Academic Medical Center, Amsterdam, the Netherlands), Laboratory of Zoonoses (Pasteur Institute in Saint Petersburg, Saint Petersburg, Russia), Institute for Medical Research (Malaysia), Faculty of Medicine and Health Sciences (University Putra Malaysia, Malaysia), and Leptospirosis Research and Expertise Unit (Institut Pasteur Nouvelle-Calédonie, Nouméa, New Caledonia), Kimron Veterinary Institute (Israel). We also downloaded genomes from our previous studies including isolates from the collections of Lao-Oxford-Mahosot Hospital-Wellcome Trust-Research Unit (LOMWRU) (Microbiology Laboratory, Mahosot Hospital, Vientiane, Lao People's Democratic Republic), Unidad Mixta Pasteur-Instituto Nacional de Investigación Agropecuaria (Institut Pasteur of Montevideo, Montevideo, Uruguay), Centre Hospitalier de Mayotte (France), and Department of Mycology-Bacteriology (Institute of Tropical Medicine Pedro Kourí, Havana, Cuba) [1,2,13–15] as well as genomes from the NCBI database. Information on strains and genomes used in this study are indicated in [S1 and S2 Tables](#).

Whole-genome sequencing

Illumina sequencing was performed from extracted genomic DNAs of exponential-phase cultures using a MagNA Pure 96 Instrument (Roche, Meylan, France). Next-generation sequencing (NGS) was performed using Nextera XT DNA Library Preparation kit and the NextSeq 500 sequencing systems (Illumina, San Diego, CA, USA) at the Mutualized Platform for

Microbiology (P2M) at Institut Pasteur. CLC Genomics Workbench 9 software (Qiagen, Hilden, Germany) was used for analyses. The generated contig sequences together with the sample metadata are available in BIGSdb hosted at the Institut Pasteur ([//bigsdb.pasteur.fr/leptospira/](http://bigsdb.pasteur.fr/leptospira/)). We also downloaded additional genome sequences of *Leptospira* isolates from the NCBI database (S1 Table). Only genomes meeting quality requirements, such as i) sequencing coverage >30x, ii) number of contigs <600, iii) cumulative contigs length within the typical range of *Leptospira* genomes (3.6–6Mb), iv) GC content within the typical range of *Leptospira* genomes (35–48%), and v) <100 uncalled cgMLST alleles out of the 545 pre-defined core genes, were selected for further analyses.

Genomic analyses

Comparative analyses of the pangenome were performed using two software: Roary version 3.11.2 [16], and a combination of COG and OMCL algorithms in GET_HOMOLOGUES version 20190411 [12]. Both methods yielded a similar number of gene clusters. In the Roary analysis, a 60% identity cut-off was applied to define gene clusters (option -i 60), and no other parameters were modified. Among the Roary outputs, a tab-separated file containing the number of genes in the pangenome was used to create a graph depicting the variation in the number of gene clusters as a function of the number of genomes analyzed. Roary iterated 10 times, calculating the number of new genes added as each genome was sequentially incorporated into the analysis. This graph facilitated a quick determination of whether the pangenome was open or closed and allowed for the calculation of the α coefficient in Heap's Law ($n = \kappa N^\gamma$, with $\gamma = 1 - \alpha$) [17]. On the other hand, GET_HOMOLOGUES was used to infer the pangenome distribution in cloud-, shell-, soft-core-, and core-genome. This was achieved by generating a tab-separated pangenome matrix file that included the number of all the clusters identified by both COG and OMCL algorithms. The matrix represented the intersection of the two methods and served as input for the `parse_pangenome_matrix.pl` script within GET_HOMOLOGUES, which classified the clusters as cloud (shared by up to 2 genomes), shell (shared by more than 2 genomes but less than 93% of genomes analyzed), soft-core (shared by 93–99% of genomes), or core-genes (shared by 100% of genomes). Due to the substantial number of genomes available for *L. interrogans* and *L. borgpetersenii*, as well as the redundancies observed in serogroups and serovars, representative genomes of each serogroup/serovar were selectively chosen to mitigate computational costs. Excluding genomes with redundant identities is not anticipated to result in significant alterations in the pangenome distribution.

Genome size and GC content for highly virulent *Leptospira* species were determined through DFAST annotation [18]. Individual values were plotted and grouped per species, with the mean and standard deviation displayed. Genome size and GC content were compared using the Kruskal-Wallis Rank Sum Test, for the comparison of *Leptospira* spp. and the Wilcoxon rank test, for the comparison of two phylogenetic-related groups. Post-hoc comparisons were performed using Dunn's Kruskal-Wallis Multiple Comparisons (Dunn, 1964). P-values were adjusted with the Bonferroni method. Statistical analyses were performed in R [19], using FSA package [20].

Average Nucleotide Identity (ANI) and Percentage of Conserved Proteins (POCP) were calculated for the 64 *L. santarosai* genome sequences as well as *L. interrogans* str. Fiocruz L1-130 and *L. borgpetersenii* str. M84 used as outgroups (S1 and S2 Figs). Genomes were annotated by Prokka version 1.13.7 [21]. ANI and POCP matrices were inferred using OMCL algorithm via GET_HOMOLOGUES version 20190411 [12]. Briefly, to calculate ANI, the option -A was employed along with option -a to utilize nucleotide sequences and perform BLASTN. This process generated a tab-separated file containing average percentage sequence identity

values between pairs of genomes, calculated from sequences within all identified clusters (option $t = 0$). This tab-separated file served as the input to create a symmetric matrix, where the genomes were clustered based on their ANI values. Dendrograms based on this clustering were generated on both sides of the matrix to visually represent the proximity among genomes. Similarly, POCP was calculated by including the option $-P$ and performing default BLASTP searches. This step yielded another tab-separated file, which was subsequently used to create a symmetric matrix. Analogous to the ANI matrix, the genomes were clustered based on the shared % of conserved proteins between pairs of genomes. These values are calculated as $POCP = (C_a + C_b) / (total_a + total_b)$, where C_a and C_b denote the number of conserved proteins from genome a in genome b and from genome b in genome a, respectively, normalized by the sum of total proteins in each genome. The clustering process also generated dendrograms, indicating the proximity among genomes in terms of conserved proteins.

Core genome MLST (cgMLST) typing was performed using a scheme based on 545 core genes as previously described [1]. *L. santarosai* core-genome based phylogeny was constructed using the 1288 core-genes alignment resulting from Roary analysis (60% identity cut-off, option $-i 60$). The best-fit model and the maximum-likelihood phylogenetic tree were determined by IQ-TREE version 1.6.11 [22], considering 10,000 ultrafast bootstraps [23]. *L. interrogans* str. Fiocruz L1-130 and *L. borgpetersenii* str. M84 were used as outgroups. Tree branches were transformed with the "proportional" option on FigTree software v1.4.4 (<http://tree.bio.ed.ac.uk/software/figtree/>), which adjusts branch distances according to the number of tips under each node to improve visualization of the tree. Gene presence/absence analyses among *rfb* clusters from different genomes here studied were performed by protein-level searches using BLASTP [24] and subsequent network associations by NetworkX version 2.6.2 [25]. A similarity threshold of 60% was applied, as previously described [2]. The resulting presence/absence table obtained from the network association analysis was converted into a binary CSV file, where 0 represents gene absence and 1 represents gene presence. This binary table was subjected to hierarchical clustering based on shared protein-encoding genes (options: euclidean distance, ward linkage) using available tools at [//mev.tm4.org](http://mev.tm4.org). Jaccard's similarity index was used to measure the similarity between *rfb* patterns.

Genomic data

The sequencing data generated in this study are available in the NCBI database under the BioSample accession numbers SAMN34670613, SAMN34670614, SAMN34670615, SAMN34670616, SAMN34670617, SAMN34670618, SAMN34670619, SAMN34670620, SAMN34670621, SAMN34670622, SAMN34670623, SAMN34670624, SAMN34670625, SAMN34670626, SAMN34670627, SAMN34670628, SAMN34670629, SAMN34670630. Genome sequences used in this study are also available at https://bigsdbs.pasteur.fr/cgi-bin/bigsdbs/bigsdbs.pl?db=pubmlst_leptospira_isolates&page=query&project_list=21&submit=1.

Results and discussion

Distribution of pathogenic *Leptospira* species shows that *L. santarosai* isolates are mostly from the Americas

We first investigated the geographical distribution of pathogenic *Leptospira* species using 914 genomes of isolates collected between 1928 and 2022 (S1 Table).

Species included in our study are: *L. interrogans* (n = 410), *L. borgpetersenii* (n = 264), *L. kirschneri* (n = 88), *L. mayottensis* (n = 33), *L. noguchii* (n = 31), *L. santarosai* (n = 64), *L. weilii* (n = 24); *L. alexanderi*, with only 2 isolates in our database, was not included in this study.

Strains were isolated from human (50%) and animal (49%) samples, in Europe (18%), Africa (2%), Indian Ocean (14%), Caribbean islands (6%), Central America (3%), South America (13%), North America (4%), Central Asia, South Asia, East and Western Asia (11%), Southeast Asia (14%), and Australia and the Pacific region (15%) (Fig 1). Although this study is based on the genomes available in the databases and may introduce a bias, *L. interrogans*, *L. kirschneri*, and *L. borgpetersenii* are distributed worldwide, *L. weilii* is mostly found in Asia, Australia and the Pacific region, *L. mayottensis* in the Indian Ocean, and *L. noguchii* and *L. santarosai* in America as previously shown [1].

