Infection, Genetics and Evolution 103 (2022) 105345

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Genetic basis underlying the serological affinity of leptospiral serovars from serogroups Sejroe, Mini and Hebdomadis

Eliseu Jayro de Souza Medeiros [a], Leonardo Cabral Afonso Ferreira [a], J. Miguel Ortega [b], Maria Raquel Venturim Cosate [c], Tetsu Sakamoto [a,b,*]

[a] Bioinformatics Multidisciplinary Environment (BioME), Instituto Metrópole Digital (IMD), Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil
[b] Laboratório de Biodados, Departamento de Bioquímica e Imunologia, Universidade Federal de Minas Gerais, Belo Horizonte, MG, Brazil
[c] MassBiologics, University of Massachusetts Medical School, Boston, USA

ARTICLE INFO

Keywords:
Cross-reactions
Leptospirosis
Lipopolysaccharides
O antigens
rfb locus

ABSTRACT

Leptospirosis is a widely distributed zoonosis caused by pathogenic strains of bacteria of the genus *Leptospira* (Phylum Spirochaetes). Its agents are commonly classified based on their antigenic characteristics into serogroups and serovars, which are relevant for epidemiologic studies and vaccine development. Serological tests are considered laborious and require a specialized infrastructure. Some molecular methods have been proposed to accelerate these procedures, but they still can not replace the immunological tests, thus requiring a further understanding of the genetic basis underlying the serological classification. In this work, we focused on elucidating the genetic factors determinant for the serogroup Sejroe, which is one of the most prevalent serogroups in livestock. For this, we conducted a comparative analysis using >700 leptospiral genomic samples available in the public database. The analysis showed that the genes comprising the *rfb* locus are the main genetic factors associated with the serological classification. Samples from the serogroup Sejroe have an *rfb* locus with a conserved gene composition that differs from most other serogroups. Hebdomadis and Mini were the only serogroups whose samples have an *rfb* locus with similar gene composition to those from serogroup Sejroe, corroborating with the serological affinity shared by them. Finally, we could determine a small region in the *rfb* locus in which each of those three serogroups can be distinguished by its gene composition. This is the first work that uses an extensive repertoire of genomic data of leptospiral samples to elucidate the molecular basis of the serological classification and open the road to more reliable strategies based on molecular methods for serodiagnosis.

1. Introduction

Leptospirosis is an important zoonosis with a worldwide distribution caused by pathogenic species of bacteria of the *Leptospira* genus (Faine et al., 1999). This genus is composed of gram-negative bacteria characterized by its long spiral-shaped cell body. It also presents high motility conferred by the two periplasmic flagella. Among pathogenic species, *Leptospira interrogans* (*Li*) is the most prevalent species associated with human and animal diseases (Ko et al., 2009; Levett, 2001).

The first taxonomic studies of this genus classified its members into two large groups according to their virulence into pathogens (*L. interrogans*) or saprophytes (*L. biflexa*) (Levett, 2001). Later, phylogenetic analysis verified that members of this genus form three well-defined clades that cluster together species with similar pathogenic levels (Perolat et al., 1998; Fouts et al., 2016). These three clades are commonly referred to as saprophytic, intermediate, and pathogenic. More recently, with the description of more species of this genus, especially of the saprophytic lineage, a new tree comprising 64 species of the genus *Leptospira* revealed a fourth clade that groups new saprophytic species in addition to the three clades characterized previously (Vincent et al., 2019).

Besides the taxonomic classification, leptospiral samples are also classified based on their antigenic characteristics in serovars (srv). There are >300 serovars described so far (Picardeau, 2017) that can be

Abbreviations: SHM, Sejroe/Hebdomadis/Mini; SVR, SHM rfb variable region.

* Corresponding author at: Bioinformatics Multidisciplinary Environment (BioME), Instituto Metrópole Digital (IMD), Universidade Federal do Rio Grande do Norte, Natal, RN, Brazil.
E-mail address: tetsu@imd.ufrn.br (T. Sakamoto).

https://doi.org/10.1016/j.meegid.2022.105345
Received 25 April 2022; Received in revised form 6 July 2022; Accepted 27 July 2022
Available online 30 July 2022

grouped into >30 serogroups (srg, Brenner et al., 1999) based in the agglutination affinities for related serovars. The major antigen involved in the serological classification is the lipopolysaccharide (LPS) exposed on the cell surface. Differences in the sugar composition and orientation of LPS are the key features that distinguish the different serovars (Adler, 2015). It is noteworthy that the serological classification has no correspondence with taxonomic classification (Faine et al., 1999; Kalambaheti et al., 1999; Mitchison et al., 1997). We could find different *Leptospira* species classified in the same serovar, e.g. the strains of srv Hardjo (srg Sejroe) has representatives of species *Li* and *L. borgpetersenii* (*Lb*) (Faine et al., 1999), and also a *Leptospira* species classified in different serovars. LPS is synthesized by genes that are clustered in a genomic region called the *rfb* locus (Mitchison et al., 1997) and differences in its gene composition suggest being the source of the serovar diversity in *Leptospira* spp. (de la Peña-Moctezuma et al., 1999, 2001).

Classifying samples into serovar/serogroup has a key role in epidemiological studies for contributing to a successful control measure of the disease through vaccination programs (Adler and de la Peña Moctezuma, 2010). The Microscopic Agglutination Test (MAT) is the gold standard for serodiagnosis (Terpstra et al., 2003). However, it is considered laborious and requires a specialized laboratory infrastructure. In this sense, molecular methods are attractive, as some of them, such as PCR, can be applied for an early diagnosis and do not require the isolation and cultivation of these bacteria (Di Azevedo and Lilenbaum, 2021; Schreier et al., 2013). In the context of leptospirosis, molecular methods are widely applied for taxonomic identification by sequencing the MLST genes (Ahmed et al., 2015; Boonsilp et al., 2013; Thaipadungpanit et al., 2007) or 16*S* rRNA (Morey et al., 2006); and for detection of pathogenic strains by amplifying the *ompL1* (Reitstetter, 2006) or *lipL32* gene (Cheemaa et al., 2007). Several molecular methods were also proposed for serological classification, such as those based on pulsed-field gel electrophoresis (Herrmann et al., 1992), detection of the variable number of tandem repeat (VNTR, Salaün et al., 2006; Majed et al., 2005), restriction fragment length polymorphism (RFLP, Jung et al., 2015), and quantification of IS (Cosate et al., 2017; Zuerner and Bolin, 1997), but they are still inefficient to replace serological tests (Marquez et al., 2017).

Comparative analyses showed a similarity in the genetic composition of this locus in *Li* and *Lb* strains of srv Hardjo and some of their genes have no orthologs in strains of other serogroups (Bulach et al., 2000; de la Peña-Moctezuma et al., 1999, 2001). However, since these first analyses were restricted to a portion of the *rfb* locus and a few samples, some uncertainties arise as (1) if there are other genomic regions important for serological classification, (2) if the gene composition among samples of the same serovar/serogroup is conserved, and (3) if leptospiral strains of other known serogroups may have similar genetic composition. To suggest a more reliable strategy based on molecular methods and to further understand the complex nature of the genetic basis underlying the serological diversity, we aimed to extend the previous analyses by using an extensive repertoire of genomic data with over 700 leptospiral strains available in public databases. We focused our study on finding the genetic bases which confer agglutination affinity among serovars of srg Sejroe, which is the prevalent leptospiral serogroup in several farm animals and manifests symptoms of difficult detection (Cosate et al., 2017; da Silva Pinto et al., 2016; de Oliveira et al., 2018; Polo et al., 2019; Tagliabue et al., 2016).

