

Genome features of *Leptospira interrogans* serovar Copenhageni

A.L.T.O. Nascimento¹,
S. Verjovski-Almeida²,
M.A. Van Sluys³,
C.B. Monteiro-Vitorello⁵,
L.E.A. Camargo⁵,
L.A. Digiampietri⁶,
R.A. Harstkeerl⁷,
P.L. Ho¹, M.V. Marques⁴,
M.C. Oliveira³,
J.C. Setubal^{6,10},
D.A. Haake^{8,9} and
E.A.L. Martins¹

¹Centro de Biotecnologia, Instituto Butantan, São Paulo, SP, Brasil
²Instituto de Química, ³Instituto de Biociências and
⁴Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, SP, Brasil
⁵Escola Superior de Agricultura Luiz de Queiroz, Universidade de São Paulo, Piracicaba, SP, Brasil
⁶Laboratório de Bioinformática, Instituto de Computação, Universidade Estadual de Campinas, Campinas, SP, Brasil
⁷Koninklijk Instituut voor de Tropen/Royal Tropical Institute (KIT), KIT Biomedical Research, Amsterdam, The Netherlands
⁸The David Geffen School of Medicine at University of California, Los Angeles, CA, USA
⁹Division of Infectious Diseases, Veterans Affairs Greater Los Angeles Healthcare System, Los Angeles, CA, USA
¹⁰Virginia Bioinformatics Institute and Department of Computer Science, Virginia Polytechnic Institute and State University Bioinformatics 1, Blacksburg, VA, USA

Abstract

We report novel features of the genome sequence of *Leptospira interrogans* serovar Copenhageni, a highly invasive spirochete. *Leptospira* species colonize a significant proportion of rodent populations worldwide and produce life-threatening infections in mammals. Genomic sequence analysis reveals the presence of a competent transport system with 13 families of genes encoding for major transporters including a three-member component efflux system compatible with the long-term survival of this organism. The leptospiral genome contains a broad array of genes encoding regulatory system, signal transduction and methyl-accepting chemotaxis proteins, reflecting the organism's ability to respond to diverse environmental stimuli. The identification of a complete set of genes encoding the enzymes for the cobalamin biosynthetic pathway and the novel coding genes related to lipopolysaccharide biosynthesis should bring new light to the study of *Leptospira* physiology. Genes related to toxins, lipoproteins and several surface-exposed proteins may facilitate a better understanding of the *Leptospira* pathogenesis and may serve as potential candidates for vaccine.

Key words

- *Leptospira interrogans*
- Outer membrane proteins
- Lipopolysaccharides
- Transport
- Regulatory systems

Correspondence

A.L.T.O. Nascimento
Centro de Biotecnologia,
Instituto Butantan,
Av. Vital Brazil, 1500
05503-900 São Paulo, SP
Brasil
Fax: +55-11-3726-1505
E-mail: tabet@butantan.gov.br

Research supported by FAPESP
and CNPq.

Received 20 February, 2004

Accepted March 8, 2004

Introduction

Spirochetes are motile, helically shaped bacteria which include the genera *Leptospira*, *Leptonema*, *Borrelia* and *Treponema*. *Borrelia* and *Treponema* are the causative agents of Lyme disease, relapsing fever and syphi-

lis. *Leptospira* consists of a genetically diverse group of pathogenic and non-pathogenic or saprophytic species (1). Leptospirosis is a widespread zoonotic disease: transmission to humans occurs through contact with domestic or wild animal reservoirs or an environment contaminated by their urine.

