Université du Québec
Institut National de la Recherche Scientifique
Centre Armand-Frappier Santé Biotechnologie

Identification of genetic determinants associated with host adaptation and serovar determination in bacteria of the genus *Leptospira*

Par

Cecilia Jacqueline Nieves Álvarez

Thèse présenté(e) pour l'obtention du grade de
Philosophiæ Doctor (Ph.D.) en Biologie

Jury d'évaluation

Président du jury et examinateur interne	Charles Dozois INRS – Centre Armand-Frappier Santé Biotechnologie
Examinateur externe	Abdoulaye Baniré Diallo Université du Québec à Montréal
Examinateur externe	Roman Thibeaux Institut Pasteur – Nouvelle-Calédonie
Directeur de recherche	Frédéric Veyrier INRS – Centre Armand-Frappier Santé Biotechnologie

Acknowledgments

I would like to express my deepest gratitude to my advisor, Frédéric Veyrier, for his guidance, invaluable support, and mentorship throughout my doctoral journey. His expertise, insightful feedback, and encouragement have been instrumental in shaping my research and academic development.

I extend my heartfelt appreciation to the members of my examining committee for their generosity and time in agreeing to evaluate this thesis. Their expertise is undoubtedly invaluable in ensuring the quality and rigor of my research.

I am immensely grateful to all the collaborators for their invaluable contributions and dedication to the different scientific publications included in this thesis. Their expertise and insights have played a pivotal role in the achieved outcomes and working with them has been an enriching experience.

I would like to acknowledge the members of Veyrier's laboratory for their collaboration, camaraderie, and intellectual stimulation. I am grateful to each member, past and present, for their invaluable contributions, fruitful discussions, and unwavering support.

To my husband, Noé, words cannot adequately express my appreciation for his love, encouragement, and understanding throughout this challenging journey. His patience, belief in me, and continuous support have been my constant source of strength.

To my loyal companion dog, Brisa, who has been by my side throughout the writing of my Bachelor, Master, and PhD theses. Her wagging tail and comforting presence have brought me solace during the long hours of research and writing.

I am grateful to my family for their love, support, and belief in my abilities. Their encouragement has been an endless source of motivation. Thank you for always being there for me.

I would also like to express special gratitude to my dear friends, both in Uruguay and from around the world. I am grateful for the countless moments we shared, and the bonds we formed.

This thesis would not have been possible without all of you. Thank you from the bottom of my heart.

Abstract

Leptospira, a bacterial genus within the phylum *Spirochaetes,* comprises both free-living and pathogenic organisms. The pathogenic strains are responsible for leptospirosis, a disease that affects various hosts, including humans. Recent advancements in genomics have led to a doubling in the number of identified species, prompting changes in *Leptospira* classification.

This thesis offers a comprehensive review of the taxonomic classification of *Leptospira*, focusing on general genome features and recent genomic findings. It also presents a detailed genomic characterization of all 68 species reported until 2021, highlighting acquired and lost genes associated with the emergence of pathogenic leptospires. Moreover, this thesis also covers amino acid permutations linked to the transition from saprophytic to pathogenic forms, which have not been studied before.

The intricate serovar classification is also addressed. Serovar identity is linked to the composition of the carbohydrate portion of the surface-exposed lipopolysaccharide *O*-antigen, with its biosynthesis-encoding genes lying within the *rfb* cluster. This thesis provides a thorough comparative analysis of the *rfb* clusters across various serovars, including 12 pathogenic strains of *L. noguchii* sequenced here. The analyses reveal a consistent correspondence between gene composition and serovar identity, which could potentially facilitate direct genetic typing of *Leptospira* serovars.

Lastly, this thesis presents a phylogenomic analysis of the pathogenic species *L. santarosai*, showing distinctive genomic features between *L. santarosai* and other pathogenic species. These findings suggest an ancient speciation of pathogens and their adaptation to diverse ecological niches. The analysis also shows a high genetic diversity of *L. santarosai* strains derived from both patients and animals, with clonal groups showing strong associations with specific geographical areas. This provides valuable insights into the genomic diversity, evolutionary history, and epidemiology of leptospirosis in the Americas and globally.

Overall, this thesis emphasizes the relevance of accurately identifying *Leptospira* strains for epidemiological, phylogenetic, and diagnostic purposes. Furthermore, it underscores the significance of genomic analyses in enhancing our understanding of the diversity and evolution of this bacterial genus.

Résumé

Leptospira, un genre bactérien du phylum *Spirochaetes*, comprend des organismes libres et pathogènes. Les souches pathogènes sont responsables de la leptospirose, une maladie qui affecte divers hôtes, y compris les humains. Les avancées récentes en génomique ont doublé le nombre d'espèces identifiées, entraînant des modifications de la classification de *Leptospira*.

Cette thèse présente une revue exhaustive de la classification taxonomique de *Leptospira*, centrée sur les caractéristiques générales du génome et les découvertes récentes en génomique. Elle offre aussi une caractérisation détaillée des génomes des 68 espèces recensées jusqu'en 2021, soulignant les gènes acquis et perdus associés à l'émergence de leptospires pathogènes. De plus, cette thèse couvre les permutations des acides aminés liées à la transition des saprophytes à un mode de vie pathogène, un aspect qui n'avait pas été étudié auparavant.

La classification complexe des sérovars est également abordée. L'identité des sérovars est liée à la composition de la partie glucidique de l'antigène *O*-lipopolysaccharide exposé en surface, dont les gènes de biosynthèse se trouvent dans le cluster *rfb*. Cette thèse propose une analyse comparative approfondie du cluster *rfb* dans différents sérovars, dont 12 souches pathogènes de *L. noguchii* séquencées ici. Les analyses révèlent une concordance entre la composition des gènes et l'identité des sérovars, ce qui peut faciliter le typage génétique direct des sérovars de *Leptospira*.

Enfin, cette thèse présente une analyse phylogénomique de l'espèce pathogène *L. santarosai*, qui met en évidence des caractéristiques génomiques distinctes entre *L. santarosai* et les autres espèces pathogènes. Ces résultats suggèrent une spéciation ancienne des pathogènes et leur adaptation à diverses niches écologiques. L'analyse montre également une grande diversité génétique des souches de *L. santarosai* provenant de patients et d'animaux, avec des groupes clonaux présentant de fortes associations avec des zones géographiques spécifiques. Cela offre des informations précieuses sur la diversité génomique, l'histoire évolutive et l'épidémiologie de la leptospirose en Amérique et dans le monde.

En résumé, cette thèse souligne l'importance d'identifier avec précision les souches de *Leptospira* à des fins épidémiologiques, phylogénétiques et diagnostiques. De plus, elle met en évidence le rôle crucial des analyses génomiques dans l'amélioration de notre compréhension de la diversité et de l'évolution de ce genre bactérien.

Thesis outline

Chapter 1 is **introductory** and includes a general description of *Leptospira* and leptospirosis. General morphological features of the bacteria are highlighted, along with a brief taxonomic description. The principal pathogenicity determinants are presented, such as several outer membrane proteins, the lipopolysaccharide, and the flagellar machinery. Towards the end, it emphasizes the relevance of this thesis and provides the **hypothesis and study objectives**.

Chapter 2 is a **book chapter titled "Taxonomy and Phylogenomics of *Leptospira*"**. It covers from the general description of pre-genomic classification of *Leptospira* to a classification based entirely on genomes. It explores typing methods, starting from the classical serology-based approaches to those based on sequencing and phylogenetic inference using specific genes. Genomic architecture and gene composition in different taxonomic groups are also discussed. This chapter is currently **accepted (in press) for publication** in ***Phylogenomics: Foundations, Methods, and Pathogen Analysis (Elsevier, Academic Press).***

Chapter 3 includes the results of **comparative genomics** among d**ifferent taxonomic groups of *Leptospira* spp.** ($P1^+$, $P1^-$, P2, S1, S2). It includes a functional characterization of their pangenome, as well as the representation of pseudogenes and paralogy and their relationship with the **emergence of pathogenic *Leptospira***. Additionally, amino acid changes related to the development of a highly virulent phenotype are analyzed. All **these results have not been published or submitted yet** but some of them are part of an ongoing collaboration that specifically examines changes associated with the gain of virulence in $P1^+$ (manuscript in preparation).

Chapter 4 is a **research article titled "Horizontal transfer of the *rfb* cluster in *Leptospira* is a genetic determinant of serovar identity"**. It incudes an initial comparative genome analysis of *L. noguchii* strains from different serovars, which served as the starting point for expanding the study of serovar variability in other species. A comparative genomics analysis of *rfb* clusters and links to serovar determination are presented. Analysis of synteny within and around the cluster is also conducted, along with an examination of other genomic features. This article was submitted to ***Life Science Alliance*** on April 12, 2022, and **published online** on December 9, 2022.

Chapter 5 is a **research article titled "Phylogenomics of *Leptospira santarosai,* a prevalent pathogenic species in the Americas".** It focuses on the study of genomic features of *L. santarosai,* considering previously reported genomes and new ones from clinical isolates. A comparative analysis with other highly virulent species is also performed. Association between serogroups/serovars and specific hosts/geographical areas is evaluated, along with comparative analyses of the *rfb* genes to infer the distribution of serovars in uncharacterized strains. This article was submitted to ***PLoS Neglected Tropical Diseases*** on September 15, 2023, and **published online** on November 2, 2023.

Chapter 6 starts with a **general discussion** that encompasses the three main topics presented in the thesis. Firstly, it covers the genetic events that determine the evolution of *Leptospira* towards the emergence of pathogenic species. Secondly, it discusses the link between specific gene content and serovar determination, as well as the association between serovar and infection of particular hosts. Thirdly, it delves into insights on the genetic variability and the evolution towards host-dependent species on highly virulent *Leptospira*. The end of the chapter comprises the general **conclusions and perspectives** that have emerged from this thesis or parallel analyses that have been conducted.

Although not included in the thesis, these are other articles in which I have contributed during my PhD:

- Veyrier FJ, **Nieves C**, Lefrancois LH, Trigui H, Vincent AT, Behr MA. RskA Is a Dual Function Activator-Inhibitor That Controls SigK Activity Across Distinct Bacterial Genera. Front Microbiol. 2020 Sep 9;11:558166. doi: 10.3389/fmicb.2020.558166. PMID: 33013790; PMCID: PMC7509140.

Author contributions

FV and MB, study designed | FV, **CN**, LL, HT, and AV, experiments and analysis | FV, **CN**, AV and MB, writing and editing.

- Grillová L, Robinson MT, Chanthongthip A, Vincent AT, **Nieves C**, Oppelt J, Mariet JF, Lorioux C, Vongsouvath M, Mayxay M, Phonemeexay O, Rattanavong S, Phommasone K, Douangnouvong A, Šmajs D, Veyrier FJ, Newton PN, Picardeau M. Genetic diversity of *Leptospira* isolates in Lao PDR and genome analysis of an outbreak strain. PLoS Negl Trop Dis. 2021 Dec 28;15(12):e0010076. doi: 10.1371/journal.pntd.0010076. PMID: 34962921; PMCID: PMC8746763.

Author contributions

LG and MP, conceptualization | MTR and FJV, data curation | LG, MTR, ATV, **CN**, JO, JFM, CL and FJV, formal analysis | DS, FJV, PNN and MP, funding acquisition | MTR, AC, MV, MM, OP, SR, KP and AD, investigation | LG, **CN**, FJV and MP, methodology | PNN, project administration | KP, PNN and MP, resources | ATV and JO, software | PNN and MP, supervision | LG and MP, writing (original draft) | MTR and PNN, writing (review & editing).

- Nyongesa S*, Weber PM*, Bernet È, Pulido F, **Nieves C**, Nieckarz M, Delaby M, Viehboeck T, Krause N, Rivera-Millot A, Nakamura A, Vischer NOE, vanNieuwenhze M, Brun YV, Cava F, Bulgheresi S, Veyrier FJ. Evolution of longitudinal division in multicellular bacteria of the *Neisseriaceae* family. Nat Commun. 2022 Aug 22;13(1):4853. doi: 10.1038/s41467-022-32260-w. PMID: 35995772; PMCID: PMC9395523.

* Equal contributors

Author contributions

SN and PMW, most experiments, visualization and formal analysis, writing and revision | EB, FP, MN, MD, **CN**, TV, NK, ARM and AN, some experiments, formal analysis and revision | NOEV, ImageJ analysis tools (ObjectJ, Fil-Tracer) | MV, materials | YB, funding and analysis tools | FC, funding, formal analysis and revision | SB, conceptualization, supervision, funding, resources, writing and revision | FJV, experiments, formal analysis, conceptualization, supervision, funding, resources, writing and revision.

Manuscripts in preparation:

- Giraud-Gatineau A, **Nieves C**, Benaroudj N, Veyrier FJ, Picardeau M. Deciphering the emergence and virulence traits of pathogenic *Leptospira* spp.

- Giraud-Gatineau A*, Ayachit G*, **Nieves C**, Dgabo C, Bourhy K, Pulido F, Benaroudj N, Picardeau M, Veyrier FJ. Pan-species transcriptomic analysis reveals adaptation against oxidative stress for the highly virulent *Leptospira* species.

* Equal contributors

Table of contents

List of figures

List of tables

Abbreviations

ANI	Average nucleotide identity
CAAT	Cross-agglutinin absorption test
CDS	Coding sequence
cgMLST	core-genome MLST
COG	Clusters of orthologous groups
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
DDH	DNA-DNA hybridization
DNA	Deoxyribonucleic acid
ECF	Extracellular function sigma factor
GO	Gene ontology
HGT	Horizontal gene transfer
HK	Histidine kinase
HMM	Hidden Markov Models
IM	Inner membrane
IS	Insertion sequence
KDO	2-keto-deoxymanno-octulosonic acid
KEGG	Kyoto Encyclopedia of Genes and Genomes
LPS	Lipopolysaccharide
MAC	Membrane attack complex
MAT	Microscopic agglutination test
MLST	Multi-locus sequence typing
MLVA	Multiple-locus variable number of tandem repeats analysis
mRNA	messenger RNA
NCBI	National Center for Biotechnology Information
OM	Outer membrane
OMP	Outer membrane protein
ORF	Open reading frame
P	Pathogenic
PCR	Polymerase chain reaction
PF	Periplasmic flagellum
PFAM	Protein families database
PFGE	Pulse-field gel electrophoresis
PG	Peptidoglycan

POCP	Percentage of conserved proteins
RAPD	Random amplified polymorphic DNA
REA	Restriction enzyme analysis
RNA	Ribonucleic acid
RR	Response regulator
rRNA	ribosomal RNA
S	Saprophytic
SB	Spherical body
SNP	Single nucleotide polymorphism
T3SS	Type 3 Secretion Systems
TCS	Two-component system
TLR	Toll-like receptor
TPase	Transposase
tRNA	Transfer RNA
VM	Virulence-modyfing
VNTR	Variable number of tandem repeats
WGS	Whole genome sequence/Whole genome sequencing

Chapter 1. Introduction/Research hypothesis and objectives

1.1 Leptospirosis: general overview

Leptospirosis is a zoonotic disease of worldwide distribution caused by pathogenic bacteria of the genus *Leptospira* (Levett, 2001). These bacteria can infect a broad spectrum of animals, including both domestic and wild mammals (Adler, 2015), but also amphibians (Everard et al., 1988; Gravekamp et al., 1991), fish (Mgode et al., 2014) and reptiles (Paz et al., 2019; Bauso et al., 2020; Rodamilans et al., 2020). The humidity and temperature in tropical and subtropical countries create favorable conditions for the survival of leptospires in the environment and its concomitant transmission among hosts, resulting in a higher disease incidence (Levett, 2001; Bharti et al., 2003).

Human infection occurs when individuals come into direct or indirect contact with urine from infected animals (Levett, 2001). It is commonly associated with occupational (e.g., farmers, veterinarians) or recreational (e.g., aquatic sports) exposures. Additionally, urban environments with a high density of rodents, which are known as one of the main carriers of leptospires, also play a role in the transmission of the disease. About 1.000.000 new cases are reported annually, in addition to 60.000 deaths around the world (Costa et al., 2015). However, the underestimation of its morbidity and mortality contributes to considering leptospirosis as a neglected disease, along with other tropical diseases such as leishmaniasis and severe dengue (Costa et al., 2015). In Canada, human leptospirosis is rare, although zoonotic transmission from dogs and racoons has been reported in Ontario and Quebec (Warshawsky et al., 2000; Prescott, 2008). In fact, circulation of *Leptospira spp*. among these two hosts is presumed to be high, but the unawareness of leptospirosis and the few cases reported in humans lead to its underdiagnosis (Prescott, 2008).

Symptoms of infection are diverse, from subclinical to severe cases, and depend on multiple factors such as the host and its immune system, the initial bacterial load, or the infecting species and serovar. Serovariety is of particular interest, as several serovars have been associated with the infection of specific hosts, resulting in the categorization of the latter into i) maintenance hosts, and ii) accidental hosts (Levett, 2001). In fact, certain hosts may be considered as maintenance hosts for some serovars but accidental for others (Levett, 2001). The first are affected by a chronic infection, with typically mild or no symptomatology. In this context, the infecting serovar, which

is “adapted” to the host, colonizes the proximal renal tubule and it is excreted in the urine, contaminating soil and water. Thus, maintenance hosts constitute the main source of bacterial dissemination in the transmission cycle (**Figure 1.1**). Some of the most well-known serovar-host associations of global importance include Hardjo in cattle, Pomona in sheep, Icterohaemorrhagiae in brown rats, and Bratislava in pigs, although other associations have also been noted (Levett, 2001). On the other hand, accidental hosts such as humans, dogs and hamsters, among others, develop an acute disease that can cause significant renal and hepatic complications, or massive hemoptysis (Adler and de la Peña Moctezuma, 2010), ultimately resulting in death.

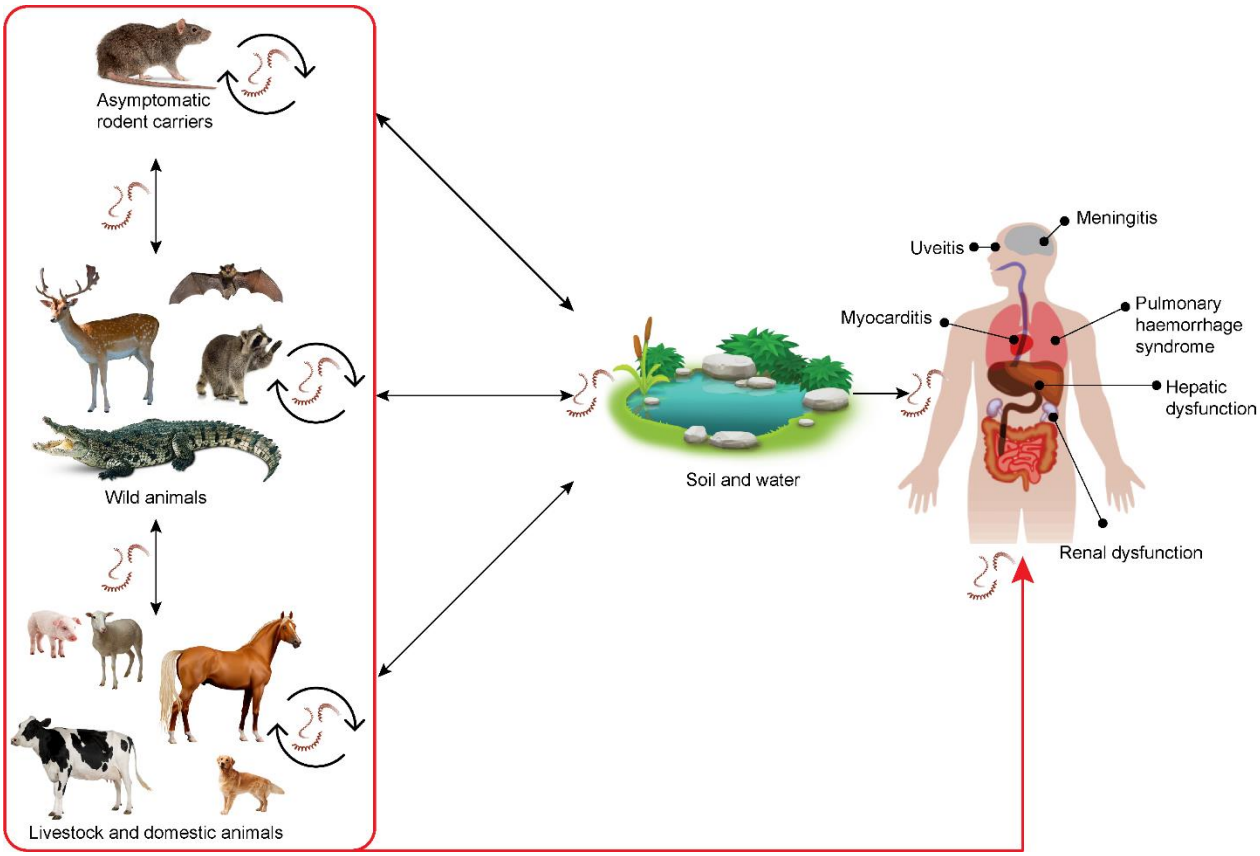

Figure 1.1. Transmission cycle of *Leptospira*.

The transmission cycle of *Leptospira spp*. includes maintenance and accidental hosts. Leptospires penetrate through lesions in the skin, enter the bloodstream and spread to different tissues. In accidental hosts, symptomatology is diverse, from fever and jaundice to renal and hepatic dysfunction, pulmonary haemorrhage, myocarditis, uveitis, and meningoencephalitis. Adapted from (Ko et al., 2009).

1.2 *Leptospira*: the causative agent of leptospirosis

1.2.1 General features

Description of members of the genus *Leptospira* date back over 100 years. They were first observed in 1907 by Arthur Stimson, who named the bacterium as *Spirochaeta interrogans* due to its question-marked shape. Years later, in 1916, the first isolate of *Leptospira sp*. was obtained from blood of a patient diagnosed with Weil's disease.

These spirochetal bacteria, whose classification will be discussed later, are typically 0.1-0.3 μm in diameter, and vary in length from 6-20 μm, with one or both ends hook-shaped (**Figure 1.2A**) (Levett, 2001; Adler and de la Peña Moctezuma, 2010). When isolated, the length and morphology change over time, becoming longer and less motile as the culture progresses. However, in long-standing cultures deprived of nutrients, they tend to become spherical (Adler, 2015).

Their motility is due to the presence of two endoflagella, one on each pole of the cell, whose filaments are in the periplasmic space (**Figure 1.2B**). The rotation of each flagellum provides movement around a central axis and translation, extremely important for pathogenicity (Xu et al., 2020). Particularly, swimming motion allows bacterial migration in fluids, while crawling helps them to navigate over the host cells once attached. Hence, these translating forms require one hooked end and the other spiral serving as a propellant, while non-translating forms may feature either both hooked or spiral ends. The implication of motility in virulence will be discussed in more detail later in this chapter.

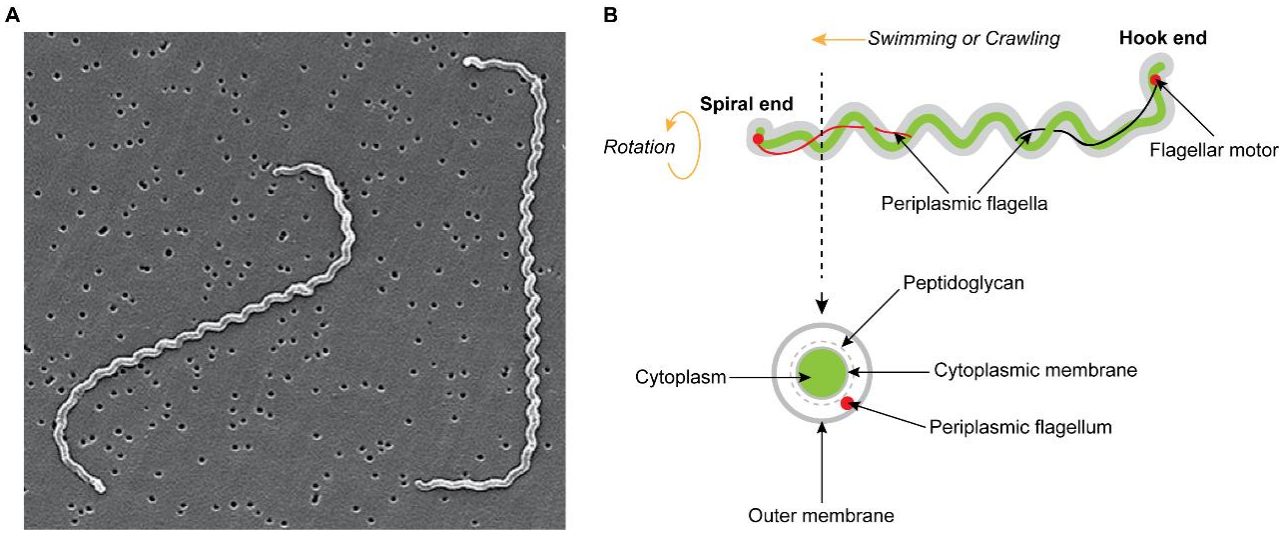

Figure 1.2. Morphological traits underlying *Leptospira* motility.

(**A**) Scanning electron micrograph of *Leptospira interrogans* displaying the typical hooked ends (Adler, 2015). (**B**) Schematic representation of cell structure showing the location of the flagellar filaments responsible for different bacterial movements. Adapted from (Xu et al., 2020).

The leptospiral cell envelope resembles that observed in Gram-negative bacteria, i.e., consisting of a double membrane separated by a peptidoglycan layer (PG) and decorated externally with lipopolysaccharide (LPS) (**Figure 1.3**). The PG, located in the periplasmic space, is largely responsible for the helical morphology (Slamti et al., 2011) along with other cytoskeletal proteins. The outer membrane (OM) is different from other pathogenic spirochetes, possibly because of its transmission cycle that encompasses free and host-associated stages. Lipoproteins, LPS, and various transmembrane proteins are among the major components found in the OM of *Leptospira* (Haake and Zückert, 2015). LPS is certainly the dominant component on the surface, albeit displaying noticeable variations in length between pathogenic and saprophytic species (**Figure 1.3E** and **F**) (Raddi et al., 2012). Indeed, this heterogeneity correlates with the role of LPS in virulence (Werts et al., 2001; Nahori et al., 2005), along with the extremely variable genetic composition of the locus involved in its biosynthesis (Fouts et al., 2016), to be discussed in further sections.

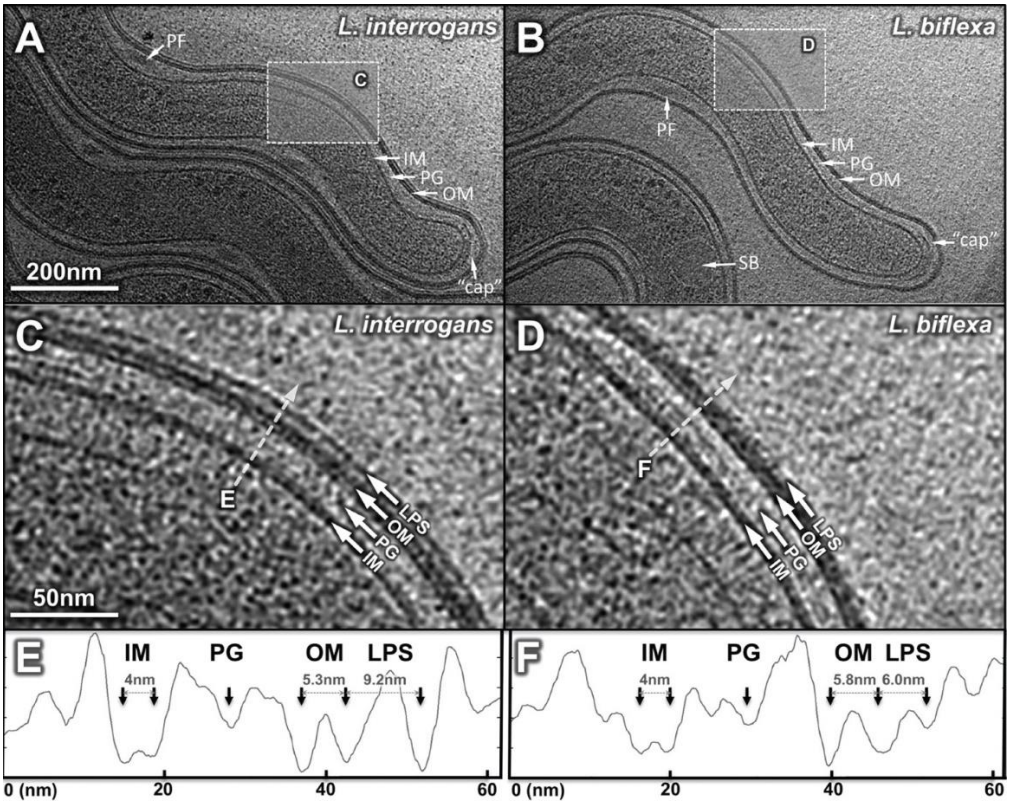

Figure 1.3. Cell envelope of pathogenic and saprophytic *Leptospira*.

Electron cryotomography from intact cells of *L. interrogans* **(A)** and *L. biflexa* **(B)**. Constituent elements are indicated: inner membrane (IM), peptidoglycan layer (PG), outer membrane (OM), periplasmic flagellum (PF), nutrient storage spherical body (SB) and the "cap" at the end of the cell involved in adhesion processes. Magnification of a cell envelope section from *L. interrogans* **(C)** and *L. biflexa* **(D)** shows the lipopolysaccharide (LPS) extending from the external leaflet of OM. The density profiles across the membranes in panels E and F outline the dimensions of each layer in *L. interrogans* and *L. biflexa*, respectively. Taken from (Raddi et al., 2012).

1.2.2 Taxonomy

Bacteria within the genus *Leptospira* belong to the phylum *Spirochaetes*, class *Spirochaetes*, order *Spirochaetales*. The latter consists of three bacterial families, including *Leptospiraceae* which is composed of the genera *Leptonema*, *Turneriella* and *Leptospira*.

Originally, the genus *Leptospira* consisted exclusively of two species: one comprising the pathogenic strains, *L. interrogans sensu lato*, and the other including the saprophytic members, *L. biflexa sensu lato* (Levett, 2001). This classification, based on the mere fact of whether the bacteria cause disease, became progressively complex, resulting in the differentiation into pathogens, intermediates, and saprophytes, with multiple species in each category. Although intermediates

emerged as strains with moderate pathogenicity in both humans and animals, the infection of animal models for acute leptospirosis with these species has been shown to fail in reproducing the disease, making their definition unclear (Thibeaux et al., 2018).

Fortunately, the use of comparative genomics in taxonomic classification has become more relevant, circumventing any need for experimental approaches to determine the virulence of strains, which in *Leptospira* is highly variable from host to host. The latest update of *Leptospira* classification based solely on phylogenetic relatedness resulted in 64 species belonging to two major groups, pathogenic (P) and saprophytic (S), separated into two subgroups each (**Figure 1.4**) (Vincent et al., 2019). Additional species have been identified (Korba et al., 2021), either pathogenic or saprophytic, while some other species were reclassified (Casanovas-Massana et al., 2021). The reconstruction of a phylogeny considering these modifications, as well as a detailed description of the taxonomy and phylogenomics of *Leptospira*, will be presented in Chapter 2.

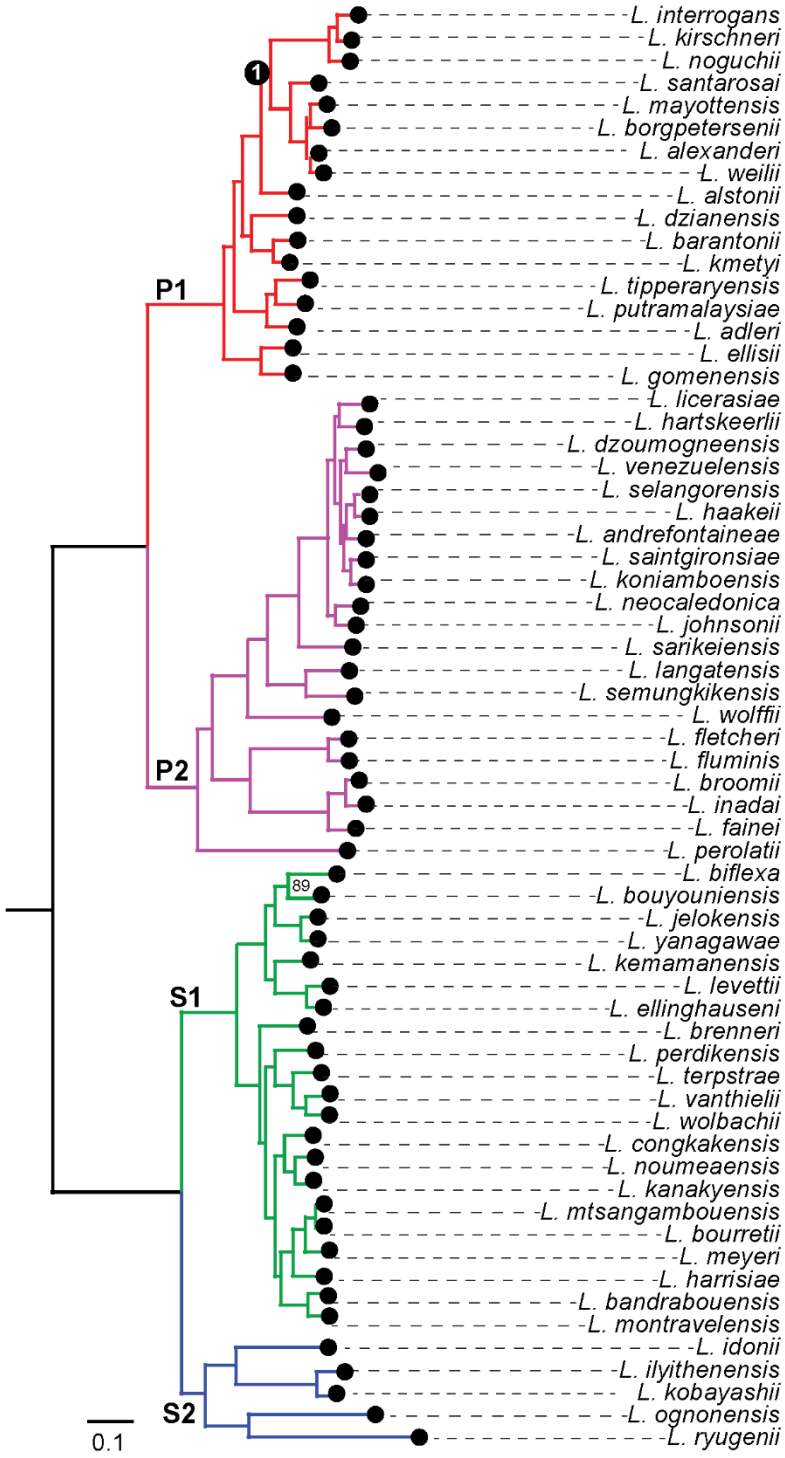

Figure 1.4. Phylogenetic distribution among *Leptospira* species.

Phylogenetic tree based on soft-core genes (present in at least 95% of strains). The main subclades are indicated with different colours: P1 (red), P2 (purple), S1 (green), and S2 (blue). P and S stands for pathogenic and saprophytic species, respectively. Node 1 includes pathogenic species frequently involved in human leptospirosis. Adapted from (Vincent et al., 2019).

Leptospira strains are also further separated into a vast number of serovars, exceeding 300 (Picardeau, 2017). This classification relies on the expression of surface-exposed epitopes of the LPS, as well as their carbohydrate composition and spatial arrangement (Adler and de la Peña Moctezuma, 2010). Moreover, serovars antigenically related are grouped into serogroups, adding some complexity. Thus, a single species may comprise several serogroups, and the same serogroup may also be found in more than one species. These serogroups might include more than one serovar, therefore, different species can belong to identical serovars (**Table 1.1**).

Table 1.1. Serovars found within *Leptospira* serogroups and their distribution among species.

Collected from the Leptospirosis Reference Centre of the Amsterdam University Medical Centre (Department of Medical Microbiology and Infection Prevention) (https://leptospira.amsterdamumc.org/leptospira-library/leptospira-strains/).

Species	Serogroup	Serovar
L. biflexa	**Andaman**	**Andamana**
L. borgpetersenii | *L. interrogans* | *L. kirschneri* | *L. noguchii*	**Australis**	**Australis | Bajan | Barbudensis | Bratislava | Fugis | Hawain | Jalna | Lora | Muenchen | Nicaragua | Peruviana | Pina | Ramisi | Rushan | Soteropolitana**
L. borgpetersenii | *L. interrogans* | *L. kirschneri* | *L. noguchii* | *L. santarosai*	**Autumnalis**	**Alice | Autumnalis | Bangkinang | Bim | Bulgarica | Butembo | Carlos | Erinaceiauriti | Fortbragg | Lambwe | Mooris | Mujunkumi | Nanla | Rachmati | Srebarna | Weerasinghe**
L. borgpetersenii | *L. mayottensis* | *L. santarosai*	**Ballum**	**Arborea | Ballum | Castellonis | Guangdong | Kenya | Peru**
L. interrogans | *L. kirschneri* | *L. noguchii* | *L. santarosai*	**Bataviae**	**Argentiniensis | Balboa | Bataviae | Brasiliensis | Claytoni | Djatzi | Kobbe | Losbanos | Paidjan | Rioja | Santarosa**

L. interrogans | *L. kirschneri*	**Canicola**	**Bafani | Benjamini | Bindjei | Broomi | Canicola | Galtoni | Jonsis | Kamituga | Kuwait | Malaya | Portlandvere | Schueffneri | Sumneri**
L. borgpetersenii | *L. weilii*	**Celledoni**	**Anhoa | Celledoni | Hainan | Mengdeng | Whitcombi**
L. wolbachii	**Codice**	**Codice**
L. kirschneri | *L. santarosai*	**Cynopteri**	**Cynopteri | Tingomaria**
L. interrogans | *L. kirschneri* | *L. noguchii*	**Djasiman**	**Agogo | Buenos Aires | Djasiman | Gurungi | Huallaga | Sentot**
| *L. interrogans* | *L. kirschneri* | *L. santarosai*	**Grippotyphosa**	**Canalzonae | Dadas | Grippotyphosa | Huanuco | Linhai | Muelleri | Ratnapura | Valbuzzi | Vanderhoedeni**
L. alexanderi | *L. borgpetersenii* | *L. idonii* | *L. interrogans* | *L. kirschneri* | *L. santarosai*	**Hebdomadis**	**Borincana | Goiano | Hebdomadis | Jules | Kabura | Kambale | Kremastos | Manzhuang | Maru | Nona | Sanmartini | Worsfoldi**
L. vanthielii	**Holland**	**Holland**
L. broomii | *L. fainei*	**Hurstbridge**	**Hurstbridge**
L. borgpetersenii | *L. interrogans* | *L. kirschneri*	**Icterohaemorrhagiae**	**Birkini | Bogvere | Copenhageni | Dakota | Gem | Hongchon | Hualin | Icterohaemorrhagiae | Lai | Mankarso | Mwogolo | Naam | Ndahambukuje | Ndambari | Smithi | Sokoine | Tonkini | Yeonchon**
L. licerasiae	**Iquitos**	**Varillal**
L. alexanderi | *L. borgpetersenii* | *L. meyeri* | *L. santarosai* | *L. weilii*	**Javanica**	**Arenal | Ceylonica | Coxi | Dehong | Fluminense | Javanica | Mengla | Mengma | Mengrun | Menoni | Poi | Sofia | Sorexjalna | Vargonicas | Yaan | Zhenkang**
L. interrogans | *L. noguchii*	**Louisiana**	**Lanka | Louisiana | Orleans**
L. inadai	**Lyme**	**Lyme**
L. alexanderi | *L. weilii*	**Manhao**	**Lichuan | Lincang | Lushui | Manhao | Qingshui**

L. alexanderi | *L. borgpetersenii* | *L. interrogans* | *L. mayottensis* | *L. santarosai* | *L. weilii*	**Mini**	**Beye | Georgia | Hekou | Mini | Perameles | Ruparupae | Szwajizak | Tabaquite | Yunnan**
L. noguchii	**Panama**	**Cristobali | Magnus | Panama**
L. borgpetersenii | *L. interrogans* | *L. kirschneri* | *L. noguchii* | *L. santarosai*	**Pomona**	**Altoduro | Kunming | Mozdok | Pomona | Proechimys | Tropica | Tsaratsovo**
L. borgpetersenii | *L. interrogans* | *L. noguchii* | *L. santarosai* | *L. weilii*	**Pyrogenes**	**Abramis | Alexi | Biggis | Camlo | Costa Rica | Guaratuba | Hamptoni | Kwale | Manilae | Menglian | Myocastoris | Nigeria | Princestown | Pyrogenes | Robinsoni | Varela | Zanoni**
L. alstonii | *L. interrogans* | *L. weilii*	**Ranarum**	**Evansi | Pinchang | Ranarum**
L. interrogans | *L. santarosai* | *L. weilii*	**Sarmin**	**Cuica | Machiguenga | Rio | Sarmin | Waskurin | Weaveri**
L. interrogans	**Sehgali**	**Portblairi**
L. borgpetersenii | *L. interrogans* | *L. santarosai*	**Sejroe**	**Balcanica | Caribe | Dikkeni | Geyaweera | Gorgas | Guaricura | Haemolytica | Hardjo | Istrica | Jin | Medanensis | Nyanza | Polonica | Recreo | Ricardi | Roumanica | Saxkoebing | Sejroe | Trinidad | Wolffi**
L. biflexa | *L. meyeri* | *L. yanagawae*	**Semaranga**	**Patoc | Saopaulo | Semaranga**
L. noguchii | *L. santarosai*	**Shermani**	**Aguaruna | Babudieri | Carimagua | Luis | Shermani**
L. alexanderi | *L. borgpetersenii* | *L. kmetyi* | *L. santarosai* | *L. weilii*	**Tarassovi**	**Atchafalaya | Atlantae | Bakeri | Banna | Bravo | Chagres | Corredores | Darien | Gatuni | Gengma | Guidae | Kanana | Kaup | Kisuba | Langati | Malaysia | Mengpeng | Mogden | Navet | Rama | Sulzerae | Tarassovi | Topaz | Tunis | Vughia | Yunxian**

Since serovar identity is associated to LPS, mainly the *O*-antigen, and the same serovar can be shared by different species, a horizontal transfer of genes involved in its biosynthesis has been suggested (Haake et al., 2004). However, such hypothesis remains largely unexplored.

Unlike species classification, which can be conducted by comparative genomics, serogroup and serovar determination are performed by time-consuming and costly serological methods (Levett, 2001; Adler, 2015). Avoidance of these techniques prompted the design of PCR protocols for the detection of few serogroups (Cai et al., 2010), however, it has not been widely developed.

1.2.3 Genetic diversity

The diversity of phenotypes observed within the genus *Leptospira* has prompted extensive research into the genetic elements underlying this variability. Numerous studies have focused on comparing saprophytes with pathogens to identify genes associated with pathogenicity. Initial investigations primarily compared well-known pathogens, such as *L. interrogans* and *L. borgpetersenii*, with the saprophyte *L. biflexa* (Ren et al., 2003; Bulach et al., 2006; Picardeau et al., 2008). While many identified genes encoded hypothetical proteins, complicating their characterization, others encoded proteins with established roles in adhesion, such as LigA and LigB (Figueira et al., 2011), or proteins involved in host cell damage, such as sphingomyelinases (Lee et al., 2002). The genome sequencing of additional species has facilitated the extension of these investigations, enabling the functional characterization across the genus. Comparative analysis of the four taxonomic groups, P1, P2, S1, and S2, has identified 16 functional categories that are significantly enriched in at least one subclade (Vincent et al., 2019). This observation suggests specific adaptations within each group, influenced by their distinct lifestyles. A comprehensive examination of comparative genomics in Leptospira will be presented in Chapter 2.

The exploration of genetic diversity extends beyond the comprehension of the emergence of pathogenic species or functional disparities among distinct phenotypes. The examination of genetic variability has also unveiled divergence among closely related serovars, exemplified by Copenhageni and Icterohaemorrhagiae, both belonging to the same serogroup (Icterohaemorrhagiae). Comparative studies have demonstrated that the sole disparity in their genomes is a thymine insertion in strains of the Icterohaemorrhagiae serovar. That insertion occurs within a gene encoding a protein involved in LPS biosynthesis, generating a premature stop codon, and possibly explaining the differentiation between these serovars (Santos et al., 2018). Horizontal gene transfer (HGT) events also serve as a mechanism for generating gene diversity. For instance, in *Leptospira*, HGT has contributed to the emergence of subtypes within specific serovars, such as

Hardjoprajitno and Hardjobovis subtypes of the Hardjo serovar (De la Peña-Moctezuma et al., 1999). It is hypothesized that the progenitor of Hardjoprajitno possessed a *O*-antigen biosynthetic cluster more closely related to that of the Copenhageni serovar and horizontally acquired 13 open reading frames (ORFs) towards the 5' end from the Hardjo serovar. This acquisition resulted in two serologically indistinguishable subtypes. Finally, recombination events play a crucial role in the genetic variability of *Leptospira*. Such is the case of the bacterial immunoglobulin-like (Big) domains of *lig* genes, which participate in the recognition of host extracellular matrix proteins. Recombination events between Big domains of *lig* genes from different species have enhanced their variability, potentially facilitating adaptation to new targets and hosts (McBride et al., 2009).

Thus, *Leptospira* generates genetic variability through several mechanisms, such as single nucleotide polymorphisms (SNPs), HGT, and recombination, among others. This genetic variability not only explains the evolution of pathogenic species but also underscores the remarkable adaptability of *Leptospira* to changing environments.

1.2.4 Pathogenesis and its determinants

Bacteria and host may interact in different manners, and the type of interaction depends on the benefits they obtain from each other. Mutualism represents a win-win situation, whereas commensalism is characterized by bacterial benefits without affecting the host. The most detrimental scenario for the host involves parasitic or pathogenic interactions, where bacteria promote an alteration in the physiological state of the host, leading to disease development. The pathogenesis of infectious diseases depends on various factors, including host susceptibility determined by physiological and immunological conditions, host resistance, which may be compromised by underlying diseases, and intrinsic features of the bacteria (Peterson, 2010). The ability of bacteria to produce disease is dependent on their virulence, which is modulated by multiple factors, such as proteins or complex structures. All these collectively known as virulence factors (Casadevall and Pirofski, 2000).

Multiple virulence factors have been proposed in pathogenic *Leptospira* (Murray, 2015). These include outer membrane proteins (OMPs), which may facilitate interactions with the host; LPS, directly linked to toxicity, as occurs in Gram-negative bacteria; and the flagellar machinery (**Figure 1.5**) (Ko et al., 2009; Adler, 2015; Wunder Jr et al., 2016). However, the cellular or even

molecular basis of *Leptospira* pathogenicity is not understood, as they do not share classical virulence factors with other pathogenic bacteria. The overrepresentation of hypothetical proteins raises speculation about novel and unique pathogenicity mechanisms (Adler, 2015).

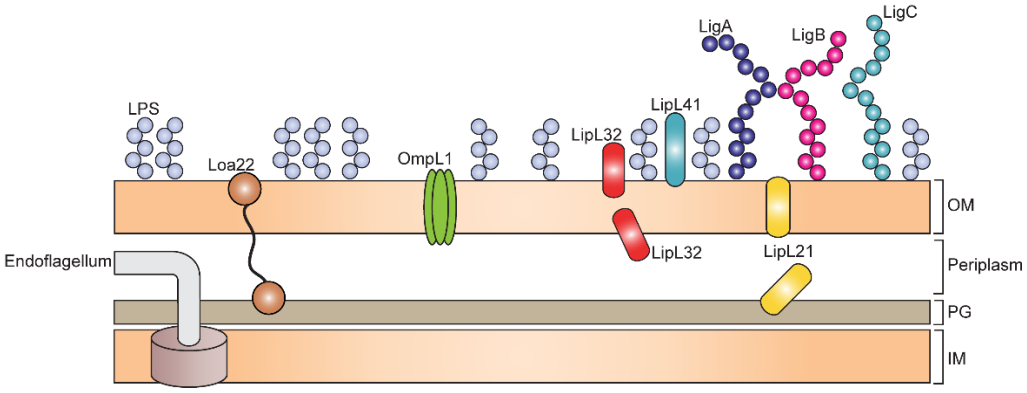

Figure 1.5. Virulence factors in the leptospiral cell envelope.

Several OMPs are highlighted, as well as the LPS and endoflagellum in the periplasmic space. Loa22 interacts with peptidoglycan (PG) through a linker, whereas LipL21 exists either associated with the outer membrane (OM), or bound to PG. Location of LipL32 is still controversial, and while it is stated as surface-exposed, there is also evidence indicating its presence in the internal leaflet of the OM. Adapted from (Ko et al., 2009; Haake and Matsunaga, 2010; Grassmann et al., 2017; Santecchia et al., 2020).

1.2.4.1 Outer membrane proteins

OMPs are involved in multiple functions, such as maintenance of cellular structure, interaction with specific targets in the host, or import and export of nutrients or toxic agents for the bacterial cell, among others. Their location constitutes the first contact with the host, serving as an anchorage, but also interfacing the host immune response (Pinne and Haake, 2009); through these interactions bacteria may develop strategies to gain binding to host cells and evade the immune system, placing OMPs as important contributors to pathogenesis. Identified OMPs, although still functionally not fully characterized, include OmpL1, LipL21, LipL32, LipL41, Loa22 and several from the Lig (leptospiral immunoglobulin-like) family (**Figure 1.5**) (Haake et al., 2004; Haake and Zückert, 2015).

In addition to being conserved in pathogenic *Leptospira*, several of these proteins have demonstrated a direct association with the infective cycle. OmpL1, LipL32 and LipL41 are the primary antigens eliciting a humoral response in human leptospirosis (Guerreiro et al., 2001). Furthermore, immunization of hamsters (used as an acute model) with a combination of OmpL1 and LipL41, provides a significant protection level upon infection, with survival rates above 70%.

Thus, also demonstrating their recognition by the host's immune system (Haake et al., 1999). Not surprisingly, these proteins present a mosaic pattern in their sequences derived from partial horizontal transfer of their coding genes, which is the result of the evolutionary pressures they undergo (Haake et al., 2004). Surface-exposed loops of these proteins display increased amino acid sequence variability, likely representing host-associated adaptations (Haake et al., 2004).

The involvement of LipL32 in the infectious cycle is unclear. Its overexpression during infection has been proven in convalescent patients (Haake et al., 2000; Lessa-Aquino et al., 2013) and animal models, as well as its interaction with TLR2 from the immune system (Yang et al., 2006; Chang et al., 2016). However, transposon disrupted LipL32 mutants have also shown the ability to bind components of the extracellular matrix and successfully infect animal models (Murray et al., 2009). Presence of other lipoproteins and functional redundancy may explain this phenotype. Indeed, peptide chromatography and mass spectrometry revealed that an *L. interrogans* LipL32 mutant undergoes a significant shift in the proteome profile, with a marked increase in the abundance of LipL45, LipL31, LigB and LipL41 (Fernandes et al., 2022).

LipL21, one of the highly expressed leptospiral outer membrane proteins, interacts with several host cell components, including elastin, laminin, collagen I and IV, and E-cadherin (Takahashi et al., 2021), and is implicated in immune evasion mechanisms. Specifically, it can bind to PG and can also prevent the recognition of PG by Nucleotide Oligomerization Domain (NOD)-like receptors. This suggests a mechanism in which LipL21 facilitates escape from innate host responses (Ratet et al., 2017). Moreover, it has also been revealed that LipL21 interferes with host oxidative stress by targeting a neutrophil-derived peroxidase (Vieira et al., 2018), thus playing a central role in the survival strategy of *Leptospira*.

Loa22 also associates with PG through its OmpA domain (Koizumi and Watanabe, 2003). Although a structural role has been proposed by maintaining a physical linkage between the OM and PG, it has been shown that the Loa22-PG association is required to trigger inflammatory responses via Loa22-TLR2 interaction. Two specific amino acid residues, an aspartic acid at position 122 and an arginine at position 143 within the OmpA domain, are responsible for the attachment to PG. Mutation of these residues prevents Loa22-TLR2 interaction and results in reduced production of IL-8 and TNF-α (Hsu et al., 2021). While its role in virulence is well-

established and its disruption impairs infection (Ristow et al., 2007), the presence of a 50% identical homolog in *L. biflexa* introduces controversy when defining it as a virulence factor.

Finally, the Lig family has been widely associated with a virulent phenotype. Members of this family promote adhesion, complement evasion and hemostasis modulation by interaction with host coagulation factors (Haake and Matsunaga, 2021). Their distribution is not equal in all pathogenic strains, especially LigA and LigC, while LigB is ubiquitous among all pathogenic *Leptospira*. Overlapping functions of LigA and LigB have been proposed, as simultaneous knock-down of their coding genes in *L. interrogans* results in a complete loss of virulence in hamsters. However, if the expression of at least one of them is maintained at certain levels, virulence is retained (Pappas and Picardeau, 2015). Recombination events between these three Lig proteins have been linked to evasion mechanisms and extended targeting of novel substrates (Mcbride et al., 2009).

1.2.4.2 Lipopolysaccharide

As previously mentioned, *Leptospira* possesses a cell wall resembling that of Gram-negative bacteria due to the presence of lipopolysaccharide in its outer membrane. While *Leptonema* and *Turneriella*, which belong to the *Leptospiraceae* family as *Leptospira*, have not been extensively studied for their LPS composition, a manual examination of their genomes (performed personally) reveals genes involved in *O*-antigen biosynthesis, providing further support for the presence of LPS in these bacteria. In contrast, two other clinically relevant spirochetes, namely *Borrelia burgdorferi* (Lyme disease) and *Treponema pallidum* (Syphilis), lack LPS in their outer membrane (Zückert, 2019). Genomic analyses of these organisms do not reveal any genes associated with LPS biosynthesis (Fraser et al., 1997, 1998), further indicating their inability to produce LPS. Instead, *B. burgdorferi* and *T. pallidum* predominantly rely on OMPs as the major components of their OM (**Figure 1.6**). *Borrelia* exhibits a greater diversity of lipoproteins compared to *Treponema*, which act as the immunodominant antigens and play a significant role in the infectious process (Kelesidis, 2014). *T. pallidum*, on the other hand, possesses a low density of exposed proteins in its OM that would enable the evasion of the host immune system, rendering it a “silent” pathogen. Notably, proinflammatory lipoproteins of *T. pallidum* are situated beneath the cell surface. It has been suggested that upon uptake and degradation of treponemes, these

lipoproteins are released and subsequently interact with and activate the host immune system (Kelesidis, 2014).

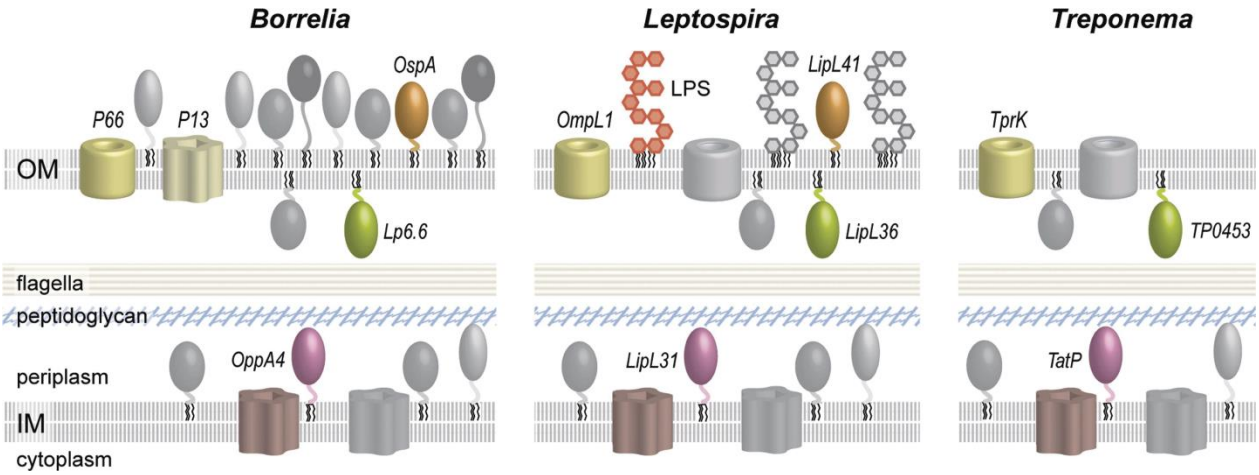

Figure 1.6. Cell envelope of *Borrelia, Leptospira* and *Treponema*.

Among the similarities shared by the three spirochetes, each of them possesses both an inner and outer membrane, which are separated by the periplasmic space. This space contains the peptidoglycan and the flagella in all three spirochetes. Notable differences can be observed as well. *Borrelia* presents a greater number of OMPs, with some lipoproteins surface-exposed (such as OspA), while others are oriented towards the inner side of the OM (Lp6.6). *Leptospira* also displays a significant number of OMPs, including both surface-exposed and non-exposed lipoproteins. Moreover, unlike *Borrelia* and *Treponema*, *Leptospira*'s OM primarily consists of LPS as the dominant component on the surface leaflet. Finally, *Treponema* expresses only a limited number of OMPs, and most of them are not exposed on the surface. Taken from (Zückert, 2019).

LPS comprises three components: lipid A, which interacts with the OM and is composed by carbohydrates and conserved fatty acids; *O*-antigen, the outermost part of LPS; and the core oligosaccharide, that links *O*-antigen and lipid A (**Figure 1.7**) (Wang and Quinn, 2010). This is the same organization observed in *Leptospira* (Patra et al., 2015).

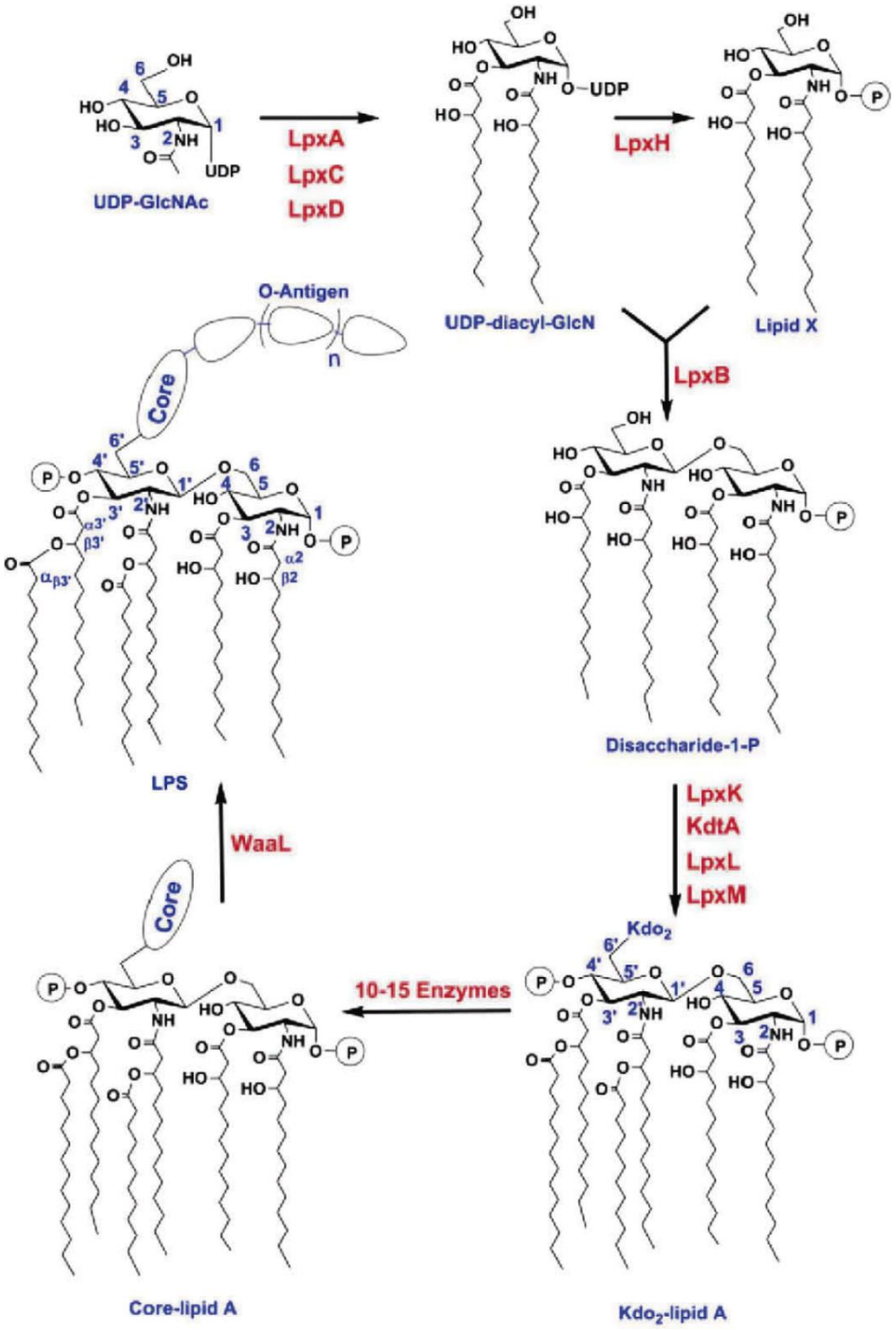

Figure 1.7. Structure and biosynthesis of LPS.

The structural organization of LPS and biosynthetic pathway in *E. coli* is depicted as an example. Enzymes involved in each step are highlighted in red, while substrates are indicated in blue. The detailed structure is provided only for lipid A, while the core-oligosaccharide and *O*-antigen are simplified due to the numerous variations encountered in these components. The core region typically consists of 10-15 monosaccharides, while the *O*-antigen usually contains a few of them that are repeated multiple times. Taken from (Wang and Quinn, 2010).

The complex structure of LPS is critical for the viability of virtually all Gram-negative bacteria, providing selective permeability to the OM and offering protection. Furthermore, the challenging environments faced by the bacteria, such as in the context of infection, may alter the LPS structure thus facilitating bacterial adaptation (Sperandeo et al., 2009). These modifications are not limited to the exposed component of LPS but also include the innermost part. For instance, lipid A can undergo the removal of phosphate groups (by LpxE, LpxF), addition of long fatty acid chains (by LpxXL), or methylation of specific phosphate groups (by LmtA) (Wang and Quinn, 2010). Some of these modifications are regulated by two-component systems that sense specific environmental changes, such as a decrease in Mg^{2+} (PhoQ/PhoP) or an increase in Fe^{3+} levels (PrmB/PrmA) (Wang and Quinn, 2010).

LPS is the primary component on the OM in *Leptospira*, and the main target of humoral immunity against infection (Bulach et al., 2000), which has made it a strong candidate for vaccine development (Matsuo, 2000). In addition, its role in bacterial colonization and dissemination, both essential for pathogenicity, has been demonstrated (Nally et al., 2005). Leptospiral LPS exhibits much lower endotoxicity than other Gram-negative bacteria, probably as a result of specific structural traits (Que-Gewirth et al., 2004). Analysis of *L. interrogans* serovar Lai revealed the presence of all *lpx* orthologous genes related to lipid A biosynthesis, but further structure elucidation showed discrepancies from typical lipid A (**Figure 1.8**). Among the distinctive hallmarks, *Leptospira* lipid A presents 2,3-diamino-2,3-dideoxy-glucopyranose (GlcN3N) instead of glucosamine, which is well supported by the presence of genes related to the synthesis of Glc3N3 precursors. This disaccharide backbone possesses amide-linked acyl chains, unsaturated fatty acids, and a methylated phosphate group at position 1, something not found in any other lipid A. Position 4', on the other hand, which is usually substituted with a phosphate, has no modification of any kind. Absence of this phosphate group is likely preventing lipid A recognition by TLR4 in human innate immune system (Scior et al., 2013), which may represent an escape mechanism.

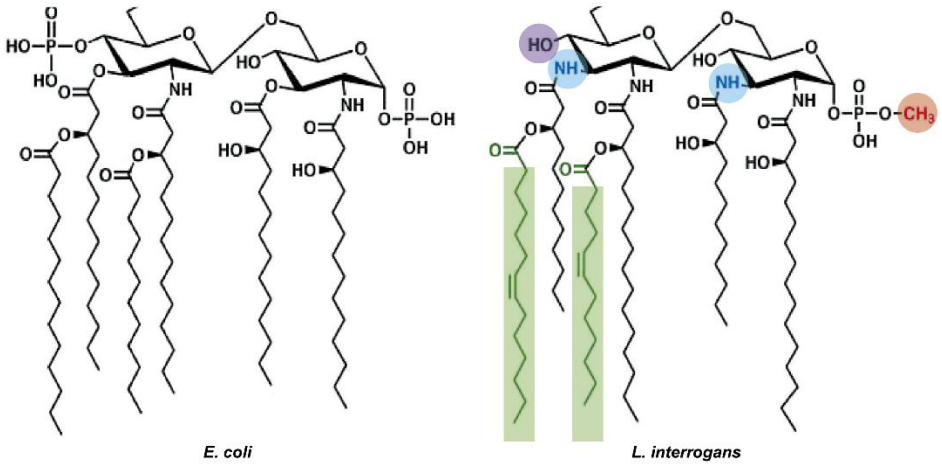

Figure 1.8. Differences between the lipid A structure from *E. coli* and *Leptospira.*

Differences are indicated on *Leptospira* lipid A using colors. These include the presence of a methyl group in the phosphate at position 1 (highlighted in red); amino groups at positions 3 and 3' (highlighted in blue); absence of a phosphate group at position 4' (highlighted in purple); and presence of unsaturated fatty acids (highlighted in green). Adapted from (Werts, 2010)

The core portion of the LPS includes 2-keto-deoxymanno-octulosonic acid (KDO). Studies have shown that the amount of KDO varies between pathogenic and intermediate strains. For example, *L. licerasiae* strain VAR010 has approximately three times less KDO compared to *L. interrogans* strain Fiocruz L1-130 (Patra et al., 2015). While it is suggested that other sugars or alternative forms of KDO may be responsible for linking the *O*-antigen and lipid A in intermediate strains (Patra et al., 2015), the core-oligosaccharide has not been extensively studied throughout the *Leptospira* genus.

Moving to the outermost portion of LPS, the *O*-antigen constitutes the most relevant part in leptospiral biology, since serovar classification is supported by the high variability of its carbohydrate portion (Picardeau, 2017). This structural variability results from the gene diversity in the *O*-antigen biosynthetic gene cassette. The genes implicated in the *O*-antigen biosynthesis reside within the *rfb* cluster and are mainly divided in three groups: i) protein-coding genes with involvement in *O*-antigen nucleotide sugar precursors biosynthesis; ii) genes encoding glycosyltransferases that transfer the precursors to a carrier lipid located in the inner membrane; and iii) genes coding for *O*-antigen processing proteins involved in translocation and polymerization in the periplasmic space (**Figure 1.9**). Once synthesized, the *O*-antigen is transported and assembled at the OM surface (Samuel and Reeves, 2003).

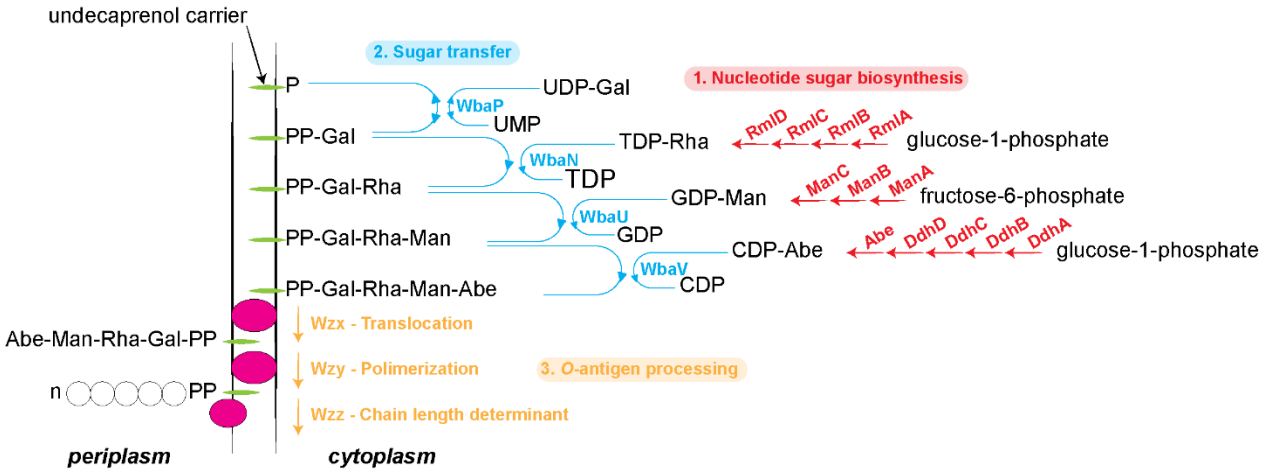

Figure 1.9. *O*-antigen biosynthesis.

Proteins involved in i) nucleotide sugar biosynthesis (highlighted in red); ii) sugar transfer (sky blue); and iii) *O*-antigen processing (orange) are found (based on the best-studied *O*-antigen gene clusters from *E. coli* and *S. enterica*). The first two steps occur entirely in the cytoplasm, while the *O*-antigen processing takes place at the periplasmic side of the inner membrane. In *Leptospira interrogans*, the *O*-antigen is assembled via *wzy*-dependent pathways (Haake and Zückert, 2015). Adapted from (Samuel and Reeves, 2003).

Bioinformatics analysis on the leptospiral *rfb* cluster has shown variations not only from one serovar to another, but also differences in cluster length between P1, P2 and S1 species (**Figure 1.10**) (Fouts et al., 2016). A more comprehensive analysis, exploring the link between *rfb* gene content and serovar designation, has been conducted within the scope of this thesis, as shown in Chapters 4 and 5.

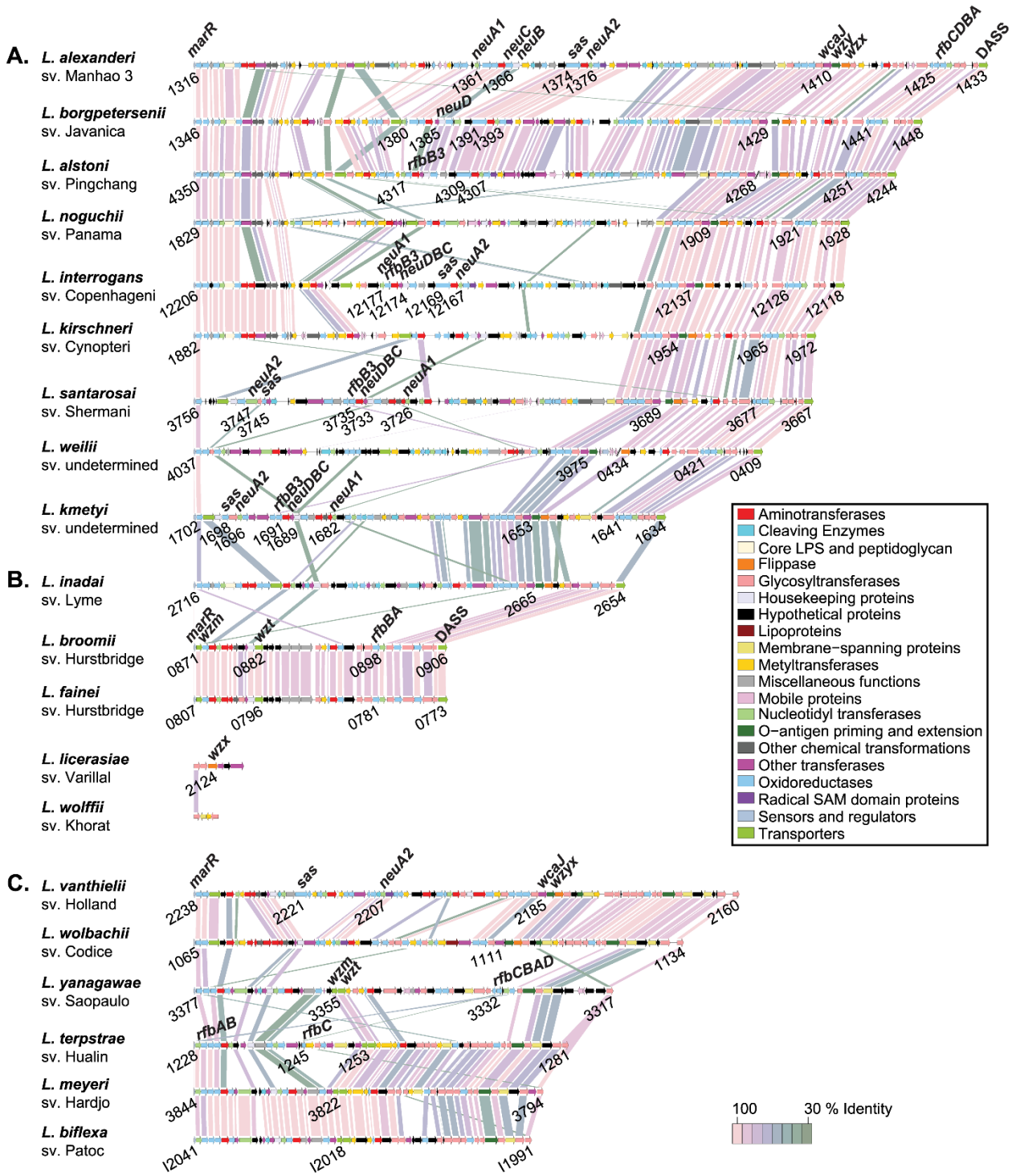

Figure 1.10. *Leptospira rfb* gene clusters.

O-antigen biosynthetic gene clusters of P1 (**A**), P2 (**B**), and S1 (**C**) species. Starting (*marR*) and ending (*DASS*) genes are well conserved, especially in P1 species. CDS are coloured according to their functionality as indicated in the insert. Identity percentages of matching CDS by Blastp are represented with different colours as outlined on the bottom right scale. Taken from (Fouts et al., 2016).

In addition to serovar designation, it is thought that *O*-antigen variation could be advantageous as a mechanism of immune evasion. Introduction of LPS antiserum into a leptospiral culture leads to LPS-altered mutants (Haake and Zückert, 2015). Expression levels of *O*-antigen have also been associated with acute or chronic disease, with the latter typically asymptomatic. When guinea pigs (acute model) are infected, the amount of *O*-antigen expressed by *Leptospira* is strikingly lower compared to infected rats (chronic model), suggesting that the regulation of *O*-antigen levels may be modulating the disease outcome (Nally et al., 2005). This differential regulation could be explained through a distinctive immune system activation. Although the immune response against LPS is typically TLR4-dependent, it has been shown that leptospiral LPS is recognized by TLR2, either in human- or mice-derived monocytes/macrophages (Werts et al., 2001). Nevertheless, stimulation of these cells with lipid A of *Leptospira* triggers a TLR4-mediated response in mice but not in human (Nahori et al., 2005), raising the question of whether the *O*-antigen is responsible for this atypical TLR2 activation. Indeed, stimulation of mice bone marrow-derived macrophages with a *L. interrogans* mutant producing a shorter LPS induces a TLR4 response via TRIF (TBK1-IKKε-IRF3) activation, whereas a wild-type LPS escapes this pathway, which is important for antimicrobial responses (Bonhomme et al., 2020). Similarly, *L. biflexa,* which has a shorter LPS compared to *L. interrogans*¸ induces TLR4-TRIF responses, suggesting that the *O*-antigen may alter CD14 binding (Bonhomme et al., 2020), which is required for TLR4-TRIF activation. Thus, the regulation of *O*-antigen expression levels, impacting on LPS length, could potentially contribute to the development of chronic infections.

1.2.4.3 Flagellar machinery

Although the flagellum structure is generally well conserved across bacteria, there are also specific features that distinguish the wide spectrum of bacterial diversity. In this regard, the spirochetal endoflagella (**Figure 1.11**) display distinctive hallmarks (San Martin et al., 2022): i) the basal body is composed of inner membrane- and peptidoglycan-associated proteins, while the outer membrane remains a bystander; ii) the periplasmic hook is highly crosslinked via a lysinoalanine adduct formation, which is not required for flagellar assembly but for motility (Miller et al., 2016); and iii) the filament is not exclusively composed of a single flagellin protein, but rather by a network of proteins forming an intricate structure. Despite these shared features, there

are some genus-specific traits. For instance, the basal body in the cytoplasm of *Leptospira* is larger in comparison to other spirochetes, such as *B. burgdorferi*, highlighted by the presence of additional proteins (San Martin et al., 2022). The number of certain proteins also varies. FliG, involved in torsional strength, has two paralogs in all the spirochetes, but three in *Leptospira*. More importantly, several of these *Leptospira*-specific proteins have shown relevance in virulence.

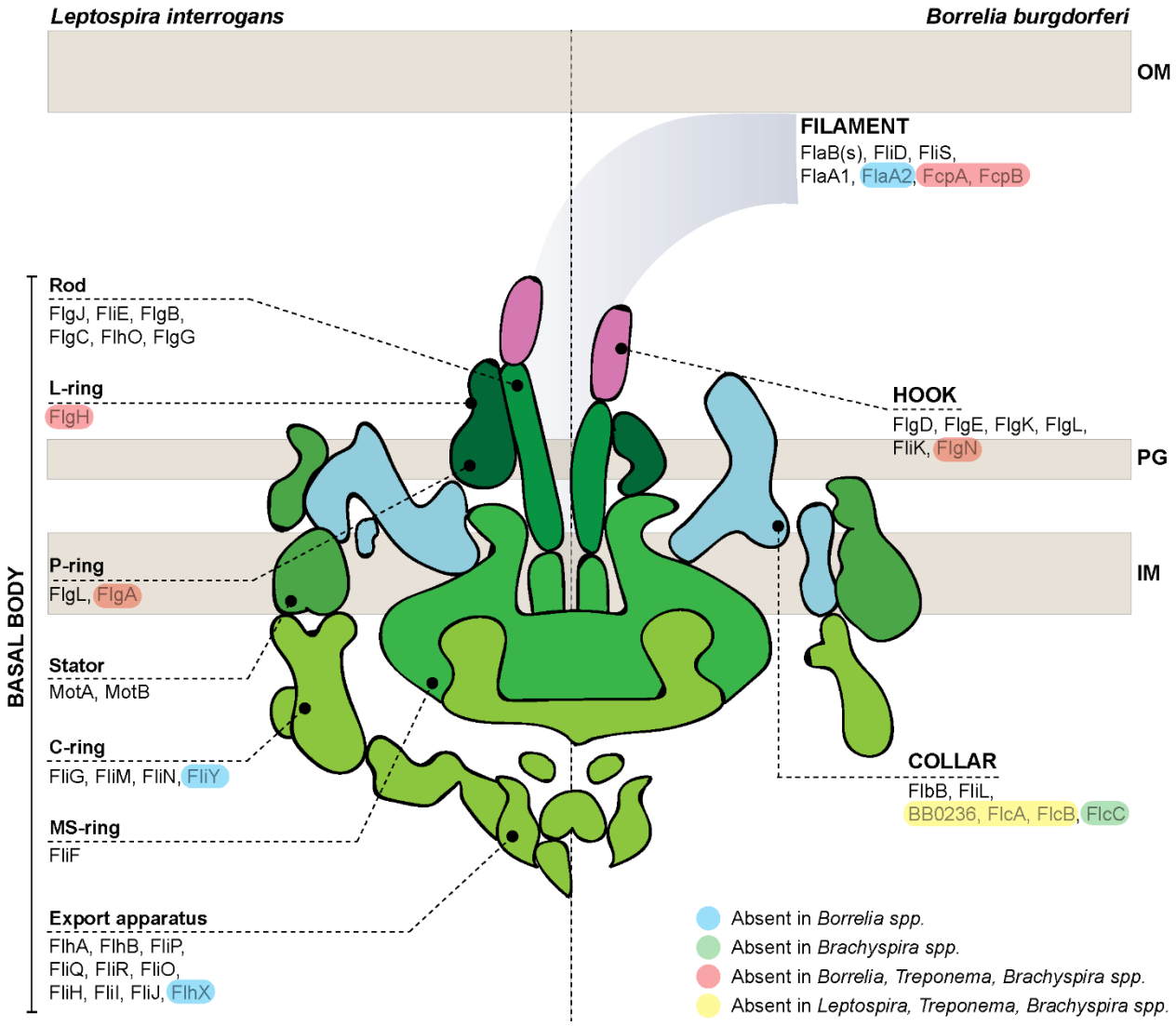

Figure 1.11. Representation of endoflagellar machinery in *Leptospira* and *Borrelia spp*.

Differential proteins are highlighted in colours: absent in *Borrelia spp*. (sky blue), or *Brachyspira spp*. (green), only present in *Leptospira spp*. (red), or exclusively found in *Borrelia spp*. (yellow). Protein redundancy is not represented. Adapted from (San Martin et al., 2022).

The *fla* genes involved in the filament synthesis are key components in *Leptospira* motility. FlaA1 and FlaA2 both present in *L. interrogans,* but the second one absent in *B. burgdorferi,* have distinctive roles. Generation of a *flaA2* mutant in *L. interrogans* sv Lai results in a markedly reduced motility, with absence of hooked or helical ends, and concomitantly inability to perform translational movement. Moreover, this mutant is completely attenuated in hamsters. Conversely, *flaA1* mutant does not affect the hooked-spiral end morphology, retains an intermediate motility with occasional translational movement, and has the ability to cause death within 5-7 days post-infection in the same animal model. So, besides a requirement for motility, FlaA2 also clearly affects virulence (Lambert et al., 2012).

The Flagellar-coiling protein A, FcpA, is an utmost component of the flagellar sheath highly conserved and restricted to *Leptospira*. Studies have shown that *L. interrogans fcpA*- strains are unable to perform translational motility and infect hamsters, which is reversed upon *fcpA* complementation. Furthermore, the absence of FcpA also reduces the expression of other filament-associated proteins (FlaA1, FlaA2), suggesting a modulator role in flagellum protein composition and an intricate interaction among proteins (Wunder Jr et al., 2016).

However, proteins other than filament constituent proteins are also important for motility and virulence. A non-motile *L. interrogans* strain carrying a stop codon in *fliM* showed alteration in the hook-shaped ends and an avirulent phenotype in hamsters. The wild-type phenotype was completely recovered after *fliM* complementation. Most importantly, the presence of a full-length FliM showed essentiality for flagella formation, since the mutant displayed a thinner and shorter flagellum, which connects with decreased amount of FlaA2 and FlaB. No differences in the transcription of these genes were found, therefore the effect of FliM impacts directly on the assembly process (Fontana et al., 2016).

FliY, also from the C-ring as FliM, does not affect the flagellar assembly, but inactivation of *fliY* in *L. interrogans* resulted in impaired adherence to macrophages and reduced ability to cause apoptosis. Motility was also affected, but not the cell morphology. Furthermore, inactivation of this gene affected the transcription of *fliP* and *fliQ* from the export apparatus. Given the link of the latter two to Type 3 Secretion Systems (T3SS), it is speculated that the export of adhesion- or cytotoxicity-associated proteins may be impacted (Liao et al., 2009).

It becomes evident that motility relying on periplasmic flagella is an essential virulence factor, but the mechanism behind this is not completely understood. Rotation of the periplasmic flagella (PF) under the outer membrane would be responsible for the rolling of the cell body, and concomitantly the propulsion of spirochetes (Nakamura, 2022). In *Leptospira spp.*, the movement is governed by the morphology of both ends, which is ultimately determined by the coiled shape of the PF located at each extremity (Goldstein and Charon, 1990). Swimming, which allows cell propulsion and needs one spiral- and the other hooked-end, can switch to a reversal twist to change direction. *Leptospira* swimming has been described as slippery motion, with enhanced efficiency in high-viscosity environment (Takabe, 2013), as well as an increase in back-and-forth movement that reduces net migration and favours colonization (Nakamura, 2022). Crawling, on the other hand, depends on the motility of the PF but is carried out by rotation of the cell body without the contribution of the spiral ends, and with no translational movement. It is mediated by adhesive components in the outer membrane being dragged along as the cell body rotates (Tahara et al., 2018). Indeed, crawling is enhanced by anti-LPS antibodies, implying that inhibition of attachment promotes this movement. Moreover, adherent and crawling populations of pathogenic *Leptospira* on renal proximal tubule epithelial cells increased after 24 h of infection, while non-pathogenic species did not change over time (Sebastián et al., 2021). Thus, adhesion and crawling are likely facilitating the interaction with tissues surfaces, until the bacteria reach the intracellular junctions and disseminate across host cell barriers by disrupting their integrity. Since LPS is one of the main constituents of the OM and plays a key role in serovar determination, its heterogeneity may impact crawling motility. However, the wide diversity of external adhesins and their potential association with crawling should be considered, as this may uncover mechanisms of host-specificity (Xu et al., 2020).

1.3 Hypothesis and objectives

Leptospira is a heterogeneous bacterial genus. Comparative genomics has been insightful by uncovering specific genes contributing to pathogenesis and has even provided clues to specific host interactions. However, with the recent explosion in the number of new species, it is necessary to update the characterization of the *Leptospira* pangenome.

1.3.1 Hypothesis

There is a large unexplored genetic variability that underlines the evolution of pathogenic *Leptospira,* and that would explain the adaptive process from free-living to host-associated bacteria. In addition, variability in specific genes would account for the vast repertoire of serovars, as well as their link with host-preference.

1.3.2 Objectives

1.3.2.1 General objective

To further explore the genetic repertoire of *Leptospira* in order to enhance our understanding of host interactions and the evolution towards obligate pathogens.

1.3.2.2 Specific objectives

- To provide an updated overview of all reported *Leptospira* species, with emphasis on taxonomy and general genomic features (Chapter 2 – Article 1/Book Chapter).
- To better characterize changes in the gene repertoire associated with the emergence of pathogenic *Leptospira* (Chapter 3 – unpublished results)
- To comprehend the genetic basis of serovar identity and host-preference (Chapter 4 – Article 2)
- To gain insights within intraspecies evolution in highly virulent *Leptospira* (Chapter 5 – Article 3)

Chapter 2. Book chapter on *Leptospira* genomics (Article 1)

Taxonomy and phylogenomics of *Leptospira*

Cecilia Nieves[1], Samuel Garcia Huete[2], Frédéric Veyrier[1*], Mathieu Picardeau[2*]

[1]Bacterial Symbionts Evolution, Centre Armand-Frappier Santé Biotechnologie, Institut National de la Recherche Scientifique, Université du Québec, Laval, QC, Canada. [2]Institut Pasteur, Université de Paris, Unité Biologie des Spirochètes, F-75015, Paris, France.

*Correspondence mathieu.picardeau@pasteur.fr; frederic.veyrier@inrs.ca

Title of journal

Phylogenomics: Foundations, Methods, and Pathogen Analysis. Elsevier (Academic Press).

Accepted. Available May 1, 2024.

Author contributions

C. Nieves wrote the original draft and participated in subsequent rounds of editing. This author also conducted most of the bioinformatics analyses presented herein, including the phylogeny of 68 species of *Leptospira* and the associated matrix with ANI values; the comparison of *rfb* clusters among different species of the same serovar revealing a consistent *rfb* genetic fingerprint; and the distribution of the *Leptospira* pangenome across different taxonomic groups, represented as a Venn diagram.

SGH, writing, bioinformatic analysis; FV and MP, critical reviewing, editing.

2.1 Abstract

Accurate identification of bacterial species is critical for epidemiologic, phylogenetic, and diagnostic purposes. The classification of leptospires has undergone several changes following the discovery of the leptospirosis agent more than a century ago. With the genomic era, the number of described *Leptospira* species has more than doubled in recent years. Genotyping, including whole-genome based typing, allows discrimination of isolates at the sub-species level. Access to genomes has also facilitated the study of gene repertoire and virulence-associated factors. Herein we review the taxonomic classification of the genus *Leptospira* and highlight recent findings based on the analysis of genomes.

2.2 Taxonomy/Phylogeny

2.2.1 The phylum of *Spirochaetes*

The phylum of *Spirochaetes* is comprised of diderm bacteria that were first recognized for having in common a unique morphology. Most of them are helical and possess internal flagella responsible for motility. The phylum of *Spirochaetes* is composed of a single class, the *Spirochaetia*, and four orders: *Brachyspirales*, *Brevinematales*, *Leptospirales* and *Spirochaetales* (Gupta et al., 2013). The *Spirochaetales* is the largest, and includes several families and genera such as *Borrelia*, *Treponema, Sphaerochaeta, Alkalispirochaeta*, and *Spirochaeta*, among others. The order *Brachyspirales* comprises a single family (*Brachyspiraceae*) that contains spirochetes of the genus *Brachyspira*. The order of *Brevinematales* consists of two families: *Brevinemataceae* and *Longinemaceae*, with a single species described in both families to date. Finally, there is only one family within the order *Leptospirales*, the family *Leptospiraceae*, which is constituted by the genera *Turneriella, Leptonema,* and *Leptospira*. The first two genera are single-species and include free-living bacteria. *Leptospira,* on the other hand, includes both free-living saprophytes as well as pathogenic species. *Spirochaetes* include bacteria that are ubiquitous and can adapt and colonize a wide range of environments. *Borrelia* spp., including *B. burgdorferi* and *B. hermsii* (causative agents of Lyme disease and Relapsing fever, respectively), are arthropod-borne pathogens of man, other mammals, and birds. *Brachyspira* spp., mainly *B. pilosicoli* and *B. aalborgi*, are the aethiological agents of Human intestinal spirochaetosis, although they can also infect pigs and

poultry (Pandey et al., 2020). The genus *Treponema*, which includes the pathogen *T. pallidum* (agent of Syphilis), is also composed of anaerobic spirochetes found in the hindguts of termites and wood-eating cockroaches. The phylogenetic structure of representative species within the *Spirochaetes* phylum is shown in **Figure 2.1**.

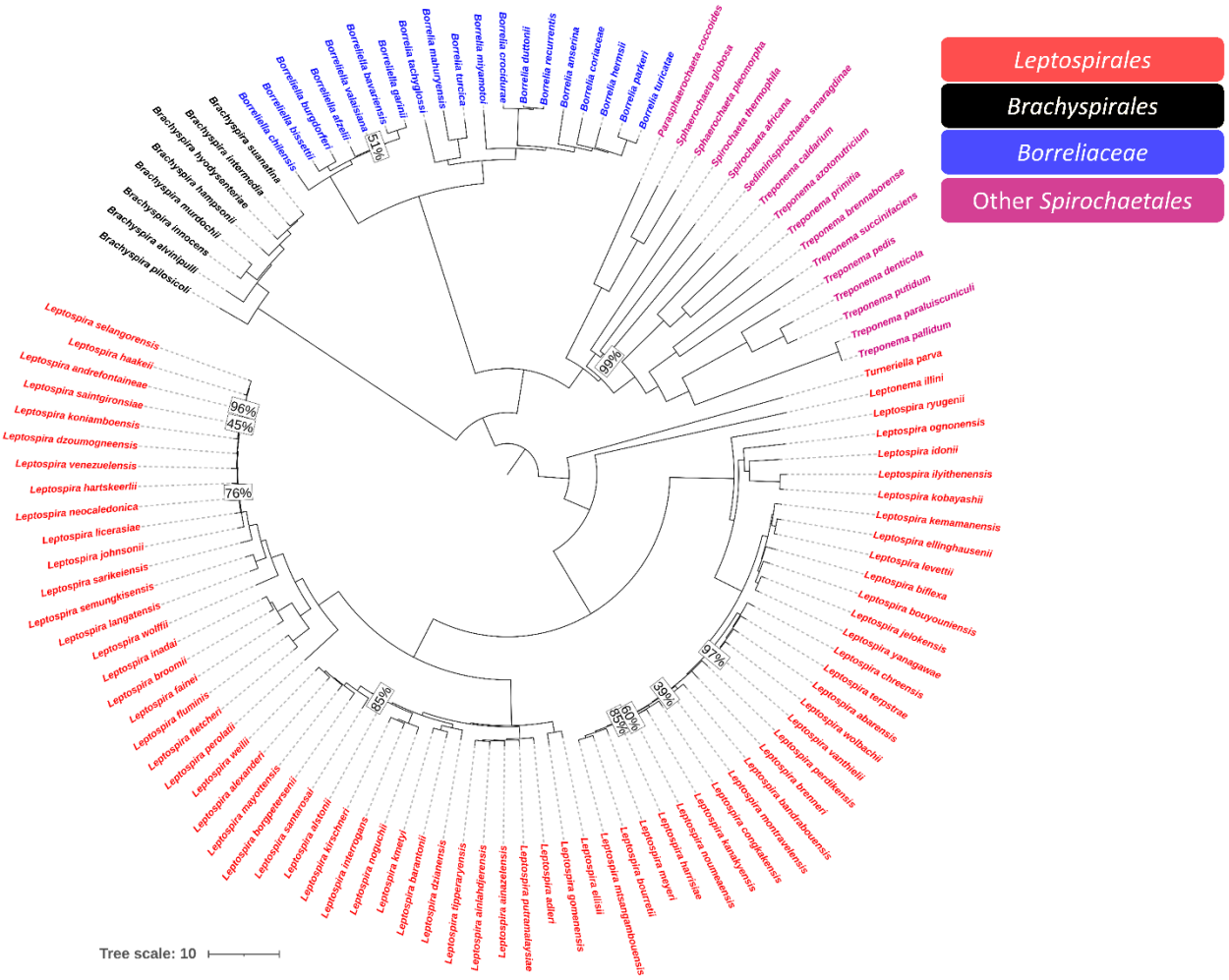

Figure 2.1. Phylogenetic relatedness of the most representative species of the *Spirochaetes* phylum.

Phylogenetic tree based on the weighted least-squares distance of 113 genomes of representative species of the *Spirochaetes* phylum. Species names are coloured according to the different families (see legend) and branch support values are only indicated when lower than 100%. The tree was visualized using iTOL version 6.5.8 (Letunic and Bork, 2019).

2.2.2 Pre-genomics classification of the genus *Leptospira*

Early bacterial classification was often based on morphological and physiological observations. More than a century ago, the genus *Leptospira* was classified (Wolbach and Binger, 1914; Noguchi, 1918) as helical, thin, and motile bacteria that existed as either pathogens or saprophytes. The use of electron microscopy in the 1960s revealed that leptospires have a cell diameter of 0.15 to 0.3 μm and a length ranging from 10 to 20 μm (Hovind-Hougen, 1979) with two endoflagella (or periplasmic flagella) that reside within the periplasmic space and are attached at the subterminal part at each end of the cell (Paster and Dewhirst, 2000). Initially, the genus was divided into *L. interrogans sensu lato* for the pathogenic strains and *L. biflexa sensu lato* for the saprophytic strains. Differentiation of pathogenic from saprophytic species is difficult because these bacteria are morphologically similar and there is substantial variation in biochemical characteristics within the two groups of species, such as growth temperature (only saprophytic *Leptospira* can grow at 13°C) and growth in presence of the purine analogue 8-azaguanine (only saprophytic *Leptospira* are resistant to 8-azaguanine) (Cerqueira and Picardeau, 2009).

Over the years, *Leptospira* classification has gained in complexity, whereby it is not only identified at the species level, but also separated into serogroups, and further subdivided into serovars, based on lipopolysaccharide (LPS) antigenic relatedness. There are currently 26 different serogroups (Ahmed et al., 2015) and more than 300 serovars that can be found across multiple species (Adler and de la Peña Moctezuma, 2010). Up to now, serogroup and serovar identification (discussed below) is still entirely performed by expensive and technically challenging serological methods, which hampers their widespread use. However, although some attempts to switch from serological to molecular techniques have been done, serological and genotypic classification have little correlation (Cerqueira and Picardeau, 2009); for example, the same serovars may be found in different *Leptospira* species.

In the 1990s, the application of DNA-DNA hybridization (DDH) allowed the speciation of members of *Leptospira* based on the accepted definition of >70% hybridization under defined high stringency conditions (Rosselló-Móra et al., 2011). DDH measures the degree of identity through annealing stability of double-stranded DNA resulting from mixing denatured DNAs, which are later incubated in conditions where renaturation is only possible for complementary sequences. DDH has historically been used as the gold-standard for circumscribing prokaryotic species. Using

this approach, the original pathogenic species *Leptospira interrogans sensu lato* split into six different species (Cerqueira and Picardeau, 2009). Thereafter, 20 species within the genus were confirmed by this method (Cerqueira and Picardeau, 2009).

Phylogenetic analysis based on ribosomal RNA (rRNA) genes has also been used in taxonomic classification for decades due to their ubiquitous distribution across bacterial lineages (Kim and Chun, 2014). Moreover, the existence of a vast repository of 16S rRNA sequences covering all known species, simplifies *in silico* comparisons (Kim and Chun, 2014). Stackebrandt and Goebel have initially proposed a <97% 16S rRNA gene sequence identity threshold for *species* delineation (Stackebrandt and Goebel, 1994), but depending on the groups of bacteria, different values are in practical use. *Leptospira* species, which possess between one and two copies of the 16S rRNA, have high sequence identity, between 86 and 100% (Vincent et al., 2019). For example, a single base difference out of ~1,300 bp differentiated strains of *L. interrogans* and *L. kirschneri*, indicating a high degree of species conservation among the genus (Morey et al., 2006). Phylogeny based on *rrs* (16S rRNA) sequences have disclosed that *Leptospira* is one of the deepest evolutionary clades within the spirochete branch (Ahmed and Grobusch, 2012). Dendrograms generated from 16S rRNA gene sequences revealed three clades within the genus *Leptospira* (Schmid et al., 1986; Paster et al., 1991; Matthias et al., 2008). This phylogeny generated groups that were consistent with the ecology and physiology of the species: the slow-growing pathogens which are capable of infecting and causing disease in humans and animals, the fast-growing saprophytes which are environmental bacteria that do not infect humans and animals, and the so-called "intermediate" group which are species that have been isolated from humans and animals and may cause mild clinical manifestations (Levett, 2001; Bharti et al., 2003). Amplification and sequencing of single genes, such as *gyrB* (DNA gyrase subunit B) (Slack et al., 2006), *rpoB* (DNA-directed RNA polymerase subunit beta) (Scola et al., 2006), *secY* (protein translocase subunit SecY) (Perez and Goarant, 2010; Bourhy et al., 2013; Hamond et al., 2016), *lfb1*(Perez and Goarant, 2010) or combination thereof, as in multi-locus sequence typing (MLST) schemes (Ahmed et al., 2006; Cerqueira et al., 2010; Boonsilp et al., 2013; Varni et al., 2014), have shown even greater discriminatory power than the highly conserved 16S rRNA. Among the soft-core genes of the genus *Leptospira*, *ppk*, which encodes for a polyphosphate kinase, was able to reproduce the same phylogeny as the one based on 1371 orthologous genes, underscoring its high discriminatory power and potential use for typing at the sub-species level (Vincent et al., 2019).

Specific probes containing single nucleotide polymorphisms (SNP) have been used to differentiate between pathogenic, intermediate, and saprophytic strains. The combination of these has also allowed the identification of several relevant species such as *L. interrogans, L. borgpetersenii, L. kirschneri* and *L. noguchii*, at high sensitivity (~50 genomic copies /reaction) (Ahmed et al., 2010). These SNP were distributed among several genes currently included in MLST schemes, later discussed, like *lipL41* (outer membrane lipoprotein LipL41), *lipL32* (outer membrane lipoprotein LipL32), *secY, rrs* and *adk* (adenylate kinase).

2.2.3 Genome-based classification of the genus *Leptospira*

As stated before, the *Leptospira* genus was initially comprised of two species, classified by their ability or inability to produce disease: *L. interrogans sensu lato*, which enclosed the pathogenic strains, and *L. biflexa sensu lato*, containing the saprophytic members (Levett, 2001). Over the years, several species have emerged within each group, as well as an additional group including strains with moderate pathogenicity. These so-called intermediate leptospires are mostly environmental isolates that can be isolated from patients or animals (Perolat et al., 1998; Matthias et al., 2008; Puche et al., 2018), but such pathogenicity has been difficult to demonstrate in animal models (Thibeaux et al., 2018).

The most recent *Leptospira* classification is entirely based on phylogenomic relatedness (Vincent et al., 2019), representing a breakthrough, as it circumvents assumptions of virulence that are often uncharacterized. Thus, *Leptospira* species are currently separated into two clades (P and S) and further subdivided into four subclades: P1, P2, S1 and S2, and finally P1 is further subdivided into $P1^+$ or high-virulence and $P1^-$ or low-virulence strains (**Figure 2.2**). The $P1^+$ group includes *L. interrogans, L. kirschneri, L. noguchii, L. santarosai, L. mayottensis, L. borgpetersenii, L. alexanderi* and *L. weilii*, the only species associated with acute infection and severe outcomes in human leptospirosis. $P1^-$ (low-virulence) members, on the other hand, have been largely isolated from the environment, except for *L. alstonii* (amphibians) and *L. tipperaryensis* (shrews), and are unable to reproduce disease or bacterial colonization in animal models (Vincent et al., 2019). P2 comprises the formerly termed intermediate group (Guglielmini et al., 2019), which, like P1-low-virulence, fail to reproduce disease in animal models. Lastly, S1 and S2 members comprise saprophytic species, which from now on will be referred collectively as S. Having outlined the

different groups, it is worth highlighting that 30 out of 64 species presented by Vincent and collaborators were identified and characterized for the first time in that study: four species belonging to the P1 group; ten additional species identified as P2 members; and sixteen strains within the subclades S1 and S2 (Vincent et al., 2019). Since that phylogeny, two species have been renamed: *L. dzianensis* and *L. putramalaysiae* were identified almost simultaneously with *L. yasudae* and *L. stimsonii* (Vincent et al., 2019; Casanovas-Massana et al., 2020). Average nucleotide identity (ANI) comparison revealed that *L. dzianensis* and *L. yasudae* were the same species, as were *L. putramalaysiae* and *L. stimsonii*. Hence, they were reclassified as *L. yasudae* and *L. stimsonii*, respectively (Casanovas-Massana et al., 2021). Four additional species have been identified, *L. ainlahdjerensis* and *L. ainazelensis* from P1 clade, and *L. abararensis* and *L. chreensis* from S1 clade (Korba et al., 2021). Currently, a total of 68 species are distributed across the *Leptospira* genus (**Figure 2.2**).

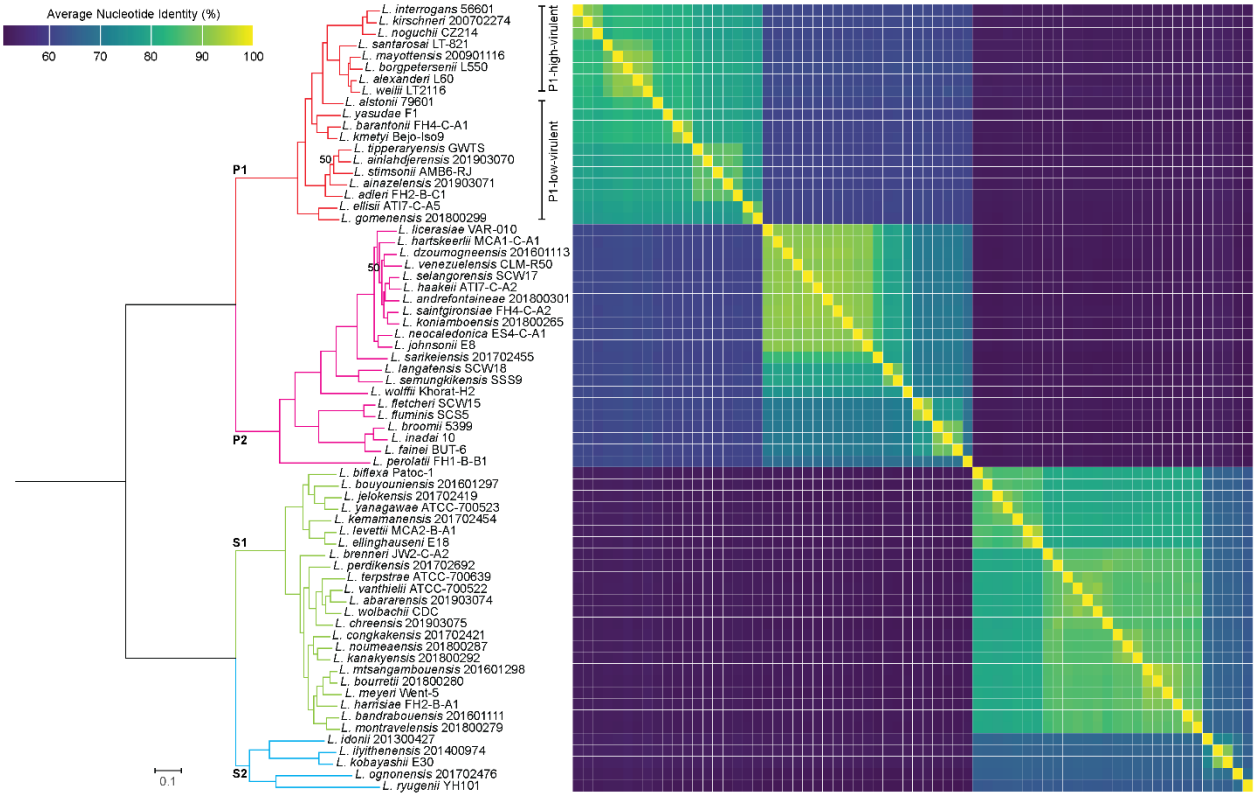

Figure 2.2. Maximum-likelihood phylogenetic tree of *Leptospira* species.

Phylogenetic tree based on soft-core genes (present in at least 95% of the genomes). Branches are highlighted as per the main subclades, P1, P2, S1 and S2. Bootstrap values are only indicated when differing from 100 (separation between *L. stimsonii* AMB6-RJ and the branch including *L. tipperaryensis* GWTS; separation between *L. hartskeerlii* MCA1-C-A1 and the branch that includes *L. dzoumogneensis* 201601113). The matrix on the right represents the ANI values of all genomic sequences.

ANI computations clearly delineate clusters, in line with the phylogenetic distribution. Among the P1 group, ANI ranges from 76.4 to 92.8%, with the highly virulent *L. interrogans, L. kirschneri* and *L. noguchii* showing close similarity (~89-91%). P2 strains are slightly more diverse, ranging from 67.6 to 93.2% of nucleotide identity to each other, and from 60.1 to 62.6% compared to P1 group. S1 exhibits similar variation as P1, with less interspecies diversity (~78.8-94.5%), but further distanced from the P members (51.2-53.3%). Finally, the smallest clade, S2, also showed a wide ANI interval from 65.2 to 90.5%, and between 50.9 to 67.1% compared to the other groups.

When comparing the percentage of conserved proteins (POCP), intra-group variation is observed. Strains within the P1 cluster are undoubtedly the most variable in terms of protein repertoire, ranging from 52.9 to 93.2%, whereas the remaining clusters showed less variability (P2=74.7-94.5%; S1=79-92%; S2=72.6-86.7%). Differences in protein conservation levels might predominantly result from lifestyle factors. In other words, the P members which can survive in soil and water until finding a host (except for the strictly host-adapted *L. borgpetersenii*) require greater adaptability to multiple habitats than S members, which are exclusively free-living (Ko et al., 2009; Xue et al., 2009). Such capacity to adapt to different scenarios (environment *vs.* host) certainly influences the protein content, especially when the targeted host spectrum is as broad as that of P1 strains of *Leptospira.* Indeed, this is consistent with the idea that P strains have tailored their genomes from a free-living to a host-associated lifestyle through expansion of certain protein families (Thibeaux et al., 2018).

2.2.4 Genotyping

The advent of genomics and the large volumes of accessible data has provided a more solid foundation to search for DNA sequences useful in typing. An example is MLST typing, whose development has been intensively addressed throughout the years. The first MLST scheme developed for *Leptospira* typing included six genes, namely *adk, icdA* (isocitrate dehydrogenase), *lipL32*, *rrs*, *secY*, and *lipL41* (Ahmed et al., 2006). Of them, *rrs* was highly conserved, while the most diverse was *icdA,* accounting for 51 different alleles among 120 isolates analyzed. Indeed, *icdA,* along with *secY,* showed the highest rates of synonymous substitution, highlighting their contribution in strain divergence (Ahmed et al., 2006). Phylogenetic relatedness using this six loci-based MLST proved to be useful for species-level clustering, as serogroups and serovars from

different species do not fall into the same group. In addition, *secY*, which has been used as a single gene in typing given its high discriminatory power, has also been sequenced from blood (Perez and Goarant, 2010; Bourhy et al., 2013; Grillová et al., 2020) and urine samples (Hamond et al., 2016), being valuable for epidemiological purposes. The second MLST scheme generated included seven instead of six genes, originally developed by Thaipadungpanit et al., 2007, and later modified by Boonsilp et al., 2013. The genes included were *glmU* (UDP-N-acetylglucosamine pyrophosphorylase), *pntA* (NAD(P)(+) transhydrogenase), *sucA* (2-oxoglutarate dehydrogenase), *tpiA* (triosephosphate isomerase), *pfkB* (1-phosphofructokinase), *caiB* (acyl-CoA transferase), and *mreA* (rod-shape determining protein RodA). Phylogenetic analysis of all seven concatenated genes strongly supported species assignment, with 100% accuracy, even resolving ambiguous phylogenies by *rrs* (Boonsilp et al., 2013). An alternative scheme with seven genes, consisting of a combination of the first two, is also currently applied (*adk, glmU, icdA, lipL32, lipL41, mreA* and *pntA*). This scheme proved to be useful not only for confirmation at species level, but also intra-species discrimination, as several *L. interrogans* strains were further separated (Varni et al., 2014). Some alternative schemes have emerged, for example, the combination of *ligB* (immunoglobulin-like protein B), *secY, rpoB* and *lipL41* (Cerqueira et al., 2010). After testing all types of combinations, it was concluded that *ligB* does not add discriminatory power, while complete resolution was achieved after inclusion of *rpoB* and *lipL41*. Yet, some problems remain in discriminating between *L. interrogans* and *L. kirschneri*. Recently, a core genome MLST (cgMLST) was created (Guglielmini et al., 2019), based on the gene repertoire across hundreds of genomes with representative P1, P2, S1 and S2 strains. The final scheme included 545 genes, and the resulting phylogenetic tree showed two major clades, separating P from S. The structure shown by the phylogeny revealed several subgroups within subclades. The discriminatory power of the cgMLST classification outperformed previous MLSTs, reaching 99.9%. Moreover, cgMLST was able to bring together highly related serovars into the same clonal group.

ANI comparisons have become more relevant. Particularly considering that it renders DDH results with high fidelity (Rosselló-Móra et al., 2011): 94% ANI was estimated to correlate with 70% DNA-DNA reassociation (Konstantinidis and Tiedje, 2005), although this cut-off was later adjusted to 95% for species determination (Richter and Rosselló-Móra, 2009). ANI does not rely on individual genes, but all conserved genes between two strains. This avoids variation in terms of evolutionary rates or HGT events having an impact when performing phylogenetic analysis. ANI

has been determined between *Leptospira* species and has endorsed the current distribution of 68 species (**Figure 2.2**). Furthermore, ANI has confirmed previous misattributed taxonomic identities (Guglielmini et al., 2019; Vincent et al., 2019), as well as the consolidation as the same species of strains identified by independent laboratories, and initially named as different species (Casanovas-Massana et al., 2021).

2.2.5 Serogroups, serovars and other clonal lineages identification

As mentioned, *Leptospira* species are divided into serogroups and serovars based on antigenic differences of surface-exposed LPS (Adler and de la Peña Moctezuma, 2010). The subtle variation of the carbohydrate portion determines the antigenic diversity at serovar level. Antigenically related serovars are grouped into serogroups. Serogroups do not constitute a taxonomic category but are useful for epidemiological purposes. From a practical standpoint, identity at serogroup level is simpler to establish, and provides a primary description of circulating strains. Serotyping at serovar level is important because most vaccine formulations are based on specific serovars circulating in a particular geographical area. In either case, serogroup or serovar identification is performed by cumbersome serological methods. The serogroup is determined following the agglutination of the strain with a panel of rabbit antisera raised against reference serovars representing the main *Leptospira* serogroups. The strain thus belongs to the serogroup corresponding to the antiserum with the highest titer of agglutination which is defined by the highest dilution of the antiserum giving a 50% agglutination of the leptospires in comparison with the control (Faine et al., 1999; Guedes et al., 2021). On the other hand, serovar is assessed by the cross-agglutinin absorption test (CAAT), which is the gold standard method for characterization at this level (Adler, 2015; Guedes et al., 2021). It requires maintenance of a large panel of reference serovars and corresponding hyperimmune sera, as well as the pure culture of reference strains and the strain to be characterized. The CAAT also requires the production of rabbit antiserum of the strain to be identified to conduct tests. After determining the serogroup of the strain under investigation by agglutination (see above), the rabbit antiserum of the unknown strain is tested against the reference serovars which belong to the same serogroup. Next step is to evaluate in a crossmatch manner those strains that were reactive with the antiserum derived from the strain to be identified. In other words, reaction of each reference strain against the antiserum of the incognita

strain will be compared with the reaction of the incognita strain against the sera of each of these reference strains. Two strains are considered to belong to different serovars if, after cross absorption with the heterologous antigen, 10% or more of the homologous titer remains in at least one of the two antisera. As can be seen, serological typing is not simple, and heavily relies on the experience of the technician. These technical difficulties have prompted the exploration of other options, especially those involving molecular biology or genomic techniques.

Restriction enzyme analysis (REA) of chromosomal DNA has been valuable for strain characterization. In such approaches, genomic DNA is extracted, digested by endonuclease(s) and the restriction fragments are separated by gel electrophoresis to give a characteristic banding pattern (Marshall et al., 1981; Tamai et al., 1988). It should be emphasized that serovars with a high degree of serological cross-reactivity are not distinguished by this technique (Thiermann et al., 1985), constituting a major weakness. Subsequent coupling between REA and Southern blot has improved sensitivity (Zuerner and Bolin, 1990). This combination resulted in the identification of new serovars that were not detected by using REA alone. Specific DNA probes have also been used to differentiate between clinically relevant serovars as *L. interrogans* Hardjo subtype Hardjo-bovis, affecting cattle, and Hardjo-prajitno (LeFebvre, 1987; Van Eys et al., 1988). Probes based on repetitive sequences have been able to detect subtle differences at genomic organization level and could successfully separate *Leptospira* into genetically related groups (Pacciarini et al., 1992), showing no cross-reactivity with other bacterial species.

Alternative techniques based on fingerprinting have also been widely used. Pulse-field gel electrophoresis (PFGE), which uses non-uniform and differently oriented electrical pulses, is suitable for separating large DNA fragments and provides higher resolution than conventional electrophoresis (Ahmed and Grobusch, 2012). Its application across a range of leptospiral serogroups successfully matched serovar identity, except for few cases, and resulted appropriate for reference and field strains. When 175 isolates from different countries and hosts were tested, 78% were identified at serovar level, and an additional 15% were believed to be new serovars. More importantly, representative isolates were confirmed by CAAT, and a 100% correlation was observed (Galloway and Levett, 2010). Random amplified polymorphic DNA (RAPD) is also a fingerprint-based technique. It relies on amplification by low-stringency PCR using a single primer. This technique has been used to identify serovars (Corney et al., 1993; Ramadass et al.,

1997), but also subtypes within Hardjo-bovis (Corney et al., 1993). Although RAPD results have been backed up by REA and CAAT, it fails to identify several serovars (Awad-masalmeh et al., 2012).

Other methods based on the amplification of repetitive elements have arisen. This type of PCR has been extensively used to determine strain-specific patterns distributed in the bacterial genome. Such is the example of the targeted amplification of BOX and ERIC (Bilung et al., 2018), two repetitive elements on bacterial genomes that have been used for molecular typing. Amplification of these elements in *Leptospira* showed genetic diversity across isolates but did not yield clear genotype-host correlation in most of the cases (Bilung et al., 2018). MLVA, multiple-locus variable number of tandem repeats analysis, is another technique based on amplification of repeated sequences and has shown to be a simple and rapid method for typing purposes. In *Leptospira* characterization, Salaün and colleagues' work has represented an undeniable achievement in serovar prediction, with the sole limitation of focusing exclusively on three species: *L. interrogans, L. borgpetersenii* and *L. kirschneri* (Salaün et al., 2006). By targeting five loci consisting of tandem repeats, the expected number of repeats per allele (based on the amplicon size obtained by PCR) was tabulated for each serovar. In total, 53 serovars of *L. interrogans,* 23 of *L. kirschneri* and 23 of *L. borgpetersenii* were evaluated. The protocol was further validated with clinical samples. The combination of VNTR-4, VNTR-7 and VNTR-10 provided discrimination of 92 and 90% for serovars in *L. interrogans* and *L. kirschneri,* while the combination of VNTR-10, VNTR-Lb4 and VNTR-Lb5 dissected 60% of serovars within *L. borgpetersenii* (Salaün et al., 2006). A similar approach was designed for *L. santarosai* serovars, with VNTR-S1, VNTR-S2 and VNTR-S9 (Hamond et al., 2015). None of these repeated regions were amplified in other relevant species, i.e., *L. interrogans, L. kirschneri, L. borgpetersenii, L. noguchii,* or *L. weilli.* Furthermore, MLVA patterns were in accordance with *secY* genotyping.

2.2.6 Reporting a new *Leptospira* species

The description of a new species should follow the requirements described in the minutes of the Subcommittee on the Taxonomy of *Leptospira* (Levett et al., 2021). A key requirement for proposing a new *Leptospira* species should be the isolation of at least one representative strain with any information that relates to the environment from which the strain was isolated as well as

phenotypic differences (growth characteristics at 13°C, 30°C, 37°C, in presence of 8-azaguanine, etc.) with existing species within the genus. The serogroup should be determined by using the microscopic agglutination test (MAT) method, using representative serovars of the major *Leptospira* serogroups. The type strain should be deposited in an international culture collection and in at least two Reference Centers or WHO Collaborating Centers for Leptospirosis. DDH should be optional and replaced by the availability of a complete or draft genome and the ANI values with other described species. For whole genome sequencing (WGS), it is important to sequence DNA of a clonal culture derived from a single colony isolated in solid medium to prevent a mixed population.

BOX 2.1 Gene network association for serovar determination

Serovar identity depends on antigenic determinants related to LPS. Special focus has been placed on genes at the *rfb* locus (Peña-Moctezuma et al., 1999; Fouts et al., 2016), which are involved in the biosynthesis of the outermost part of LPS, the *O*-antigen.

Several attempts of PCR-based genotyping directed to *rfb* genes have been performed. Early studies tackled amplification of the *wzy* gene, which encodes the *O*-antigen polymerase. Comparison of *wzy* genes in different strains yielded some clustering between the most genetically related serogroups, but did not clearly correlate with classical serogroup or serovar classification (Wangroongsarb et al., 2007). The fact that *wzy* shares a high identity between different serogroups may explain these results (Cai et al., 2010). Specific genes for serogroups Canicola, Autumnalis, Grippotyphosa, Hebdomadis, Icterohaemorrhagiae and Sejroe have also been predicted. Individual amplification of these genes, coding mostly for glycosyltransferases, have allowed the correct identification of strains at serogroup level. The only exception was constituted by some strains of serogroup Sejroe, which might have undergone recombination and lost the genes (Cai et al., 2010). Although targeting a small set of serogroups, this is a good example for PCR-based serogroup identification. Finally, six *rfb* ORFs showing variable levels of polymorphisms have also been targeted, but genotyping remained ambiguous. Despite being able to produce serovar-specific patterns *in silico*, non-specific bands were observed when the PCR was performed. Some clusters were distinguished, but not at serovar level, and while some cases clustered within serogroups, this was not always the case. Furthermore, the PCR was not assessed in different species of the same serovar (Bezerra Da Silva et al., 2011).