2. Material and methods

2.1. Genomic and serological data source

Genomic data of the *Leptospira* samples used in this work were obtained from NCBI RefSeq Genome Database. Genomic data from a total of 722 samples of Leptospires were downloaded (in January 2021) and analyzed. The quality of a genome assembly was assessed by BUSCO (v5, Simão et al., 2015), using the Spirochaetes ortholog dataset, and by manually inspecting the fragmentation level in the region of interest. Data on the serological classification of each sample was obtained from the Bacterial Isolate Genome Sequence Database (http://bigsdb.pasteur.fr/). Samples analyzed in this work cover 19 leptospiral species and 26 leptospiral serogroups (Supplementary Table 1).

2.2. Comparative and functional analysis

Comparative genomics analysis was performed to identify genomic regions of *Leptospira* that could be associated with the srg Sejroe. This was conducted by using the Proteome Comparison Tool from PATRIC (Davis et al., 2020), which performs a sequence similarity analysis of a reference proteome with up to six other proteomes. We selected *Li* str. Norma (srg Sejroe, Cosate et al., 2015) as the reference, and six other leptospiral strains of Sejroe (*Li.* str. Hardjoprajitno, *Lb* str. Sponselee, and *Lb* str. L550) and non-Sejroe (*Lb* str. 56604, *Li.* str. 56601, *Li* str. Fiocruz L1–130) serogroups. Genomic regions containing genes with higher similarity value among samples of srg Sejroe than of non-Sejroe serogroup were depicted for further analysis. Proteins encoded in depicted regions were functionally annotated using the ARGOT webserver (Lavezzo et al., 2016).

To verify the presence and the conservedness of those proteins in the 722 leptospiral samples from the NCBI RefSeq Genome Database, we submitted them to a sequence similarity search using the BLAST algorithm (blastp, Altschul et al., 1990) against the proteome of each leptospiral sample available at the NCBI database. The identity score between the query and subject proteins was calculated as the product of the percentage of identity and the query coverage of the first high-scoring segment pair (HSP). Graphical maps of circular genome and gene organization schemes were generated on the PATRIC web resource (Davis et al., 2020).

2.3. Primer designing

Primers to amplify genes of interest were designed using Primer-BLAST web server (Ye et al., 2012). For this, we firstly retrieved the DNA sequences of a target gene from genomes of all *Leptospira* strains deposited at RefSeq representative genomes. Then, we submitted them as a group of sequences to the Primer-BLAST with default parameters. Parameters for primer pair specificity checking were set to perform the database search for unintended primer templates in genome sequences of genus *Leptospira* which are part of Refseq representative genomes database.

3. Results

3.1. rfb locus has genes associated with serovars of srg Sejroe

The first screening for genomic regions that could be associated with serovars of srg Sejroe was performed by analyzing the sequence similarity of the proteome of *Li* str. Norma (srv Hardjo) with six other leptospiral strains (*Li* and *Lb*) of srg Sejroe (three strains) and of other serogroups (three strains). In this comparison, we could depict a single genomic region in which several genes of *Li* str. Norma has a higher similarity with other strains of srg Sejroe than with those from other serogroups (Fig. 1A, Supplementary Table 2). This region is part of the *rfb* locus, which is known to comprise genes involved in the biosynthesis of the O-antigen unit of the LPS structure (Fig. 1B). The genes of locus *rfb* were also depicted after repeating this analysis using other strains of srg Sejroe and non-Sejroe serogroups (Supplementary Fig. 1).

The *rfb* locus of *Li* str. Norma has 84 kb and is composed of 75 genes (Supplementary Table 3). To improve readability, the genes were sequentially numbered from 1 to 75 and referred to by their number and the "ORF" prefix (ORF1 to ORF75). This locus is flanked by genes encoding for a copper-binding protein and a ribosomal protein S6 and has the genes MarR family EPS-associated transcriptional regulator

(ORF1) and DASS family sodium-coupled anion symporter (ORF75) determining its start and end, respectively. Among the genes in the locus, we could find the gene cluster of dTDP-rhamnose biosynthesis (rfbABCD, ORF68 to ORF71) and genes related to the Wzx/Wzy-dependent pathway (ORF58 and ORF59). Most of the other genes in the locus encode for enzymes involved in sugar biosynthetic activities. Among them, the most frequent functional group was transferase, which includes methyl-, glycosyl-, and nucleotidyltransferases. Other

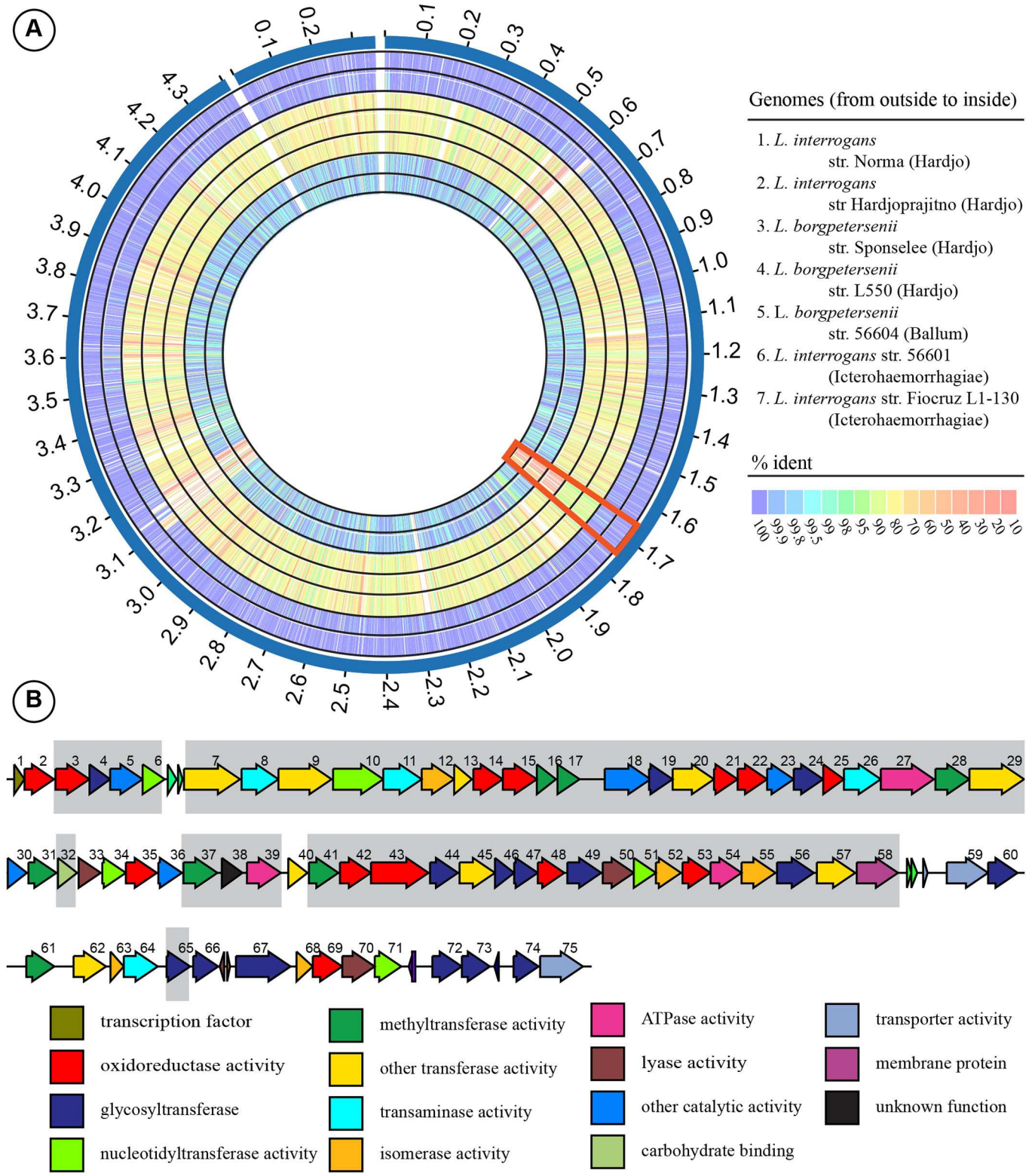

Fig. 1. A) Whole-genome sequence similarity analysis of *Li* str. Norma (srg Sejroe) with Sejroe and non-Sejroe leptospiral strains. The cluster of genes in which *Li* str. Norma shares a higher similarity with strains of srg Sejroe is highlighted in red. B) Gene organization of *rfb* locus in *Li* str. Norma. Genes in gray boxes are those specific to or share a higher similarity among samples of the srg Sejroe. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

functional groups found were oxidoreductases, transaminases, lyases, and isomerases (Fig. 1B, Supplementary Table 4). In this analysis, 50 out of 75 genes were exclusive to or shared a higher similarity with genes from srg Sejroe samples, suggesting a unique gene content and organization of *rfb* locus for serovars of this serogroup.