Infection produces a wide spectrum of clinical manifestations. The early phase of illness is characterized by fever, chills, headache, and severe myalgias. The disease progresses in 5 to 15% of the clinical infections to produce severe multisystem complications such as jaundice, renal failure and hemorrhagic manifestations (2). In developed countries, leptospirosis is associated with recreational activities (1) while in developing countries it produces large urban epidemics with mortality mainly during the rainy season (3). Leptospirosis also represents a major economic problem producing abortions, stillbirths, infertility, failure to thrive, reduced milk production, and death in animals such as cows, pigs, sheep, goats, horses, and dogs (1). Environmental control measures are difficult to implement because of the long-term survival of pathogenic leptospires in soil and water and the abundance of wild and domestic animal reservoirs (1). *Leptospira* are classified according to serovar status - more than 200 pathogenic serovars have been identified. Structural heterogeneity in lipopolysaccharide (LPS) moieties appears to be the basis for the large degree of antigenic variation observed among serovars (1). The development of vaccines has been pursued as a strategy for the prevention of leptospirosis. At present, vaccines are based on inactivated whole cell or membrane preparations of pathogenic leptospires which induce immune responses against leptospiral LPS (1). However, these vaccines do not induce long-term protection against infection and do not provide cross-protective immunity against heterologous leptospiral serovars. Protein antigens conserved among pathogenic serovars may contribute to overcoming the limitations of the currently available vaccines.

The genome sequence of *Leptospira interrogans* serovar Lai was recently published (4) and comparative genome analysis with *L. interrogans* serovar Copenhageni has been performed. We report here new

features of the *L. interrogans* serovar Copenhageni that should contribute to understanding the molecular mechanisms of leptospiral physiology, pathogenesis and facilitate the identification of candidates for broad-range vaccines.

Material and Methods

The sequenced strain, Fiocruz L1-130, was isolated as described by Nascimento et al. (5). The sequencing strategy adopted follows the basic outline of the *Xylella* genome project (6). Library construction, sequencing, assembly, and finishing were carried out by the Agronomical and Environmental Genomes consortium [<http://aeg.lbi.ic.unicamp.br>] and by Instituto Butantan. The genome was assembled using phrap from shotgun reads, cosmid reads and PCR-product sequences. Scaffolding was performed using domestic software. Finishing criteria are based on consensus base phred quality of at least 20 and consensus base covered by at least one read sequence of each DNA strand (6). The first base of the sequence was chosen based on our hypothesis for the origin of the replication locus, which was in turn based on the presence of certain genes and on GC-skew variation. Genome annotation and comparative genomics were done as previously described (7). Detection of potentially surface-exposed integral membrane proteins was carried out as described by Nascimento et al. (5). Sequences from 16S rDNA were manually assembled using ESEE 3.2. Phylogenetic analyses were performed based on two matrices (34 taxa and 1255 positions; 24 taxa and 1375 positions) using the program PAUP 4.0b8 (8). Divergence time was estimated based on 1445 positions of the 16S rRNA sequences. A constant rate of 1 to 2% per 50 million years was assumed (9).

The sequences have been deposited in Genbank under accession numbers AE016823 (chromosome I) and AE016824 (chromosome II).

Results and Discussion

Genome analysis

The *Leptospira* genome consists of two circular chromosomes with a total of 4,627,366 base pairs (bp), chromosome I with 4,277,185 bp and chromosome II with 350,181 bp (5). Circular representations of both chromosomes are depicted in Figure 1. The origin of replication of the large replicon was identified between the *dnaA* and *dnaN* genes, as in other bacterium genomes (10). GC nucleotide skew ($G - C/G + C$) analysis (11) confirmed the origin of replication of the large replicon and indicated two putative sites for the small replicon.

As previously described (12), rRNA genes in *L. interrogans* are not organized into operons, as in most other bacteria, but are scattered over the chromosome I (Figure 1). *L. interrogans* serovar Copenhageni has one *rrf* gene, two *rri* genes and two *rrs* genes coding for 5S, 23S and 16S rRNA, respectively. As in other parasitic strains, *L. interrogans* serovar Copenhageni has only one *rrf* (5S) gene, which is located close to the origin replication region as described before for other strains of *L. interrogans* (12). Comparing the complete *rrs* (16S) sequences for *L. interrogans*, serovars Copenhageni, Lai and Canicola the identity among the sequences is of 99.9 to 100%. The *rrf* (5S) sequence identity comparing Lai and Copenhageni is 100% and the *rri* (23S) is 99%. Based on ribosomal genes, Copenhageni and Lai are closely related, as supported by the whole genome comparison (5).

