



**HAL**  
open science

# Core genome sequencing and genotyping of *Leptospira interrogans* in clinical samples by target capture sequencing

Linda Grillova, Thomas Cokelaer, Jean-François Mariet, Juliana Pipoli da Fonseca, Mathieu Picardeau

► **To cite this version:**

Linda Grillova, Thomas Cokelaer, Jean-François Mariet, Juliana Pipoli da Fonseca, Mathieu Picardeau. Core genome sequencing and genotyping of *Leptospira interrogans* in clinical samples by target capture sequencing. *BMC Infectious Diseases*, 2023, 23 (1), pp.157. 10.1186/s12879-023-08126-x . pasteur-04120601

**HAL Id: pasteur-04120601**

**<https://pasteur.hal.science/pasteur-04120601>**

Submitted on 7 Jun 2023

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH

Open Access



# Core genome sequencing and genotyping of *Leptospira interrogans* in clinical samples by target capture sequencing

Linda Grillova<sup>1,4†</sup>, Thomas Cokelaer<sup>2,3†</sup>, Jean-François Mariet<sup>1</sup>, Juliana Pipoli da Fonseca<sup>2</sup> and Mathieu Picardeau<sup>1\*</sup>

## Abstract

**Background** The life-threatening pathogen *Leptospira interrogans* is the most common agent of leptospirosis, an emerging zoonotic disease. However, little is known about the strains that are currently circulating worldwide due to the fastidious nature of the bacteria and the difficulty to isolate cultures. In addition, the paucity of bacteria in blood and other clinical samples has proven to be a considerable challenge for directly genotyping the agent of leptospirosis directly from patient material. Our understanding of the genetic diversity of strains during human infection is therefore limited.

**Methods** Here, we carried out hybridization capture followed by Illumina sequencing of the core genome directly from 20 clinical samples that were PCR positive for pathogenic *Leptospira* to elucidate the genetic diversity of currently circulating *Leptospira* strains in mainland France.

**Results** Capture with RNA probes covering the *L. interrogans* core genome resulted in a 72 to 13,000-fold increase in pathogen reads relative to standard sequencing without capture. Variant analysis of the genomes sequenced from the biological samples using 273 *Leptospira* reference genomes was then carried out to determine the genotype of the infecting strain. For samples with sufficient coverage (19/20 samples with coverage > 8x), we could unambiguously identify *L. interrogans* serovars Icterohaemorrhagiae and Copenhageni (14 samples), *L. kirschneri* serovar Grippotyphosa (4 samples), and *L. interrogans* serovar Pyrogenes (1 sample) as the infecting strains.

**Conclusions** We obtained high-quality genomic data with suitable coverage for confident core genome genotyping of the agent of leptospirosis for most of our clinical samples. The recovery of the genome of the serovars Icterohaemorrhagiae and Copenhageni directly from multiple clinical samples revealed low adaptive diversification of the core genes during human infection. The ability to generate culture-free genomic data opens new opportunities for better understanding of the epidemiology of this fastidious pathogen and pathogenesis of this neglected disease.

**Keywords** Leptospirosis, *Leptospira*, Genotyping, Genomics

<sup>†</sup>Linda Grillova and Thomas Cokelaer contributed equally to this work

\*Correspondence:

Mathieu Picardeau  
mpicard@pasteur.fr

Full list of author information is available at the end of the article



© The Author(s) 2023. **Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>. The Creative Commons Public Domain Dedication waiver (<http://creativecommons.org/publicdomain/zero/1.0/>) applies to the data made available in this article, unless otherwise stated in a credit line to the data.

## Introduction

Leptospirosis is a globally distributed zoonosis responsible for more than one million severe cases and 60,000 deaths per year, with the highest incidence in tropical countries [1]. The agent of leptospirosis belongs to the genus *Leptospira*, which is composed of 68 species and more than 300 serovars [2, 3]. The strains responsible for leptospirosis in humans or animals belong to one of the eight pathogenic *Leptospira* species described to date. Among them, *L. interrogans* is the most frequently encountered worldwide [4] and several studies have shown that strains belonging to the Icterohaemorrhagiae serogroup (*L. interrogans* serogroup Icterohaemorrhagiae), of which the main reservoir is the rat, are responsible for the most severe forms of the disease [5–8].

Pathogenic leptospires are slow-growing bacteria that require a rich culture medium susceptible to contamination by other organisms. Isolation from biological samples is therefore tedious, especially as the bacteria can be present in low concentrations in blood and urine. During the course of infection, the bacteria are present in the blood during the first week after the onset of symptoms, with their concentration determined by qPCR ranging from  $10^2$  to  $10^6$  *Leptospira*/mL for the peak of leptospiremia [9, 10]. A decreasing number of leptospires are then found in the blood 6–7 days after the onset of symptoms until *Leptospira* nucleic acid is no longer detectable [10, 11]. *Leptospira* can also be detected in urine after symptom onset for a longer duration than blood. However, the bacteria may not be consistently present in the urine during the infection and both their concentration and the duration of their excretion are poorly defined.

The identification of circulating strains in a particular region is essential for establishing appropriate control and prevention measures such as development of vaccines, control of potential reservoirs, information for the general population, etc. Typing of the clinical isolates can also be important for identifying strains or virulence factors associated with disease severity. However, as indicated above, culture isolation is challenging.

Given the value of whole genomes for phylogenetic, epidemiological, and biological studies, there is an increasing interest in obtaining the genomic sequences of pathogens from clinical samples. This is particularly true for pathogens that are found in low quantities in the host organism and difficult to culture, as for pathogenic *Leptospira*. Single alleles, such as *rrs* [12], *ligB* [13], *lfb1* [14], and *secY* [15–17], can be directly amplified from the samples and sequenced for subtyping but this approach provides only low-level resolution and does not allow discrimination between serovars and closely related strains. Multi-locus sequence typing (MLST) schemes using several alleles can also be used for direct typing from clinical

samples, but this can result in incomplete allelic profiles [18–20] and they provide limited genetic information on the infecting strain. We recently developed a core genome MLST (cgMLST) scheme based on 545 genes that are highly conserved across the *Leptospira* genus [4]. Our cgMLST scheme allows the identification of pathogenic species, serogroups, and closely related serovars. However, this highly discriminatory approach requires culture isolation of clinical strains. Direct sequencing from clinical samples is hampered by high human host DNA contamination. Illumina sequencing of the cerebrospinal fluid of a patient with neuroleptospirosis showed, for example, only 0.016% of the sequence reads corresponding to the bacterial agent of leptospirosis [21]. Due to the low number of pathogenic microbes in clinical samples, several culture-independent genome sequencing methods have been recently developed using host depletion and/or microbial enrichment approaches. Targeted DNA enrichment, which relies on reference genomes of the target bacteria, has thus been used to retrieve the DNA of bacterial pathogens, such as *Chlamydia trachomatis* [22], *Mycobacterium tuberculosis* [23], and *Treponema pallidum* [24–26], from clinical samples.

Here, we describe a method utilizing biotinylated RNA probes designed specifically for *L. interrogans* DNA to capture the *Leptospira* core genomes defined by our cgMLST scheme [4] directly from routine diagnostic samples. This study demonstrates, for the first time, the successful and accurate high-coverage sequencing of *Leptospira* genomes directly from biological samples.