3.2. *rfb* locus gene content in serovars of serogroups Mini and Hebdomadis is similar to those of srg Sejroe

To verify the specificity of the gene content of the *rfb* locus observed in the samples from srg Sejroe in the previous analysis, we performed a similarity analysis with BLAST using the protein sequence of genes comprising the *rfb* locus of *Li* str. Norma as query against proteomes of other leptospiral samples available in the NCBI database (Fig. 2, Supplementary Table 5). In this analysis, a total of 382 leptospiral samples with serogroup identified were analyzed. The samples covered 17 species and 26 serogroups. BUSCO completeness score among those leptospiral samples ranged from 75.3% to 100% with a mean score of 99.10%. 369 leptospiral samples (96.6%) showed a completeness score above 95%.

The samples of srg Sejroe analyzed in this work (42 samples) were taxonomically classified into one of three species: *Li* (8 samples in total; 4 of srv Hardjo, three of srv Medanensis, and one of srv Geyaweera), *Lb* (33 samples in total; 28 of srv Hardjo-bovis, and 5 of srv Balcanica), or *L. santarosai* (*Ls;* 1 sample of srv Guaricura). The similarity analysis of the genes comprising the *rfb* locus showed that all samples have a similar gene composition to that of *Li* str. Norma, despite the taxonomic diversity found in this serogroup. Of the 50 genes that were observed to be distinctive in this serogroup in the previous analysis, 36 were present in all samples. Among the srg Sejroe, 16 samples contained the *rfb* locus in a single contig. In these samples, the locus contained an average of ~73 genes, which is a number close to that found in the str. Norma. However, it was found that two of these samples, GCF_001618585.1 and GCF_001618525.1, contained a low number of genes compared to the average (57 and 58 genes). Analyzing their *rfb* locus, we found that a region composed of 20 genes was replaced by transposable elements. Additionally, we found that these genes were present in a genomic region outside the *rfb* locus, suggesting that genes involved in the LPS biosynthesis may be in more than one genomic region.

In addition to the samples from the srg Sejroe, the analysis also showed nine samples from other serogroups that presented a gene content of the *rfb* locus similar to that found among the srg Sejroe samples (Fig. 2, Supplementary Table 5). Among them were all samples from srg Mini (6 samples in total, one of srv Mini, two of srv Szwajizak, and three with srv undetermined) and Hebdomadis (2 samples in total, one of srv Hebdomadis, and one with srv undetermined), and a single sample from srg Bataviae (srv Bataviae). Since all samples from srg Mini and srg Hebdomadis analyzed share a similar gene content to srg Sejroe, this suggests that this *rfb* locus gene composition is also characteristic of samples of those two serogroups. This is also endorsed by the fact that different taxonomic groups can be found among samples from srg Mini and srg Hebdomadis, which cover, respectively, 5 (*Li*, *Lb*, *Ls*, *Lk*: *L. kirschneri*, and *Lm*: *L. mayottensis*) and 2 (*Li* and *Lw*: *L. weilii*) distinct species. We could not suggest the same for the srg Bataviae, since only one (*Li* str. UT234) of the 12 samples from this serogroup showed similarity in the gene composition of the *rfb* locus. As this single sample behaves as an exception among the others, it was not included in the following analyzes.

Among the 36 genes of the *rfb* locus found in all samples of the srg Sejroe, 30 were also found with high identity (> 80%) in the samples from srg Hebdomadis and srg Mini with high BUSCO scores (> 95%). In this analysis, only two samples, one from srg Mini and one from srg Hebdomadis, were not included because of their low BUSCO scores (75.3% and 80.8%, respectively), and their assemblies were highly fragmented in the *rfb* locus. Since those 30 genes are found only in serovars of serogroups Sejroe/Hebdomadis/Mini (SHM), we could consider them as interesting targets for molecular identification to distinguish SHM samples from other serogroups. We will refer from now on to the genetic composition of the *rfb* locus shared by serovars of these three serogroups as SHM rfb.

3.3. SHM rfb has a variable region that can discern samples between serogroups Sejroe, Hebdomadis, and Mini

Differences in the gene composition among the SHM rfb were verified to determine the genes associated with their serological classification in serogroups. In this analysis, we identified that the region between the ORF13 (GNAT family *N*-acetyltransferase) and ORF18 (DUF4910 domain-containing protein) show different patterns of gene composition, varying according to their serogroup classification (Fig. 3A). We refer to this region as the SHM rfb variable region (SVR). Sequence similarity analysis of proteins encoded in the SVR of a sample from srg Sejroe showed that the ORF14 (zinc-binding alcohol dehydrogenase) and ORF16 (class I SAM-dependent methyltransferase) are present in all samples of srg Sejroe (42 samples) and absent in samples from srg Hebdomadis and Mini (Fig. 3B). The ORF15 (Gfo/Idh/MocA family oxidoreductase genes) and ORF17 (a second class I SAM-dependent methyltransferase) were also specific for srg Sejroe, but the absence of these genes in one sample (*Li* srv Geyaweera str. 1L-Int) indicates that they could not be essential for the characterization of this serogroup (Fig. 3B).

To determine the genes in the SVR region that could be characteristic of srg Mini and srg Hebdomadis, we also performed a similarity analysis with BLAST but using the protein sequences of *rfb* locus from representative samples of those serogroups: *Lb* srv Mini str. 201000851 for srg Mini (Fig. 3C) and *Li* srv Hebdomadis str. R499 for srg Hebdomadis (Fig. 3D). The SVR of srg Mini and srg Hebdomadis samples have a carbamoyltransferase (ORFM1 and ORFH1) starting this region in five of six samples from srg Mini samples and one of two samples from srg Hebdomadis. Samples that did not show this gene, *Lk* str. 200801925 (srv undetermined; srg Mini) and *Lw* str. Ecochallenge (srv undetermined; srg Hebdomadis), have a fragmented assembly in this region and, therefore, it was not possible to confirm its absence. The presence of this gene in samples from both serogroups makes them more similar to each other than to the srg Sejroe. Following this gene, four samples from srg Mini presented a hypothetical protein (ORFM2), a class I SAM-dependent methyltransferase (ORFM3), which differs from that found in srg Sejroe, and a transferase hexapeptide repeat protein (ORFM4). These genes were shown to be specific for this serogroup. In the samples from srg Hebdomadis, another class I SAM-dependent methyltransferase (ORFH2), which differs from those found in samples from srg Mini and srg Sejroe, was specific for this serogroup. The domain-containing protein methyltransferase protein (ORFH3) also appeared to be a determinant for the classification of srg Hebdomadis. However, by checking this protein in the *Lk* str. Ecochallenge, there is a 381 bp deletion that removed its methyltransferase domain. We also noticed that among the samples from srg Mini, one of them (*Lk* srv undetermined str. 200803703) has a gene highly similar to that found specifically in srg Hebdomadis (ORFM2).

3.4. Classification of leptospiral strain with undetermined serogroup

An expressive amount of leptospiral samples (340 samples) have the genome sequenced, but do not have the serogroup determined. Even without their serological classification, we took advantage of this data to verify the coverage of the SHM rfb within the *Leptospira* genus and to verify the conservedness of the genes in the SVR that are determinants for the serological classification. For this, we performed a similarity analysis using the genes that comprise the *rfb* locus of a representative sample of each SHM serogroup (*Li* str. Norma for srg Sejroe, *Li* str. R499 for srg Hebdomadis, and *Lb* str. 201000851 srg Mini) as queries and the proteome of each sample of an unknown serogroup as the subject.