The Spirochaetes are divided into three major phylogenetic groups, or families: Spirochaetaceae, which includes *Borrelia* and *Treponema* among others; Brachyspiraceae and Leptospiraceae, which contain two genera, *Leptospira* and *Leptonema* (13). Phylogenetic analyses based on 16S rDNA sequences, using *Leptonema* as an outgroup, show that *Leptospira* are split into two well-

supported monophyletic groups (Figure 2), one of them formed by the pathogenic strains (e.g., *L. interrogans*) and the other formed by the non-pathogenic strains (e.g., *L. biflexa*). At the base of the clade of the pathogenic strains, *L. inadai* and *L. fainei* form a well-supported assemblage. Similar results were obtained by Postic et al. (14) based on 16S rDNA analyses. In these analyses *L. interrogans* serovars formed a well-supported monophyletic cluster closely related to *L. kirchneri* (Figure 2B). Considering a constant divergence rate of 1 to 2% per 50 million years for the 16S rDNA (9), separation time between the two main assemblages (*L. interrogans* versus *L. biflexa*) was estimated at 590 to 295 million years.

L. interrogans belongs to a growing number of multichromosomal prokaryotes, including *Vibrio cholerae* (15) and *Ralstonia solanacearum* (16). The small replicon of *L. interrogans* was previously suggested to be a second chromosome based upon the localization of the *metF* gene which encodes an essential biosynthetic pathway enzyme, methylene tetrahydrofolate reductase (17). The genome sequence reveals that genes encoding enzymes for metabolic pathways, such as glycolysis and the tricarboxylic acid cycle, as well as the enzymes for biosynthesis pathways of amino acid and co-factor are also distributed between the two chromosomes. Sequence analysis of chromosome II shows that an almost complete operon of genes coding for the protoheme biosynthesis pathway is present (*hemAIBCENYH*). Although no homologue of the gene coding for uroporphyrinogen III synthetase (*hemD*) was found, experimental evidence has shown that the *hemC* gene is able to cope with *hemD* activity (18). Therefore, *L. interrogans* has the ability to synthesize protoheme *de novo*. In addition, 13 genes clustered in chromosome II coding for the cobalamin biosynthesis pathway were identified (*cobC, cobD, cbiP, cobP, cobB, cobO, cobM, cobJ, cbiG, cobI, cobL, cobH, cobF*) (Figure 3). Orthologues of

