Genomic Analysis of a New Serovar of *Leptospira weilii* Serogroup Manhao

Yinghua Xu[1†], Huajun Zheng[2,3†], Ying Zhang[1†], Yuezhu Wang[3], Jinlong Zhang[1], Zhe Li[1], Shenghui Cui[1], Xiaofang Xin[1], Qiang Ye[1], Yung-Fu Chang[4] and Junzhi Wang[1]*

[1] Key Laboratory of the Ministry of Health for Research on Quality and Standardization of Biotech Products, National Institutes for Food and Drug Control, Beijing, China, [2] Key Laboratory of Reproduction Regulation of NPFPC, Shanghai Institute of Planned Parenthood Research, IRD, Fudan University, Shanghai, China, [3] Shanghai-MOST Key Laboratory of Health and Disease Genomics, Chinese National Human Genome Center at Shanghai, Zhangjiang Hi-Tech Park, Shanghai, China, [4] Department of Population Medicine and Diagnostic Sciences, Cornell University, Ithaca, NY, USA

Leptospirosis, caused by pathogenic *Leptospira* spp., is recognized as an important emerging zoonotic disease throughout the world. In this study, multiple approaches were used to characterize the recently discovered serovar Heyan strain L231. This strain can infect guinea pigs and belonged to the pathogenic species *L. weilii*. Genome sequencing analysis revealed the draft genome of 4.2 M bp with a G+C content of 40.67% for strain L231, and a total of 4,794 ORFs were identified. The strain L231 genome was found to have a larger LPS biosynthesis locus than that of strains *L. interrogans* serovar Lai and *L. borgpetersenii* serovar Hardjobovis. Phylogenomic reconstructions showed that the evolutionary position of *L. weilii* serovar Heyan was different from that of other serovars from serogroup Manhao. These findings may lead us to a better understanding of *Leptospira* pathogenesis and evolution.

Keywords: Leptospirosis, serogroup Manhao, new serovar, pathogenesis, genome

INTRODUCTION

Leptospirosis, caused by pathogenic species of *Leptospira*, is recognized as an emerging zoonotic disease of global importance (Bharti et al., 2003). Humans are infected through contact with animal reservoirs or contaminated environment (i.e., soil, sewage, or water). It is estimated that there are approximately 1.03 million global leptospirosis cases each year, resulting in 58,900 deaths. In China, the earliest leptospirosis case was recorded in 1920s. So far, more than 2.5 million cases and over 20,000 deaths have been reported. In the past 60 years, 10 outbreaks of leptospirosis with incidence of more than 10 cases per 100,000 have occurred in China (Shi and Jiang, 2000).

To develop better control strategies for infectious diseases, such as new vaccines and diagnostic tests, further characterization of each pathogen is required. *Leptospira* strains are identified by serogroup, and the serogroups are further classified into different serovars using a conventional serological classification system (Kmety and Dikken, 1993). Although serogroups have no taxonomic value, they are useful for clinical and epidemiological studies (Levett, 2001). Until now, more than 250 pathogenic serovars belonging to 24 serogroups have been reported. Due to the variety of geographic and climatic conditions in China, pathogenic leptospiral serovar diversity is far greater than that of other countries, and more than 70 serovars arranged in 18 serogroups of pathogenic *Leptospira* from a variety of hosts have been isolated and identified and some of them (i.e., Manhao) are only found in China (Kmety and Dikken, 1993).

Recently, a new serovar (Heyan) of serogroup Manhao from a patient in Yunnan, China has been reported (Qin et al., 2004). The development of new molecular techniques has allowed for classification of the genus *Leptospira* into species, which may provide a better way of categorizing pathogenic *Leptospira* strains than that of traditional serological methods. Based on the classical DNA-DNA hybridization studies and 16S rRNA gene phylogeny, pathogenic *Leptospira* are now classified into nine species, *L. interrogans*, *L. borgpetersenii*, *L. kirschneri*, *L. alexanderi*, *L. alstonii*, *L. kmetyi*, *L. noguchii*, *L. santarosai*, and *L. weilii* (Brenner et al., 1999; Morey et al., 2006). A novel pathogenic species, designated *L. mayottensis* has recently been reported (Bourhy et al., 2014). Furthermore, pulsed-field gel electrophoresis, multiple-loci variable number tandem repeat analysis, and multi-locus sequence typing have also been developed for typing pathogenic *Leptospira* strains (Herrmann et al., 1992; Salaun et al., 2006; Xu et al., 2015; Zhang et al., 2015). More importantly, genome sequencing has been also used to analyze several *Leptospira* species and a high-level plasticity among pathogenic *Leptospira* genomes was revealed, which has improved our understanding of genetic diversity at the genome level (Ren et al., 2003; Bulach et al., 2006; Fouts et al., 2016; Xu et al., 2016). To further understand the molecular features and origin of those unique serogroups in China, multiple molecular approaches were used to characterize new *Leptospira* serovar Heyan in this study. In addition, the virulence of this new serovar was evaluated in the guinea pig challenge model.

MATERIALS AND METHODS

Leptospira Strains and Culture

Serovar Heyan strain L231 was originally isolated from a clinical patient in Yunnan, China (Qin et al., 2004). *L. interrogans* serogroup Icterohaemorrhagiae serovar Icterohaemorrhagiae strain Lai was obtained from the collection maintained by the National Reference Laboratory for *Leptospira*, at the National Institutes of Food and Drug Control (NIFDC), Beijing, China (Qin, 1999). All leptospiral strains were cultivated at 28°C on modified TEPCKNN medium (10 mM Na_2HPO4, 4 mM KH_2PO4, pH 7.2; NIFDC, Beijing) supplemented with 10% rabbit serum. Genomic DNA was extracted using a Wizard Genomic DNA Purification Kit (Promega, Southampton, UK) following the manufacturer's instructions. The extracted DNA was stored at −70°C until used.

Virulence Testing and Histopathology

The animal experiments were authorized by the Animal Ethic Committee of NIFDC (License No. 2014-B-003), and conducted in compliance with Regulations on Management of Laboratory Animal (Ministry of Science and Technology) 1988, China.

The virulence of the isolate was evaluated using a guinea pig challenge model as previously described in the Chinese Pharmacopeia with minor modifications (Npc, 2010). Briefly, three Hartley female guinea pigs (180–200 g, from NIFDC), were inoculated intraperitoneally with 1 mL of leptospiral strain L231 containing 1×10^9 organisms. Animals were injected with the same concentration of *L. interrogans* serovar Icterohaemorrhagiae strain Lai as a positive control while three guinea pigs were administrated with modified TEPCKNN medium as a negative control. Clinical symptoms were monitored by daily observations during the study period and moribund animals were euthanized immediately. Kidneys, lungs, and liver were collected and stored in 10% formalin for histopathological examination.

16S rRNA Analysis

16S rRNA analysis of strain L231 was performed as described previously (Merien et al., 1992). Briefly, the 16s rRNA gene was amplified with the primers LA (5′-GGCGGCGCGTCTT AAACATG-3′) and LB (5′-TTCCCCCCATTGAGCAAGATT-3′). The assembled sequence was then aligned against other 16S rRNA sequences available in the GenBank using BLAST[1].