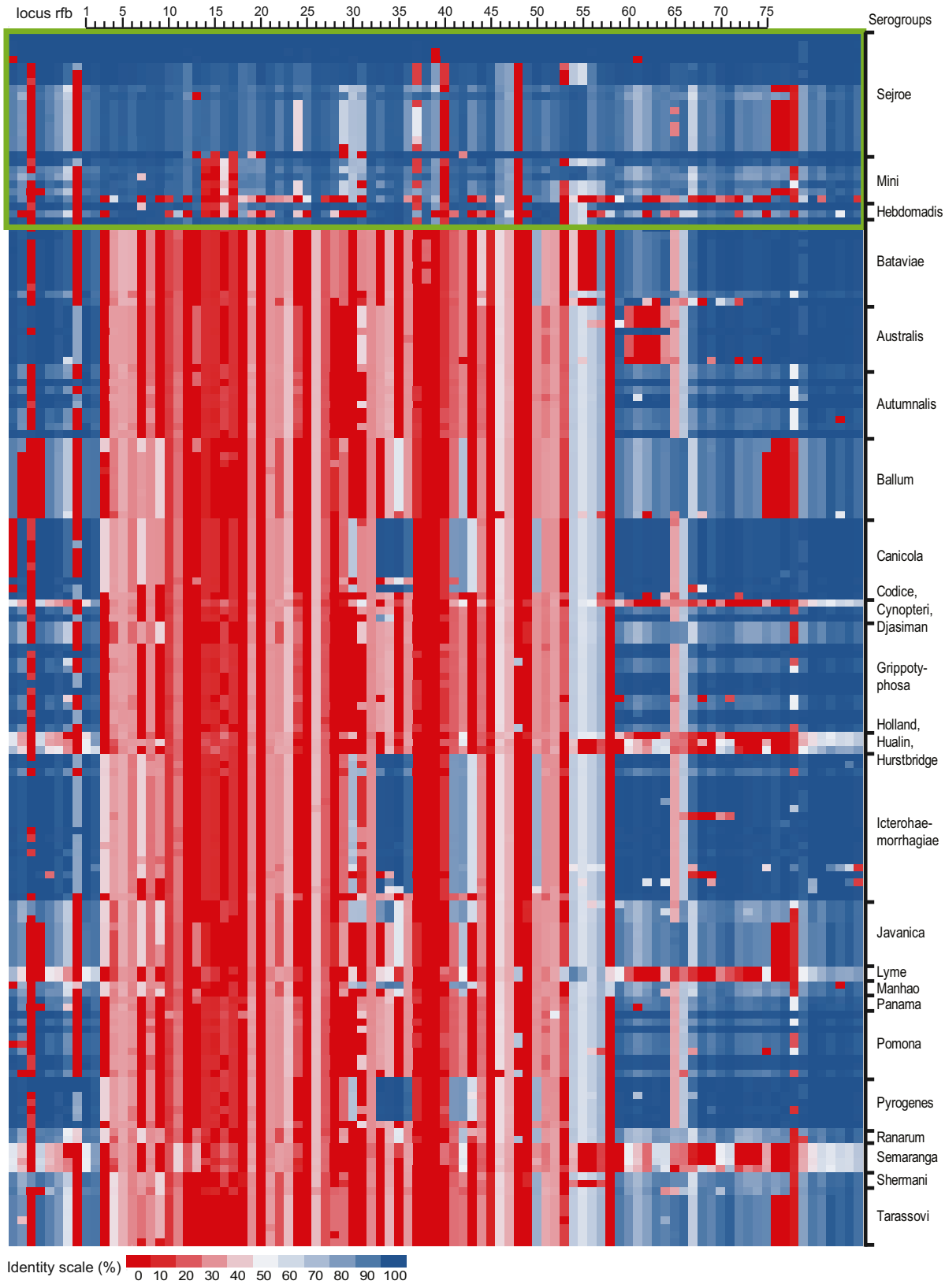

Fig. 2. Sequence similarity analysis of the 75 proteins composing the *rfb* locus of *Li* str. Norma (srg Sejroe) with the other leptospiral strains. Strains that share similar genetic composition with *Li* str. Norma are delimited in the green rectangle. The complete table is available in Supplementary Table 5. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

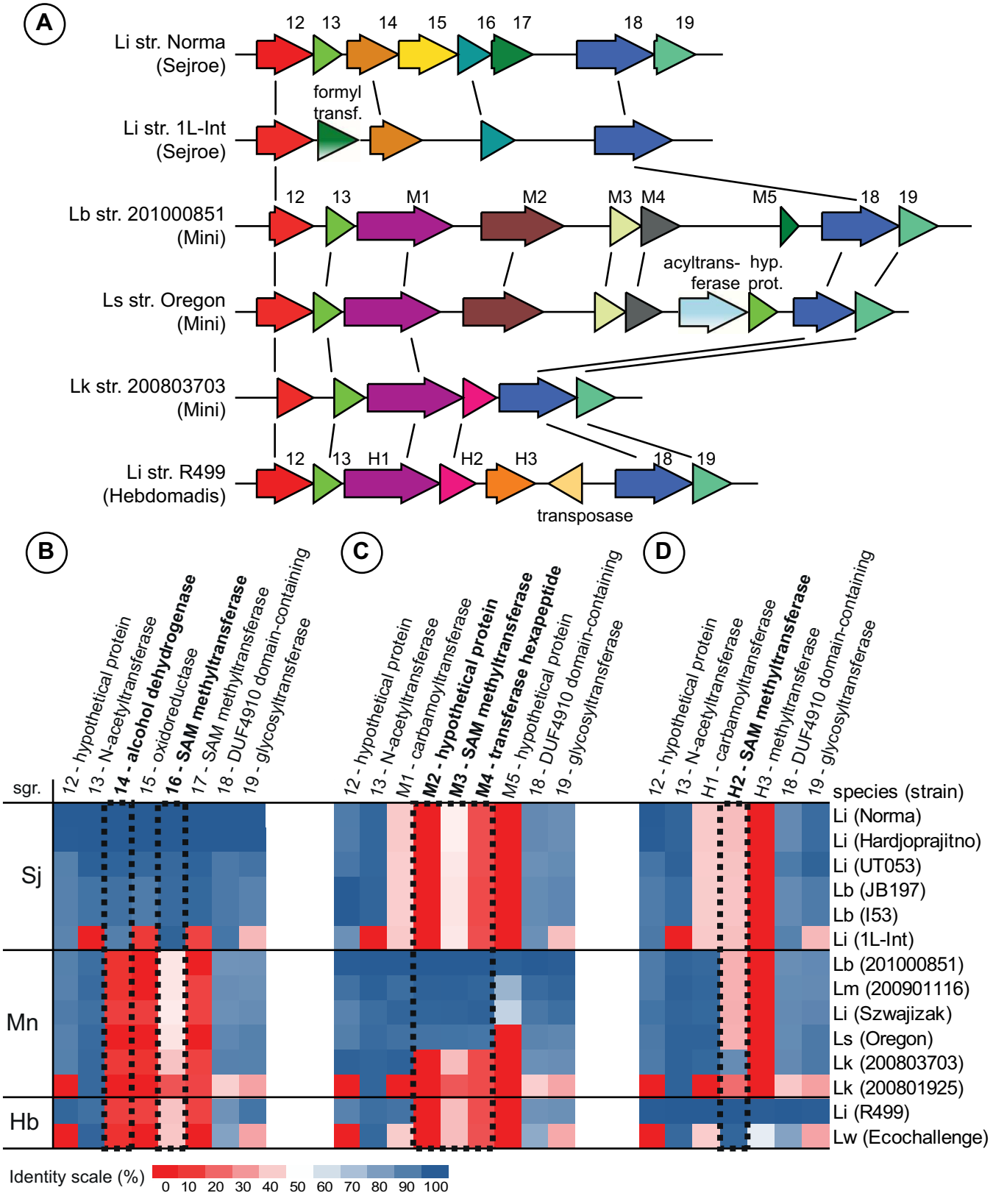

Fig. 3. A) Diversity on the genetic composition of SVR in strains of srgs SHM. Genes considered as putative orthologs have the same colour and are linked by a line. B) to D) Sequence similarity analysis using the proteins composing the SVR (columns) from B) *Li* str. Norma (sgr Sejroe, Sj), C) *Lb* str. 201000851 (sgr Mini, Mn), and D) *Li* str. R499 (sgr Hebdomadis, Hb) as query against the proteome of other leptospiral strains of srgs SHM (lines). Highlighted columns are proteins considered as determinants for the discernment of the srgs SHM. Legend: Li: *L. interrogans*, Lb: *L. borgpetersenii*, Lk: *L. kirschneri*, Ls: *L. santarosai*, Lm: *L. mayottensis*, Lw: *L. weilii*.