## Methods

### Samples

Twenty routine diagnostic samples (12 blood, 1 serum, and 7 urine samples) testing positive for Leptospirosis by real-time PCR in the French National Reference Center (NRC, Institut Pasteur) were analyzed in this study (Table 1); there was no attempt of culture isolation for these samples. Total DNA was extracted using DNeasy Blood and Tissue DNA extraction and QIAamp kits (Qiagen) and PCR was performed by real-time PCR using *lfb1* as a target [27]. Sequencing of the PCR products of *lfb1* enabled identification of the *Leptospira* species. *Leptospira interrogans*- or the *L. interrogans*-related species *L. kirschneri*-infected samples with a Ct value  $\leq 38$  were further selected for this study (Table 1).

### SureSelect<sup>XT</sup> target enrichment

In total, 42,117 custom-made SureSelect 120-mer RNA baits (total probe size 1459 Mb) based on 130 *L. interrogans* genome sequences (Additional file 2: Table S1) spanning the 545 core genes previously defined for our

**Table 1** Samples used in this study and identification of the infecting *Leptospira* strain

ID	Source	Patient	Sex	Age	Ct value <sup>a</sup>	Days <sup>b</sup>	Nreads <sup>c</sup>	Bacteria (%) <sup>d</sup>	Species <sup>e</sup>	Serogroup <sup>e</sup>	Serovar <sup>e</sup>	Strain <sup>e</sup>	Reference genome (id) <sup>f</sup>	Number of SNPs
1	Blood	A	M	60	31	Unknown	11,223,500	30	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	4
2	Urine	B	M	61	30	6	9,523,661	72	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	4
3	Blood	C	M	41	33	Unknown	11,662,248	20	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	2
4	Blood	D	M	42	20	2	19,405,989	99	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	3
5	Blood	E	M	49	34	2	2,817,867	21	<i>L. interrogans</i>	Pyrogenes	Zanoni	LT2156	228	1272
6	Serum	F	M	40	34	4	6,936,356	13	<i>L. kirschneri</i>	Grippotyphosa	Grippotyphosa	201402975	117	6
7	Blood	G	M	54	27	4	13,037,685	90	<i>L. kirschneri</i>	Grippotyphosa	Grippotyphosa	201402975	117	15
8	Blood	H	M	20	32	Unknown	8,739,223	65	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	4
9	Blood	I	M	69	29	3	12,665,561	65	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	6
10	Blood	J	M	74	28	0	12,989,647	94	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	3
11	Blood	K	M	36	32	4	9,589,816	90	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	5
12	Urine	L	M	24	31	Unknown	10,827,346	65	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	4
13	Urine	M	M	73	31	5	11,947,042	90	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	7
14	Urine	N	M	39	35	8	29,832,813	33	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	3
15	Urine	O	M	59	34	Unknown	12,556,161	29	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	4
16	Blood	P	M	52	37	3	10,420,132	62	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	6
17	Blood	Q	M	44	25	5	1,350,830	98	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	4
18	Blood	R	M	22	38	4	1,0980,69	13	<i>L. kirschneri</i>	NA	NA	NA	NA	NA
19	Urine	S	M	33	35	5	1,124,989	35	<i>L. interrogans</i>	Icterohaemorrhagiae	Icterohaemorrhagiae	Verdun	106	0
20	Urine	T	M	28	32	4	1,776,330	64	<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	L1-130	246	2

NA non applicable

<sup>a</sup> Ct value (real-time PCR Lfb1)

<sup>b</sup> Days of sample collection since symptom onset

<sup>c</sup> Number of illumina reads

<sup>d</sup> Proportion of reads from bacteria (%), other reads correspond to human reads

<sup>e</sup> Identification of the infecting strain based on the closest homologue of reference genomes

<sup>f</sup> Identification number (id) of the reference genome in BIGSdb ([www.bigsdb.pasteur.fr](http://www.bigsdb.pasteur.fr))

cgMLST scheme [4] were designed and synthesized by Agilent Technologies.

For all samples, libraries were prepared following the Sureselect Xt HS Target Enrichment System for Illumina from Agilent Technologies. For pre-capture library preparation, between 2 and 200 ng total gDNA was used. Briefly, samples were mechanically sheared using a Covaris E220, repaired, and the ends A-tailed for barcoded Illumina adapter ligation. Ligated samples were amplified for 14 cycles and library quality was assessed using the Fragment Analyzer HS NGS migration kit. Libraries were captured individually, as recommended by Agilent. Captured libraries were pooled and sequenced using Illumina sequencing technology (Miniseq or Nextseq 500 sequencers). We also sequenced five libraries (samples 3, 4, 5, 7, and 8) before capture on an Illumina Miniseq sequencer to assess the efficiency of the capture method on patient samples.

### Sequence analysis

Prior to mapping, reads were trimmed of the adapters. The variant calling pipeline described herein includes an additional trimming step performed during the mapping step, which removes bases with a Phred score < 30. The estimated depth of coverage was computed using the genome of the *L. interrogans* serovar *Copenhageni* strain Fiocruz L1-130 as a reference. Taxonomic analysis based on Kraken2 [28] using human, bacterial, and viral databases allowed us to classify an average of 99% of the reads with a minimum of 97.2% for one sample.

We generated a database of 273 genome sequences of representative pathogenic *Leptospira* strains from our publicly accessible online genome database <https://bigsd.b.pasteur.fr/leptospira/>, which is based on the software framework Bacterial Isolate Genome Sequence Database [4]. Our allele database includes genomes from eight pathogenic species from 23 serogroups and 59 serovars isolated from patients from 40 countries (Additional file 2: Table S2).

Variant calling analysis was performed using the variant calling pipeline (v0.10.0) from the Sequana project [29] ([https://github.com/sequana/variant\\_calling](https://github.com/sequana/variant_calling)). Default parameters were used except for the minimum frequency set to 0.5. The mapping step was performed using BWA (v0.7.17) [30] and variant calling was performed using freebayes (v1.3.2) [31]. Annotations were also included in the final HTML report by using prokka [32] annotation of the core genomes. All VCF (variant calling format) files (20 samples times 273 strains) were processed to gather the number of SNPs, INDELS, and MNPs in each sample and each strain using the VCF files and the Python notebooks used to process them are

available on Zenodo [<https://doi.org/10.5281/zenodo.7584745>].

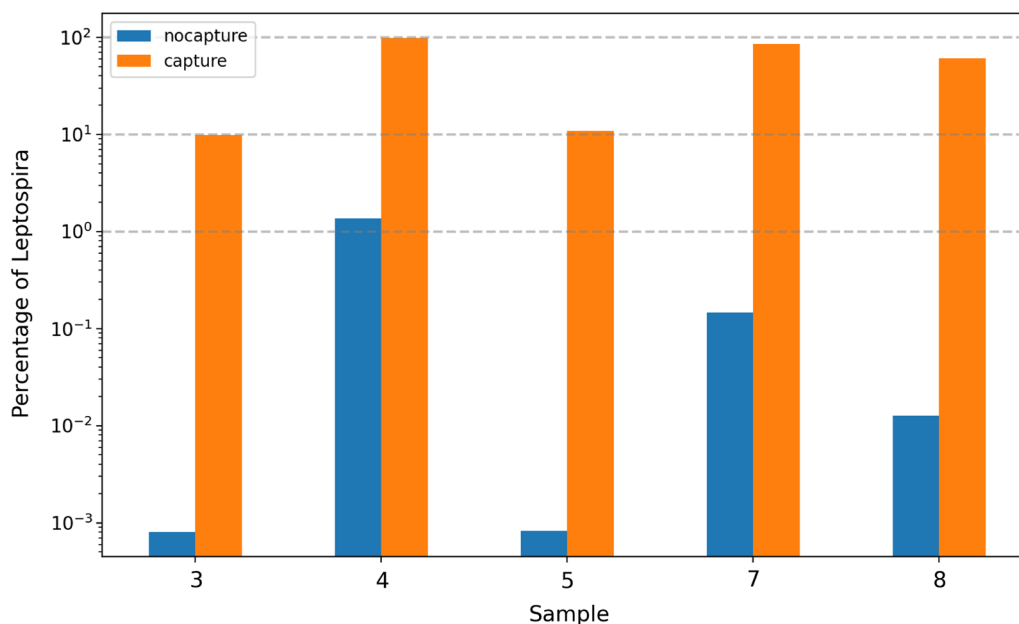
Variants were removed from subsequent analysis if one of the following conditions was met: (i) a frequency of the alternate < 0.5 (minor variants), (ii) a strand balance < 0.2 (or > 0.8), indicating an unbalanced count of forward or reverse reads supporting the variant, or (iii) coverage < 10. For sample 18 (highest Ct), we allowed the depth to be as low as 4.

