



OPEN ACCESS

EDITED BY

Tao Lin,
Baylor College of Medicine, United States

REVIEWED BY

Karina Caimi,
National Scientific and Technical Research
Council (CONICET), Argentina
Claudia Toma,
University of the Ryukyus, Japan

*CORRESPONDENCE

Ana L. T. O. Nascimento

✉ ana.nascimento@butantan.gov.br

Aline F. Teixeira

✉ aline.rteixeira@fundacaobutantan.org.br

†These authors have contributed equally to
this work

†These authors share senior authorship

RECEIVED 06 September 2024

ACCEPTED 15 November 2024

PUBLISHED 13 December 2024

CITATION

Foltran BB, Gaspar JP, Silva IRM, Pires HM,
Andrade FB, Costa GM, Paixao JEL,
Fernandes LGV, Teixeira AF and
Nascimento ALTO (2024) New insights into
the putative role of leucine-rich repeat
proteins of *Leptospira interrogans* and
their participation in host cell invasion:
an *in silico* analysis.
Front. Cell. Infect. Microbiol. 14:1492352.
doi: 10.3389/fcimb.2024.1492352

COPYRIGHT

© 2024 Foltran, Gaspar, Silva, Pires, Andrade,
Costa, Paixao, Fernandes, Teixeira and
Nascimento. This is an open-access article
distributed under the terms of the [Creative
Commons Attribution License \(CC BY\)](#). The
use, distribution or reproduction in other
forums is permitted, provided the original
author(s) and the copyright owner(s) are
credited and that the original publication in
this journal is cited, in accordance with
accepted academic practice. No use,
distribution or reproduction is permitted
which does not comply with these terms.

New insights into the putative role of leucine-rich repeat proteins of *Leptospira interrogans* and their participation in host cell invasion: an *in silico* analysis

Bruno B. Foltran^{1,2†}, João P. Gaspar^{1,2†}, Igor R. M. Silva^{1,2†},
Henrique M. Pires^{1,2}, Fernanda B. Andrade¹,
Giovanna M. Costa^{1,3}, Julia E. L. Paixao¹, Luis G. V. Fernandes⁴,
Aline F. Teixeira^{1*} and Ana L. T. O. Nascimento^{1†*}

¹Laboratório de Desenvolvimento de Vacinas, Instituto Butantan, São Paulo, Brazil, ²Programa de Pós-Graduação Interunidades em Biotecnologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, Brazil, ³Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo, Ribeirão Preto, Brazil, ⁴Infectious Bacterial Disease Research Unit, USDA Agricultural Research Service, National Animal Disease Center, Ames, IA, United States

Pathogenic *Leptospira* are spirochetes that cause leptospirosis, a worldwide zoonotic disease. Leptospirosis affects humans and animals, with approximately 1 million human infections and 60,000 deaths per year. The diversity of leptospiral strains and serovars allied to the fact that pathogenesis is not yet fully understood, make the development of an effective vaccine against leptospirosis a challenge. Outer membrane and secreted proteins are considered potential antigens since they play a vital role in mediating interactions with host molecules. Several domains or motifs have been reported to participate in the leptospiral infection process. Among them, leucine-rich repeat (LRR) proteins have been highlighted as attractive multipurpose proteins, exhibiting a broad spectrum of ligands and having a putative role in bacterial pathogenesis. Indeed, genome annotation of leptospiral species pointed out that LRR proteins are predominant in pathogenic strains, a feature that corroborates this hypothesis. A few LRR proteins of *L. santarosai*, *L. borgpetersenii* and *L. interrogans* have been studied and their possible role in virulence was proposed. Yet, a mechanistic and broad investigation of LRR proteins was not fully performed. In this review, a comprehensive *in silico* analysis of 21 LRR proteins of *L. interrogans* was performed in relation to structure, function, dynamics and virulent potential that will contribute to understanding the key role of these domains in the underlying mechanisms of leptospiral infection.

KEYWORDS

Leptospira, leptospirosis, LRR proteins, pathogenesis, *in silico* analysis

1 Introduction

The genus *Leptospira* comprises pathogenic and saprophytic species. Pathogenic *Leptospira* spp. are the etiological agents of leptospirosis, a widespread zoonosis, while saprophytic bacteria are free-living environmental organisms. To date, classification based on whole genome sequencing has identified more than 68 species of *Leptospira* described (Vincent et al., 2019; Fernandes et al., 2022), within pathogens (P) and saprophytes (S). These clades are further subdivided into subclades. The subclade P1 is divided in P1+, which includes high virulence pathogens and P1- composed by low-virulence pathogens. Species belong to P1+ group are isolated of mammals and responsible to cause infections in humans and animals. The subclade P2 includes those formerly described as the intermediate and similar the P1- group, being mostly of the species environmental isolates. The subclades S1 and S2 include saprophyte species, where the latter constitutes a new subclade defined after the isolation of new saprophytic species from environmental samples (Vincent et al., 2019; Giraud-Gatineau et al., 2024). Lipopolysaccharide (LPS) structural heterogeneity is the basis for the degree of antigenic variation observed among more than 250 identified serovars. Infection in humans occurs through direct contact with wild or domestic animals' urine or indirectly by exposure to contaminated soil or water (Adler and de la Peña Moctezuma, 2010).

The initial phase of leptospirosis is characterized by nonspecific flu-like symptoms such as fever, chills, headache and myalgia. The disease can evolve to a severe condition known as Weil's syndrome, corresponding to 5 – 15% of the reported cases (Bharti et al., 2003), and to leptospirosis pulmonary hemorrhage syndrome (LPHS), which is another severe manifestation of the disease that has been globally reported (McBride et al., 2005)

Although numerous efforts have been made to understand the pathogenic mechanisms involved during leptospiral infection, pathways associated with pathogenesis are not yet fully elucidated. To mitigate disease burden, the development of effective vaccines and diagnostic tests are critical. Outer membrane and secreted proteins are considered potential antigens since they can play a pivotal role in mediating interactions with glycosaminoglycans (GAGs), extracellular matrix (ECM) components, serum components and cell receptors such as cadherins and integrins, and they are readily accessible to host immune defenses. Thus, their characterization is an interesting strategy for elucidating the pathogenesis mechanism of *Leptospira* spp.

Another interesting way to narrow down the identification of proteins involved in bacterial pathogenesis is through the presence of domains and motifs in their structure, especially those that occur mostly in pathogens. Proteins comprising LRR motifs are very interesting, ubiquitous proteins that are present in procaryotes and eukaryotes, including humans (Ng et al., 2011). LRR motifs comprise 20 to 29 amino acid residues having a high proportion of leucine residues (Matsushima et al., 2000; Dolan et al., 2007), and they are found in versatile proteins with diverse functions and cellular locations. It has been reported that LRRs form a flexible

framework that can adapt a diversity of interactions, permitting an association of a broad group of ligands (Bell et al., 2003; Kedzierski et al., 2004). Several proteins having these motifs were found to be associated with innate immunity, such as in PAMPs (pathogen associated molecular patterns), and in Toll-like and NOD-like receptors (Inohara et al., 2005; Ng and Xavier, 2011). More recently, a study using mouse macrophages expressing an endogenous NLRP3 mutant lacking the LRR domain showed that removal of this domain reduced NLRP3 inflammasome stimulation in mouse macrophages (Duan et al., 2022). The majority of plant and animal immune receptors have an LRR domain, and various functions are attributed to this motif (Padmanabhan et al., 2009). Beyond innate immunity, they participate in wide-ranging functional processes, such as DNA repair, cell adhesion, signal transduction, development, transcription, and RNA processing (Ng et al., 2011).

From the microbial viewpoint, the cell surface-associated protein (BspA) of the human periodontopathogen *Bacteroides forsythus* encompasses 14 complete repeats of 23 amino acid residues with homology to LRR motifs. This LRR protein is predicted to mediate the binding of bacteria to ECM components and to clotting factors, and it probably has a role in bacterial colonization (Sharma et al., 1998). One of the most studied bacterial proteins containing LRR domains is represented by a family of 9 proteins called internalins of *Listeria monocytogenes*, a food-borne bacterium that causes gastroenteritis, meningitis, or abortion. Members of this family are associated with mammalian host cell invasion (Marino et al., 2000.) through interaction with different host receptors to induce infection of human cells, involving complex mechanisms that ultimately result in disease (Iretton et al., 2021). Other pathogens having virulence factors containing LRR domains are *Yersinia pestis* (YopM) (Leung et al., 1990), *Salmonella* Typhimurium (SspH1, SspH2 and SlrP) (Miao et al., 1999), and *Shigella flexneri* (IpaH) (Hartman et al., 1990; Fernandez-Prada et al., 2000)

In *Treponema pallidum*, the syphilis spirochete, an LRR protein (TpLRR) was identified, and although its function remains to be determined, the presence of LRR domains participating in protein-protein and/or protein-lipid interactions could facilitate the associations between molecules of the *T. pallidum* cell envelope (Shevchenko et al., 1997). LrrA, the cell surface-associated LRR, of *T. denticola*, was reported to play a role in the attachment to and infiltration of human epithelial cells, and in coaggregation with another periodontal pathogen *Tannerella forsythensis*. These properties of LrrA suggest this protein to be an important virulence factor for these oral spirochetes (Ikegami et al., 2004).

Genome annotation of *L. interrogans* identified at least 20 encoding LRR-containing proteins, while *L. borgpetersenii* has at least 5, and the saprophyte *L. biflexa* genome bears only one annotated LRR protein-encoding gene (Picardeau, 2017; Eshghi et al., 2019). In pathogenic *Leptospira* spp., several LRR proteins were studied. In *L. santarosai* serovar Shermani, the crystal structure of LRR20 protein, encoded by the gene LSS_11580, was resolved and the recombinant protein (rLRR20) bound to human epithelial cadherin (E-cadherin) (Hsu et al., 2021). In addition, the

recombinant LRR38 encoded by LSS_01692 interacted with fibronectin, collagen IV, and Toll-like receptor 2 (TLR2); furthermore, rLRR38 induced inflammation involving the NF- κ B and MAPK signal transduction pathways, suggesting its participation during infection (Hsu et al., 2021). Leptospiral LRR proteins were identified from the genome of *L. borgpetersenii* serogroup Sejroe, the main agent of bovine leptospirosis, and two recombinant proteins, rKU_Sej_LRR_2012M (2012) and rhKU_Sej_LRR_2271 (2271), had their immune protective activities evaluated in a hamster model of leptospirosis. rKU_Sej_LRR_2271 showed promising results, inducing high protective efficacy and tissue clearance after heterologous challenge, and it was suggested to be a potential vaccine candidate against animal leptospirosis (Prapong et al., 2022). The crystal structures of 4 LRR proteins of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 were solved, LIC_12234, LIC_10831, LIC_11098 and LIC_12759 (Miras et al., 2015), and two of them, rLIC_10831 and rLIC_11098 proteins, were further characterized; rLIC_10831 is a vascular endothelial (VE)- and E-cadherin-binding protein (Eshghi et al., 2019), while rLIC_11098 showed a broader spectrum of interaction with host components (Gaspar et al., 2024).