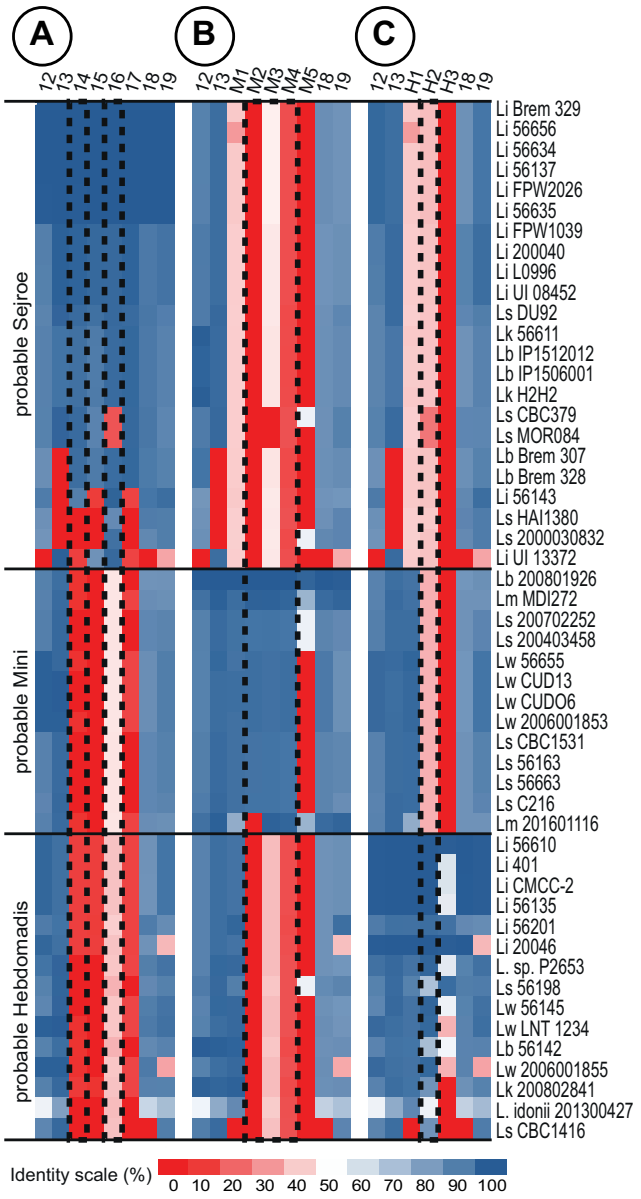

Fig. 4. Sequence similarity analysis of proteins composing the SVR against strains with undetermined srg. The proteins from A) *Li* str. Norma (sgr. Sejroe), B) *Lb* str. 201000851 (sgr. Mini), and C) *Li* str. R499 (sgr. Hebdomadis) were used as references.

Among the 340 samples, we verified that 51 of them, which cover 7 distinct species, had the SHM rfb (Supplementary Table 5).

The similarity analysis of proteins comprising the SVR allowed us to suggest the probable serogroup to which those samples should belong to (Fig. 4). The number of samples classified as of srg Sejroe, srg Hebdomadis, and srg Mini was 23 (covering the species *Li*, *Lb*, *Ls*, and *Lk*), 15 (*Li*, *Ls*, *Lw*, *Lb*, *Lk*, and *L. idonii*), and 13 (*Lb*, *Lm*, *Ls*, and *Lw*) samples, respectively. It is noteworthy to mention that there was neither the mixture nor the absence of patterns of the gene composition of SVR previously verified (Fig. 3A) in these samples. Some samples among those classified as probable Sejroe with continuous assembly did not present one of the ORFs defined as determinants for their classification in the srg Sejroe. Two samples (*Ls* str. HAI1380 and *Ls* str. 2000030832) did not show ORF14 and two other samples (*Ls* str. MOR084 and *Ls* str. CBC379) did not show ORF16. A serological test for those samples is needed to verify if, even in the absence of one of those genes, the sample will be classified as srg Sejroe. We also performed the sequence similarity analysis using protein sequences of SVR from other samples of SHM serogroups as queries. The classification obtained was the same, thus providing greater support to the previous results (Supplementary Table 6). Finally, for those genes in SVR used for the classification procedure, we designed primers and verified them *in silico* regarding their specificity in amplifying only in those samples of SHM serogroups or in those classified as such. All designed primers are summarized in Supplementary Table 7.

4. Discussion

LPS is known to be the main antigen that defines the serovar in *Leptospira* (Adler and de la Peña Moctezuma, 2010; Fraga et al., 2011). Its structure is catalyzed by enzymes orthologous to *Escherichia coli lpx* genes, which synthesize the saccharolipid anchor (or lipid A, Ren et al., 2003), and by enzymes encoded in the *rfb* locus (de la Peña-Moctezuma et al., 1999), which synthesize the distal polysaccharide portion (or O-antigen). Comparative analysis performed on the whole genome of the *Li* str. Norma with other samples from the srg Sejroe and non-Sejroe serogroups showed that many genes associated with the serovars of srg Sejroe are part of the *rfb* locus (Fig. 1). The first studies on comparative analysis of *rfb* locus in this serogroup (Bulach et al., 2000; de la Peña-Moctezuma et al., 1999, 2001) have already shown the intriguing association of these genes with its serological classification. However, these analyses were performed only for the 31 genes of the 3′ terminal region of the *rfb* locus and for a restricted number of samples. In this work, we verified that most of the genes in the *rfb* locus associated with the serovars of srg Sejroe were not previously reported and are located at the beginning of the locus (Fig. 1B). The similarity analysis showed that the *rfb* locus of samples from the srg Sejroe have a conserved genetic composition (Fig. 2). Since cross-reactions in the MAT assay is common among samples from the same serogroup (Faine et al., 1999), we could suggest that the gene composition of the *rfb* locus reflects on the final structure of the O-antigen and that samples with similar composition tend to generate similar O-antigen. This finding is also supported by the observation of a similar gene composition of the *rfb* locus between two samples of the srv Hurstbridge, which has representatives in two distinct species (*L. broomii* and *L. fainei*, Fouts et al., 2016).

An aspect that may prevent the serological identification by MAT of samples belonging to srg Sejroe is that the cross-reaction between some serovars of this serogroup is less common (Balboni et al., 2022; Khamassi Khbou et al., 2017); e.g. the srv Saxkoebing and srv Sejroe, both of srg Sejroe, seem to partially cross-react with srv Hardjo, a strain commonly used to detect strains of srg Sejroe. For this reason, some authors recommend the inclusion of additional strains in the MAT panel to avoid underestimation of the occurrence of this serogroup in epidemiological studies (Balboni et al., 2022). In this sense, the evidence of unique genes in *rfb* locus shared by samples belonging to srg Sejroe analyzed in this work, which cover 3 different species (*Li*, *Lb*, and *Ls*) and 5 serovars (Hardjo, Medanensis, Geyaweera, Balcanica and Guaricura), demonstrates that molecular approach detecting those genes could identify a broad spectrum of samples of srg Sejroe. Unfortunately, there is no genomic data regarding the srv Saxkoebing and srv Sejroe in the public database. Thus, accessing the *rfb* locus profile, either by genomic or PCR analysis, in both serovars is of great interest to further support our findings.