Leptospirosis is endemic in most countries of South and Central America, as well as in the Caribbean region [9,26–28]. In addition, most outbreaks of leptospirosis have been reported in the Latin America and the Caribbean region [29], where the disease is widespread in domestic and wild animals [12]. However, comprehensive data concerning human and animal leptospirosis remain largely scarce in most American countries [30]. We previously studied the genomes of *L. noguchii* isolated from human and animals in America [2] but our knowledge of *L. santarosai*, the other prevalent species in America, is rather limited.

Here, we sequenced 18 *L. santarosai* strains, including twelve strains that were isolated in Costa Rica in 2020–2021 from patients. The ANI and POCP values were calculated for the 64 *L. santarosai* strains further confirming they all belong to the same species (S1 and S2 Figs). Of the 64 *L. santarosai* strains in our genome database, 28 were isolated in South America (Brazil, Colombia, Ecuador, Peru), 21 in Central America (Costa Rica, Panama), 6 in North America (US; not including Puerto Rico), 7 in the Caribbean region (Martinique, Guadeloupe, Trinidad and Tobago and Puerto Rico), and only two strains were isolated outside the Americas (China and Democratic Republic of the Congo) (Fig 1 and S2 Table). Of note, *L. santarosai* has not been isolated in Uruguay, where a large number of *Leptospira* strains have been isolated from cattle [15].

L. santarosai strains in our study were isolated from humans (n = 38), bovine (n = 12), rodents (n = 8, including rats, spiny rats, capybara and muskrat), opossum (n = 2), dog (n = 1), goat (n = 1), pig (n = 1), and racoon (n = 1) (S2 Table).

Previous studies have shown that *L. santarosai* can be detected from different sources in many countries of America and the Caribbean region. It is the predominant species in humans, rodents and dogs in Peru and Colombia [31,32]. In Peru, it has additionally been found in rural environmental water samples (but not in urban samples), as well as in association with pigs and cattle [33]. In Brazil, *L. santarosai* has been isolated from dogs [34], cattle [35], goats [36], and capybaras [37]. Moreover, it has also been identified in patients in French Guiana [38], Guadeloupe [39] and the US [40].

Only a few reports have described the existence of *L. santarosai* outside the American continent. Some years after the original description of *L. santarosai* [11], Brenner *et al.* listed 65 *L. santarosai* strains, of which only three were isolated from outside America [41]. One *L. santarosai* strain was isolated from a patient in Sri-Lanka in 1966 but has never been reported in this country afterwards [42, 43]. The other two strains were isolated in Denmark and Indonesia but, again, *L. santarosai* has not been subsequently isolated in these countries. More recently, a strain belonging to *L. santarosai* serogroup Grippotyphosa was isolated from a patient in India and its genome sequenced [44]. However, because of highly fragmented genome (884 contigs) and missing genomic data (135 uncalled cgMLST alleles), the cluster assignment was not possible for this isolate and we removed its genome from our analysis.

The serogroup Shermani, is commonly reported in serological surveys in animals in Asia [45–48]. Unfortunately, there is no evidence that the infecting strains described in these studies were *L. santarosai* or another species such as *L. noguchii* and *L. inadai* which also contain serovars from the serogroup Shermani [49]. Finally, the other country outside Americas where

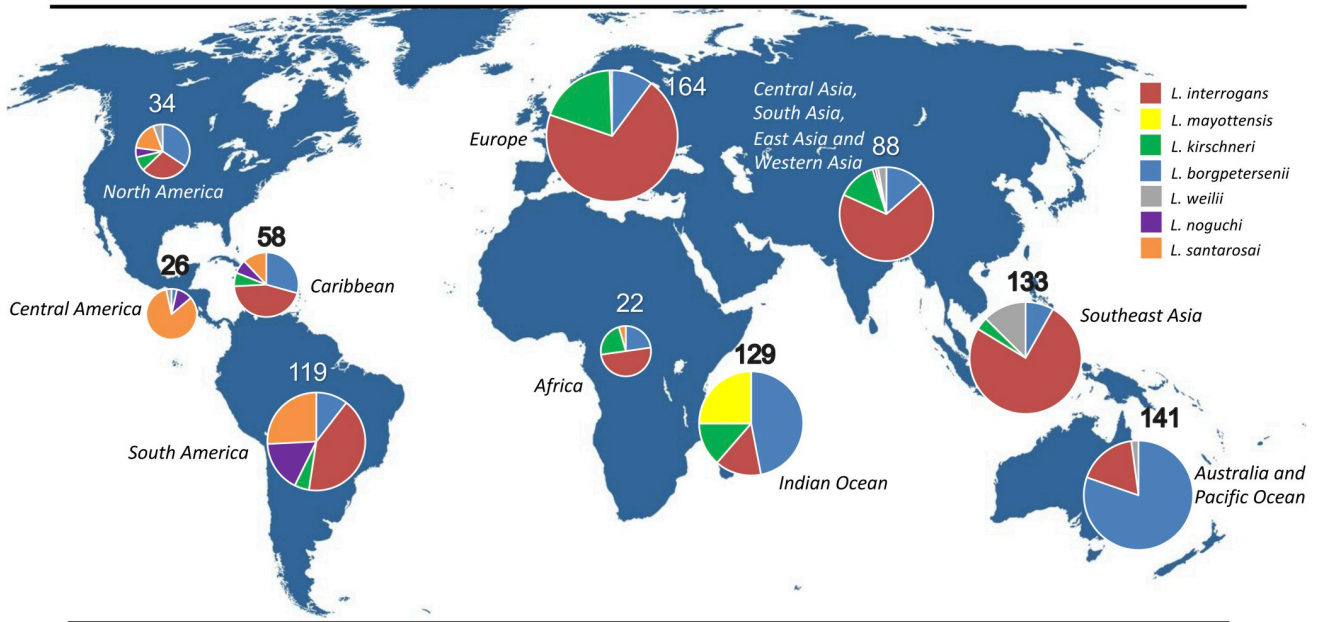


Fig 1. Geographic origins of the most frequent pathogenic *Leptospira* species in our genome database (n = 914). Each pie chart corresponds to a given world region. As shown in our map, *L. santarosai* (n = 64) is mostly found in America (North America, Central America, South America and the Caribbean islands). The base layer of the map is freely available from outline-world-map.com.

[//doi.org/10.1371/journal.pntd.0011733.g001](https://doi.org/10.1371/journal.pntd.0011733.g001)

L. santarosai was reported in Taiwan in East Asia. Serogroup Shermani, presumably belonging to *L. santarosai*, is predominant among patients with severe leptospirosis in Taiwan [50]. However, only one *L. santarosai* strain, strain CCF, has been isolated from a patient with leptospirosis in Taiwan [51] and this strain, for which we do not have the complete genome [52], is no longer available (personal communication of Prof Chih-Wei Yang).

Genome analysis shows species-specific features and an open pangenome for most pathogenic *Leptospira* species

Phylogenetic analysis of the eight pathogenic *Leptospira* species using the saprophyte *L. biflexa* as the outgroup shows two distinct groups as previously shown [8, 53]. One phylogenetic group constituted by *L. santarosai*, *L. mayottensis*, *L. borgpetersenii*, *L. alexanderi* and *L. weilii* (group I), and another one with *L. interrogans*, *L. kirschneri* and *L. noguchii* (group II) (Fig 2A). The genome size and GC content vary widely among pathogenic species usually correlating with these two phylogenetic subgroups (Fig 2B and S3 and S4 Tables). Notably, the genome size of the group I (3.96 ± 0.17 Mb) is significantly smaller to species from the group II (4.63 ± 0.32 Mb; Wilcoxon rank test, $W = 4985$, $p < 2.2e-16$; Fig 2B, left panel). Inversely, guanine+cytosine content (G+C%) is higher in the group I (40.4 ± 0.7 G+C%) than the group II (35.5 ± 0.43 %; Wilcoxon rank test, $W = 207668$, $p < 2.2e-16$; Fig 2B, right panel). Interestingly, with the exclusion of *L. weilii*, the genome sizes were not significantly different between species of the group I (Dunn's test, all pair comparisons with p adjusted >0.05), supporting their characterization as a monophyletic group. However, the G+C% content did differ significantly between *Leptospira* spp. within group I. (Dunn's test, all pair comparisons with p adjusted value ≤ 0.04). Variation was larger within the species of group II, as *L. interrogans* shows a larger genome size than *L. kirschneri* (Dunn's test, p adjusted ≤ 0.0000003), whereas G+C% was significantly different between *L. interrogans* and *L. kirschneri*, and *L. interrogans*