### Results

Targeted enrichment was applied to 20 biological samples, consisting of blood, serum, and urine samples, from leptospirosis patients (Table 1). *Leptospira* isolates were collected from patients living in mainland France but patient E mentioned having traveled to the Philippines 2 weeks before the onset of symptoms. Among the 20 patients, all were males (71.2%) and the median (range) age was 46 (20–74) years. Patients presented with median (range) of 3.9 (0–8) days of acute illness and most of them exhibited fever, headache and myalgia; (patient Q died).

Library preparation, hybridization, and subsequent enrichment were carried out on samples using the SureSelect Target Enrichment System (Agilent Technologies) [33] and custom designed RNA baits. We compared the proportion of reads mapped to the *Leptospira* reference genomes with or without the SureSelect system for five samples to better evaluate the efficiency of *Leptospira* capture (Fig. 1). The percentages of reads mapped to the *Leptospira* reference genomes for samples prepared without the target-enrichment steps were 0.0008% (sample 3), 1.36% (sample 4), 0.0008% (sample 5), 0.15% (sample 7), and 0.013% (sample 8). The percentages of reads mapped to the *Leptospira* genomes for the same samples prepared using the SureSelect system jumped to 10%, 98%, 11%, 86%, and 61%, respectively. Thus, capture increased the proportion of *Leptospira* by several orders of magnitude (72–13,000).

Almost all of the bacterial reads were assigned to the family *Leptospiraceae* (Additional file 1: Fig. S1, S2); viral content was marginal (< 0.1%) (Additional file 1: Fig. S2). The average depth of coverage in the 20 target-enriched samples was 590 ×, ranging from 0.5 × (sample 18) to 6000 × (sample 4) (Additional file 1: Fig. S3), hence leading to a large standard deviation of 1320. Coverage across the genomes was computed using the Sequana coverage tool [34] to more precisely characterize the genomic variations in the different samples (Additional file 1: Fig. S4). The enrichment results, together with the average coverage (Additional file 1: Fig. S3), highlights several key points. Most samples had coverage above 50 ×, except samples 3, 5, 6, and 18. The coverage of samples 3 and 5 was still sufficient, with 42 × and 28 ×, respectively.



**Fig. 1** Percentage of *Leptospira* reads found in samples 3, 4, 5, 7, and 8 before (blue bars) and after hybridization capture using *Leptospira* RNA probes (orange bars)

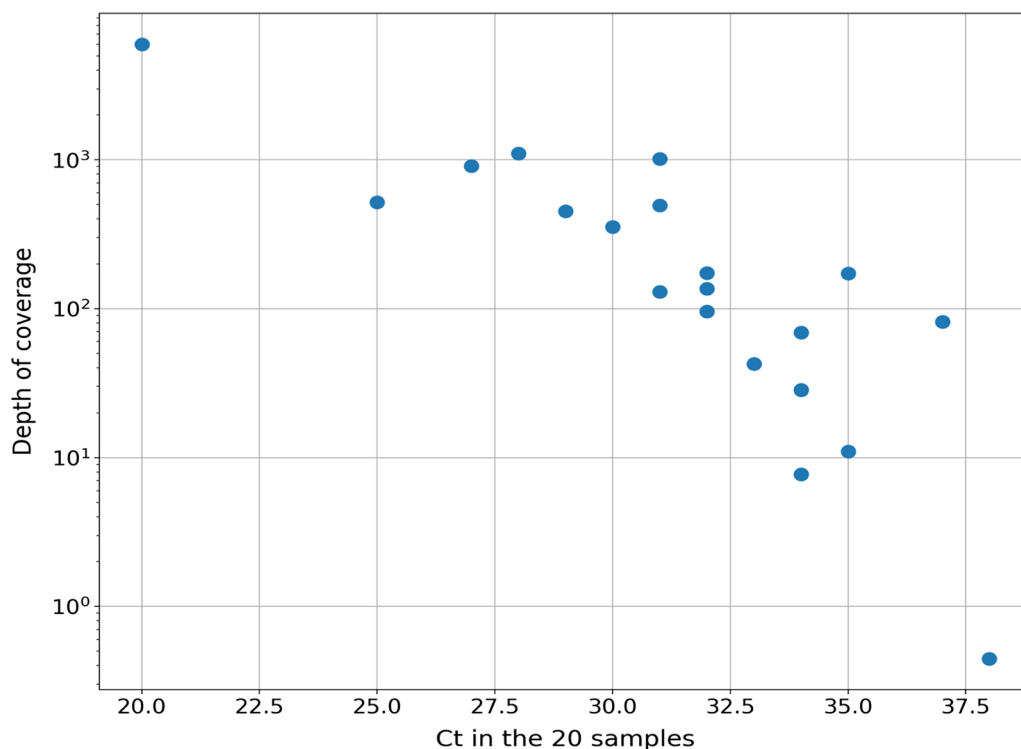
Sample 6 had a low coverage of  $8 \times$ . Finally, sample 18 was more problematic, as its coverage was below  $1 \times$ . It was also possible to assess the breadth of coverage (percentage of bases covered by at least one read) using *L. interrogans* serogroup serovar Copengageni (id246) as a reference (Additional file 1: Fig. S4); it was above 99.5% for most samples, except for sample 6, which had a breadth of coverage of 85%, and sample 18, for which the coverage was only about 3%.

The Ct values of real-time PCR targeting the pathogen-specific target *lfb1* ranged between 20 and 38, corresponding to 105 bacteria/ $\mu$ l to less than 1 bacterium/ $\mu$ l [27]. There was a good correlation between the Ct values and the proportion of mapped *Leptospira* reads and depth coverage (Fig. 2). Thus, the six samples with more than 90% *Leptospira* reads (samples 4, 7, 10, 11, 13, and 17) had Ct values  $< 32$  (Table 1).