By comparing the *rfb* locus among samples from other serogroups, we found that its gene composition in serovars of srg Sejroe is highly similar to those found in serovars of srg Mini and srg Hebdomadis, suggesting that they share an antigenic similarity. This corroborates with the historical records of cross-reactions involving serovars of these three serogroups which made them part of the same serogroup (Hebdomadis) until 1982 (Stallman, 1984). Serovars of other serogroups are also known to show cross-reactivity on MAT assays, such as the Icterohemorrhagiae/Canicola serogroups (Jorge et al., 2011; Lang and

Morse, 1959). Similar analyses should confirm whether samples from these serogroups share a similar gene composition of the *rfb* locus. Similarities in the gene composition of the *rfb* locus between samples from different serogroups have also been observed in *L. alenxaderi* srv Manhao 3 (srg Manhao), *Lb* srv Javanica (srg Javanica), and *L. alstonii* srv Pingchang (srg Ranarum) (Fouts et al., 2016), suggesting that serovars of these serogroups also share a serological affinity.

The findings of this work can contribute to elaborate strategies to classify samples into the serogroups analyzed. The similarity analysis showed that at most 30 genes in the *rfb* locus are conserved and specific to SHM serogroups, which could be targeted for PCR amplification to identify samples belonging to one of these serogroups. Despite the high similarity of the *rfb* locus between samples from the SHM serogroups, we could find a small region, called here SVR, whose gene composition varies according to each serogroup (Fig. 3). Genes in the SVR identified to be specific to one serogroup could be targeted for a more precise serological classification. They were the ORF14 and ORF16 for srg Sejroe; ORFM2, ORFM3 and ORFM4 for srg Mini; and ORFH2 for srg Hebdomadis. Some authors have already suggested some genes to be used as PCR targets to aid in the serological identification of leptospiral samples. Cai et al. (2010) suggested the gene encoding Gfo/Idh/MocA family oxidoreductase (ORF15) to identify the serovars of srg Sejroe and carbamoyltransferase (ORFH1) to identify the serovars of srg Hebdomadis. Bezerra da Silva et al. (2011) use the gene encoding glycosyltransferase family 4 protein (ORF49) to identify srv Hardjo (srg Sejroe). Note that the genes suggested by Cai et al., 2010 are genes located in the SVR, however, some Sejroe samples do not have ORF15, and orthologs of ORFH1 are also found in samples of srg Mini (ORFM1). As for the ORF49 suggested by Bezerra da Silva et al. (2011), since it is conserved among SHM serogroups, it could not attend its first purpose of identifying specifically the srv Hardjo.

Few samples deviated from the pattern established in this work to characterize the SHM serogroups. *Li* str. UT234 is classified as srg Bataviae, but its *rfb* locus is more similar to those from srg Sejroe than from any other samples from the srg Bataviae. Some might suggest that this sample has experienced a recent genetic event that changed its serological classification to srg Bataviae. So, we performed a comparative genomic analysis between the sample in question (*Li* str. UT234) with other samples from srg Bataviae and srg Sejroe. This analysis, however, did not evidence genes that could support this event (data not shown). As we did not find in the literature any register of serological tests involving this sample for checking, we do not discard the possibility of a misannotation. The second sample is *Lk* str. 200801925 which is classified as srg Mini but does not have the three determinant genes of this serogroup (ORFM2, ORFM3, and ORFM4). By investigating the literature about this sample, we found that it cross-reacted with srg Mini and srg Hebdomadis samples (Bourhy et al., 2010, 2012). Since its gene composition of SVR is similar to those of the srg Hebdomadis, we suggest that this sample be reclassified to srg Hebdomadis.

5. Conclusion

The extensive use of leptospiral genomic data available in public databases allowed us to contemplate the pattern of conservedness and specificity on the genetic composition of the *rfb* locus of samples from SHM serogroups. We could rely on the genetic structure of the *rfb* locus, at least for serovars of SHM serogroups, to develop a reliable molecular strategy for the serological classification in serogroups of leptospiral samples. Our next step would be to make a similar analysis covering other leptospiral serogroups and finally to elaborate a molecular test based on these analyses.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2022.105345.

Funding

This study was financed in part by Coordenação de Aperfeiçoamento de Pessoal de Nível Superior - Brasil (CAPES) - Finace Code 001.

CRediT authorship contribution statement

Eliseu Jayro de Souza Medeiros: Formal analysis, Investigation. **Leonardo Cabral Afonso Ferreira:** Formal analysis, Investigation. **J. Miguel Ortega:** Resources. **Maria Raquel Venturim Cosate:** Conceptualization, Writing – review & editing. **Tetsu Sakamoto:** Supervision, Conceptualization, Writing – original draft.

Declaration of Competing Interest

None

Acknowledgments

We thank the Institut Pasteur teams for the curation and maintenance of BIGSdb-Pasteur databases at http://bigsdb.pasteur.fr/ and the BIOME/IMD/UFRN and CEPAD/ICB/UFMG teams for the support on the computational resource.

References

Adler, B., 2015. History of leptospirosis and leptospira. Curr. Top. Microbiol. Immunol. 387, 1–9. https://doi.org/10.1007/978-3-662-45059-8_1.

Adler, B., de la Peña Moctezuma, A., 2010. Leptospira and leptospirosis. Vet. Microbiol., Zoonoses: Adv. Perspect. 140, 287–296. https://doi.org/10.1016/j.vetmic.2009.03.012.

Ahmed, A., Ferreira, A.S., Hartskeerl, R.A., 2015. Multilocus sequence typing (MLST): markers for the traceability of pathogenic Leptospira strains. Methods Mol. Biol. Clifton NJ 1247, 349–359. https://doi.org/10.1007/978-1-4939-2004-4_25.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.

Balboni, A., Mazzotta, E., Boniotti, M.B., Bertasio, C., Bellinati, L., Lucchese, L., Battilani, M., Ceglie, L., Marchione, S., Esposito, G., Natale, A., 2022. Outbreak of Leptospira borgpetersenii serogroup sejroe infection in kennel: the role of dogs as sentinel in specific environments. Int. J. Environ. Res. Public Health 19, 3906. https://doi.org/10.3390/ijerph19073906.

Bezerra da Silva, J., Carvalho, E., Hartskeerl, R.A., Ho, P.L., 2011. Evaluation of the use of selective PCR amplification of LPS biosynthesis genes for molecular typing of leptospira at the serovar level. Curr. Microbiol. 62, 518–524. https://doi.org/10.1007/s00284-010-9738-7.

Boonsilp, S., Thaipadungpanit, J., Amornchai, P., Wuthiekanun, V., Bailey, M.S., Holden, M.T.G., Zhang, C., Jiang, X., Koizumi, N., Taylor, K., Galloway, R., Hoffmaster, A.R., Craig, S., Smythe, L.D., Hartskeerl, R.A., Day, N.P., Chantratita, N., Feil, E.J., Aanensen, D.M., Spratt, B.G., Peacock, S.J., 2013. A single multilocus sequence typing (MLST) scheme for seven pathogenic Leptospira species. PLoS Negl. Trop. Dis. 7, e1954 https://doi.org/10.1371/journal.pntd.0001954.

Bourhy, P., Collet, L., Clément, S., Huerre, M., Ave, P., Giry, C., Pettinelli, F., Picardeau, M., 2010. Isolation and characterization of new leptospira genotypes from patients in mayotte (Indian Ocean). PLoS Negl. Trop. Dis. 4, e724 https://doi.org/10.1371/journal.pntd.0000724.