In this paper, we discuss the *in silico* analysis of 21 LRR proteins of *L. interrogans* in terms of structure, function, dynamics and virulent potential and provide new insights into their participation in pathogenicity mechanisms of pathogenic *Leptospira* spp.

2 Methods

2.1 *In silico* analysis of operon structure, cellular localization, signal peptide and conserved domains in the LRR proteins

The amino acid sequences of the 21 LRR proteins from *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 (Nascimento et al., 2004; Miras et al., 2015) were obtained from the NCBI database. Protein sequences were analyzed by SMART web server (<http://smart.embl-heidelberg.de>) (Letunic et al., 2021) to determine the presence of conserved domains. Signal peptide and lipoprotein predictions were assessed by SMART, SignalP v. 5.0 (<https://services.healthtech.dtu.dk/services/SignalP-5.0/>) and LipoP v. 1.0 (<https://services.healthtech.dtu.dk/services/LipoP-1.0/>) (Juncker et al., 2003) web servers. PSORTb 3.0 (<https://www.psорт.org/psортb/>) (Yu et al., 2010) and Cello v. 2.5 (<http://cello.life.nctu.edu.tw>) (Yu et al., 2004) web servers were used to determine cellular localization. The physicochemical parameters predicted molecular mass and theoretical pI were analyzed by the ProtParam tool, available at the website (<https://web.expasy.org/protparam/>) (Gasteiger et al., 2005). Operon analysis of 21 LRR sequences was performed by using the MicrobesOnline program (<http://www.microbesonline.org/operons/gnc267671.html>) from the genome of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 and the found data was compared with the results *in vitro* obtained by RNAseq (Zhukova et al., 2017).

2.2 Proteomics data analysis

The data of the LRR proteins from *L. interrogans* proteome analysis were also collected. For this purpose, the studies of Malmstrom et al. (2009) and (Sarma et al., 2021) were used. Malmstrom and collaborators used a strategy that combines mass spectrometry-based proteomic methods to determine the number of protein copies per cell, while Sarma et al. (2021) estimated the relative abundance of the proteins in the sample by using the iBAQ (intensity-based absolute quantification) algorithm.

2.3 Virulence prediction of LRR proteins and vaccine target prediction

VirulentPred 2.0 (<https://bioinfo.icgeb.res.in/virulent2/predict.html>) was used to predict bacterial virulence factors. This web server is based on a two-layer cascade support vector machine, responsible for predicting virulent proteins in reliance on a repertoire of experimentally validated virulent protein sequences (Sharma et al., 2023). The VaxiJen v2.0 web server (<https://www.ddg-pharmfac.net/vaxijen/VaxiJen/VaxiJen.html>) was used to predict possible protective antigens regardless of alignment, allowing the classification of antigens only by the physicochemical properties of the proteins (Doytchinova and Flower, 2007).

2.4 Structural analysis

Structural analysis of LRR proteins of *Leptospira* spp. were modelled using CollabFold v15.5 (Mirdita et al., 2022), which is based on the AlphaFold2 protein structure prediction algorithm - AlphaFold2/CollabFold (<https://colab.research.google.com/github/sokrypton/ColabFold/blob/main/AlphaFold2.ipynb>) (Jumper et al., 2021; Mirdita et al., 2022; Varadi et al., 2024). For comparison, the crystal structures of LIC_10831 (Miras et al., 2015; Eshghi et al., 2019) and the virulent factor InlB (PDB 1D0B) of *Listeria monocytogenes*, were used. The structures were aligned in the PyMOL Molecular Graphics System, Version 2.5.5 Schrödinger, LLC software (<https://pymol.org/-page-top>) using CE algorithm, and the root mean square deviation was used to classify proteins as homologous (under 1.0), similar (from 1.0 to 3.0) and different (above 3.0). Sequences were aligned using MAFFT algorithm in Jalview software (Waterhouse et al., 2009), for identification of homologous regions, and by the Basic Local Alignment Search Tool (BLAST) web service (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) (Altschup et al., 1990), to obtain query cover and percentage identity.

2.5 Prediction of molecular, biological and cellular function

Gene Ontology search was performed using Argot 2.5 web server (<https://www.medcomp.medicina.unipd.it/Argot2-5/>)

(Lavezzo et al., 2016), with a threshold of 200, annotating IDs and hits for molecular function, biological process and cellular component.

2.6 Conservation analysis of LRR proteins among *Leptospira* spp.

BLAST was used to align the LRR sequences of *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 against different species of *Leptospira* from the subclades P1, P2, S1 and S2. The coverage and identity percentages were used to calculate a conservation value (*c*-value) among the orthologs of LRR proteins. The *c*-value was expressed as an index between 0.0 (non-present) and 1.0 (conserved) (Lopes et al., 2019; Nascimento Filho et al., 2024), and was calculated as follows:

$$c \text{ value} = \frac{\text{identical} + \text{similar amino acids}}{2} \cdot \text{coverage}$$

C-values were used to generate the heatmap image, and data analysis was carried out using Seaborn version 0.13.2 (Waskom, 2021), NumPy version 1.24.0 (van der Walt et al., 2011), Panda's version 2.2.2 (Reback et al., 2020) and Matplotlib version 3.9.0 (Hunter, 2007) packages in Python 3.12.

3 Results and discussion

3.1 Conservation of LRR proteins among pathogenic, intermediate and saprophytic *Leptospira* species

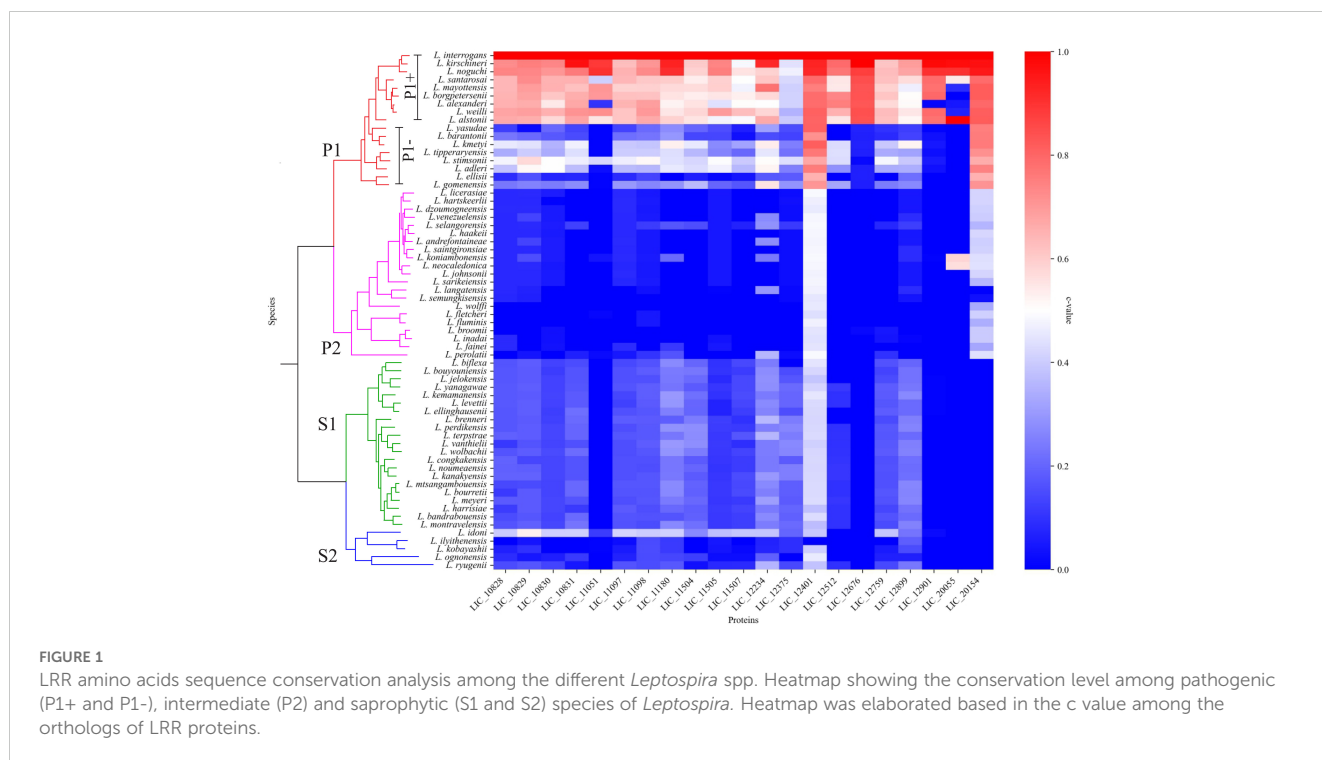
The 64 *Leptospira* species classified according to (Vincent et al., 2019) were used to determine the conservation of LRR proteins among the leptospiral pathogenic, intermediate and saprophytic groups. A total of 21 amino acid sequences were identified with the LRR domain encoded by *L. interrogans* strain Fiocruz L1-130 genes, as shown in Table 1. Sequences were compared using BLASTp against the predicted protein database of each *Leptospira* species, and a heatmap was created to visualize the conservation level among the different groups. As observed in Figure 1, the LRR proteins are well conserved among the pathogenic groups (subclade P1), exhibiting a higher *c*-value when compared with intermediate (subclade P2) and saprophytic strains (subclade S1 and S2), but only LIC_12401 and LIC_20154 showed identity greater than 70% for all pathogenic species (see Supplementary Table S1), except for *L. gomenensis*. Moreover, LIC_20154 is found only in pathogenic (*c*-value of 0.79) and intermediate species (*c*-value of 0.39), which can indicate an active role in the infection process. The coverage rate for most intermediate strains was very low, which shows a reduced identity among the sequences. For saprophytic strains, most LRR sequences showed a conservation level below 40% (Supplementary Table S1).