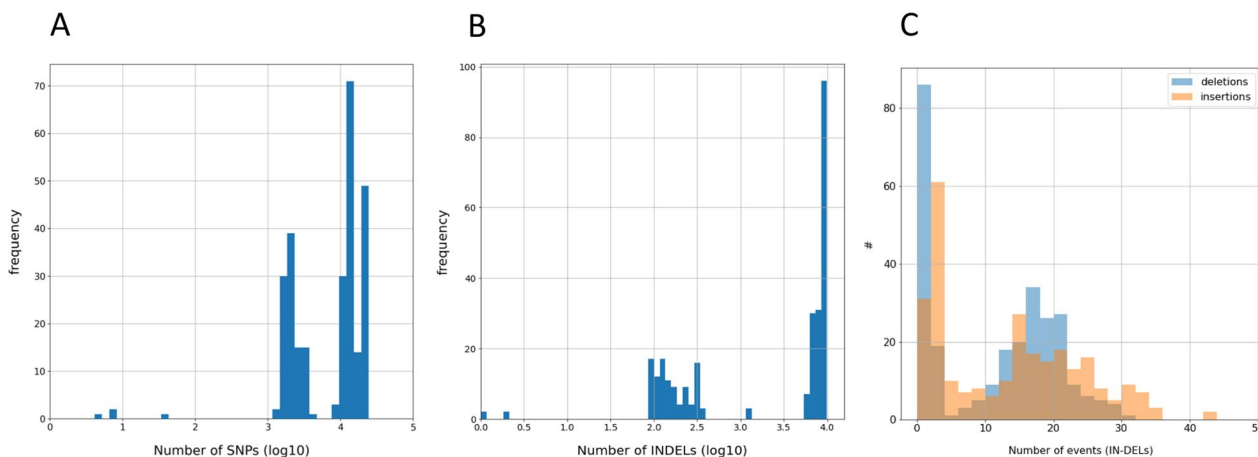
The variant calling approach was performed to identify the genotype of each sample. We searched for the closest genome for a given sample using a database of 273 genomes of pathogenic *Leptospira* strains from different species, serogroups, and serovars originating from various geographic areas (Additional file 2: Table S2) by minimizing the distance between the raw sequencing data of the sample and the *Leptospira* reference genomes. The distance used was the count of high-quality variants found in a given sample relative to the different strains, as explained below. As the capture was designed using probes covering the core genomes only, we solely

considered the core genome of the 273 strains; the average core genome length was  $574.7 \pm 8.5$  kb. Although coverage was uneven in some samples, with the presence of spikes (excess coverage in short regions, low frequency trend in sample 20) (Additional file 1: Fig. S4), it was generally high enough for variant calling analysis. We first examined the number of SNPs. The distribution of SNPs across all genomes and samples was highly variable, with values ranging from 0 to 23,000 SNPs (average of 10,000). The SNP count histogram across the 273 strains is shown in Additional file 1: Fig. S5, in which 95% of the strains show a count above 100, whereas a few strains had SNP counts below 10 (Fig. 3A; Additional file 1: Fig. S5).

Interestingly, most of the examined samples had less than 10 SNPs relative to the reference genomes of *L. interrogans* serovar (sv) Icterohaemorrhagiae strain RGA (id97), *L. interrogans* sv Icterohaemorrhagiae strain Verdun (id106), and *L. interrogans* sv Copenhageni strain Fiocruz L1-130 (id246) (Table 2). These three strains are phylogenetically related (Fig. 4) and belong to *L. interrogans* serogroup (sg) Icterohaemorrhagiae. In particular, *L. interrogans* sv Copenhageni (id246) appeared to be the strain with the smallest number of SNPs in most samples (if we ignore samples 5, 6, and 7 that were distant from the serogroup Icterohaemorrhagiae, and sample 18, which had no SNPs due to low coverage). Using the minimum number of SNPs as the criteria for assignment, 15 of the 19 samples were assigned to *L. interrogans* sv Copenhageni (id246). Sample 19 had 2 SNPs in id246



**Fig. 2** Final depth of coverage as a function of the measured Ct for the 20 samples. The depth of coverage is based on the *L. interrogans* core genome. The correlation coefficient between the depth of coverage and Ct values was 0.76



**Fig. 3** Analysis of multiple nucleotide polymorphisms. Histograms of the number of single-nucleotide polymorphisms (SNPs) (A), INDELS, including multiple nucleotide polymorphisms (MNPs) (B), and insertions and deletions, excluding MNPs (C) found in sample 1 across all 273 *Leptospira* genomes

and none in id106 and was thus assigned to *L. interrogans* sv Icterohaemorrhagiae (id106). Sample 5 was assigned to *L. interrogans* serovar Zanoni (id228) from serogroup Pyrogenes with large number of SNPs (1272); the second best hit among the 273 reference genomes was another Pyrogenes strain (id11). Samples 6 and 7 were assigned to *L. kirschneri* sv Grippotyphosa (id117) with 6 and 15

SNPs respectively (Table 2). Sample 18 was excluded from the analysis due to a low average. Nevertheless, if we decreased the required depth for variant calling to  $4 \times$  then the number of SNPs rose to approximately 20, on average, across all genomes (Additional file 1: Fig. S5). One had no SNPs (*L. kirschneri* sg Grippotyphosa; id149), although 15 other strains had 1 or 2 SNPs. These

**Table 2** SNPs, INDELS, and MNPs found in the 20 samples

Id	Species	Serovar	Strain	Samples																					
				1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20		
				Single nucleotide polymorphisms																					
246	<i>L. interrogans</i>	Copenhageni	Fiocruz L1-130	4	4	2	3	1737	3972	19916	4	6	3	5	4	7	3	4	6	4	2	2	2		
106	<i>L. interrogans</i>	Icterohaemorrhagiae	Verdun LP	6	6	5	5	1740	3972	19917	6	8	5	7	6	9	5	6	8	6	2	0	5		
97	<i>L. interrogans</i>	Icterohaemorrhagiae	RGA	7	8	6	7	1731	3929	19758	7	10	7	8	8	11	6	7	9	8	2	5	6		
11	<i>L. interrogans</i>	Pyrogenes	2006006973	1544	1550	1488	1559	1703	3867	19333	1543	1552	1553	1545	1546	1530	1534	1555	2	648	1537				
228	<i>L. interrogans</i>	Pyrogenes	LT2156	1477	1467	1424	1493	1272	3073	14816	1476	1480	1446	1165	1482	1439	1479	1466	1469	1487	2	671	1475		
117	<i>L. kirschneri</i>	Grippotyphosa	201402975	20737	20875	20027	21044	19254	6	15	20709	20907	20967	20786	20883	20955	20789	20522	20605	20923	0	8419	20697		
				Deletions																					
246	<i>L. interrogans</i>	Copenhageni	Fiocruz L1-130	0	0	0	0	0	2	22	0	0	0	0	0	0	0	0	0	0	0	0	0		
106	<i>L. interrogans</i>	Icterohaemorrhagiae	Verdun LP	0	0	0	0	0	2	22	0	0	0	0	0	0	0	0	0	0	0	0	0		
97	<i>L. interrogans</i>	Icterohaemorrhagiae	RGA	0	0	0	0	0	2	22	0	0	0	0	0	0	0	0	0	0	0	0	0		
11	<i>L. interrogans</i>	Pyrogenes	2006006973	1	1	1	1	1	2	19	1	1	1	1	1	1	1	1	1	1	0	0	1		
228	<i>L. interrogans</i>	Pyrogenes	LT2156	0	0	0	0	0	3	18	0	0	0	0	0	0	0	0	0	0	0	0	0		
117	<i>L. kirschneri</i>	Grippotyphosa	201402975	17	18	13	20	12	0	0	18	20	20	0	18	20	20	18	20	18	13	16	20	4	17
				Insertions																					
246	<i>L. interrogans</i>	Copenhageni	Fiocruz L1-130	0	0	0	0	2	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
106	<i>L. interrogans</i>	Icterohaemorrhagiae	Verdun LP	1	1	1	1	3	0	14	1	1	1	1	1	1	1	1	1	1	0	0	2		
97	<i>L. interrogans</i>	Icterohaemorrhagiae	RGA	0	0	0	0	2	0	13	0	0	0	0	0	0	0	0	0	0	0	0	0	1	
11	<i>L. interrogans</i>	Pyrogenes	2006006973	3	3	3	3	4	1	13	3	3	3	3	3	3	3	3	3	3	3	0	4		
228	<i>L. interrogans</i>	Pyrogenes	LT2156	6	6	5	6	6	2	11	5	6	6	4	6	6	6	6	5	6	0	3	6		
117	<i>L. kirschneri</i>	Grippotyphosa	201402975	24	26	15	27	15	1	1	24	26	27	25	26	27	23	19	18	26	0	4	22		
				Multiple nucleotide polymorphisms																					
246	<i>L. interrogans</i>	Copenhageni	Fiocruz L1-130	0	0	0	0	124	862	6395	0	0	0	0	0	0	0	1	0	0	0	0	1		
106	<i>L. interrogans</i>	Icterohaemorrhagiae	Verdun LP	0	0	0	0	124	862	6393	0	0	0	0	0	0	0	1	0	0	0	0	0		
97	<i>L. interrogans</i>	Icterohaemorrhagiae	RGA	0	0	0	0	123	850	6345	0	0	0	0	0	0	0	1	0	0	0	0	0		
11	<i>L. interrogans</i>	Pyrogenes	2006006973	87	88	112	89	114	842	6259	89	90	91	90	92	93	90	97	91	89	0	42	89		
228	<i>L. interrogans</i>	Pyrogenes	LT2156	85	92	102	84	100	669	4745	87	87	90	81	85	96	85	88	90	83	0	37	86		
117	<i>L. kirschneri</i>	Grippotyphosa	201402975	7032	7060	6987	7135	6702	1	3	7029	7064	7108	7038	7071	7086	7032	6990	7002	7042	0	2341	7044		