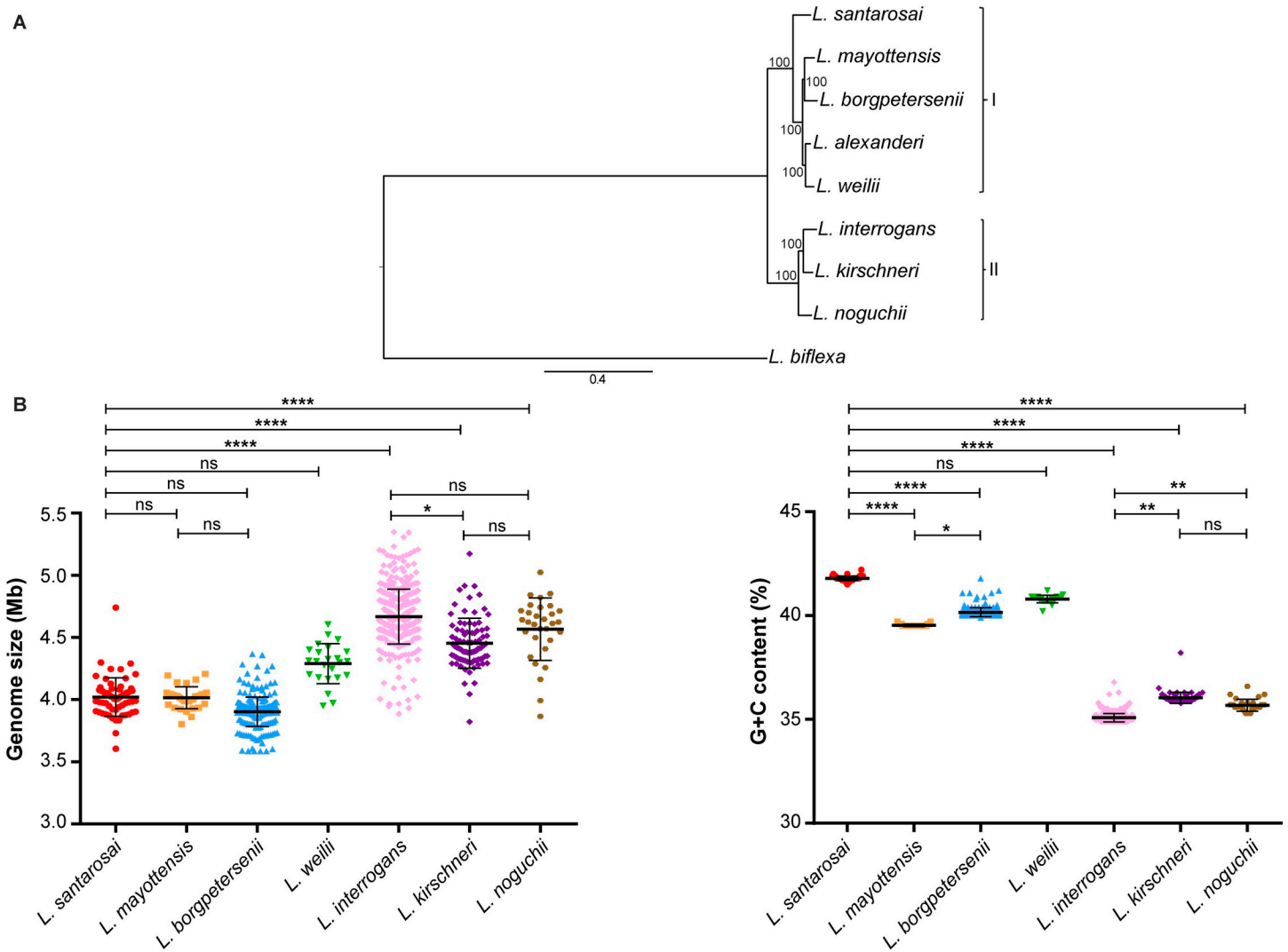


Fig 2. Genome characteristics of pathogenic *Leptospira*. A. Phylogenetic tree built using IQTREE on the alignment of 645 soft-core genes (60% identity cut-off, present in at least 95% of the genomes). The bootstraps values are indicated in the tree (as a percentage). B. The variation of size and G+C% are indicated for *L. interrogans* (n = 410), *L. borgpetersenii* (n = 264), *L. kirschneri* (n = 88), *L. mayottensis* (n = 33), *L. noguchii* (n = 31), *L. santarosai* (n = 64) and *L. weillii* (n = 24); *L. alexanderi*, with only 2 isolates in our database, was not included in this study. Statistical differences between *L. santarosai* and other pathogenic species were determined with the non-parametric ANOVA test (Kruskal-Wallis).

[//doi.org/10.1371/journal.pntd.0011733.g002](https://doi.org/10.1371/journal.pntd.0011733.g002)

and *L. noguchii* (Dunn’s test, all pair comparisons with *p* adjusted value ≤ 0.0002). These disparities in GC content and genome size may be a response to long-term niche adaptation of pathogens which emerged hundred million years ago at the same time as the appearance of mammals [54].

Most pathogenic species exhibit a strong enrichment (>3X) of accessory genes (Fig 3). However, it is important to notice that *L. mayottensis* stands as an exception, where the analysis of 33 strains showed a distribution of 1,949 and 2,284 genes constituting the core and the accessory genomes, respectively (1.2X enrichment), which probably resides in the fact that *L. mayottensis* is restricted to the Indian Ocean (Mayotte and Madagascar) and its specific adaptation to tenrec, described as the main reservoir of this pathogenic species [55,56].

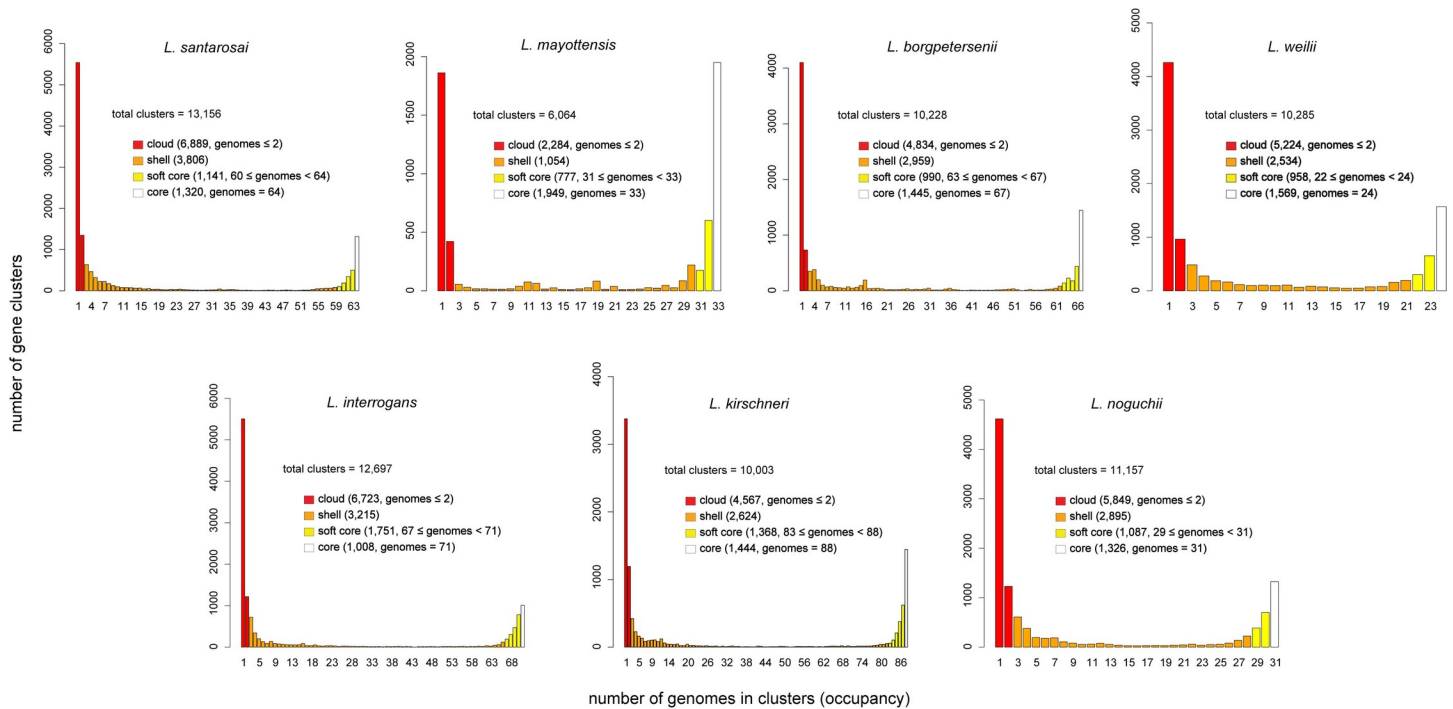


Fig 3. Pangenome distribution in four categories (cloud, shell, soft core and core) for *Leptospira* pathogenic species. Analyses done with GET_HOMOLOGUES showing the U-shaped distribution of pangenome from *L. santarosai*, *L. mayottensis*, *L. borgpetersenii*, *L. weilii*, *L. interrogans*, *L. kirschneri* and *L. noguchii*. For *L. interrogans* and *L. borgpetersenii*, a subset of representative genomes of all sizes and from all geographic locations was selected to reduce computational costs and avoid redundancy.

[//doi.org/10.1371/journal.pntd.0011733.g003](https://doi.org/10.1371/journal.pntd.0011733.g003)

Concerning specifically *L. santarosai*, it harbors an open pangenome suggesting a great diversity in its gene repertoire (S3 Fig). Our analysis shows a strong enrichment ($\approx 5X$) of gene clusters that are unique to a maximum of two genomes (6,889) compared to core genome gene clusters (1,320) (Fig 3).