Genome Sequencing and Assembly

Genome sequencing of strain L231 was performed using Illumina Hiseq 2500 (Illumina, Little Chesterford, Essex). The Illumina data were *de novo* assembled using the program Velvet 1.2.03 (Zerbino and Birney, 2008) with the following custom parameters: hash-length was 81–111 and coverage cut-off was 30. The draft genome assemblies were aligned in a pair-wise fashion using Mauve (Darling et al., 2004), and using *L. interrogans* serovar Icterohaemorrhagiae strain Lai as a reference. Details for the genomic assembly were summarized in **Table 1**. This genome shotgun project has been deposited at DDBJ/ENA/GenBank under the accession MSFX00000000. The version described in this paper is version MSFX01000000.

Genome Analysis

Putative protein-coding sequences were predicted with the GeneMark program (Besemer et al., 2001) and functional

---

[1]http://www.ncbi.nlm.nih.gov/BLAST

TABLE 1 | Summary of *L. weilii* serovar Heyan strain L231 genome sequencing, assembly and annotation*.

Strain name	L231
Contig number	208
N50	54,534 bp
Size of assembled genomes	4,268,389 bp
Maximum contig length	167,804 bp
Sequencing depth	118-folds
Assembly ratio	97.90%
Number of predicted genes	4,794
Gene average length	750 bp
GC content	40.67%
tRNA genes	37

N50 is used to describe the length of contigs, and the size sum of contigs larger than N50 (54,534 bp) account for half of the genome size. Assembly ratio represents that 97.9% of all the sequenced reads was assembled into this genome (208 contigs).

annotation was performed by searching against the NCBI non-redundant protein database. Clusters of Orthologous Group (COGs) assignment (Tatusov et al., 2000) were conducted by RPS-BLAST using the NCBI CDD library (Marchler-Bauer et al., 2007). The metabolic pathways were constructed based on Kyoto Encyclopedia of Genes and Genomes (KEGG) database (Kanehisa et al., 2004). To further understand the evolution and pathogenesis of *L. weilii*, previously published genomes from several spirochetal species were also included in the phylogenetic analysis, such as those of non-pathogenic (*L. biflexa*), intermediate (*L. licerasiae*), and other pathogenic *Leptospira* species (Ren et al., 2003; Nascimento et al., 2004; Bulach et al., 2006; Picardeau et al., 2008; Zhong et al., 2011; Ricaldi et al., 2012; Lehmann et al., 2013; Xu et al., 2016) (**Table 2**). Genome alignments were performed using Mauve (Darling et al., 2004), and Pan-genome analysis was conducted with PGAP 1.11 using the MultiParanoid method (Zhao et al., 2012) with the following parameters: intraspecies coverage, 50%; intraspecies identity, 50%; interspecies coverage 50%; and interspecies identity 20%. A preliminary phylogenetic tree was constructed using PHYML (Guindon and Gascuel, 2003) and the concatenated protein sequences of 1359 putative orthologs selected by sequence similarity (Supplementary Table S1). To avoid errors derived from concatenating genes with different evolutionary pressures, we further selected genes from 235 ortholog groups that had been validated as having a neutral or nearly neutral evolution (Llanes et al., 2016). An alignment of the concatenated amino acid sequence of those genes was used to build a refined maximum likelihood tree for further

investigating the evolutionary position of *L. weilii* serogroup Manhao.

RESULTS AND DISCUSSION

Virulence Analysis

To evaluate the virulence of the *Leptospira* isolates, the guinea pig model of leptospirosis were used in this study (Merien et al., 1998; Nally et al., 2004). The results showed that distinct clinical symptoms such as decreased activity ruffled hair coat and hemorrhage from the mouth in some animals were observed on day 4 post-infection from challenged group. No abnormalities were found in the negative control guinea pigs. While most of challenged-animals appeared moribund on day 7 post-infection, guinea pigs from all groups were sacrificed at day 7. Post-mortem analysis showed multiple areas of hemorrhaging visible on the surfaces of the lungs and extensive hemorrhaging on peritoneal surfaces (data not shown). The histopathological analysis further demonstrated that these infected animals developed typical leptospirosis pathological changes, including significant intra-alveolar hemorrhage, edema of alveolar septa and inflammatory infiltrate in the lungs, renal epithelium degeneration, interstitial nephritis, and glomerular capillary congestion in the kidneys, hepatic cell vacuolar degeneration and inflammatory infiltration (**Figure 1**).

It has been reported that *Leptospira* serogroups have specific host preferences; rats are the reservoir for serogroup Icterohaemorrhagiae while house mice for serogroup Ballum

TABLE 2 | *Leptospira* strain included in the comparative genomic analysis.

| Strain name	Species	Serovar	Genebank accession.	Lifestyle
56105	*L. weilii*	Sarmin	SRX673804	Pathogenic
56145	*L. weilii*	Worsfoldi	SRX673805	Pathogenic
56621	*L. weilii*	Anhoa	SRX673806	Pathogenic
56622	*L. weilii*	Hainan	SRX673807	Pathogenic
56646	*L. weilii*	Whitcombi	SRX673808	Pathogenic
56655	*L. weilii*	Mini	SRX673809	Pathogenic
56674	*L. weilii*	Menglian	SRX673811	Pathogenic
56679	*L. weilii*	Liangshan	SRX673810	Pathogenic
56601	*L. interrogans*	Icterohaemorrhagiae	NC_004342.2, NC_004343.2	Pathogenic
L550	*L. borgpetersenii*	Hardjobovis	NC_008508.1, NC_008509.1	Pathogenic
3522 CT	*L. kirschneri*	Cynopteri	AHMN00000000.2	Pathogenic
1342KT	*L. santarosai*	Shermani	AOHB00000000.2	Pathogenic
CZ 214T	*L. noguchii*	Panama	AKWY00000000.2	Pathogenic
L 60^T	*L. alexanderi*	Manhao	AHMT00000000.2	Pathogenic
56643	*L. alexanderi*	Weaveri	SRX673674	Pathogenic
56650	*L. alexanderi*	Nanla	SRX673672	Pathogenic
56159	*L. alexanderi*	Erinacei auriti	SRX672003	Pathogenic
56640	*L. alexanderi*	Banna	SRX673675	Pathogenic
56659	*L. alexanderi*	Lichuan	SRX673676	Pathogenic
80–412	*L. alstonii*	Pingchang	AOHD00000000.2	Pathogenic
Bejo-Iso9T	*L. kmetyi*	Malaysia	AHMP00000000.2	Pathogenic
Ames	*L. biflexa*	Patoc	NC_010842.1, NC_010845.1,NC_010846.1	Non-pathogenic
VAR 010	*L. licerasiae*	Varillal	NZ_AHOO00000000.2	Intermediate
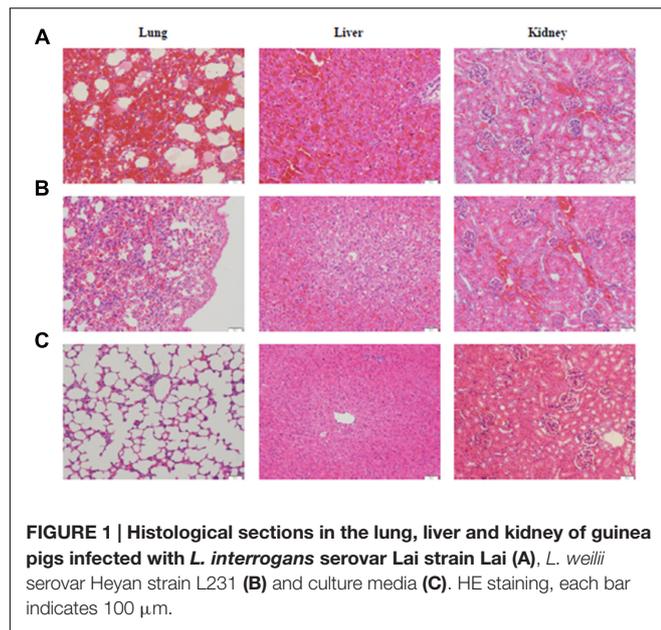
FIGURE 1 | Histological sections in the lung, liver and kidney of guinea pigs infected with *L. interrogans* serovar Lai strain Lai (A), *L. weilii* serovar Heyan strain L231 **(B)** and culture media **(C)**. HE staining, each bar indicates 100 μm.