The *Leptospira* sequences from samples were compared using a database of 273 core genomes of pathogenic *Leptospira* strains to find the closest homologue

strains are close to each other in the phylogenetic tree (Fig. 4) and belong to *L. kirschneri*. In particular, id117 has only 1 SNP, and was also the best hit for samples 6 and 7 (Table 2).

To confirm these results, we also examined other types of variants: insertions, deletions, and multiple nucleotide polymorphisms (MNPs). The counts of insertions and deletions are hereafter summarized together as the number of INDELS. The number of INDELS in sample 1 varied from 0 up to 10,000 (Fig. 3B, C). Using the minimum of the total count of INDELS and MNPs across genomes, we found results similar to those found with the SNPs analysis. Sample 6 had several best hits: id117, id110, and id700 that led to no deletions, one insertion and one MNP each. Interestingly, id117, id110, id700 belong to serogroup Grippotyphosa and they are next to each other in the phylogenetic branch. For sample 7, id117 was the closest with one insertion and 2 MNPs. Other reference genomes from serogroup Grippotyphosa (id110, id315, and id700) were closeby with only 1 or 2 additional INDELS/MNPs. For sample 5, id228 had the minimum number of INDELS/MNPs with 0 deletions, 6 insertions and 100 MNPs. All other samples, which were close to id246 in terms of SNPs, had no INDELS or MNPs in id246, id97 and id106 (except sample 15, which had 1 MNP and sample 20 with several insertions or MNPs) (see Table 2 for details). Overall, the studies of SNPs, INDELS and MNPs converge to provide a robust assignation for each sample.

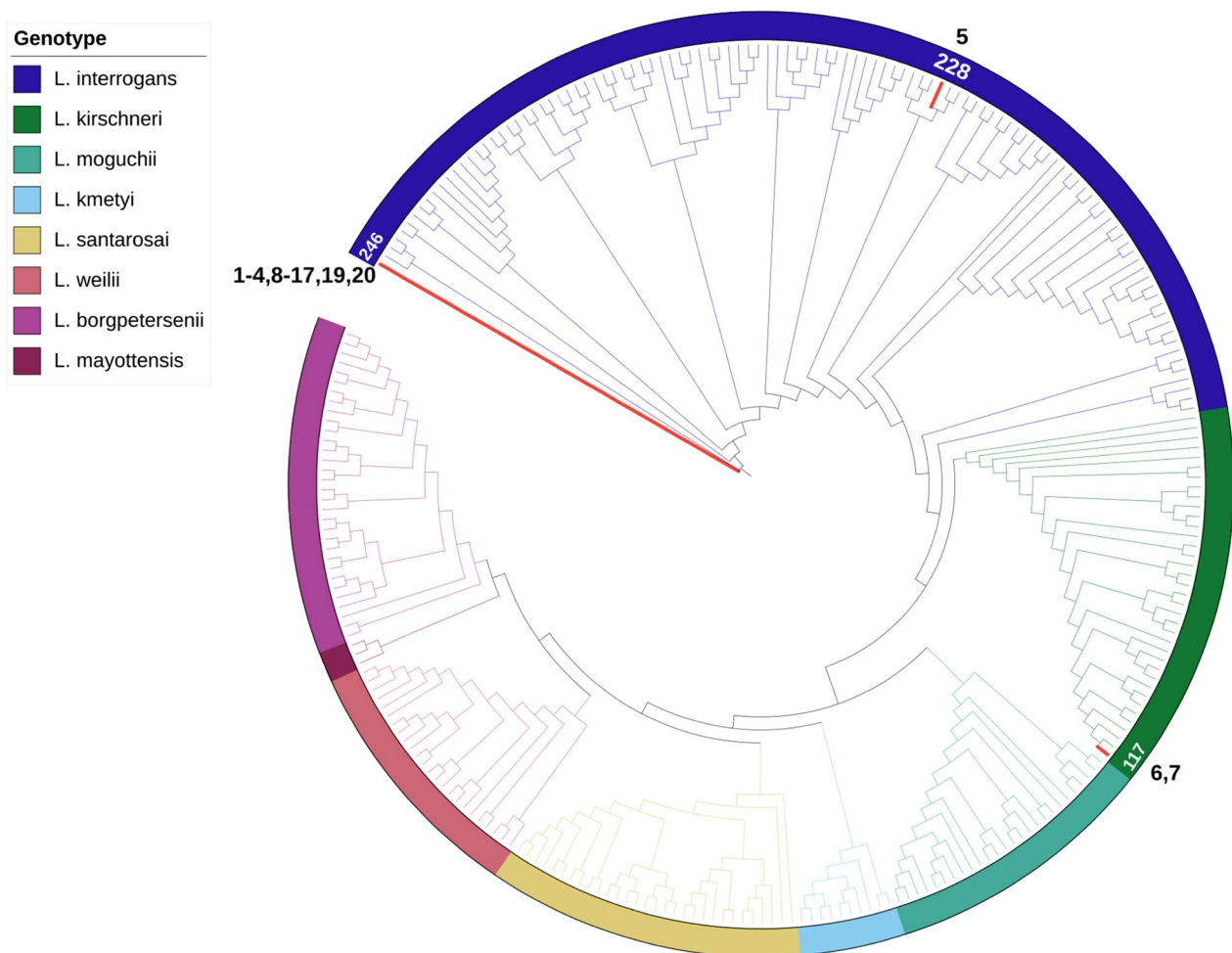
We further analyzed the SNPs and INDELS identified in coding regions for samples 1–4, 8–17 and 19–20 which were assigned to *L. interrogans* serovars Copenhageni and Icterohaemorrhagiae (Tables 1, 2). Comparison

of genome of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 with the sequences of the 545 core genes of the samples resulted in the identification of 2 to 7 SNPs which were distributed in 22 different genes (Table 3). SNPs in genes LIC11311, LIC13481 and LIC12955 were conserved in most if not all samples (Table 3); other SNPs were sample specific. Of the identified mutations in the core genes, 9 were synonymous and 13 were non-synonymous. The vast majority of genes with non-synonymous SNPs are annotated as involved in various biological processes such as amino acid transport and metabolism, energy production and conversion, lipid transport and metabolism, transcription, translation, ribosomal structure and biogenesis, etc. (Additional file 2: Table S3). We also noted 1 insertion causing a frameshift in LIC11604 which encodes a hypothetical protein (Additional file 2: Table S3).