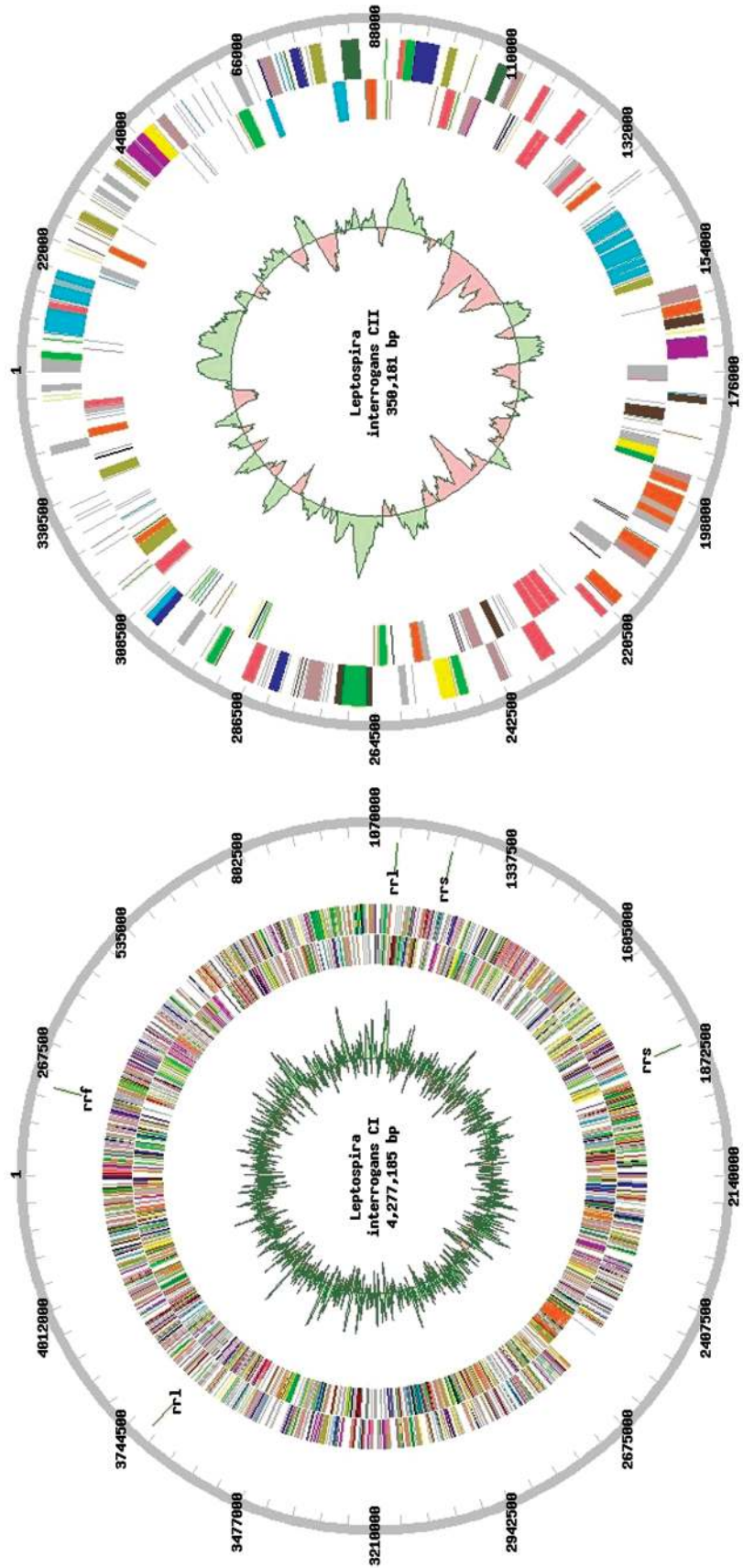


Figure 1. Circular maps of the two chromosomes in *Leptospira interrogans*. Circular representation of the *L. interrogans* serovar Copenhageni genome. The two large and small chromosomes are presented. Circles 1 and 2 (from the outside in), all predicted protein-coding regions (forward and reverse strand, respectively) color-coded by category 1; 2; 3; G + C content. Numbers on the outer circle are base pairs. *rrl*, *rrs* and *rrf* are 23S, 16S and 5S, respectively. Note: the two replicons are not drawn to the same scale. Chromosome II (CII) is 12 times smaller than chromosome I (CI).