High genetic diversity of *L. santarosai* strains

To further investigate the genetic diversity of *L. santarosai* isolates, we used a core genome MLST (cgMLST) scheme [1] (Fig 4). The species *L. santarosai* ($n = 64$) were divided into 55 cgMLST clonal groups (cgCGs) showing a high intraspecies genetic diversity (Fig 4) as shown in previous studies [29,32,39,57,58]. Among the 55 cgCGs, none is composed of more than 3 strains (S1 Table), and none is composed of both human and animal strains. We cannot therefore identify transmission of *L. santarosai* clones between different hosts. There is a wide range of possible reservoirs for *L. santarosai* in the Americas. Some countries in the region are among the largest cattle producers in the world so these animals could be important reservoirs for human infections. The Americas also exhibit a great biodiversity, so many species of wild animals such as rodents, marsupials, and domestic animals, such as dogs may be involved in transmission cycles. Among the 15 strains isolated from patients in Costa Rica, only two (id1256 and id1260) exhibit the same clonal group further confirming the great diversity of strains even within one small country.

Unfortunately, analyses to identify associations of *Leptospira* genotypes to particular epidemiological variables (host reservoir, disease outcome, etc.) cannot be performed with our small sample size. However, we could determine some phylogeographic lineages. A clear

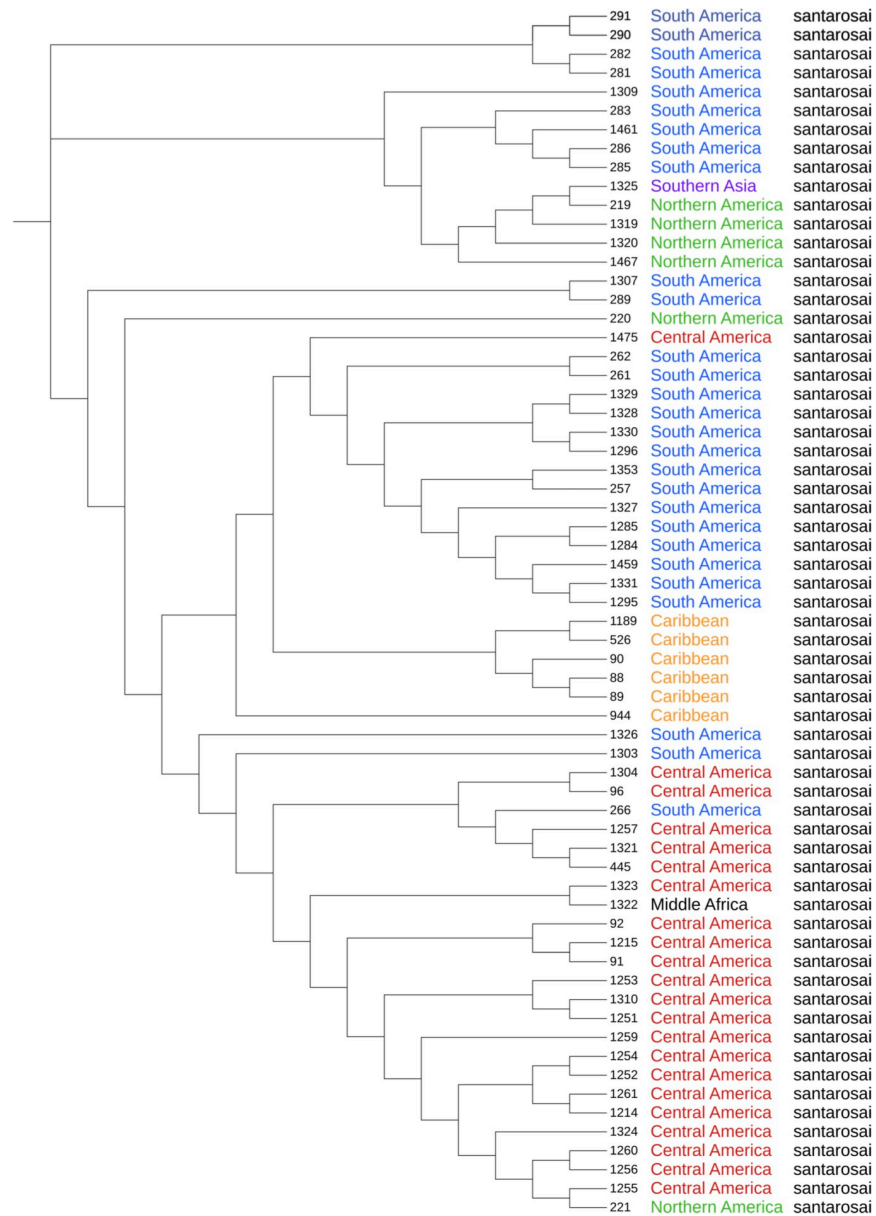


Fig 4. Phylogenetic tree of *L. santarosai* strains. Maximum-likelihood phylogeny based on the variable sites of the cgMLST scheme consisting of 545 core genes showing the distribution of species, serogroups and geographic origins. The species *L. santarosai* (n = 64) were divided into 55 cgMLST clonal groups belonging to 9 different serogroups; for 20 strains the serogroup is unknown or undetermined. Colors indicate strains isolated from the same geographic region (Central America in red, South America in blue, North America in green, Caribbean in orange, Southern Asia in purple, and Middle Africa in black). Branch lengths were not used to ease readability of groups and isolates.

[//doi.org/10.1371/journal.pntd.0011733.g004](https://doi.org/10.1371/journal.pntd.0011733.g004)

geographical separation of the clonal groups was observed for strains from (i) Central America, comprising isolates from Costa Rica (15; all human strains), Panama (6) but also one strain from South America (Colombia); South America which was further divided in two divergent groups, (ii) one containing strains from Brazil (10; mostly bovine strains) and the other (iii) including strains from Peru (9); and (iv) Caribbean islands, with strains from Guadeloupe (2), Martinique (2), Trinidad (1) and Puerto Rico (1) (Fig 4). This suggests ancestral presence of this species in these different countries and further separated evolution with no or low

geographic diffusion. On the contrary, previous phylogenetic analyses of *L. noguchii* [2] and *L. interrogans* [1] did not reveal a correlation of genotype with geographical distribution.

Diversity of serovar and O-antigen-encoding locus in *L. santarosai*

Further identification of *L. santarosai* strains was performed at the level of serovar and serogroup. Serogroups Grippytyphosa (8), Shermani (6), Tarassovi (8), Mini (5), Javanica (3), Pyrogenes (7), Sejroe (3), Hebdomadis (2), and Sarmin (2) were identified by serogrouping, but serovar/serogroup could not be assigned for 20 strains (no agglutination in serogrouping analyses or absence of serotyping information in the genome database). The three strains of the Javanica serogroup have been identified as belonging to a new serovar, Arenal, circulating in patients in Costa Rica [59,60].

We recently showed that the analysis of the gene content of the LPS O-antigen-encoding cluster, or *rfb* cluster, correlates with *Leptospira* serovar and serogroup identity [2]. We then analyzed the gene composition of the *rfb* cluster of our set of *L. santarosai* strains, including 20 strains of unknown serogroup, in comparison to the *rfb* from reference strains of known serovars/serogroups (Fig 5). Strains were organized after hierarchical clustering considering presence/absence of the *rfb* genes. As previously shown, well-typed reference serovars from the same serogroup share the same patterns [2] but serovars from different serogroups may also share a similar genetic fingerprint: this is the case, for example, between serovars Copenhageni, Lai (both from serogroup Icterohaemorrhagiae) and serovar Canicola (serogroup Canicola), or serovars belonging to serogroups Grippytyphosa, Cynopteri and Autumnalis or between serovars from serogroups Mini, Sejroe and Hebdomadis. Four major clusters of genetic fingerprints can be distinguished (Fig 5) and were further confirmed using Jaccard's similarity index (S4 Fig): cluster 1, containing reference strains from serogroups Australis, Grippytyphosa, Cynopteri and Autumnalis; cluster 2, which includes more variability of genetic fingerprints with reference strains from serogroups Pomona, Javanica, Pyrogenes, Celledoni, Icterohaemorrhagiae, Canicola and Sarmin; cluster 3, composed only of reference strains from serogroups Tarassovi and Shermani; and cluster 4, that comprises reference strains from serogroups Sejroe, Hebdomadis and Mini (Fig 5). The 64 *L. santarosai* strains were then assigned to the different cluster according to *rfb* gene composition (S1 Table). Cluster 1 contains all *L. santarosai* strains identified as belonging to serogroup Grippytyphosa (id1259, id1257, id1296, id445, id1328, id1329, id1330, id1253), as well as four strains of unknown serogroups (id220, id1321, id286, id90) showing patterns related to Grippytyphosa. Interestingly, id220, id1321 and id286 were isolated at three different geographic locations: North (id220), Central (id1321) and South America (id286). The three strains are phylogenetically distant, but exhibit an identical *rfb* cluster gene composition, suggesting that the *rfb* genomic island, important for bacterial persistence and adaptation to specific reservoir hosts, has disseminated in unrelated *L. santarosai* strains. Strains from the serogroups Sarmin (id1459, id1461), Javanica (id91, id1310, id92; all from the new serovar Arenal recently described in Costa Rica), Pyrogenes (id1307, id1260, id1475, id1215, id1256, id526, id1189) belong to Cluster 2. Cluster 2 also includes two strains (id285 and id257) from an unknown serogroup exhibiting Sarmin/Pyrogenes/Javanica-like patterns. Cluster 3 comprises all *L. santarosai* strains belonging to serogroups Shermani (id96, id261, id262, id1304, id1252, id1255) and Tarassovi (id266, id289, id1295, id1303, id1467, id1254, id1214, id1261), along with strains of unknown serogroups (id1323, id1331, id282, id1320, id1324) showing patterns related to serogroups Tarassovi and Shermani and one strain (id944) with many *rfb* genes not present in other genomes, suggesting that it may belong to a new serovar/serogroup or a serovar/serogroup not included in our analysis. Finally, strains from cluster 4 belong to the antigenically related serogroups Mini,