Recently, a comparative analysis among different *Leptospira* serovars has been carried out, focusing on the gene repertoire of the *rfb* locus (Nieves et al., in revision). In this analysis, a bioinformatics pipeline allowing pairwise comparisons of a genomic dataset followed by network association through NetworkX (Hagberg et al., 2008), was used. The pipeline results in the generation of presence/absence matrices that are useful in determining genes associated with specific populations. Thus, a unique genetic fingerprint in the *rfb* linked to serovar irrespective of species has been established (**Figure B2.1**). The analysis was extended by comparing an artificially generated *pan-rfb* against whole genome sequences, which yielded consistent results. Further exploration of such matrices would be useful to design PCR-based protocols for serovar identification.

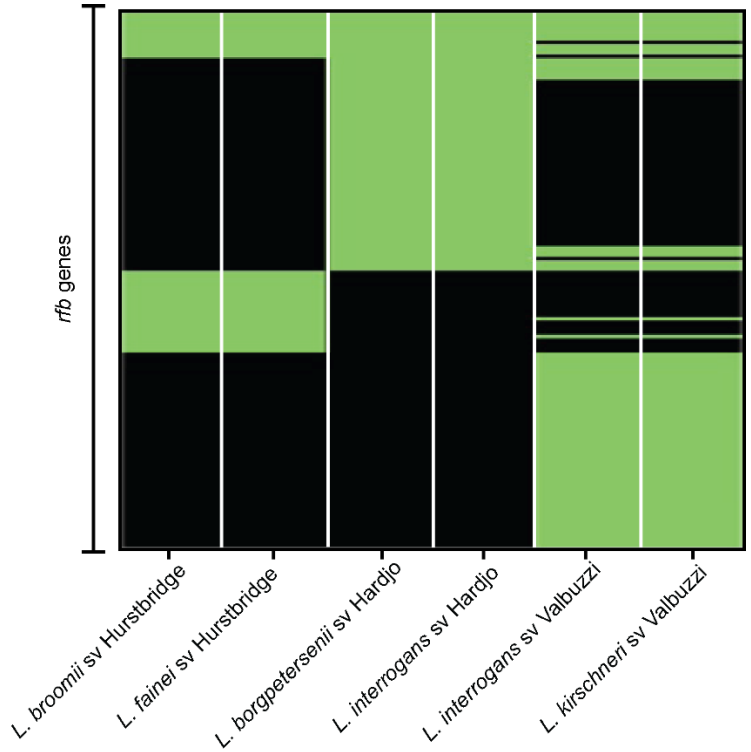

Figure B2.1. Gene presence/absence matrix of *rfb* loci from different *Leptospira* species with shared serovar identities. Rows correspond to individual genes. Presence is represented in green, while black indicates absence. Columns correspond to different *Leptospira* species, grouped by serovar as indicated in the labelling. Strains compared are *L. broomii* sv Hurstbridge str 5399, *L. fainei* sv Hurstbridge str BUT6, *L. borgpetersenii* sv Hardjo str L550, *L. interrogans* sv Hardjo str Hardjoprajitno, *L. interrogans* sv Valbuzzi str Valbuzzi, and *L. kirschneri* sv Valbuzzi str Duyster.

2.3 Genome architecture and gene repertoire

2.3.1 Replicon organization

It has been almost 20 years since the first sequencing of *Leptospira* genomes (Ren et al., 2003; Nascimento et al., 2004). At that time, two representative genomes of *L. interrogans* serogroup Icterohaemorrhagiae, of major relevance in human leptospirosis, were sequenced and analyzed. Both genomes showed an ANI index of 95%, while this percentage increased up to 99% when only orthologous genes were considered. Two circular chromosomes were observed: a large one of approximately 4.3 Mbp (chromosome I, CI) and a small one of 3.5 kbp (chromosome II, CII). Although both replicons were mostly collinear, a large inversion with flanking insertion sequence (IS) elements was evidenced on CI (Nascimento et al., 2004). Over the following years, other P and S species were also sequenced, revealing essentially the same genome architecture with slight differences in the sizes of chromosome I and II (Bulach et al., 2006; Picardeau et al., 2008; Chou et al., 2012; Ricaldi et al., 2012; Nally et al., 2016; Nakao et al., 2021). Comparative genomics studies have also provided a comprehensive overview of genomic features across the species within the genus (Fouts et al., 2016; Vincent et al., 2019). On average, *Leptospira* has a genome size of 4.3 Mbp, ranging from 3.7 Mbp in *L. fletcheri* strain SSW15 (Vincent et al., 2019) to 4.8 Mbp in *L. stimsonii* strain AMB6-RJ (Casanovas-Massana et al., 2020), P2 and P1 species, respectively. GC content is also variable, 40.7% on average, but as low as 35.5% in *L. noguchii* (Fouts et al., 2016) and reaching 47.7% in *L. fluminis* (Vincent et al., 2019). It is worth noting that both P1 and P2 clades display a large variation in terms of GC content (Vincent et al., 2019), which may account for differences in the accessory genome across the species. While there are species whose survival and replication requires association with a host (i.e., *L. borgpetersenii*) (Bulach et al., 2006), most of them persist in the environment as part of their infective cycle. This ability of adapting to either free-living or within hosts is responsible for shaping the genetic content. Other elements found in the *Leptospira* genome are plasmids. Several have been characterized, such as p74 in the saprophyte *L. biflexa* serovar Patoc strain Patoc I (Picardeau et al., 2008), or different plasmids distributed in *L. interrogans* strains (Zhu et al., 2014; Zhu et al., 2015), or in *L. borgpetersenii* (Wang et al., 2015). As for the size, those described in the literature averaged 65 kbp, varying from 54,986 bp -lcp3 in *L. interrogans* serovar Linhai strain 56609 (Zhu et al., 2015) - to 74,981 bp -pGui1 in *L. interrogans* serovar Canicola strain Gui44 (Zhu et al., 2014). GC content is variable, around 40% for those from *L. borgpetersenii* serovar Ballum strain 56604, and 36% in

those from *L. interrogans* and *L. biflexa*. The presence of some of these plasmids has been confirmed in other pathogenic strains within the genus (Wang et al., 2015), but their role in the adaptation and evolution processes of *Leptospira* remains unexplored.

2.3.2 General features

As mentioned above, the genome size of *Leptospira* is approximately 4.3 Mbp, albeit quite variable across the genus. The largest genomes are among the P group, yet their size is more dispersed, with some of them being as small as the S members. The GC content follows the same tendency, higher and sparser in P, and lower and equally distributed in S (Vincent et al., 2019). Previous studies have already shown such variability even in strains of the same species (Xu et al., 2015). Genome size range within the genus *Leptospira* is much larger than in other zoonotic pathogens (Xu et al., 2015), possibly underlining complex adaptive processes. Dispersion within the P group is stronger in the P1 clade, where the two subclades evolving divergently present indeed differences in GC content, coding ratio and percentage of pseudogenes (Vincent et al., 2019). As a matter of fact, some P species, including the well-known *L. santarosai, L. weilii* and *L. borgpetersenii*, are characterized by an enrichment of genes involved in replication, recombination, and repair, with genes encoding transposases and recombinases (Vincent et al., 2019). This is consistent with previous observations where genome reduction in *L. borgpetersenii* was ascribed to recombination events driven by the higher number of transposases as compared to *L. interrogans* and *L. biflexa* (Bulach et al., 2006).

Replication elements have been identified on both chromosomes: the *dnaA* locus (only on CI) upstream of replication origins, and the *parAB* operons on either side of them (Ren et al., 2003; Bulach et al., 2006; Picardeau et al., 2008). CI possesses most of the housekeeping genes, such as ribosomal and transfer RNA (Bulach et al., 2006), but CII also harbours essential genes, like those implicated in some amino acid biosynthetic pathways (Zuerner et al., 1993; Bourhy and Saint Girons, 2000). A cluster of genes for *de novo* haem biosynthesis has also been located on CII of *L. interrogans* strain Lai (Ren et al., 2003). Just as an entire biosynthetic cluster is localized in CII, it is noteworthy that certain metabolic pathways are dispersed between the two chromosomes; this is the case of the methionine pathway: *metX* and *metY* operating in the first steps are in CI, while *metF* is in CII, thus revealing both chromosomes to be functionally interconnected (Bourhy and

Saint Girons, 2000). Transfer RNAs (tRNA) are located on CI and present a small variation in number (35-40). It was previously suggested that the low tRNA copy number would explain the slow growth of P strains (Ren et al., 2003), but since it is similar in S (Picardeau et al., 2008), that hypothesis is not well-supported. As for ribosomal RNAs (rRNA), also located and disperse on CI rather than in clusters as in other bacteria, range from 4 to 7 (Ricaldi et al., 2012), although there are strains with up to 9 rRNA genes (i.e., *L. kirschneri* serovar Mozdok Type 2 strain) (Inácio et al., 2021).

Transposases and IS transposase-like elements are also an important feature of the leptospiral genome, playing a key role in their plasticity. These elements are responsible, for example, of changes in the genetic organization of *Leptospira*, as well as the alteration of gene function by its disruption (Nascimento et al., 2004; Bulach et al., 2006). Their number across the genus shows remarkable fluctuations from one species to another, and even between serogroups of the same species. The saprophyte *L. biflexa* strain Patoc I, has only five of these elements in its genome, whereas in *L. interrogans* strain Fiocruz LI-130 are seven times more abundant. It is even higher in *L. borgpetersenii* strain L550, where their total number is approximately 160, a five-fold increase compared to *L. interrogans* (Picardeau et al., 2008). Recently, it has been shown that the high content of transposases and integrases is a specific hallmark of P1-high-virulence strains, along with a lower percentage of coding sequences and higher proportions of pseudogenes (Vincent et al., 2019). These observations raised the hypothesis that, after the ecological habitat transition from free-living to a symbiotic lifestyle, accompanied by an initial gene expansion, the genomes of these species are now stabilizing, shifting to a greater dependence on specific niches (Vincent et al., 2019). That is clearly the example of *L. borgpetersenii*, a host-restricted P1-high-virulence species, in which the genomic reduction has been linked to IS-mediated events (Bulach et al., 2006). In other words, the genome of *L. borgpetersenii* might have already reached such steady state where it can only survive by infecting a host. More importantly, niche specialization occurring alongside genome decay has been evidenced in other pathogens (Siguier et al., 2014), strengthening this assumption in *Leptospira.*

Prophages are also present in *Leptospira* genomes. The relevance of these elements modulating bacterial gene expression or conferring some metabolic advantages when bacteria cope with hostile environments, is well known (Fortier and Sekulovic, 2013). Initial analyses including

20 *Leptospira* species revealed 14 major prophage regions, with LE1-like prophage being shared by several genomes. Prophages were mostly found in P1 and P2 members, underlining their evolutionary relevance in these species (Fouts et al., 2016). A more recent analysis, comprising more than 200 genomes of species from all clades, i.e. P1, P2, S1 and S2, and even from the P1-high-virulence and P1-low-virulence subclades, uncovered a total of 227 prophages distributed throughout the genus (Ndela et al., 2021). Most of them (about 26%) present in the P1-high-virulence strains. *Leptospira weilii* had the most prophage elements with 63% of strains carrying from one to four prophages, located on CI or CII. Flanking sequences consisted of six base pairs directly repeated, which is compatible with a transposition event mediated by the B3/Tn552 transposase family (Ndela et al., 2021). *Leptospiraceae* prophages show a genomic organization typical of B3 bacteriophages, but surprisingly harbor a different set of core genes compared to known transposable phages. Only three orthologous proteins were shared with the latter: the transposases TnpA and TnpB, and GemA of unclear function. The late regulator Mor/C was absent in all leptospiral prophages. Apart from *L. weilii*, prophages seem to be circulating among *L. alexanderi, L. borgpetersenii, L. interrogans, L. kirschneri, L. mayottensis, L. noguchii, L. santarosai* (all from P1-high-virulence clade) and *L. stimsonii* (P1-low-virulence). Of note, six groups of prophages are detected in several *Leptospira* species within different clades, just as most species are infected by prophages of several groups (Ndela et al., 2021), which suggests a horizontal transfer may be occurring.

Plasmids play a major role in terms of adaptation and evolution, although its role in *Leptospira* has not been broadly studied. The first plasmid to be characterized was p74 in *L. biflexa* strain Patoc I (Picardeau et al., 2008). Unexpectedly, this plasmid exhibited some genes typically located in CI of P members and relevant for their viability, thereby implying that p74 is essential for *L. biflexa* survival. Other plasmids have been identified in P strains, but so far, p74 is unique to the saprophytic *L. biflexa.* As a common feature, genes involved in plasmid segregation, namely *parAB* system or *rep* genes, are found in leptospiral plasmids. Post-segregational killing systems, such as toxin/antitoxin genes that ensure vertical transmission of these elements have also been identified, like those present in pGui1 and pGui2 from *L. interrogans* strain Gui44 (Zhu et al., 2014). Although genes involved in pathogenicity, regulatory functions or some conferring specific drug or heavy metal resistance have been identified, most of them encode hypothetical proteins (Zhu et al., 2014). Similar findings apply to plasmids characterized in other *L. interrogans* strains.

For example, lcp1 and lcp2 of *L. interrogans* strain 56609 have a high abundance of hypothetical proteins. Exception is made by lcp3 also from the same strain, where 73% of the coding sequences are phage-related, including functions such as bacteriophage replication, regulation and packaging, among others (Zhu et al., 2015). Absence of an integrase prompted to postulate lcp3 as an inducible prophage. Presence of plasmids has also been demonstrated in other P species, such as lbp1 and lbp2 in *L. borgpetersenii* strain 56604. Notably, the GC content in these plasmids is higher (around 40%) than those found in *L. interrogans* strain Gui44 (pGui1=35%, pGui2=33%), *L. interrogans* strain 56609 (lcp1=35%, lcp2=35%, lcp3=39%), *L. biflexa* strain Patoc I (p74=37%) or *L. interrogans* strain 56601 (Laicp=35%), all of them averaging (63691 $\pm$ 7005) bp in size (Wang et al., 2015). Interspecies distribution of these plasmids has also been evaluated: *rep* genes of lcp1, lcp2 and lcp3 were found by blastp in *L. weilii, L. borgpetersenii, L. santarosai, L. interrogans, L. kirschneri,* and *L. noguchii* (Wang et al., 2015), suggesting their circulation between P species. A summary of plasmids and their general features from complete genomes reported in the National Center for Biotechnology Information, NCBI, is presented in **Table 2.1**. As can be seen, their number is variable, with up to 7 plasmids detected in a single strain. Further assessment should be performed to infer plasmid content and their transmission across the genus, to evaluate its contribution on *Leptospira* evolution.

Table 2.1. Plasmid distribution across complete genomes of *Leptospira.*

General features of the main chromosomes as well as of the different plasmids in each strain are shown, as indicated by the column headings. Feature prediction was performed using DFAST prokaryotic genome annotation pipeline (Tanizawa et al., 2018).

		Host	Geographical location	Collection date	Replicon	Size (bp)	C+G content (%)	Protein Coding (%)	Total CDs	CDs for hypothetical proteins	IS elements	tRNAs	rRNAs
L. biflexa	serovar Patoc strain Patoc 1 (Ames) (GCA_000017605.1)	Free living	ND	ND	CI	3,603,977	38.9	92	3,368	1,580	2	35	6
					CII	277,995	39.3	90.7	266	138	0	0	0
					p74	74,117	37.5	82.5	52	33	0	0	0
	serovar Patoc strain Patoc 1 (Paris) (GCA_000017685.1)	Free living	ND	ND	CI	3,599,677	38.9	92	3,363	1,565	2	35	6
					CII	277,655	39.3	91	265	137	0	0	0
					p74	74,116	37.5	83.4	53	33	0	0	0
L. borgpetersenii	serovar Ballum strain 56604 (GCA_001444465.1)	ND	China: west region	1978	CI	3,550,837	40.2	79.7	3,056	1,373	8	38	5
					CII	361,762	40.2	75.1	322	162	2	0	0
					lbp1	65,435	41	71.8	48	29	0	1	0
					lbp2	59,545	39.7	70.2	41	18	0	0	0
L. interrogans	serovar Bataviae strain 1489 (GCA_014858865.1)	*Homo sapiens*	ND	2014	CI	4,431,091	35.2	74.7	3,556	1,595	50	37	5
					CII	353,717	35.1	74.5	282	104	1	0	0
					p1	140,106	35	76.8	136	117	0	0	0
					p2	93,306	34.9	72.2	77	64	0	0	0
					p3	83,109	33.2	78.9	83	70	0	0	0
					p4	76,580	36.1	87.6	80	73	0	0	0
					p5	69,692	35.4	64.9	68	54	0	0	0

					p6	62,980	36.5	66.7	54	42	0	0	0
					p7	39,186	33.6	73.5	30	26	0	0	0
	serovar Bataviae strain 1548 (GCA_014858935.1) **(*)**	*Homo sapiens*	ND	2015	CI	4,354,848	35.2	74.5	3,436	1,478	23	37	5
					CII	393,564	35	75	325	142	1	0	0
					p1	80,435	34.6	66.4	72	55	0	0	0
					p2	51,265	34.6	72.3	44	32	0	0	0
	serovar Bataviae strain D64 (GCA_010287845.2)	*Canis lupus*	ND	2017	CI	4,355,742	35.2	74.5	3,435	1,477	23	37	5
					CII	359,451	35.1	74.1	288	108	1	0	0
					p1	80,435	34.6	63.2	73	55	0	0	0
					p2	51,223	34.6	77	47	31	0	0	0
	serovar Canicola strain 611 (GCA_008831465.1)	ND	ND	1950	CI	4,255,595	35	74.7	3,362	1,408	18	37	5
					CII	357,285	35.1	76.1	304	117	2	0	0
					p1	75,928	34.7	67.3	59	46	0	0	0
					p2	66,534	33.4	71.4	63	57	0	0	0
	serovar Canicola strain 782 (GCA_014858915.1)	*Homo sapiens*	ND	2014	CI	4,556,406	35.3	74.7	3,683	1,680	43	37	5
					CII	354,515	35.1	74.6	283	103	1	0	0
					p1	101,385	34.6	78.9	86	71	1	0	0
					p2	90,872	34.3	73.6	75	67	0	0	0
					p3	77,059	33.8	72.1	69	48	0	0	0
					p4	74,118	33.7	66.6	74	53	0	0	0
					p5	53,819	38.5	88	80	77	0	0	0
		ND	ND	2004	CI	4,259,066	35	74.7	3,351	1,391	17	37	5

serovar Canicola strain LJ178 (GCA_008831445.1)				CII	357,037	35.1	75.7	301	114	2	0	0
				p1	75,855	34.7	67	58	44	0	0	0
				p2	66,530	33.4	72.7	60	54	0	0	0
serovar Hardjo strain L53 (GCA_008118365.1)	*Bos taurus*	ND	ND	CI	4,355,568	35	74.1	3,471	1,497	23	37	5
				CII	349,935	35.1	75	284	109	1	0	0
				p1	36,895	34.7	71.1	32	23	0	0	0
serovar Icterohaemorrhagiae strain Langkawi (GCA_014858895.1)	*Homo sapiens*	ND	2004	CI	4,367,462	35.1	74.6	3,426	1,453	16	37	5
				CII	395,410	35.6	76.6	344	159	2	0	0
				p1	63,414	34.7	60.8	56	39	0	0	0
				p2	59,030	40.1	85.5	94	91	0	0	0
serovar Linhai strain 56609 (GCA_000941035.1)	ND	ND	ND	CI	4,331,770	35	73.9	3,385	1,424	20	37	5
				CII	404,857	35.1	75.3	336	128	2	0	0
				lcp1	67,282	35.9	70.3	70	59	0	0	0
				lcp2	56,757	34.7	57.6	44	30	0	0	0
				lcp3	54,986	39.4	83.8	85	82	0	0	0
serovar Manilae strain UP-MMC-NIID HP (GCA_001047655.1)	ND	ND	2012	CI	4,238,922	35	74.5	3,313	1,383	5	36	5
				CII	358,377	34.9	76.4	299	115	0	0	0
				pLIMHP1	70,055	34.5	61.7	62	41	0	0	0
serovar Manilae strain UP-MMC-NIID LP (GCA_001047635.1)	ND	ND	2012	CI	4,238,972	35	74.5	3,308	1,377	5	37	5
				CII	358,378	34.9	76.4	298	113	0	0	0
				pLIMLP1	70,055	34.5	61.7	62	41	0	0	0
strain FMAS_AP1 (GCA_022453705.1)	ND	Sri Lanka: Anuradhapura	ND	CI	4,281,936	35	74.5	3,329	1,378	29	37	5

				CII	357,388	35	75.8	295	113	2	0	0
				pLiSL23	83,121	34.3	68.6	73	53	0	0	0
				pLiSL24	64,981	34.3	72.3	62	50	0	0	0
	strain FMAS_AP5 (GCA_022343805.1)	ND	Sri Lanka: Anuradhapura	2018	CI	4,478,874	35.1	74.7	3,597	1,579	41	37	5
					CII	370,896	34.9	74.5	300	113	2	0	0
					pLiSL10	86,491	34.3	66.6	67	47	0	0	0
					pLiSL11	64,130	43.9	84.5	99	96	0	0	0
					pLiSL12	115,951	34.6	85	112	93	0	0	0
	strain FMAS_AP6 (GCA_022343785.1)	ND	Sri Lanka: Anuradhapura	2018	CI	4,525,852	35.2	74.8	3,635	1,610	41	37	5
					CII	370,932	35.1	74.8	296	109	2	0	0
					pLiSL3	99,828	34.3	86.8	105	94	0	0	0
					pLiSL4	97,130	34.7	65.1	93	72	0	0	0
					pLiSL5	96,317	40.1	87.5	139	136	0	0	0
	strain FMAS_AP7 (GCA_022343765.1)	ND	Sri Lanka: Anuradhapura	ND	CI	4,298,104	35	74.5	3,339	1,376	29	37	5
					CII	368,617	35.1	75.9	305	115	2	0	0
					pLiSL7	100,329	34.2	65	94	67	0	0	0
					pLiSL8	64,942	34.6	71.7	59	45	0	0	0
	strain FMAS_AW1 (GCA_005222625.1)	ND	Sri Lanka: Awissawella	2017	CI	4,497,709	35.2	74.2	3,558	1,555	47	37	5
					CII	360,383	35.1	75.1	294	112	2	0	0
					pLiSL2	129,376	35.2	74.6	130	112	0	0	0
					pLiSL3	84,267	34.3	67	85	71	2	0	0
	strain FMAS_AW2 (GCA_022488445.1)	ND	Sri Lanka: Awissawella	2017	CI	4,634,565	35.3	74.4	3,679	1,642	53	40	5

				CII	359,426	35.1	74.8	291	110	2	0	0
				pLiSL35	142,315	34.9	74.9	136	117	0	0	0
				pLiSL36	84,077	34.2	60.2	70	55	1	0	0
				pLiSL37	71,951	37.2	76.8	84	78	0	0	0
	strain FMAS_AW3 (GCA_022453725.1)	ND	Sri Lanka: Awissawella	2017	CI	4,526,110	35.2	74.9	3,683	1,688	32	37	5
					CII	356,676	35.1	74.7	288	108	3	0	0
					pLiSL30	85,479	34.7	64.6	64	46	0	0	0
					pLiSL31	89,746	34.4	85.6	91	80	0	0	0
					pLiSL32	63,090	43.7	80.7	93	90	0	0	0
	strain FMAS_KG1 (GCA_022453745.1)	ND	Sri Lanka: Kegalle	2017	CI	4,458,643	35.2	74.4	3,574	1,593	59	37	5
					CII	380,003	35.2	75.2	322	139	3	0	0
					pLiSL25	109,384	36.7	73	88	54	1	0	0
					pLiSL26	82,993	35.4	75.2	67	59	0	0	0
					pLiSL27	84,061	34.1	84.5	87	76	0	0	0
					pLiSL28	88,750	34.8	62.6	74	55	0	0	0
					pLiSL29	96,233	35.2	63	87	69	2	0	0
					pLiSL34	140,032	37.1	76.6	187	168	0	0	0
	strain FMAS_KG2 (GCA_022559725.1)	ND	Sri Lanka: Kegalle	2017	CI	4,721,761	35.4	75	3,838	1,738	89	38	5
					CII	447,944	35.9	76.5	376	175	6	0	0
					pLiSL43	81,332	34	84.6	90	80	0	0	0
					pLiSL44	70,292	35.8	73.1	64	50	1	0	0
					pLiSL45	38,575	38.4	85.8	28	26	0	0	0

				pLiSL46	76,132	36.1	70.9	63	44	4	0	0
				pLiSL47	85,419	35.1	72.9	83	76	0	0	0
strain FMAS_KW1 (GCA_005222565.1)	ND	Sri Lanka: Karawanalla	2017	CI	4,306,144	35	74.3	3,372	1,409	46	37	5
				CII	355,745	35	75.6	291	114	1	0	0
				pLiLS1	78,201	34.3	63	79	57	2	0	0
strain FMAS_PD1 (GCA_022488465.1)	ND	Sri Lanka: Peradeniya	2017	CI	4,503,505	35.2	74.8	3,613	1,563	86	37	5
				CII	376,848	35.2	76.1	321	138	5	0	0
				pLiSL38	73,983	34.2	72.8	44	36	0	0	0
				pLiSL39	73,333	35.2	62	75	52	1	0	0
				pLiSL40	91,282	37	89.7	127	125	0	0	0
				pLiSL41	74,595	35	66.4	55	40	1	0	0
				pLiSL42	74,147	36	67.3	56	44	2	0	0
strain FMAS_PN2 (GCA_022343745.1)	ND	Sri Lanka: Polonnaruwa	2018	CI	4,302,184	35	74.7	3,437	1,492	11	38	5
				CII	368,724	35.1	75.1	304	120	0	0	0
				pLiSL	63,559	35.2	63.5	46	32	0	0	0
strain FMAS_RT2 (GCA_022343825.1)	ND	Sri Lanka: Rathnapura	2017	CI	4,342,805	35.1	74.5	3,380	1,428	24	37	5
				CII	409,481	35.6	76.5	350	159	1	0	0
				pLiSL13	84,167	34.3	72.1	96	77	0	0	0
				pLiSL14	84,083	34.4	70.8	82	68	0	0	0
				pLiSL15	94,947	34.6	66.5	86	70	0	0	0
strain KR40 (GCA_023158895.1)	*Equus caballus*	Italy	1997	CI	4,318,893	35.1	72.9	3,962	2,343	25	35	5
				CII	354,085	35.1	73.3	326	180	2	0	0

					pI_URK_3	123,777	35	65.9	121	101	0	0	0
	strain N116 (GCA_023515975.1)	*Bos taurus*	Belgium	2016	CI	4,312,952	35	74.1	3,422	1,467	25	37	5
					CII	354,191	35	74.7	286	107	2	0	0
					pI_URK_4a	77,345	36.7	67.3	51	45	0	0	0
					pI_URK_4b	64,666	35.1	61.3	54	32	0	0	0
					pI_URK_4c	38,354	33.5	72.3	26	21	0	0	0
	strain UI29382 (GCA_021378355.1)	ND	Laos: Vientiane	2014	CI	4,311,919	35	74.5	3,419	1,483	27	37	5
					CII	355,097	35.1	75	297	120	2	0	0
					p1	60,147	35.2	55	59	47	0	0	0
L. kobayashii	strain E30 (GCA_003114835.3)	Free living	ND	2014	CI	3,985,339	40.7	91.8	3,608	1,655	0	36	6
					CII	274,191	40.9	89.5	244	122	0	0	0
					pE30-1	45,413	39.3	73.4	36	16	0	0	0
L. mayottensis	strain 200901116 (GCA_000306675.3)	*Homo sapiens*	Mayotte	ND	CI	3,815,263	39.7	80.7	3,174	1,342	33	37	5
					CII	307,019	39.4	79.5	263	119	5	0	0
					p1_L200901116	94,000	36.4	78.8	87	76	2	0	0
	strain MDI272 (GCA_003429525.1)	*Tenrecidae*	Mayotte: Tsoundzou	2012	CI	3,861,885	39.6	80.5	3,227	1,394	50	37	5
					CII	332,647	39.2	79.6	285	138	5	0	0
					p_Lmay_tenre	52,509	39.2	64.6	73	70	0	0	0
L. noguchii	strain 201601331 (GCA_022820235.1)	*Homo sapiens*	ND	ND	CI	4,334,850	35.7	74	3,447	1,477	40	37	5
					CII	350,933	35.7	76.9	293	115	0	0	0
					p1_L201601331	65,245	35.9	65.2	46	32	0	0	0
					p2_L201601331	38,153	33.4	57	33	24	0	0	0

					p3_L201601331	28,484	35.8	34.1	26	25	0	0	0
	strain Barbudensis (GCA_022820405.1)	*Amphibia*	ND	ND	CI	4,408,823	35.5	75.9	3,883	2,065	12	37	5
					CII	359,178	35.3	76.8	337	117	0	0	0
					p1_Lbarbudensis	47,641	42.2	45.2	65	65	0	0	0
					p2_Lbarbudensis	44,440	42.5	46	63	61	0	0	0
	strain IP1512017 (GCA_022820045.1)	*Bos taurus*	Uruguay	2015	CI	4,297,194	35.6	73.3	3,375	1,451	11	37	5
					CII	370,404	36	77.1	327	144	5	0	0
					p1_LIP1512017	178,101	38.3	74.3	139	122	0	0	0
					p2_LIP1512017	67,330	37.2	76.6	48	42	1	0	0
					p3_LIP1512017	43,769	35.4	73	33	29	0	0	0
	strain IP1605021 (GCA_022819915.1)	*Bos taurus*	Uruguay	2016	CI	4,438,826	35.8	73.4	3,555	1,600	13	37	5
					CII	395,164	36.2	79.5	368	189	3	0	0
					p1_LIP1605021	169,819	38.1	75	136	115	0	0	0
					p2_LIP1605021	47,268	36.4	67.3	29	23	0	0	0
	strain IP1611024 (GCA_022819715.1)	*Bos taurus*	Uruguay	2016	CI	4,248,034	35.6	74	3,329	1,402	6	37	5
					CII	351,624	35.6	76.6	291	104	0	0	0
					p1_LIP1611024	89,758	32.4	83.1	93	84	0	0	0
					p2_LIP1611024	74,082	34.6	67	61	47	0	0	0
					p3_LIP1611024	57,448	34.3	61.2	50	39	0	0	0
					p4_LIP1611024	41,294	32.4	71.2	30	22	0	0	0
	strain IP1703027 (GCA_022819565.1)	*Bos taurus*	Uruguay	2017	CI	4,329,282	35.7	73.7	3,408	1,477	25	37	5
					CII	338,457	35.7	76.4	275	100	1	0	0

					p1_LIP1703027	83,272	35.9	71.2	52	40	0	0	0
					p2_LIP1703027	61,325	35.7	73.1	44	34	0	0	0
	strain IP1705032 (GCA_022819445.1)	*Bos taurus*	Uruguay	2017	CI	4,342,451	35.7	74	3,435	1,513	3	37	5
					CII	343,634	35.6	77.3	286	107	0	0	0
					p1_LIP1705032	58,999	35.9	66.1	50	43	0	0	0
	strain IP1709037 (GCA_022819425.1)	*Bos taurus*	Uruguay	2017	CI	4,199,394	35.7	74.1	3,354	1,442	12	37	5
					CII	376,569	36.1	76.8	326	143	1	1	0
					p1_LIP1709037	90,519	35.3	73.5	73	59	2	0	0
					p2_LIP1709037	74,078	34.5	65.7	59	44	0	0	0
	strain IP1712055 (GCA_022819405.1)	*Bos taurus*	Uruguay	2017	CI	4,329,115	35.7	73.6	3,455	1,560	25	37	5
					CII	338,447	35.7	76.2	277	104	1	0	0
					p1_LIP1712055	83,272	35.9	71.2	52	40	0	0	0
					p2_LIP1712055	61,322	35.7	71.5	44	35	0	0	0
					p3_LIP1712055	35,952	41.3	53.1	34	34	0	0	0
					p4_LIP1712055	26,771	41.1	5.7	7	7	0	0	0
					p5_LIP1712055	24,253	42	71.5	38	38	0	0	0
					p6_LIP1712055	13,148	40.3	36.8	11	10	0	0	0
	strain IP1804061 (GCA_022820565.1)	*Bos taurus*	Uruguay	2018	CI	4,322,667	35.7	73.4	3,447	1,524	14	37	5
					CII	330,637	35.8	76	280	109	2	0	0
					p1_LIP1804061	127,018	37.3	70.8	100	82	0	0	0
					p2_LIP1804061	100,275	36.3	74.1	71	61	0	0	0
					p3_LIP1804061	51,374	35.2	72.1	37	30	0	0	0

(*) Although for *L. interrogans* serovar Bataviae strain 1548 (GCA_014858935.1), CII originally appears as a plasmid on NCBI, it is likely due to an incorrect annotation, since it aligns perfectly with others CII, and presents genes already reported on that chromosome of *Leptospira* (i.e., haem biosynthetic cluster).

2.3.3 Gene repertoire across *Leptospira* species

Several studies profiling the gene repositories of *Leptospira* from P and S groups were originally focused on the comparison of *L. interrogans*, *L. borgpetersenii*, and *L. biflexa*. Examination of these species disclosed a shared set of 2052 genes, and around 650 pathogen-specific with no homologues in *L. biflexa* (Ren et al., 2003; Bulach et al., 2006; Picardeau et al., 2008). Unfortunately, nearly 60% of these genes encode hypothetical proteins, with no counterparts in other bacteria, raising the speculation that *Leptospira* may harbor novel pathogenicity determinants. Of those that can be listed are the *lig* genes encoding immunoglobulin-like repeat proteins that play a role in adhesion. Heterologous expression of LigA and LigB in *L. biflexa* improved the binding to the mammalian host fibronectin, either cellular or plasma protein, as well as to laminin, while showing no effect on binding to other extracellular matrix proteins (Figueira et al., 2011). Other pathogen-specific genes include a collagenase (LA_0872), which is overexpressed during infection and has been identified as a virulence factor (Kassegne et al., 2014), and sphingomyelinases, such as *sphH* (LA_3540), confirmed as pore-forming protein that damages the membrane of mammalian epithelial cells (Lee et al., 2002). Additional virulence factors are postulated to be key players at different stages of the *Leptospira* infectious process (Murray, 2015); factors such as FlaA2, FliY and FcpA, involved in flagella functionality, are crucial for bacterial dissemination. Indeed, mutants in the genes coding for these proteins display impaired motility and reduced virulence in acute infection models (Liao et al., 2009; Lambert et al., 2012; Wunder Jr et al., 2016). Other factors have a major role on bacterial persistence within the host; the catalase KatE, was shown to be involved in hydrogen peroxide resistance, thus counteracting phagocytic production of reactive oxygen species. Moreover, *katE* mutants in *L. interrogans* serovar Pomona or Manilae are unable to kill hamsters, which show no signs of disease (Eshghi et al., 2012). The ClpB chaperone is also linked to oxidative resistance and has an essential role in virulence; upon *clpB* complementation, the bacteria restore normal growth under different stress conditions, including oxidative exposure (Lourdault et al., 2011). Lastly, multiple other virulence-associated elements remain functionally undefined. For instance, Loa22 is thought to have a structural role given its abundance and location on the leptospiral cell envelope, but even so, its mutant exhibits moderate attenuation, through a mechanism not fully understood. LPS also needs further characterization, structurally and functionally, since many proteins involved in its biosynthesis are of unknown function (Murray, 2015). Selected mutants have shown a marked attenuation of

virulence, failing to generate symptoms in acute models of infection (Murray et al., 2010; Srikram et al., 2011). Host colonization is also compromised (Marcsisin et al., 2013), but the underlying mechanism is unclear. It should be noted that several of these virulence factors are also present in S members, conflicting their definition as a virulence factor. However, these determinants overlap normal metabolic functions, including some that may play a role in environmental survival. Also of note, even factors found only in P species differ from serovar to serovar, in some of which there is no association to virulence (e.g., *clpB*) (Lourdault et al., 2011). Beyond presence of specific genes, redundancy is also a trait in P strains: while paralogues constitute only 1.5% in *L. biflexa* strain Patoc I, they represent around 5% in *L. interrogans* strain Fiocruz L1-130 and *L. interrogans* strain 56601, and approximately 10% in *L. borgpetersenii* strain L550 and *L. borgpetersenii* strain JB197 (Picardeau et al., 2008). Further studies corroborate that occurrence of paralogs is even more pronounced in P1-high-virulence strains as compared to P1-low-virulence and P2 groups (Thibeaux et al., 2018). Leucine-rich domains (PF13855), found for instance in leptospiral proteins interacting with human epithelial cadherins (Eshghi et al., 2019), provide a clear demonstration thereof. A large number of these domains are present within P1-high-virulence species, such as *L. interrogans* (111), *L. kirschneri* (76), or *L. noguchii* (56), but sparsely or not even found in others of reduced virulence like the P2 species *L. licerasiae* (0), *L. wolffii* (2) or *L. broomii* (1) (Xu et al., 2015). Indeed, gene duplication along with horizontal gene transfer have shown a strong correlation in the stepwise acquisition of virulence determinants, thus constituting major events in the emergence of the P clade (Xu et al., 2015).

As more genomes have been sequenced, more comprehensive comparisons have been performed. Such studies have revealed that, for example, the P1 and P2 clades had virtually the same number of group-specific genes, with about 30% of their gene repertoire shared by both groups (**Figure 2.3**). Furthermore, P1 members carry more species-specific genes, with *L. noguchii* strain CZ214 having the largest number of exclusive genes (Fouts et al., 2016), which may explain the ability of this species to infect multiple types of hosts (Silva et al., 2009). P1 species show an enrichment in motile and extrachromosomal elements when analyzing functional categories, whereas S species are characterized by a predominance of energy metabolism, signal transduction and regulation functions, along with others transport-related categories (Fouts et al., 2016). Discriminatory genes between P1-high-virulence, P1-low-virulence and P2 strains have also been postulated. Although several genes encode for hypothetical proteins, some exclusively present in

P1-high-virulence strains can be listed (Thibeaux et al., 2018): LipL36, a major outer membrane protein whose expression is downregulated during infection (Haake et al., 1998), likely associated with immune evasion mechanisms; esterase/lipase (LA_2505), also downregulated at 37°C, again possibly related to immune escape strategies (Lo et al., 2009); several receptors, such as the TonB-dependent receptor LA_2641, associated with iron acquisition, a key survival strategy, and which showed to be up-regulated when interacting with macrophages (Xue et al., 2010); and the PilZ domain-containing protein LA_0142, which is downregulated in a leptospiral transposon mutant that showed attenuated virulence in a hamster model of infection (Eshghi et al., 2014). Some metabolic pathways are overrepresented in P1-high-virulence strains: chemotaxis and cell motility, cell wall/membrane biogenesis, and post-translational modifications (Thibeaux et al., 2018), that may be required for colonization and invasion of a host. Further comparison of representative species across the genus showed significant differences in amino acid, vitamin, and carbohydrate metabolism. For example, the machinery for cobalamin (vitamin B12) autotrophy is present in P but not S species, prompting the hypothesis of its production being necessary during infection if sequestered by the host (Fouts et al., 2016).

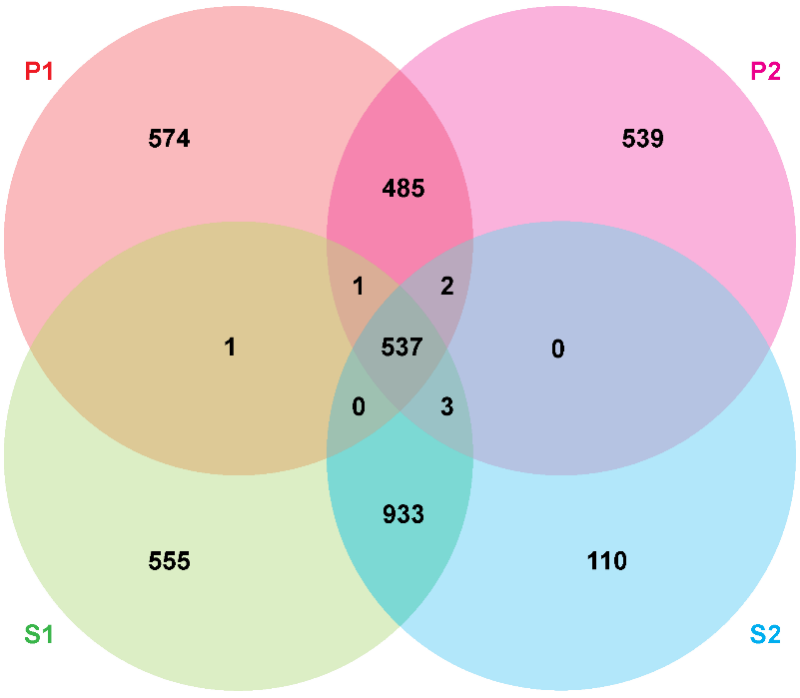

Figure 2.3. Venn diagram of gene distribution across all groups of *Leptospira*.

Pangenome distribution of *Leptospira* considering the 68 species shown in Figure 2.2 (60% identity cut-off).

A core-genome based phylogenetic analysis of the *Leptospiraceae* family using *Leptonema illini* as outgroup, showed that P leptospires were more distantly related to the most common ancestor than S species, thereby implying that virulent attributes were independently acquired during evolution (Xu et al., 2015), even between P1 and P2 species, where there are differences in the accessory genome and protein domains influencing their level of virulence (Thibeaux et al., 2018). Gene loss in the emergence of pathogenicity includes metabolic pathways and signaling elements, specifically two-component systems (TCS), which supports a shift in adaptation from a complex free-living habitat, towards a parasitic lifestyle. Indeed, the open profile exhibited by leptospiral pangenome, as well as the signs of horizontal gene transfer (e.g., protein domains in *Leptospirales* are closer to gram-positive bacteria than to other *Spirochaetales*) (Wang et al., 2022), correlate with soil microbiota cohabiting with other microorganisms, ultimately supporting a free-living ancestor (Vincent et al., 2019). The biodiversity of *Leptospira* and their hosts suggests that host-adaptation is shaping *Leptospira* evolution, which correlates with the enormous number of accessory genes in P1 members (Vincent et al., 2019), underlining specific adaptations.

BOX 2.2 DNA decoration also speaks for itself

Modification of nucleotide bases provides additional information that may explain a variety of DNA functions. Three types of methylation are identified in bacterial genomes: N6-methyladenine (6mA), 4-methylcytosine (4mC), and 5-methylcytosine (5mC), with 6mA and 4mC the most common. Regulation of gene expression is a major player, with direct involvement in defense system, cell cycle, or modulating virulence. Since DNA methylation is reversible and leaves DNA structure unaltered, it can be modulated by extracellular signals. Thus, processes promoting bacterial survival, adhesin expression, or pili formation (Seong et al., 2021) can be affected by something as seemingly harmless as a methyl group.

Long-read and single-molecule sequencing techniques have allowed the study of a wide range of DNA modifications. PacBio performs better at detecting 4mC and 6mA methylations, while Nanopore is more sensitive when it comes to 5mC methylation (Gouil and Keniry, 2019). Development and refinement of these platforms has opened a new window to understand regulatory processes that may not be explained by screening of presence/absence of genes.

DNA methylation in *Leptospira* remains understudied. A recent investigation, in which Gaultney and co-workers characterized the role of an orphan 4mC methylase in *L. interrogans* virulence, is the only direct evidence of epigenetics impact in *Leptospira* biology. This methylase, named LomA, showed to be present in P strains with no homologues in S strains. Transposon mutation of *lomA* in *L. interrogans* resulted in retarded growth, decreased motility, and incapacity to kill hamsters. LomA activity affected the expression of several genes, including an extracellular function (ECF) sigma factor. Methylation of a CTAG motif within the promoter of this sigma factor led to its downregulation, resulting in the downregulation of three lipoproteins exclusively present in P species, while absence of methylation leads to its overexpression and impaired virulence (Gaultney et al., 2021).

Methylases and methylated genes repertoires need to be further studied across the genus. Higher priority should be placed on the P group, where epigenetics may explain the variability in virulence or host-tropism. As P2 species are more heterogeneous in terms of virulence, an increased diversity of methylases/methylated sites may be anticipated. Considering the genetic background of *Leptospira*, it would not be surprising to find intra-species variation as well. Still a long way to go, and much to learn from this bacterial genus.

2.4 Concluding remarks

Techniques such as DDH and CAAT for the identification of species and serovar, respectively, will soon be obsolete and not used in any laboratory in the near future. The time has now arrived to use genome sequences for the purpose of species and subspecies delineation (Richter and Rosselló-Móra, 2009). For *Leptospira* typing, there have been significant advances: from species-level classification by comparison of individual genes, or sets of genes (e.g., MLST schemes), to prediction/clustering of specific serovars. In this regard, the development of a cgMLST scheme based on more than 500 core-genes distributed in several species from P and S clades has allowed clustering of highly related serovars (Guglielmini et al., 2019), which is beyond the performance of other MLST schemes. The major constraint in serovar determination by comparative genomics is the scarcity of sequenced serovars, barely 20% of the total reported. Furthermore, many serovars have only one representative strain, which also impedes the drawing of solid conclusions. Greater representativeness of sequenced serovars becomes imperative. This

will certainly be a unique opportunity to move from purely serological to molecular typing. Thus, characterization of *Leptospira* strains will be simplified and costs reduced. Likewise, it would also be of value to unveil serovar-specific traits, which could provide insight into the molecular basis of serovar adaptation to particular hosts.

Through the continued progress achieved in sequencing platforms, the collection of genomic data has also led to a detailed understanding of the genetic information carried by microorganisms, thus explaining the emergence of multiple phenotypes. Hence, genomics has been useful in the comparison of pathogenic *vs.* saprophytic organisms and has revealed specific determinants underlying pathogenesis, or even genes directly related to pathogen-host interaction, essential to understand the development of disease. The first *Leptospira* genomes sequenced corresponded to *L. interrogans,* which is largely overrepresented in public databases. With the sequencing of other species, genus-specific features came to light: organization into two main chromosomes, a larger one of about 4 Mbp, and a smaller one of 3 kbp. Despite CI harboring most of the essential genes, CII also has genes of relevance for the bacterial survival. There are even genes involved in the same metabolic pathway distributed in both chromosomes, showing their functional interrelation. Other evolutionary important elements, such as prophages or plasmids, have been addressed, although not in detail. There are some studies on the distribution of these elements across different species, but a more exhaustive exploration is needed. It would be interesting, for example, to establish the origin of the plasmids found in *Leptospira,* as well as to deepen the study carried out by Wang et al., 2015, but considering a broader plasmid repertoire. A characterization regarding their gene content would also be crucial to evaluate the presence of specific genetic determinants conferring some kind of advantage, such as resistance to antibiotics, which has not been addressed too much within the genus. The great challenge, as is often the case for *Leptospira*, is that most of the genes present in the plasmids code for hypothetical proteins, hindering the analysis. Obtaining complete genomes from environmental strains, which compete with a great variety of other bacterial species, would be a good starting point to evaluate the presence of these transferable elements.

Chapter 3. Intragenus evolution: a look into the gene transition from strict free-living to host-specialized organisms

(Unpublished results)

3.1 Abstract

The genus *Leptospira* is composed of a wide range of phenotypes, including non-pathogens and pathogens with different degrees of virulence, ranging from avirulent to highly virulent. Previous studies have focused on determining the genetic content or genomic features of pathogenic species to establish what defines the occurrence of pathogenicity. However, most of these analyses comprise only a few representative species from each clade, and sometimes not even all clades are represented. Here, a thorough genomic characterization of the *Leptospira* genus, including all 68 species, has been performed. Examination of the whole gene repertoire provided a comprehensive overview of the distribution of functional categories across the clades, which helps to understand how these bacteria are found in different niches. Comparison between groups also revealed a certain gene flux throughout evolution, identifying both acquired and lost genes in the emergence of pathogenic *Leptospira*. More importantly, and not studied before, amino acid permutations associated with the appearance of pathogens were evaluated. Thus, genes that have accumulated mutations resulting in amino acid replacements from S to $P1^+$ were defined. Among these, the flagellar protein FlaA2, which is key to motility and virulence, was detected. Interestingly, a comparison of gene expression between $P1^+$ and $P1^-$ (a parallel collaborative project not included in this thesis) revealed a 1.5-fold upregulation of *flaA2* in $P1^+$ species, prompting speculation about a potential association between its misregulation and the accumulated mutations. However, experimental assessment is required. Additionally, amino acid changes explaining differences in virulence between $P1^-$ and $P1^+$ species were identified. Notably, two pairs of TCS mutated in both the histidine-kinase (HK) and the response regulator (RR) were detected, and one of these TCS has previously been associated with chemotaxis. Analysis of gene expression between $P1^+$ and $P1^-$ also showed differences in these cases, with a 2-fold upregulation of the HK and RR genes of both TCS in $P1^+$ species, once again suggesting a potential influence of the mutations on the expression of these genes. Undoubtedly, their further investigation holds the potential to uncover key factors contributing to the acquisition of virulence in *Leptospira*.

3.2 Introduction

Association between bacteria and host can occur in different ways and evolve over time, which is why it is considered as a dynamic equilibrium (Lederberg, 2000). The enormous diversity and complexity of molecular interactions taking place between bacteria and host are behind the nature of this relationship and drive bacterial evolution.

In pathogenic associations, the microorganism causes changes in the physiological state of the host that can lead to the development of disease. This capacity to produce disease is a measure of its virulence, which is governed by genetic factors in the microorganism, collectively known as virulence factors. These factors, acquired through different mechanisms (horizontal gene transfer, pathogenicity islands, etc.) (Balasubramanian et al., 2022), allow the pathogen to manipulate host cells and facilitate infection. Therefore, identifying these genetic determinants is essential to understand how a microorganism becomes a pathogen and acquires a virulent phenotype. Furthermore, the identification of these genes helps to define targets of interest that could potentially be used to prevent the establishment of infection.

The diverse niches in which *Leptospira* can survive and be associated with, as well as its diversity in terms of virulence, make this genus an interesting model for studying the transition from free-living to host-associated species. In Chapter 2, several genetic elements that are differentially present in P (pathogenic) and S (non-pathogenic) members were mentioned. Genes acquired by $P1^+$ members (highly virulent) include those encoding proteins that bind to extracellular matrix elements in the host or proteins responsible for damaging the membrane of mammalian epithelial cells (e.g. sphingomyelases) (Lee et al., 2002). Furthermore, the acquisition of factors involved in flagellar functionality (Murray, 2015) has been reported. Features such as gene redundancy and horizontal gene transfer have also been analyzed. Notably, gene redundancy has shown more prominence in P compared to S species (Picardeau et al., 2008), and it is directly related to the gain of virulence, as it is more extensive in $P1^+$ strains (Thibeaux et al., 2018). HGT events also have an impact on gene acquisition during the emergence of pathogenic *Leptospira*, with approximately 30% of genes obtained by horizontal transfer (Xu et al., 2015). Gene loss has also been implicated in the emergence of pathogenicity, particularly genes related to metabolic pathways and signaling elements, providing evidence of lifestyle adaptation (Thibeaux et al., 2018; Vincent et al., 2019). **Figure 3.1** provides a comprehensive overview of the reported changes that

would play a significant role in the emergence of $P1^+$ species, as well as the events associated with genome reduction and the subsequent development of strict host dependence in *L. borgpetersenii* (Bulach et al., 2006).

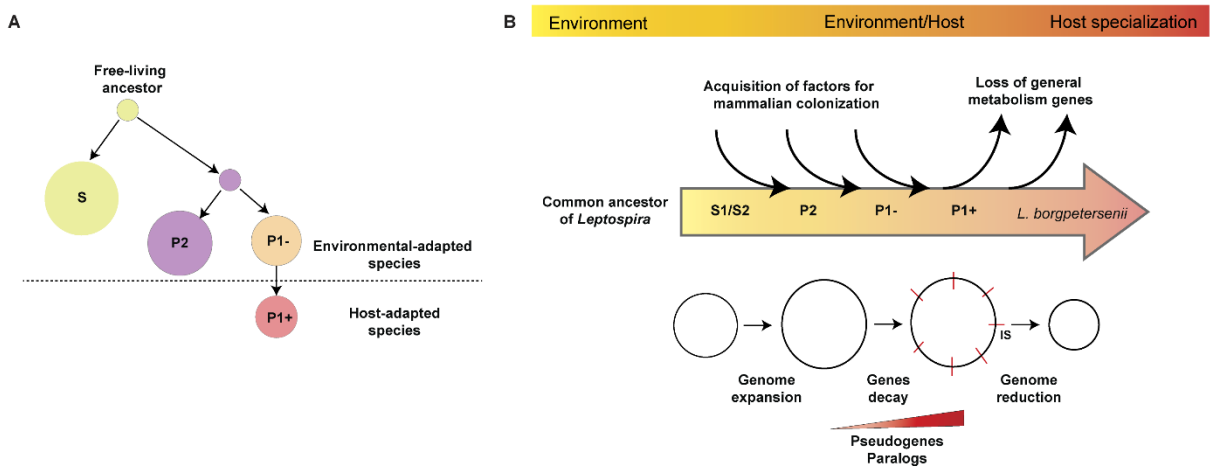

Figure 3.1. Evolution of *Leptospira*.

(A) Current hypothesis depicting a free-living ancestor. The circles representing different groups are proportionate to the number of species included in each of them (S=28, P2=21, $P1^-$=11, $P1^+$=8). **(B)** Events associated with the emergence of pathogenic *Leptospira*: acquisition of virulence factors, followed by a massive loss of genes involved in general metabolism. Appearance of highly virulent *Leptospira* is also accompanied by an increase of pseudogenes in metabolic pathways and gene duplication of other advantageous genes for host colonization. The proposed evolutionary scenario for *L. borgpetersenii*, in which initial pseudogenization and subsequent genome reduction mediated by IS elements led to its strict host dependence, is also depicted.

In this chapter, broader pangenome analyses are conducted across the 68 species presented in Chapter 2. These analyses include an overview of the functional categories represented in each group ($P1^+$, $P1^-$, P2, S1, S2) as well as other genomic features (paralogy, pseudogenes). The goal is to obtain a deeper understanding of the genetic basis that underlies pathogenicity. Both the gain and loss of genes, along with amino acid changes related to the occurrence of pathogenicity, are also evaluated. Furthermore, STRING and functional analysis are performed on these genes (acquired, lost, mutated) to determine which pathways are affected.

3.3 Results

3.3.1 Overview of protein-coding genes in *Leptospira*

The complete repertoire of protein-coding genes for 68 *Leptospira* species spanning the five subclades across the phylogeny ($P1^+$, $P1^-$, P2, S1 and S2, see Chapter 2) was compared. The distribution of these coding sequences among the groups is shown in **Figure 3.2A.** A total of 480 genes constituted the leptospiral core-genome, which is consistent with the findings presented in Chapter 2 (**Figure 2.3**). Of note, the slight differences between the Venn diagram shown here and that in Chapter 2 are the result of performing the calculation with different bioinformatic tools. As can be observed from **Figure 3.2**, the number of clade-specific genes normalized by the total number of species in each group is similar for the four major clades, namely P1, P2, S1 and S2, with a gene/genome ratio of 26, 24, 22 and 18, respectively. However, the number of $P1^+$ or $P1^-$-specific genes is notably lower, accounting for only 4% of the genes exclusively identified in the P1 clade. The genes unique to $P1^+$ encode an amidase (LA_2200) that is completely absent in $P1^-$, two putative lipoproteins (LA_1569 and LA_3834) with homologs in $P1^-$ but less than 60% similarity to those of $P1^+$ species, and two uncharacterized enzymes (LA_0589 and LA_0620). Although the two putative lipoproteins and the two uncharacterized proteins were not detected by identity cut-off, they have homologs in *L. alstonii* of over 60% similarity compared to *L. interrogans* str. 56601. It is important to note that while the majority of $P1^-$ species have been obtained from environmental samples (Vincent et al., 2019), *L. alstonii*, which is typically included within this cluster, has been isolated from amphibians. In fact, its phylogenomic position (see **Figure 2.2**) falls between $P1^+$ and $P1^-$, suggesting an intermediate genotype that bridges both groups. On the other hand, a more general comparison of P *vs* S showed a slightly lower count of clade-specific genes in P (1,417) compared to S (1,490). However, the S-core genome is 2-fold larger than the P-core genome, implying an important gene loss in the evolutionary shift from S to P, as suggested previously (Xu et al., 2015).

Contrasting with the Protein Families Database (PFAM), an enrichment of ATPases (AAA, ATPase, HATPase_C), ATP-binding cassette transporters (ABC_tran), and protein domains involved in signal transduction, such as methyl-accepting chemotaxis protein domain (MCP signal), histidine kinases (HisKA), Per-Arnt-Sim (PAS) domain, and TCS (Response_reg), was revealed in P2, S1 and S2 members (**Figure 3.2B**). All these domains with direct involvement in

general metabolism, or nutrient uptake, or secretion of toxins/antimicrobial agents, or even signaling in response to environmental changes. Conversely, $P1^+$ species exhibited a higher representation of nucleases (S1) and transposases (DEDD_Tnp_IS110, Transposase_20), as previously reported (Thibeaux et al., 2018), which may reflect the relevance of genomic rearrangements in the acquisition of virulence.

Finally, the analysis of Clusters of Orthologous Groups (COG) (**Figure 3.2C**) indicated that $P1^+$ species display a sharp reduction in the transport and metabolism of amino acids, lipids and inorganic ions, as well as in the biosynthesis of secondary metabolites. These findings may suggest a loss of non-required pathways when host specialization occurs. Transcription and signal transduction mechanisms are also less represented in $P1^+$, implying a more specialized regulation of gene expression and signaling pathways when adapting to a more narrowly defined host-specific associations. On the contrary, DNA replication, recombination and repair functionalities are increased in $P1^+$, which aligns with the findings from PFAM analysis and further supports the influence of genomic rearrangements in the tailoring of pathogenic genomes.

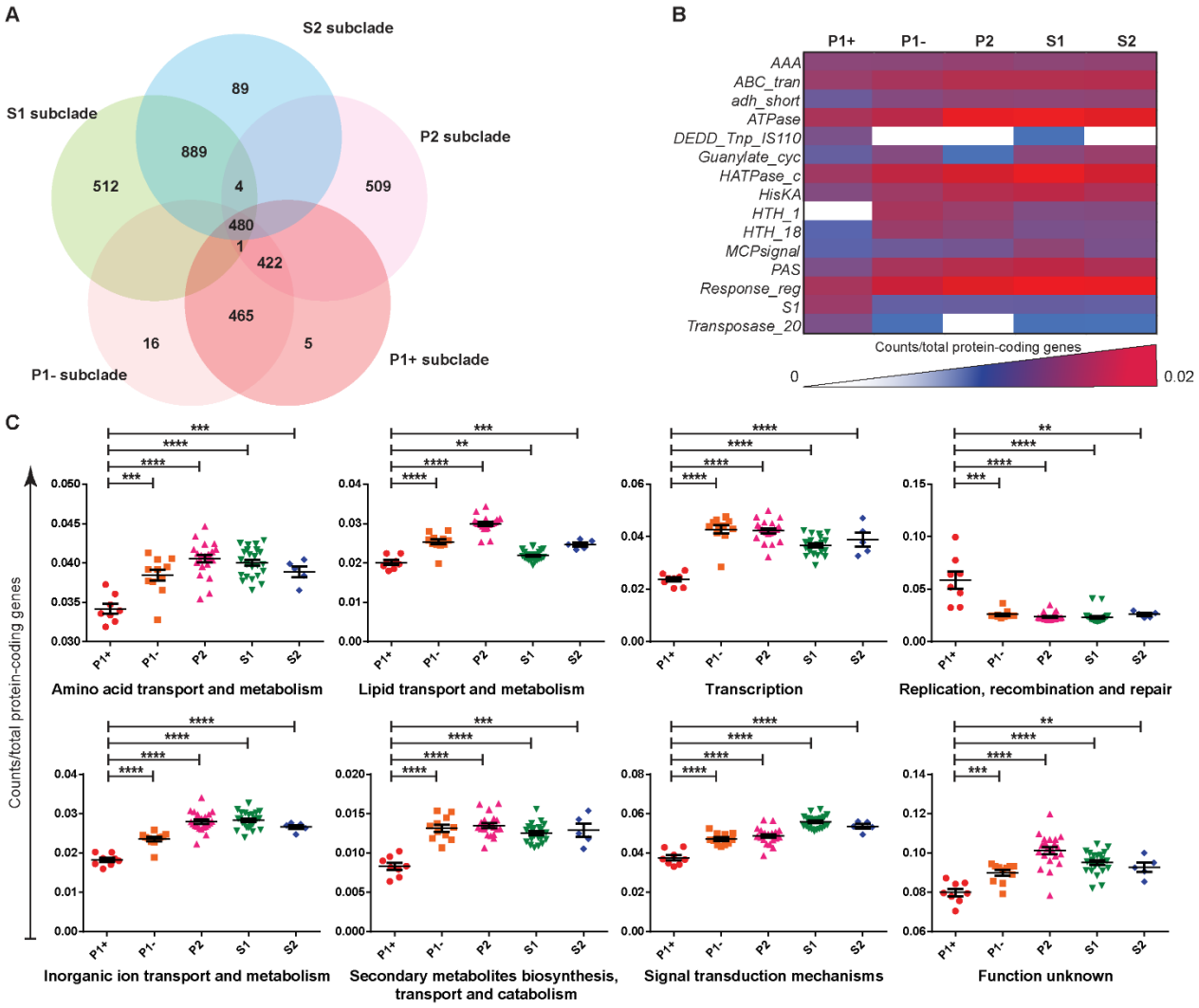

Figure 3.2. Characterization of *Leptospira* protein-coding gene repertoire.

(A) Pangenome distribution across the 68 *Leptospira* species. For group-specific genes, only the number of genes shared by all species within the group is indicated. Strain-specific genes are not included in the diagram. **(B)** Distribution of the main PFAM categories of protein-coding genes among clusters. Top 10 PFAM categories from each subclade were selected and compared between subclades. Average frequencies per subclade are coloured according to the bottom scale. **(C)** COG functional categories of protein-coding genes. Only categories in which $P1^+$ and $P1^-$ showed significant differences are depicted. Graphs represent the ratio of protein-coding genes in each category normalized by the total number of protein-coding genes.

Previous studies (Thibeaux et al., 2018) have explored paralogy in *Leptospira*, but they were limited to a smaller subset of species, specifically focusing on strains within the P group. In this chapter, gene duplication was assessed by considering all the clusters across the phylogeny. The results revealed a clear overrepresentation of paralogous genes in $P1^+$ species (**Figure 3.3A** and **B**, left panel), which aligns with previous reports (Thibeaux et al., 2018). However, a few

exceptions within the P1^{-} and S1 clusters were also observed. PFAM analysis of these paralogues showed an enrichment of transposase-related domains (**Figure 3.3B**, right panel), further highlighting the relevance of transposition events in pathogenic *Leptospira* and corroborating the results from the analysis of the whole gene repertoire.

Functional analysis of these paralogues indicated that several categories are more prevalent in P1^{+} species (**Figure 3.3C**). Once again, DNA replication, recombination and repair processes are prominent, along with paralogs involved in transcription, post-translational modifications, signal transduction and defense mechanisms. It is interesting to note that, while the analysis of the entire gene set (**Figure 3.2C**) showed a reduced presence of genes related to transcription and signal transduction in P1^{+} species, the analysis of paralogs showed the opposite trend. This could potentially suggest that even though these genes are generally less abundant in P1^{+} genomes, there might have been an expansion of a particular subset among them that confers an advantage for host colonization.

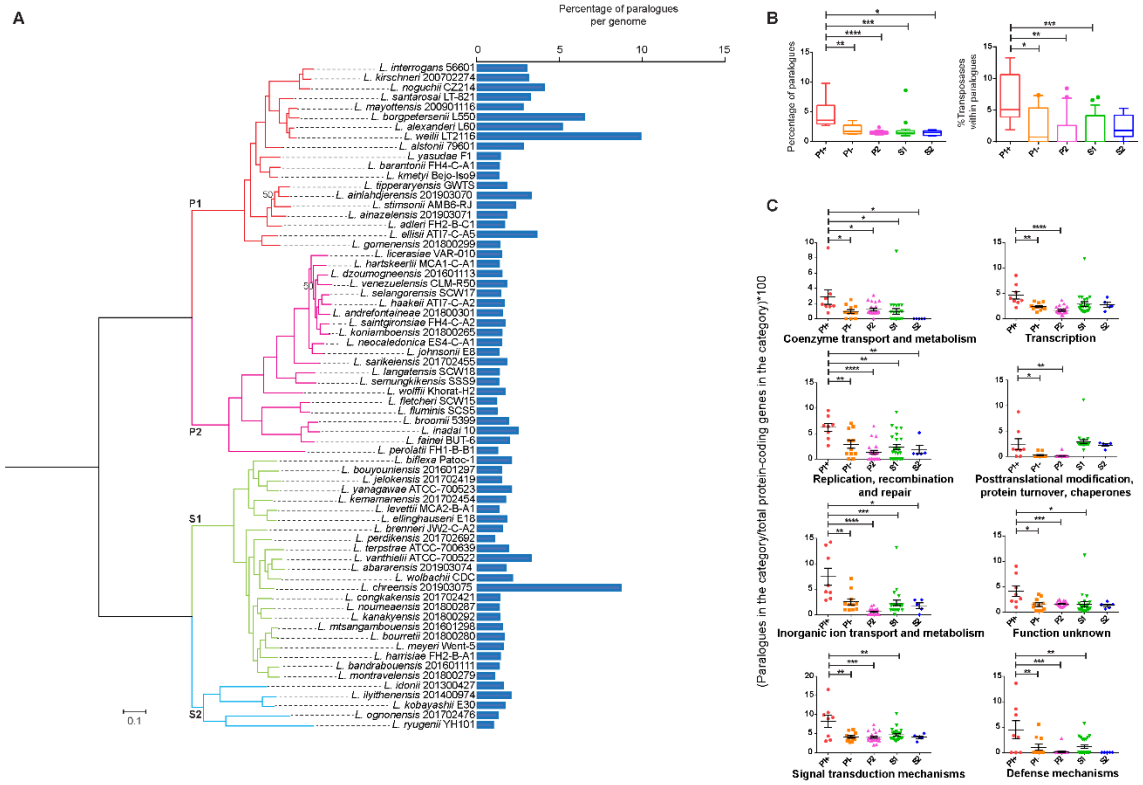

Figure 3.3. Paralogues distribution across *Leptospira* species.

(A) Blue bars adjacent to the soft-core genome-based phylogeny (see **Figure 2.2**) represent the percentage of paralogues per genome. **(B)** Tukey distribution of percentage of paralogues (on the left), and PFAM transposase-related paralogs (on the right) per group. The latter are represented as the percentage of PFAM transposase-related paralogs normalized by total of paralog genes in

each subclade. **(C)** Paralogues COG categories in which $P1^+$ and $P1^-$ showed significant differences. Graphs represent the percentage of paralogues in each category normalized by the total number of protein-coding genes in the same category.

Similarly, the distribution of pseudogenes was analyzed across all 68 species. In line with previous findings (Vincent et al., 2019), there is a clear overrepresentation of pseudogenes in P1 species, particularly in $P1^+$ species. However, there are few exceptions observed in other groups, such as the $P1^-$ *L. ellisii* (**Figure 3.4A**). After classifying the pseudogenes within P1-core or P1-cloud genome, it was revealed that most of them belong to the cloud genome (**Figure 3.4B**). This would suggest that these pseudogenes do not have an impact on conserved gene functionalities, but rather play a role in the evolution at the intraspecies level. Furthermore, the distribution of pseudogenes within the P1-core genome is not uniform across species (**Figure 3.4B**). In fact, *L. interrogans* does not possess any pseudogenes in the core genome, and those shared by several species were rare.

The COG analysis for functional profiling revealed enrichment of few categories in $P1^+$ species. It is important to note that the difference is not very pronounced, and there is a large dispersion between strains. Alongside the pseudogenes of unknown function, there is a prevalence of pseudogenes related to energy production/conversion, lipid transport and metabolism and post-translational modifications (**Figure 3.4C**). Thus, the decrease previously observed in $P1^+$ genes linked to lipid metabolism (**Figure 3.2C**) could potentially be ascribed to the gradual accumulation of mutations throughout the course of evolution. These mutations may have led to the formation of pseudogenes, rendering them inactive and no longer required. It is equally striking that $P1^+$ species exhibited an enrichment of both pseudogenes and paralogs related to post-translational modifications. This raises the hypothetical possibility of a dynamic interplay between the overrepresentation and deactivation of protein-modifying systems, which could have contributed to the emergence of pathogenic *Leptospira*.

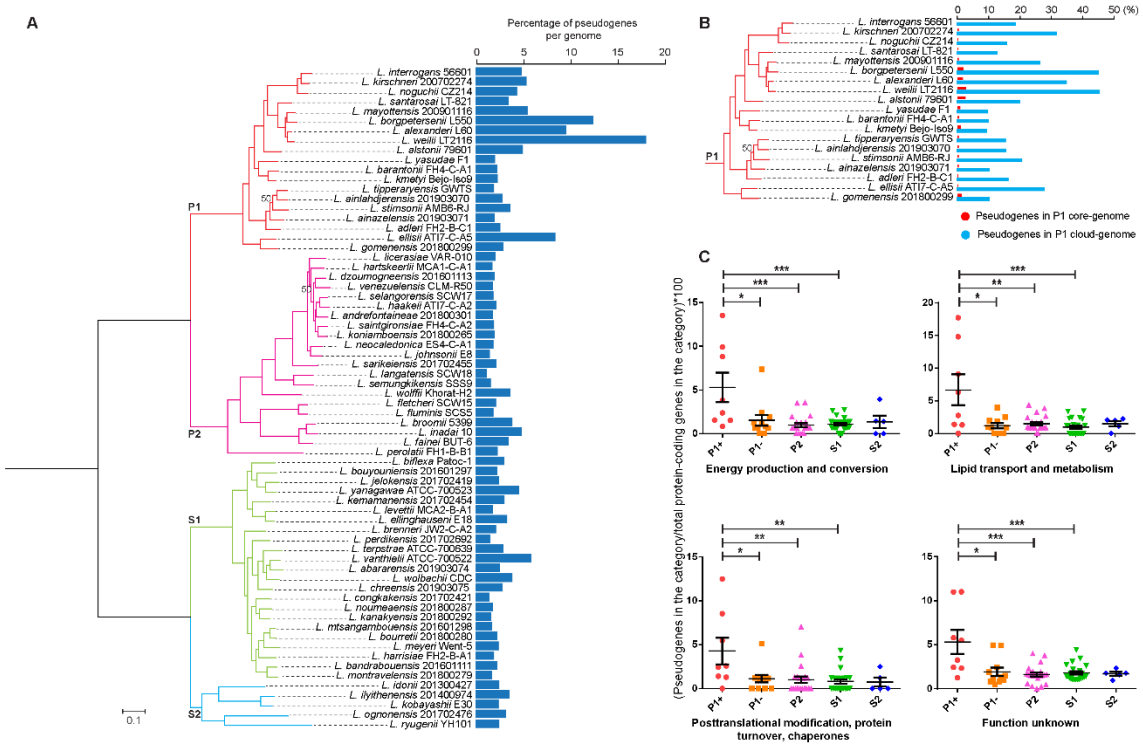

Figure 3.4. Pseudogenes distribution across *Leptospira* species.

(A) Blue bars adjacent to the soft-core genome-based phylogeny (see **Figure 2.2**) represent the percentage of pseudogenes per genome. **(B)** Percentage of pseudogenes corresponding to P1-core (orange) or P1-cloud genes (sky-blue). *L. interrogans* str 56601 did not present any pseudogene in the P1-core genome. **(C)** COG categories of pseudogenes in which $P1^+$ and $P1^-$ showed significant differences. Graphs represent the percentage of pseudogenes in each category normalized by the total number of protein-coding genes in the same category.

3.3.2 Gene acquisition and loss in pathogenic *Leptospira*

As indicated in **Figure 3.2A**, the S-core genome was 2-fold larger than P-core genome, suggesting major gene loss in the evolutionary shift from S to P, as previously implied (Xu et al., 2015).

To investigate acquired and lost genes, a comparative analysis with less stringent criteria was conducted, following a similar approach as previously employed to evaluate the emergence of *M. abscessus* (Bryant et al., 2021). Gene acquisition was defined as the presence of a protein sequence (60% similarity cut-off) in 80% of species within a specific group of interest, while being present in no more than 20% of species in the non-targeted group. Likewise, gene loss was defined as the absence of a protein sequence in 80% of species within a targeted group, while being absent

in no more than 20% of species in the non-targeted group. By employing this less stringent methodology, it is possible to infer the enrichment of genes in specific sublineages. The objective of this analysis was to gradually evaluate the acquisition and conservation of genes (or the loss of generally conserved leptospiral genes) in the most recent common ancestor of P, P1, and finally $P1^+$ sublineages.

As can be observed in **Figure 3.5**, the overall result of this analysis reveals certain gene flow in gene acquisition and loss. In general terms, the emergence of pathogens (S to P transition) is characterized by a notable gene loss, occurring at an approximate ratio of 2:1 in comparison to gene acquisition. Conversely, the transition from P2 to P1, which leads to a gain in virulence, is distinguished by a higher number of acquired genes, surpassing lost genes by a factor of seven. Finally, in the appearance of highly virulent $P1^+$, the acquired and lost genes are similar in quantity. Overall, the gene flow at this stage is less significant than the cumulative sum of genes lost and gained in other stages of *Leptospira* evolution, possibly reflecting more subtle and specific changes towards host-adaptation in $P1^+$ species. Each of these transitions will be further examined below in terms of functional categories and affected pathways.

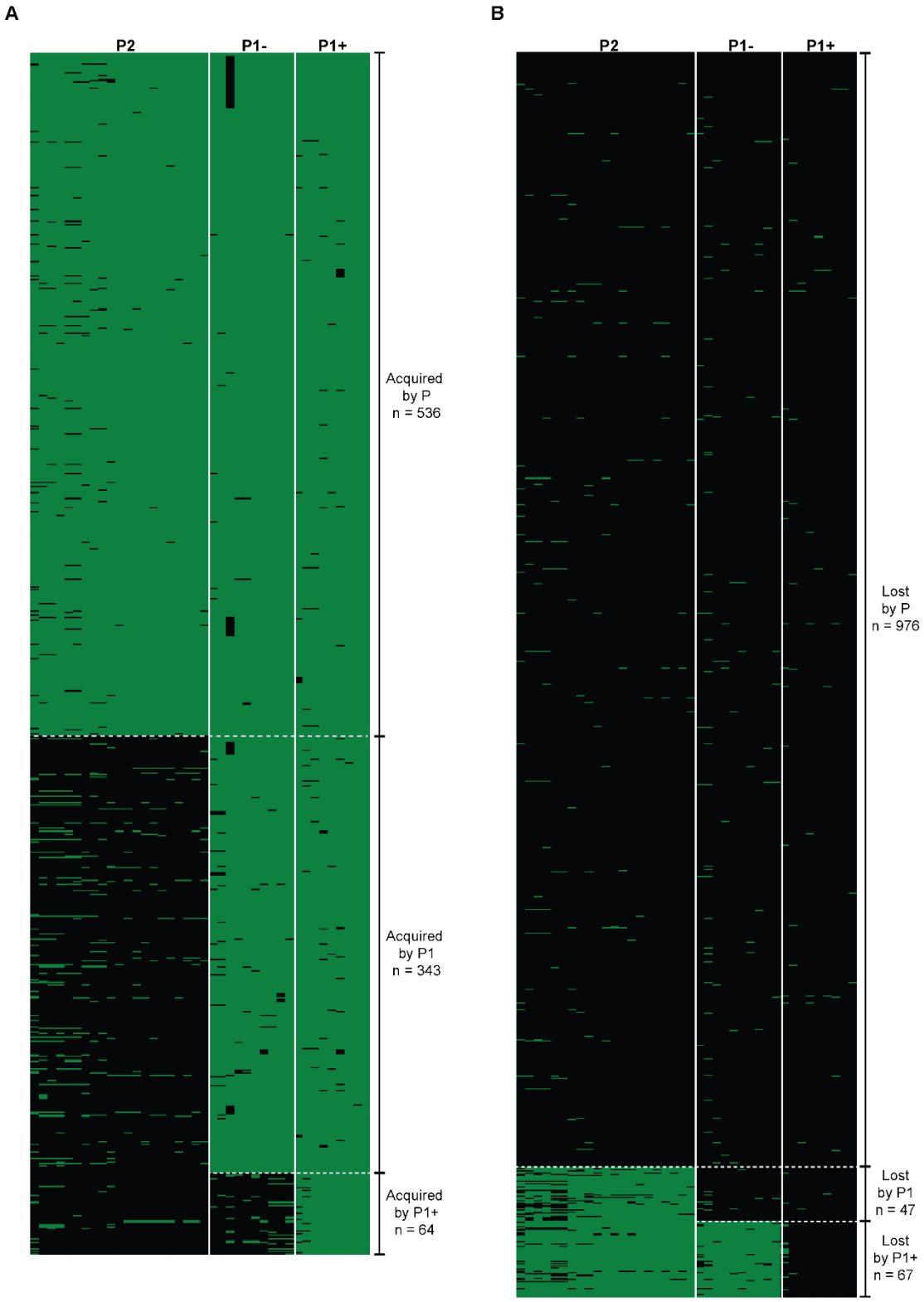

Figure 3.5. Overview of gene gain and loss in *Leptospira*.

(A) Genes acquired and **(B)** lost in the emergence of pathogenic *Leptospira* and in the gradual acquisition of virulence. Individual genes are represented by horizontal lines, green meaning presence (60% similarity cut-off), and black absence. The number of genes

acquired or lost in each transition (S→P→P1→P1$^+$) is indicated (n). In the matrices shown in both (**A**) and (**B**), P1$^+$ includes *L. alstonii*, which despite being considered as P1$^-$, has a genetic content more similar to P1$^+$.

i) Genes acquired by P2 and conserved in P1$^-$ and P1$^+$

The comparison of S and P species yielded a total of 536 protein-coding genes found in at least 80% P and sharing between 60-100% similarity. The complete list of these candidates can be found in **Table S3.2**, **Appendix I**.

As is typical of *Leptospira*, many genes identified in this search encoded hypothetical proteins. However, it is also worth noting the presence of specific lipoproteins associated exclusively with pathogenic leptospires such as LipL41 (LA_0616) or LipL32 (LA_2637) (Haake, 2000); fibronectin binding protein (LA_0322), which has been used as target gene for detection of pathogenic *Leptospira* (Merien et al., 2005; Bourhy et al., 2011); anti-sigma factor antagonists (LA_0348, LA_2461), regulators that have been previously reported to be more abundant especially in P1 (Fouts et al., 2016)*;* and even a DNA methyltransferase (LA_1547), corresponding to LIMLP_11030 in *L. interrogans* str. UP-MMC-NIID, and which has recently been shown to confer an epigenetic modification that affects virulence in *L. interrogans* (Gaultney et al., 2021).

The STRING analysis of these 536 protein-coding genes, using a high confidence cut-off for interaction score, resulted in 32 interactions. However, no functional enrichment was determined. Among the connections detected, association between *kdtA* (LA_1477) and *lpxK* (LA_3695) is shown, both of which are involved in lipid A biosynthesis. Another interesting connection is the one between *fliJ* (LA_2593), which encodes for an endoflagellar biosynthesis chaperone, and *fliO* (LA_2612), both belonging to the export apparatus of the leptospiral endoflagella. While all these genes are present in S, their homologs do not meet the utilized cut-off. Differences in lipid A, part of the LPS (known virulence factor), and the flagellar export apparatus, which is crucial for the self-assembly flagellar structure, could potentially favor the emergence of pathogens.

ii) Genes acquired by P1$^-$ and conserved in P1$^+$

Comparative analysis, now between species of S-P2 and P1 resulted in 343 protein-coding genes conserved in at least 80% of P1 species. These candidates are listed in **Table S3.3**, **Appendix I**.

Again, the number of genes encoding hypothetical proteins precludes drawing conclusions, but once more, lipoproteins were detected among the hits (LA_3415, LA_3446); a peroxiredoxin-like protein (LA_0734), homologous to a 2-Cys thioredoxin peroxidase (LA_2809) whose role in oxidative stress resistance has been suggested (Xue et al., 2010); and a cytoplasmic membrane protein (LA_3114), recently reported to be secreted in response to a temperature change from 30°C to 37°C in *L. interrogans* str. RGA (Thiangtrongjit et al., 2022), suggesting potential involvement in infection.

The STRING analysis of these 343 protein-coding genes, using a high confidence cut-off for interaction score, resulted in three interactions. The connections included the exonucleases LA_0294, LA_0965 and LA_0966, the last two also known as RecC and RecB, which are involved in DNA repair and recombination. Additionally, the genes *folC* (LA_0637) and the pyrimidine deaminase LA_1144 were found in association, both connected to folate metabolism and thus, important for acid nucleic synthesis and bacterial growth. Functional enrichment revealed a prevalence of diguanylate cyclase activity, with LA_1709, LA_2927, LA_2930, LA_2931, LA_2932, LA_2933, LA_4065, LB_238 and LB_240, belonging to GGDEF protein family. Diguanylate cyclase is involved in intracellular signalling processes in bacteria, such as motility, or biofilm formation (Ryjenkov et al., 2005), important in pathogenesis.

iii) Genes acquired by P1$^+$

Finally, comparison between S-P2-P1$^-$ and P1$^+$ species showed 64 protein-coding genes acquired by at least 80% of P1$^+$ species (**Table S3.4**, **Appendix I**).

In agreement with previous results (**Figure 3.2A**), the amidase (LA_2200) was once again detected as exclusive of P1$^+$ species. As mentioned earlier, the other four candidates (LA_0589, LA_0620, LA_1569 and LA_3834), which were initially defined as P1$^+$-specific based on a 60% identity cut-off, presented homologous sequences in *L. alstonii* (P1$^-$) with a similarity > 60%, but below this threshold in all the other species.

The STRING analysis of these 64 protein-coding genes (medium confidence cut-off, interaction score), produced one interaction, involving the methylase CheR (LA_1743), and the methylesterase CheB2 (LA_1744). Both of them associated with chemotaxis and part of the Che protein network responsible for regulating bacterial flagella rotation (Lahoz-Beltra et al., 2014).

iv) Genes lost by P2 and absent in P1$^-$ and P1$^+$

Looking at the overall loss of genes in the transition from S to P, 976 genes were specific to S species (**Table S3.5**, **Appendix I**). These S-specific genes represent 1.8-fold more genes than the P-specific ones determined earlier (536), which closely aligns with the difference observed between the S-core and P-core genomes (**Figure 3.2**).

Despite the large number of genes encoding hypothetical proteins, several predicted products were promptly searched in the KEGG BRITE database. Some of these corresponded to transporters involved in signaling and cellular processes, such as the NAD(P)H dehydrogenase (LEPBI_I2124), the cation exchanger (LEPBI_I0691), the permease (LEPBI_I1475) and the periplasmic binding protein of ABC-type transport system (LEPBI_I1476); these last two seem to be part of a TCS. Also, a number of candidates implicated in regulation, including transcriptional regulators of TetR family (LEPBI_I3106; LEPBI_I0835), or the anti-anti-sigma factor (LEPBI_I2282), previously identified as S-specific (Fouts et al., 2016). Genes involved in the biosynthesis of unsaturated fatty acids (LEPBI_I3104, LEPBI_I3105) are also noted. Not surprisingly, all these genes are related to biosynthetic pathways and transport systems, which are more diverse and numerous in free-living organisms.

The STRING analysis (high confidence cut-off of interaction score) produced 108 interactions, with no detected functional enrichment. Remarkable connections include genes related to oxidative response, *sodB* (LEPBI_I0027) and *katG* (LEPBI_I2495), which are completely absent (0% similarity) in P species. Additionally, several flagellar proteins, including FliJ (LEPBI_I0953), FliO (LEPBI_I2551) and FlgA (LEPBI_I1533), all belonging to the basal body of the flagella, and which exhibited a 50% similar counterpart in P species. Furthermore, an interaction between a PilZ domain containing-protein (LEPBI_II0088) and FliJ was determined. Given that PilZ domains serve as receptors for the c-di-GMP second messenger (Xiao et al., 2018), and the complete absence of LEPBI_II0088 in P species, this network of interactions may represent an S-specific mechanism for regulating motility.

v) Genes lost by P1$^-$ and absent in P1$^+$

Comparative analysis between species of S-P2 and P1 resulted in 47 protein-coding genes that were lost in at least 80% of P1 species. These candidates are listed in **Table S3.6**, **Appendix I**.

Among the notable genes, several response regulators stand out: transcriptional regulators (LEPBI_I0743, LEPBI_I1427, LEPBI_I3033, LEPBI_II0242), TCS response regulators (LEPBI_I2977, LEPBI_II0151), and anti-sigma factor antagonists (LEPBI_I1321, LEPBI_I1741). This highlights a significant loss of sensor systems.

The STRING analysis (high confidence cut-off, interaction score) revealed only five connections, with no functional enrichment observed. These connections include enzymes, such hydrolases (LEPBI_I3177, LEPBI_II0169), transferases (LEPBI_I1779), and oxidoreductases (LEPBI_I1752), along with other hypothetical proteins.

vi) Genes lost by P1$^+$

Finally, the comparison of S-P2-P1$^-$ against P1$^+$ revealed 67 protein-coding genes that were absent in at least 80% of the strains in the latter group. These candidates can be found in the **Table S3.7**, **Appendix I**.

Within the lost genes, once again many involved in signaling were identified: TCS response regulators (LEPBI_I1549, LEPBI_I2526), histidine kinase sensor protein (LEPBI_I1855), PPM-type phosphatase domain-containing protein (LEPBI_I2680), other transcriptional regulators (LEPBI_I2741, LEPBI_I3424). Also, several transporters (LEPBI_I2408, LEPBI_II0173), as well as various proteins with catalytic function.

The STRING analysis (high confidence cut-off interaction score) resulted only in two connections: *eutB* (LEPBI_I2493) with *eutC* (LEPBI_I2492), and a dehydrogenase (LEPBI_II0111) with a hypothetical protein (LEPBI_II0112). The former pair is involved in the conversion of ethanolamine to acetaldehyde and ammonia, with the latter playing a critical role in synthesizing vital components such as amino acids, among others.

3.3.2.1 Distribution of COG categories among acquired and lost genes in the evolution towards $P1^+$

The functional COG analysis of acquired and lost genes easily allowed comparison of the functions most affected in each evolutionary event ($P \rightarrow P1 \rightarrow P1^+$) (**Figure 3.6**). Thus, it is readily apparent that genes related to defense mechanisms were acquired in the emergence of pathogenic species from the beginning, but they are not evident in the pathway to the appearance of $P1^+$. Conversely, this category was affected by gene loss in the stepwise acquisition of virulence. Regarding vesicular transport systems, it is easily noticeable that more genes were lost in the appearance of $P1^+$ than those gained. The same pattern is observed for genes involved in lipid, coenzyme, carbohydrate, nucleotide, and amino acid transport and metabolism. All these categories with a noted reduction during the emergence of pathogens. Interestingly, the category related to motility is significantly overrepresented in the acquired genes linked to the development of a more virulent phenotype. This supports the idea of a reinforcement of the flagellar apparatus in $P1^+$ species, which leads to increased motility that is crucial for enhanced virulence.

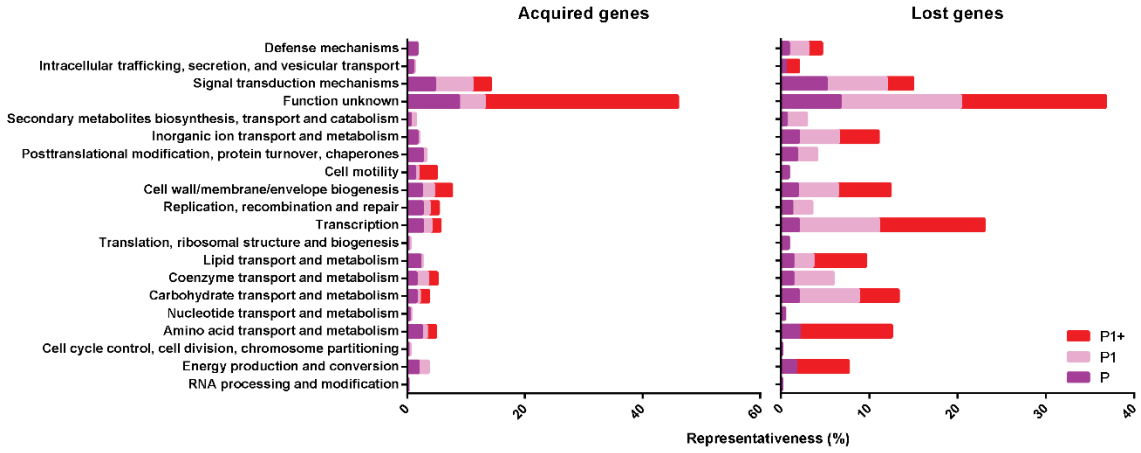

Figure 3.6. COG categories of acquired and lost genes in pathogenic *Leptospira*.

COG functional categories of protein-coding genes for acquired (left panel) and lost genes (right panel) during the stepwise evolution of P species (P→P1→P1+). The representativeness was determined by calculating the ratio of acquired or lost genes within each category to the total number of genes acquired or lost in each evolutionary event.

3.3.3 Amino acid changes involved in the appearance of $P1^+$

In addition to gene gain/loss, amino acid changes also play a role in the emergence of pathogenic microorganisms and their evolution in a number of ways (Woodford and Ellington, 2007; Laleye and Abolnik, 2020; Bentham et al., 2021). With this in mind, amino acid changes acquired by P2 and conserved in P1, by $P1^-$ and conserved in $P1^+$, and those specifically acquired by $P1^+$, were evaluated.

The overall comparison between S and P included a total of 476 protein-coding genes with amino acid permutations that were conserved in all P members and different from S (**Table S3.8**, **Appendix I**). The top three functional categories based on COG were related to translation, ribosomal structure and biogenesis, followed by energy production and conversion, and finally amino acid transport and metabolism (**Figure 3.7**). The STRING analysis (higher confidence of interaction score) resulted in a total of 3083 interactions and enrichment of several functional categories. Notably, some of the more overrepresented Gene Ontonoly (GO) categories include the small ribosomal subunit (*rpsS, rpsE, rpsD, rpsB, rpsG, rpsL*), translational elongation (*fusA, tufB, smpB, lepA, tsf, hflX, efp*), and bacterial-type flagellum filament (LA_2017, *flaB, flgL*, LA_2418, LA_4308), among others (**Table S3.9**, **Appendix I**).

The number of amino acid-mutated protein-coding genes that appeared in $P1^-$ and were conserved in $P1^+$ was very similar, with 423 candidates identified (**Table S3.10**, **Appendix I**). Most of these genes were the same than those detected in the transition from S to P, but permutations were in different positions. Evaluation of COG functional categories once again revealed the predominance of the same three categories mentioned earlier (**Figure 3.7**), indicating a preferential mutation of genes associated with specific functions. The STRING analysis (higher confidence of interaction score) resulted in a total of 2297 interactions. The top three GO categories were the AMP biosynthetic process (*adk, purA, purB, apt*), aminoacyl-tRNA metabolism involved in translational fidelity (*ileS, proS, leuS, valS*), and translational elongation (*fusA, tufB, smpB, lepA, tsf, hflX*) (**Table S3.11**, **Appendix I**).

Finally, the permutations that appeared in $P1^+$ involved 45 protein-coding sequences (not including *L. alstonii* in $P1^-$) (**Table S3.12**, **Appendix I**). The most affected functional categories were translation, ribosomal structure and biogenesis, amino acid transport and metabolism, and unknown function. However, there were other notable categories that also stood out, such as signal transduction mechanisms, cell wall biogenesis, and transcription (**Figure 3.7**). The STRING analysis (medium confidence of interaction score) resulted in a total of 27 interactions. Several proteins were classified under the GO category of cellular anatomical entity, and a number of them presented ATP- or nucleotide-binding domains (*trpS, pyrB, ruvB, yhbG, clpX, lon*, LA_3630, *hflX, metG, coaD*) (**Table S3.13**, **Appendix I**).

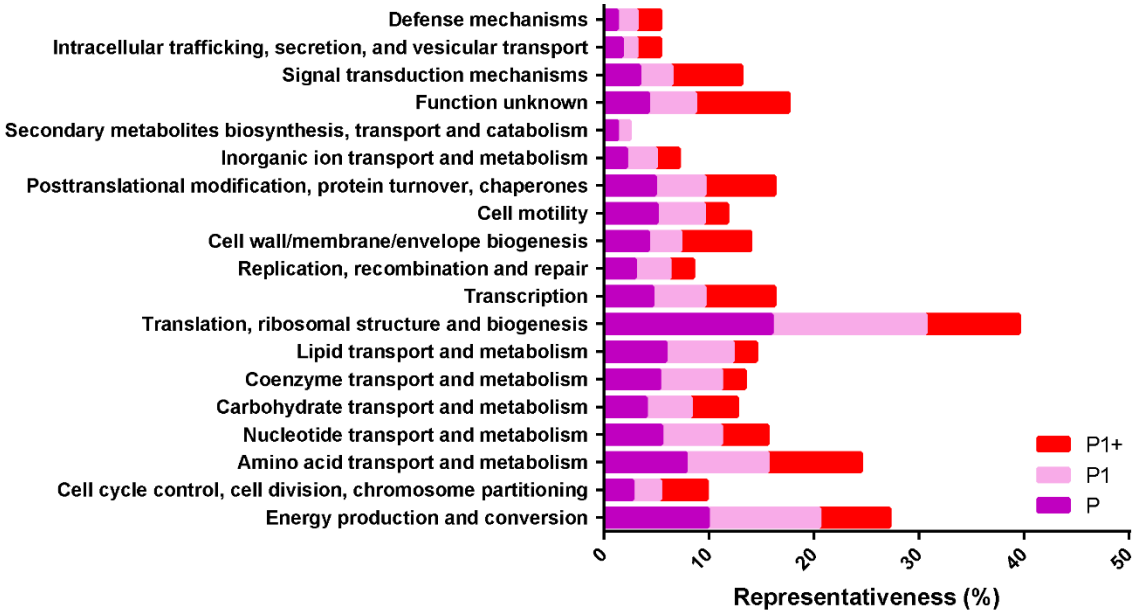

Figure 3.7. COG categories of candidates with amino acid changes linked to the emergence of pathogenic *Leptospira*.
COG functional categories of candidates with amino acid permutations that appeared during the stepwise evolution of P species (P→P1→P1+). The representativeness was determined by calculating the ratio of mutated candidates within each category to the total number of mutated candidates in each evolutionary event.

Lastly, amino acid permutations between $P1^-$ to $P1^+$ were examined to identify specific changes associated with increased virulence and host specialization in the latter (**Table S3.14**, **Appendix I**). This comparison revealed two TCS with the sensing and effector partners mutated (LA_2421/LA_2422 and LA_2540/LA_2541). The predicted permutations were confirmed by protein sequence alignment and were effectively conserved throughout $P1^+$ and different in $P1^-$ (**Figure S3.1-S3.4**, **Appendix I**). These findings become relevant considering that TCS have been directly linked to virulence in *Leptospira* (Adhikarla et al., 2018). Moreover, a transposon mutant in a gene encoding a sensor protein (*lb139*), which displayed reduced motility and attenuated virulence in hamsters, showed downregulation of LA_2421 and LA_2422 (Eshghi et al., 2014), along with other genes. This observation provides additional evidence supporting the putative involvement of this TCS in leptospiral virulence.

3.4 Discussion

In recent years, there has been a significant increase in the number of newly identified *Leptospira* species, leading to a refinement of the phylogeny structure. In addition to the original classification of saprophytic (S) and pathogenic (P) species, five subclades are now distinguished: S1 and S2, containing non-pathogens, and P2, $P1^-$ and $P1^+$, corresponding to avirulent, low-virulence and high-virulence pathogenic species, respectively (Guglielmini et al., 2019; Vincent et al., 2019). The wide range of phenotypes within a single genus provides an excellent opportunity to characterize the stepwise transition between groups and gain insights into the mechanisms underlying the emergence of pathogens. In this regard, several studies have attempted comparative analyses of genes associated to specific groups to infer gene acquisition or loss, but considering a small number of representative species (Xu et al., 2015; Fouts et al., 2016; Thibeaux et al., 2018).

A current hypothesis suggests that the common ancestor belongs to a free-living species, and pathogenic members have arisen sequentially and independently, mainly through HGT events (Vincent et al., 2019c). Indeed, both HGT and gene duplication have played a significant role in gene acquisition in P species (Xu et al., 2015), although overall more genes were lost than gained in the emergence of pathogens (Xu et al., 2015). The results obtained here support the last statement, where the S-core genome is on average twice the size of the P-core genome (**Figure 3.2**); a further confirmation was provided after performing a presence/absence analysis and determining S- and P-specific genes (ratio ~1.8). Among the genes lost by pathogenic *Leptospira*, transcriptional regulators such as anti-sigma factors or anti-anti-sigma factors, and response regulators such as TCS, were identified. This observation indicates a significant loss in signaling systems during the evolution towards pathogenic species. Previous studies have shown a lower number of TCS in pathogenic species (Fouts et al., 2016), supporting the findings presented here. Among the TCS previously reported, there is a notable presence of those containing PAS domains (Fouts et al., 2016), which are involved in the response to oxygen and redox changes (Taylor and Zhulin, 1999). In fact, PFAM analysis conducted to assess the distribution of protein domains among different functional groups revealed an overrepresentation of PAS domains in environmental species (**Figure 3.2B**). Protein-protein interaction analysis also aligns with these observations. Among the interactions detected in the transition from S to P, the axis *sodB-katG* was identified, where *sodB* encodes for a superoxide dismutase and *katG* for a peroxidase, both

completely absent in pathogenic species. Superoxide dismutase acts on superoxide, generating hydrogen peroxide and oxygen, and catalase transforms this hydrogen peroxide into water and more oxygen. Thus, all together suggesting specific mechanisms against oxidative stress that vary according to the reactive oxygen species that saprophytes and pathogens may encounter in their respective niches, with saprophytes equipped with a broader repertoire to respond to a wider range of environmental signals.

Functional characterization of the missing genes in P revealed predominantly a loss in biosynthetic pathways and transporters, as previously described (Fouts et al., 2016). Moreover, a general decrease in metabolism-related functions in the P2 to $P1^+$ transition has been observed (Thibeaux et al., 2018), also evident here, with a marked underrepresentation of categories linked to amino acid, lipid, and inorganic ion transport and metabolism, among others (**Figure 3.2**). Since $P1^+$ leptospires can infect a wide range of hosts, and colonize both the kidney and liver, which have important roles in amino acid (Bröer, 2008) and lipid metabolism (Alves-Bezerra and Cohen, 2017), respectively, it is not rare that these pathways are less represented. On the contrary, it is expected that $P1^+$ species adapt their machinery to utilize resources from the host, suppressing metabolic routes that are no longer needed. Pseudogenes formation may account for some of the lost genes during evolution. An increase of pseudogenes in $P1^+$ strains was previously reported (Vincent et al., 2019), and corroborated here (**Figure 3.4**), where lipid metabolism is one of the functional categories in which $P1^+$ pseudogenes have more representation compared to the other groups. However, the distribution of pseudogenes in each category was extremely variable across species, especially in $P1^+$, leading to speculation that they may play a role in intraspecies evolution, somewhat supported by the predominance of pseudogenes within the P1-cloud genome.

Several genes acquired by pathogenic *Leptospira* could also be listed, such as those encoding the previously described lipoproteins LipL41 and LipL32 (Haake, 2000); these two lipoproteins have been identified among the most abundant proteins in the outer membrane of pathogenic *Leptospira*. Although their disruption by transposons does not affect virulence in animal models (Murray et al., 2009; King et al., 2013), likely due to the functional redundancy of other lipoproteins (Fernandes et al., 2022), they play crucial roles in pathogenicity. In the case of LipL32, it has demonstrated interaction with TLR2 of the immune system (Yang et al., 2006; Chang et al., 2016). The LipL32-TLR2 interaction in proximal tubule cells increases the expression

of proinflammatory cytokine genes in mice model, resulting in nephritis (Yang et al., 2006). On the other hand, it has been demonstrated that LipL41 forms a supramolecular assembly and binds to hemin (Lin et al., 2013), suggesting a possible role as an iron storage protein. This is particularly significant in pathogenic microorganisms, as iron is essential for their growth and virulence.

Additional genes were acquired by pathogenic species: other putative lipoproteins not yet characterized (LA_2030, LA_2266); several regulators (LA_0348, LA_2461); and even DNA-modifying enzymes (LA_1547), with a proven impact on *L. interrogans* virulence (Gaultney et al., 2021). In the study conducted by Gaultney and colleagues, LA_1547 was also exclusively identified in the P clade. A transposon mutant of this methylase in *L. interrogans*, exhibited dysregulation in signal transduction pathways, LPS biosynthesis, and membrane proteins. Importantly, this mutant did not demonstrate virulence in a hamster model of infection. No correlation was observed between the misregulation of genes and their direct methylation, as assessed by the presence of the target motif CTAG in the genes. However, the study revealed that the overexpression of three lipoproteins lacking CTAG motifs in their sequences was controlled by the activity of an ECF-sigma factor that harbored a CTAG motif in its promoter. Indeed, the absence of the methylase resulted in the upregulation of the ECF sigma factor and the three lipoproteins (Gaultney et al., 2021), demonstrating the significance of acquiring specific genes in regulatory cascades and virulence. Among the $P1^+$-specific candidates, the amidase (LA_2200) was detected. Interestingly, this amidase was downregulated along with a number of other genes in an avirulent mutant of *L. interrogans* (Eshghi et al., 2014), exposing a complex network of transcriptional regulation directly influencing virulence. In addition, upregulation of this amidase at physiologic osmolarity has been reported (Matsunaga et al., 2007), pointing to a possible connection with the infection process.

Gene acquisition was also evaluated in terms of paralogy. As already reported (Xu et al., 2015; Thibeaux et al., 2018), $P1^+$ species have higher abundance of paralogs (**Figure 3.3**), with larger representation of those transposase-related, as described before (Thibeaux et al., 2018; Vincent et al., 2019). Functional categories related to transcription, post-translational modifications, signal transduction and defense mechanisms were better represented in $P1^+$. However, as emphasized for pseudogenes, the broad distribution of each category within $P1^+$ species hints towards implication in intraspecies evolution.

Lastly, amino acid changes in association with the emergence of pathogenic *Leptospira,* to date not analyzed at this scale, were determined. Functional analysis of genes mutated in the emergence of P, showed preponderance of translation, ribosomal structure and biogenesis, energy production and conversion, and amino acid transport and metabolism. The categories related to energy production and amino acid metabolism were particularly affected in terms of lost genes, placing them as metabolic pathways that were preferably shut down in pathogenic *Leptospira.* When specifically examining the changes related to the gain of virulence from $P1^-$ to $P1^+$, some flagellar proteins became apparent. One of these encodes a protein constituent of the flagellar basal body, MotA (LA_3577), while the other does for a protein of the flagellar filament, FlaA2 (LA_3380). It has been shown that mutation of *flaA2* alters the hook-shaped ends and results in impaired translational motility. Moreover, its interruption impeded the production of FlaA2 but also FlaA1, another protein forming the flagellar filament; overall leading to an avirulent mutant in *L. interrogans* (Lambert et al., 2012), with no variation in the flagella sheath morphology. Taken together, the amino acid changes on FlaA2 evidenced here could have promoted a gain in motility, ultimately contributing to a more virulent phenotype.

Continuing with the differences between $P1^-$ and $P1^+$ species, numerous predicted proteins involved in regulation and signaling were found to be affected, many of which belonged to TCS. These systems have been linked to virulence, as exemplified by the Lvr system, where the mutation of either *lvrA* or *lvrB* causes a drastic attenuation of *L. interrogans* in a hamster infection model. Furthermore, mutation of *lvrB* or *lvrAB* affects the transcriptional regulation of over 800 genes (Adhikarla et al., 2018), including several related to motility. Here, two TCS pairs (LA_2421/LA_2422 and LA_2540/LA_2541) were identified, each with a single amino acid permutation in the histidine kinase and another in the response regulator totally conserved in all $P1^+$ and different in $P1^-$ (**Figure S3.1-S3.4**, **Appendix I**). The LA_2540/LA_2541 system has not been characterized yet, but LA_2421/LA_2422 has been related to chemotaxis (Lambert et al., 2015). Interruption of LA_2421 affects the change of direction during swimming motion, but no change in virulence nor the speed and trajectory in media of different viscosities. Since LA_2421 is the first gene in the chemotaxis operon in *L. interrogans*, of which some genes interact with the flagellar machinery and control the change in rotational sense, the role of this TCS in signal transduction regulating chemotaxis seems unambiguous (Lambert et al., 2015). A role in virulence is not ruled out when infection routes other than intraperitoneal are involved (Lambert et al., 2015),

especially considering that the back-and-forth movement promotes colonization (Nakamura, 2022). In addition, both LA_2421 and LA_2422 were downregulated along with a hundred genes, including the amidase LA_2200, after interruption of a gene coding for a sensor protein on chromosome II (LB_139) of *L. interrogans*. As mentioned earlier, the rearrangement in the transcriptional regulation caused by the LB_139 mutation resulted in avirulence (Eshghi et al., 2014). Interestingly, LB_139 appeared as a mutated candidate exclusively in the $P1^-/P1^+$ comparison, when not considering *L. alstonii*. LB_139 has a C-terminal PPM-type phosphatase-like domain (corroborated by InterPro), which plays key roles in signaling pathways. Differential gene expression between $P1^+$ and $P1^-$ strains is currently being analyzed in a parallel collaborative project. Surprisingly, the HK and RR of both TCS were upregulated 2-fold in $P1^+$ (data not shown, manuscript in preparation), suggesting a possible effect of the permutations in the expression of these genes. LB_139, on the other hand, present a log2(FC) of 3.3, which means that it is upregulated by approximately 8-fold in $P1^+$. Undoubtedly, replacing amino acids present in the $P1^+$ species by the ones associated to $P1^-$ would provide valuable insights into the effect of these changes in relation to virulence and discern any impaired pathways.

3.5 Materials and Methods

3.5.1 Genome data sets

Representative genomes for the 68 *Leptospira* species were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) and are summarized in **Table S3.1** (**Appendix I**).

3.5.2 Pangenome analyses

All genomes were annotated by Prokka 1.14.16 (Seemann, 2014), and the resulting annotation files were used in the subsequent analyses. Pangenome distribution was inferred using GET_HOMOLOGUES version 20190411 (Contreras-Moreira and Vinuesa, 2013) via Annotator (*in-house* software under development) (cut-off 60% identity). Protein-coding genes for each species were categorized according to the Clusters of Orthologous Genes (COG) and Protein families (PFAM) databases using eggNOG mapper (options *--evalue 0.001 --score 60 --pident 40* --

query_cover 20 --subject_cover 20 --target_orthologs all --pfam_realign denovo) (Huerta-Cepas et al., 2017). Representativeness of COG categories per species was calculated as the ratio of protein-coding genes in each category normalized by the total number of protein-coding genes. PFAM categories contribution was calculated as the sum of counts per category normalized by the total number of protein-coding genes of each species. These frequencies were later used to calculate average frequencies as the sum of individual frequencies/number of strains per subclade.

Paralogs were determined by Roary 3.11.2 (option *-s*) (Page et al., 2015), and categorized by comparison against PFAM and COG databases via eggNOG mapper, as described above. Percentage of paralogs was calculated by considering the total number of proteins in each genome. Representativeness of paralogs transposase-related in each cluster was calculated as a fraction of paralogous genes in that category normalized by the total number of paralogs per cluster. Distribution of paralogs in the different COG categories was represented as a percentage of the total protein-coding genes classified in the same category.

Pseudogene prediction was performed using Pseudofinder version 1.1.0 (option *–annotate*) (Syberg-Olsen et al., 2022). Search was done against the UniProt/TrEMBL protein database by alignment through DIAMOND (option *–diamond* on Pseudofinder command line) (Buchfink et al., 2015). Percentage of pseudogenes was calculated by considering the total number of proteins in each genome. Pseudogenes in P1-core or P1-cloud genome were determined by protein-protein searches in two independent databases generated by tools from the Blast+ package (Camacho et al., 2009). One database including coding sequences present in all the P1 species here analyzed (P1-core genome), and the second one containing coding sequences encountered in at least two P1 species (P1-cloud genome). The percentage of pseudogenes included in the P1-core genome was calculated considering the total number of genes in the P1-core genome. In the case of P1-cloud pseudogenes, since not all the species have the same number of genes in the P1-cloud genome, the percentage was calculated considering the number of P1-cloud genes per species. Protein-coding sequences of the identified pseudogenes were also inspected by eggNOG mapper for its functional characterization, as described above, and representativeness of COG categories was expressed as the percentage of pseudogenes in each category considering the total protein-coding genes classified in the same category.

All graphs and statistical analysis between groups (parametric/unpaired t tests) were performed using GraphPad Prism version 6.0.

3.5.3 Gene gain/loss analyses

Gene insertion and deletion was evaluated by an *in-house* developed software, MycoHIT (Veyrier et al., 2009). Tblastn searches (default E-value = -10) of all 68 *Leptospira* species were independently performed against reference protein-coding sequences of *L. interrogans* str. 56601 and *L. biflexa* str. Patoc 1 (Paris) to respectively evaluate acquisition or loss of genes in the emergence of pathogenic *Leptospira*. Presence/absence of genes was defined based on 60% similarity cut-off in at least 80% of target group. Furthermore, presence/absence of these genes should not exceed 20% in other groups. For example, for a gene to be considered "present" in $P1^+$, it needed a similarity $\geq$ 60% in at least 7 out of 8 species included in that group. Additionally, that gene could not be present in more than 20% of the species included in $P1^-$/P2/S1/S2 (< 12 species). The same criteria apply for "absence" as well. Comparisons between several subgroups were defined: P *vs* S; P1 *vs* P2S; $P1^+$ *vs* $P1^-$P2S. Given the diversity observed throughout the *Leptospira* genus, a 60% similarity cut-off was considered prudent when assessing, for instance, genes shared among different clades, such as P2 and P1. This is the clearest case; both groups consist of pathogenic species, but P2 is mostly found in the environment, while P1 is more closely associated with hosts. Setting a higher cut-off could potentially result in overlooking certain genes associated with the emergence of pathogens. Moreover, many pathogen-specific genes previously described in the literature, were defined using this cut-off, serving as a validation control for the analyses conducted here.

Clusters of Orthologous Genes (COG) were determined using eggNOG mapper with the same parameters stated before. Protein interactions and functional gene ontology (GO) enrichment were determined by STRING v11.5 (Szklarczyk et al., 2021), using *Leptospira interrogans* serovar Lai str 56601 as the reference. A high confidence interaction score (0.700) was used when the list of proteins exceeded 100, just to capture the most relevant interactions and prevent them from being overshadowed by less robust ones. Otherwise, a medium confidence score was applied (0.400). The level of confidence is indicated each time.

3.5.4 Amino acid permutation analyses

Amino acid changes associated with different phenotypes were investigated using an *in-house* developed software, CapriB (Guerra et al., 2020). Tblastn searches (default E-value = -10) of all 68 *Leptospira* species were performed against reference protein-coding sequences of *L. interrogans* str. 56601. Comparisons between several subgroups were defined: P *vs* S; P1 *vs* P2S; $P1^{+}$ *vs* $P1^{-}$P2S; $P1^{+}$ *vs* $P1^{-}$. Conserved amino acid changes within the same group but different in the other (option *I vs D*) in protein-coding genes shared by the compared groups (60% identity cut-off) were analyzed. All permutations are supported by two scores (Grantham's distance and exchange-ability score), that measure the potential effect of the identified changes.

Mutated protein-coding sequences were analyzed in terms of COG functional categories, protein-protein interactions and GO enrichment, as described for the gene gain/loss analyses.

Mutations on specific candidates were checked by protein sequence alignment using Mafft version 7.407 (Katoh and Standley, 2013), and subsequent visualization in MView from European Molecular Biology Laboratory-European Bioinformatics Institute (EMBL-EBI, https://www.ebi.ac.uk) (Madeira et al., 2022). Functional domain screening was performed through InterPro also from EMBL-EBI.

Chapter 4. Research article on genetic signature of *Leptospira* serovars (Article 2)

Horizontal transfer of the *rfb* cluster in *Leptospira* is a genetic determinant of serovar identity

Cecilia Nieves,[1] Antony T. Vincent,[1,2] Leticia Zarantonelli[3], Mathieu Picardeau,[4,5] Frédéric J. Veyrier,[1#] and Alejandro Buschiazzo[3,5#]

[1]Bacterial Symbionts Evolution, Centre Armand-Frappier Santé Biotechnologie, Institut National de la Recherche Scientifique, Université du Québec, Laval, QC, Canada; [2]Département des sciences animales, Faculté des sciences de l'agriculture et de l'alimentation, Université Laval, 2425, rue de l'Agriculture, Quebec City, QC, Canada G1V 0A6; [3]Laboratory of Molecular & Structural Microbiology, Institut Pasteur de Montevideo, 11400, Montevideo, Uruguay; [4]Institut Pasteur, Université Paris Cité, CNRS UMR 6047, Biology of Spirochetes Unit, Paris, France; [5]Integrative Microbiology of Zoonotic Agents, Pasteur International Joint Research Unit, Paris/Montevideo

Corresponding authors: frederic.veyrier@inrs.ca, alebus@pasteur.edu.uy

Title of journal

Life Science Alliance, EMBO-Rockefeller University-Cold Spring Harbor Laboratory Press
DOI 10.26508/lsa.202201480 | Submitted 12 April 2022 | Revised 22 November 2022 | Accepted 23 November 2022 | Published online 9 December 2022

Author contributions

C. Nieves participated in the investigation, conceptualization of the study, and the design of the methodology. This author conducted the majority of bioinformatics and data analyses, generated figures, drafted the original manuscript, and contributed to additional rounds of editing.

FJV and AB, study conceptualization; ATV, and FJV, methodology design; ATV, FJV, and AB, data analyses; FJV and ATV, investigation; FJV experiments and bioinformatic analyses supervision; LZ, MP, FJV, and AB, resources acquisition; LZ, MP, FJV, and AB, writing, reviewing, and editing; FJV and AB, figures and final manuscript version with input from all co-authors; FJV, MP and AB, funding acquisition.

4.1 Abstract

Leptospira, a phylogenetically diverse bacterial genus within the *Spirochaetes* phylum, comprises numerous species with a wide range of lifestyles: from free-living to highly virulent pathogenic organisms that cause disease to a broad spectrum of hosts including humans. *Leptospira* displays marked intraspecific variation among strains, leading to a convoluted definition of serovars that crisscrosses the species classification. Associated to host adaptation and virulence, serovar identity is linked to the composition of the *O*-antigen carbohydrate portion of surface-exposed lipopolysaccharide. The latter is a highly antigenic molecule, with many of its biosynthesis-encoding genes residing within the *rfb* chromosomic gene cluster. However, the genetic basis of such intra-species variability is not fully understood, hampering mechanistic insights of pathogenicity, and imposing practical problems to diagnostics and typing methods that still rely on cumbersome serologic procedures. We now show that the gene content of the *rfb* cluster strongly correlates with *Leptospira* serovar designation. By determining whole genome sequences of twelve strains of the pathogenic species *L. noguchii*, including different serogroups, a comprehensive comparative analysis reveals that the *rfb* cluster undergoes extensive horizontal gene transfer. Extremely variable in length, the *rfb* clusters from *L. noguchii* and other *Leptospira* species exhibit a univocal correspondence between gene composition and serovar identity. This work paves the way to devising straightforward genetic typing of *Leptospira* serovars and to pinpoint specific genes within the distinct *rfb* clusters, encoding host-specific virulence traits. Further research shall unveil the molecular mechanism of *rfb* transfer among *Leptospira* strains and species.

4.2 Introduction

Leptospirosis is a bacterial disease that affects humans and animals. Despite being one of the most extended zoonoses worldwide, leptospirosis remains a neglected and underdiagnosed febrile illness. Pathogenic *Leptospira* species, the etiological agents of leptospirosis, infect a broad spectrum of hosts, with a global annual incidence of 1 million human cases and approximately 60,000 deaths (Costa et al., 2015; Torgerson et al., 2015). Leptospirosis constitutes an important case model within the “One Health” perspective (Jancloes

et al., 2014), being a zoonotic disease that spreads among symptomatic and asymptomatic hosts, with transmission strongly influenced by environmental conditions (Mwachui et al., 2015).

Spirochetes belonging to the genus *Leptospira* have traditionally been divided into three groups: pathogens, intermediates, and saprophytes (Ko et al., 2009). This classification relied on bacterial virulence, isolation from infected hosts and phylogeny. Recently, expanded and more elaborate phylogenetic analyses resulted in a comprehensive new classification scheme, beyond the species' infectious capacity. The 68 species of *Leptospira* that have been identified so far (Vincent et al., 2019; Korba et al., 2021), are thus classified in two major clades: P (pathogenic) and S (saprophytic), each of which subdivided into two sub-clades (Vincent et al., 2019).

Leptospira noguchii belongs to the first sub-clade within the P group (P1), that comprises the most important species causing human and animal disease such as *L. interrogans* and *L. borgpetersenii* (Vincent et al., 2019). Many human infections by *L. noguchii* have been recorded since 1940 (Gochenour et al., 1952; Fraser et al., 1973), but it was not until 1987 that it was recognized as a distinct species (Yasuda et al., 1987), baptized after the Japanese bacteriologist Hideyo Noguchi, himself responsible for choosing the genus' name. Despite the clinical importance of *L. noguchii* and its extended geographical distribution, it has received far less attention compared to other pathogenic *Leptospira* species. Noticeably, no finished or closed whole genome sequence (WGS) of *L. noguchii* are currently available, data that would otherwise boost the power of comparative genomics analyses.

Why could *L. noguchii* WGS contribute with novel insights into leptospirosis? Comprising the largest reported genome among *Leptospira* spp. (Fouts et al., 2016), *L. noguchii* also ranks amongst the species with more genes and predicted proteins of the entire genus, particularly compared to the closely related pathogens *L. interrogans* and *L. borgpetersenii*, both of which have been studied with much more detail. Besides from humans, *L. noguchii* has been isolated from armadillos, cattle, sheep, dogs, frogs, and opossums, among others (Silva et al., 2007; Silva et al., 2009), demonstrating its remarkably high adaptability to infecting a very broad range of hosts. Human infections by *L. noguchii* have been reported in geographic areas where the same strains were previously detected in other hosts (Silva et al., 2009; Flores et al., 2017), confirming *L. noguchii*'s capacity for zoonotic transmission. Found predominantly in the

Americas, and more rarely in Asia (Guglielmini et al., 2019), *L. noguchii* exhibits high genetic diversity among circulating strains (Martins et al., 2015; Hamond et al., 2016; Zarantonelli et al., 2018), with no apparent correlation between genotypes and hosts or geographic distribution (Loureiro et al., 2020). Particularly in South America, systematic field studies of infected animal hosts reveal a much larger diversity of *L. noguchii* serovars than that encountered for *L. interrogans* and *L. borgpetersenii* (Zarantonelli et al., 2018).