(Levett, 2001; Ko et al., 2009). Furthermore, previous studies also found that two strains of *L. borgpetersenii* serovar Hardjobovis have different capacities for infecting animal models in virulence studies, suggesting that the virulence of *Leptospira* spp. is strain dependent (Faine et al., 1999). *Leptospira* serogroup Manhao, is only found in China, which includes Lincang, Lushui, Qingshui, Manhao, Lichuan, and Heyan serovars. Until now, all isolates of serogroup Manhao except only one (porcine) were from human patients (Qin et al., 2004). Animal experiments showed that human isolate L231 of *Leptospira* serovar Heyan was a low virulence strain in guinea pigs requiring a high dosage of inoculation (1×10^9) to induce mortality, and is able to infect rodents, suggesting that some rodent species may serve as animal reservoirs of serogroup Manhao in China. Serogroup Icterohaemorrhagiae is the most predominant epidemic-causing serogroup (among 18) with more than 70 serovars (Zhang et al., 2015). Some serovars from this serogroup are only found in China (Qin et al., 2004; Zhang et al., 2012). To understand comprehensively the pathogenesis of leptospirosis and develop a better control strategy for this disease, further studies are required to investigate the epidemiology, maintenance host and the virulence mechanisms of the pathogenic *Leptospira* serogroups, especially for those found exclusively in China.

DNA Sequencing and Preliminary Characterization of Genomic Features

Leptospira strain L231 was characterized by 16S rRNA gene sequencing. The amplification and sequencing of the 16S rRNA gene indicated that the strain L231 belonged to the pathogenic species *L. weilii*, which was different from other serovars of serogroup Manhao (i.e., species of serovars Manhao and Lichuan was *L. alexanderi*; Brenner et al., 1999; Fouts et al., 2016). LPS is the major antigen involved in *Leptospira* serological classification (Kmety and Dikken, 1993; de la Pena-Moctezuma
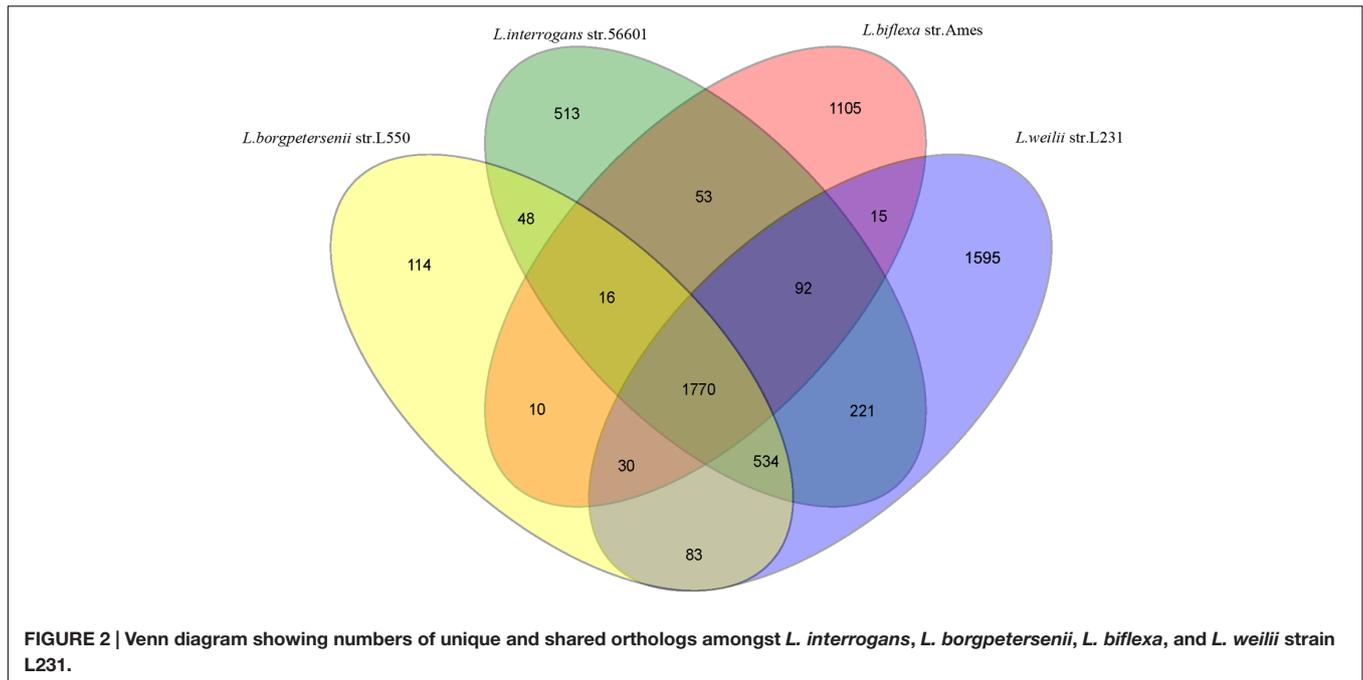
et al., 2001), whereas species differentiation is closely associated with DNA heterogeneity (Brenner et al., 1999). Recently, it has been reported that bacterial strains within the same serovar grouping may belong to distinct species, having evolved through genetic drift, homologous recombination or horizontal gene transfer (de la Pena-Moctezuma et al., 1999; Xu et al., 2016).

To further characterize the new serovar uncovered in this study, genome sequencing technique with high resolution was used. The draft genome was assembled into 208 contigs with a total length of 4,268,389 bp, 300 kb larger than the smallest *Leptospira* genome (*L. borgpetersenii* with an average genome size of 3.88 Mb). The genomic analysis showed that the G+C content was 40.67%. Gene prediction on the strain L231 genome showed that 84.2% of the total sequence comprises coding regions, containing 4,794 ORFs with an average size of 750 bp. Among all of the protein-coding genes (CDSs) in strain L231, 46.1% were assigned to a functional category of COGs (Supplementary Table S2). Further analysis showed that the L231 genome shares 2617 orthologs with *L. interrogans*, 2417 with *L. borgpetersenii* and 1907 with *L. biflexa*. A total of 1770 orthologs were present in all genomes analyzed (**Figure 2**). It was also determined that 1,595 ortholog clusters of strain L231 share no orthologs with the other three species, with 1,358 of these (78.4%) annotated as hypothetical proteins (data not shown).