3.2 Features of the *L. interrogans* LRR proteins

The amino acid sequences of the 21 putative LRR proteins from *L. interrogans* (Miras et al., 2015) were analyzed *in silico* and their characteristics including number of amino acids, molecular mass, isoelectric point (pI), cellular localization and presence of domains and signal peptide are compiled in Table 1. The LRR proteins showed heterogeneous features. Amino acid composition varies from 122 to 1616 amino acids, which results in proteins with a range of molecular mass from 14 to 184 kDa. The majority of LRR proteins exhibits an isoelectric point above pH 7.0, which could suggest the importance of basic amino acid in the protein sequence. The isoelectric point of proteins produced by pathogenic bacteria is a crucial factor that can affect solubility, stability, interaction with host cells, evasion of the immune system, enzymatic activity and biofilm formation, all of which are fundamental aspects for bacterial pathogenesis (Schoorjans Stekhoven et al., 2008). The number of LRR units in each protein is also variable, ranging from 2 to 20 LRR domains, as observed by SMART program. A representative scheme of LRR domains can be visualized in Figure 2. In comparison to the prediction performed by Miras et al. (2015), the number of these domains was very similar. However, LIC_12676, LIC_20055 and LIC_20154 were not identified as LRR by SMART, as previously observed (Miras et al., 2015). In addition to LRR domains, the LRR proteins LIC_11051, LIC_12676 and LIC_12901 have a WGR domain and LIC_12901 also possesses a DUF4132 domain, which was previously reported by Sarma et al. (2021). Analysis of signal peptide was performed combining the results generated by SignalP 5.0 and LipoP 1.0. The results produced by the programs were divergent, since LipoP 1.0 was only able to recognize a signal peptide in 3 proteins (LIC_11098, LIC_11505, and LIC_11507), which were all identified as signal peptidase sequence SPI. The results obtained disagree with the SignalP 5.0 analysis, which predicted a SPII for LIC_11098 and LIC_11505. For LIC_11507, the prediction was the same, characterizing it as a SPI protein. Besides LIC_11507, the software did not identify any signal peptide for the LIC_10829, LIC_11051, LIC_11180, LIC_11504, LIC_12234, LIC_12375, LIC_12401, LIC_12676, LIC_12901, and LIC_20055. Comparing our results and the analysis performed by Miras and colleagues (2015), which also used the SignalP 3.0 version, some inconsistencies could be evidenced probably because of the database update of this program. When cellular localization analysis was performed using PSORTb, most LRR proteins were assigned as extracellular, where only LIC_20154, LIC_20055 and LIC_12676 were predicted as cytoplasmic proteins, which corroborates signal peptide analysis, except for LIC_20154. The predictions performed by Cello were diverse, as more than one location was predicted for most of the proteins (see Table 1). These distinct results may be due to the different database pool used by the programs. The confirmation of cellular location performed by *in vitro* assays was demonstrated to be extracellular only for LIC_10831, LIC_11098, LIC_11051 and LIC_11505 (Eshghi et al., 2019; Gaspar et al., 2024; Foltran et al., 2024). To verify whether the LRR sequences are in

TABLE 1 General features of LRR- proteins from *L. interrogans*.

<i>L. interrogans</i> Flocruz L1-130 Gene Locus	TrEMBL	ProtParam					SMART prediction		Signal P 5.0 Signal peptide (aa)	LipoP 1.0
		#AA	Mol mass (Da)	pI	PSORTb Localization	Cello Localization	LRR domains	Other domains		
LIC_10828	Q72U36	378	44104.89	9.12	EC	CYT, OM	12	X	1 to 24 (SPII)	Absent
LIC_10829	Q72U35	402	46312.73	9.23	EC	EC	17	X	Absent	Absent
LIC_10830	Q72U34	521	60165.64	5.85	EC	CYT, EC	17	X	1 to 24 (SPII)	Absent
LIC_10831	Q72U33	377	44197.53	8.94	EC	EC or CYT	13	X	1 to 29 (SPI)	Absent
LIC_11051	Q72TH0	685	78363.38	5.18	EC	OM or EC	7	WGR	Absent	Absent
LIC_11097	Q72TC4	413	48214.35	9.09	EC	CYT or OM	13	X	1 to 38 (SPI)	Absent
LIC_11098	Q72TC3	426	48971.95	8.97	EC	EC or CYT	16	X	1 to 24 (SPII)	1 to 29 (SPI)
LIC_11180	Q72T41	288	33433.13	9.38	EC	CYT	8	X	Absent	Absent
LIC_11504	Q72S80	266	31060.53	9.01	EC	CYT	6	X	Absent	Absent
LIC_11505	Q72S79	572	66278.22	8.82	EC	OM or CYT	17	X	1 to 24 (SPII)	1 to 29 (SPI)
LIC_11507	Q72S77	500	58441.47	9.04	EC	CYT	13	X	1 to 26 (SPI)	1 to 26 (SPI)
LIC_12234	Q72Q78	167	19190.18	5.35	EC	CYT	5	X	Absent	Absent
LIC_12375	Q72PU2	301	35232.28	10.03	EC	CYT or OM	6	X	Absent	Absent
LIC_12401	Q72PR6	217	24581.75	8.65	EC	CYT or OM	3	X	Absent	Absent
LIC_12512	Q72PF9	122	14155.89	9.26	Unknown	CYT	2	X	1 to 23 (SPI)	Absent
LIC_12676	Q72P01	657	76255.87	5.75	CYT	CYT	0	WGR	Absent	Absent
LIC_12759	Q72NS0	423	48762.96	8.82	EC	OM or CYT	14	X	1 to 23 (SPI)	Absent
LIC_12899	Q72ND5	272	31028.95	6.14	EC	CYT or PRP	7	X	1 to 17 (SPII)	Absent
LIC_12901	Q72ND3	1616	184691.17	6.21	EC	OM or EC	6	WGR and DUF4132	Absent	Absent
LIC_20055	Q75FX6	291	33310.26	9.00	CYT	CYT	0	X	Absent	Absent
LIC_20154	Q75FM8	241	28261.02	8.75	CYT	CYT	0	X	1 to 21 (SPII)	Absent



operon, a computational analysis for locating operon structures in *L. interrogans* genome was performed. Of all LRR, the sequences LIC_11097, LIC_11098 together with LIC_11096, the LIC_11180 linked to LIC_11179 and LIC_11181 and LIC_12401 in association to LIC_12400 and LIC_12402 were identified in operon. Interestingly, the genes “LIC_11096, LIC_11097, LIC_11098” and “LIC_11179, LIC_11180, LIC_11181” appear in a mechanism of translational coupling, as observed for the VapBC-1 module of *L. interrogans*, which result in a translation interdependence process (Damiano et al., 2024). Despite their sequential organization, LIC_10828, LIC_10829, LIC_10830, LIC_10831, and LIC_10504 and LIC_10505, these genes are not located in an operon, whereas each sequence is spaced by approximately 200pb, which possibly make the independent transcription process. These in silico analysis agree with the RNAseq analysis by Zhukova et al. (2017).

3.3 Proteome analysis of LRR proteins

Experimental data obtained from whole cell and extracellular proteome of *L. interrogans* are highlighted in Table 2. Malmstrom et al. (2009) identified all LRR proteins, but only LIC_11051 was quantified with an expected 309 copies/cell. When the copies/spectral count number was calculated, LIC_10828, LIC_11051, LIC_11098 and LIC_12676 were quantified, reaching 27, 132, 75 and 30 copies/spectral count, respectively. In Sarma et al. (2021), LIC_10828, LIC_10831, LIC_11051, LIC_11098, LIC_12676, LIC_12759 and LIC_12901 were identified from Triton X-114 fraction, which contain intracellular outer membrane and secreted proteins. The protein amount was estimated by iBAQ, with the range being from 2.4 to 52.8 iBAQ (Table 2). In the

secretory fraction, only LIC_11051 and LIC_12901 were detected, reaching 1.2 and 0.1 iBAQ, respectively. Taking into account that the most abundant leptospiral protein, LipL32, is found as having 38,050 copies/cell, according to Malmstrom et al. (2009) and 6388.2 iBAQ in the Triton X-114 fraction (Sarma et al., 2021), and proteins less abundant, such as, FlaB (2,038 copies/cell and 391.5 iBAQ) and DnaK (3,037 copies/cell and 4,128.8 iBAQ), according to Malmstrom et al. (2009) and Sarma et al. (2021), respectively, the data indicate that the LRR proteins occur in low copy numbers in *L. interrogans*.

3.4 Antigenicity and virulent factor prediction

Aiming to evaluate the potential of leptospiral LRR proteins as virulent factors and their potential as antigen molecules, the VirulentPred and VaxiJen v2.0 web servers were used. The virulence prediction identified all proteins as a virulent factor, except LIC_10830, assigned as non-virulent (Table 3). Although the involvement of proteins containing LRR domains in virulence of pathogens is expected, their roles in *L. interrogans* virulence has not been fully explored. *In vitro* data regarding interactions with cell receptors and host components are available for LIC_10831, LIC_11098, LIC_11051 and LIC_11505, suggesting their potential role in the leptospiral pathogenicity (Eshghi et al., 2019; Gaspar et al., 2024; Foltran et al., 2024). From the data obtained with these proteins, it seems that the interaction with host components correlates with the higher number of LRR domains present in the proteins (>7 LRR domains). When the LRR sequences were analyzed regarding their antigenicity, most of them were

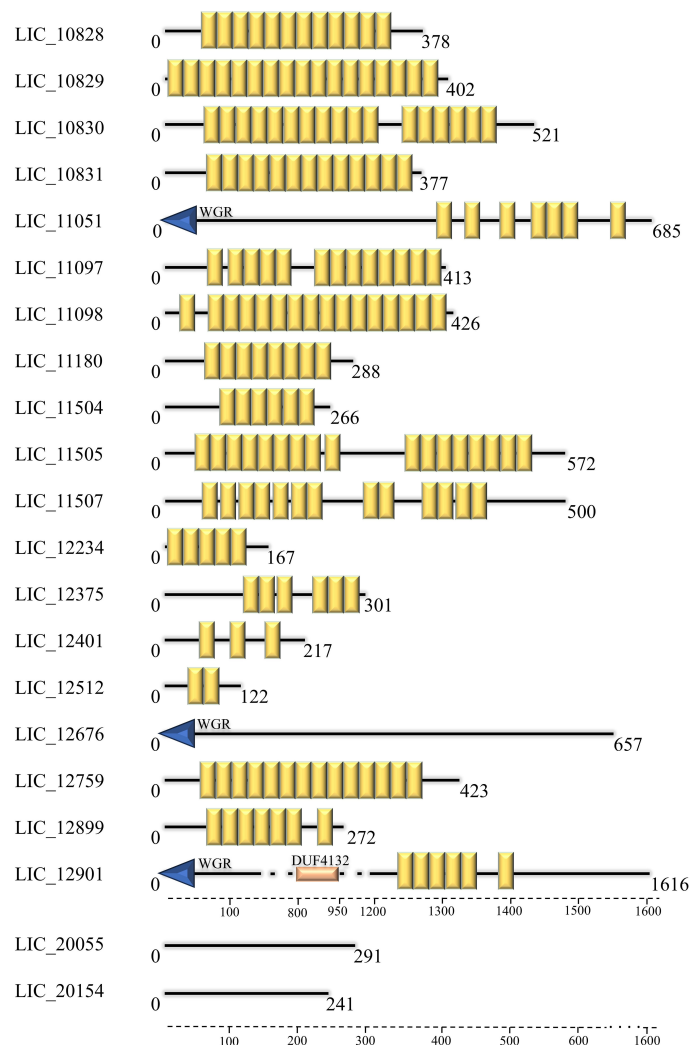


FIGURE 2
 Representative scheme and distribution of LRR domains. Schematic representation of LRR proteins from *L. interrogans* according to SMART webserver (<http://smart.embl-heidelberg.de>). Each sequence shown in yellow represents a LRR domain; WGR and DUF domains are also demonstrated. Scale is representative of the amino acids sequence.

