



Published in final edited form as:

Infect Genet Evol. 2009 March ; 9(2): 196. doi:10.1016/j.meegid.2008.10.012.

Genetic diversity of the Leptospiral immunoglobulin-like (Lig) genes in pathogenic *Leptospira* spp.

Alan J.A. McBride^{a,1}, Gustavo M. Cerqueira^{b,1}, Marc A. Suchard^{c,d,e}, Angela N. Moreira^b, Richard L. Zuerner^f, Mitermayer G. Reis^a, David A. Haake^{g,h}, Albert I. Ko^{a,i}, and Odir A. Dellagostin^{b,*}

^aCentro de Pesquisa Gonçalo Moniz, Fundação Oswaldo Cruz, Salvador, BA, Brazil

^bCentro de Biotecnologia, Universidade Federal de Pelotas, Pelotas, RS, Brazil

^cDepartment of Biomathematics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

^dDepartment of Human Genetics, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

^eDepartment of Biostatistics, UCLA School of Public Health, Los Angeles, CA, USA

^fNational Animal Disease Centre, Agricultural Research Service, U.S. Department of Agriculture, Ames, Iowa, USA

^gVeterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA, USA

^hDepartment of Medicine, David Geffen School of Medicine at UCLA, Los Angeles, CA, USA

ⁱDivision of International Medicine and Infectious Disease, Weill Medical College of Cornell University, New York, NY, USA

Abstract

Recent serologic, immunoprotection, and pathogenesis studies identified the Lig proteins as key virulence determinants in interactions of leptospiral pathogens with the mammalian host. We examined the sequence variation and recombination patterns of *ligA*, *ligB*, and *ligC* among 10 pathogenic strains from five *Leptospira* species. All strains were found to have intact *ligB* genes and genetic drift accounting for most of the *ligB* genetic diversity observed. The *ligA* gene was found exclusively in *L. interrogans* and *L. kirschneri* strains, and was created from *ligB* by a two-step partial gene duplication process. The aminoterminal domain of LigB and the LigA paralog were essentially identical ($98.5 \pm 0.8\%$ mean identity) in strains with both genes. Like *ligB*, *ligC* gene variation also followed phylogenetic patterns, suggesting an early gene duplication event. However, *ligC* is a pseudogene in several strains, suggesting that LigC is not essential for virulence. Two *ligB* genes and one *ligC* gene had mosaic compositions and evidence for recombination events between related *Leptospira* species was also found for some *ligA* genes. In conclusion, the results presented here indicate that Lig diversity has important ramifications for the selection of Lig polypeptides for use in diagnosis and as vaccine candidates. This sequence information will aid the identification of highly conserved regions within the Lig proteins and improve upon the performance characteristics of the Lig proteins in diagnostic assays and in subunit vaccine formulations with the potential to confer heterologous protection.

© 2008 Elsevier B.V. All rights reserved.

*Corresponding author. Tel.: +55 53 3275 7350; fax: +55 53 3275 7555. E-mail address: odir@ufpel.edu.br (O.A. Dellagostin)..

¹The first two authors contributed equally to this work.

Appendix A. Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.meegid.2008.10.012.

Keywords

Leptospirosis; Lig; Pathogenesis; Molecular evolution; Sequence analysis

1. Introduction

Pathogenic spirochaetes belonging to the genus *Leptospira* are the agents of leptospirosis, which is considered to be the most widespread zoonosis in the world (Faine et al., 1999; Levett, 2001; Bharti et al., 2003). Susceptible animals, including humans, are infected by direct contact with urine from a reservoir host, usually rats or other rodents, or indirectly through contaminated water. Transmission occurs via dermal abrasions or inoculation of the mucous or conjunctival membranes (Faine et al., 1999). In the majority of infected individuals, leptospirosis is a self-limited disease characterized by flu-like symptoms (Faine et al., 1999). However, hepatorenal manifestations, as observed in Weil's disease, are frequent complications and are associated with significant (10–15%) mortality (Bharti et al., 2003; McBride et al., 2005). In addition, leptospirosis causes severe pulmonary haemorrhage syndrome (SPHS), for which case fatality is >50% (Segura et al., 2005; Gouveia et al., 2008). Leptospirosis is considered to be an emerging infectious disease in endemic regions of Asia (Karande et al., 2003, 2005; LaRocque et al., 2005; Yanagihara et al., 2007; Peacock and Newton, 2008) and Latin America (Ko et al., 1999; Sarkar et al., 2002; Romero et al., 2003; Johnson et al., 2004) and is a major public health concern in poverty stricken regions of the world (McBride et al., 2005; Ganoza et al., 2006; Riley et al., 2007).

The *Leptospira* genus is sub-classified into 18 genomospecies that includes both saprophytic and pathogenic species (Levett, 2001; Levett et al., 2006; Matthias et al., 2008). Classification based on serologic methods has identified ~300 serovars, of which more than 200 are considered to be pathogenic (Faine et al., 1999; Levett, 2001; Bharti et al., 2003). The availability of genomic sequence data from five *Leptospira* strains, *L. interrogans* serovars Lai (Ren et al., 2003) and Copenhageni (Nascimento et al., 2004), *L. borgpetersenii* serovar Hardjo strains L550 and JB197 (Bulach et al., 2006), and the saprophyte *L. biflexa* serovar Patoc I (Picardeau et al., 2008), is driving the discovery of new diagnostic tools and vaccines for leptospirosis. Considerable effort has been expended towards identifying conserved surface-exposed antigenic determinants that could improve diagnosis and provide heterologous protection via subunit or DNA vaccines.

A number of leptospiral outer membrane proteins (OMPs) have been characterized (Cullen et al., 2005), including OmpL1 (Haake et al., 1993), LipL41 (Shang et al., 1996), LipL36 (Haake et al., 1998), the major outer membrane protein, LipL32 (Haake et al., 2000), LipL21 (Cullen et al., 2003), LipL46 (Matsunaga et al., 2006), LenA (Verma et al., 2006), and the OmpA-like proteins Loa22 (Koizumi and Watanabe, 2003) and Omp52 (Hsieh et al., 2005). However, their performance in diagnostic assays for acute leptospirosis or as vaccine candidates has been problematic (Haake et al., 1999; Branger et al., 2001; Flannery et al., 2001; Guerreiro et al., 2001). LigA and LigB, belonging to a family of leptospiral immunoglobulin-like (Lig) proteins, appear to be promising antigens (Palaniappan et al., 2002; Matsunaga et al., 2003). The gene encoding a third Lig protein, *ligC*, was identified as a pseudogene in *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa (Matsunaga et al., 2003), but was found to be intact in *L. interrogans* serovar Lai (Ren et al., 2003). The Lig proteins contain a series of bacterial immunoglobulin-like (Big) repeat domains that were originally identified in virulence determinants from *Escherichia coli* and *Yersinia pseudotuberculosis* (Hamburger et al., 1999; Luo et al., 2000).

The *lig* genes are of great interest because emerging serologic, vaccine, and pathogenesis studies indicate that Lig proteins are key virulence determinants involved in host—pathogen interactions. Lig proteins mediate interaction with multiple host extracellular matrix proteins, including fibronectin, fibrinogen, collagen, and laminin (Choy et al., 2007). Several studies have provided evidence that the Lig proteins are protective immunogens in animal models of leptospirosis (Koizumi and Watanabe, 2004; Palaniappan et al., 2006; Silva et al., 2007). In addition, we recently demonstrated that a recombinant polypeptide containing Big domains 2–6 from LigB was able to protect hamsters against homologous challenge by *L. interrogans* serovar Copenhageni (unpublished data). Virulent forms of *L. interrogans* and *L. kirschneri* strains express higher levels of Lig proteins than culture-attenuated forms (Matsunaga et al., 2003). Lig expression is strongly induced by shifting the osmolarity from low levels used in EMJH culture medium to osmolarity levels found in host tissues (Matsunaga et al., 2005). Up-regulation during early host infection is consistent with the strong serologic response to Lig proteins observed during acute leptospirosis (Croda et al., 2007).

Considering the large number of pathogenic *Leptospira* serovars and the broad distribution of leptospiral host reservoirs, the potential effect of selective pressure on the genetic diversity of the Lig proteins was unclear. Given the potential of the Lig proteins as diagnostic antigens and vaccine candidates, we examined their sequence diversity in the serovars most often associated with leptospirosis.

2. Materials and methods

2.1. *Leptospira* strains and culture conditions

Virulent leptospiral strains (Table 1) were obtained from culture collections maintained by the authors. The isolation conditions of a number of the strains used in this study were previously described (Ko et al., 1999; Haake et al., 2002; Silva et al., 2008). The identity of each of the strains used in this study was confirmed by 16S rRNA gene sequencing (Hookey et al., 1993) and serogrouping based on the microscopic agglutination test (MAT) (Cole et al., 1973). Strains were cultured in liquid Ellinghausen—McCullough—Johnson—Harris modified tween 80-bovine albumin medium (Ellinghausen and McCullough, 1965; Johnson and Harris, 1967) at 30 °C. Virulence of each *Leptospira* strain was evaluated in the hamster model of lethal leptospirosis as previously described (Silva et al., 2008).

2.2. Genomic DNA extraction

Leptospiral genomic DNA was extracted from 7-day-old cultures using the GFX Genomic Blood DNA Purification Kit (GE Healthcare) according to the instructions provided for Gram-negative bacteria. The concentration was determined by absorbance at 260 nm and the quality of the genomic DNA was confirmed by agarose gel electrophoresis.