Similar to other pathogenic and slow-growing species (Ren et al., 2003; Bulach et al., 2006), a total of 37 transfer RNA (tRNA) genes were revealed in strain L231 genome, representing all 20 amino acids. In contrast, the fast-growing *L. biflexa* has 35 tRNA genes (Picardeau et al., 2008). The finding further suggests that the growth rate of *Leptospira* is not restricted by low number of tRNA genes, but rather is associated with distinct metabolic capacities amongst *Leptospira* spp. (Ren et al., 2003; Picardeau et al., 2008). To further analyze other non-coding RNA elements in the strain L231 genome, riboswitch predictions were performed using Riboswitch Scanner (Mukherjee and Sengupta, 2016). In agreement with the previous studies (Ricaldi et al., 2012; Fouts et al., 2016), three riboswitches associated with transport and vitamin biosynthesis were found. Of these, two elements are located upstream of *L231_gene1384* (encoding a putative TonB-dependent receptor) and *L231_gene3285* (encoding a putative sirohydrochlorin cobaltochelatase), respectively. It is reported that a complete genetic support for vitamin B12 autotrophy exists in pathogenic but not saprophyte *Leptospira*, suggesting the synthesis of B12 *de novo* may be critical for *Leptospira* to infect mammals in the face of B12 sequestration by host (Fouts et al., 2016). Although the mechanistic details of these riboswitches are still unknown, these non-coding RNAs seems to be involved in the regulation of transport and biosynthesis of vitamin B12 in pathogenic *Leptospira*, facilitating adaptation to different host environments. Furthermore, another riboswitch, found in strain L231, is upstream of *L231_gene2603* (encoding a hydroxymethylpyrimidine phosphate synthase, ThiC), which exists in almost all reported leptospiral genomes.

Energy Metabolism

Based on the KEGG analysis, the stain L231 had the potential to synthesize 17 amino acids; methionine, histidine and asparagine

FIGURE 2 | Venn diagram showing numbers of unique and shared orthologs amongst *L. interrogans*, *L. borgpetersenii*, *L. biflexa*, and *L. weilii* strain L231.

being the exceptions. Although some enzymes for *de novo* synthesis are not apparently present, methionine could be synthesized from homocysteine, histidine could be synthesized from an L-histidinol intermediate and isoleucine could be synthesized from 2-Oxobutanoate (Crout et al., 1980; Winkler and Ramos-Montañez, 2009; Ferla and Patrick, 2014). For the tricarboxylic acid (TCA) cycle, the gene for pyruvate carboxylase synthesis from pyruvate was not found in L231. Malate dehydrogenase (mdh; *L231_gene702*) and citrate synthase (*L231_gene1259*) were identified, the products of which catalyzed the synthesis of oxaloacetate from malate and citrate (Panchenko and Vinogradov, 1991).
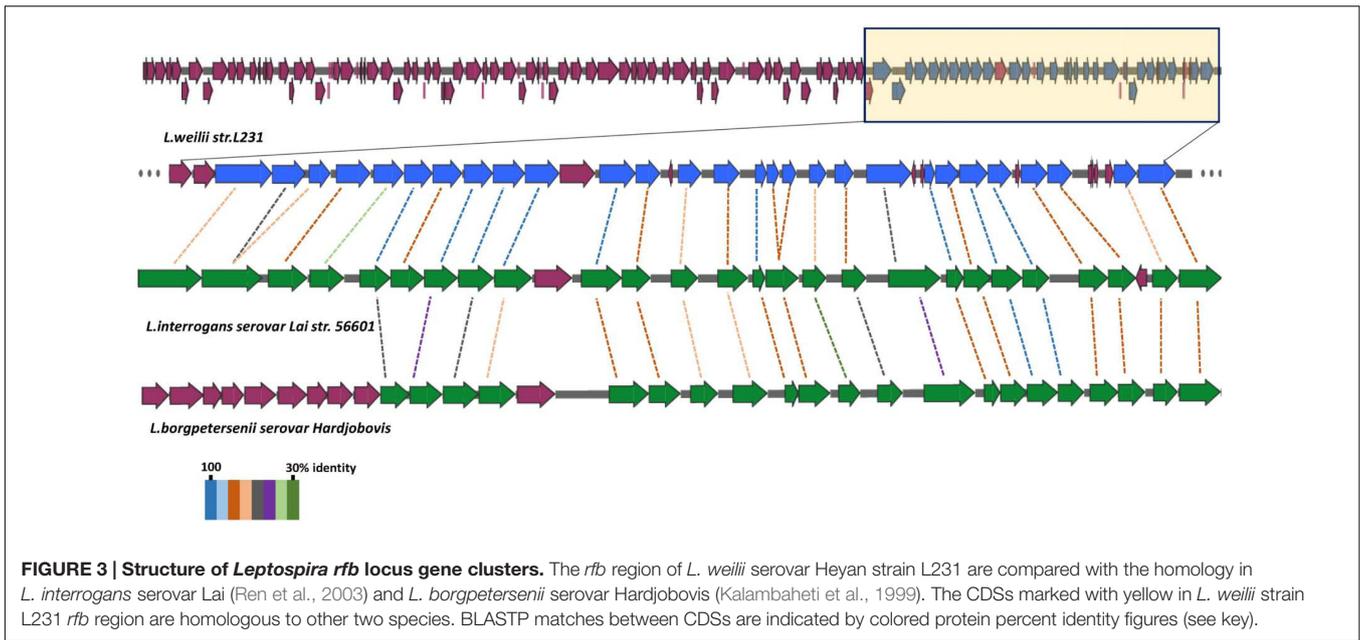
Similar to other pathogenic *Leptospira* species (Ren et al., 2003; Nascimento et al., 2004), genes involved in the bacterial glycerol metabolism were found in the strain L231 genome. Those genes encode glycerol-3-phosphate dehydrogenase (*L231_gene224*), glycerol-3-phosphate transporter (*L231_gene320*), glycerol kinase (*L231_gene721*), and 1-acyl-sn-glycerol-3-phosphate acyltransferase (*L231_gene79*). These findings suggest that that glycerol and fatty acids may be generated from phospholipid degradation (Nascimento et al., 2004). It is well known that glycerol is the central structural component of the major classes of biological lipids, triglycerides and phosphatidyl-rich phospholipids, and also serves as an important intermediate in carbohydrate and lipid metabolism (Thomas et al., 1990).

LPS Biosynthetic System

Many studies have demonstrated that leptospiral LPS plays an important role on host immunity responses and in serovar determination in *Leptospira* (Kmety and Dikken, 1993; de la Pena-Moctezuma et al., 1999, 2001; Faine et al., 1999). Changes in genes of the LPS biosynthetic pathway (*rfb*) locus are associated

with serovar diversity among *Leptospira* spp. (de la Pena-Moctezuma et al., 1999, 2001).

In agreement with the previous findings (Fouts et al., 2016), a larger *rfb* locus (spanning approximately 111 kb was observed in the strain L231 genome (Supplementary Table S3), when compared with those of *L. interrogans* serovar Lai (36.0 kb) and *L. borgpetersenii* serovar Hardjobovis (36.7 kb; Mitchison et al., 1997; Kalambaheti et al., 1999; Ren et al., 2003). While loci from *L. interrogans* and *L. borgpetersenii*, respectively, contain 91 and 76 genes, strain L231 *rfb* locus contains 120 genes (**Figure 3**, Supplementary Figure S1 and Supplementary Table S3). Comparison of the nucleotide sequence of the *rfb* locus of serogroup Manhao serovar Manhao strain L60^T (121 kb) and Heyan strain L231 revealed more than 91.5% identity in the 73 kb homologous regions, including 81 pairs of orthologous genes (**Figure 3** and Supplementary Figure S1), hence reflecting the antigenic/serologic specificity of serogroup Manhao. Although most of the unique genes of *rfb* locus in strain L231 were hypothetical proteins, three genes encoding aminotransferase PglE were found (Supplementary Table S3). PglE belongs to the aspartate aminotransferase superfamily, and is involved in a novel sugar N,N′-diacetylbacillosamine biosynthesis, which contributes to bacterial membrane protein glycosylation (Riegert et al., 2015). Furthermore, those differences may occur in different serovars from the same serogroup, providing further evidence that different *Leptospira* genospecies or strains may be classified as being in the same serogroup or serovar (de la Pena-Moctezuma et al., 1999). Similar to other pathogenic *Leptospira* species, a complete dTDP-rhamnose biosynthesis gene cluster encoding *rfb ABCD* was present in strain L231 genome (*L231_gene1714-L231_gene1717*). Previous studies demonstrate that LPS of pathogenic *L. interrogans* contain high amounts of rhamnose (Vinh et al., 1989) whereas LPS of intermediately