TABLE 2 Proteome analysis of leptospiral LRR proteins; in grey, proteins characterized by Miras et al. (2015).

<i>L. interrogans</i> Fiocruz L1-130 Gene Locus	TrEMBL	Malmstrom et al. (2009)		Sarma et al. (2021)				
		Copies/ cell	Copies/ spectral counts	Triton X- 114 Fractions	Wash	Secretory	Abundance	Pathogenicity
LIC_10828	Q72U36	-	27	10.9	3.6	not found	0.3	P
LIC_10829	Q72U35	-	-	-	-	-	-	-
LIC_10830	Q72U34	-	-	-	-	-	-	-
LIC_10831	Q72U33	-	-	2.4	0.1	not found	0.1	P
LIC_11051	Q72TH0	309	132	25.1	0.4	1.2	0.1	P
LIC_11097	Q72TC4	-	37	-	-	-	-	-
LIC_11098	Q72TC3	-	75	26.2	5.5	not found	0.2	-

(Continued)

TABLE 2 Continued

<i>L. interrogans</i> Fiocruz L1-130 Gene Locus	TrEMBL	Malmström et al. (2009)		Sanna et al. (2021)				
		Copies/ cell	Copies/ spectral counts	Triton X- 114 Fractions	Wash	Secretory	Abundance	Pathogenicity
LIC_11180	Q72T41	-	-	-	-	-	-	-
LIC_11504	Q72S80	-	-	-	-	-	-	-
LIC_11505	Q72S79	-	-	-	-	-	-	-
LIC_11507	Q72S77	-	-	-	-	-	-	-
LIC_12234	Q72Q78	-	-	-	-	-	-	-
LIC_12375	Q72PU2	-	-	-	-	-	-	-
LIC_12401	Q72PR6	-	24	-	-	-	-	-
LIC_12512	Q72PF9	-	-	-	-	-	-	-
LIC_12676	Q72P01	-	30	14.3	0.1	not found	0	P
LIC_12759	Q72NS0	-	226	52.8	11.3	not found	0.2	P
LIC_12899	Q72ND5	-	-	-	-	-	-	-
LIC_12901	Q72ND3	-	64	15.4	0	0.1	0	P
LIC_20055	Q75FX6	-	-	-	-	-	-	-
LIC_20154	Q75FM8	-	-	-	-	-	-	-
LipL32	Q72SM7	38050	3793	6388.2	895.4	542.7	0.2	P
FlaB	Q72R59	2038	1156	391.5	8.5	7	0	P
DnaK	P61442	3037	3622	4128.8	390.3	39.6	0.1	NP

identified as non-antigenic. Only LIC_11051, LIC_11504, LIC_12401, LIC_12676, LIC_12901 and LIC_20154 showed scores above 0.4, being classified as possible antigens. Despite this prediction, the recombinant proteins LIC_10831 and LIC_12234 were able to stimulate specific antibodies in rabbit (Eshghi et al., 2019), while LIC_11098 and LIC_11505 were capable of eliciting a robust IgG antibody response in a mouse model (Gaspar et al., 2024; Foltran et al., 2024), denoting the antigenic potential of these LRR proteins. However, the prospective of these proteins as vaccine antigen were not yet evaluated. Only LBJ_2271 from *L. borgpetersenii*, which contain 74.19% identity with LIC_12401 from *L. interrogans*, was evaluated as recombinant vaccine candidate and showed a 75% protective efficacy in immunized hamsters submitted to heterologous challenge with *L. interrogans*. Despite these promising results, sterilizing immunity was not achieved (Praopong et al., 2022).

3.5 Comparative analysis of leptospiral LRR protein structure

Proteins containing LRR domains display curved solenoid structures that may present distinct types of folding being that

there are many types of solenoid folds in this protein family. Mainly, the concave side of the LRR domains is marked by a parallel β -sheet, and on the convex side, α -helices, 3_{10} helices, polyproline II helices, β -turns and short β -strands can be found (Bella et al., 2008). It seems that this topology contributes to ligand interaction, and such modules of binding were already characterized in the LRR protein domains (Price et al., 1998; Marino et al., 1999; Kobe and Kajava, 2001; Schubert et al., 2002). Aiming to analyze the leptospiral LRR proteins regarding their structure, we compared the LRR proteins of *L. interrogans* with LIC_10831, whose crystal structure has been solved, containing a distinct binding pocket that may be responsible for interaction with human E-cadherin. Thus, amino acid sequences were submitted to the AlphaFold program, and in Figure 3, it is possible to observe a spatial coincidence among the structures, even showing a different number of repeats units, denoting that the leptospiral LRRs seems to display a similar curvature radius. For some structures, such as LIC_10828, LIC_10829 and LIC_11097, which share close LRR domain quantities, the overlap was almost identical. In general, the LRR proteins possess a terminal structure that protects the hydrophobic core of the α/β solenoid. It has been observed that the four leptospiral proteins which had their solved structures (LIC_12234, LIC_10831, LIC_11098 and LIC_12759) have the same structural topology, which suggests that not only

TABLE 3 Antigenicity and virulent factor prediction.

Server	UniProt	ViulentPred2	VaxiJen v2.0	
Protein	Cod.	Prediction	Prediction	Score
LIC_10828	Q72U36	Virulent	ANTIGEN	0.1129
LIC_10829	Q72U35	Virulent	NON-ANTIGEN	0.1689
LIC_10830	Q72U34	Non-virulent	NON-ANTIGEN	0.1999
LIC_10831	Q72U33	Virulent	NON-ANTIGEN	0.2206
LIC_11051	Q72TH0	Virulent	ANTIGEN	0.4062
LIC_11097	Q72TC4	Virulent	NON-ANTIGEN	0.1817
LIC_11098	Q72TC3	Virulent	NON-ANTIGEN	0.1267
LIC_11180	Q72T41	Virulent	NON-ANTIGEN	0.1961
LIC_11504	Q72S80	Virulent	ANTIGEN	0.4189
LIC_11505	Q72S79	Virulent	NON-ANTIGEN	0.3018
LIC_11507	Q72S77	Virulent	NON-ANTIGEN	0.2944
LIC_12234	Q72Q78	Virulent	NON-ANTIGEN	0.0469
LIC_12375	Q72PU2	Virulent	NON-ANTIGEN	0.2964
LIC_12401	Q72PR6	Virulent	ANTIGEN	0.5094
LIC_12512	Q72PF9	Virulent	NON-ANTIGEN	0.1143
LIC_12676	Q72P01	Virulent	ANTIGEN	0.4507
LIC_12759	Q72NS0	Virulent	NON-ANTIGEN	0.3229
LIC_12899	Q72ND5	Virulent	NON-ANTIGEN	0.2898
LIC_12901	Q72ND3	Virulent	ANTIGEN	0.4569
LIC_20055	Q75FX6	Virulent	NON-ANTIGEN	0.1856
LIC_20154	Q75FM8	Virulent	ANTIGEN	0.4609

the stability of the α/β solenoid conformation is maintained but also that specific binding sites are promoted (Miras et al., 2015). Analyzing the N-terminal and C-terminal regions of other leptospiral LRR proteins, we can observe that all proteins share a similar structural topology. However, only LIC_10828, LIC_10829, LIC_10830, LIC_10831, LIC_11097, LIC_11098, LIC_11507 and LIC_12899 overlap identically in the N-terminal region, while for the C-terminal, the only proteins that do not show the same superposition are LIC_12901 and LIC_12676. Both regions contain some conserved residues that are important to maintain the structure, but the C-terminal portion has a greater number of conserved residues among the leptospiral LRRs, mainly arginine and lysine. We have also compared the leptospiral LRR protein structures with internalin B (InlB) from *Listeria monocytogenes*, which is one of the most studied LRR proteins (Marino et al., 1999). Superposition of InlB with the 21 *L. interrogans* LRR proteins shows a similar structural profile (Figure 4). Although the repeat unit of the InlB is smaller compared to most leptospiral LRRs, the same curvature radius was observed. LIC_12234, which shares close LRR

number to InlB, overlapped almost identically, differing from other LRR proteins showing a similar LRR domain number. As in *Listeria monocytogenes*, InlB belongs to the protein family that is associated with invasion of mammalian host cells (Marino et al., 2000), and it is possible that LIC_12234 may be involved in similar activity (see below).

3.6 Analysis of molecular, biological and cellular function by Gene Ontology

Aiming to compare the molecular function, the cellular components and the biological processes involved with the leptospiral LRR proteins, the 21 amino acid sequences were analyzed by the Gene Ontology program. As observed in Figure 5A and Table 4, analysis related to the molecular function identified 17 proteins associated with GTP binding, which is a regulator protein responsible for cell signaling. That function was also identified in the LRR proteins InlB from *Listeria monocytogenes* and SLR from *S. pyogenes* (Supplementary Table S2). Among the GTP-binding protein families in eukaryotes, the proteins of the Rho family play a pivotal role in host cell cytoskeleton dynamics. It participates in signal transduction pathways related to B and T lymphocyte activities, leukocyte migration and phagocytosis and phagocyte degranulation (Aktories, 2011; El Masri and Delon, 2021). The participation of proteins that mimic or covalently modify the host's GTP-binding protein regulator has already been described for some pathogenic bacteria. For example, included are YopE from *Yersinia* spp. and SopE and SopE2 from *S. Typhimurium*, which act as activators of GTP-binding regulators, and YopT from *Yersinia* spp., which cleaves the host GTPases (Just et al., 2001; Aktories, 2011); 15 of the LRR proteins were identified to have adenylate cyclase activity, which is the enzyme responsible for synthesizing cyclic AMP (cAMP) and 14 were described as ubiquitin-protein transferase activity, which is involved in the regulation of other proteins. In the LRR proteins of other pathogens, this transferase activity related to ubiquitin-protein complex was identified in YopM from *Y. pestis*, SlrP from *S. Typhimurium* and LrrA from *T. denticola* (Supplementary Table S2). Seven proteins showing transferase activity were also identified. A low number of LRR proteins were associated with protein kinase activity, lyase activity, phosphoprotein, phosphatase activity, cadherin binding, DNA ligase (ATP) activity, ligase activity and ATP binding. Ligase activity was also observed in YopM from *Y. pestis* and SlrP from *S. Typhimurium* (Wei et al., 2016; Bullones-Bolaños et al., 2024). Additionally, YopM from *Y. pestis*, LrrG from *S. agalactiae*, LrrA from *T. denticola* and TpLRR from *T. pallidum* were associated with ion metallic binding particularly with zinc (Ikegami et al., 2004; Uittenbogaard et al., 2012; Sullivan et al., 2021). Although none of the LRR proteins from *L. interrogans* were associated with ion binding function, it has been described that these proteins are known to bind to divalent cations, such as Zn^{2+} and Ca^{2+} (Miras et al., 2015).