Bourhy, P., Collet, L., Lernout, T., Zinini, F., Hartskeerl, R.A., van der Linden, H., Thiberge, J.M., Diancourt, L., Brisse, S., Giry, C., Pettinelli, F., Picardeau, M., 2012. Human Leptospira isolates circulating in mayotte (Indian Ocean) have unique serological and molecular features. J. Clin. Microbiol. 50, 307–311. https://doi.org/10.1128/JCM.05931-11.

Brenner, D.J., Kaufmann, A.F., Sulzer, K.R., Steigerwalt, A.G., Rogers, F.C., Weyant, R.S., 1999. 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. 49 (Pt 2), 839–858. https://doi.org/10.1099/00207713-49-2-839.

Bulach, D.M., Kalambaheti, T., de la Peña-Moctezuma, A., Adler, B., 2000. Lipopolysaccharide biosynthesis in Leptospira. J. Mol. Microbiol. Biotechnol. 2, 375–380.

Cai, C.-S., Zhu, Y.-Z., Zhong, Y., Xin, X.-F., Jiang, X.-G., Lou, X.-L., He, P., Qin, J.-H., Zhao, G.-P., Wang, S.-Y., Guo, X.-K., 2010. Development of O-antigen gene cluster-specific PCRs for rapid typing six epidemic serogroups of Leptospira in China. BMC Microbiol. 10, 67. https://doi.org/10.1186/1471-2180-10-67.

Cheemaa, P.S., Srivastava, S.K., Amutha, R., Singh, S., Singh, H., Sandey, M., 2007. Detection of pathogenic leptospires in animals by PCR based on lipL21 and lipL32 genes. Indian J. Exp. Biol. 45, 568–573.

Cosate, M.R.V., Soares, S.C., Mendes, T.A., Raittz, R.T., Moreira, E.C., Leite, R., Fernandes, G.R., Haddad, J.P.A., Ortega, J.M., 2015. Whole-genome sequence of

Leptospira interrogans serovar Hardjo subtype Hardjoprajitno Strain Norma, isolated from cattle in a leptospirosis outbreak in Brazil. Genome Announc. 3 https://doi.org/10.1128/genomeA.01302-15 e01302-15.

Cosate, M.R.V., Sakamoto, T., de Oliveira Mendes, T.A., Moreira, É.C., Regis da Silva, C. G., Brasil, B.S.A.F., Oliveira, C.S.F., de Azevedo, V.A., Ortega, J.M., Leite, R.C., Haddad, J.P., 2017. Molecular typing of Leptospira interrogans serovar Hardjo isolates from leptospirosis outbreaks in Brazilian livestock. BMC Vet. Res. 13, 1–12. https://doi.org/10.1186/s12917-017-1081-9.

da Silva Pinto, P., Libonati, H., Penna, B., Lilenbaum, W., 2016. A systematic review on the microscopic agglutination test seroepidemiology of bovine leptospirosis in Latin America. Trop. Anim. Health Prod. 48, 239–248. https://doi.org/10.1007/s11250-015-0954-9.

Davis, J.J., Wattam, A.R., Aziz, R.K., Brettin, T., Butler, R., Butler, R.M., Chlenski, P., Conrad, N., Dickerman, A., Dietrich, E.M., Gabbard, J.L., Gerdes, S., Guard, A., Kenyon, R.W., Machi, D., Mao, C., Murphy-Olson, D., Nguyen, M., Nordberg, E.K., Olsen, G.J., Olson, R.D., Overbeek, J.C., Overbeek, R., Parrello, B., Pusch, G.D., Shukla, M., Thomas, C., VanOeffelen, M., Vonstein, V., Warren, A.S., Xia, F., Xie, D., Yoo, H., Stevens, R., 2020. The PATRIC bioinformatics resource center: expanding data and analysis capabilities. Nucleic Acids Res. 48, D606–D612. https://doi.org/10.1093/nar/gkz943.

de la Peña-Moctezuma, A., Bulach, D.M., Kalambaheti, T., Adler, B., 1999. Comparative analysis of the LPS biosynthetic loci of the genetic subtypes of serovar Hardjo: Leptospira interrogans subtype Hardjoprajitno and Leptospira borgpetersenii subtype Hardjobovis. FEMS Microbiol. Lett. 177, 319–326.

de la Peña-Moctezuma, A., Bulach, D.M., Adler, B., 2001. Genetic differences among the LPS biosynthetic loci of serovars of Leptospira interrogans and Leptospira borgpetersenii. FEMS Immunol. Med. Microbiol. 31, 73–81. https://doi.org/10.1111/j.1574-695X.2001.tb01589.x.

de Oliveira, P.R.F., Soares, L.B.F., de Borges, J.M., de Barrosa, N.C., Langoni, H., Brandespim, D.F., Junior, J.W.P., Mota, R.A., 2018. Occurrence of serological reactions for serogroup Sejroe (CTG and Prajtino) in female buffalo in the state of Pernambuco, Brazil. Braz. J. Microbiol. 49, 795–800. https://doi.org/10.1016/j.bjm.2018.02.007.

Di Azevedo, M.I.N., Lilenbaum, W., 2021. An overview on the molecular diagnosis of animal leptospirosis. Lett. Appl. Microbiol. 72, 496–508. https://doi.org/10.1111/lam.13442.

Faine, S., Adler, B., Bolin, C., Perolat, P., 1999. "Leptospira" and leptospirosis. MediSci.

Fouts, D.E., Matthias, M.A., Adhikarla, H., Adler, B., Amorim-Santos, L., Berg, D.E., Bulach, D., Buschiazzo, A., Chang, Y.-F., Galloway, R.L., Haake, D.A., Haft, D.H., Hartskeerl, R., Ko, A.I., Levett, P.N., Matsunaga, J., Mechaly, A.E., Monk, J.M., Nascimento, A.L.T., Nelson, K.E., Palsson, B., Peacock, S.J., Picardeau, M., Ricaldi, J. N., Thaipandungpanit Jr., J., Wunder Jr., Elsio A., Yang, X.F., Zhang, J.-J., Vinetz, J. M., 2016. What makes a bacterial species pathogenic?: Comparative genomic analysis of the Genus Leptospira. PLoS Negl. Trop. Dis. 10, e0004403 https://doi.org/10.1371/journal.pntd.0004403.

Fraga, T.R., Barbosa, A.S., Isaac, L., 2011. Leptospirosis: aspects of innate immunity, immunopathogenesis and immune evasion from the complement system. Scand. J. Immunol. 73, 408–419. https://doi.org/10.1111/j.1365-3083.2010.02505.x.

Herrmann, J.L., Bellenger, E., Perolat, P., Baranton, G., Saint Girons, I., 1992. Pulsed-field gel electrophoresis of NotI digests of leptospiral DNA: a new rapid method of serovar identification. J. Clin. Microbiol. 30, 1696–1702. https://doi.org/10.1128/jcm.30.7.1696-1702.1992.

Jorge, R.S.P., Ferreira, F., Ferreira Neto, J.S., de Vasconcellos, S.A., de Lima, E.S., de Morais, Z.M., de Souza, G.O., 2011. Exposure of free-ranging wild carnivores, horses and domestic dogs to Leptospira spp in the northern Pantanal. Brazil. Mem. Inst. Oswaldo Cruz 106, 441–444. https://doi.org/10.1590/S0074-02762011000400009.

Jung, L.R.C., Bomfim, M.R.Q., Kroon, E.G., Nunes, Á.C., 2015. Identification of Leptospira serovars by RFLP of the RNA polymerase beta subunit gene (rpoB). Braz. J. Microbiol. 46, 465–476. https://doi.org/10.1590/S1517-838246220120018.

Kalambaheti, T., Bulach, D.M., Rajakumar, K., Adler, B., 1999. Genetic organization of the lipopolysaccharide O-antigen biosynthetic locus of Leptospira borgpetersenii serovar Hardjobovis. Microb. Pathog. 27, 105–117. https://doi.org/10.1006/mpat.1999.0285.