FIGURE 3 | Structure of *Leptospira rfb* locus gene clusters. The *rfb* region of *L. weilii* serovar Heyan strain L231 are compared with the homology in *L. interrogans* serovar Lai (Ren et al., 2003) and *L. borgpetersenii* serovar Hardjobovis (Kalambaheti et al., 1999). The CDSs marked with yellow in *L. weilii* strain L231 *rfb* region are homologous to other two species. BLASTP matches between CDSs are indicated by colored protein percent identity figures (see key).

pathogenic *L. licerasiae,* occasionally causing human infection, has a small amount of rhamnose (Patra et al., 2015). Furthermore, compared with those of intermediately pathogenic *Leptospira*, the chemical composition of pathogenic *L. interrogans* serovar Copenhageni LPS comprises of some specific sugars, such as fucose and acetylglucosamine (Patra et al., 2015). It is well known that LPS is involved in pathogenesis of Gram-negative bacteria, including pathogenic *Leptospira* species. Composition and structural complexity of LPS residues in pathogenic *Leptospira* species are thought to be greater than those of intermediate and non-pathogenic species (Murray et al., 2010; Patra et al., 2015). Due to large variations in *rfb* gene clusters among different pathogenic species, more experimental investigations are needed to comparatively characterize the LPS of other pathogenic species, particularly those species with larger *rfb* gene cluster, to better understand mechanisms of pathogenesis and colonization of pathogenic leptospires.

Cell Motility, Putative Virulence Factors and Protein Secretion

Bacterial motility has long been shown to be a central factor in enabling pathogens to penetrate host tissue barriers during infection, and it has also been suggested that it plays a crucial role in the pathogenesis of *Leptospira* (Liao et al., 2009). We identified 32 CDSs encoding motility-related proteins. Most CDSs (22/32) encoded putative proteins involved in flagellar biosynthesis and export, while the remaining CDSs (10/32) were predicted to encode proteins from bacterial chemotaxis.

Potential virulence factors in the strain L231 genome were identified using BLAST search in the VFDB database. A total of 407 putative virulence factors are presented in Supplementary Table S4. Some of these proteins (such as hemolysins and flagella) have been demonstrated to be virulence factors in pathogenic *Leptospira* species, although most are homologous to

TABLE 3 | Comparison of *L. weilii* serovar Heyan strain L231 with other *Leptospira* species*.

Species	Genomic similarity	Gene identity	Protein identity
L. interrogans	43.20%	82.6% (2507)	84.7% (2618)
L. borgpetersenii	61.50%	91.9% (2465)	92.3% (2550)
L. kirschneri	43.30%	82.9% (2490)	84.9% (2581)
L. alexanderi	77.10%	95.3% (2699)	95.8% (2701)
L. noguchii	42.30%	82.7% (2508)	85.1% (2603)
L. santarosai	59.30%	89.2% (2587)	90.6% (2603)
L. alstonii	57.40%	85.5% (2611)	87.2% (2648)
L. kmetyi	40.50%	83.5% (2461)	85.1% (2584)
L. biflexa	0.12%	81.6% (213)	51.4% (2119)
L. licerasiae	1.10%	80.9% (764)	62.9% (2344)

Number in brackets represented number of gene or protein orthologs.

proteins that had been shown to be virulent in other pathogens. Furthermore, previous studies report that some lipoproteins may be essential for leptospiral virulence, which play important role in infection (Koizumi and Watanabe, 2003; Ristow et al., 2007). We also identified 48 putative lipoproteins and three OmpA homologs (Supplementary Table S5). It is suggested that further experiments should be conducted to investigate the function of potential virulence in *L. weilii* strain L231 infection.

Currently, at least eight different pathways have been reported in Gram-negative bacteria for secretion of proteins across the cell envelope (Pallen et al., 2003; Kostakioti et al., 2005). Eleven genes, separately involved in Type II and the general secretory (Sec)-signal recognition particle (SRP) systems, were revealed in the genome of strain L231. The Type II secretion genes are organized in a single cluster (*L231_gene1861-1865*) in strain L231. This translocation system is Sec-dependent, and is responsible for the extracellular transport of a large number of toxins and hydrolytic enzymes (Kostakioti et al., 2005; Haake and Zuckert, 2015).
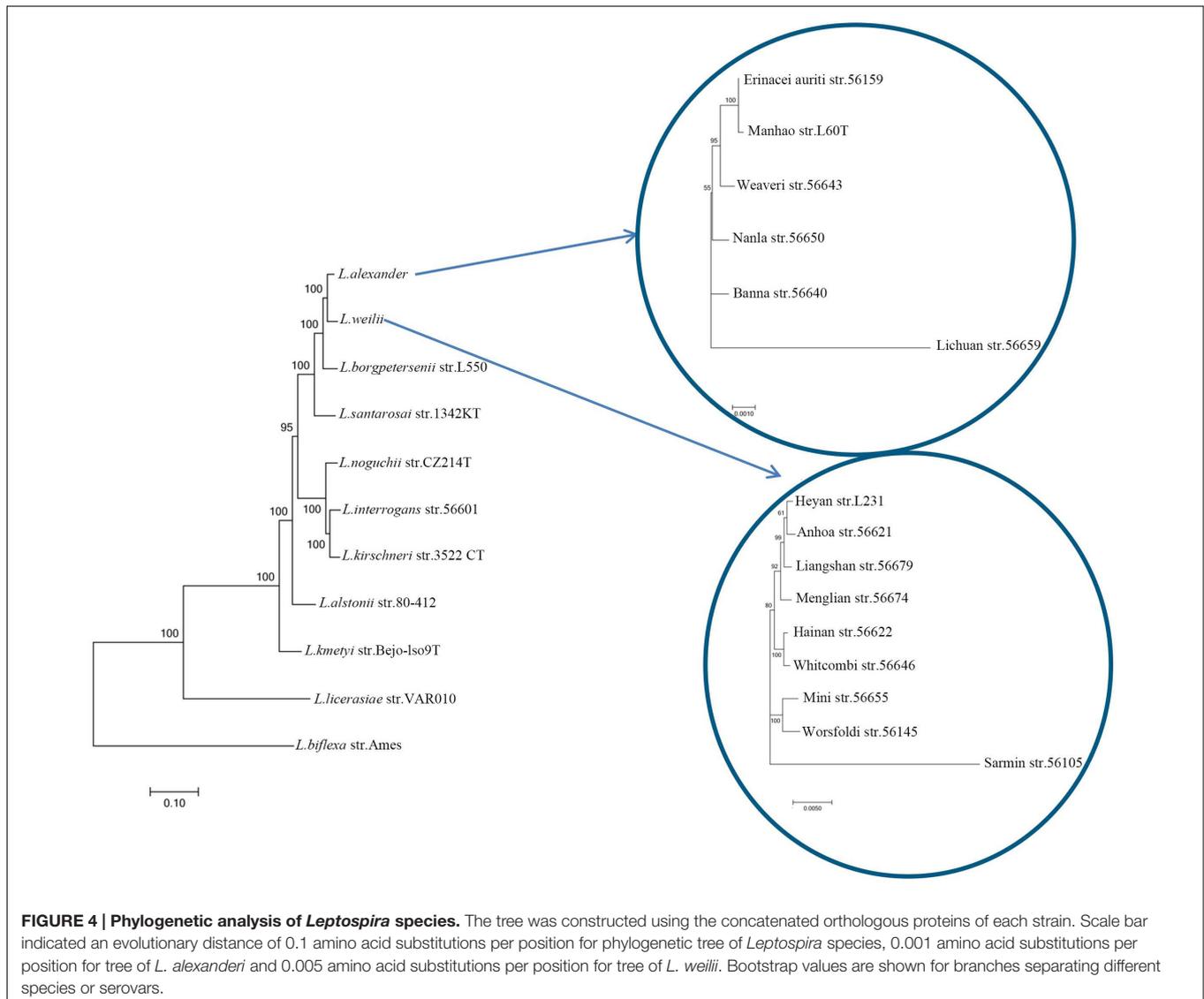
These results suggested that strain L231 utilize Type II secretory systems to export proteins from the cytoplasm. Furthermore, we identified 26 genes involved in ATP-binding cassette (ABC) transporters, but only four sets of complete ABC type clusters for transporting lipopolysaccharides, lipoproteins and sulfates.