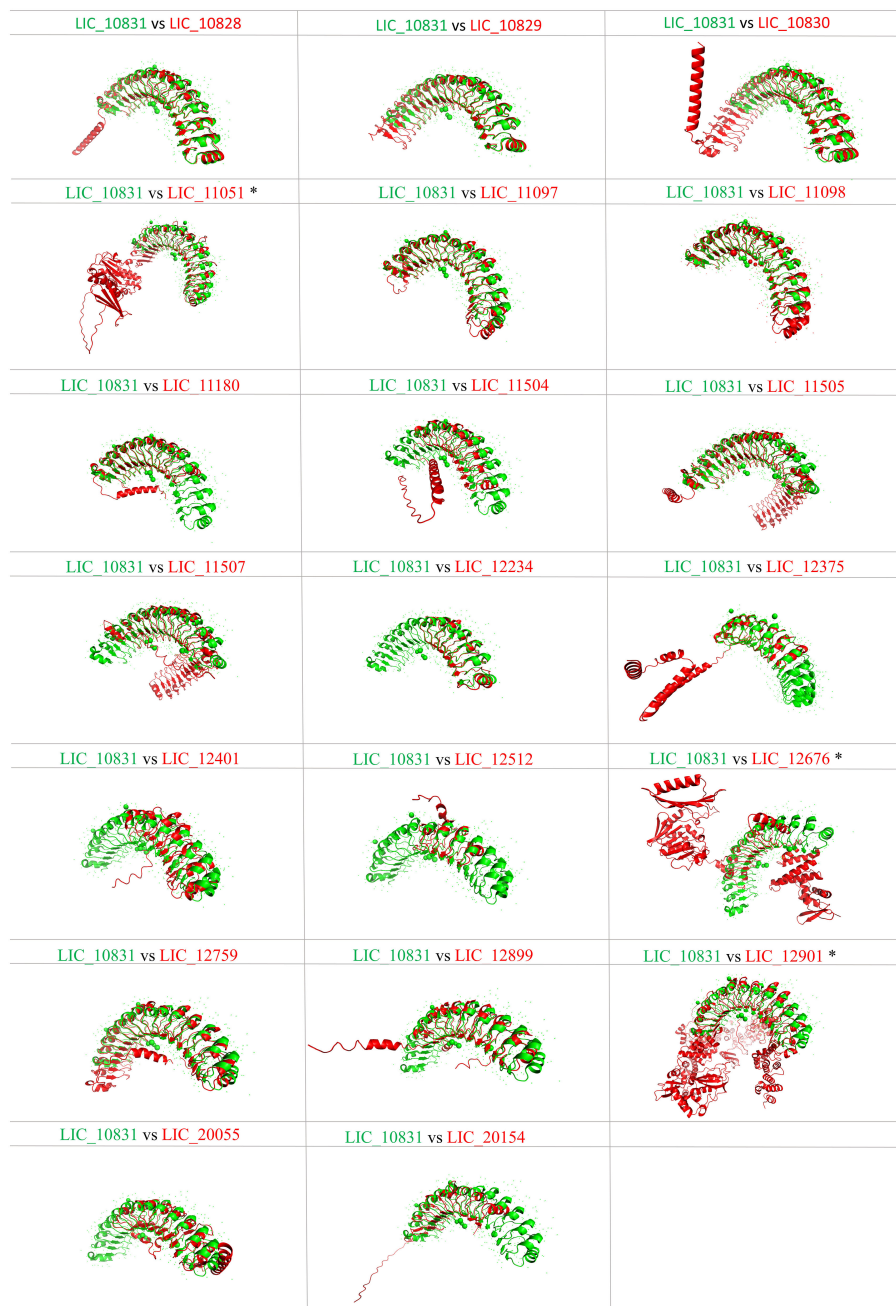


FIGURE 3

Tertiary structure prediction of *L. interrogans* LRR proteins and spatial alignment with LIC_10831. Amino acid sequences of proteins LIC_10828, LIC_10829, LIC_10830, LIC_11051, LIC_11097, LIC_11098, LIC_11180, LIC_11504, LIC_11507, LIC_11505, LIC_12234, LIC_12375, LIC_12401, LIC_12512, LIC_12676, LIC_12759, LIC_12899, LIC_12901, LIC_20055, LIC_20154 were submitted to the AlphaFold2 program, and the best score was selected. Superposition of LIC_10831 (green) with the other LRR proteins (red) was performed by PyMOL software. (*) refers to the presence of WGR domain.

When the cellular components were evaluated, 16 LRR proteins were described as intracellular, 6 of them as membrane proteins and 5 proteins as integral component of the membrane. InlB, YopM, SLR, LrrG, LrrA and TplRR were membrane associated, while all other LRR proteins were also classified as integral components of the membrane, except YopM. Only 2 LRR proteins have been identified as protein of the nucleus, and one protein as nucleoplasm, cell junction, cytoplasm and scrib-APC-beta-catenin complex (Figure 5B; Table 4). In relation

to biological process, 17 proteins were associated with small GTPase-mediated signal transduction. That function was also observed for InlB from *Listeria monocytogenes* and SLR from *S. pyogenes*. A total of 7 proteins seemed to be involved with cAMP biosynthetic processes, while 9 proteins were related to the protein ubiquitination process. This ubiquitination process seems to occur for the proteins YopM, SlrP and LrrA (Ikegami et al., 2004; Wei et al., 2016; Bullones-Bolaños et al., 2024). Ubiquitination is a reversible post-translational process, usually



FIGURE 4

Tertiary structure prediction of *L. interrogans* LRR proteins and spatial alignment with InlB of *L. monocytogenes*. Amino acid sequences of proteins LIC_10828, LIC_10829, LIC_10830, LIC_10831, LIC_11051, LIC_11097, LIC_11098, LIC_11180, LIC_11504, LIC_11505, LIC_12234, LIC_12375, LIC_12401, LIC_12512, LIC_12676, LIC_12759, LIC_12899, LIC_12901, LIC_20055, LIC_20154. Superposition of InlB (green) with the other LRR proteins (red) was performed by PyMOL software. (*) refers to the presence of WGR domain.

associated with protein-protein interactions, innate immune signaling, proteasome regulation, cell autophagy and xenophagy (Ashida et al., 2014; Vozandychova et al., 2021). To survive in the host cell, some bacteria have to subvert the host cell machinery. For example, it has been shown that *S. Typhimurium* produces 2 effector proteins, SseL and AvrA, that have deubiquitinating function and are responsible for inhibiting the NF- κ B signaling pathway, which is related to the expression of proinflammatory cytokine genes. Also, *S. pyogenes* produces an effector protein, SpeB, which is a cysteine protease that

degrades components of the ubiquitination process, thus avoiding bacterial ubiquitination and host cell autophagy, and *Listeria monocytogenes* possess InlK, that has the ability to bind a host cytoplasmatic protein, avoiding ubiquitination and consequently the xenophagy process (Vozandychova et al., 2021). Nine proteins were identified in signal transduction and 8 in the protein phosphorylation processes. Only one LRR protein out of 21 analyzed sequences were identified as DNA ligand or involved in the recombination and repair processes (Figure 5C; Table 4).

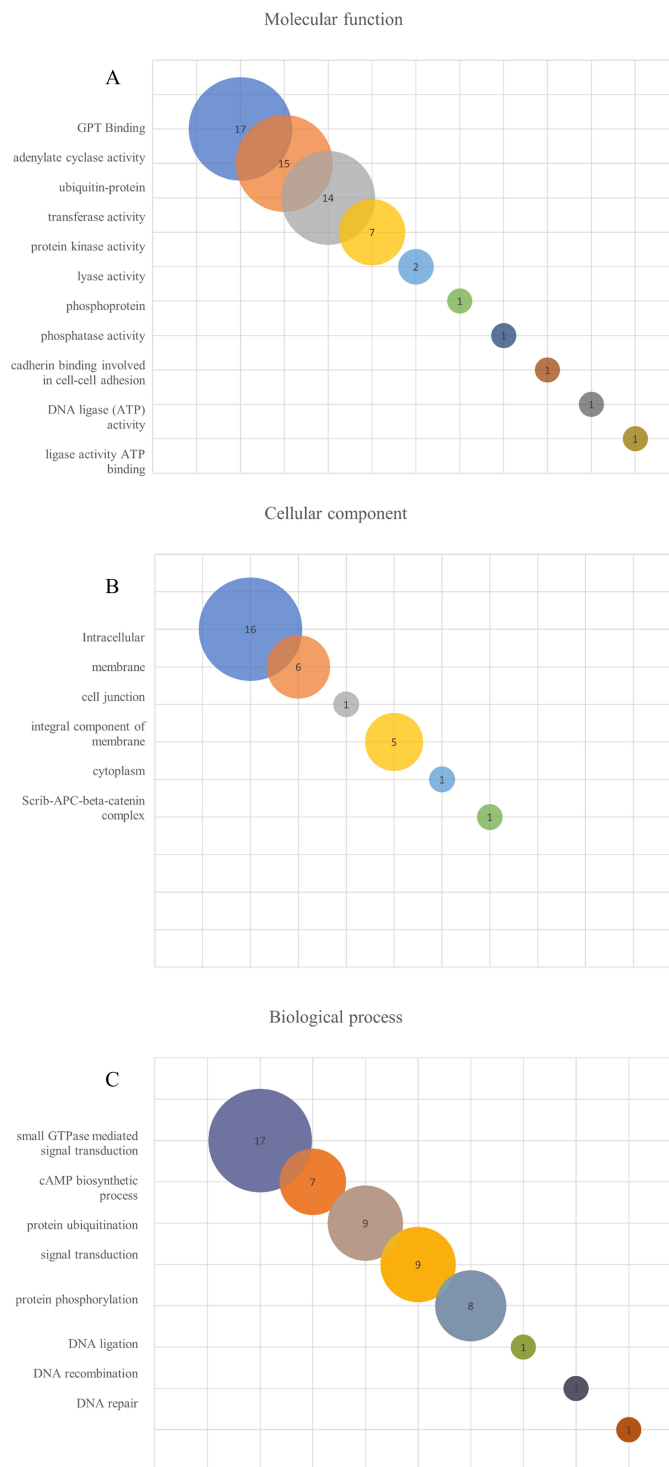


FIGURE 5 Gene ontology (GO) enrichment analysis of *L. interrogans* LRR proteins. The search was performed using Argot2.5 tool and a threshold of 200 was considered to determine hits for molecular function (A), cellular component (B) and biological process (C) among the LRR proteins.