Khamassi Khbou, M., Haouala, K., Benzarti, M., 2017. High frequency of seropositivity of Leptospira in cattle in North Tunisia. Vet. Med. Sci. 3, 13–21. https://doi.org/10.1002/vms3.52.

Ko, A.I., Goarant, C., Picardeau, M., 2009. Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen. Nat. Rev. Microbiol. 7, 736–747. https://doi.org/10.1038/nrmicro2208.

Lang, R.W., Morse, E.V., 1959. Serologic cross reactions among Leptospirae observed with sera from animals infected with Leptospira Pomona. J. Immunol. 82, 471–476.

Lavezzo, E., Falda, M., Fontana, P., Bianco, L., Toppo, S., 2016. Enhancing protein function prediction with taxonomic constraints–The Argot2.5 web server. Methods San Diego Calif 93, 15–23. https://doi.org/10.1016/j.ymeth.2015.08.021.

Levett, P.N., 2001. Leptospirosis. Clin. Microbiol. Rev. 14, 296–326. https://doi.org/10.1128/CMR.14.2.296-326.2001.

Majed, Z., Bellenger, E., Postic, D., Pourcel, C., Baranton, G., Picardeau, M., 2005. Identification of variable-number tandem-repeat loci in Leptospira interrogans sensu stricto. J. Clin. Microbiol. 43, 539–545. https://doi.org/10.1128/JCM.43.2.539-545.2005.

Marquez, A., Djelouadji, Z., Lattard, V., Kodjo, A., 2017. Overview of laboratory methods to diagnose leptospirosis and to identify and to type leptospires. Int. Microbiol. Off. J. Span. Soc. Microbiol. 20, 184–193. https://doi.org/10.2436/20.1501.01.302.

Mitchison, M., Bulach, D.M., Vinh, T., Rajakumar, K., Faine, S., Adler, B., 1997. Identification and characterization of the dTDP-rhamnose biosynthesis and transfer genes of the lipopolysaccharide-related rfb locus in Leptospira interrogans serovar Copenhageni. J. Bacteriol. 179, 1262–1267.

Morey, R.E., Galloway, R.L., Bragg, S.L., Steigerwalt, A.G., Mayer, L.W., Levett, P.N., 2006. Species-specific identification of Leptospiraceae by 16S rRNA gene sequencing. J. Clin. Microbiol. 44, 3510–3516. https://doi.org/10.1128/JCM.00670-06.

Perolat, P., Chappel, R.J., Adler, B., Baranton, G., Bulach, D.M., Billinghurst, M.L., Letocart, M., Merien, F., Serrano, M.S., 1998. Leptospira fainei sp. nov., isolated from pigs in Australia. Int. J. Syst. Bacteriol. 48 (Pt 3), 851–858. https://doi.org/10.1099/00207713-48-3-851.

Picardeau, M., 2017. Virulence of the zoonotic agent of leptospirosis: still terra incognita? Nat. Rev. Microbiol. 15, 297–307. https://doi.org/10.1038/nrmicro.2017.5.

Polo, N., Machado, G., Rodrigues, R., Nájera Hamrick, P., Munoz-Zanzi, C., Pereira, M. M., Bercini, M., Timm, L.N., Schneider, M.C., 2019. A one health approach to investigating Leptospira serogroups and their spatial distributions among humans and animals in Rio Grande do Sul, Brazil, 2013–2015. Trop. Med. Infect. Dis. 4, 42. https://doi.org/10.3390/tropicalmed4010042.

Reitstetter, R.E., 2006. Development of species-specific PCR primer sets for the detection of Leptospira. FEMS Microbiol. Lett. 264, 31–39. https://doi.org/10.1111/j.1574-6968.2006.00431.x.

Ren, S.-X., Fu, G., Jiang, X.-G., Zeng, R., Miao, Y.-G., Xu, H., Zhang, Y.-X., Xiong, H., Lu, G., Lu, L.-F., Jiang, H.-Q., Jia, J., Tu, Y.-F., Jiang, J.-X., Gu, W.-Y., Zhang, Y.-Q., Cai, Z., Sheng, H.-H., Yin, H.-F., Zhang, Y., Zhu, G.-F., Wan, M., Huang, H.-L., Qian, Z., Wang, S.-Y., Ma, W., Yao, Z.-J., Shen, Y., Qiang, B.-Q., Xia, Q.-C., Guo, X.-K., Danchin, A., Saint Girons, I., Somerville, R.L., Wen, Y.-M., Shi, M.-H., Chen, Z., Xu, J.-G., Zhao, G.-P., 2003. Unique physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing. Nature 422, 888–893. https://doi.org/10.1038/nature01597.

Salaün, L., Mérien, F., Gurianova, S., Baranton, G., Picardeau, M., 2006. Application of multilocus variable-number tandem-repeat analysis for molecular typing of the agent of leptospirosis. J. Clin. Microbiol. 44, 3954–3962. https://doi.org/10.1128/JCM.00336-06.

Schreier, S., Doungchawee, G., Chadsuthi, S., Triampo, D., Triampo, W., 2013. Leptospirosis: current situation and trends of specific laboratory tests. Expert. Rev. Clin. Immunol. 9, 263–280. https://doi.org/10.1586/eci.12.110.

Simão, F.A., Waterhouse, R.M., Ioannidis, P., Kriventseva, E.V., Zdobnov, E.M., 2015. BUSCO: assessing genome assembly and annotation completeness with single-copy orthologs. Bioinformatics 31, 3210–3212. https://doi.org/10.1093/bioinformatics/btv351.

Stallman, N.D., 1984. Minutes of the meeting. Int. J. Syst. Bacteriol. 34, 258–259.

Tagliabue, S., Figarolli, B.M., D'Incau, M., Foschi, G., Gennero, M.S., Giordani, R., Giordani, R., Natale, A., Papa, P., Ponti, N., Scaltrito, D., Spadari, L., Vesco, G., Ruocco, L., 2016. Serological surveillance of leptospirosis in Italy: two-year national data (2010−2011). Vet. Ital. 52, 129–138. https://doi.org/10.12834/VetIt.58.169.2.

Terpstra, W.J., Organization, World Health, Society, International Leptospirosis (Eds.), 2003. Human Leptospirosis: Guidance for Diagnosis, Surveillance and Control. World Health Organization, Geneva.

Thaipadungpanit, J., Wuthiekanun, V., Chierakul, W., Smythe, L.D., Petkanchanapong, W., Limpaiboon, R., Apiwatanaporn, A., Slack, A.T., Suputtamongkol, Y., White, N.J., Feil, E.J., Day, N.P.J., Peacock, S.J., 2007. A dominant clone of Leptospira interrogans associated with an outbreak of human leptospirosis in Thailand. PLoS Negl. Trop. Dis. 1, e56 https://doi.org/10.1371/journal.pntd.0000056.

Vincent, A.T., Schiettekatte, O., Goarant, C., Neela, V.K., Bernet, E., Thibeaux, R., Ismail, N., Mohd Khalid, M.K.N., Amran, F., Masuzawa, T., Nakao, R., Amara Korba, A., Bourhy, P., Veyrier, F.J., Picardeau, M., 2019. Revisiting the taxonomy and evolution of pathogenicity of the genus Leptospira through the prism of genomics. PLoS Negl. Trop. Dis. 13, e0007270 https://doi.org/10.1371/journal.pntd.0007270.

Ye, J., Coulouris, G., Zaretskaya, I., Cutcutache, I., Rozen, S., Madden, T.L., 2012. Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction. BMC Bioinformat. 13, 134. https://doi.org/10.1186/1471-2105-13-134.

Zuerner, R.L., Bolin, C.A., 1997. Differentiation of Leptospira interrogans isolates by IS1500 hybridization and PCR assays. J. Clin. Microbiol. 35, 2612–2617. https://doi.org/10.1128/jcm.35.10.2612-2617.1997.