2.3. PCR detection of *lig* genes

A PCR-based screening method for the detection of the *lig* genes was developed. Degenerate primers (Table S4) were designed to amplify *ligA*, *ligB* and *ligC* based on sequence information deposited in GenBank. Additional primers were designed to amplify *ligB* from *L. weilii* strains due to the heterogeneity of *ligB* from this species compared to the other leptospiral species. The DNA sequence of each *lig* gene was aligned using AlignX (Vector NTI, Invitrogen) and, based on homology between the sequences, suitable regions were identified and primers, degenerate where appropriate, were designed. PCR (30 cycles) was carried out using recombinant Taq DNA polymerase (Invitrogen) following the protocol provided. PCR products were analysed by agarose gel electrophoresis. The expected product size for *ligA* was estimated to be 211 bp, 536 bp for *ligB* (1076 bp for *ligB* from *L. noguchii* and 1625 bp for

ligB from *L. weillii*) and 248 bp for *ligC*. Quality control PCR amplification of the 16S rDNA and *lipL32* genes was used to verify the quality of the genomic DNA.

2.4. Southern blotting

Southern blotting was carried out as described previously (Sambrook and Russell, 2001). Briefly, 3 µg of genomic DNA was digested with 20 units of BamHI (Invitrogen) and separated by agarose gel electrophoresis. DNA was transferred from the gel to a positively charged Hybond-N nylon membrane (GE Healthcare) with a vacuum blotter (Bio-Rad). Probes to each of the *lig* genes were based on pooled PCR products amplified using the primers described in Table S4 and labelled using the ECL Direct Nucleic Acid Labelling and Detection System (GE Healthcare) as described in the protocol provided. Prehybridization was carried out at 42 °C for 1 h in hybridization buffer supplemented with 0.5 M NaCl and 5% blocking agent. Hybridization was carried out overnight at 42 °C in roller bottles. Following hybridization, the membrane was washed twice for 10 min at 55 °C in wash solution (0.4% SDS, 0.5× SSC). Finally, the membrane was washed twice in 2× SSC, 5 min per wash at room temperature. After incubation with ECL detection reagents, hybridization products were detected by exposure of the membrane to Hyperfilm ECL X-ray film (GE Healthcare).

2.5. DNA sequencing

Dependant on their presence, full-length *ligA* (~3.7 kb), *ligB* (~5.7 kb) and *ligC* (~5.9 kb) genes from *L. interrogans* serovars Pomona and Canicola, *L. noguchii* serogroup Bataviae and *L. weillii* serogroup Hebdomadis were amplified using Elongase Mix (Invitrogen), which contains a proof-reading polymerase, and subsequently cloned using the TOPO-TA cloning kit (Invitrogen). Each gene sequence was determined by direct sequencing of PCR products amplified using the cloned genes and internal primers (Table S4). Ambiguous bases or those different to previously published sequences were resolved by direct sequencing of PCR products amplified from genomic DNA. Each base was sequenced on both strands a minimum of two times resulting in each base being sequenced a minimum of four times. Previously unpublished sequences were submitted to GenBank and assigned accession numbers EU700267 to EU700275.

2.6. DNA and protein sequence analysis

Raw DNA sequences were analysed for the quality of base calling using Phred (Ewing et al., 1998), the 5' and 3' ends were trimmed accordingly using FinchTV (Geospiza Inc.) and each contig was assembled using Contig Express (Vector NTI, Invitrogen) at the default settings. Consensus DNA sequences were exported and coding sequences (CDS) identified using Vector NTI. CDS were aligned using AlignX (Vector NTI, Invitrogen), based on the ClustalW algorithm (Thompson et al., 1994), at the default settings. Protein sequences were generated by translation of the CDS and were aligned using AlignX at the default settings. The number of observed synonymous (S_d), and nonsynonymous (N_d) substitutions in the *lig* genes were calculated using Syn-SCAN (Gonzales et al., 2002). The evolutionary history relating the *ligA*, *ligB*, and *ligC* genes from a total of 14 strains was examined using a statistical phylogenetic approach. Alignments of the genes were based on their amino acid translations to maintain reading frames and then reverse translated back to their nucleotide sequences using the Geneious Pro 3.0.1 software package (Drummond et al., 2007). Possible intragenic recombination was examined using the DualBrothers plugin for Geneious. Default settings were used with the following exceptions: chain length 220,000, subsampling frequency 100 and burn-in length 10,000. DualBrothers is a recombination detection algorithm based on a phylogenetic dual multiple change-point model (MCP) (Suchard et al., 2002, 2003; Minin et al., 2005). The MCP model allows for changes in evolutionary relationships and rates across sites in a multiple sequence alignment by assuming that the sites separate into an unknown

number of contiguous segments, each with possibly different topologies or mutation processes. Differing evolutionary topologies on either side of a break-point suggests recombination (Li et al., 1988). The DualBrothers implementation takes a Bayesian approach that employs Markov chain Monte Carlo to simulate from the posterior distribution of model parameters. One strength of the Bayesian model is that it can measure uncertainty in break-point locations, determine the most likely parental sequences of the putative recombinant and assess the statistical significance of recombination simultaneously; this simultaneous approach avoids the pitfall inherent in sequential testing for recombination found in many recombination detection programs (Suchard et al., 2002). For each putative recombinant, break-points were inferred and parental representatives and the significance of recombination using the approach of Suchard et al. (2003) were assessed.

Statistical analysis was carried out using Graphpad Prism 5.01 (GraphPad Software, San Diego, CA, USA). The Mann—Whitney test was used to calculate differences between populations and two-tailed P values <0.05 were considered significant.

3. Results

3.1. The *lig* genes in pathogenic *Leptospira* spp.

A PCR screening assay demonstrated that *ligB* was present in all 10 strains studied, *ligA* in five strains and *ligC* in seven strains (Table 1). Using the *lig* sequences deposited in GenBank, primers were designed to screen for the presence of *lig* genes in the leptospiral species most commonly associated with human leptospirosis (Table 1). *L. interrogans* serovars Copenhageni Fiocruz L1-130, Pomona Kennewicki, Canicola Kito and *L. kirschneri* serovar Grippotyphosa RM52 were found to contain three *lig* genes (Fig. 1). *L. interrogans* serovar Lai 56601 and *L. weilii* serogroup Hebdomadis Eco-Challenge were found to contain *ligB* and *ligC* while *L. noguchii* serogroup Bataviae Cascata and *L. borgpetersenii* serovar Hardjo strains JB197 and L550 contained only *ligB* (Fig. 1). Of note, in place of the *ligA* gene in *L. interrogans* serovar Lai and the two *L. borgpetersenii* serovar Hardjo genomes there were transposases (Fig. 1). In serovar Lai, upstream of *ligB* and where *ligA* would normally be located, two transposase genes were identified, Tn8 and an integrase, and a *ccrB* ortholog (encodes camphor resistance in bacteria). *L. borgpetersenii* serovar Hardjo strain JB197 contained a transposase, IS1533, and the C-terminal of the *ccrB* ortholog. *L. borgpetersenii* serovar Hardjo L550 contained an IS1533 pseudogene and the C-terminal region of the *ccrB* gene. The JB197 strain has a genome layout similar to that of serovar Lai with a genome inversion (Ren et al., 2003), while the genome organization of strain L550 more closely resembles that of the serovars Copenhageni, Pomona and Canicola.

To confirm the integrity of the *lig* genes in pathogenic *Leptospira* spp. they were cloned and sequenced from *L. interrogans* serovar Canicola, serovar Pomona, *L. noguchii* serogroup Bataviae and *L. weilii* serogroup Hebdomadis. The presence or absence of the *lig* genes was also confirmed by Southern blotting (data not shown). The *ligA* and *ligB* genes in serovars Pomona and Canicola were of the expected sizes and genome locations, with *ligA* situated upstream of *ligB* in both strains. We previously reported that the *ligC* gene was a pseudogene in *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa (Matsunaga et al., 2003). In this study we found *ligC* to be present in *L. interrogans* serovars Pomona and Canicola, and *L. weilii* serogroup Hebdomadis, but absent in *L. noguchii* serogroup Bataviae.

3.2. Lig sequence conservation

LigB was the only *lig* gene ortholog found in all 10 pathogenic strains of *Leptospira* spp. studied (Table 1) and was compared by estimating pair-wise sequence distances. LigB was found to be significantly more variable than *ligA* ($P < 0.05$) and *ligC* ($P < 0.005$), ranging from 67.9 to

99.9% (mean $80.2 \pm 13.1\%$) and 62.6 to 99.9% (mean $77.3 \pm 15.3\%$) for pair-wise DNA and amino acid sequence identity, respectively (Table S1). The base substitutions in *ligB* were found throughout the gene, although there was increased variability within a region of the carboxyterminal domain (amino acids 1518–1669). When phylogenetic groupings were analysed, *L. interrogans* serovars Copenhageni, Canicola, Lai and Pomona had a 95.4 and 94.4% mean pair-wise DNA and amino acid sequence identity, respectively. The mean number of observed synonymous substitutions (S_d) was 77.8, the mean number of observed nonsynonymous substitutions (N_d) was 65.5 and the ratio of synonymous to nonsynonymous substitutions (d_S/d_N) was 4.01. Inclusion of *L. kirschneri* serovar Grippotyphosa in this grouping reduced, but not significantly, the mean identity observed to 91.0 and 90.6% for DNA and amino acid sequences, respectively ($d_S/d_N = 4.88$). The two *L. borgpetersenii* serovar Hardjo strains and the *L. weilii* Eco-Challenge strain had a 90.9 and 90.2% mean DNA and amino acid sequence identity, respectively, and the majority of the base substitutions were synonymous (mean $S_d = 182.7$, mean $N_d = 125.7$, mean $d_S/d_N = 4.24$). LigB from the *L. noguchii* strain did not fall within either of these two groupings and inclusion in either group increased the levels of sequence variability by 20–30% (Table S1).