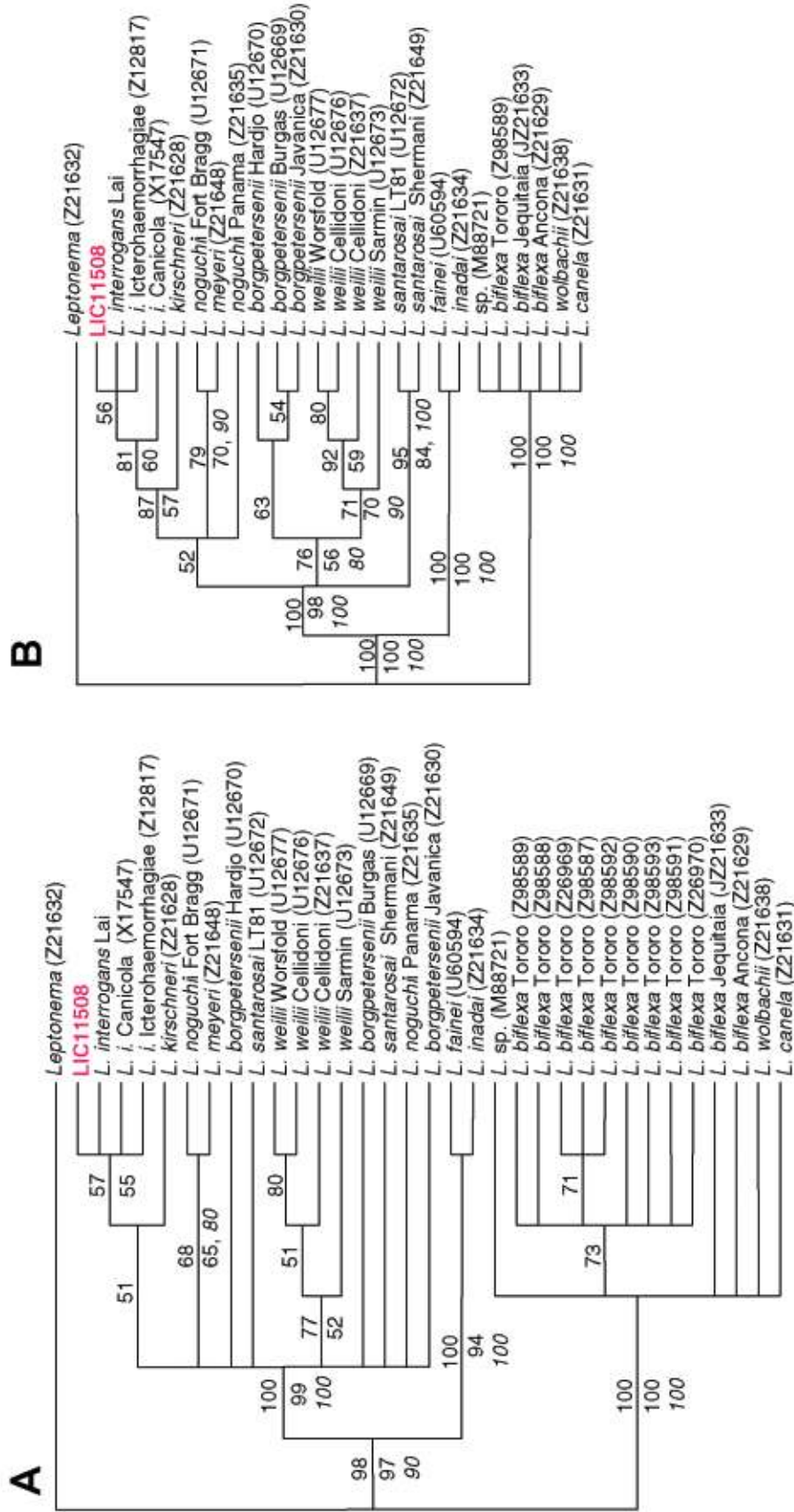


Figure 2. Phylogenetic analysis of *Leptosira interrogans*. Consensus phylogenetic distance tree constructed using 16S rDNA sequences. Numbers on the branches are bootstrap values (2000 replicates) for distance (top), parsimony (below the branch) and maximum likelihood (10 replicates, below the branch, italics). GenBank accession numbers are in parentheses. *Leptonemais* used as outgroup. A, Based on a matrix with 34 taxa and 1255 positions; B, based on a matrix with 24 taxa and 1375 positions.