The presence of many serovars (serological variants) is a common attribute within *Leptospira* species. With >300 serovars having been reported, their classification into serogroups has been instrumental, clustering together related serovars that express overlapping antigenic determinants (Bharti et al., 2003). Probably related to variable structures of the surface-exposed lipopolysaccharide (LPS) antigen on the bacterial cell wall (Adler and de la Peña Moctezuma, 2010), different serovars trigger distinct antibody responses during infection. Interestingly, a number of known serovar-host associations have been pinpointed, leading to the concept of serovar adaptation and chronicity of infection for particular hosts, also correlating to more acute virulence when non-adapted serovars accidentally infect heterologous hosts (Ellis, 2015). Despite the relevance of this phenomenon in terms of epidemiology and clinical outcomes, the molecular mechanisms that underlie serovar determination are not fully understood (Peña-Moctezuma et al., 1999; Adler and de la Peña Moctezuma, 2010; Fouts et al., 2016). A connection between serovar determination and gene content has been proposed (Bulach et al., 2000; Bulach et al, 2006; Santos et al., 2018), but not demonstrating a direct, biunivocal link among each of the many different serovars and a defined set of genes (genetic presence/absence profiles). Such unequivocal link has also been hampered by the scarcity of precise serovar identification for most reported isolates, and the lack of finished whole genome sequencing data, particularly so for *L. noguchii* and other understudied species.

By sequencing the genomes of twelve *L. noguchii* strains (10 closed genomes, 2 drafts), we undertook an extensive comparative genomics approach, uncovering underlying reasons for *Leptospira* phenotypic complexity. We now reveal (i) the detailed genomic features and plasmid repertoire of *L. noguchii* and its phylogenetic structure; (ii) that the cluster comprising most of the LPS-synthesis enzyme encoding genes, known as *rfb* (Patra et al., 2015; Picardeau, 2017), exhibits clear signs of horizontal gene transfer (HGT) among different *Leptospira* species; and,

(iii) that serovar identity is univocally linked to the presence/absence of specific genes within this *rfb* cluster.

In sum, this work constitutes the first report of complete genomes of *L. noguchii*, which allowed a comprehensive analysis of its genetic variability. Remarkably, after comparing with known serovars of different *Leptospira* species, it was possible to reveal serovar-specific genetic fingerprints encoded within a horizontally-transferred gene cluster, paving the way towards genome-based serotyping and further molecular studies of the HGT mechanisms at play.

4.3 Results

4.3.1 *L. noguchii* whole genome sequences: general features

Whole genomes from twelve *L. noguchii* strains (**Table 4.1**) were sequenced using a long-read sequencing approach (PacBio® technology). These strains were isolated from different hosts at four distant geographic locations in Central and South America: Barbados (2 isolates from amphibian hosts), Guadeloupe island/France (1, human), Uruguay (8, cattle) and Venezuela (1, human). Exhibiting an average genome size of 4,863,036 ± 99,185 bp, all larger than other well-studied species like *L. interrogans* (~4.6 Mb) and *L. borgpetersenii* (~3.9 Mb). Most genomes reached a finished status, with 3 to 8 contigs (**Table 4.1**) corresponding to chromosomes 1 (Chr1) and 2 (Chr2), plus a variable number of plasmids. The whole genomes from strains ‘bajan’ and ‘201102933’ were the only ones not closed, albeit rendering very high-quality draft sequences (five contigs in the case of strain bajan, and six for strain 201102933, with N50s of 2,827,750 bp and 3,551,682 bp respectively).

Table 4.1. *L. noguchii* whole genome sequences.

The 'Strain' column provides information about the serogroup of each strain (AUS=Australis, AUT=Autumnalis, PYR=Pyrogenes, ND=Not determined) and its country of origin (BRB=Barbados, GLP=Guadeloupe; URY=Uruguay, VEN=Venezuela), in parentheses. The other columns provide information as described in the column headings. "TPases" stands for transposases.

	Strain	Host	Replicon	Size (bp)	CDSs	rRNA	tRNA	CRISPRs	TPases	GC (%)	Coding Ratio (%)	Accession Number
Finished genomes												
	barbudensis (AUS/BRB)	Amphibian	Chr1	4,408,823	3,883	5	37	6	101	35.5	75.9	CP091967
			Chr2	359,178	337	-	-	-	3	35.3	76.8	CP091968
			p1	47,641	65	-	-	-	0	42.2	45.2	CP091969
			p2	44,440	63	-	-	-	0	42.5	46	CP091970
			Total	***4,860,082***								
	IP1512017 (ND/URY)	Cattle	Chr1	4,297,194	3,375	5	37	8	104	35.6	73.3	CP091957
			Chr2	370,404	327	-	-	-	14	36	77.1	CP091958
			p1	178,101	139	-	-	-	7	38.3	74.3	CP091959
			p2	67,330	48	-	-	-	3	37.2	76.6	CP091960
			p3	43,769	33	-	-	-	0	35.4	73	CP091961
			Total	***4,956,798***								
	IP1605021 (PYR/URY)	Cattle	Chr1	4,438,826	3,555	5	37	10	121	35.8	73.4	CP091953
			Chr2	395,164	368	-	-	-	10	36.2	79.5	CP091954
			p1	169,819	136	-	-	-	11	38.1	75	CP091955
			p2	47,268	29	-	-	-	3	36.4	67.3	CP091956
			Total	***5,051,077***								
	IP1611024 (AUS/URY)	Cattle	Chr1	4,248,034	3,329	5	37	10	83	35.6	74	CP091947
			Chr2	351,624	291	-	-	-	10	35.6	76.6	CP091948
			p1	89,758	93	-	-	-	0	32.4	83.1	CP091949
			p2	74,081	60	-	-	-	5	34.6	67	CP091950
			p3	57,447	50	-	-	-	3	34.3	61.2	CP091951

		p4	41,294	30	-	-	-	1	32.4	71.2	CP091952
		Total	***4,862,238***								
IP1703027 (ND/URY)	Cattle	Chr1	4,329,282	3,408	5	37	12	94	35.7	73.7	CP091943
		Chr2	338,457	275	0	0	1	3	35.7	76.4	CP091944
		p1	83,272	52	-	-	-	1	35.9	71.2	CP091945
		p2	61,325	44	-	-	-	3	35.7	73.1	CP091946
		Total	***4,812,336***								
IP1705032 (AUT/URY)	Cattle	Chr1	4,342,451	3,435	5	37	10	98	35.7	74	CP091940
		Chr2	343,634	286	-	-	-	4	35.6	77.3	CP091941
		p1	58,982	49	-	-	-	0	35.9	66.1	CP091942
		Total	***4,745,067***								
IP1709037 (AUT/URY)	Cattle	Chr1	4,199,394	3,354	5	37	8	81	35.7	74.1	CP091936
		Chr2	376,569	326	-	1	-	3	36.1	76.8	CP091937
		p1	90,518	74	-	-	-	6	35.3	73.5	CP091938
		p2	74,070	61	-	-	-	5	34.5	65.7	CP091939
		Total	***4,740,551***								
IP1712055 (ND/URY)	Cattle	Chr1	4,329,115	3,455	5	37	12	93	35.7	73.6	CP091928
		Chr2	338,447	277	-	-	1	3	35.7	76.2	CP091929
		p1	83,272	52	-	-	-	1	35.9	71.2	CP091930
		p2	61,322	44	-	-	-	3	35.7	71.5	CP091931
		p3	35,952	34	-	-	-	0	41.3	53.1	CP091932
		p4	26,771	7	-	-	-	0	41.1	5.7	CP091933
		p5	24,253	38	-	-	-	0	42	71.5	CP091934
		p6	13,148	11	-	-	-	0	40.3	36.8	CP091935
		Total	***4,912,280***								
IP1804061 (ND/URY)	Cattle	Chr1	4,322,667	3,447	5	37	9	93	35.7	73.4	CP092112
		Chr2	330,637	280	-	-	-	16	35.8	76	CP092113
		p1	127,018	100	-	-	-	5	37.3	70.8	CP092114
		p2	100,275	71	-	-	-	0	36.3	74.1	CP092115

			p3	51,374	37	-	-	-	2	35.2	72.1	CP092116
			Total	***4,931,971***								
	201601331 (ND/VEN)	Human	Chr1	4,334,850	3,447	5	37	8	134	35.7	74	CP091962
			Chr2	350,933	293	-	-	-	5	35.7	76.8	CP091963
			p1	65,245	46	-	-	-	1	35.9	65.2	CP091964
			p2	38,153	33	-	-	-	2	33.4	57	CP091965
			p3	28,484	26	-	-	-	1	35.8	34.1	CP091966
			Total	***4,817,665***								
Draft genomes												
	bajan (AUS/BRB)	Amphibian	Five (5) contigs	4,850,434	4,326	5	37	6	104	35.7	76.4	JAKNBP000000000
	201102933 (AUS/GLP)	Human	Six (6) contigs	4,697,964	3,869	5	38	6	103	35.5	76.9	JAKNBO000000000

The average nucleotide identity (ANI) among all sequenced strains in this study (using *L. noguchii* sv Panama strain CZ214 as reference) was >95% (**Table S4.1** and **Figure S4.1, Appendix II**), consistent with previous determinations based on 16S rDNA sequence (Zarantonelli et al., 2018). Similar identity figures with all reported draft WGSs from *L. noguchii* strains further confirm the taxonomic determination, as well as ANIs ≤90% with respect to closely related species such as *L. interrogans* and *L. kirschneri*. However, the percentage of conserved proteins (POCP) among the twelve sequenced strains did not exceed 98.1% (**Figure S4.2**, **Appendix II**), uncovering a significant intra-species phenotypic complexity. Especially diverse in terms of protein repertoire is the cluster comprising the three Caribbean strains (bajan, barbudensis and 2011029331), which consistently exhibit highest ANI figures amongst them. Regarding replicon content, *L. noguchii* displays the same genetic organization as reported for other *Leptospira* species (Picardeau et al., 2008), with two chromosomes and a variable number of plasmids (**Table 4.1**). In consistency with their larger genome size, *L. noguchii* also exhibits more CDSs (~4000 on average) compared to other *Leptospira* species (Picardeau et al., 2008), with other features such as number of tRNA genes and GC content being similar. The number of predicted Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR) sequences was more variable, as were those of transposases and IS transposase-like CDSs, ranging from 95 to 145 (**Table 4.1**), in any case much more numerous than in *L. interrogans* (26 in sv Copenhageni strain Fiocruz) or in the saprophyte *L. biflexa* (8 in sv Patoc strain Ames) (Picardeau et al., 2008).

Analysis of the *L. noguchii* pangenome of the 12 strains sequenced in this study (**Figure S4.3**, **Appendix II**) showed an open profile (Heap's law parameter $\alpha = 0.36$ (Tettelin et al., 2008)), confirming the high genetic variability among *L. noguchii* strains. Out of the 7963 genes that constitute the pangenome, only 2671 were found in almost all the strains thus comprising the core genome. Indeed, the cloud genome was defined by 2183 genes (accessory genes), uncovering a rich array of unique attributes that distinguish strains.

To further explore the properties of such strain-specific genes, likely underlying phenotypic variability, profile Hidden Markov Models (HMM) were calculated for each one of the genes present in only one of the strains and absent in all others. These HMM profiles were then mapped (Aramaki et al., 2020) onto the Kyoto Encyclopedia of Genes and Genomes (KEGG) database to investigate whether these variant-specific genes are enriched in particular

biochemical functions, or instead randomly distributed (**Table S4.2**). Among the four top-ranking pathways –(i) metabolic pathways, (ii) biosynthesis of nucleotide sugars, (iii) amino sugar and nucleotide sugar metabolism, and (iv) *O*-antigen nucleotide sugar biosynthesis– a clear enrichment is observed in functions related to carbohydrate metabolism, and glycoside modification and synthesis. This KEGG-mapping analysis was systematically extended to accessory genes present only in two strains, three, and further, confirming the importance of variations in carbohydrate-related metabolism as a main source of strain-specific genetic variability. Of note, a number of carbohydrate-related genes, including several that encode LPS-biosynthesis enzymes, were present only in particular groups of strains. Examples such as UDP-glucuronate 4-epimerase (GalE (Bulach et al., 2000), or 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (KdsC (Biswas et al., 2009; Valvano, 2015), among others, cluster according to serogroup identity.

4.3.2 Plasmid repertoire in *L. noguchii*

Besides the two chromosomes, *L. noguchii* strains harbor a variable number of plasmids (**Tables 4.1** and **S4.3**, first sheet), ranging from only one, to as much as six replicons. The plasmid repertoire is unique to each strain. A network association analysis was performed to compare the plasmid-encoded proteins in different *L. noguchii* strains (**Figure 4.1A**). Some plasmids showed identical or nearly identical presence/absence patterns of protein-encoding genes, suggesting that plasmids may be transferred among strains. For example, the two plasmids of strain IP1703027, bear identical gene composition compared to two of the plasmids in IP1712055 (p1 and p2); plasmids p1 from strains IP1512017 and IP1605021, and plasmids p3 from IP1512017 and IP1804061, share many genes.

Extending the analysis of plasmid repertoires to other strains and *Leptospira* species (**Table S4.3**, second sheet) revealed no core- or even softcore-genes, highlighting the extreme plasmid diversity in *Leptospira*. A network association analysis considering this extended set of plasmids, revealed species-specific clustering of plasmid sequences, again uncovering cases of strong similarity/identity in the arrays of protein-encoding genes between different strains (**Figure 4.1B**). For instance, plasmids p1/p2, along with p2/p3 from distinct *L. interrogans* strains D64 and 1548, exhibit nearly identical profiles. The same was observed for plasmids p1

and p2 from *L. interrogans* strains 611, Gui44 and LJ178; as well as comparing plasmids pD13 and pDO6 from *L. weilii*. Overall, plasmids from *L. noguchii* clustered together and did not show similar patterns with those from other species, except just in one case. Plasmid p1 from strain IP1611024, shared a significant number of protein-encoding genes with plasmids p1 from *L. mayottensis* str 200901116, pLmayMDI222 from *L. mayottensis* str MDI222, and p4 from *L. interrogans* str 1489 (**Figure 4.1B**, red square).

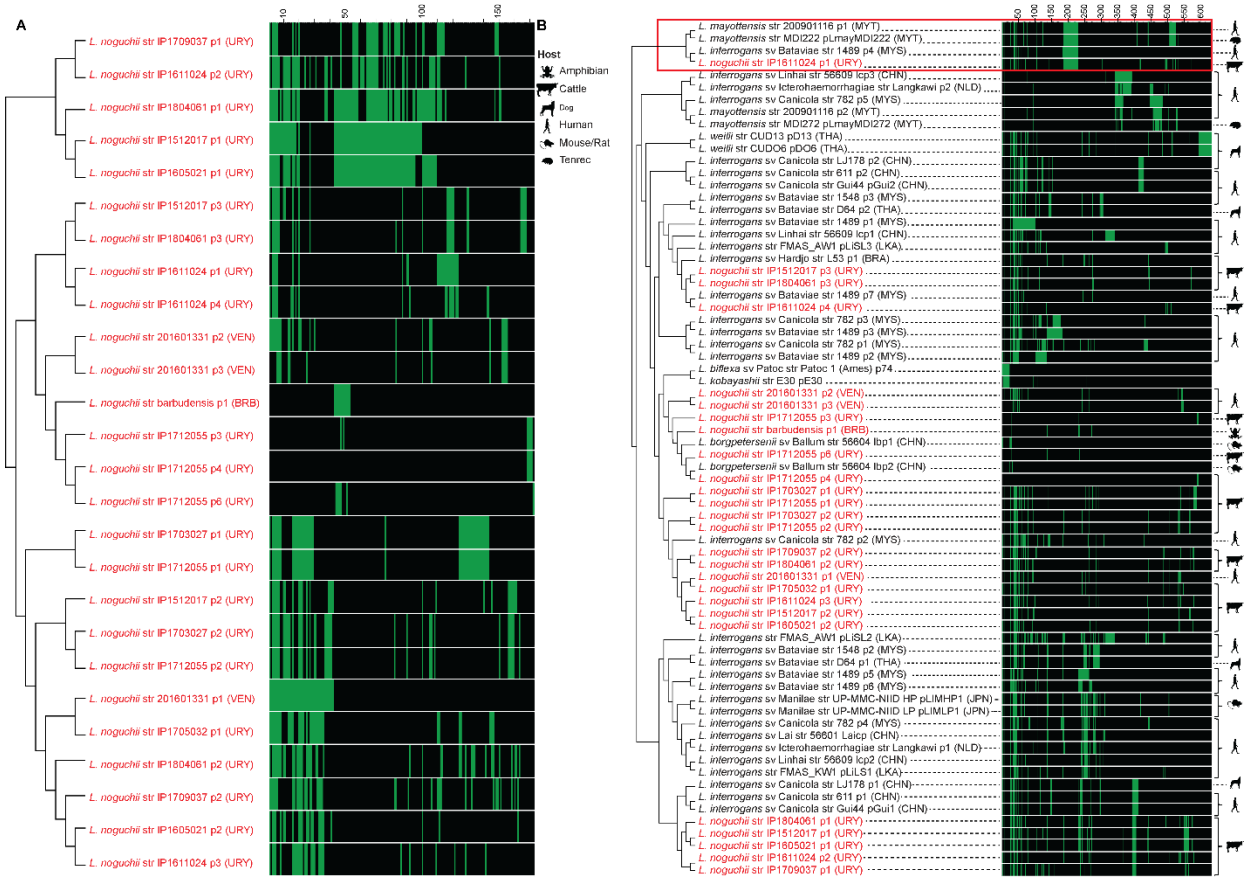

Figure 4.1. Conservation of protein-encoding genes across *Leptospira* plasmids.

Sequence network association analysis by hierarchical clustering, based on presence/absence of plasmidic protein-encoding genes. Matrices on the right of each clustering depict individual genes with vertical lines, green meaning that the gene is present in that strain (similarity cut-off 60%), and black meaning absence. Scales on the top of the matrices indicate the number of different genes being compared. **(A)** Plasmids from *L. noguchii* strains sequenced in this study. Plasmid names are indicated right after the strain designation, and country of origin in parentheses (BRB=Barbados; URY=Uruguay; VEN=Venezuela). **(B)** Plasmids from different *Leptospira* species. Plasmid names as in (A), and country of origin in parentheses (BRA=Brazil; BRB=Barbados; CHN=China; JPN=Japan; LKA=Sri Lanka; MYS=Malaysia; MYT=Mayotte; NLD=Netherlands; THA=Thailand; URY=Uruguay; VEN=Venezuela). Strains sequenced in this study are highlighted in red. The hosts from which they were isolated are indicated with cartoons on the right side of the matrix. The red square encloses plasmids from different *Leptospira* species sharing a large number of protein-encoding genes.

A functional/biological analysis of *L. noguchii* plasmid-encoded genes is difficult, as most of the constituent proteins are hypothetical (**Table S4.3**, third sheet). A first general inquiry about the potential link of plasmid identity and environmental factors, did not result in a clear association. Neither the geographical locations of strains, nor the infected host from which they were isolated, showed clear-cut connections with plasmids and their protein-encoding gene compositions. Plasmid-borne virulence factors and antibiotic resistance determinants were also explored, recognizing two genes encoding putative multidrug efflux proteins of the Resistance-nodulation-division (RND) family (Nishino et al., 2007), MtdA and MtdB, in plasmids p1 (from strains IP1512017, IP1605021, IP1709037 and IP1804061), and p2 (from IP1611024). The limited identity with *bona fide* antibiotic resistance proteins did not allow for conclusive antibiotic specificity and/or functional prediction. These genes were always found as a cluster in *Leptospira* plasmids, with *mtdA* followed by *mtdB* and *cusA*, the latter encoding a cation efflux pump. The extended network association analysis including also other *Leptospira* species (**Table S4.3**, fourth sheet) revealed this same cluster in plasmid p1 from three *L. interrogans* strains (611, Gui44 and LJ178). Further work is needed to uncover the biological role of these proteins in *Leptospira*, especially considering that antibiotic resistance is not a usual feature in Spirochetes. Functional analysis of clusters of orthologous genes (COG) showed a few categories to be absent from plasmidic genes in *L. noguchii* and other *Leptospira* species: (i) RNA processing and modification; (ii) chromatin structure and dynamics; (iii) carbohydrate metabolism and transport; (iv) nuclear structure; (v) cytoskeleton; and (vi) general function prediction. On the other hand, among the most represented COG categories were those related to (i) unknown function, (ii) replication and repair, (iii) transcription, and (iv) signal transduction. Surprisingly, the functional category linked to amino acid metabolism and transport was completely absent in *L. noguchii* plasmids, in stark contrast to other *Leptospira* species.

4.3.3 *Leptospira noguchii* phylogeny

The complexity and high diversity of this species' pangenome could also be related to the adaptation of *L. noguchii* to different hosts and geographic locations. To uncover such potential genotype/phenotype associations, the phylogenetic structure of *L. noguchii* was

explored in greater detail by analyzing the genomes from eleven different geographic locations and nine types of hosts (**Figure 4.2** and **Table S4.4**) applying a maximum-likelihood approach.

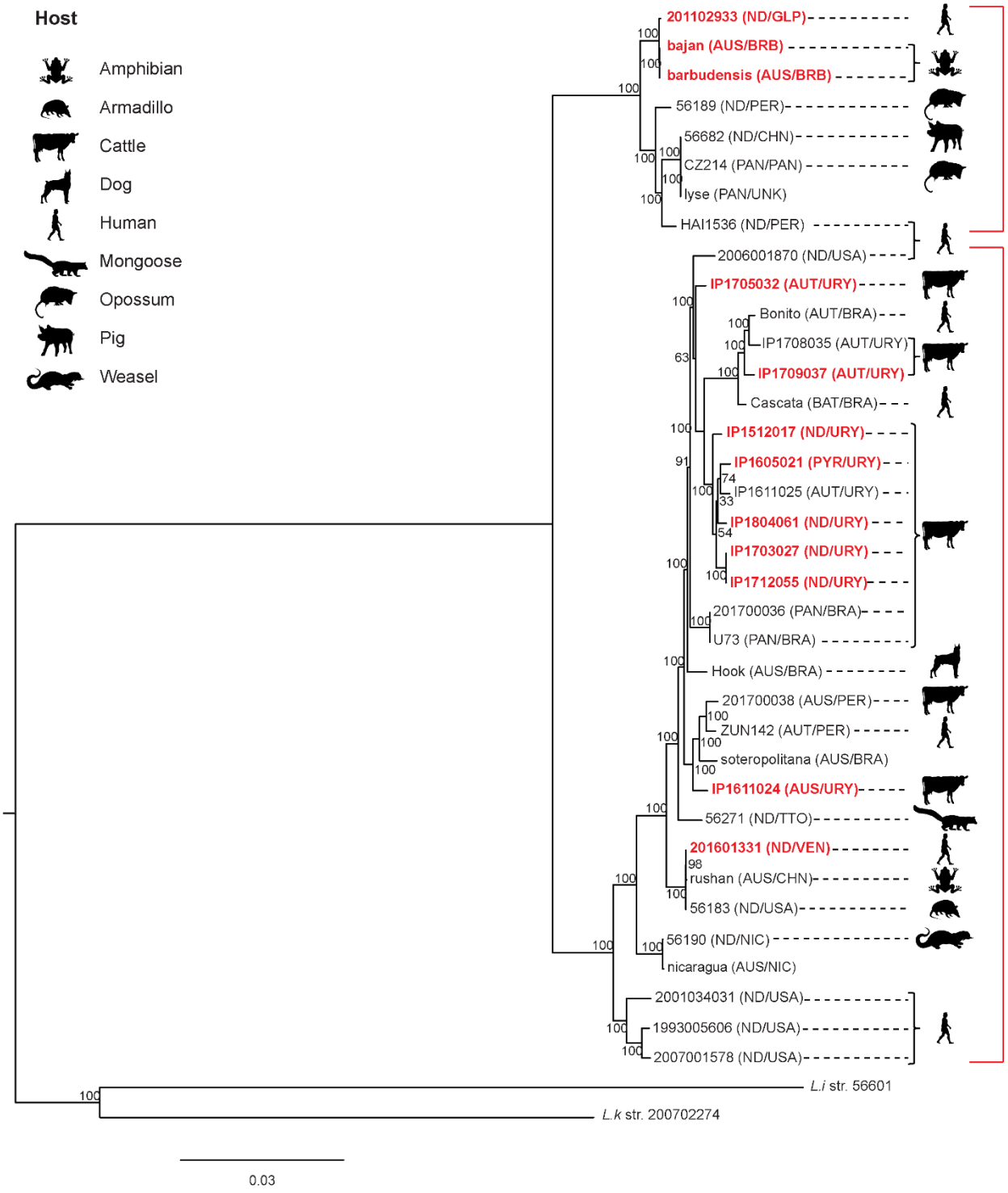

Figure 4.2. Phylogenetic tree of *Leptospira noguchii*.

The maximum-likelihood phylogenetic tree is based on the softcore genes (present in more than 95% of the genomes). *L. interrogans* str 56601 and *L. kirschneri* str 200702274 were included as outgroups. The serogroup of each strain (AUS=Australis; AUT=Autumnalis; BAT=Bataviae; ND=not determined; PAN=Panama; PYR=Pyrogenes) and its country of origin (BRB=Barbados; BRA=Brazil; CHN=China; GLP=Guadeloupe; NIC=Nicaragua; PAN=Panama; PER=Peru; TTO=Trinidad & Tobago; USA=United States of America; URY=Uruguay; UNK=unknown; VEN=Venezuela) are indicated in parentheses. Strains sequenced in this study are outlined in bold red. The hosts from which they were isolated are indicated with cartoons (some hosts were not specified in the original reports). Two clades can be distinguished, highlighted with red brackets.

Two clades could be distinguished, which do not correlate to geographic distribution nor to host. Further studies to increase the number of strains shall enable a more conclusive statement. The Uruguayan strains isolated from cattle cluster within one of the clades, but they are not phylogenetically distant from other host species including humans, other mammals and even amphibians. Focusing the analysis on *L. noguchii* strains isolated from human infections, the distribution is once again extremely broad throughout the phylogenetic tree, indicative of transmission among different reservoirs (**Fig 4.2**).

4.3.4 Genetic variability of the *rfb* cluster in *L. noguchii*

Considering that serological variability is a particularly relevant phenotypic trait in *Leptospira* (Adler and de la Peña Moctezuma, 2010), that the *L. noguchii* phylogenetic structure did not reveal clear genotype/phenotype associations including host tropism, and that *L. noguchii* strain-specific accessory genes were found to be highly enriched in carbohydrate pathways and LPS biosynthesis, a more detailed analysis focusing on particular genomic regions was done. Genes coding for the cell wall LPS biosynthesis, have been linked to serovariation, and in *Leptospira* tend to concentrate within a gene cluster known as *rfb* (Peña-Moctezuma et al., 1999; Fouts et al., 2016). Access to complete, finished whole genome sequences, is particularly relevant to analyze delimited loci, avoiding inaccurate gene composition descriptions that result from fuzzy boundaries and/or incompleteness (Denton et al., 2014). Exploiting the WGS of *L. noguchii* strains that we are now reporting, a reliable evaluation of gene diversity related to LPS biosynthesis is feasible.

The core moiety of LPS known as lipid A (Raetz and Whitfield 2002; Que-Gewirth et al. 2004), is synthesized by several enzymes encoded in a cluster of 13 genes: *lpxA, lpxC, lpxD1, lpxD2, lpxB1, lpxB2, lpxK, kdtA, kdsB1, kdsB2, lnt, kdsA* and *htrB*. The composition of this gene

cluster was almost identical in all the strains analyzed, including genomes reported in this work as well as those from other *Leptospira* strains of known serovar identity (**Table S4.5**, first sheet). This result confirms previous reports analyzing several different pathogenic *Leptospira* species (Fouts et al. 2016).

A second component of LPS is the core oligosaccharide (Raetz & Whitfield, 2002), whose biosynthesis starts by addition of 3-deoxy-D-*manno*-oct-2-ulosonic acid (Kdo) molecules to the lipid A moiety, subsequently incorporating heptoses and further modifications (Bertani and Ruiz, 2018). Comparison of the genes coding for core synthesis enzymes (such as WaaA that adds Kdo molecules; RfaC, -D, -E, and -F that attach ADP-L-*glycero*-β-D-manno-heptose intermediates; other glycosyltransferases like RfaG, which append glucose units to the heptoses; etc) among the different strains (**Table S4.5**, first sheet) revealed no variation in terms of differential presence of genes. The only difference concerned *L. interrogans* sv Weerasinghe, in which three genes coding for WaaA isoforms were found (~99% identical among them).

Finally, the outermost section of the LPS known as the *O*-antigen, is the most variable part of the structure and most exposed to interact with the environment (Raetz and Whitfield, 2002). The *O*-antigen is an oligosaccharide comprising a fairly large number of diverse monosaccharides, synthesized and oligomerized together by the action of several enzymes, most of which are encoded in the *rfb* cluster in *Leptospira* (Mitchison et al., 1997). The genetic composition of the *rfb* clusters of the 12 *L. noguchii* genomes revealed a striking variability (**Table 4.2**). High *rfb* variability had previously been described comparing different *Leptospira* species (Fouts et al., 2016), hereby confirmed within a single species.

Table 4.2. Overall composition of *rfb* clusters in *L. noguchii*.

The 'Strain' column informs, in parentheses, about the serogroup of each strain (AUS=Australis, AUT=Autumnalis, PYR=Pyrogenes, ND=Not determined) and its country of origin (BRB=Barbados, GLP=Guadeloupe; URY=Uruguay, VEN=Venezuela).

Strain	Size (bp)	CDSs	GC content (%)
IP1512017 (ND/URY)	84,097	74	33.7
IP1605021 (PYR/URY)	113,593	110	34.0
IP1611024 (AUS/URY)	106,137	97	32.8
IP1703027 (ND/URY)	82,932	74	33.7
IP1705032 (AUT/URY)	93,324	83	31.9
IP1709037 (AUT/URY)	96,628	90	32.4
IP1712055 (ND/URY)	82,927	77	33.7
IP1804061 (ND/URY)	82,821	74	33.7
201102933 (ND/GLP)	103,442	100	33.8
201601331 (ND/VEN)	99,122	89	32.2
bajan (AUS/BRB)	104,988	105	33.8
barbudensis (AUS/BRB)	103,476	103	33.8

Further insights were obtained by aligning the *rfb* clusters of these twelve *L. noguchii* strains, highlighting the synteny among their constituent genes (**Figure 4.3A**).

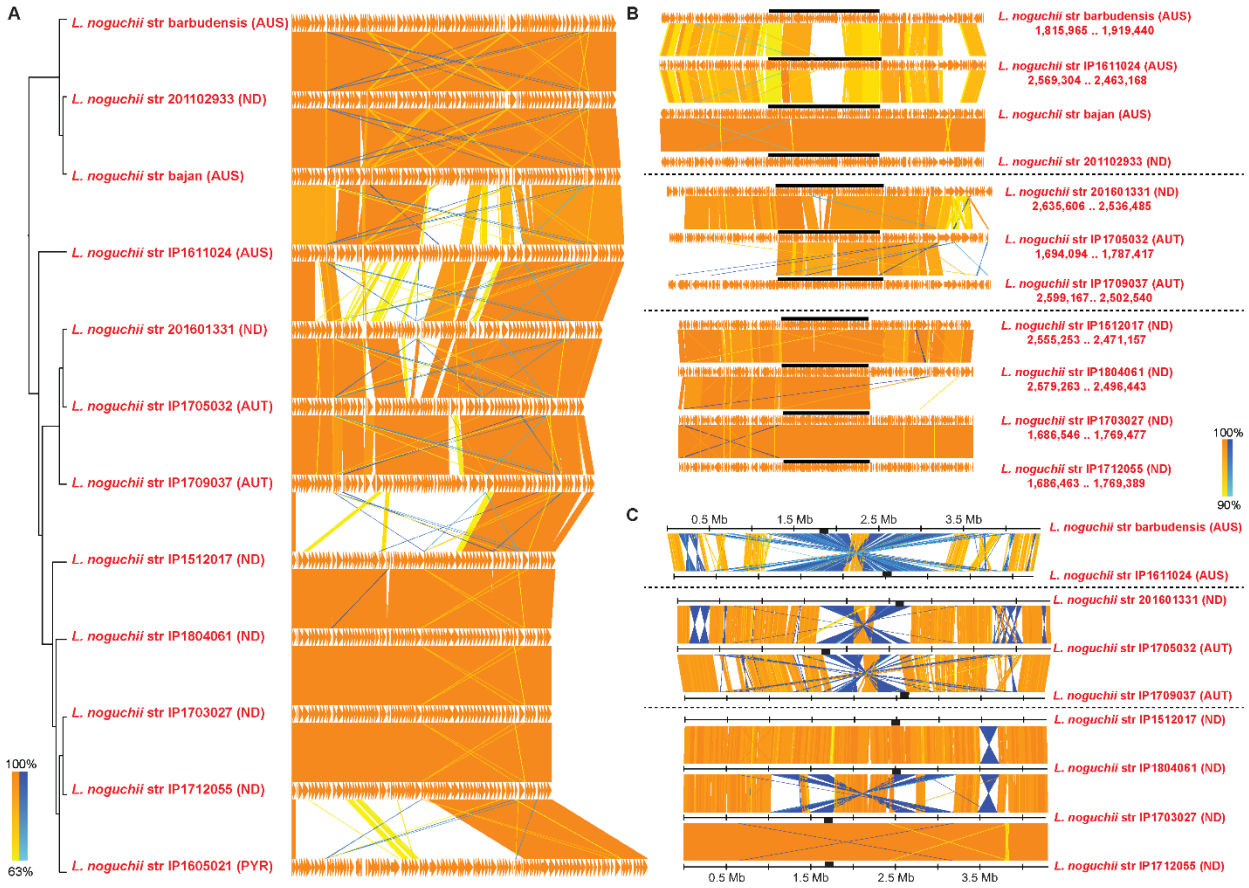

Figure 4.3. Comparison of *rfb* clusters in *L. noguchii* shows highly variable gene composition.

(A) softcore-*rfb* genes from *L. noguchii* strains aligned and clustered considering a 60% identity cut-off and gene presence in at least 60% of the *rfb* clusters analyzed (left). Serogroup identity is indicated in parentheses (AUS=Australis; AUT=Autumnalis; PYR=Pyrogenes; ND=Not determined). To the right, *rfb* clusters gene content and shared synteny are depicted. Homologous regions are linked with orange lines (same orientation) and blue lines (inverted regions), from lighter to darker colors according to identity level (as marked by the lower left scale index). **(B)** Close-up of *rfb* clusters (highlighted as black bars) from *L. noguchii* strains that belong to the same serogroup, or with highly similar *rfb* clusters, including 100,000 extra bp flanking at each side. The genomic coordinates of *rfb*-delimiting genes *marR* and *sdcS* (in base pairs, with *dnaA* at position 0) are indicated below the strains' names for those strains with closed/finished genomes and their delimitations within the alignments are highlighted as black bars. Nucleotide alignment was performed considering a 90% identity cut-off. Homologous regions are linked with orange lines (same orientation) and blue lines (inverted regions), from lighter to darker colors according to identity level (as marked by the lower right scale index). **(C)** Nucleotide alignments (cut-off 90% identity) of entire chromosome 1 from *L. noguchii* strains that possess highly similar *rfb* clusters (barbudensis *vs* IP1611024; 201601331 *vs* IP1705032/IP1709037; and, IP1512017 *vs* IP1804061/IP1703027/IP1712055). Color references as in (B). The *rfb* clusters are highlighted as black bars.

Delimiting the boundary ends of the *rfb* cluster, genes coding for a transcriptional regulator (MarR) and a sodium/sulfate symporter (SdcS) were consistently found, as in other pathogenic *Leptospira* species (Fouts et al., 2016). The 3' end is more conserved, toward this

end a gene sub-cluster is located, composed by *rfbC, rfbD, rfbB* and *rfbA*, which encode enzymes involved in the dTDP-rhamnose biosynthesis, implicated in LPS assembly in pathogenic strains (Mitchison et al., 1997). The order of appearance of these four genes was conserved in all strains. Only strain IP1605021 (serogroup Pyrogenes) presented an extra copy of *rfbC* within the *rfb* cluster but separated from the dTDP-rhamnose biosynthesis sub-cluster. Systematically, the *rfb* cluster was found in preferential locations within chromosome 1, approximately at ~1.75 Mb and ~2.50 Mb from the origin (**Figure 4.3B**) and run in opposite senses comparing locations 1 *vs* 2. The regions that flank the *rfb* clusters are conserved (**Figure 4.3B**), although this depends on the location site. An interesting example of this is illustrated in cases where the same *rfb* cluster is identified in one genomic site or the other in different strains (*e.g.* strains IP1705032 *vs* IP1709037, or IP1804061 *vs* IP1703027, see **Figure 4.3B**). A detailed examination of both *rfb* locations revealed an explanation to this feature: genomic inversions are coincident with the *rfb* clusters being located at one site or the other (**Figure 4.3C**). Although larger in some genomes, such inversions do not implicate the entire genomic range between ~1.75 Mb and ~2.50 Mb, wherein colinear regions are also observed. Interestingly, in some of the strains the *rfb* cluster is located precisely at the boundary where the inversion occurs, thereby explaining why on those cases there is only one conserved *rfb* flanking region (**Figure 4.3B**). Of note, insertion sequence (IS) transposase-like CDSs were found at, or near the boundaries of the *rfb* cluster (**Figure S4.4**, **Appendix II** and **Table S4.5** fourth and fifth sheets). Genomic rearrangements involving inversions have been reported in other *Leptospira* species (Nascimento et al., 2004; Ndela et al., 2021), some of which are indeed IS-mediated (Nascimento et al., 2004).

4.3.5 The *rfb* cluster shows hallmarks of horizontal gene transfer

Signs of horizontal gene transfer (HGT) were readily identified when analyzing the *rfb* cluster of genes in *L. noguchii*, as well as in other *Leptospira* species. A significant decrease in the GC content was systematically observed at the cluster position in all *L. noguchii* genomes reported in this study (**Figure 4.4A**). Moreover, the deviated GC content that identifies *rfb* clusters as islands within *L. noguchii* chromosome 1, was further confirmed extending these analyses to ten additional *Leptospira* species for which complete whole genome sequences are

available (**Figure S4.5**, **Appendix II**). A conspicuously low GC-content region corresponds with the position of the *rfb* cluster in all cases, in several of the species being the only such deviated segment, while additional ones are present in other cases as well. *L. interrogans* is the species that displays the least pronounced decrease, although the deviation is still evident. In *L. biflexa* a second nearby segment exhibits a noticeable GC-content decrease other than the *rfb* cluster itself (**Figure S4.5**, **Appendix II**).

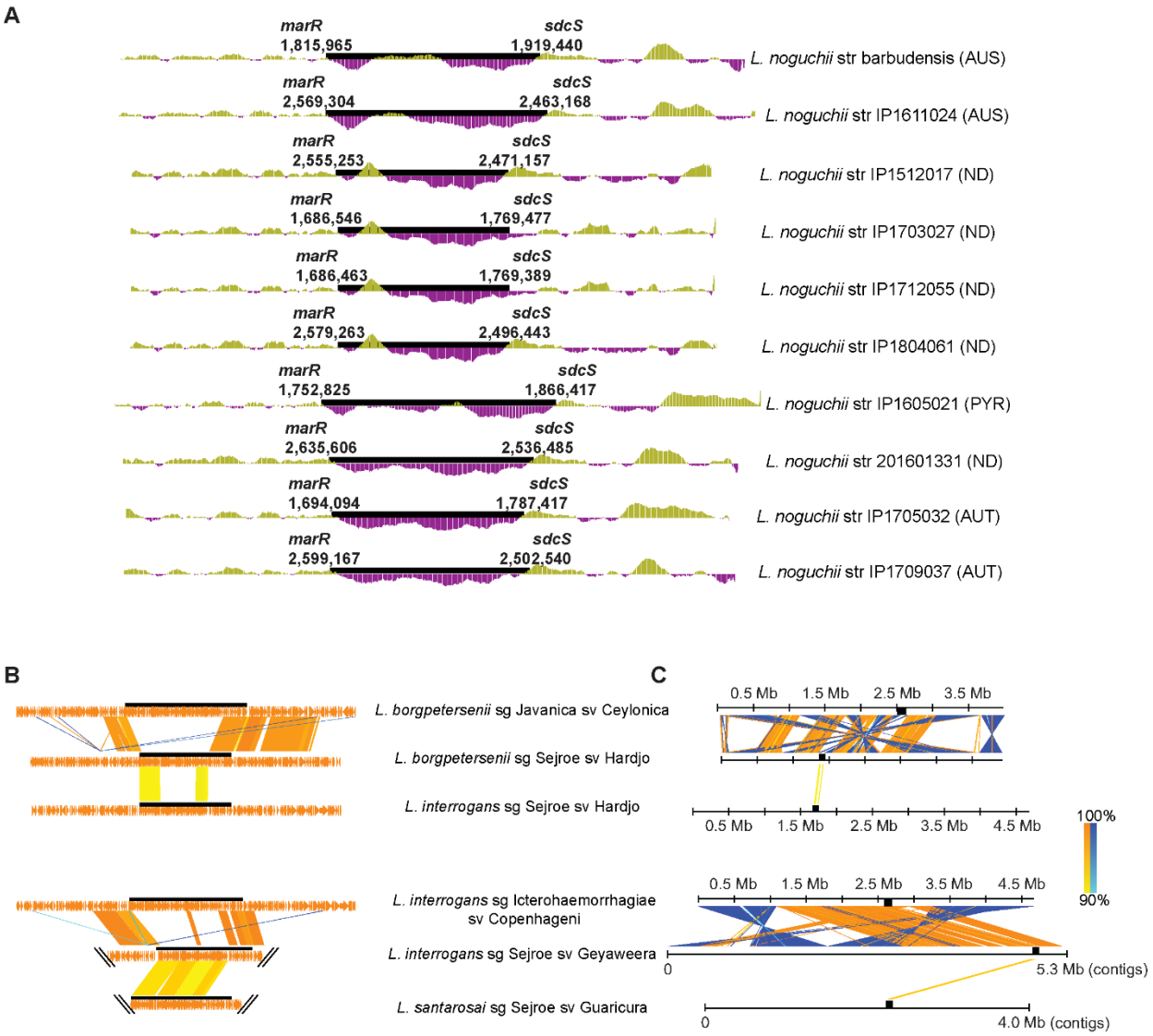

Figure 4.4. Genomic details of the *rfb* cluster in *Leptospira* reveal hallmarks of HGT.

(A) Close-up of each of the *rfb* clusters (shown as black bars), including 100,000 extra bp flanking at each side. GC % (calculated every 1,000 bp) is plotted, with purple or green indicating, respectively, a reduced or increased percentage compared to the average found in the whole chromosome 1. Relative position of genes *marR* and *sdcS* delimiting the *rfb* cluster, are indicated in base pairs

(*dnaA* at the origin). **(B)** Nucleotide alignment (90% identity cut-off) of the *rfb* clusters (black bars) and 100,000 extra base pairs flanking at each side, comparing more distant species all belonging to different serovars within serogroup Sejroe (*L. borgpetersenii* sv Hardjo str L550, *L. interrogans* sv Hardjo str Hardjoprajitno, *L. interrogans* sv Geyaweera str 1L-int, and *L. santarosai* sv Guaricura str M4/98). Unrelated serovars (*L. borgpetersenii* sv Ceylonica str Piyasena and *L. interrogans* sv Copenhageni str Fiocruz L1-130) were also added to compare species-specific conservation, outside of the *rfb* cluster. Homologous genes are linked with orange lines (if they share the same orientation) and blue lines (for inverted orientations), from lighter to darker colors according to identity percentage as marked by the right scale index. Of note, the genomes from *L. interrogans* sv Geyaweera and *L. santarosai* sv Guaricura are draft. In these cases, contig boundaries are indicated with parallel black slashes, as the analysis could not be reliably extended beyond those limits. **(C)** Nucleotide alignment (90% identity cut-off) over the entire chromosome 1 (depicted as straight black lines), comparing the same set of species as in (B). The position of *rfb* clusters is indicated with black square blocks. Reference colors as in (B).

A second feature pointing to HGT of *rfb*, is the decreased sequence conservation of flanking regions. This is more difficult to observe when only *L. noguchii* genomes are compared (**Figure 4.3B** and **C**), because of the overall higher conservation within a species. However, the *rfb*-flanking regions' variability becomes evident when comparing different species of the same serovar and serogroup. In these cases, high percentage of identity (>90%) is only observed within the *rfb* cluster, and not in the immediate surrounding segments (**Figure 4.4B**), nor in other parts of the whole genome (**Figure 4.4C**), a clear sign that a large extension of the *rfb* cluster is being horizontally transferred.

Certain genomic rearrangements, including inversions, can also be related to HGT (Oliveira et al., 2017), further highlighting the relevance of the above-mentioned inversions observed in *L. noguchii* strains (**Figure 4.3C**). And lastly, even though the positions of the *rfb* clusters have a slight variation among chromosomes 1 from different *L. noguchii* strains and *Leptospira* species (**Figure 4.3C** and **4.4C**), they do locate at restricted positions, both when they are on the sense strand at position 1, or on the antisense at position 2. Considering all the evidence together, our data strongly suggest that genes within the *rfb* cluster are horizontally transferred among different strains and species of *Leptospira*.

4.3.6 *Leptospira* strains with identical/similar serologic identity, display identical/similar *rfb* cluster gene composition

Considering that the *rfb* cluster uncovers clear signs of horizontal transfer, and comprises a great genetic variability, we then wished to explore whether the specific *rfb* cluster present in a given strain, serves as a genetic signature underlying serovar identity. A number of observations support such hypothesis. Four of the Uruguayan strains for which the serogroup identity could not be assigned –due to undetectable agglutination by standard serogroup-specific antisera panels (Zarantonelli et al., 2018)–, presented remarkably similar *rfb* clusters, inviting to posit that they may belong to the same serogroup/serovar (**Figure 4.3**). On the other hand, strains IP1705032 and IP1709037, both belonging to serogroup Autumnalis, were indeed grouped together according to their *rfb* gene composition (**Figure 4.3A**). It must be stressed however, that the Uruguayan strains have not been assigned to specific serovars yet (Zarantonelli et al., 2018). We thus extended this analysis using genomic data from a wide diversity of strains of different *Leptospira* species with known serovar identities, indeed confirming that the serovar/*rfb* identity link holds (**Figure 4.5A**). By constructing matrices where presence/absence of *rfb* genes are crossed with different strains, serovar-specific patterns or signatures were unambiguously uncovered (**Figure 4.5**).

Very few genes were conserved in all the *rfb* clusters from different serovars, consistent with the previous observations (**Table S4.6**). A detailed analysis by serogroup unveiled several important observations: (i) a few serovars belonging to the same serogroup showed indistinguishable patterns (*e.g.* Bratislava *vs* Lora, Ceylonica *vs* Javanica, Copenhageni *vs* Icterohaemorrhagiae and Lai *vs* Naam); extreme relatedness had been previously reported for Copenhageni/Icterohaemorrhagiae (Santos et al., 2018), where only one indel frameshift in a single LPS-biosynthesis gene explains their differentiation; (ii) differential profiles were evident within most serogroups, featuring genes that may discriminate serovars; (iii) *L. noguchii* strains did not show similar patterns to other species with well-typed serovars, not even to those belonging to shared serogroups, suggesting that these *L. noguchii* strains may represent novel serovars; and, (iv) the 4 non-agglutinating *L. noguchii* strains showing an identical *rfb* composition (except for slight differences in IP1512017), did not share many genes with other serogroups, suggesting they perhaps belong to a new serogroup altogether.

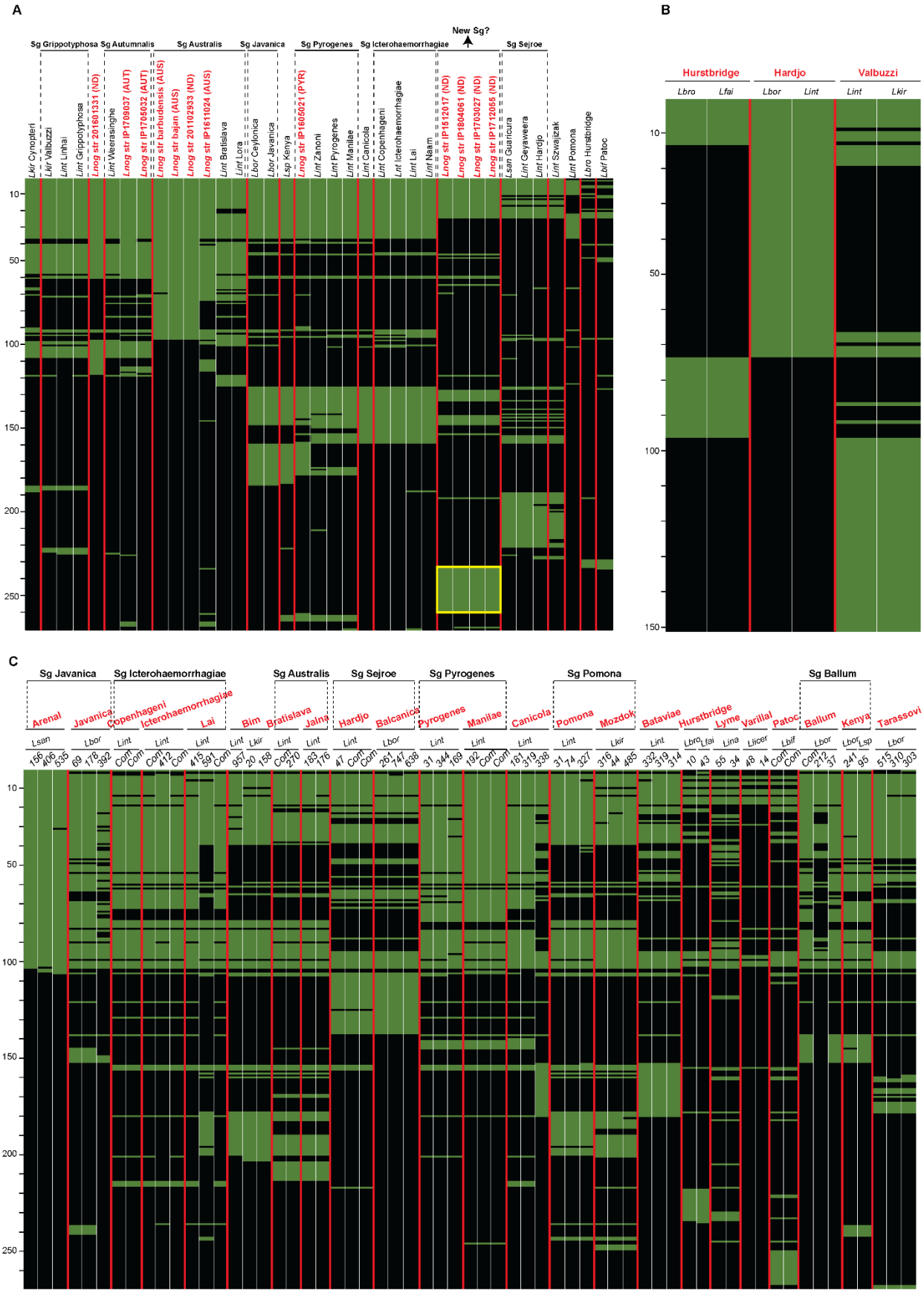
A
Sg Grippotyphosa
Sg Autumnalis
Sg Australis
Sg Javanica
Sg Pyrogenes
Sg Icterohaemorrhagiae
New Sg?
Sg Sejroe
Lkir Cynopteri
Lkir Valbuzzi
Lint Linhai
Lint Grippotyphosa
Lnog str 201601331 (ND)
Lint Weerasinghe
Lnog str IP1708037 (AUT)
Lnog stt. IP1705032 (AUT)
Lnog str barbudensis (AUS)
Lnog str bajan (AUS)
Lnog str 201102933 (ND)
Lnog str IP1611024 (AUS)
Lint Bratislava
Lint Lora
Lbor Ceylonica
Lbor Javanica
Lsp Kenya
Lnog str IP1605021 (PYR)
Lint Zanoni
Lint Pyrogenes
Lint Manilae
Lint Canicola
Lint Copenhageni
Lnt Icterohaemorrhagiae
Lint Lai
Lint Naam
Lnog stt IP1512017 (ND)
Lnog str IP1804061 (ND)
Lnog str IP1703027 (ND)
Lnog str IP1712055 (ND)
Lsan Guaricura
Lint Geyaweera
Lint Hardjo
Lint Szwajizak
Lint Pomona
Lbro Hurstbridge
Lbif Patoc
10
50
100
150
200
250
B
Hurstbridge
Hardjo
Valbuzzi
Lbro
Lfai
Lbor
Lint
Lint
Lkir
10
50
100
150
C
Sg Javanica
Sg Icterohaemorrhagiae
Sg Australis
Sg Sejroe
Sg Pyrogenes
Sg Pomona
Sg Ballum
Arenal
Javanica
Copenhageni
Icterohaemorrhagiae
Lai
Bim
Bratislava
Jalna
Hardjo
Balcanica
Pyrogenes
Manilae
Canicola
Pomona
Mozdok
Bataviae
Hurstbridge
Lyme
Varillal
Patoc
Ballum
Kenya
Tarassovi
Lsan
Lbor
Lint
Com
Lkir
Lbro
Lfai
Lina
Llicer
Lbif
Lbor
Lsp
156 406 535
69 178 392
412 415 591
957 20 158
270 163 176
47 261 747 638
31 344 169 192
181 319 338
31 74 327
316 44 485
332 319 314
10 43
55 34
46 14
212 37
241 95
515 310 303
10
50
100
150
200
250

Figure 4.5. Gene presence/absence matrices of *rfb* clusters from different *Leptospira* strains and species, covering a range of distinct serogroup/serovar identities.

Horizontal lines correspond to individual genes or set of genes grouped according to their percentage of similarity (cut-off 60%), green meaning presence, and black absence. Scales on the left of the matrices indicate the number of different genes being compared. Columns correspond to different *Leptospira* spp. serovars as indicated on the columns' labels. **(A)** Strains were organized after hierarchical clustering considering presence/absence of *rfb* genes. Strain names marked in red are those sequenced in this study and their serogroup identity is indicated between parentheses. Serogroups comprising several serovars are indicated in dotted brackets and bold lettering. The yellow square indicates genes exclusively present in non-agglutinating *L. noguchii* strains. **(B)** The representation of three serovars (labeled in red), each one corresponding to two different *Leptospira* species as marked, are shown side-by-side with the same matrix representation as in panel (A). **(C)** Comparison of the *rfb* cluster in whole genomes from different species (Lsan=*L. santarosai*; Lint=*L. interrogans*; Lkir=*L. kirschneri*; Lbor=*L. borgpetersenii*; Lbro=*L. bromii*; Lfai=*L. fainei*; Lina=*L. inadai*; Llicer=*L. liceriasae*; Lbif=*L. biflexa*; Lsp=*Leptospira* sp.), presented side by side by grouping species with the same serovar (labeled in bold red) and serogroup (enclosed in dotted brackets and bold lettering). The numbers at the top of each column correspond to the number of contigs for draft genomes, while complete/finished genomes are marked as *Com*.

Worth highlighting is the fact that serovars belonging to the same serogroup, shared all, or the vast majority of genes among their *rfb* clusters. In this regard, Australis was the most variable serogroup (**Figure 4.5A**), with strains barbudensis, bajan and 201102933 almost clonal, and IP1611024 more closely related to serovars Bratislava and Lora. Of note, the serogroup of strain 201102933 is not known, but the proximity to barbudensis and bajan, and the clustering with other strains of serogroup Australis, strongly suggests that strain 201102933 belongs to Australis, a hypothesis-driven prediction amenable for future testing.

In further support of serovar-specific *rfb* genetic signatures, is the presence of 26 *rfb* genes only present in, and shared among, those Uruguayan strains which were not amenable to serogroup assignment (genes framed within a yellow square in **Figure 4.5A**). Almost identical gene arrays were recognized among the shared signature gene sets. After a search of orthologous protein sequences (Blast best hits Accession Numbers are indicated in **Table S4.6**), among the shared signature genes several were found to encode carbohydrate-active enzymes such as UDP-glucuronate 4-epimerase [EC:5.1.3.6], GDP/UDP-N,N'-diacetylbacillosamine 2-epimerase [EC:3.2.1.184] N,N'-diacetyl-legionaminate synthase [EC:2.5.1.101] and CMP-N,N'-diacetyl-legionaminic acid synthase [EC:2.7.7.82] (using PFAM and KEGG Mapper). These enzymes likely play key roles in generating the –yet to be determined– serovar-distinctive structures of LPS *O*-antigens.

The comparative analyses described above (**Figure 4.5A**) have limitations, since not all serovars are represented. Either due to a lack of information in the characterization of isolates, or because not enough finished and good-quality draft genomes are available, information loss is inevitable, eventually hampering reliable reconstructions of the relevant *rfb* genetic clusters. To further address these issues, and as a means of testing the decisive role of gene composition in serovar determination, representative genomes corresponding to identical serovar but belonging to different *Leptospira* species, were analyzed in detail. Seven reliably determined serovars were found to belong to different species. In four of them the *rfb* clusters were split into more than one contig, so that for the sake of maximum reliability, only strains corresponding to three serovars were used: Hurstbridge (from *L. broomii* and *L. fainei* strains), Hardjo (*L. borgpetersenii* and *L. interrogans*) and Valbuzzi (*L. interrogans* and *L. kirschneri*) (**Table S4.5**, second sheet). The gene presence/absence matrix calculated for the *rfb* clusters from this subgroup (**Figure 4.5B)**, readily confirms that specific groups of *rfb* genes are associated to each serovar (**Table S4.6**, second sheet), irrespective of the species. Despite the previously stated constraint in terms of genomic fragmentation, a "pan-*rfb"* was created considering the genes resulting from the analysis in **Figure 4.5A**. The comparison of this artificial *rfb* against whole genome sequences from different serovars, including more than one strain in each case, showed the same genetic pattern for several examples (**Figure 4.5C** and **Table S4.6**, third sheet). These findings constitute a solid starting point to define a comprehensive set of serovar-specific genetic signatures, eventually revising the current protocols for serogroup and serovar assignments, which are extremely useful in clinical and epidemiologic work.

4.4 Discussion

L. noguchii strains are pathogenic members of the genus *Leptospira*, of worldwide distribution and, together with *L. interrogans*, one of the species most involved in human leptospirosis (Vincent et al., 2019). Beyond human infections, *L. noguchii* has been detected or isolated from different hosts (Silva et al., 2007; Martins et al., 2015; Barragan et al., 2016; Zarantonelli et al., 2018), exhibiting a singular host-adaptation capacity, and being one of the few *Leptospira* species isolated not only from different mammals but also from amphibians (Everard et al., 1988; Gravekamp et al., 1991). However, *L. noguchii* is still a poorly

characterized species, with scarce information about the circulating serovars. A few draft genomes of *L. noguchii* have been published (Moreno et al., 2015; Nieves et al., 2019), but no complete genomes had been reported so far.

The twelve *L. noguchii* strains that we have now sequenced, were selected such that a broad range of serogroup variants were included, as well as different hosts including cattle, human and amphibians. These genomes were analyzed in terms of their core genome and strain-specific genes and put in a broader context by studying their phylogenetic relationship with other *L. noguchii* strains. Special attention was given to analyzing the genetic composition of the *rfb* cluster, which is notoriously linked to serovar determination (Patra et al., 2015; Picardeau, 2017). Moreover, this analysis was extended to other serovars from different *Leptospira* species.

The genome architecture in terms of replicon content and its organization is overall similar to that found in other *Leptospira* species (Picardeau et al., 2008), exhibiting two main chromosomes and a diverse repertoire of plasmids. On average, *L. noguchii* has a larger genome (4.8 Mb) compared to other relevant pathogens such as *L. interrogans* (4.6 Mb) and *L. borgpetersenii* (3.9 Mb), with a larger array of accessory genes that might be involved in *L. noguchii*'s remarkable adaptability. Indeed, approximately 30% of the genes from the strains sequenced in this study, correspond to accessory genome, with a clear enrichment in carbohydrate biochemistry pathways. Concerning the plasmid repertoire, some strains showed identical or nearly identical protein-encoding plasmidic genes (**Figure 4.1**), suggesting that these replicons may be transferred between strains. Such exchange among different species may be evolutionarily ancient, considering the large number of shared proteins that can be identified in several cases (*e.g. L. mayottensis* str 200901116 p1, *L. mayottensis* str MDI222 pLmayMDI222, *L. interrogans* sv Bataviae str 1489 p4 and *L. noguchii* str IP1611024 p1). On the other hand, plasmid-borne gene variability seems larger in some *Leptospira* species (**Table S4.3**, first sheet), perhaps reflecting a varying contribution of horizontal plasmid acquisition as a source of adaptation. COG analyses revealed no variation in functional representation between plasmids from *L. noguchii vs* those from other species, except for the absence of genes related to amino acid metabolism and transport in *L. noguchii* plasmids.

Phylogenetic analyses using *L. noguchii* WGS data (**Figure 4.2)** did not reveal a correlation of genotype with geographical distribution nor with host specificity traits. This is

consistent with previous reports (Loureiro et al., 2020), now further including the set of ten finished genomes and two high-quality drafts that were not previously available. This led us to focus on the study of genes associated with LPS synthesis, to explore genotype-phenotype relationships that could underlie the rich phenotypic complexity exhibited among and within *Leptospira* species, which is expressed as disparate serovars/serogroups. It is known that the expression of surface epitopes in the LPS is a major determinant for serovar identity, particularly concerning the LPS's composition and spatial arrangement of sugars (Adler and de la Peña Moctezuma, 2010). The *rfb* cluster, which harbors most of the genes encoding enzymes for the LPS *O*-antigen synthesis, has been studied in *Leptospira* species of different serovars (de la Peña-Moctezuma et al., 1999; Kalambaheti et al., 1999; Bulach et al., 2000; Casas et al., 2005; Fouts et al., 2016), showing great variability in terms of genetic content and synteny. LPS *O*-antigens are frequently synthesized through a so called Wzx/Wzy-dependent pathway, implicating an *O*-flippase (Wzx) and an *O*-polymerase (Wzy). This biosynthesis system is complex and includes several proteins highly conserved in Gram-negative and Gram-positive bacteria that possess glycan polymers on their cell surfaces, such as LPS *O*-antigen, spore coats, enterobacterial common antigen and outer capsules (Islam and Lam, 2014). *Leptospira* spp. appear to conserve the Wzx/Wzy-dependent *O*-antigen biosynthesis pathway, exhibiting orthologues of the genes encoding these two enzymes within the *rfb* cluster (Nascimento et al., 2004), even though variations of the typical pathway are anticipated based upon the sporadic presence of *wzy* and *wzz* genes (Fouts et al., 2016). However, a definite association between the presence of genes and serovar determination has not been shown. Some studies in *Leptospira* have been insightful, albeit focused into identifying pathogenicity determinants via characterization of LPS products (Patra et al., 2015; Vanithamani et al., 2021), but limited to the examination of subsets and not all the genes present within *rfb* clusters.

By including twelve genome sequences from *L. noguchii*, and systematically extending the analysis to all *Leptospira* species with assigned serovar identity, we have now substantiated a definite and biunivocal link between *rfb* gene composition and serogroup/serovar identity (**Figure 4.5**). The requirement of closed genomes in studies that depend on the analysis of gene presence/absence cannot be overstated, as even minor assembly mistakes, often introduced in multiple contig draft genomes, can lead to major misinterpretations. Many available draft genomes suffer from incompleteness artifacts due to contig closure errors (Kosugi et al., 2015; Lu et al.,

2020). Progress in the elucidation of the structure of different LPS variants from *Leptospira* can be anticipated by using NMR, as done with other bacterial genera (Fontana et al., 2014). Such an approach critically depends on accurate whole genome sequence information, to single out ambiguous alternatives from the carbohydrates' NMR spectra according to the specific set of glycosyltransferases present in the genome. The LPS structures shall thus be conclusive about their dependence on distinct sets of genes present in the genomes, rather than on regulation of gene expression. A serovar-specific genetic fingerprint such as the one we are now reporting, shall be instrumental to shifting from serologic techniques to simpler and more accurate PCR-based serovar determination protocols. Considering the clinical impact that different *Leptospira* serovars exhibit, associated to distinct host adaptation and virulence phenotypes, such molecular genetic approaches have been attempted, but only at the level of serogroup (Cai et al., 2010) or for some serovars (Bezerra da Silva et al., 2011) and suffering from considerable cross-detection among strains. As further, and reliably complete *rfb* gene clusters from finished/closed genomes become available, more precise gene composition assessments among *Leptospira* serovars are expected to be made.

Added to this serovar-specific signature found in the *rfb* cluster, strong indications that the *rfb* functions as a large genomic island, dispersed via events of horizontal gene transfer (HGT), were uncovered. Hallmarks of HGT include (i) the characteristic low GC-content within the *rfb* clusters, ostensibly distinct from its flanking DNA segments', (ii) signs of genomic rearrangements (inversions) in the cluster's surroundings, and (iii) the fact that identical/nearly identical *rfb* clusters were found in different *Leptospira* species. A similar HGT scenario of genomic islands has been described in different classes of Proteobacteria, concerning LPS(*O*-antigen)- and capsule polysaccharide(K-antigen)-encoding loci (Thrane et al., 2015; Huszczynski et al., 2020; Bian et al., 2021; Buzzanca et al., 2021). Such loci have been observed to locate at highly plastic regions of the genomes from enterobacteria such as *Escherichia* and *Salmonella*, *Pseudomonas*, *Vibrio* and *Aliarcobacter*, among other genera, exhibiting clear evidence of HGT underlying locus exchange.

Taking all the evidence together, our results strongly suggest that serovar identity can change in *Leptospira* by horizontal gene transfer (HGT) of a part or even the entire *rfb* cluster, acting as an LPS *O*-antigen-encoding genomic island. This HGT phenomenon seems to occur

within and among species from the *Leptospira* genus, contributing to population diversity and adaptability. This observation is consistent with the large variation of gene composition among different *rfb* clusters, with more, or less genes being transferred in different cases, together with the ill-defined downstream limit of the island that had already been reported (Fouts et al., 2016). The molecular HGT mechanism explaining *rfb* exchange in *Leptospira* remains to be determined, potentially by homologous recombination, IS elements –which are indeed observed surrounding and within the locus– or phages (Wang and Quinn, 2010).

4.5 Materials and Methods

4.5.1 DNA extraction, sequencing, and assembly

Genomic DNA was extracted from 100 mL of a 10^8 bacteria/mL culture of each isolate using QIAGEN Genomic-tip 100/G and Genomic DNA Buffer Set (QIAGEN). PacBio SMRT sequencing was performed with RSII technology (McGill University-Genome Quebec, Canada; Eurofins, Germany). *De novo* assembly was performed with HGAP v.4 (Chin et al., 2013) available on SMRT Link v.7 (default parameters, except, min. subread length: 500; estimated genome size: 4.8 Mb), Canu (https://github.com/marbl/canu) (Koren et al., 2017), Unicycler (https://github.com/rrwick/Unicycler) (Wick et al., 2017), or Trycycler (https://github.com/rrwick/Trycycler) (Wick et al., 2021). The polishing step was run on SMRT Link v.7 using the Resequencing application (default parameters).

4.5.2 Genome data sets

Genomes included in the phylogenetic analyses were downloaded from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) or from the Institut Pasteur Bacterial Isolate Genome Sequence Database (https://bigsdb.pasteur.fr/leptospira/). The metadata for all isolates, including for those sequenced in this study, are summarized in **Table S4.4**. Genomes from strains of known serovar used for the *rfb* cluster comparisons were downloaded from Patric (https://www.patricbrc.org), and their metadata are summarized in **Table S4.5**. Plasmid sequences from other *Leptospira* species were also obtained from GenBank and their associated metadata are summarized in **Table S4.3**, along with the general features of plasmids from the

strains sequenced in this study. Of note, plasmids p2 from strain 'barbudensis' and p5 from 'IP1712055' were not included in the network analysis (**Figure 4.1** and **Table S4.3**, third and fourth sheets) in order to simplify the comparison, as they show high similarities with plasmids p1 and p3 from those strains, respectively. Nevertheless, all genbank plasmid sequence files are included as supplemental material.

4.5.3 Phylogenetic analyses

All 38 genome sequences (including *L. interrogans* str. 56601 and *L. kirschneri* str. 200702274 as outgroups) were annotated using Prokka version 1.13.7 (Seemann, 2014). Orthology between the coding sequences were inferred using the combination of the two algorithms COG and OMCL through GET_HOMOLOGUES version 20190411 (Contreras-Moreira and Vinuesa, 2013). The sequences of orthologous genes single copy and corresponding to the softcore (sequences present in more than 95% of the genomes) were aligned using Mafft version 7.407 (Katoh and Standley, 2013). The resulting alignments were filtered using BMGE version 1.12 (Criscuolo and Gribaldo, 2010) and concatenated in a partitioned supermatrix using AMAS (Borowiec, 2016). The best-fit model of each partition and the maximum-likelihood phylogeny was performed using IQ-TREE version 1.6.11 (Nguyen et al, 2015) and 10,000 ultrafast bootstraps (Hoang et al., 2018). The same protocol was followed for the construction of the phylogenetic tree, but sequences were codon aligned using TranslatorX version 1.1 (Abascal et al., 2010).

The ANI index was calculated using the OrthoANIu algorithm available at EZBioCloud (https://www.ezbiocloud.net/tools/ani) (Yoon et al., 2017).

4.5.4 Genomic analyses

Comparative pangenome analysis was performed using Roary version 3.11.2 (Page et al., 2015). By combining the use of Blast+ (Camacho et al., 2009) and KEGG (Aramaki et al., 2020; Kanehisa and Sato, 2020) it was possible to assign functions of selected core and accessory genes (**Table S4.2**). *Leptospira* genomes with assigned serovars and less than 500 contigs (**Table S4.5**) were downloaded, and their *rfb* clusters, lipid A and core oligosaccharide

biosynthesis-encoding clusters analyzed. Location of sites and gene cluster sequence extraction were done with the bioinformatics tools included in Emboss 6.6.0 (Rice et al., 2000) and then annotated using Prokka version 1.13.7. Softcore *rfb* alignments were performed with Mafft through Roary version 3.11.2, considering 60% identity and gene presence in at least 60% of strains included in the analysis. The resulting alignment was then used to calculate the phylogenetic tree using IQ-TREE version 1.6.11. The synteny of *rfb* clusters, as well as pairwise genome comparisons to determine conservation, were inferred and represented using Easyfig 2.2.2 (Sullivan et al., 2011). Gene presence/absence analyses among *rfb* clusters from different serovars were performed by protein-level cross-matching and subsequent network associations. Briefly, a pairwise comparison of each *rfb* cluster with one another was conducted using blastp. The network connection was thereafter established using the previously generated blast files and NetworkX version 2.6.2 (Hagberg et al., 2008) with 60% similarity threshold, thus proteins having ≥60% similarity were grouped together generating the gene presence/absence matrices. Close-up plots of GC-content along linearized sequence fragments was performed with DNAplotter (Carver et al., 2009). Network association analysis of plasmid protein-encoding genes repertoire was carried out as described for the *rfb* cluster, considering a 60% similarity cut-off. Hierarchical clustering according to the shared protein-encoding genes (options used: Euclidean distance, ward linkage) was performed using available tools at https://mev.tm4.org. Functional annotation was done using eggnog-mapper v2 (Huerta-Cepas et al., 2017). Transposase positions across chromosome 1 of *L. noguchii* strains were obtained from the Prokka annotation, and then represented using the online version of shinyCircos (https://venyao.xyz/shinycircos/) (Yu et al., 2018).

4.6 Data availability

The genome sequences have been deposited in DDBJ/ENA/GenBank under the BioProject PRJNA803166. Specifically, with the following accession numbers: *L. noguchii* strain barbudensis CP091967-CP091970; *L. noguchii* strain 201601331 CP091962-CP091966; *L. noguchii* strain IP1512017 CP091957-CP091961; *L. noguchii* strain IP1605021 CP091953-CP091956; *L. noguchii* strain IP1611024 CP091947-CP091952; *L. noguchii* strain IP1703027 CP091943-CP091946; *L. noguchii* strain IP1705032 CP091940-CP091942; *L. noguchii* strain

IP1709037 CP091936-CP091939; *L. noguchii* strain IP1712055 CP091928-CP091935; *L. noguchii* strain IP1804061 CP092112-CP092116; *L. noguchii* strain bajan JAKNBP000000000; *L. noguchii* strain 201102933 JAKNBO000000000.

Acknowledgments

This research was supported by Institut Pasteur through grant PTR 30-2017 (M.P., A.B., F.J.V.); by Institut Pasteur & Institut Pasteur de Montevideo through their Pasteur International Joint Research Units program "Integrative Microbiology of Zoonotic Agents" (ImiZA) (M.P., A.B.); and by the Natural Sciences and Engineering Research Council of Canada (NSERC) discovery grant (RGPIN-2016-04940) (F.J.V.). C.N. Received a Ph.D. studentship Calmette & Yersin from the Institut Pasteur International Network. F.J.V. received a Junior 1 and Junior 2 research scholar salary award from the Fonds de Recherche du Québec - Santé. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript. We thank Howard Takiff and Lizeth Caraballo (Venezuelan Institute of Scientific Investigation) for providing a *L. noguchii* isolate; and Gregorio Iraola and Ignacio Ferrés for helpful discussions.

Chapter 5. Research article on intraspecies genetic diversity in *Leptospira* (Article 3)

Phylogenomics of *Leptospira santarosai*, a prevalent pathogenic species in the Americas

Diana Chinchilla[1,2]*, Cecilia Nieves[3]*, Ricardo Gutiérrez[1,2], Vallier Sordoillet[4], Frédéric J. Veyrier[3], Mathieu Picardeau[4]

[1]Instituto Costarricense de Investigación y Enseñanza en Nutrición y Salud (INCIENSA), La Unión, Cartago, Costa Rica. [2]Instituto Costarricense de Investigación y Enseñanza en Nutrición y Salud (INCIENSA), Centro Nacional de Referencia de Bacteriología, Cartago, Tres Ríos, Costa Rica [3]Bacterial Symbionts Evolution, Centre Armand-Frappier Santé Biotechnologie, Institut National de la Recherche Scientifique, Université du Québec, Laval, QC, Canada. [4]Institut Pasteur, Université Paris Cité, CNRS UMR 6047, Biology of Spirochetes Unit, Paris, France.

*contributed equally to this work

Title of journal

PLoS Neglected Tropical Diseases, DOI 10.1371/journal.pntd.0011733 | Submitted 15 August 2023 | Accepted 17 October 2023 | Published 2 November 2023

Author contributions

C. Nieves conducted the majority of the bioinformatics analyses and generated most of the figures. This author was also actively engaged in critically reviewing and editing the document through multiple rounds of corrections.

DC, writing, editing, bacterial isolation; RG, editing, critical reviewing, supervision; VS, sequencing, some bioinformatic analysis; FV, critical reviewing, editing, supervision; MP writing (original draft), critical reviewing, editing, funding, resources.

5.1 Abstract

5.1.1 Background

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

5.1.2 Methodology/principal findings

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

5.1.3 Conclusions/significance

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

5.2 Author summary

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

5.3 Introduction

Leptospira is a highly heterogeneous bacterial genus divided into pathogenic and saprophytic species and then further divided into more than 300 serovars, which are defined according to structural heterogeneity of the lipopolysaccharide (LPS) *O*-antigen. Nowadays, strain identification is mainly based on genome analysis, and core genome multilocus sequence typing (cgMLST) (Guglielmini et al., 2019) enables identification of the species and below. Recent studies have also shown that whole-genome sequences can be used for predicting *Leptospira* serotypes on the basis of the *rfb* locus which contains the genes for the *O*-antigen biosynthesis (Nieves et al., 2022; Medeiros et al., 2022). This approach offers a promising alternative to the conventional serotyping method, which is laborious, time-consuming, expensive and requires a high level of expertise.

Over the past decade, the number of *Leptospira* species described has rapidly extended from 22 in 2014 to 69 in 2022 (Fernandes et al., 2022), largely due to the use of improved protocols for culture isolation from the environment (Chakraborty et al., 2011; Thibeaux et al., 2018) and the generalization of next generation sequencing (Land et al., 2015). Among the genus *Leptospira*, eight species (*L. interrogans*, *L. kirschneri*, *L. noguchii*, *L. santarosai*, *L. mayottensis*, *L. borgpetersenii*, *L. alexanderi* and *L. weilii*), which diverged after a specific node of evolution, constitute the most virulent group of pathogenic species (Vincent et al., 2019). These *Leptospira* species are the causative agents of leptospirosis in both humans and animals, leading to a high

disease burden in tropical countries (Costa et al., 2019) and major economic losses in the livestock sector (Noguera et al., 2022).

Our previous analysis of the distribution of pathogenic *Leptospira* species showed that *L. interrogans* is the most frequently encountered and globally distributed species (Guglielmini et al., 2019). This cosmopolitan species is also by far the most studied in terms of virulence, and molecular epidemiology, among other aspects. On the contrary, to date, very little is known about the geographical distribution, reservoirs, genomic features and virulence factors of pathogenic species other than *L. interrogans*. In the same analysis, we showed that some pathogenic species were geographically restricted (Guglielmini et al., 2019). Thus, only limited reports have described the existence of *L. santarosai* outside the American continent. *L. santarosai*, named after Carlos A. Santa Rosa, a Brazilian veterinary microbiologist who pioneered the study of leptospirosis in Brazil, was first described in 1987 (Yasuda et al., 1987). *L. santarosai* is predominant in many countries from Central and South America.

In the present study, we first performed an analysis of the pangenome in pathogenic *Leptospira* species and then took a closer look at the genetic diversity of *L. santarosai* including a set of strains recently isolated from patients in Costa Rica, which is an endemic country for leptospirosis (Carvajal and Fagerstrom, 2017).

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

5.4 Materials and methods

5.4.1 Ethics statement

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

of the study. Human samples were anonymized, and collection of the samples was conducted according to the Declaration of Helsinki.

5.4.2 Strains

Isolates sequenced in this study (n = 153) were obtained from the collections of the French National Reference Center for Leptospirosis (Institut Pasteur, Paris, France), Laboratorio de Genética Molecular (Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela), Institut Pasteur of Alger (Algiers, Algeria), Institute of Veterinary Bacteriology (University of Bern, Switzerland), Molecular Epidemiology and Public Health Laboratory (School of Veterinary Sciences, Massey University, New Zealand), Instituto de Higiene (Facultad de Medicina, Universidad de la República, Montevideo, Uruguay), Universidade Federal Fluminense (Rio de Janeiro, Brazil), Faculty of Veterinary Medicine (University of Zagreb, Croatia), National Collaborating Centre for Reference and Research on Leptospirosis (Academic Medical Center, Amsterdam, the Netherlands), Laboratory of Zoonoses (Pasteur Institute in Saint Petersburg, Saint Petersburg, Russia), Institute for Medical Research (Malaysia), Faculty of Medicine and Health Sciences (University Putra Malaysia, Malaysia), and Leptospirosis Research and Expertise Unit (Institut Pasteur Nouvelle-Calédonie, Nouméa, New Caledonia), Kimron Veterinary Institute (Israel). We also downloaded genomes from our previous studies including isolates from the collections of Lao-Oxford-Mahosot Hospital-Wellcome Trust-Research Unit (LOMWRU) (Microbiology Laboratory, Mahosot Hospital, Vientiane, Lao People's Democratic Republic), Unidad Mixta Pasteur-Instituto Nacional de Investigación Agropecuaria (Institut Pasteur of Montevideo, Montevideo, Uruguay), Centre Hospitalier de Mayotte (France), and Department of Mycology-Bacteriology (Institute of Tropical Medicine Pedro Kourí, Havana, Cuba) (Zarantonelli et al., 2018; Guglielmini et al., 2019; Noda et al., 2020; Grillová et al., 2021; Nieves et al., 2022) as well as genomes from the NCBI database. Information on strains and genomes used in this study are indicated in **Tables S5.1** and **S5.2**.

5.4.3 Whole-genome sequencing

Illumina sequencing was performed from extracted genomic DNAs of exponential-phase cultures using a MagNA Pure 96 Instrument (Roche, Meylan, France). Next-generation sequencing (NGS) was performed using Nextera XT DNA Library Preparation kit and the NextSeq 500 sequencing systems (Illumina, San Diego, CA, USA) at the Mutualized Platform for Microbiology (P2M) at Institut Pasteur. CLC Genomics Workbench 9 software (Qiagen, Hilden, Germany) was used for analyses. The generated contig sequences together with the sample metadata are available in BIGSdb hosted at the Institut Pasteur (https://bigsdb.pasteur.fr/leptospira/). We also downloaded additional genome sequences of *Leptospira* isolates from the NCBI database (**Table S5.1**). Only genomes meeting quality requirements, such as i) sequencing coverage >30x, ii) number of contigs (<600), iii) cumulative contigs length within the typical range of *Leptospira* genomes (3.6-6Mb), iv) GC content within the typical range of *Leptospira* genomes (35-48%), and v) <100 uncalled cgMLST alleles out of the 545 pre-defined core genes, were selected for further analyses.

5.4.4 Genomic analyses

Comparative analyses of the pangenome were performed using two software: Roary version 3.11.2 (Page et al., 2015), and a combination of COG and OMCL algorithms in GET_HOMOLOGUES version 20190411 (Contreras-Moreira and Vinuesa, 2013). Both methods yielded a similar number of gene clusters. In the Roary analysis, a 60% identity cut-off was applied to define gene clusters (option -i 60), and no other parameters were modified. Among the Roary outputs, a tab-separated file containing the number of genes in the pangenome was used to create a graph depicting the variation in the number of gene clusters as a function of the number of genomes analyzed. Roary iterated 10 times, calculating the number of new genes added as each genome was sequentially incorporated into the analysis. This graph facilitated a quick determination of whether the pangenome was open or closed and allowed for the calculation of the α coefficient in Heap's Law ($n = \kappa N\gamma$, with $\gamma = 1 - \alpha$) (Tettelin et al., 2008). On the other hand, GET_HOMOLOGUES was used to infer the pangenome distribution in cloud-, shell-, soft-core-, and core-genome. This was achieved by generating a tab-separated pangenome matrix file that included the number of all the clusters identified by both COG and OMCL algorithms. The matrix represented the intersection of the two methods and served as input for the

parse_pangenome_matrix.pl script within GET_HOMOLOGUES, which classified the clusters as cloud (shared by up to 2 genomes), shell (shared by more than 2 genomes but less than 93% of genomes analyzed), soft-core (shared by 93–99% of genomes), or core-genes (shared by 100% of genomes). Due to the substantial number of genomes available for *L. interrogans* and *L. borgpetersenii*, as well as the redundancies observed in serogroups and serovars, representative genomes of each serogroup/serovar were selectively chosen to mitigate computational costs. Excluding genomes with redundant identities is not anticipated to result in significant alterations in the pangenome distribution.

Genome size and GC content for highly virulent *Leptospira* species were determined through DFAST annotation (Tanizawa et al., 2018). Individual values were plotted and grouped per species, with the mean and standard deviation displayed. Genome size and GC content were compared using the Kruskal-Wallis Rank Sum Test, for the comparison of *Leptospira* spp. and the Wilcoxon rank test, for the comparison of two phylogenetic-related groups. Post-hoc comparisons were performed using Dunn's Kruskal-Wallis Multiple Comparisons (Dunn, 1964). P-values were adjusted with the Bonferroni method. Statistical analyses were performed in R (RCoreTeam, 2016), using FSA package (Ogle et al., 2023).

Average Nucleotide Identity (ANI) and Percentage of Conserved Proteins (POCP) were calculated for the 64 *L. santarosai* genome sequences as well as *L. interrogans* str. Fiocruz L1-130 and *L. borgpetersenii* str. M84 used as outgroups (**Figure S5.1** and **S5.2**). Genomes were annotated by Prokka version 1.13.7 (Seemann, 2014). ANI and POCP matrices were inferred using OMCL algorithm via GET_HOMOLOGUES version 20190411 (Contreras-Moreira and Vinuesa, 2013). Briefly, to calculate ANI, the option -A was employed along with option -a to utilize nucleotide sequences and perform BLASTN. This process generated a tab-separated file containing average percentage sequence identity values between pairs of genomes, calculated from sequences within all identified clusters (option t = 0). This tab-separated file served as the input to create a symmetric matrix, where the genomes were clustered based on their ANI values. Dendrograms based on this clustering were generated on both sides of the matrix to visually represent the proximity among genomes. Similarly, POCP was calculated by including the option -P and performing default BLASTP searches. This step yielded another tab-separated file, which was subsequently used to create a symmetric matrix. Analogous to the ANI matrix, the genomes were clustered based on the

shared % of conserved proteins between pairs of genomes. These values are calculated as POCP = $(C_a + C_b)/(total_a + total_b)$, where C_a and C_b denote the number of conserved proteins from genome a in genome b and from genome b in genome a, respectively, normalized by the sum of total proteins in each genome. The clustering process also generated dendrograms, indicating the proximity among genomes in terms of conserved proteins.

Core genome MLST (cgMLST) typing was performed using a scheme based on 545 core genes as previously described (Guglielmini et al., 2019). *L. santarosai* core-genome based phylogeny was constructed using the 1288 core-genes alignment resulting from Roary analysis (60% identity cut-off, option -i 60). The best-fit model and the maximum-likelihood phylogenetic tree were determined by IQ-TREE version 1.6.11 (Nguyen et al., 2015), considering 10,000 ultrafast bootstraps (Hoang et al., 2018). *L. interrogans* str Fiocruz L1-130 and *L. borgpetersenii* str. M84 were used as outgroups. Tree branches were transformed with the "proportional" option on FigTree software v1.4.4 (http://tree.bio.ed.ac.uk/software/figtree/), which adjusts branch distances according to the number of tips under each node to improve visualization of the tree. Gene presence/absence analyses among *rfb* clusters from different genomes here studied were performed by protein-level searches using BLASTP (Camacho et al., 2009) and subsequent network associations by NetworkX version 2.6.2 (Hagberg et al., 2018). A similarity threshold of 60% was applied, as previously described (Nieves et al., 2022). The resulting presence/absence table obtained from the network association analysis was converted into a binary CSV file, where 0 represents gene absence and 1 represents gene presence. This binary table was subjected to hierarchical clustering based on shared protein-encoding genes (options: euclidean distance, ward linkage) using available tools at https://mev.tm4.org. Jaccard's similarity index was used to measure the similarity between *rfb* patterns.

5.4.5 Genomic data

The sequencing data generated in this study are available in the NCBI database under the BioSample accession numbers SAMN34670613, SAMN34670614, SAMN34670615, SAMN34670616, SAMN34670617, SAMN34670618, SAMN34670619, SAMN34670620, SAMN34670621, SAMN34670622, SAMN34670623, SAMN34670624, SAMN34670625,

SAMN34670626, SAMN34670627, SAMN34670628, SAMN34670629, SAMN34670630. Genome sequences used in this study are also available at https://bigsdb.pasteur.fr/leptospira/.

5.5 Results and Discussion

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

We first investigated the geographical distribution of pathogenic *Leptospira* species using 914 genomes of isolates collected between 1928 and 2022 (**Table S5.1**).

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

Strains were isolated from human (50%) and animal (49%) samples, in Europe (18 %), Africa (2 %), Indian Ocean (14%), Caribbean islands (6%), Central America (3%), South America (13%), North America (4%), Central Asia, South Asia, East and Western Asia (11%), Southeast Asia (14 %), and Australia and the Pacific region (15%) (**Figure 5.1**). Although this study is based on the genomes available in the databases and may introduce a bias, *L. interrogans*, *L. kirschneri*, and *L. borgpetersenii* are distributed worldwide, *L. weilii* is mostly found in Asia, Australia and the Pacific region, *L. mayottensis* in the Indian Ocean, and *L. noguchii* and *L. santarosai* in America as previously shown (Guglielmini et al., 2019).

Leptospirosis is endemic in most countries of South and Central America, as well as in the Caribbean region (Hotez et al., 2008; Costa et al., 2015; Peters e al., 2017; Schneider et al., 2017). In addition, most outbreaks of leptospirosis have been reported in the Latin America and the Caribbean region (Munoz-Zanzi et al., 2020), where the disease is widespread in domestic and wild animals (Carvajal and Fagerstrom, 2017). However, comprehensive data concerning human and animal leptospirosis remain largely scarce in most American countries (Pereira et al., 2018). We previously studied the genomes of *L. noguchii* isolated from human and animals in America (Nieves et al., 2022) but our knowledge of *L. santarosai*, the other prevalent species in America, is rather limited.

Here, we sequenced 18 *L. santarosai* strains, including twelve strains that were isolated in Costa Rica in 2020–2021 from patients. The ANI and POCP values were calculated for the 64 *L. santarosai* strains further confirming they all belong to the same species (**Figure S5.1** and **S5.2**). Of the 64 *L. santarosai* strains in our genome database, 28 were isolated in South America (Brazil, Colombia, Ecuador, Peru), 21 in Central America (Costa Rica, Panama), 6 in North America (US; not including Puerto Rico), 7 in the Caribbean region (Martinique, Guadeloupe, Trinidad and Tobago and Puerto Rico), and only two strains were isolated outside the Americas (China and Democratic Republic of the Congo) (**Figure 5.1** and **Table S5.2**). Of note, *L. santarosai* has not been isolated in Uruguay, where a large number of *Leptospira* strains have been isolated from cattle (Zarantonelli et al., 2018).

L. santarosai strains in our study were isolated from humans (n = 38), cattle (n = 12), rodents (n = 8, including rats, spiny rats, capybara and muskrat), opossum (n = 2), dog (n = 1), goat (n = 1), pig (n = 1), and racoon (n = 1) (**Table S5.2**).

Previous studies have shown that *L. santarosai* can be detected from different sources in many countries of America and the Caribbean region. It is the predominant species in humans, rodents and dogs in Peru and Colombia (Peláez-Sánchez et al., 2017; Delgado et al., 2022). In Peru, it has additionally been found in rural environmental water samples (but not in urban samples), as well as in association with pigs and cattle (Ganoza et al., 2006). In Brazil, *L. santarosai* has been isolated from dogs (Miotto et al., 2016), cattle (Kremer et al., 2015), goats (Lilenbaum et al., 2015), and capybaras (Moreno et al., 2016). Moreover, it has also been identified in patients in French Guiana (Kallel et al., 2020), Guadeloupe (Bourhy et al., 2013) and the US (Wilson et al., 2014).

Only a few reports have described the existence of *L. santarosai* outside the American continent. Some years after the original description of *L. santarosai* (Yasuda et al., 1987), Brenner et al. listed 65 *L. santarosai* strains, of which only three were isolated from outside America (Brenner et al., 1999). One *L. santarosai* strain was isolated from a patient in Sri-Lanka in 1966 but has never been reported in this country afterwards (Naotunna et al., 2016; Jayasundara et al., 2015). The other two strains were isolated in Denmark and Indonesia but, again, *L. santarosai* has not been subsequently isolated in these countries. More recently, a strain belonging to *L. santarosai* serogroup Grippotyphosa was isolated from a patient in India and its genome sequenced (Lata et al., 2020). However, because of highly fragmented genome (884 contigs) and missing genomic

data (135 uncalled cgMLST alleles), the cluster assignment was not possible for this isolate and we removed its genome from our analysis.

The serogroup Shermani, is commonly reported in serological surveys in animals in Asia (Oni et al., 2004; Suwancharoen et al., 2013; Suut et al., 2016; Alamuri et al., 2019). Unfortunately, there is no evidence that the infecting strains described in these studies were *L. santarosai* or another species such as *L. noguchii* and *L. inadai* which also contain serovars from the serogroup Shermani (Levett, 2001). Finally, the other country outside Americas where *L. santarosai* was reported is Taiwan in East Asia. Serogroup Shermani, presumably belonging to *L. santarosai*, is predominant among patients with severe leptospirosis in Taiwan (Wang et al., 2020). However, only one *L. santarosai* strain, strain CCF, has been isolated from a patient with leptospirosis in Taiwan (Hsieh et al., 2005) and this strain, for which we do not have the complete genome (Chou et al., 2014), is no longer available (personal communication of Prof Chih-Wei Yang).

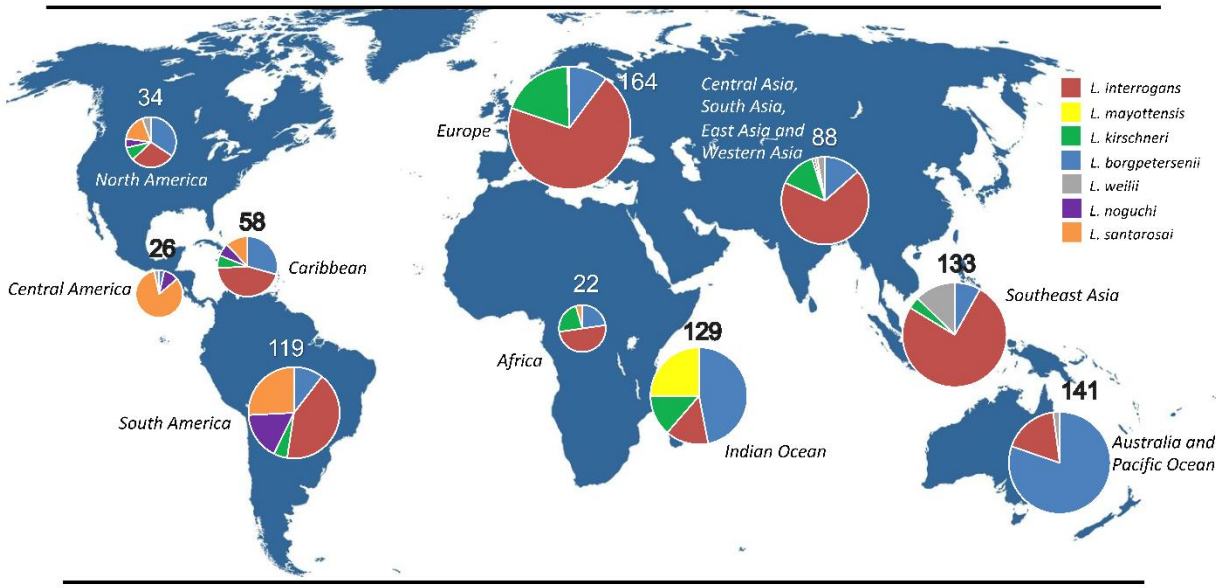

Figure 5.1. Geographic origins of the most frequent pathogenic *Leptospira* species in our genome database (n=914). Each pie chart corresponds to a given world region. As shown in our map, *L. santarosai* (n=64) is mostly found in America (North America, Central America, South America and the Caribbean islands). The base layer of the map is freely available from outline-world-map.com.

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

Phylogenetic analysis of the eight pathogenic *Leptospira* species using the saprophyte *L. biflexa* as the outgroup shows two distinct groups as previously shown (Xu et al., 2016; Vincent et al., 2019). One phylogenetic group constituted by *L. santarosai*, *L. mayottensis*, *L. borgpetersenii*, *L. alexanderi* and *L. weilii* (group I), and another one with *L. interrogans*, *L. kirschneri* and *L. noguchii* (group II) (**Figure 5.2A**). The genome size and GC content vary widely among pathogenic species usually correlating with these two phylogenetic subgroups (**Figure 5.2B** and **Tables S5.3** and **S5.4**). Notably, the genome size of the group I (3.96 ± 0.17 Mb) is significantly smaller to species from the group II (4.63 ± 0.32 Mb; Wilcoxon rank test, W = 4985, p < 2.2e-16; **Figure 5.2B**, left panel). Inversely, guanine+cytosine content (G+C%) is higher in the group I (40.4 ± 0.7 G+C%) than the group II (35.5 ± 0.43%; Wilcoxon rank test, W = 207668, p < 2.2e-16; **Figure 5.2B**, right panel). Interestingly, with the exclusion of *L. weilii*, the genome sizes were not significantly different between species of the group I (Dunn's test, all pair comparisons with p-adjusted > 0.05), supporting their characterization as a monophyletic group. However, the G+C% content did differ significantly between *Leptospira* spp. within group I (Dunn's test, all pair comparisons with p-adjusted value ≤ 0.04). Variation was larger within the species of group II, as *L. interrogans* shows a larger genome size than *L. kirschneri* (Dunn's test, p-adjusted ≤ 0.0000003), whereas G+C% was significantly different between *L. interrogans* and *L. kirschneri*, and *L. interrogans* and *L. noguchii* (Dunn's test, all pair comparisons with p-adjusted value ≤ 0.0002). These disparities in GC content and genome size may be a response to long-term niche adaptation of pathogens which emerged hundred million years ago at the same time as the appearance of mammals (Kurilung et al., 2019).

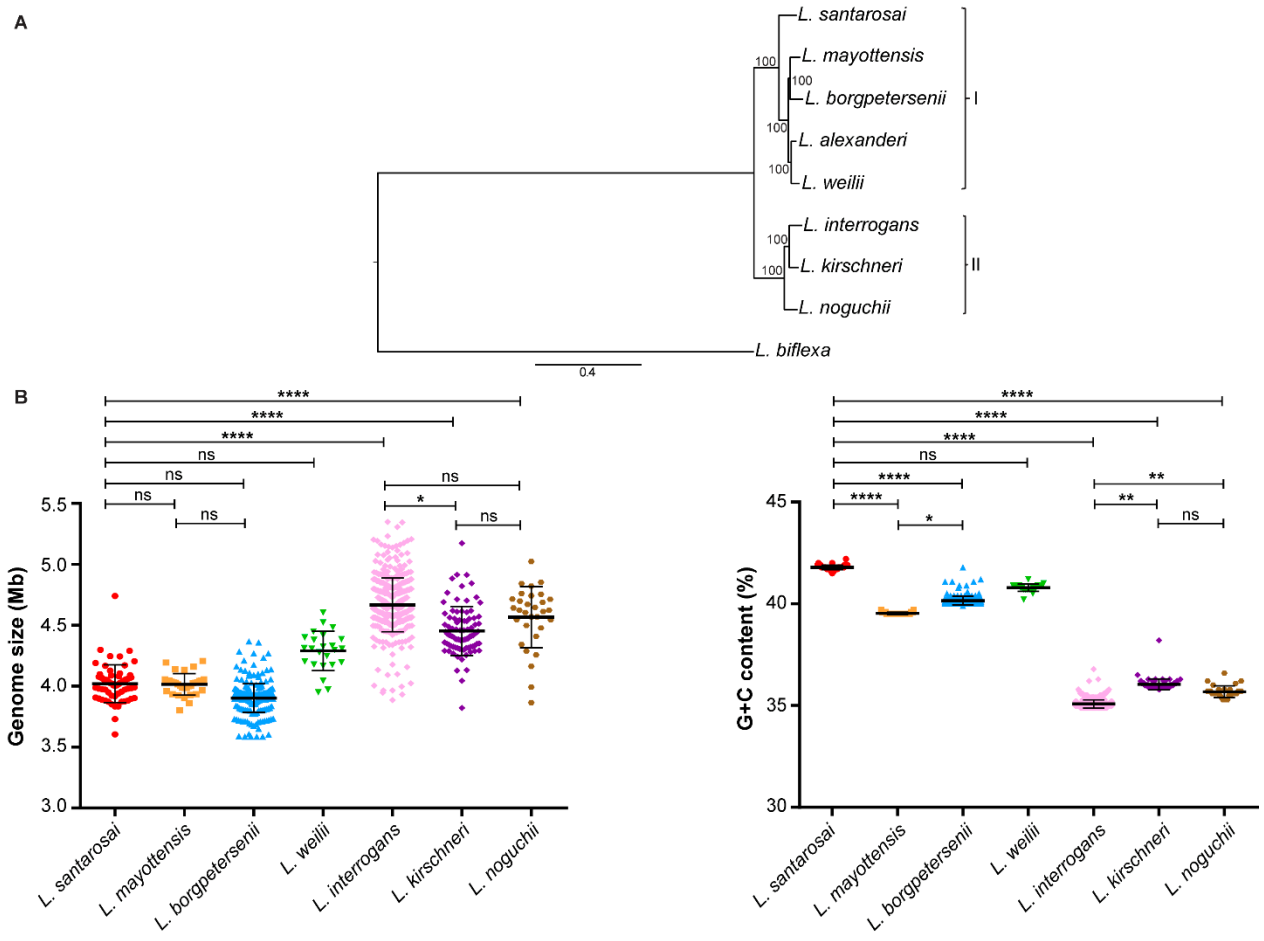

Figure 5.2. Genome characteristics of pathogenic *Leptospira*.

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

Most pathogenic species exhibit a strong enrichment (>3X) of accessory genes (**Figure 5.3**). However, it is important to notice that *L. mayottensis* stands as an exception, where the analysis of 33 strains showed a distribution of 1,949 and 2,284 genes constituting the core and the accessory genomes, respectively (1.2X enrichment), which probably resides in the fact that *L. mayottensis* is restricted to the Indian Ocean (Mayotte and Madagascar) and its specific adaptation to tenrec, described as the main reservoir of this pathogenic species (Bourhy et al., 2014; Lagadec et al., 2016). Concerning specifically *L. santarosai*, it harbors an open pangenome suggesting a

great diversity in its gene repertoire (**Figure S5.3**). Our analysis shows a strong enrichment (≈5X) of gene clusters that are unique to a maximum of two genomes (6,889) compared to core genome gene clusters (1,320) (**Figure 5.3**).

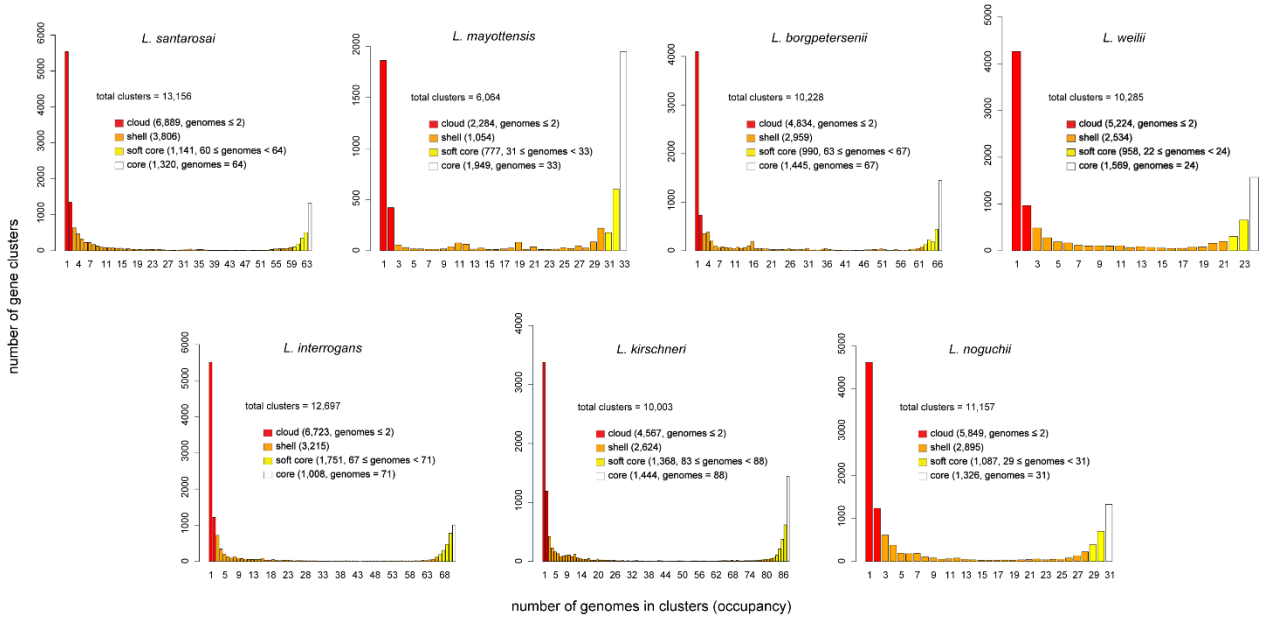

Figure 5.3. Pangenome distribution in four categories (cloud, shell, soft core and core) for *Leptospira* pathogenic species. Analyses done with GET_HOMOLOGUES showing the U-shaped distribution of pangenome from *L. santarosai, L. mayottensis, L. borgpetersenii, L. weilii, L. interrogans, L. kirschneri* and *L. noguchii.* For *L. interrogans* and *L. borgpetersenii*, a subset of representative genomes of all sizes and from all geographic locations was selected to reduce computational costs and avoid redundancy.

5.5.3 High genetic diversity of *L. santarosai* strains

To further investigate the genetic diversity of *L. santarosai* isolates, we used a core genome MLST (cgMLST) scheme (Guglielmini et al., 2019) (**Figure 5.4**). The species *L. santarosai* (n = 64) were divided into 55 cgMLST clonal groups (cgCGs) showing a high intraspecies genetic diversity (**Figure 5.4**) as shown in previous studies (Bourhy et al., 2013; Hamond et al., 2015; Peláez-Sánchez et al., 2017; Jaeger et al., 2019; Munoz-Zanzi et al., 2020). Among the 55 cgCGs, none is composed of more than 3 strains (**Table S5.1**), and none is composed of both human and animal strains. We cannot therefore identify transmission of *L. santarosai* clones between different hosts. There is a wide range of possible reservoirs for *L. santarosai* in the Americas. Some countries in the region are among the largest cattle producers in the world so these animals could

be important reservoirs for human infections. The Americas also exhibit a great biodiversity, so many species of wild animals such as rodents, marsupials, and domestic animals, such as dogs may be involved in transmission cycles. Among the 15 strains isolated from patients in Costa Rica, only two (id1256 and id1260) exhibit the same clonal group further confirming the great diversity of strains even within one small country.

Unfortunately, analyses to identify associations of *Leptospira* genotypes to particular epidemiological variables (host reservoir, disease outcome, etc.) cannot be performed with our small sample size. However, we could determine some phylogeographic lineages. A clear geographical separation of the clonal groups was observed for strains from (i) Central America, comprising isolates from Costa Rica (15; all human strains), Panama (6) but also one strain from South America (Colombia); South America which was further divided in two divergent groups, (ii) one containing strains from Brazil (10; mostly bovine strains) and the other (iii) including strains from Peru (9); and (iv) Caribbean islands, with strains from Guadeloupe (2), Martinique (2), Trinidad (1) and Puerto Rico (1) (**Figure 5.4**). This suggests ancestral presence of this species in these different countries and further separated evolution with no or low geographic diffusion. On the contrary, previous phylogenetic analyses of *L. noguchii* (Nieves et al., 2022) and *L. interrogans* (Guglielmini et al., 2019) did not reveal a correlation of genotype with geographical distribution.

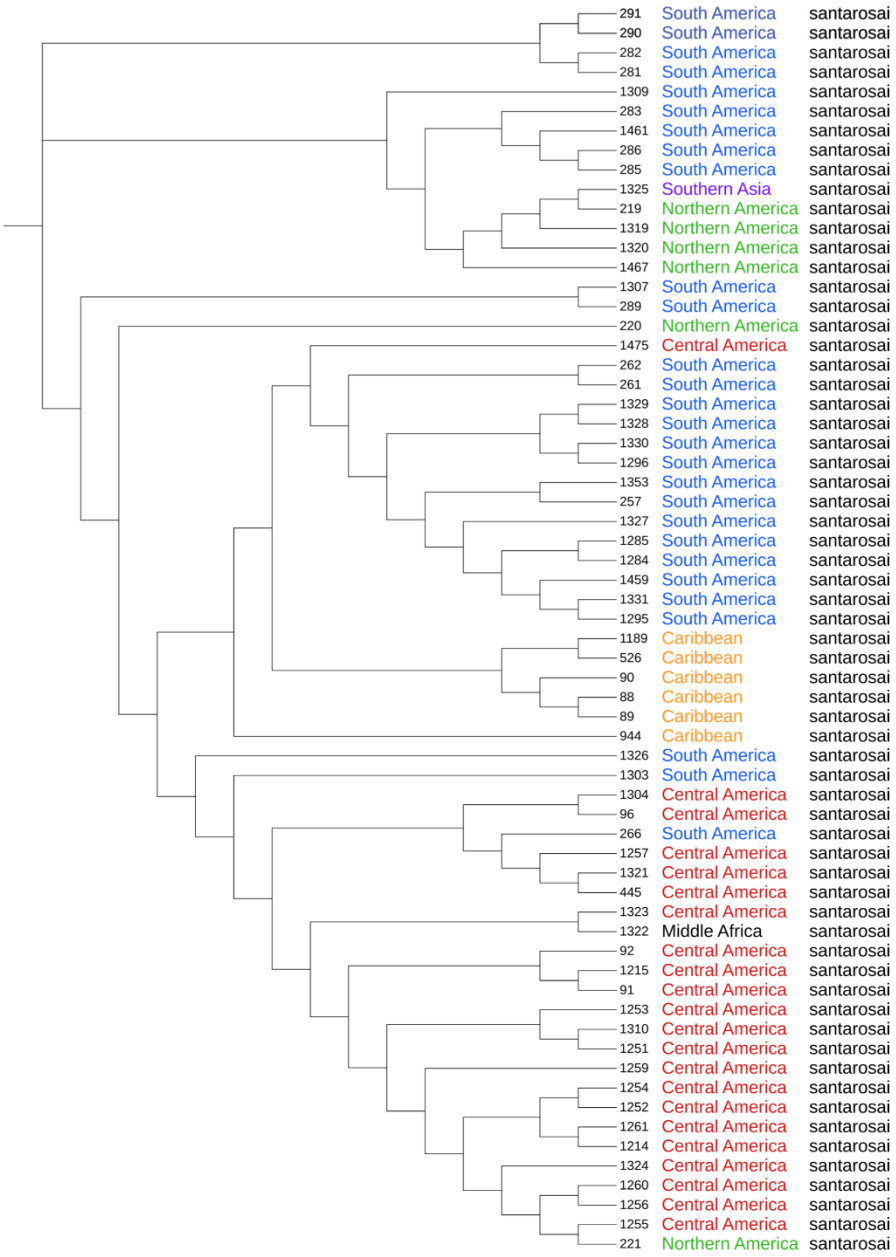

Figure 5.4. Phylogenetic tree of *L. santarosai* strains.

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

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

Further identification of *L. santarosai* strains was performed at the level of serovar and serogroup. Serogroups Grippotyphosa (8), Shermani (6), Tarassovi (8), Mini (5), Javanica (3), Pyrogenes (7), Sejroe (3), Hebdomadis (2), and Sarmin (2) were identified by serogrouping, but serovar/serogroup could not be assigned for 20 strains (no agglutination in serogrouping analyses or absence of serotyping information in the genome database). The three strains of the Javanica serogroup have been identified as belonging to a new serovar, Arenal, circulating in patients in Costa Rica (Valverde et al., 2008, 2013).

We recently showed that the analysis of the gene content of the LPS *O*-antigen–encoding cluster, or *rfb* cluster, correlates with *Leptospira* serovar and serogroup identity (Nieves et al., 2022). We then analyzed the gene composition of the *rfb* cluster of our set of *L. santarosai* strains, including 20 strains of unknown serogroup, in comparison to the *rfb* from reference strains of known serovars/ serogroups (**Figure 5.5**). Strains were organized after hierarchical clustering considering presence/absence of the *rfb* genes. As previously shown, well-typed reference serovars from the same serogroup share the same patterns (Nieves et al., 2022) but serovars from different serogroups may also share a similar genetic fingerprint: this is the case, for example, between serovars Copenhageni, Lai (both from serogroup Icterohaemorrhagiae) and serovar Canicola (serogroup Canicola), or serovars belonging to serogroups Grippotyphosa, Cynopteri and Autumnalis or between serovars from serogroups Mini, Sejroe and Hebdomadis. Four major clusters of genetic fingerprints can be distinguished (**Figure 5.5**) and were further confirmed using Jaccard's similarity index (**Figure S5.4**): cluster 1, containing reference strains from serogroups Australis, Grippotyphosa, Cynopteri and Autumnalis; cluster 2, which includes more variability of genetic fingerprints with reference strains from serogroups Pomona, Javanica, Pyrogenes, Celledoni, Icterohaemorrhagiae, Canicola and Sarmin; cluster 3, composed only of reference strains from serogroups Tarassovi and Shermani; and cluster 4, that comprises reference strains from serogroups Sejroe, Hebdomadis and Mini (**Figure 5.5**). The 64 *L. santarosai* strains were then assigned to the different cluster according to *rfb* gene composition (**Table S5.1**). Cluster 1 contains all *L. santarosai* strains identified as belonging to serogroup Grippotyphosa (id1259, id1257, id1296, id445, id1328, id1329, id1330, id1253), as well as four strains of unknown serogroups (id220, id1321, id286, id90) showing patterns related to Grippotyphosa. Interestingly,

id220, id1321 and id286 were isolated at three different geographic locations: North (id220), Central (id1321) and South America (id286). The three strains are phylogenetically distant, but exhibit an identical *rfb* cluster gene composition, suggesting that the *rfb* genomic island, important for bacterial persistence and adaptation to specific reservoir hosts, has disseminated in unrelated *L. santarosai* strains. Strains from the serogroups Sarmin (id1459, id1461), Javanica (id91, id1310, id92; all from the new serovar Arenal recently described in Costa Rica), Pyrogenes (id1307, id1260, id1475, id1215, id1256, id526, id1189) belong to Cluster 2. Cluster 2 also includes two strains (id285 and id257) from an unknown serogroup exhibiting Sarmin/Pyrogenes/Javanica-like patterns. Cluster 3 comprises all *L. santarosai* strains belonging to serogroups Shermani (id96, id261, id262, id1304, id1252, id1255) and Tarassovi (id266, id289, id1295, id1303, id1467, id1254, id1214, id1261), along with strains of unknown serogroups (id1323, id1331, id282, id1320, id1324) showing patterns related to serogroups Tarassovi and Shermani and one strain (id944) with many *rfb* genes not present in other genomes, suggesting that it may belong to a new serovar/serogroup or a serovar/serogroup not included in our analysis. Finally, strains from cluster 4 belong to the antigenically related serogroups Mini, Hebdomadis and Sejroe which share a similar *rfb* gene composition (Medeiros et al., 2022). *L. santarosai* strains from cluster 4 share similar genetic fingerprints and belong to serogroups Mini (id1325, id1319, id219, id89, id88), Sejroe (id1327, id1285, id1284), and Hebdomadis (id1251, id1322) (Carvajal and Fagerstrom, 2017), in addition to strains of unknown serogroups (id1309, id221, id1353, id281, id283, id1326) and strains id290 and id291 with more variability of genetic fingerprints.

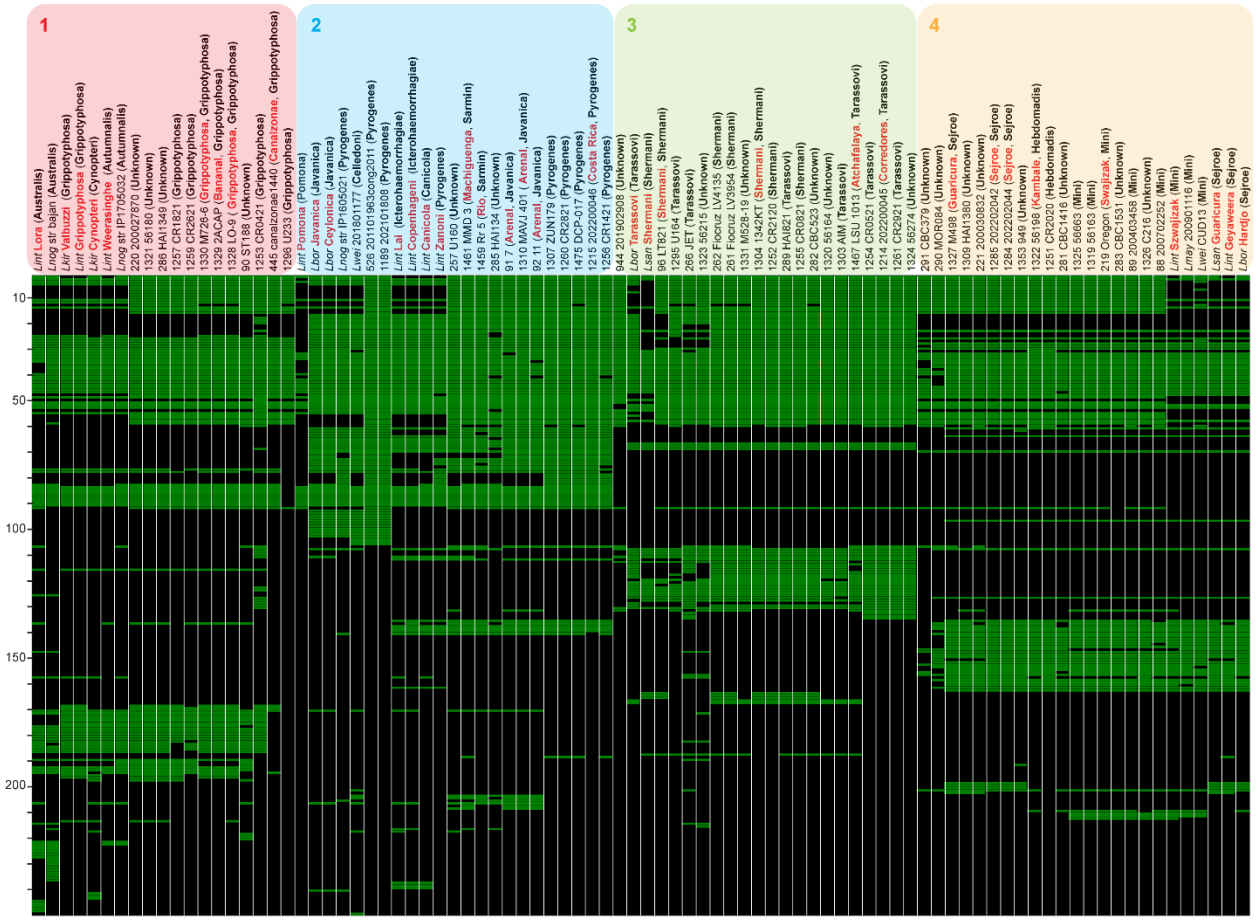

Figure 5.5. Gene presence/absence matrices of *rfb* clusters from different *Leptospira* strains and species, covering a range of distinct serogroup/serovar identities compared to *L. santarosai* strains.

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

Our analysis of the gene composition of the *rfb* cluster is consistent with previous studies showing that *L. santarosai* only contains serovars from serogroups Shermani, Hebdomadis, Tarassovi, Pyrogenes, Autumnalis, Bataviae, Mini, Grippotyphosa, Sejroe, Pomona, Javanica, and

Sarmin (Levett, 2001), as well as probable new serovars. However, it must be noted that most of the *L. santarosai* genomic assemblies analyzed here are fragmented (average contig number, 149), which could lead to inaccurate serotype assignations. Further studies should include less fragmented genomes, ideally closed genomes, for a correct interpretation of the *rfb* patterns and to better identify the genes determining the serovar identity. As previously shown (Guglielmini et al., 2019) most serogroups had a polyphyletic distribution. Thus, isolates from the serogroups Tarassovi and Grippotyphosa did not all cluster together in the phylogenetic tree based on either cgMLST alleles (**Figure 5.4**) or core-genes (**Figure S5.5**).

5.6 Conclusion

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

Acknowledgements

We thank Vincent Enouf and the team of the core facility P2M (Institut Pasteur, Mutualized Platform for Microbiology) for genomic sequencing and the staff of the French National Reference Center for Leptospirosis for processing some of the samples. We thank Federica Palma for her help with BIGSdb. We thank Dr Howard Takiff and Lizeth Caraballo (Laboratorio de Genética Molecular, CMBC, Instituto Venezolano de Investigaciones Científicas, Caracas, Venezuela), Anissa Amara Korba (Institut Pasteur of Alger, Algiers, Algeria), Prof Walter Lilembaum (Universidade Federal Fluminense, Rio de Janeiro, Brazil), Sabrina Rodrigues Campos (Institute of Veterinary Bacteriology, Bern, Switzerland), Dr Paulina Meny (Instituto de Higiene, Facultad de Medicina, Montevideo, Uruguay), Dr Josipa Habus and Dr Nenad Turk (Faculty of Veterinary Medicine, University of Zagreb, Croatia), Prof Jackie Benschop (Molecular Epidemiology and Public Health Laboratory, School of Veterinary Sciences, Massey University, New Zealand), Dr Zarantonelli (Unidad Mixta Pasteur-Instituto Nacional de Investigación Agropecuaria, Institut Pasteur de Montevideo), Dr Vasantha Kumari Neela (Faculty of Medicine and Health Sciences, Universiti Putra Malaysia, Malaysia), Dr Mohammad Ridhuan Mohd Ali (Institute for Medical Research, Malaysia), Dr Shlomo Blum (Kimron Veterinary Institute, Israel), Dr Nikolai Tokarevich (Laboratory of Zoonoses, Pasteur Institute in Saint Petersburg, Saint Petersburg, Russia), Dr Angelica Delgado (National Reference Laboratory of Bacterial Zoonoses, Lima, Peru), Dr Marga Goris (National Collaborating Centre for Reference and Research on Leptospirosis, Academic Medical Center, Amsterdam, the Netherlands) and Prof Chih-Wei Yang (Chang Gung University, Taiwan) for sharing strains and/or information on some of the strains used in our study.