LigA from five strains (Table 1) was highly conserved, ranging from 85.4 to 99.8% (mean $91.8 \pm 5.9\%$) and 80.5 to 99.8% (mean $88.9 \pm 6.8\%$) DNA and amino acid sequence identity, respectively (Table S2). The carboxyterminal region of LigA, comprising Big domain repeats 11–13 (see Fig. 2a), exhibited the highest level of variability, with 78.1 and 71.2% mean pair-wise DNA and amino acid identity, respectively. Although LigA sequence conservation was high, the majority of the single-nucleotide polymorphisms resulted in nonsynonymous amino acid substitutions (mean $S_d = 119.7$, mean $N_d = 174.9$, mean $d_S/d_N = 2.89$). Of note, the serovar Pomona LigA amino acid sequence previously deposited in GenBank ((Palaniappan et al., 2002), accession number AAN52495), contained a region (amino acids 740–774) of very low identity (<10%) compared to the other LigA sequences, specifically, point-insertions at nucleotide positions 2218, 2315 and 2322. We resequenced the *ligA* gene from the serovar Pomona Kennewicki PO-06-047 strain, accession number EU700270, and did not find the point-insertions.

The *ligC* gene was present in 7 of the 10 strains evaluated in this study. Furthermore, it was the most conserved Lig protein. LigC DNA and amino acid sequence identity ranged from 77.0 to 100% (mean $91.2 \pm 8.9\%$) and 83.6 to 99.9% (mean $94.1 \pm 6.2\%$), respectively (Table S3). Similar to *ligB*, the base substitutions are spread throughout the gene. The majority of the base substitutions were nonsynonymous (mean $S_d = 96.1$, mean $N_d = 342.9$, mean $d_S/d_N = 1.03$). Previously we found that the *ligC* genes from *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa appear to be pseudogenes. Serovar Copenhageni contained a point mutation while serovar Grippotyphosa contained a frame-shift mutation, both of which resulted in stop codons (Matsunaga et al., 2003). The major difference between the various *ligC* sequences was a gap towards the middle of *ligC* (nucleotides 2839–3104) from the Eco-Challenge strain, as there was one less Big repeat domain in the *L. weilii ligC*.

3.3. Phylogenetic analysis of Lig domains

Translation-based alignment of nucleotide sequences of the immunoglobulin-like domains of LigA, LigB, and LigC of *L. interrogans* serovar Copenhageni revealed clustering of LigB domains 5–10 with LigA domains 8–13 (Fig. 2b). The same relatedness pattern was found for the Big domains of LigA, B, and C of *L. kirschneri* serovar Grippotyphosa (data not shown). This result suggests that the *ligA* gene was derived from the *ligB* gene in a partial gene duplication event (Fig. 2a). The *ligA* genes were found only in *L. interrogans* and *L. kirschneri* strains, indicating that the partial gene duplication event occurred in a progenitor of the *L. interrogans*—*L. kirschneri* branch of the leptospiral evolutionary tree. LigC domains

were more distantly related than the LigA and LigB domains, consistent with the more ancient origin of *ligC* (see below).

3.4. Phylogenetic analysis of *ligB*

The relatedness of the 10 full-length *ligB* sequences is presented in Fig. 3a. The *ligB* sequences of *L. borgpetersenii* and *L. weilii* cluster together in one region of the tree while the *L. interrogans* and *L. kirschneri* sequences cluster together in a separate region of the tree, with the *ligB* sequence of *L. noguchii* occupying an intermediate position. This relatedness pattern is consistent with a phylogenetic tree based on 16S rDNA sequences (Haake et al., 2004). Analysis using the multiple change-point (MCP) model revealed that all *ligB* genes except two were phylogenetically clonal (no evidence of rearrangements). The *ligB* sequence of *L. interrogans* serovar Copenhageni was found to have two *L. kirschneri*-like regions: a 153 nucleotide region in Big domain 11, and a 500 nucleotide region in the carboxyterminal domain (Fig. 3b). The *ligB* sequence of *L. kirschneri* serovar Grippotyphosa was found to have a large 1300 nucleotide rearrangement containing an *L. interrogans* serovar Lai-like sequence in the region coding for the carboxyterminal domain (Fig. 3c).

3.5. Phylogenetic analysis of *ligA*

There were two different phylogenetic trees for the 5' (nucleotides 1–2820) and 3' regions (nucleotides 2821–3675) of the *ligA* gene sequences. The phylogenetic tree for the 5' region (Fig. 4a) showed low levels of sequence non-identity (ranging from 1 to 9%) for the five strains. In contrast, the tree for the 3' region (Fig. 4b) revealed that while the *ligA* sequences of the Canicola and Pomona strains are 100% identical, their *ligA* sequences were 33% non-identical to the Grippotyphosa and Copenhageni *ligA* sequences, a 4–10-fold increase in sequence non-identity. One possible interpretation of this difference in sequence variation for the two different regions of the *ligA* gene is the recent acquisition the 3' *ligA* region, encoding the repeats 11–13 (Fig. 4c).

3.6. Phylogenetic analysis of *ligC*

In contrast to the distribution of the *ligA* genes, the *ligC* genes were found not only in *L. interrogans* and *L. kirschneri*, but also in *L. weilii*. The sequence identity between the *ligB* and *ligC* sequences of *L. interrogans* and *L. weilii* was similar (69 and 78%, respectively), suggesting that the *ligB* and *ligC* genes coevolved with leptospiral evolution, rather than the *L. weilii* *ligC* gene representing a more recent horizontal acquisition. MCP analysis of the seven *ligC* genes revealed a single tree structure consistent with leptospiral evolution throughout, except for an *L. interrogans*-like 640 nucleotide insertion (4660–5300) into the carboxyterminal domain coding region of the *ligC* gene of *L. weilii* serovar Eco-Challenge (Fig. 5). In addition, the *ligC* gene of *L. weilii* serovar Eco-Challenge was found to lack one Big repeat domain (domain 11).

4. Discussion

Genome sequencing studies have demonstrated that pathogenic *Leptospira* spp. contain *ligB* together with up to two *lig* paralogs, *ligA* and *ligC* (Ren et al., 2003; Nascimento et al., 2004; Bulach et al., 2006), while they are absent from the non-pathogenic saprophyte, *L. biflexa* (Matsunaga et al., 2003; Picardeau et al., 2008). However, little was known about the distribution of the *lig* genes among pathogenic leptospiral strains or their interrelationships. Our study demonstrates that an intact *ligB* gene is found in all leptospiral pathogens studied to date, suggesting an important, or perhaps essential, role in virulence. The *ligA* gene appears to have been derived from *ligB* by a two-step partial gene duplication process. LigC is structurally similar to LigB and *ligC* gene variation also follows phylogenetic patterns, suggesting an early gene duplication event. However, the role of LigC in virulence is less clear, as some strains

have lost *ligC*, while in others, such as *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa, *ligC* is a pseudogene. Of note, we were unable to demonstrate any association between the number of *lig* genes, their diversity, or the hosts from which they were isolated and the degree of virulence in the hamster model (data not shown). There is strong evidence that LigA is expressed during infection (Palaniappan et al., 2002; Matsunaga et al., 2003; Koizumi and Watanabe, 2004; Silva et al., 2007; Srimanote et al., 2008) and yet several virulent pathogenic *Leptospira* strains do not contain *ligA* (Table 1). A possible explanation is that *ligA* and *ligB* are involved in virulence but that both copies are not required, an example of gene redundancy. Indeed, a recent report showed that a *ligB* knockout in *L. interrogans* strain Fiocruz L1-130 did not alter the virulence of the *ligB*⁻ strain (Croda et al., 2008). As LigA was expressed in the *ligB*⁻ strain it would appear that LigA can replace LigB during infection, although the role of the Lig proteins in virulence is not yet clear.

Phylogenetic analysis based on LigB, which is conserved in all strains, sorted the strains into three distinct groups: (i) *L. kirschneri* and *L. interrogans*; (ii) *L. borgpetersenii* and *L. weilii*; and (iii) *L. noguchii* (Fig. 1). Amino acid sequence alignment of LigB reveals that overall the carboxyterminal domain is the most conserved region (>60% mean identity), although there are short, highly variable regions within this domain. However, when aligned by phylogenetic group the aminoterminal region (Big domains 1–12) exhibited a similar level of identity when compared with the carboxyterminal domain (>90% mean identity). LigA, although only present in *L. interrogans* and *L. kirschneri* strains, demonstrated a high level of conservation of Big domains 1–10 (89% mean identity), while the full-length LigA had a mean identity of 80%. LigC was the most conserved of the Lig proteins (90% mean identity) however, it appears to be a pseudogene in *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa (Matsunaga et al., 2003) and is absent in *L. borgpetersenii* and *L. noguchii* strains (this work). The most widely used method for identifying non-functional genes is the d_S/d_N test, which compares the rate of synonymous to nonsynonymous mutations (Nei and Kumar, 2000). Pseudogenes have no functional restrictions as they are not expressed and are therefore expected to have a d_S/d_N ratio that does not differ significantly from one. Analysis of *ligC* revealed a mean d_S/d_N ratio of 1.03 suggesting that *ligC* may have lost or is in the process of losing its role in virulence and strains that contain this gene may be subject to a genome reduction event in the future.