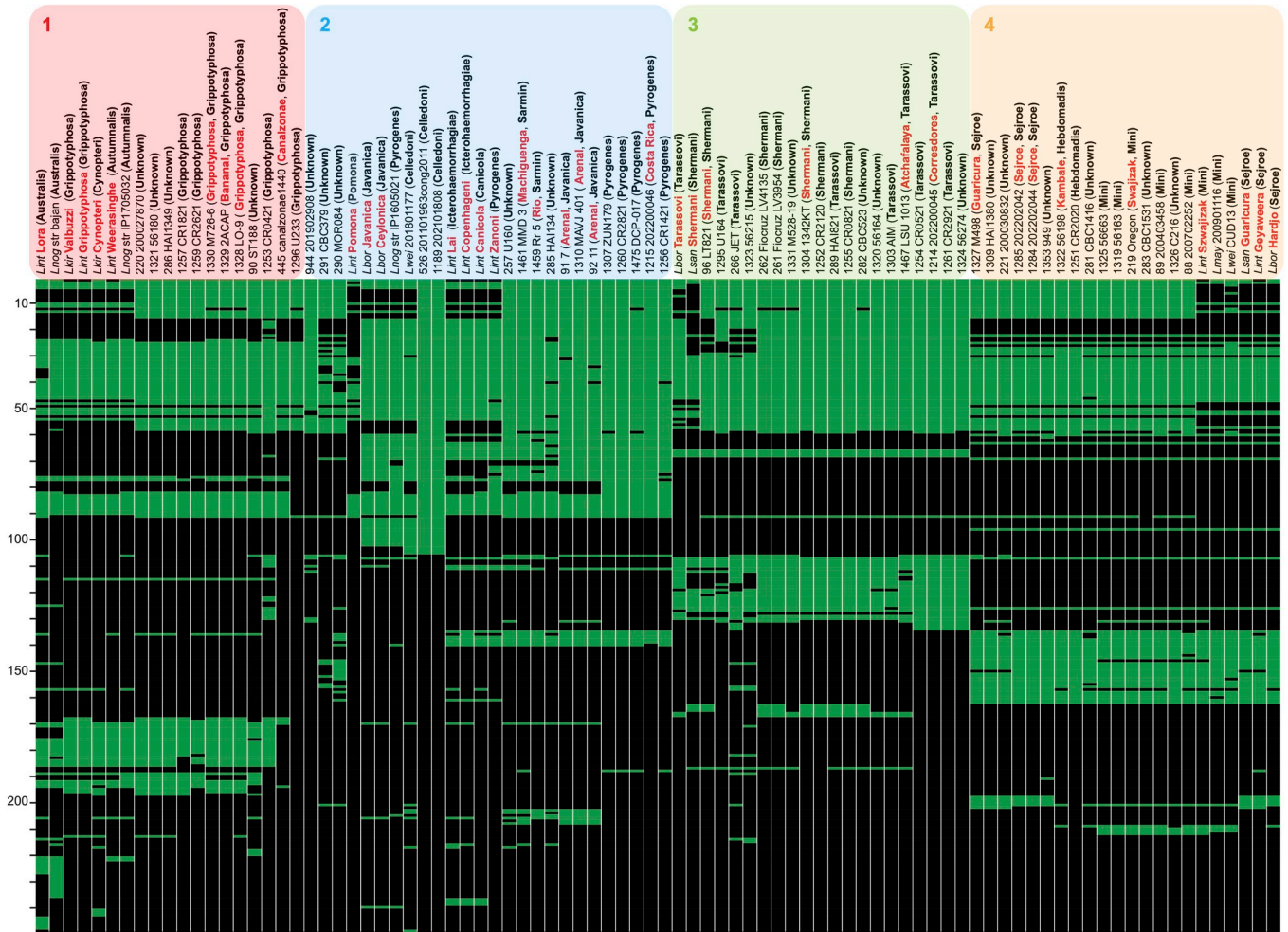


Fig 5. Gene presence/absence matrix of *rfb* clusters from different *Leptospira* strains and species, covering a range of distinct serogroup/serovar identities compared to *L. santarosai* strains. Horizontal lines correspond to individual genes or set of genes grouped according to their percentage of similarity (cut-off 60%), green meaning presence, and black absence. Scales on the left of the matrices indicate the number of different genes being compared. Columns correspond to different *Leptospira* strains as indicated on the columns' labels. Those whose name begins with the species, correspond to the *rfb* cluster of known serovar/serogroup strains (serovar indicated in red, serogroup in brackets and in bold). Names of the *L. santarosai* strains analyzed in the present study begin with their ID number followed by the name of the strain. In brackets, serovar is indicated in red, if assigned, and serogroup in bold. For the latter strains, the whole genome was used and compared against the other *rfb* and a pan-*rfb* reference. Strains were organized after hierarchical clustering considering presence/absence of *rfb* genes. Four major clusters can be evidenced: 1) including serogroups Australis, Grippotyphosa, Cynopteri, and Autumnalis; 2) the most variable, comprising serogroups Tarassovi, Pomona, Javanica, Pyrogenes, Celledoni, Icterohaemorrhagiae, Canicola and Sarmin; 3) composed only of serogroups Tarassovi and Shermani; and 4) including serogroups Sejroe, Hebdomadis and Mini.

[/doi.org/10.1371/journal.pntd.0011733.g005](https://doi.org/10.1371/journal.pntd.0011733.g005)

Hebdomadis and Sejroe which share a similar *rfb* gene composition [3]. *L. santarosai* strains from cluster 4 share similar genetic fingerprints and belong to serogroups Mini (id1325, id1319, id219, id89, id88), Sejroe (id1327, id1285, id1284), and Hebdomadis (id1251, id1322) [12], in addition to strains of unknown serogroups (id1309, id221, id1353, id281, id283, id1326) and strains id290 and id291 with more variability of genetic fingerprints.

Our analysis of the gene composition of the *rfb* cluster is consistent with previous studies showing that *L. santarosai* only contains serovars from serogroups Shermani, Hebdomadis, Tarassovi, Pyrogenes, Autumnalis, Bataviae, Mini, Grippotyphosa, Sejroe, Pomona, Javanica, and Sarmin [49], as well as probable new serovars. However, it must be noted that most of the *L. santarosai* genomic assemblies analyzed here are fragmented (average contig number, 149),

which could lead to inaccurate serotype assignments. Further studies should include less fragmented genomes, ideally closed genomes, for a correct interpretation of the *rfb* patterns and to better identify the genes determining the serovar identity. As previously shown [1] most serogroups had a polyphyletic distribution. Thus, isolates from the serogroups Tarassovi and Grippytyphosa did not all cluster together in the phylogenetic tree based on either cgMLST alleles (Fig 4) or core-genes (S5 Fig).

Conclusion

In conclusion, genome analyses showed species-specific genome size and GC content and an open pangenome in pathogenic species, with the exception of *L. mayottensis*. Taken together, these analyses suggest an ancient speciation of pathogens and their adaptation to diverse niches resulting in a great genotypic and phenotypic diversity across species. We also showed that despite the limited geographic distribution of *L. santarosai* to America, this species exhibits its great diversity and an open pangenome. This study represents the largest and most detailed analysis of the genetic and serotype diversity of this pathogen to date, thus providing a comprehensive analysis of this pathogenic species. Our collection of *L. santarosai* exhibits an over-representation of isolates from America and more genomes representing undersampled regions and different animal reservoirs will be necessary to better understand the evolutionary history, epidemiology, and population dynamics of *L. santarosai*. We discovered a large genetic diversity among isolates from both human and animal samples, with no apparent transmission from one host to another, although circulation of strains that share the same serogroup was evident in multiple hosts. Outbreak investigations performed at the local level would likely improve the identification of animal reservoirs. These results will improve our understanding of the dissemination of genotypes in specific geographic regions and update the knowledge of strains circulating in America for effective disease surveillance.

Supporting information

S1 Table. Geographic origins of the most frequent pathogenic *Leptospira* species in our genome database (n = 914).

(XLSX)

S2 Table. List of *L. santarosai* strains used in this study and genome characteristics.

(XLSX)

S3 Table. Dunn Kruskal-Wallis multiple comparisons results between the G+C% content of *Leptospira* genomes.

(DOCX)

S4 Table. Dunn Kruskal-Wallis multiple comparisons results between the genome size (length) of *Leptospira* genomes.