Phages and CRISPR-Cas System

Prophages are commonly found in bacterial genomes, where they play an important role in the evolution of the host bacteria (Fortier and Sekulovic, 2013). Phage_Finder predictive analysis revealed one prophage (27,454 bp) in strain L231 genome (Supplementary Table S6), which have been also found in the genomes of other pathogenic, or intermediate *Leptospira* species (Fouts et al., 2016). In order to be able to survive in the host, intermediate and pathogenic *Leptospira* species have to adapt to a variety of growth environments and conditions (Xu et al., 2016). It is speculated that the resistance mechanisms and metabolic advantages provided by prophages facilitate pathogen adaptation

and evolution into a variety of different genospecies, allowing colonization of a wide variety of different mammalian hosts.

Clustered regularly interspaced short palindromic repeats (CRISPR) and CRISPR-associated (Cas) genes are found in many sequenced bacterial genomes, which contribute to resistance to exogenous genetic elements, such as plasmids and phages, as well as regulation of gene expression (Louwen et al., 2014). In the genome of strain L231, three CRISPRs with length of 756, 1,123 and 333 bp were revealed (Supplementary Table S7). Previous studies have shown that all nine pathogenic *Leptospira* species and some intermediate species have potential CRISPRs-Cas systems, whereas non-pathogenic species seem to lack these sequences, suggesting that specific CRISPRs-Cas systems among pathogenic species may play a role in invasion and evasion of host immune systems (Fouts et al., 2016). Furthermore, it was reported that pathogen subspecies of bacteria with variant CRISPR-Cas systems show substantial differences with respect to virulence (Louwen et al., 2014). Therefore, further studies

FIGURE 4 | Phylogenetic analysis of *Leptospira* species. The tree was constructed using the concatenated orthologous proteins of each strain. Scale bar indicated an evolutionary distance of 0.1 amino acid substitutions per position for phylogenetic tree of *Leptospira* species, 0.001 amino acid substitutions per position for tree of *L. alexanderi* and 0.005 amino acid substitutions per position for tree of *L. weilii*. Bootstrap values are shown for branches separating different species or serovars.

aimed at the characterization of CRISPRs-Cas systems should be performed among pathogenic *Leptospira* spp., which will certainly provide deeper insight into the mechanisms leading to control of virulence via CRISPRs-Cas system.

Comparative and Phylogenomic Analysis

Genomic analysis showed that the genome sequence similarity varied from less than 1% to more than 77% for pair-wise comparisons of strain L231 and other pathogenic, intermediate, and non-pathogenic species (**Table 3**). Furthermore, greater variation on predicted protein levels was found among different pathogenic *Leptospira* species. For example, the similarity between predicted homologs at the amino acid level in the *L. weilii* strain L231 genome was 85.1 and 92.3% in comparison with *L. noguchii* and *L. borgpetersenii*, respectively (**Table 3**). These observations further suggested high genomic plasticity among pathogenic *Leptospira* spp. (Xu et al., 2016).

To further understand the molecular evolution of *L. weilii*, the pan-genome for *L. weilii* species was calculated based on nine available genomes. Accumulation curve for the pan-genome suggested it to be open, with a $\gamma = 0.3486$. According to Heap's Law, the pan-genome is considered open when the power law exponents $\gamma > 0$ (Supplementary Figure S2). The result revealed that the size of the *L. weilii* pan-genome is relatively large, comprising 9,639 genes. The large pan-genome in *L. weilii* species may thus provide more insights into bacterial evolution because pathogenic *Leptospira* species adapt to distinct environments via multiple ways of exchanging genetic material (Bulach et al., 2006). Further analysis identified 2,771 annotated core genes in *L. weilii* species (Supplementary Table S8), accounting for only 28.7% of total pan-genes. In agreement with our previous studies (Xu et al., 2016), some known virulence factors including bacterial toxins and surface proteins were found in the core genomes of *L. weilii* species. In addition, many of the core genes were also involved in metabolic pathways, together reflecting that conserved core genes may be closely related to bacterial pathogenesis.

A preliminary phylogenetic tree built with concatenated sequences of putative orthologs (Supplementary Figure S3), shows that pathogenic, intermediate and non-pathogenic species are separately clustered, which is consistent with previous phylogenetic hypotheses based on 16S rRNA gene sequence analysis (Brenner et al., 1999). Different from the cluster formed by those closely related *L. interrogans*, *L. kirschneri* and *L. noguchii*, *L. weilii* seems to be closer to *L. borgpetersenii* and *L. alexanderi*. Previous studies have reported that neutral synonymous sites from genes have a reasonably better phylogenetic signal (Massey et al., 2008). To further explore the origin of *L. weilii* serogroup Manhao and its evolutionary relationship with other member of the genus *Leptospira*, a set of 235 ortholog groups predicted to be subjected to neutral evolution (with an overall dN/dS ratio between 0.2–2.0; Llanes et al., 2016) were used to reconstruct a refined phylogenetic tree (**Figure 4**). The topology of this new tree is in agreement with that of Supplementary Figure S3, but individual *L. weilii* and *L. alexanderi* specimens can now be subdivided. The phylogenetic tree shows three separate subgroups within the *L. weilii* clade. The evolutionary position of serovar Heyan (strain L231) is closer to serovar Anhoa (strain 56621). The topology also revealed that other serovars from serogroup Manhao, namely serovar Lichuan (strain 56659) and serovar Manhao (strain L60^T), belong to *L. alexanderi* rather than to *L. weilii*. These findings further support the hypothesis that the presence of similar *rfb* loci in different *Leptospira* species may be closely associated with gene acquisition via lateral gene transfer (Llanes et al., 2016).

CONCLUSION

In this study, our experiment evidences indicated a new serovar Heyan strain L231 can infect rodents, suggesting that the monitoring of specific host reservoir of serogroup Manhao involved in transmission should be strengthened for better control of leptospirosis in China. Genomic analysis provided comprehensive information about strain L231 molecular features. Distinctive LPS biosynthesis locus of serovar Heyan provides further evidence for the unique antigenic/serologic specificity of serogroup Manhao. More importantly, our comparative genomic analysis confirmed a high genomic plasticity among pathogenic species, indicating that different species may undergo a distinct evolutionary process to adapt to different hosts or environmental niches. Unlike the other two serovars of serogroup Manhao, serovar Heyan seems to be more closely related to serovar Anhoa, within the *L. weilii* clade. All these findings have important implications for understanding evolution of pathogenic *Leptospira* in China.