Evidence is presented for horizontal recombination events affecting all three *lig* genes. The results indicate that several types of evolutionary mechanisms have been acting on the *lig* genes, including genetic drift, gene duplication, and horizontal gene transfer. We find that genetic drift accounts for most of *ligB* genetic diversity, suggesting that *ligB* was acquired early during the evolution of leptospires from free-living saprophytes to colonizers of host tissues. Although the level of DNA and amino acid sequence diversity for the LigB was similar to what had been observed previously for the gene encoding the porin, OmpL1, it would have been difficult to anticipate what level of diversity to expect without a sequence analysis study of this type. Our previous study found surprisingly large differences in the rates of sequence variation among genes encoding surface-exposed leptospiral outer membrane proteins (Haake et al., 2004). Sequences of *lipL32* genes encoding the major outer membrane protein were highly conserved (3.1% DNA and 0.9% amino acid sequence non-identity). Strong *lipL32* sequence conservation was not anticipated because LipL32 is highly immunogenic; >95% of patients with leptospirosis have an antibody response to LipL32 (Flannery et al., 2001). In contrast to *lipL32*, genes encoding OmpL1 exhibited significantly higher rates of sequence variation (14.9% DNA and 9.6% amino acid sequence non-identity). Rates of sequence variation and recombination for the *lig* genes were comparable to those observed for the *ompL1* genes. However, there were some notable differences compared to the earlier study. For example, DNA sequence variability was higher than amino acid sequence variability for *lipL32*, *lipL41*, and *ompL1*. In contrast, the rates of DNA and amino acid sequence variation for the

lig genes were comparable, indicating that *lig* genes had a higher overall rate of nonsynonymous sequence changes.

Our analysis revealed evidence for a second mechanism of sequence diversity for all three *lig* genes: recombination events and horizontal DNA transfer between related bacterial species. Two *ligB* genes and one *ligC* gene were found to be mosaics. The *ligB* gene of *L. interrogans* serovar Copenhageni was found to contain two *L. kirschneri*-like insertions: one insertion in Big domain 11 and a second insertion in the carboxyterminal domain (relative probability of recombination event >1000:1). The *ligB* gene of *L. kirschneri* strain RM52 and the *ligC* gene of *L. weilii* strain Eco-Challenge included *L. interrogans*-like insertions in the regions encoding their carboxyterminal domains (relative probability of recombination event >1000:1). We previously reported mosaicism for 20% of genes encoding the outer membrane protein, OmpL1. Mosaicism does not affect all leptospiral outer membrane proteins, as no evidence of recombination events was found for the gene encoding the major outer membrane lipoprotein, LipL32, and only one recombination event affecting a second outer membrane lipoprotein, LipL41, from 38 different strains representing six pathogenic *Leptospira* spp. (Haake et al., 2004). In the case of *ligA*, three of five strains appear to have acquired the same DNA encoding their last three *ligA* Big domains. Phylogenetic comparison of the transferred DNA encoding the exogenous *ligA* Big domains 11–13 with all known *lig* Big domains shows that they are most closely related to their *ligA* orthologs in *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa (data not shown). However, insufficient *ligA* sequence data is available to determine the phylogenetic origin of this exogenous *ligA* DNA.

Of note, we found that the first six Big domains from *ligB* and the first six Big domains from *ligA* were essentially identical ($98.5 \pm 0.8\%$ mean identity). This is an important observation for the future development of diagnostic reagents. Recombinant LigB polypeptides containing Big domains 2–6 from *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa were evaluated as antigens for the diagnosis of leptospirosis. Sensitivity and specificity were reported to be >90 and >97%, respectively, during the acute-phase of leptospirosis (Croda et al., 2007). Furthermore, there was no evidence of major genetic rearrangements in this region. The lack of genetic drift within the identical regions of *ligB* and *ligA* is evidence of selective pressure, intragenic recombination or gene conversion. The remaining Big domains (7–12 in *ligB* and 7–13 in *ligA*) were considerably more variable ($34.2 \pm 1.6\%$ mean identity). This is an important observation, as our previous findings indicate that these Big domains are involved in binding extracellular matrix proteins and fibrinogen (Choy et al., 2007). Together these findings support a role for the Lig proteins during the transmission of leptospirosis. The carboxyterminal Big domains may have evolved to recognise specific host extracellular matrix proteins. This region was subject to horizontal recombination between *Leptospira* spp. suggesting that increased variability in the carboxyterminal Big domains of *ligA* and *ligB* may have contributed towards the adaptation to novel hosts and potentially accounting for the extensive serovar-host specificity that typifies leptospirosis.

Our study had several potential limitations. The status of the *ligB* gene of *L. interrogans* serovar Manilae remains to be clarified since the *ligB* gene sequence has been reported to lack the carboxyterminal domain due to a point mutation that created a stop codon (Koizumi and Watanabe, 2004). Due to the difficulties inherent in sequencing the *lig* genes it is possible that the *ligB* sequence for serovar Manilae contains an erroneous stop codon in the carboxyterminal domain. Bacterial genes containing indels resulting in premature stop codons are defined as pseudogenes and are not functional (Ochman and Davalos, 2006). Yet, experimental data suggested that LigB was expressed in *L. interrogans* serovar Manilae. Koizumi and Watanabe (2004) showed that sera from leptospirosis patients specifically recognized recombinant LigB cloned from serovar Manilae. The partial *ligB* sequence (accession number AB098517) demonstrated >95% identity with the same region in the other *ligB* orthologs (data not shown).

However, we were unable to obtain the serovar Manilae strain to evaluate this possibility and therefore excluded the serovar Manilae *lig* gene sequences from further analyses. Nevertheless, sequence data from other serovars from the same Icterohaemorrhagiae serogroup (serovars Lai and Copenhageni) as well as other strains from the same species (serovars Canicola and Pomona) provided consistent findings for *ligB* (see Table 1). Sequence alignment suggests that the previously sequenced *ligA* from *L. interrogans* serovar Pomona strain Kennewicki (GenBank sequence AAN52495) contains a highly variable region (<10% amino acid identity) due to three indels that alter the coding sequence of this region. These indels are either representative of strain-to-strain variation within serovar Pomona Kennewicki or of PCR artefacts. The *ligA* sequence data in this study was derived from three independent PCR products sequenced in both directions for the region in question. Each base was therefore sequenced a minimum of six times to rule out the possibility of PCR artefacts and sequence errors. As the strains are probably different, our strain was isolated from an aborted swine foetus while the other was isolated from a case of equine recurrent uveitis (Palaniappan et al., 2002), more serovar Pomona strains would need to be sequenced to resolve this issue. A further limitation is that we were only able to include 10 *Leptospira* strains in our analysis however, all of these strains were considered virulent.

To date, Lig research has focused mainly on the Big domains, which are highly antigenic and are the focus of most of the anti-Lig antibody response during leptospirosis infection. LigA and LigB repeats have been shown to provide protective immunity in animal models of leptospirosis (Koizumi and Watanabe, 2004; Palaniappan et al., 2006; Silva et al., 2007). Lastly, Big domains 7–13 of LigA and domains 7–12 of LigB have been shown to function in binding to host extracellular matrix proteins (Choy et al., 2007). The results presented here serve to highlight the potential importance of the carboxyterminal domains of LigB and to prompt studies investigating its cellular location and function. Sequence analysis of *lig* genes from multiple strains is important in the ongoing efforts to develop new Lig-based serologic tests and vaccines. LigB has been found in all strains studied, making it an ideal candidate for vaccine and diagnostic applications. Furthermore, sequence conservation of LigB from *L. interrogans*, which contains the serovars most important to public health, was high, >96%. The sequence information provided here will improve the performance characteristics of the Lig proteins in diagnostic and subunit vaccine formulations. Knowledge of sequence variations among Lig serodiagnostic antigens should prompt cross-reaction studies using sera from different human and animal populations and possibly inclusion of homologous antigens from antigenically distinct strains. Future Lig vaccine studies should evaluate cross-protection using antigens and challenge strains with different *lig* sequences in order to assess the significance of sequence variation on immunoprotection.

Acknowledgments

This work was supported by Bio-Manguinhos, Oswaldo Cruz Foundation, Brazilian Ministry of Health (grant 09224-7), Brazilian National Research Council (grants 01.06.0298.00 3773/2005, 420067/2005, 554788/2006 and 473006/2006-5); Public Health Service grants AI-34431 (to D.A.H.) and R01 AI034431 (to A.I.K.) from the National Institute of Allergy and Infectious Diseases, the VA Medical Research Funds (to D.A.H.) and grant D43 TW00919 (to A.I.K.) from the Fogarty International Center, National Institutes of Health. M.A.S. is supported by an Alfred P. Sloane Research Fellowship and a John Simon Guggenheim Memorial Fellowship.