(DOCX)

S1 Fig. Average Nucleotide Identity (ANI) among *L. santarosai* strains. *L. interrogans* strain Fiocruz L1-130 and *L. borgpetersenii* strain M84 were included as outgroups. The ANI percentages are depicted as colors of the square matrix elements, according to the scale shown in the upper left insert. The names of the 64 *L. santarosai* strains are indicated on the right of the matrix. Clustering shown on the left side (and upper side, symmetric) of the matrix table, was performed by GET_HOMOLOGUES version 20190411.

(PDF)

S2 Fig. Percentage of Conserved Proteins (POCP) among *L.santarosai* strains. *L. interrogans* strain Fiocruz L1-130 and *L. borgpetersenii* strain M84 were included as reference of distinct species. POCP percentages are represented as colors of the square matrix elements, according to the scale shown in the upper left insert. The names of the 64 *L. santarosai* strains are indicated on the right of the matrix. Clustering shown on the left side (and upper side, symmetric) of the matrix table, was performed by GET_HOMOLOGUES version 20190411. (PDF)

S3 Fig. Pangenome analysis of *L.santarosai* strains. The graph was obtained with Roary 3.11.2 (yielding a total of 13,168 gene clusters). The pangenome of *L. santarosai* presents an open profile, which was further verified by Heap's law [3], $n = \kappa N^\gamma$; considering a total of 13,168 genes (n) in the pangenome (according to Roary) and the 64 genomes (N) included, the observed curve allows for non-linear fitting with a constant $\kappa = 2851.4$. Thus, $1 - \gamma = \alpha = 0.63$. An α value < 1 indicates an open profile. Same result was obtained considering $n = 13,156$ (GET_HOMOLOGUES). (PDF)

S4 Fig. Jaccard similarity matrix for the comparative analyses of the *rfb* clusters. Jaccard similarity matrix is organized according to the gene presence/absence matrix of *rfb* clusters shown in Fig 5. To perform the Jaccard similarity index calculation, each strain (column) was converted into a vector of 0 (gene absence) and 1 (gene presence), and then used to do pairwise comparisons between the vectors using a Python pipeline via NumPy. Jaccard similarity is represented in colors, according to the scale shown in the lower right insert. (PDF)

S5 Fig. Core- genome based phylogeny of *L.santarosai*. Phylogenetic tree based on the sequences of 1288 core-genes of *L. santarosai*. A core-gene alignment was obtained by Roary (60% identity cut-off), and then used to perform the phylogeny. The best-fit model and the maximum-likelihood phylogenetic tree were determined by IQ-TREE version 1.6.11 [5], considering 10,000 ultrafast bootstraps [6]. *L. interrogans* strain Fiocruz L1-130 and *L. borgpetersenii* strain M84 were included as outgroups. The serogroup of each strain is indicated in parenthesis, as well as the world region (AF = Africa; CA = Central America; CB = Caribbean; NA = North America; SA = South America; SAs = South Asia). Bootstraps values other than 100% are shown. (PDF)

Acknowledgments

We thank Vincent Enouf and the team of the core facility P2M (Institut Pasteur, Mutualized Platform for Microbiology) for genomic sequencing and the staff of the French National Reference Center for Leptospirosis for processing some of the samples. We thank Federica Palma for her help with BIGSdb. We thank Dr Howard Takiff and Lizeth Caraballo (Laboratorio de Genética Molecular, CMBC, Instituto Venezolano de Investigaciones Científica, Caracas, Venezuela), Anissa Amara Korba (Institut Pasteur of Alger, Algiers, Algeria), Prof Walter Lilembaum (Universidade Federal Fluminense, Rio de Janeiro, Brazil), Sabrina Rodrigues Campos (Institute of Veterinary Bacteriology, Bern, Switzerland), Dr Paulina Meny (Instituto de Higiene, Facultad de Medicina, Montevideo, Uruguay), Dr Josipa Habus and Dr Nenad Turk (Faculty of Veterinary Medicine, University of Zagreb, Croatia), Prof Jackie Benschop (Molecular Epidemiology and Public Health Laboratory, School of Veterinary Sciences, Massey University, New Zealand), Dr Zarantonelli (Unidad Mixta Pasteur-Instituto Nacional de

Investigación Agropecuaria, Institut Pasteur de Montevideo), Dr Vasantha Kumari Neela (Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Malaysia), Dr Mohammad Ridhuan Mohd Ali (Institute for Medical Research, Malaysia), Dr Shlomo Blum (Kimron Veterinary Institute, Israel), Dr Nikolai Tokarevich (Laboratory of Zoonoses, Pasteur Institute in Saint Petersburg, Saint Petersburg, Russia), Dr Angelica Delgado (National Reference Laboratory of Bacterial Zoonoses, Lima, Peru), Dr Marga Goris (National Collaborating Centre for Reference and Research on Leptospirosis, Academic Medical Center, Amsterdam, the Netherlands) and Prof Chih-Wei Yang (Chang Gung University, Taiwan) for sharing strains and/or information on some of the strains used in our study.

Author Contributions

Conceptualization: Ricardo Gutiérrez, Frédéric J. Veyrier.

Data curation: Cecilia Nieves, Ricardo Gutiérrez, Vallier Sordoillet, Mathieu Picardeau.

Formal analysis: Cecilia Nieves, Ricardo Gutiérrez, Vallier Sordoillet.

Funding acquisition: Frédéric J. Veyrier, Mathieu Picardeau.

Investigation: Diana Chinchilla, Cecilia Nieves, Frédéric J. Veyrier, Mathieu Picardeau.

Methodology: Diana Chinchilla, Cecilia Nieves, Frédéric J. Veyrier.

Software: Frédéric J. Veyrier.

Supervision: Mathieu Picardeau.

Validation: Diana Chinchilla, Frédéric J. Veyrier, Mathieu Picardeau.

Visualization: Frédéric J. Veyrier.

Writing – original draft: Mathieu Picardeau.

Writing – review & editing: Diana Chinchilla, Cecilia Nieves, Ricardo Gutiérrez, Frédéric J. Veyrier, Mathieu Picardeau.

References

1. Guglielmini J, Bourhy P, Schiettekatte O, Zinini F, Brisse S, Picardeau M. Genus-wide *Leptospira* core genome multilocus sequence typing for strain taxonomy and global surveillance. *PLoS Negl Trop Dis*. 2019; 13:e0007374. [//doi.org/10.1371/journal.pntd.0007374](https://doi.org/10.1371/journal.pntd.0007374) PMID: 31026256
2. Nieves C, Vincent AT, Zarantonelli L, Picardeau M, Veyrier FJ, Buschiazio A. Horizontal transfer of the rfb cluster in *Leptospira* is a genetic determinant of serovar identity. *Life Science Alliance*. 2022; 6:e202201480. [//doi.org/10.26508/lsa.202201480](https://doi.org/10.26508/lsa.202201480) PMID: 36622346
3. Medeiros EJS, Ferreira LCA, Ortega JM, Cosate MRV, Sakamoto T. Genetic basis underlying the serological affinity of leptospiral serovars from serogroups Sejroe, Mini and Hebdomadis. *Infect Genet Evol* 2022; 103:105345. [//doi.org/10.1016/j.meegid.2022.105345](https://doi.org/10.1016/j.meegid.2022.105345) PMID: 35917899
4. Fernandes LGV, Stone NE, Roe CC, Goris MGA, van der Linden H, Sahl JW, et al. *Leptospira sanjuanensis* sp. nov., a pathogenic species of the genus *Leptospira* isolated from soil in Puerto Rico. *Int J Syst Evol Microbiol*. 2022; 72:005560.
5. Chakraborty A, Miyahara S, Villanueva SY, Saito M, Gloriani NG, Yoshida S. A novel combination of selective agents for isolation of *Leptospira* species. *Microbiol Immunol*. 2011; 55:494–501. [//doi.org/10.1111/j.1348-0421.2011.00347.x](https://doi.org/10.1111/j.1348-0421.2011.00347.x) PMID: 21545510
6. Thibeaux R, Girault D, Bierque E, Soupé-Gilbert ME, Rettinger A, Douyère A, et al. Biodiversity of Environmental *Leptospira*: Improving Identification and Revisiting the Diagnosis. *Front Microbiol* 2018; 9:816. [//doi.org/10.3389/fmicb.2018.00816](https://doi.org/10.3389/fmicb.2018.00816) PMID: 29765361
7. Land M, Hauser L, Jun SR, Nookaew I, Leuze MR, Ahn TH, et al. Insights from 20 years of bacterial genome sequencing. *Funct Integr Genomics*. 2015; 15:141–61. [//doi.org/10.1007/s10142-015-0433-4](https://doi.org/10.1007/s10142-015-0433-4) PMID: 25722247