cobGKN genes, known to be involved in the cobalamin pathway (19) were not found. However, there are two predicted coding sequences inside this operon in chromosome II that could perform these steps. One has an oxidoreductase NAD-binding domain (LIC20133) and the other is a [2Fe-2S] ferredoxin involved in electron transfer (LIC20135). In addition, other genes present in the genome coding for reductases such as LIC11145, LIC13354, LIC12391, and LIC10522 could also cope with these activities. The presence of *cysG* in chromosome I, a gene that encodes a multifunctional protein with methylase, oxidase and ferrocyclase activities (20), may also be a cobalt-inserting enzyme in the B₁₂ pathway. Other genes involved in this biosynthesis pathway were found in chromosome I (*cysG/hemX/cobA, cobT/cobU, cobS*) (Figure 3). In fact, experimental evidence has recently shown that leptospire can grow in medium in the absence of B₁₂ (Hartskeerl RA, unpublished results). This is contrary to the previous statement that *L. interrogans* is unable to synthesize B₁₂, and that is why it is an essential component of the EMJH semi-synthetic medium (1,4). Thus, *L. interrogans*, unlike the spirochetes *Borrelia burgdorferi* and *Treponema pallidum*, have the complete repertoire of genes for *de novo* synthesis of protoheme and cobalamin. The functional link between the two replicons supports the view that the small replicon is in fact a second chromosome.

Transport

Among the 220 *Leptospira* transport proteins, we found 148 proteins comprising 108 different major primary and secondary transporters (Figure 4), as defined by Paulsen et al. (21). There are 34 major primary ATP-driven transporters including 30 ABC-transporters (53 proteins), the largest protein family in *L. interrogans*, as usually is the case for bacteria (22). The ABC superfamily con-

tains both uptake and efflux transport systems, and ATP hydrolysis energizes the transport. The porters of the ABC superfamily consist of two integral membrane proteins (that determine specificity of the transported solute) and two cytoplasmic ATP-hydrolyzing proteins present as homodimers or heterodimers. The uptake systems (but not the efflux systems) additionally possess extra-cytoplasmic solute-binding receptors (one or more per system), which in Gram-negative bacteria are found in the periplasm. We found 21 ABC efflux systems, including drug and heavy metal export and detoxification, lipoprotein-releasing and hemolysin export systems. The remaining 9 are ABC uptake systems, including iron, sulfate, nickel, phosphate, dipeptide, amino acid, and carbohydrate uptake. There is one F-type ATP-synthase system (8 proteins), as mentioned above, and 3 P-type cation-transport ATPases (5 proteins).

We found 59 secondary electrochemical potential-driven transporters (65 proteins) including the largest family of secondary transporters, the resistance-nodulation-cell division superfamily with 12 members, 5 of which belong to the heavy metal efflux family (with 7 proteins) and 7 to the multiple drug efflux family (with 8 proteins). These secondary efflux transporters are energized by proton-motive force and show the widest substrate specificity among all known multi-drug pumps, ranging from most of the currently used antibiotics, disinfectants, dyes, and detergents to simple solvents. The second largest secondary transporter family is the major facilitator superfamily with 9 members including drug:cation antiporters (8 proteins) and a glycerol-phosphate:inorganic phosphate antiporter (1 protein). Additional secondary transporters include three sodium-solute symporters (one sodium-glucose co-transporter), three sodium-bile-acid symporters, and symporters for phosphate, di-tripeptides, glutamate, amino acids, uracil, sulfate as well as an ammonium:potassium