References

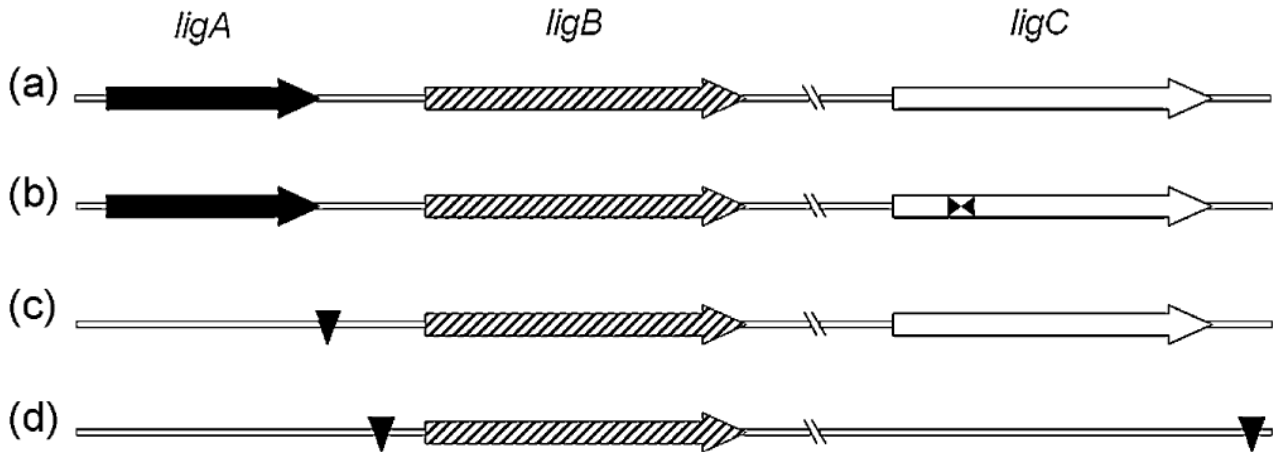
- Bharti AR, Nally JE, Ricaldi JN, Matthias MA, Diaz MM, Lovett MA, Levett PN, Gilman RH, Willig MR, Gotuzzo E, Vinetz JM. Leptospirosis: a zoonotic disease of global importance. *Lancet Infect. Dis* 2003;3:757–771. [PubMed: 14652202]
- Branger C, Sonrier C, Chatrenet B, Klonjowski B, Ruvoen-Clouet N, Aubert A, Andre-Fontaine G, Eloit M. Identification of the hemolysis-associated protein 1 as a cross-protective immunogen of

- Leptospira interrogans* by adenovirus-mediated vaccination. *Infect. Immun* 2001;69:6831–6838. [PubMed: 11598056]
- Bulach DM, Zuerner RL, Wilson P, Seemann T, McGrath A, Cullen PA, Davis J, Johnson M, Kuczek E, Alt DP, Peterson-Burch B, Coppel RL, Rood JI, Davies JK, Adler B. Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proc. Natl. Acad. Sci. U.S.A* 2006;103:14560–14565. [PubMed: 16973745]
- Choy HA, Kelley MM, Chen TL, Moller AK, Matsunaga J, Haake DA. Physiological osmotic induction of *Leptospira interrogans* adhesion: LigA and LigB bind extracellular matrix proteins and fibrinogen. *Infect. Immun* 2007;75:2441–2450. [PubMed: 17296754]
- Cole JR Jr, Sulzer CR, Pursell AR. Improved microtechnique for the leptospiral microscopic agglutination test. *Appl. Microbiol* 1973;25:976–980. [PubMed: 4736794]
- Croda J, Ramos JG, Matsunaga J, Queiroz A, Homma A, Riley LW, Haake DA, Reis MG, Ko AI. *Leptospira* immunoglobulin-like proteins as a serodiagnostic marker for acute leptospirosis. *J. Clin. Microbiol* 2007;45:1528–1534. [PubMed: 17360842]
- Croda J, Figueira CP, Wunder EA Jr, Santos CS, Reis MG, Ko AI, Picardeau M. Targeted mutagenesis in pathogenic leptospira: disruption of the LigB gene does not affect virulence in animal models of leptospirosis. *Infect. Immun* 2008;76:5826–5833. [PubMed: 18809657]
- Cullen PA, Haake DA, Bulach DM, Zuerner RL, Adler B. LipL21 is a novel surface-exposed lipoprotein of pathogenic *Leptospira* species. *Infect. Immun* 2003;71:2414–2421. [PubMed: 12704111]
- Cullen PA, Xu X, Matsunaga J, Sanchez Y, Ko AI, Haake DA, Adler B. Surfaceome of *Leptospira* spp. *Infect. Immun* 2005;73:4853–4863. [PubMed: 16040999]
- Drummond, AJ.; Ashton, B.; Cheung, M.; Heled, J.; Kearse, M.; Moir, R.; Stones-Havas, S.; Thierer, T.; Wilson, A. Geneious v3.5. 2007. Available from <http://www.geneious.com/>
- Ellinghausen HC Jr, McCullough WG. Nutrition of leptospira pomona and growth of 13 other serotypes: fractionation of oleic albumin complex and a medium of bovine albumin and polysorbate 80. *Am. J. Vet. Res* 1965;26:45–51. [PubMed: 14266934]
- Ewing B, Hillier L, Wendl MC, Green P. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. *Genome Res* 1998;8:175–185. [PubMed: 9521921]
- Faine, SB.; Adler, B.; Bolin, C.; Perolat, P. *Leptospira* and Leptospirosis. MediSci; Melbourne, Australia: 1999.
- Flannery B, Costa D, Carvalho FP, Guerreiro H, Matsunaga J, da Silva ED, Ferreira AGP, Riley LW, Reis MG, Haake DA, Ko AI. Evaluation of recombinant *Leptospira* antigen-based enzyme-linked immunosorbent assays for the serodiagnosis of leptospirosis. *J. Clin. Microbiol* 2001;39:3303–3310. [PubMed: 11526167]
- Ganoza CA, Matthias MA, Collins-Richards D, Brouwer KC, Cunningham CB, Segura ER, Gilman RH, Gotuzzo E, Vinetz JM. Determining risk for severe leptospirosis by molecular analysis of environmental surface waters for pathogenic *Leptospira*. *PLoS Med* 2006;3:e308. [PubMed: 16933963]
- Gonzales MJ, Dugan JM, Shafer RW. Synonymous-non-synonymous mutation rates between sequences containing ambiguous nucleotides (Syn-SCAN). *Bioinformatics* 2002;18:886–887. [PubMed: 12075026]
- Gouveia EL, Metcalfe J, de Carvalho AL, Aires TS, Villasboas-Bisneto JC, Queiroz A, Santos AC, Salgado K, Reis MG, Ko AI. Leptospirosis-associated severe pulmonary hemorrhagic syndrome, Salvador, Brazil. *Emerg. Infect. Dis* 2008;14:505–508. [PubMed: 18325275]
- Guerreiro H, Croda J, Flannery B, Mazel M, Matsunaga J, Galvao RM, Levett PN, Ko AI, Haake DA. Leptospiral proteins recognized during the humoral immune response to leptospirosis in humans. *Infect. Immun* 2001;69:4958–4968. [PubMed: 11447174]
- Haake DA, Champion CI, Martinich C, Shang ES, Blanco DR, Miller JN, Lovett MA. Molecular cloning and sequence analysis of the gene encoding *OmpL1*, a transmembrane outer membrane protein of pathogenic *Leptospira* spp. *J. Bacteriol* 1993;175:4225–4234. [PubMed: 8320237]
- Haake DA, Martinich C, Summers TA, Shang ES, Pruetz JD, McCoy AM, Mazel MK, Bolin CA. Characterization of leptospiral outer membrane lipoprotein LipL36: downregulation associated with late-log-phase growth and mammalian infection. *Infect. Immun* 1998;66:1579–1587. [PubMed: 9529084]

- Haake DA, Mazel MK, McCoy AM, Milward F, Chao G, Matsunaga J, Wagar EA. Leptospiral outer membrane proteins OmpL1 and LipL41 exhibit synergistic immunoprotection. *Infect. Immun* 1999;67:6572–6582. [PubMed: 10569777]
- Haake DA, Chao G, Zuerner RL, Barnett JK, Barnett D, Mazel M, Matsunaga J, Levett PN, Bolin CA. The leptospiral major outer membrane protein LipL32 is a lipoprotein expressed during mammalian infection. *Infect. Immun* 2000;68:2276–2285. [PubMed: 10722630]
- Haake DA, Dundoo M, Cader R, Kubak BM, Hartskeerl RA, Sejvar JJ, Ashford DA. Leptospirosis, water sports, and chemoprophylaxis. *Clin. Infect. Dis* 2002;34:e40–43. [PubMed: 11941571]
- Haake DA, Suchard MA, Kelley MM, Dundoo M, Alt DP, Zuerner RL. Molecular evolution and mosaicism of leptospiral outer membrane proteins involves horizontal DNA transfer. *J. Bacteriol* 2004;186:2818–2828. [PubMed: 15090524]
- Hamburger ZA, Brown MS, Isberg RR, Bjorkman PJ. Crystal structure of invasins: a bacterial integrin-binding protein. *Science* 1999;286:291–295. [PubMed: 10514372]
- Hookey JV, Bryden J, Gatehouse L. The use of 16S rDNA sequence analysis to investigate the phylogeny of Leptospiroaceae and related spirochaetes. *J. Gen. Microbiol* 1993;139:2585–2590. [PubMed: 7506293]
- Hsieh WJ, Chang YF, Chen CS, Pan MJ. Omp52 is a growth-phase-regulated outer membrane protein of *Leptospira santarosai* serovar Shermani. *FEMS Microbiol. Lett* 2005;243:339–345. [PubMed: 15686833]
- Johnson MA, Smith H, Joeph P, Gilman RH, Bautista CT, Campos KJ, Cespedes M, Klatsky P, Vidal C, Terry H, Calderon MM, Coral C, Cabrera L, Parmar PS, Vinetz JM. Environmental exposure and leptospirosis, Peru. *Emerg. Infect. Dis* 2004;10:1016–1022. [PubMed: 15207052]
- Johnson RC, Harris VG. Differentiation of pathogenic and saprophytic leptospires. I. Growth at low temperatures. *J. Bacteriol* 1967;94:27–31. [PubMed: 6027998]
- Karande S, Bhatt M, Kelkar A, Kulkarni M, De A, Varaiya A. An observational study to detect leptospirosis in Mumbai, India, 2000. *Arch. Dis. Child* 2003;88:1070–1075. [PubMed: 14670771]
- Karande S, Gandhi D, Kulkarni M, Bharadwaj R, Pol S, Thakare J, De A. Concurrent outbreak of leptospirosis and dengue in Mumbai, India, 2002. *J. Trop. Pediatr* 2005;51:174–181. [PubMed: 15831670]
- Ko AI, Reis M, Galvao, Dourado C.M. Ribeiro, Johnson WD Jr. Riley LW, Salvador Leptospirosis Study Group. Urban epidemic of severe leptospirosis in Brazil. *Lancet* 1999;354:820–825. [PubMed: 10485724]
- Koizumi N, Watanabe H. Molecular cloning and characterization of a novel leptospiral lipoprotein with OmpA domain. *FEMS Microbiol. Lett* 2003;226:215–219. [PubMed: 14553914]
- Koizumi N, Watanabe H. Leptospiral immunoglobulin-like proteins elicit protective immunity. *Vaccine* 2004;22:1545–1552. [PubMed: 15063580]
- LaRocque RC, Breiman RF, Ari MD, Morey RE, Janan FA, Hayes JM, Hossain MA, Brooks WA, Levett PN. Leptospirosis during dengue outbreak, Bangladesh. *Emerg. Infect. Dis* 2005;11:766–769. [PubMed: 15890136]
- Levett PN. Leptospirosis. *Clin. Microbiol. Rev* 2001;14:296–326. [PubMed: 11292640]
- Levett PN, Morey RE, Galloway RL, Steigerwalt AG. *Leptospira broomii* sp. nov., isolated from humans with leptospirosis. *Int. J. Syst. Evol. Microbiol* 2006;56:671–673. [PubMed: 16514048]
- Li WH, Tanimura M, Sharp PM. Rates and dates of divergence between AIDS virus nucleotide sequences. *Mol. Biol. Evol* 1988;5:313–330. [PubMed: 3405075]
- Luo Y, Frey EA, Pfuetschner RA, Creagh AL, Knoechel DG, Haynes CA, Finlay BB, Strynadka NC. Crystal structure of enteropathogenic *Escherichia coli* intimin-receptor complex. *Nature* 2000;405:1073–1077. [PubMed: 10890451]
- Matsunaga J, Barocchi MA, Croda J, Young TA, Sanchez Y, Siqueira I, Bolin CA, Reis MG, Riley LW, Haake DA, Ko AI. Pathogenic *Leptospira* species express surface-exposed proteins belonging to the bacterial immunoglobulin superfamily. *Mol. Microbiol* 2003;49:929–945. [PubMed: 12890019]
- Matsunaga J, Sanchez Y, Xu X, Haake DA. Osmolarity, a key environmental signal controlling expression of leptospiral proteins LigA and LigB and the extracellular release of LigA. *Infect. Immun* 2005;73:70–78. [PubMed: 15618142]