8. Vincent AT, Schiettekatte S, Goarant C, Neela VK, Bernet E, Thibeaux R, et al. Revisiting the taxonomy and evolution of pathogenicity of the genus *Leptospira* through the prism of genomics. *PLoS Negl Trop Dis*. 2019; 13:e0007270. [//doi.org/10.1371/journal.pntd.0007270](https://doi.org/10.1371/journal.pntd.0007270) PMID: 31120895
9. Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, et al. Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLoS Negl Trop Dis* 2015; 9:e0003898. [//doi.org/10.1371/journal.pntd.0003898](https://doi.org/10.1371/journal.pntd.0003898) PMID: 26379143
10. Noguera LPZ, Charypkhan D, Hartnack S, Torgerson PR, Rügge SR. The dual burden of animal and human zoonoses: A systematic review. *PLoS Negl Trop Dis*. 2022; 16:e0010540.
11. Yasuda PH, Steigerwalt AG, Sulzer KR, Kaufmann AF, Rogers FC, Brenner DJ. Deoxyribonucleic acid relatedness between serogroups and serovars in the family Leptospiraceae with proposals for seven new *Leptospira* species. *Int J Syst Bacteriol*. 1987; 37:407–15.
12. Carvajal M.P, Fagerstrom KA. Epidemiology of Leptospirosis in Costa Rica 2011–2015. *Curr Trop Med Rep* 2017; 4: 41–46. [//doi.org/10.1007/s40475-017-0102-x](https://doi.org/10.1007/s40475-017-0102-x) PMID: 28603685
13. Grillová L, Robinson MT, Chanthongthip A, Vincent AT, Nieves C, Oppelt J, et al. Genetic diversity of *Leptospira* isolates in Lao PDR and genome analysis of an outbreak strain. *PLoS Negl Trop Dis*. 2021; 15:e0010076. [//doi.org/10.1371/journal.pntd.0010076](https://doi.org/10.1371/journal.pntd.0010076) PMID: 34962921
14. Noda AA, Grillová L, Mariet JF, Paiffer NB, Ruiz YC, Rodríguez I, et al. A first insight into the genomic diversity of *Leptospira* strains isolated from patients in Cuba. *PLoS ONE*. 2020; 15:e0229673. [//doi.org/10.1371/journal.pone.0229673](https://doi.org/10.1371/journal.pone.0229673) PMID: 32106244
15. Zarantonelli L, Suanes A, Meny P, Buroni F, Nieves C, Salaberry X, et al. Isolation of pathogenic *Leptospira* strains from naturally infected cattle in Uruguay reveals high serovar diversity, and uncovers a relevant risk for human leptospirosis. *PLoS Negl Trop Dis*. 2018; 12:e0006694. [//doi.org/10.1371/journal.pntd.0006694](https://doi.org/10.1371/journal.pntd.0006694) PMID: 30212451
16. Page AJ, Cummins CA, Hunt M, Wong VK, Reuter S, Holden MTG, et al. Roary: rapid large-scale prokaryote pan genome analysis *Bioinformatics*. 2015; 31:3691–83.
17. Tettelin H, Riley D, Cattuto C, Medini D. Comparative genomics: the bacterial pan-genome. *Curr Opin Microbiol*. 2008; 11(5):472–7. Epub 2008/12/18. [//doi.org/10.1016/j.mib.2008.09.006](https://doi.org/10.1016/j.mib.2008.09.006) PMID: 19086349.
18. Tanizawa Y, Fujisawa T, Nakamura Y. DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics*. 2018; 34:1037–9. [//doi.org/10.1093/bioinformatics/btx713](https://doi.org/10.1093/bioinformatics/btx713) PMID: 29106469
19. RCoreTeam. R: A Language and Environment for Statistical Computing, R Foundation for Statistical Computing. [urldefense.com/v3/___/www.R-project.org/_;!!JFdNOqOXpB6UZW0!vWw9-j2tv0GgtJ6F_apAzEF2CZFpPiAxxw1-s6WRRDLaLLwEEzeK3dvrB33h9kG7X5dCRLI7GYTOuBmdxFBzN4Nz-mc0H\\$a](https://urldefense.com/v3/___/www.R-project.org/_;!!JFdNOqOXpB6UZW0!vWw9-j2tv0GgtJ6F_apAzEF2CZFpPiAxxw1-s6WRRDLaLLwEEzeK3dvrB33h9kG7X5dCRLI7GYTOuBmdxFBzN4Nz-mc0H$a), editor. 2016.
20. Ogle DH, Doll JC, Wheeler AP, Dinno A. FSA: Simple Fisheries Stock Assessment Methods. R package version 0.9.4. 2023; urldefense.com/v3/___/CRAN.R-project.org/package=FSA_;!!JFdNOqOXpB6UZW0!vWw9-j2tv0GgtJ6F_apAzEF2CZFpPiAxxw1-s6WRRDLaLLwEEzeK3dvrB33h9kG7X5dCRLI7GYTOuBmdxFBzN4DoX66kY
21. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014; 15:2068–9. [//doi.org/10.1093/bioinformatics/btu153](https://doi.org/10.1093/bioinformatics/btu153) PMID: 24642063
22. Nguyen LT, Schmidt HA, vonHaeseler A, Minh BQ. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* 2015; 32:268–74. [//doi.org/10.1093/molbev/msu300](https://doi.org/10.1093/molbev/msu300) PMID: 25371430
23. Hoang DT, Chernomor O, von Haeseler A, Minh BQ, Vinh LS. UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Mol Biol Evol*. 2018; 35:518–22. [//doi.org/10.1093/molbev/msx281](https://doi.org/10.1093/molbev/msx281) PMID: 29077904
24. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, et al. BLAST+: architecture and applications. *BMC Bioinformatics* 2009; 10:421. [//doi.org/10.1186/1471-2105-10-421](https://doi.org/10.1186/1471-2105-10-421) PMID: 20003500
25. Hagberg AA, Schult DA, Swart P, editors. Exploring Network Structure, Dynamics, and Function using NetworkX. Proceedings of the 7th Python in Science Conference 2018.
26. Schneider MC, Leonel DG, Hamrick PN, de Caldas EP, Velásquez RT, Mendigaña Paez FA, et al. Leptospirosis in Latin America: exploring the first set of regional data. *Rev Panam Salud Publica*. 2017; 41 (e81). [//doi.org/10.26633/RPSP.2017.81](https://doi.org/10.26633/RPSP.2017.81) PMID: 31384245
27. Hotez PJ, Bottazzi ME, Franco-Paredes C, Ault SK, Periago MR. The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. *PLoS Negl Trop Dis* 2008; 2:e300. [//doi.org/10.1371/journal.pntd.0000300](https://doi.org/10.1371/journal.pntd.0000300) PMID: 18820747