Chapter 6. Closing remarks

6.1 General discussion

In recent years, the identification and characterization of new species has led to a refinement of the *Leptospira* phylogeny. Five clusters including non-pathogens and pathogens with varying degrees of virulence are now clearly discerned. Comparative genomics analyses have identified gene determinants associated with the occurrence of pathogenesis, but a comprehensive analysis covering all species within the genus has not been performed yet. The aim of this thesis was to characterize the extensive variability encountered in *Leptospira*. Starting by exploring differences in the genetic repertoire that accompanied the emergence of pathogens; then continuing with determinants of serovar identity, which is often associated with host-preference; and finally focusing on intraspecies phylogenomics and the differences between highly virulent species that would explain specific adaptations. Of note, *L. sanjuanensis* (Fernandes et al., 2022) was the only species not included, as it was reported after completion of the analyses presented here. Based on the results that were obtained, key points within the three major subjects aforesaid will be discussed below.

What genetic events determined the evolution towards a virulent phenotype?

The analysis performed in Chapter 3 correlates well with previous reports (Xu et al., 2015), and points to a larger loss of genes rather than acquisition during pathogen emergence. The missing genes were involved in biosynthetic pathways and transport, which seems to be a repeated pattern in reductive evolution (Ochman and Moran, 2001) towards the appearance of an obligate pathogen. The wealth of metabolic intermediates in the host triggers molecular mechanisms in the microorganism that favour the elimination of genes that become useless and are not essential for microorganism survival (Ochman and Moran, 2001). So far, the only strictly host-dependent species within the genus is *L. borgpetersenii*. Comparative genomics has suggested that genome decay in *L. borgpetersenii* is the result of IS-mediated events and would be the main cause for its speciation (Bulach et al., 2006). This genome reduction rendered a diminished representation of genes related to metabolic and transport functions when compared to *L. interrogans* (Bulach et al.,

2006), consistent with the findings here. Eventually, the shaping of pathogenic genomes driven by reductive evolution stabilizes and no longer reduces, but lost genes are not recovered, hence, becoming obligate pathogens (Ochman and Moran, 2001). Furthermore, gene loss has been associated to appearance of species with an enhanced virulence. In the genus *Rickettsia* (Diop et al., 2018), Gram-negative bacteria, highly virulent species showed larger genome reduction than that observed in less virulent or non-pathogenic species. Moreover, few or no acquired genes were detected in the most virulent compared to closely related phenotypes. Such is the case for *Leptospira.* Here, only five genes were distinguishing when strictly comparing $P1^+$ and $P1^-$ species. Taken together, all this hints at an ongoing genome reduction in $P1^+$ species that will eventually stabilize as it did in *L. borgpetersenii* ($P1^+$), and ultimately result in an entirely host-associated lifestyle, as already suggested (Vincent et al., 2019).

On the other hand, gene acquisition also plays a role in the appearance of pathogenic *Leptospira.* Acquired genes included those encoding lipoproteins, or proteins involved in signaling, or participating in the response to oxidative stress. All of which perform pivotal functions in host interaction and virulence. The relevance of some of the genes here identified, particularly those associated with the emergence of highly virulent *Leptospira*, has been assessed through both *in vitro* and *in vivo* experiments by our collaborators. The general approach involved introducing *L. interrogans* ($P1^+$) genes into *L. adleri* and *L. dzianensis* (currently *L. yasudae*) (both $P1^-$). For instance, the insertion of the lipoprotein LA_2020 (LIMLP_09380) has been found to play a pivotal role in bacterial survival. Following a 2-hour incubation with normal human serum, *L. adleri* and *L. dzianensis* strains with LA_2020 insertion exhibited an approximate 50% increase in survival compared to their wild-type counterparts. Furthermore, evaluation of Membrane Attack Complex (MAC) deposition through immunofluorescence revealed a reduction of around 15% and 30% for knock-in strains of *L. adleri* and *L. dzianensis,* respectively. Collectively, these findings suggest the involvement of LA_2020 in immune evasion mechanisms, preventing bacterial osmolysis and promoting survival. Another illustrative example is provided by the collagenase LA_0872 (LIMLP_03665). Knock-in strains showed enhanced colonization efficiency within hamster tissues, even as early as 1-day post-infection, when compared to wild-type strains. The implication of this collagenase in *L. interrogans* virulence has been previously reported. A *L. interrogans ΔcolA* strain showed a notable decrease in lung, liver, and kidney colonization, along with attenuated virulence in hamsters. Additionally, it displayed impaired transcytosis in HEK293 and HUVEC

monolayers (Kassegne et al., 2014). Finally, a gene that encodes a hypothetical protein (LA_1402/LIMLP_11655) belonging to the PF07598 family, which includes virulence-modifying (VM) proteins (Chaurasia et al., 2022), has been found to play a significant role in NFκB activation. Inserting *L. interrogans* LA_1402 into *L. adleri* and *L. dzianensis* led to a decrease in the expression of NFκB-regulated genes in infected human macrophages (measured at 6 hours post-infection), comparable to the levels observed after infection with *L. interrogans*. This positions LA_1402 as key player in immune evasion, promoting lower levels of inflammation and diminishing the recruitment of immune cells. Among the acquired genes resulting from the analysis presented in Chapter 3, a significant proportion consisted of hypothetical proteins, which may even mask new mechanisms of pathogenicity, as suggested (Adler, 2015). Previous studies have pointed to specific TCS acquired by pathogenic *Leptospira* which showed homology with genes of *Legionella pneumophila, Yersinia pestis* and *Mesorhizobium loti,* supporting events of lateral gene transfer with other pathogens (Xu et al., 2015). However, these TCS were not found in the present study, suggesting that, although present in the species analyzed by Xu and collaborators, are likely not ubiquitous in all the pathogenic species. Among the mechanisms facilitating gene gain in pathogenic *Leptospira*, HGT and gene duplication have shown a significant implication (Thibeaux et al., 2018; Xu et al., 2015). The number of paralogs predicted in $P1^+$ here, supports these previous findings, yet their representation varies between species. Redundant genes can be subject to independent selective pressures, develop differential functionalities, and overall provide specific advantages (Ghosh and O'Connor, 2017). Interestingly, there are four species that outnumbered paralogs: *L. weilii, L. borgpetersenii, L. alexanderi* and *L. noguchii.* Besides *L. weilii,* which has been found in multiple hosts (Corney et al., 2008; Javati et al., 2022; Kurilung et al., 2021; Safiee et al., 2022), and environmental samples (Kurilung et al., 2017; Mason et al., 2016), both *L. alexanderi* and *L. noguchii* are largely associated with hosts (Brenner et al., 1999; Monroy et al., 2021; Silva et al., 2007; Silva et al., 2009). While only *L. borgpetersenii* has been described as an obligate pathogen, it could be that gene duplication is facilitating the shift towards host dependence in the other species as well; indeed, overrepresentation of paralogs involved in signal transduction and defense mechanisms among $P1^+$ may represent specialization in survival within the host. The impact of paralogy on bacterial virulence has been proven before. For example, the *virB6* gene cluster in *Anaplasma phagocytophilum* (obligate intracellular tick-borne bacteria) encoding a type IV secretion system, present four paralogs of *virB6* that are transcribed during infection.

Transposon interruption of just one of them, resulted in downregulation of all four and reduced proliferation in human and tick cells (Crosby et al., 2020). Several genes encoding for leptospiral flagellar proteins and potentially linked to virulence have experienced duplication. Leptospiral FliG from the C-ring, for example, presents three paralogs and interruption of one of them (70% homolog of *fliG2* in *B. burgdorferi*) leads to a decrease in bacterial motility (Fernandes et al., 2019). FlaA and FlaB from the flagellar filament also have multiple copies, some of which directly affect virulence, such as *flaA2* (Lambert et al., 2012). That said, the results here along with previous evidence position paralogy as one of the possible mechanisms underlying virulence gain in the appearance of $P1^+$ species in *Leptospira.*

Lastly, amino acid permutations linked to the emergence of a virulent phenotype were also numerous. The number of protein-coding genes involved in signal transduction harbouring mutations hints at a profound restructuring of signaling pathways that could directly affect virulence. In *P. aeruginosa* a few amino acid substitutions in the histidine kinase *bfmS* of the BfmRS were implicated in the phenotypic change from highly virulent to reduced virulence strains, with a clear reduction in the occurrence of quorum sensing and increase of biofilm production (Cao et al., 2020). More importantly, these mutations were found in *P. aeruginosa* strains from cystic fibrosis patients (Cao et al., 2020), suggesting a clear contribution to the establishment of chronic infection. One of the leptospiral TCS identified as mutated here, previously showed association with chemotaxis, ultimately resulting in reduced motility (Lambert et al., 2015). Although no connection with virulence was determined under intraperitoneal infection conditions, its impact of virulence is not ruled out. Given that the mutations identified here are close to the sensor domain of the HK and the effector domain of the RR, an effect on signal detection and transduction could be suggested. Coincidentally, both HK and RR are 2-fold downregulated in $P1^-$ strains (collaborative project, data not shown), which could also indicate that these mutations are affecting its expression and overall functionality. While the system may still be functional, the reduced expression could impede its capacity to process signals effectively. Another interesting example of amino acid permutations distinctly found in $P1^+$ species is provided by PerRA (LA_1857), which plays a role in the response to oxidative stress. RNAseq analysis within the same collaborative project mentioned previously, revealed differential expression of the PerRA regulon between $P1^-$ and $P1^+$ species, aligning with an augmented resistance to peroxide in the latter. Previous studies have demonstrated that PerRA acts as a metal-dependent repressor of gene expression under

normal conditions (Lo et al., 2010; Zavala-Alvarado et al., 2021), since its inactivation led to upregulation of several genes, such as *katE,* which encodes a catalase. Sequence analysis of PerRA in $P1^+$ and $P1^-$ species revealed a mutation at position 89, which is located within the DNA binding site and metal coordination region of PerRA. $P1^+$ species possess a histidine residue (H89), while the majority of $P1^-$ species have an asparagine (N89), except for *L. alstonii* (H89), and *L. ainazelensis* (glycine, G89). Initial amino acid permutation analysis conducted in Chapter 3 did not identify LA_1857 among the candidates. However, after adjusting the search parameters and excluding *L. alstonii* and *L. ainazelensis,* it was detected, linking the mutation with the emergence of a highly virulent phenotype. Genetic engineering experiments performed by our collaborators showed that *L. interrogans ⊿perRA* complemented with i) $perRA_{P1+}$, ii) $perRA_{P1+\text{-}H89N}$ or iii) $perRA_{P1-}$ did not exhibit differences in *perRA* expression. However, mutant (ii) exhibited a 1.4-fold decrease in catalase activity, while mutant (iii) displayed a 5-fold decrease in presence of hydrogen peroxide compared to mutant (i). These results highlight the partial contribution of the identified mutation to the overall function of the repressor in response to oxidative stress. Nevertheless, it is clear that other features of $P1^-$ PerRA are influencing the regulation of gene expression in PerRA regulon, warranting further investigation.

LPS and serovar: largely defines host specificity?

Historically, LPS and specifically the *O*-antigen composition, has been associated to serovar determination. Since serovar identity can be shared by different species, HGT of genes included in the *rfb* cluster has been suggested (Haake et al., 2004), although never extensively explored. Here, a specific genetic fingerprint of *rfb* genes showed connection with serovar identity (**Figure 4.5**). Moreover, antigenically related serovars, i.e., from the same serogroup, clustered together based on shared protein-coding genes. Interestingly, some serovars within specific serogroups were not distinguished by gene pattern, suggesting that other specific events rather than gene composition, could also establish serovar identity (such as subtle mutations in the genes). That has been evidenced in serovars Copenhageni and Icterohaemorrhagiae which differ by a single indel frameshift in an LPS-related gene (Santos et al., 2018). Analysis of different *rfb* clusters revealed strong indications of HGT events, with a decrease in GC content within the cluster, as well as genomic rearrangement in the flanking regions (**Figure 4.4**). Horizontal transfer of *O*-

antigen-related genes is a common phenomenon among Gram-negative bacteria, and is responsible for the diversity encountered within this biosynthetic cluster (Samuel and Reeves, 2003). Such transfer has been evidenced, for example, between *Shigella* strains and *E. coli* (Pupo et al., 2000) as well as other species, illustrating the genomic plasticity of this region. Performing a blast search of the 12 *L. noguchii rfb* nucleotide sequences presented in Chapter 4, a high coverage and identity with other leptospiral *rfb* from pathogenic species, mainly *L. kirschneri, L. santarosai* and *L. interrogans,* was observed. Overlapping regions with other serovars were also identified, indicating the existence of more widely distributed genes, as shown in the presence/absence matrices (**Figure 4.5**). By excluding *Leptospira* from the blast searches, regions with minimal coverage but high identity matched other bacteria. Predominantly the same bacterial species or genus were found among the top hits for the 12 *rfb* analyzed. These included *Enterobacter asburiae,* originally found in human urine and respiratory sources, among others (Brenner et al., 1986); *Staphylococcus chromogenes,* isolated from dairy cows with mastitis (dos Santos et al., 2016); *Buttiauxella ferragutiae,* isolated from soil and water (Muller et al. , 1996), but also from wild ruminants (Lauková et al., 2018); *Bacteroidetes,* that represent a quarter of the bacterial population of the human colon (Wexler, 2007); *Tenacibaculum dicentrarchi*, sourced from sea bass (Piñeiro-Vidal et al., 2012); *Legionella lansingensis,* first isolated from the human lower respiratory tract (Thacker et al., 1992); *Anoxybacillus gonensis,* found in environmental samples (Belduz et al., 2003); and *Aeribacillus pallidus,* closely related to *Anoxybacillus,* and obtained from hot springs (Filippidou et al., 2015). Most of these bacteria are strongly associated with hosts in which *Leptospira* has been reported. More importantly, *Leptospira* has been isolated or detected in urine, lower respiratory tract and even milk from cattle, demonstrating that the microenvironment offered by each host propitiates the encounter with certain symbionts, commensals or pathogens, and facilitates HGT events. HGT with bacteria primarily found in soil and water are also evident, but not unexpected since most of pathogenic *Leptospira* can survive in the environment as part of the transmission cycle (Ko et al., 2009). In contrast, blast searches of the *L. biflexa rfb* excluding *Leptospira,* yielded sequences of low coverage coming from *Chryseobacterium glacei,* isolated from glacier surfaces (Pal et al., 2018); *Olleya sp.,* a marine bacteria that has also been recovered from fishes and green alga (Nedashkovskaya et al., 2017); *Flavobacterium sediminilitoris,* recovered from tidal flat (Park et al., 2018); and *Leptolyngbya boryana,* a cyanobacteria found in freshwater (Shimura et al., 2015). All these with environmental niches where *L. biflexa* of a free-living lifestyle circulates.

Surprisingly, when performing blast against *Leptospira*, a small region encompassing a glycosyltransferase family 2 protein towards the 3' of the *rfb* showed high identity with pathogenic *Leptospira*, clearly demonstrating co-habitat of pathogenic and free-living species. Overall, although there were very small regions with high identity towards other bacterial genera, leptospiral *rfb* experiences HGT almost exclusively within the genus. Genes related to *O*-antigen biosynthesis were probably acquired from bacteria circulating in the environment a long time ago and have evolved in response to the different niches faced by *Leptospira*, leaving no clear trace of a common ancestor.

As for associating the serovars with specific hosts, the results obtained here were inconclusive. Although most of the *L. noguchii* strains analyzed in Chapter 4 were not characterized at the serovar level, there were strains sharing serogroup isolated from different hosts. Likewise for strains harboring an almost identical *rfb*, both in synteny and in presence/absence *rfb* gene pattern (**Figure 4.3** and **4.5**). However, *L. noguchii* might not be the best model to demonstrate host-preference given its high adaptability, as already discussed. Host preference is also defining specific disease outcomes. For example, serovar Hardjo is frequently associated to chronic infection (asymptomatic) in cattle (Adler and de la Peña Moctezuma, 2010), which act as maintenance carrier facilitating bacterial dissemination through the urine. Symptomatology is an important aspect to consider when evaluating serovar-host associations. Most of the strains sequenced within the framework of this thesis were obtained from cases of acute leptospirosis. This is associated with non-adapted serovars. Unfortunately, clinical information is lacking for the other isolates included in the *L. noguchii* phylogeny. This should be improved and considered for similar studies in the future. In addition to the investigation of molecular aspects underlying serovar host-adaptation, other bacterial features have been explored. In recent reports, leptospiral adhesion and crawling were evaluated in kidney epithelial cell lines from different hosts and correlated with the disease outcome, i.e., asymptomatic *vs* severe leptospirosis, ultimately related to host adapted *vs* non-adapted serovars, respectively. *L. interrogans* serovar Manilae showed less adhesion/crawling in epithelial cells from rat (asymptomatic) compared to human (severe outcome) cells (Xu et al., 2020). Non-pathogenic *L. biflexa* showed even less adhesion/crawling, further supporting the idea of crawling as a determinant for pathogenesis. Since LPS is postulated a major adhesin potentially involved in crawling (Xu et al., 2020), and it has been shown that, for example, *L. biflexa* has a shorter LPS compared to *L. interrogans* (Raddi et al., 2012), it is tempting to postulate that the

amount of LPS is directly proportional to crawling, and therefore to disease severity. However, infection of guinea pigs (acute model) and rats (chronic) both with *L. interrogans* serovar Copenhageni showed reduced levels of *O*-antigen in the acute model of infection (Nally et al., 2005). Taken together, it suggests a more complex remodeling of the leptospiral outer membrane; it could be that LPS levels are regulating the expression of other adhesins, or simply exposing hidden pathogen-associated adhesins on the surface that define the interaction with the host.

How to interpret genetic variability in highly virulent species?

In Chapter 4 and 5, intraspecies variability of two $P1^+$ members, *L. noguchii* and *L. santarosai*, respectively, was evaluated. *L. noguchii* strains showed more variability in terms of ANI (**Figure S4.1**, **Appendix II**), ranging between 97-100%, whereas *L. santarosai* strains were far more indistinguishable, with virtually all of them sharing about 100% identity (**Figure S5.1**, **Appendix III**). Such variability could be a mere consequence of the host repertoire of the strains analyzed here. In the case of *L. noguchii*, although there were several isolates from cattle, there was more host diversity, hinting at the high adaptability of this species. As for *L. santarosai*, even though the strains came from several hosts, 60% of the isolates were obtained from humans, and secondly from cattle (15%), suggesting a possible association between *L. santarosai* infection and farming activities. However, the serogroups circulating in the same geographical areas within farm animals and humans were diverse and would not support a strong zoonotic transmission of *L. santarosai* (**Figure 5.4**). On the contrary, the phylogeny of *L. noguchii* showed that restricting by geographic location, strains from the same serogroup circulate between dogs and cattle, or cattle and humans, for example (**Figure 4.2**). The protein repertoire was also variable across strains of both species and two separated clusters were observed (**Figure S4.2** and **S5.2**, **Appendix II** and **III**, respectively). *L. noguchii* strains from serogroup Australis infecting amphibians and human were separated from the rest, perhaps denoting specialized repertoires for infection and transmission between these hosts. The same applied to *L. santarosai*, with a cluster that included domestic/farm animal and human strains. Hence, although not demonstrated by the phylogeny, the zoonotic aspect in *L. santarosai* might be there, but was not reflected in the dataset analyzed here. Another aspect to note is the larger representativeness of serogroups in *L. santarosai* (**Figure 5.4** and **5.5**) compared to *L. noguchii* (**Figure 4.5**), which somehow could be related to host-preference.

In other words, serogroups identified in *L. noguchii,* as shown in the phylogeny, may not display host specificity, implying enhanced adaptability to jump from one host to another. In contrast, serogroups found in *L. santarosai* were different between hosts, which may reflect serovars/serogroups constrained to infection of specific hosts. However, a larger dataset and further identification of unknown serogroups are needed to test this hypothesis.

When compared to other $P1^+$ strains, two clusters became evident. One containing *L. santarosai, L. mayottensis, L. borgpetersenii* and *L. weilii* (hereafter referred as Group I), and the other comprising *L. interrogans, L. kirschneri,* and *L. noguchii* (henceforth termed Group II). The same clustering is observed when considering genome size and GC content. Group I showed reduced genome size and higher GC content compared to Group II (**Figure 6.1**). Reduction in genome size may be representing the transition to obligate symbionts, as occurred in *L. borgpetersenii* (Bulach et al., 2006), which happens to be in Group I. In the reductive evolution experienced by obligate pathogens, there is propensity to cumulate pseudogenes and mobile elements that upon prolonged host-pathogen association are prone to removal, especially in genes no longer required in the nutritionally rich host environment (Bobay and Ochman, 2017). Conversely, bacteria with larger genomes often harbor more accessory genes, ultimately reflecting ecologically diversified niches. However, overrepresentation of accessory genes occurred in almost all $P1^+$, not only in Group II (**Figure 5.3**). The scattered genome size of *L. interrogans, L. kirschneri* and *L. noguchii* strains may be revealing an ongoing reductive evolution within each population, not yet stabilized. Note that some genomes within this cluster have reached a genome size comparable to those in *L. borgpetersenii* (**Figure 5.2**). On the other hand, the GC content has also been positively correlated with optimal growth temperatures (Musto et al., 2004). In line with the idea of an ongoing transition to obligate pathogens in which Group I is more advanced, a lower GC content in Group II was anticipated, and consistent with the importance that the environment still has in their transmission cycle (water and soil may have lower temperatures than a host). In fact, the average GC content in non-pathogenic *Leptospira* is lower than in pathogenic members (Vincent et al., 2019).

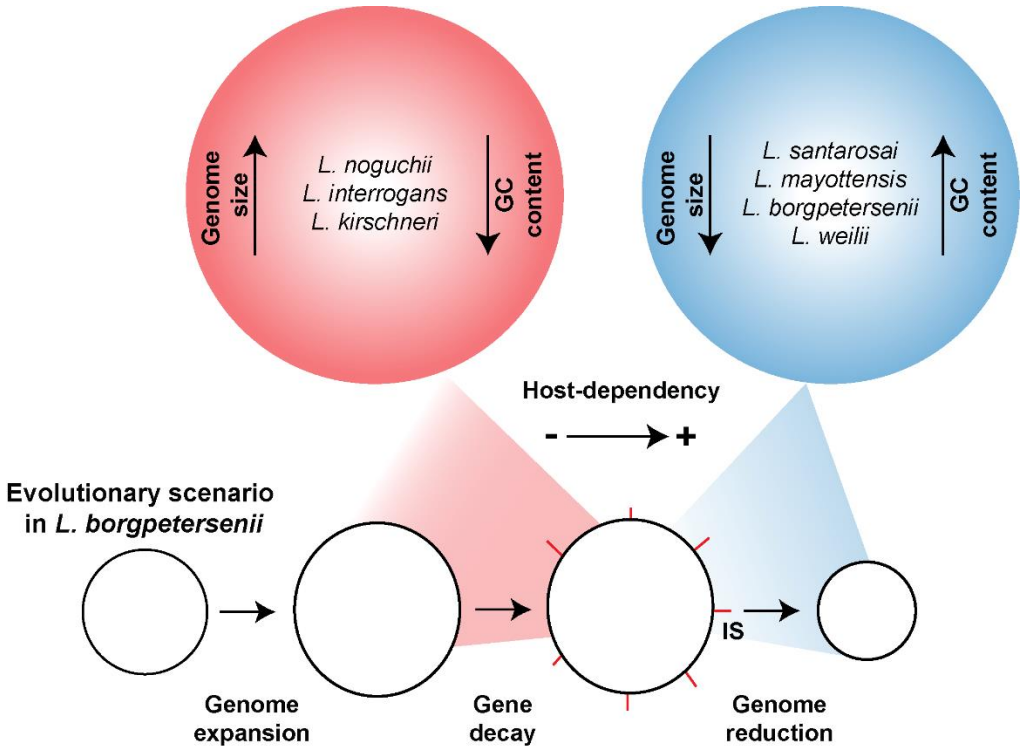

Figure 6.1. Genome features and evolution towards host-dependency in P1^{+} species.

Difference in sizes between both groups may reflect two distinct stages in evolution, following the evolutionary model of *L. borgpetersenii*. On one hand, the group including *L. noguchii*, *L. interrogans* and *L. kirschneri*, with a larger genome size and having already undergone the initial expansion, could be situated in the gene decay phase. In fact, the results presented here support the loss of metabolic genes in these species. Moreover, the lower GC content may still indicate an alternation between host-association and free-living stages, as discussed in the main text. In contrast, the group including *L. santarosai*, *L. mayottensis*, *L. borgpetersenii* and *L. weilii*, with a smaller genome size, could be transitioning towards complete host-dependency. *L. borgpetersenii* is the only one that has reached a stable level of genomic reduction and strict host dependence.

Differences in GC content may also indicate differential codon usage (Nishio et al., 2003). Indeed, a minor bias in codon usage in Group II turns into manifest (**Figure 6.2**). For example, this cluster uses more TTA for leucine, ATT for isoleucine, and TCT for serine, which differ from the codons used in Group I. Even *L. santarosai* exhibits some differences, suggesting genomic peculiarities that have not yet been fully addressed. In any case, since codon bias certainly has a role in multiple cellular processes, including transcription and translation (Parvathy et al., 2022), a deeper study of *Leptospira* codon usage could facilitate the understanding of ongoing evolution, and even provide insight into specific adaptations.

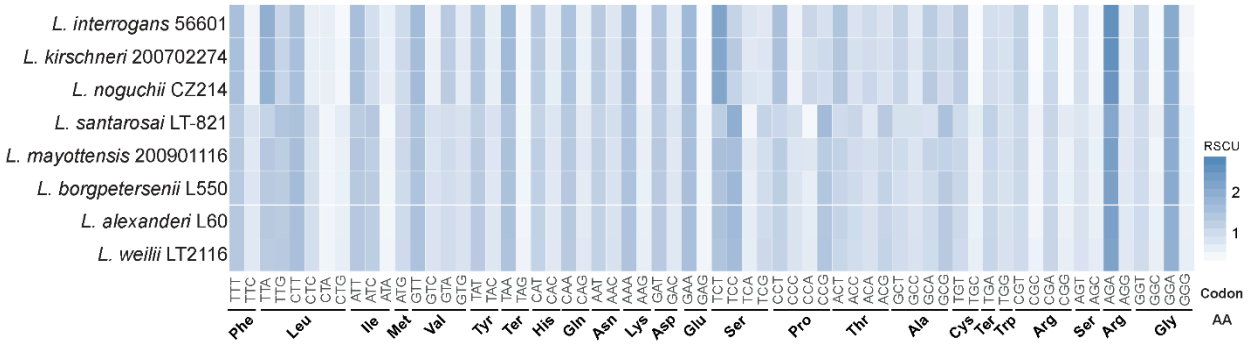

Figure 6.2. Codon usage in *Leptospira* P1^{+} species.

Relative synonymous codon usage (RSCU, scale at the right side of the matrix) is the ratio of the observed codon frequency compared to the expected frequency considering an equal use of codons for the same amino acid. Amino acids associated to each codon are indicated by their three-letter code (Phe=phenylalanine; Leu=leucine; Ile=isoleucine; Met=methionine; Val=valine; Tyr=tyrosine; Ter=termination; His=histidine; Gln=glutamine; Asn=asparagine; Lys=lysine; Asp=aspartate; Glu=glutamate; Ser=serine; Pro=proline; Thr=threonine; Ala=alanine; Cys=cysteine; Trp=tryptophan; Arg=arginine; Gly=glycine).

6.2 Conclusions and perspectives

In association with the emergence of pathogenic *Leptospira*, several pathogen-specific determinants were identified, as well as a few of them strictly linked to an increased virulence phenotype, exclusively present in $P1^+$ species. Genes that were lost in the appearance of pathogenic *Leptospira* were also determined, with enrichment in those related to metabolism. More importantly, besides gene acquisition and loss, amino acid substitutions associated with a virulent phenotype were discerned for the first time, including several response regulators previously linked to virulence.

Undoubtedly, the 68 species presented here are only a part of the existing diversity in *Leptospira*, and an update may be needed sooner than expected. An aspect not yet widely studied is the impact of extrachromosomal elements on *Leptospira* evolution. Few reports mentioned in Chapter 2, evaluated bacteriophages and plasmids in *Leptospira*, but not exhaustively. Even in Chapter 4, *L. noguchii* plasmids were compared against other *Leptospira* plasmids. Some plasmid exchangeability was apparent but needs further exploration. To do so, obtaining more closed genomes is non-negotiable. Another question that emerges from the results obtained here and that deserves further study is the extent to which the amino acid substitutions identified in TCS are involved in virulence. Taking advantage of ongoing collaborations, the introduction of mutations found in $P1^-$ into *L. interrogans* ($P1^+$) could be attempted to evaluate their effect in an acute model of infection, such as hamster.

As for serovar identity determinants, a univocal link with the presence/absence of specific genes in the *rfb* cluster was here reported for the first time. Different species belonging to the same serovar and showing an identical gene pattern supports horizontal gene transfer, which was also backed up by other analysis. This HGT phenomenon largely occurs among species of *Leptospira* genus, contributing to its diversity and adaptability. On the other hand, serovar host preference could not be established in the model of choice for the study, due to the great adaptability of *L. noguchii*. Co-occurrence of same serogroup infecting human and other mammals in a given location supports a non-host-specific association.

This study opens a window to molecular-based typing, targeting the amplification of set of serovar-specific genes. However, not all the serovars have been sequenced, and some of them, although sequenced, could not be included here due to genome fragmentation. That said, closed genomes of additional serovars are needed to expand this analysis. This is part of an ongoing collaborative project, where genomes of previously not included serovars will be sequenced and used to develop a serovar prediction tool (under development by a postdoc in F. Veyrier's laboratory). Regarding host-preference, although *L. noguchii* fits to non-host-specific association, some preliminary RNA-sequencing results were promising. The log2(fold-change)-based multidimensional scaling plot showed a well-defined clustering according to host. Among the differentially expressed genes, candidates with pivotal roles in cell-cell interaction, or virulence were identified, which could define host-specificity. Since only three bovine and three amphibian isolates and one human isolate were compared, no solid conclusions could be drawn. RNA-seq experiments covering more strains per host, as well as additional hosts, would strengthen the analysis.

Finally, the results shown in Chapter 5 point to a high diversity of *L. santarosai* serovars circulating mainly in the Americas, a comparable scenario to that observed for *L. noguchii* in Chapter 4. Most importantly, the comparison within the $P1^+$ cluster revealed two separate groups sharing specific genomic features: one including *L. santarosai, L. mayottensis, L. borgpetersenii* and *L. weilii* (Group I), with a clearly reduced genome size and higher GC content than the one composed by *L. noguchii*, *L. interrogans* and *L. kirschneri* (Group II). *L. santarosai* clusters together with *L. borgpetersenii* and the observed characteristics fit with an ongoing reductive evolution. On the contrary, most of the population in *L. interrogans, L. kirschneri* and *L. noguchii* is still in the expansion phase, prior to genome reduction.

In light of these results, it becomes evident that further characterization of $P1^+$ species is needed. The preliminary results on codon usage invite to further explore this aspect, e.g., identify genes where there is differential codon usage, as well as a thorough pangenome characterization, including several strains per species. This type of analysis would contribute to the understanding of intraspecies evolution and species-specific adaptations in highly virulent *Leptospira*.

References

Abascal, F., Zardoya, R., & Telford, M. J. (2010). TranslatorX: multiple alignment of nucleotide sequences guided by amino acid translations. *Nucleic Acids Research*, *38*. https://doi.org/10.1093/NAR/GKQ291

Adhikarla, H., Wunder, E. A., Mechaly, A. E., Mehta, S., Wang, Z., Santos, L., … Ko, A. I. (2018). Lvr, a Signaling System That Controls Global Gene Regulation and Virulence in Pathogenic Leptospira. *Frontiers in Cellular and Infection Microbiology*, *8*. https://doi.org/10.3389/fcimb.2018.00045

Adler, B., & de la Peña Moctezuma, A. (2010). Leptospira and leptospirosis. *Veterinary Microbiology*, *140*(3–4), 287–296. https://doi.org/10.1016/j.vetmic.2009.03.012

Adler, B. et al. (2015). *Leptospira and Leptospirosis*. (B. Adler, Ed.), *Current Topics in Microbiology and Immunology* (Vol. 387). Berlin, Heidelberg: Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-662-45059-8

Ahmed, A., & P. Grobusch, M. (2012). Molecular Approaches in the Detection and Characterization of Leptospira. *Journal of Bacteriology & Parasitology*, *03*(02). https://doi.org/10.4172/2155-9597.1000133

Ahmed, A., Anthony, R. M., & Hartskeerl, R. A. (2010). A simple and rapid molecular method for Leptospira species identification. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *10*(7), 955–962. https://doi.org/10.1016/J.MEEGID.2010.06.002

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

Ahmed, N., Devi, S. M., de los Á Valverde, M., Vijayachari, P., Machang'u, R. S., Ellis, W. A., & Hartskeerl, R. A. (2006). Multilocus sequence typing method for identification and genotypic classification of pathogenic Leptospira species. *Annals of Clinical Microbiology and Antimicrobials*, *5*(1), 28. https://doi.org/10.1186/1476-0711-5-28

Alamuri, A., Thirumalesh, S. R. A., Kumari, S. S., Kumar, K. V., Roy, P., & Balamurugan, V.

(2019). Seroprevalence and distribution of serogroup-specific pathogenic Leptospira antibodies in cattle and buffaloes in the state of Andhra Pradesh, India. Veterinary world, 12(8), 1212–1217. https://doi.org/10.14202/vetworld.2019.1212-1217

Alves-Bezerra, M., & Cohen, D. E. (2017). Triglyceride Metabolism in the Liver. In *Comprehensive Physiology* (pp. 1–22). Wiley. https://doi.org/10.1002/cphy.c170012

Aramaki, T., Blanc-Mathieu, R., Endo, H., Ohkubo, K., Kanehisa, M., Goto, S., & Ogata, H. (2020). KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics*, *36*(7), 2251. https://doi.org/10.1093/BIOINFORMATICS/BTZ859

Awad-masalmeh, M., Resch, G., Bakoss, P., & Jarekova, J. (2012). DNA relatedness and serotyping of Leptospira strains. *Bratislavske Lekarske Listy*, *113*(2), 70–72. https://doi.org/10.4149/BLL_2012_017

Balasubramanian, D., López-Pérez, M., Grant, T.-A., Ogbunugafor, C. B., & Almagro-Moreno, S. (2022). Molecular mechanisms and drivers of pathogen emergence. *Trends in Microbiology*, *30*(9), 898–911. https://doi.org/10.1016/j.tim.2022.02.003

Barragan, V., Chiriboga, J., Miller, E., Olivas, S., Birdsell, D., Hepp, C., … Pearson, T. (2016). High Leptospira Diversity in Animals and Humans Complicates the Search for Common Reservoirs of Human Disease in Rural Ecuador. *PLoS Neglected Tropical Diseases*, *10*(9). https://doi.org/10.1371/JOURNAL.PNTD.0004990

Bauso, J., Simoncini, M. S., Chiani, Y., Schmeling, M. F., Larriera, A., Vanasco, N. B., & Piña, C. I. (2020). Presence of Leptospira spp. in Caiman latirostris (Crocodylia, Alligatoridae) populations in Santa Fe, Argentina. *Heliyon*, *6*(5), e03837. https://doi.org/10.1016/j.heliyon.2020.e03837

Belduz, A. O., Dulger, S., & Demirbag, Z. (2003). Anoxybacillus gonensis sp. nov., a moderately thermophilic, xylose-utilizing, endospore-forming bacterium. *International Journal of Systematic and Evolutionary Microbiology*, *53*(5), 1315–1320. https://doi.org/10.1099/ijs.0.02473-0

Bentham, A. R., Petit-Houdenot, Y., Win, J., Chuma, I., Terauchi, R., Banfield, M. J., … Langner, T. (2021). A single amino acid polymorphism in a conserved effector of the multihost blast

fungus pathogen expands host-target binding spectrum. *PLOS Pathogens*, *17*(11), e1009957. https://doi.org/10.1371/journal.ppat.1009957

Bertani, B., & Ruiz, N. (2018). Function and biogenesis of lipopolysaccharides. *EcoSal Plus*, *8*(1). https://doi.org/10.1128/ECOSALPLUS.ESP-0001-2018

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. *Current Microbiology*, *62*(2), 518–524. https://doi.org/10.1007/s00284-010-9738-7

Bharti, A. R., Nally, J. E., Ricaldi, J. N., Matthias, M. A., Diaz, M. M., Lovett, M. A., … Vinetz, J. M. (2003). Figures from Leptospirosis - a zoonotic disease of global importance. *The Lancet Infectious Diseases*, *3*(12), 757–771.

Bian, S., Zeng, W., Li, Q., Li, Y., Wong, N. K., Jiang, M., … Li, L. (2021). Genetic Structure, Function, and Evolution of Capsule Biosynthesis Loci in Vibrio parahaemolyticus. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/FMICB.2020.546150/FULL

Bilung, L. M., Pui, C. F., Su'Ut, L., & Apun, K. (2018). Evaluation of BOX-PCR and ERIC-PCR as Molecular Typing Tools for Pathogenic Leptospira. *Disease Markers*, *2018*. https://doi.org/10.1155/2018/1351634

Biswas, T., Yi, L., Aggarwal, P., Wu, J., Rubin, J. R., Stuckey, J. A., … Tsodikov, O. V. (2009). The Tail of KdsC: conformational changes control the activity of a haloacid dehalogenase superfamily phosphatase*. *The Journal of Biological Chemistry*, *284*(44), 30594. https://doi.org/10.1074/JBC.M109.012278

Bobay, L.-M., & Ochman, H. (2017). The Evolution of Bacterial Genome Architecture. *Frontiers in Genetics*, *8*. https://doi.org/10.3389/fgene.2017.00072

Bonhomme, D., Santecchia, I., Vernel-Pauillac, F., Caroff, M., Germon, P., Murray, G., … Werts, C. (2020). Leptospiral LPS escapes mouse TLR4 internalization and TRIF-associated antimicrobial responses through O antigen and associated lipoproteins. *PLOS Pathogens*, *16*(8), e1008639. https://doi.org/10.1371/journal.ppat.1008639

Boonsilp, S., Thaipadungpanit, J., Amornchai, P., Wuthiekanun, V., Bailey, M. S., Holden, M. T. G., … Peacock, S. J. (2013). A single multilocus sequence typing (MLST) scheme for seven

pathogenic Leptospira species. *PLoS Neglected Tropical Diseases*, *7*(1), e1954. https://doi.org/10.1371/journal.pntd.0001954

Borowiec, M. L. (2016). AMAS: A fast tool for alignment manipulation and computing of summary statistics. *PeerJ*, *2016*(1). https://doi.org/10.7717/PEERJ.1660/SUPP-2

Bourhy, P., Bremont, S., Zinini, F., Giry, C., & Picardeau, M. (2011). Comparison of Real-Time PCR Assays for Detection of Pathogenic Leptospira spp. in Blood and Identification of Variations in Target Sequences. *Journal Of Clinical Microbiology*, *49*(6), 2154–2160. https://doi.org/10.1128/JCM.02452-10

Bourhy, P., Collet, L., Brisse, S., & Picardeau, M. (2014). Leptospira mayottensis sp. nov., a pathogenic species of the genus Leptospira isolated from humans. International journal of systematic and evolutionary microbiology, 64(Pt 12), 4061–4067. https://doi.org/10.1099/ijs.0.066597-0

Bourhy, P., Herrmann Storck, C., Theodose, R., Olive, C., Nicolas, M., Hochedez, P., … Picardeau, M. (2013). Serovar Diversity of Pathogenic Leptospira Circulating in the French West Indies. *PLoS Neglected Tropical Diseases*, *7*(3), e2114. https://doi.org/10.1371/journal.pntd.0002114

Bourhy, P., & Saint Girons, I. (2000). Localization of the Leptospira interrogans metF gene on the CII secondary chromosome. *FEMS Microbiology Letters*, *191*(2), 259–263. https://doi.org/10.1016/S0378-1097(00)00393-1

Brenner, D J, McWhorter, A. C., Kai, A., Steigerwalt, A. G., & Farmer, J. J. (1986). Enterobacter asburiae sp. nov., a new species found in clinical specimens, and reassignment of Erwinia dissolvens and Erwinia nimipressuralis to the genus Enterobacter as Enterobacter dissolvens comb. nov. and Enterobacter nimipressuralis comb. nov. *Journal of Clinical Microbiology*, *23*(6), 1114–1120. https://doi.org/10.1128/jcm.23.6.1114-1120.1986

Brenner, Don 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. *International Journal of Systematic and Evolutionary Microbiology*, *49*(2), 839–858. https://doi.org/10.1099/00207713-49-2-839

Bröer, S. (2008). Amino Acid Transport Across Mammalian Intestinal and Renal Epithelia. *Physiological Reviews*, *88*(1), 249–286. https://doi.org/10.1152/physrev.00018.2006

Bryant, J. M., Brown, K. P., Burbaud, S., Everall, I., Belardinelli, J. M., Rodriguez-Rincon, D., … Floto, R. A. (2021). Stepwise pathogenic evolution of Mycobacterium abscessus. *Science*, *372*(6541). https://doi.org/10.1126/science.abb8699

Buchfink, B., Xie, C., & Huson, D. H. (2015). Fast and sensitive protein alignment using DIAMOND. *Nature Methods*, *12*(1), 59–60. https://doi.org/10.1038/nmeth.3176

Bulach, D. M, Seemann, T., Zuerner, R., & Adler, B. (2006). The organization of Leptospira at a genomic level. *Physical Education and Sport for Children and Youth with Special Needs Researches – Best Practices – Situation*, 109–123. https://doi.org/10.2/JQUERY.MIN.JS

Bulach, Dieter M., Kalambaheti, T., De La Peña-Moctezuma, A., & Adler, B. (2000). Functional Analysis of Genes in the rfb Locus of Leptospira borgpetersenii Serovar Hardjo Subtype Hardjobovis. *Infection and Immunity*, *68*(7), 3793. https://doi.org/10.1128/IAI.68.7.3793-3798.2000

Bulach, Dieter M, Kalambaheti, T., De La Peña-Moctezuma, A., & Adler, B. (2000). Lipopolysaccharide Biosynthesis in Leptospira. *J. Mol. Microbiol. Biotechnol*, *2*(4), 375–380. Retrieved from https://www.caister.com/jmmb/v/v2/v2n4/06.pdf

Bulach, Dieter M, Zuerner, R. L., Wilson, P., Seemann, T., McGrath, A., Cullen, P. a, … Adler, B. (2006). Genome reduction in Leptospira borgpetersenii reflects limited transmission potential. *Proceedings of the National Academy of Sciences of the United States of America*, *103*(39), 14560–14565. https://doi.org/10.1073/pnas.0603979103

Buzzanca, D., Botta, C., Ferrocino, I., Alessandria, V., Houf, K., & Rantsiou, K. (2021). Functional pangenome analysis reveals high virulence plasticity of Aliarcobacter butzleri and affinity to human mucus. *Genomics*, *113*(4), 2065–2076. https://doi.org/10.1016/J.YGENO.2021.05.001

Cai, C. S., Zhu, Y. Z., Zhong, Y., Xin, X. F., Jiang, X. G., Lou, X. L., … Guo, X. K. (2010). Development of O-antigen gene cluster-specific PCRs for rapid typing six epidemic serogroups of Leptospira in China. *BMC Microbiology*, *10*, 67. https://doi.org/10.1186/1471-2180-10-67

Camacho, C., Coulouris, G., Avagyan, V., Ma, N., Papadopoulos, J., Bealer, K., & Madden, T. L. (2009). BLAST+: architecture and applications. *BMC Bioinformatics*, *10*, 421. https://doi.org/10.1186/1471-2105-10-421

Cao, Q., Yang, N., Wang, Y., Xu, C., Zhang, X., Fan, K., … Lan, L. (2020). Mutation-induced remodeling of the BfmRS two-component system in Pseudomonas aeruginosa clinical isolates. *Science Signaling*, *13*(656). https://doi.org/10.1126/scisignal.aaz1529

Carvajal, M. P., & Fagerstrom, K. A. (2017). Epidemiology of Leptospirosis in Costa Rica 2011-2015. Current tropical medicine reports, 4(2), 41–46. https://doi.org/10.1007/s40475-017-0102-x

Carver, T., Thomson, N., Bleasby, A., Berriman, M., & Parkhill, J. (2009). DNAPlotter: circular and linear interactive genome visualization. *Bioinformatics*, *25*(1), 119. https://doi.org/10.1093/BIOINFORMATICS/BTN578

Casadevall, A., & Pirofski, L. A. (2000). Host-pathogen interactions: basic concepts of microbial commensalism, colonization, infection, and disease. *Infection and Immunity*, *68*(12), 6511–6518. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11083759

Casanovas-Massana, A., Hamond, C., Santos, L. A., de Oliveira, D., Hacker, K. P., Balassiano, I., … Wunder, E. A. (2020). Leptospira yasudae sp. Nov. and Leptospira stimsonii sp. nov., two new species of the pathogenic group isolated from environmental sources. *International Journal of Systematic and Evolutionary Microbiology*, *70*(3), 1450–1456. https://doi.org/10.1099/IJSEM.0.003480

Casanovas-Massana, A., Vincent, A. T., Bourhy, P., Neela, V. K., Veyrier, F. J., Picardeau, M., & Wunder, E. A. (2021). Leptospira dzianensis and Leptospira putramalaysiae are later heterotypic synonyms of Leptospira yasudae and Leptospira stimsonii. *International Journal of Systematic and Evolutionary Microbiology*, *71*(3), 4713. https://doi.org/10.1099/IJSEM.0.004713

Casas, E. M. C., Ríos, R. R., Bulach, D., Adler, B., & de la Peña-Moctezuma, A. (2005). [Evaluation of the transcription of locus rfb for the biosynthesis of LPS in Leptospira interrogans subtype hardjoprajitno]. *Revista Cubana de Medicina Tropical*, *57*(1), 43–44. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/17966475

Cerqueira, G. M., McBride, A. J. A., Hartskeerl, R. A., Ahmed, N., Dellagostin, O. A., Eslabão, M. R., & Nascimento, A. L. T. O. (2010). Bioinformatics describes novel Loci for high resolution discrimination of leptospira isolates. *PloS One*, *5*(10). https://doi.org/10.1371/JOURNAL.PONE.0015335

Cerqueira, G. M., & Picardeau, M. (2009). A century of Leptospira strain typing. *Infection, Genetics and Evolution : Journal of Molecular Epidemiology and Evolutionary Genetics in Infectious Diseases*, *9*(5), 760–768. https://doi.org/10.1016/J.MEEGID.2009.06.009

Chakraborty, A., Miyahara, S., Villanueva, S. Y., Saito, M., Gloriani, N. G., & Yoshida, S. (2011). A novel combination of selective agents for isolation of Leptospira species. Microbiology and immunology, 55(7), 494–501. https://doi.org/10.1111/j.1348-0421.2011.00347.x

Chang, M.-Y., Cheng, Y.-C., Hsu, S.-H., Ma, T.-L., Chou, L.-F., Hsu, H.-H., … Yang, C.-W. (2016). Leptospiral outer membrane protein LipL32 induces inflammation and kidney injury in zebrafish larvae. https://doi.org/10.1038/srep27838

Chaurasia, R., Salovey, A., Guo, X., Desir, G., & Vinetz, J. M. (2022). Vaccination With Leptospira interrogans PF07598 Gene Family-Encoded Virulence Modifying Proteins Protects Mice From Severe Leptospirosis and Reduces Bacterial Load in the Liver and Kidney. *Frontiers in Cellular and Infection Microbiology*, *12*. https://doi.org/10.3389/fcimb.2022.926994

Chin, C. S., Alexander, D. H., Marks, P., Klammer, A. A., Drake, J., Heiner, C., … Korlach, J. (2013). Nonhybrid, finished microbial genome assemblies from long-read SMRT sequencing data. *Nature Methods 2013 10:6*, *10*(6), 563–569. https://doi.org/10.1038/nmeth.2474

Chou, L. F., Chen, T. W., Ko, Y. C., Pan, M. J., Tian, Y. C., Chiu, C. H., Tang, P., Hung, C. C., & Yang, C. W. (2014). Potential impact on kidney infection: a whole-genome analysis of Leptospira santarosai serovar Shermani. Emerging microbes & infections, 3(11), e82. https://doi.org/10.1038/emi.2014.78

Chou, L.-F., Chen, Y.-T., Lu, C.-W., Ko, Y.-C., Tang, C.-Y., Pan, M.-J., … Yang, C.-W. (2012). Sequence of Leptospira santarosai serovar Shermani genome and prediction of virulence-associated genes. *Gene*, *511*(2), 364–370. https://doi.org/10.1016/j.gene.2012.09.074

Contreras-Moreira, B., & Vinuesa, P. (2013). GET_HOMOLOGUES, a Versatile Software

Package for Scalable and Robust Microbial Pangenome Analysis. *Applied and Environmental Microbiology*, *79*(24), 7696. https://doi.org/10.1128/AEM.02411-13

Corney, B. G., Colley, J., Djordjevic, S. P., Whittington, R., & Graham, G. C. (1993). Rapid identification of some Leptospira isolates from cattle by random amplified polymorphic DNA fingerprinting. *Journal of Clinical Microbiology*, *31*(11), 2927–2932. https://doi.org/10.1128/jcm.31.11.2927-2932.1993

Corney, B. G., Slack, A. T., Symonds, M. L., Dohnt, M. F., McClintock, C. S., McGowan, M. R., & Smythe, L. D. (2008). Leptospira weilii serovar Topaz, a new member of the Tarassovi serogroup isolated from a bovine source in Queensland, Australia. *International Journal Of Systematic And Evolutionary Microbiology*, *58*(10), 2249–2252. https://doi.org/10.1099/ijs.0.64884-0

Costa, F., Hagan, J. E., Calcagno, J., Kane, M., Torgerson, P., Martinez-Silveira, M. S., … Ko, A. I. (2015). Global Morbidity and Mortality of Leptospirosis: A Systematic Review. *PLoS Neglected Tropical Diseases*, *9*(9). https://doi.org/10.1371/JOURNAL.PNTD.0003898

Criscuolo, A., & Gribaldo, S. (2010). BMGE (Block Mapping and Gathering with Entropy): a new software for selection of phylogenetic informative regions from multiple sequence alignments. *BMC Evolutionary Biology*, *10*(1), 210. https://doi.org/10.1186/1471-2148-10-210

Crosby, F. L., Munderloh, U. G., Nelson, C. M., Herron, M. J., Lundgren, A. M., Xiao, Y.-P., … Barbet, A. F. (2020). Disruption of VirB6 Paralogs in Anaplasma phagocytophilum Attenuates Its Growth. *Journal of Bacteriology*, *202*(23). https://doi.org/10.1128/JB.00301-20

Delgado, M. A., Cáceres, O. A., Calderón, J. E., Balda, L., Sotil, G., & Céspedes, M. J. (2022). New Genetic Variants of Leptospira spp Characterized by MLST from Peruvian Isolates. Journal of tropical medicine, 2022, 4184326. https://doi.org/10.1155/2022/4184326

Denton, J. F., Lugo-Martinez, J., Tucker, A. E., Schrider, D. R., Warren, W. C., & Hahn, M. W. (2014). Extensive error in the number of genes inferred from draft genome assemblies. *PLoS Computational Biology*, *10*(12). https://doi.org/10.1371/JOURNAL.PCBI.1003998

Diop, A., Raoult, D., & Fournier, P.-E. (2018). Rickettsial genomics and the paradigm of genome

reduction associated with increased virulence. *Microbes and Infection*, *20*(7–8), 401–409. https://doi.org/10.1016/j.micinf.2017.11.009

dos Santos, D. C., Lange, C. C., Avellar-Costa, P., dos Santos, K. R. N., Brito, M. A. V. P., & Giambiagi-deMarval, M. (2016). Staphylococcus chromogenes, a Coagulase-Negative Staphylococcus Species That Can Clot Plasma. *Journal of Clinical Microbiology*, *54*(5), 1372–1375. https://doi.org/10.1128/JCM.03139-15

Ellis, W. A. (2015). Animal Leptospirosis. *Current Topics in Microbiology and Immunology*, *387*, 99–137. https://doi.org/10.1007/978-3-662-45059-8_6

Eshghi, A., Becam, J., Lambert, A., Sismeiro, O., Dillies, M. A., Jagla, B., … Picardeau, M. (2014). A putative regulatory genetic locus modulates virulence in the pathogen Leptospira interrogans. *Infection and Immunity*, *82*(6), 2542–2552. https://doi.org/10.1128/IAI.01803-14

Eshghi, A., Gaultney, R. A., England, P., Brûlé, S., Miras, I., Sato, H., … Picardeau, M. (2019). An extracellular <scp> *Leptospira interrogans* </scp> leucine-rich repeat protein binds human E- and VE-cadherins. *Cellular Microbiology*, *21*(2), 139–148. https://doi.org/10.1111/cmi.12949

Eshghi, A., Lourdault, K., Murray, G. L., Bartpho, T., Sermswan, R. W., Picardeau, M., … Cameron, C. E. (2012). Leptospira interrogans Catalase Is Required for Resistance to H 2 O 2 and for Virulence. *Infection and Immunity*, *80*(11), 3892–3899. https://doi.org/10.1128/IAI.00466-12

Everard, C. O., Carrington, D., Korver, H., & Everard, J. D. (1988). Leptospires in the marine toad (Bufo marinus) on Barbados. *Journal of Wildlife Diseases*, *24*(2), 334–338. https://doi.org/10.7589/0090-3558-24.2.334

Faine, S., Adler, B., Bolin, C., & Perolat, P. (1999). *Leptospira and Leptospirosis* (2nd Editio). Melbourne: Medisci Press.

Fernandes, L. G. V., Guaman, L. P., Vasconcellos, S. A., Heinemann, M. B., Picardeau, M., & Nascimento, A. L. T. O. (2019). Gene silencing based on RNA-guided catalytically inactive Cas9 (dCas9): a new tool for genetic engineering in Leptospira. *Scientific Reports*, *9*(1), 1839. https://doi.org/10.1038/s41598-018-37949-x

Fernandes, Luis G. V., Putz, E. J., Stasko, J., Lippolis, J. D., Nascimento, A. L. T. O., & Nally, J.

E. (2022). Evaluation of LipL32 and LigA/LigB Knockdown Mutants in Leptospira interrogans Serovar Copenhageni: Impacts to Proteome and Virulence. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.799012

Fernandes, Luis G. V., Stone, N. E., Roe, C. C., Goris, M. G. A., van der Linden, H., Sahl, J. W., … Nally, J. E. (2022). Leptospira sanjuanensis sp. nov., a pathogenic species of the genus Leptospira isolated from soil in Puerto Rico. *International Journal of Systematic and Evolutionary Microbiology*, *72*(10). https://doi.org/10.1099/ijsem.0.005560

Figueira, C., Croda, J., Choy, H. A., Haake, D. A., Reis, M. G., Ko, A. I., & Picardeau, M. (2011). Heterologous expression of pathogen-specific genes ligA and ligB in the saprophyte Leptospira biflexa confers enhanced adhesion to cultured cells and fibronectin. *BMC Microbiology*, *11*(1), 129. https://doi.org/10.1186/1471-2180-11-129

Filippidou, S., Jaussi, M., Junier, T., Wunderlin, T., Jeanneret, N., Regenspurg, S., … Junier, P. (2015). Genome Sequence of Aeribacillus pallidus Strain GS3372, an Endospore-Forming Bacterium Isolated in a Deep Geothermal Reservoir. *Genome Announcements*, *3*(4). https://doi.org/10.1128/genomeA.00981-15

Flores, B. J., Pérez-Sánchez, T., Fuertes, H., Sheleby-Elías, J., Múzquiz, J. L., Jirón, W., … Halaihel, N. (2017). A cross-sectional epidemiological study of domestic animals related to human leptospirosis cases in Nicaragua. *Acta Tropica*, *170*, 79–84. https://doi.org/10.1016/J.ACTATROPICA.2017.02.031

Fontana, Carolina, Lundborg, M., Weintraub, A., & Widmalm, G. (2014). Rapid structural elucidation of polysaccharides employing predicted functions of glycosyltransferases and NMR data: application to the O-antigen of Escherichia coli O59. *Glycobiology*, *24*(5), 450–457. https://doi.org/10.1093/GLYCOB/CWU011

Fontana, Célia, Lambert, A., Benaroudj, N., Gasparini, D., Gorgette, O., Cachet, N., … Picardeau, M. (2016). Analysis of a Spontaneous Non-Motile and Avirulent Mutant Shows That FliM Is Required for Full Endoflagella Assembly in Leptospira interrogans. *PLOS ONE*, *11*(4), e0152916. https://doi.org/10.1371/journal.pone.0152916

Fortier, L.-C., & Sekulovic, O. (2013). Importance of prophages to evolution and virulence of bacterial pathogens. *Virulence*, *4*(5), 354–365. https://doi.org/10.4161/viru.24498

Fouts, D. E., Matthias, M. A., Adhikarla, H., Adler, B., Amorim-Santos, L., Berg, D. E., … Vinetz, J. M. (2016). What Makes a Bacterial Species Pathogenic?:Comparative Genomic Analysis of the Genus Leptospira. *PLOS Neglected Tropical Diseases*, *10*(2), e0004403. https://doi.org/10.1371/journal.pntd.0004403

Fraser, C. M., Casjens, S., Huang, W. M., Sutton, G. G., Clayton, R., Lathigra, R., … Venter, J. C. (1997). Genomic sequence of a Lyme disease spirochaete, Borrelia burgdorferi. *Nature*, *390*(6660), 580–586. https://doi.org/10.1038/37551

Fraser, C. M., Norris, S. J., Weinstock, G. M., White, O., Sutton, G. G., Dodson, R., … Venter, J. C. (1998). Complete Genome Sequence of Treponema pallidum , the Syphilis Spirochete. *Science*, *281*(5375), 375–388. https://doi.org/10.1126/science.281.5375.375

Fraser, D. W., Glosser, J. W., Francis, D. P., Phillips, C. J., Feeley, J. C., & Sulzer, C. R. (1973). Leptospirosis Caused by Serotype *fort-Bragg*. *Annals of Internal Medicine*, *79*(6), 786. https://doi.org/10.7326/0003-4819-79-6-786

Galloway, R. L., & Levett, P. N. (2010). Application and validation of PFGE for serovar identification of Leptospira clinical isolates. *PLoS Neglected Tropical Diseases*, *4*(9). https://doi.org/10.1371/journal.pntd.0000824

Ganoza, C. A., Matthias, M. A., Collins-Richards, D., Brouwer, K. C., Cunningham, C. B., Segura, E. R., Gilman, R. H., Gotuzzo, E., & Vinetz, J. M. (2006). Determining risk for severe leptospirosis by molecular analysis of environmental surface waters for pathogenic Leptospira. PLoS medicine, 3(8), e308. https://doi.org/10.1371/journal.pmed.0030308

Gaultney, R. A., Vincent, A. T., Lorioux, C., Coppeé, J. Y., Sismeiro, O., Varet, H., … Picardeau, M. (2021). 4-Methylcytosine DNA modification is critical for global epigenetic regulation and virulence in the human pathogen Leptospira interrogans. *Nucleic Acids Research*, *48*(21), 12102–12115. https://doi.org/10.1093/nar/gkaa966

Ghosh, S., & O'Connor, T. J. (2017). Beyond Paralogs: The Multiple Layers of Redundancy in Bacterial Pathogenesis. *Frontiers in Cellular and Infection Microbiology*, *7*. https://doi.org/10.3389/fcimb.2017.00467

Gochenour, W. S., Smadel, J. E., Jackson, E. B., Evans, L. B., & Yager, R. H. (1952). Leptospiral etiology of Fort Bragg fever. *Public Health Reports*, *67*(8), 811.

https://doi.org/10.2307/4588208

Goldstein, S. F., & Charon, N. W. (1990). Multiple-exposure photographic analysis of a motile spirochete. *Proceedings of the National Academy of Sciences*, *87*(13), 4895–4899. https://doi.org/10.1073/pnas.87.13.4895

Gouil, Q., & Keniry, A. (2019). Latest techniques to study DNA methylation. *Essays in Biochemistry*, *63*(6), 639–648. https://doi.org/10.1042/EBC20190027

Grassmann, A. A., Souza, J. D., & McBride, A. J. A. (2017). A Universal Vaccine against Leptospirosis: Are We Going in the Right Direction? *Frontiers in Immunology*, *8*(MAR). https://doi.org/10.3389/FIMMU.2017.00256

Gravekamp, C., Korver, H., Montgomery, J., Everard, C. O. R., Carrington, D., Ellis, W. A., & Terpstra, W. J. (1991). Leptospires Isolated from Toads and Frogs on the Island of Barbados. *Zentralblatt Für Bakteriologie*, *275*(3), 403–411. https://doi.org/10.1016/S0934-8840(11)80308-7

Grillová, L., Angermeier, H., Levy, M., Giard, M., Lastère, S., & Picardeau, M. (2020). Circulating genotypes of Leptospira in French Polynesia : An 9-year molecular epidemiology surveillance follow-up study. *PLOS Neglected Tropical Diseases*, *14*(9), e0008662. https://doi.org/10.1371/journal.pntd.0008662

Guedes, I. B., Souza, G. O. De, Fernandes, J., & Castro, D. P. (2021). Usefulness of the Ranking Technique in the Microscopic Agglutination Test (MAT) to Predict the Most Likely Infecting Serogroup of Leptospira, *8*(March), 1–6. https://doi.org/10.3389/fvets.2021.654034

Guerra Maldonado, J. F., Vincent, A. T., Chenal, M., & Veyrier, F. J. (2020). CAPRIB: a user-friendly tool to study amino acid changes and selection for the exploration of intra-genus evolution. *BMC Genomics*, *21*(1). https://doi.org/10.1186/S12864-020-07232-3

Guerreiro, H., Croda, J. L., Flannery, B., Mazel, M., Matsunaga, J., Galv, M., … Haake, D. A. (2001). Leptospiral Proteins Recognized during the Humoral Immune Response to Leptospirosis in Humans. *INFECTION AND IMMUNITY*, *69*(8), 4958–4968. https://doi.org/10.1128/IAI.69.8.4958–4968.2001

Guglielmini, J., Bourhy, P., Schiettekatte, O., Zinini, F., Brisse, S., & Picardeau, M. (2019). Genus-wide Leptospira core genome multilocus sequence typing for strain taxonomy and global

surveillance. *PLoS Neglected Tropical Diseases*, *13*(4). https://doi.org/10.1371/JOURNAL.PNTD.0007374

Gupta, R. S., Mahmood, S., Adeolu, M., & Ly, B. (2013). A phylogenomic and molecular signature based approach for characterization of the phylum Spirochaetes and its major clades : proposal for a taxonomic revision of the phylum, *4*(July), 1–18. https://doi.org/10.3389/fmicb.2013.00217

Haake, D. A. (2000). Spirochaetal lipoproteins and pathogenesis. *Microbiology*, *146 (Pt 7*(7), 1491–1504. https://doi.org/10.1099/00221287-146-7-1491

Haake, D. A., Chao, G., Zuerner, R. L., Barnett, J. K., Barnett, D., Mazel, M., … Bolin, C. A. (2000). The Leptospiral Major Outer Membrane Protein LipL32 Is a Lipoprotein Expressed during Mammalian Infection. *INFECTION AND IMMUNITY*, *68*(4), 2276–2285. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC97414/pdf/ii002276.pdf

Haake, D. A., Martinich, C., Summers, T. A., Shang, E. S., Pruetz, J. D., McCoy, A. M., … Bolin, C. A. (1998). Characterization of Leptospiral Outer Membrane Lipoprotein LipL36: Downregulation Associated with Late-Log-Phase Growth and Mammalian Infection. *Infection and Immunity*, *66*(4), 1579–1587. https://doi.org/10.1128/iai.66.4.1579-1587.1998

Haake, D. A., & Matsunaga, J. (2010). Leptospira: a spirochaete with a hybrid outer membrane. *Molecular Microbiology*, no-no. https://doi.org/10.1111/j.1365-2958.2010.07262.x

Haake, D. A., & Matsunaga, J. (2021). Leptospiral Immunoglobulin-Like Domain Proteins: Roles in Virulence and Immunity. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/fimmu.2020.579907

Haake, D. A., Mazel, M. K., Mccoy, A. M., Milward, F., Chao, G., Matsunaga, J., & Wagar, E. A. (1999). Leptospiral outer membrane proteins OmpL1 and LipL41 exhibit synergistic immunoprotection. *Infection and Immunity*, *67*(12), 6572–6582. https://doi.org/10.1128/IAI.67.12.6572-6582.1999

Haake, D. A., Suchard, M. A., Kelley, M. M., Dundoo, M., Alt, D. P., & Zuerner, R. L. (2004). Molecular Evolution and Mosaicism of Leptospiral Outer Membrane Proteins Involves Horizontal DNA Transfer. *Journal Of Bacteriology*, *186*(9), 2818–2828. https://doi.org/10.1128/JB.186.9.2818–2828.2004

Haake, D. A., & Zückert, W. R. (2015). The Leptospiral Outer Membrane (pp. 187–221). https://doi.org/10.1007/978-3-662-45059-8_8

Hagberg, A., Swart, P., S Chult, D. (2008). Exploring network structure, dynamics, and function using NetworkX.

Hamond, C., Pestana, C. P., Medeiros, M. A., & Lilenbaum, W. (2016). Genotyping of Leptospira directly in urine samples of cattle demonstrates a diversity of species and strains in Brazil. *Epidemiology and Infection*, *144*(01), 72–75. https://doi.org/10.1017/S0950268815001363

Hamond, C., Pinna, M., Medeiros, M. A., Bourhy, P., Lilenbaum, W., & Picardeau, M. (2015). A multilocus variable number tandem repeat analysis assay provides high discrimination for genotyping Leptospira santarosai strains. *Journal of Medical Microbiology*, *64*, 507–512. https://doi.org/10.1099/jmm.0.000045

Hoang, D. T., Chernomor, O., Von Haeseler, A., Minh, B. Q., & Vinh, L. S. (2018). UFBoot2: Improving the Ultrafast Bootstrap Approximation. *Molecular Biology and Evolution*, *35*(2), 518. https://doi.org/10.1093/MOLBEV/MSX281

Hotez, P. J., Bottazzi, M. E., Franco-Paredes, C., Ault, S. K., & Periago, M. R. (2008). The neglected tropical diseases of Latin America and the Caribbean: a review of disease burden and distribution and a roadmap for control and elimination. PLoS neglected tropical diseases, 2(9), e300. https://doi.org/10.1371/journal.pntd.0000300

Hovind-Hougen, K. (1979). Leptospiraceae, a New Family to Include Leptospira Noguchi 1917 and Leptonema gen. nov. *International Journal of Systematic Bacteriology*, *29*(3), 245–251. https://doi.org/10.1099/00207713-29-3-245

Hsieh, W. J., Chang, Y. F., Chen, C. S., & Pan, M. J. (2005). Omp52 is a growth-phase-regulated outer membrane protein of Leptospira santarosai serovar Shermani. FEMS microbiology letters, 243(2), 339–345. https://doi.org/10.1016/j.femsle.2004.12.021

Hsu, S.-H., Chang, M.-Y., Lin, S.-M., Ko, Y.-C., Chou, L.-F., Tian, Y.-C., … Yang, C.-W. (2021). Peptidoglycan mediates Leptospira outer membrane protein Loa22 to toll-like receptor 2 for inflammatory interaction: a novel innate immune recognition. *Scientific Reports*, *11*(1), 1064. https://doi.org/10.1038/s41598-020-79662-8

Huerta-Cepas, J., Forslund, K., Coelho, L. P., Szklarczyk, D., Jensen, L. J., von Mering, C., &

Bork, P. (2017). Fast Genome-Wide Functional Annotation through Orthology Assignment by eggNOG-Mapper. *Molecular Biology and Evolution*, *34*(8), 2115–2122. https://doi.org/10.1093/molbev/msx148

Huszczynski, S. M., Hao, Y., Lam, J. S., & Khursigara, C. M. (2020). Identification of the Pseudomonas aeruginosa O17 and O15 O-Specific Antigen Biosynthesis Loci Reveals an ABC Transporter-Dependent Synthesis Pathway and Mechanisms of Genetic Diversity. *Journal of Bacteriology*, *202*(19). https://doi.org/10.1128/JB.00347-20

Inácio, J., Muñoz-Mérida, A., Ahmed, A. A., Rocha, T., Mesquita, J. R., Thompson, G., … Ferreira, A. S. (2021). Draft Genome Sequence of a Leptospira kirschneri Serovar Mozdok Type 2 Strain Isolated from a Horse in Portugal. *Microbiology Resource Announcements*, *10*(28). https://doi.org/10.1128/MRA.00217-21

Islam, S. T., & Lam, J. S. (2014). Synthesis of bacterial polysaccharides via the Wzx/Wzy-dependent pathway. *Canadian Journal of Microbiology*, *60*(11), 697–716. https://doi.org/10.1139/CJM-2014-0595

Jaeger, L. H., Pestana, C. P., Correia, L., Carvalho-Costa, F. A., Medeiros, M. A., & Lilenbaum, W. (2019). Novel MLST sequence types of pathogenic Leptospira spp.: Opening the black box of animal leptospirosis in Brazil. Acta tropica, 196, 135–141. https://doi.org/10.1016/j.actatropica.2019.05.025

Jancloes, M., Bertherat, E., Schneider, C., Belmain, S. R., Munoz-Zanzi, C., Hartskeerl, R. A., … Benschop, J. (2014). Towards a " One Health " Strategy against Leptospirosis. *Planet@Risk*, *2*(3), 204–206. Retrieved from http://planet-risk.org/index.php/pr/article/view/94

Javati, S., Guernier-Cambert, V., Jonduo, M., Robby, S., Kimopa, J., Maure, T., … Horwood, P. F. (2022). Diversity of Leptospira spp. in bats and rodents from Papua New Guinea. *Transboundary and Emerging Diseases*, *69*(6), 4048–4054. https://doi.org/10.1111/tbed.14725

Jayasundara, D., Senavirathna, I., Warnasekara, J., Gamage, C., Siribaddana, S., Kularatne, S. A. M., Matthias, M., Mariet, J. F., Picardeau, M., Agampodi, S., & M Vinetz, J. (2021). 12 Novel clonal groups of Leptospira infecting humans in multiple contrasting epidemiological contexts in Sri Lanka. PLoS neglected tropical diseases, 15(3), e0009272.

https://doi.org/10.1371/journal.pntd.0009272

Kalambaheti, T., Bulach, D. M., Rajakumar, K., & Adler, B. (1999). Genetic organization of the lipopolysaccharide O-antigen biosynthetic locus of Leptospira borgpetersenii serovar Hardjobovis. *Microbial Pathogenesis*, *27*(2), 105–117. https://doi.org/10.1006/MPAT.1999.0285

Kallel, H., Bourhy, P., Mayence, C., Houcke, S., Hommel, D., Picardeau, M., Caro, V., & Matheus, S. (2020). First report of human Leptospira santarosai infection in French Guiana. Journal of infection and public health, 13(8), 1181–1183. https://doi.org/10.1016/j.jiph.2020.03.020

Kanehisa, M., & Sato, Y. (2020). KEGG Mapper for inferring cellular functions from protein sequences. *Protein Science : A Publication of the Protein Society*, *29*(1), 28–35. https://doi.org/10.1002/PRO.3711

Kassegne, K., Hu, W., Ojcius, D. M., Sun, D., Ge, Y., Zhao, J., … Yan, J. (2014). Identification of collagenase as a critical virulence factor for invasiveness and transmission of pathogenic leptospira species. *Journal of Infectious Diseases*, *209*(7), 1105–1115. https://doi.org/10.1093/infdis/jit659

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*(4), 772–780. https://doi.org/10.1093/MOLBEV/MST010

Kelesidis, T. (2014). The Cross-Talk between Spirochetal Lipoproteins and Immunity. *Frontiers in Immunology*, *5*. https://doi.org/10.3389/fimmu.2014.00310

Kim, M., & Chun, J. (2014). 16S rRNA Gene-Based Identification of Bacteria and Archaea using the EzTaxon Server (pp. 61–74). https://doi.org/10.1016/bs.mim.2014.08.001

Ko, A. I., Goarant, C., & Picardeau, M. (2009). Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nature Reviews. Microbiology*, *7*(10), 736–747. https://doi.org/10.1038/nrmicro2208

Koizumi, N., & Watanabe, H. (2003). Molecular cloning and characterization of a novel leptospiral lipoprotein with OmpA domain. *FEMS Microbiology Letters*, *226*(2), 215–219. https://doi.org/10.1016/S0378-1097(03)00619-0

Konstantinidis, K. T., & Tiedje, J. M. (2005). Genomic insights that advance the species definition for prokaryotes. *Proceedings of the National Academy of Sciences*, *102*(7), 2567–2572. https://doi.org/10.1073/pnas.0409727102

Korba, A. A., Lounici, H., Kainiu, M., Vincent, A. T., Mariet, J. F., Veyrier, F. J., … Picardeau, M. (2021). Leptospira ainlahdjerensis sp. Nov., leptospira ainazelensis sp. nov., leptospira abararensis sp. nov. and leptospira chreensis sp. nov., four new species isolated from water sources in algeria. *International Journal of Systematic and Evolutionary Microbiology*, *71*(12), 005148. https://doi.org/10.1099/

Koren, S., Walenz, B. P., Berlin, K., Miller, J. R., Bergman, N. H., & Phillippy, A. M. (2017). Canu: scalable and accurate long-read assembly via adaptive k-mer weighting and repeat separation. *Genome Research*, *27*(5), 722–736. https://doi.org/10.1101/GR.215087.116

Kosugi, S., Hirakawa, H., & Tabata, S. (2015). GMcloser: closing gaps in assemblies accurately with a likelihood-based selection of contig or long-read alignments. *Bioinformatics (Oxford, England)*, *31*(23), 3733–3741. https://doi.org/10.1093/BIOINFORMATICS/BTV465

Kremer, F. S., Eslabão, M. R., Provisor, M., Woloski, R. D., Ramires, O. V., Moreno, L. Z., Moreno, A. M., Hamond, C., Lilenbaum, W., & Dellagostin, O. A. (2015). Draft Genome Sequences of Leptospira santarosai Strains U160, U164, and U233, Isolated from Asymptomatic Cattle. Genome announcements, 3(4), e00910-15. https://doi.org/10.1128/genomeA.00910-15

Kurilung, A., Chanchaithong, P., Lugsomya, K., Niyomtham, W., Wuthiekanun, V., & Prapasarakul, N. (2017). Molecular detection and isolation of pathogenic Leptospira from asymptomatic humans, domestic animals and water sources in Nan province, a rural area of Thailand. *Research in Veterinary Science*, *115*, 146–154. https://doi.org/10.1016/J.RVSC.2017.03.017

Kurilung, A., Keeratipusana, C., Horiike, T., Suriyaphol, P., Hampson, D. J., & Prapasarakul, N. (2019). Chronology of emergence of the genus Leptospira and over-representation of gene families enriched by vitamin B2, B12 biosynthesis, cell adhesion and external encapsulating structure in L. interrogans isolates from asymptomatic dogs. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 73, 7–12. https://doi.org/10.1016/j.meegid.2019.04.005

Kurilung, A., Perreten, V., & Prapasarakul, N. (2021). Comparative Genomic Analysis and a Novel Set of Missense Mutation of the Leptospira weilii Serogroup Mini From the Urine of Asymptomatic Dogs in Thailand. *Frontiers in Microbiology*, *12*. https://doi.org/10.3389/fmicb.2021.731937

Lagadec, E., Gomard, Y., Le Minter, G., Cordonin, C., Cardinale, E., Ramasindrazana, B., Dietrich, M., Goodman, S. M., Tortosa, P., & Dellagi, K. (2016). Identification of Tenrec ecaudatus, a Wild Mammal Introduced to Mayotte Island, as a Reservoir of the Newly Identified Human Pathogenic Leptospira mayottensis. PLoS neglected tropical diseases, 10(8), e0004933. https://doi.org/10.1371/journal.pntd.0004933

Lahoz-Beltra, R., Navarro, J., & MarijuÃ¡n, P. C. (2014). Bacterial computing: a form of natural computing and its applications. *Frontiers in Microbiology*, *5*. https://doi.org/10.3389/fmicb.2014.00101

Laleye, A. T., & Abolnik, C. (2020). Emergence of highly pathogenic H5N2 and H7N1 influenza A viruses from low pathogenic precursors by serial passage in ovo. *PLOS ONE*, *15*(10), e0240290. https://doi.org/10.1371/journal.pone.0240290

Lambert, A., Picardeau, M., Haake, D. A., Sermswan, R. W., Srikram, A., Adler, B., & Murray, G. A. (2012). FlaA Proteins in Leptospira interrogans Are Essential for Motility and Virulence but Are Not Required for Formation of the Flagellum Sheath. *Infection and Immunity*, *80*(6), 2019–2025. https://doi.org/10.1128/IAI.00131-12

Lambert, A., Wong Ng, J., & Picardeau, M. (2015). Gene inactivation of a chemotaxis operon in the pathogen Leptospira interrogans. *FEMS Microbiology Letters*, *362*(3), 1–8. https://doi.org/10.1093/femsle/fnu054

Land, M., Hauser, L., Jun, S. R., Nookaew, I., Leuze, M. R., Ahn, T. H., Karpinets, T., Lund, O., Kora, G., Wassenaar, T., Poudel, S., & Ussery, D. W. (2015). Insights from 20 years of bacterial genome sequencing. Functional & integrative genomics, 15(2), 141–161. https://doi.org/10.1007/s10142-015-0433-4

Lata, K. S., Vaghasia, V., Bhairappanavar, S. B., Kumar, S., Ayachit, G., Patel, S., & Das, J. (2020). Whole genome sequencing and de novo assembly of three virulent Indian isolates of Leptospira. Infection, genetics and evolution : journal of molecular epidemiology and

evolutionary genetics in infectious diseases, 85, 104579. https://doi.org/10.1016/j.meegid.2020.104579

Lauková, A., Pogány Simonová, M., Kubašová, I., Miltko, R., Bełzecki, G., & Strompfová, V. (2018). Sensitivity to antimicrobials of faecal Buttiauxella spp. from roe and red deer (Capreolus capreolus, Cervus elaphus) detected with MALDI-TOF mass spectrometry. *Polish Journal of Veterinary Sciences*, *21*(3), 543–547. https://doi.org/10.24425/124288

Lederberg, J. (2000). Infectious History. *Science*, *288*(5464), 287–293. https://doi.org/10.1126/science.288.5464.287

Lee, S. H., Kim, S., Park, S. C., & Kim, M. J. (2002). Cytotoxic activities of Leptospira interrogans hemolysin SphH as a pore-forming protein on mammalian cells. *Infection and Immunity*, *70*(1), 315–322. https://doi.org/10.1128/IAI.70.1.315-322.2002

LeFebvre, R. B. (1987). DNA probe for detection of the Leptospira interrogans serovar hardjo genotype hardjo-bovis. *Journal of Clinical Microbiology*, *25*(11), 2236–2238. https://doi.org/10.1128/JCM.25.11.2236-2238.1987

Lessa-Aquino, C., Rodrigues, C. B., Pablo, J., Sasaki, R., Jasinskas, A., Liang, L., … Felgner, P. L. (n.d.). Identification of Seroreactive Proteins of Leptospira interrogans Serovar Copenhageni Using a High-Density Protein Microarray Approach. https://doi.org/10.1371/journal.pntd.0002499

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4 : recent updates and, *47*(April), 256–259. https://doi.org/10.1093/nar/gkz239

Levett, P N. (2001). Leptospirosis. *Clinical Microbiology*, *14*(2), 296–326. https://doi.org/10.1128/CMR.14.2.296

Levett, Paul N, Picardeau, M., Call, M., Order, T. O., Chappel, R., Galloway, R., … Secre-, P. L. (2021). International Committee on Systematics of Prokaryotes Subcommittee on the taxonomy of Leptospiraceae Minutes of the closed meeting , 10 July 2019 , Vancouver , British The minutes of the meeting held in Palmerston North , New, 3–5. https://doi.org/10.1099/ijsem.0.005002

Liao, S., Sun, A., Ojcius, D. M., Wu, S., Zhao, J., & Yan, J. (2009). Inactivation of the fliY gene encoding a flagellar motor switch protein attenuates mobility and virulence of Leptospira

interrogansstrain Lai. *BMC Microbiology*, *9*(1), 253. https://doi.org/10.1186/1471-2180-9-253

Lilenbaum, W., Kremer, F., Ristow, P., Dellagostin, O., Bourhy, P., Hartskeerl, R., & Vasconcellos, S. (2015). Molecular characterization of the first leptospires isolated from goats in Brazil. Brazilian journal of microbiology : [publication of the Brazilian Society for Microbiology], 45(4), 1527–1530. https://doi.org/10.1590/s1517-83822014000400050

Lo, M., Cordwell, S. J., Bulach, D. M., & Adler, B. (2009). Comparative transcriptional and translational analysis of leptospiral outer membrane protein expression in response to temperature. *PLoS Neglected Tropical Diseases*, *3*(12). https://doi.org/10.1371/journal.pntd.0000560

Lo, M., Murray, G. L., Khoo, C. A., Haake, D. A., Zuerner, R. L., & Adler, B. (2010). Transcriptional Response of Leptospira interrogans to Iron Limitation and Characterization of a PerR Homolog. *Infection and Immunity*, *78*(11), 4850–4859. https://doi.org/10.1128/IAI.00435-10

Lourdault, K., Cerqueira, G. M., Wunder, E. A., & Picardeau, M. (2011). Inactivation of clpB in the Pathogen Leptospira interrogans Reduces Virulence and Resistance to Stress Conditions. *Infection and Immunity*, *79*(9), 3711–3717. https://doi.org/10.1128/IAI.05168-11

Loureiro, A. P., Jaeger, L. H., Di Azevedo, M. I. N., Miraglia, F., Moreno, L. Z., Moreno, A. M., … Lilenbaum, W. (2020). Molecular epidemiology of Leptospira noguchii reveals important insights into a One Health context. *Transboundary and Emerging Diseases*, *67*(1), 276–283. https://doi.org/10.1111/

Lu, P., Jin, J., Li, Z., Xu, Y., Hu, D., Liu, J., & Cao, P. (2020). PGcloser: Fast Parallel Gap-Closing Tool Using Long-Reads or Contigs to Fill Gaps in Genomes. *Evolutionary Bioinformatics Online*, *16*. https://doi.org/10.1177/1176934320913859

Madeira, F., Pearce, M., Tivey, A. R. N., Basutkar, P., Lee, J., Edbali, O., … Lopez, R. (2022). Search and sequence analysis tools services from EMBL-EBI in 2022. *Nucleic Acids Research*, *50*(W1), W276–W279. https://doi.org/10.1093/nar/gkac240

Marcsisin, R. A., Bartpho, T., Bulach, D. M., Srikram, A., Sermswan, R. W., Adler, B., & Murray, G. L. (2013). Use of a high-throughput screen to identify Leptospira mutants unable to

colonize the carrier host or cause disease in the acute model of infection. *Journal of Medical Microbiology*, *62*(10), 1601–1608. https://doi.org/10.1099/jmm.0.058586-0

Marshall, R. B., Wilton, B. E., & Robinson, A. J. (1981). Identification of Leptospira serovars by restriction-endonuclease analysis. *Journal of Medical Microbiology*, *14*(1), 163–166. https://doi.org/10.1099/00222615-14-1-163/CITE/REFWORKS

Martins, G., Loureiro, A. P., Hamond, C., Pinna, M. H., Bremont, S., Bourhy, P., & Lilenbaum, W. (2015). First isolation of Leptospira noguchii serogroups Panama and Autumnalis from cattle. *Epidemiology and Infection*, *143*(7), 1538–1541. https://doi.org/10.1017/S0950268814002416

Mason, M. R., Encina, C., Sreevatsan, S., & Muñoz-Zanzi, C. (2016). Distribution and Diversity of Pathogenic Leptospira Species in Peri-domestic Surface Waters from South Central Chile. *PLOS Neglected Tropical Diseases*, *10*(8), e0004895. https://doi.org/10.1371/journal.pntd.0004895

Matsunaga, J., Lo, M., Bulach, D. M., Zuerner, R. L., Adler, B., & Haake, D. A. (2007). Response of Leptospira interrogans to Physiologic Osmolarity: Relevance in Signaling the Environment-to-Host Transition. *Infection and Immunity*, *75*(6), 2864–2874. https://doi.org/10.1128/IAI.01619-06

Matthias, M. A., Ricaldi, J. N., Cespedes, M., Diaz, M. M., Galloway, R. L., Saito, M., … Vinetz, J. M. (2008). Human Leptospirosis Caused by a New, Antigenically Unique Leptospira Associated with a Rattus Species Reservoir in the Peruvian Amazon. *PLoS Neglected Tropical Diseases*, *2*(4), e213. https://doi.org/10.1371/journal.pntd.0000213

Mcbride, A. J. A., Cerqueira, G. M., Suchard, M. A., Moreira, A. N., Zuerner, R. L., Reis, M. G., … Dellagostin, O. A. (2009). Genetic diversity of the Leptospiral immunoglobulin-like (Lig) genes in pathogenic Leptospira spp. NIH Public Access. *Infect Genet Evol*, *9*(2). https://doi.org/10.1016/j.meegid.2008.10.012

Medeiros, E. J. S., Ferreira, L. C. A., Ortega, J. M., Cosate, M. R. V., & Sakamoto, T. (2022). Genetic basis underlying the serological affinity of leptospiral serovars from serogroups Sejroe, Mini and Hebdomadis. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases, 103, 105345.

https://doi.org/10.1016/j.meegid.2022.105345

Merien, F., Portnoi, D., Bourhy, P., Charavay, F., Berlioz-Arthaud, A., & Baranton, G. (2005). A rapid and quantitative method for the detection of *Leptospira* species in human leptospirosis. *FEMS Microbiology Letters*, *249*(1), 139–147. https://doi.org/10.1016/j.femsle.2005.06.011

Mgode, G. F., Mhamphi, G. G., Katakweba, A. S., & Thomas, M. (2014). Leptospira infections in freshwater fish in Morogoro Tanzania: a hidden public health threat. *Tanzania Journal of Health Research*, *16*(2). https://doi.org/10.4314/THRB.V16I2.7

Miller, M. R., Miller, K. A., Bian, J., James, M. E., Zhang, S., Lynch, M. J., … Charon, N. W. (2016). Spirochaete flagella hook proteins self-catalyse a lysinoalanine covalent crosslink for motility. *Nature Microbiology*, *1*(10), 16134. https://doi.org/10.1038/nmicrobiol.2016.134

Miotto, B. A., Moreno, L. Z., Guilloux, A. G. A., Sousa, G. O., Loureiro, A. P., Moreno, A. M., Lilenbaum, W., Vasconcellos, S. A., Heinemann, M. B., & Hagiwara, M. K. (2016). Molecular and serological characterization of the first Leptospira santarosai strain isolated from a dog. Acta tropica, 162, 1–4. https://doi.org/10.1016/j.actatropica.2016.06.007

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. *Journal Of Bacteriology*, *179*(4), 1262–1267. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC178824/pdf/1791262.pdf

Monroy, F. P., Solari, S., Lopez, J. Á., Agudelo-Flórez, P., & Peláez Sánchez, R. G. (2021). High Diversity of Leptospira Species Infecting Bats Captured in the Urabá Region (Antioquia-Colombia). *Microorganisms*, *9*(9), 1897. https://doi.org/10.3390/microorganisms9091897

Moreno, L. Z., Loureiro, A. P., Miraglia, F., Matajira, C. E. C., Kremer, F. S., Eslabao, M. R., … Moreno, A. M. (2015). Draft Genome Sequence of Brazilian Leptospira noguchii Serogroup Panama Strain U73, Isolated from Cattle. *Genome Announcements*, *3*(5). https://doi.org/10.1128/GENOMEA.01179-15

Moreno, L. Z., Miraglia, F., Marvulo, M. F., Silva, J. C., Paula, C. D., Costa, B. L., Morais, Z. M., Ferreira, F., Neto, J. S., Dellagostin, O. A., Hartskeerl, R. A., Vasconcellos, S. A., & Moreno, A. M. (2016). Characterization of Leptospira santarosai Serogroup Grippotyphosa Serovar

Bananal Isolated from Capybara (Hydrochaeris hydrochaeris) in Brazil. Journal of wildlife diseases, 52(3), 688–693. https://doi.org/10.7589/2015-09-245

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. *Journal of Clinical Microbiology*, *44*(10), 3510–3516. https://doi.org/10.1128/JCM.00670-06

Muller, H. E., Brenner, D. J., Fanning, G. R., Grimont, P. A. D., & Kampfer, P. (1996). Emended Description of Buttiauxella agrestis with Recognition of Six New Species of Buttiauxella and Two New Species of Kluyvera: Buttiauxella ferragutiae sp. nov., Buttiauxella gaviniae sp. nov., Buttiauxella brennerae sp. nov., Buttiauxella izardii sp. *International Journal of Systematic Bacteriology*, *46*(1), 50–63. https://doi.org/10.1099/00207713-46-1-50

Munoz-Zanzi, C., Groene, E., Morawski, B. M., Bonner, K., Costa, F., Bertherat, E., & Schneider, M. C. (2020). A systematic literature review of leptospirosis outbreaks worldwide, 1970-2012. Revista panamericana de salud publica = Pan American journal of public health, 44, e78. https://doi.org/10.26633/RPSP.2020.78

Murray, G. L. (2015). The Molecular Basis of Leptospiral Pathogenesis. In *Current topics in microbiology and immunology* (Vol. 387, pp. 139–185). https://doi.org/10.1007/978-3-662-45059-8_7

Murray, G. L., Srikram, A., Henry, R., Hartskeerl, R. A., Sermswan, R. W., & Adler, B. (2010). Mutations affecting Leptospira interrogans lipopolysaccharide attenuate virulence. *Molecular Microbiology*, *78*(3), 701–709. https://doi.org/10.1111/j.1365-2958.2010.07360.x

Murray, G. L., Srikram, A., Hoke, D. E., Wunder, E. A., Henry, R., Lo, M., … Adler, B. (2009). Major Surface Protein LipL32 Is Not Required for Either Acute or Chronic Infection with Leptospira interrogans . *INFECTION AND IMMUNITY*, *77*(3), 952–958. https://doi.org/10.1128/IAI.01370-08

Musto, H., Naya, H., Zavala, A., Romero, H., Alvarez-Valín, F., & Bernardi, G. (2004). Correlations between genomic GC levels and optimal growth temperatures in prokaryotes. *FEBS Letters*, *573*(1–3), 73–77. https://doi.org/10.1016/j.febslet.2004.07.056

Mwachui, M. A., Crump, L., Hartskeerl, R., Zinsstag, J., & Hattendorf, J. (2015). Environmental and Behavioural Determinants of Leptospirosis Transmission: A Systematic Review. *PLOS*

Neglected Tropical Diseases, *9*(9), e0003843. https://doi.org/10.1371/journal.pntd.0003843

Nahori, M.-A., Fournié-Amazouz, E., Que-Gewirth, N. S., Balloy, V., Chignard, M., Raetz, C. R. H., … Werts, C. (2005). Differential TLR recognition of leptospiral lipid A and lipopolysaccharide in murine and human cells. *Journal of Immunology (Baltimore, Md. : 1950)*, *175*(9), 6022–6031. https://doi.org/10.4049/JIMMUNOL.175.9.6022

Nakamura, S. (2022). Motility of the Zoonotic Spirochete Leptospira: Insight into Association with Pathogenicity. *International Journal of Molecular Sciences*, *23*(3), 1859. https://doi.org/10.3390/ijms23031859

Nakao, R., Masuzawa, T., Nakamura, S., & Koizumi, N. (2021). Complete Genome Sequence of Leptospira kobayashii Strain E30, Isolated from Soil in Japan. *Microbiology Resource Announcements*, *10*(45). https://doi.org/10.1128/MRA.00907-21

Nally, J. E., Bayles, D. O., Hurley, D., Fanning, S., McMahon, B. J., & Arent, Z. (2016). Complete Genome Sequence of Leptospira alstonii Serovar Room22 Strain GWTS #1. *Genome Announcements*, *4*(6). https://doi.org/10.1128/genomeA.01230-16

Nally, J. E., Chow, E., Fishbein, M. C., Blanco, D. R., & Lovett, M. A. (2005). Changes in Lipopolysaccharide O Antigen Distinguish Acute versus Chronic Leptospira interrogans Infections. *Infection And Immunity*, *73*(6), 3251–3260. https://doi.org/10.1128/IAI.73.6.3251–3260.2005

Nascimento, A. L. T. O., Ko, A. I., Martins, E. A. L., Monteiro-Vitorello, C. B., Ho, P. L., Haake, D. A., … Giglioti, E. A. (2004). Comparative Genomics of Two Leptospira interrogans Serovars Reveals Novel Insights into Physiology and Pathogenesis. *Journal Of Bacteriology*, *186*(7), 2164–2172. https://doi.org/10.1128/JB.186.7.2164–2172.2004

Naotunna, C., Agampodi, S. B., & Agampodi, T. C. (2016). Etiological agents causing leptospirosis in Sri Lanka: A review. Asian Pacific journal of tropical medicine, 9(4), 390–394. https://doi.org/10.1016/j.apjtm.2016.03.009

Ndela, E. O., Enault, F., & Toussaint, A. (2021). Transposable Prophages in Leptospira: An Ancient, Now Diverse, Group Predominant in Causative Agents of Weil's Disease. *International Journal of Molecular Sciences Article*. https://doi.org/10.3390/ijms222413434

Nedashkovskaya, O. I., Kim, S.-G., Zhukova, N. V., & Mikhailov, V. V. (2017). Olleya algicola

sp. nov., a marine bacterium isolated from the green alga Ulva fenestrata. *International Journal of Systematic and Evolutionary Microbiology*, *67*(7), 2205–2210. https://doi.org/10.1099/ijsem.0.001926

Nguyen, L. T., Schmidt, H. A., Von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*(1), 268–274. https://doi.org/10.1093/MOLBEV/MSU300

Nieves, C., Ferrés, I., Díaz-Viraqué, F., Buschiazzo, A., Zarantonelli, L., & Iraola, G. (2019). Draft Genome Sequences of 40 Pathogenic Leptospira Strains Isolated from Cattle in Uruguay. *Microbiology Resource Announcements*, *8*(47). https://doi.org/10.1128/MRA.00893-19

Nieves, C., Vincent, A. T., Zarantonelli, L., Picardeau, M., Veyrier, F. J., & Buschiazzo, A. (2022). Horizontal transfer of the rfb cluster in Leptospira is a genetic determinant of serovar identity. Life science alliance, 6(2), e202201480. https://doi.org/10.26508/lsa.202201480

Nishino, K., Nikaido, E., & Yamaguchi, A. (2007). Regulation of Multidrug Efflux Systems Involved in Multidrug and Metal Resistance of Salmonella enterica Serovar Typhimurium. *Journal of Bacteriology*, *189*(24), 9066–9075. https://doi.org/10.1128/JB.01045-07

Nishio, Y., Nakamura, Y., Kawarabayasi, Y., Usuda, Y., Kimura, E., Sugimoto, S., … Gojobori, T. (2003). Comparative Complete Genome Sequence Analysis of the Amino Acid Replacements Responsible for the Thermostability of Corynebacterium efficiens. *Genome Research*, *13*(7), 1572–1579. https://doi.org/10.1101/gr.1285603

Noda, A. A., Grillová, L., Mariet, J. F., Paiffer, N. B., Ruiz, Y. C., Rodríguez, I., Echevarría, E., Obregón, A. M., Lienhard, R., & Picardeau, M. (2020). A first insight into the genomic diversity of Leptospira strains isolated from patients in Cuba. PloS one, 15(2), e0229673. https://doi.org/10.1371/journal.pone.0229673

Noguera Z, L. P., Charypkhan, D., Hartnack, S., Torgerson, P. R., & Rüegg, S. R. (2022). The dual burden of animal and human zoonoses: A systematic review. PLoS neglected tropical diseases, 16(10), e0010540. https://doi.org/10.1371/journal.pntd.0010540

Noguchi, H. (1918). Morphological Characteristics And Nomenclature Of Leptospira (Spirochæta) Icterohæmorrhagiæ (Inada And Ido). *Journal of Experimental Medicine*, *27*(5), 575–592. https://doi.org/10.1084/jem.27.5.575

Ochman, H., & Moran, N. A. (2001). Genes lost and genes found: Evolution of bacterial pathogenesis and symbiosis. *Science*, *292*(5519), 1096–1098. https://doi.org/10.1126/science.1058543

Ogle, D. H., Doll, J. C., Wheeler, A. P., Dinno, A. (2023). FSA: Simple Fisheries Stock Assessment Methods. R package version 0.9.5, https://CRAN.R-project.org/package=FSA.

Oliveira, P. H., Touchon, M., Cury, J., & Rocha, E. P. C. (2017). The chromosomal organization of horizontal gene transfer in bacteria. *Nature Communications*, *8*(1), 841. https://doi.org/10.1038/s41467-017-00808-w

Oni, O., Sujit, K., Kasemsuwan, S., Sakpuaram, T., & Pfeiffer, D. U. (2007). Seroprevalence of leptospirosis in domesticated Asian elephants (Elephas maximus) in north and west Thailand in 2004. The Veterinary record, 160(11), 368–371. https://doi.org/10.1136/vr.160.11.368

Pacciarini, M. L., Savio, M. L., Tagliabue, S., & Rossi, C. (1992). Repetitive Sequences Cloned from Leptospira interrogans Serovar hardjo Genotype hardjoprajitno and Their Application to Serovar Identification. *Journal Of Clinical Microbiology*, 1243–1249.

Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., … Parkhill, J. (2015). Roary: rapid large-scale prokaryote pan genome analysis. *Bioinformatics*, *31*(22), 3691–3693. https://doi.org/10.1093/bioinformatics/btv421

Pal, M., Kumari, M., Kiran, S., Salwan, R., Mayilraj, S., Chhibber, S., & Gulati, A. (2018). Chryseobacterium glaciei sp. nov., isolated from the surface of a glacier in the Indian trans-Himalayas. *International Journal of Systematic and Evolutionary Microbiology*, *68*(3), 865–870. https://doi.org/10.1099/ijsem.0.002600

Pandey, A., Humbert, M. V., Jackson, A., Passey, J. L., Hampson, D. J., Cleary, D. W., … Christodoulides, M. (2020). Evidence of homologous recombination as a driver of diversity in Brachyspira pilosicoli. https://doi.org/10.1099/mgen.0.000470

Pappas, C. J., & Picardeau, M. (2015). Control of Gene Expression in Leptospira spp. by Transcription Activator-Like Effectors Demonstrates a Potential Role for LigA and LigB in Leptospira interrogans Virulence. *Applied and Environmental Microbiology*, *81*(22), 7888–7892. https://doi.org/10.1128/AEM.02202-15

Park, S., Choi, J., Choi, S. J., & Yoon, J.-H. (2018). Flavobacterium sediminilitoris sp. nov.,

isolated from a tidal flat. *International Journal of Systematic and Evolutionary Microbiology*, *68*(2), 630–635. https://doi.org/10.1099/ijsem.0.002555

Parvathy, S. T., Udayasuriyan, V., & Bhadana, V. (2022). Codon usage bias. *Molecular Biology Reports*, *49*(1), 539–565. https://doi.org/10.1007/s11033-021-06749-4

Paster, B. J., & Dewhirst, F. E. (2000). Phylogenetic foundation of spirochetes. *Journal of Molecular Microbiology and Biotechnology*, *2*(4), 341–344. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/11075904

Paster, B. J., Dewhirst, F. E., Weisburg, W. G., Tordoff, L. A., Fraser, G. J., Hespell, R. B., … Woese, C. R. (1991). Phylogenetic analysis of the spirochetes. *Journal of Bacteriology*, *173*(19), 6101–6109. https://doi.org/10.1128/jb.173.19.6101-6109.1991

Patra, K. P., Choudhury, B., Matthias, M. M., Baga, S., Bandyopadhya, K., & Vinetz, J. M. (2015). Comparative analysis of lipopolysaccharides of pathogenic and intermediately pathogenic Leptospira species. https://doi.org/10.1186/s12866-015-0581-7

Paz, L. N., Hamond, C., Dias, C. S., Curvelo, V. P., Medeiros, M. A., Oriá, A. P., & Pinna, M. H. (2019). Detection of Leptospira spp. in Captive Broad-Snouted Caiman (Caiman latirostris). *EcoHealth*, *16*(4), 694–700. https://doi.org/10.1007/S10393-019-01452-0

Peláez Sánchez, R. G., Quintero, J. Á. L., Pereira, M. M., & Agudelo-Flórez, P. (2017). High-Resolution Melting Curve Analysis of the 16S Ribosomal Gene to Detect and Identify Pathogenic and Saprophytic Leptospira Species in Colombian Isolates. The American journal of tropical medicine and hygiene, 96(5), 1031–1038. https://doi.org/10.4269/ajtmh.16-0312

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 Microbiology Letters*, *177*(2), 319–326. https://doi.org/10.1111/J.1574-6968.1999.TB13749.X

Pereira, M. M., Schneider, M. C., Munoz-Zanzi, C., Costa, F., Benschop, J., Hartskeerl, R., Martinez, J., Jancloes, M., & Bertherat, E. (2018). A road map for leptospirosis research and health policies based on country needs in Latin America. Revista panamericana de salud publica = Pan American journal of public health, 41, e131.

https://doi.org/10.26633/RPSP.2017.131

Perez, J., & Goarant, C. (2010). Rapid Leptospira identification by direct sequencing of the diagnostic PCR products in New Caledonia. *BMC Microbiology*, *10*(1), 325. https://doi.org/10.1186/1471-2180-10-325

Perolat, P., Chappel, R. J., Adler, B., Baranton, G., Bulach, D. M., Billinghurst, M. L., … Serrano, M. S. (1998). Leptospira fainei sp. nov., isolated from pigs in Australia. *International Journal of Systematic Bacteriology*, *48*(3), 851–858. https://doi.org/10.1099/00207713-48-3-851

Peters, A., Vokaty, A., Portch, R., & Gebre, Y. (2017). Leptospirosis in the Caribbean: a literature review. Revista panamericana de salud publica = Pan American journal of public health, 41, e166. https://doi.org/10.26633/RPSP.2017.166

Peterson, J. (2010). Bacterial Pathogenesis. In *Bacteria versus Antibacterial Agents* (pp. 33–42). Washington, DC, USA: ASM Press. https://doi.org/10.1128/9781555817794.ch2

Picardeau, M. (2017). Virulence of the zoonotic agent of leptospirosis: still terra incognita? *Nature Reviews. Microbiology*, *15*(5), 297–307. https://doi.org/10.1038/NRMICRO.2017.5

Picardeau, M., Bulach, D. M., Bouchier, C., Zuerner, R. L., Zidane, N., Wilson, P. J., … Adler, B. (2008). Genome sequence of the saprophyte Leptospira biflexa provides insights into the evolution of Leptospira and the pathogenesis of leptospirosis. *PloS One*, *3*(2), e1607. https://doi.org/10.1371/journal.pone.0001607

Piñeiro-Vidal, M., Gijón, D., Zarza, C., & Santos, Y. (2012). Tenacibaculum dicentrarchi sp. nov., a marine bacterium of the family Flavobacteriaceae isolated from European sea bass. *International Journal of Systematic and Evolutionary Microbiology*, *62*(2), 425–429. https://doi.org/10.1099/ijs.0.025122-0

Pinne, M., & Haake, D. A. (2009). A Comprehensive Approach to Identification of Surface-Exposed, Outer Membrane-Spanning Proteins of Leptospira interrogans. https://doi.org/10.1371/journal.pone.0006071

Prescott, J. (2008). Public health: Canine leptospirosis in Canada: a veterinarian's perspective. *CMAJ : Canadian Medical Association Journal*, *178*(4), 397. https://doi.org/10.1503/CMAJ.071092

Puche, R., Ferrés, I., Caraballo, L., Rangel, Y., Picardeau, M., Takiff, H., & Iraola, G. (2018). Leptospira venezuelensis sp. nov., a new member of the intermediate group isolated from rodents, cattle and humans. *International Journal of Systematic and Evolutionary Microbiology*, *68*(2), 513–517. https://doi.org/10.1099/ijsem.0.002528

Pupo, G. M., Lan, R., & Reeves, P. R. (2000). Multiple independent origins of Shigella clones of Escherichia coli and convergent evolution of many of their characteristics. *Proceedings of the National Academy of Sciences*, *97*(19), 10567–10572. https://doi.org/10.1073/pnas.180094797

Que-Gewirth, N. L. S., Ribeiro, A. A., Kalb, S. R., Cotter, R. J., Bulach, D. M., || B. A., … Raetz, C. R. H. (2004). A Methylated Phosphate Group and Four Amide-linked Acyl Chains in Leptospira interrogans Lipid A. The Membrane Anchor of an Unusual Lipopolysaccharide that Activates TLR2. *J Biol Chem*, *279*(24), 25420–25429. https://doi.org/10.1074/jbc.M400598200

Raddi, G., Morado, D. R., Yan, J., Haake, D. A., Yang, X. F., & Liu, J. (2012). Three-Dimensional Structures of Pathogenic and Saprophytic Leptospira Species Revealed by Cryo-Electron Tomography. *Journal of Bacteriology*, *194*(6), 1299. https://doi.org/10.1128/JB.06474-11

Raetz, C. R. H., & Whitfield, C. (2002). Lipopolysaccharide endotoxins. *Annual Review of Biochemistry*, *71*, 635–700. https://doi.org/10.1146/Annurev.Biochem.71.110601.135414

Ramadass, P., Meerarani, S., Venkatesha, M. D., Senthilkumar, A., & Nachimuthu, K. (1997). Characterization of leptospiral serovars by randomly amplified polymorphic DNA fingerprinting. *International Journal of Systematic Bacteriology*, *47*(2), 575–576. https://doi.org/10.1099/00207713-47-2-575

Ratet, G., Santecchia, I., Fanton d'Andon, M., Vernel-Pauillac, F., Wheeler, R., Lenormand, P., … Werts, C. (2017). LipL21 lipoprotein binding to peptidoglycan enables Leptospira interrogans to escape NOD1 and NOD2 recognition. *PLoS Pathogens*, *13*(12). https://doi.org/10.1371/Journal.Ppat.1006725

R Core Team (2018) R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna. https://www.R-project.org

Ren, S.-X., Fu, G., Jiang, X.-G., Zeng, R., Miao, Y.-G., Xu, H., … Zhao, G.-P. (2003). Unique

physiological and pathogenic features of Leptospira interrogans revealed by whole-genome sequencing. *Nature*, *422*(6934), 888–893. https://doi.org/10.1038/nature01597

Ricaldi, J. N., Fouts, D. E., Selengut, J. D., Harkins, D. M., Patra, K. P., Moreno, A., … Matthias, M. A. (2012). Whole Genome Analysis of Leptospira licerasiae Provides Insight into Leptospiral Evolution and Pathogenicity. https://doi.org/10.1371/journal.pntd.0001853

Rice, P., Longden, L., & Bleasby, A. (2000). EMBOSS: the European Molecular Biology Open Software Suite. *Trends in Genetics : TIG*, *16*(6), 276–277. https://doi.org/10.1016/S0168-9525(00)02024-2

Richter, M., & Rosselló-Móra, R. (2009). Shifting the genomic gold standard for the prokaryotic species definition. *Proceedings of the National Academy of Sciences*, *106*(45), 19126–19131. https://doi.org/10.1073/pnas.0906412106

Ristow, P., Bourhy, P., McBride, F. W. da C., Figueira, C. P., Huerre, M., Ave, P., … Picardeau, M. (2007). The OmpA-Like Protein Loa22 Is Essential for Leptospiral Virulence. *PLoS Pathogens*, *3*(7), e97. https://doi.org/10.1371/journal.ppat.0030097

Rodamilans, G. M., Fonseca, M. S., Paz, L. N., Fernandez, C. C., Biondi, I., Lira-da-Silva, R. M., … Portela, R. D. (2020). Leptospira interrogans in wild Boa constrictor snakes from Northeast Brazil peri-urban rainforest fragments. *Acta Tropica*, *209*. https://doi.org/10.1016/J.ACTATROPICA.2020.105572

Rosselló-Móra, R., Urdiain, M., & López-López, A. (2011). DNA–DNA Hybridization (pp. 325–347). https://doi.org/10.1016/B978-0-12-387730-7.00015-2

Ryjenkov, D. A., Tarutina, M., Moskvin, O. V., & Gomelsky, M. (2005). Cyclic Diguanylate Is a Ubiquitous Signaling Molecule in Bacteria: Insights into Biochemistry of the GGDEF Protein Domain. *Journal of Bacteriology*, *187*(5), 1792–1798. https://doi.org/10.1128/JB.187.5.1792-1798.2005

Safiee, A. W. M., Mohd Ali, M. R., Zoqratt, M. Z. H. M., Siew, T. H., Chuan, C. W., Huey, L. L., … Ismail, N. (2022). Putative Pathogenic Genes of Leptospira interrogans and Leptospira weilii Isolated from Patients with Acute Febrile Illness. *Tropical Medicine and Infectious Disease*, *7*(10), 284. https://doi.org/10.3390/tropicalmed7100284

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. *Journal of Clinical Microbiology*, *44*(11), 3954–3962. https://doi.org/10.1128/JCM.00336-06

Samuel, G., & Reeves, P. (2003). Biosynthesis of O-antigens: genes and pathways involved in nucleotide sugar precursor synthesis and O-antigen assembly. *Carbohydrate Research*, *338*(23), 2503–2519. https://doi.org/10.1016/j.carres.2003.07.009

San Martin, F., Fule, L., Iraola, G., Buschiazzo, A., & Picardeau, M. (2022). Diving into the complexity of the spirochetal endoflagellum. *Trends in Microbiology*. https://doi.org/10.1016/j.tim.2022.09.010

Santecchia, I., Ferrer, M. F., Vieira, M. L., Gómez, R. M., & Werts, C. (2020). Phagocyte Escape of Leptospira: The Role of TLRs and NLRs. *Frontiers in Immunology*, *11*. https://doi.org/10.3389/FIMMU.2020.571816

Santos, L. A., Adhikarla, H., Yan, X., Wang, Z., Fouts, D. E., Vinetz, J. M., … Wunder, E. A. (2018). Genomic Comparison Among Global Isolates of L. interrogans Serovars Copenhageni and Icterohaemorrhagiae Identified Natural Genetic Variation Caused by an Indel. *Frontiers in Cellular and Infection Microbiology*, *8*(JUN). https://doi.org/10.3389/FCIMB.2018.00193

Schmid, G. P., Steere, A. C., Kornblatt, A. N., Kaufmann, A. F., Moss, C. W., Johnson, R. C., … Brenner, D. J. (1986). Newly recognized Leptospira species ("Leptospira inadai" serovar lyme) isolated from human skin. *Journal of Clinical Microbiology*, *24*(3), 484–486. https://doi.org/10.1128/jcm.24.3.484-486.1986

Schneider, M. C., Leonel, D. G., Hamrick, P. N., de Caldas, E. P., Velásquez, R. T., Mendigaña Paez, F. A., González Arrebato, J. C., Gerger, A., Maria Pereira, M., & Aldighieri, S. (2017). Leptospirosis in Latin America: exploring the first set of regional data. Revista panamericana de salud publica = Pan American journal of public health, 41, e81. https://doi.org/10.26633/RPSP.2017.81

Scior, T., Lozano-Aponte, J., Figueroa-Vazquez, V., Yunes-Rojas, J. A., Zähringer, U., & Alexander, C. (2013). Three-Dimensional Mapping Of Differential Amino Acids Of Human, Murine, Canine And Equine TLR4/MD-2 Receptor Complexes Conferring Endotoxic Activation By Lipid A, Antagonism By Eritoran And Species-Dependent Activities Of Lipid

IVA In The Mammalian LPS Sensor System. *Computational and Structural Biotechnology Journal*, *7*(9), e201305003. https://doi.org/10.5936/csbj.201305003

Scola, B. La, Bui, L. T. M., Baranton, G., Khamis, A., & Raoult, D. (2006). Partial rpoB gene sequencing for identification of Leptospira species. *FEMS Microbiology Letters*, *263*(2), 142–147. https://doi.org/10.1111/j.1574-6968.2006.00377.x

Sebastián, I., Okura, N., Humbel, B. M., Xu, J., Hermawan, I., Matsuura, C., … Toma, C. (2021). Disassembly of the apical junctional complex during the transmigration of <scp> *Leptospira interrogans* </scp> across polarized renal proximal tubule epithelial cells. *Cellular Microbiology*, *23*(9). https://doi.org/10.1111/cmi.13343

Seemann, T. (2014). Prokka: rapid prokaryotic genome annotation. *Bioinformatics*, *30*(14), 2068–2069. https://doi.org/10.1093/bioinformatics/btu153

Seong, H. J., Han, S. W., & Sul, W. J. (2021). Prokaryotic DNA methylation and its functional roles. *Journal of Microbiology*, *59*(3), 242–248. https://doi.org/10.1007/s12275-021-0674-y

Shimura, Y., Hirose, Y., Misawa, N., Osana, Y., Katoh, H., Yamaguchi, H., & Kawachi, M. (2015). Comparison of the terrestrial cyanobacterium Leptolyngbya sp. NIES-2104 and the freshwater Leptolyngbya boryana PCC 6306 genomes. *DNA Research*, *22*(6), 403–412. https://doi.org/10.1093/dnares/dsv022

Siguier, P., Gourbeyre, E., & Chandler, M. (2014). Bacterial insertion sequences: their genomic impact and diversity. *FEMS Microbiology Reviews*, *38*(5), 865–891. https://doi.org/10.1111/1574-6976.12067

Silva, É. F., Brod, C. S., Cerqueira, G. M., Bourscheidt, D., Seyffert, N., Queiroz, A., … Dellagostin, O. A. (2007). Isolation of Leptospira noguchii from sheep. *Veterinary Microbiology*, *121*(1–2), 144–149. https://doi.org/10.1016/J.VETMIC.2006.11.010

Silva, É. F., Cerqueira, G. M., Seyffert, N., Seixas, F. K., Hartwig, D. D., Athanazio, D. A., … Dellagostin, O. A. (2009). Leptospira noguchii and human and animal leptospirosis, Southern Brazil. *Emerging Infectious Diseases*, *15*(4), 621–623. https://doi.org/10.3201/EID1504.071669

Slack, A. T., Symonds, M. L., Dohnt, M. F., & Smythe, L. D. (2006). Identification of pathogenic Leptospira species by conventional or real-time PCR and sequencing of the DNA gyrase

subunit B encoding gene. *BMC Microbiology*, *6*. https://doi.org/10.1186/1471-2180-6-95

Slamti, L., De Pedro, M. A., Guichet, E., & Picardeau, M. (2011). Deciphering Morphological Determinants of the Helix-Shaped Leptospira. *JOURNAL OF BACTERIOLOGY*, *193*(22), 6266–6275. https://doi.org/10.1128/JB.05695-11

Sperandeo, P., Dehò, G., & Polissi, A. (2009). The lipopolysaccharide transport system of Gram-negative bacteria. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids*, *1791*(7), 594–602. https://doi.org/10.1016/j.bbalip.2009.01.011

Srikram, A., Zhang, K., Bartpho, T., Lo, M., Hoke, D. E., Sermswan, R. W., … Murray, G. L. (2011). Cross-protective Immunity Against Leptospirosis Elicited by a Live, Attenuated Lipopolysaccharide Mutant. *The Journal of Infectious Diseases*, *203*(6), 870–879. https://doi.org/10.1093/infdis/jiq127

Stackebrandt, E., & Goebel, B. M. (1994). Taxonomic Note: A Place for DNA-DNA Reassociation and 16S rRNA Sequence Analysis in the Present Species Definition in Bacteriology. *International Journal of Systematic and Evolutionary Microbiology*, *44*(4), 846–849. https://doi.org/10.1099/00207713-44-4-846

Sullivan, M. J., Petty, N. K., & Beatson, S. A. (2011). Easyfig: a genome comparison visualizer. *Bioinformatics (Oxford, England)*, *27*(7), 1009–1010. https://doi.org/10.1093/Bioinformatics/BTR039

Suut, L., Mazlan, M. N., Arif, M. T., Yusoff, H., Abdul Rahim, N. A., Safii, R., & Suhaili, M. R. (2016). Serological Prevalence of Leptospirosis Among Rural Communities in the Rejang Basin, Sarawak, Malaysia. Asia-Pacific journal of public health, 28(5), 450–457. https://doi.org/10.1177/1010539516648003

Suwancharoen, D., Chaisakdanugull, Y., Thanapongtharm, W., & Yoshida, S. (2013). Serological survey of leptospirosis in livestock in Thailand. Epidemiology and infection, 141(11), 2269–2277. https://doi.org/10.1017/S0950268812002981

Syberg-Olsen, M. J., Garber, A. I., Keeling, P. J., McCutcheon, J. P., & Husnik, F. (2022). Pseudofinder: Detection of Pseudogenes in Prokaryotic Genomes. *Molecular Biology and Evolution*, *39*(7). https://doi.org/10.1093/molbev/msac153

Szklarczyk, D., Gable, A. L., Lyon, D., Junge, A., Wyder, S., Huerta-Cepas, J., … Mering, C. von.