AUTHOR CONTRIBUTIONS

JW and YX conceived and designed the study. YX, YZ, JZ, ZL, SC, XX, and QY performed the experiments. HZ, YW, Y-FC, and YX conducted the bioinformatic analyses of the data. JW, YX, Y-FC, and HZ wrote the paper. All authors read and approved the final manuscript.

FUNDING

ACKNOWLEDGMENT

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: http://journal.frontiersin.org/article/10.3389/fmicb.2017.00149/full#supplementary-material

REFERENCES

Besemer, J., Lomsadze, A., and Borodovsky, M. (2001). GeneMarkS: a self-training method for prediction of gene starts in microbial genomes. Implications for finding sequence motifs in regulatory regions. *Nucleic Acids Res.* 29, 2607–2618.

Bharti, A. R., Nally, J. E., Ricaldi, J. N., Matthias, M. A., Diaz, M. M., Lovett, M. A., et al. (2003). Leptospirosis: a zoonotic disease of global importance. *Lancet Infect. Dis.* 3, 757–771. doi: 10.1016/S1473-3099(03)00830-2

Bourhy, P., Collet, L., Brisse, S., and Picardeau, M. (2014). *Leptospira mayottensis* sp. nov., a pathogenic species of the genus *Leptospira* isolated from humans. *Int. J. Syst. Evol. Microbiol.* 64, 4061–4067. doi: 10.1099/ijs.0.066597-0

Brenner, D. J., Kaufmann, A. F., Sulzer, K. R., Steigerwalt, A. G., Rogers, F. C., and Weyant, R. S. (1999). Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for *Leptospira alexanderi* sp. nov. and four new *Leptospira* genomospecies. *Int. J. Syst. Bacteriol.* 49(Pt 2), 839–858. doi: 10.1099/00207713-49-2-839

Bulach, D. M., Zuerner, R. L., Wilson, P., Seemann, T., Mcgrath, A., Cullen, P. A., et al. (2006). Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proc. Natl. Acad. Sci. U.S.A.* 103, 14560–14565. doi: 10.1073/pnas.0603979103

Crout, D. H., Gregorio, M. V., Muller, U. S., Komatsubara, S., Kisumi, M., and Chibata, I. (1980). Stereochemistry of the conversions of L-threonine and D-threonine into 2-oxobutanoate by the L-threonine and D-threonine dehydratases of *Serratia marcescens*. *Eur. J. Biochem.* 106, 97–105. doi: 10.1111/j.1432-1033.1980.tb06000.x

Darling, A. C., Mau, B., Blattner, F. R., and Perna, N. T. (2004). Mauve: multiple alignment of conserved genomic sequence with rearrangements. *Genome Res.* 14, 1394–1403. doi: 10.1101/gr.2289704

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

de la Pena-Moctezuma, A., Bulach, D. M., Kalambaheti, T., and Adler, B. (1999). Comparative analysis of the LPS biosynthetic loci of the genetic subtypes of serovar Hardjo: *Leptospira interrogans* subtype Hardjoprajitno and *Leptospira borgpetersenii* subtype Hardjobovis. *FEMS Microbiol. Lett.* 177, 319–326. doi: 10.1016/S0378-1097(99)00333-X

Faine, S., Adler, B., Bolin, C., and Perolat, P. (1999). "Leptospira and leptospirosis," in *Current Topics in Microbiology and Immunology*, 2nd Edn, ed. B. Adler (Berlin: Springer).

Ferla, M. P., and Patrick, W. M. (2014). Bacterial methionine biosynthesis. *Microbiology* 160, 1571–1584. doi: 10.1099/mic.0.077826-0

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

Fouts, D. E., Matthias, M. A., Adhikarla, H., Adler, B., Amorim-Santos, L., Berg, D. E., et al. (2016). What makes a bacterial species pathogenic?:Comparative genomic analysis of the genus *Leptospira*. *PLoS Negl. Trop. Dis.* 10:e0004403. doi: 10.1371/journal.pntd.0004403

Guindon, S., and Gascuel, O. (2003). A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Syst. Biol.* 52, 696–704. doi: 10.1080/10635150390235520

Haake, D. A., and Zuckert, W. R. (2015). The leptospiral outer membrane. *Curr. Top. Microbiol. Immunol.* 387, 187–221. doi: 10.1007/978-3-662-45059-8_8

Herrmann, J. L., Bellenger, E., Perolat, P., Baranton, G., and Saint Girons, I. (1992). Pulsed-field gel electrophoresis of NotI digests of leptospiral DNA: a new rapid method of serovar identification. *J. Clin. Microbiol.* 30, 1696–1702.

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

Kanehisa, M., Goto, S., Kawashima, S., Okuno, Y., and Hattori, M. (2004). The KEGG resource for deciphering the genome. *Nucleic Acids Res.* 32, D277–D280. doi: 10.1093/nar/gkh063

Kmety, E., and Dikken, H. (1993). *Classification of the Species Leptospira interrogans and History of its Serovars*. Groningen: Groningen University Press, 104.

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

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

Kostakioti, M., Newman, C. L., Thanassi, D. G., and Stathopoulos, C. (2005). Mechanisms of protein export across the bacterial outer membrane. *J. Bacteriol.* 187, 4306–4314. doi: 10.1128/JB.187.13.4306-4314.2005

Lehmann, J. S., Fouts, D. E., Haft, D. H., Cannella, A. P., Ricaldi, J. N., Brinkac, L., et al. (2013). Pathogenomic inference of virulence-associated genes in *Leptospira interrogans*. *PLoS Negl. Trop. Dis.* 7:e2468. doi: 10.1371/journal.pntd.0002468

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

Liao, S., Sun, A., Ojcius, D. M., Wu, S., Zhao, J., and Yan, J. (2009). Inactivation of the fliY gene encoding a flagellar motor switch protein attenuates mobility and virulence of *Leptospira interrogans* strain Lai. *BMC Microbiol.* 9:253. doi: 10.1186/1471-2180-9-253

Llanes, A., Restrepo, C. M., and Rajeev, S. (2016). Whole genome sequencing allows better understanding of the evolutionary history of *Leptospira interrogans* serovar Hardjo. *PLoS ONE* 11:e0159387. doi: 10.1371/journal.pone.0159387

Louwen, R., Staals, R. H., Endtz, H. P., Van Baarlen, P., and Van Der Oost, J. (2014). The role of CRISPR-Cas systems in virulence of pathogenic bacteria. *Microbiol. Mol. Biol. Rev.* 78, 74–88. doi: 10.1128/MMBR.00039-13

Marchler-Bauer, A., Anderson, J. B., Derbyshire, M. K., Deweese-Scott, C., Gonzales, N. R., Gwadz, M., et al. (2007). CDD: a conserved domain database for interactive domain family analysis. *Nucleic Acids Res.* 35, D237–D240. doi: 10.1093/nar/gkl951

Massey, S. E., Churbanov, A., Rastogi, S., and Liberles, D. A. (2008). Characterizing positive and negative selection and their phylogenetic effects. *Gene* 418, 22–26. doi: 10.1016/j.gene.2008.03.017

Merien, F., Amouriaux, P., Perolat, P., Baranton, G., and Saint Girons, I. (1992). Polymerase chain reaction for detection of *Leptospira* spp. in clinical samples. *J. Clin. Microbiol.* 30, 2219–2224.