28. Peters A, Vokaty A, Portch R, Gebre Y. Leptospirosis in the Caribbean: a literature review. *Rev Panam Salud Publica* 2017; 41:e166. [//doi.org/10.26633/RPSP.2017.166](https://doi.org/10.26633/RPSP.2017.166) PMID: 31384278
29. Munoz-Zanzi C, Groene E, Morawski BM, Bonner K, Costa F, Bertherat E, et al. A systematic literature review of leptospirosis outbreaks worldwide, 1970–2012. *Rev Panam Salud Publica*. 2020; 44:e78. [//doi.org/10.26633/RPSP.2020.78](https://doi.org/10.26633/RPSP.2020.78) PMID: 32684917
30. Pereira MM, Schneider MC, Munoz-Zanzi C, Costa F, Benschop J, Hartskeerl RA, et al. A road map for leptospirosis research and health policies based on country needs in Latin America. *Rev Panam Salud Publica*. 2018; 41:e131. [//doi.org/10.26633/RPSP.2017.131](https://doi.org/10.26633/RPSP.2017.131) PMID: 29466519
31. Delgado MA, Cáceres OA, Calderón JE, Balda L, Sotil G, Céspedes MJ. New Genetic Variants of *Leptospira* spp Characterized by MLST from Peruvian Isolates. *J Trop Med*. 2022; 2022:4184326. [//doi.org/10.1155/2022/4184326](https://doi.org/10.1155/2022/4184326) PMID: 36249734
32. Peláez Sánchez RG, Quintero JÁL, Pereira MM, Agudelo-Flórez P. High-Resolution Melting Curve Analysis of the 16S Ribosomal Gene to Detect and Identify Pathogenic and Saprophytic *Leptospira* Species in Colombian Isolates. *Am J Trop Med Hyg* 2017; 96:1031–8. [//doi.org/10.4269/ajtmh.16-0312](https://doi.org/10.4269/ajtmh.16-0312) PMID: 28500802
33. Ganoza CA, Matthias MA, Collins-Richards D, Brouwer KC, Cunningham CB, Segura ER, et al. Determining risk for severe leptospirosis by molecular analysis of environmental surface waters for pathogenic *Leptospira*. *PLoS Med*. 2006; 3:e308. [//doi.org/10.1371/journal.pmed.0030308](https://doi.org/10.1371/journal.pmed.0030308) PMID: 16933963
34. Miotto BA, Moreno LZ, Guilloux AGA, Sousa GO, Loureiro AP, Moreno AM, et al. Molecular and serological characterization of the first *Leptospira santarosai* strain isolated from a dog. *Acta Trop*. 2016; 162:1–4. [//doi.org/10.1016/j.actatropica.2016.06.007](https://doi.org/10.1016/j.actatropica.2016.06.007) PMID: 27282095
35. Kremer FS, Eslabão MR, Provisor M, Woloski RD, Ramires OV, Moreno LZ, et al. Draft Genome Sequences of *Leptospira santarosai* Strains U160, U164, and U233, Isolated from Asymptomatic Cattle. *Genome Announc*. 2015; 3:e00910–15. [//doi.org/10.1128/genomeA.00910-15](https://doi.org/10.1128/genomeA.00910-15) PMID: 26272577
36. Lilienbaum W, Kremer FS, Ristow P, Dellagostin OA, Bourhy P, Hartskeerl RA, et al. Molecular characterization of the first leptospires isolated from goats in Brazil. *Braz J Microbiol* 2015; 45:1527–30.
37. Moreno LZ, Miraglia F, Marvulo MF, Silva JC, Paula CD, Costa BL, et al. Characterization of *Leptospira santarosai* Serogroup Grippotyphosa Serovar Bananal Isolated from Capybara (*Hydrochaeris hydrochaeris*) in Brazil. *J Wildl Dis*. 2016; 52:688–93. [//doi.org/10.7589/2015-09-245](https://doi.org/10.7589/2015-09-245) PMID: 27258408
38. Kallel H, Bourhy P, Mayence C, Houcke S, Hommel D, Picardeau M, et al. First report of human *Leptospira santarosai* infection in French Guiana. *J Infect Public Health*. 2020; 13:1181–3. [//doi.org/10.1016/j.jiph.2020.03.020](https://doi.org/10.1016/j.jiph.2020.03.020) PMID: 32307314
39. Bourhy P, Herrmann-Storck C, Theodose R, Olive C, Nicolas M, Hochedez P, et al. Serovar diversity of pathogenic *Leptospira* circulating in the French West Indies. *PLoS Negl Trop Dis*. 2013; 7:e2114. [//doi.org/10.1371/journal.pntd.0002114](https://doi.org/10.1371/journal.pntd.0002114) PMID: 23516654
40. Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu GX, et al. Actionable Diagnosis of Neuroleptospirosis by Next-Generation Sequencing. *N Engl J Med*. 2014; 370:2408–17. [//doi.org/10.1056/NEJMoa1401268](https://doi.org/10.1056/NEJMoa1401268) PMID: 24896819
41. Brenner DJ, Kaufmann AF, Sulzer KR, Steigerwalt AG, Rogers FC, Weyant RS. Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for *Leptospira alexanderi* sp. nov. and four new *Leptospira* genomospecies. *Int J Syst Bacteriol*. 1999; 49:839–58. [//doi.org/10.1099/00207713-49-2-839](https://doi.org/10.1099/00207713-49-2-839) PMID: 10319510
42. Naotunna C, Agampodi SB, Agampodi TC. Etiological agents causing leptospirosis in Sri Lanka: A review. *Asian Pac J Trop Med*. 2016; 9:390–4. [//doi.org/10.1016/j.apjtm.2016.03.009](https://doi.org/10.1016/j.apjtm.2016.03.009) PMID: 27086159
43. Jayasundara D, Senavirathna I, Warnasekara J, Gamage C, Siribaddana S, Kularatne SAM, et al. 12 Novel clonal groups of *Leptospira* infecting humans in multiple contrasting epidemiological contexts in Sri Lanka. *PLoS Negl Trop Dis*. 2021; 15:e0009272.
44. Lata KS, Vaghasia V, Bhairappanavar SB, Kumar S, Ayachit G, Patel S, et al. Whole genome sequencing and de novo assembly of three virulent Indian isolates of *Leptospira*. *Infect Genet Evol*. 2020; 85:104579. [//doi.org/10.1016/j.meegid.2020.104579](https://doi.org/10.1016/j.meegid.2020.104579) PMID: 33017688
45. Oni O, Sujit K, Kasemsuwan S, Sakpuaram T, Pfeiffer DU. Seroprevalence of leptospirosis in domesticated Asian elephants (*Elephas maximus*) in north and west Thailand in 2004. *Vet Rec* 2007; 160:368–71. [//doi.org/10.1136/vr.160.11.368](https://doi.org/10.1136/vr.160.11.368) PMID: 17369477
46. Suwancharoen D, Chaisakdanugull Y, Thanapongtharm W, Yoshida S. Serological survey of leptospirosis in livestock in Thailand. *Epidemiol Infect*. 2013; 141:2269–77. [//doi.org/10.1017/S0950268812002981](https://doi.org/10.1017/S0950268812002981) PMID: 23308397
47. Alamuri A, Thirumalesh SRA, Kumari SS, Kumar KV, Roy P, Balamurugan V. Seroprevalence and distribution of serogroup-specific pathogenic *Leptospira* antibodies in cattle and buffaloes in the state of

- Andhra Pradesh, India. *Vet World* 2019; 12:1212–7. [//doi.org/10.14202/vetworld.2019.1212-1217](https://doi.org/10.14202/vetworld.2019.1212-1217) PMID: 31641299
48. Suut L, Mazlan MN, Arif MT, Yusoff H, Abdul RNA, Safii R, et al. Serological Prevalence of Leptospirosis Among Rural Communities in the Rejang Basin, Sarawak, Malaysia. *Asia Pac J Public Health*. 2016; 28:450–7. [//doi.org/10.1177/1010539516648003](https://doi.org/10.1177/1010539516648003) PMID: 27183976
 49. Levett PN. Leptospirosis. *Clin Microbiol Rev*. 2001; 14:296–326.
 50. Wang HK, Lee MH, Chen YC, Hsueh PR, Chang SC. Factors associated with severity and mortality in patients with confirmed leptospirosis at a regional hospital in northern Taiwan. *J Microbiol Immunol Infect* 2020; 53:307–14. [//doi.org/10.1016/j.jmii.2018.05.005](https://doi.org/10.1016/j.jmii.2018.05.005) PMID: 29934034
 51. Hsieh WJ, Chang YF, Chen CS, Pan MJ. *Omp52* is a growth-phase-regulated outer membrane protein of *Leptospira santarosai* serovar Shermani. *FEMS Microbiol Lett*. 2005; 243:339–45. [//doi.org/10.1016/j.femsle.2004.12.021](https://doi.org/10.1016/j.femsle.2004.12.021) PMID: 15686833
 52. Chou LF, Chen TW, Ko YC, Pan MJ, Tian YC, Chiu CH, et al. Potential impact on kidney infection: a whole-genome analysis of *Leptospira santarosai* serovar Shermani. *Emerg Microbes Infect* 2014; 3:e82. [//doi.org/10.1038/emi.2014.78](https://doi.org/10.1038/emi.2014.78) PMID: 26038504
 53. Xu Y, Zhu Y, Wang Y, Chang YF, Zhang Y, Jiang X, et al. Whole genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic *Leptospira*. *Sci Rep*. 2016; 6:20020. [//doi.org/10.1038/srep20020](https://doi.org/10.1038/srep20020) PMID: 26833181
 54. Kurilung A, Keeratipusana C, Horiike T, Suriyaphol P, Hampson DJ, Prapasarakul N. Chronology of emergence of the genus *Leptospira* and over-representation of gene families enriched by vitamin B2, B12 biosynthesis, cell adhesion and external encapsulating structure in *L. interrogans* isolates from asymptomatic dogs. *Infect Genet Evol* 2019; 73:7–12. [//doi.org/10.1016/j.meegid.2019.04.005](https://doi.org/10.1016/j.meegid.2019.04.005) PMID: 30974265
 55. Lagadec E, Gomard Y, Le Minter G, Cordonin C, Cardinale E, Ramasindrazana B, et al. Identification of *Tenrec ecaudatus*, a Wild Mammal Introduced to Mayotte Island, as a Reservoir of the Newly Identified Human Pathogenic *Leptospira mayottensis*. *PLoS Negl Trop Dis* 2016; 10:e0004933. [//doi.org/10.1371/journal.pntd.0004933](https://doi.org/10.1371/journal.pntd.0004933) PMID: 27574792
 56. Bourhy P, Collet L, Brisse S, Picardeau M. *Leptospira mayottensis* sp. nov., a pathogenic species of the genus *Leptospira* isolated from humans. *Int J Syst Evol Microbiol*. 2014; 64:4061–7. [//doi.org/10.1099/ijs.0.066597-0](https://doi.org/10.1099/ijs.0.066597-0) PMID: 25249563
 57. Hamond C, Pinna M, Medeiros MA, Bourhy P, Lilenbaum W, Picardeau M. A multilocus variable number tandem repeat analysis assay provides high discrimination for genotyping *Leptospira santarosai* strains. *J Med Microbiol*. 2015; 64:507–12. [//doi.org/10.1099/jmm.0.000045](https://doi.org/10.1099/jmm.0.000045) PMID: 25721051
 58. Jaeger LH, Pestana CP, Correia L, Carvalho-Costa FA, Medeiros MA, Lilenbaum W. Novel MLST sequence types of pathogenic *Leptospira* spp.: Opening the black box of animal leptospirosis in Brazil. *Acta Trop*. 2019; 196:135–41. [//doi.org/10.1016/j.actatropica.2019.05.025](https://doi.org/10.1016/j.actatropica.2019.05.025) PMID: 31121146
 59. Valverde MA, Ramírez JM, Montes de Oca LG, Goris MG, Ahmed N, Hartskeerl RA. Arenal, a new *Leptospira* serovar of serogroup Javanica, isolated from a patient in Costa Rica. *Infect Genet Evol* 2008; 8:529–33. [//doi.org/10.1016/j.meegid.2008.02.008](https://doi.org/10.1016/j.meegid.2008.02.008) PMID: 18417425
 60. Valverde MA, Goris MGA, González V, Anchia ME, Ahmed A, Hartskeerl RA. New serovars of *Leptospira* isolated from patients in Costa Rica: implications for public health. *J Med Microbiol*. 2013; 62:1263–71. [//doi.org/10.1099/jmm.0.058545-0](https://doi.org/10.1099/jmm.0.058545-0) PMID: 23788592