(2019). STRING v11: protein–protein association networks with increased coverage, supporting functional discovery in genome-wide experimental datasets. *Nucleic Acids Research*, *47*(D1), D607–D613. https://doi.org/10.1093/nar/gky1131

Tahara, H., Takabe, K., Sasaki, Y., Kasuga, K., Kawamoto, A., Koizumi, N., & Nakamura, S. (2018). The mechanism of two-phase motility in the spirochete Leptospira: Swimming and crawling. Retrieved from https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5976277/pdf/aar7975.pdf

Takabe, K., Nakamura, S., Ashihara, M., & Kudo, S. (2013). Effect of osmolarity and viscosity on the motility of pathogenic and saprophytic Leptospira. *Microbiology and Immunology*, *57*(3), 236–239. https://doi.org/10.1111/1348-0421.12018

Takahashi, M. B., Teixeira, A. F. & Nascimento, A. L. T. O. (2021). The leptospiral LipL21 and LipL41 proteins exhibit a broad spectrum of interactions with host cell components. Virulence, 12(1), 2798–2813. https://doi.org/10.1080/21505594.2021.1993427

Tamai, T., Sada, E., & Kobayashi, Y. (1988). Restriction Endonuclease DNA Analysis of Leptospira interrogans Serovars Icterohaemorrhagiae and Copenhageni. *Microbiology and Immunology*, *32*(9), 887–894. https://doi.org/10.1111/J.1348-0421.1988.TB01450.X

Tanizawa, Y., Fujisawa, T., & Nakamura, Y. (2018). DFAST: a flexible prokaryotic genome annotation pipeline for faster genome publication. *Bioinformatics (Oxford, England)*, *34*(6), 1037–1039. https://doi.org/10.1093/bioinformatics/btx713

Tettelin, H., Riley, D., Cattuto, C., & Medini, D. (2008). Comparative genomics: the bacterial pan-genome. *Current Opinion in Microbiology*, *11*(5), 472–477. https://doi.org/10.1016/J.MIB.2008.09.006

Thacker, W. L., Dyke, J. W., Benson, R. F., Havlichek, D. H., Robinson-Dunn, B., Stiefel, H., … Brenner, D. J. (1992). Legionella lansingensis sp. nov. isolated from a patient with pneumonia and underlying chronic lymphocytic leukemia. *Journal of Clinical Microbiology*, *30*(9), 2398–2401. https://doi.org/10.1128/jcm.30.9.2398-2401.1992

Thiangtrongjit, T., Reamtong, O., Phumratanapraphin, W., & Limmathurotsakul, D. (2022). Extracellular Secretomes of Leptospira Interrogans Serovar Ic- terohemorrhagiae Respond to the In Vivo Mimic of Physiologi- cal Osmolarity and Temperature Transition, (October), 1–

19. https://doi.org/10.20944/preprints202210.0099.v1

Thibeaux, R., Girault, D., Bierque, E., Soupé-Gilbert, M. E., Rettinger, A., Douyère, A., Meyer, M., Iraola, G., Picardeau, M., & Goarant, C. (2018). Biodiversity of Environmental Leptospira: Improving Identification and Revisiting the Diagnosis. Frontiers in microbiology, 9, 816. https://doi.org/10.3389/fmicb.2018.00816

Thibeaux, R., Iraola, G., Ferrés, I., Bierque, E., Girault, D., Soupé-Gilbert, M.-E., … Goarant, C. (2018). Deciphering the unexplored Leptospira diversity from soils uncovers genomic evolution to virulence. *Microbial Genomics*, *4*(1). https://doi.org/10.1099/mgen.0.000144

Thiermann, A. B., Handsaker, A. L., Moseley,'t And, S. L., & Kingscote2, B. (1985). New Method for Classification of Leptospiral Isolates Belonging to Serogroup Pomona by Restriction Endonuclease Analysis: Serovar kennewicki. *JOURNAL OF CLINICAL MICROBIOLOGY*, 585–587.

Thrane, S. W., Taylor, V. L., Freschi, L., Kukavica-Ibrulj, I., Boyle, B., Laroche, J., … Jelsbaka, L. (2015). The Widespread Multidrug-Resistant Serotype O12 Pseudomonas aeruginosa Clone Emerged through Concomitant Horizontal Transfer of Serotype Antigen and Antibiotic Resistance Gene Clusters. *MBio*, *6*(5). https://doi.org/10.1128/MBIO.01396-15

Torgerson, P. R., Hagan, J. E., Costa, F., Calcagno, J., Kane, M., Martinez-Silveira, M. S., … Abela-Ridder, B. (2015). Global Burden of Leptospirosis: Estimated in Terms of Disability Adjusted Life Years. *PLoS Neglected Tropical Diseases*, *9*(10). https://doi.org/10.1371/JOURNAL.PNTD.0004122

Valvano, M. A. (2015). Genetics and Biosynthesis of Lipopolysaccharide. In *Molecular Medical Microbiology* (pp. 55–89). Elsevier. https://doi.org/10.1016/B978-0-12-397169-2.00004-4

Valverde, M. L. A., Goris, M. G. A., González, V., Anchia, M. E., Díaz, P., Ahmed, A., & Hartskeerl, R. A. (2013). New serovars of Leptospira isolated from patients in Costa Rica: implications for public health. Journal of medical microbiology, 62(Pt 9), 1263–1271. https://doi.org/10.1099/jmm.0.058545-0

Valverde, M.deL., Ramírez, J. M., Montes de Oca, L. G., Goris, M. G., Ahmed, N., & Hartskeerl, R. A. (2008). Arenal, a new Leptospira serovar of serogroup Javanica, isolated from a patient in Costa Rica. Infection, genetics and evolution : journal of molecular epidemiology and

evolutionary genetics in infectious diseases, 8(5), 529–533. https://doi.org/10.1016/j.meegid.2008.02.008

Van Eys, G. J. J. M., Zaal, J., Schoone, G. J., & Terpstra, W. J. (1988). DNA hybridization with hardjobovis-specific recombinant probes as a method for type discrimination of Leptospira interrogans serovar hardjo. *Journal of General Microbiology*, *134*(3), 567–574. https://doi.org/10.1099/00221287-134-3-567

Vanithamani, S., Akino Mercy, C. S., Kanagavel, M., Sumaiya, K., Bothammal, P., Saranya, P., … Natarajaseenivasan, K. (2021). Biochemical analysis of leptospiral LPS explained the difference between pathogenic and non-pathogenic serogroups. *Microbial Pathogenesis*, *152*, 104738. https://doi.org/10.1016/J.MICPATH.2021.104738

Varni, V., Ruybal, P., Lauthier, J. J., Tomasini, N., Brihuega, B., Koval, A., & Caimi, K. (2014). Reassessment of MLST schemes for Leptospira spp. typing worldwide. *Infection, Genetics and Evolution*, *22*, 216–222. https://doi.org/10.1016/j.meegid.2013.08.002

Veyrier, F., Pletzer, D., Turenne, C., & Behr, M. A. (2009). Phylogenetic detection of horizontal gene transfer during the step-wise genesis of Mycobacterium tuberculosis. *BMC Evolutionary Biology*, *9*, 196. https://doi.org/10.1186/1471-2148-9-196

Vieira, M. L., Teixeira, A. F., Pidde, G., Ching, A. T. C., Tambourgi, D. V., Nascimento, A. L. T. O., & Herwald, H. (2018). Leptospira interrogans outer membrane protein LipL21 is a potent inhibitor of neutrophil myeloperoxidase. *Virulence*, *9*(1), 414–425. https://doi.org/10.1080/21505594.2017.1407484

Vincent, A. T., Schiettekatte, O., Goarant, C., Neela, V. K., Bernet, E., Thibeaux, R., … Picardeau, M. (2019). Revisiting the taxonomy and evolution of pathogenicity of the genus Leptospira through the prism of genomics. *PLoS Neglected Tropical Diseases*, *13*(5). https://doi.org/10.1371/JOURNAL.PNTD.0007270

Wang, B., Zhou, J., Jin, Y., Hu, M., Zhao, Y., Wang, X., … Ren, H. (2022). A Standardized Framework for Better Understanding of Phenotypic Differences within Bacterial Phyla Based on Protein, (May).

Wang, H. K., Lee, M. H., Chen, Y. C., Hsueh, P. R., & Chang, S. C. (2020). Factors associated with severity and mortality in patients with confirmed leptospirosis at a regional hospital in

northern Taiwan. Journal of microbiology, immunology, and infection, 53(2), 307–314. https://doi.org/10.1016/j.jmii.2018.05.005

Wang, X., & Quinn, P. J. (2010). Endotoxins: lipopolysaccharides of gram-negative bacteria. *Sub-Cellular Biochemistry*, *53*, 3–25. https://doi.org/10.1007/978-90-481-9078-2_1

Wang, X., & Quinn, P. J. (2010). Lipopolysaccharide: Biosynthetic pathway and structure modification. *Progress in Lipid Research*, *49*(2), 97–107. https://doi.org/10.1016/j.plipres.2009.06.002

Wang, Y., Zhuang, X., Zhong, Y., Zhang, C., Zhang, Y., Zeng, L., … Qin, J. (2015). Distribution of Plasmids in Distinct Leptospira Pathogenic Species. *PLOS Neglected Tropical Diseases*, *9*(11), e0004220. https://doi.org/10.1371/journal.pntd.0004220

Wangroongsarb, P., Chanket, T., Gunlabun, K., Long, D. H., Satheanmethakul, P., Jetanadee, S., … Kalambaheti, T. (2007). Molecular typing of Leptospira spp. based on putative O-antigen polymerase gene (wzy), the benefit over 16S rRNA gene sequence. *FEMS Microbiology Letters*, *271*(2), 170–179. https://doi.org/10.1111/J.1574-6968.2007.00711.X

Warshawsky, B., Lindsay, L. R., & Artsob, H. (2000). Leptospira infections in trappers from Ontario. *The Canadian Journal of Infectious Diseases*, *11*(1), 47. https://doi.org/10.1155/2000/392419

Werts, C. (2010). Leptospirosis: A toll road from B lymphocytes. *Chang Gung Medical Journal*, *33*(6), 591–601.

Werts, C., Tapping, R. I., Mathison, J. C., Chuang, T.-H., Kravchenko, V., Saint Girons, I., … Ulevitch, R. J. (2001). Leptospiral lipopolysaccharide activates cells through a TLR2-dependent mechanism. *Nature Immunology*, *2*(4), 346–352. https://doi.org/10.1038/86354

Wexler, H. M. (2007). Bacteroides : the Good, the Bad, and the Nitty-Gritty. *Clinical Microbiology Reviews*, *20*(4), 593–621. https://doi.org/10.1128/CMR.00008-07

Wick, R. R., Judd, L. M., Cerdeira, L. T., Hawkey, J., Méric, G., Vezina, B., … Holt, K. E. (2021). Trycycler: consensus long-read assemblies for bacterial genomes. *Genome Biology*, *22*(1), 1–17. https://doi.org/10.1186/s13059-021-02483-z

Wick, R. R., Judd, L. M., Gorrie, C. L., & Holt, K. E. (2017). Unicycler: Resolving bacterial

genome assemblies from short and long sequencing reads. *PLoS Computational Biology*, *13*(6). https://doi.org/10.1371/JOURNAL.PCBI.1005595

Wilson, M. R., Naccache, S. N., Samayoa, E., Biagtan, M., Bashir, H., Yu, G., Salamat, S. M., Somasekar, S., Federman, S., Miller, S., Sokolic, R., Garabedian, E., Candotti, F., Buckley, R. H., Reed, K. D., Meyer, T. L., Seroogy, C. M., Galloway, R., Henderson, S. L., Gern, J. E., … Chiu, C. Y. (2014). Actionable diagnosis of neuroleptospirosis by next-generation sequencing. The New England journal of medicine, 370(25), 2408–2417. https://doi.org/10.1056/NEJMoa1401268

Wolbach, S. B., & Binger, C. A. (1914). Notes on a filterable Spirochete from fresh Water. Spirocheta biflexa (new Species). *The Journal of Medical Research*, *30*(1), 23-26.1. Retrieved from http://www.ncbi.nlm.nih.gov/pubmed/19972157

Woodford, N., & Ellington, M. J. (2007). The emergence of antibiotic resistance by mutation. *Clinical Microbiology and Infection*, *13*(1), 5–18. https://doi.org/10.1111/j.1469-0691.2006.01492.x

Wunder Jr, E. A., Figueira, C. P., Benaroudj, N., Hu, B., Tong, B. A., Trajtenberg, F., … Ko, A. I. (2016). A novel flagellar sheath protein, FcpA, determines filament coiling, translational motility and virulence for the Leptospira spirochete HHS Public Access. *Mol Microbiol*, *101*(3), 457–470. https://doi.org/10.1111/mmi.13403

Xiao, G., Kong, L., Che, R., Yi, Y., Zhang, Q., Yan, J., & Lin, X. (2018). Identification and Characterization of c-di-GMP Metabolic Enzymes of Leptospira interrogans and c-di-GMP Fluctuations After Thermal Shift and Infection. *Frontiers in Microbiology*, *9*. https://doi.org/10.3389/fmicb.2018.00764

Xu, J., Koizumi, N., & Nakamura, S. (2020). Crawling Motility on the Host Tissue Surfaces Is Associated With the Pathogenicity of the Zoonotic Spirochete Leptospira. *Frontiers in Microbiology*, *11*. https://doi.org/10.3389/FMICB.2020.01886/FULL

Xu, Y., Zhu, Y., Wang, Y., Chang, Y.-F., Zhang, Y., Jiang, X., … Wang, J. (2015). Whole genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic Leptospira. *Nature Publishing Group*. https://doi.org/10.1038/srep20020

Xue, F., Dong, H., Wu, J., Wu, Z., Hu, W., Sun, A., … Yan, J. (2010). Transcriptional responses

of Leptospira interrogans to host innate immunity: significant changes in metabolism, oxygen tolerance, and outer membrane. *PLoS Neglected Tropical Diseases*, *4*(10), e857. https://doi.org/10.1371/journal.pntd.0000857

Xue, F., Yan, J., & Picardeau, M. (2009). Evolution and pathogenesis of Leptospira spp.: lessons learned from the genomes. *Microbes and Infection*, *11*(3), 328–333. https://doi.org/10.1016/j.micinf.2008.12.007

Yang, C.-W., Hung, C.-C., Wu, M.-S., Tian, Y.-C., Chang, C.-T., Pan, M.-J., & Vandewalle, A. (2006). Toll-like receptor 2 mediates early inflammation by leptospiral outer membrane proteins in proximal tubule cells. *Kidney International*, *695000119*, 815–822. https://doi.org/10.1038/sj.ki.5000119

Yasuda, P. H., Steigerwalt, A. G., Sulzer, K. R., Kaufmann, A. F., Rogers, F., & Brenner, D. J. (1987). Deoxyribonucleic Acid Relatedness between Serogroups and Serovars in the Family Leptospiraceae with Proposals for Seven New Leptospira Species. *International Journal of Systematic and Evolutionary Microbiology*, *37*(4), 407–415. https://doi.org/10.1099/00207713-37-4-407

Yoon, S. H., Ha, S. min, Lim, J., Kwon, S., & Chun, J. (2017). A large-scale evaluation of algorithms to calculate average nucleotide identity. *Antonie van Leeuwenhoek*, *110*(10), 1281–1286. https://doi.org/10.1007/S10482-017-0844-4

Yu, Y., Ouyang, Y., & Yao, W. (2018). Genome analysis shinyCircos : an R / Shiny application for interactive creation of Circos plot, *34*(November 2017), 1229–1231. https://doi.org/10.1093/bioinformatics/btx763

Zarantonelli, L., Suanes, A., Meny, P., Buroni, F., Nieves, C., Salaberry, X., … Grupo de Trabajo Interinstitucional de Leptospirosis Consortium, on behalf of the G. de T. I. de L. (2018). Isolation of pathogenic Leptospira strains from naturally infected cattle in Uruguay reveals high serovar diversity, and uncovers a relevant risk for human leptospirosis. *PLoS Neglected Tropical Diseases*, *12*(9), e0006694. https://doi.org/10.1371/journal.pntd.0006694

Zavala-Alvarado, C., G. Huete, S., Vincent, A. T., Sismeiro, O., Legendre, R., Varet, H., … Benaroudj, N. (2021). The oxidative stress response of pathogenic Leptospira is controlled by two peroxide stress regulators which putatively cooperate in controlling virulence. *PLOS*

Pathogens, *17*(12), e1009087. https://doi.org/10.1371/journal.ppat.1009087

Zhu, W.-N., Huang, L.-L., Zeng, L.-B., Zhuang, X.-R., Chen, C.-Y., Wang, Y.-Z., … Guo, X.-K. (2014). Isolation and Characterization of Two Novel Plasmids from Pathogenic Leptospira interrogans Serogroup Canicola Serovar Canicola Strain Gui44. *PLoS Neglected Tropical Diseases*, *8*(8), e3103. https://doi.org/10.1371/journal.pntd.0003103

Zhu, W., Wang, J., Zhu, Y., Tang, B., Zhang, Y., He, P., … Qin, J. (2015). Identification of three extra-chromosomal replicons in Leptospira pathogenic strain and development of new shuttle vectors. *BMC Genomics*, *16*(1), 90. https://doi.org/10.1186/s12864-015-1321-y

Zückert, W. R. (2019). Protein Secretion in Spirochetes. *Microbiology Spectrum*, *7*(3). https://doi.org/10.1128/microbiolspec.PSIB-0026-2019

Zuerner, R. L., & Bolin, C. A. (1990). Nucleic acid probe characterizes Leptospira interrogans serovars by restriction fragment length polymorphisms. *Veterinary Microbiology*, *24*(3–4), 355–366. https://doi.org/10.1016/0378-1135(90)90183-V

Zuerner, Richard L, Herrmann, J. L., & Saint, I. (1993). Comparison of Genetic Maps for Two Leptospira interrogans Serovars Provides Evidence for Two Chromosomes and Intraspecies Heterogeneity, *175*(17), 5445–5451.

Appendix I. Supplementary material of Chapter 3

Table S3.1. General features of the 68 *Leptospira* species.

Strain	Length (bp)	GC (%)
L. abararensis 201903074	4,176,760	39.09
L. adleri FH2-B-C1	4,388,198	43.82
L. ainazelensis 201903071	4,864,687	42.61
L. ainlahdjerensis 201903070	4,843,892	42.54
L. alexanderi L60	4,223,825	40.20
L. alstonii 79601	4,436,408	42.48
L. andrefontaineae 201800301	4,281,887	39.93
L. bandrabouensis 201601111	4,027,579	37.88
L. barantonii FH4-C-A1	4,392,226	43.96
L. biflexa Patoc 1 (Paris)	3,951,448	38.89
L. borgpetersenii L550	3,931,782	40.23
L. bourretii 201800280	4,235,722	38.24
L. bouyouniensis 201601297	4,100,043	37.06
L. brenneri JW2_C_A2	4,113,151	38.34
L. broomii 5399	4,395,904	42.99
L. chreensis 201903075	4,489,308	39.77
L. congkakensis 201702421	4,003,841	38.23
L. dzoumogneensis 201601113	4,118,320	40.99
L. ellinghauseni E18	4,193,578	37.35
L. ellisii ATI7-C-A5	4,314,057	47.80
L. fainei BUT 6	4,287,324	43.53
L. fletcheri SSW15	3,733,663	47.35
L. fluminis SCS5	3,746,420	47.70
L. gomenensis 201800299	4,284,018	46.13
L. haakeii ATI7-C-A2	4,191,631	39.81
L. harrisiae FH2-B-A1	3,946,761	37.86
L. hartskeerlii MCA1_C_A1	4,048,089	40.47
L. idonii 201300427	4,088,029	41.15
L. ilyithenensis 201400974	4,211,893	40.54
L. inadai 10	4,457,871	44.61
L. interrogans 56601	4,698,134	35.02
L. jelokensis 201702419	4,124,627	38.90
L. johnsonii E8	4,063,992	41.26
L. kanakyensis 201800292	4,138,228	38.50
L. kemamanensis 201702454	3,769,371	38.90
L. kirschneri 200702274	4,317,431	35.89
L. kmetyi Bejo-Iso9	4,418,431	44.82
L. kobayashii E30	4,294,786	40.68

L. koniamboensis 201800265	4,319,728	38.97
L. langatensis SSW18	4,089,386	44.79
L. levettii MCA2_B_A1	3,875,926	37.61
L. licerasiae VAR 010	4,211,147	41.13
L. mayottensis 200901116	4,135,276	39.49
L. meyeri Went 5	4,188,061	38.03
L. montravelensis 201800279	4,083,426	37.40
L. mtsangambouensis 201601298	4,083,871	38.20
L. neocaledonica ES4-C-A1	4,213,808	40.17
L. noguchii CZ214	4,710,917	35.54
L. noumeaensis 201800287	4,108,400	38.29
L. ognonensis 201702476	3,988,113	39.66
L. perdikensis 201702692	4,001,979	38.50
L. perolatii FH1-B-B1	3,985,918	42.36
L. ryugenii YH101	3,951,395	39.92
L. saintgironsiae FH4-C-A2	4,083,319	39.11
L. santarosai LT 821	3,983,611	41.82
L. sarikeiensis 201702455	4,373,171	40.27
L. selangorensis SSW17	4,191,189	39.99
L. semungkikensis SSS9	3,944,076	42.76
L. stimsonii AMB6-RJ	4,745,109	42.65
L. terpstrae ATCC 700639	4,092,188	38.21
L. tipperaryensis GWTS	4,591,888	42.42
L. vanthielii ATCC 700522	4,232,327	38.90
L. venezuelensis CLM-R50	4,308,482	39.19
L. weilii LT2116	4,323,312	40.46
L. wolbachii CDC	4,081,981	39.15
L. wolffii Khorat-H2	4,400,835	45.58
L. yanagawae ATCC 700523	4,056,240	38.23
L. yasudae F10	4,444,566	45.49

Table S3.2. Genes acquired by P2 and conserved in P1$^-$ and P1$^+$.

L. interrogans serovar Lai strain 56601 used as reference.

Locus tag	Product	Locus tag	Product
LA_0004	Uncharacterized protein	LA_2316	Uncharacterized protein
LA_0013	Transcriptional regulator	LA_2353	Uncharacterized protein
LA_0014	DNA-directed DNA polymerase delta subunit	LA_2365	Uncharacterized protein
LA_0031	Urate_ox_N domain-containing protein	LA_2371	Uncharacterized protein
LA_0032	Hypothetical lipoprotein	LA_2379	Mg chelatase subunit chII
LA_0034	Uncharacterized protein	LA_2381	UPF0102 protein LA_2381
LA_0036	33 kDa chaperonin	LA_2386	Ribonuclease

LA_0039	Metal-dependent molecular chaperone	LA_2389	Ribosome maturation factor RimM
LA_0043	TPR-repeat-containing protein	LA_2399	Uncharacterized protein
LA_0048	Uncharacterized protein	LA_2406	OstA-like_N domain-containing protein
LA_0049	Methyl-accepting chemotaxis protein	LA_2413	Cell wall-associated hydrolase/lipoprotein
LA_0055	Uncharacterized protein	LA_2421	Histidine kinase sensor protein
LA_0059	TPR-repeat-containing protein	LA_2454	Uncharacterized protein
LA_0065	drug:Na+ antiporter of the multi antimicrobial extrusion family	LA_2456	Uncharacterized protein
LA_0068	Uncharacterized protein	LA_2461	anti-Sigma factor antagonist
LA_0075	Uncharacterized protein	LA_2462	ATP-binding protein of an ABC transporter complex
LA_0077	Uncharacterized protein	LA_2463	ABC-type transport system involved in gliding motility
LA_0078	Uncharacterized protein	LA_2464	ABC-type transport system involved in gliding motility
LA_0080	Hydrolase	LA_2478	Hypothetical lipoprotein
LA_0095	Uncharacterized protein	LA_2519	Uncharacterized protein
LA_0097	Uncharacterized protein	LA_2528	GGDEF domain receiver component of a two-component response regulator
LA_0098	tRNA (uracil-5-)-methyltransferase	LA_2532	Histidine kinase and response regulator hybrid protein
LA_0103	Putative lipoprotein	LA_2537	LipL45-related protein
LA_0104	Metallo-beta-lactamase	LA_2540	Signal transduction histidine kinase
LA_0109	RNA 2',3'-cyclic phosphodiesterase	LA_2544	Alpha/beta hydrolase superfamily protein
LA_0113	Cystathionine gamma-synthase	LA_2548	Serine/threonine phosphatase containing a PAS domain
LA_0115	Uncharacterized protein	LA_2554	Phosphate sodium symporter
LA_0116	Uncharacterized protein	LA_2574	Putative methyl-accepting chemotaxis transmembrane protein
LA_0128	Uncharacterized protein	LA_2578	FeoA-like protein
LA_0129	Uncharacterized protein	LA_2582	M23 family metalloendopeptidase
LA_0135	Uncharacterized protein	LA_2584	Sacchrp_dh_NADP domain-containing protein
LA_0137	Hypothetical lipoprotein	LA_2593	Endoflagellar biosynthesis chaperone

LA_0138	TPR-repeat-containing protein	LA_2595	Uncharacterized protein
LA_0142	PilZ domain-containing protein	LA_2601	Uncharacterized protein
LA_0143	Uncharacterized protein	LA_2612	Flagellar protein
LA_0145	tRNA(Ile)-lysidine synthase	LA_2617	Uncharacterized protein
LA_0149	SET domain-containing protein	LA_2623	Glutathione transferase
LA_0150	ATPase and permease components of ABC-type multidrug transport system	LA_2634	ParB-like protein
LA_0151	ACT domain-containing protein	LA_2637	LipL32
LA_0171	Valine--pyruvate aminotransferase	LA_2643	Adenylate/guanylate cyclase
LA_0184	Uncharacterized protein	LA_2659	Biotin carboxylase
LA_0231	Glycosyltransferase	LA_2662	Ribonuclease BN
LA_0233	Glycosyltransferase	LA_2666	Endoflagellar basal body P-ring biosynthesis protein
LA_0234	Uncharacterized protein	LA_2672	ATP-cone domain-containing protein
LA_0248	Uncharacterized protein	LA_2691	Predicted N6-adenine-specific DNA methylase
LA_0253	Putative lipoprotein	LA_2699	D-alanine--D-alanine ligase
LA_0254	Uncharacterized protein	LA_2703	Germane domain-containing protein
LA_0271	Uncharacterized protein	LA_2704	Two-component response regulator
LA_0272	Uncharacterized protein	LA_2705	Hypothetical lipoprotein
LA_0276	Uncharacterized protein	LA_2718	Cytoplasmic membrane protein
LA_0279	Uncharacterized protein	LA_2726	ABM domain-containing protein
LA_0299	Carbon starvation protein A	LA_2753	Uncharacterized protein
LA_0300	ADP-ribose pyrophosphatase	LA_2780	ATP synthase subunit delta
LA_0301	OmpA family protein	LA_2787	Uncharacterized protein
LA_0302	Uncharacterized protein	LA_2790	Transcriptional regulator
LA_0312	Cell wall hydrolase	LA_2793	Magnesium transport protein CorA
LA_0316	Transcriptional regulator with HTH domain	LA_2796	Uncharacterized protein
LA_0320	Thioredoxin	LA_2797	FecR domain-containing protein
LA_0322	Fibronectin binding protein	LA_2800	Uncharacterized protein
LA_0343	GFA domain-containing protein	LA_2813	Methyl-accepting chemotaxis protein

LA_0348	Antagonist of anti-sigma factor	LA_2827	Signal transduction protein containing REC and EAL domains
LA_0351	TPR-repeat-containing protein	LA_2829	Signal receiver component of two-component system
LA_0353	Uncharacterized protein	LA_2832	Predicted esterase of the alpha-beta hydrolase
LA_0379	Uncharacterized protein	LA_2845	Transcriptional regulator
LA_0389	TPR_REGION domain-containing protein	LA_2866	Uncharacterized protein
LA_0396	ArnT-like protein	LA_2872	Uncharacterized protein
LA_0399	Paraoxonase	LA_2887	Ferric uptake regulator
LA_0402	ATP-binding protein of an ABC transporter complex	LA_2901	M23 family metalloendopeptidase
LA_0405	Dimethyladenosine transferase	LA_2936	LipL45-related lipoprotein
LA_0427	Polyphosphate kinase 2	LA_2948	Uncharacterized protein
LA_0433	AraC_E_bind domain-containing protein	LA_2950	HtrA2
LA_0441	Uncharacterized protein	LA_2954	Membrane associated acid phosphatase
LA_0443	RNA polymerase sigma subunit	LA_2967	YCII domain-containing protein
LA_0444	LipL45-related protein	LA_2970	Uncharacterized protein
LA_0447	Uncharacterized protein	LA_2983	ABC transporter integral membrane protein
LA_0462	Uncharacterized protein	LA_2987	Uncharacterized protein
LA_0466	FOG:HEAT repeat protein	LA_2992	Uncharacterized protein
LA_0471	Uncharacterized protein	LA_3019	TPM_phosphatase domain-containing protein
LA_0478	Uncharacterized protein	LA_3067	TolB-related protein
LA_0482	Uncharacterized protein	LA_3068	Putative lipoprotein
LA_0488	DNA helicase UvrD	LA_3069	Uncharacterized protein
LA_0495	Uncharacterized protein	LA_3075	LigB-like protein
LA_0496	Uncharacterized protein	LA_3076	Uncharacterized protein
LA_0506	Uncharacterized protein	LA_3078	Sterol desaturase
LA_0511	Transcriptional regulator	LA_3079	Uncharacterized protein
LA_0517	Uncharacterized protein	LA_3091	Uncharacterized protein
LA_0518	Predicted acetyltransferase	LA_3093	Uncharacterized protein
LA_0537	Predicted hydrolase	LA_3097	LipL71
LA_0549	Uncharacterized protein	LA_3100	Short-chain dehydrogenase
LA_0551	Phosphoribosyl-AMP cyclohydrolase	LA_3110	Potassium-transporting ATPase KdpC subunit
LA_0573	Uncharacterized protein	LA_3119	FGE-sulfatase domain-containing protein
LA_0574	Cytoplasmic membrane protein	LA_3120	Zinc-bindin carboxypeptidase

LA_0595	Glutathione S-transferase-like protein	LA_3121	Permease
LA_0605	SET family protein	LA_3139	Cytochrome c peroxidase
LA_0606	UPF0316 protein LA_0606	LA_3149	TonB-dependent outer membrane hemin receptor
LA_0610	Cell division protein FtsQ	LA_3150	Uncharacterized protein
LA_0615	Uncharacterized protein	LA_3200b	Putative lipoprotein
LA_0616	LipL41	LA_3342	Uncharacterized protein
LA_0630	Oxidoreductase-related protein	LA_3344	Uncharacterized protein
LA_0635	S-layer-like protein	LA_3352	Uncharacterized protein
LA_0650	Intramembrane serine protease	LA_3356	Glutathione transferase
LA_0654	Uncharacterized protein	LA_3357	Histidine kinase sensor protein
LA_0659	Uncharacterized protein	LA_3360	Phosphoglycerate mutase-related protein
LA_0694	Sodium/glucose cotransport protein	LA_3361	Uncharacterized protein
LA_0736	Elongation factor G	LA_3371	Glucans biosynthesis protein
LA_0773	DUF5009 domain-containing protein	LA_3392	Uncharacterized protein
LA_0779	Fatty acid desaturase	LA_3397	Uncharacterized protein
LA_0785	Predicted Zn-dependent hydrolase	LA_3398	O-methyltransferase
LA_0786	Arabinose kinase	LA_3400	Zn-dependent peptidase
LA_0789	Predicted SAM-dependent methyltransferase	LA_3403	Hypothetical lipoprotein
LA_0801	Uncharacterized protein	LA_3410	YceI domain-containing protein
LA_0803	Uncharacterized protein	LA_3432	Uncharacterized protein
LA_0815	Histidine kinase sensor protein	LA_3433	Uncharacterized protein
LA_0818	Uncharacterized protein	LA_3444	Penicillin binding protein
LA_0819	Uncharacterized protein	LA_3450	Regulatory protein
LA_0820	Uncharacterized protein	LA_3451	Serine phosphatase RsbU
LA_0821	N6-adenine-specific methylase	LA_3469	Putative lipoprotein
LA_0823	Alpha/beta hydrolase-related protein	LA_3472	Predicted nucleic-acid-binding protein
LA_0827	Predicted thioesterase	LA_3522	Uncharacterized protein
LA_0832	biotin--[acetyl-CoA-carboxylase] ligase	LA_3523	ABC-type methionine transporter
LA_0842	Hypothetical lipoprotein	LA_3544	Uncharacterized protein
LA_0846	Uncharacterized protein	LA_3545	Uncharacterized protein
LA_0850	Predicted ribosomal protein	LA_3551	tRNA 5-methylaminomethyl-2-thiouridine biosynthesis bifunctional protein MnmC

LA_0861	Anti-sigma factor antagonist	LA_3559	NifU-related protein
LA_0888	NADH-quinone oxidoreductase subunit J	LA_3562	ABC-type transport system involved in Fe-S cluster assembly, permease component
LA_0899	MaoC-like acyl dehydratase	LA_3564	Protein required for attachment to host cells
LA_0900	Transcriptional regulator of MarR family	LA_3565	Glycerol kinase
LA_0908	Uncharacterized protein	LA_3571a	Uncharacterized protein
LA_0912	TPR-repeat-containing protein	LA_3582	Gamma-glutamyl carboxylase-like protein
LA_0927	TPR_REGION domain-containing protein	LA_3586	Uncharacterized protein
LA_0932	Lysophospholipase	LA_3603	Sulfatase
LA_0954	Uncharacterized protein	LA_3610	Serine phosphatase RsbU, regulator of sigma subunit
LA_0971	Short chain dehydrogenase	LA_3613	2-dehydropantoate 2-reductase
LA_0990	TMEM189_B_dmain domain-containing protein	LA_3615	OmpA family protein
LA_1008	Cofac_haem_bdg domain-containing protein	LA_3623	Glutamine amidotransferase
LA_1013	Uncharacterized protein	LA_3628	Anthranilate synthase component 1
LA_1031	Uncharacterized protein	LA_3657	Two-component response regulator
LA_1073	Uncharacterized protein	LA_3668	uracil-DNA glycosylase
LA_1080	Aminopeptidase N	LA_3669	Putative lipoprotein
LA_1086	TPR-repeat lipoprotein	LA_3672	Hydrolase or acyltransferase
LA_1103	Uncharacterized protein	LA_3680	Uncharacterized protein
LA_1123	Predicted exonuclease	LA_3693	ABC-type dipeptide transport system periplasmic component
LA_1126	Erythronate-4-phosphate dehydrogenase	LA_3695	Tetraacyldisaccharide 4'-kinase
LA_1161	Fatty acid transport protein	LA_3744	Uncharacterized protein
LA_1172	Uncharacterized protein	LA_3749	Uncharacterized protein
LA_1182	Histidinol-phosphate aminotransferase	LA_3778	LigB-like protein
LA_1188	Uncharacterized protein	LA_3789	Uncharacterized protein
LA_1190	TPR_REGION domain-containing protein	LA_3797	Uncharacterized protein
LA_1196	Uncharacterized protein	LA_3809	Uncharacterized protein
LA_1197	Serine phosphatase RsbU	LA_3811	Polyamine aminopropyltransferase 1
LA_1209	Uncharacterized protein	LA_3825	Transcriptional regulator
LA_1210	Peroxiredoxin-like protein	LA_3833	Uncharacterized protein

LA_1247	5-formyltetrahydrofolate cyclo-ligase	LA_3839	Phospholipid binding protein
LA_1276	Sulfatase	LA_3845	AcrR family transcriptional regulator
LA_1286	Uncharacterized protein	LA_3854	Uncharacterized protein
LA_1306	Uncharacterized protein	LA_3855	Pirin-like protein
LA_1312	Uncharacterized protein	LA_3861	Uncharacterized protein
LA_1326	Hypothetical lipoprotein	LA_3868	Histidine kinase sensor protein
LA_1332	Ankyrin repeat-containing protein	LA_3869	Response regulator containing a signal receiver domain and a DNA-binding domain
LA_1334	Oxidoreductase	LA_3883	uracil-DNA glycosylase
LA_1347	Signal receiver component of two-component system	LA_3903	Metallo-beta-lactamase superfamily hydrolase
LA_1348	Histidine kinase sensor protein	LA_3912	Uncharacterized protein
LA_1351	Capsule biosynthesis protein CapA (poly-gamma-glutamate biosynthesis protein)	LA_3916	Uncharacterized protein
LA_1358	Ankyrin repeat-containing protein	LA_3924	NAD(P)H steroid dehydrogenase
LA_1365	Predicted membrane protein involved in D-alaninealginate export/acetyltransferase of MBOAT family	LA_3927	TolC family protein
LA_1389	Uncharacterized protein	LA_3930	UDP-N-acetylenolpyruvoylglucosamine reductase
LA_1397	Export protein	LA_3938	Uncharacterized protein
LA_1398	Uncharacterized protein	LA_3943	Uncharacterized protein
LA_1413	Endo/exonuclease/phosphatase domain-containing protein	LA_3946	Putative lipoprotein
LA_1420	DUF1987 domain-containing protein	LA_3957	3-deoxy-D-arabino-heptulosonate 7-phosphate (DAHP) synthase
LA_1421	Uncharacterized protein	LA_3970	Uncharacterized protein
LA_1422	Serine/threonine kinase with GAF and PP2C domains	LA_3980	Glycerophosphodiester phosphodiesterase
LA_1423	3-oxoacyl-[acyl-carrier-protein] synthase	LA_3993	Uncharacterized protein
LA_1424	3-oxoacid CoA-transferase	LA_4009	Cholesterol oxidase precursor
LA_1428	Serine/threonine phosphatase	LA_4017	Penicillin-binding protein

LA_1429	Uncharacterized protein	LA_4040	Inorganic pyrophosphatase
LA_1430	Beta-ketoacyl synthase	LA_4052	Short-chain dehydrogenase
LA_1439	Uncharacterized protein	LA_4053	Uncharacterized protein
LA_1448	Putative lipoprotein	LA_4056	Ankyrin repeat-containing protein
LA_1455	Uncharacterized protein	LA_4059	Uncharacterized protein
LA_1456	DNA repair protein RadC	LA_4068	Uncharacterized protein
LA_1470	TPR_REGION domain-containing protein	LA_4083	Hypothetical lipoprotein
LA_1476	Dehalogenase-like hydrolase	LA_4084	Hypothetical lipoprotein
LA_1477	3-deoxy-D-manno-octulosonic acid transferase	LA_4087	Putative lipoprotein
LA_1478	Uncharacterized protein	LA_4088	Uncharacterized protein
LA_1490	Ribosomal large subunit pseudouridine synthase	LA_4096	Uncharacterized protein
LA_1508	Uncharacterized protein	LA_4099	Uncharacterized protein
LA_1512	Uncharacterized protein	LA_4106	Uncharacterized protein
LA_1522	Uncharacterized protein	LA_4108	Uncharacterized protein
LA_1528	Two-component response regulator	LA_4123	Uncharacterized protein
LA_1533	Thymidylate synthase	LA_4152	Uncharacterized protein
LA_1547	DNA methyltransferase	LA_4160	DUF2779 domain-containing protein
LA_1550	Uncharacterized protein	LA_4170	Uncharacterized protein
LA_1554	PLDc_N domain-containing protein	LA_4179	Uncharacterized protein
LA_1564	Hsp20/alpha crystallin molecular chaperone	LA_4180	Putative aminotransferase
LA_1586	Transketolase C-terminal subunit	LA_4200	L-amino acid oxidase
LA_1667	Sodium:sulfate symporter	LA_4205	Adenosylcobinamide-GDP ribazoletransferase
LA_1674	Poly-gamma-glutamate synthesis protein	LA_4207	VWFA domain-containing protein
LA_1681	Phosphate starvation-inducibleprotein	LA_4208	Uncharacterized protein
LA_1685	DNA repair protein RecO	LA_4209	Uncharacterized protein
LA_1690	Uncharacterized protein	LA_4216	Ferredoxin-like sulfite reductase
LA_1710	Histidine kinase sensor protein	LA_4232	Uncharacterized protein
LA_1711	Histidine kinase sensor protein	LA_4246	Alkaline phosphatase
LA_1712	Signal peptidase I	LA_4247	Hydrolase
LA_1715	Uncharacterized protein	LA_4249	Methyl-accepting chemotaxis protein
LA_1721	DNA topoisomerase 1	LA_4256	Mannosyltransferase
LA_1874	Uncharacterized protein	LA_4262	Putative lipoprotein

LA_1899	Aldo/keto oxidoreductase	LA_4273	Formate hydrogenase subunit B
LA_1900	Uncharacterized protein	LA_4286	Uncharacterized protein
LA_1910	Uncharacterized protein	LA_4287	ABC transporter ATP-binding protein
LA_1912	Hypothetical lipoprotein	LA_4288	ATP-binding protein of an ABC transporter complex
LA_1915	TPR-repeat-containing protein	LA_4289	ABC-type multidrug transport system, permease component
LA_1917	Uncharacterized protein	LA_4293	Uncharacterized protein
LA_1919	DNA-binding transcriptional activator	LA_4314	Uncharacterized protein
LA_1924	Uncharacterized protein	LA_4315	Uncharacterized protein
LA_1946	Uncharacterized protein	LA_4318	Uncharacterized protein
LA_1949	Uncharacterized protein	LA_4319	Hypothetical lipoprotein
LA_1954	Uncharacterized protein	LB_008	Haloacid dehalogenase-like protein
LA_1957	Putative lipoprotein	LB_020	Protoporphyrinogen oxidase
LA_1968	Uncharacterized protein	LB_034	Short-chain dehydrogenase
LA_1971	Uncharacterized protein	LB_047	Conserved hypothetical lipoprotein
LA_1974	Aspartate/glutamate leucyltransferase	LB_050	Uncharacterized protein
LA_1988	Ribosomal RNA small subunit methyltransferase E	LB_052	DUF58 domain-containing protein
LA_1996	Uncharacterized protein	LB_056	BatC
LA_2030	Putative lipoprotein	LB_060	Uncharacterized protein
LA_2034	Putative hydrolase	LB_062	Antagonist of anti-sigma factor
LA_2054	3-hydroxyisobutyrate dehydrogenase	LB_066	Uncharacterized protein
LA_2083	LipL45-related lipoprotein	LB_068	Fatty-acid desaturase
LA_2089	O-antigen ligase/O-antigen polymerase	LB_075	Uncharacterized protein
LA_2091	Uncharacterized protein	LB_078	Thioredoxin domain-containing protein
LA_2093	ABC transporter ATP-binding protein	LB_098	Polysaccharide deacetylase
LA_2098	DNA and RNA helicase	LB_099	Uncharacterized protein
LA_2101	RNA polymerase ECF-type sigma factor	LB_120	Uncharacterized protein
LA_2106	Glutamate--cysteine ligase	LB_121	Short-chain dehydrogenase/reductase SDR
LA_2116	Zn-dependent hydrolase	LB_122	XerD-related integrase
LA_2117	Anti-sigma factor antagonist	LB_141	LipL45-related protein

LA_2120	Uncharacterized protein	LB_167	Uncharacterized protein
LA_2122	Sigma factor regulatory protein	LB_178	Putative serine protease
LA_2131	Uncharacterized protein	LB_182	Uncharacterized protein
LA_2141	ATP-dependent dethiobiotin synthetase BioD	LB_201	Endopeptidase La
LA_2150	Uncharacterized protein	LB_203	Uncharacterized protein
LA_2176	Uncharacterized protein	LB_211	Uncharacterized protein
LA_2177	Long-chain-fatty-acid CoA ligase	LB_217	Lipoprotein
LA_2184	Predicted Zn-dependent protease	LB_218	Uncharacterized protein
LA_2185	Predicted Zn-dependent protease	LB_242	LipL45-related lipoprotein
LA_2203	Uncharacterized protein	LB_250	Hypothetical lipoprotein
LA_2211	Periplasmic serine protease	LB_251	Uncharacterized protein
LA_2213	Uncharacterized protein	LB_255	acetyl-CoA hydrolase/transferase
LA_2222	Sensor histidine kinase and response regulator of a two-component complex	LB_257	Uncharacterized protein
LA_2223	Sensor histidine kinase of a two-component response regulator	LB_258	Cysteine protease
LA_2227	Uncharacterized protein	LB_263	Uncharacterized protein
LA_2248	Uncharacterized protein	LB_271	Permease
LA_2257	Uncharacterized protein	LB_272	Integral membrane protein TerC
LA_2258	Uncharacterized protein	LB_273	methylmalonyl-CoA mutase
LA_2262	Alkylglycerone-phosphate synthase	LB_285	Metallophos domain-containing protein
LA_2266	Putative lipoprotein	LB_299	Hypothetical lipoprotein
LA_2267	TPR-repeat-containing protein	LB_307	Uncharacterized protein
LA_2276	HNH family endonuclease	LB_312	S-adenosylmethionine decarboxylase
LA_2278	J domain-containing protein	LB_312a	S-adenosylmethionine decarboxylase alpha chain
LA_2279	Uncharacterized protein	LB_319	Uncharacterized protein
LA_2284	Uncharacterized protein	LB_321	Uncharacterized protein
LA_2290	Acetyltransferase of MBOAT family	LB_322	Histidine kinase sensor protein
LA_2300	Uncharacterized protein	LB_325	Transcriptional regulator
LA_2301	Uncharacterized protein	LB_330	Uncharacterized protein
LA_2302	Uncharacterized protein	LB_332	Uncharacterized protein
LA_2303	Short chain dehydrogenase	LB_340	Uncharacterized protein
LA_2312	Thioredoxin	LB_341	Uncharacterized protein

LA_2315	Inositol monophosphatase family protein	LB_342	Uncharacterized protein

Table S3.3. Genes acquired by P1^{-} and conserved in P1^{+}.

L. interrogans serovar Lai strain 56601 used as reference.

Locus tag	Product	Locus tag	Product
LA_0012	Uncharacterized protein	LA_2527	Glycerate kinase
LA_0016	Uncharacterized protein	LA_2531	Uncharacterized protein
LA_0018	Uncharacterized protein	LA_2533	Histidine kinase sensor protein
LA_0019	Hypothetical lipoprotein	LA_2538	Uncharacterized protein
LA_0022	Putative lipoprotein	LA_2600	Uncharacterized protein
LA_0038	ATPase	LA_2656	Uncharacterized protein
LA_0045	3-dehydroquinate dehydratase	LA_2671	Uncharacterized protein
LA_0091	Uncharacterized protein	LA_2715	Sensor histidine kinase of a two component response regulator
LA_0107	Uncharacterized protein	LA_2717	Histidine kinase sensor protein
LA_0110	Para-aminobenzoate synthase component I	LA_2719	Uncharacterized protein
LA_0117	Uncharacterized protein	LA_2722	Hypothetical lipoprotein
LA_0193	Uncharacterized protein	LA_2764	Putative lipoprotein
LA_0195	Transcriptional regulator	LA_2767	Uncharacterized protein
LA_0224	Predicted amidohydrolase	LA_2770	Uncharacterized protein
LA_0227	Uncharacterized protein	LA_2774	SLBB domain-containing protein
LA_0284	Uncharacterized protein	LA_2783a	Uncharacterized protein
LA_0286	Uncharacterized protein	LA_2801	Uncharacterized protein
LA_0289	Adenylate/guanylate cyclase	LA_2803	TPR-repeat-containing protein
LA_0290	Ankyrin repeat-containing protein	LA_2815	Putative lipoprotein
LA_0293	DNA double-strand break repair rad50 ATPase	LA_2817	YceI domain-containing protein
LA_0294	Exonuclease	LA_2824	Uncharacterized protein
LA_0297	CBS-domain-containing membrane protein	LA_2825a	Uncharacterized protein
LA_0350	Putative lipoprotein	LA_2830	Histidine kinase sensor protein
LA_0357	Dienelactone hydrolase family protein	LA_2831	Histidine kinase sensor protein
LA_0368	Uncharacterized protein	LA_2859	Uncharacterized protein
LA_0371	Uncharacterized protein	LA_2867	Cytoplasmic membrane protein
LA_0374	Uncharacterized protein	LA_2877	Putative lipoprotein
LA_0375	Uncharacterized protein	LA_2899	Methyltransferase

LA_0392	Phosphagen kinase C-terminal domain-containing protein	LA_2920	Uncharacterized protein
LA_0394	Integrase/recombinase	LA_2926	GGDEF family protein
LA_0397	Lipase/esterase	LA_2927	GGDEF family protein
LA_0403	Permease component of an ABC transporter complex	LA_2929	GGDEF family protein
LA_0413	Uncharacterized protein	LA_2930	GGDEF family protein
LA_0415	SAM-dependent O-methyltransferase	LA_2931	GGDEF family protein
LA_0417	SAM-dependent O-methyltransferase	LA_2932	GGDEF family protein
LA_0429	ANK_REP_REGION domain-containing protein	LA_2933	GGDEF family protein
LA_0442	Uncharacterized protein	LA_2952	Uncharacterized protein
LA_0460	Uncharacterized protein	LA_2969	Glycosyltransferase
LA_0479	Uncharacterized protein	LA_2971	Uncharacterized protein
LA_0523	tRNA pseudouridine synthase D	LA_2972	Putative lipoprotein
LA_0532	Uncharacterized protein	LA_2973	Putative lipoprotein
LA_0565	Adenylate/guanylate cyclase	LA_2975	Hypothetical lipoprotein
LA_0575	Uncharacterized protein	LA_2986	Uncharacterized protein
LA_0599	Signal transduction protein	LA_3015	Uncharacterized protein
LA_0617	Uncharacterized protein	LA_3016	Uncharacterized protein
LA_0631	Uncharacterized protein	LA_3039	SH3b domain-containing protein
LA_0637	Dihydrofolate synthase	LA_3102	Plug domain-containing protein
LA_0644	Uncharacterized protein	LA_3113	Serine/threonine kinase with GAF domain
LA_0703	Molybdate metabolism regulator	LA_3114	Cytoplasmic membrane protein
LA_0728	Uncharacterized protein	LA_3115	Uncharacterized protein
LA_0734	Peroxiredoxin-like protein	LA_3118	Putative lipoprotein
LA_0735	Uncharacterized protein	LA_3152	CopG-like transcriptional regulator
LA_0792	Uncharacterized protein	LA_3165	Uncharacterized protein
LA_0793	Uncharacterized protein	LA_3168	Cytoplasmic membrane protein
LA_0802	TPR-repeat-containing protein	LA_3243	Uncharacterized protein
LA_0808	Serine protease	LA_3258	TonB dependent receptor
LA_0816	Response regulator containing a signal receiver domain and a DNA-binding domain	LA_3307	Undecaprenyl-phosphate alpha-N-acetylglucosaminyltransferase
LA_0817	Uncharacterized protein	LA_3353	Uncharacterized protein

LA_0858	Putative lipoprotein	LA_3358	CopG-like transcriptional regulator
LA_0875	Hsa	LA_3370	Surface antigen OrfC lipoprotein
LA_0876	ECF-like sigma factor SigE	LA_3390	Uncharacterized protein
LA_0878	FecR domain-containing protein	LA_3402	Uncharacterized protein
LA_0879	Uncharacterized protein	LA_3415	Putative lipoprotein
LA_0898	Uncharacterized protein	LA_3438	Uncharacterized protein
LA_0903	Methylase	LA_3440	Putative lipoprotein
LA_0906	Uncharacterized protein	LA_3446	Putative lipoprotein
LA_0913a	Uncharacterized protein	LA_3448	Uncharacterized protein
LA_0914	Uncharacterized protein	LA_3456	Metalloendopeptidase
LA_0939	ADP-dependent (S)-NAD(P)H-hydrate dehydratase	LA_3462	Carbon-nitrogen hydrolase
LA_0965	Exodeoxyribonuclease v gamma chain	LA_3463	Uncharacterized protein
LA_0966	Exodeoxyribonuclease V beta chain	LA_3471	Predicted periplasmic lipoprotein
LA_0991	Uncharacterized protein	LA_3491	Uncharacterized protein
LA_0992	Mechanosensitive ion channel	LA_3497	WGR domain-containing protein
LA_1010	Uncharacterized protein	LA_3506	Methyl-accepting chemotaxis transducer transmembrane protein
LA_1065	Uncharacterized protein	LA_3507	Cytochrome c
LA_1066	Uncharacterized protein	LA_3508	Methylamine utilization protein
LA_1067	Uncharacterized protein	LA_3509	Cytochrome c peroxidase
LA_1069	Hydrolase	LA_3533	Uncharacterized protein
LA_1087	Uncharacterized protein	LA_3535	Uncharacterized protein
LA_1088	Uncharacterized protein	LA_3552	Uncharacterized protein
LA_1099	SH3b domain-containing protein	LA_3587	Acetyltransferase
LA_1104	DnaJ-related protein	LA_3611	Putative lipoprotein
LA_1121	Uncharacterized protein	LA_3637	Uncharacterized protein
LA_1141a	Uncharacterized protein	LA_3640	Uncharacterized protein
LA_1144	Pyrimidine deaminase, riboflavin biosynthesis	LA_3641	Putative acetyltransferase
LA_1213	Uncharacterized protein	LA_3651	Uncharacterized protein
LA_1282	Transcriptional regulator	LA_3751	Uncharacterized protein
LA_1283	Uncharacterized protein	LA_3762	Uncharacterized protein
LA_1297	Uncharacterized protein	LA_3777	Putative fluoride ion transporter CrcB
LA_1310	Uncharacterized protein	LA_3779	Putative lipoprotein
LA_1312a	Uncharacterized protein	LA_3780	Hypothetical lipoprotein
LA_1324	Leucine-rich-repeat protein	LA_3781	Putative lipoprotein

LA_1366	Uncharacterized protein	LA_3798	Uncharacterized protein
LA_1382	Acetyltransferase	LA_3838	Uncharacterized protein
LA_1384	Putative lipoprotein	LA_3848	Ig-like domain-containing protein
LA_1396	Uncharacterized protein	LA_3849	Putative lipoprotein
LA_1410	Lipoate protein ligase A	LA_3856	Uncharacterized protein
LA_1452	Uncharacterized protein	LA_3862	SPOR domain-containing protein
LA_1473	Uncharacterized protein	LA_3871a	Uncharacterized protein
LA_1479	Uncharacterized protein	LA_3873	Uncharacterized protein
LA_1486	Putative lipoprotein	LA_3890	Uncharacterized protein
LA_1499	Cytoplasmic membrane protein	LA_3907	Uncharacterized protein
LA_1504	M23 family metalloendopeptidase	LA_3909	Signal transduction protein containing GGDEF and EAL domains
LA_1517a	Uncharacterized protein	LA_3910	Uncharacterized protein
LA_1524	Uncharacterized protein	LA_3966	Uncharacterized protein
LA_1549	Beta-lactamase regulatory protein 1	LA_3974	Exonuclease
LA_1552	Histidine kinase sensor protein	LA_3994	Fe-S oxidoreductase-like protein
LA_1660	dTDP-4-dehydrorhamnose reductase	LA_3997	Small ribosomal subunit biogenesis GTPase RsgA
LA_1664	dTDP-rhamnosyl transferase	LA_3998	Cholesterol oxidase
LA_1666	Glycosyl transferase	LA_3999	Choline dehydrogenase
LA_1686	Uncharacterized protein	LA_4000	JerF
LA_1709	Signal receiver component of two-component system	LA_4008	Adenylate/guanylate cyclase
LA_1725	Uncharacterized protein	LA_4010	Uncharacterized protein
LA_1733	Uncharacterized protein	LA_4011	Uncharacterized protein
LA_1734	Antisigma factor antagonist-related protein	LA_4013	Co/Zn/Cd efflux pump
LA_1770	Transcriptional regulator	LA_4026	Uncharacterized protein
LA_1774	Uncharacterized protein	LA_4030	Putative lipoprotein
LA_1873	Uncharacterized protein	LA_4055	Malate permease
LA_1885	Uncharacterized protein	LA_4064	Uncharacterized protein
LA_1889	1-aminocyclopropane-1-carboxylate deaminase	LA_4065	Two-component response regulator
LA_1908	PaaI thioesterase	LA_4069	Uncharacterized protein
LA_1922	Uncharacterized protein	LA_4110a	Uncharacterized protein
LA_1937	CopG-like transcriptional regulator	LA_4113	Uncharacterized protein
LA_1945	Uncharacterized protein	LA_4121	Uncharacterized protein

LA_1961	Uncharacterized protein	LA_4127	Sensor histidine kinase of a two component response regulator
LA_1962	Uncharacterized protein	LA_4153	DUF4340 domain-containing protein
LA_1973	Uncharacterized protein	LA_4182	Hypothetical lipoprotein
LA_1978	PMT_2 domain-containing protein	LA_4187	Uncharacterized protein
LA_1981	Uncharacterized protein	LA_4191	Uncharacterized protein
LA_1982	O-antigen polymerase-like protein	LA_4206	Phosphoglycerate mutase
LA_1984	Uncharacterized protein	LA_4226	Uncharacterized protein
LA_1997	Glycosyltransferase	LA_4235	Uncharacterized protein
LA_1998	Polysaccharide deacetylase	LA_4236	DNA mismatch repair protein ATPase component
LA_2026	Uncharacterized protein	LA_4261	dTDP-4-dehydrorhamnose 3,5-epimerase
LA_2032	CopG-like transcriptional regulator	LA_4285	Predicted glycosyl hydrolase
LA_2041	Uncharacterized protein	LA_4292	Uncharacterized protein
LA_2052	DUF2157 domain-containing protein	LA_4305	Putative lipoprotein
LA_2053	Uncharacterized protein	LA_4325	Uncharacterized protein
LA_2064	Ribosomal RNA small subunit methyltransferase E	LB_018	Uncharacterized protein
LA_2065	Uncharacterized protein	LB_030	Intercellular adhesion protein C
LA_2088	Uncharacterized protein	LB_053	Uncharacterized protein
LA_2094	FHA domain-containing protein	LB_057	BatD
LA_2100	Uncharacterized protein	LB_072	Uncharacterized protein
LA_2123	Uncharacterized protein	LB_080	Uncharacterized protein
LA_2127	Uncharacterized protein	LB_096	Methyltransf_11 domain-containing protein
LA_2155	Uncharacterized protein	LB_109	Uncharacterized protein
LA_2160	Uncharacterized protein	LB_110	Putative outer membrane protein
LA_2167	Putative lipoprotein	LB_130	AraC-family transcriptional regulator
LA_2168	Uncharacterized protein	LB_133	Putative diguanylate phosphodiesterase
LA_2180	Uncharacterized protein	LB_134	Uncharacterized protein
LA_2183	Uncharacterized protein	LB_137	Uncharacterized protein
LA_2190	2-dehydropantoate 2-reductase	LB_138	Uncharacterized protein
LA_2192	SprT-like domain-containing protein	LB_139	Serine phosphatase RsbU
LA_2195	Uncharacterized protein	LB_143	LipL45-related protein

LA_2196	Uncharacterized protein	LB_169	STAS domain-containing protein
LA_2259	Uncharacterized protein	LB_190	Uncharacterized protein
LA_2268	Uncharacterized protein	LB_192	Uncharacterized protein
LA_2308	Uncharacterized protein	LB_194	Putative lipoprotein
LA_2311	Uncharacterized protein	LB_197	Uncharacterized protein
LA_2330	Uncharacterized protein	LB_199	Uncharacterized protein
LA_2385	Uncharacterized protein	LB_205	Uncharacterized protein
LA_2419	Uncharacterized protein	LB_213	ExbD-related biopolymer transport protein
LA_2424	Antagonist of anti-sigma factor	LB_219	Uncharacterized protein
LA_2437	Uncharacterized protein	LB_238	GGDEF family protein
LA_2460	Methyltransferase	LB_240	GGDEF family protein
LA_2465	DUF4340 domain-containing protein	LB_280	Uncharacterized protein
LA_2467	Uncharacterized protein	LB_287	Yip1 domain-containing protein
LA_2468	Uncharacterized protein	LB_309	Uncharacterized protein
LA_2471	Uncharacterized protein	LB_358	Uncharacterized protein
LA_2472	Uncharacterized protein	LB_364	Response regulator
LA_2486	Uncharacterized protein		
LA_2488	Uncharacterized protein		
LA_2505	Putative esterase/lipase		

Table S3.4. Genes acquired by P1^{+}.

L. interrogans serovar Lai strain 56601 used as reference.

Locus tag	Product	Locus tag	Product
LA_0283	Uncharacterized protein	LA_1569	Putative lipoprotein
LA_0328	Acetyltransferase	LA_1601	FdtA-like protein
LA_0492	LipL36	LA_1653	Glycosyltransferase
LA_0494	Putative lipoprotein	LA_1657	Glycosyl transferase
LA_0578	Uncharacterized protein	LA_1665	Uncharacterized protein
LA_0588	Uncharacterized protein	LA_1743	Methylase of chemotaxis methyl-accepting protein
LA_0589	Uncharacterized protein	LA_1744	Putative protein-glutamate methylesterase
LA_0591	Uncharacterized protein	LA_1771	Uncharacterized protein
LA_0598	Transcriptional regulator	LA_1793	Transposase
LA_0620	Uncharacterized protein	LA_1808	Transposase
LA_0707	Transposase	LA_1830	Transposase
LA_0728a	Hypothetical lipoprotein	LA_1944	Transposase
LA_0769	Uncharacterized protein	LA_2020	Lipoprotein
LA_0795	Uncharacterized protein	LA_2154a	Uncharacterized protein
LA_0812	Uncharacterized protein	LA_2169	Lipoprotein
LA_0835	Uncharacterized protein	LA_2200	Amidase
LA_0836	Uncharacterized protein	LA_2529	Uncharacterized protein
LA_0872	Microbial collagenase	LA_2628	Uncharacterized protein

LA_0873	Cytoplasmic membrane protein	LA_2641	Ferrichrome-iron receptor
LA_0905	Hypothetical lipoprotein	LA_3050	Hemolysin
LA_0934	Uncharacterized protein	LA_3271	Uncharacterized protein
LA_0985	WGR domain-containing protein	LA_3322	Putative lipoprotein
LA_1027	Sphingomyelinase C precursor	LA_3323	Leucine-rich repeat protein
LA_1029	Sphingomyelinase C precursor	LA_3338	Putative lipoprotein
LA_1092	Uncharacterized protein	LA_3387	Uncharacterized protein
LA_1183	Uncharacterized protein	LA_3388	Uncharacterized protein
LA_1353	Cytoplasmic membrane protein	LA_3490	Uncharacterized protein
LA_1400	Ricin B-type lectin domain-containing protein	LA_3540	Sphingomyelinase C precursor
LA_1402	Uncharacterized protein	LA_3834	Putative lipoprotein
LA_1454	Uncharacterized protein	LA_4004	Sphingomyelinase C precursor
LA_1567	Putative lipoprotein	LA_4135	Putative lipoprotein
LA_1568	Putative lipoprotein	LB_102	Haloacid dehalogenase-like protein

Table S3.5. Genes lost by P2 and absent in P1$^-$ and P1$^+$.

L. biflexa serovar Patoc strain Patoc1 (Paris) used as reference.

Locus tag	Product	Locus tag	Product
LEPBI_I0004	Uncharacterized protein	LEPBI_I2070	Uncharacterized protein
LEPBI_I0008	PilZ domain-containing protein	LEPBI_I2072	Uncharacterized protein
LEPBI_I0010	Uncharacterized protein	LEPBI_I2073	Putative phospholipase D
LEPBI_I0011	Putative transcriptional regulator	LEPBI_I2074	Uncharacterized protein
LEPBI_I0012	DNA_pol3_delta domain-containing protein	LEPBI_I2075	Uncharacterized protein
LEPBI_I0013	Uncharacterized protein	LEPBI_I2078	Signal peptidase I
LEPBI_I0015	Uncharacterized protein	LEPBI_I2079	Putative two-component sensor
LEPBI_I0018	PGA_cap domain-containing protein	LEPBI_I2087	Uncharacterized protein
LEPBI_I0019	Uncharacterized protein	LEPBI_I2090	Uncharacterized protein
LEPBI_I0021	Peptidase_M22 domain-containing protein	LEPBI_I2091	DNA repair protein RecO
LEPBI_I0022	Uncharacterized protein	LEPBI_I2101	Uncharacterized protein
LEPBI_I0023	Uncharacterized protein	LEPBI_I2102	PGA_cap domain-containing protein
LEPBI_I0025	Uncharacterized protein	LEPBI_I2103	Uncharacterized protein
LEPBI_I0027	Superoxide dismutase	LEPBI_I2104	Putative TPR-repeat-containing protein putative signal peptide
LEPBI_I0030	Uncharacterized protein	LEPBI_I2105	Uncharacterized protein
LEPBI_I0032	Uncharacterized protein	LEPBI_I2107	CHAT domain-containing protein
LEPBI_I0036	Uncharacterized protein	LEPBI_I2108	Uncharacterized protein
LEPBI_I0044	Putative carboxylesterase 2	LEPBI_I2109	Putative RNA polymerase sigma-E factor

LEPBI_I0051	Peptidase_M23 domain-containing protein	LEPBI_I2110	Uncharacterized protein
LEPBI_I0053	Response regulatory domain-containing protein	LEPBI_I2112	Uncharacterized protein
LEPBI_I0055	Uncharacterized protein	LEPBI_I2115	Uncharacterized protein
LEPBI_I0065	Uncharacterized protein	LEPBI_I2119	Putative two-component sensor histidine kinase
LEPBI_I0066	ACT domain-containing protein	LEPBI_I2122	Uncharacterized protein
LEPBI_I0070	DUF4139 domain-containing protein	LEPBI_I2124	NAD(P)H dehydrogenase (Quinone) putative signal peptide
LEPBI_I0093	Putative sensor protein	LEPBI_I2134	Uncharacterized protein
LEPBI_I0095	Putative beta-phosphoglucomutase	LEPBI_I2141	Uncharacterized protein
LEPBI_I0096	Pseudouridine synthase (Uracil hydrolyase)	LEPBI_I2143	Uncharacterized protein
LEPBI_I0099	EAL domain-containing protein	LEPBI_I2144	Putative glycoside hydrolase, family 57
LEPBI_I0117	Uncharacterized protein	LEPBI_I2145	Uncharacterized protein
LEPBI_I0138	Uncharacterized protein	LEPBI_I2148	Uncharacterized protein
LEPBI_I0145	Uncharacterized protein	LEPBI_I2149	Uncharacterized protein
LEPBI_I0153	Uncharacterized protein	LEPBI_I2152	Putative ferric uptake regulation protein, fur family (Ferric uptake regulator)
LEPBI_I0158	Uncharacterized protein	LEPBI_I2157	Uncharacterized protein
LEPBI_I0159	Uncharacterized protein	LEPBI_I2271	Uncharacterized protein
LEPBI_I0160	Uncharacterized protein	LEPBI_I2272	Pseudouridine synthase
LEPBI_I0161	Uncharacterized protein	LEPBI_I2273	ABC-type transport system, permease ATP binding protein
LEPBI_I0167	STAS domain-containing protein	LEPBI_I2274	Methyl-accepting transducer domain-containing protein
LEPBI_I0169	Uncharacterized protein	LEPBI_I2279	Putative glycosyl transferase, family 39 putative membrane protein
LEPBI_I0170	VWFA domain-containing protein	LEPBI_I2280	Uncharacterized protein
LEPBI_I0182	Putative response regulator, LuxR family	LEPBI_I2282	Putative anti-anti-sigma regulatory factor
LEPBI_I0183	TPR_REGION domain-containing protein	LEPBI_I2285	S-adenosylmethionine decarboxylase proenzyme
LEPBI_I0184	Putative conserved protein	LEPBI_I2286	Uncharacterized protein
LEPBI_I0188	Acyl-CoA desaturase putative membrane protein	LEPBI_I2287	Putative methyl-accepting chemotaxis protein (MCP) putative membrane protein
LEPBI_I0190	Uncharacterized protein	LEPBI_I2288	Uncharacterized protein
LEPBI_I0191	LEA_2 domain-containing protein	LEPBI_I2289	Putative leucine-reach repeat protein
LEPBI_I0201	Uncharacterized protein	LEPBI_I2292	STAS domain-containing protein
LEPBI_I0202	Biotin_lipoyl_2 domain-containing protein	LEPBI_I2293	Uncharacterized protein
LEPBI_I0203	Putative ABC-type transport system putative membrane protein	LEPBI_I2319	Uncharacterized protein
LEPBI_I0204	Putative ABC-type transport system	LEPBI_I2320	Uncharacterized protein

LEPBI_I0205	YceI domain-containing protein	LEPBI_I2323	Putative alpha amylase
LEPBI_I0224	Putative permease putative membrane protein	LEPBI_I2327	Putative ATPase
LEPBI_I0228	Uncharacterized protein	LEPBI_I2332	Putative metalloendopeptidase putative signal peptide
LEPBI_I0229	Uncharacterized protein	LEPBI_I2337	Uncharacterized protein
LEPBI_I0231	SCP domain-containing protein	LEPBI_I2339	Putative two-component regulatory response regulator, LuxR family
LEPBI_I0239	WYL domain-containing protein	LEPBI_I2340	Uncharacterized protein
LEPBI_I0241	Uncharacterized protein	LEPBI_I2342	Putative ATPase
LEPBI_I0242	Putative sodium:solute symporter putative membrane protein	LEPBI_I2353	Putative para-aminobenzoate synthase component I (ADC synthase)
LEPBI_I0243	Uncharacterized protein	LEPBI_I2358	Uncharacterized protein
LEPBI_I0244	Peptidase_M23 domain-containing protein	LEPBI_I2360	Uncharacterized protein
LEPBI_I0247	Uncharacterized protein	LEPBI_I2367	Uncharacterized protein
LEPBI_I0248	Thioredoxin domain-containing protein	LEPBI_I2369	Uncharacterized protein
LEPBI_I0254	Sigma-54 factor interaction domain-containing protein	LEPBI_I2370	Uncharacterized protein
LEPBI_I0255	OmpA-like domain-containing protein	LEPBI_I2373	Uncharacterized protein
LEPBI_I0263	Uncharacterized protein	LEPBI_I2376	Uncharacterized protein
LEPBI_I0264	Uncharacterized protein	LEPBI_I2378	DNA helicase
LEPBI_I0272	Phospho-2-dehydro-3-deoxyheptonate aldolase	LEPBI_I2381	Uncharacterized protein
LEPBI_I0276	Uncharacterized protein	LEPBI_I2394	5-formyltetrahydrofolate cyclo-ligase
LEPBI_I0278	Uncharacterized protein	LEPBI_I2397	Uncharacterized protein
LEPBI_I0279	Uncharacterized protein	LEPBI_I2401	Uncharacterized protein
LEPBI_I0280	Probable membrane transporter protein	LEPBI_I2403	Putative transcriptional regulator
LEPBI_I0282	Uncharacterized protein	LEPBI_I2404	Putative alpha/beta hydrolase putative signal peptide
LEPBI_I0283	Putative polyketide biosynthesis associated protein	LEPBI_I2410	Putative ATP-dependent Clp protease adaptor protein C
LEPBI_I0285	Uncharacterized protein	LEPBI_I2412	Putative adenylate or guanylate cyclase putative membrane protein
LEPBI_I0287	ANK_REP_REGION domain-containing protein	LEPBI_I2415	HDOD domain-containing protein
LEPBI_I0288	Uncharacterized protein	LEPBI_I2423	Hydrogenase-4 subunit C putative membrane protein
LEPBI_I0289	Uncharacterized protein	LEPBI_I2425	ABC-type transport system, permease putative membrane protein
LEPBI_I0294	Putative membrane-bound lytic murein transglycosylase	LEPBI_I2427	Uncharacterized protein
LEPBI_I0299	Uncharacterized protein	LEPBI_I2431	Uncharacterized protein
LEPBI_I0301	DUF1330 domain-containing protein	LEPBI_I2439	Putative voltage-gated chloride channel protein putative membrane protein putative signal peptide

LEPBI_I0303	Putative permease of the major facilitator superfamily putative signal peptide	LEPBI_I2441	PlsC domain-containing protein
LEPBI_I0314	Cell division protein FtsQ	LEPBI_I2442	Putative two-component response regulator
LEPBI_I0317	Uncharacterized protein	LEPBI_I2444	HTH tetR-type domain-containing protein
LEPBI_I0318	Uncharacterized protein	LEPBI_I2447	Uncharacterized protein
LEPBI_I0319	Uncharacterized protein	LEPBI_I2450	Putative serine phosphatase RsbU regulator putative membrane protein putative signal peptide
LEPBI_I0320	Uncharacterized protein	LEPBI_I2452	Putative two-component regulator
LEPBI_I0323	Uncharacterized protein	LEPBI_I2456	Putative hydroxypyruvate reductase
LEPBI_I0324	Putative phosphoserine phosphatase	LEPBI_I2458	Putative two-component sensor protein putative membrane protein
LEPBI_I0330	AsmA_2 domain-containing protein	LEPBI_I2463	2-dehydropantoate 2-reductase
LEPBI_I0334	Putative DNA mismatch repair protein putative membrane protein	LEPBI_I2464	Uncharacterized protein
LEPBI_I0343	AB hydrolase-1 domain-containing protein	LEPBI_I2466	Putative cAMP-binding protein
LEPBI_I0354	Uncharacterized protein	LEPBI_I2467	Putative type I phosphodiesterase/nucleotide pyrophosphatase
LEPBI_I0357	ABC-type transport system, ATPase	LEPBI_I2468	Uncharacterized protein
LEPBI_I0358	Putative associated RTX toxin transporter	LEPBI_I2469	3-dehydroquinate synthase
LEPBI_I0359	ABC-type transport system, ATPase and permease	LEPBI_I2470	Uncharacterized protein
LEPBI_I0360	Putative integral outer membrane protein TolC, efflux pump component putative signal peptide	LEPBI_I2471	Putative metallo-dependent hydrolase
LEPBI_I0365	Putative negative transcriptional regulator	LEPBI_I2472	Uncharacterized protein
LEPBI_I0368	Uncharacterized protein	LEPBI_I2474	Putative transcriptional regulator, TetR family
LEPBI_I0400	Putative sensor protein putative membrane protein	LEPBI_I2475	Putative alcohol dehydrogenase, zinc-containing
LEPBI_I0427	Uncharacterized protein	LEPBI_I2477	Putative hemolysin-III related protein putative membrane protein putative signal peptide
LEPBI_I0457	Uncharacterized protein	LEPBI_I2479	EF-hand domain-containing protein
LEPBI_I0463	Putative methyl-accepting chemotaxis protein putative signal peptide	LEPBI_I2485	Neutral ceramidase
LEPBI_I0471	Putative integrase/recombinase	LEPBI_I2488	Putative transcriptional regulator
LEPBI_I0472	Uncharacterized protein	LEPBI_I2489	Putative enzyme
LEPBI_I0475	Uncharacterized protein	LEPBI_I2490	Putative lipoprotein
LEPBI_I0477	Uncharacterized protein	LEPBI_I2495	Catalase-peroxidase

LEPBI_I0489	Uncharacterized protein	LEPBI_I2496	High-potential iron-sulfur protein
LEPBI_I0491	Uncharacterized protein	LEPBI_I2501	Uncharacterized protein
LEPBI_I0497	Putative penicillin-binding protein, transpeptidase	LEPBI_I2502	Uncharacterized protein
LEPBI_I0503	Putative sensor protein with PAS domain	LEPBI_I2508	Putative adenylate cyclase putative membrane protein
LEPBI_I0507	Uncharacterized protein	LEPBI_I2524	Uncharacterized protein
LEPBI_I0508	Uncharacterized protein	LEPBI_I2536	Uncharacterized protein
LEPBI_I0513	Uncharacterized protein	LEPBI_I2538	Putative hydrolase, alpha/beta superfamily putative signal peptide
LEPBI_I0514	Uncharacterized protein	LEPBI_I2539	Uncharacterized protein
LEPBI_I0519	Putative hydrolase, alpha/beta family	LEPBI_I2542	Uncharacterized protein
LEPBI_I0526	Uncharacterized protein	LEPBI_I2551	Putative flagellar biogenesis protein FliO
LEPBI_I0527	Putative penicillin-binding protein, transpeptidase putative signal peptide	LEPBI_I2555	Uncharacterized protein
LEPBI_I0529	Uncharacterized protein	LEPBI_I2557	Putative arsenate reductase-like protein
LEPBI_I0530	Uncharacterized protein	LEPBI_I2560	OmpA-like domain-containing protein
LEPBI_I0532	Uncharacterized protein	LEPBI_I2561	Uncharacterized protein
LEPBI_I0543	Uncharacterized protein	LEPBI_I2562	Uncharacterized protein
LEPBI_I0544	Uncharacterized protein	LEPBI_I2565	Putative thiosulfate sulfurtransferase putative signal peptide
LEPBI_I0545	Fibronectin type-III domain-containing protein	LEPBI_I2567	O-succinylhomoserine sulfhydrylase
LEPBI_I0560	Putative lipoprotein	LEPBI_I2572	TPR_REGION domain-containing protein
LEPBI_I0561	Uncharacterized protein	LEPBI_I2575	Phosphagen kinase C-terminal domain-containing protein
LEPBI_I0563	Uncharacterized protein	LEPBI_I2578	Uncharacterized protein
LEPBI_I0564	ATP_bind_3 domain-containing protein	LEPBI_I2579	Uncharacterized protein
LEPBI_I0565	Putative phosphoserine phosphatase RsbU (Sigma factor sigB regulation protein RsbU)	LEPBI_I2580	Uncharacterized protein
LEPBI_I0577	Uncharacterized protein	LEPBI_I2587	Uncharacterized protein
LEPBI_I0582	Uncharacterized protein	LEPBI_I2588	Uncharacterized protein
LEPBI_I0589	DUF4468 domain-containing protein	LEPBI_I2594	Uncharacterized protein
LEPBI_I0590	Uncharacterized protein	LEPBI_I2600	Putative thioredoxin-disulfide reductase
LEPBI_I0604	tRNA-dihydrouridine synthase	LEPBI_I2602	Uncharacterized protein
LEPBI_I0608	Uncharacterized protein	LEPBI_I2626	Putative methyl-accepting chemotaxis protein (MCP) putative membrane protein
LEPBI_I0609	Putative NADPH-dependent FMN reductase	LEPBI_I2632	Uncharacterized protein
LEPBI_I0615	Putative oxidoreductase	LEPBI_I2636	Uncharacterized protein
LEPBI_I0620	Uncharacterized protein	LEPBI_I2638	Putative acyltransferase
LEPBI_I0621	Uncharacterized protein	LEPBI_I2639	Putative serine phosphatase

LEPBI_I0623	Putative short-chain dehydrogenase, SDR family	LEPBI_I2640	Cupin_2 domain-containing protein
LEPBI_I0626	ANK_REP_REGION domain-containing protein	LEPBI_I2641	Putative short chain dehydrogenase
LEPBI_I0634	Uncharacterized protein	LEPBI_I2642	Putative signal transduction histidine kinase putative membrane protein
LEPBI_I0638	Uncharacterized protein	LEPBI_I2646	Putative serine/threonine protein kinase
LEPBI_I0639	Uncharacterized protein	LEPBI_I2648	Uncharacterized protein
LEPBI_I0650	Uncharacterized protein	LEPBI_I2650	ABC-type transport system, permease, putative sperimidine/putrescine transport protein PotC
LEPBI_I0651	Uncharacterized protein	LEPBI_I2653	Putative spermidine/putrescine-binding periplasmic protein PotF/PotD putative signal peptide
LEPBI_I0652	Putative alpha-L-glutamate ligase	LEPBI_I2655	4-aminobutyrate aminotransferase
LEPBI_I0655	Uncharacterized protein	LEPBI_I2656	Putative betaine aldehyde dehydrogenase
LEPBI_I0663	Putative triacylglycerol lipase	LEPBI_I2658	Uncharacterized protein
LEPBI_I0664	DUF2779 domain-containing protein	LEPBI_I2660	Uncharacterized protein
LEPBI_I0665	WYL domain-containing protein	LEPBI_I2662	Uncharacterized protein
LEPBI_I0667	Uncharacterized protein	LEPBI_I2663	Uncharacterized protein
LEPBI_I0678	Uncharacterized protein	LEPBI_I2666	Uncharacterized protein
LEPBI_I0681	Putative penicillin binding protein, beta-lactamase class C	LEPBI_I2669	Putative competence-related protein
LEPBI_I0686	Uncharacterized protein	LEPBI_I2671	Uncharacterized protein
LEPBI_I0687	Putative oxidoreductase, short-chain dehydrogenase family	LEPBI_I2673	Uncharacterized protein
LEPBI_I0688	Ribosomal-protein-alanine acetyltransferase	LEPBI_I2677	Uncharacterized protein
LEPBI_I0691	Putative cation exchanger putative membrane protein	LEPBI_I2679	Putative methyltransferase
LEPBI_I0692	Uncharacterized protein	LEPBI_I2681	Cytochrome c oxidase polypeptide IVB (Cytochrome aa3 subunit 4B Caa-3605 subunit 4B)
LEPBI_I0693	Lysine decarboxylase	LEPBI_I2688	Uncharacterized protein
LEPBI_I0697	Uncharacterized protein	LEPBI_I2689	Putative membrane protein
LEPBI_I0698	Uncharacterized protein	LEPBI_I2690	Putative phosphoserine phosphatase putative membrane protein
LEPBI_I0700	GST N-terminal domain-containing protein	LEPBI_I2693	Uncharacterized protein
LEPBI_I0701	Putative peptidase, M23B family putative signal peptide	LEPBI_I2696	Uncharacterized protein
LEPBI_I0702	Putative peptidase, M23B family putative signal peptide	LEPBI_I2704	Uncharacterized protein
LEPBI_I0703	Uncharacterized protein	LEPBI_I2706	PlsC domain-containing protein

LEPBI_I0704	Uncharacterized protein	LEPBI_I2707	Uncharacterized protein
LEPBI_I0705	SH3b domain-containing protein	LEPBI_I2712	Uncharacterized protein
LEPBI_I0706	Uncharacterized protein	LEPBI_I2715	Thiamine-phosphate synthase
LEPBI_I0713	Putative phosphoserine phosphatase putative membrane protein	LEPBI_I2717	Uncharacterized protein
LEPBI_I0715	Putative two-component sensor protein	LEPBI_I2720	Uncharacterized protein
LEPBI_I0716	Uncharacterized protein	LEPBI_I2722	Uncharacterized protein
LEPBI_I0727	Uncharacterized protein	LEPBI_I2724	Uncharacterized protein
LEPBI_I0731	Uncharacterized protein	LEPBI_I2732	Uncharacterized protein
LEPBI_I0732	Putative phospho-N-acetylmuramoyl-pentapeptide-transferase, glycosyltransferase family 4	LEPBI_I2737	Putative transcriptional regulator, AraC family
LEPBI_I0749	Uncharacterized protein	LEPBI_I2739	Putative heavy-metal binding protein putative copper binding protein CopP
LEPBI_I0755	Putative transcriptional regulator, AraC family putative membrane protein	LEPBI_I2742	Uncharacterized protein
LEPBI_I0758	MlaD domain-containing protein	LEPBI_I2743	Uncharacterized protein
LEPBI_I0760	Uncharacterized protein	LEPBI_I2744	Uncharacterized protein
LEPBI_I0761	Putative sterol desaturase putative membrane protein	LEPBI_I2745	Uncharacterized protein
LEPBI_I0768	Uncharacterized protein	LEPBI_I2746	Uncharacterized protein
LEPBI_I0769	Uncharacterized protein	LEPBI_I2750	Uncharacterized protein
LEPBI_I0770	Uncharacterized protein	LEPBI_I2754	Uncharacterized protein
LEPBI_I0780	Uncharacterized protein	LEPBI_I2758	Uncharacterized protein
LEPBI_I0782	Uncharacterized protein	LEPBI_I2759	CoA_binding domain-containing protein
LEPBI_I0783	DUF1554 domain-containing protein	LEPBI_I2778	Putative hydrolase, HAD superfamily, subfamily IA, variant 1
LEPBI_I0784	Aldehyde dehydrogenase	LEPBI_I2782	Putative hydrolase, alpha/beta hydrolase superfamily
LEPBI_I0785	Uncharacterized protein	LEPBI_I2787	Uncharacterized protein
LEPBI_I0788	Uncharacterized protein	LEPBI_I2796	LamGL domain-containing protein
LEPBI_I0791	Uncharacterized protein	LEPBI_I2816	Uncharacterized protein
LEPBI_I0800	Uncharacterized protein	LEPBI_I2835	Uncharacterized protein
LEPBI_I0809	Uncharacterized protein	LEPBI_I2839	Uncharacterized protein
LEPBI_I0810	Uncharacterized protein	LEPBI_I2840	Uncharacterized protein
LEPBI_I0812	Putative acyltransferase, MBOAT family putative membrane protein	LEPBI_I2844	Uncharacterized protein
LEPBI_I0813	Uncharacterized protein	LEPBI_I2848	Uncharacterized protein
LEPBI_I0816	Uncharacterized protein	LEPBI_I2849	Putative ferric uptake regulation protein, Fur family (Ferric uptake regulator)
LEPBI_I0819	PHS	LEPBI_I2851	Uncharacterized protein
LEPBI_I0820	Putative permease, DMT superfamily putative membrane protein	LEPBI_I2853	Uncharacterized protein
LEPBI_I0821	Uracil-DNA glycosylase	LEPBI_I2854	Uncharacterized protein

LEPBI_I0822	Uncharacterized protein	LEPBI_I2858	Uncharacterized protein
LEPBI_I0829	Putative sodium/galactoside symporter	LEPBI_I2863	Uncharacterized protein
LEPBI_I0835	Putative transcriptional regulator, TetR family	LEPBI_I2864	Uncharacterized protein
LEPBI_I0836	FolC bifunctional protein: Folylpolyglutamate synthase/Tetrahydrofolate ynthase	LEPBI_I2867	Uncharacterized protein
LEPBI_I0841	Uncharacterized protein	LEPBI_I2869	Putative nucleoside-diphosphate-sugar epimerase
LEPBI_I0848	ABC-type transport system, periplasmic component	LEPBI_I2870	Uncharacterized protein
LEPBI_I0850	Uncharacterized protein	LEPBI_I2873	Uncharacterized protein
LEPBI_I0851	Putative acetyl-CoA hydrolase/transferase	LEPBI_I2878	Putative two-component sensor protein
LEPBI_I0853	Methyltransf_11 domain-containing protein	LEPBI_I2887	Uncharacterized protein
LEPBI_I0861	Putative two-component sensor protein kinase putative membrane protein	LEPBI_I2888	Uncharacterized protein
LEPBI_I0867	TPR_REGION domain-containing protein	LEPBI_I2890	Putative adenylate cyclase, family 3
LEPBI_I0872	Putative rhomboid-like protein putative membrane protein	LEPBI_I2892	Uncharacterized protein
LEPBI_I0873	Uncharacterized protein	LEPBI_I2897	TPR-repeat-containing protein putative signal peptide
LEPBI_I0880	Ferrous iron transport protein FeoA	LEPBI_I2898	PEGA domain-containing protein
LEPBI_I0884	Apolipoprotein N-acyltransferase	LEPBI_I2903	Uncharacterized protein
LEPBI_I0893	Putative regulator	LEPBI_I2905	Putative ATPase
LEPBI_I0899	Uncharacterized protein	LEPBI_I2906	Metallophos domain-containing protein
LEPBI_I0902	Putative Mn2+ and Fe2+ transporter putative membrane protein	LEPBI_I2908	Uncharacterized protein
LEPBI_I0903	Na(+)/H(+) antiporter (Sodium/proton antiporter	LEPBI_I2909	Putative arylsulfatase
LEPBI_I0904	Putative sigma factor	LEPBI_I2921	Uncharacterized protein
LEPBI_I0906	Uncharacterized protein	LEPBI_I2935	HTH araC/xylS-type domain-containing protein
LEPBI_I0907	Uncharacterized protein	LEPBI_I2939	Adenosylcobinamide-GDP ribazoletransferase
LEPBI_I0909	Putative periplasmic component of the Tol biopolymer transport system	LEPBI_I2940	Putative phosphoglycerate/bisphosphoglycerate mutase
LEPBI_I0915	Putative two-component sensor molecule putative membrane protein	LEPBI_I2942	Putative hemerythrin HHE cation binding protein

LEPBI_I0918	Putative chemotactic two-component histidine kinase with CheY motif	LEPBI_I2944	Putative adenylate/guanylate cyclase putative membrane protein
LEPBI_I0922	Methyl-accepting chemotaxis protein (MCP) putative signal peptide	LEPBI_I2945	Putative adenylate cyclase, family 3 or guanylate cyclase
LEPBI_I0924	Uncharacterized protein	LEPBI_I2946	Flagellar motor switch protein FliG
LEPBI_I0932	Uncharacterized protein	LEPBI_I2948	Putative two-component sensor protein putative membrane protein putative signal peptide
LEPBI_I0933	Uncharacterized protein	LEPBI_I2949	7TMR-DISM_7TM domain-containing protein
LEPBI_I0938	Uncharacterized protein	LEPBI_I2964	Putative TPR-repeats-containing protein
LEPBI_I0940	Tetraacyldisaccharide 4'-kinase	LEPBI_I2969	DUF374 domain-containing protein
LEPBI_I0941	tRNA 5-methylaminomethyl-2-thiouridine biosynthesis bifunctional protein MnmC	LEPBI_I2971	PilZ domain-containing protein
LEPBI_I0942	Methyltransf_30 domain-containing protein	LEPBI_I2974	Uncharacterized protein
LEPBI_I0948	Uncharacterized protein	LEPBI_I2979	Putative adenylate cyclase putative membrane protein putative signal peptide
LEPBI_I0949	Uncharacterized protein	LEPBI_I2990	Glycogen synthase
LEPBI_I0953	Putative flagellar protein FliJ	LEPBI_I2995	Sigma70_r2 domain-containing protein
LEPBI_I0954	Putative two-component response regulator putative membrane protein putative signal peptide	LEPBI_I3000	Uncharacterized protein
LEPBI_I0956	Uncharacterized protein	LEPBI_I3002	Putative transcriptional regulator, TetR family
LEPBI_I0960	Uncharacterized protein	LEPBI_I3003	Sulfatase domain-containing protein
LEPBI_I0963	Uncharacterized protein	LEPBI_I3004	Putative penicillin-binding protein (Peptidoglycan glycosyltransferase)
LEPBI_I0971	Putative beta-lactamase II (Penicillinase Cephalosporinase)	LEPBI_I3005	Uncharacterized protein
LEPBI_I0972	Uncharacterized protein	LEPBI_I3006	Uncharacterized protein
LEPBI_I0975	Uncharacterized protein	LEPBI_I3008	Alkyl hydroperoxide reductase subunit F
LEPBI_I0977	Uncharacterized protein	LEPBI_I3010	Putative bacterial regulatory protein, LysR family
LEPBI_I0978	Thioredoxin domain-containing protein	LEPBI_I3011	Putative phosphoserine phosphatase putative membrane protein putative signal peptide
LEPBI_I0981	PMT_2 domain-containing protein	LEPBI_I3018	Uncharacterized protein
LEPBI_I0982	Uncharacterized protein	LEPBI_I3025	PLDc_N domain-containing protein
LEPBI_I0983	Uncharacterized protein	LEPBI_I3027	Putative transcriptional regulator, TetR family
LEPBI_I0985	DUF4384 domain-containing protein	LEPBI_I3029	Putative permease

LEPBI_I0986	FecR domain-containing protein	LEPBI_I3030	Uncharacterized protein
LEPBI_I0998	Uncharacterized protein	LEPBI_I3031	TPM_phosphatase domain-containing protein
LEPBI_I1005	Uncharacterized protein	LEPBI_I3037	Putative acyltransferase
LEPBI_I1011	Uncharacterized protein	LEPBI_I3039	Uncharacterized protein
LEPBI_I1012	Uncharacterized protein	LEPBI_I3042	Uncharacterized protein
LEPBI_I1017	Uncharacterized protein	LEPBI_I3044	PPM-type phosphatase domain-containing protein
LEPBI_I1022	Uncharacterized protein	LEPBI_I3047	Uncharacterized protein
LEPBI_I1030	Uncharacterized protein	LEPBI_I3051	Putative ankyrin-like protein
LEPBI_I1037	Uncharacterized protein	LEPBI_I3054	Uncharacterized protein
LEPBI_I1042	Putative phosphoserine phosphatase RsbP putative membrane protein putative signal peptide	LEPBI_I3056	Uncharacterized protein
LEPBI_I1045	Uncharacterized protein	LEPBI_I3072	Uncharacterized protein
LEPBI_I1049	Uncharacterized protein	LEPBI_I3077	Uncharacterized protein
LEPBI_I1051	Uncharacterized protein	LEPBI_I3079	Uncharacterized protein
LEPBI_I1068	Putative short chain dehydrogenase	LEPBI_I3081	Uncharacterized protein
LEPBI_I1069	Putative two-component response regulator putative membrane protein	LEPBI_I3094	Uncharacterized protein
LEPBI_I1071	Putative NAD(FAD)-dependent dehydrogenase	LEPBI_I3098	Uncharacterized protein
LEPBI_I1072	OmpA-like domain-containing protein	LEPBI_I3102	Putative carboxymethylenebutenolidase
LEPBI_I1073	Probable malate:quinone oxidoreductase	LEPBI_I3104	Putative fatty acid desaturase
LEPBI_I1075	Putative methyl-accepting chemotaxis protein (MCP) putative membrane protein	LEPBI_I3105	Putative ferredoxin--NAD(+) reductase
LEPBI_I1080	Putative adenylate cyclase putative membrane protein	LEPBI_I3106	Putative transcriptional regulator, TetR family
LEPBI_I1089	Uncharacterized protein	LEPBI_I3119	Uncharacterized protein
LEPBI_I1099	Uncharacterized protein	LEPBI_I3122	Uncharacterized protein
LEPBI_I1103	Putative metallo-dependent phosphatase	LEPBI_I3124	Putative phosphate transporter putative membrane protein putative signal peptide
LEPBI_I1104	Putative transcriptional regulator, TetR family	LEPBI_I3125	Putative sodium-dependent phosphate transport protein
LEPBI_I1110	Uncharacterized protein	LEPBI_I3137	Fumarate hydratase class I
LEPBI_I1115	Uncharacterized protein	LEPBI_I3151	Uncharacterized protein
LEPBI_I1117	Putative acyltransferase putative membrane protein putative signal peptide	LEPBI_I3154	Putative regulatory protein putative membrane protein putative signal peptide
LEPBI_I1118	Putative amine oxidase (Flavin-containing)	LEPBI_I3155	Putative regulatory protein
LEPBI_I1123	Putative haloacid dehalogenase-like hydrolase	LEPBI_I3162	Uncharacterized protein
LEPBI_I1126	Putative dimethyladenosine transferase	LEPBI_I3163	Putative sulfatase family protein

LEPBI_I1127	Putative competence protein putative membrane protein putative signal peptide	LEPBI_I3170	D-alanyl-D-alanine dipeptidase
LEPBI_I1135	Uncharacterized protein	LEPBI_I3171	Putative vancomycin B-type resistance protein VanW
LEPBI_I1136	Putative adenylate cyclase putative membrane protein	LEPBI_I3172	Uncharacterized protein
LEPBI_I1145	Putative alpha/beta hydrolase	LEPBI_I3173	Uncharacterized protein
LEPBI_I1147	Putative transcriptional regulator, MarR family	LEPBI_I3185	Uncharacterized protein
LEPBI_I1152	LysM domain-containing protein	LEPBI_I3186	Uncharacterized protein
LEPBI_I1155	Uncharacterized protein	LEPBI_I3192	Putative cytochrome c with heme-binding site putative signal peptide
LEPBI_I1157	Putative thiol-disulfide isomerase and thioredoxin	LEPBI_I3193	Putative cysteine protease putative signal peptide
LEPBI_I1159	Uncharacterized protein	LEPBI_I3194	Uncharacterized protein
LEPBI_I1161	Uncharacterized protein	LEPBI_I3195	Uncharacterized protein
LEPBI_I1163	DNA topoisomerase I	LEPBI_I3198	Putative PA-phosphatase related phosphoesterase
LEPBI_I1165	Protoporphyrinogen oxidase (PPO)	LEPBI_I3200	Uncharacterized protein
LEPBI_I1171	Glutamyl-tRNA reductase (GluTR)	LEPBI_I3202	Putative serine protease putative signal peptide
LEPBI_I1174	Uncharacterized protein	LEPBI_I3204	Uncharacterized protein
LEPBI_I1176	NERD domain-containing protein	LEPBI_I3207	Metallophos domain-containing protein
LEPBI_I1177	Uncharacterized protein	LEPBI_I3218	Uncharacterized protein
LEPBI_I1180	Sulfite reductase [NADPH] flavoprotein alpha-component (SIR-FP)	LEPBI_I3221	Uncharacterized protein
LEPBI_I1181	Sulfite reductase [NADPH] hemoprotein beta-component	LEPBI_I3224	Uncharacterized protein
LEPBI_I1189	Putative flagellar protein	LEPBI_I3229	Uncharacterized protein
LEPBI_I1192	Uncharacterized protein	LEPBI_I3237	Uncharacterized protein
LEPBI_I1193	Uncharacterized protein	LEPBI_I3240	Putative TPR-repeat-containing protein putative signal peptide
LEPBI_I1196	UDP-N-acetylmuramoylalanine--D-glutamate ligase	LEPBI_I3249	Uncharacterized protein
LEPBI_I1199	Putative two-component sensor putative membrane protein	LEPBI_I3252	Uncharacterized protein
LEPBI_I1249	Uncharacterized protein	LEPBI_I3253	Uncharacterized protein
LEPBI_I1250	Uncharacterized protein	LEPBI_I3256	Putative aminotransferase
LEPBI_I1251	Uncharacterized protein	LEPBI_I3258	Uncharacterized protein
LEPBI_I1254	Uncharacterized protein	LEPBI_I3260	Putative metalloendopeptidase
LEPBI_I1258	Putative methyl-accepting chemotaxis protein putative membrane protein	LEPBI_I3262	Uncharacterized protein
LEPBI_I1262	Uncharacterized protein	LEPBI_I3263	Putative glycosyltransferase
LEPBI_I1263	Uncharacterized protein	LEPBI_I3265	Putative glycosyltransferase

LEPBI_I1264	ABC-type transport system, periplasmic adhesin putative signal peptide	LEPBI_I3267	Uncharacterized protein
LEPBI_I1265	ABC-type transport system, ATPase	LEPBI_I3268	Putative TelA-like protein
LEPBI_I1266	ABC-type transport system, permease putative membrane protein	LEPBI_I3269	Putative integrase
LEPBI_I1274	Ribosomal RNA small subunit methyltransferase E	LEPBI_I3271	PMT_2 domain-containing protein
LEPBI_I1276	Uncharacterized protein	LEPBI_I3273	BirA bifunctional protein biotin operon repressor/Biotin--[acetyl-CoA-carboxylase] synthetase
LEPBI_I1277	Uncharacterized protein	LEPBI_I3274	Putative short-chain dehydrogenase/reductase, SDR family putative signal peptide
LEPBI_I1278	Uncharacterized protein	LEPBI_I3276	Uncharacterized protein
LEPBI_I1279	Uncharacterized protein	LEPBI_I3277	Putative response receiver
LEPBI_I1280	Putative sulfatase putative membrane protein	LEPBI_I3278	Putative chemotaxis protein CheC
LEPBI_I1283	Uncharacterized protein	LEPBI_I3280	Uncharacterized protein
LEPBI_I1284	Uncharacterized protein	LEPBI_I3281	Putative subtilisin-like serine protease
LEPBI_I1311	Uncharacterized protein	LEPBI_I3284	Uncharacterized protein
LEPBI_I1313	Inorganic pyrophosphatase (Pyrophosphate phospho-hydrolase PPase)	LEPBI_I3285	Uncharacterized protein
LEPBI_I1317	Uncharacterized protein	LEPBI_I3286	Uncharacterized protein
LEPBI_I1324	Putative flagellar motor switch protein	LEPBI_I3294	PPM-type phosphatase domain-containing protein
LEPBI_I1330	Uncharacterized protein	LEPBI_I3295	Uncharacterized protein
LEPBI_I1332	3-deoxy-D-manno-octulosonic acid transferase	LEPBI_I3298	Uncharacterized protein
LEPBI_I1334	Uncharacterized protein	LEPBI_I3299	FecR domain-containing protein
LEPBI_I1339	Uncharacterized protein	LEPBI_I3300	Flavin-dependent thymidylate synthase
LEPBI_I1343	Putative ribosomal large subunit pseudouridine synthase	LEPBI_I3302	Uncharacterized protein
LEPBI_I1346	Uncharacterized protein	LEPBI_I3305	Phosphoribosyl-AMP cyclohydrolase
LEPBI_I1347	Uncharacterized protein	LEPBI_I3306	Putative adenylate cyclase putative membrane protein
LEPBI_I1352	Uncharacterized protein	LEPBI_I3310	Methyltransf_12 domain-containing protein
LEPBI_I1356	Uncharacterized protein	LEPBI_I3311	Uncharacterized protein
LEPBI_I1357	ABC-type transport system, ATPase	LEPBI_I3312	Beta-lactamase domain-containing protein
LEPBI_I1365	Uncharacterized protein	LEPBI_I3313	Uncharacterized protein
LEPBI_I1379	Putative helicase	LEPBI_I3314	Uncharacterized protein
LEPBI_I1381	Uncharacterized protein	LEPBI_I3315	Uncharacterized protein
LEPBI_I1390	Uncharacterized protein	LEPBI_I3329	Uncharacterized protein
LEPBI_I1400	Uncharacterized protein	LEPBI_I3330	Uncharacterized protein
LEPBI_I1403	Riboflavin biosynthesis protein RibD	LEPBI_I3331	Uncharacterized protein
LEPBI_I1417	Uncharacterized protein	LEPBI_I3336	Uncharacterized protein

LEPBI_I1420	Putative phosphoserine phosphatase	LEPBI_I3337	Uncharacterized protein
LEPBI_I1425	4-alpha-glucanotransferase	LEPBI_I3348	Uncharacterized protein
LEPBI_I1436	Uncharacterized protein	LEPBI_I3349	Uncharacterized protein
LEPBI_I1447	Uncharacterized protein	LEPBI_I3351	Conserved pirin-related protein
LEPBI_I1449	SPOR domain-containing protein	LEPBI_I3353	Uncharacterized protein
LEPBI_I1453	Putative bifunctional protein: transcriptional regulator/amino transferase	LEPBI_I3359	ANK_REP_REGION domain-containing protein
LEPBI_I1454	Putative aspartate aminotransferase (Transaminase A ASPAT)	LEPBI_I3367	UDP-N-acetylenolpyruvoylglucosamine reductase
LEPBI_I1455	Uncharacterized protein	LEPBI_I3375	Uncharacterized protein
LEPBI_I1457	Uncharacterized protein	LEPBI_I3382	Uncharacterized protein
LEPBI_I1468	Uncharacterized protein	LEPBI_I3383	Uncharacterized protein
LEPBI_I1472	Uncharacterized protein	LEPBI_I3384	Uncharacterized protein
LEPBI_I1473	Uncharacterized protein	LEPBI_I3395	OB_aCoA_assoc domain-containing protein
LEPBI_I1474	Putative transport system putative membrane protein	LEPBI_I3404	Uncharacterized protein
LEPBI_I1475	ABC-type transport system, permease putative membrane protein	LEPBI_I3409	TonB_C domain-containing protein
LEPBI_I1476	ABC-type transport system, periplasmic binding protein	LEPBI_I3412	3-dehydroquinate dehydratase
LEPBI_I1477	Uncharacterized protein	LEPBI_I3413	Uncharacterized protein
LEPBI_I1481	Putative glutathione S-transferase	LEPBI_I3417	Aminopeptidase N, peptidase M1 family
LEPBI_I1482	Uncharacterized protein	LEPBI_I3418	Uncharacterized protein
LEPBI_I1498	Putative ABC-type transport system, permease putative membrane protein	LEPBI_I3420	Uncharacterized protein
LEPBI_I1500	Uncharacterized protein	LEPBI_I3422	HTH cro/C1-type domain-containing protein
LEPBI_I1502	Putative esterase	LEPBI_I3426	Uncharacterized protein
LEPBI_I1503	Putative lipoprotein putative signal peptide	LEPBI_I3429	Uncharacterized protein
LEPBI_I1505	Uncharacterized protein	LEPBI_I3430	Uncharacterized protein
LEPBI_I1506	Uncharacterized protein	LEPBI_I3431	Uncharacterized protein
LEPBI_I1507	Peptidase_M48 domain-containing protein	LEPBI_I3435	Uncharacterized protein
LEPBI_I1514	Uncharacterized protein	LEPBI_I3442	SGNH_hydro domain-containing protein
LEPBI_I1516	PPK2 domain-containing protein	LEPBI_I3443	Uncharacterized protein
LEPBI_I1517	PPK2 domain-containing protein	LEPBI_I3444	Uncharacterized protein
LEPBI_I1518	Putative transcriptional regulator	LEPBI_I3446	Uncharacterized protein
LEPBI_I1519	Uncharacterized protein	LEPBI_I3447	Big_5 domain-containing protein
LEPBI_I1520	Uncharacterized protein	LEPBI_I3449	Uncharacterized protein

LEPBI_I1521	ADP-dependent (S)-NAD(P)H-hydrate dehydratase	LEPBI_I3455	Putative feruloyl esterase
LEPBI_I1533	Flagellar protein FlgA putative signal peptide	LEPBI_I3456	Putative epoxide hydrolase
LEPBI_I1537	Putative ribonuclease BN putative membrane protein	LEPBI_I3457	Uncharacterized protein
LEPBI_I1550	Uncharacterized protein	LEPBI_I3458	Putative RNA methyltransferase, TrmA family
LEPBI_I1551	Putative L-fuculose-phosphate aldolase	LEPBI_I3472	ABC-type transport system, ATP binding protein putative membrane protein putative signal peptide
LEPBI_I1560	Uncharacterized protein	LEPBI_Ia1014	Uncharacterized protein
LEPBI_I1561	Uncharacterized protein	LEPBI_Ia1510	Putative histidine triad (HIT) protein
LEPBI_I1563	Uncharacterized protein	LEPBI_Ia2473	Putative regulator protein, TetR family
LEPBI_I1568	Uncharacterized protein	LEPBI_II0006	Uncharacterized protein
LEPBI_I1587	Putative sensor protein	LEPBI_II0008	Histidine kinase sensor protein
LEPBI_I1590	L-threonine dehydratase	LEPBI_II0010	Hypothetical BatD protein putative von Willebrand factor, type A domain containing protein
LEPBI_I1595	Ribosomal RNA small subunit methyltransferase E	LEPBI_II0011	Hypothetical BatB protein putative von Willebrand factor, type A domain containing protein
LEPBI_I1596	Uncharacterized protein	LEPBI_II0013	Uncharacterized protein
LEPBI_I1600	Putative acetyltransferase	LEPBI_II0014	DUF58 domain-containing protein
LEPBI_I1601	RNHCP domain-containing protein	LEPBI_II0016	Uncharacterized protein
LEPBI_I1611	Putative hydrolase putative signal peptide	LEPBI_II0018	Putative hydrolase
LEPBI_I1616	Uncharacterized protein	LEPBI_II0019	Uncharacterized protein
LEPBI_I1618	Uncharacterized protein	LEPBI_II0020	Uncharacterized protein
LEPBI_I1622	Dipeptide transport system ATP-binding protein	LEPBI_II0023	Uncharacterized protein
LEPBI_I1623	FHA domain-containing protein	LEPBI_II0027	Uncharacterized protein
LEPBI_I1624	Uncharacterized protein	LEPBI_II0029	Acyl_transf_3 domain-containing protein
LEPBI_I1629	Putative cyclic nucleotide-binding protein	LEPBI_II0031	Uncharacterized protein
LEPBI_I1636	PAPS (Adenosine 3'-phosphate 5'-phosphosulfate) 3'(2'),5'-bisphosphate nucleotidase	LEPBI_II0034	Uncharacterized protein
LEPBI_I1638	Uncharacterized protein	LEPBI_II0036	Putative DNA-binding protein (HU-like protein)
LEPBI_I1639	Putative thioredoxin-related protein	LEPBI_II0037	Uncharacterized protein
LEPBI_I1641	Uncharacterized protein	LEPBI_II0038	Putative alkyldihydroxyacetonephosphate synthase
LEPBI_I1648	OstA-like_N domain-containing protein	LEPBI_II0039	HTH tetR-type domain-containing protein
LEPBI_I1655	Uncharacterized protein	LEPBI_II0040	DAGKc domain-containing protein

LEPBI_I1662	Ribosome maturation factor RimM	LEPBI_II0045	Uncharacterized protein
LEPBI_I1665	Ribonuclease	LEPBI_II0049	Uncharacterized protein
LEPBI_I1669	Uncharacterized protein	LEPBI_II0050	Uncharacterized protein
LEPBI_I1671	Putative ATPase with chaperone activity	LEPBI_II0054	Uncharacterized protein
LEPBI_I1679	General secretion pathway protein H	LEPBI_II0055	Uncharacterized protein
LEPBI_I1685	Uncharacterized protein	LEPBI_II0056	DsbD_2 domain-containing protein
LEPBI_I1687	YkuD domain-containing protein	LEPBI_II0057	Uncharacterized protein
LEPBI_I1688	Putative flavin-containing monooxygenase	LEPBI_II0059	Uncharacterized protein
LEPBI_I1693	Uncharacterized protein	LEPBI_II0065	Uncharacterized protein
LEPBI_I1694	Uncharacterized protein	LEPBI_II0066	Glutamate synthase (NADPH)
LEPBI_I1696	Uncharacterized protein	LEPBI_II0077	Uncharacterized protein
LEPBI_I1701	Putative glycosyltransferase	LEPBI_II0078	Uncharacterized protein
LEPBI_I1708	Putative protease TldD	LEPBI_II0079	Uncharacterized protein
LEPBI_I1709	Putative protease TldE (PmbA)	LEPBI_II0081	Uncharacterized protein
LEPBI_I1711	Uncharacterized protein	LEPBI_II0082	Uncharacterized protein
LEPBI_I1715	Uncharacterized protein	LEPBI_II0088	PilZ domain-containing protein
LEPBI_I1718	Uncharacterized protein	LEPBI_II0089	Uncharacterized protein
LEPBI_I1720	Uncharacterized protein	LEPBI_II0095	Methylmalonyl-COA mutase small subunit
LEPBI_I1727	Uncharacterized protein	LEPBI_II0096	Uncharacterized protein
LEPBI_I1728	Sodium:solute symporter family protein putative membrane protein putative signal peptide	LEPBI_II0105	Uncharacterized protein
LEPBI_I1734	Uncharacterized protein	LEPBI_II0107	Uncharacterized protein
LEPBI_I1737	Uncharacterized protein	LEPBI_II0117	Uncharacterized protein
LEPBI_I1738	Putative peptidase, M48 family	LEPBI_II0118	Uncharacterized protein
LEPBI_I1739	Putative two-component sensor protein	LEPBI_II0122	Uncharacterized protein
LEPBI_I1742	Uncharacterized protein	LEPBI_II0126	HATPase_c domain-containing protein
LEPBI_I1743	Putative membrane protein	LEPBI_II0132	Uncharacterized protein
LEPBI_I1744	Uncharacterized protein	LEPBI_II0133	Putative helicase, ATP-dependent
LEPBI_I1745	Uncharacterized protein	LEPBI_II0135	Uncharacterized protein
LEPBI_I1769	Uncharacterized protein	LEPBI_II0142	FecR domain-containing protein
LEPBI_I1771	AsmA domain-containing protein	LEPBI_II0144	Putative soluble pyridine nucleotide transhydrogenase (NAD(P)(+) transhydrogenase [B-specific])
LEPBI_I1774	Putative ribosomal protein	LEPBI_II0147	Uncharacterized protein
LEPBI_I1786	Putative enoyl-CoA hydratase	LEPBI_II0153	Putative biopolymer transport protein ExbD/TolR putative signal peptide
LEPBI_I1792	THUMP domain-containing protein	LEPBI_II0156	DUF4842 domain-containing protein
LEPBI_I1796	Putative D-alanine--d-alanine ligase	LEPBI_II0157	Putative 4-coumarate--CoA ligase

LEPBI_I1799	Germane domain-containing protein	LEPBI_II0162	Uncharacterized protein
LEPBI_I1800	Putative GGDEF family protein putative membrane protein	LEPBI_II0168	AB hydrolase-1 domain-containing protein
LEPBI_I1803	LysM domain-containing protein	LEPBI_II0172	Uncharacterized protein
LEPBI_I1804	Uncharacterized protein	LEPBI_II0180	Uncharacterized protein
LEPBI_I1805	PPM-type phosphatase domain-containing protein	LEPBI_II0182	Uncharacterized protein
LEPBI_I1807	TPR_REGION domain-containing protein	LEPBI_II0187	Putative methyl-accepting chemotaxis protein tlpA, putative transmembrane protein
LEPBI_I1808	TPR_REGION domain-containing protein	LEPBI_II0188	Putative outermembrane protein
LEPBI_I1809	Uncharacterized protein	LEPBI_II0191	Uncharacterized protein
LEPBI_I1812	Uncharacterized protein	LEPBI_II0196	Putative divalent cation transport-related protein
LEPBI_I1816	Putative chaperone protein DnaJ	LEPBI_II0197	Uncharacterized protein
LEPBI_I1817	Uncharacterized protein	LEPBI_II0199	Putative 3-oxoacyl-[acyl-carrier protein] reductase
LEPBI_I1819	PUA domain-containing protein	LEPBI_II0203	Uncharacterized protein
LEPBI_I1826	Uncharacterized protein	LEPBI_II0207	Putative two-component sensor histidine kinase
LEPBI_I1827	FecR domain-containing protein	LEPBI_II0208	ATP-dependent RNA helicase, DEAD-box family (DeaD)
LEPBI_I1833	Uncharacterized protein	LEPBI_II0231	ZIP zinc transporter family protein
LEPBI_I1836	Putative acyl-CoA dehydrogenase	LEPBI_II0232	Uncharacterized protein
LEPBI_I1838	Putative petidase S49, protease IV family	LEPBI_II0233	Uncharacterized protein
LEPBI_I1842	Glycerophosphoryl diester phosphodiesterase (Glycerophosphodiester phosphodiesterase)	LEPBI_II0235	Uncharacterized protein
LEPBI_I1843	Uncharacterized protein	LEPBI_II0236	Uncharacterized protein
LEPBI_I1844	Uncharacterized protein	LEPBI_II0239	Uncharacterized protein
LEPBI_I1846	Histidinol-phosphate aminotransferase	LEPBI_II0240	Uncharacterized protein
LEPBI_I1849	Putative heat shock protein IbpB, Hsp20 family	LEPBI_II0246	Putative integral membrane transport protein PnuC
LEPBI_I1851	Uncharacterized protein	LEPBI_II0248	Uncharacterized protein
LEPBI_I1858	Uncharacterized protein	LEPBI_II0250	FlgO domain-containing protein
LEPBI_I1873	Putative outer membrane protein, OmpA domain putative membrane protein	LEPBI_II0251	Uncharacterized protein
LEPBI_I1875	Putative cyclic-nucleotide-gated cation channel putative membrane protein	LEPBI_II0253	Uncharacterized protein
LEPBI_I1876	Putative ParA family protein	LEPBI_II0255	Uncharacterized protein
LEPBI_I1877	YopX domain-containing protein	LEPBI_II0260	Uncharacterized protein

LEPBI_I1878	Uncharacterized protein	LEPBI_II0261	Uncharacterized protein
LEPBI_I1880	Aminopeptidase N, peptidase M1 family	LEPBI_II0262	Uncharacterized protein
LEPBI_I1886	Uncharacterized protein	LEPBI_II0263	Uncharacterized protein
LEPBI_I1887	Putative amino acid permease	LEPBI_II0264	Uncharacterized protein
LEPBI_I1888	Uncharacterized protein	LEPBI_II0265	Hypothetical methyl-accepting chemotaxis protein putative membrane protein
LEPBI_I1892	Putative hydrolase, alpha/beta superfamily putative signal peptide	LEPBI_II0281	Putative transcriptional regulator
LEPBI_I1895	Uncharacterized protein	LEPBI_p0002	Putative chromosome partitioning protein ParB
LEPBI_I1896	Uncharacterized protein	LEPBI_p0003	Putative bacteriophage protein
LEPBI_I1898	Putative outer membrane protein putative signal peptide	LEPBI_p0004	DUF1653 domain-containing protein
LEPBI_I1900	Hydroxyacylglutathione hydrolase	LEPBI_p0007	Putative transcriptional regulator
LEPBI_I1902	Endonuclease I putative signal peptide	LEPBI_p0010	Putative phosphoserine phosphatase putative membrane protein
LEPBI_I1908	Putative glutathione S-transferase protein	LEPBI_p0013	ABC-type hemin transport system, ATPase
LEPBI_I1916	Putative NifU-like domain protein	LEPBI_p0014	ABC-type hemin transport system, permease putative membrane protein
LEPBI_I1922	Glucose-1-phosphate adenylyltransferase small subunit	LEPBI_p0019	ATP-dependent exoDNAse (Exonuclease V), alpha subunit, helicase superfamily I member
LEPBI_I1931	ABC-type transport system, ATP binding	LEPBI_p0020	DNA 3'-5' helicase
LEPBI_I1933	Uncharacterized protein	LEPBI_p0021	Exodeoxyribonuclease V gamma chain
LEPBI_I1937	Putative histidine kinase sensor protein	LEPBI_p0023	Putative transcription regulator putative membrane protein
LEPBI_I1967	Putative elongation factor G	LEPBI_p0046	histidine kinase
LEPBI_I1993	Uncharacterized protein	LEPBI_p0048	Putative dihydrofolate reductase
LEPBI_I2044	Uncharacterized protein	LEPBI_p0051	Putative rRNA methylase
LEPBI_I2045	Putative tRNA guanosine-2'-O-methyltransferase	LEPBI_p0053	Putative GGDEF/response regulator receiver domain protein
LEPBI_I2051	Uncharacterized protein	LEPBI_p0054	Lipoprotein
LEPBI_I2054	Uncharacterized protein	LEPBI_p0055	Lipoprotein
LEPBI_I2055	Putative bacterial sensor	LEPBI_p0057	Uncharacterized protein
LEPBI_I2059	Uncharacterized protein	LEPBI_p0058	Putative glycosyl transferase, group 2 family protein putative membrane protein putative signal peptide
LEPBI_I2063	Uncharacterized protein	LEPBI_p0060	Lipoprotein
LEPBI_I2064	Uncharacterized protein	LEPBI_p0061	Putative bacteriophage protein
LEPBI_I2065	Uncharacterized protein	LEPBI_pa0017	Heme-binding protein HmuY
LEPBI_I2067	Putative glutathione S-transferase related protein		
LEPBI_I2069	LysM domain-containing protein		

Table S3.6. Genes lost by P1$^-$ and absent in P1$^+$.

L. biflexa serovar Patoc strain Patoc1 (Paris) used as reference.

Locus tag	Product	Locus tag	Product
LEPBI_I0298	Cation efflux protein putative membrane protein	LEPBI_I1741	Anti-sigma factor antagonist
LEPBI_I0369	Putative chromate transporter putative membrane protein	LEPBI_I1751	ABC1 domain-containing protein
LEPBI_I0441	Uncharacterized protein	LEPBI_I1752	Putative glucose-fructose oxidoreductase
LEPBI_I0455	Uncharacterized protein	LEPBI_I1779	Probable nicotinate-nucleotide adenylyltransferase
LEPBI_I0493	Putative polyglutamyl synthetase, CapC family putative membrane protein	LEPBI_I2882	Putative short-chain dehydrogenase/reductase, SDR family
LEPBI_I0558	Putative biotin binding protein	LEPBI_I2923	Uncharacterized protein
LEPBI_I0624	Putative cyclic-nucleotide binding protein	LEPBI_I2977	Putative two-component response regulator
LEPBI_I0636	Cofac_haem_bdg domain-containing protein	LEPBI_I3033	Putative transcriptional regulator, AraC family putative membrane protein
LEPBI_I0742	GSA	LEPBI_I3161	OmpA-like domain-containing protein
LEPBI_I0743	Putative transcriptional regulator	LEPBI_I3177	Putative Zn-dependent hydrolase
LEPBI_I0744	MFS domain-containing protein	LEPBI_I3182	Putative DNA repair system specific for alkylated DNA
LEPBI_I0745	Putative xylulokinase	LEPBI_I3208	Putative OsmC-like protein
LEPBI_I0826	ATP-dependent dethiobiotin synthetase BioD	LEPBI_I3323	Putative arylesterase
LEPBI_I1019	Cytokinin riboside 5'-monophosphate phosphoribohydrolase	LEPBI_I3343	Putative drug/sodium antiporter, Multi Antimicrobial Extrusion family (MATE)
LEPBI_I1094	Uncharacterized protein	LEPBI_I3386	Putative alpha-glucosidase
LEPBI_I1119	Putative chemotaxis protein histidine kinase putative membrane protein putative signal peptide	LEPBI_I3452	Putative glycosyltransferase
LEPBI_I1270	B12-binding domain-containing protein	LEPBI_I3453	Uncharacterized protein
LEPBI_I1321	Anti-sigma factor antagonist	LEPBI_II0151	Putative two-component response regulator

LEPBI_I1426	Uncharacterized protein	LEPBI_II0169	Putative N-carbamoyl-D-amino acid hydrolase
LEPBI_I1427	Putative transcriptional regulator	LEPBI_II0242	Putative transcriptional regulator
LEPBI_I1437	Putative cAMP-binding protein	LEPBI_p0012	Hemin degradation protein HemS
LEPBI_I1438	Putative cAMP-binding protein	LEPBI_p0015	ABC-type Fe3+-hydroxamate transport system, periplasmic component
LEPBI_I1458	PPM-type phosphatase domain-containing protein	LEPBI_p0018	Putative TonB-dependent outer membrane receptor
LEPBI_I1635	Uncharacterized protein		

Table S3.7. Genes lost by P1^{+}.