- Matsunaga J, Werneid K, Zuerner RL, Frank A, Haake DA. LipL46 is a novel surface-exposed lipoprotein expressed during leptospiral dissemination in the mammalian host. *Microbiology* 2006;152:3777–3786. [PubMed: 17159228]
- Matthias MA, Ricaldi JN, Cespedes M, Diaz MM, Galloway RL, Saito M, Steigerwalt AG, Patra KP, Ore CV, Gotuzzo E, Gilman RH, Levett PN, Vinetz JM. Human leptospirosis caused by a new, antigenically unique *Leptospira* associated with a *Rattus* species reservoir in the Peruvian Amazon. *PLoS Negl. Trop. Dis* 2008;2:e213. [PubMed: 18382606]
- McBride AJ, Athanazio DA, Reis MG, Ko AI. Leptospirosis. *Curr. Opin. Infect. Dis* 2005;18:376–386. [PubMed: 16148523]
- Minin VN, Dorman KS, Fang F, Suchard MA. Dual multiple change-point model leads to more accurate recombination detection. *Bioinformatics* 2005;21:3034–3042. [PubMed: 15914546]
- Nascimento AL, Ko AI, Martins EA, Monteiro-Vitorello CB, Ho PL, Haake DA, Verjovski-Almeida S, Hartskeerl RA, Marques MV, Oliveira MC, Menck CF, Leite LC, Carrer H, Coutinho LL, Degraive WM, Dellagostin OA, El-Dorry H, Ferro ES, Ferro MI, Furlan LR, Gamberini M, Giglioti EA, Goes-Neto A, Goldman GH, Goldman MH, Harakava R, Jeronimo SM, Junqueira-de-Azevedo IL, Kimura ET, Kuramae EE, Lemos EG, Lemos MV, Marino CL, Nunes LR, de ORC, Pereira GG, Reis MS, Schriefer A, Siqueira WJ, Sommer P, Tsai SM, Simpson AJ, Ferro JA, Camargo LE, Kitajima JP, Setubal JC, Van SMA. Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *J. Bacteriol* 2004;186:2164–2172. [PubMed: 15028702]
- Nei, M.; Kumar, S. *Molecular Evolution and Phylogenetics*. Oxford University Press US; New York: 2000.
- Ochman H, Davalos LM. The nature and dynamics of bacterial genomes. *Science* 2006;311:1730–1733. [PubMed: 16556833]
- Palaniappan RU, Chang YF, Jusuf SS, Artiushin S, Timoney JF, McDonough SP, Barr SC, Divers TJ, Simpson KW, McDonough PL, Mohammed HO. Cloning and molecular characterization of an immunogenic LigA protein of *Leptospira interrogans*. *Infect. Immun* 2002;70:5924–5930. [PubMed: 12379666]
- Palaniappan RU, McDonough SP, Divers TJ, Chen CS, Pan MJ, Matsumoto M, Chang YF. Immunoprotection of recombinant leptospiral immunoglobulin-like protein A against *Leptospira interrogans* serovar Pomona infection. *Infect. Immun* 2006;74:1745–1750. [PubMed: 16495547]
- Peacock SJ, Newton PN. Public health impact of establishing the cause of bacterial infections in rural Asia. *Trans. R. Soc. Trop. Med. Hyg* 2008;102:5–6. [PubMed: 17619030]
- Picardeau M, Bulach DM, Bouchier C, Zuerner RL, Zidane N, Wilson PJ, Creno S, Kuczek ES, Bommezzadri S, Davis JC, McGrath A, Johnson MJ, Boursaux-Eude C, Seemann T, Rouy Z, Coppel RL, Rood JI, Lajus A, Davies JK, Medigue C, Adler B. Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. *PLoS ONE* 2008;3:e1607. [PubMed: 18270594]
- Ren SX, Fu G, Jiang XG, Zeng R, Miao YG, Xu H, Zhang YX, Xiong H, Lu G, Lu LF, Jiang HQ, Jia J, Tu YF, Jiang JX, Gu WY, Zhang YQ, Cai Z, Sheng HH, Yin HF, Zhang Y, Zhu GF, Wan M, Huang HL, Qian Z, Wang SY, Ma W, Yao ZJ, Shen Y, Qiang BQ, Xia QC, Guo XK, Danchin A, Saint GI, Somerville RL, Wen YM, Shi MH, Chen Z, Xu JG, Zhao GP. Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature* 2003;422:888–893. [PubMed: 12712204]
- Riley LW, Ko AI, Unger A, Reis MG. Slum health: diseases of neglected populations. *BMC Int. Health Hum. Rights* 2007;7:2. [PubMed: 17343758]
- Romero EC, Bernardo CC, Yasuda PH. Human leptospirosis: a twenty-nine-year serological study in Sao Paulo, Brazil. *Rev. Inst. Med. Trop. Sao Paulo* 2003;45:245–248. [PubMed: 14743663]
- Sambrook, J.; Russell, DW. *Molecular Cloning: A Laboratory Manual*. Cold Spring Harbor Laboratory Press; Cold Spring Harbor, NY: 2001.
- Sarkar U, Nascimento SF, Barbosa R, Martins R, Nuevo H, Kalafanos I, Grunstein I, Flannery B, Dias J, Riley LW, Reis MG, Ko AI. Population-based case-control investigation of risk factors for leptospirosis during an urban epidemic. *Am. J. Trop. Med. Hyg* 2002;66:605–610. [PubMed: 12201599]

- Segura ER, Ganoza CA, Campos K, Ricaldi JN, Torres S, Silva H, Cespedes MJ, Matthias MA, Swancutt MA, Linan R, Lopez, Gotuzzo E, Guerra H, Gilman RH, Vinetz JM. Clinical spectrum of pulmonary involvement in leptospirosis in a region of endemicity, with quantification of leptospiral burden. *Clin. Infect. Dis* 2005;40:343–351. [PubMed: 15668855]
- Shang ES, Summers TA, Haake DA. Molecular cloning and sequence analysis of the gene encoding LipL41, a surface-exposed lipoprotein of pathogenic *Leptospira* species. *Infect. Immun* 1996;64:2322–2330. [PubMed: 8675344]
- Silva EF, Medeiros MA, McBride AJ, Matsunaga J, Esteves GS, Ramos JG, Santos CS, Croda J, Homma A, Dellagostin OA, Haake DA, Reis MG, Ko AI. The terminal portion of leptospiral immunoglobulin-like protein LigA confers protective immunity against lethal infection in the hamster model of leptospirosis. *Vaccine* 2007;25:6277–6286. [PubMed: 17629368]
- Silva EF, Santos CS, Athanzio DA, Seyffert N, Seixas FK, Cerqueira GM, Fagundes MQ, Brod CS, Reis MG, Dellagostin OA, Ko AI. Characterization of virulence of *Leptospira* isolates in a hamster model. *Vaccine* 2008;26:3892–3896. [PubMed: 18547690]
- Srimanote P, Wongdeethai N, Jieanampunkul P, Samonkiert S, Leepiyasakulchai C, Kalambaheti T, Prachayasittikul V. Recombinant LigA for leptospirosis diagnosis and ligA among the *Leptospira* spp. clinical isolates. *J. Microbiol. Methods* 2008;72:73–81. [PubMed: 18079011]
- Suchard MA, Weiss RE, Dorman KS, Sinsheimer JS. Oh brother, where art thou? A Bayes factor test for recombination with uncertain heritage. *Syst. Biol* 2002;51:715–728. [PubMed: 12396586]
- Suchard MA, Weiss RE, Dorman KS, Sinsheimer JS. Inferring spatial phylogenetic variation along nucleotide sequences: a multiple change-point model. *J. Am. Stat. Assoc* 2003;98:427–437.
- Thompson JD, Higgins DG, Gibson TJ. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 1994;22:4673–4680. [PubMed: 7984417]
- Verma A, Hellwage J, Artiushin S, Zipfel PF, Kraiczy P, Timoney JF, Stevenson B. LfhA, a novel factor H-binding protein of *Leptospira interrogans*. *Infect. Immun* 2006;74:2659–2666. [PubMed: 16622202]
- Yanagihara Y, Villanueva SY, Yoshida S, Okamoto Y, Masuzawa T. Current status of leptospirosis in Japan and Philippines. *Comp. Immunol. Microbiol. Infect. Dis* 2007;30:399–413. [PubMed: 17614131]