4 Concluding remarks

LRRs are versatile binding domains present in a diversity of proteins that can accommodate a vast type of ligands. Because LRRs are mainly involved in protein-protein interaction, it can be inferred

that they are associated with host-pathogen interactions and are used by either the host or pathogen. The comprehensive analysis focused on 21 LRR proteins of *L. interrogans* indicates that they have the potential to have a diverse role in bacteria, which can be extended to other pathogenic *Leptospira* spp. by their orthologous

TABLE 4 Gene ontology (GO) analysis of *L. interrogans* LRR proteins showing the prediction for molecular function, cellular component and biological process among the leptospiral LRR proteins.

Protein	Molecular Function			Biological Process			Cellular Component		
	ID	Name	Score	ID	Name	Score	ID	Name	Score*
LIC_10828	GO:0005525	GTP binding	1.770	GO:0007264	small GTPase mediated signal transduction	2.801	GO:0005622	intracellular membrane	1.110 206
	GO:0004016	adenylate cyclase activity	619	GO:0006171	cAMP biosynthetic process	893	GO:0016020		
	GO:0004842	ubiquitin-protein transferase activity	521	GO:0016567	protein ubiquitination	398			
				GO:0007165	signal transduction	258			
LIC_10829	GO:0005525	GTP binding	1.677	GO:0007264	small GTPase mediated signal transduction	2.241	GO:0005622	intracellular nucleus nucleoplasm cell junction integral component of membrane membrane	835 253 248 230 228 210
	GO:0004016	adenylate cyclase activity	450	GO:0007165	signal transduction	411	GO:0005634		
	GO:0004842	ubiquitin-protein transferase activity	403	GO:0006468	protein phosphorylation	251	GO:0005654		
	GO:0004672	protein kinase activity	253				GO:0030054		
							GO:0016021		
LIC_10830	GO:0005525	GTP binding	1.265	GO:0007264	small GTPase mediated signal transduction	2.411	GO:0005622	intracellular integral component of membrane membrane	969 609 470
	GO:0004672	protein kinase activity	482	GO:0006468	protein phosphorylation	595	GO:0016021		
	GO:0004842	ubiquitin-protein transferase activity	381	GO:0007165	signal transduction	465	GO:0016020		
	GO:0004016	adenylate cyclase activity	361						
LIC_10831	GO:0005525	GTP binding	1.727	GO:0007264	small GTPase mediated signal transduction	2.691	GO:0005622	intracellular	1.183 - - - - -
	GO:0004016	adenylate cyclase activity	860	GO:0006171	cAMP biosynthetic process	1.194			
	GO:0004842	ubiquitin-protein transferase activity	391	GO:0006468	protein phosphorylation	289			
	GO:0016829	lyase activity	330	GO:0007165	signal transduction	261			
	GO:0004672	protein kinase activity	221	GO:0016567	protein ubiquitination	232			
LIC_11051	GO:0005525	GTP binding	751	GO:0007264	small GTPase mediated signal transduction	1.045	GO:0005622	intracellular	544 - -
	GO:0004016	adenylate cyclase activity	402	GO:0006171	cAMP biosynthetic process	515			
				GO:0006468	protein phosphorylation	216			
LIC_11097	GO:0005525	GTP binding	1.958	GO:0007264	small GTPase mediated signal transduction	3.066	GO:0005622	intracellular integral component of membrane membrane	1.101 319 221
	GO:0004016	adenylate cyclase activity	505	GO:0016567	protein ubiquitination	335	GO:0016021		
	GO:0004842	ubiquitin-protein transferase activity	441	GO:0007165	signal transduction	258	GO:0016020		
LIC_11098	GO:0005525	GTP binding	2.103	GO:0007264	small GTPase mediated signal transduction	3.231	GO:0005622	intracellular integral component of membrane membrane	1.271 409 284
	GO:0004016	adenylate cyclase activity	570	GO:0016567	protein ubiquitination	385	GO:0016021		
	GO:0004842	ubiquitin-protein transferase activity	555	GO:0007165	signal transduction	359	GO:0016020		
	GO:0004672	protein kinase activity	206	GO:0006468	protein phosphorylation	227			
LIC_11180	GO:0005525	GTP binding	947	GO:0007264	small GTPase mediated signal transduction	1.117	GO:0005622	Intracellular	262
	GO:0004016	adenylate cyclase activity	523						
	GO:0004842	ubiquitin-protein transferase activity	298						
LIC_11504	GO:0005525	GTP binding	1.322	GO:0007264	small GTPase mediated signal transduction	2.172	GO:0005622	Intracellular	524 - - -
	GO:0004016	adenylate cyclase activity	527	GO:0006171	cAMP biosynthetic process	741			
	GO:0004842	ubiquitin-protein transferase activity	268	GO:0016567	protein ubiquitination	404			
	GO:0016829	lyase activity	204						

(Continued)

TABLE 4 Continued

Protein	Molecular Function			Biological Process			Cellular Component		
	ID	Name	Score	ID	Name	Score	ID	Name	Score*
LIC_11505	GO:0005525	GTP binding	867	GO:0007264	small GTPase mediated signal transduction	1.686	GO:0005622	intracellular	744
	GO:0004672	protein kinase activity	528	GO:0006468	protein phosphorylation	594	GO:0016021	integral component of membrane	448
	GO:0004842	ubiquitin-protein transferase activity	400	GO:0007165	signal transduction	525	GO:0016020	membrane	311
	GO:0004721	phosphoprotein phosphatase activity	228				GO:0005737	cytoplasm	220
	GO:0005634					GO:0005634	nucleus	210	
LIC_11507	GO:0005525	GTP binding	990	GO:0007264	small GTPase mediated signal transduction	1.379	GO:0005622	Intracellular	754
	GO:0004016	adenylate cyclase activity	346	GO:0007165	signal transduction	384			
	GO:0004842	ubiquitin-protein transferase activity	264	GO:0006468	protein phosphorylation	273			
	GO:0004672	protein kinase activity	234						
LIC_12334	GO:0005525	GTP binding	1.129	GO:0007264	small GTPase mediated signal transduction	1.747	GO:0005622	Intracellular	475
	GO:0004016	adenylate cyclase activity	400	GO:0006171	cAMP biosynthetic process	537			
	GO:0004842	ubiquitin-protein transferase activity	331	GO:0016567	protein ubiquitination	429			
LIC_12375	GO:0005525	GTP binding	1.037	GO:0007264	small GTPase mediated signal transduction	1.927	GO:0005622	Intracellular	484
	GO:0004016	adenylate cyclase activity	474	GO:0006171	cAMP biosynthetic process	707	-	-	-
	GO:0004842	ubiquitin-protein transferase activity	251	GO:0016567	protein ubiquitination	348			
LIC_12401	GO:0098641	cadherin binding involved in cell-cell adhesion	302	-	-	-	-	-	-
LIC_12512	-	-	-	-	-	-	-	-	-
LIC_12676	GO:0003910	DNA ligase (ATP) activity	855	GO:0006266	DNA ligation	390	-	-	-
	GO:0016874	ligase activity	467	GO:0006310	DNA recombination	274			
	GO:0005524	ATP binding	339	GO:0006281	DNA repair	262			
LIC_12759	GO:0005525	GTP binding	1.554	GO:0007264	small GTPase mediated signal transduction	2.374	GO:0005622	Intracellular	1.174
	GO:0004842	ubiquitin-protein transferase activity	509	GO:0007165	signal transduction	378			
	GO:0004016	adenylate cyclase activity	400	GO:0016567	protein ubiquitination	308			
	GO:0004672	protein kinase activity	226	GO:0006468	protein phosphorylation	245			
LIC_12899	GO:0005525	GTP binding	1.275	GO:0007264	small GTPase mediated signal transduction	2.237	GO:0005622	Intracellular	613
	GO:0004842	ubiquitin-protein transferase activity	443	GO:0006171	cAMP biosynthetic process	613			
	GO:0004016	adenylate cyclase activity	420	GO:0016567	protein ubiquitination	526			
LIC_12901	GO:0005525	GTP binding	882	GO:0007264	small GTPase mediated signal transduction	728	GO:0034750	Scrib-APC-beta-catenin complex	220
	GO:0004016	adenylate cyclase activity	422	-	-	-	GO:0035748	myelin sheath abaxonal region	208
	GO:0005622					GO:0005622	intracellular	205	
LIC_20055	-	-	-	-	-	-	-	-	-
LIC_20154	GO:0005525	GTP binding	385	GO:0007264	small GTPase mediated signal transduction	356	-	-	-

* It was considered a threshold of 200.

(-) indicates that GO could not associate any data with the protein sequence.

proteins. The results suggest that they could be important for cellular invasion and bacterial survival by altering the host cells' signaling pathways, suggesting a potential role in infection. Taken together, the data indicate that leptospiral LRR proteins warrant further studies to elucidate their participation in pathophysiological mechanisms, including the selection of targets for mutagenesis in pathogenic *Leptospira*, as a proof of concept of LRR proteins' role in virulence.

Author contributions

BF: Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review & editing. JG: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing, Writing – original draft. IS: Data curation, Formal analysis, Investigation, Software, Validation, Visualization, Writing – review & editing, Writing – original draft. HP: Data curation, Formal analysis, Investigation, Methodology, Visualization, Writing – review & editing, Writing – original draft. FA: Data curation, Formal analysis, Investigation, Validation, Visualization, Writing – review & editing, Writing – original draft. GC: Data curation, Investigation, Validation, Visualization, Writing – review & editing, Writing – original draft. JP: Data curation, Formal analysis, Investigation, Visualization, Writing – review & editing, Writing – original draft. LF: Data curation, Formal analysis, Investigation, Visualization, Writing – original draft, Writing – review & editing. AT: Conceptualization, Data curation, Formal analysis, Methodology, Visualization, Writing – original draft, Writing – review & editing. AN: Conceptualization, Formal analysis, Funding acquisition, Project administration, Supervision, Visualization, Writing – original draft, Writing – review & editing.