L. biflexa serovar Patoc strain Patoc1 (Paris) used as reference.

Locus tag	Product	Locus tag	Product
LEPBI_I0028	Anthranilate synthase component I	LEPBI_I2518	Uncharacterized protein
LEPBI_I0277	Uncharacterized protein	LEPBI_I2526	Putative two-component response regulator
LEPBI_I0300	Acetyl-CoA hydrolase	LEPBI_I2563	Putative glycosyl transferase
LEPBI_I0340	Alkyl hydroperoxide reductase AhpD	LEPBI_I2564	MULTIHEME_CYTC domain-containing protein
LEPBI_I0341	Uncharacterized protein	LEPBI_I2570	HlyD_D23 domain-containing protein
LEPBI_I0342	Methionine aminopeptidase	LEPBI_I2635	Uncharacterized protein
LEPBI_I0373	N-acetyltransferase domain-containing protein	LEPBI_I2637	Putative strictosidine synthase putative signal peptide
LEPBI_I0406	Uncharacterized protein	LEPBI_I2680	PPM-type phosphatase domain-containing protein
LEPBI_I0592	Permease, MFS superfamily	LEPBI_I2741	Transcriptional activator/repressor,heavy-metal dependent, MerR family
LEPBI_I0602	Uncharacterized protein	LEPBI_I2868	Uncharacterized protein
LEPBI_I0754	Putative sterol desaturase family protein putative membrane protein	LEPBI_I2871	Arsenical resistance operon repressor
LEPBI_I0855	Uncharacterized protein	LEPBI_I2875	Putative NADH oxidoreductase
LEPBI_I0857	Putative cation efflux system, RND family membrane protein	LEPBI_I2934	Long-chain acyl-CoA synthetase (AMP-forming)
LEPBI_I0858	Uncharacterized protein	LEPBI_I2996	Putative AraC-type transcriptional regulator
LEPBI_I0859	Uncharacterized protein	LEPBI_I2997	NAD(P)-bd_dom domain-containing protein
LEPBI_I0860	Uncharacterized protein	LEPBI_I3052	Uncharacterized protein

LEPBI_I0863	Uncharacterized protein	LEPBI_I3178	Gamma-glutamyltranspeptidase (Precursor)
LEPBI_I1070	Putative RNA polymerase sigma-70-like factor	LEPBI_I3220	Uncharacterized protein
LEPBI_I1111	Long-chain acyl-CoA synthetase, AMP-forming (Acyl-CoA synthetase)	LEPBI_I3346	Putative NADPH dehydrogenase
LEPBI_I1228	Putative transcriptional regulator, LysR family	LEPBI_I3356	Putative nitroreductase
LEPBI_I1378	Guanylate cyclase domain-containing protein	LEPBI_I3370	TGc domain-containing protein
LEPBI_I1451	Putative triacylglycerol lipase putative signal peptide	LEPBI_I3371	Alpha-E domain-containing protein
LEPBI_I1549	Two-component response regulator	LEPBI_I3372	TGc domain-containing protein
LEPBI_I1768	TPR_REGION domain-containing protein	LEPBI_I3373	Alpha-E domain-containing protein
LEPBI_I1855	Putative histidine kinase sensor protein putative membrane protein	LEPBI_I3374	Uncharacterized protein
LEPBI_I2050	TPR_REGION domain-containing protein	LEPBI_I3424	Putative transcriptional regulator
LEPBI_I2159	Uncharacterized protein	LEPBI_I3438	Uncharacterized protein
LEPBI_I2160	Uncharacterized protein	LEPBI_II0111	Putative dehydrogenase
LEPBI_I2161	Putative DNA polymerase	LEPBI_II0112	Uncharacterized protein
LEPBI_I2305	PilZ domain-containing protein	LEPBI_II0113	Cyclopropane-fatty-acyl-phospholipid synthase
LEPBI_I2408	Ammonium transporter	LEPBI_II0173	RND divalent metal cation efflux transporter
LEPBI_I2457	Putative thioredoxin-disulfide reductase putative membrane protein	LEPBI_II0209	Putative cyclopropane-fatty-acyl-phospholipid synthase
LEPBI_I2492	Ethanolamine ammonia-lyase light chain	LEPBI_II0241	Uncharacterized protein
LEPBI_I2493	Ethanolamine ammonia-lyase, heavy chain (Ethanolamine ammonia-lyase large subunit)		

Table S3.8. Proteins that exhibit AA changes in P2 and the mutations are conserved in P1$^-$ and P1$^+$.

L. interrogans serovar Lai strain 56601 used as reference.

Locus tag	Product	Locus tag	Product
LA_0005	DNA gyrase subunit B	LA_2337	chromosome-partitioning ATPase
LA_0006	DNA gyrase subunit A	LA_2345	ATP-dependent protease ATP-binding subunit HslU
LA_0020	short-chain dehydrogenase	LA_2346	ATP-dependent protease peptidase subunit

LA_0021	translation factor Sua5	LA_2347	integrase/recombinase XerD
LA_0025	flagellar motor switch protein G	LA_2350	isopropylmalate/homocitrate/citramalate synthase
LA_0106	long-chain-fatty-acid CoA ligase	LA_2360	ribonucleotide-diphosphate reductase subunit alpha
LA_0243	cytochrome c oxidase polypeptide I	LA_2373	type II secretory pathway component protein F
LA_0249	transcriptional regulator	LA_2374	type II secretory pathway ATPase protein E
LA_0250	PIN domain-containing integral membrane protein	LA_2375	type II secretory pathway component protein D
LA_0251	phosphoenolpyruvate carboxykinase	LA_2380	hypothetical protein
LA_0256	aspartyl/glutamyl-tRNA amidotransferase subunit B	LA_2387	50S ribosomal protein L19
LA_0258	DNA polymerase III subunit alpha	LA_2388	tRNA (guanine-N(1)-)-methyltransferase
LA_0273	lipoprotein releasing system permease	LA_2391	30S ribosomal protein S16
LA_0274	lipoprotein releasing system LolD	LA_2394	ribulose-5-phosphate 3-epimerase
LA_0280	cAMP-binding protein	LA_2400	two-component system response regulator
LA_0296	Zn-dependent alcohol dehydrogenase	LA_2403	HPr kinase/phosphorylase
LA_0309	cAMP-binding protein	LA_2405	ABC transporter ATP-binding protein
LA_0313	elongation factor G	LA_2408	2-dehydro-3-deoxyphosphooctonate aldolase
LA_0319	phosphoribose diphosphate:decaprenyl-phosphate phosphoribosyltransferase	LA_2409	CTP synthetase
LA_0332	6-carboxytetrahydropterin synthase QueD	LA_2417	endoflagellar hook-filament protein
LA_0334	proton-translocating transhydrogenase subunit alpha	LA_2418	endoflagellar filament core protein
LA_0335	NAD(P)(+) transhydrogenase subunit beta	LA_2432	biotin carboxylase
LA_0347	flagellar basal-body rod protein FlgB	LA_2433	acetyl-CoA carboxylase
LA_0360	glycine dehydrogenase	LA_2436	hypothetical protein
LA_0361	glycine cleavage system protein H	LA_2438	peptide deformylase
LA_0362	glycine cleavage system aminomethyltransferase T	LA_2457	methionyl aminopeptidase
LA_0369	ribose-5-phosphate isomerase	LA_2458	hypothetical protein

LA_0391	ATP-dependent protease ClpA	LA_2469	MCP methylation inhibitor
LA_0395	anti-sigma factor antagonist	LA_2483	integrase/recombinase XerD
LA_0408	tryptophanyl-tRNA synthetase	LA_2494	thioredoxin reductase
LA_0411	electron transfer flavoprotein subunit alpha	LA_2506	imidazole glycerol phosphate synthase subunit HisF
LA_0412	electron transfer flavoprotein subunit beta	LA_2507	aspartyl/glutamyl-tRNA amidotransferase subunit A
LA_0414	isovaleryl-CoA dehydrogenase	LA_2513	transcription-repair coupling factor
LA_0439	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	LA_2558	ATP-dependent protease ATP-binding subunit ClpX
LA_0457	acetyl-CoA acetyltransferase	LA_2559	ATP-dependent Clp protease proteolytic subunit
LA_0463	6,7-dimethyl-8-ribityllumazine synthase	LA_2569	acetolactate synthase small subunit
LA_0547	hypothetical protein	LA_2570	acetolactate synthase large subunit
LA_0552	periplasmic solute-binding protein	LA_2572	TPR-repeat-containing protein
LA_0560	acyl-CoA dehydrogenase	LA_2575	hypothetical protein
LA_0611	cell division protein FtsA	LA_2576	hypothetical protein
LA_0612	cell division protein FtsZ	LA_2589	flagellar assembly protein H
LA_0618	quinolinate synthetase	LA_2591	flagellar MS-ring protein
LA_0627	glycosyltransferase	LA_2592	endoflagellar biosynthesis/type III secretory pathway ATPase
LA_0629	hypothetical protein	LA_2602	polymerase
LA_0645	hypothetical protein	LA_2603	RNA polymerase sigma factor WhiG
LA_0649	excinuclease ABC subunit B	LA_2605	ParA protein
LA_0653	anti-sigma factor antagonist	LA_2607	endoflagellar biosynthesis protein
LA_0666	cytochrome c peroxidase	LA_2610	endoflagellar biosynthesis pathway protein
LA_0671	citrate synthase	LA_2611	flagellar biosynthesis protein FliP
LA_0693	aspartate kinase	LA_2627	MaoC family protein
LA_0720	hypothetical protein	LA_2631	ATP-dependent Clp protease adaptor protein ClpS
LA_0727	carbamoyl phosphate synthase large subunit	LA_2632	hypothetical protein
LA_0737	elongation factor Tu	LA_2654	co-chaperonin GroES
LA_0739	50S ribosomal protein L3	LA_2655	molecular chaperone GroEL

LA_0741	50S ribosomal protein L23	LA_2667	flagellar basal body rod protein FlgG
LA_0742	50S ribosomal protein L2	LA_2674	metal-dependent amidase/aminoacylase/carboxypeptidase
LA_0743	30S ribosomal protein S19	LA_2690	bacterioferritin
LA_0744	50S ribosomal protein L22	LA_2701	L-lysine 2,3-aminomutase
LA_0745	30S ribosomal protein S3	LA_2706	thiamine biosynthesis enzyme
LA_0746	50S ribosomal protein L16	LA_2712	Fe-S oxidoreductase
LA_0748	30S ribosomal protein S17	LA_2733	peptide chain release factor 3
LA_0749	50S ribosomal protein L14	LA_2739	hypothetical protein
LA_0750	50S ribosomal protein L24	LA_2751	hypothetical protein
LA_0751	50S ribosomal protein L5	LA_2754	rod shape-determining protein
LA_0752	30S ribosomal protein S14	LA_2755	transpeptidase/penicillin-binding protein
LA_0753	30S ribosomal protein S8	LA_2756	cell shape-determining protein
LA_0754	50S ribosomal protein L6	LA_2759	rod shape-determining protein MreB
LA_0756	30S ribosomal protein S5	LA_2776	ATP synthase F0F1 subunit beta
LA_0759	preprotein translocase subunit SecY	LA_2778	ATP synthase F0F1 subunit gamma
LA_0760	adenylate kinase	LA_2779	ATP synthase F0F1 subunit alpha
LA_0762	30S ribosomal protein S13	LA_2808	ABC transporter permease
LA_0763	30S ribosomal protein S11	LA_2809	peroxiredoxin
LA_0764	30S ribosomal protein S4	LA_2821	hypothetical protein
LA_0765	DNA-directed RNA polymerase subunit alpha	LA_2840	ornithine carbamoyltransferase
LA_0776	L,L-diaminopimelate aminotransferase	LA_2841	citrate lyase
LA_0778	aspartate carbamoyltransferase	LA_2848	flagellar hook protein FlgE
LA_0790	citrate synthase	LA_2876	hypothetical protein
LA_0799	cell shape determination protein	LA_2891	Na+/H+ antiporter
LA_0810	Holliday junction DNA helicase RuvB	LA_2894	transcriptional regulator
LA_0811	hypothetical protein	LA_2916	DNA repair protein RadA
LA_0828	acetyl-CoA acetyltransferase	LA_2917	ribonuclease D
LA_0833	pantothenate kinase	LA_2924	pyruvate kinase
LA_0839	anti-sigma factor antagonist	LA_2947	thiosulfate sulfurtransferase

LA_0841	23S rRNA (adenine(2503)-C(2))-methyltransferase RlmN	LA_2956	methylmalonyl-CoA mutase
LA_0845	2-isopropylmalate synthase	LA_2996	alcohol dehydrogenase
LA_0851	50S ribosomal protein L27	LA_3004	prolipoprotein diacylglyceryl transferase
LA_0852	GTPase ObgE	LA_3052	aminotransferase
LA_0854	gamma-glutamyl phosphate reductase	LA_3054	hypothetical protein
LA_0887	NADH dehydrogenase (ubiquinone) chain K	LA_3070	anti-sigma-factor antagonist
LA_0889	NADH dehydrogenase (ubiquinone) chain H	LA_3080	adenylosuccinate lyase
LA_0890	NADH dehydrogenase (ubiquinone) chain F	LA_3083	hypothetical protein
LA_0892	NADH dehydrogenase (ubiquinone) chain D	LA_3085	guanosine polyphosphate pyrophosphohydrolase/synthetase
LA_0895	NADH dehydrogenase (ubiquinone) chain A	LA_3094	ferric uptake regulator
LA_0909	methylglyoxal synthase	LA_3095	queuine tRNA-ribosyltransferase
LA_0942	transcription elongation factor NusA	LA_3096	anti-sigma factor antagonist
LA_0947	polynucleotide phosphorylase	LA_3135	acetyl-CoA acetyltransferase
LA_1019	hypothetical protein	LA_3143	acyl-CoA dehydrogenase
LA_1020	50S ribosomal protein L31	LA_3144	acyl-CoA transferase
LA_1021	transcription termination factor Rho	LA_3158a	hypothetical protein
LA_1032	enoyl-CoA hydratase	LA_3219	hypothetical protein
LA_1044	cell shape determination protein	LA_3265	hypothetical protein
LA_1047	diaminopimelate decarboxylase	LA_3266	Fe-S-cluster-containing hydrogenase
LA_1101	succinyl-CoA synthetase subunit alpha	LA_3268	hypothetical protein
LA_1102	succinyl-CoA synthetase subunit beta	LA_3285	1-deoxy-D-xylulose-5-phosphate synthase
LA_1105	(dimethylallyl)adenosine tRNA methylthiotransferase	LA_3286	adenylate cyclase
LA_1110	adenylosuccinate synthetase	LA_3289	tryptophan synthase subunit beta
LA_1112	glycerol-3-phosphate dehydrogenase	LA_3290	prolyl-tRNA synthetase
LA_1132	hypothetical protein	LA_3294	undecaprenyl pyrophosphate synthase
LA_1137	hypothetical protein	LA_3295	ribosome-recycling factor
LA_1138	2-methylthioadenine synthetase	LA_3296	uridylate kinase
LA_1142	preprotein translocase subunit SecD	LA_3297	elongation factor Ts

LA_1147	bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II	LA_3298	30S ribosomal protein S2
LA_1156	sulfate ABC transporter permease	LA_3300	cell shape determination protein
LA_1158	ABC transporter ATP-binding protein	LA_3302	phosphopantetheinyl transferase
LA_1189	sigma 54 modulation protein	LA_3306	4-hydroxy-tetrahydrodipicolinate synthase
LA_1205	RNA polymerase ECF-type sigma factor	LA_3349	cytochrome c-type biogenesis protein
LA_1221	membrane carboxypeptidase/penicillin-binding protein	LA_3350	cytochrome c biogenesis protein
LA_1222	dihydrolipoamide acetyltransferase	LA_3354	Zn-dependent alcohol dehydrogenase
LA_1223	dihydrolipoamide dehydrogenase	LA_3363	acyl-CoA dehydrogenase
LA_1224	2-oxoglutarate dehydrogenase subunit E1	LA_3378	transcription elongation factor
LA_1240	threonyl-tRNA synthetase	LA_3380	endoflagellar filament sheath protein
LA_1242	translation initiation factor IF-3	LA_3406	nucleotide-binding protein
LA_1244	50S ribosomal protein L20	LA_3408	30S ribosomal protein S9
LA_1253	two-component system response regulator	LA_3409	50S ribosomal protein L13
LA_1255	segregation/condensation protein B	LA_3416	30S ribosomal protein S7
LA_1260	30S ribosomal protein S1	LA_3417	30S ribosomal protein S12
LA_1262	ATP phosphoribosyltransferase	LA_3419	DNA-directed RNA polymerase subunit beta'
LA_1263	50S ribosomal protein L32	LA_3420	DNA-directed RNA polymerase subunit beta
LA_1264	glycerol-3-phosphate acyltransferase PlsX	LA_3421	50S ribosomal protein L7/L12
LA_1293	aspartate/tyrosine/aromatic aminotransferase	LA_3422	50S ribosomal protein L10
LA_1295	hypothetical protein	LA_3423	50S ribosomal protein L1
LA_1302	SsrA-binding protein	LA_3424	50S ribosomal protein L11
LA_1309	chromosome segregation protein	LA_3425	transcription antiterminator
LA_1313	glutamine synthetase	LA_3426	preprotein translocase subunit SecE
LA_1319	hypothetical protein	LA_3442	peroxiredoxin
LA_1323	phosphoribosylformylglycinamidine synthase II	LA_3459	polyphosphate kinase
LA_1325	isoleucyl-tRNA synthetase	LA_3461	FAD/FMN-containing dehydrogenase

LA_1378	GTP-binding protein BipA	LA_3484	short-chain dehydrogenase
LA_1388	glycyl-tRNA synthetase	LA_3498	phosphate ABC transporter regulatory protein
LA_1392	ribokinase	LA_3560	selenocysteine lyase
LA_1409	serine hydroxymethyltransferase	LA_3563	ABC transporter ATP-binding protein
LA_1417	lactoylglutathione lyase	LA_3572	anti-sigma factor antagonist
LA_1438	antimicrobial peptide ABC transporter ATP-binding protein	LA_3573	3-deoxy-manno-octulosonate cytidylyltransferase
LA_1447	LexA repressor	LA_3575	flagellar basal body protein FliL
LA_1451	CDP-diacylglycerol--serine O-phosphatidyltransferase	LA_3576	flagellar motor protein MotB
LA_1459	UDP-glucose 6-dehydrogenase	LA_3577	endoflagellar motor protein A
LA_1463	5-(carboxyamino)imidazole ribonucleotide mutase	LA_3579	glycosyltransferase
LA_1471	membrane-bound proton-translocating pyrophosphatase	LA_3596	ATP-dependent Lon protease
LA_1483	response regulator	LA_3597	pyridoxal phosphate-dependent aminotransferase
LA_1487	tRNA-specific 2-thiouridylase MnmA	LA_3598	ferritin
LA_1488	glycolate oxidase	LA_3606	hypothetical protein
LA_1511	L-aspartate oxidase	LA_3607	acyl dehydratase MaoC
LA_1517	host factor I related protein	LA_3627	acyl-CoA dehydrogenase
LA_1518	mannose-1-phosphate guanylyltransferase	LA_3630	multidrug ABC transporter ATPase/permease
LA_1521	chemotaxis protein CheW	LA_3634	GTP-binding protein
LA_1532	fructose-bisphosphate aldolase	LA_3636	cyclic amidohydrolase
LA_1543	heavy metal efflux pump	LA_3684	ABC-F family ATPase
LA_1678	50S ribosomal protein L9	LA_3694	multidrug ABC transporter ATPase/permease
LA_1680	aspartyl-tRNA synthetase	LA_3703	heat shock operon regulator HrcA
LA_1682	phosphate starvation-inducible protein	LA_3705	molecular chaperone DnaK
LA_1692	two-component system response regulator	LA_3706	molecular chaperone DnaJ
LA_1696	triosephosphate isomerase	LA_3709	oxidoreductase
LA_1703	phosphoglycerate kinase	LA_3714	leucyl-tRNA synthetase
LA_1704	glyceraldehyde 3-phosphate dehydrogenase	LA_3747	threonine synthase

LA_1713	carbonic anhydrase/acetyltransferase	LA_3763	valyl-tRNA synthetase
LA_1719	cysteine synthase	LA_3767	peptide chain release factor 2
LA_1729	aspartate alpha-decarboxylase	LA_3783	adenosine deaminase
LA_1855	endoflagellar distal basal body rod protein	LA_3793	hemolysin
LA_1865	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase	LA_3794	hypothetical protein
LA_1866	asparaginyl-tRNA synthetase	LA_3803	acetyl-CoA carboxylase
LA_1880	transcriptional regulator	LA_3806	ammonium transporter
LA_1888	GTP-binding protein LepA	LA_3807	nitrogen regulatory protein P-II
LA_1896	succinate dehydrogenase/fumarate reductase iron-sulfur subunit	LA_3819	ATP-dependent Zn protease
LA_1897	succinate dehydrogenase flavoprotein subunit	LA_3820	peptidyl-tRNA hydrolase
LA_1911	D-3-phosphoglycerate dehydrogenase	LA_3822	ribose-phosphate pyrophosphokinase
LA_1927	hypothetical protein	LA_3905	pyridoxal phosphate-dependent aminotransferase
LA_1929	CAP family transcriptional factor	LA_3911	hypothetical protein
LA_1930	acyl-CoA dehydrogenase	LA_3918	methionyl-tRNA synthetase
LA_1935	hypothetical protein	LA_3936	phosphopantothenoylcysteine decarboxylase
LA_1948	two-component system response regulator	LA_3939	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
LA_1951	phosphopyruvate hydratase	LA_3941	phosphoesterase
LA_1953	ATP-dependent protease ClpP	LA_3961	hypothetical protein
LA_1960	preprotein translocase subunit SecA	LA_3964	GDP-mannose 4,6-dehydratase
LA_1964	multidrug efflux pump	LA_3975	acylphosphatase
LA_1977	FAD-binding oxidoreductase	LA_3978	lysine methyltransferase
LA_1986	inosine-5'-monophosphate dehydrogenase	LA_3982	hypothetical protein
LA_1995	lysyl-tRNA synthetase	LA_4023	hypothetical protein

LA_2009	pyruvate dehydrogenase subunit beta	LA_4034	carbonic anhydrase/acetyltransferase
LA_2010	pyruvate dehydrogenase subunit alpha	LA_4036	ATP-dependent RNA helicase
LA_2017	endoflagellar filament core protein	LA_4138	3-hydroxyacyl-CoA dehydrogenase
LA_2019	flagellin protein	LA_4139	acetyl-CoA acetyltransferase
LA_2048	phospho-N-acetylmuramoyl-pentapeptide-transferase	LA_4150	phenylalanyl-tRNA synthetase subunit alpha
LA_2056	ABC transporter ATP-binding protein	LA_4155	ABC transporter permease
LA_2057	ABC transporter permease	LA_4165	argininosuccinate synthase
LA_2061	homoserine O-acetyltransferase	LA_4167	phosphopantetheine adenylyltransferase
LA_2072	HflC membrane associated protease	LA_4168	nucleoside diphosphate kinase
LA_2081	flagellar motor switch protein FliM	LA_4190	glutamyl-tRNA synthetase
LA_2081a	ferredoxin family protein	LA_4193	DNA gyrase subunit A
LA_2095	isopropylmalate isomerase large subunit	LA_4194	DNA gyrase subunit B
LA_2113	ABC transporter permease	LA_4228	ABC transporter permease
LA_2114	ABC transporter ATP-binding protein	LA_4229	dipeptide ABC transporter permease
LA_2118	metal-dependent hydrolase	LA_4242	ketol-acid reductoisomerase
LA_2119	glycerol kinase	LA_4248	purine nucleoside phosphorylase
LA_2129	DNA/RNA helicase	LA_4254	acetyl-CoA synthetase
LA_2130	aspartate aminotransferase	LA_4255	GTP cyclohydrolase I
LA_2139	malate dehydrogenase	LA_4308	flagellar hook-associated protein FlgL
LA_2143	biotin synthase	LA_4309	flagellar hook-associated protein FlgK
LA_2151	two-component system response regulator	LA_4327	ferredoxin-NADP reductase
LA_2152	3-isopropylmalate dehydrogenase	LA_4329	transcriptional coactivator-like protein
LA_2162	50S ribosomal protein L28	LA_4339	seryl-tRNA synthetase
LA_2172	glycosyltransferase	LA_4340	Mg-dependent DNase
LA_2173	lipopolysaccharide core biosynthesis protein	LA_4342	cell shape determination protein
LA_2174	hypothetical protein	LA_4347	ParA
LA_2179	recombinase A	LA_4359	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA
LA_2188	ATPase	LA_4360	branched-chain amino acid aminotransferase

LA_2202	2-isopropylmalate synthase	LB_013	glutamate-1-semialdehyde aminotransferase
LA_2207	hypothetical protein	LB_027	chromosome partitioning protein ParB
LA_2212	excinuclease ABC subunit A	LB_074	methylmalonyl-CoA mutase
LA_2226	fructose-1,6-bisphosphatase	LB_082	3-ketoacyl-ACP reductase
LA_2228	hypothetical protein	LB_106	adenosylhomocysteinase
LA_2229	30S ribosomal protein S21	LB_108	B12-dependent methionine synthase
LA_2232	RNA polymerase sigma-70 factor	LB_111	diphosphate--fructose-6-phosphate 1-phosphotransferase
LA_2237	tyrosyl-tRNA synthetase	LB_114	acetylglutamate kinase
LA_2280	endoflagellar biosynthesis chaperone	LB_170	capsular polysaccharide biosynthesis protein
LA_2283	bifunctional phosphoribosylaminoimid azolecarboxamide formyltransferase/IMP cyclohydrolase	LB_176	adenine phosphoribosyltransferase
LA_2286	translaldolase	LB_208	alginate o-acetyltransferase
LA_2297	hypothetical protein	LB_212	hypothetical protein
LA_2306	UDP-3-O-[3-hydroxymyristoyl] N-acetylglucosamine deacetylase	LB_262	UDP-glucose pyrophosphorylase
LA_2309	long-chain-fatty-acid CoA ligase	LB_274	methylmalonyl-CoA mutase
LA_2318	hypothetical protein	LB_282	transport protein ExbD
LA_2319	ExbD-like biopolymer transport protein	LB_286	glutamate synthase
LA_2323	hypothetical protein	LB_294	elongation factor P
LA_2324	3-ketoacyl-ACP reductase	LB_311	spermidine synthase
LA_2325	transcriptional regulator	LB_327	aconitate hydratase
LA_2326	nucleoside-diphosphate-sugar epimerase	LB_333	two-component system response regulator
LA_2329	GTP-dependent nucleic acid-binding protein EngD	LB_353	pyruvate kinase
LA_2335	3-polyprenyl-4-hydroxybenzoate decarboxylase	LB_355	aspartate-semialdehyde dehydrogenase
LA_2336	hypothetical protein	LB_365	chromosome partitioning protein ParA

Table S3.9. GO functional enrichment of proteins with AA changes in P2 and mutations conserved in $P1^-$ and $P1^+$.

GO categories are organized from most to least represented, considering the ratio between the mutated proteins within a GO process and the total number of proteins within that process in the reference used (*L. interrogans* serovar Lai strain 56601).

GO ID	Description	Observed gene count	Reference gene count	Matching proteins in the network
GO:0015935	Small ribosomal subunit	6	6	rpsS, rpsE, rpsD, rpsB, rpsG, rpsL
GO:0006414	Translational elongation	7	8	fusA, tufB, smpB, lepA, tsf, hflX, efp
GO:0009420	Bacterial-type flagellum filament	5	6	LA_2017, flaB, flgL, LA_2418, LA_4308
GO:0006165	Nucleoside diphosphate phosphorylation	9	11	gpmI, fbaB, tpiA, pgk, eno, PykF, ndk, pfkA, Pyk2
GO:0006096	Glycolytic process	8	10	gpmI, fbaB, tpiA, pgk, eno, PykF, pfkA, Pyk2
GO:0044391	Ribosomal subunit	11	14	rplB, rpsS, rplV, rplN, rpsE, rpsD, rpmF, rpsB, rpsG, rpsL, rplA
GO:0006099	Tricarboxylic acid cycle	8	11	LA_0671, LA_0790, sucD, sucC, sucA, LA_1896, mdh, acnA
GO:0006418	tRNA aminoacylation for protein translation	15	21	trpS, thrS, ileS, grs1, aspS, asnS, lysU, tyrS, proS, leuS, valS, metG, pheS, glnS, serS
GO:0009060	Aerobic respiration	9	13	cyoB, LA_0671, LA_0790, sucD, sucC, sucA, LA_1896, mdh, acnA
GO:0004812	aminoacyl-tRNA ligase activity	15	22	trpS, thrS, ileS, grs1, aspS, asnS, lysU, tyrS, proS, leuS, valS, metG, pheS, glnS, serS
GO:0005198	Structural molecule activity	44	65	rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, rpmF, rplI, LA_1855, LA_2017, flaB, rpmB, rpsU, fliS, rplS, rpsP, flgL, LA_2418, LA_2667, rpsB, rpsI, rplM, rpsG, rpsL, rplL, rplA, rplK, LA_4308, FlgK

GO:0003735	Structural constituent of ribosome	35	53	rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, rpmF, rplI, rpmB, rpsU, rplS, rpsP, rpsB, rpsI, rplM, rpsG, rpsL, rplL, rplA, rplK
GO:0019843	rRNA binding	27	41	rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rlmN, rplT, bipA, rplI, rpsG, rpsL, rplJ, rplA, rplK
GO:0006412	Translation	67	103	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, rplI, aspS, asnS, lepA, lysU, rpmB, rpsU, tyrS, rplS, rpsP, def, LA_2507, prfC, proS, frr, tsf, rpsB, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, hflX, leuS, valS, LA_3767, LA_3820, metG, pheS, glnS, serS, efp
GO:0000049	tRNA binding	13	20	rplP, rplE, rpsM, rlmN, thrS, ileS, bipA, trmU, rpsG, rpsL, rplA, pheS, glnS
GO:0009201	Ribonucleoside triphosphate biosynthetic process	7	11	pyrG, fliI, atpD, atpG, atpA, pyrH, ndk
GO:0005840	Ribosome	38	60	rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, LA_1260, rpmF, rplI, rpmB, rpsU, rplS, rpsP, LA_2631, rpsB, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK
GO:0016052	Carbohydrate catabolic process	10	16	rpiB, gpmI, fbaB, tpiA, pgk, eno, LA_2119, PykF, pfkA, Pyk2
GO:0045333	Cellular respiration	11	19	cyoB, LA_0671, LA_0790, nuoK, sucD, sucC, sucA, LA_1896, mdh, LA_3266, acnA
GO:0008135	Translation factor activity, RNA binding	8	14	fusA, tufB, infC, lepA, prfC, tsf, LA_3767, efp
GO:0046034	ATP metabolic process	13	23	gpmI, nuoK, fbaB, tpiA, pgk, eno, fliI, atpD, atpG, atpA, PykF, pfkA, Pyk2

GO:0043604	Amide biosynthetic process	71	126	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, rplI, aspS, panD, asnS, lepA, lysU, bioB, rpmB, rpsU, tyrS, rplS, rpsP, def, LA_2507, prfC, proS, frr, tsf, rpsB, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, hflX, leuS, valS, LA_3767, LA_3820, metG, pheS, glnS, acsA, folE, serS, efp
GO:0006417	Regulation of translation	11	20	fusA, tufB, LA_1189, infC, lepA, prfC, tsf, rplA, hflX, LA_3767, efp
GO:0015980	Energy derivation by oxidation of organic compounds	12	22	cyoB, LA_0671, LA_0790, nuoK, sucD, sucC, sucA, LA_1896, mdh, KamA, LA_3266, acnA
GO:0043603	Cellular amide metabolic process	74	136	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, glyA, rplI, aspS, panD, folD, asnS, lepA, lysU, bioB, rpmB, rpsU, tyrS, rplS, rpsP, def, LA_2507, prfC, proS, frr, tsf, rpsB, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, hflX, leuS, valS, LA_3767, LA_3820, metG, LA_3936, pheS, glnS, acsA, folE, serS, efp
GO:1990204	Oxidoreductase complex	8	15	gcvP, gcvH, LA_1112, LA_1222, sucA, guaB, nrdA, LA_3939
GO:0071973	Bacterial-type flagellum-dependent cell motility	14	28	LA_0025, FlgB, LA_1855, LA_2017, flaB, fliM, fliS, flgL, LA_2418, fliF, LA_2667, flgE, LA_3575, LA_4308
GO:0043228	Non-membrane-bounded organelle	59	118	gyrB, LA_0006, LA_0025, FlgB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, LA_1260, rpmF, rplI, LA_1855, LA_2017, flaB, fliM, rpmB, rpsU, rplS, rpsP, flgL, LA_2418, LA_2458, fliF, fliQ, fliP, LA_2631, LA_2667, flgE, rpsB, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, LA_3575, LA_4193, LA_4194, LA_4308, FlgK
GO:0043232	Intracellular non-membrane-	43	86	gyrB, LA_0006, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, LA_1260, rpmF, rplI, rpmB,

	bounded organelle			rpsU, rplS, rpsP, LA_2458, LA_2631, rpsB, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, LA_4193, LA_4194
GO:0009288	Bacterial-type flagellum	16	32	LA_0025, FlgB, LA_1855, LA_2017, flaB, fliM, flgL, LA_2418, fliF, fliQ, fliP, LA_2667, flgE, LA_3575, LA_4308, FlgK
GO:0097588	Archaeal or bacterial-type flagellum-dependent cell motility	17	35	LA_0025, FlgB, LA_1855, LA_2017, flaB, fliM, LA_2151, fliS, flgL, LA_2418, fliF, LA_2667, flgE, LA_3575, LA_3576, LA_3577, LA_4308
GO:0009084	Glutamine family amino acid biosynthetic process	8	17	carB, pyrB, proA, glnA, argF, argG, argB, gltB
GO:0019693	Ribose phosphate metabolic process	33	71	gpmI, carB, adk, pyrB, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, eno, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, pyrH, pyrC, prsA, LA_3936, coaD, ndk, acsA, pfkA, apt, Pyk2
GO:0006163	Purine nucleotide metabolic process	29	63	gpmI, adk, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, folD, eno, guaB, purH, LA_2318, fliI, atpD, atpG, atpA, PykF, purB, spoT, LA_3936, coaD, ndk, acsA, pfkA, apt, Pyk2
GO:0009150	Purine ribonucleotide metabolic process	27	59	gpmI, adk, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, eno, purH, LA_2318, fliI, atpD, atpG, atpA, PykF, purB, spoT, LA_3936, coaD, ndk, acsA, pfkA, apt, Pyk2
GO:0009259	Ribonucleotide metabolic process	32	70	gpmI, carB, adk, pyrB, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, eno, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, pyrH, pyrC, LA_3936, coaD, ndk, acsA, pfkA, apt, Pyk2
GO:0009425	Bacterial-type flagellum basal body	10	22	LA_0025, FlgB, LA_1855, fliM, fliF, fliQ, fliP, LA_2667, flgE, LA_3575
GO:0046390	Ribose phosphate biosynthetic process	24	53	carB, adk, pyrB, coaX, purA, LA_1323, purE, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, purB, spoT, pyrH, pyrC, prsA, LA_3936, coaD, ndk, acsA, apt
GO:0072521	Purine-containing compound	30	67	gpmI, adk, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, folD, eno, guaB, purH, LA_2318, fliI, atpD, atpG, atpA, PykF, purB,

	metabolic process			spoT, LA_3936, coaD, ndk, mtnP, acsA, pfkA, apt, Pyk2
GO:0140101	Catalytic activity, acting on a tRNA	21	47	gatB, trpS, rlmN, thrS, ileS, grs1, aspS, asnS, lysU, tyrS, trmD, LA_2507, tgt, proS, leuS, valS, LA_3820, metG, pheS, glnS, serS
GO:0009260	Ribonucleotide biosynthetic process	23	52	carB, adk, pyrB, coaX, purA, LA_1323, purE, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, purB, spoT, pyrH, pyrC, LA_3936, coaD, ndk, acsA, apt
GO:0006164	Purine nucleotide biosynthetic process	19	43	adk, coaX, purA, LA_1323, purE, folD, purH, LA_2318, fliI, atpD, atpG, atpA, purB, spoT, LA_3936, coaD, ndk, acsA, apt
GO:0010467	Gene expression	86	195	LA_0021, gatB, fusA, trpS, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpoA, rlmN, rpmA, nusA, pnp, rpmE, rho, LA_1105, rimO, LA_1205, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, trmU, rplI, aspS, asnS, lepA, lysU, rpmB, rpsU, rpoD, tyrS, rplS, trmD, rpsP, def, LA_2507, LA_2603, prfC, LA_2917, tgt, proS, frr, tsf, rpsB, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, hflX, leuS, valS, LA_3767, LA_3820, metG, pheS, glnS, serS, gidA, efp
GO:0006091	Generation of precursor metabolites and energy	27	62	cyoB, etfA, etfB, gpmI, LA_0666, LA_0671, LA_0790, nuoK, sucD, sucC, LA_1222, sucA, fbaB, tpiA, pgk, LA_1896, eno, mdh, mipB, rpe, KamA, PykF, LA_3265, LA_3266, pfkA, acnA, Pyk2
GO:0072522	Purine-containing compound biosynthetic process	20	46	adk, coaX, purA, LA_1323, purE, folD, purH, LA_2318, fliI, atpD, atpG, atpA, purB, spoT, LA_3936, coaD, ndk, mtnP, acsA, apt
GO:1902494	Catalytic complex	22	51	gcvP, gcvH, ribH, uvrB, ruvB, sucD, sucC, LA_1112, LA_1222, sucA, eno, guaB, uvrA, hslU, hslV, nrdA, LA_2432, LA_2433, ilvH, ppk, LA_3803, LA_3939
GO:0003723	RNA binding	49	114	LA_0021, fusA, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rlmN, nusA, pnp, rho, thrS, infC, rplT, LA_1260, smpB, ileS, bipA, trmU, hfq, rplI, lepA, tyrS, prfC, tsf, rpsG, rpsL, rplJ, rplA, rplK, LA_3767, pheS, glnS, efp, acnA

GO:0006399	tRNA metabolic process	24	56	LA_0021, trpS, rlmN, LA_1105, rimO, thrS, ileS, grs1, trmU, aspS, asnS, lysU, tyrS, trmD, LA_2917, tgt, proS, leuS, valS, metG, pheS, glnS, serS, gidA
GO:0009152	Purine ribonucleotide biosynthetic process	18	42	adk, coaX, purA, LA_1323, purE, purH, LA_2318, fliI, atpD, atpG, atpA, purB, spoT, LA_3936, coaD, ndk, acsA, apt
GO:0051246	Regulation of protein metabolic process	12	28	fusA, tufB, LA_1189, infC, lepA, LA_2072, prfC, tsf, rplA, hflX, LA_3767, efp
GO:0019842	Vitamin binding	15	35	dapL, sucA, LA_1293, glyA, LA_2130, IlvB, icmF, LA_3052, dxs, CsdB, thrC, hemL, LB_074, metH, Mcm3
GO:0046395	Carboxylic acid catabolic process	11	26	gcvP, gcvH, gcvT, LA_0457, LA_0828, LA_1032, KamA, LA_3135, LA_3936, fadB, LA_4139
GO:0009066	Aspartate family amino acid metabolic process	10	24	lysC, folD, met2, abgB, KamA, dapA, thrC, mtnP, metH, asd
GO:0006520	Cellular amino acid metabolic process	55	134	gcvP, gcvH, gcvT, trpS, lysC, carB, dapL, pyrB, proA, lysA, thrS, hisG, LA_1293, glnA, ileS, grs1, glyA, aspS, LA_1719, panD, folD, asnS, lysU, met2, leuC, LA_2130, LA_2152, tyrS, pyrG, hisF, ilvH, abgB, KamA, argF, sseA, LA_3052, trpB, proS, dapA, CsdB, leuS, thrC, valS, metG, pheS, argG, glnS, ilvC, mtnP, serS, ilvE, metH, argB, gltB, asd
GO:0016874	Ligase activity	34	86	gatB, trpS, carB, pyrB, sucD, sucC, purA, thrS, glnA, LA_1323, ileS, grs1, aspS, asnS, lysU, tyrS, pyrG, LA_2432, LA_2433, LA_2507, atpD, atpG, atpA, proS, leuS, valS, LA_3803, metG, LA_3936, pheS, argG, glnS, acsA, serS
GO:0009156	Ribonucleoside monophosphate biosynthetic process	11	28	carB, adk, pyrB, purA, LA_1323, purE, purH, purB, pyrC, prsA, apt
GO:0009117	Nucleotide metabolic process	39	100	gpmI, nadA, carB, adk, pyrB, coaX, purA, LA_1222, LA_1323, purE, nadB, fbaB, tpiA, pgk, folD, eno, guaB, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT,

				pyrH, pyrC, add, prsA, LA_3936, coaD, ndk, acsA, pfkA, apt, LB_212, Pyk2
GO:0009165	Nucleotide biosynthetic process	27	70	nadA, carB, adk, pyrB, coaX, purA, LA_1323, purE, nadB, folD, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, purB, spoT, pyrH, pyrC, prsA, LA_3936, coaD, ndk, acsA, apt
GO:0017111	Nucleoside-triphosphatase activity	43	112	gyrB, LA_0006, LA_0025, LA_0273, LA_0274, fusA, ftsZ, uvrB, tufB, ruvB, obg, rho, cysA, LA_1295, bipA, LA_1438, lepA, ttg2A, fliM, LA_2114, recA, LA_2188, uvrA, ychF, hslU, yhbG, mfd, fliH, fliF, fliI, LA_2605, prfC, radA, sufC, lon, LA_3630, hflX, LA_3684, LA_3694, hflB, LA_4036, LA_4193, LA_4194
GO:1901605	Alpha-amino acid metabolic process	41	107	gcvP, gcvH, gcvT, lysC, carB, dapL, pyrB, proA, lysA, hisG, LA_1293, glnA, glyA, LA_1719, panD, folD, met2, leuC, LA_2130, LA_2152, pyrG, hisF, ilvH, abgB, KamA, argF, sseA, LA_3052, trpB, dapA, CsdB, thrC, argG, ilvC, mtnP, serS, ilvE, metH, argB, gltB, asd
GO:1901607	Alpha-amino acid biosynthetic process	33	87	lysC, carB, dapL, pyrB, proA, hisG, LA_1293, glnA, glyA, LA_1719, panD, folD, met2, leuC, LA_2130, LA_2152, hisF, ilvH, abgB, argF, LA_3052, trpB, dapA, thrC, argG, ilvC, mtnP, serS, ilvE, metH, argB, gltB, asd
GO:0016887	ATPase activity	31	82	gyrB, LA_0006, LA_0273, LA_0274, uvrB, ruvB, rho, cysA, LA_1295, LA_1438, ttg2A, LA_2114, recA, LA_2188, uvrA, ychF, hslU, yhbG, mfd, fliI, LA_2605, radA, sufC, lon, LA_3630, LA_3684, LA_3694, hflB, LA_4036, LA_4193, LA_4194
GO:0034645	Cellular macromolecule biosynthetic process	96	255	gatB, dnaE, fusA, trpS, mltG, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpoA, rpmA, nusA, rpmE, rho, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, rpmF, smpB, ileS, grs1, lexA, manC, rplI, aspS, asnS, lepA, lysU, rfe, rpmB, LA_2173, LA_2174, rpsU, rpoD, tyrS, nrdA, rplS, rpsP, kdsA, def, LA_2507, LA_2603, prfC, rodA, LA_2755, lgt, proS, frr, tsf, rpsB, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, LA_3573, LA_3597, hflX, dnaJ, leuS, valS, LA_3767, LA_3820, LA_3905, metG, pheS, glnS, serS, LB_170, LB_208, efp

GO:0043436	Oxoacid metabolic process	94	250	LA_0106, gcvP, gcvH, gcvT, trpS, gpmI, LA_0457, nadA, lysC, carB, dapL, pyrB, LA_0828, proA, LA_1032, lysA, LA_1222, thrS, hisG, plsX, LA_1293, glnA, ileS, grs1, glyA, ugd, fbaB, aspS, tpiA, pgk, LA_1719, panD, folD, asnS, eno, lysU, met2, leuC, LA_2118, LA_2130, mdh, bioB, LA_2152, tyrS, LA_2309, kdsA, pyrG, LA_2432, LA_2433, hisF, ilvH, IlvB, LA_2627, abgB, KamA, argF, citE, PykF, sseA, LA_3052, LA_3135, LA_3268, trpB, proS, acpS, dapA, ppk, CsdB, LA_3606, LA_3607, leuS, thrC, valS, LA_3803, metG, LA_3936, fadB, LA_4139, pheS, argG, glnS, ilvC, mtnP, acsA, folE, serS, ilvE, metH, pfkA, argB, gltB, acnA, Pyk2, asd
GO:0006631	Fatty acid metabolic process	12	32	LA_0106, LA_0457, LA_0828, LA_1032, plsX, LA_2309, LA_2432, LA_3135, acpS, fadB, LA_4139, acnA
GO:0019752	Carboxylic acid metabolic process	92	248	LA_0106, gcvP, gcvH, gcvT, trpS, gpmI, LA_0457, nadA, lysC, carB, dapL, pyrB, LA_0828, proA, LA_1032, lysA, LA_1222, thrS, hisG, plsX, LA_1293, glnA, ileS, grs1, glyA, ugd, fbaB, aspS, tpiA, pgk, LA_1719, panD, folD, asnS, eno, lysU, met2, leuC, LA_2118, LA_2130, mdh, bioB, LA_2152, tyrS, LA_2309, kdsA, pyrG, LA_2432, LA_2433, hisF, ilvH, IlvB, LA_2627, abgB, KamA, argF, citE, PykF, sseA, LA_3052, LA_3135, trpB, proS, acpS, dapA, CsdB, LA_3606, LA_3607, leuS, thrC, valS, LA_3803, metG, LA_3936, fadB, LA_4139, pheS, argG, glnS, ilvC, mtnP, acsA, folE, serS, ilvE, metH, pfkA, argB, gltB, acnA, Pyk2, asd
GO:0044271	Cellular nitrogen compound biosynthetic process	123	334	gatB, dnaE, fusA, LA_0332, trpS, ribH, nadA, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, pyrB, coaX, rpmA, nusA, rpmE, rho, purA, ribB, LA_1205, thrS, infC, rplT, LA_1260, rpmF, smpB, LA_1323, ileS, grs1, ugd, purE, nadB, manC, rplI, aspS, panD, folD, asnS, lepA, lysU, LA_2118, bioB, rpmB, LA_2207, rpsU, rpoD, tyrS, purH, LA_2318, rplS, rpsP, pyrG, def, LA_2507, fliI, LA_2603, prfC, atpD, atpG, atpA, purB, spoT, tgt, dxs, trpB, proS, frr, pyrH, tsf, rpsB, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, LA_3573, hflX, pyrC, leuS, valS, LA_3767, LA_3820, prsA, metG, LA_3936, gmd, pheS, coaD, ndk, glnS,

				mtnP, acsA, folE, serS, hemL, apt, LB_212, efp, speE
GO:0009059	Macromolecule biosynthetic process	97	264	gatB, dnaE, fusA, trpS, mltG, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpoA, rpmA, nusA, rpmE, rho, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, rpmF, smpB, ileS, grs1, lexA, ugd, manC, rplI, aspS, asnS, lepA, lysU, rfe, rpmB, LA_2173, LA_2174, rpsU, rpoD, tyrS, nrdA, rplS, rpsP, kdsA, def, LA_2507, LA_2603, prfC, rodA, LA_2755, lgt, proS, frr, tsf, rpsB, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, LA_3573, LA_3597, hflX, dnaJ, leuS, valS, LA_3767, LA_3820, LA_3905, metG, pheS, glnS, serS, LB_170, LB_208, efp
GO:0043226	Organelle	65	177	gyrB, LA_0006, LA_0025, LA_0332, FlgB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, LA_1260, rpmF, LA_1295, rplI, LA_1855, acoA, LA_2017, flaB, fliM, rpmB, rpsU, ubiX, rplS, rpsP, flgL, LA_2418, LA_2458, fliF, fliQ, fliP, LA_2631, LA_2667, flgE, rpsB, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, LA_3575, LA_3627, LA_3978, LA_4193, LA_4194, LA_4308, FlgK
GO:0006082	Organic acid metabolic process	95	261	LA_0106, gcvP, gcvH, gcvT, trpS, gpmI, LA_0457, nadA, lysC, carB, dapL, pyrB, LA_0828, proA, LA_1032, lysA, LA_1222, thrS, hisG, plsX, LA_1293, glnA, ileS, grs1, glyA, ugd, fbaB, aspS, tpiA, pgk, LA_1719, panD, folD, asnS, eno, lysU, met2, leuC, LA_2118, LA_2130, mdh, bioB, LA_2152, tyrS, LA_2309, kdsA, pyrG, LA_2432, LA_2433, hisF, ilvH, IlvB, LA_2627, abgB, KamA, argF, citE, PykF, sseA, LA_3052, LA_3135, LA_3268, trpB, proS, acpS, dapA, ppk, CsdB, LA_3606, LA_3607, leuS, thrC, valS, LA_3803, metG, LA_3936, fadB, LA_4139, pheS, argG, glnS, ilvC, mtnP, acsA, folE, serS, ilvE, metH, pfkA, argB, LB_208, gltB, acnA, Pyk2, asd
GO:1901566	Organonitrogen compound biosynthetic process	141	390	gatB, fusA, LA_0332, trpS, ribH, mltG, nadA, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, dapL, pyrB, coaX, rpmA, proA, rpmE, lysA, purA, ribB, LA_1221, thrS, infC, rplT, LA_1260, hisG, rpmF, LA_1293, smpB, glnA, LA_1323, ileS, grs1, glyA, purE,

				nadB, rplI, aspS, LA_1719, panD, folD, asnS, lepA, LA_1911, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, LA_2207, rpsU, tyrS, purH, LA_2318, rplS, rpsP, pyrG, def, hisF, LA_2507, ilvH, fliI, abgB, prfC, rodA, LA_2755, atpD, atpG, atpA, argF, lgt, LA_3052, purB, spoT, tgt, dxs, trpB, proS, frr, pyrH, tsf, rpsB, dapA, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, hflX, pyrC, leuS, thrC, valS, LA_3767, LA_3820, metG, LA_3936, pheS, argG, coaD, ndk, glnS, ilvC, mtnP, acsA, folE, serS, ilvE, hemL, metH, argB, apt, gltB, efp, speE, asd
GO:0055086	Nucleobase-containing small molecule metabolic process	48	133	LA_0332, gpmI, nadA, carB, adk, pyrB, coaX, purA, LA_1222, LA_1323, ugd, purE, nadB, manC, fbaB, tpiA, pgk, folD, eno, guaB, LA_2207, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, tgt, pyrH, LA_3573, pyrC, add, prsA, LA_3936, gmd, coaD, ndk, mtnP, acsA, pfkA, apt, LB_212, GalF, Pyk2
GO:0016462	Pyrophosphatase activity	45	127	gyrB, LA_0006, LA_0025, LA_0273, LA_0274, fusA, ftsZ, uvrB, tufB, ruvB, obg, rho, cysA, LA_1295, bipA, LA_1438, ovp1, lepA, ttg2A, fliM, LA_2114, recA, LA_2188, uvrA, ychF, hslU, yhbG, mfd, fliH, fliF, fliI, LA_2605, prfC, radA, sufC, lon, LA_3630, hflX, LA_3684, LA_3694, hflB, LA_4036, LA_4193, LA_4194, LB_212
GO:0032787	Monocarboxylic acid metabolic process	28	81	LA_0106, gpmI, LA_0457, LA_0828, LA_1032, plsX, fbaB, tpiA, pgk, panD, eno, bioB, LA_2309, LA_2432, LA_2627, KamA, PykF, LA_3135, acpS, LA_3606, LA_3607, LA_3936, fadB, LA_4139, acsA, pfkA, acnA, Pyk2
GO:0032991	Protein-containing complex	47	136	gcvP, gcvH, ribH, uvrB, rplB, rpsS, rplV, rplN, rpsE, rpsD, ruvB, sucD, sucC, LA_1112, cysA, LA_1222, sucA, rpmF, lexA, eno, guaB, ttg2B, LA_2113, uvrA, hslU, hslV, nrdA, gspD, yhbG, LA_2432, LA_2433, LA_2458, LA_2507, ilvH, fliI, atpD, atpG, atpA, citE, rpsB, rpsG, rpsL, rplA, ppk, sufC, LA_3803, LA_3939
GO:0043229	Intracellular organelle	49	142	gyrB, LA_0006, LA_0332, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, LA_1260, rpmF, LA_1295, rplI, acoA, rpmB, rpsU, ubiX, rplS, rpsP, LA_2458, LA_2631, rpsB, rpsI,

				rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, LA_3627, LA_3978, LA_4193, LA_4194
GO:0009123	Nucleoside monophosphate metabolic process	12	35	carB, adk, pyrB, purA, LA_1323, purE, purH, purB, pyrH, pyrC, prsA, apt
GO:0032549	Ribonucleoside binding	15	44	fusA, ftsZ, tufB, obg, purA, ribB, bipA, LA_1483, manC, lepA, ychF, prfC, rpoB, hflX, folE
GO:0016818	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	46	135	gyrB, LA_0006, LA_0025, LA_0273, LA_0274, fusA, ftsZ, uvrB, tufB, ruvB, obg, rho, cysA, LA_1295, bipA, LA_1438, ovp1, lepA, ttg2A, fliM, LA_2114, recA, LA_2188, uvrA, ychF, hslU, yhbG, mfd, fliH, fliF, fliI, LA_2605, prfC, radA, sufC, lon, LA_3630, hflX, LA_3684, LA_3694, hflB, acyP, LA_4036, LA_4193, LA_4194, LB_212
GO:0044267	Cellular protein metabolic process	74	220	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, rplI, aspS, asnS, lepA, lysU, LA_2118, rpmB, rpsU, tyrS, hslV, rplS, rpsP, hprK, def, map, LA_2507, prfC, proS, frr, tsf, rpsB, ccmE, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, lon, hflX, leuS, valS, LA_3767, LA_3820, metG, LA_3978, pheS, glnS, serS, efp
GO:0040011	Locomotion	19	57	LA_0025, FlgB, LA_1521, LA_1855, LA_2017, flaB, fliM, LA_2151, fliS, flgL, LA_2418, LA_2469, fliF, LA_2667, flgE, LA_3575, LA_3576, LA_3577, LA_4308
GO:0000287	Magnesium ion binding	21	63	obg, pnp, sucC, purA, LA_1323, ovp1, eno, lysU, LA_2152, fbp, hprK, IlvB, PykF, dxs, uppS, acpS, rpoC, prsA, pheS, ilvC, Pyk2
GO:0034654	Nucleobase-containing compound biosynthetic process	47	143	dnaE, LA_0332, nadA, carB, adk, rpoA, pyrB, coaX, nusA, rho, purA, LA_1205, LA_1323, ugd, purE, nadB, manC, folD, LA_2207, rpoD, purH, LA_2318, pyrG, fliI, LA_2603, atpD, atpG, atpA, purB, spoT, tgt, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, prsA, LA_3936, gmd, coaD, ndk, mtnP, acsA, serS, apt, LB_212
GO:0015031	Protein transport	14	43	LA_0273, secY, secD, secA, LA_2319, gspF, gspD, fliH, fliI, flhA, fliQ, fliP, secE, ExbD3

GO:0019637	Organophosphate metabolic process	50	157	gpmI, nadA, carB, adk, pyrB, coaX, purA, LA_1112, LA_1222, plsX, LA_1323, pssA, purE, nadB, fbaB, tpiA, pgk, gapA, folD, eno, guaB, LA_2119, purH, mipB, lpxC, LA_2318, rpe, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, dxs, pyrH, pyrC, add, prsA, LA_3936, LA_3939, coaD, ndk, acsA, folE, pfkA, apt, LB_212, Pyk2
GO:0044281	Small molecule metabolic process	146	459	LA_0106, pckA, LA_0319, LA_0332, gcvP, gcvH, gcvT, rpiB, trpS, gpmI, LA_0457, ribH, nadA, lysC, carB, adk, dapL, pyrB, LA_0828, coaX, proA, mgsA, LA_1032, lysA, purA, ribB, LA_1222, thrS, hisG, plsX, LA_1293, glnA, LA_1323, ileS, grs1, glyA, ugd, purE, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, LA_1719, panD, folD, asnS, LA_1911, eno, guaB, lysU, acoB, acoA, met2, leuC, LA_2118, LA_2119, LA_2130, mdh, bioB, LA_2152, LA_2207, fbp, tyrS, purH, LA_2309, LA_2318, kdsA, pyrG, LA_2432, LA_2433, hisF, ilvH, IlvB, fliI, LA_2627, abgB, KamA, thiH, LA_2712, atpD, atpG, atpA, argF, citE, PykF, sseA, icmF, LA_3052, purB, spoT, tgt, LA_3135, LA_3268, dxs, trpB, proS, pyrH, acpS, dapA, LA_3354, ppk, CsdB, LA_3573, LA_3606, LA_3607, LA_3627, pyrC, leuS, thrC, valS, add, LA_3803, prsA, metG, LA_3936, gmd, fadB, LA_4139, pheS, argG, coaD, ndk, glnS, ilvC, mtnP, acsA, folE, serS, ilvE, LB_074, sam1, metH, pfkA, argB, apt, LB_208, LB_212, GalF, Mcm3, gltB, acnA, Pyk2, asd
GO:0005515	Protein binding	19	60	LA_0334, rpoA, LA_1253, scpB, grs1, LA_1521, LA_1692, gspD, atoC, LA_2458, clpX, LA_2631, groL, greA, LA_3498, LA_3576, dnaK, dnaJ, asd
GO:0036094	Small molecule binding	165	524	gyrB, LA_0006, LA_0021, pckA, gatB, LA_0273, LA_0274, fusA, LA_0334, pntB, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, lysC, carB, tufB, adk, dapL, pyrB, ruvB, coaX, obg, proA, nuoF, nuoD, rho, sucD, sucC, purA, LA_1112, ribB, cysA, LA_1221, LA_1223, sucA, thrS, LA_1253, hisG, LA_1293, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1438, ugd, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, gapA, asnS, lepA, sdhA, LA_1911, LA_1930, secA, LA_1977, lysU, ttg2A, LA_2114, LA_2119, ssl2, LA_2130, LA_2152, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, nrdA,

				gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, IlvB, fliI, LA_2605, groS, groL, prfC, LA_2755, mreB, atpD, atpG, atpA, argF, radA, LA_2917, PykF, icmF, LA_3052, LA_3143, dxs, proS, pyrH, LA_3363, LA_3406, rpoB, ppk, glcD, CsdB, sufC, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, thrC, valS, LA_3794, glnK, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, acsA, folE, serS, LA_4347, gidA, hemL, LB_074, metH, pfkA, argB, LB_212, Mcm3, Pyk2, asd, ParA4
GO:0006996	Organelle organization	16	51	gyrB, LA_0006, ruvB, rplT, fliS, mfd, fliH, fliI, flhA, fliQ, fliP, hflX, LA_3978, LA_4193, LA_4194, FlgK
GO:0015833	Peptide transport	16	52	LA_0273, secY, secD, secA, LA_2319, gspF, gspD, fliH, fliI, flhA, fliQ, fliP, secE, LA_4228, LA_4229, ExbD3
GO:0046394	Carboxylic acid biosynthetic process	42	137	nadA, lysC, carB, dapL, pyrB, proA, hisG, plsX, LA_1293, glnA, glyA, ugd, LA_1719, panD, folD, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, kdsA, LA_2432, hisF, ilvH, abgB, argF, LA_3052, trpB, acpS, dapA, thrC, argG, ilvC, mtnP, folE, serS, ilvE, metH, argB, gltB, asd
GO:0000166	Nucleotide binding	145	474	gyrB, LA_0006, LA_0021, pckA, gatB, LA_0273, LA_0274, fusA, LA_0334, pntB, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, lysC, carB, tufB, adk, pyrB, ruvB, coaX, obg, proA, nuoF, nuoD, rho, sucD, sucC, purA, LA_1112, ribB, cysA, LA_1223, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, ugd, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, gapA, asnS, lepA, sdhA, LA_1911, LA_1930, secA, LA_1977, lysU, ttg2A, LA_2114, LA_2119, ssl2, LA_2152, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, nrdA, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groS, groL, prfC, mreB, atpD, atpG, atpA, radA, LA_2917, PykF, LA_3143, proS, pyrH, LA_3363, LA_3406, ppk, glcD, sufC, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, glnK, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, acsA,

				folE, serS, LA_4347, gidA, pfkA, argB, LB_212, Pyk2, asd, ParA4
GO:0043168	Anion binding	142	466	gyrB, LA_0006, LA_0021, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, lysC, carB, tufB, adk, dapL, pyrB, ruvB, coaX, obg, nuoF, rho, sucC, purA, ribB, cysA, LA_1221, LA_1223, sucA, thrS, LA_1253, hisG, LA_1293, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1438, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, asnS, lepA, sdhA, LA_1930, secA, LA_1977, lysU, ttg2A, LA_2114, LA_2119, ssl2, LA_2130, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, IlvB, fliI, LA_2605, groS, groL, prfC, LA_2755, mreB, atpD, atpG, atpA, argF, radA, PykF, LA_3052, LA_3143, dxs, proS, pyrH, LA_3363, ppk, glcD, CsdB, sufC, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, thrC, valS, LA_3794, hflB, prsA, metG, LA_3936, gmd, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, acsA, folE, serS, LA_4347, gidA, hemL, pfkA, argB, Pyk2, ParA4
GO:0032553	Ribonucleotide binding	114	382	gyrB, LA_0006, LA_0021, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, ftsZ, uvrB, lysC, carB, tufB, adk, pyrB, ruvB, coaX, obg, nuoF, rho, sucC, purA, ribB, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, asnS, lepA, secA, lysU, ttg2A, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groS, groL, prfC, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, ppk, sufC, lon, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_3936, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, acsA, folE, serS, LA_4347, pfkA, argB, Pyk2, ParA4
GO:0032555	Purine ribonucleotide binding	111	375	gyrB, LA_0006, LA_0021, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, ftsZ, uvrB, lysC, carB, tufB, adk, pyrB, ruvB, coaX, obg, rho, sucC, purA, ribB, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, LA_1483, trmU, manC, aspS, phoH,

				LA_1692, pgk, asnS, lepA, secA, lysU, ttg2A, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groS, groL, prfC, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, ppk, sufC, lon, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, acsA, folE, serS, LA_4347, pfkA, argB, Pyk2, ParA4
GO:0035639	Purine ribonucleoside triphosphate binding	111	375	gyrB, LA_0006, LA_0021, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, ftsZ, uvrB, lysC, carB, tufB, adk, pyrB, ruvB, coaX, obg, rho, sucC, purA, ribB, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, LA_1483, trmU, manC, aspS, phoH, LA_1692, pgk, asnS, lepA, secA, lysU, ttg2A, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groS, groL, prfC, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, ppk, sufC, lon, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, acsA, folE, serS, LA_4347, pfkA, argB, Pyk2, ParA4
GO:0090407	Organophosphate biosynthetic process	34	115	nadA, carB, adk, pyrB, coaX, purA, LA_1112, plsX, LA_1323, pssA, purE, nadB, folD, purH, lpxC, LA_2318, pyrG, fliI, atpD, atpG, atpA, purB, spoT, dxs, pyrH, pyrC, prsA, LA_3936, LA_3939, coaD, ndk, acsA, folE, apt
GO:0097367	Carbohydrate derivative binding	116	394	gyrB, LA_0006, LA_0021, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, ftsZ, uvrB, lysC, carB, tufB, adk, pyrB, ruvB, coaX, obg, nuoF, rho, sucC, purA, ribB, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, asnS, lepA, secA, lysU, ttg2A, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groS, groL, prfC, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, rpoB, ppk, sufC, lon, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG,

				LA_3936, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, acsA, folE, serS, LA_4347, pfkA, argB, LB_212, Pyk2, ParA4
GO:0005524	ATP binding	98	335	gyrB, LA_0006, LA_0021, pckA, gatB, LA_0273, LA_0274, LA_0391, trpS, uvrB, lysC, carB, adk, pyrB, ruvB, coaX, rho, sucC, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, grs1, rbsK, LA_1438, trmU, aspS, phoH, LA_1692, pgk, asnS, secA, lysU, ttg2A, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groS, groL, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, ppk, sufC, lon, LA_3630, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, acsA, serS, LA_4347, pfkA, argB, Pyk2, ParA4
GO:0016053	Organic acid biosynthetic process	43	147	nadA, lysC, carB, dapL, pyrB, proA, hisG, plsX, LA_1293, glnA, glyA, ugd, LA_1719, panD, folD, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, kdsA, LA_2432, hisF, ilvH, abgB, argF, LA_3052, trpB, acpS, dapA, thrC, argG, ilvC, mtnP, folE, serS, ilvE, metH, argB, LB_208, gltB, asd
GO:0010468	Regulation of gene expression	37	128	fusA, LA_0395, LA_0653, LA_0720, tufB, LA_0839, nusA, pnp, LA_1189, LA_1205, infC, LA_1253, lexA, hfq, LA_1692, lepA, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3070, LA_3094, LA_3096, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, efp, LB_333
GO:1901576	Organic substance biosynthetic process	187	648	pckA, gatB, dnaE, fusA, LA_0319, LA_0332, trpS, ribH, mltG, nadA, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, coaX, rpmA, proA, mgsA, nusA, rpmE, rho, lysA, purA, LA_1112, ribB, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, LA_1293, smpB, glnA, LA_1323, ileS, grs1, glyA, lexA, pssA, ugd, purE, nadB, manC, rplI, aspS, tpiA, LA_1719, panD, folD, asnS, lepA, LA_1911, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, LA_2172, LA_2173, LA_2174, LA_2207, fbp, rpsU, rpoD, tyrS, purH, lpxC, LA_2318, LA_2326, nrdA, rplS, rpsP, kdsA, pyrG, LA_2432, def, hisF, LA_2507, ilvH, fliI, LA_2603, abgB, thiH, prfC, rodA, LA_2755,

				atpD, atpG, atpA, LA_2821, argF, lgt, LA_3052, purB, spoT, tgt, dxs, trpB, proS, uppS, frr, pyrH, tsf, rpsB, acpS, dapA, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, ppk, LA_3573, LA_3579, LA_3597, hflX, pyrC, dnaJ, leuS, thrC, valS, LA_3767, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, pheS, argG, coaD, ndk, glnS, ilvC, mtnP, acsA, folE, serS, ilvE, hemL, metH, argB, LB_170, apt, LB_208, LB_212, gltB, efp, speE, asd
GO:0044249	Cellular biosynthetic process	184	638	gatB, dnaE, fusA, LA_0319, LA_0332, trpS, ribH, mltG, nadA, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, coaX, rpmA, proA, mgsA, nusA, rpmE, rho, lysA, purA, LA_1112, ribB, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, LA_1293, smpB, glnA, LA_1323, ileS, grs1, glyA, lexA, pssA, ugd, purE, nadB, manC, rplI, aspS, LA_1719, panD, folD, asnS, lepA, LA_1911, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, LA_2172, LA_2173, LA_2174, LA_2207, rpsU, rpoD, tyrS, purH, lpxC, LA_2318, LA_2326, nrdA, rplS, rpsP, kdsA, pyrG, LA_2432, def, hisF, LA_2507, ilvH, fliI, LA_2603, abgB, thiH, prfC, rodA, LA_2755, atpD, atpG, atpA, LA_2821, argF, lgt, LA_3052, purB, spoT, tgt, dxs, trpB, proS, uppS, frr, pyrH, tsf, rpsB, acpS, dapA, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, ppk, LA_3573, LA_3579, LA_3597, hflX, pyrC, dnaJ, leuS, thrC, valS, LA_3767, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, pheS, argG, coaD, ndk, glnS, ilvC, mtnP, acsA, folE, serS, ilvE, hemL, metH, argB, LB_170, apt, LB_208, LB_212, gltB, efp, speE, asd
GO:2000112	Regulation of cellular macromolecule biosynthetic process	36	126	fusA, LA_0395, LA_0653, LA_0720, tufB, LA_0839, nusA, LA_1189, LA_1205, infC, LA_1253, lexA, hfq, LA_1692, lepA, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3070, LA_3094, LA_3096, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, efp, LB_333
GO:0005622	Intracellular	289	1017	gyrB, LA_0006, LA_0020, LA_0021, pckA, dnaE, LA_0280, LA_0309, fusA, LA_0332, gcvP, gcvH, gcvT, LA_0391, trpS, etfB, gpmI, LA_0457, ftsZ, nadA, uvrB, LA_0671, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE,

				rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, rpmA, obg, proA, mgsA, nusA, pnp, LA_1019, rpmE, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, ribB, LA_1189, LA_1222, LA_1223, sucA, thrS, infC, rplT, LA_1253, scpB, LA_1260, hisG, rpmF, plsX, LA_1293, LA_1295, smpB, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, ugd, purE, trmU, nadB, hfq, LA_1521, fbaB, rplI, aspS, phoH, LA_1692, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, asnS, LA_1911, LA_1929, eno, LA_1953, secA, lysU, acoA, met2, leuC, LA_2118, LA_2130, mdh, bioB, LA_2151, LA_2152, rpmB, LA_2173, recA, LA_2188, LA_2202, uvrA, fbp, rpsU, rpoD, tyrS, fliS, purH, mipB, LA_2297, lpxC, LA_2324, ychF, ubiX, hslU, hslV, xerC, LA_2350, nrdA, rplS, trmD, rpsP, rpe, atoC, hprK, kdsA, pyrG, LA_2432, LA_2433, def, map, LA_2458, LA_2483, trxB, hisF, LA_2507, mfd, LA_2559, ilvH, IlvB, fliH, fliI, LA_2603, LA_2605, LA_2631, groS, groL, abgB, bfr, prfC, mreB, atpD, atpG, atpA, sufB, ahpC, LA_2821, argF, citE, LA_2894, radA, LA_2917, PykF, eutG, LA_3052, purB, LA_3094, tgt, LA_3135, caiB, dxs, trpB, proS, uppS, frr, pyrH, tsf, rpsB, acpS, dapA, LA_3354, LA_3406, rpsI, rplM, rpsG, rpsL, rpoB, rplL, rplJ, rplA, rplK, nusG, Bcp, LA_3484, LA_3498, CsdB, sufC, LA_3573, lon, dps, LA_3627, hflX, pyrC, LA_3684, dnaK, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3803, glnK, LA_3820, prsA, metG, LA_3939, gmd, LA_3978, LA_4034, LA_4036, LA_4139, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, mtnP, acsA, folE, LA_4327, serS, tatD, LA_4347, gidA, ilvE, hemL, ParB3, LB_082, sam1, metH, pfkA, argB, apt, LB_212, GalF, gltB, efp, speE, acnA, LB_333, Pyk2, asd, ParA4
GO:0140098	Catalytic activity, acting on RNA	25	89	gatB, trpS, rpoA, rlmN, thrS, ileS, grs1, aspS, asnS, lysU, tyrS, trmD, LA_2507, LA_2917, tgt, proS, rpoC, rpoB, leuS, valS, LA_3820, metG, pheS, glnS, serS
GO:0019538	Protein metabolic process	83	297	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, LA_1221, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, lexA, rplI, aspS, asnS, lepA, LA_1953, lysU, LA_2118, rpmB, rpsU, tyrS, hslU, hslV, rplS, rpsP,

				hprK, def, map, LA_2507, LA_2559, LA_2631, prfC, lgt, LA_3054, proS, frr, tsf, rpsB, ccmE, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, lon, hflX, leuS, valS, LA_3767, hflB, LA_3820, metG, LA_3978, pheS, glnS, serS, efp
GO:1901564	Organonitrogen compound metabolic process	176	636	gatB, fusA, LA_0332, gcvP, gcvH, gcvT, trpS, gpmI, ribH, mltG, nadA, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, dapL, pyrB, coaX, rpmA, proA, rpmE, lysA, purA, ribB, LA_1221, LA_1222, thrS, infC, rplT, LA_1260, hisG, rpmF, LA_1293, smpB, glnA, LA_1323, ileS, grs1, glyA, lexA, purE, nadB, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1719, panD, folD, asnS, lepA, LA_1911, eno, LA_1953, guaB, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, LA_2207, rpsU, tyrS, purH, LA_2318, hslU, hslV, rplS, rpsP, hprK, pyrG, def, map, hisF, LA_2507, LA_2559, ilvH, fliI, LA_2631, abgB, KamA, prfC, rodA, LA_2755, atpD, atpG, atpA, argF, PykF, sseA, icmF, lgt, LA_3052, LA_3054, purB, spoT, tgt, dxs, trpB, proS, frr, pyrH, tsf, rpsB, dapA, ccmE, rpsI, rplM, rpsG, rpsL, rplL, rplJ, rplA, rplK, CsdB, lon, hflX, pyrC, leuS, thrC, valS, LA_3767, hflB, LA_3820, prsA, metG, LA_3936, LA_3978, pheS, argG, coaD, ndk, glnS, ilvC, mtnP, acsA, folE, serS, ilvE, hemL, LB_074, metH, pfkA, argB, apt, LB_212, Mcm3, gltB, efp, speE, Pyk2, asd
GO:0080090	Regulation of primary metabolic process	38	138	fusA, LA_0395, LA_0653, LA_0720, tufB, LA_0839, nusA, LA_1189, LA_1205, infC, LA_1253, lexA, hfq, LA_1692, lepA, LA_2072, rpoD, LA_2325, atoC, hprK, mfd, LA_2603, prfC, LA_3070, LA_3094, LA_3096, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, efp, LB_333
GO:0071705	Nitrogen compound transport	19	69	LA_0273, secY, secD, secA, LA_2319, gspF, gspD, fliH, fliI, flhA, fliQ, fliP, ccmF, secE, LA_3794, LA_3806, LA_4228, LA_4229, ExbD3
GO:0005829	Cytosol	53	194	gcvP, gcvT, ruvB, LA_0845, mgsA, sucD, sucC, ribB, sucA, purE, hfq, LA_1713, folD, eno, leuC, LA_2173, LA_2202, fliS, purH, hslU, hslV, LA_2350, nrdA, rpe, hprK, pyrG, def, ilvH, LA_2605, bfr, sufB, radA, PykF, eutG, purB, tgt, dxs, uppS, LA_3406, rpoB,

				nusG, sufC, add, glnK, LA_4034, ilvC, ilvE, metH, LB_212, gltB, acnA, Pyk2, asd
GO:0005737	Cytoplasm	243	898	gyrB, LA_0006, LA_0020, LA_0021, pckA, dnaE, LA_0280, LA_0309, fusA, LA_0332, gcvP, gcvH, gcvT, LA_0391, trpS, etfB, gpmI, LA_0457, ftsZ, nadA, uvrB, LA_0671, lysC, LA_0720, carB, tufB, adk, rpoA, dapL, pyrB, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, obg, proA, mgsA, nusA, pnp, LA_1019, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, ribB, LA_1189, LA_1222, LA_1223, sucA, thrS, infC, LA_1253, scpB, hisG, plsX, LA_1293, smpB, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, ugd, purE, trmU, nadB, hfq, LA_1521, fbaB, aspS, phoH, LA_1692, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, asnS, LA_1911, LA_1929, eno, LA_1953, secA, lysU, met2, leuC, LA_2118, LA_2130, mdh, bioB, LA_2151, LA_2152, LA_2173, recA, LA_2188, LA_2202, uvrA, fbp, rpoD, tyrS, fliS, purH, mipB, LA_2297, lpxC, LA_2324, ychF, ubiX, hslU, hslV, xerC, LA_2350, nrdA, trmD, rpe, atoC, hprK, kdsA, pyrG, LA_2432, LA_2433, def, map, LA_2483, trxB, hisF, mfd, LA_2559, ilvH, IlvB, fliH, fliI, LA_2603, LA_2605, LA_2631, groS, groL, abgB, bfr, prfC, mreB, sufB, ahpC, LA_2821, argF, citE, LA_2894, radA, LA_2917, PykF, eutG, LA_3052, purB, LA_3094, tgt, LA_3135, caiB, dxs, trpB, proS, uppS, frr, pyrH, tsf, acpS, dapA, LA_3354, LA_3406, rpoB, nusG, Bcp, LA_3484, LA_3498, CsdB, sufC, LA_3573, lon, dps, LA_3627, hflX, pyrC, LA_3684, dnaK, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3803, glnK, LA_3820, prsA, metG, LA_3939, gmd, LA_4034, LA_4036, LA_4139, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, mtnP, acsA, folE, LA_4327, serS, tatD, LA_4347, gidA, ilvE, hemL, LB_082, sam1, metH, pfkA, argB, apt, LB_212, GalF, gltB, efp, speE, acnA, LB_333, Pyk2, asd, ParA4
GO:0051171	Regulation of nitrogen compound metabolic process	37	137	fusA, LA_0395, LA_0653, LA_0720, tufB, LA_0839, nusA, LA_1189, LA_1205, infC, LA_1253, lexA, hfq, LA_1692, lepA, LA_2072, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3070, LA_3094, LA_3096, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, efp, LB_333

GO:1901575	Organic substance catabolic process	32	119	gcvP, gcvH, gcvT, rpiB, gpmI, LA_0457, LA_0828, pnp, LA_1032, LA_1112, LA_1222, fbaB, tpiA, pgk, eno, acoB, acoA, LA_2119, hslV, LA_2631, KamA, PykF, LA_3135, lon, hflB, LA_3936, LA_3939, fadB, LA_4139, pfkA, LB_212, Pyk2
GO:0016070	RNA metabolic process	34	127	LA_0021, trpS, rpoA, rlmN, nusA, pnp, rho, LA_1105, rimO, LA_1205, thrS, ileS, grs1, trmU, aspS, asnS, lysU, rpoD, tyrS, trmD, LA_2603, LA_2917, tgt, proS, rpoC, rpoB, nusG, leuS, valS, metG, pheS, glnS, serS, gidA
GO:0060255	Regulation of macromolecule metabolic process	38	142	fusA, LA_0395, LA_0653, LA_0720, tufB, LA_0839, nusA, pnp, LA_1189, LA_1205, infC, LA_1253, lexA, hfq, LA_1692, lepA, LA_2072, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3070, LA_3094, LA_3096, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, efp, LB_333
GO:1901135	Carbohydrate derivative metabolic process	62	232	LA_0332, gpmI, mltG, carB, adk, pyrB, coaX, purA, LA_1112, LA_1221, LA_1222, LA_1323, ugd, purE, manC, fbaB, tpiA, pgk, eno, rfe, LA_2119, LA_2173, LA_2174, LA_2207, purH, mipB, lpxC, LA_2318, LA_2326, rpe, kdsA, pyrG, fliI, rodA, LA_2755, atpD, atpG, atpA, PykF, purB, spoT, tgt, dxs, pyrH, LA_3573, LA_3597, pyrC, prsA, LA_3905, LA_3936, LA_3939, gmd, coaD, ndk, mtnP, acsA, pfkA, LB_170, apt, LB_212, GalF, Pyk2
GO:1901137	Carbohydrate derivative biosynthetic process	48	180	LA_0332, mltG, carB, adk, pyrB, coaX, purA, LA_1112, LA_1221, LA_1323, ugd, purE, manC, rfe, LA_2173, LA_2174, LA_2207, purH, lpxC, LA_2318, LA_2326, kdsA, pyrG, fliI, rodA, LA_2755, atpD, atpG, atpA, purB, spoT, tgt, dxs, pyrH, LA_3573, LA_3597, pyrC, prsA, LA_3905, LA_3936, LA_3939, gmd, coaD, ndk, mtnP, acsA, LB_170, apt
GO:0044283	Small molecule biosynthetic process	60	226	pckA, LA_0319, LA_0332, ribH, nadA, lysC, carB, dapL, pyrB, proA, mgsA, ribB, hisG, plsX, LA_1293, glnA, glyA, ugd, tpiA, LA_1719, panD, folD, LA_1911, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, LA_2207, fbp, kdsA, LA_2432, hisF, ilvH, abgB, thiH, argF, LA_3052, spoT, tgt, dxs, trpB, acpS, dapA, thrC, argG, ndk, ilvC,

				mtnP, folE, serS, ilvE, metH, argB, apt, LB_208, LB_212, gltB, asd
GO:0055114	Oxidation-reduction process	72	274	LA_0020, cyoB, LA_0334, pntB, gcvP, etfA, etfB, ivd, LA_0457, LA_0560, LA_0666, LA_0671, LA_0790, LA_0828, proA, nuoK, nuoH, nuoF, nuoD, nuoA, LA_1032, sucD, sucC, LA_1112, LA_1223, sucA, ugd, lldD, nadB, gapA, folD, LA_1896, sdhA, LA_1911, LA_1930, LA_1977, guaB, acoB, acoA, mdh, LA_2152, mipB, LA_2324, nrdA, rpe, trxB, bfr, KamA, ahpC, eutG, LA_3135, LA_3143, LA_3265, LA_3266, LA_3268, LA_3354, LA_3363, Bcp, glcD, LA_3484, dps, LA_3627, mviM, LA_3939, fadB, LA_4139, ilvC, LA_4327, LB_082, gltB, acnA, asd
GO:0097159	Organic cyclic compound binding	225	857	gyrB, LA_0006, LA_0021, cyoB, pckA, gatB, LA_0273, LA_0274, LA_0280, LA_0309, fusA, LA_0334, pntB, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, LA_0666, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, ruvB, coaX, rlmN, obg, proA, nuoF, nuoD, nusA, pnp, rho, sucD, sucC, purA, LA_1112, LA_1137, ribB, cysA, LA_1205, LA_1221, LA_1223, sucA, thrS, infC, rplT, LA_1253, LA_1260, hisG, LA_1293, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1438, lexA, ugd, LA_1483, trmU, lldD, hfq, manC, rplI, aspS, phoH, LA_1692, pgk, gapA, asnS, lepA, sdhA, LA_1911, LA_1930, secA, LA_1977, lysU, ttg2A, LA_2114, LA_2119, ssl2, LA_2130, LA_2152, recA, LA_2188, uvrA, rpoD, tyrS, LA_2318, LA_2325, ychF, mrp, hslU, xerC, nrdA, gspE, atoC, hprK, yhbG, pyrG, LA_2458, LA_2483, LA_2507, mfd, clpX, IlvB, fliI, LA_2603, LA_2605, groS, groL, prfC, LA_2755, mreB, atpD, atpG, atpA, radA, LA_2917, PykF, icmF, LA_3052, LA_3094, LA_3143, LA_3265, LA_3268, dxs, proS, pyrH, tsf, ccmE, ccmF, LA_3363, greA, LA_3406, rpsG, rpsL, rpoC, rpoB, rplJ, rplA, rplK, nusG, ppk, glcD, CsdB, sufC, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, hrcA, dnaK, dnaJ, leuS, thrC, valS, LA_3767, glnK, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, acsA, folE, serS, LA_4347, gidA, hemL, ParB3, LB_074, metH, pfkA, argB,

				LB_212, Mcm3, efp, acnA, LB_333, Pyk2, asd, ParA4
GO:1901363	Heterocyclic compound binding	225	857	gyrB, LA_0006, LA_0021, cyoB, pckA, gatB, LA_0273, LA_0274, LA_0280, LA_0309, fusA, LA_0334, pntB, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, LA_0666, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, ruvB, coaX, rlmN, obg, proA, nuoF, nuoD, nusA, pnp, rho, sucD, sucC, purA, LA_1112, LA_1137, ribB, cysA, LA_1205, LA_1221, LA_1223, sucA, thrS, infC, rplT, LA_1253, LA_1260, hisG, LA_1293, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1438, lexA, ugd, LA_1483, trmU, lldD, hfq, manC, rplI, aspS, phoH, LA_1692, pgk, gapA, asnS, lepA, sdhA, LA_1911, LA_1930, secA, LA_1977, lysU, ttg2A, LA_2114, LA_2119, ssl2, LA_2130, LA_2152, recA, LA_2188, uvrA, rpoD, tyrS, LA_2318, LA_2325, ychF, mrp, hslU, xerC, nrdA, gspE, atoC, hprK, yhbG, pyrG, LA_2458, LA_2483, LA_2507, mfd, clpX, IlvB, fliI, LA_2603, LA_2605, groS, groL, prfC, LA_2755, mreB, atpD, atpG, atpA, radA, LA_2917, PykF, icmF, LA_3052, LA_3094, LA_3143, LA_3265, LA_3268, dxs, proS, pyrH, tsf, ccmE, ccmF, LA_3363, greA, LA_3406, rpsG, rpsL, rpoC, rpoB, rplJ, rplA, rplK, nusG, ppk, glcD, CsdB, sufC, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, hrcA, dnaK, dnaJ, leuS, thrC, valS, LA_3767, glnK, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, acsA, folE, serS, LA_4347, gidA, hemL, ParB3, LB_074, metH, pfkA, argB, LB_212, Mcm3, efp, acnA, LB_333, Pyk2, asd, ParA4
GO:0019222	Regulation of metabolic process	40	153	fusA, LA_0395, LA_0653, LA_0720, tufB, LA_0839, nusA, pnp, LA_1189, LA_1205, infC, LA_1253, lexA, hfq, LA_1692, lepA, LA_2072, rpoD, LA_2325, atoC, hprK, mfd, LA_2603, prfC, LA_3070, LA_3094, LA_3096, tsf, greA, rplA, nusG, LA_3498, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, efp, LB_333
GO:0034641	Cellular nitrogen compound	165	634	gyrB, LA_0006, LA_0021, LA_0250, gatB, dnaE, fusA, LA_0332, trpS, gpmI, ribH, nadA, uvrB, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE,

	metabolic process			rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, pyrB, ruvB, coaX, rlmN, rpmA, nusA, pnp, rpmE, rho, LA_1105, purA, rimO, ribB, LA_1205, LA_1222, thrS, infC, rplT, LA_1260, hisG, rpmF, smpB, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, panD, folD, asnS, lepA, LA_1935, eno, guaB, lysU, LA_2118, bioB, rpmB, recA, LA_2207, uvrA, rpsU, rpoD, tyrS, purH, LA_2318, xerC, rplS, trmD, rpsP, pyrG, def, LA_2483, hisF, LA_2507, mfd, fliI, LA_2603, prfC, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, dxs, trpB, proS, frr, pyrH, tsf, rpsB, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, LA_3573, hflX, pyrC, LA_3684, leuS, valS, LA_3767, add, LA_3820, prsA, metG, LA_3936, gmd, pheS, coaD, ndk, glnS, LA_4193, LA_4194, mtnP, acsA, folE, serS, tatD, gidA, hemL, pfkA, apt, LB_212, GalF, efp, speE, Pyk2
GO:0031323	Regulation of cellular metabolic process	37	143	fusA, LA_0395, LA_0653, LA_0720, tufB, LA_0839, nusA, LA_1189, LA_1205, infC, LA_1253, lexA, hfq, LA_1692, lepA, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3070, LA_3094, LA_3096, tsf, greA, rplA, nusG, LA_3498, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, efp, LB_333
GO:0016829	Lyase activity	25	99	pckA, LA_0332, mltG, mgsA, LA_1032, lysA, fbaB, panD, eno, leuC, LA_2318, ubiX, hisF, LA_2627, citE, purB, trpB, dapA, LA_3606, LA_3607, thrC, LA_3936, gmd, fadB, acnA
GO:0043167	Ion binding	202	806	gyrB, LA_0006, LA_0021, cyoB, LA_0250, pckA, gatB, LA_0273, LA_0274, fusA, LA_0332, LA_0391, trpS, etfA, ivd, gpmI, LA_0560, ftsZ, nadA, uvrB, LA_0666, lysC, carB, tufB, adk, dapL, pyrB, ruvB, coaX, rlmN, obg, nuoF, pnp, LA_1019, rho, sucC, LA_1105, purA, rimO, ribB, cysA, LA_1221, LA_1223, sucA, thrS, LA_1253, hisG, LA_1293, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, LA_1438, ovp1, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, asnS, lepA, LA_1896, sdhA, LA_1930, LA_1948, eno, secA, LA_1977, lysU, ttg2A, leuC, LA_2114, LA_2118, LA_2119, ssl2, LA_2130, bioB, LA_2151, LA_2152, recA, LA_2188, uvrA, fbp, tyrS, LA_2297, lpxC, LA_2318, ychF, mrp, hslU, hslV, nrdA, gspF, gspE, rpe, atoC, hprK, yhbG, pyrG,

				def, map, trxB, LA_2507, mfd, clpX, IlvB, fliI, LA_2605, groS, groL, bfr, KamA, thiH, LA_2712, prfC, LA_2755, mreB, atpD, atpG, atpA, argF, citE, radA, PykF, icmF, eutG, LA_3052, LA_3094, tgt, LA_3143, LA_3265, LA_3266, LA_3268, dxs, proS, uppS, pyrH, acpS, ccmE, LA_3354, LA_3363, rpoC, ppk, glcD, CsdB, sufC, lon, dps, LA_3627, LA_3630, hflX, pyrC, LA_3684, LA_3694, dnaK, dnaJ, leuS, thrC, valS, add, LA_3794, hflB, prsA, metG, LA_3936, gmd, LA_4036, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, acsA, folE, LA_4327, serS, tatD, LA_4347, gidA, hemL, LB_074, metH, pfkA, argB, LB_212, Mcm3, gltB, acnA, Pyk2, ParA4
GO:0018130	Heterocycle biosynthetic process	58	232	dnaE, LA_0332, ribH, nadA, carB, adk, rpoA, pyrB, coaX, proA, nusA, rho, purA, ribB, LA_1205, hisG, LA_1323, ugd, purE, nadB, manC, folD, LA_2118, bioB, LA_2207, rpoD, purH, LA_2318, pyrG, hisF, fliI, LA_2603, atpD, atpG, atpA, purB, spoT, tgt, dxs, trpB, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, prsA, LA_3936, gmd, coaD, ndk, mtnP, acsA, folE, serS, hemL, apt, LB_212
GO:0019438	Aromatic compound biosynthetic process	53	219	dnaE, LA_0332, nadA, carB, adk, rpoA, pyrB, coaX, nusA, rho, purA, LA_1205, hisG, LA_1323, ugd, purE, nadB, manC, folD, LA_2207, rpoD, purH, LA_2318, pyrG, hisF, fliI, LA_2603, atpD, atpG, atpA, purB, spoT, tgt, dxs, trpB, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, prsA, LA_3936, gmd, coaD, ndk, mtnP, acsA, folE, serS, hemL, apt, LB_212
GO:1901362	Organic cyclic compound biosynthetic process	58	240	dnaE, LA_0332, ribH, nadA, carB, adk, rpoA, pyrB, coaX, proA, nusA, rho, purA, ribB, LA_1205, hisG, LA_1323, ugd, purE, nadB, manC, folD, LA_2118, bioB, LA_2207, rpoD, purH, LA_2318, pyrG, hisF, fliI, LA_2603, atpD, atpG, atpA, purB, spoT, tgt, dxs, trpB, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, prsA, LA_3936, gmd, coaD, ndk, mtnP, acsA, folE, serS, hemL, apt, LB_212
GO:0016491	Oxidoreductase activity	59	247	LA_0020, cyoB, LA_0334, pntB, gcvP, etfA, etfB, ivd, LA_0560, LA_0666, proA, nuoK, nuoH, nuoF, nuoD, nuoA, LA_1112, LA_1223, sucA, ugd, lldD, nadB, gapA, folD, LA_1896, sdhA, LA_1911, LA_1930, LA_1977, guaB, acoB, acoA, mdh, LA_2152, LA_2324, nrdA, trxB, bfr, ahpC,

				eutG, LA_3143, LA_3265, LA_3266, LA_3268, LA_3354, LA_3363, Bcp, glcD, LA_3484, dps, LA_3627, mviM, LA_3939, fadB, ilvC, LA_4327, LB_082, gltB, asd
GO:0006355	Regulation of transcription, DNA-templated	25	105	LA_0395, LA_0653, LA_0720, LA_0839, nusA, LA_1205, LA_1253, lexA, hfq, LA_1692, rpoD, LA_2325, atoC, mfd, LA_2603, LA_3070, LA_3094, LA_3096, greA, nusG, LA_3572, hrcA, LA_3911, LA_4036, LB_333
GO:0005488	Binding	283	1214	gyrB, LA_0006, LA_0021, cyoB, LA_0250, pckA, gatB, LA_0273, LA_0274, LA_0280, LA_0309, fusA, LA_0332, LA_0334, pntB, LA_0391, trpS, etfA, ivd, gpmI, LA_0560, ftsZ, nadA, uvrB, LA_0666, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, secY, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, ruvB, coaX, rlmN, obg, proA, nuoK, nuoH, nuoF, nuoD, nuoA, nusA, pnp, LA_1019, rho, sucD, sucC, LA_1105, purA, LA_1112, LA_1137, rimO, ribB, cysA, LA_1189, LA_1205, LA_1221, LA_1223, sucA, thrS, infC, rplT, LA_1253, scpB, LA_1260, hisG, LA_1293, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, LA_1438, lexA, ugd, ovp1, LA_1483, trmU, lldD, hfq, manC, LA_1521, rplI, aspS, phoH, LA_1692, pgk, gapA, asnS, lepA, LA_1896, sdhA, LA_1911, LA_1929, LA_1930, LA_1948, eno, secA, LA_1977, lysU, ttg2A, leuC, LA_2114, LA_2118, LA_2119, ssl2, LA_2130, bioB, LA_2151, LA_2152, recA, LA_2188, uvrA, fbp, rpoD, tyrS, LA_2297, lpxC, LA_2318, LA_2325, ychF, mrp, hslU, hslV, xerC, nrdA, gspF, gspE, gspD, rpe, atoC, hprK, yhbG, pyrG, def, map, LA_2458, LA_2483, trxB, LA_2507, mfd, clpX, IlvB, fliI, LA_2603, LA_2605, LA_2631, groS, groL, bfr, KamA, thiH, LA_2712, prfC, LA_2755, mreB, atpD, atpG, atpA, argF, citE, radA, LA_2917, PykF, icmF, eutG, LA_3052, LA_3094, tgt, LA_3143, LA_3265, LA_3266, LA_3268, dxs, proS, uppS, frr, pyrH, tsf, acpS, ccmE, ccmF, LA_3354, LA_3363, greA, LA_3406, rpsG, rpsL, rpoC, rpoB, rplJ, rplA, rplK, nusG, ppk, glcD, LA_3498, CsdB, sufC, LA_3576, lon, dps, LA_3627, LA_3630, hflX, pyrC, LA_3684, LA_3694, hrcA, dnaK, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3794, glnK, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4036, pheS, argG, coaD, ndk,

				glnS, LA_4193, LA_4194, ilvC, acsA, folE, LA_4327, serS, tatD, LA_4347, gidA, hemL, ParB3, LB_074, metH, pfkA, argB, LB_212, Mcm3, gltB, efp, acnA, LB_333, Pyk2, asd, ParA4
GO:0016043	Cellular component organization	31	134	gyrB, LA_0006, mltG, ftsZ, ruvB, LA_1019, LA_1221, rplT, LA_1295, rfe, fliS, mfd, fliH, fliI, flhA, fliQ, fliP, prfC, rodA, LA_2755, sufB, frr, ccmE, ccmF, sufC, hflX, LA_3767, LA_3978, LA_4193, LA_4194, FlgK
GO:0003676	Nucleic acid binding	89	388	gyrB, LA_0006, LA_0021, fusA, uvrB, LA_0720, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpoA, ruvB, rlmN, nusA, pnp, rho, LA_1137, LA_1205, thrS, infC, rplT, LA_1253, LA_1260, smpB, smc, ileS, bipA, lexA, trmU, hfq, rplI, aspS, LA_1692, asnS, lepA, lysU, ssl2, recA, uvrA, rpoD, tyrS, LA_2325, xerC, atoC, LA_2458, LA_2483, mfd, LA_2603, prfC, radA, LA_2917, LA_3094, tsf, greA, rpsG, rpsL, rpoC, rpoB, rplJ, rplA, rplK, nusG, lon, hrcA, LA_3767, LA_4036, pheS, glnS, LA_4193, LA_4194, ParB3, efp, acnA, LB_333
GO:0043169	Cation binding	98	428	gyrB, cyoB, LA_0250, pckA, LA_0332, gpmI, nadA, LA_0666, carB, pyrB, coaX, rlmN, obg, nuoF, pnp, LA_1019, sucC, LA_1105, purA, rimO, ribB, sucA, thrS, glnA, LA_1323, ileS, LA_1417, ovp1, LA_1483, LA_1896, eno, lysU, leuC, LA_2118, bioB, LA_2151, LA_2152, uvrA, fbp, LA_2297, lpxC, ychF, hslV, gspF, rpe, hprK, pyrG, def, map, clpX, IlvB, bfr, KamA, thiH, LA_2712, citE, radA, PykF, icmF, eutG, LA_3094, tgt, LA_3265, LA_3266, LA_3268, dxs, uppS, acpS, ccmE, LA_3354, rpoC, ppk, dps, hflX, pyrC, dnaJ, add, LA_3794, hflB, prsA, metG, LA_3936, pheS, ndk, glnS, LA_4194, ilvC, acsA, folE, tatD, LB_074, metH, pfkA, LB_212, Mcm3, gltB, acnA, Pyk2
GO:0044237	Cellular metabolic process	293	1297	gyrB, LA_0006, LA_0021, LA_0106, cyoB, LA_0250, gatB, dnaE, fusA, LA_0319, LA_0332, gcvP, gcvH, gcvT, rpiB, trpS, etfA, etfB, gpmI, LA_0457, ribH, mltG, nadA, uvrB, LA_0666, LA_0671, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, LA_0790, ruvB, LA_0828, coaX,

				rlmN, LA_0845, rpmA, proA, nuoK, mgsA, nusA, pnp, LA_1019, rpmE, rho, LA_1032, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, ribB, LA_1205, LA_1221, LA_1222, sucA, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, LA_1293, smpB, glnA, LA_1323, ileS, grs1, rbsK, glyA, lexA, pssA, ugd, purE, trmU, lldD, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, asnS, lepA, LA_1896, LA_1911, LA_1935, eno, guaB, lysU, acoB, acoA, rfe, met2, leuC, LA_2118, LA_2119, LA_2130, mdh, bioB, LA_2152, rpmB, LA_2172, LA_2173, LA_2174, recA, LA_2202, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, purH, mipB, lpxC, LA_2309, LA_2318, LA_2326, ubiX, mrp, hslV, xerC, LA_2350, nrdA, rplS, trmD, rpsP, rpe, hprK, kdsA, pyrG, LA_2432, LA_2433, def, map, LA_2483, hisF, LA_2507, mfd, ilvH, IlvB, LA_2572, fliI, LA_2603, LA_2627, abgB, KamA, thiH, LA_2712, prfC, rodA, LA_2755, atpD, atpG, atpA, sufB, LA_2821, argF, citE, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, purB, spoT, tgt, LA_3135, caiB, LA_3265, LA_3266, LA_3268, dxs, trpB, proS, uppS, frr, pyrH, tsf, rpsB, acpS, dapA, ccmE, LA_3354, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, ppk, CsdB, sufC, LA_3573, LA_3579, lon, LA_3597, LA_3606, LA_3607, LA_3627, hflX, pyrC, LA_3684, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3803, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4034, fadB, LA_4139, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, mtnP, acsA, folE, serS, tatD, gidA, ilvE, hemL, LB_074, sam1, metH, pfkA, argB, LB_170, apt, LB_208, LB_212, GalF, Mcm3, gltB, efp, speE, acnA, Pyk2, asd
GO:0046872	Metal ion binding	96	425	gyrB, cyoB, LA_0250, pckA, LA_0332, gpmI, nadA, LA_0666, carB, pyrB, coaX, rlmN, obg, nuoF, pnp, LA_1019, sucC, LA_1105, purA, rimO, ribB, thrS, glnA, LA_1323, ileS, LA_1417, ovp1, LA_1483, LA_1896, eno, lysU, leuC, LA_2118, bioB, LA_2151, LA_2152, uvrA, fbp, LA_2297, lpxC, ychF, hslV, gspF, rpe, hprK, pyrG, def, map, clpX, IlvB, bfr, KamA, thiH, LA_2712, citE, radA, PykF, icmF, eutG, LA_3094, tgt, LA_3265, LA_3266, LA_3268, dxs, uppS, acpS, ccmE, LA_3354, rpoC, ppk, dps, hflX, pyrC, dnaJ, add, hflB, prsA, metG, LA_3936, pheS, ndk, glnS, LA_4194, ilvC, acsA, folE,

				tatD, LB_074, metH, pfkA, LB_212, Mcm3, gltB, acnA, Pyk2
GO:0006807	Nitrogen compound metabolic process	218	966	gyrB, LA_0006, LA_0021, LA_0250, gatB, dnaE, fusA, LA_0332, gcvP, gcvH, gcvT, trpS, gpmI, ribH, mltG, nadA, uvrB, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, ruvB, coaX, rlmN, rpmA, proA, nusA, pnp, rpmE, rho, lysA, LA_1105, purA, rimO, ribB, LA_1205, LA_1221, LA_1222, thrS, infC, rplT, LA_1260, hisG, rpmF, LA_1293, smpB, glnA, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1719, panD, folD, asnS, lepA, LA_1911, LA_1935, eno, LA_1953, guaB, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, recA, LA_2207, uvrA, rpsU, rpoD, tyrS, purH, LA_2318, hslU, hslV, xerC, rplS, trmD, rpsP, hprK, pyrG, def, map, LA_2483, hisF, LA_2507, mfd, LA_2559, ilvH, fliI, LA_2603, LA_2631, abgB, KamA, prfC, rodA, LA_2755, atpD, atpG, atpA, argF, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3268, dxs, trpB, proS, frr, pyrH, tsf, rpsB, dapA, ccmE, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, CsdB, LA_3573, lon, hflX, pyrC, LA_3684, leuS, thrC, valS, LA_3767, add, hflB, LA_3820, prsA, metG, LA_3936, gmd, LA_3978, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, mtnP, acsA, folE, serS, tatD, gidA, ilvE, hemL, LB_074, metH, pfkA, argB, apt, LB_212, GalF, Mcm3, gltB, efp, speE, Pyk2, asd
GO:0044238	Primary metabolic process	248	1104	gyrB, LA_0006, LA_0021, LA_0106, LA_0250, pckA, gatB, dnaE, fusA, LA_0332, gcvP, gcvH, gcvT, rpiB, trpS, gpmI, LA_0457, nadA, uvrB, LA_0671, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, LA_0790, ruvB, LA_0828, coaX, rlmN, rpmA, proA, nusA, pnp, rpmE, rho, LA_1032, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, LA_1189, LA_1205, LA_1221, LA_1222, sucA, thrS, infC, rplT, LA_1260, hisG, rpmF, plsX, LA_1293, smpB, glnA, LA_1323, ileS, grs1, rbsK, glyA, lexA, pssA, ugd, purE, trmU, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, LA_1719, panD, folD, asnS, lepA, LA_1896,

				LA_1935, eno, LA_1953, guaB, lysU, met2, leuC, LA_2118, LA_2119, LA_2130, mdh, LA_2152, rpmB, LA_2172, LA_2173, LA_2174, recA, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, purH, mipB, lpxC, LA_2309, LA_2318, LA_2326, hslU, hslV, xerC, rplS, trmD, rpsP, rpe, hprK, kdsA, pyrG, LA_2432, LA_2433, def, map, LA_2483, hisF, LA_2507, mfd, LA_2559, ilvH, fliI, LA_2603, LA_2631, LA_2632, abgB, KamA, prfC, atpD, atpG, atpA, argF, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3135, dxs, trpB, proS, frr, pyrH, tsf, rpsB, acpS, dapA, ccmE, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, CsdB, LA_3573, lon, LA_3597, LA_3627, hflX, pyrC, LA_3684, leuS, thrC, valS, LA_3767, add, LA_3803, hflB, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, fadB, LA_4139, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, mtnP, acsA, serS, tatD, gidA, ilvE, LB_074, metH, pfkA, argB, LB_170, apt, LB_212, GalF, Mcm3, gltB, efp, acnA, Pyk2, asd
GO:0006139	Nucleobase-containing compound metabolic process	98	445	gyrB, LA_0006, LA_0021, LA_0250, dnaE, LA_0332, trpS, gpmI, nadA, uvrB, carB, adk, rpoA, pyrB, ruvB, coaX, rlmN, nusA, pnp, rho, LA_1105, purA, rimO, LA_1205, LA_1222, thrS, LA_1323, ileS, grs1, lexA, ugd, purE, trmU, nadB, manC, fbaB, aspS, tpiA, pgk, folD, asnS, LA_1935, eno, guaB, lysU, recA, LA_2207, uvrA, rpoD, tyrS, purH, LA_2318, xerC, trmD, pyrG, LA_2483, mfd, fliI, LA_2603, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, proS, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, leuS, valS, add, prsA, metG, LA_3936, gmd, pheS, coaD, ndk, glnS, LA_4193, LA_4194, mtnP, acsA, serS, tatD, gidA, pfkA, apt, LB_212, GalF, Pyk2
GO:0044260	Cellular macromolecule metabolic process	120	556	gyrB, LA_0006, gatB, dnaE, fusA, trpS, mltG, uvrB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpoA, ruvB, rlmN, rpmA, nusA, pnp, rpmE, rho, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, rpmF, smpB, ileS, grs1, lexA, manC, rplI, aspS, asnS, lepA, lysU, rfe, LA_2118, rpmB, LA_2173, LA_2174, recA, uvrA, rpsU, rpoD, tyrS, hslV, xerC, nrdA, rplS, trmD, rpsP, hprK, kdsA, def, map, LA_2483, LA_2507, mfd, LA_2603, prfC, rodA, LA_2755, radA, lgt, proS, frr, tsf,

				rpsB, ccmE, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, LA_3573, lon, LA_3597, hflX, LA_3684, dnaJ, leuS, valS, LA_3767, LA_3820, LA_3905, metG, LA_3978, pheS, glnS, LA_4193, LA_4194, serS, tatD, LB_170, LB_208, efp
GO:0071840	Cellular component organization or biogenesis	35	163	gyrB, LA_0006, mltG, ftsZ, ruvB, rlmN, obg, LA_1019, LA_1221, rplT, LA_1295, bipA, rfe, fliS, mfd, fliH, fliI, flhA, fliQ, fliP, prfC, rodA, LA_2755, sufB, frr, ccmE, ccmF, rplJ, sufC, hflX, LA_3767, LA_3978, LA_4193, LA_4194, FlgK
GO:0051179	Localization	59	279	LA_0025, cyoB, LA_0273, LA_0274, FlgB, secY, secD, cysU, cysA, LA_1438, ovp1, LA_1543, LA_1855, secA, LA_1964, LA_2017, flaB, ttg2B, fliM, LA_2113, LA_2114, LA_2151, LA_2188, LA_2207, fliS, LA_2319, gspF, gspD, yhbG, flgL, LA_2418, fliH, fliF, fliI, flhA, fliQ, fliP, LA_2667, bfr, atpD, atpG, atpA, flgE, LA_2891, ccmF, secE, LA_3498, LA_3575, LA_3576, LA_3577, LA_3630, LA_3694, LA_3794, LA_3806, LA_4155, LA_4228, LA_4229, LA_4308, ExbD3
GO:0009987	Cellular process	374	1773	gyrB, LA_0006, LA_0021, LA_0025, LA_0106, cyoB, LA_0249, LA_0250, gatB, dnaE, fusA, LA_0319, LA_0332, FlgB, gcvP, gcvH, gcvT, rpiB, LA_0391, trpS, etfA, etfB, gpmI, LA_0457, ribH, mltG, ftsA, ftsZ, nadA, LA_0627, uvrB, LA_0666, LA_0671, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, secY, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, rpmA, obg, proA, nuoK, mgsA, nusA, pnp, LA_1019, rpmE, rho, LA_1032, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, secD, ribB, cysU, cysA, LA_1205, LA_1221, LA_1222, LA_1223, sucA, thrS, infC, rplT, LA_1253, scpB, LA_1260, hisG, rpmF, plsX, LA_1293, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, lexA, pssA, ugd, purE, ovp1, LA_1483, trmU, lldD, nadB, manC, LA_1521, fbaB, LA_1543, rplI, aspS, LA_1692, tpiA, pgk, gapA, LA_1713, LA_1719, panD, LA_1855, folD, asnS, lepA, LA_1896, LA_1911, LA_1935, LA_1948, eno, secA, LA_1964, guaB, lysU, acoB, acoA, LA_2017, flaB, rfe, met2, fliM, leuC, LA_2113, LA_2114, LA_2118, LA_2119, LA_2130, mdh, bioB, LA_2151, LA_2152, rpmB, LA_2172,

				LA_2173, LA_2174, recA, LA_2188, LA_2202, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, fliS, purH, mipB, lpxC, LA_2309, LA_2318, LA_2319, LA_2326, ubiX, mrp, hslU, hslV, xerC, LA_2350, nrdA, gspF, gspE, gspD, rplS, trmD, rpsP, rpe, atoC, hprK, yhbG, kdsA, pyrG, flgL, LA_2418, LA_2432, LA_2433, def, map, LA_2458, LA_2483, hisF, LA_2507, mfd, clpX, ilvH, IlvB, LA_2572, fliH, fliF, fliI, LA_2603, LA_2605, flhA, fliQ, fliP, LA_2627, groS, groL, LA_2667, abgB, bfr, KamA, thiH, LA_2712, prfC, rodA, LA_2755, atpD, atpG, atpA, sufB, ahpC, LA_2821, argF, citE, flgE, LA_2891, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, purB, spoT, tgt, LA_3135, caiB, LA_3265, LA_3266, LA_3268, dxs, trpB, proS, uppS, frr, pyrH, tsf, rpsB, acpS, dapA, ccmE, ccmF, LA_3354, LA_3406, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, secE, Bcp, ppk, LA_3498, CsdB, sufC, LA_3573, LA_3575, LA_3576, LA_3577, LA_3579, lon, LA_3597, dps, LA_3606, LA_3607, LA_3627, LA_3630, hflX, pyrC, LA_3684, LA_3694, dnaK, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3803, LA_3806, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4034, fadB, LA_4139, pheS, LA_4155, argG, coaD, ndk, glnS, LA_4193, LA_4194, LA_4228, LA_4229, ilvC, mtnP, acsA, folE, LA_4308, FlgK, serS, tatD, LA_4347, gidA, ilvE, hemL, ParB3, LB_074, sam1, metH, pfkA, argB, LB_170, apt, LB_208, LB_212, GalF, Mcm3, ExbD3, gltB, efp, speE, acnA, LB_333, Pyk2, asd, ParA4
GO:0006793	Phosphorus metabolic process	63	301	gpmI, nadA, lysC, carB, adk, pyrB, coaX, nuoK, purA, LA_1112, LA_1222, plsX, LA_1323, rbsK, pssA, ugd, purE, nadB, fbaB, tpiA, pgk, gapA, folD, eno, guaB, LA_2119, fbp, purH, mipB, lpxC, LA_2318, LA_2326, rpe, hprK, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, dxs, pyrH, ppk, LA_3573, hflX, pyrC, add, prsA, LA_3936, LA_3939, gmd, coaD, ndk, acsA, folE, pfkA, argB, apt, LB_212, GalF, Pyk2
GO:0071704	Organic substance metabolic process	280	1349	gyrB, LA_0006, LA_0021, LA_0106, LA_0250, pckA, gatB, dnaE, fusA, LA_0319, LA_0332, gcvP, gcvH, gcvT, rpiB, trpS, gpmI, LA_0457, ribH, mltG, nadA, uvrB, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK,

				rpsD, rpoA, dapL, pyrB, ruvB, LA_0828, coaX, rlmN, rpmA, proA, mgsA, nusA, pnp, rpmE, rho, LA_1032, lysA, LA_1105, purA, LA_1112, rimO, ribB, LA_1205, LA_1221, LA_1222, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, LA_1293, smpB, smc, glnA, LA_1323, ileS, grs1, rbsK, glyA, lexA, pssA, ugd, purE, trmU, lldD, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, asnS, lepA, LA_1911, LA_1935, eno, LA_1953, guaB, lysU, acoB, acoA, rfe, met2, leuC, LA_2118, LA_2119, LA_2130, mdh, bioB, LA_2152, rpmB, LA_2172, LA_2173, LA_2174, recA, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, purH, mipB, lpxC, LA_2309, LA_2318, LA_2326, hslU, hslV, xerC, nrdA, rplS, trmD, rpsP, rpe, hprK, kdsA, pyrG, LA_2432, LA_2433, def, map, LA_2483, hisF, LA_2507, mfd, LA_2559, ilvH, IlvB, fliI, LA_2603, LA_2627, LA_2631, LA_2632, abgB, KamA, thiH, LA_2712, prfC, rodA, LA_2755, atpD, atpG, atpA, LA_2821, argF, citE, LA_2894, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3135, caiB, LA_3268, dxs, trpB, proS, uppS, frr, pyrH, tsf, rpsB, acpS, dapA, ccmE, LA_3354, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, ppk, CsdB, LA_3573, LA_3579, lon, LA_3597, LA_3606, LA_3607, LA_3627, hflX, pyrC, LA_3684, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3803, hflB, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4034, fadB, LA_4139, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, mtnP, acsA, folE, serS, tatD, gidA, ilvE, hemL, LB_074, metH, pfkA, argB, LB_170, apt, LB_208, LB_212, GalF, Mcm3, gltB, efp, speE, acnA, Pyk2, asd
GO:0046483	Heterocycle metabolic process	112	541	gyrB, LA_0006, LA_0021, LA_0250, dnaE, LA_0332, trpS, gpmI, ribH, nadA, uvrB, carB, adk, rpoA, pyrB, ruvB, coaX, rlmN, proA, nusA, pnp, rho, LA_1105, purA, rimO, ribB, LA_1205, LA_1222, thrS, hisG, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, folD, asnS, LA_1935, eno, guaB, lysU, LA_2118, bioB, recA, LA_2207, uvrA, rpoD, tyrS, purH, LA_2318, xerC, trmD, pyrG, LA_2483, hisF, mfd, fliI, LA_2603, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, dxs, trpB, proS, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, leuS, valS, add, prsA, metG, LA_3936, gmd, pheS, coaD,

				ndk, glnS, LA_4193, LA_4194, mtnP, acsA, folE, serS, tatD, gidA, hemL, metH, pfkA, apt, LB_212, GalF, Pyk2
GO:0006796	Phosphate-containing compound metabolic process	58	281	gpmI, nadA, lysC, carB, adk, pyrB, coaX, nuoK, purA, LA_1112, LA_1222, plsX, LA_1323, rbsK, pssA, purE, nadB, fbaB, tpiA, pgk, gapA, folD, eno, guaB, LA_2119, fbp, purH, mipB, lpxC, LA_2318, rpe, hprK, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, dxs, pyrH, ppk, hflX, pyrC, add, prsA, LA_3936, LA_3939, coaD, ndk, acsA, folE, pfkA, argB, apt, LB_212, Pyk2
GO:0008152	Metabolic process	333	1614	gyrB, LA_0006, LA_0020, LA_0021, LA_0106, cyoB, LA_0250, pckA, gatB, dnaE, fusA, LA_0319, LA_0332, LA_0334, pntB, gcvP, gcvH, gcvT, rpiB, trpS, etfA, etfB, ivd, gpmI, LA_0457, ribH, mltG, LA_0560, nadA, uvrB, LA_0666, LA_0671, lysC, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, rpmA, proA, nuoK, nuoH, nuoF, nuoD, nuoA, mgsA, nusA, pnp, LA_1019, rpmE, rho, LA_1032, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, ribB, LA_1189, LA_1205, LA_1221, LA_1222, LA_1223, sucA, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, LA_1293, smpB, smc, glnA, LA_1323, ileS, grs1, rbsK, glyA, lexA, pssA, ugd, purE, trmU, lldD, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, asnS, lepA, LA_1896, sdhA, LA_1911, LA_1930, LA_1935, eno, LA_1953, LA_1977, guaB, lysU, acoB, acoA, rfe, met2, leuC, LA_2118, LA_2119, LA_2130, mdh, bioB, LA_2152, rpmB, LA_2172, LA_2173, LA_2174, recA, LA_2202, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, purH, mipB, LA_2297, lpxC, LA_2309, LA_2318, LA_2324, LA_2326, ubiX, mrp, hslU, hslV, xerC, LA_2350, nrdA, rplS, trmD, rpsP, rpe, hprK, kdsA, pyrG, LA_2432, LA_2433, def, map, LA_2483, trxB, hisF, LA_2507, mfd, LA_2559, ilvH, IlvB, LA_2572, fliI, LA_2603, LA_2627, LA_2631, LA_2632, abgB, bfr, KamA, thiH, LA_2712, prfC, rodA, LA_2755, atpD, atpG, atpA, sufB, ahpC, LA_2821, argF, citE, LA_2894, radA, LA_2917, PykF, sseA, icmF, eutG, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3135, LA_3143, caiB, LA_3219, LA_3265, LA_3266, LA_3268,

				dxs, trpB, proS, uppS, frr, pyrH, tsf, rpsB, acpS, dapA, ccmE, LA_3354, LA_3363, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, Bcp, ppk, glcD, LA_3484, CsdB, sufC, LA_3573, LA_3579, lon, LA_3597, dps, LA_3606, LA_3607, LA_3627, hflX, pyrC, LA_3684, dnaJ, mviM, leuS, thrC, valS, LA_3767, add, LA_3803, hflB, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4034, fadB, LA_4139, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, mtnP, acsA, folE, LA_4327, serS, tatD, gidA, ilvE, hemL, LB_074, LB_082, sam1, metH, pfkA, argB, LB_170, apt, LB_208, LB_212, GalF, Mcm3, gltB, efp, speE, acnA, Pyk2, asd
GO:0110165	Cellular anatomical entity	421	2078	gyrB, LA_0006, LA_0020, LA_0021, LA_0025, LA_0106, cyoB, LA_0250, pckA, gatB, dnaE, LA_0273, LA_0274, LA_0280, LA_0296, LA_0309, fusA, LA_0319, LA_0332, LA_0334, pntB, FlgB, gcvP, gcvH, gcvT, rpiB, LA_0391, trpS, etfA, etfB, ivd, gpmI, LA_0457, mltG, LA_0560, ftsA, ftsZ, nadA, LA_0627, uvrB, LA_0666, LA_0671, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, secY, adk, rpsM, rpsK, rpsD, rpoA, dapL, pyrB, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, rpmA, obg, proA, nuoK, nuoH, nuoF, nuoD, nuoA, mgsA, nusA, pnp, LA_1019, rpmE, rho, lysA, sucD, sucC, LA_1105, purA, LA_1112, LA_1137, rimO, secD, ribB, cysU, cysA, LA_1189, LA_1205, LA_1221, LA_1222, LA_1223, sucA, thrS, infC, rplT, LA_1253, scpB, LA_1260, hisG, rpmF, plsX, LA_1293, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, LA_1438, pssA, ugd, purE, ovp1, LA_1483, trmU, lldD, nadB, hfq, LA_1521, fbaB, LA_1543, rplI, aspS, phoH, LA_1692, tpiA, pgk, gapA, LA_1713, LA_1719, panD, LA_1855, folD, asnS, lepA, LA_1896, sdhA, LA_1911, LA_1929, LA_1930, LA_1935, LA_1948, eno, LA_1953, secA, LA_1964, LA_1977, lysU, acoB, acoA, LA_2017, flaB, rfe, ttg2A, ttg2B, met2, LA_2072, fliM, leuC, LA_2113, LA_2114, LA_2118, LA_2119, ssl2, LA_2130, mdh, bioB, LA_2151, LA_2152, rpmB, LA_2172, LA_2173, LA_2174, recA, LA_2188, LA_2202, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, fliS, purH, mipB, LA_2297, lpxC, LA_2309, LA_2318,

				LA_2319, LA_2324, LA_2325, LA_2326, ychF, ubiX, mrp, hslU, hslV, xerC, LA_2350, nrdA, gspF, gspE, gspD, rplS, trmD, rpsP, rpe, atoC, hprK, yhbG, kdsA, pyrG, flgL, LA_2418, LA_2432, LA_2433, def, map, LA_2458, LA_2483, trxB, hisF, mfd, clpX, LA_2559, ilvH, IlvB, LA_2572, fliH, fliF, fliI, LA_2603, LA_2605, flhA, fliQ, fliP, LA_2627, LA_2631, groS, groL, LA_2667, abgB, bfr, thiH, LA_2712, prfC, rodA, LA_2755, LA_2756, mreB, atpD, atpG, atpA, sufB, ahpC, LA_2821, argF, citE, flgE, LA_2891, LA_2894, radA, LA_2917, PykF, sseA, icmF, eutG, lgt, LA_3052, LA_3054, purB, spoT, LA_3094, tgt, LA_3135, LA_3143, caiB, LA_3219, LA_3265, LA_3266, LA_3268, dxs, trpB, proS, uppS, frr, pyrH, tsf, rpsB, acpS, dapA, ccmE, ccmF, LA_3354, LA_3363, LA_3406, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, secE, Bcp, glcD, LA_3484, LA_3498, CsdB, sufC, LA_3573, LA_3575, LA_3576, LA_3577, LA_3579, lon, LA_3597, dps, LA_3606, LA_3607, LA_3627, LA_3630, hflX, pyrC, LA_3684, LA_3694, hrcA, dnaK, dnaJ, mviM, leuS, thrC, valS, LA_3767, add, LA_3794, LA_3803, LA_3806, glnK, hflB, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4034, LA_4036, LA_4139, pheS, LA_4155, argG, coaD, ndk, glnS, LA_4193, LA_4194, LA_4228, LA_4229, ilvC, mtnP, acsA, folE, LA_4308, FlgK, LA_4327, serS, tatD, LA_4347, gidA, ilvE, hemL, ParB3, LB_074, LB_082, sam1, metH, pfkA, argB, LB_170, apt, LB_208, LB_212, GalF, Mcm3, ExbD3, gltB, efp, speE, acnA, LB_333, Pyk2, asd, ParA4
GO:1901360	Organic cyclic compound metabolic process	112	555	gyrB, LA_0006, LA_0021, LA_0250, dnaE, LA_0332, trpS, gpmI, ribH, nadA, uvrB, carB, adk, rpoA, pyrB, ruvB, coaX, rlmN, proA, nusA, pnp, rho, LA_1105, purA, rimO, ribB, LA_1205, LA_1222, thrS, hisG, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, folD, asnS, LA_1935, eno, guaB, lysU, LA_2118, bioB, recA, LA_2207, uvrA, rpoD, tyrS, purH, LA_2318, xerC, trmD, pyrG, LA_2483, hisF, mfd, fliI, LA_2603, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, dxs, trpB, proS, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, leuS, valS, add, prsA, metG, LA_3936, gmd, pheS, coaD, ndk, glnS, LA_4193, LA_4194, mtnP, acsA,

				folE, serS, tatD, gidA, hemL, metH, pfkA, apt, LB_212, GalF, Pyk2
GO:0006725	Cellular aromatic compound metabolic process	107	532	gyrB, LA_0006, LA_0021, LA_0250, dnaE, LA_0332, trpS, gpmI, nadA, uvrB, carB, adk, rpoA, pyrB, ruvB, coaX, rlmN, nusA, pnp, rho, LA_1105, purA, rimO, LA_1205, LA_1222, thrS, hisG, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, folD, asnS, LA_1935, eno, guaB, lysU, recA, LA_2207, uvrA, rpoD, tyrS, purH, LA_2318, xerC, trmD, pyrG, LA_2483, hisF, mfd, fliI, LA_2603, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, dxs, trpB, proS, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, leuS, valS, add, prsA, metG, LA_3936, gmd, pheS, coaD, ndk, glnS, LA_4193, LA_4194, mtnP, acsA, folE, serS, tatD, gidA, hemL, metH, pfkA, apt, LB_212, GalF, Pyk2
GO:0043170	Macromolecule metabolic process	137	713	gyrB, LA_0006, LA_0021, LA_0250, gatB, dnaE, fusA, trpS, mltG, uvrB, tufB, rplC, rplW, rplB, rpsS, rplV, rpsC, rplP, rpsQ, rplN, rplX, rplE, rpsN, rpsH, rplF, rpsE, rpsM, rpsK, rpsD, rpoA, ruvB, rlmN, rpmA, nusA, pnp, rpmE, rho, LA_1105, rimO, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, rpmF, smpB, smc, ileS, grs1, lexA, ugd, trmU, manC, rplI, aspS, asnS, lepA, LA_1935, LA_1953, lysU, rfe, LA_2118, rpmB, LA_2173, LA_2174, recA, uvrA, rpsU, rpoD, tyrS, hslU, hslV, xerC, nrdA, rplS, trmD, rpsP, hprK, kdsA, def, map, LA_2483, LA_2507, mfd, LA_2559, LA_2603, LA_2631, prfC, rodA, LA_2755, radA, LA_2917, lgt, LA_3054, tgt, proS, frr, tsf, rpsB, ccmE, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplL, rplJ, rplA, rplK, nusG, LA_3573, lon, LA_3597, hflX, LA_3684, dnaJ, leuS, valS, LA_3767, hflB, LA_3820, LA_3905, metG, LA_3978, pheS, glnS, LA_4193, LA_4194, serS, tatD, gidA, LB_170, LB_208, efp
GO:0003824	Catalytic activity	302	1592	gyrB, LA_0006, LA_0020, LA_0021, LA_0025, LA_0106, cyoB, LA_0250, pckA, gatB, dnaE, LA_0273, LA_0274, LA_0296, fusA, LA_0319, LA_0332, LA_0334, pntB, gcvP, gcvT, rpiB, trpS, etfA, etfB, ivd, gpmI, LA_0457, ribH, mltG, LA_0560, ftsZ, nadA, LA_0627, uvrB, LA_0666, LA_0671, lysC, carB, tufB, rplB, adk, rpoA, dapL, pyrB, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, obg, proA, nuoK, nuoH, nuoF, nuoD, nuoA, mgsA, pnp, rho, LA_1032,

				lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, ribB, cysA, LA_1221, LA_1222, LA_1223, sucA, thrS, hisG, plsX, LA_1293, LA_1295, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, LA_1438, lexA, pssA, ugd, purE, ovp1, LA_1483, trmU, lldD, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, asnS, lepA, LA_1896, sdhA, LA_1911, LA_1930, LA_1935, LA_1948, eno, LA_1953, LA_1977, guaB, lysU, acoB, acoA, rfe, ttg2A, met2, LA_2072, fliM, leuC, LA_2114, LA_2118, LA_2119, ssl2, LA_2130, mdh, bioB, LA_2152, LA_2172, LA_2173, LA_2174, recA, LA_2188, LA_2202, uvrA, fbp, tyrS, purH, mipB, LA_2297, lpxC, LA_2309, LA_2318, LA_2324, LA_2326, ychF, ubiX, mrp, hslU, hslV, xerC, LA_2350, nrdA, trmD, rpe, hprK, yhbG, kdsA, pyrG, LA_2432, LA_2433, def, map, LA_2469, LA_2483, trxB, hisF, LA_2507, mfd, LA_2559, ilvH, IlvB, fliH, fliF, fliI, LA_2605, LA_2627, LA_2632, abgB, bfr, KamA, thiH, prfC, rodA, atpD, atpG, atpA, ahpC, argF, citE, radA, LA_2917, PykF, sseA, icmF, eutG, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3135, LA_3143, caiB, LA_3219, LA_3265, LA_3266, LA_3268, dxs, trpB, proS, uppS, pyrH, acpS, dapA, LA_3354, LA_3363, rpoC, rpoB, Bcp, ppk, glcD, LA_3484, CsdB, sufC, LA_3573, LA_3579, lon, LA_3597, dps, LA_3606, LA_3607, LA_3627, LA_3630, hflX, pyrC, LA_3684, LA_3694, mviM, leuS, thrC, valS, add, LA_3803, hflB, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, LA_3941, gmd, acyP, LA_3978, LA_4034, LA_4036, fadB, LA_4139, pheS, argG, coaD, ndk, glnS, LA_4193, LA_4194, ilvC, mtnP, acsA, folE, LA_4327, serS, tatD, ilvE, hemL, LB_074, LB_082, sam1, metH, pfkA, argB, LB_170, apt, LB_208, LB_212, GalF, Mcm3, gltB, speE, acnA, Pyk2, asd
GO:0016740	Transferase activity	90	491	LA_0021, dnaE, LA_0319, gcvT, LA_0457, ribH, mltG, nadA, LA_0627, LA_0671, lysC, rplB, adk, rpoA, dapL, LA_0790, LA_0828, coaX, rlmN, pnp, LA_1105, rimO, LA_1221, LA_1222, hisG, plsX, LA_1293, rbsK, glyA, pssA, LA_1483, trmU, manC, pgk, LA_1719, rfe, met2, LA_2119, LA_2130, bioB, LA_2172, LA_2173, LA_2174, purH, mipB, trmD, hprK, kdsA, hisF, ilvH, thiH, rodA, argF, PykF, sseA, lgt, LA_3052, spoT, tgt, LA_3135, caiB, LA_3219, dxs, uppS,

	pyrH, acpS, rpoC, rpoB, ppk, CsdB, LA_3573, LA_3579, LA_3597, prsA, LA_3905, LA_3978, LA_4139, coaD, ndk, mtnP, ilvE, hemL, metH, pfkA, argB, apt, LB_208, GalF, speE, Pyk2

Table S3.10. Proteins that exhibit AA changes in P1$^-$ and the mutations are conserved in P1$^+$.