**Discussion**

Genomics studies are proving to be important for the characterization of pathogen diversity and pathogenicity, yet the fastidious growth of *Leptospira* and the low abundance of *Leptospira* in clinical samples has presented a challenge for such studies. Thus, it can take up to four months of incubation for a primary culture to become positive [35, 36]. In addition, certain *Leptospira* serovars require additional culture media supplements for their growth [37]. Our data demonstrate, for the first time, the suitability of target-capture technology for purifying very low quantities of *Leptospira* DNA from complex DNA populations in which the host genome is in vast excess. We show the successful enrichment of *Leptospira* DNA by the significant increase in the ratio of bacteria:human





**Fig. 4** Phylogeny of *Leptospira* genomes and the sequenced samples. The *Leptospira* sequences from samples were compared using a database of 273 core genomes of pathogenic *Leptospira* strains (Additional file 1: Table S2) to find the closest homologue. Samples 1–4, 8–17, 19, 20 had a minimal number of single nucleotide polymorphisms (SNPs), deletions, insertions, and multiple nucleotide polymorphisms (MNPs) relative to id246 (*L. interrogans* sv Copenhageni). Samples 6 and 7 are closely related to id117 (*L. kirschneri* sv Grippyphosa) and Sample 5 is closely related to id228 (*L. interrogans* sv Pyrogenes). The phylogenetic tree was generated using multiple alignment (mafft v7.490) and RaxML software to infer the underlying phylogenetic tree. Visualization was performed via the iTOL web service (<https://itol.embl.de>)

DNA post-hybridization in a subset of samples. The Ct strongly correlated with capture efficiency. In six samples with a high leptospiral burden (qPCR Cts between 20 and 31 or 105 bacteria/ $\mu$ l to approximately 102 bacteria/ $\mu$ l), *Leptospira* reads accounted for >90% of the total reads. Targeted enrichment was applied to blood, serum, and urine samples from leptospirosis patients a few days after symptom onset (0–8 days, average of 3.9 days), showing that the method can be applied to routine diagnostic samples. We show that enrichment of *L. interrogans* reads provides sequencing data that match the quality and quantity of data obtained via sequencing from cultures, with coverage above  $50\times$  for 16 of the 20 samples, providing an opportunity to compare *Leptospira* strains from routine diagnostic samples with greater resolution

than previously possible. Today, this approach is still relatively expensive, currently costing approximately \$300 per sample in our laboratory, but as next-generation sequencing costs continue to decline, this approach should become more affordable and accessible. To reduce the cost in future studies, samples can be barcoded and pooled before enrichment, thus enabling multiplexing of hybridization reactions. This approach was already been proven to considerably decrease the cost [24, 38].

The use of specific probes for *L. interrogans* is justified by the cosmopolitan nature of this species, which is found worldwide [4]. The species *L. interrogans* also hosts the most pathogenic serovars, such as those belonging to the Icterohaemorrhagiae serogroup [5–8]. Finally, *L. interrogans* is particularly appropriate for

**Table 3** Genes showing SNPs and INDELS

Sample	Genes showing mutations									
	LIC10121	LIC10339	LIC10484	LIC10580	LIC10842	LIC10874	LIC12841	LIC11070	LIC11265	LIC11311
	<i>rsgA</i>	<i>clpC</i>	<i>thrC</i>	<i>msbA</i>	<i>dapA</i>	<i>RS04520</i>	<i>dapL</i>	<i>dhaT</i>		
1										623
2										623
3										623
4										623
8										623
9			1123	629						623
10										623
11	372									623
12										623
13		2207				293	220			623
14										623
15										623
16								930	107	623
17					902					623
19										623
20										623

Sample	Genes showing mutations											
	LIC11335	LIC11374	LIC12764	LIC11455	LIC11604	LIC11518	LIC11570	LIC11791	LIC12870	LIC12955	LIC10974	LIC13481
	<i>groL</i>	<i>fliR</i>	<i>ilvB</i>	<i>mfd</i>			<i>pulD</i>	<i>recG</i>	<i>rplB</i>	<i>fruA</i>	<i>fadA</i>	<i>ycfH</i>
1		12								30		721
2			666							30		721
3												721
4										30		721
8								1373		30		721
9							1819			30		721
10										30		721
11										30	549	721
12				1984						30		721
13						1928				30		721
14										30		721
15									704	30		721
16	1586									30		721
17										30		721
19												721
20					675							721

The genome of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (id246) was compared with the sequences from samples which were assigned to *L. interrogans* serovars Copenhageni and Icterohaemorrhagiae (Tables 1, 2). The position of mutations is indicated. More informations on the identified mutations can be found in Additional file 1: Table S3.

the use of target enrichment, as it has a relatively well-characterized clonal nature and *L. interrogans* strains from different origins show high genetic relatedness [4]. The specificity of the target enrichment probe sets was confirmed by our ability to specifically target *L. interrogans* (17/20). Furthermore, we were also able to

target *L. kirschneri* (3/20), which is the closest species phylogenetically to *L. interrogans* (Table 1).

More importantly, this enrichment method effectively captures regions of diversity in the *Leptospira* core genome, which enables precise molecular typing of infecting strains. Comparison of these assembled

sequences to the pathogenic *Leptospira* reference core genomes revealed only a limited number of SNPs for most of the samples. Remarkably, the number of SNPs for most samples was as low as a few (e.g., 4 SNPs in sample 1 on strain Fiocruz L1-130), even for samples with a low sequencing yield. Analyzing the SNPs, INDELS, or MNPs independently also provided coherent results, leading to robust assignment. There is however an exception with sample 5 which presents more than 1,200 SNPs with the closest reference genome which belongs to serogroup Pyrogenes. Interestingly this sample was collected from a patient who travelled in the Philippines where Pyrogenes is the most prevalent serogroup in both humans and animals [18]. However, the high number of SNPs reported for this sample suggests that the genome of the infecting strain is not present in our database. Further work will need to isolate and sequence more strains from patients to provide a better picture of the strains that are circulating worldwide.

Because we have a significant number of samples from patients infected with *L. interrogans* serovars Copenhageni and Icterohaemorrhagiae, we can analyze the genetic diversity of the core genes among these samples. Direct sequencing from clinical samples allows to get rid of the numerous mutations that occur during in vitro passages as previously described [39–42]. Previous comparative genomic studies of in vitro cultures from *L. interrogans* serovars Copenhageni and Icterohaemorrhagiae from different origin revealed that the genomes of these two serovars are highly conserved [43]. Similarly, we found a low proportion of mutations among the core genes during human infection and we did not identify any bias toward any particular biological function. One of the limitations of our study is that we analyze only a part of the genome (575/4600 kb or 12.5%) and we don't have access to the accessory genome which usually includes a number of virulence functions. In the near future, custom synthesized RNA probe sets could be designed to span the entire chromosome of *L. interrogans*. This will provide insights on the bias that may be introduced by culture, as previously shown for the spirochete *Treponema pallidum* [44], as well as increase our understanding of genetic diversity among strains and its impact on immune evasion, persistence and disease outcome.

## Supplementary Information

The online version contains supplementary material available at <https://doi.org/10.1186/s12879-023-08126-x>.