Funding

The author(s) declare financial support was received for the research, authorship, and/or publication of this article. This work

References

- Adler, B., and de la Peña Moctezuma, A. (2010). Leptospira and leptospirosis. *Vet. Microbiol.* 140, 287–296. doi: 10.1016/j.vetmic.2009.03.012
- Aktories, K. (2011). Bacterial protein toxins that modify host regulatory GTPases. *Nat. Rev. Microbiol.* 9, 487–498. doi: 10.1038/nrmicro2592
- Altschup, S. F., Gish, W., Miller, W., Myers, E. W., and Lipman, D. J. (1990). Basic local alignment search tool. *J. Mol. Biol.* 215, 403–410. doi: 10.1016/S0022-2836(05)80360-2
- Ashida, H., Kim, M., and Sasakawa, C. (2014). Exploitation of the host ubiquitin system by human bacterial pathogens. *Nat. Rev. Microbiol.* 12, 399–413. doi: 10.1038/nrmicro3259
- Bell, J. K., Mullen, G. E. D., Leifer, C. A., Mazzoni, A., Davies, D. R., and Segal, D. M. (2003). Leucine-rich repeats and pathogen recognition in Toll-like receptors. *Trends Immunol.* 24, 528–533. doi: 10.1016/S1471-4906(03)00242-4
- Bella, J., Hindle, K. L., McEwan, P. A., and Lovell, S. C. (2008). The leucine-rich repeat structure. *Cellular and Molecular Life Sciences. Cell Mol. Life Sci.* 65, 2307–2333. doi: 10.1007/s00018-008-8019-0
- Bharti, A. R., Nally, J. E., Ricaldi, J. N., Matthias, M. A., Diaz, M. M., Lovett, M. A., et al. (2003). Leptospirosis: a zoonotic disease of global importance. *Lancet Infect. Dis.* 3, 757–771. doi: 10.1016/S1473-3099(03)00830-2
- Bullones-Bolaños, A., Martín-Muñoz, P., Vallejo-Grijalba, C., Bernal-Bayard, J., and Ramos-Morales, F. (2024). Specificities and redundancies in the NEL family of bacterial E3 ubiquitin ligases of *Salmonella enterica* serovar Typhimurium. *Front. Immunol.* 15. doi: 10.3389/fimmu.2024.1328707
- Damiano, D. K., Azevedo, B. O. P., Fernandes, G. S. C., Teixeira, A. F., Gonçalves, V. M., Nascimento, A. L. T. O., et al. (2024). The toxin of vapbc-1 toxin-antitoxin module from *Leptospira interrogans* is a ribonuclease that does not arrest bacterial growth but affects cell viability. *Microorganisms* 12, 1660–1674. doi: 10.3390/microorganisms12081660
- Dolan, J., Walshe, K., Alsbury, S., Hokamp, K., O'Keefe, S., Okafuji, T., et al. (2007). The extracellular Leucine-rich repeat superfamily; a comparative survey and analysis of evolutionary relationships and expression patterns. *BMC Genomics* 8, 320–344. doi: 10.1186/1471-2164-8-320

was financially supported by Fundação de Amparo à Pesquisa do Estado de São Paulo (FAPESP), São Paulo, Brazil (grants 2019/17488-2, 2022/16481-7, 2022/08985-5, 2022/12101-5, 2023/08486-1 to ALTON, BBF, IRMS, FBA and JPG, respectively), Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq) São Paulo, Brazil (grants304445/2021-5 and 126322/2023-6 to ALTON and GMC), and Fundação Butantan, Sao Paulo, Brazil. The funders had no role in the study design, collection and data analysis, decision of the publication or preparation of this manuscript.

Acknowledgments

We are deeply indebted to Dr. Al Leyva, BS, for his critical reading and English editing of this manuscript.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fcimb.2024.1492352/full#supplementary-material>

- Doytchinova, I. A., and Flower, D. R. (2007). VaxiJen: A server for prediction of protective antigens, tumour antigens and subunit vaccines. *BMC Bioinf.* 8, 1–7. doi: 10.1186/1471-2105-8-4
- Duan, Y., Wang, J., Cai, J., Kelley, N., and He, Y. (2022). The leucine-rich repeat (LRR) domain of NLRP3 is required for NLRP3 inflammasome activation in macrophages. *J. Biol. Chem.* 298, 102717–102728. doi: 10.1016/j.jbc.2022.102717
- El Masri, R., and Delon, J. (2021). RHO GTPases: from new partners to complex immune syndromes. *Nat. Rev. Immunol.* 21, 499–513. doi: 10.1038/s41577-021-00500-7
- Eshghi, A., Gaultney, R. A., England, P., Brùlé, S., Miras, I., Sato, H., et al. (2019). An extracellular *Leptospira* interrogans leucine-rich repeat protein binds human E- and VE-cadherins. *Cell Microbiol.* 21, 12949–12978. doi: 10.1111/cmi.12949
- Fernandes, L. G. V., Stone, N. E., Roe, C. C., Goris, M. G. A., van der Linden, H., Sahl, J. W., et al. (2022). *Leptospira sanjuanensis* sp. nov., a pathogenic species of the genus *Leptospira* isolated from soil in Puerto Rico. *Int. J. Syst. Evol. Microbiol.* 72, 1–9. doi: 10.1099/ijsem.0.005560
- Fernandez-Prada, C. M., Hoover, D. L., Tall, B. D., Hartman, A. B., Kopelowitz, J., and Venkatesan, M. M. (2000). *Shigella flexneri* ipaH 7.8 facilitates escape of virulent bacteria from the endocytic vacuoles of mouse and human macrophages. *Infect. Immun.* 68, 3608–3619. doi: 10.1128/IAI.68.6.3608-3619.2000
- Foltran, B. B., Texeira, A. F., Romero, E. C., Fernandes, L. G. V., and Nascimento, A. L. T. O. (2024). Leucine-rich repeat proteins of *Leptospira* interrogans that interact to host glycosaminoglycans and integrins. *Front. Microbiol.* 15. doi: 10.3389/fmicb.2024.1497712
- Gaspar, J. P., Takahashi, M. B., Teixeira, A. F., and Nascimento, A. L. T. O. (2024). In silico analysis and functional characterization of a leucine-rich repeat protein of *Leptospira* interrogans. *Int. J. Med. Microbiol.* 316, 151633. doi: 10.1016/j.ijmm.2024.151633
- Gasteiger, E., Hoogland, C., Gattiker, A., Duvaud, S., Wilkins, M. R., Appel, R. D., et al. (2005). *Protein identification and analysis tools on the exPASy server The proteomics protocols handbook* (Totowa, NJ: Humana Press), 571–607. doi: 10.1385/1-59259-890-0:571
- Giraud-Gatineau, A., Nieves, C., Harrison, L. B., Benaroudj, N., Veyrier, F. J., and Picardeau, M. (2024). Evolutionary insights into the emergence of virulent *Leptospira* spirochetes. *PLoS Pathog.* 20. doi: 10.1371/journal.ppat.1012161
- Hartman, A. B., Venkatesan, M., Oaks, E. V., Buysse, J. M., Buysse, J. M., Stover, C. K., et al. (1990). Sequence and Molecular Characterization of a Multicopy Invasion Plasmid Antigen Gene, ipaH, of *Shigella flexneri*. *J. Bacteriol.* 172, 1905–1915. doi: 10.1128/jb.172.4.1905-1915.1990
- Hsu, S.-H., Chou, L.-F., Hong, C.-H., Chang, M.-Y., Tsai, C.-Y., Tian, Y.-C., et al. (2021). Crosstalk between E-Cadherin/ β -Catenin and NF- κ B Signaling Pathways: The Regulation of Host-Pathogen Interaction during Leptospirosis Crosstalk between E-Cadherin/ β -Catenin and NF- κ B Signaling Pathways: The Regulation of Host-Pathogen Interaction during. *Leptospirosis. Int. J. Mol. Sci.* 22, 13132–13155. doi: 10.3390/ijms
- Ikegami, A., Honma, K., Sharma, A., and Kuramitsu, H. K. (2004). Multiple functions of the leucine-rich repeat protein LrrA of *Treponema denticola*. *Infect. Immun.* 72, 4619–4627. doi: 10.1128/IAI.72.8.4619-4627.2004
- Inohara, N., Chamailard, M., McDonald, C., and Nuñez, G. (2005). NOD-LRR proteins: Role in host-microbial interactions and inflammatory disease. *Annu. Rev. Biochem.* 74, 355–383. doi: 10.1146/annurev.biochem.74.082803.133347
- Ireton, K., Mortuza, R., Gyanwali, G. C., Gianfelice, A., and Hussain, M. (2021). Role of internalin proteins in the pathogenesis of *Listeria monocytogenes*. *Mol. Microbiol.* 116, 1407–1419. doi: 10.1111/mmi.14836
- Jumper, J., Evans, R., Pritzel, A., Green, T., Figurnov, M., Ronneberger, O., et al. (2021). Highly accurate protein structure prediction with AlphaFold. *Nature* 596, 583–589. doi: 10.1038/s41586-021-03819-2
- Juncker, A. S., Willenbrock, H., von Heijne, G., Brunak, S., Nielsen, H., and Krogh, A. (2003). Prediction of lipoprotein signal peptides in Gram-negative bacteria. *Protein Sci.* 12, 1652–1662. doi: 10.1110/ps.0303703
- Just, I., Hofmann, F., Genth, H., and Gerhard, R. (2001). Bacterial protein toxins inhibiting low-molecular-mass GTP-binding proteins. *Int. J. Med. Microbiol.* 291, 243–250. doi: 10.1078/1438-4221-00127
- Kedzierski, L., Montgomery, J., Curtis, J., and Handman, E. (2004). Leucine-rich repeats in host-pathogen interactions. *Arch. Immunol. Ther. Exp. (Warsz)* 52, 104–112. Available at: <http://www.ncbi.nlm.nih.gov/pubmed/15179324>.
- Kobe, B., and Kajava, A. V. (2001). The leucine-rich repeat as a protein recognition motif. *Curr. Opin. Struct. Biol.* 11, 725–732. doi: 10.1016/S0959-440X(01)00266-4
- Lavezzo, E., Falda, M., Fontana, P., Bianco, L., and Toppo, S. (2016). Enhancing protein function prediction with taxonomic constraints - The Argot2.5 web server. *Methods* 93, 15–23. doi: 10.1016/j.ymeth.2015.08.021
- Letunic, I., Khedkar, S., and Bork, P. (2021). SMART: Recent updates, new developments and status in 2020. *Nucleic Acids Res.* 49, D458–D460. doi: 10.1093/nar/gkaa937
- Leung, K. Y., Reisner, B. S., and Straley, S. C. (1990). YopM inhibits platelet aggregation and is necessary for virulence of *Yersinia pestis* in mice. *Infect Immun.* 58, 3262–3271. doi: 10.1128/iai.58.10.3262-3271.1990
- Lopes, A. P. Y., Azevedo, B. O. P., Emidio, R. C., Damiano, D. K., Nascimento, A. L. T. O., and Barazzone, G. C. (2019). In silico analysis of genetic vappc profiles from the toxin-antitoxin type ii vappc modules among pathogenic, intermediate, and non-pathogenic *leptospira*. *Microorganisms* 7, 56–74. doi: 10.3390/microorganisms7020056
- Malmstrom, J., Beck, M., Schmidt, A., Lange, V., Deutsch, E. W., and Aebersold, R. (2009). Proteome-wide cellular protein concentrations of the human pathogen *Leptospira interrogans*. *Nature* 460, 762–765. doi: 10.1038/nature08184
- Marino, M., Braun, L., Cossart, P., and Ghosh, P. (1999). Structure of the inlB leucine-rich repeats, a domain that triggers host cell invasion by the bacterial pathogen *L. monocytogenes*. *Mol. Cell* 4, 1063–1072. doi: 10.1016/S1097-2765(00)80234-8
- Marino, M., Braun, L., Cossart, P., and Ghosh, P. (2000). A framework for interpreting the leucine-rich repeats of the *Listeria internalins*. *Proc. Natl. Acad. Sci.* 97, 8784–8788. doi: 10.1073/pnas.97.16.8784
- Matsushima, N., Ohyanagi, T., Tanaka, T., and Kretsinger, R. H. (2000). Super-motifs and evolution of tandem leucine-rich repeats within the small proteoglycans - Biglycan, decorin, lumican, fibromodulin, PRELP, keratan, osteoadherin, epiphycan, and osteoglycin. *Proteins: Structure Funct. Genet.* 38, 210–225. doi: 10.1002/(SICI)1097-0134(20000201)38:2<210::AID-PROT9>3.0.CO;2-1
- McBride, A. J., Athanazio, D. A., Reis, M. G., and Ko, A. I. (2005). Leptospirosis. *Curr. Opin. Infect. Dis.* 18, 376–386. doi: 10.1097/01.qco.0000178824.05715.2c
- Miao, E. A., Scherer, C. A., Tsois, R. M., Kingsley, R. A., Adams, L. G., Bäumlner, A. J., et al. (1999). Salmonella typhimurium leucine-rich repeat proteins are targeted to the SPI1 and SPI2 type III secretion systems. *Mol. Microbiol.* 34, 850–864. doi: 10.1046/j.1365-2958.1999.01651.x
- Miras, I., Saul, F., Nowakowski, M., Weber, P., Haouz, A., Shepard, W., et al. (2015). Structural characterization of a novel subfamily of leucine-rich repeat proteins from the human pathogen *Leptospira interrogans*. *Acta Crystallogr. D Biol. Crystallogr.* 71, 1351–1359. doi: 10.1107/S139900471500704X
- Mirdita, M., Schütze, K., Moriwaki, Y., Heo, L., Ovchinnikov, S., and Steinegger, M. (2022). ColabFold: making protein folding accessible to all. *Nat. Methods* 19, 679–682. doi: 10.1038/s41592-022-01488-1
- Nascimento, A. L. T. O., Verjovski-Almeida, S., Van Sluys, M. A., Monteiro-Vitorello, C. B., Camargo, L. E. A., Digimapietri, L. A., et al. (2004). Genome features of *Leptospira interrogans* serovar Copenhageni. *Braz. J. Med. Biol. Res.* 37, 459–477. doi: 10.1590/S0100-879X2004000400003
- Nascimento Filho, E. G., Vieira, M. L., Dias, M., Mendes, M. A., Sanchez, F. B., Setubal, J. C., et al. (2024). Global proteome of the saprophytic strain *Leptospira biflexa* and comparative analysis with pathogenic strain *Leptospira interrogans* uncover new pathogenesis mechanisms. *J. Proteomics* 297, 105125. doi: 10.1016/j.jpro.2024.105125
- Ng, A. C. Y., Eisenberg, J. M., Heath, R. J. W., Huett, A., Robinson, C. M., Nau, G. J., et al. (2011). Human leucine-rich repeat proteins: A genome-wide bioinformatic categorization and functional analysis in innate immunity. *Proc. Natl. Acad. Sci. U.S.A.* 108, 4631–4638. doi: 10.1073/pnas.1000093107
- Ng, A., and Xavier, R. J. (2011). Leucine-rich repeat (LRR) proteins: Integrators of pattern recognition and signaling in immunity. *Autophagy* 7, 1082–1084. doi: 10.4161/auto.7.9.16464
- Padmanabhan, M., Cournoyer, P., and Dinesh-Kumar, S. P. (2009). The leucine-rich repeat domain in plant innate immunity: A wealth of possibilities. *Cell Microbiol.* 11, 191–198. doi: 10.1111/j.1462-5822.2008.01260.x
- Picardeau, M. (2017). Virulence of the zoonotic agent of leptospirosis: Still terra incognita? *Nat. Rev. Microbiol.* 15, 297–307. doi: 10.1038/nrmicro.2017.5
- Prapong, S., Tansiri, Y., Sritrakul, T., Sripattanakul, S., Sopitthummakun, A., Katzenmeier, G., et al. (2022). *Leptospira borgpetersenii* leucine-rich repeat proteins provide strong protective efficacy as novel leptospiral vaccine candidates. *Trop. Med. Infect. Dis.* 8, 6. doi: 10.3390/tropicalmed8010006
- Price, S. R., Evans, P. R., and Nagai, K. (1998). Crystal structure of the spliceosomal U2B'-U2A' protein complex bound to a fragment of U2 small nuclear RNA. *Nature* 394, 645–650. doi: 10.1038/29234
- Sarma, A., Gunasekaran, D., Rex, D. A. B., Sikha, T., Phukan, H., Kiran, K. M., et al. (2021). Extracellular proteome analysis shows the abundance of histidine kinase sensor protein, DNA helicase, putative lipoprotein containing peptidase M75 domain and peptidase C39 domain protein in *leptospira interrogans* grown in EMJH medium. *Pathogens* 10, 852. doi: 10.3390/pathogens10070852
- Schubert, W.-D., Urbanke, C., Ziehm, T., Beier, V., Machner, M. P., Domann, E., et al. (2002). Structure of internalin, a major invasion protein of *listeria monocytogenes*, in complex with its human receptor E-cadherin. *Cell* 111, 825–836. doi: 10.1016/S0092-8674(02)01136-4
- Schuermans Stekhoven, F. M. A. H., Gorissen, M. H. A. G., and Flik, G. (2008). The isoelectric point, a key to understanding a variety of biochemical problems: A minireview. *Fish Physiol. Biochem.* 34, 1–8. doi: 10.1007/s10695-007-9145-6
- Sharma, A., Garg, A., Ramana, J., and Gupta, D. (2023). VirulentPred 2.0: An improved method for prediction of virulent proteins in bacterial pathogens. *Protein Sci.* 32, 4808–4818. doi: 10.1002/pro.4808
- Sharma, A., Sojar, H. T., Glurich, I., Honma, K., Kuramitsu, H. K., and Genco, R. J. (1998). Cloning, expression, and sequencing of a cell surface antigen containing a leucine-rich repeat motif from *Bacteroides forsythus* ATCC 43037. *Infect. Immun.* 66, 5703–5710. doi: 10.1128/IAI.66.12.5703-5710.1998
- Shevchenko, D. V., Akins, D. R., Robinson, E., Li, M., Popova, T. G., Cox, D. L., et al. (1997). Molecular characterization and cellular localization of tpLRR, a processed leucine-rich repeat protein of *treponema pallidum*, the syphilis spirochete. *J. Bacteriol.* 179, 3188–3195. doi: 10.1128/jb.179.10.3188-3195.1997
- Sullivan, M. J., Goh, K. G. K., and Ulett, G. C. (2021). Cellular management of zinc in group B streptococcus supports bacterial resistance against metal intoxication and promotes disseminated infection. *mSphere* 6, 105–121. doi: 10.1128/mSphere.00105-21