L. interrogans serovar Lai strain 56601 used as reference.

Locus tag	Product	Locus tag	Product
LA_0005	DNA gyrase subunit B	LA_2324	3-ketoacyl-ACP reductase
LA_0006	DNA gyrase subunit A	LA_2325	transcriptional regulator
LA_0020	short-chain dehydrogenase	LA_2326	nucleoside-diphosphate-sugar epimerase
LA_0106	long-chain-fatty-acid CoA ligase	LA_2329	GTP-dependent nucleic acid-binding protein EngD
LA_0243	cytochrome c oxidase polypeptide I	LA_2335	3-polyprenyl-4-hydroxybenzoate decarboxylase
LA_0249	transcriptional regulator	LA_2336	hypothetical protein
LA_0251	phosphoenolpyruvate carboxykinase	LA_2337	chromosome-partitioning ATPase
LA_0256	aspartyl/glutamyl-tRNA amidotransferase subunit B	LA_2345	ATP-dependent protease ATP-binding subunit HslU
LA_0258	DNA polymerase III subunit alpha	LA_2346	ATP-dependent protease peptidase subunit
LA_0273	lipoprotein releasing system permease	LA_2347	integrase/recombinase XerD
LA_0274	lipoprotein releasing system LolD	LA_2350	isopropylmalate/homocitrate/citramalate synthase
LA_0280	cAMP-binding protein	LA_2360	ribonucleotide-diphosphate reductase subunit alpha
LA_0296	Zn-dependent alcohol dehydrogenase	LA_2373	type II secretory pathway component protein F
LA_0309	cAMP-binding protein	LA_2374	type II secretory pathway ATPase protein E
LA_0313	elongation factor G	LA_2375	type II secretory pathway component protein D
LA_0319	phosphoribose diphosphate:decaprenyl-phosphate phosphoribosyltransferase	LA_2387	50S ribosomal protein L19
LA_0332	6-carboxytetrahydropterin synthase QueD	LA_2388	tRNA (guanine-N(1)-)-methyltransferase
LA_0334	proton-translocating transhydrogenase subunit alpha	LA_2391	30S ribosomal protein S16
LA_0335	NAD(P)(+) transhydrogenase subunit beta	LA_2394	ribulose-5-phosphate 3-epimerase

LA_0347	flagellar basal-body rod protein FlgB	LA_2400	two-component system response regulator
LA_0360	glycine dehydrogenase	LA_2403	HPr kinase/phosphorylase
LA_0361	glycine cleavage system protein H	LA_2405	ABC transporter ATP-binding protein
LA_0362	glycine cleavage system aminomethyltransferase T	LA_2409	CTP synthetase
LA_0369	ribose-5-phosphate isomerase	LA_2417	endoflagellar hook-filament protein
LA_0391	ATP-dependent protease ClpA	LA_2418	endoflagellar filament core protein
LA_0395	anti-sigma factor antagonist	LA_2432	biotin carboxylase
LA_0408	tryptophanyl-tRNA synthetase	LA_2433	acetyl-CoA carboxylase
LA_0411	electron transfer flavoprotein subunit alpha	LA_2436	hypothetical protein
LA_0412	electron transfer flavoprotein subunit beta	LA_2438	peptide deformylase
LA_0414	isovaleryl-CoA dehydrogenase	LA_2457	methionyl aminopeptidase
LA_0439	2,3-bisphosphoglycerate-independent phosphoglycerate mutase	LA_2469	MCP methylation inhibitor
LA_0457	acetyl-CoA acetyltransferase	LA_2494	thioredoxin reductase
LA_0463	6,7-dimethyl-8-ribityllumazine synthase	LA_2507	aspartyl/glutamyl-tRNA amidotransferase subunit A
LA_0547	hypothetical protein	LA_2513	transcription-repair coupling factor
LA_0552	periplasmic solute-binding protein	LA_2558	ATP-dependent protease ATP-binding subunit ClpX
LA_0560	acyl-CoA dehydrogenase	LA_2559	ATP-dependent Clp protease proteolytic subunit
LA_0611	cell division protein FtsA	LA_2569	acetolactate synthase small subunit
LA_0612	cell division protein FtsZ	LA_2570	acetolactate synthase large subunit
LA_0618	quinolinate synthetase	LA_2572	TPR-repeat-containing protein
LA_0627	glycosyltransferase	LA_2576	hypothetical protein
LA_0629	hypothetical protein	LA_2589	flagellar assembly protein H
LA_0645	hypothetical protein	LA_2591	flagellar MS-ring protein
LA_0649	excinuclease ABC subunit B	LA_2592	endoflagellar biosynthesis/type III secretory pathway ATPase
LA_0653	anti-sigma factor antagonist	LA_2602	polymerase
LA_0666	cytochrome c peroxidase	LA_2603	RNA polymerase sigma factor WhiG

LA_0671	citrate synthase	LA_2605	ParA protein
LA_0693	aspartate kinase	LA_2607	endoflagellar biosynthesis protein
LA_0720	hypothetical protein	LA_2627	MaoC family protein
LA_0727	carbamoyl phosphate synthase large subunit	LA_2631	ATP-dependent Clp protease adaptor protein ClpS
LA_0737	elongation factor Tu	LA_2632	hypothetical protein
LA_0739	50S ribosomal protein L3	LA_2655	molecular chaperone GroEL
LA_0741	50S ribosomal protein L23	LA_2667	flagellar basal body rod protein FlgG
LA_0742	50S ribosomal protein L2	LA_2690	bacterioferritin
LA_0744	50S ribosomal protein L22	LA_2712	Fe-S oxidoreductase
LA_0746	50S ribosomal protein L16	LA_2727	dioxygenase
LA_0749	50S ribosomal protein L14	LA_2733	peptide chain release factor 3
LA_0750	50S ribosomal protein L24	LA_2739	hypothetical protein
LA_0751	50S ribosomal protein L5	LA_2749	6-pyruvoyl tetrahydrobiopterin synthase
LA_0752	30S ribosomal protein S14	LA_2754	rod shape-determining protein
LA_0753	30S ribosomal protein S8	LA_2755	transpeptidase/penicillin-binding protein
LA_0756	30S ribosomal protein S5	LA_2756	cell shape-determining protein
LA_0759	preprotein translocase subunit SecY	LA_2759	rod shape-determining protein MreB
LA_0760	adenylate kinase	LA_2776	ATP synthase F0F1 subunit beta
LA_0762	30S ribosomal protein S13	LA_2778	ATP synthase F0F1 subunit gamma
LA_0763	30S ribosomal protein S11	LA_2779	ATP synthase F0F1 subunit alpha
LA_0764	30S ribosomal protein S4	LA_2808	ABC transporter permease
LA_0765	DNA-directed RNA polymerase subunit alpha	LA_2809	peroxiredoxin
LA_0776	L,L-diaminopimelate aminotransferase	LA_2821	hypothetical protein
LA_0790	citrate synthase	LA_2840	ornithine carbamoyltransferase
LA_0810	Holliday junction DNA helicase RuvB	LA_2848	flagellar hook protein FlgE
LA_0811	hypothetical protein	LA_2876	hypothetical protein
LA_0828	acetyl-CoA acetyltransferase	LA_2891	Na+/H+ antiporter
LA_0833	pantothenate kinase	LA_2894	transcriptional regulator
LA_0841	23S rRNA (adenine(2503)-C(2))-methyltransferase RlmN	LA_2916	DNA repair protein RadA

LA_0845	2-isopropylmalate synthase	LA_2917	ribonuclease D
LA_0851	50S ribosomal protein L27	LA_2924	pyruvate kinase
LA_0852	GTPase ObgE	LA_2947	thiosulfate sulfurtransferase
LA_0854	gamma-glutamyl phosphate reductase	LA_2956	methylmalonyl-CoA mutase
LA_0889	NADH dehydrogenase (ubiquinone) chain H	LA_2996	alcohol dehydrogenase
LA_0890	NADH dehydrogenase (ubiquinone) chain F	LA_3004	prolipoprotein diacylglyceryl transferase
LA_0891	NADH dehydrogenase (ubiquinone) chain E	LA_3052	aminotransferase
LA_0892	NADH dehydrogenase (ubiquinone) chain D	LA_3054	hypothetical protein
LA_0894	NADH dehydrogenase subunit B	LA_3080	adenylosuccinate lyase
LA_0895	NADH dehydrogenase (ubiquinone) chain A	LA_3083	hypothetical protein
LA_0909	methylglyoxal synthase	LA_3085	guanosine polyphosphate pyrophosphohydrolase/synthetase
LA_0942	transcription elongation factor NusA	LA_3094	ferric uptake regulator
LA_0947	polynucleotide phosphorylase	LA_3095	queuine tRNA-ribosyltransferase
LA_0969	ABC transporter ATP-binding protein	LA_3135	acetyl-CoA acetyltransferase
LA_0970	ABC transporter permease	LA_3143	acyl-CoA dehydrogenase
LA_1019	hypothetical protein	LA_3144	acyl-CoA transferase
LA_1020	50S ribosomal protein L31	LA_3146	cation transporter
LA_1021	transcription termination factor Rho	LA_3158a	hypothetical protein
LA_1032	enoyl-CoA hydratase	LA_3219	hypothetical protein
LA_1044	cell shape determination protein	LA_3265	hypothetical protein
LA_1047	diaminopimelate decarboxylase	LA_3266	Fe-S-cluster-containing hydrogenase
LA_1101	succinyl-CoA synthetase subunit alpha	LA_3268	hypothetical protein
LA_1102	succinyl-CoA synthetase subunit beta	LA_3285	1-deoxy-D-xylulose-5-phosphate synthase
LA_1105	(dimethylallyl)adenosine tRNA methylthiotransferase	LA_3286	adenylate cyclase
LA_1110	adenylosuccinate synthetase	LA_3289	tryptophan synthase subunit beta
LA_1112	glycerol-3-phosphate dehydrogenase	LA_3290	prolyl-tRNA synthetase
LA_1132	hypothetical protein	LA_3294	undecaprenyl pyrophosphate synthase
LA_1137	hypothetical protein	LA_3296	uridylate kinase

LA_1138	2-methylthioadenine synthetase	LA_3297	elongation factor Ts
LA_1142	preprotein translocase subunit SecD	LA_3302	phosphopantetheinyl transferase
LA_1147	bifunctional 3,4-dihydroxy-2-butanone 4-phosphate synthase/GTP cyclohydrolase II	LA_3349	cytochrome c-type biogenesis protein
LA_1156	sulfate ABC transporter permease	LA_3350	cytochrome c biogenesis protein
LA_1157	sulfate ABC transporter permease	LA_3354	Zn-dependent alcohol dehydrogenase
LA_1158	ABC transporter ATP-binding protein	LA_3363	acyl-CoA dehydrogenase
LA_1189	sigma 54 modulation protein	LA_3378	transcription elongation factor
LA_1205	RNA polymerase ECF-type sigma factor	LA_3380	endoflagellar filament sheath protein
LA_1221	membrane carboxypeptidase/penicillin-binding protein	LA_3406	nucleotide-binding protein
LA_1222	dihydrolipoamide acetyltransferase	LA_3408	30S ribosomal protein S9
LA_1223	dihydrolipoamide dehydrogenase	LA_3409	50S ribosomal protein L13
LA_1224	2-oxoglutarate dehydrogenase subunit E1	LA_3416	30S ribosomal protein S7
LA_1240	threonyl-tRNA synthetase	LA_3417	30S ribosomal protein S12
LA_1242	translation initiation factor IF-3	LA_3419	DNA-directed RNA polymerase subunit beta'
LA_1244	50S ribosomal protein L20	LA_3420	DNA-directed RNA polymerase subunit beta
LA_1253	two-component system response regulator	LA_3423	50S ribosomal protein L1
LA_1255	segregation/condensation protein B	LA_3424	50S ribosomal protein L11
LA_1260	30S ribosomal protein S1	LA_3425	transcription antiterminator
LA_1262	ATP phosphoribosyltransferase	LA_3442	peroxiredoxin
LA_1263	50S ribosomal protein L32	LA_3459	polyphosphate kinase
LA_1264	glycerol-3-phosphate acyltransferase PlsX	LA_3461	FAD/FMN-containing dehydrogenase
LA_1295	hypothetical protein	LA_3484	short-chain dehydrogenase
LA_1302	SsrA-binding protein	LA_3498	phosphate ABC transporter regulatory protein
LA_1309	chromosome segregation protein	LA_3560	selenocysteine lyase
LA_1313	glutamine synthetase	LA_3572	anti-sigma factor antagonist

LA_1323	phosphoribosylformylglycinamidine synthase II	LA_3573	3-deoxy-manno-octulosonate cytidylyltransferase
LA_1325	isoleucyl-tRNA synthetase	LA_3575	flagellar basal body protein FliL
LA_1378	GTP-binding protein BipA	LA_3576	flagellar motor protein MotB
LA_1388	glycyl-tRNA synthetase	LA_3577	endoflagellar motor protein A
LA_1392	ribokinase	LA_3579	glycosyltransferase
LA_1409	serine hydroxymethyltransferase	LA_3596	ATP-dependent Lon protease
LA_1417	lactoylglutathione lyase	LA_3597	pyridoxal phosphate-dependent aminotransferase
LA_1438	antimicrobial peptide ABC transporter ATP-binding protein	LA_3598	ferritin
LA_1447	LexA repressor	LA_3606	hypothetical protein
LA_1451	CDP-diacylglycerol--serine O-phosphatidyltransferase	LA_3607	acyl dehydratase MaoC
LA_1459	UDP-glucose 6-dehydrogenase	LA_3627	acyl-CoA dehydrogenase
LA_1463	5-(carboxyamino)imidazole ribonucleotide mutase	LA_3630	multidrug ABC transporter ATPase/permease
LA_1471	membrane-bound proton-translocating pyrophosphatase	LA_3634	GTP-binding protein
LA_1483	response regulator	LA_3636	cyclic amidohydrolase
LA_1487	tRNA-specific 2-thiouridylase MnmA	LA_3684	ABC-F family ATPase
LA_1488	glycolate oxidase	LA_3694	multidrug ABC transporter ATPase/permease
LA_1511	L-aspartate oxidase	LA_3703	heat shock operon regulator HrcA
LA_1518	mannose-1-phosphate guanylyltransferase	LA_3705	molecular chaperone DnaK
LA_1521	chemotaxis protein CheW	LA_3706	molecular chaperone DnaJ
LA_1532	fructose-bisphosphate aldolase	LA_3709	oxidoreductase
LA_1543	heavy metal efflux pump	LA_3714	leucyl-tRNA synthetase
LA_1678	50S ribosomal protein L9	LA_3747	threonine synthase
LA_1680	aspartyl-tRNA synthetase	LA_3763	valyl-tRNA synthetase
LA_1682	phosphate starvation-inducible protein	LA_3767	peptide chain release factor 2
LA_1692	two-component system response regulator	LA_3783	adenosine deaminase
LA_1696	triosephosphate isomerase	LA_3793	hemolysin
LA_1703	phosphoglycerate kinase	LA_3794	hypothetical protein
LA_1704	glyceraldehyde 3-phosphate dehydrogenase	LA_3819	ATP-dependent Zn protease

LA_1713	carbonic anhydrase/acetyltransferase	LA_3820	peptidyl-tRNA hydrolase
LA_1719	cysteine synthase	LA_3822	ribose-phosphate pyrophosphokinase
LA_1729	aspartate alpha-decarboxylase	LA_3905	pyridoxal phosphate-dependent aminotransferase
LA_1865	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase	LA_3911	hypothetical protein
LA_1888	GTP-binding protein LepA	LA_3918	methionyl-tRNA synthetase
LA_1896	succinate dehydrogenase/fumarate reductase iron-sulfur subunit	LA_3936	phosphopantothenoylcysteine decarboxylase
LA_1897	succinate dehydrogenase flavoprotein subunit	LA_3939	NAD(P)H-dependent glycerol-3-phosphate dehydrogenase
LA_1927	hypothetical protein	LA_3941	phosphoesterase
LA_1929	CAP family transcriptional factor	LA_3961	hypothetical protein
LA_1930	acyl-CoA dehydrogenase	LA_3964	GDP-mannose 4,6-dehydratase
LA_1935	hypothetical protein	LA_3975	acylphosphatase
LA_1948	two-component system response regulator	LA_3978	lysine methyltransferase
LA_1951	phosphopyruvate hydratase	LA_3982	hypothetical protein
LA_1960	preprotein translocase subunit SecA	LA_4034	carbonic anhydrase/acetyltransferase
LA_1977	FAD-binding oxidoreductase	LA_4036	ATP-dependent RNA helicase
LA_1986	inosine-5'-monophosphate dehydrogenase	LA_4138	3-hydroxyacyl-CoA dehydrogenase
LA_1995	lysyl-tRNA synthetase	LA_4139	acetyl-CoA acetyltransferase
LA_2009	pyruvate dehydrogenase subunit beta	LA_4150	phenylalanyl-tRNA synthetase subunit alpha
LA_2010	pyruvate dehydrogenase subunit alpha	LA_4155	ABC transporter permease
LA_2019	flagellin protein	LA_4165	argininosuccinate synthase
LA_2048	phospho-N-acetylmuramoyl-pentapeptide-transferase	LA_4167	phosphopantetheine adenylyltransferase
LA_2057	ABC transporter permease	LA_4193	DNA gyrase subunit A
LA_2061	homoserine O-acetyltransferase	LA_4194	DNA gyrase subunit B

LA_2072	HflC membrane associated protease	LA_4219	sulfate adenylyltransferase subunit 1
LA_2081	flagellar motor switch protein FliM	LA_4220	sulfate adenylyltransferase subunit 2
LA_2081a	ferredoxin family protein	LA_4228	ABC transporter permease
LA_2095	isopropylmalate isomerase large subunit	LA_4229	dipeptide ABC transporter permease
LA_2113	ABC transporter permease	LA_4242	ketol-acid reductoisomerase
LA_2114	ABC transporter ATP-binding protein	LA_4248	purine nucleoside phosphorylase
LA_2118	metal-dependent hydrolase	LA_4254	acetyl-CoA synthetase
LA_2119	glycerol kinase	LA_4255	GTP cyclohydrolase I
LA_2129	DNA/RNA helicase	LA_4308	flagellar hook-associated protein FlgL
LA_2130	aspartate aminotransferase	LA_4309	flagellar hook-associated protein FlgK
LA_2143	biotin synthase	LA_4329	transcriptional coactivator-like protein
LA_2151	two-component system response regulator	LA_4339	seryl-tRNA synthetase
LA_2152	3-isopropylmalate dehydrogenase	LA_4340	Mg-dependent DNase
LA_2154	phosphomannomutase	LA_4359	tRNA uridine 5-carboxymethylaminomethyl modification enzyme GidA
LA_2162	50S ribosomal protein L28	LA_4360	branched-chain amino acid aminotransferase
LA_2172	glycosyltransferase	LB_013	glutamate-1-semialdehyde aminotransferase
LA_2173	lipopolysaccharide core biosynthesis protein	LB_027	chromosome partitioning protein ParB
LA_2179	recombinase A	LB_074	methylmalonyl-CoA mutase
LA_2188	ATPase	LB_082	3-ketoacyl-ACP reductase
LA_2202	2-isopropylmalate synthase	LB_106	adenosylhomocysteinase
LA_2207	hypothetical protein	LB_108	B12-dependent methionine synthase
LA_2212	excinuclease ABC subunit A	LB_114	acetylglutamate kinase
LA_2226	fructose-1,6-bisphosphatase	LB_170	capsular polysaccharide biosynthesis protein
LA_2229	30S ribosomal protein S21	LB_176	adenine phosphoribosyltransferase
LA_2232	RNA polymerase sigma-70 factor	LB_208	alginate o-acetyltransferase
LA_2237	tyrosyl-tRNA synthetase	LB_262	UDP-glucose pyrophosphorylase
LA_2280	endoflagellar biosynthesis chaperone	LB_274	methylmalonyl-CoA mutase

LA_2283	bifunctional phosphoribosylaminoimid azolecarboxamide formyltransferase/IMP cyclohydrolase	LB_286	glutamate synthase
LA_2286	translaldolase	LB_311	spermidine synthase
LA_2297	hypothetical protein	LB_327	aconitate hydratase
LA_2309	long-chain-fatty-acid CoA ligase	LB_333	two-component system response regulator
LA_2318	hypothetical protein	LB_353	pyruvate kinase
LA_2319	ExbD-like biopolymer transport protein	LB_365	chromosome partitioning protein ParA
LA_2323	hypothetical protein		

Table S3.11. GO functional enrichment of proteins with AA changes in P1^{-} and mutations conserved in P1^{+}.

GO categories are organized from most to least represented, considering the ratio between the mutated proteins within a GO process and the total number of proteins within that process in the reference used (*L. interrogans* serovar Lai strain 56601).

GO ID	Description	Observed gene count	Reference gene count	Matching proteins in the network
GO:0006167	AMP biosynthetic process	4	4	adk, purA, purB, apt
GO:0106074	aminoacyl-tRNA metabolism involved in translational fidelity	4	4	ileS, proS, leuS, valS
GO:0006414	Translational elongation	6	8	fusA, tufB, smpB, lepA, tsf, hflX
GO:0006096	Glycolytic process	7	10	gpmI, fbaB, tpiA, pgk, eno, PykF, Pyk2
GO:0044391	Ribosomal subunit	9	14	rplB, rplV, rplN, rpsE, rpsD, rpmF, rpsG, rpsL, rplA
GO:0006099	Tricarboxylic acid cycle	7	11	LA_0671, LA_0790, sucD, sucC, sucA, LA_1896, acnA
GO:0006418	tRNA aminoacylation for protein translation	13	21	trpS, thrS, ileS, grs1, aspS, lysU, tyrS, proS, leuS, valS, metG, pheS, serS
GO:0009060	Aerobic respiration	8	13	cyoB, LA_0671, LA_0790, sucD, sucC, sucA, LA_1896, acnA

GO:0000049	tRNA binding	12	20	rplP, rplE, rpsM, rlmN, thrS, ileS, bipA, trmU, rpsG, rpsL, rplA, pheS
GO:1990204	Oxidoreductase complex	9	15	gcvP, gcvH, nuoE, LA_1112, LA_1222, sucA, guaB, nrdA, LA_3939
GO:0004812	aminoacyl-tRNA ligase activity	13	22	trpS, thrS, ileS, grs1, aspS, lysU, tyrS, proS, leuS, valS, metG, pheS, serS
GO:0016052	Carbohydrate catabolic process	9	16	rpiB, gpmI, fbaB, tpiA, pgk, eno, LA_2119, PykF, Pyk2
GO:0005198	Structural molecule activity	36	65	rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, rpmF, rplI, flaB, rpmB, rpsU, fliS, rplS, rpsP, flgL, LA_2418, LA_2667, rpsI, rplM, rpsG, rpsL, rplA, rplK, LA_4308, FlgK
GO:0003735	Structural constituent of ribosome	29	53	rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, rpmF, rplI, rpmB, rpsU, rplS, rpsP, rpsI, rplM, rpsG, rpsL, rplA, rplK
GO:0006635	Fatty acid beta-oxidation	6	11	LA_0457, LA_0828, LA_1032, LA_3135, fadB, LA_4139
GO:0009201	Ribonucleoside triphosphate biosynthetic process	6	11	pyrG, fliI, atpD, atpG, atpA, pyrH
GO:0006412	Translation	56	103	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, rplI, aspS, lepA, lysU, rpmB, rpsU, tyrS, rplS, rpsP, def, LA_2507, prfC, proS, tsf, rpsI, rplM, rpsG, rpsL, rplA, rplK, hflX, leuS, valS, LA_3767, LA_3820, metG, pheS, serS
GO:0019843	rRNA binding	22	41	rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rlmN, rplT, bipA, rplI, rpsG, rpsL, rplA, rplK
GO:0005840	Ribosome	31	60	rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, LA_1260, rpmF, rplI, rpmB, rpsU, rplS, rpsP, LA_2631, rpsI, rplM, rpsG, rpsL, rplA, rplK

GO:0006417	Regulation of translation	10	20	fusA, tufB, LA_1189, infC, lepA, prfC, tsf, rplA, hflX, LA_3767
GO:0003924	GTPase activity	9	18	fusA, ftsZ, tufB, obg, bipA, lepA, prfC, hflX, cysN
GO:0008135	Translation factor activity, RNA binding	7	14	fusA, tufB, infC, lepA, prfC, tsf, LA_3767
GO:0046034	ATP metabolic process	11	23	gpmI, fbaB, tpiA, pgk, eno, fliI, atpD, atpG, atpA, PykF, Pyk2
GO:0043604	Amide biosynthetic process	60	126	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, rplI, aspS, panD, lepA, lysU, bioB, rpmB, rpsU, tyrS, rplS, rpsP, def, LA_2507, prfC, proS, tsf, rpsI, rplM, rpsG, rpsL, rplA, rplK, hflX, leuS, valS, LA_3767, LA_3820, metG, pheS, acsA, folE, serS
GO:0045333	Cellular respiration	9	19	cyoB, LA_0671, LA_0790, sucD, sucC, sucA, LA_1896, LA_3266, acnA
GO:0072329	Monocarboxylic acid catabolic process	7	15	LA_0457, LA_0828, LA_1032, LA_3135, LA_3936, fadB, LA_4139
GO:0043603	Cellular amide metabolic process	63	136	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, glyA, rplI, aspS, panD, folD, lepA, lysU, bioB, rpmB, rpsU, tyrS, rplS, rpsP, def, LA_2507, prfC, proS, tsf, rpsI, rplM, rpsG, rpsL, rplA, rplK, hflX, leuS, valS, LA_3767, LA_3820, metG, LA_3936, pheS, acsA, folE, serS
GO:1902494	Catalytic complex	23	51	gcvP, gcvH, ribH, uvrB, ruvB, nuoE, sucD, sucC, LA_1112, LA_1222, sucA, eno, guaB, uvrA, hslU, hslV, nrdA, LA_2432, LA_2433, ilvH, ppk, LA_3939, cysN
GO:0006163	Purine nucleotide metabolic process	27	63	gpmI, adk, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, folD, eno, guaB, purH, LA_2318, fliI, atpD, atpG, atpA, PykF, purB, spoT, LA_3936, coaD, acsA, apt, Pyk2
GO:0009150	Purine ribonucleotide	25	59	gpmI, adk, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, eno, purH, LA_2318,

	metabolic process			fliI, atpD, atpG, atpA, PykF, purB, spoT, LA_3936, coaD, acsA, apt, Pyk2
GO:0019693	Ribose phosphate metabolic process	30	71	gpmI, carB, adk, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, eno, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, pyrH, pyrC, prsA, LA_3936, coaD, acsA, apt, Pyk2
GO:0006164	Purine nucleotide biosynthetic process	18	43	adk, coaX, purA, LA_1323, purE, folD, purH, LA_2318, fliI, atpD, atpG, atpA, purB, spoT, LA_3936, coaD, acsA, apt
GO:0072521	Purine-containing compound metabolic process	28	67	gpmI, adk, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, folD, eno, guaB, purH, LA_2318, fliI, atpD, atpG, atpA, PykF, purB, spoT, LA_3936, coaD, mtnP, acsA, apt, Pyk2
GO:0046390	Ribose phosphate biosynthetic process	22	53	carB, adk, coaX, purA, LA_1323, purE, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, purB, spoT, pyrH, pyrC, prsA, LA_3936, coaD, acsA, apt
GO:0009259	Ribonucleotide metabolic process	29	70	gpmI, carB, adk, coaX, purA, LA_1222, LA_1323, purE, fbaB, tpiA, pgk, eno, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, pyrH, pyrC, LA_3936, coaD, acsA, apt, Pyk2
GO:0072522	Purine-containing compound biosynthetic process	19	46	adk, coaX, purA, LA_1323, purE, folD, purH, LA_2318, fliI, atpD, atpG, atpA, purB, spoT, LA_3936, coaD, mtnP, acsA, apt
GO:0043232	Intracellular non-membrane-bounded organelle	35	86	gyrB, LA_0006, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, LA_1260, rpmF, rplI, rpmB, rpsU, rplS, rpsP, LA_2631, rpsI, rplM, rpsG, rpsL, rplA, rplK, LA_4193, LA_4194
GO:0009152	Purine ribonucleotide biosynthetic process	17	42	adk, coaX, purA, LA_1323, purE, purH, LA_2318, fliI, atpD, atpG, atpA, purB, spoT, LA_3936, coaD, acsA, apt
GO:0140101	Catalytic activity, acting on a tRNA	19	47	gatB, trpS, rlmN, thrS, ileS, grs1, aspS, lysU, tyrS, trmD, LA_2507, tgt, proS, leuS, valS, LA_3820, metG, pheS, serS

GO:0009260	Ribonucleotide biosynthetic process	21	52	carB, adk, coaX, purA, LA_1323, purE, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, purB, spoT, pyrH, pyrC, LA_3936, coaD, acsA, apt
GO:0097588	Archaeal or bacterial-type flagellum-dependent cell motility	14	35	FlgB, flaB, fliM, LA_2151, fliS, flgL, LA_2418, fliF, LA_2667, flgE, LA_3575, LA_3576, LA_3577, LA_4308
GO:0009069	Serine family amino acid metabolic process	8	20	gcvP, gcvH, gcvT, glyA, LA_1719, sseA, CsdB, serS
GO:0019842	Vitamin binding	14	35	dapL, sucA, glyA, LA_2130, IlvB, icmF, LA_3052, dxs, CsdB, thrC, hemL, LB_074, metH, Mcm3
GO:0051246	Regulation of protein metabolic process	11	28	fusA, tufB, LA_1189, infC, lepA, LA_2072, prfC, tsf, rplA, hflX, LA_3767
GO:0071973	Bacterial-type flagellum-dependent cell motility	11	28	FlgB, flaB, fliM, fliS, flgL, LA_2418, fliF, LA_2667, flgE, LA_3575, LA_4308
GO:0006351	Transcription, DNA-templated	9	23	rpoA, nusA, rho, LA_1205, rpoD, LA_2603, rpoC, rpoB, nusG
GO:0043228	Non-membrane-bounded organelle	46	118	gyrB, LA_0006, FlgB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, LA_1260, rpmF, rplI, flaB, fliM, rpmB, rpsU, rplS, rpsP, flgL, LA_2418, fliF, LA_2631, LA_2667, flgE, rpsI, rplM, rpsG, rpsL, rplA, rplK, LA_3575, LA_4193, LA_4194, LA_4308, FlgK
GO:0046395	Carboxylic acid catabolic process	10	26	gcvP, gcvH, gcvT, LA_0457, LA_0828, LA_1032, LA_3135, LA_3936, fadB, LA_4139
GO:0010467	Gene expression	74	195	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpoA, rlmN, rpmA, nusA, pnp, rpmE, rho, LA_1105, rimO, LA_1205, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, trmU, rplI, aspS, lepA, lysU, rpmB, rpsU, rpoD, tyrS, rplS, trmD, rpsP, def, LA_2507, LA_2603, prfC, LA_2917, tgt,

				proS, tsf, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, hflX, leuS, valS, LA_3767, LA_3820, metG, pheS, serS, gidA
GO:0006399	tRNA metabolic process	21	56	trpS, rlmN, LA_1105, rimO, thrS, ileS, grs1, trmU, aspS, lysU, tyrS, trmD, LA_2917, tgt, proS, leuS, valS, metG, pheS, serS, gidA
GO:0006631	Fatty acid metabolic process	12	32	LA_0106, LA_0457, LA_0828, LA_1032, plsX, LA_2309, LA_2432, LA_3135, acpS, fadB, LA_4139, acnA
GO:0017111	Nucleoside-triphosphatase activity	42	112	gyrB, LA_0006, LA_0273, LA_0274, fusA, ftsZ, uvrB, tufB, ruvB, obg, LA_0969, rho, cysA, LA_1295, bipA, LA_1438, lepA, fliM, LA_2114, recA, LA_2188, uvrA, ychF, hslU, yhbG, mfd, fliH, fliF, fliI, LA_2605, prfC, radA, lon, LA_3630, hflX, LA_3684, LA_3694, hflB, LA_4036, LA_4193, LA_4194, cysN
GO:0006091	Generation of precursor metabolites and energy	23	62	cyoB, etfA, etfB, gpmI, LA_0666, LA_0671, LA_0790, sucD, sucC, LA_1222, sucA, fbaB, tpiA, pgk, LA_1896, eno, mipB, rpe, PykF, LA_3265, LA_3266, acnA, Pyk2
GO:0016887	ATPase activity	30	82	gyrB, LA_0006, LA_0273, LA_0274, uvrB, ruvB, LA_0969, rho, cysA, LA_1295, LA_1438, LA_2114, recA, LA_2188, uvrA, ychF, hslU, yhbG, mfd, fliI, LA_2605, radA, lon, LA_3630, LA_3684, LA_3694, hflB, LA_4036, LA_4193, LA_4194
GO:0032549	Ribonucleoside binding	16	44	fusA, ftsZ, tufB, obg, purA, ribB, bipA, LA_1483, manC, lepA, ychF, prfC, rpoB, hflX, cysN, folE
GO:0009165	Nucleotide biosynthetic process	25	70	nadA, carB, adk, coaX, purA, LA_1323, purE, nadB, folD, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, purB, spoT, pyrH, pyrC, prsA, LA_3936, coaD, acsA, apt
GO:0009156	Ribonucleoside monophosphate biosynthetic process	10	28	carB, adk, purA, LA_1323, purE, purH, purB, pyrC, prsA, apt
GO:0032774	RNA biosynthetic process	10	28	rpoA, nusA, rho, LA_1205, rpoD, LA_2603, rpoC, rpoB, nusG, serS
GO:0003723	RNA binding	40	114	fusA, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rlmN, nusA, pnp, rho, thrS, infC,

				rplT, LA_1260, smpB, ileS, bipA, trmU, rplI, lepA, tyrS, prfC, tsf, rpsG, rpsL, rplA, rplK, LA_3767, pheS, acnA
GO:0009117	Nucleotide metabolic process	35	100	gpmI, nadA, carB, adk, coaX, purA, LA_1222, LA_1323, purE, nadB, fbaB, tpiA, pgk, folD, eno, guaB, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, pyrH, pyrC, add, prsA, LA_3936, coaD, acsA, apt, Pyk2
GO:0000287	Magnesium ion binding	22	63	obg, pnp, sucC, purA, LA_1323, ovp1, eno, lysU, LA_2152, LA_2154, fbp, hprK, IlvB, PykF, dxs, uppS, acpS, rpoC, prsA, pheS, ilvC, Pyk2
GO:0016874	Ligase activity	30	86	gatB, trpS, carB, sucD, sucC, purA, thrS, glnA, LA_1323, ileS, grs1, aspS, lysU, tyrS, pyrG, LA_2432, LA_2433, LA_2507, atpD, atpG, atpA, proS, leuS, valS, metG, LA_3936, pheS, argG, acsA, serS
GO:0005525	GTP binding	15	43	fusA, ftsZ, tufB, obg, purA, ribB, bipA, LA_1483, manC, lepA, ychF, prfC, hflX, cysN, folE
GO:0009288	Bacterial-type flagellum	11	32	FlgB, flaB, fliM, flgL, LA_2418, fliF, LA_2667, flgE, LA_3575, LA_4308, FlgK
GO:0006520	Cellular amino acid metabolic process	46	134	gcvP, gcvH, gcvT, trpS, lysC, carB, dapL, proA, lysA, thrS, hisG, glnA, ileS, grs1, glyA, aspS, LA_1719, panD, folD, lysU, met2, leuC, LA_2130, LA_2152, tyrS, pyrG, ilvH, argF, sseA, LA_3052, trpB, proS, CsdB, leuS, thrC, valS, metG, pheS, argG, ilvC, mtnP, serS, ilvE, metH, argB, gltB
GO:0016462	Pyrophosphatase activity	43	127	gyrB, LA_0006, LA_0273, LA_0274, fusA, ftsZ, uvrB, tufB, ruvB, obg, LA_0969, rho, cysA, LA_1295, bipA, LA_1438, ovp1, lepA, fliM, LA_2114, recA, LA_2188, uvrA, ychF, hslU, yhbG, mfd, fliH, fliF, fliI, LA_2605, prfC, radA, lon, LA_3630, hflX, LA_3684, LA_3694, hflB, LA_4036, LA_4193, LA_4194, cysN
GO:0055086	Nucleobase-containing small molecule metabolic process	45	133	LA_0332, gpmI, nadA, carB, adk, coaX, purA, LA_1222, LA_1323, ugd, purE, nadB, manC, fbaB, tpiA, pgk, folD, eno, guaB, LA_2207, purH, LA_2318, pyrG, fliI, LA_2749, atpD, atpG, atpA, PykF, purB, spoT, tgt, pyrH, LA_3573, pyrC, add, prsA, LA_3936, gmd, coaD, mtnP, acsA, apt, GalF, Pyk2

GO:0044271	Cellular nitrogen compound biosynthetic process	110	334	gatB, dnaE, fusA, LA_0332, trpS, ribH, nadA, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, coaX, rpmA, nusA, rpmE, rho, purA, ribB, LA_1205, thrS, infC, rplT, LA_1260, rpmF, smpB, LA_1323, ileS, grs1, ugd, purE, nadB, manC, rplI, aspS, panD, folD, lepA, lysU, LA_2118, bioB, rpmB, LA_2207, rpsU, rpoD, tyrS, purH, LA_2318, rplS, rpsP, pyrG, def, LA_2507, fliI, LA_2603, prfC, LA_2749, atpD, atpG, atpA, purB, spoT, tgt, dxs, trpB, proS, pyrH, tsf, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, LA_3573, hflX, pyrC, leuS, valS, LA_3767, LA_3820, prsA, metG, LA_3936, gmd, pheS, coaD, mtnP, acsA, folE, serS, hemL, apt, speE
GO:0016818	Hydrolase activity, acting on acid anhydrides, in phosphorus-containing anhydrides	44	135	gyrB, LA_0006, LA_0273, LA_0274, fusA, ftsZ, uvrB, tufB, ruvB, obg, LA_0969, rho, cysA, LA_1295, bipA, LA_1438, ovp1, lepA, fliM, LA_2114, recA, LA_2188, uvrA, ychF, hslU, yhbG, mfd, fliH, fliF, fliI, LA_2605, prfC, radA, lon, LA_3630, hflX, LA_3684, LA_3694, hflB, acyP, LA_4036, LA_4193, LA_4194, cysN
GO:0034645	Cellular macromolecule biosynthetic process	83	255	gatB, dnaE, fusA, trpS, mltG, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpoA, rpmA, nusA, rpmE, rho, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, rpmF, smpB, ileS, grs1, lexA, manC, rplI, aspS, lepA, lysU, rfe, rpmB, LA_2173, rpsU, rpoD, tyrS, nrdA, rplS, rpsP, def, LA_2507, LA_2603, prfC, rodA, LA_2755, lgt, proS, tsf, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, LA_3573, LA_3597, hflX, dnaJ, leuS, valS, LA_3767, LA_3820, LA_3905, metG, pheS, serS, LB_170, LB_208
GO:0032991	Protein-containing complex	44	136	gcvP, gcvH, ribH, uvrB, rplB, rplV, rplN, rpsE, rpsD, ruvB, nuoE, LA_0970, sucD, sucC, LA_1112, cysA, LA_1222, sucA, rpmF, lexA, eno, guaB, ttg2B, LA_2113, uvrA, hslU, hslV, nrdA, gspD, yhbG, LA_2432, LA_2433, LA_2507, ilvH, fliI, atpD, atpG, atpA, rpsG, rpsL, rplA, ppk, LA_3939, cysN
GO:0032787	Monocarboxylic acid metabolic process	26	81	LA_0106, gpmI, LA_0457, LA_0828, LA_1032, plsX, fbaB, tpiA, pgk, panD, eno, bioB, LA_2309, LA_2432, LA_2627, PykF, LA_3135, acpS, LA_3606, LA_3607, LA_3936, fadB, LA_4139, acsA, acnA, Pyk2

GO:0043436	Oxoacid metabolic process	80	250	LA_0106, gcvP, gcvH, gcvT, trpS, gpmI, LA_0457, nadA, lysC, carB, dapL, LA_0828, proA, LA_1032, lysA, LA_1222, thrS, hisG, plsX, glnA, ileS, grs1, glyA, ugd, fbaB, aspS, tpiA, pgk, LA_1719, panD, folD, eno, lysU, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, tyrS, LA_2309, pyrG, LA_2432, LA_2433, ilvH, IlvB, LA_2627, argF, PykF, sseA, LA_3052, LA_3135, LA_3268, trpB, proS, acpS, ppk, CsdB, LA_3606, LA_3607, leuS, thrC, valS, metG, LA_3936, fadB, LA_4139, pheS, argG, ilvC, mtnP, acsA, folE, serS, ilvE, metH, argB, gltB, acnA, Pyk2
GO:0009059	Macromolecule biosynthetic process	84	264	gatB, dnaE, fusA, trpS, mltG, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpoA, rpmA, nusA, rpmE, rho, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, rpmF, smpB, ileS, grs1, lexA, ugd, manC, rplI, aspS, lepA, lysU, rfe, rpmB, LA_2173, rpsU, rpoD, tyrS, nrdA, rplS, rpsP, def, LA_2507, LA_2603, prfC, rodA, LA_2755, lgt, proS, tsf, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, LA_3573, LA_3597, hflX, dnaJ, leuS, valS, LA_3767, LA_3820, LA_3905, metG, pheS, serS, LB_170, LB_208
GO:1901605	Alpha-amino acid metabolic process	34	107	gcvP, gcvH, gcvT, lysC, carB, dapL, proA, lysA, hisG, glnA, glyA, LA_1719, panD, folD, met2, leuC, LA_2130, LA_2152, pyrG, ilvH, argF, sseA, LA_3052, trpB, CsdB, thrC, argG, ilvC, mtnP, serS, ilvE, metH, argB, gltB
GO:1901566	Organonitrogen compound biosynthetic process	123	390	gatB, fusA, LA_0332, trpS, ribH, mltG, nadA, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, dapL, coaX, rpmA, proA, rpmE, lysA, purA, ribB, LA_1221, thrS, infC, rplT, LA_1260, hisG, rpmF, smpB, glnA, LA_1323, ileS, grs1, glyA, purE, nadB, rplI, aspS, LA_1719, panD, folD, lepA, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, LA_2207, rpsU, tyrS, purH, LA_2318, rplS, rpsP, pyrG, def, LA_2507, ilvH, fliI, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, argF, lgt, LA_3052, purB, spoT, tgt, dxs, trpB, proS, pyrH, tsf, rpsI, rplM, rpsG, rpsL, rplA, rplK, hflX, pyrC, leuS, thrC, valS, LA_3767, LA_3820, metG, LA_3936, pheS, argG,

				coaD, ilvC, mtnP, acsA, folE, serS, ilvE, hemL, metH, argB, apt, gltB, speE
GO:0034654	Nucleobase-containing compound biosynthetic process	45	143	dnaE, LA_0332, nadA, carB, adk, rpoA, coaX, nusA, rho, purA, LA_1205, LA_1323, ugd, purE, nadB, manC, folD, LA_2207, rpoD, purH, LA_2318, pyrG, fliI, LA_2603, LA_2749, atpD, atpG, atpA, purB, spoT, tgt, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, prsA, LA_3936, gmd, coaD, mtnP, acsA, serS, apt
GO:0019752	Carboxylic acid metabolic process	78	248	LA_0106, gcvP, gcvH, gcvT, trpS, gpmI, LA_0457, nadA, lysC, carB, dapL, LA_0828, proA, LA_1032, lysA, LA_1222, thrS, hisG, plsX, glnA, ileS, grs1, glyA, ugd, fbaB, aspS, tpiA, pgk, LA_1719, panD, folD, eno, lysU, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, tyrS, LA_2309, pyrG, LA_2432, LA_2433, ilvH, IlvB, LA_2627, argF, PykF, sseA, LA_3052, LA_3135, trpB, proS, acpS, CsdB, LA_3606, LA_3607, leuS, thrC, valS, metG, LA_3936, fadB, LA_4139, pheS, argG, ilvC, mtnP, acsA, folE, serS, ilvE, metH, argB, gltB, acnA, Pyk2
GO:0009123	Nucleoside monophosphate metabolic process	11	35	carB, adk, purA, LA_1323, purE, purH, purB, pyrH, pyrC, prsA, apt
GO:0006082	Organic acid metabolic process	81	261	LA_0106, gcvP, gcvH, gcvT, trpS, gpmI, LA_0457, nadA, lysC, carB, dapL, LA_0828, proA, LA_1032, lysA, LA_1222, thrS, hisG, plsX, glnA, ileS, grs1, glyA, ugd, fbaB, aspS, tpiA, pgk, LA_1719, panD, folD, eno, lysU, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, tyrS, LA_2309, pyrG, LA_2432, LA_2433, ilvH, IlvB, LA_2627, argF, PykF, sseA, LA_3052, LA_3135, LA_3268, trpB, proS, acpS, ppk, CsdB, LA_3606, LA_3607, leuS, thrC, valS, metG, LA_3936, fadB, LA_4139, pheS, argG, ilvC, mtnP, acsA, folE, serS, ilvE, metH, argB, LB_208, gltB, acnA, Pyk2
GO:1901607	Alpha-amino acid biosynthetic process	27	87	lysC, carB, dapL, proA, hisG, glnA, glyA, LA_1719, panD, folD, met2, leuC, LA_2130, LA_2152, ilvH, argF, LA_3052, trpB, thrC, argG, ilvC, mtnP, serS, ilvE, metH, argB, gltB
GO:0043226	Organelle	53	177	gyrB, LA_0006, LA_0332, FlgB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT,

				LA_1260, rpmF, LA_1295, rplI, acoA, flaB, fliM, rpmB, rpsU, ubiX, rplS, rpsP, flgL, LA_2418, fliF, LA_2631, LA_2667, LA_2749, flgE, rpsI, rplM, rpsG, rpsL, rplA, rplK, LA_3575, LA_3627, LA_3978, LA_4193, LA_4194, LA_4308, FlgK
GO:0043229	Intracellular organelle	42	142	gyrB, LA_0006, LA_0332, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, rplT, LA_1260, rpmF, LA_1295, rplI, acoA, rpmB, rpsU, ubiX, rplS, rpsP, LA_2631, LA_2749, rpsI, rplM, rpsG, rpsL, rplA, rplK, LA_3627, LA_3978, LA_4193, LA_4194
GO:0036094	Small molecule binding	153	524	gyrB, LA_0006, pckA, gatB, LA_0273, LA_0274, fusA, LA_0334, pntB, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, lysC, carB, tufB, adk, dapL, ruvB, coaX, obg, proA, nuoF, nuoD, LA_0969, rho, sucD, sucC, purA, LA_1112, ribB, cysA, LA_1221, LA_1223, sucA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1438, ugd, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, gapA, lepA, sdhA, LA_1930, secA, LA_1977, lysU, LA_2114, LA_2119, ssl2, LA_2130, LA_2152, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, nrdA, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, IlvB, fliI, LA_2605, groL, LA_2727, prfC, LA_2755, mreB, atpD, atpG, atpA, argF, radA, LA_2917, PykF, icmF, LA_3052, LA_3143, dxs, proS, pyrH, LA_3363, LA_3406, rpoB, ppk, glcD, CsdB, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, thrC, valS, LA_3794, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, ilvC, acsA, folE, serS, gidA, hemL, LB_074, metH, argB, Mcm3, Pyk2, ParA4
GO:0019637	Organophosphate metabolic process	45	157	gpmI, nadA, carB, adk, coaX, purA, LA_1112, LA_1222, plsX, LA_1323, pssA, purE, nadB, fbaB, tpiA, pgk, gapA, folD, eno, guaB, LA_2119, purH, mipB, LA_2318, rpe, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, dxs, pyrH, pyrC, add, prsA, LA_3936, LA_3939, coaD, acsA, folE, apt, Pyk2
GO:0044267	Cellular protein metabolic process	63	220	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, rplI, aspS, lepA, lysU, LA_2118, rpmB, rpsU,

				tyrS, hslV, rplS, rpsP, hprK, def, map, LA_2507, prfC, proS, tsf, ccmE, rpsI, rplM, rpsG, rpsL, rplA, rplK, lon, hflX, leuS, valS, LA_3767, LA_3820, metG, LA_3978, pheS, serS
GO:0043168	Anion binding	133	466	gyrB, LA_0006, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, lysC, carB, tufB, adk, dapL, ruvB, coaX, obg, nuoF, LA_0969, rho, sucC, purA, ribB, cysA, LA_1221, LA_1223, sucA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1438, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, lepA, sdhA, LA_1930, secA, LA_1977, lysU, LA_2114, LA_2119, ssl2, LA_2130, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, IlvB, fliI, LA_2605, groL, prfC, LA_2755, mreB, atpD, atpG, atpA, argF, radA, PykF, LA_3052, LA_3143, dxs, proS, pyrH, LA_3363, ppk, glcD, CsdB, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, thrC, valS, LA_3794, hflB, prsA, metG, LA_3936, gmd, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, acsA, folE, serS, gidA, hemL, argB, Pyk2, ParA4
GO:0005515	Protein binding	17	60	LA_0334, rpoA, LA_1253, scpB, grs1, LA_1521, LA_1692, gspD, atoC, clpX, LA_2631, groL, greA, LA_3498, LA_3576, dnaK, dnaJ
GO:0000166	Nucleotide binding	134	474	gyrB, LA_0006, pckA, gatB, LA_0273, LA_0274, fusA, LA_0334, pntB, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, lysC, carB, tufB, adk, ruvB, coaX, obg, proA, nuoF, nuoD, LA_0969, rho, sucD, sucC, purA, LA_1112, ribB, cysA, LA_1223, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, ugd, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, gapA, lepA, sdhA, LA_1930, secA, LA_1977, lysU, LA_2114, LA_2119, ssl2, LA_2152, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, nrdA, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groL, LA_2727, prfC, mreB, atpD, atpG, atpA, radA, LA_2917, PykF, LA_3143, proS, pyrH, LA_3363, LA_3406, ppk, glcD, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_4036,

				pheS, argG, coaD, LA_4193, LA_4194, cysN, ilvC, acsA, folE, serS, gidA, argB, Pyk2, ParA4
GO:0044282	Small molecule catabolic process	13	46	gcvP, gcvH, gcvT, rpiB, gpmI, LA_0457, LA_0828, LA_1032, LA_2119, LA_3135, LA_3936, fadB, LA_4139
GO:0044281	Small molecule metabolic process	129	459	LA_0106, pckA, LA_0319, LA_0332, gcvP, gcvH, gcvT, rpiB, trpS, gpmI, LA_0457, ribH, nadA, lysC, carB, adk, dapL, LA_0828, coaX, proA, mgsA, LA_1032, lysA, purA, ribB, LA_1222, thrS, hisG, plsX, glnA, LA_1323, ileS, grs1, glyA, ugd, purE, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, LA_1719, panD, folD, eno, guaB, lysU, acoB, acoA, met2, leuC, LA_2118, LA_2119, LA_2130, bioB, LA_2152, LA_2207, fbp, tyrS, purH, LA_2309, LA_2318, pyrG, LA_2432, LA_2433, ilvH, IlvB, fliI, LA_2627, LA_2712, LA_2749, atpD, atpG, atpA, argF, PykF, sseA, icmF, LA_3052, purB, spoT, tgt, LA_3135, LA_3268, dxs, trpB, proS, pyrH, acpS, LA_3354, ppk, CsdB, LA_3573, LA_3606, LA_3607, LA_3627, pyrC, leuS, thrC, valS, add, prsA, metG, LA_3936, gmd, fadB, LA_4139, pheS, argG, coaD, ilvC, mtnP, acsA, folE, serS, ilvE, LB_074, sam1, metH, argB, apt, LB_208, GalF, Mcm3, gltB, acnA, Pyk2
GO:0040011	Locomotion	16	57	FlgB, LA_1521, flaB, fliM, LA_2151, fliS, flgL, LA_2418, LA_2469, fliF, LA_2667, flgE, LA_3575, LA_3576, LA_3577, LA_4308
GO:0032553	Ribonucleotide binding	106	382	gyrB, LA_0006, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, ftsZ, uvrB, lysC, carB, tufB, adk, ruvB, coaX, obg, nuoF, LA_0969, rho, sucC, purA, ribB, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, lepA, secA, lysU, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groL, prfC, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, ppk, lon, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_3936, LA_4036, pheS, argG, coaD,

				LA_4193, LA_4194, cysN, acsA, folE, serS, argB, Pyk2, ParA4
GO:0032555	Purine ribonucleotide binding	103	375	gyrB, LA_0006, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, ftsZ, uvrB, lysC, carB, tufB, adk, ruvB, coaX, obg, LA_0969, rho, sucC, purA, ribB, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, LA_1483, trmU, manC, aspS, phoH, LA_1692, pgk, lepA, secA, lysU, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groL, prfC, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, ppk, lon, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, acsA, folE, serS, argB, Pyk2, ParA4
GO:0035639	Purine ribonucleoside triphosphate binding	103	375	gyrB, LA_0006, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, ftsZ, uvrB, lysC, carB, tufB, adk, ruvB, coaX, obg, LA_0969, rho, sucC, purA, ribB, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, LA_1483, trmU, manC, aspS, phoH, LA_1692, pgk, lepA, secA, lysU, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groL, prfC, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, ppk, lon, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, acsA, folE, serS, argB, Pyk2, ParA4
GO:0006996	Organelle organization	14	51	gyrB, LA_0006, ruvB, rplT, fliS, mfd, fliH, fliI, flhA, hflX, LA_3978, LA_4193, LA_4194, FlgK
GO:0097367	Carbohydrate derivative binding	107	394	gyrB, LA_0006, pckA, gatB, LA_0273, LA_0274, fusA, LA_0391, trpS, ftsZ, uvrB, lysC, carB, tufB, adk, ruvB, coaX, obg, nuoF, LA_0969, rho, sucC, purA, ribB, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, LA_1438, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, lepA, secA, lysU, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI,

				LA_2605, groL, prfC, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, rpoB, ppk, lon, LA_3630, hflX, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_3936, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, acsA, folE, serS, argB, Pyk2, ParA4
GO:0090407	Organophosphate biosynthetic process	31	115	nadA, carB, adk, coaX, purA, LA_1112, plsX, LA_1323, pssA, purE, nadB, folD, purH, LA_2318, pyrG, fliI, atpD, atpG, atpA, purB, spoT, dxs, pyrH, pyrC, prsA, LA_3936, LA_3939, coaD, acsA, folE, apt
GO:0005524	ATP binding	90	335	gyrB, LA_0006, pckA, gatB, LA_0273, LA_0274, LA_0391, trpS, uvrB, lysC, carB, adk, ruvB, coaX, LA_0969, rho, sucC, cysA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, grs1, rbsK, LA_1438, trmU, aspS, phoH, LA_1692, pgk, secA, lysU, LA_2114, LA_2119, ssl2, recA, LA_2188, uvrA, tyrS, LA_2318, ychF, mrp, hslU, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, fliI, LA_2605, groL, mreB, atpD, atpG, atpA, radA, PykF, proS, pyrH, ppk, lon, LA_3630, LA_3684, LA_3694, dnaK, dnaJ, leuS, valS, hflB, prsA, metG, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, acsA, serS, argB, Pyk2, ParA4
GO:0005829	Cytosol	51	194	gcvP, gcvT, ruvB, LA_0845, mgsA, sucD, sucC, ribB, sucA, purE, LA_1713, folD, eno, leuC, LA_2154, LA_2173, LA_2202, fliS, purH, hslU, hslV, LA_2350, nrdA, rpe, hprK, pyrG, def, ilvH, LA_2605, bfr, LA_2727, sufB, radA, PykF, eutG, purB, tgt, dxs, uppS, LA_3406, rpoB, nusG, add, LA_4034, cysN, ilvC, ilvE, metH, gltB, acnA, Pyk2
GO:0140098	Catalytic activity, acting on RNA	23	89	gatB, trpS, rpoA, rlmN, thrS, ileS, grs1, aspS, lysU, tyrS, trmD, LA_2507, LA_2917, tgt, proS, rpoC, rpoB, leuS, valS, LA_3820, metG, pheS, serS
GO:0044249	Cellular biosynthetic process	163	638	gatB, dnaE, fusA, LA_0319, LA_0332, trpS, ribH, mltG, nadA, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, coaX, rpmA, proA, mgsA, nusA, rpmE, rho, lysA, purA, LA_1112, ribB, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, smpB, glnA, LA_1323, ileS, grs1, glyA, lexA, pssA, ugd, purE, nadB, manC, rplI, aspS, LA_1719, panD, folD, lepA, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB,

				LA_2172, LA_2173, LA_2207, rpsU, rpoD, tyrS, purH, LA_2318, LA_2326, nrdA, rplS, rpsP, pyrG, LA_2432, def, LA_2507, ilvH, fliI, LA_2603, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, LA_2821, argF, lgt, LA_3052, purB, spoT, tgt, dxs, trpB, proS, uppS, pyrH, tsf, acpS, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, ppk, LA_3573, LA_3579, LA_3597, hflX, pyrC, dnaJ, leuS, thrC, valS, LA_3767, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, pheS, argG, coaD, cysN, cysD, ilvC, mtnP, acsA, folE, serS, ilvE, hemL, metH, argB, LB_170, apt, LB_208, gltB, speE
GO:0055114	Oxidation-reduction process	70	274	LA_0020, cyoB, LA_0334, pntB, gcvP, etfA, etfB, ivd, LA_0457, LA_0560, LA_0666, LA_0671, LA_0790, LA_0828, proA, nuoH, nuoF, nuoE, nuoD, nuoB, nuoA, LA_1032, sucD, sucC, LA_1112, LA_1223, sucA, ugd, lldD, nadB, gapA, folD, LA_1896, sdhA, LA_1930, LA_1977, guaB, acoB, acoA, LA_2152, mipB, LA_2324, nrdA, rpe, trxB, bfr, LA_2727, ahpC, eutG, LA_3135, LA_3143, LA_3265, LA_3266, LA_3268, LA_3354, LA_3363, Bcp, glcD, LA_3484, dps, LA_3627, mviM, LA_3939, fadB, LA_4139, cysD, ilvC, LB_082, gltB, acnA
GO:0046394	Carboxylic acid biosynthetic process	35	137	nadA, lysC, carB, dapL, proA, hisG, plsX, glnA, glyA, ugd, LA_1719, panD, folD, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, LA_2432, ilvH, argF, LA_3052, trpB, acpS, thrC, argG, ilvC, mtnP, folE, serS, ilvE, metH, argB, gltB
GO:0005622	Intracellular	258	1017	gyrB, LA_0006, LA_0020, pckA, dnaE, LA_0280, LA_0309, fusA, LA_0332, gcvP, gcvH, gcvT, LA_0391, trpS, etfB, gpmI, LA_0457, ftsZ, nadA, uvrB, LA_0671, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, rpmA, obg, proA, mgsA, nusA, pnp, LA_1019, rpmE, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, ribB, LA_1189, LA_1222, LA_1223, sucA, thrS, infC, rplT, LA_1253, scpB, LA_1260, hisG, rpmF, plsX, LA_1295, smpB, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, ugd, purE, trmU, nadB, LA_1521, fbaB, rplI, aspS, phoH, LA_1692, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, LA_1929, eno, secA, lysU, acoA, met2, leuC, LA_2118, LA_2130, bioB, LA_2151,

				LA_2152, LA_2154, rpmB, LA_2173, recA, LA_2188, LA_2202, uvrA, fbp, rpsU, rpoD, tyrS, fliS, purH, mipB, LA_2297, LA_2324, ychF, ubiX, hslU, hslV, xerC, LA_2350, nrdA, rplS, trmD, rpsP, rpe, atoC, hprK, pyrG, LA_2432, LA_2433, def, map, trxB, LA_2507, mfd, LA_2559, ilvH, IlvB, fliH, fliI, LA_2603, LA_2605, LA_2631, groL, bfr, LA_2727, prfC, LA_2749, mreB, atpD, atpG, atpA, sufB, ahpC, LA_2821, argF, LA_2894, radA, LA_2917, PykF, eutG, LA_3052, purB, LA_3094, tgt, LA_3135, caiB, dxs, trpB, proS, uppS, pyrH, tsf, acpS, LA_3354, LA_3406, rpsI, rplM, rpsG, rpsL, rpoB, rplA, rplK, nusG, Bcp, LA_3484, LA_3498, CsdB, LA_3573, lon, dps, LA_3627, hflX, pyrC, LA_3684, dnaK, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3820, prsA, metG, LA_3939, gmd, LA_3978, LA_4034, LA_4036, LA_4139, pheS, argG, coaD, LA_4193, LA_4194, cysN, cysD, ilvC, mtnP, acsA, folE, serS, tatD, gidA, ilvE, hemL, ParB3, LB_082, sam1, metH, argB, apt, GalF, gltB, speE, acnA, LB_333, Pyk2, ParA4
GO:1901576	Organic substance biosynthetic process	164	648	pckA, gatB, dnaE, fusA, LA_0319, LA_0332, trpS, ribH, mltG, nadA, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, coaX, rpmA, proA, mgsA, nusA, rpmE, rho, lysA, purA, LA_1112, ribB, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, smpB, glnA, LA_1323, ileS, grs1, glyA, lexA, pssA, ugd, purE, nadB, manC, rplI, aspS, tpiA, LA_1719, panD, folD, lepA, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, LA_2172, LA_2173, LA_2207, fbp, rpsU, rpoD, tyrS, purH, LA_2318, LA_2326, nrdA, rplS, rpsP, pyrG, LA_2432, def, LA_2507, ilvH, fliI, LA_2603, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, LA_2821, argF, lgt, LA_3052, purB, spoT, tgt, dxs, trpB, proS, uppS, pyrH, tsf, acpS, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, ppk, LA_3573, LA_3579, LA_3597, hflX, pyrC, dnaJ, leuS, thrC, valS, LA_3767, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, pheS, argG, coaD, ilvC, mtnP, acsA, folE, serS, ilvE, hemL, metH, argB, LB_170, apt, LB_208, gltB, speE

GO:0010468	Regulation of gene expression	32	128	fusA, LA_0395, LA_0653, LA_0720, tufB, nusA, pnp, LA_1189, LA_1205, infC, LA_1253, lexA, LA_1692, lepA, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3094, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, LB_333
GO:0046914	Transition metal ion binding	18	72	gpmI, nuoB, LA_1019, ribB, ileS, bioB, uvrA, clpX, bfr, LA_3268, LA_3354, rpoC, dps, pyrC, dnaJ, hflB, folE, metH
GO:0009058	Biosynthetic process	166	672	pckA, gatB, dnaE, fusA, LA_0319, LA_0332, trpS, ribH, mltG, nadA, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, coaX, rpmA, proA, mgsA, nusA, rpmE, rho, lysA, purA, LA_1112, ribB, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, smpB, glnA, LA_1323, ileS, grs1, glyA, lexA, pssA, ugd, purE, nadB, manC, rplI, aspS, tpiA, LA_1719, panD, folD, lepA, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, LA_2172, LA_2173, LA_2207, fbp, rpsU, rpoD, tyrS, purH, LA_2318, LA_2326, nrdA, rplS, rpsP, pyrG, LA_2432, def, LA_2507, ilvH, fliI, LA_2603, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, LA_2821, argF, lgt, LA_3052, purB, spoT, tgt, dxs, trpB, proS, uppS, pyrH, tsf, acpS, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, ppk, LA_3573, LA_3579, LA_3597, hflX, pyrC, dnaJ, leuS, thrC, valS, LA_3767, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, pheS, argG, coaD, cysN, cysD, ilvC, mtnP, acsA, folE, serS, ilvE, hemL, metH, argB, LB_170, apt, LB_208, gltB, speE
GO:2000112	Regulation of cellular macromolecule biosynthetic process	31	126	fusA, LA_0395, LA_0653, LA_0720, tufB, nusA, LA_1189, LA_1205, infC, LA_1253, lexA, LA_1692, lepA, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3094, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, LB_333
GO:0005737	Cytoplasm	220	898	gyrB, LA_0006, LA_0020, pckA, dnaE, LA_0280, LA_0309, fusA, LA_0332, gcvP, gcvH, gcvT, LA_0391, trpS, etfB, gpmI, LA_0457, ftsZ, nadA, uvrB, LA_0671, lysC, LA_0720, carB, tufB, adk, rpoA, dapL, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, obg, proA, mgsA, nusA, pnp, LA_1019, lysA, sucD, sucC, LA_1105,

				purA, LA_1112, rimO, ribB, LA_1189, LA_1222, LA_1223, sucA, thrS, infC, LA_1253, scpB, hisG, plsX, smpB, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, ugd, purE, trmU, nadB, LA_1521, fbaB, aspS, phoH, LA_1692, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, LA_1929, eno, secA, lysU, met2, leuC, LA_2118, LA_2130, bioB, LA_2151, LA_2152, LA_2154, LA_2173, recA, LA_2188, LA_2202, uvrA, fbp, rpoD, tyrS, fliS, purH, mipB, LA_2297, LA_2324, ychF, ubiX, hslU, hslV, xerC, LA_2350, nrdA, trmD, rpe, atoC, hprK, pyrG, LA_2432, LA_2433, def, map, trxB, mfd, LA_2559, ilvH, IlvB, fliH, fliI, LA_2603, LA_2605, LA_2631, groL, bfr, LA_2727, prfC, LA_2749, mreB, sufB, ahpC, LA_2821, argF, LA_2894, radA, LA_2917, PykF, eutG, LA_3052, purB, LA_3094, tgt, LA_3135, caiB, dxs, trpB, proS, uppS, pyrH, tsf, acpS, LA_3354, LA_3406, rpoB, nusG, Bcp, LA_3484, LA_3498, CsdB, LA_3573, lon, dps, LA_3627, hflX, pyrC, LA_3684, dnaK, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3820, prsA, metG, LA_3939, gmd, LA_4034, LA_4036, LA_4139, pheS, argG, coaD, LA_4193, LA_4194, cysN, cysD, ilvC, mtnP, acsA, folE, serS, tatD, gidA, ilvE, hemL, LB_082, sam1, metH, argB, apt, GalF, gltB, speE, acnA, LB_333, Pyk2, ParA4
GO:0016053	Organic acid biosynthetic process	36	147	nadA, lysC, carB, dapL, proA, hisG, plsX, glnA, glyA, ugd, LA_1719, panD, folD, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, LA_2432, ilvH, argF, LA_3052, trpB, acpS, thrC, argG, ilvC, mtnP, folE, serS, ilvE, metH, argB, LB_208, gltB
GO:1901137	Carbohydrate derivative biosynthetic process	44	180	LA_0332, mltG, carB, adk, coaX, purA, LA_1112, LA_1221, LA_1323, ugd, purE, manC, rfe, LA_2173, LA_2207, purH, LA_2318, LA_2326, pyrG, fliI, LA_2749, rodA, LA_2755, atpD, atpG, atpA, purB, spoT, tgt, dxs, pyrH, LA_3573, LA_3597, pyrC, prsA, LA_3905, LA_3936, LA_3939, gmd, coaD, mtnP, acsA, LB_170, apt
GO:0016070	RNA metabolic process	31	127	trpS, rpoA, rlmN, nusA, pnp, rho, LA_1105, rimO, LA_1205, thrS, ileS, grs1, trmU, aspS, lysU, rpoD, tyrS, trmD, LA_2603, LA_2917, tgt, proS, rpoC, rpoB, nusG, leuS, valS, metG, pheS, serS, gidA
GO:1901575	Organic substance	29	119	gcvP, gcvH, gcvT, rpiB, gpmI, LA_0457, LA_0828, pnp, LA_1032, LA_1112,

	catabolic process			LA_1222, fbaB, tpiA, pgk, eno, acoB, acoA, LA_2119, hslV, LA_2631, PykF, LA_3135, lon, hflB, LA_3936, LA_3939, fadB, LA_4139, Pyk2
GO:1901564	Organonitrogen compound metabolic process	154	636	gatB, fusA, LA_0332, gcvP, gcvH, gcvT, trpS, gpmI, ribH, mltG, nadA, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, dapL, coaX, rpmA, proA, rpmE, lysA, purA, ribB, LA_1221, LA_1222, thrS, infC, rplT, LA_1260, hisG, rpmF, smpB, glnA, LA_1323, ileS, grs1, glyA, lexA, purE, nadB, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1719, panD, folD, lepA, eno, guaB, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, LA_2207, rpsU, tyrS, purH, LA_2318, hslU, hslV, rplS, rpsP, hprK, pyrG, def, map, LA_2507, LA_2559, ilvH, fliI, LA_2631, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, argF, PykF, sseA, icmF, lgt, LA_3052, LA_3054, purB, spoT, tgt, dxs, trpB, proS, pyrH, tsf, ccmE, rpsI, rplM, rpsG, rpsL, rplA, rplK, CsdB, lon, hflX, pyrC, leuS, thrC, valS, LA_3767, hflB, LA_3820, prsA, metG, LA_3936, LA_3978, pheS, argG, coaD, ilvC, mtnP, acsA, folE, serS, ilvE, hemL, LB_074, metH, argB, apt, Mcm3, gltB, speE, Pyk2
GO:1901135	Carbohydrate derivative metabolic process	56	232	LA_0332, gpmI, mltG, carB, adk, coaX, purA, LA_1112, LA_1221, LA_1222, LA_1323, ugd, purE, manC, fbaB, tpiA, pgk, eno, rfe, LA_2119, LA_2173, LA_2207, purH, mipB, LA_2318, LA_2326, rpe, pyrG, fliI, LA_2749, rodA, LA_2755, atpD, atpG, atpA, PykF, purB, spoT, tgt, dxs, pyrH, LA_3573, LA_3597, pyrC, prsA, LA_3905, LA_3936, LA_3939, gmd, coaD, mtnP, acsA, LB_170, apt, GalF, Pyk2
GO:0080090	Regulation of primary metabolic process	33	138	fusA, LA_0395, LA_0653, LA_0720, tufB, nusA, LA_1189, LA_1205, infC, LA_1253, lexA, LA_1692, lepA, LA_2072, rpoD, LA_2325, atoC, hprK, mfd, LA_2603, prfC, LA_3094, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, LB_333
GO:0019538	Protein metabolic process	71	297	gatB, fusA, trpS, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpmA, rpmE, LA_1221, thrS, infC, rplT, LA_1260, rpmF, smpB, ileS, grs1, lexA, rplI, aspS, lepA, lysU, LA_2118, rpmB, rpsU, tyrS, hslU, hslV, rplS, rpsP, hprK, def, map, LA_2507, LA_2559, LA_2631, prfC, lgt, LA_3054, proS, tsf,

				ccmE, rpsI, rplM, rpsG, rpsL, rplA, rplK, lon, hflX, leuS, valS, LA_3767, hflB, LA_3820, metG, LA_3978, pheS, serS
GO:0097159	Organic cyclic compound binding	204	857	gyrB, LA_0006, cyoB, pckA, gatB, LA_0273, LA_0274, LA_0280, LA_0309, fusA, LA_0334, pntB, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, LA_0666, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, ruvB, coaX, rlmN, obg, proA, nuoF, nuoD, nusA, pnp, LA_0969, rho, sucD, sucC, purA, LA_1112, LA_1137, ribB, cysA, LA_1205, LA_1221, LA_1223, sucA, thrS, infC, rplT, LA_1253, LA_1260, hisG, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1438, lexA, ugd, LA_1483, trmU, lldD, manC, rplI, aspS, phoH, LA_1692, pgk, gapA, lepA, sdhA, LA_1930, secA, LA_1977, lysU, LA_2114, LA_2119, ssl2, LA_2130, LA_2152, recA, LA_2188, uvrA, rpoD, tyrS, LA_2318, LA_2325, ychF, mrp, hslU, xerC, nrdA, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, IlvB, fliI, LA_2603, LA_2605, groL, LA_2727, prfC, LA_2755, mreB, atpD, atpG, atpA, radA, LA_2917, PykF, icmF, LA_3052, LA_3094, LA_3143, LA_3265, LA_3268, dxs, proS, pyrH, tsf, ccmE, ccmF, LA_3363, greA, LA_3406, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, ppk, glcD, CsdB, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, hrcA, dnaK, dnaJ, leuS, thrC, valS, LA_3767, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, ilvC, acsA, folE, serS, gidA, hemL, ParB3, LB_074, metH, argB, Mcm3, acnA, LB_333, Pyk2, ParA4
GO:1901363	Heterocyclic compound binding	204	857	gyrB, LA_0006, cyoB, pckA, gatB, LA_0273, LA_0274, LA_0280, LA_0309, fusA, LA_0334, pntB, LA_0391, trpS, etfA, ivd, LA_0560, ftsZ, uvrB, LA_0666, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, ruvB, coaX, rlmN, obg, proA, nuoF, nuoD, nusA, pnp, LA_0969, rho, sucD, sucC, purA, LA_1112, LA_1137, ribB, cysA, LA_1205, LA_1221, LA_1223, sucA, thrS, infC, rplT, LA_1253, LA_1260, hisG, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1438, lexA, ugd, LA_1483, trmU, lldD, manC, rplI, aspS, phoH, LA_1692, pgk, gapA, lepA, sdhA, LA_1930, secA,

				LA_1977, lysU, LA_2114, LA_2119, ssl2, LA_2130, LA_2152, recA, LA_2188, uvrA, rpoD, tyrS, LA_2318, LA_2325, ychF, mrp, hslU, xerC, nrdA, gspE, atoC, hprK, yhbG, pyrG, LA_2507, mfd, clpX, IlvB, fliI, LA_2603, LA_2605, groL, LA_2727, prfC, LA_2755, mreB, atpD, atpG, atpA, radA, LA_2917, PykF, icmF, LA_3052, LA_3094, LA_3143, LA_3265, LA_3268, dxs, proS, pyrH, tsf, ccmE, ccmF, LA_3363, greA, LA_3406, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, ppk, glcD, CsdB, lon, LA_3627, LA_3630, hflX, LA_3684, LA_3694, hrcA, dnaK, dnaJ, leuS, thrC, valS, LA_3767, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, ilvC, acsA, folE, serS, gidA, hemL, ParB3, LB_074, metH, argB, Mcm3, acnA, LB_333, Pyk2, ParA4
GO:0018130	Heterocycle biosynthetic process	55	232	dnaE, LA_0332, ribH, nadA, carB, adk, rpoA, coaX, proA, nusA, rho, purA, ribB, LA_1205, hisG, LA_1323, ugd, purE, nadB, manC, folD, LA_2118, bioB, LA_2207, rpoD, purH, LA_2318, pyrG, fliI, LA_2603, LA_2749, atpD, atpG, atpA, purB, spoT, tgt, dxs, trpB, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, prsA, LA_3936, gmd, coaD, mtnP, acsA, folE, serS, hemL, apt
GO:0043167	Ion binding	190	806	gyrB, LA_0006, cyoB, pckA, gatB, LA_0273, LA_0274, fusA, LA_0332, LA_0391, trpS, etfA, ivd, gpmI, LA_0560, ftsZ, nadA, uvrB, LA_0666, lysC, carB, tufB, adk, dapL, ruvB, coaX, rlmN, obg, nuoF, nuoE, nuoB, pnp, LA_0969, LA_1019, rho, sucC, LA_1105, purA, rimO, ribB, cysA, LA_1221, LA_1223, sucA, thrS, LA_1253, hisG, LA_1295, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, LA_1438, ovp1, LA_1483, trmU, lldD, manC, aspS, phoH, LA_1692, pgk, lepA, LA_1896, sdhA, LA_1930, LA_1948, eno, secA, LA_1977, lysU, leuC, LA_2114, LA_2118, LA_2119, ssl2, LA_2130, bioB, LA_2151, LA_2152, LA_2154, recA, LA_2188, uvrA, fbp, tyrS, LA_2297, LA_2318, ychF, mrp, hslU, hslV, nrdA, gspF, gspE, rpe, atoC, hprK, yhbG, pyrG, def, map, trxB, LA_2507, mfd, clpX, IlvB, fliI, LA_2605, groL, bfr, LA_2712, prfC, LA_2749, LA_2755, mreB, atpD, atpG, atpA, argF, radA, PykF, icmF, eutG, LA_3052, LA_3094, tgt, LA_3143, LA_3265, LA_3266, LA_3268, dxs, proS, uppS, pyrH, acpS, ccmE, LA_3354, LA_3363, rpoC, ppk, glcD, CsdB, lon, dps,

				LA_3627, LA_3630, hflX, pyrC, LA_3684, LA_3694, dnaK, dnaJ, leuS, thrC, valS, add, LA_3794, hflB, prsA, metG, LA_3936, gmd, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, ilvC, acsA, folE, serS, tatD, gidA, hemL, LB_074, metH, argB, Mcm3, gltB, acnA, Pyk2, ParA4
GO:0051171	Regulation of nitrogen compound metabolic process	32	137	fusA, LA_0395, LA_0653, LA_0720, tufB, nusA, LA_1189, LA_1205, infC, LA_1253, lexA, LA_1692, lepA, LA_2072, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3094, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, LB_333
GO:0060255	Regulation of macromolecule metabolic process	33	142	fusA, LA_0395, LA_0653, LA_0720, tufB, nusA, pnp, LA_1189, LA_1205, infC, LA_1253, lexA, LA_1692, lepA, LA_2072, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3094, tsf, greA, rplA, nusG, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, LB_333
GO:0016829	Lyase activity	23	99	pckA, LA_0332, mltG, mgsA, LA_1032, lysA, fbaB, panD, eno, leuC, LA_2318, ubiX, LA_2627, LA_2749, purB, trpB, LA_3606, LA_3607, thrC, LA_3936, gmd, fadB, acnA
GO:0034641	Cellular nitrogen compound metabolic process	147	634	gyrB, LA_0006, gatB, dnaE, fusA, LA_0332, trpS, gpmI, ribH, nadA, uvrB, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, ruvB, coaX, rlmN, rpmA, nusA, pnp, rpmE, rho, LA_1105, purA, rimO, ribB, LA_1205, LA_1222, thrS, infC, rplT, LA_1260, hisG, rpmF, smpB, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, panD, folD, lepA, LA_1935, eno, guaB, lysU, LA_2118, bioB, rpmB, recA, LA_2207, uvrA, rpsU, rpoD, tyrS, purH, LA_2318, xerC, rplS, trmD, rpsP, pyrG, def, LA_2507, mfd, fliI, LA_2603, prfC, LA_2749, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, dxs, trpB, proS, pyrH, tsf, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, LA_3573, hflX, pyrC, LA_3684, leuS, valS, LA_3767, add, LA_3820, prsA, metG, LA_3936, gmd, pheS, coaD, LA_4193, LA_4194, mtnP, acsA, folE, serS, tatD, gidA, hemL, apt, GalF, speE, Pyk2

GO:0016491	Oxidoreductase activity	57	247	LA_0020, cyoB, LA_0334, pntB, gcvP, etfA, etfB, ivd, LA_0560, LA_0666, proA, nuoH, nuoF, nuoE, nuoD, nuoB, nuoA, LA_1112, LA_1223, sucA, ugd, lldD, nadB, gapA, folD, LA_1896, sdhA, LA_1930, LA_1977, guaB, acoB, acoA, LA_2152, LA_2324, nrdA, trxB, bfr, LA_2727, ahpC, eutG, LA_3143, LA_3265, LA_3266, LA_3268, LA_3354, LA_3363, Bcp, glcD, LA_3484, dps, LA_3627, mviM, LA_3939, fadB, ilvC, LB_082, gltB
GO:1901362	Organic cyclic compound biosynthetic process	55	240	dnaE, LA_0332, ribH, nadA, carB, adk, rpoA, coaX, proA, nusA, rho, purA, ribB, LA_1205, hisG, LA_1323, ugd, purE, nadB, manC, folD, LA_2118, bioB, LA_2207, rpoD, purH, LA_2318, pyrG, fliI, LA_2603, LA_2749, atpD, atpG, atpA, purB, spoT, tgt, dxs, trpB, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, prsA, LA_3936, gmd, coaD, mtnP, acsA, folE, serS, hemL, apt
GO:0019222	Regulation of metabolic process	35	153	fusA, LA_0395, LA_0653, LA_0720, tufB, nusA, pnp, LA_1189, LA_1205, infC, LA_1253, lexA, LA_1692, lepA, LA_2072, rpoD, LA_2325, atoC, hprK, mfd, LA_2603, prfC, LA_3094, tsf, greA, rplA, nusG, LA_3498, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, LB_333
GO:0019438	Aromatic compound biosynthetic process	50	219	dnaE, LA_0332, nadA, carB, adk, rpoA, coaX, nusA, rho, purA, LA_1205, hisG, LA_1323, ugd, purE, nadB, manC, folD, LA_2207, rpoD, purH, LA_2318, pyrG, fliI, LA_2603, LA_2749, atpD, atpG, atpA, purB, spoT, tgt, dxs, trpB, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, prsA, LA_3936, gmd, coaD, mtnP, acsA, folE, serS, hemL, apt
GO:0031323	Regulation of cellular metabolic process	32	143	fusA, LA_0395, LA_0653, LA_0720, tufB, nusA, LA_1189, LA_1205, infC, LA_1253, lexA, LA_1692, lepA, rpoD, LA_2325, atoC, mfd, LA_2603, prfC, LA_3094, tsf, greA, rplA, nusG, LA_3498, LA_3572, hflX, hrcA, LA_3767, LA_3911, LA_4036, LB_333
GO:0044283	Small molecule biosynthetic process	50	226	pckA, LA_0319, LA_0332, ribH, nadA, lysC, carB, dapL, proA, mgsA, ribB, hisG, plsX, glnA, glyA, ugd, tpiA, LA_1719, panD, folD, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, LA_2207, fbp, LA_2432, ilvH, LA_2749, argF, LA_3052, spoT, tgt, dxs, trpB, acpS, thrC, argG, ilvC, mtnP, folE, serS, ilvE, metH, argB, apt, LB_208, gltB

GO:0043169	Cation binding	92	428	gyrB, cyoB, pckA, LA_0332, gpmI, nadA, LA_0666, carB, coaX, rlmN, obg, nuoF, nuoE, nuoB, pnp, LA_1019, sucC, LA_1105, purA, rimO, ribB, sucA, thrS, glnA, LA_1323, ileS, LA_1417, ovp1, LA_1483, LA_1896, eno, lysU, leuC, LA_2118, bioB, LA_2151, LA_2152, LA_2154, uvrA, fbp, LA_2297, ychF, hslV, gspF, rpe, hprK, pyrG, def, map, clpX, IlvB, bfr, LA_2712, LA_2749, radA, PykF, icmF, eutG, LA_3094, tgt, LA_3265, LA_3266, LA_3268, dxs, uppS, acpS, ccmE, LA_3354, rpoC, ppk, dps, hflX, pyrC, dnaJ, add, LA_3794, hflB, prsA, metG, LA_3936, pheS, LA_4194, ilvC, acsA, folE, tatD, LB_074, metH, Mcm3, gltB, acnA, Pyk2
GO:0005488	Binding	258	1214	gyrB, LA_0006, cyoB, pckA, gatB, LA_0273, LA_0274, LA_0280, LA_0309, fusA, LA_0332, LA_0334, pntB, LA_0391, trpS, etfA, ivd, gpmI, LA_0560, ftsZ, nadA, uvrB, LA_0666, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, secY, adk, rpsM, rpsK, rpsD, rpoA, dapL, ruvB, coaX, rlmN, obg, proA, nuoH, nuoF, nuoE, nuoD, nuoB, nuoA, nusA, pnp, LA_0969, LA_1019, rho, sucD, sucC, LA_1105, purA, LA_1112, LA_1137, rimO, ribB, cysA, LA_1189, LA_1205, LA_1221, LA_1223, sucA, thrS, infC, rplT, LA_1253, scpB, LA_1260, hisG, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, LA_1438, lexA, ugd, ovp1, LA_1483, trmU, lldD, manC, LA_1521, rplI, aspS, phoH, LA_1692, pgk, gapA, lepA, LA_1896, sdhA, LA_1929, LA_1930, LA_1948, eno, secA, LA_1977, lysU, leuC, LA_2114, LA_2118, LA_2119, ssl2, LA_2130, bioB, LA_2151, LA_2152, LA_2154, recA, LA_2188, uvrA, fbp, rpoD, tyrS, LA_2297, LA_2318, LA_2325, ychF, mrp, hslU, hslV, xerC, nrdA, gspF, gspE, gspD, rpe, atoC, hprK, yhbG, pyrG, def, map, trxB, LA_2507, mfd, clpX, IlvB, fliI, LA_2603, LA_2605, LA_2631, groL, bfr, LA_2712, LA_2727, prfC, LA_2749, LA_2755, mreB, atpD, atpG, atpA, argF, radA, LA_2917, PykF, icmF, eutG, LA_3052, LA_3094, tgt, LA_3143, LA_3265, LA_3266, LA_3268, dxs, proS, uppS, pyrH, tsf, acpS, ccmE, ccmF, LA_3354, LA_3363, greA, LA_3406, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, ppk, glcD, LA_3498, CsdB, LA_3576, lon, dps, LA_3627, LA_3630, hflX, pyrC, LA_3684,

				LA_3694, hrcA, dnaK, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3794, hflB, prsA, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4036, pheS, argG, coaD, LA_4193, LA_4194, cysN, ilvC, acsA, folE, serS, tatD, gidA, hemL, ParB3, LB_074, metH, argB, Mcm3, gltB, acnA, LB_333, Pyk2, ParA4
GO:0046872	Metal ion binding	90	425	gyrB, cyoB, pckA, LA_0332, gpmI, nadA, LA_0666, carB, coaX, rlmN, obg, nuoF, nuoE, nuoB, pnp, LA_1019, sucC, LA_1105, purA, rimO, ribB, thrS, glnA, LA_1323, ileS, LA_1417, ovp1, LA_1483, LA_1896, eno, lysU, leuC, LA_2118, bioB, LA_2151, LA_2152, LA_2154, uvrA, fbp, LA_2297, ychF, hslV, gspF, rpe, hprK, pyrG, def, map, clpX, IlvB, bfr, LA_2712, LA_2749, radA, PykF, icmF, eutG, LA_3094, tgt, LA_3265, LA_3266, LA_3268, dxs, uppS, acpS, ccmE, LA_3354, rpoC, ppk, dps, hflX, pyrC, dnaJ, add, hflB, prsA, metG, LA_3936, pheS, LA_4194, ilvC, acsA, folE, tatD, LB_074, metH, Mcm3, gltB, acnA, Pyk2
GO:0006139	Nucleobase-containing compound metabolic process	90	445	gyrB, LA_0006, dnaE, LA_0332, trpS, gpmI, nadA, uvrB, carB, adk, rpoA, ruvB, coaX, rlmN, nusA, pnp, rho, LA_1105, purA, rimO, LA_1205, LA_1222, thrS, LA_1323, ileS, grs1, lexA, ugd, purE, trmU, nadB, manC, fbaB, aspS, tpiA, pgk, folD, LA_1935, eno, guaB, lysU, recA, LA_2207, uvrA, rpoD, tyrS, purH, LA_2318, xerC, trmD, pyrG, mfd, fliI, LA_2603, LA_2749, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, proS, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, leuS, valS, add, prsA, metG, LA_3936, gmd, pheS, coaD, LA_4193, LA_4194, mtnP, acsA, serS, tatD, gidA, apt, GalF, Pyk2
GO:0044237	Cellular metabolic process	262	1297	gyrB, LA_0006, LA_0106, cyoB, gatB, dnaE, fusA, LA_0319, LA_0332, gcvP, gcvH, gcvT, rpiB, trpS, etfA, etfB, gpmI, LA_0457, ribH, mltG, nadA, uvrB, LA_0666, LA_0671, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, rpmA, proA, mgsA, nusA, pnp, LA_1019, rpmE, rho, LA_1032, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, ribB, LA_1205, LA_1221, LA_1222, sucA, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, smpB, glnA, LA_1323, ileS, grs1, rbsK, glyA, lexA, pssA, ugd, purE, trmU, lldD, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1713,

				LA_1719, panD, folD, lepA, LA_1896, LA_1935, eno, guaB, lysU, acoB, acoA, rfe, met2, leuC, LA_2118, LA_2119, LA_2130, bioB, LA_2152, rpmB, LA_2172, LA_2173, recA, LA_2202, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, purH, mipB, LA_2309, LA_2318, LA_2326, ubiX, mrp, hslV, xerC, LA_2350, nrdA, rplS, trmD, rpsP, rpe, hprK, pyrG, LA_2432, LA_2433, def, map, LA_2507, mfd, ilvH, IlvB, LA_2572, fliI, LA_2603, LA_2627, LA_2712, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, sufB, LA_2821, argF, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, purB, spoT, tgt, LA_3135, caiB, LA_3265, LA_3266, LA_3268, dxs, trpB, proS, uppS, pyrH, tsf, acpS, ccmE, LA_3354, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, ppk, CsdB, LA_3573, LA_3579, lon, LA_3597, LA_3606, LA_3607, LA_3627, hflX, pyrC, LA_3684, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4034, fadB, LA_4139, pheS, argG, coaD, LA_4193, LA_4194, cysN, cysD, ilvC, mtnP, acsA, folE, serS, tatD, gidA, ilvE, hemL, LB_074, sam1, metH, argB, LB_170, apt, LB_208, GalF, Mcm3, gltB, speE, acnA, Pyk2
GO:0006807	Nitrogen compound metabolic process	193	966	gyrB, LA_0006, gatB, dnaE, fusA, LA_0332, gcvP, gcvH, gcvT, trpS, gpmI, ribH, mltG, nadA, uvrB, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, ruvB, coaX, rlmN, rpmA, proA, nusA, pnp, rpmE, rho, lysA, LA_1105, purA, rimO, ribB, LA_1205, LA_1221, LA_1222, thrS, infC, rplT, LA_1260, hisG, rpmF, smpB, glnA, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1719, panD, folD, lepA, LA_1935, eno, guaB, lysU, rfe, met2, leuC, LA_2118, LA_2130, bioB, LA_2152, rpmB, recA, LA_2207, uvrA, rpsU, rpoD, tyrS, purH, LA_2318, hslU, hslV, xerC, rplS, trmD, rpsP, hprK, pyrG, def, map, LA_2507, mfd, LA_2559, ilvH, fliI, LA_2603, LA_2631, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, argF, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3268, dxs, trpB, proS, pyrH, tsf, ccmE, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, CsdB, LA_3573, lon, hflX, pyrC, LA_3684, leuS, thrC, valS, LA_3767, add, hflB, LA_3820,

				prsA, metG, LA_3936, gmd, LA_3978, pheS, argG, coaD, LA_4193, LA_4194, ilvC, mtnP, acsA, folE, serS, tatD, gidA, ilvE, hemL, LB_074, metH, argB, apt, GalF, Mcm3, gltB, speE, Pyk2
GO:0044238	Primary metabolic process	220	1104	gyrB, LA_0006, LA_0106, pckA, gatB, dnaE, fusA, LA_0332, gcvP, gcvH, gcvT, rpiB, trpS, gpmI, LA_0457, nadA, uvrB, LA_0671, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, LA_0790, ruvB, LA_0828, coaX, rlmN, rpmA, proA, nusA, pnp, rpmE, rho, LA_1032, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, LA_1189, LA_1205, LA_1221, LA_1222, sucA, thrS, infC, rplT, LA_1260, hisG, rpmF, plsX, smpB, glnA, LA_1323, ileS, grs1, rbsK, glyA, lexA, pssA, ugd, purE, trmU, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, LA_1719, panD, folD, lepA, LA_1896, LA_1935, eno, guaB, lysU, met2, leuC, LA_2118, LA_2119, LA_2130, LA_2152, LA_2154, rpmB, LA_2172, LA_2173, recA, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, purH, mipB, LA_2309, LA_2318, LA_2326, hslU, hslV, xerC, rplS, trmD, rpsP, rpe, hprK, pyrG, LA_2432, LA_2433, def, map, LA_2507, mfd, LA_2559, ilvH, fliI, LA_2603, LA_2631, LA_2632, prfC, LA_2749, atpD, atpG, atpA, argF, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3135, dxs, trpB, proS, pyrH, tsf, acpS, ccmE, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, CsdB, LA_3573, lon, LA_3597, LA_3627, hflX, pyrC, LA_3684, leuS, thrC, valS, LA_3767, add, hflB, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, fadB, LA_4139, pheS, argG, coaD, LA_4193, LA_4194, ilvC, mtnP, acsA, serS, tatD, gidA, ilvE, LB_074, metH, argB, LB_170, apt, GalF, Mcm3, gltB, acnA, Pyk2
GO:0003676	Nucleic acid binding	77	388	gyrB, LA_0006, fusA, uvrB, LA_0720, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpoA, ruvB, rlmN, nusA, pnp, rho, LA_1137, LA_1205, thrS, infC, rplT, LA_1253, LA_1260, smpB, smc, ileS, bipA, lexA, trmU, rplI, aspS, LA_1692, lepA, lysU, ssl2, recA, uvrA, rpoD, tyrS, LA_2325, xerC, atoC, mfd, LA_2603, prfC, radA, LA_2917, LA_3094, tsf, greA, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, lon, hrcA, LA_3767,

				LA_4036, pheS, LA_4193, LA_4194, ParB3, acnA, LB_333
GO:0051179	Localization	54	279	cyoB, LA_0273, LA_0274, FlgB, secY, LA_0969, LA_0970, secD, cysU, cysW, cysA, LA_1438, ovp1, LA_1543, secA, flaB, ttg2B, fliM, LA_2113, LA_2114, LA_2151, LA_2188, LA_2207, fliS, LA_2319, gspF, gspD, yhbG, flgL, LA_2418, fliH, fliF, fliI, flhA, LA_2667, bfr, atpD, atpG, atpA, flgE, LA_2891, LA_3146, ccmF, LA_3498, LA_3575, LA_3576, LA_3577, LA_3630, LA_3694, LA_3794, LA_4155, LA_4228, LA_4229, LA_4308
GO:0044260	Cellular macromolecule metabolic process	106	556	gyrB, LA_0006, gatB, dnaE, fusA, trpS, mltG, uvrB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpoA, ruvB, rlmN, rpmA, nusA, pnp, rpmE, rho, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, rpmF, smpB, ileS, grs1, lexA, manC, rplI, aspS, lepA, lysU, rfe, LA_2118, rpmB, LA_2173, recA, uvrA, rpsU, rpoD, tyrS, hslV, xerC, nrdA, rplS, trmD, rpsP, hprK, def, map, LA_2507, mfd, LA_2603, prfC, rodA, LA_2755, radA, lgt, proS, tsf, ccmE, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, LA_3573, lon, LA_3597, hflX, LA_3684, dnaJ, leuS, valS, LA_3767, LA_3820, LA_3905, metG, LA_3978, pheS, LA_4193, LA_4194, serS, tatD, LB_170, LB_208
GO:0046483	Heterocycle metabolic process	103	541	gyrB, LA_0006, dnaE, LA_0332, trpS, gpmI, ribH, nadA, uvrB, carB, adk, rpoA, ruvB, coaX, rlmN, proA, nusA, pnp, rho, LA_1105, purA, rimO, ribB, LA_1205, LA_1222, thrS, hisG, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, folD, LA_1935, eno, guaB, lysU, LA_2118, bioB, recA, LA_2207, uvrA, rpoD, tyrS, purH, LA_2318, xerC, trmD, pyrG, mfd, fliI, LA_2603, LA_2749, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, dxs, trpB, proS, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, leuS, valS, add, prsA, metG, LA_3936, gmd, pheS, coaD, LA_4193, LA_4194, mtnP, acsA, folE, serS, tatD, gidA, hemL, metH, apt, GalF, Pyk2
GO:0006793	Phosphorus metabolic process	57	301	gpmI, nadA, lysC, carB, adk, coaX, purA, LA_1112, LA_1222, plsX, LA_1323, rbsK, pssA, ugd, purE, nadB, fbaB, tpiA, pgk, gapA, folD, eno, guaB, LA_2119, fbp, purH, mipB, LA_2318, LA_2326, rpe, hprK, pyrG,

				fliI, atpD, atpG, atpA, PykF, purB, spoT, dxs, pyrH, ppk, LA_3573, hflX, pyrC, add, prsA, LA_3936, LA_3939, gmd, coaD, acsA, folE, argB, apt, GalF, Pyk2
GO:0009987	Cellular process	335	1773	gyrB, LA_0006, LA_0106, cyoB, LA_0249, gatB, dnaE, fusA, LA_0319, LA_0332, FlgB, gcvP, gcvH, gcvT, rpiB, LA_0391, trpS, etfA, etfB, gpmI, LA_0457, ribH, mltG, ftsA, ftsZ, nadA, LA_0627, uvrB, LA_0666, LA_0671, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, secY, adk, rpsM, rpsK, rpsD, rpoA, dapL, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, rpmA, obg, proA, mgsA, nusA, pnp, LA_0969, LA_0970, LA_1019, rpmE, rho, LA_1032, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, secD, ribB, cysU, cysW, cysA, LA_1205, LA_1221, LA_1222, LA_1223, sucA, thrS, infC, rplT, LA_1253, scpB, LA_1260, hisG, rpmF, plsX, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, lexA, pssA, ugd, purE, ovp1, LA_1483, trmU, lldD, nadB, manC, LA_1521, fbaB, LA_1543, rplI, aspS, LA_1692, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, lepA, LA_1896, LA_1935, LA_1948, eno, secA, guaB, lysU, acoB, acoA, flaB, rfe, met2, fliM, leuC, LA_2113, LA_2114, LA_2118, LA_2119, LA_2130, bioB, LA_2151, LA_2152, rpmB, LA_2172, LA_2173, recA, LA_2188, LA_2202, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, fliS, purH, mipB, LA_2309, LA_2318, LA_2319, LA_2326, ubiX, mrp, hslU, hslV, xerC, LA_2350, nrdA, gspF, gspE, gspD, rplS, trmD, rpsP, rpe, atoC, hprK, yhbG, pyrG, flgL, LA_2418, LA_2432, LA_2433, def, map, LA_2507, mfd, clpX, ilvH, IlvB, LA_2572, fliH, fliF, fliI, LA_2603, LA_2605, flhA, LA_2627, groL, LA_2667, bfr, LA_2712, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, sufB, ahpC, LA_2821, argF, flgE, LA_2891, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, purB, spoT, tgt, LA_3135, caiB, LA_3146, LA_3265, LA_3266, LA_3268, dxs, trpB, proS, uppS, pyrH, tsf, acpS, ccmE, ccmF, LA_3354, LA_3406, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, Bcp, ppk, LA_3498, CsdB, LA_3573, LA_3575, LA_3576, LA_3577, LA_3579, lon, LA_3597, dps, LA_3606, LA_3607, LA_3627, LA_3630, hflX, pyrC, LA_3684, LA_3694, dnaK, dnaJ, leuS, thrC, valS, LA_3767, add, LA_3820, prsA, LA_3905,

				metG, LA_3936, LA_3939, gmd, LA_3978, LA_4034, fadB, LA_4139, pheS, LA_4155, argG, coaD, LA_4193, LA_4194, cysN, cysD, LA_4228, LA_4229, ilvC, mtnP, acsA, folE, LA_4308, FlgK, serS, tatD, gidA, ilvE, hemL, ParB3, LB_074, sam1, metH, argB, LB_170, apt, LB_208, GalF, Mcm3, gltB, speE, acnA, LB_333, Pyk2, ParA4
GO:0008152	Metabolic process	304	1614	gyrB, LA_0006, LA_0020, LA_0106, cyoB, pckA, gatB, dnaE, fusA, LA_0319, LA_0332, LA_0334, pntB, gcvP, gcvH, gcvT, rpiB, trpS, etfA, etfB, ivd, gpmI, LA_0457, ribH, mltG, LA_0560, nadA, uvrB, LA_0666, LA_0671, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, rpmA, proA, nuoH, nuoF, nuoE, nuoD, nuoB, nuoA, mgsA, nusA, pnp, LA_1019, rpmE, rho, LA_1032, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, ribB, LA_1189, LA_1205, LA_1221, LA_1222, LA_1223, sucA, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, smpB, smc, glnA, LA_1323, ileS, grs1, rbsK, glyA, lexA, pssA, ugd, purE, trmU, lldD, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, lepA, LA_1896, sdhA, LA_1930, LA_1935, eno, LA_1977, guaB, lysU, acoB, acoA, rfe, met2, leuC, LA_2118, LA_2119, LA_2130, bioB, LA_2152, LA_2154, rpmB, LA_2172, LA_2173, recA, LA_2202, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, purH, mipB, LA_2297, LA_2309, LA_2318, LA_2324, LA_2326, ubiX, mrp, hslU, hslV, xerC, LA_2350, nrdA, rplS, trmD, rpsP, rpe, hprK, pyrG, LA_2432, LA_2433, def, map, trxB, LA_2507, mfd, LA_2559, ilvH, IlvB, LA_2572, fliI, LA_2603, LA_2627, LA_2631, LA_2632, bfr, LA_2712, LA_2727, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, sufB, ahpC, LA_2821, argF, LA_2894, radA, LA_2917, PykF, sseA, icmF, eutG, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3135, LA_3143, caiB, LA_3219, LA_3265, LA_3266, LA_3268, dxs, trpB, proS, uppS, pyrH, tsf, acpS, ccmE, LA_3354, LA_3363, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, Bcp, ppk, glcD, LA_3484, CsdB, LA_3573, LA_3579, lon, LA_3597, dps, LA_3606, LA_3607, LA_3627, hflX, pyrC, LA_3684, dnaJ, mviM, leuS, thrC, valS, LA_3767, add, hflB, LA_3820, prsA, LA_3905, metG, LA_3936,

				LA_3939, gmd, LA_3978, LA_4034, fadB, LA_4139, pheS, argG, coaD, LA_4193, LA_4194, cysN, cysD, ilvC, mtnP, acsA, folE, serS, tatD, gidA, ilvE, hemL, LB_074, LB_082, sam1, metH, argB, LB_170, apt, LB_208, GalF, Mcm3, gltB, speE, acnA, Pyk2
GO:1901360	Organic cyclic compound metabolic process	103	555	gyrB, LA_0006, dnaE, LA_0332, trpS, gpmI, ribH, nadA, uvrB, carB, adk, rpoA, ruvB, coaX, rlmN, proA, nusA, pnp, rho, LA_1105, purA, rimO, ribB, LA_1205, LA_1222, thrS, hisG, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, folD, LA_1935, eno, guaB, lysU, LA_2118, bioB, recA, LA_2207, uvrA, rpoD, tyrS, purH, LA_2318, xerC, trmD, pyrG, mfd, fliI, LA_2603, LA_2749, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, dxs, trpB, proS, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, leuS, valS, add, prsA, metG, LA_3936, gmd, pheS, coaD, LA_4193, LA_4194, mtnP, acsA, folE, serS, tatD, gidA, hemL, metH, apt, GalF, Pyk2
GO:0006796	Phosphate-containing compound metabolic process	52	281	gpmI, nadA, lysC, carB, adk, coaX, purA, LA_1112, LA_1222, plsX, LA_1323, rbsK, pssA, purE, nadB, fbaB, tpiA, pgk, gapA, folD, eno, guaB, LA_2119, fbp, purH, mipB, LA_2318, rpe, hprK, pyrG, fliI, atpD, atpG, atpA, PykF, purB, spoT, dxs, pyrH, ppk, hflX, pyrC, add, prsA, LA_3936, LA_3939, coaD, acsA, folE, argB, apt, Pyk2
GO:0071704	Organic substance metabolic process	249	1349	gyrB, LA_0006, LA_0106, pckA, gatB, dnaE, fusA, LA_0319, LA_0332, gcvP, gcvH, gcvT, rpiB, trpS, gpmI, LA_0457, ribH, mltG, nadA, uvrB, lysC, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, adk, rpsM, rpsK, rpsD, rpoA, dapL, ruvB, LA_0828, coaX, rlmN, rpmA, proA, mgsA, nusA, pnp, rpmE, rho, LA_1032, lysA, LA_1105, purA, LA_1112, rimO, ribB, LA_1205, LA_1221, LA_1222, thrS, infC, rplT, scpB, LA_1260, hisG, rpmF, plsX, smpB, smc, glnA, LA_1323, ileS, grs1, rbsK, glyA, lexA, pssA, ugd, purE, trmU, lldD, nadB, manC, fbaB, rplI, aspS, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, lepA, LA_1935, eno, guaB, lysU, acoB, acoA, rfe, met2, leuC, LA_2118, LA_2119, LA_2130, bioB, LA_2152, LA_2154, rpmB, LA_2172, LA_2173, recA, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, purH, mipB, LA_2309, LA_2318, LA_2326, hslU, hslV, xerC, nrdA, rplS, trmD, rpsP, rpe, hprK, pyrG, LA_2432,

				LA_2433, def, map, LA_2507, mfd, LA_2559, ilvH, IlvB, fliI, LA_2603, LA_2627, LA_2631, LA_2632, LA_2712, prfC, LA_2749, rodA, LA_2755, atpD, atpG, atpA, LA_2821, argF, LA_2894, radA, LA_2917, PykF, sseA, icmF, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3135, caiB, LA_3268, dxs, trpB, proS, uppS, pyrH, tsf, acpS, ccmE, LA_3354, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, ppk, CsdB, LA_3573, LA_3579, lon, LA_3597, LA_3606, LA_3607, LA_3627, hflX, pyrC, LA_3684, dnaJ, leuS, thrC, valS, LA_3767, add, hflB, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4034, fadB, LA_4139, pheS, argG, coaD, LA_4193, LA_4194, ilvC, mtnP, acsA, folE, serS, tatD, gidA, ilvE, hemL, LB_074, metH, argB, LB_170, apt, LB_208, GalF, Mcm3, gltB, speE, acnA, Pyk2
GO:0006725	Cellular aromatic compound metabolic process	98	532	gyrB, LA_0006, dnaE, LA_0332, trpS, gpmI, nadA, uvrB, carB, adk, rpoA, ruvB, coaX, rlmN, nusA, pnp, rho, LA_1105, purA, rimO, LA_1205, LA_1222, thrS, hisG, LA_1323, ileS, grs1, glyA, lexA, ugd, purE, trmU, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, folD, LA_1935, eno, guaB, lysU, recA, LA_2207, uvrA, rpoD, tyrS, purH, LA_2318, xerC, trmD, pyrG, mfd, fliI, LA_2603, LA_2749, atpD, atpG, atpA, radA, LA_2917, PykF, purB, spoT, tgt, dxs, trpB, proS, pyrH, rpoC, rpoB, nusG, LA_3573, pyrC, leuS, valS, add, prsA, metG, LA_3936, gmd, pheS, coaD, LA_4193, LA_4194, mtnP, acsA, folE, serS, tatD, gidA, hemL, metH, apt, GalF, Pyk2
GO:0110165	Cellular anatomical entity	382	2078	gyrB, LA_0006, LA_0020, LA_0106, cyoB, pckA, gatB, dnaE, LA_0273, LA_0274, LA_0280, LA_0296, LA_0309, fusA, LA_0319, LA_0332, LA_0334, pntB, FlgB, gcvP, gcvH, gcvT, rpiB, LA_0391, trpS, etfA, etfB, ivd, gpmI, LA_0457, mltG, LA_0560, ftsA, ftsZ, nadA, LA_0627, uvrB, LA_0666, LA_0671, lysC, LA_0720, carB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, secY, adk, rpsM, rpsK, rpsD, rpoA, dapL, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, rpmA, obg, proA, nuoH, nuoF, nuoE, nuoD, nuoB, nuoA, mgsA, nusA, pnp, LA_0969, LA_0970, LA_1019, rpmE, rho, lysA, sucD, sucC, LA_1105, purA, LA_1112, LA_1137, rimO, secD, ribB, cysU, cysW, cysA, LA_1189, LA_1205, LA_1221, LA_1222,

				LA_1223, sucA, thrS, infC, rplT, LA_1253, scpB, LA_1260, hisG, rpmF, plsX, LA_1295, smpB, smc, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, LA_1438, pssA, ugd, purE, ovp1, LA_1483, trmU, lldD, nadB, LA_1521, fbaB, LA_1543, rplI, aspS, phoH, LA_1692, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, lepA, LA_1896, sdhA, LA_1929, LA_1930, LA_1935, LA_1948, eno, secA, LA_1977, lysU, acoB, acoA, flaB, rfe, ttg2B, met2, LA_2072, fliM, leuC, LA_2113, LA_2114, LA_2118, LA_2119, ssl2, LA_2130, bioB, LA_2151, LA_2152, LA_2154, rpmB, LA_2172, LA_2173, recA, LA_2188, LA_2202, LA_2207, uvrA, fbp, rpsU, rpoD, tyrS, fliS, purH, mipB, LA_2297, LA_2309, LA_2318, LA_2319, LA_2324, LA_2325, LA_2326, ychF, ubiX, mrp, hslU, hslV, xerC, LA_2350, nrdA, gspF, gspE, gspD, rplS, trmD, rpsP, rpe, atoC, hprK, yhbG, pyrG, flgL, LA_2418, LA_2432, LA_2433, def, map, trxB, mfd, clpX, LA_2559, ilvH, IlvB, LA_2572, fliH, fliF, fliI, LA_2603, LA_2605, flhA, LA_2627, LA_2631, groL, LA_2667, bfr, LA_2712, LA_2727, prfC, LA_2749, rodA, LA_2755, LA_2756, mreB, atpD, atpG, atpA, sufB, ahpC, LA_2821, argF, flgE, LA_2891, LA_2894, radA, LA_2917, PykF, sseA, icmF, eutG, lgt, LA_3052, LA_3054, purB, spoT, LA_3094, tgt, LA_3135, LA_3143, caiB, LA_3146, LA_3219, LA_3265, LA_3266, LA_3268, dxs, trpB, proS, uppS, pyrH, tsf, acpS, ccmE, ccmF, LA_3354, LA_3363, LA_3406, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, Bcp, glcD, LA_3484, LA_3498, CsdB, LA_3573, LA_3575, LA_3576, LA_3577, LA_3579, lon, LA_3597, dps, LA_3606, LA_3607, LA_3627, LA_3630, hflX, pyrC, LA_3684, LA_3694, hrcA, dnaK, dnaJ, mviM, leuS, thrC, valS, LA_3767, add, LA_3794, hflB, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, gmd, LA_3978, LA_4034, LA_4036, LA_4139, pheS, LA_4155, argG, coaD, LA_4193, LA_4194, cysN, cysD, LA_4228, LA_4229, ilvC, mtnP, acsA, folE, LA_4308, FlgK, serS, tatD, gidA, ilvE, hemL, ParB3, LB_074, LB_082, sam1, metH, argB, LB_170, apt, LB_208, GalF, Mcm3, gltB, speE, acnA, LB_333, Pyk2, ParA4
GO:0003824	Catalytic activity	281	1592	gyrB, LA_0006, LA_0020, LA_0106, cyoB, pckA, gatB, dnaE, LA_0273, LA_0274, LA_0296, fusA, LA_0319, LA_0332,

				LA_0334, pntB, gcvP, gcvT, rpiB, trpS, etfA, etfB, ivd, gpmI, LA_0457, ribH, mltG, LA_0560, ftsZ, nadA, LA_0627, uvrB, LA_0666, LA_0671, lysC, carB, tufB, rplB, adk, rpoA, dapL, LA_0790, ruvB, LA_0828, coaX, rlmN, LA_0845, obg, proA, nuoH, nuoF, nuoE, nuoD, nuoB, nuoA, mgsA, pnp, LA_0969, rho, LA_1032, lysA, sucD, sucC, LA_1105, purA, LA_1112, rimO, ribB, cysA, LA_1221, LA_1222, LA_1223, sucA, thrS, hisG, plsX, LA_1295, glnA, LA_1323, ileS, bipA, grs1, rbsK, glyA, LA_1417, LA_1438, lexA, pssA, ugd, purE, ovp1, LA_1483, trmU, lldD, nadB, manC, fbaB, aspS, tpiA, pgk, gapA, LA_1713, LA_1719, panD, folD, lepA, LA_1896, sdhA, LA_1930, LA_1935, LA_1948, eno, LA_1977, guaB, lysU, acoB, acoA, rfe, met2, LA_2072, fliM, leuC, LA_2114, LA_2118, LA_2119, ssl2, LA_2130, bioB, LA_2152, LA_2154, LA_2172, LA_2173, recA, LA_2188, LA_2202, uvrA, fbp, tyrS, purH, mipB, LA_2297, LA_2309, LA_2318, LA_2324, LA_2326, ychF, ubiX, mrp, hslU, hslV, xerC, LA_2350, nrdA, trmD, rpe, hprK, yhbG, pyrG, LA_2432, LA_2433, def, map, LA_2469, trxB, LA_2507, mfd, LA_2559, ilvH, IlvB, fliH, fliF, fliI, LA_2605, LA_2627, LA_2632, bfr, LA_2727, prfC, LA_2749, rodA, atpD, atpG, atpA, ahpC, argF, radA, LA_2917, PykF, sseA, icmF, eutG, lgt, LA_3052, LA_3054, purB, spoT, tgt, LA_3135, LA_3143, caiB, LA_3219, LA_3265, LA_3266, LA_3268, dxs, trpB, proS, uppS, pyrH, acpS, LA_3354, LA_3363, rpoC, rpoB, Bcp, ppk, glcD, LA_3484, CsdB, LA_3573, LA_3579, lon, LA_3597, dps, LA_3606, LA_3607, LA_3627, LA_3630, hflX, pyrC, LA_3684, LA_3694, mviM, leuS, thrC, valS, add, hflB, LA_3820, prsA, LA_3905, metG, LA_3936, LA_3939, LA_3941, gmd, acyP, LA_3978, LA_4034, LA_4036, fadB, LA_4139, pheS, argG, coaD, LA_4193, LA_4194, cysN, cysD, ilvC, mtnP, acsA, folE, serS, tatD, ilvE, hemL, LB_074, LB_082, sam1, metH, argB, LB_170, apt, LB_208, GalF, Mcm3, gltB, speE, acnA, Pyk2
GO:0016740	Transferase activity	83	491	dnaE, LA_0319, gcvT, LA_0457, ribH, mltG, nadA, LA_0627, LA_0671, lysC, rplB, adk, rpoA, dapL, LA_0790, LA_0828, coaX, rlmN, pnp, LA_1105, rimO, LA_1221, LA_1222, hisG, plsX, rbsK, glyA, pssA, LA_1483, trmU, manC, pgk, LA_1719, rfe, met2, LA_2119, LA_2130, bioB, LA_2172,

				LA_2173, purH, mipB, trmD, hprK, ilvH, rodA, argF, PykF, sseA, lgt, LA_3052, spoT, tgt, LA_3135, caiB, LA_3219, dxs, uppS, pyrH, acpS, rpoC, rpoB, ppk, CsdB, LA_3573, LA_3579, LA_3597, prsA, LA_3905, LA_3978, LA_4139, coaD, cysN, mtnP, ilvE, hemL, metH, argB, apt, LB_208, GalF, speE, Pyk2
GO:0043170	Macromolecule metabolic process	120	713	gyrB, LA_0006, gatB, dnaE, fusA, trpS, mltG, uvrB, tufB, rplC, rplW, rplB, rplV, rplP, rplN, rplX, rplE, rpsN, rpsH, rpsE, rpsM, rpsK, rpsD, rpoA, ruvB, rlmN, rpmA, nusA, pnp, rpmE, rho, LA_1105, rimO, LA_1205, LA_1221, thrS, infC, rplT, scpB, LA_1260, rpmF, smpB, smc, ileS, grs1, lexA, ugd, trmU, manC, rplI, aspS, lepA, LA_1935, lysU, rfe, LA_2118, rpmB, LA_2173, recA, uvrA, rpsU, rpoD, tyrS, hslU, hslV, xerC, nrdA, rplS, trmD, rpsP, hprK, def, map, LA_2507, mfd, LA_2559, LA_2603, LA_2631, prfC, rodA, LA_2755, radA, LA_2917, lgt, LA_3054, tgt, proS, tsf, ccmE, rpsI, rplM, rpsG, rpsL, rpoC, rpoB, rplA, rplK, nusG, LA_3573, lon, LA_3597, hflX, LA_3684, dnaJ, leuS, valS, LA_3767, hflB, LA_3820, LA_3905, metG, LA_3978, pheS, LA_4193, LA_4194, serS, tatD, gidA, LB_170, LB_208

Table S3.12. Proteins that exhibit AA changes in P1^{+}.