**Additional file 1: Fig. S1.** Proportion of reads of the 20 samples assigned to bacteria genomes (blue bars) and proportion of reads assigned specifically to the Leptospiraceae family (orange bars). **Fig S2** Taxonomic assignment of the reads of the 20 samples by Kraken2. **Fig. S3** Coverage depth of the *L. interrogans* core genome across samples. The coverage depth

on average for all samples is above 50X (top dashed red line) in most samples (sample 3 coverage equal 42X). Sample 5 and 6 have low coverage of 28X and 8X. Sample 18 coverage is below 1X. Black bars give the  $\pm 1$  standard deviation. **Fig. S4** Depth coverage of the 20 samples across the *L. interrogans* core genome. **Fig. S5** Histogram of SNP counts found in the 20 samples across all 273 genomes. For each of the 20 samples, we called variants using the 273 references of *Leptospira* (Additional file 2: Table S2) independently. For each genome, we obtained a set of SNPs that were filtered out to remove low quality variants (frequency below 10, uneven strand balance). Reference genomes distant from a sample led to thousands of SNPs while closely-related genomes led to less than 10 SNPs. Minimizing the count gives the closest genome to a given sample. Number of SNPs found in the 20 samples across all 273 genomes. Most genomes have more than 10,000 SNPs while only a few exhibit SNPs below 100.

**Additional file 2: Table S1** *L. interrogans* strains used for the design of probes. **Table S2** Database of *Leptospira* core genomes used for variant calling. **Table S3** Genes showing SNPs and INDELS. The genome of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (id246) was compared with the sequences from samples which were assigned to *L. interrogans* serovars Copenhageni and Icterohaemorrhagiae (Tables 1, 2)

## Acknowledgements

We thank the staff of the National Reference Center for Leptospirosis (Pascale Bourhy, Céline Lorient, and Farida Zinini) for support and processing of some of the samples.

## Author contributions

Conceived the study: LG, MP, and TC. Study implementation and procedures: JPF, TC, LG, and JFM. Data analysis: JFM, LG, MP, and TC. Wrote the paper: MP. All authors read and approved the final manuscript.

## Funding

This work was supported by the Institut Pasteur through grant PTR 30-2017 and Santé Publique France to MP. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. TC and JP (Biomics Platform, C2RT, Institut Pasteur, Paris, France), supported by France Génomique (ANR-10-INBS-09) and IBISA.

## Availability of data and materials

The raw sequencing data have been deposited in array express at EMBL-EBI (<https://www.ebi.ac.uk/>) with the accession number E-MTAB-11667.

## Declarations

### Ethics approval and consent to participate

Ethical approval was given by the Institutional Review Board of the Institut Pasteur. Written informed consent from patients was not required, as the study was conducted as part of the routine diagnosis of the French NRC for Leptospirosis, and no additional clinical specimens were collected for the purpose of the study. The need for informed consent was waived by the Institutional Review Board of the Institut Pasteur because of the retrospective nature of the study. Human samples were anonymized, and human sequences were removed from the data before submission to the database. Collection of the samples was conducted according to the Declaration of Helsinki.

### Consent for publication

Not applicable.

### Competing interests

All authors declare that they have no competing interests.

### Author details

<sup>1</sup>Biology of Spirochetes Unit, French National Reference Centre for Leptospirosis, CNRS UMR 6047, Institut Pasteur, Université Paris Cité, 75015 Paris, France. <sup>2</sup>Institut Pasteur, Université Paris Cité, Plate-Forme Technologique Biomics, 75015 Paris, France. <sup>3</sup>Département Biologie Computationnelle, Institut Pasteur, Université Paris Cité, Hub de Bioinformatique et Biostatistique, 75015 Paris,

France. <sup>4</sup>Present Address: Parasites and Microbes Programme, Wellcome Sanger Institute, Hinxton, UK.

Received: 22 May 2022 Accepted: 28 February 2023  
Published online: 14 March 2023