**Fig. 1.**

Genomic organization of the *lig* genes in *Leptospira* spp. The individual *lig* genes are represented by arrows, the presence of a stop codon is indicated by the ▶◀ symbol and transposase genes are designated by the ▼ symbol. The *ligA* (solid arrow) and *ligB* (hatched arrow) genes are located within 0.9–1.4 kb of each other in the same orientation within the genome in an operon-like arrangement. The *ligC* gene (open arrow) is distally located on the opposite side of the genome (~715 kb distant). Genomic organization in the *Leptospira* strains fell into one of four distinct groups: (a) full-length intact *lig* genes: *L. interrogans* serovar Canicola and serovar Pomona; (b) intact *ligA*, *ligB* and *ligC* pseudogene: *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa; (c) intact *ligB* and *ligC*: *L. interrogans* serovar Lai and *L. weilii* serogroup Hebdomadis; (d) intact *ligB* only: *L. borgpetersenii* serovar Hardjo and *L. noguchii* serogroup Bataviae strains.

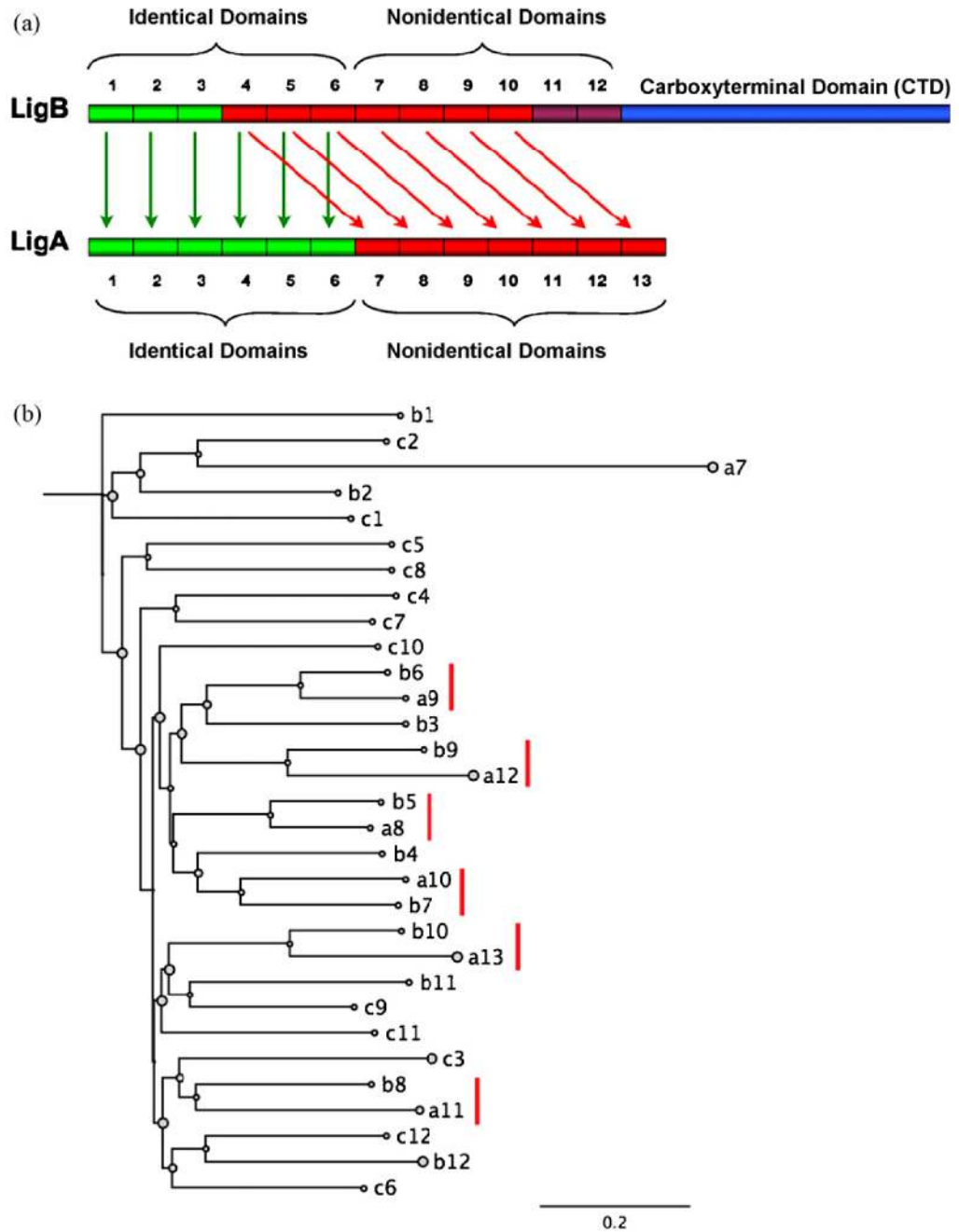


Fig. 2. (a) Proposed mechanism of *ligA* gene creation. A two-step partial gene duplication process is proposed: in the first step (green arrows), the first six Big domains of *ligA* were derived from the first six domains of *ligB*. In the second step (red arrows), Big domains 7–13 of *ligA* were derived from the domains 4–10 of *ligB*. Ordering of steps remains unknown. The individual Big domains comprising the *lig* genes are numbered, and the carboxyterminal domain of *ligB* is indicated. (b) Phylogenetic tree of the Lig domains. The tree is built from a translation-based alignment of nucleotide sequences of the immunoglobulin-like domains of LigA (a1–13), LigB (b1–12), and LigC (c1–12) of *L. interrogans* serovar Copenhageni. Vertical red lines indicate clustering of corresponding LigB domains 5–10 with LigA domains 8–13 in a “+3”

relationship. LigC domains are more irregularly related than the LigA and LigB domains (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