L. interrogans serovar Lai strain 56601 used as reference. *L. alstonii* not included in the analysis.

Locus tag	Product	Locus tag	Product
LA_0106	long-chain-fatty-acid CoA ligase	LA_2739	hypothetical protein
LA_0243	cytochrome c oxidase polypeptide I	LA_2759	rod shape-determining protein MreB
LA_0250	PIN domain-containing integral membrane protein	LA_2891	Na+/H+ antiporter
LA_0273	lipoprotein releasing system permease	LA_2894	transcriptional regulator
LA_0280	cAMP-binding protein	LA_2996	alcohol dehydrogenase
LA_0408	tryptophanyl-tRNA synthetase	LA_3052	aminotransferase
LA_0627	glycosyltransferase	LA_3085	guanosine polyphosphate pyrophosphohydrolase/synthetase
LA_0778	aspartate carbamoyltransferase	LA_3095	queuine tRNA-ribosyltransferase
LA_0810	Holliday junction DNA helicase RuvB	LA_3380	endoflagellar filament sheath protein

LA_0894	NADH dehydrogenase subunit B	LA_3596	ATP-dependent Lon protease
LA_1047	diaminopimelate decarboxylase	LA_3597	pyridoxal phosphate-dependent aminotransferase
LA_1156	sulfate ABC transporter permease	LA_3630	multidrug ABC transporter ATPase/permease
LA_1260	30S ribosomal protein S1	LA_3634	GTP-binding protein
LA_1532	fructose-bisphosphate aldolase	LA_3636	cyclic amidohydrolase
LA_1880	transcriptional regulator	LA_3703	heat shock operon regulator HrcA
LA_1948	two-component system response regulator	LA_3918	methionyl-tRNA synthetase
LA_2048	phospho-N-acetylmuramoyl-pentapeptide-transferase	LA_3982	hypothetical protein
LA_2323	hypothetical protein	LA_4023	hypothetical protein
LA_2347	integrase/recombinase XerD	LA_4167	phosphopantetheine adenylyltransferase
LA_2375	type II secretory pathway component protein D	LB_170	capsular polysaccharide biosynthesis protein
LA_2405	ABC transporter ATP-binding protein	LB_327	aconitate hydratase
LA_2558	ATP-dependent protease ATP-binding subunit ClpX	LB_355	aspartate-semialdehyde dehydrogenase
LA_2576	hypothetical protein		

Table S3.13. GO functional enrichment of proteins with AA changes in P1^{+}.

GO categories are organized from most to least represented, considering the ratio between the mutated proteins within a GO process and the total number of proteins within that process in the reference used (*L. interrogans* serovar Lai strain 56601).

GO ID	Description	Observed gene count	Reference gene count	Matching proteins in the network
GO:0110165	Cellular anatomical entity	38	2078	LA_0106, cyoB, LA_0250, LA_0273, LA_0280, trpS, LA_0627, pyrB, ruvB, nuoB, lysA, cysU, LA_1260, fbaB, LA_1948, rfe, xerC, gspD, yhbG, clpX, mreB, LA_2891, LA_2894, eutG, LA_3052, spoT, tgt, lon, LA_3597, LA_3630, hflX, pyrC, hrcA, metG, coaD, LB_170, acnA, asd
KW-0067*	ATP-binding	9	188	trpS, pyrB, ruvB, yhbG, clpX, lon, LA_3630, metG, coaD
KW-0547*	Nucleotide-binding	10	212	trpS, pyrB, ruvB, yhbG, clpX, lon, LA_3630, hflX, metG, coaD

*Uniprot Keywords based on UniprotKB/Swiss-Prot entries

Table S3.14. Proteins that exhibit AA changes exclusively between P1$^-$ and P1$^+$.

L. interrogans serovar Lai strain 56601 used as reference. *L. alstonii* not included in the analysis.

Locus tag	Product	Locus tag	Product
LA_0001	chromosomal replication initiation protein A	LA_2436	hypothetical protein
LA_0002	DNA polymerase III subunit beta	LA_2439	heavy metal efflux pump
LA_0005	DNA gyrase subunit B	LA_2459	hypothetical protein
LA_0012	hypothetical protein	LA_2461	anti-Sigma factor antagonist
LA_0017	septum formation inhibitor Maf	LA_2468	hypothetical protein
LA_0021	translation factor Sua5	LA_2474	polymerase
LA_0106	long-chain-fatty-acid CoA ligase	LA_2480	anti-sigma factor antagonist
LA_0240	hypothetical protein	LA_2486	hypothetical protein
LA_0241	SCO1/SenC/PrrC family protein	LA_2489	MBOAT family D-alaninealginate export/acetyltransferase
LA_0243	cytochrome c oxidase polypeptide I	LA_2490	hypothetical protein
LA_0250	PIN domain-containing integral membrane protein	LA_2505	esterase/lipase
LA_0261	RNA polymerase ECF-type sigma factor	LA_2506	imidazole glycerol phosphate synthase subunit HisF
LA_0266	hypothetical protein	LA_2509	glycosyltransferase
LA_0268	hypothetical protein	LA_2510	hypothetical protein
LA_0271	hypothetical protein	LA_2511	undecaprenol kinase
LA_0272	hypothetical protein	LA_2512	cytoplasmic membrane lipoprotein LipL31
LA_0273	lipoprotein releasing system permease	LA_2519	hypothetical protein
LA_0279	hypothetical protein	LA_2528	two-component system response regulator
LA_0280	cAMP-binding protein	LA_2532	fused two-component system sensor histidine kinase/response regulator
LA_0286	hypothetical protein	LA_2536	guanylate cyclase
LA_0292	tRNA (uracil-5-)-methyltransferase	LA_2537	hypothetical protein
LA_0301	OmpA family protein	LA_2538	hypothetical protein
LA_0306	hypothetical protein	LA_2539	ABC transporter ATPase
LA_0308	protoporphyrinogen oxidase	LA_2540	signal transduction histidine kinase
LA_0310	hypothetical protein	LA_2541	fused two-component system sensor histidine kinase/response regulator
LA_0311	metal-dependent hydrolase	LA_2548	serine/threonine phosphatase

LA_0312	cell wall hydrolase	LA_2550	FKBP-type peptidylprolyl isomerase
LA_0314	potassium uptake transporter	LA_2554	phosphate sodium symporter
LA_0321	hypothetical protein	LA_2558	ATP-dependent protease ATP-binding subunit ClpX
LA_0325	Zn-dependent alcohol dehydrogenase	LA_2560	FKBP-type peptidylprolyl isomerase
LA_0326	geranylgeranyl pyrophosphate synthase	LA_2574	methyl-accepting chemotaxis transmembrane protein
LA_0327	hemolysin A	LA_2576	hypothetical protein
LA_0348	anti-sigma factor antagonist	LA_2580	phosphoglycerol transferase
LA_0350	hypothetical protein	LA_2584	hypothetical protein
LA_0351	TPR-repeat-containing protein	LA_2602	polymerase
LA_0362	glycine cleavage system aminomethyltransferase T	LA_2604	hypothetical protein
LA_0365	hypothetical protein	LA_2605	ParA protein
LA_0368	hypothetical protein	LA_2616	sodium:alanine symporter/phosphatidylserine decarboxylase
LA_0370	hypothetical protein	LA_2617	hypothetical protein
LA_0375	hypothetical protein	LA_2621	short-chain dehydrogenase
LA_0379	hypothetical protein	LA_2623	glutathione transferase
LA_0389	hypothetical protein	LA_2625	oligopeptidase A
LA_0392	hypothetical protein	LA_2631	ATP-dependent Clp protease adaptor protein ClpS
LA_0397	lipase/esterase	LA_2656	hypothetical protein
LA_0400	hypothetical protein	LA_2661	methylase/methyltransferase
LA_0401	phosphohistidine phosphatase	LA_2662	ribonuclease BN
LA_0404	HD family protein	LA_2674	metal-dependent amidase/aminoacylase/carboxypeptidase
LA_0405	16S rRNA (adenine(1518)-N(6)/adenine(1519)-N(6))-dimethyltransferase	LA_2689	acetyl-CoA acetyltransferase
LA_0408	tryptophanyl-tRNA synthetase	LA_2692	acetylglutamate kinase
LA_0409	hypothetical protein	LA_2702	phosphoglycolate phosphatase
LA_0441	hypothetical protein	LA_2704	two-component system response regulator
LA_0446	hypothetical protein	LA_2712	Fe-S oxidoreductase
LA_0459	N-acetylmuramoyl-L-alanine amidase	LA_2722	hypothetical protein
LA_0465	TPR-repeat-containing protein	LA_2728	transcriptional regulator

LA_0474	permease	LA_2733	peptide chain release factor 3
LA_0477	hypothetical protein	LA_2738	hypothetical protein
LA_0478	hypothetical protein	LA_2739	hypothetical protein
LA_0484	transcriptional regulator	LA_2740	pseudouridylate synthase
LA_0496	hypothetical protein	LA_2741	hypothetical protein
LA_0500	hypothetical protein	LA_2742	hypothetical protein
LA_0505	hypothetical protein	LA_2746	hypothetical protein
LA_0506	hypothetical protein	LA_2755	transpeptidase/penicillin-binding protein
LA_0512	UDP-3-O-[3-hydroxymyristoyl] glucosamine N-acyltransferase	LA_2759	rod shape-determining protein MreB
LA_0515	flavoprotein	LA_2761	exopolyphosphatase
LA_0517	hypothetical protein	LA_2770	hypothetical protein
LA_0520	hypothetical protein	LA_2771	MBOAT family D-alaninealginate export/acetyltransferase
LA_0533	pyridoxal phosphate-dependent aminotransferase	LA_2772	cGMP-specific phosphodiesterase
LA_0534	hypothetical protein	LA_2788	signal peptidase I
LA_0537	hydrolase	LA_2796	hypothetical protein
LA_0544	hypothetical protein	LA_2797	hypothetical protein
LA_0546	hypothetical protein	LA_2800	hypothetical protein
LA_0548	hypothetical protein	LA_2809	peroxiredoxin
LA_0550	cation/multidrug efflux pump	LA_2813	methyl-accepting chemotaxis protein
LA_0552	periplasmic solute-binding protein	LA_2824	hypothetical protein
LA_0565	guanylate cyclase	LA_2826	hypothetical protein
LA_0576	hypothetical protein	LA_2827	signal transduction protein
LA_0579	bifunctional permease/carbonic anhydrase	LA_2830	two-component system sensor histidine kinase
LA_0595	glutathione S-transferase	LA_2847	response regulator
LA_0599	signal transduction protein	LA_2857	hypothetical protein
LA_0602	NAD/FAD-binding protein	LA_2867	cytoplasmic membrane protein
LA_0607	ABC transporter	LA_2871	polymerase
LA_0616	hypothetical protein	LA_2872	hypothetical protein
LA_0617	hypothetical protein	LA_2891	Na+/H+ antiporter
LA_0625	ATP-dependent DNA helicase RecQ	LA_2892	hypothetical protein
LA_0627	glycosyltransferase	LA_2894	transcriptional regulator
LA_0630	oxidoreductase	LA_2899	methyltransferase

LA_0634	dipeptide/oligopeptide ABC transporter permease	LA_2900	hypothetical protein
LA_0635	hypothetical protein	LA_2901	M23 family metalloendopeptidase
LA_0639	hypothetical protein	LA_2903	Holliday junction DNA helicase RuvA
LA_0644	hypothetical protein	LA_2918	alpha-galactosidase
LA_0647	rRNA methylase	LA_2921	two-component system response regulator
LA_0656	mannose-6-phosphate isomerase	LA_2922	hypothetical protein
LA_0662	endoflagellar motor protein A	LA_2936	hypothetical protein
LA_0665	metallo-beta-lactamase superfamily protein	LA_2947	thiosulfate sulfurtransferase
LA_0666	cytochrome c peroxidase	LA_2949	serine protease
LA_0674	oxidoreductase	LA_2950	hypothetical protein
LA_0735	hypothetical protein	LA_2951	two-component system sensor histidine kinase\response regulator
LA_0736	elongation factor G	LA_2969	glycosyltransferase
LA_0778	aspartate carbamoyltransferase	LA_2970	hypothetical protein
LA_0788	SAM-dependent methyltransferase	LA_2973	hypothetical protein
LA_0791	transglycosylase	LA_2974	cytochrome c peroxidase
LA_0792	hypothetical protein	LA_2977	tRNA (uracil-5-)-methyltransferase
LA_0798	hypothetical protein	LA_2980	stage II sporulation protein
LA_0801	hypothetical protein	LA_2983	ABC transporter permease
LA_0802	TPR-repeat-containing protein	LA_2989	NH(3)-dependent NAD(+) synthetase
LA_0803	hypothetical protein	LA_2996	alcohol dehydrogenase
LA_0810	Holliday junction DNA helicase RuvB	LA_2999	single-stranded DNA exonuclease
LA_0815	two-component system sensor histidine kinase	LA_3003	M23 family metalloendopeptidase
LA_0816	response regulator	LA_3011	cell division protein FtsK
LA_0818	hypothetical protein	LA_3016	hypothetical protein
LA_0819	hypothetical protein	LA_3052	aminotransferase
LA_0825	transcriptional regulator	LA_3067	hypothetical protein
LA_0840	indole-3-glycerol phosphate synthase	LA_3069	hypothetical protein
LA_0855	nicotinate-nucleotide adenylyltransferase	LA_3078	sterol desaturase
LA_0866	HD family protein	LA_3080	adenylosuccinate lyase
LA_0894	NADH dehydrogenase subunit B	LA_3084	nickel ABC transporter periplasmic protein OppA
LA_0900	MarR family transcriptional regulator	LA_3085	guanosine polyphosphate pyrophosphohydrolase/synthetase

LA_0903	methylase	LA_3086	nicotinate-nucleotide pyrophosphorylase
LA_0908	hypothetical protein	LA_3092	hypothetical protein
LA_0909	methylglyoxal synthase	LA_3095	queuine tRNA-ribosyltransferase
LA_0912	TPR-repeat-containing protein	LA_3097	hypothetical protein
LA_0913	NHL repeat protein	LA_3113	serine/threonine kinase
LA_0915	histone deacetylase	LA_3114	cytoplasmic membrane protein
LA_0920	hypothetical protein	LA_3115	hypothetical protein
LA_0927	hypothetical protein	LA_3120	zinc-bindin carboxypeptidase
LA_0929	enolase	LA_3121	permease
LA_0940	hypothetical protein	LA_3146	cation transporter
LA_0941	hypothetical protein	LA_3172	aldo/keto reductase
LA_0943	translation initiation factor IF-2	LA_3200b	hypothetical protein
LA_0945	tRNA pseudouridine synthase B	LA_3206	two-component system response regulator
LA_0946	30S ribosomal protein S15	LA_3210	hypothetical protein
LA_0954	hypothetical protein	LA_3214	hypothetical protein
LA_0956	glutamate synthase subunit alpha	LA_3217	hypothetical protein
LA_0957	outer membrane efflux protein TolC	LA_3224	hypothetical protein
LA_0964	ATP-dependent DNA helicase RecQ	LA_3234	guanylate cyclase
LA_0966	exodeoxyribonuclease V subunit beta	LA_3247	periplasmic protein TonB
LA_1008	hypothetical protein	LA_3255	hypothetical protein
LA_1009	membrane carboxypeptidase/penicillin -binding protein	LA_3262	hypothetical protein
LA_1036	fused two-component system sensor histidine kinase/response regulator	LA_3293	CDP-diglyceride synthetase
LA_1043	M23 family metalloendopeptidase	LA_3303	hypothetical protein
LA_1046	hypothetical protein	LA_3304	hypothetical protein
LA_1047	diaminopimelate decarboxylase	LA_3305	4-hydroxy-tetrahydrodipicolinate reductase
LA_1073	hypothetical protein	LA_3345	hypothetical protein
LA_1080	aminopeptidase	LA_3353	hypothetical protein
LA_1082	ADP-ribose pyrophosphatase	LA_3357	two-component system sensor histidine kinase
LA_1097	hypothetical protein	LA_3362	aminoglycoside phosphotransferase
LA_1099	hypothetical protein	LA_3365	hydroxyethylthiazole kinase

LA_1100	cytoplasmic membrane protein	LA_3371	glucans biosynthesis protein
LA_1107	hypothetical protein	LA_3378	transcription elongation factor
LA_1112	glycerol-3-phosphate dehydrogenase	LA_3380	endoflagellar filament sheath protein
LA_1123	exonuclease	LA_3396	hypothetical protein
LA_1133	hypothetical protein	LA_3397	hypothetical protein
LA_1136	outer membrane lipoprotein-sorting protein	LA_3398	O-methyltransferase
LA_1155	sulfate ABC transporter substrate-binding protein	LA_3399	metallo-beta-lactamase
LA_1156	sulfate ABC transporter permease	LA_3400	Zn-dependent peptidase
LA_1166	pyridoxamine-phosphate oxidase	LA_3412	thiamine monophosphate kinase
LA_1189	sigma 54 modulation protein	LA_3432	hypothetical protein
LA_1190	hypothetical protein	LA_3433	hypothetical protein
LA_1195	hypothetical protein	LA_3434	N-acetylmuramoyl-L-alanine amidase
LA_1202	hypothetical protein	LA_3436	3-demethylubiquinone-9 3-methyltransferase
LA_1203	MBOAT family D-alaninealginate export/acetyltransferase	LA_3438	hypothetical protein
LA_1210	peroxiredoxin	LA_3444	penicillin-binding protein
LA_1220	metal-dependent hydrolase	LA_3449	riboflavin biosynthesis protein
LA_1224	2-oxoglutarate dehydrogenase subunit E1	LA_3450	regulatory protein
LA_1246	cell division protein ZapA	LA_3451	serine phosphatase RsbU
LA_1251	two-component system sensor histidine kinase CheA	LA_3463	hypothetical protein
LA_1252	chemotaxis-specific methylesterase CheB	LA_3469	hypothetical protein
LA_1257	bifunctional prephenate dehydrogenase/chorismate mutase	LA_3470	thiol oxidoreductase
LA_1258	3-phosphoshikimate 1-carboxyvinyltransferase	LA_3471	hypothetical protein
LA_1260	30S ribosomal protein S1	LA_3485	esterase
LA_1286	hypothetical protein	LA_3486	hypothetical protein
LA_1307	hypothetical protein	LA_3492	hypothetical protein
LA_1312	hypothetical protein	LA_3514	serine/threonine phosphatase
LA_1325	isoleucyl-tRNA synthetase	LA_3515	heme exporter protein C
LA_1326	hypothetical protein	LA_3522	hypothetical protein
LA_1328	hypothetical protein	LA_3523	ABC transporter ATP-binding protein

LA_1332	ankyrin repeat-containing protein	LA_3553	tRNA-dihydrouridine synthase A
LA_1335	protein kinase	LA_3565	glycerol kinase
LA_1339	permease	LA_3573	3-deoxy-manno-octulosonate cytidylyltransferase
LA_1340	sodium bile acid symporter	LA_3577	endoflagellar motor protein A
LA_1357	hypothetical protein	LA_3581	dioxygenase
LA_1363	transcriptional regulator	LA_3582	gamma-glutamyl carboxylase
LA_1365	MBOAT family D-alaninealginate export/acetyltransferase	LA_3583	DNA-processing protein A
LA_1366	hypothetical protein	LA_3596	ATP-dependent Lon protease
LA_1384	hypothetical protein	LA_3597	pyridoxal phosphate-dependent aminotransferase
LA_1389	hypothetical protein	LA_3603	sulfatase
LA_1390	hypothetical protein	LA_3604	lysophospholipase
LA_1391	50S ribosomal protein L11 methyltransferase	LA_3610	serine phosphatase RsbU
LA_1397	export protein	LA_3614	signal transduction protein
LA_1412	hypothetical protein	LA_3616	divalent ion tolerance protein
LA_1419	guanylate cyclase	LA_3625	DNA polymerase I
LA_1435	hypothetical protein	LA_3629	anthranilate synthase component II
LA_1437	glucose kinase	LA_3630	multidrug ABC transporter ATPase/permease
LA_1439	hypothetical protein	LA_3631	two-component system response regulator
LA_1440	hypothetical protein	LA_3632	FOG:HEAT repeat protein
LA_1448	hypothetical protein	LA_3634	GTP-binding protein
LA_1449	carboxy-terminal processing protease	LA_3636	cyclic amidohydrolase
LA_1451	CDP-diacylglycerol--serine O- phosphatidyltransferase	LA_3652	RNA polymerase ECF-type sigma factor
LA_1456	DNA repair protein RadC	LA_3672	hydrolase/acyltransferase
LA_1464	UDP-N-acetylmuramyl pentapeptide synthase	LA_3692	membrane carboxypeptidase/penicillin -binding protein
LA_1466	oligo-1,6-glucosidase	LA_3693	dipeptide ABC transporter substrate-binding protein
LA_1476	hypothetical protein	LA_3703	heat shock operon regulator HrcA
LA_1480	PLP dependent enzyme class III	LA_3704	heat shock protein GrpE
LA_1481	pyrroline-5-carboxylate reductase	LA_3712	ABC transporter permease

LA_1487	tRNA-specific 2-thiouridylase MnmA	LA_3744	hypothetical protein
LA_1489	hypothetical protein	LA_3745	hypothetical protein
LA_1491	Na+/H+ antiporter	LA_3751	hypothetical protein
LA_1492	Na+/H+ antiporter	LA_3763	valyl-tRNA synthetase
LA_1495	hypothetical protein	LA_3789	hypothetical protein
LA_1501	hypothetical protein	LA_3794	hypothetical protein
LA_1507	hypothetical protein	LA_3809	hypothetical protein
LA_1508	hypothetical protein	LA_3811	spermidine synthase
LA_1511	L-aspartate oxidase	LA_3824	4-diphosphocytidyl-2C-methyl-D-erythritol kinase
LA_1513	carboxy-terminal processing protease	LA_3825	transcriptional regulator
LA_1516	tRNA isopentenyltransferase	LA_3829	hypothetical protein
LA_1517a	hypothetical protein	LA_3833	hypothetical protein
LA_1521	chemotaxis protein CheW	LA_3856	hypothetical protein
LA_1528	two-component system response regulator	LA_3869	response regulator
LA_1532	fructose-bisphosphate aldolase	LA_3871	acriflavine resistance protein
LA_1544	hypothetical protein	LA_3871a	hypothetical protein
LA_1549	beta-lactamase regulatory protein 1	LA_3872	periplasmic serine protease
LA_1553	hypothetical protein	LA_3873	hypothetical protein
LA_1555	multicopper oxidase-like protein	LA_3885	galactokinase
LA_1564	Hsp20/alpha crystallin molecular chaperone	LA_3893	hypothetical protein
LA_1667	sodium:sulfate symporter	LA_3916	hypothetical protein
LA_1674	poly-gamma-glutamate synthesis protein	LA_3918	methionyl-tRNA synthetase
LA_1676	single-stranded DNA-binding protein	LA_3921	UTP:GlnB (protein PII) uridylyltransferase
LA_1679	replicative DNA helicase	LA_3924	NAD(P)H steroid dehydrogenase
LA_1681	phosphate starvation-inducibleprotein	LA_3927	TolC family protein
LA_1684	metal-dependent hydrolase	LA_3929	response regulator
LA_1688	arginyl-tRNA synthetase	LA_3935	DNA/pantothenate metabolism flavoprotein
LA_1689	competence-damage inducible protein CinA	LA_3940	hypothetical protein
LA_1690	hypothetical protein	LA_3945	ATP-dependent DNA helicase RecG
LA_1694	hypothetical protein	LA_3971	potassium uptake transporter
LA_1709	two-component system response regulator	LA_3976	hypothetical protein
LA_1710	two-component system sensor histidine kinase	LA_3980	glycerophosphodiester phosphodiesterase

LA_1711	two-component system sensor histidine kinase	LA_3982	hypothetical protein
LA_1721	DNA topoisomerase	LA_3993	hypothetical protein
LA_1724	ABC transporter ATP-binding protein	LA_3994	Fe-S oxidoreductase
LA_1726	HD family protein	LA_3996	fused two-component system sensor histidine kinase/response regulator
LA_1727	hypothetical protein	LA_3998	cholesterol oxidase
LA_1733	hypothetical protein	LA_3999	choline dehydrogenase
LA_1735	MviN-like protein	LA_4008	guanylate cyclase
LA_1740	hypothetical protein	LA_4009	cholesterol oxidase precursor
LA_1861	hypothetical protein	LA_4011	hypothetical protein
LA_1862	23S rRNA (guanosine(2251)-2'-O)-methyltransferase RlmB	LA_4013	Co/Zn/Cd efflux pump
LA_1863	cysteinyl-tRNA synthetase	LA_4023	hypothetical protein
LA_1864	acetyl esterase	LA_4035	phenylalanyl-tRNA synthetase subunit beta
LA_1865	bifunctional 5,10-methylene-tetrahydrofolate dehydrogenase/ 5,10-methylene-tetrahydrofolate cyclohydrolase	LA_4152	hypothetical protein
LA_1867	hypothetical protein	LA_4159	fatty acid synthase subunit beta
LA_1875	ATP-dependent helicase	LA_4167	phosphopantetheine adenylyltransferase
LA_1880	transcriptional regulator	LA_4174	two-component system sensor histidine kinase
LA_1883	hypothetical protein	LA_4179	hypothetical protein
LA_1887	hypothetical protein	LA_4180	aminotransferase
LA_1899	aldo/keto reductase	LA_4183	endoflagellar motor protein
LA_1900	hypothetical protein	LA_4185	hypothetical protein
LA_1910	hypothetical protein	LA_4189	anti-sigma regulatory factor
LA_1912	hypothetical protein	LA_4200	L-amino acid oxidase
LA_1916	hypothetical protein	LA_4203	hypothetical protein
LA_1917	hypothetical protein	LA_4209	hypothetical protein
LA_1924	hypothetical protein	LA_4210	chloride channel protein EriC
LA_1930	acyl-CoA dehydrogenase	LA_4212	hypothetical protein
LA_1932	hypothetical protein	LA_4216	sulfite reductase subunit beta
LA_1934	hypothetical protein	LA_4226	hypothetical protein
LA_1935	hypothetical protein	LA_4227	hydrolase
LA_1947	hypothetical protein	LA_4232	hypothetical protein

LA_1948	two-component system response regulator	LA_4236	DNA mismatch repair protein ATPase component
LA_1950	hypothetical protein	LA_4246	alkaline phosphatase
LA_1955	alpha-beta hydrolase family esterase	LA_4247	hydrolase
LA_1956	hypothetical protein	LA_4249	methyl-accepting chemotaxis protein
LA_1959	bifunctional GMP synthase/glutamine amidotransferase	LA_4261	dTDP-4-dehydrorhamnose 3,5-epimerase
LA_1968	hypothetical protein	LA_4263	hypothetical protein
LA_1970	[cytidine(C)-cytidine(C)-adenosine (A)]-adding enzyme	LA_4264	glycerol-3-phosphate dehydrogenase
LA_1974	arginyl-tRNA-protein transferase	LA_4266	phosphoribosylglycinamide formyltransferase 2
LA_1981	hypothetical protein	LA_4273	formate hydrogenase subunit B
LA_1985	hypothetical protein	LA_4286	hypothetical protein
LA_1986	inosine-5'-monophosphate dehydrogenase	LA_4288	ABC transporter ATP-binding protein
LA_1987	hypothetical protein	LA_4290	dihydroorotate dehydrogenase 2
LA_1988	16S rRNA (uracil(1498)-N(3))-methyltransferase	LA_4291	hypothetical protein
LA_1995	lysyl-tRNA synthetase	LA_4293	hypothetical protein
LA_2010	pyruvate dehydrogenase subunit alpha	LA_4300	Hsp31/PfpI family protein
LA_2024	hypothetical protein	LA_4301	sodium/hydrogen antiporter
LA_2030	hypothetical protein	LA_4310	hypothetical protein
LA_2047	UDP-N-acetylmuramoylalanyl-D-glutamate--2,6-diaminopimelate ligase	LA_4314	hypothetical protein
LA_2048	phospho-N-acetylmuramoyl-pentapeptide-transferase	LA_4322	carboxypeptidase I
LA_2050	UDP-diphospho-muramoylpentapeptide beta-N-acetylglucosaminyltransferase	LA_4327	ferredoxin-NADP reductase
LA_2051	UDP-N-acetylmuramate--L-alanine ligase	LA_4335	hypothetical protein
LA_2054	3-hydroxyisobutyrate dehydrogenase	LA_4340	Mg-dependent DNase
LA_2063	cytoplasmic membrane protein	LA_4349	metallopeptidase
LA_2066	hypothetical protein	LB_005	hypothetical protein
LA_2077	peptidyl-prolyl cis-trans isomerase	LB_010	glutamyl-tRNA reductase

LA_2083	hypothetical protein	LB_016	uroporphyrinogen decarboxylase
LA_2091	hypothetical protein	LB_017	coproporphyrinogen III oxidase
LA_2094	FHA domain-containing protein	LB_018	hypothetical protein
LA_2106	gamma-glutamylcysteine synthetase	LB_020	protoporphyrinogen oxidase
LA_2107	ATP-grasp enzyme	LB_024	ferrochelatase
LA_2113	ABC transporter permease	LB_025	hypothetical protein
LA_2114	ABC transporter ATP-binding protein	LB_027	chromosome partitioning protein ParB
LA_2120	hypothetical protein	LB_034	3-oxoacyl-ACP reductase
LA_2124	MBOAT family D-alaninealginate export/acetyltransferase	LB_036	NADH dehydrogenase
LA_2132	hydrolase	LB_047	hypothetical protein
LA_2133	sodium:solute symporter	LB_050	hypothetical protein
LA_2136	hypothetical protein	LB_052	hypothetical protein
LA_2138	oxidoreductase	LB_054	von Willebrand factor type A domain-containing protein
LA_2142	adenosylmethionine--8-amino-7-oxononanoate aminotransferase BioA	LB_057	oxygen tolerance protein BatD
LA_2144	SGNH-hydrolase	LB_060	hypothetical protein
LA_2153	acetylornithine aminotransferase	LB_062	anti-anti-sigma factor
LA_2162	50S ribosomal protein L28	LB_078	hypothetical protein
LA_2166	excinuclease ABC subunit C	LB_086	3-dehydroquinate synthase
LA_2171	exopolysaccharide production protein	LB_087	hypothetical protein
LA_2177	long-chain-fatty-acid CoA ligase	LB_089	hypothetical protein
LA_2178	N-acetyl-gamma-glutamyl-phosphate reductase	LB_094	adenylate cyclase
LA_2184	Zn-dependent protease	LB_110	outermembrane protein
LA_2187	penicillin-binding protein	LB_116	hypothetical protein
LA_2193	DNA helicase UvrD	LB_121	3-oxoacyl-ACP reductase
LA_2197	2,4-dienoyl-CoA reductase	LB_125	chemotaxis protein
LA_2202	2-isopropylmalate synthase	LB_127	hypothetical protein
LA_2208	hypothetical protein	LB_130	araC family transcriptional regulator
LA_2211	periplasmic serine protease	LB_131	araC family transcriptional regulator
LA_2215	endoflagellar motor protein B	LB_138	hypothetical protein
LA_2217	magnesium transporter	LB_139	serine phosphatase RsbU
LA_2219	hypothetical protein	LB_142	hypothetical protein

LA_2223	two-component system sensor histidine kinase	LB_143	lipoprotein qlp42
LA_2227	hypothetical protein	LB_170	capsular polysaccharide biosynthesis protein
LA_2256	MBOAT family D-alaninealginate export/acetyltransferase	LB_171	hypothetical protein
LA_2260	transcriptional regulator	LB_177	serine protease
LA_2266	hypothetical protein	LB_187	hypothetical protein
LA_2267	TPR-repeat-containing protein	LB_191	TonB-dependent hemin-binding protein
LA_2268	hypothetical protein	LB_192	hypothetical protein
LA_2278	hypothetical protein	LB_196	lipoprotein
LA_2286	translaldolase	LB_199	outermembrane protein
LA_2308	hypothetical protein	LB_202	queuosine biosynthesis protein
LA_2309	long-chain-fatty-acid CoA ligase	LB_208	alginate o-acetyltransferase
LA_2314	mannosyltransferase	LB_210	alginate o-acetyltransferase
LA_2315	3'(2'),5'-bisphosphate nucleotidase CysQ	LB_211	hypothetical protein
LA_2318	hypothetical protein	LB_215	ubiquinone/menaquinone biosynthesis methlytransferase
LA_2321	DNA repair protein RecN	LB_219	hypothetical protein
LA_2323	hypothetical protein	LB_222	hypothetical protein
LA_2335	3-polyprenyl-4-hydroxybenzoate decarboxylase	LB_242	hypothetical protein
LA_2338	phosphohydrolase	LB_245	sodium:solute symporter
LA_2347	integrase/recombinase XerD	LB_250	hypothetical protein
LA_2351	Smr domain-containing protein	LB_251	hypothetical protein
LA_2353	hypothetical protein	LB_258	cysteine protease
LA_2355	permease	LB_262	UDP-glucose pyrophosphorylase
LA_2356	exodeoxyribonuclease VII large subunit	LB_263	hypothetical protein
LA_2367	ATPase	LB_264	hypothetical protein
LA_2368	type II secretion system protein K	LB_267	aminotransferase
LA_2369	type II secretory pathway component	LB_272	hypothetical protein
LA_2370	hypothetical protein	LB_279	hypothetical protein
LA_2375	type II secretory pathway component protein D	LB_285	hypothetical protein
LA_2376	type II secretory pathway component protein C	LB_289	hypothetical protein
LA_2377	M23 family metalloendopeptidase	LB_291	3-oxoacyl-ACP reductase

LA_2379	Mg chelatase subunit ChII	LB_301	hypothetical protein
LA_2381	hypothetical protein	LB_317	alanine racemase
LA_2401	two-component system sensor histidine kinase	LB_319	hypothetical protein
LA_2405	ABC transporter ATP-binding protein	LB_322	two-component system sensor histidine kinase/response regulator
LA_2406	hypothetical protein	LB_327	aconitate hydratase
LA_2409	CTP synthetase	LB_330	hypothetical protein
LA_2413	cell wall-associated hydrolase/lipoprotein	LB_332	hypothetical protein
LA_2421	two-component system sensor histidine kinase	LB_354	hypothetical protein
LA_2422	fused two-component system sensor histidine kinase/response regulator	LB_355	aspartate-semialdehyde dehydrogenase
LA_2433	acetyl-CoA carboxylase	LB_359	alginate O-acetylation protein
LA_2434	anti-sigma factor antagonist	LB_363	hypothetical protein
LA_2435	serine/threonine phosphatase	LB_364	FIS family transcriptional regulator

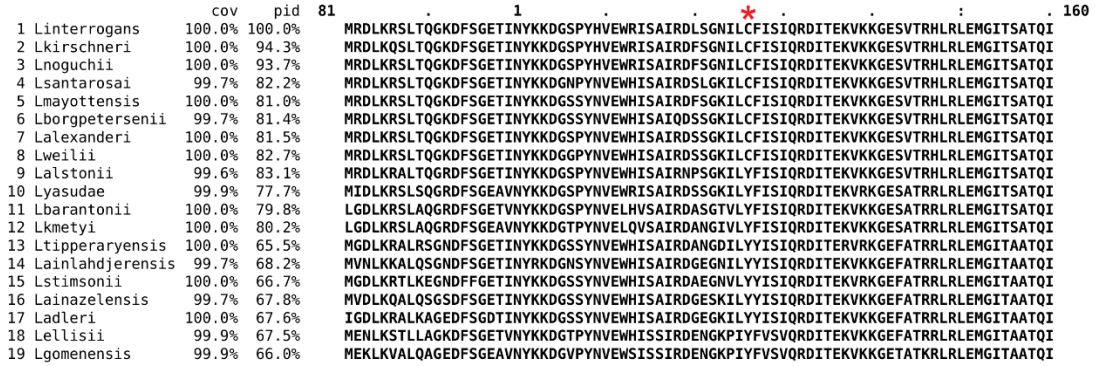

Figure S3.1. Protein sequence alignment of the histidine kinase LA_2421 between P1 species.

Amino acid changes in LA_2421 were only found in $P1^+$ *vs* $P1^-$ comparison. $P1^+$species comprise *L. interrogans, L. kirschneri, L. noguchii, L. santarosai, L. mayottensis, L. borgpetersenii, L. alexanderi, and L. weilii.* $P1^-$ includes *L. alstonii, L. yasudae, L. barantonii, L. kmetyi, L. tipperaryensis, L. ainlahdjerensis, L. stimsonii, L. ainazelensis, L. adleri, L. ellisii and L. gomenensis.* The amino acid permutation detected (Y121C, $P1^- \rightarrow P1^+$) is indicated by a red asterisk. Only the region containing the amino acid change is shown.

Figure S3.2. Protein sequence alignment of the response regulator LA_2422 between P1 species.

Amino acid changes in LA_2422 were only found in $P1^+$ *vs* $P1^-$ comparison. $P1^+$species comprise *L. interrogans, L. kirschneri, L. noguchii, L. santarosai, L. mayottensis, L. borgpetersenii, L. alexanderi, and L. weilii.* $P1^-$ includes *L. alstonii, L. yasudae, L. barantonii, L. kmetyi, L. tipperaryensis, L. ainlahdjerensis, L. stimsonii, L. ainazelensis, L. adleri, L. ellisii and L. gomenensis.* The amino acid permutation detected (S/P249F, $P1^-$→$P1^+$) is indicated by a red asterisk. Only the region containing the amino acid change is shown.

Figure S3.3. Protein sequence alignment of the histidine kinase LA_2540 between P1 species.

Amino acid changes in LA_2540 were only found in $P1^+$ *vs* $P1^-$ comparison. $P1^+$species comprise *L. interrogans, L. kirschneri, L. noguchii, L. santarosai, L. mayottensis, L. borgpetersenii, L. alexanderi, and L. weilii.* $P1^-$ includes *L. alstonii, L. yasudae, L. barantonii, L. kmetyi, L. tipperaryensis, L. ainlahdjerensis, L. stimsonii, L. ainazelensis, L. adleri, L. ellisii and L. gomenensis.* The amino acid permutation detected (R8H, $P1^-$→$P1^+$) is indicated by a red asterisk. Only the region containing the amino acid change is shown.

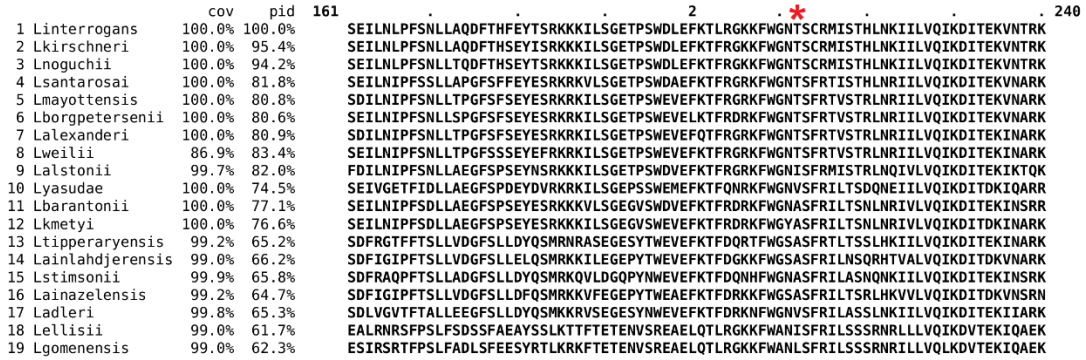

Figure S3.4. Protein sequence alignment of the response regulator LA_2541 between P1 species.

Amino acid changes in LA_2541 were only found in $P1^+$ *vs* $P1^-$ comparison. $P1^+$species comprise *L. interrogans, L. kirschneri, L. noguchii, L. santarosai, L. mayottensis, L. borgpetersenii, L. alexanderi, and L. weilii.* $P1^-$ includes *L. alstonii, L. yasudae, L. barantonii, L. kmetyi, L. tipperaryensis, L. ainlahdjerensis, L. stimsonii, L. ainazelensis, L. adleri, L. ellisii and L. gomenensis.* The amino acid permutation detected (I/V/A/L210T, $P1^-\rightarrow P1^+$) is indicated by a red asterisk. Only the region containing the amino acid change is shown.

Appendix II. Supplementary material of Chapter 4

Here are the supplemental tables and figures of the article presented in Chapter 4.

Table S4.1. ANI figures among *L. noguchii* strains.

Ordered according to phylogenetic proximity (see **Figure 4.2**). The ANI values are presented in a square matrix, where the strain names listed in the first column are, in the same order, the strains from left to right.

201102933		99.9	99.9	98.3	98.2	98.1	98.2	98.2	95.6	95.4	95.5	95.4	95.3	95.4	95.3	95.3	95.8	95.4	95.4	95.4	95.8	95.8	95.7	95.8	95.7	96.1	95.8	95.8	95.6	95.7	95.7	96.1	96.3	96.0	96.1	96.1	89.4	90.0
bajan	99.9		99.9	98.3	98.2	98.0	98.2	98.2	95.6	95.4	95.4	95.4	95.3	95.4	95.4	95.4	95.8	95.4	95.4	95.4	95.7	95.8	95.7	95.8	95.7	96.2	95.6	95.7	95.5	95.7	95.6	96.0	96.3	95.9	96.0	96.1	89.5	90.0
barbudensis	99.9	99.9		98.3	98.1	98.1	98.1	98.2	95.6	95.5	95.4	95.4	95.3	95.5	95.4	95.3	95.7	95.4	95.5	95.4	95.8	95.8	95.7	95.8	95.7	96.1	95.7	95.8	95.6	95.6	95.8	96.0	96.3	96.0	96.1	96.0	89.4	90.1
56189	98.3	98.3	98.3		98.7	98.7	98.7	98.7	95.5	95.4	95.3	95.2	95.2	95.3	95.3	95.2	95.5	95.3	95.3	95.3	95.6	95.6	95.6	95.6	95.7	96.0	95.6	95.7	95.7	95.7	95.7	96.0	96.3	96.0	96.0	96.0	89.4	90.1
56682	98.2	98.2	98.1	98.7		100.0	99.9	98.7	95.6	95.5	95.4	95.3	95.3	95.4	95.4	95.4	95.7	95.4	95.5	95.5	95.8	95.8	95.7	95.7	95.7	96.1	95.7	95.8	95.7	95.7	95.8	96.0	96.3	96.1	96.1	96.1	89.4	90.1
CZ214	98.1	98.0	98.1	98.7	100.0		99.9	98.7	95.5	95.3	95.2	95.3	95.1	95.4	95.4	95.4	95.7	95.5	95.5	95.3	95.5	95.7	95.6	95.7	95.6	96.1	95.6	95.7	95.6	95.7	95.7	96.0	96.3	96.0	96.2	96.1	89.5	90.1
lyse	98.2	98.2	98.1	98.7	99.9	99.9		98.6	95.6	95.5	95.3	95.3	95.2	95.4	95.4	95.4	95.7	95.4	95.4	95.4	95.7	95.7	95.6	95.7	95.7	96.1	95.6	95.8	95.6	95.7	95.8	96.0	96.3	96.0	96.1	96.1	89.5	90.1
HAI1536	98.2	98.2	98.2	98.7	98.7	98.7	98.6		95.4	95.3	95.3	95.2	95.2	95.3	95.3	95.3	95.8	95.4	95.3	95.3	95.6	95.7	95.5	95.5	95.4	95.9	95.5	95.6	95.6	95.7	95.7	95.9	96.3	96.0	96.0	96.1	89.4	90.3
2006001870	95.6	95.6	95.6	95.5	95.6	95.5	95.6	95.4		98.8	98	98.0	97.8	98.1	98.4	98.3	98.5	98.3	98.4	98.4	98.6	98.6	98.7	98.5	98.4	98.5	98.5	98.6	98.6	98.6	98.6	98.1	98.1	97.4	97.6	97.6	89.5	90.6
IP1705032	95.4	95.4	95.5	95.4	95.5	95.3	95.5	95.3	98.8		98.0	98.1	97.9	98.1	98.5	98.4	98.7	98.5	98.5	98.5	98.7	98.8	98.7	98.5	98.5	98.6	98.6	98.7	98.5	98.5	98.6	98.0	98.0	97.3	97.4	97.5	89.6	90.4
Bonito	95.5	95.4	95.4	95.3	95.4	95.2	95.3	95.3	98.0	98.0		99.3	99.0	98.9	98.0	98.1	98.3	98.0	98.1	98.1	97.8	98.0	98.0	97.8	97.7	98.0	97.7	97.8	97.8	97.9	97.9	97.5	97.5	97.2	97.3	97.3	89.7	90.6
IP1708035	95.4	95.4	95.4	95.2	95.3	95.3	95.3	95.2	98.0	98.1	99.3		98.9	98.9	98.1	98.1	98.3	98.0	99.0	98.1	97.9	98.0	98.0	97.8	97.7	97.9	97.8	98.0	97.7	97.9	98.0	97.6	97.6	97.2	97.4	97.3	89.7	90.5
IP1709037	95.3	95.3	95.3	95.2	95.3	95.1	95.2	95.2	97.8	97.9	99.0	98.9		98.9	97.9	97.9	98.1	97.9	98.0	97.9	97.8	97.9	97.9	97.8	97.7	97.9	97.6	97.8	97.6	97.7	97.8	97.5	97.6	97.1	97.2	97.2	89.6	90.6
Cascata	95.4	95.4	95.5	95.3	95.4	95.4	95.4	95.3	98.1	98.1	98.9	98.9	98.9		98.1	98.1	98.2	98.1	98.2	98.1	97.9	98.1	98.1	97.9	97.8	98.1	97.9	98.1	97.8	97.9	98.0	97.6	97.5	97.2	97.3	97.4	89.7	90.7
IP1512017	95.3	95.4	95.4	95.3	95.4	95.4	95.4	95.3	98.4	98.5	98.0	98.1	97.9	98.1		99.1	99.0	99.0	99.0	99.0	98.4	98.4	98.4	98.2	98.1	98.3	98.1	98.3	98.1	98.2	98.4	97.9	97.9	97.3	97.4	97.4	89.6	90.5
IP1605021	95.3	95.4	95.3	95.2	95.4	95.4	95.4	95.3	98.3	98.4	98.1	98.1	97.9	98.1	99.1		99.0	98.9	98.9	98.9	98.3	98.4	98.4	98.1	98.1	98.2	98.2	98.3	98.1	98.2	98.2	97.8	97.9	97.3	97.4	97.4	89.6	90.5
IP1611025	95.8	95.8	95.7	95.5	95.7	95.7	95.7	95.8	98.5	98.7	98.3	98.3	98.1	98.2	99.0	99.0		99.1	99.1	99.1	98.5	98.5	98.5	98.2	98.3	98.3	98.3	98.4	98.3	98.3	98.4	97.9	97.9	97.4	97.6	97.6	89.9	90.7
IP1804061	95.4	95.4	95.4	95.3	95.4	95.5	95.4	95.4	98.3	98.5	98.0	98.0	97.9	98.1	99.0	98.9	99.1		99.0	99.0	98.3	98.4	98.4	98.2	98.0	98.3	98.2	98.4	98.1	98.2	98.3	97.9	97.9	97.2	97.5	97.5	89.5	90.4
IP1703027	95.4	95.4	95.5	95.3	95.5	95.5	95.4	95.3	98.4	98.5	98.1	99.0	98.0	98.2	99.0	98.9	99.1	99.0		99.9	98.3	98.5	98.5	98.2	98.2	98.3	98.2	98.4	98.1	98.3	98.3	97.8	97.8	97.3	97.4	97.5	89.5	90.4
IP1712055	95.4	95.4	95.4	95.3	95.5	95.3	95.4	95.3	98.4	98.5	98.1	98.1	97.9	98.1	99.0	98.9	99.1	99.0	99.9		98.3	98.5	98.4	98.2	98.1	98.3	98.2	98.4	98.1	98.2	98.3	97.8	97.8	97.3	97.4	97.6	89.5	90.5
201700036	95.8	95.7	95.8	95.6	95.8	95.5	95.7	95.6	98.6	98.7	97.8	97.9	97.8	97.9	98.4	98.3	98.5	98.3	98.3	98.3		99.7	98.7	98.5	98.4	98.5	98.5	98.6	98.4	98.4	98.5	98.0	98.0	97.4	97.5	97.6	89.5	90.4
U73	95.8	95.8	95.8	95.6	95.8	95.7	95.7	95.7	98.6	98.8	98.0	98.0	97.9	98.1	98.4	98.4	98.5	98.4	98.5	98.5	99.7		98.8	98.6	98.6	98.6	98.6	98.7	98.4	98.5	98.6	98.2	98.1	97.5	97.7	97.7	89.7	90.6
Hook	95.7	95.7	95.7	95.6	95.7	95.6	95.6	95.5	98.7	98.7	98.0	98.0	97.9	98.1	98.4	98.4	98.5	98.4	98.5	98.4	98.7	98.8		98.6	98.5	98.6	98.7	98.7	98.5	98.5	98.6	98.1	98.1	97.4	97.6	97.7	89.6	90.5
201700038	95.8	95.8	95.8	95.6	95.7	95.7	95.7	95.5	98.5	98.5	97.8	97.8	97.8	97.9	98.2	98.1	98.2	98.2	98.2	98.2	98.5	98.6	98.6		99.3	98.9	98.9	98.5	98.3	98.4	98.5	98.2	98.1	97.5	97.6	97.7	89.6	90.5
ZUN142	95.7	95.7	95.7	95.7	95.7	95.6	95.7	95.4	98.4	98.5	97.7	97.7	97.7	97.8	98.1	98.1	98.3	98.0	98.2	98.1	98.4	98.6	98.5	99.3		99.0	98.9	98.6	98.3	98.3	98.4	98.1	98.1	97.3	97.5	97.6	89.5	90.4
soteropolitana	96.1	96.2	96.1	96.0	96.1	96.1	96.1	95.9	98.5	98.6	98	97.9	97.9	98.1	98.3	98.2	98.3	98.3	98.3	98.3	98.5	98.6	98.6	98.9	99.0		98.9	98.6	98.4	98.4	98.5	98.1	98.1	97.6	97.7	97.7	89.7	90.6
IP1611024	95.8	95.6	95.7	95.6	95.7	95.6	95.6	95.5	98.5	98.6	97.7	97.8	97.6	97.9	98.1	98.2	98.3	98.2	98.2	98.2	98.5	98.6	98.7	98.9	98.9	98.9		98.7	98.4	98.4	98.5	98.1	98.1	97.4	97.5	97.6	89.5	90.4
56271	95.8	95.7	95.8	95.7	95.8	95.7	95.8	95.6	98.6	98.7	97.8	98.0	97.8	98.1	98.3	98.3	98.4	98.4	98.4	98.4	98.6	98.7	98.7	98.5	98.6	98.6	98.7		98.5	98.5	98.6	98.1	98.2	97.6	97.7	97.7	89.6	90.5
201601331	95.6	95.5	95.6	95.7	95.7	95.6	95.6	95.6	98.6	98.5	97.8	97.7	97.6	97.8	98.1	98.1	98.3	98.1	98.1	98.1	98.4	98.4	98.5	98.3	98.3	98.4	98.4	98.5		99.8	99.9	98.2	98.2	97.5	97.7	97.7	89.5	90.4
rushan	95.7	95.7	95.6	95.7	95.7	95.7	95.7	95.7	98.6	98.5	97.9	97.9	97.7	97.9	98.2	98.2	98.3	98.2	98.3	98.2	98.4	98.5	98.5	98.4	98.3	98.4	98.4	98.5	99.8		99.9	98.2	98.1	97.5	97.7	97.8	89.6	90.4
56183	95.7	95.6	95.8	95.7	95.8	95.7	95.8	95.7	98.6	98.6	97.9	98.0	97.8	98.0	98.4	98.2	98.4	98.3	98.3	98.3	98.5	98.6	98.6	98.5	98.4	98.5	98.5	98.6	99.9	99.9		98.2	98.2	97.6	97.8	97.8	89.6	90.5
56190	96.1	96.0	96.0	96.0	96.0	96	96.0	95.9	98.1	98.0	97.5	97.6	97.5	97.6	97.9	97.8	97.9	97.9	97.8	97.8	98.0	98.2	98.1	98.2	98.1	98.1	98.1	98.1	98.2	98.2	98.2		99.8	97.9	98.0	98.0	89.6	90.5
nicaragua	96.3	96.3	96.3	96.3	96.3	96.3	96.3	96.3	98.1	98.0	97.5	97.6	97.6	97.5	97.9	97.9	97.9	97.9	97.8	97.8	98.0	98.1	98.1	98.1	98.1	98.1	98.1	98.2	98.2	98.1	98.2	99.8		97.9	98.0	98.0	89.8	90.5
2001034031	96.0	95.9	96.0	96.0	96.1	96.0	96.0	96.0	97.4	97.3	97.2	97.2	97.1	97.2	97.3	97.3	97.4	97.2	97.3	97.3	97.4	97.5	97.4	97.5	97.3	97.6	97.4	97.6	97.5	97.5	97.6	97.9	97.9		98.7	98.7	89.6	90.4
1993005606	96.1	96.0	96.1	96.0	96.1	96.2	96.1	96.0	97.6	97.4	97.3	97.4	97.2	97.3	97.4	97.4	97.6	97.5	97.4	97.4	97.5	97.7	97.6	97.6	97.5	97.7	97.5	97.7	97.7	97.7	97.8	98.0	98.0	98.7		99.4	89.5	90.5
2007001578	96.1	96.1	96.0	96.0	96.1	96.1	96.1	96.1	97.6	97.5	97.3	97.3	97.2	97.4	97.4	97.4	97.6	97.5	97.5	97.6	97.6	97.7	97.7	97.7	97.6	97.7	97.6	97.7	97.7	97.7	97.8	97.8	98.0	98.0	98.7	99.4		89.6	90.4
***L.i* 56601**	89.4	89.5	89.4	89.4	89.4	89.5	89.5	89.4	89.5	89.6	89.7	89.7	89.6	89.7	89.6	89.6	89.9	89.5	89.5	89.5	89.5	89.7	89.6	89.6	89.5	89.7	89.5	89.6	89.5	89.6	89.6	89.6	89.8	89.6	89.5	89.6		91.0
***L.k* 200702274**	90.0	90.0	90.1	90.1	90.1	90.1	90.1	90.3	90.6	90.4	90.6	90.5	90.6	90.7	90.5	90.5	90.7	90.4	90.4	90.5	90.4	90.6	90.5	90.5	90.4	90.6	90.4	90.5	90.4	90.4	90.5	90.5	90.5	90.4	90.5	90.4	91.0	

Table S4.2. Functional mapping of strain-specific genes.

Genes present only in one *L. noguchii* strain and absent in the others, were used to map them onto KEGG orthology (KO) groups, which are nodes in KEGG pathway maps. The search and mapping were done as described in Methods. KO groups are displayed clustered according to different pathways, themselves listed in descending order according to the number of hits for each pathway.

map01100 Metabolic pathways (16)

ko:K00721 DPM1; dolichol-phosphate mannosyltransferase [EC:2.4.1.83]
ko:K00979 kdsB; 3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase) [EC:2.7.7.38]
ko:K01790 rfbC; dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13]
ko:K01953 asnB; asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]
ko:K03270 kdsC; 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (KDO 8-P phosphatase) [EC:3.1.3.45]
ko:K08679 GAE; UDP-glucuronate 4-epimerase [EC:5.1.3.6]
ko:K10532 HGSNAT; heparan-alpha-glucosaminide N-acetyltransferase [EC:2.3.1.78]
ko:K13020 wlbA; UDP-N-acetyl-2-amino-2-deoxyglucuronate dehydrogenase [EC:1.1.1.335]
ko:K15897 pseG; UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase [EC:3.6.1.57]
ko:K18429 legG; GDP/UDP-N,N'-diacetylbacillosamine 2-epimerase (hydrolysing) [EC:3.2.1.184]
ko:K18430 legI; N,N'-diacetyllegionaminate synthase [EC:2.5.1.101]
ko:K18431 legF; CMP-N,N'-diacetyllegionaminic acid synthase [EC:2.7.7.82]
ko:K24305 lea6; UDP-2-acetamido-4-(D-alanylamino)-2,4,6-trideoxy-alpha-D-mannopyranose hydrolase
ko:K01652 E2.2.1.6L; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]
ko:K01735 aroB; 3-dehydroquinate synthase [EC:4.2.3.4]
ko:K07816 E2.7.6.5; GTP pyrophosphokinase [EC:2.7.6.5]

map01250 Biosynthesis of nucleotide sugars (10)

ko:K00979 kdsB; 3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase) [EC:2.7.7.38]
ko:K01790 rfbC; dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13]
ko:K03270 kdsC; 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (KDO 8-P phosphatase) [EC:3.1.3.45]
ko:K08679 GAE; UDP-glucuronate 4-epimerase [EC:5.1.3.6]
ko:K13020 wlbA; UDP-N-acetyl-2-amino-2-deoxyglucuronate dehydrogenase [EC:1.1.1.335]
ko:K15897 pseG; UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase [EC:3.6.1.57]

ko:K18429 legG; GDP/UDP-N,N'-diacetylbacillosamine 2-epimerase (hydrolysing) [EC:3.2.1.184]
ko:K18430 legI; N,N'-diacetyllegionaminate synthase [EC:2.5.1.101]
ko:K18431 legF; CMP-N,N'-diacetyllegionaminic acid synthase [EC:2.7.7.82]
ko:K24305 lea6; UDP-2-acetamido-4-(D-alanylamino)-2,4,6-trideoxy-alpha-D-mannopyranose hydrolase

map00520 Amino sugar and nucleotide sugar metabolism (6)

ko:K08679 GAE; UDP-glucuronate 4-epimerase [EC:5.1.3.6]
ko:K13020 wlbA; UDP-N-acetyl-2-amino-2-deoxyglucuronate dehydrogenase [EC:1.1.1.335]
ko:K15897 pseG; UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase [EC:3.6.1.57]
ko:K18429 legG; GDP/UDP-N,N'-diacetylbacillosamine 2-epimerase (hydrolysing) [EC:3.2.1.184]
ko:K18430 legI; N,N'-diacetyllegionaminate synthase [EC:2.5.1.101]
ko:K18431 legF; CMP-N,N'-diacetyllegionaminic acid synthase [EC:2.7.7.82]

map00541 O-Antigen nucleotide sugar biosynthesis (5)

ko:K01790 rfbC; dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13]
ko:K08679 GAE; UDP-glucuronate 4-epimerase [EC:5.1.3.6]
ko:K13020 wlbA; UDP-N-acetyl-2-amino-2-deoxyglucuronate dehydrogenase [EC:1.1.1.335]
ko:K15897 pseG; UDP-2,4-diacetamido-2,4,6-trideoxy-beta-L-altropyranose hydrolase [EC:3.6.1.57]
ko:K24305 lea6; UDP-2-acetamido-4-(D-alanylamino)-2,4,6-trideoxy-alpha-D-mannopyranose hydrolase

map01110 Biosynthesis of secondary metabolites (4)

ko:K01652 E2.2.1.6L; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]
ko:K01735 aroB; 3-dehydroquinate synthase [EC:4.2.3.4]
ko:K01790 rfbC; dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13]
ko:K01953 asnB; asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]

map01230 Biosynthesis of amino acids (3)

ko:K01652 E2.2.1.6L; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]

ko:K01735 aroB; 3-dehydroquinate synthase [EC:4.2.3.4]

ko:K01953 asnB; asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]

map00540 Lipopolysaccharide biosynthesis (2)

ko:K00979 kdsB; 3-deoxy-manno-octulosonate cytidylyltransferase (CMP-KDO synthetase) [EC:2.7.7.38]

ko:K03270 kdsC; 3-deoxy-D-manno-octulosonate 8-phosphate phosphatase (KDO 8-P phosphatase) [EC:3.1.3.45]

map01240 Biosynthesis of cofactors (1)

ko:K08679 GAE; UDP-glucuronate 4-epimerase [EC:5.1.3.6]

map02030 Bacterial chemotaxis (1)

ko:K03406 mcp; methyl-accepting chemotaxis protein

map00521 Streptomycin biosynthesis (1)

ko:K01790 rfbC; dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13]

map00230 Purine metabolism (1)

ko:K07816 E2.7.6.5; GTP pyrophosphokinase [EC:2.7.6.5]

map00290 Valine, leucine and isoleucine biosynthesis (1)

ko:K01652 E2.2.1.6L; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]

map00523 Polyketide sugar unit biosynthesis (1)

ko:K01790 rfbC; dTDP-4-dehydrorhamnose 3,5-epimerase [EC:5.1.3.13]

map00531 Glycosaminoglycan degradation (1)

ko:K10532 HGSNAT; heparan-alpha-glucosaminide N-acetyltransferase [EC:2.3.1.78]

map00650 Butanoate metabolism (1)

ko:K01652 E2.2.1.6L; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]

map00400 Phenylalanine, tyrosine and tryptophan biosynthesis (1)

ko:K01735 aroB; 3-dehydroquinate synthase [EC:4.2.3.4]

map00053 Ascorbate and aldarate metabolism (1)

ko:K08679 GAE; UDP-glucuronate 4-epimerase [EC:5.1.3.6]

map02020 Two-component system (1)

ko:K03406 mcp; methyl-accepting chemotaxis protein

map00970 Aminoacyl-tRNA biosynthesis (1)

ko:K01887 RARS; arginyl-tRNA synthetase [EC:6.1.1.19]

map01210 2-Oxocarboxylic acid metabolism (1)

ko:K01652 E2.2.1.6L; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]

map00770 Pantothenate and CoA biosynthesis (1)

ko:K01652 E2.2.1.6L; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]

map00250 Alanine, aspartate and glutamate metabolism (1)

ko:K01953 asnB; asparagine synthase (glutamine-hydrolysing) [EC:6.3.5.4]

map00510 N-Glycan biosynthesis (1)

ko:K00721 DPM1; dolichol-phosphate mannosyltransferase [EC:2.4.1.83]

map00660 C5-Branched dibasic acid metabolism (1)

ko:K01652 E2.2.1.6L; acetolactate synthase I/II/III large subunit [EC:2.2.1.6]

Table S4.3. Plasmid repertoire of the *L. noguchii* strains sequenced in this study, and of other *Leptospira* species.

This supplementary Excel document includes five sheets summarizing the analyses involving comparisons of plasmids. Due to their large size, it is not possible to arrange it in this annex without affecting visualization. It can be downloaded from:

Plasmid repertoire of the *L. noguchii* strains sequenced in this study, and of other *Leptospira* species.

Table S4.4. Strains used to perform *L. noguchii* phylogeny analyses.

The country of origin is indicated by a three-letter code (BRB=Barbados; BRA=Brazil; CHN=China; FRA=France; GLP=Guadeloupe; NIC=Nicaragua; PAN=Panama; PER=Peru; TTO=Trinidad & Tobago; USA=United States of America; URY=Uruguay; UNK=unknown; VEN=Venezuela).

Species	Serogroup	Serovar	Strain	Country	Host	BioSample	Assembly	Genome Status	Size (Mb)	GC (%)
L. nog	ND	ND	56183	USA	Armadillo	SAMN02947892	GCF_001569165.1	Draft	4.6	35,63
L. nog	ND	ND	56189	PER	Opossum	SAMN02947902	GCF_001569215.1	Draft	4.7	35,27
L. nog	ND	ND	56190	NIC	Weasel	SAMN02947898	GCF_001569205.1	Draft	4.5	35,46
L. nog	ND	ND	56271	TTO	Mongoose	SAMN02947958	GCF_001569225.1	Draft	4.6	35,55
L. nog	ND	ND	56682	CHN	Pig	SAMN02947960	GCF_001568325.1	Draft	4.7	35,36
L. nog	ND	ND	201102933	GLP	Human	This study		Draft	4.7	35,5
L. nog	ND	ND	201601331	VEN	Human	This study		Finished	4.8	35,7
L. nog	Panama	ND	201700036	BRA	Cattle	SAMEA5168159	ERS2975876 (SRA)	Draft	4.7	35,65
L. nog	Australis	Peruviana	201700038	PER	Cattle	SAMEA5168183	ERS2975900 (SRA)	Draft	4.7	35,59
L. nog	ND	ND	1993005606	USA	Human	SAMN02436432	GCA_000243535.3	Draft	4.9	35,9
L. nog	ND	ND	2001034031	USA	Human	SAMN02436392	GCA_000306195.2	Draft	5.0	35,76
L. nog	ND	ND	2006001870	USA	Human	SAMN02436466	GCA_000216255.3	Draft	4.8	35,56
L. nog	ND	ND	2007001578	USA	Human	SAMN02436594	GCA_000243575.3	Draft	5.0	35,8
L. nog	Australis	Bajan	bajan	BRB	Amphibian	This study		Draft	4.9	35,7
L. nog	Australis	Barbudensis	barbudensis	BRB	Amphibian	This study		Finished	4.9	35,6
L. nog	Autumnalis	ND	Bonito	BRA	Human	SAMN01801600	GCA_000346655.1	Draft	4.5	35,7
L. nog	Bataviae	ND	Cascata	BRA	Human	SAMN01162017	GCA_000350585.1	Draft	4.5	35,3

L. nog	Panama	Panama	CZ214	PAN	Opossum	SAMN02436411	GCA_000306255.2	Draft	4.7	35,5
L. nog	ND	ND	HAI1536	PER	Human	SAMN02436491	GCA_000306635.2	Draft	4.7	35,3
L. nog	Australis	ND	Hook	BRA	Dog	SAMN01162022	GCA_000350605.1	Draft	4.5	35,6
L. nog	ND	ND	IP1512017	URY	Cattle	This study		Finished	5.0	35,78
L. nog	Pyrogenes	ND	IP1605021	URY	Cattle	This study		Finished	5.1	35,88
L. nog	Australis	ND	IP1611024	URY	Cattle	This study		Finished	4.9	35,47
L. nog	Autumnalis	ND	IP1611025	URY	Cattle	SAMEA5168228	ERS2975945 (SRA)	Draft	4.3	36,21
L. nog	ND	ND	IP1703027	URY	Cattle	This study		Finished	4.8	35,75
L. nog	Autumnalis	ND	IP1705032	URY	Cattle	This study		Finished	4.7	35,7
L. nog	Autumnalis	ND	IP1708035	URY	Cattle	SAMEA5168334	ERS2976051 (SRA)	Draft	4.6	35,61
L. nog	Autumnalis	ND	IP1709037	URY	Cattle	This study		Finished	4.7	35,73
L. nog	ND	ND	IP1712055	URY	Cattle	This study		Finished	4.9	35,86
L. nog	ND	ND	IP1804061	URY	Cattle	This study		Finished	5.0	35,71
L. nog	Panama	ND	lyse	UNK	Unknown	SAMEA5168089	ERS2975806 (SRA)	Draft	4.6	35,3
L. nog	Australis	Nicaragua	nicaragua	NIC	Unknown	SAMEA5168249	ERS2975966 (SRA)	Draft	4.6	35,5
L. nog	Australis	Rushan	rushan	CHN	Amphibian	SAMEA5168202	ERS2975919 (SRA)	Draft	4.6	35,7
L. nog	Australis	Soteropolitana	soteropolitana	BRA	Unknown	SAMEA5168194	ERS2975911 (SRA)	Draft	4.3	36,1
L. nog	Panama	ND	U73	BRA	Cattle	SAMN03943765	GCA_001276935.1	Draft	4.8	35,7
L. nog	Autumnalis	Autumnalis	ZUN142	PER	Human	SAMN02436382	GCA_000244775.3	Draft	4.9	35,5
L. int	Icterohaemorrhagiae	Lai	56601	CHN	Human	SAMN02603127	GCA_000092565.1	Finished	5	35,0
L. kir	Grippotyphosa	Valbuzzi	200702274	FRA	Human	SAMN02436380	GCA_000244515.3	Draft	4	36,0

Table S4.5. *Leptospira* species and strains used to perform genetic analyses of the *rfb* cluster.

Table S4.6. *rfb* genes were used to search orthologous protein sequences: Blast best hits.

These supplementary Excel documents include several sheets summarizing the *rfb* clusters analyses. Due to their large size, it is not possible to arrange them in this annex without affecting visualization. They can be downloaded from:

Leptospira species and strains used to perform genetic analyses of the *rfb* cluster.

rfb genes were used to search orthologous protein sequences: Blast best hits' accession numbers are listed.

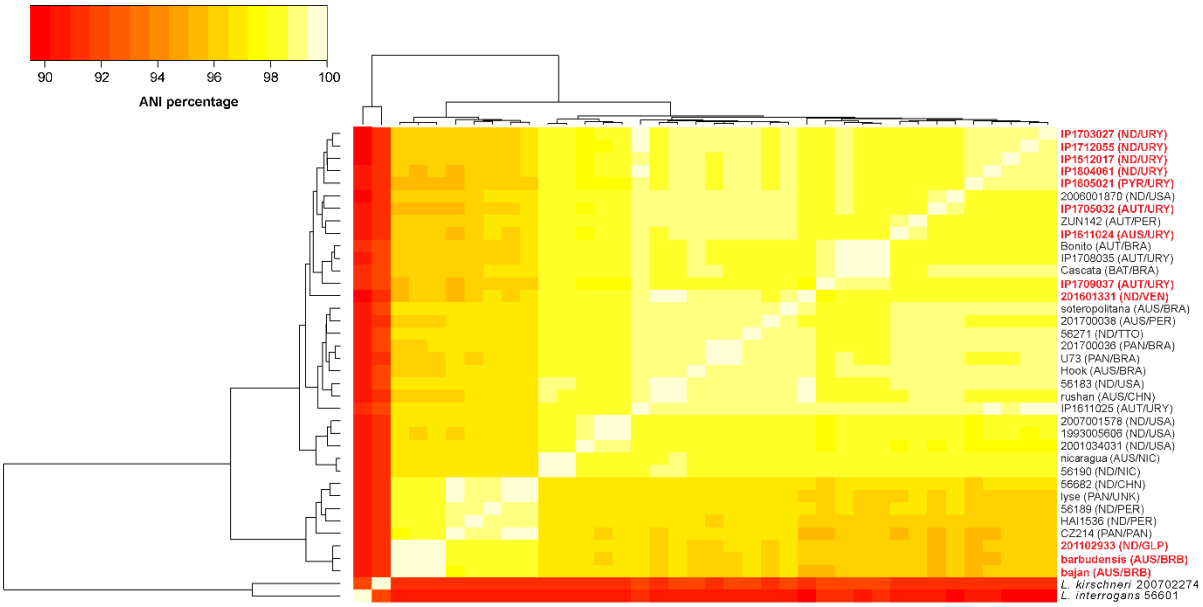

Figure S4.1. Average Nucleotide Identity (ANI) among *L. noguchii* strains with reported whole genome sequences.

L. interrogans and *L. kirschneri* were included as reference of distinct species. The ANI percentages are depicted as colors of the square matrix elements, according to the scale shown in the upper left insert. The names of the twelve strains sequenced in this study are highlighted with red fonts. Clustering shown on the left side (and upper side, symmetric) of the matrix table, was performed by GET_HOMOLOGUES version 20190411 (Contreras-Moreira and Vinuesa 2013). Each strain's serogroup (AUS=Australis; AUT=Autumnalis; BAT=Bataviae; ND=not determined; PAN=Panama; PYR=Pyrogenes), and the country from where they were obtained (BRB=Barbados; BRA=Brazil; CHN=China; GLP=Guadeloupe; NIC=Nicaragua; PAN=Panama; PER=Peru; TTO=Trinidad & Tobago; USA=United States of America; URY=Uruguay; UNK=unknown; VEN=Venezuela) are indicated in parentheses.

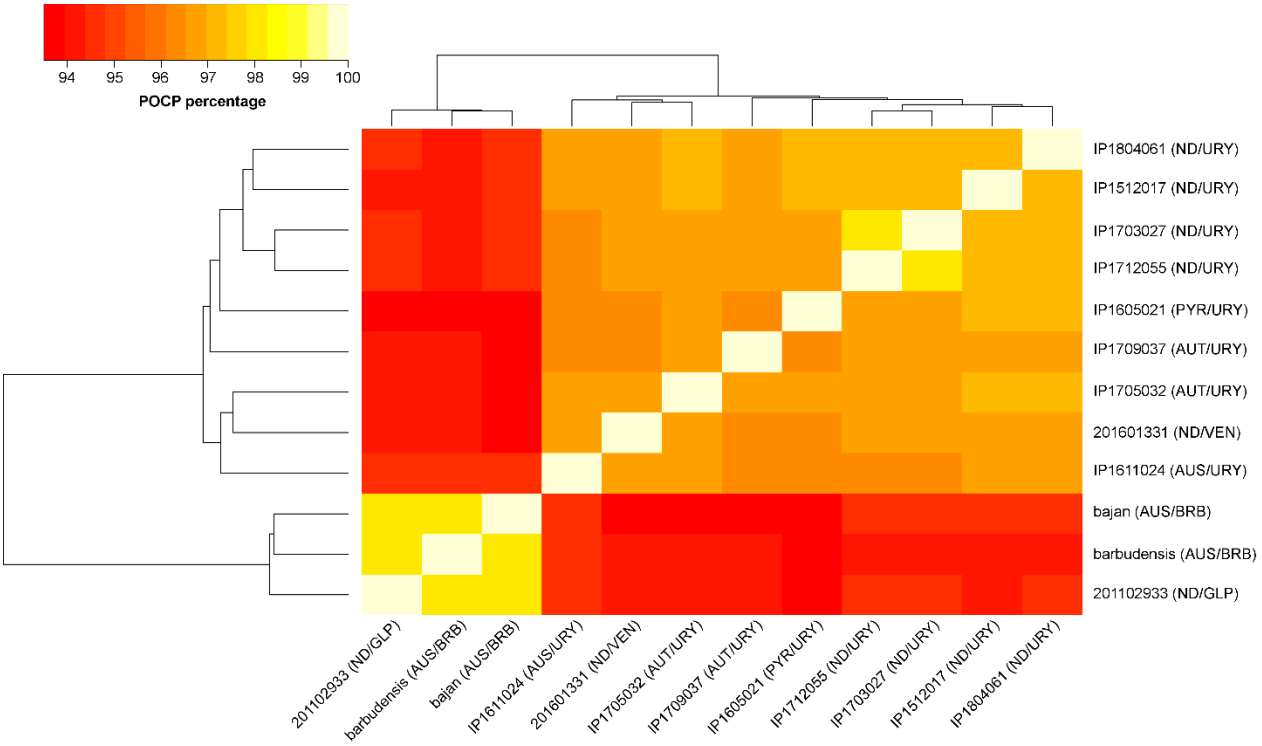

Figure S4.2. Percentage of Conserved Proteins (POCP) among *L. noguchii* strains.

POCP percentages are represented as colors of the square matrix elements, according to the scale shown in the upper left insert. The names of the twelve sequenced strains are indicated on the right and bottom. Clustering shown on the left and upper sides of the matrix, was performed by GET_HOMOLOGUES version 20190411 (Contreras-Moreira and Vinuesa 2013). Strain names identifiers as in Supplemental Fig. S4.1.

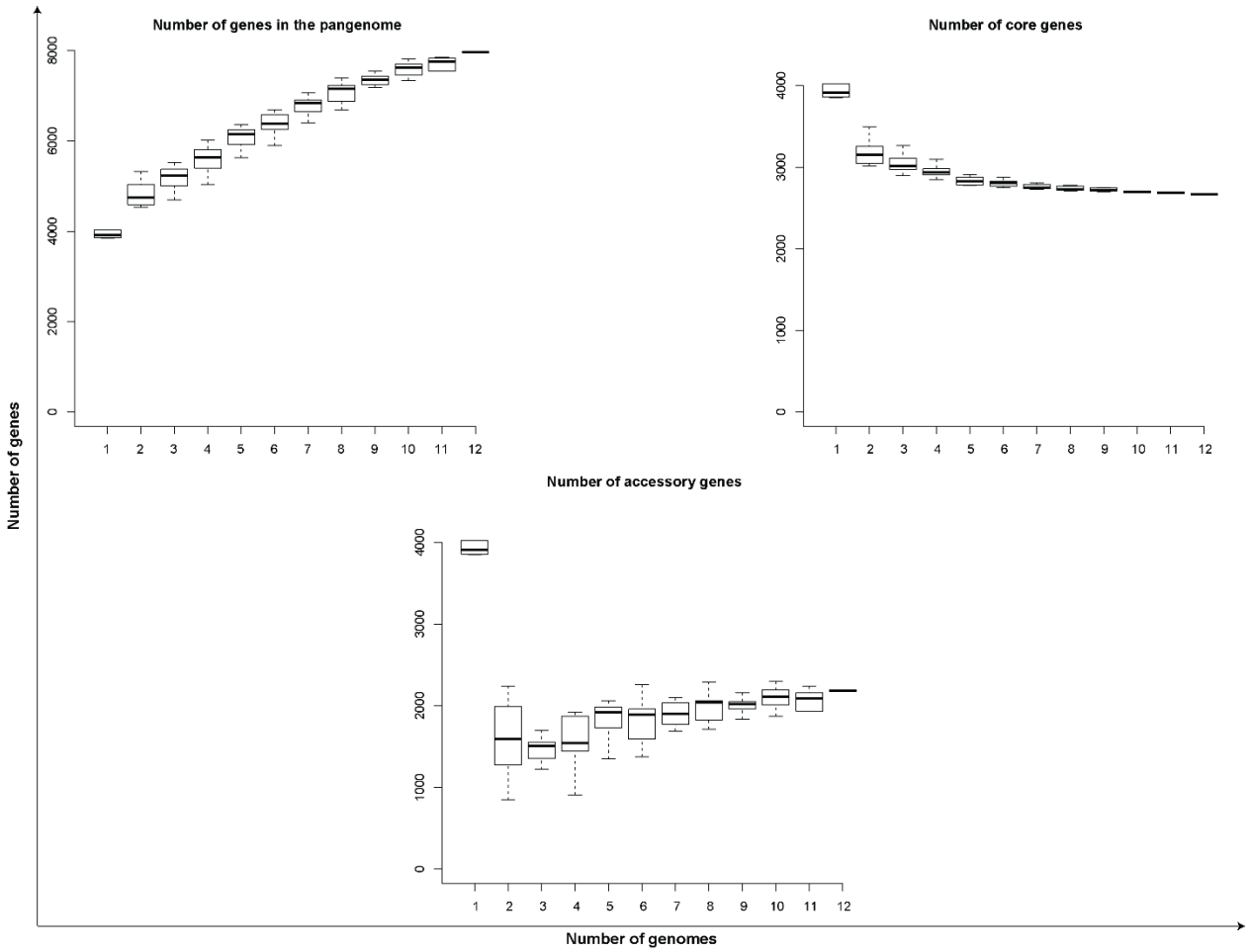

Figure S4.3. Pangenome profiling of the twelve *L. noguchii* strains sequenced in this study.

Calculations and plots were performed using Roary 3.11.2 (Page et al. 2015). The curve observed in the pangenome plot was used to fit a calculated curve according to Heaps' law (Tettelin et al. 2008), $n = \kappa N\gamma$, considering a total of 7963 genes (n) in the pangenome vs 12 genomes (N) included, the observed curve allows to fit non-linearly $\kappa = 1610.8$ and $1-\gamma = \alpha = 0.36$. $\alpha < 1$ indicates an open profile.

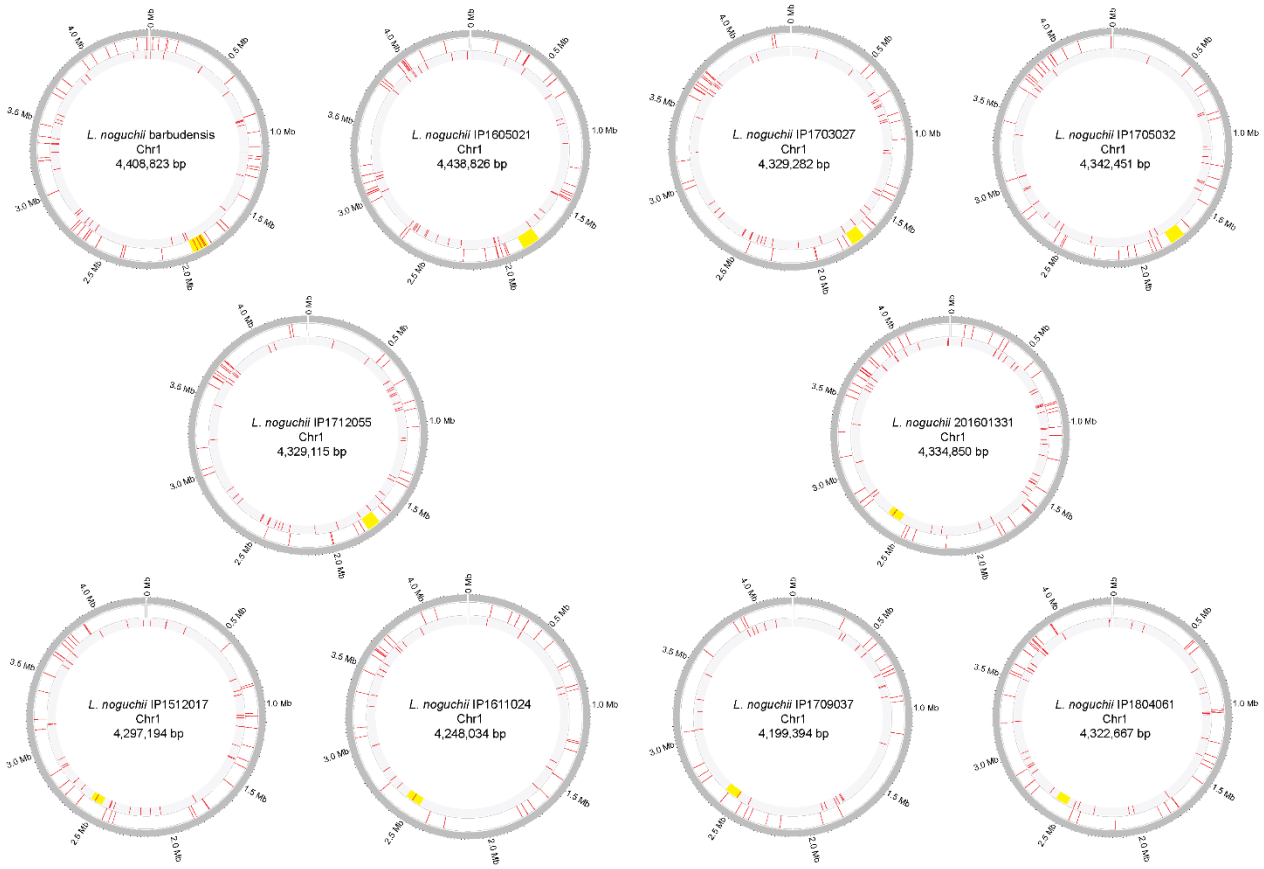

Figure S4.4. IS/transposase genes distribution across chromosome 1 from *L. noguchii* strains sequenced in this study. Transposase and transposase-like genes are depicted as red lines; the complementary and leading strands are represented respectively on the inner and the outer circle positions. The *rfb* cluster is shown as yellow blocks. This graphical representation was produced with the online version of shinyCircos (https://venyao.xyz/shinycircos/) (Yu et al. 2018).

Figure S4.5. GC-content across chromosome 1 from different *Leptospira* species.

Chromosomes 1 from different *Leptospira* species are laid down, with the *rfb* cluster depicted as a red bar on top of each one. The GC-content (calculated every 1000 bp) was plotted in colored curves using DNAplotter (Carver et al. 2009), with purple highlighting a decrease in its percentage, and green an increase compared to the average found in the whole chromosome 1.

Appendix III. Supplementary material of Chapter 5

Here are the supplemental tables and figures of the article presented in Chapter 5.

Table S5.1. Geographic origins of the most frequent pathogenic *Leptospira* species in our genome database (n = 914).

This supplementary Excel document includes the metadata of the 914 *Leptospira* genomes used in this study, separated into sheets according to the geographical region. Due to their large size, it is not possible to arrange it in this annex without affecting visualization. It can be downloaded from:

https://doi.org/10.1371/journal.pntd.0011733.s001

Table S5.2. List of *L. santarosai* strains used in this study and genome characteristics.

Some columns were omitted for the purpose of clear visualization. The original table can be downloaded from: https://doi.org/10.1371/journal.pntd.0011733.s002.

id (a)	Isolate	Serogroup	Serovar	Size (Mb)	Contigs	GC (%)	cgST (b)	*rfb* typing (c)	Isolation year	Country	Host
88	200702252	Mini	Unknown	4.1	34	41.9	69	cluster 4	ND	Guadeloupe	Human
89	200403458	Mini	Unknown	4.1	35	41.9	70	cluster 4	ND	Guadeloupe	Human
90	ST188	Unknown	Unknown	4.0	378	41.8	71	cluster 1	ND	Trinidad & Tobago	Dog
91	7	Javanica	Arenal	4.2	535	41.6	72;1154	cluster 2	2001	Costa Rica	Human
92	11	Javanica	Arenal	4.0	406	41.7	73	cluster 2	2001	Costa Rica	Human
96	LT821	Shermani	Shermani	3.9	107	41.8	76	cluster 3	ND	Panama	Rat
219	Oregon	Mini	Szwajizak	3.9	123	41.9	175	cluster 4	ND	USA	Bovine
220	2000027870	Unknown	Unknown	3.9	41	41.8	176	cluster 4	ND	USA	Human
221	2000030832	Unknown	Unknown	4.1	36	41.8	177	cluster 4	ND	USA	Human
246	Fiocruz L1-130	Icterohaemorrhagiae	Copenhageni	4.6	2	35	196;1119	cluster 1	1996	Brazil	Human
257	U160	Unknown	Unknown	4.2	204	41.7	206;1121	cluster 2	ND	Brazil	Bovine
261	Fiocruz LV3954	Shermani	Unknown	4.0	91	41.7	209	cluster 3	ND	Brazil	Human

262	Fiocruz LV4135	Shermani	Unknown	4.0	133	41.7	210	cluster 3	ND	Brazil	Human
266	JET	Tarassovi	Unknown	4.1	82	41.6	213;1130	cluster 3	ND	Colombia	Human
281	CBC1416	Unknown	Unknown	4.0	195	42.0	223	cluster 4	ND	Peru	Bovine
282	CBC523	Unknown	Unknown	4.0	52	41.9	224;1230	cluster 3	ND	Peru	Bovine
283	CBC1531	Unknown	Unknown	3.8	285	41.8	225	cluster 4	ND	Peru	Bovine
285	HAI134	Unknown	Unknown	4.1	97	41.8	227	cluster 2	ND	Peru	Human
286	HAI1349	Unknown	Unknown	3.9	114	41.7	228	cluster 1	ND	Peru	Human
289	HAI821	Tarassovi	Unknown	4.1	47	41.8	230	cluster 3	ND	Peru	Human
290	MOR084	Unknown	Unknown	3.9	120	41.8	231	cluster 4	ND	Peru	Human
291	CBC379	Unknown	Unknown	4.1	168	41.9	232	cluster 4	ND	Peru	Pig
445	1440	Grippotyphosa	Canalzonae	3.9	139	41.9	364	cluster 1	1966	Panama	spiny rat
526	201101963	Pyrogenes	Unknown	4.2	341	41.7	432	cluster 2	2011	Martinique	Human
944	201902908	Unknown	Unknown	3.9	235	42.2	719	cluster 3	2018	Martinique	Human
1189	202101808	Pyrogenes	Unknown	4.3	202	41.8	995	cluster 2	2021	Martinique	Human
1214	JICH 05	Tarassovi	Corredores	3.9	111	41.8	1018	cluster 3	2005	Costa Rica	Human
1215	inciensa 04	Pyrogenes	Costa rica	4.1	248	41.7	1019	cluster 2	2004	Costa Rica	Human
1251	CR2020	Hebdomadis	Unknown	4.1	213	41.9	1025	cluster 4	2020	Costa Rica	Human
1252	CR2120	Shermani	Unknown	4.0	123	41.8	1026;1098	cluster 3	2021	Costa Rica	Human
1253	CR0421	Grippotyphosa *	Unknown	4.3	418	41.6	1027	cluster 1	2021	Costa Rica	Human
1254	CR0521	Tarassovi	Unknown	3.9	143	41.8	1028;1099	cluster 3	2021	Costa Rica	Human
1255	CR0821	Shermani	Unknown	4.2	121	41.7	1029;1100	cluster 3	2021	Costa Rica	Human
1256	CR1421	Pyrogenes	Unknown	3.9	123	41.9	1030	cluster 2	2021	Costa Rica	Human
1257	CR1821	Grippothyphosa	Unknown	3.9	142	41.8	1034	cluster 1	2020	Costa Rica	Human
1259	CR2621	Grippothyphosa	Unknown	3.8	133	41.9	1031	cluster 1	2021	Costa Rica	Human
1260	CR2821	Pyrogenes	Unknown	4.0	122	41.9	1032	cluster 2	2021	Costa Rica	Human
1261	CR2921	Tarassovi	Unknown	4.0	116	41.7	1033	cluster 3	2021	Costa Rica	Human
1284	2015-VF237	Sejroe	Sejroe	4.0	113	41.8	870;1110;1113;1334	cluster 4	2015	Brazil	Bovine
1285	2013-VF52	Sejroe	Sejroe	4.0	115	41.8	1110	cluster 4	2013	Brazil	Bovine
1295	U164	Tarassovi	Unknown	3.7	1	41.9	1122	cluster 3	2012	Brazil	bovine

1296	U233	Grippotyphosa	Unknown	3.6	1	42.0	1123	cluster 1	2012	Brazil	bovine
1303	AIM	Tarassovi	Unknown	4.1	72	41.7	1125	cluster 3	2013	Colombia	Human
1304	1342KT	Shermani	Shermani	4.0	68	41.8	1126	cluster 3	2011	Panama	Spiny Rat
1307	ZUN179	Pyrogenes	Unknown	4.1	55	41.7	1129	cluster 2	2013	Peru	Human
1309	HAI1380	Unknown	Unknown	4.0	38	41.8	1131	cluster 4	2013	Peru	Human
1310	MAVJ 401	Javanica	Arenal	4.2	87	41.6	1132	cluster 2	2013	Costa Rica	Human
1319	56163	Mini	Unknown	3.9	210	41.9	1141	cluster 4	1952	USA	racoon
1320	56164	Unknown	Unknown	4.0	288	41.6	1142	cluster 3	1955	USA	Opossum
1321	56180	Unknown	Unknown	3.9	203	41.8	1143	cluster 1	1964	Panama	Rat
1322	56198	Hebdomadis	Kambale	4.0	335	41.8	1144	cluster 4	1952	Congo [DRC]	Human
1323	56215	Unknown	Unknown	4.1	311	41.8	1145	cluster 3	1961	Panama	Human
1324	56274	Unknown	Unknown	3.9	212	41.8	1146	cluster 3	1982	Panama	Opossum
1325	56663	Mini	Unknown	3.9	211	41.9	1147	cluster 4	1980	China	Muskrat
1326	C216	Unknown	Unknown	4.0	95	41.8	1148	cluster 4	2014	Ecuador	Human
1327	M498	Sejroe	Guaricura	4.0	57	41.9	1149	cluster 4	1998	Brazil	Bovine
1328	LO-9	Grippotyphosa	Grippotyphosa	4.1	60	41.7	1150	cluster 1	2001	Brazil	Bovine
1329	2ACAP	Grippotyphosa	Bananal	4.1	44	41.7	1151	cluster 1	2001	Brazil	Capybara
1330	M726-6	Grippotyphosa	Grippotyphosa	4.1	60	41.8	1152	cluster 1	2006	Brazil	Goat
1331	M528-19	Unknown	Unknown	4.2	45	41.7	1153	cluster 3	2008	Brazil	capybara
1353	949	Unknown	Unknown	4.7	1878	41.5	1372	cluster 4	2013	Brazil	Human
1459	Rr 5	Sarmin	Rio	4.1	179	41.9	1450	cluster 2	1973	Brazil	rat
1461	MMD 3	Sarmin	Machigueng a	4.0	133	41.8	1452	cluster 2	1970	Peru	opossum
1467	LSU 1013	Tarassovi	Atchafalaya	4.0	164	41.8	1458	cluster 3	1963	USA	opossum
1475	DCP-017	Pyrogenes		4.1	2	41.7	1465	cluster 2	ND	USA	Bovine

(a): Identification number (id) of the reference genome in BIGSdb (www.bigsdb.pasteur.fr).

(b): cgMLST sequence type (cgST)

(c): see gene presence/absence matrices of rfb clusters in Figure 5.6.

*: Serogrouping also shows strong agglutination with rabbit antisera against serogroups Djasiman, Louisiana and Cynopteri.

ND: not determined.

Table S5.3. Dunn Kruskal-Wallis multiple comparisons results between the G+C% content of *Leptospira* genomes.

Genome 1	Genome 2	Z	P unadjusted	P adjusted*	Significance?
L. borgpetersenii	*L. interrogans*	23.695601	4.00E-124	**8.405E-123**	**yes**
L. borgpetersenii	*L. kirschneri*	6.7034072	2.04E-11	**4.2759E-10**	**yes**
L. interrogans	*L. kirschneri*	-8.8324078	1.02E-18	**2.1514E-17**	**yes**
L. borgpetersenii	*L. mayottensis*	3.0745397	2.11E-03	**0.04427381**	**yes**
L. interrogans	*L. mayottensis*	-7.1513896	8.59E-13	**1.804E-11**	**yes**
L. kirschneri	*L. mayottensis*	-1.2599567	2.08E-01	1.00	ns
L. borgpetersenii	*L. noguchii*	5.4434492	5.23E-08	**1.0974E-06**	**yes**
L. interrogans	*L. noguchii*	-4.4448081	8.80E-06	**0.00018474**	**yes**
L. kirschneri	*L. noguchii*	0.998442	3.18E-01	1.00	ns
L. mayottensis	*L. noguchii*	1.8619713	6.26E-02	1.00	ns
L. borgpetersenii	*L. santarosai*	-5.0288379	4.93E-07	**1.0363E-05**	**yes**
L. interrogans	*L. santarosai*	-19.0799017	3.71E-81	**7.7901E-80**	**yes**
L. kirschneri	*L. santarosai*	-9.2837431	1.64E-20	**3.4362E-19**	**yes**
L. mayottensis	*L. santarosai*	-5.9165522	3.29E-09	**6.904E-08**	**yes**
L. noguchii	*L. santarosai*	-7.9228156	2.32E-15	**4.876E-14**	**yes**
L. borgpetersenii	*L. weilii*	-2.3345431	1.96E-02	0.41	ns
L. interrogans	*L. weilii*	-11.2352603	2.74E-29	**5.7474E-28**	**yes**
L. kirschneri	*L. weilii*	-5.7424584	9.33E-09	**1.9595E-07**	**yes**
L.mayottensis	*L. weilii*	-3.9706147	7.17E-05	**0.00150544**	**yes**
L. noguchii	*L. weilii*	-5.630652	1.80E-08	**3.7701E-07**	**yes**
L. santarosai	*L. weilii*	0.8471142	3.97E-01	1.00	ns

* p-values were adjusted with the Bonferroni method.

Table S5.4. Dunn Kruskal-Wallis multiple comparisons results between the genome size (length) of *Leptospira* genomes.

Genome 1	Genome 2	Z	P unadjusted	P adjusted*	significance?
L. borgpetersenii	*L. interrogans*	-24.5110307	1.13E-132	**2.37E-131**	**yes**
L. borgpetersenii	*L. kirschneri*	-10.6232542	2.32E-26	**4.88E-25**	**yes**
L. interrogans	*L. kirschneri*	5.266405	1.39E-07	**2.92E-06**	**yes**
L. borgpetersenii	*L. mayottensis*	-2.4309378	1.51E-02	0.32	ns
L. interrogans	*L. mayottensis*	8.1625592	3.28E-16	**6.89E-15**	**yes**
L. kirschneri	*L. mayottensis*	4.2046359	2.62E-05	**5.49E-04**	**yes**
L. borgpetersenii	*L. noguchii*	-9.0038653	2.18E-19	**4.58E-18**	**yes**
L. interrogans	*L. noguchii*	1.1581012	2.47E-01	1.00	ns
L. kirschneri	*L. noguchii*	-1.9252253	5.42E-02	1.00	ns
L. mayottensis	*L. noguchii*	-5.0390041	4.68E-07	**9.83E-06**	**yes**
L. borgpetersenii	*L. santarosai*	-2.8806962	3.97E-03	0.08	ns
L. interrogans	*L. santarosai*	11.3498105	7.43E-30	**1.56E-28**	**yes**
L. kirschneri	*L. santarosai*	5.5136552	3.51E-08	**7.38E-07**	**yes**
L. mayottensis	*L. santarosai*	0.2217722	8.24E-01	1.00	ns
L. noguchii	*L. santarosai*	5.9769778	2.27E-09	**4.77E-08**	**yes**
L. borgpetersenii	*L. weilii*	-4.397792	1.09E-05	**0.0002**	**yes**
L. interrogans	*L. weilii*	4.7032591	2.56E-06	**5.38E-05**	**yes**
L. kirschneri	*L. weilii*	1.6047652	1.09E-01	1.00	ns
L. mayottensis	*L. weilii*	-1.8217257	6.85E-02	1.00	ns
L. noguchii	*L. weilii*	2.8380825	4.54E-03	0.10	ns
L. santarosai	*L. weilii*	-2.2403587	2.51E-02	0.53	ns

* p-values were adjusted with the Bonferroni method.

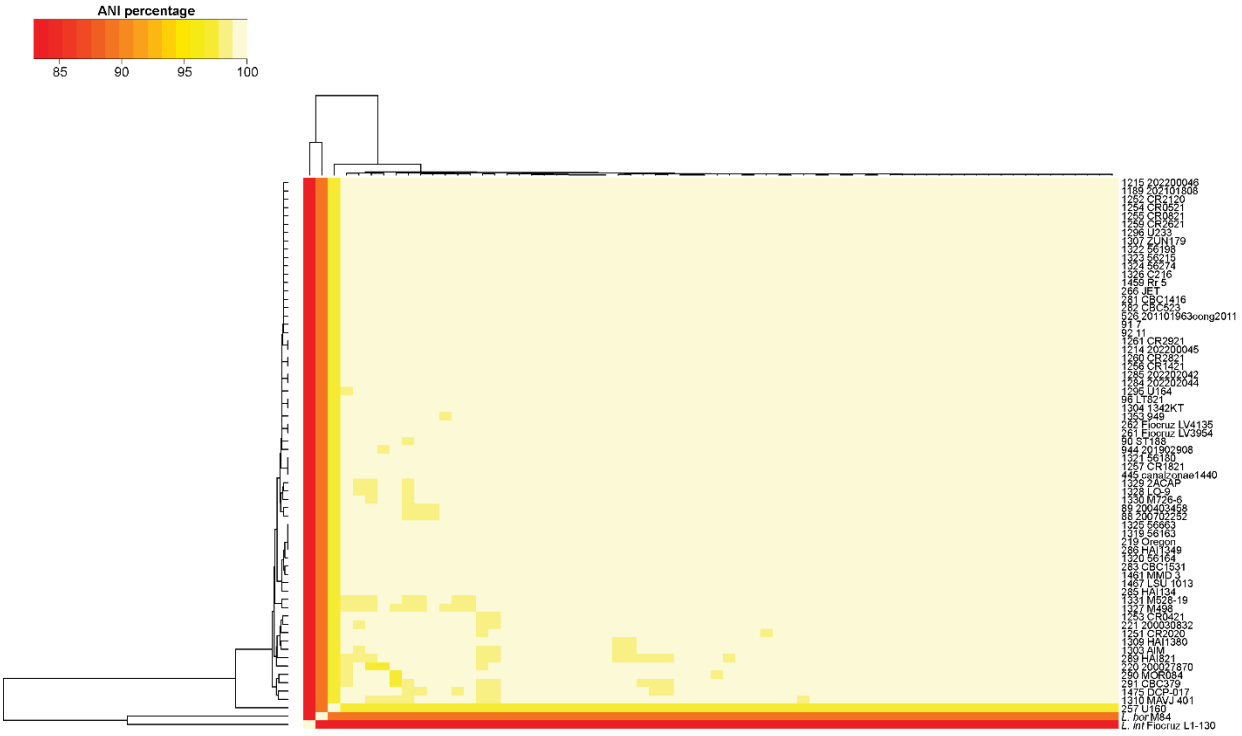

Figure S5.1 Average Nucleotide Identity (ANI) among *L. santarosai* strains.

L. interrogans strain Fiocruz L1-130 and *L. borgpetersenii* strain M84 were included as outgroups. The ANI percentages are depicted as colors of the square matrix elements, according to the scale shown in the upper left insert. The names of the 64 *L. santarosai* strains are indicated on the right of the matrix. Clustering shown on the left side (and upper side, symmetric) of the matrix table, was performed by GET_HOMOLOGUES version 20190411 (Contreras-Moreira and Vinuesa, 2013).

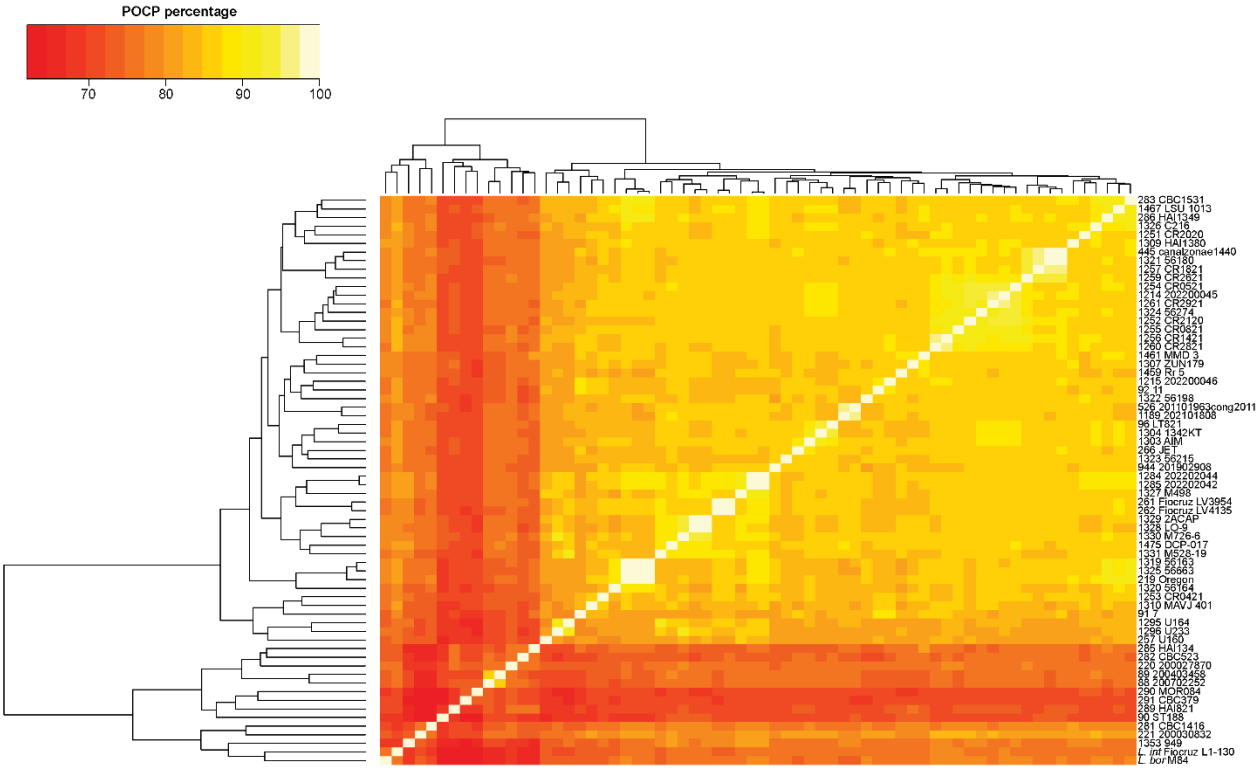

Figure S5.2. Percentage of Conserved Proteins (POCP) among *L. santarosai* strains.

L. interrogans strain Fiocruz L1-130 and *L. borgpetersenii* strain M84 were included as reference of distinct species. POCP percentages are represented as colors of the square matrix elements, according to the scale shown in the upper right insert. The names of the 64 *L. santarosai* strains are indicated on the right of the matrix. Clustering shown on the left side (and upper side, symmetric) of the matrix table, was performed by GET_HOMOLOGUES version 20190411 (Contreras-Moreira and Vinuesa, 2013).

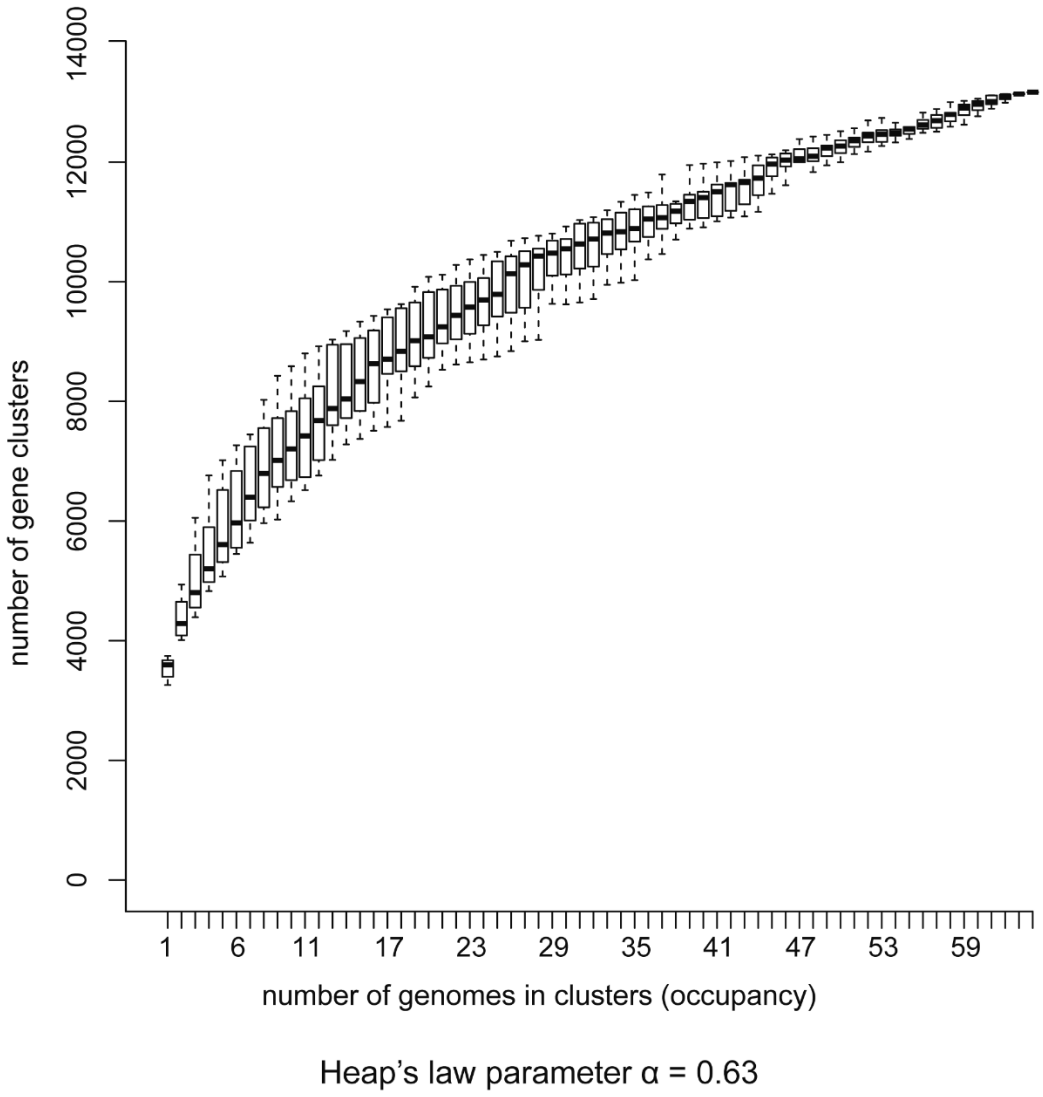

Figure S5.3. Pangenome analysis of *L. santarosai* strains.

The graph was obtained with Roary 3.11.2 (Page et al., 2015) (yielding a total of 13,168 gene clusters). The pangenome of *L. santarosai* presents an open profile, which was further verified by Heap's law (Tettelin et al., 2008), $n = \kappa N^{\gamma}$; considering a total of 13,168 genes (n) in the pangenome (according to Roary) and the 64 genomes (N) included, the observed curve allows for non-linear fitting with a constant $\kappa = 2851.4$. Thus, $1-\gamma = \alpha = 0.63$. An α value <1 indicates an open profile. Same result was obtained considering n = 13,156 (GET_HOMOLOGUES).

Figure S5.4. Jaccard similarity matrix for the comparative analyses of the *rfb* clusters.

Jaccard similarity matrix is organized according to the gene presence/absence matrix of *rfb* clusters shown in Figure 5. To perform the Jaccard similarity index calculation, each strain (column) was converted into a vector of 0 (gene absence) and 1 (gene presence), and then used to do pairwise comparisons between the vectors using a Python pipeline via NumPy (Harris et al., 2020). Jaccard similarity is represented in colors, according to the scale shown in the lower right insert.

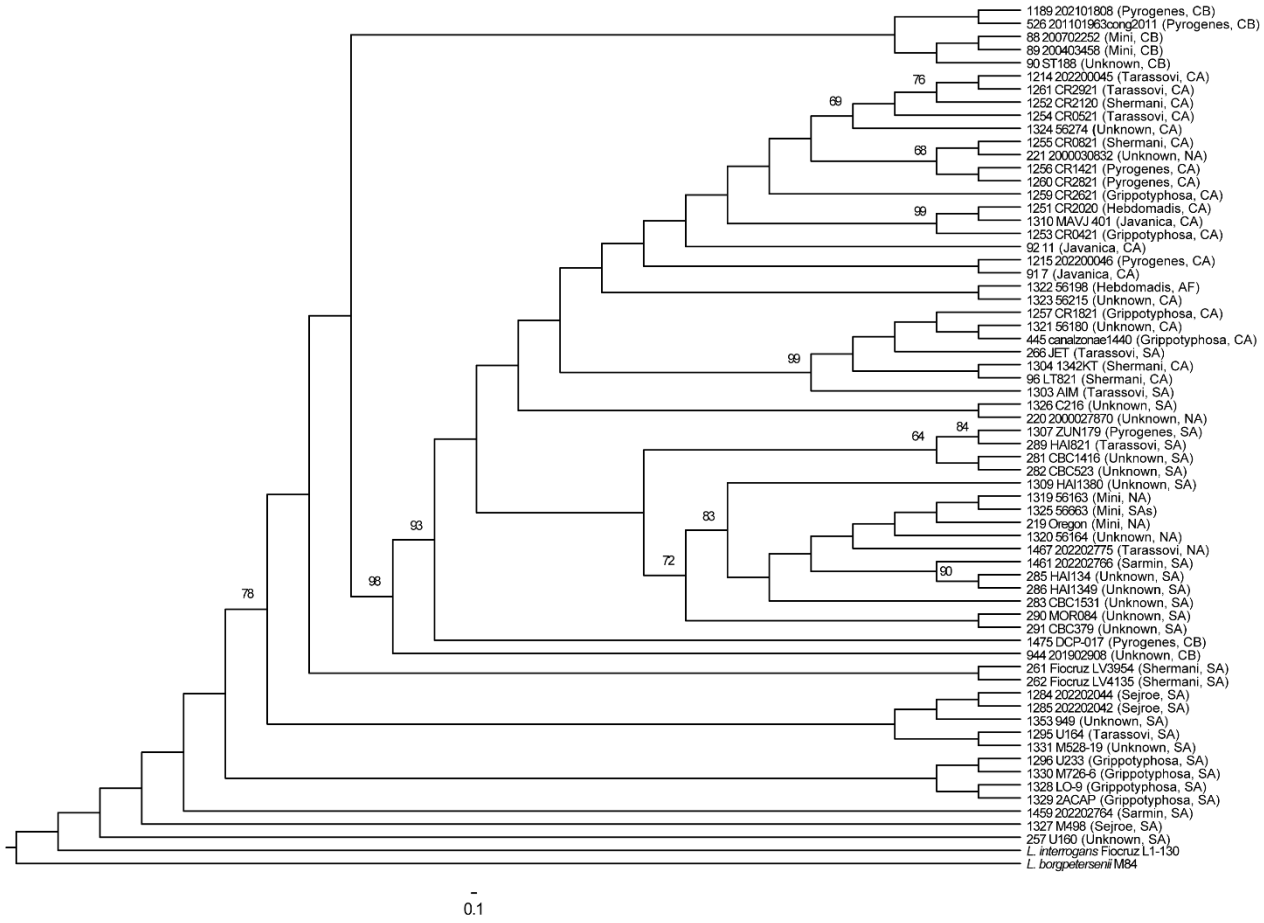

Figure S5.5. Core- genome based phylogeny of *L. santarosai.*

Phylogenetic tree based on the sequences of 1288 core-genes of *L. santarosai*. A core-gene alignment was obtained by Roary (60% identity cut-off), and then used to perform the phylogeny. The best-fit model and the maximum-likelihood phylogenetic tree were determined by IQ-TREE version 1.6.11 (Nguyen et al., 2015), considering 10,000 ultrafast bootstraps (Hoang et al., 2018). *L. interrogans* strain Fiocruz L1-130 and *L. borgpetersenii* strain M84 were included as outgroups. The serogroup of each strain is indicated in parenthesis, as well as the world region (AF=Africa; CA=Central America; CB=Caribbean; NA=North America; SA=South America; SAs=South Asia). Bootstraps values other than 100% are shown.