## References

- Costa F, Hagan JE, Calcagno J, Kane M, Torgerson P, Martinez-Silveira MS, Stein C, Abela-Ridder B, Ko AI. Global morbidity and mortality of leptospirosis: a systematic review. *PLoS Negl Trop Dis*. 2015;9:e0003898.
- Vincent AT, Schiettekatte S, Goarant C, Neela VK, Bernet E, Thibeaux R, Ismail N, Mohd Khalid MK, Amran F, Masuzawa T, 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.
- Korba AA, Lounici H, Kainiu M, Vincent AT, Mariet JF, Veyrier FJ, Goarant C, Picardeau M. *Leptospira ainahdjerensis* sp. nov., *Leptospira ainazelensis* sp. nov., *Leptospira abararensis* sp. nov. and *Leptospira chreensis* sp. nov., four new species isolated from water sources in Algeria. *Int J Syst Evol Microbiol*. 2021;71.
- 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.
- Hochedez P, Theodose R, Olive C, Bourhy P, Hurtrel G, Vignier N, Mehdaoui H, Valentino R, Martinez R, Delord JM, et al. Factors associated with severe leptospirosis, martinique, 2010–2013. *Emerg Infect Dis*. 2015;21:2221–4.
- Tubiana S, Mikulski M, Becam J, Lacassin F, Lefevre P, Gourinat AC, Goarant C, D'Ortenzio E. Risk factors and predictors of severe leptospirosis in New Caledonia. *PLoS Negl Trop Dis*. 2013;7:e1991.
- Herrmann-Storck C, Saint-Louis M, Foucand T, Lamaury I, Deloumeaux J, Baranton G, Simonetti M, Sertour N, Nicolas M, Salin J, et al. Severe leptospirosis in hospitalized patients. *Guadeloupe Emerg Infect Dis*. 2010;16(2):331–4.
- Christova I, Tasseva E, Manev H. Human leptospirosis in Bulgaria, 1989–2001: epidemiological, clinical, and serological features. *Scand J Infect Dis*. 2003;35:869–72.
- Riediger IN, Stoddard RA, Ribeiro GS, Nakatani SM, Moreira SDR, Skraba I, Biondo AW, Reis MG, Hoffmaster AR, Vinetz JM, et al. Rapid, actionable diagnosis of urban epidemic leptospirosis using a pathogenic *Leptospira* lipL32-based real-time PCR assay. *PLoS Negl Trop Dis*. 2017;11:e0005940.
- Agampodi SB, Matthias MA, Moreno AC, Vinetz JM. Utility of quantitative polymerase chain reaction in leptospirosis diagnosis: association of level of leptospiemia and clinical manifestations in Sri Lanka. *Clin Infect Dis*. 2012;54:1249–55.
- Waggoner JJ, Balassiano I, Mohamed-Hadley A, Vital-Brazil JM, Sahoo MK, Pinsky BA. Reverse-transcriptase PCR detection of leptospirosis: absence of agreement with single-specimen microscopic agglutination testing. *PLoS ONE*. 2015;10:e0132988.
- Cosson JF, Mielcarek M, Tatar C, Chaval Y, Suputtamongkol Y, Buchy P, Jitapalapong S, Herbreteau V, Morand S. Epidemiology of leptospirosis transmitted by rodents in southeast Asia. *PLoS Negl Trop Dis*. 2014;8:e2902.
- Bourhy P, Collet L, Clément S, Huerre M, Ave P, Giry C, Pettinelli F, Picardeau M. Isolation and characterization of new *Leptospira* genotypes from patients in Mayotte (Indian Ocean). *PLoS Negl Trop Dis*. 2010;4:e724.
- Perez J, Goarant C. Rapid *Leptospira* identification by direct sequencing of the diagnostic PCR products in New Caledonia. *BMC Microbiol*. 2010;10:325.
- Guernier V, Richard V, Nhan T, Rouault E, Tessier A, Musso D. *Leptospira* diversity in animals and humans in Tahiti, French Polynesia. *PLOS Neg Trop Dis*. 2017;11:e0005676.
- Mason MR, Encina C, Sreevatsan S, Muñoz-Zanzi C. Distribution and diversity of pathogenic *Leptospira* species in peri-domestic surface waters from South Central Chile. *PLoS Negl Trop Dis*. 2016;10:e0004895.
- Grillová L, Angermeier H, Levy M, Giard M, Lastère S, Picardeau M. Circulating genotypes of *Leptospira* in French Polynesia: an 9-year molecular epidemiology surveillance follow-up study. *PLoS Negl Trop Dis*. 2020;14:e0008662.
- Mendoza MV, Rivera WL. Application of simplified MLST scheme for direct typing of clinical samples from human leptospirosis cases in a tertiary hospital in the Philippines. *PLoS ONE*. 2021;16:e0258891.
- Varni V, Chiani Y, Nagel A, Ruybal P, Vanasco NB, Caimi K. Simplified MLST scheme for direct typing of *Leptospira* human clinical samples. *Pathog Glob Health*. 2018;112:203–9.
- Weiss S, Menezes A, Woods K, Chanthongthip A, Dittrich S, Opoku-Boateng A, Kimuli M, Chalker V. An extended multilocus sequence typing (MLST) scheme for rapid direct typing of leptospires from clinical samples. *PLoS Negl Trop Dis*. 2016;10:e0004996.
- Wilson MR, Naccache SN, Samayoa E, Biagtan M, Bashir H, Yu GX, Salamat SM, Somasekar S, Federman S, Miller S, et al. Actionable diagnosis of neuroleptospirosis by next-generation sequencing. *N Engl J Med*. 2014;370:2408–17.
- Christiansen MT, Brown AC, Kundu S, Tutill HJ, Williams R, Brown JR, Holdstock J, Holland MJ, Stevenson S, Dave J, et al. Whole-genome enrichment and sequencing of *Chlamydia trachomatis* directly from clinical samples. *BMC Infect Dis*. 2014;14:591.
- Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet D, Chan JZM, Depledge DP, Nikolayevskyy V, Broda A, Stone MJ, et al. Rapid whole-genome sequencing of *Mycobacterium tuberculosis* isolates directly from clinical samples. *J Clin Microbiol*. 2015;53:2230–7.
- Marks M, Fookes M, Wagner J, Butcher R, Ghinai R, Sokana O, Sarkodie YA, Lukehart SA, Solomon AW, Mabey DCW, et al. Diagnostics for yaws eradication: insights from direct next-generation sequencing of cutaneous strains of *Treponema pallidum*. *Clin Inf Dis*. 2018;6:818–24.
- Arora N, Schuenemann VJ, Jäger G, Peltzer A, Seitz A, Herbig A, Strouhal M, Grillová L, Sánchez-Busó L, Kühnert D, et al. Origin of modern syphilis and emergence of a pandemic *Treponema pallidum* cluster. *Nat Microbiol*. 2016;2:16245.
- Pinto M, Borges V, Antelo M, Pinheiro M, Nunes A, Azevedo J, Borrego MJ, Mendonça J, Carpinteiro D, Vieira L, et al. Genome-scale analysis of the non-cultivable *Treponema pallidum* reveals extensive within-patient genetic variation. *Nat Microbiol*. 2016;2:16190.
- Bourhy P, Bremont S, Zinini F, Giry C, Picardeau M. Comparison of real-time PCR assays for detection of pathogenic *Leptospira* spp. in blood and identification of variations in target sequences. *J Clin Microbiol*. 2011;49:2154–60.
- Wood DE, Lu J, Langmead B. Improved metagenomic analysis with Kraken 2. *Genome Biol*. 2019;20:257.
- Cokelaer T, Desvillechabrol D, Legendre R, Cardon M. 'Sequana': a set of snakemake NGS pipelines. *J Open Source Softw*. 2017;2:352.
- Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* 2013; 1303:3997v3992 [q-bio.GN].
- Garrison E, Marth G. Haplotype-based variant detection from short-read sequencing. *arXiv* 2012; 1207:3907 [q-bio. GN].
- Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;15:2068–9.
- Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, et al. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol*. 2009;27:182–9.
- Desvillechabrol D, Bouchier C, Kennedy S, Cokelaer T. Sequana coverage: detection and characterization of genomic variations using running median and mixture models. *Gigascience*. 2018;7:gij110.
- Wuthiekanun V, Chierakul W, Limmathurotsakul D, Smythe LD, Symonds ML, Dohnt MF, Slack AT, Limpaboon R, Suputtamongkol Y, White NJ, et al. Optimization of culture of *Leptospira* from humans with leptospirosis. *J Clin Microbiol*. 2007;45:1363–5.
- Jayasundara D, Senavirathna I, Warnasekara J, Gamage C, Siribaddana S, Kularatne SAM, Matthias M, Mariet JF, Picardeau M, Agampodi S, 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.
- Hornsby RL, Alt DP, Nally JE. Isolation and propagation of leptospires at 37 °C directly from the mammalian host. *Sci Rep*. 2020;10:9620.
- Beale MA, Marks M, Sahi SK, Tantaló LC, Nori AV, French P, Lukehart SA, Marra CM, Thomson NR. Genomic epidemiology of syphilis reveals independent emergence of macrolide resistance across multiple circulating lineages. *Nat Commun*. 2019;10:3255.

39. Zhong Y, Chang X, Cao XJ, Zhang Y, Zheng H, Zhu YZ, Cai C, Cui Z, Zhang Y, Li YY, et al. Comparative proteogenomic analysis of the *Leptospira interrogans* virulence-attenuated strain IPAV against the pathogenic strain 56601. *Cell Res*. 2011;21:1210–29.
40. Lehmann JS, Fouts DE, Haft DH, Cannella AP, Ricaldi JN, Brinkac L, Harkins D, Durkin S, Sanka R, Sutton G, et al. Pathogenomic inference of virulence-associated genes in *Leptospira interrogans*. *PLoS Negl Trop Dis*. 2013;7: e2468.
41. Lehmann JS, Corey VC, Ricaldi JN, Vinetz JM, Winzeler EA, Matthias MA. Whole genome shotgun sequencing shows selection on *Leptospira* regulatory proteins during in vitro culture attenuation. *Am J Trop Med Hyg*. 2016;94:302–13.
42. Satou K, Shimoji M, Tamotsu H, Juan A, Ashimine N, Shinzato M, Toma C, Nohara T, Shiroma A, Nakano K, et al. Complete genome sequences of low-passage virulent and high-passage avirulent variants of pathogenic *Leptospira interrogans* serovar manilae strain UP-MMC-NIID, originally isolated from a patient with severe leptospirosis, determined using PacBio single-molecule real-time technology. *Genome Announc*. 2015;3:e00882-e1815.
43. Santos LA, Adhikarla H, Yan X, Wang Z, Fouts DE, Vinetz JM, Alcantara LCJ, Hartskeerl RA, Goris MGA, Picardeau M, et al. Genomic comparison among global isolates of *L. interrogans* serovars copenhageni and icterohaemorrhagiae identified natural genetic variation caused by an indel. *Front Cell Infect Microbiol*. 2018;8:193.
44. Pinto M, Borges V, Antelo M, Pinheiro M, Nunes A, Azevedo JA, Borrego MJ, Mendonça J, Carpinteiro D, Vieira L, et al. Genome-scale analysis of the non-cultivable *Treponema pallidum* reveals extensive within-patient genetic variation. *Nature Microbiol*. 2017;2:16190.

### Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