Merien, F., Truccolo, J., Rougier, Y., Baranton, G., and Perolat, P. (1998). In vivo apoptosis of hepatocytes in guinea pigs infected with *Leptospira interrogans* serovar icterohaemorrhagiae. *FEMS Microbiol. Lett.* 169, 95–102. doi: 10.1111/j.1574-6968.1998.tb13304.x

Mitchison, M., Bulach, D. M., Vinh, T., Rajakumar, K., Faine, S., and Adler, B. (1997). Identification and characterization of the dTDP-rhamnose biosynthesis and transfer genes of the lipopolysaccharide-related rfb locus in *Leptospira interrogans* serovar Copenhageni. *J. Bacteriol.* 179, 1262–1267. doi: 10.1128/jb.179.4.1262-1267.1997

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

Mukherjee, S., and Sengupta, S. (2016). Riboswitch Scanner: an efficient pHMM-based web-server to detect riboswitches in genomic sequences. *Bioinformatics* 32, 776–778. doi: 10.1093/bioinformatics/btv640

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

Nally, J. E., Chantranuwat, C., Wu, X. Y., Fishbein, M. C., Pereira, M. M., Da Silva, J. J., et al. (2004). Alveolar septal deposition of immunoglobulin and complement parallels pulmonary hemorrhage in a guinea pig model of severe pulmonary leptospirosis. *Am. J. Pathol.* 164, 1115–1127. doi: 10.1016/S0002-9440(10)63198-7

Nascimento, A. L., Ko, A. I., Martins, E. A., Monteiro-Vitorello, C. B., Ho, P. L., Haake, D. A., et al. (2004). Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *J. Bacteriol.* 186, 2164–2172. doi: 10.1128/JB.186.7.2164-2172.2004

Npc (2010). *National Pharmacopoeia Committee,The Chinese Pharmacopoeia (2010 Version) 3rd Part*. Beijing: Chinese Medical Science and Technology Press.

Pallen, M. J., Chaudhuri, R. R., and Henderson, I. R. (2003). Genomic analysis of secretion systems. *Curr. Opin. Microbiol.* 6, 519–527. doi: 10.1016/j.mib.2003.09.005

Panchenko, M. V., and Vinogradov, A. D. (1991). Direct demonstration of enol-oxaloacetate as an immediate product of malate oxidation by the mammalian succinate dehydrogenase. *FEBS Lett.* 286, 76–78. doi: 10.1016/0014-5793(91)80944-X

Patra, K. P., Choudhury, B., Matthias, M. M., Baga, S., Bandyopadhya, K., and Vinetz, J. M. (2015). Comparative analysis of lipopolysaccharides of pathogenic and intermediately pathogenic *Leptospira* species. *BMC Microbiol.* 15:244. doi: 10.1186/s12866-015-0581-7

Picardeau, M., Bulach, D. M., Bouchier, C., Zuerner, R. L., Zidane, N., Wilson, P. J., et al. (2008). Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. *PLoS ONE* 3:e1607. doi: 10.1371/journal.pone.0001607

Qin, J. (1999). Establishment and application of Chinese *Leptospira* reference strains. *Prog. Microbiol. Immunol.* 27, 78–82.

Qin, J., Chen, M., Jiang, X., Xin, X., and Yang, Y. (2004). A new serovar in serogroup Manhao of *Leptospira interrogans*-serovar Heyan. *Chin. J. Zoonoses* 20, 474–476.

Ren, S. X., Fu, G., Jiang, X. G., Zeng, R., Miao, Y. G., Xu, H., et al. (2003). Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature* 422, 888–893. doi: 10.1038/nature01597

Ricaldi, J. N., Fouts, D. E., Selengut, J. D., Harkins, D. M., Patra, K. P., Moreno, A., et al. (2012). Whole genome analysis of *Leptospira licerasiae* provides insight into leptospiral evolution and pathogenicity. *PLoS Negl. Trop. Dis.* 6:e1853. doi: 10.1371/journal.pntd.0001853

Riegert, A. S., Young, N. M., Watson, D. C., Thoden, J. B., and Holden, H. M. (2015). Structure of the external aldimine form of PglE, an aminotransferase required for N,N'-diacetylbacillosamine biosynthesis. *Protein Sci.* 24, 1609–1616. doi: 10.1002/pro.2745

Ristow, P., Bourhy, P., Da Cruz Mcbride, F. W., Figueira, C. P., Huerre, M., Ave, P., et al. (2007). The OmpA-like protein Loa22 is essential for leptospiral virulence. *PLoS Pathog.* 3:e97. doi: 10.1371/journal.ppat.0030097

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

Shi, M., and Jiang, X. (2000). Leptospira epidemiology in the past fifty years in China. *Chin. J. Epidemiol.* 21, 228–230.

Tatusov, R. L., Galperin, M. Y., Natale, D. A., and Koonin, E. V. (2000). The COG database: a tool for genome-scale analysis of protein functions and evolution. *Nucleic Acids Res.* 28, 33–36. doi: 10.1093/nar/28.1.33

Thomas, J. R., Dwek, R. A., and Rademacher, T. W. (1990). Structure, biosynthesis, and function of glycosylphosphatidylinositols. *Biochemistry* 29, 5413–5422. doi: 10.1021/bi00475a001

Vinh, T. U., Shi, M. H., Adler, B., and Faine, S. (1989). Characterization and taxonomic significance of lipopolysaccharides of *Leptospira interrogans* serovar hardjo. *J. Gen. Microbiol.* 135, 2663–2673.

Winkler, M. E., and Ramos-Montañez, S. (2009). Biosynthesis of Histidine. *EcoSal Plus* 3, 1–34. doi: 10.1128/ecosalplus.3.6.1.9

Xu, Y., Zhang, J., Cui, S., Li, M., Zhang, Y., Xue, H., et al. (2015). Genetic stability of vaccine strains by multilocus sequence typing and pulsed-field gel electrophoresis analysis: implications for quality control of the leptospiral vaccine. *Hum. Vaccin. Immunother.* 11, 1272–1276. doi: 10.1080/21645515.2015.1020266

Xu, Y., Zhu, Y., Wang, Y., Chang, Y. F., Zhang, Y., Jiang, X., et al. (2016). Whole genome sequencing revealed host adaptation-focused genomic plasticity of pathogenic Leptospira. *Sci. Rep.* 6:20020. doi: 10.1038/srep20020

Zerbino, D. R., and Birney, E. (2008). Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res.* 18, 821–829. doi: 10.1101/gr.074492.107

Zhang, C., Wang, H., and Yan, J. (2012). Leptospirosis prevalence in Chinese populations in the last two decades. *Microbes Infect.* 14, 317–323. doi: 10.1016/j.micinf.2011.11.007

Zhang, C., Yang, H., Li, X., Cao, Z., Zhou, H., Zeng, L., et al. (2015). Molecular typing of pathogenic *Leptospira* serogroup icterohaemorrhagiae strains circulating in China during the past 50 years. *PLoS Negl. Trop. Dis.* 9:e0003762. doi: 10.1371/journal.pntd.0003762

Zhao, Y., Wu, J., Yang, J., Sun, S., Xiao, J., and Yu, J. (2012). PGAP: pan-genomes analysis pipeline. *Bioinformatics* 28, 416–418. doi: 10.1093/bioinformatics/btr655

Zhong, Y., Chang, X., Cao, X. J., Zhang, Y., Zheng, H., Zhu, Y., et al. (2011). Comparative proteogenomic analysis of the *Leptospira interrogans* virulence-attenuated strain IPAV against the pathogenic strain 56601. *Cell Res.* 21, 1210–1229. doi: 10.1038/cr.2011.46