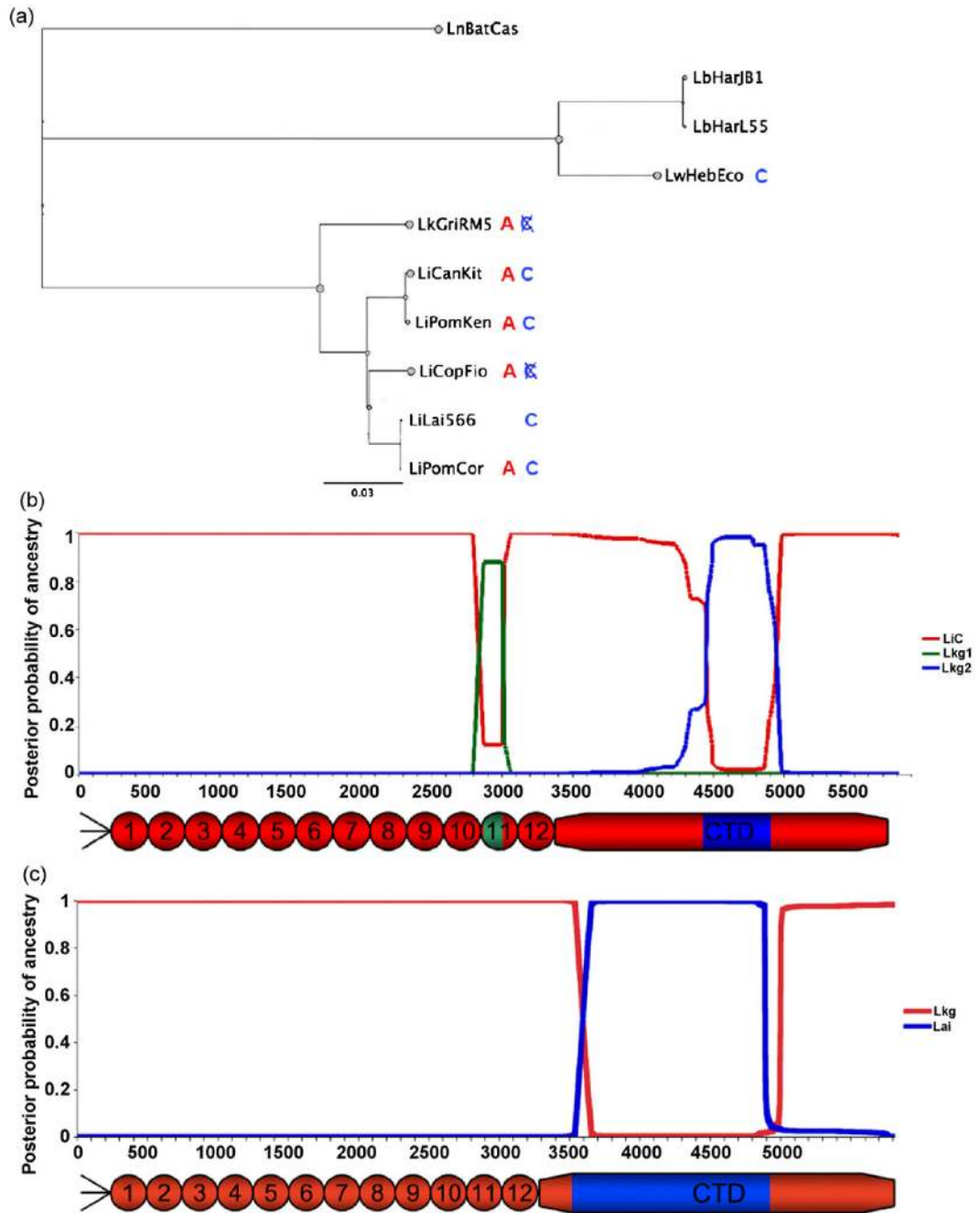


Fig. 3.
 (a) Phylogenetic tree of *ligB* sequences, the tree is built from the ten available full-length *ligB* sequences, and shows the same branching relationships as would be anticipated from the 16S-based phylogenetic tree for the five species represented. Strains with *ligA* (red) and *ligC* (blue) genes are shown. The *ligC* genes of *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa are pseudogenes. Mosaicism in the *ligB* genes, MCP analysis of the *ligB* genes of *L. interrogans* serovar Copenhageni and *L. kirschneri* serovar Grippotyphosa revealed two distinct mosaic patterns. (b) In the first pattern, the *ligB* gene of *L. interrogans* serovar Copenhageni contains two regions of *L. kirschneri*-like sequences: a 153 nucleotide region in Big domain 11 (green), and a 500 nucleotide region in the

carboxyterminal domain (blue). (c) In the second pattern, the *ligB* gene of *L. kirschneri* serovar Grippotyphosa contains a large 1300 nucleotide region (blue) containing an *L. interrogans* Lai-like sequence in the region coding for the carboxyterminal domain (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

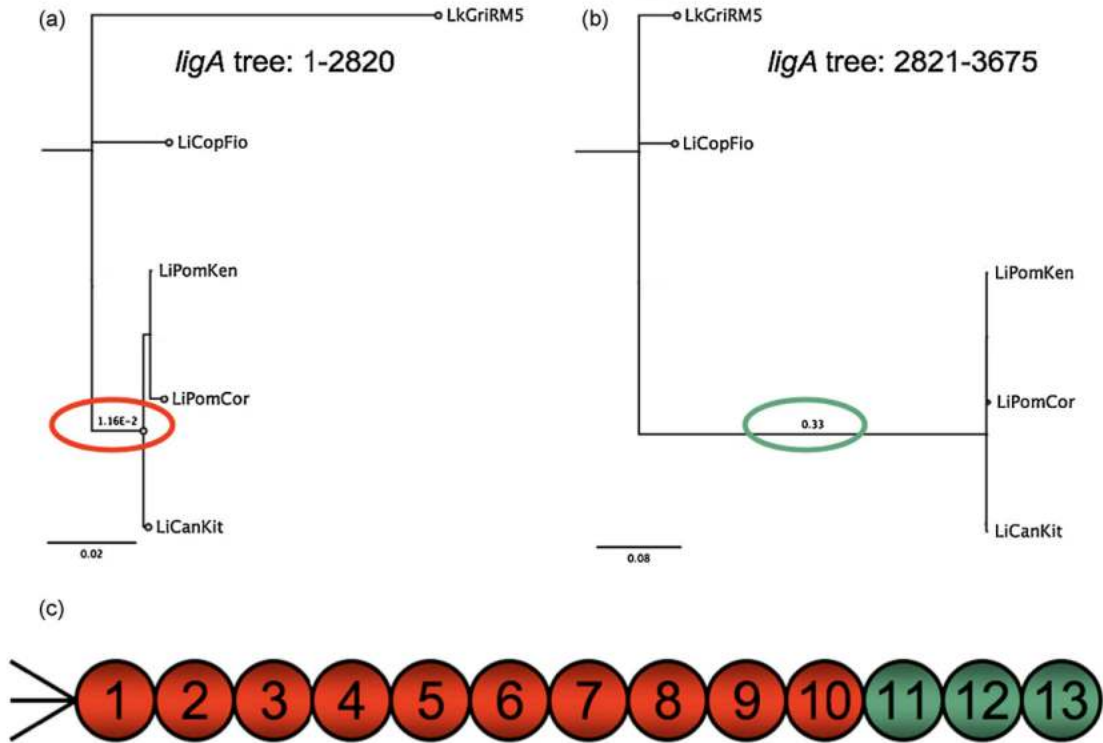


Fig. 4. Evidence for recombination in *ligA*. Comparison of five *ligA* genes revealed two different evolutionary rates for the 5' and 3' regions. (a) The phylogenetic tree for the 5' region showed low levels of sequence non-identity (ranging from 1 to 9%) for the five strains (key branch length circled in red). (b) In contrast, the tree for the 3' region revealed that while the *ligA* sequences of the Canicola and Pomona strains are 100% identical, their *ligA* sequences were 33% non-identical to the Grippytyphosa and Copenhageni *ligA* sequences (key branch length circled in green). (c) One interpretation of this result is that the last three Big domains of Canicola and Pomona strains were acquired by horizontal DNA transfer (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

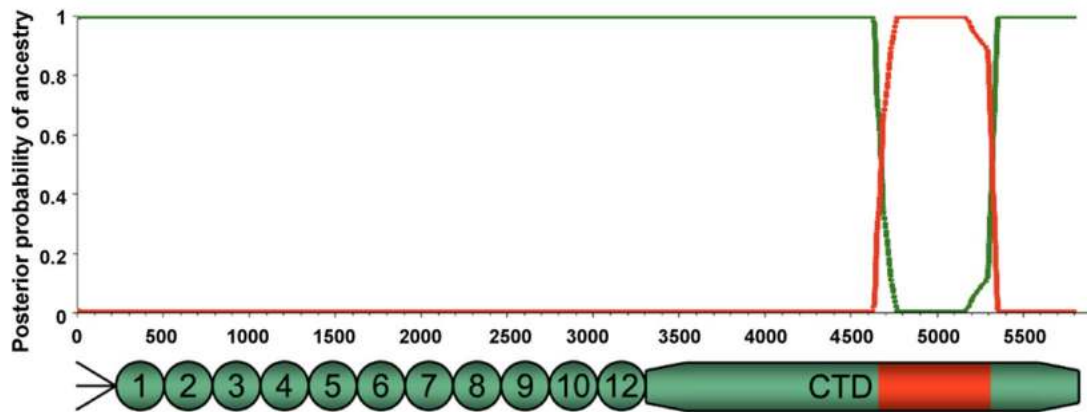


Fig. 5. Mosaicism in the *ligC* gene. MCP analysis of the *ligC* gene from the *L. weilii* serogroup Hebdomadis Eco-Challenge strain. The *ligC* gene of *L. weilii* has a 640 nucleotide interrogans-like insertion in the carboxyterminal domain (red). In addition, the *L. weilii* Eco-Challenge strain lacks Big domain 11 (For interpretation of the references to color in this figure legend, the reader is referred to the web version of the article).

Table 1

Leptospira strains, the status of their *lig* genes and level of identity compared to that of serovar Copenhageni.

<i>Leptospira</i> species	Serogroup	Serovar	Strain	lig gene status (PCR/SB/Seq)			% lig DNA sequence identity vs. Copenhageni		
				ligA	ligB	ligC	ligA	ligB	ligC
<i>L. borgpetersenii</i>		Hardjo	L550 ^a	ND/ND/-	+ND/+	ND/ND/-	NA	68.7	NA
		Hardjo	JB 197 ^a	ND/ND/-	+ND/+	ND/ND/-	NA	67.9	NA
<i>L. interrogans</i>	Icterohaemorrhagiae	Copenhageni	Fiocruz L1-130 ^b	+/+/+	+/+/+	+/+/+	100	100	100
	Icterohaemorrhagiae	Lai	56601 ^c	-ND/-	+/+/+	+/+/+	NA	97.1	99.5
	Canicola	ND	Kit0 ^d	+/+/+	+/+/+	+ND/+	90.3	96.7	98.5
	Pomona	Pomona	Kennewicki Cornell ^e	+ND/+	+/+/+	+ND/+	90.1	97.1	99.5
	Pomona	Pomona	Kennewicki PO-06-04 ^f	+/+/+	+/+/+	+ND/+	90.3	96.7	98.8
<i>L. kirschneri</i>	Grippityphosa	Grippityphosa	RM52 ^g	+/+/+	+/+/+	+/+/+	91.4	93.2	90.5
<i>L. noguchii</i>	Bataviae	ND	Cascata ^h	-/-	+/+/+	-ND/-	NA	74.3	NA
<i>L. weilii</i>	Hebdomadis	ND	Eco-Challenge ⁱ	-/-	+/+/+	+ND/+	NA	69.6	78

PCR: PCR screening assay; SB: Southern blot; Seq: gene sequenced; ND: not determined; NA: not applicable.

^aIsolated from a chronically infected bovine, Australia, genome sequenced (Bulach et al., 2006).

^bIsolated from a patient during a leptospirosis epidemic, Brazil (Ko et al., 1999), genome sequenced (Nascimento et al., 2004).

^cCausative agent of rural leptospirosis in China, genome sequenced (Ren et al., 2003).

^dIsolated from an infected canine, Brazil (Silva et al., 2008).

^eIsolated from an aborted equine foetus, USA (Palaniappan et al., 2002).

^fIsolated from an aborted swine foetus, USA.

^gIsolated from an infected swine, USA (Shang et al., 1996).

^hIsolated from a patient, Brazil (Silva et al., 2008).

ⁱIsolated from a triathlete diagnosed with leptospirosis, Malaysia (Haake et al., 2002).