- Uittenbogaard, A. M., Chelvarajan, R. L., Myers-Morales, T., Gorman, A. A., Brickey, W. J., Ye, Z., et al. (2012). Toward a molecular pathogenic pathway for *Yersinia pestis* YopM. *Front. Cell Infect. Microbiol.* 2. doi: 10.3389/fcimb.2012.00155
- van der Walt, S., Colbert, S. C., and Varoquaux, G. (2011). The numPy array: A structure for efficient numerical computation. *Comput. Sci. Eng.* 13, 22–30. doi: 10.1109/MCSE.2011.37
- Varadi, M., Bertoni, D., Magana, P., Paramval, U., Pidruchna, I., Radhakrishnan, M., et al. (2024). AlphaFold Protein Structure Database in 2024: providing structure coverage for over 214 million protein sequences. *Nucleic Acids Res.* 52, D368–D375. doi: 10.1093/nar/gkad1011
- Vincent, A. T., Schiettekatte, O., Goarant, C., Neela, V. K., Bernet, E., Thibeaux, R., et al. (2019). Revisiting the taxonomy and evolution of pathogenicity of the genus *Leptospira* through the prism of genomics. *PLoS Negl. Trop. Dis.* 13, 7270–7295. doi: 10.1371/journal.pntd.0007270
- Vozandychova, V., Stojkova, P., Hercik, K., Rehulka, P., and Stulik, J. (2021). The ubiquitination system within bacterial host–pathogen interactions. *Microorganisms* 9, 638. doi: 10.3390/microorganisms9030638
- Waskom, M. (2021). seaborn: statistical data visualization. *J. Open Source Softw.* 6, 3021. doi: 10.21105/joss.03021
- Waterhouse, A. M., Procter, J. B., Martin, D. M. A., Clamp, M., and Barton, G. J. (2009). Jalview Version 2-A multiple sequence alignment editor and analysis workbench. *Bioinformatics* 25, 1189–1191. doi: 10.1093/bioinformatics/btp033
- Wei, C., Wang, Y., Du, Z., Guan, K., Cao, Y., Yang, H., et al. (2016). The *Yersinia* Type III secretion effector YopM Is an E3 ubiquitin ligase that induced necrotic cell death by targeting NLRP3. *Cell Death Dis.* 7, e2519–e2519. doi: 10.1038/cddis.2016.413
- Yu, C.-S., Lin, C.-J., and Hwang, J.-K. (2004). Predicting subcellular localization of proteins for Gram-negative bacteria by support vector machines based on n-peptide compositions. *Protein Sci.* 13, 1402–1406. doi: 10.1110/ps.03479604
- Yu, N. Y., Wagner, J. R., Laird, M. R., Melli, G., Rey, S., Lo, R., et al. (2010). PSORTb 3.0: Improved protein subcellular localization prediction with refined localization subcategories and predictive capabilities for all prokaryotes. *Bioinformatics* 26, 1608–1615. doi: 10.1093/bioinformatics/btq249
- Zhukova, A., Fernandes, L. G., Hugon, P., Pappas, C. J., Sismeiro, O., Coppée, J. Y., et al. (2017). Genome-wide transcriptional start site mapping and sRNA identification in the pathogen *Leptospira interrogans*. *Front. Cell Infect. Microbiol.* 7. doi: 10.3389/fcimb.2017.00010