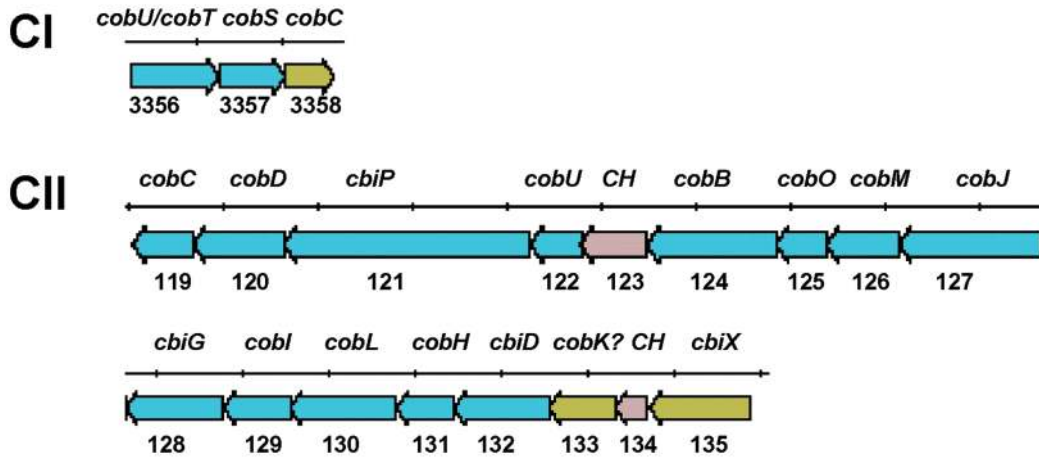
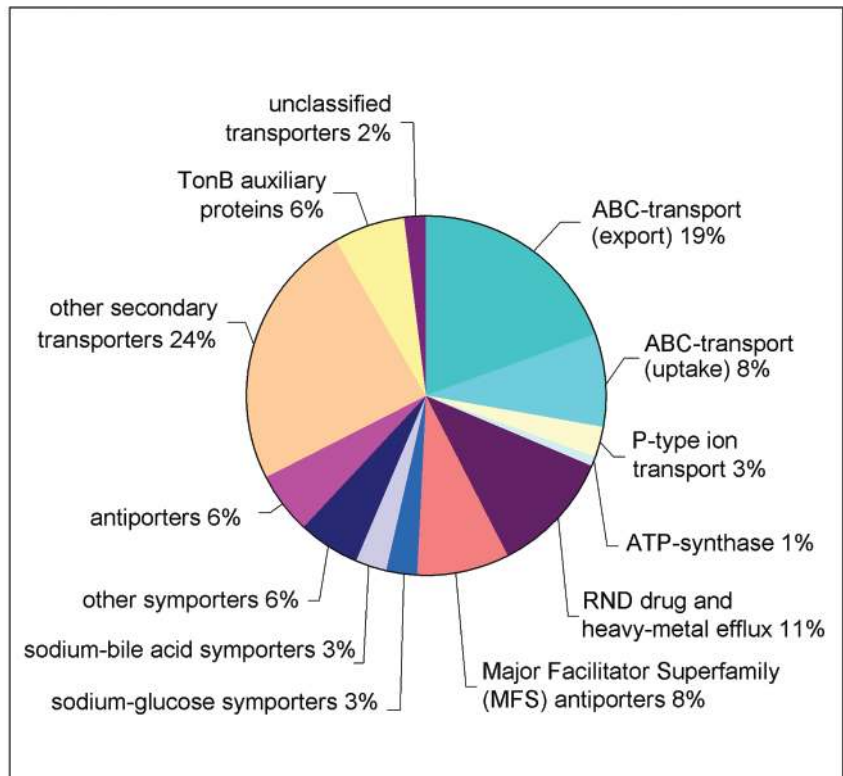


Figure 3. The cobalamin biosynthetic pathway of *Leptospira interrogans*. Genetic organization of the cobalamin biosynthetic locus in the two chromosomes (CI and CII) of *L. interrogans* serovar Copenhageni.

Figure 4. Major transporters of *Leptospira interrogans*. Major primary transporters are driven by ATP hydrolysis and include ABC-transporters, P-type ion transporters and the ATP-synthase. All other transporters indicated are the major secondary transporters, which are driven by the transmembrane electrochemical potential. RND = resistance-nodulation-cell division superfamily.



antiporter, a glycerol-3-P:Pi antiporter, two drug-sodium antiporters and two polysaccharide exporters. Secondary transporters also include 7 members of the TonB family of auxiliary proteins for energizing the outer membrane receptor-mediated active transport. *Leptospira* has many iron-transporting proteins in addition to the cobalamin/iron periplasmic binding protein component of the ABC-transport system (LIC13403) mentioned above. Four members of the outer membrane receptor family have been identified, which are probably involved in the transport of iron dicitrate (LIC10714), hemin (LIC10964), ferric hydroxamate (LIC11345) and ferrienterobactin (LIC12998). In addition, an orthologue of the ferrous iron uptake (FeoB) protein of *E. coli* was found (LIC11402). FeoB has been characterized as an Fe²⁺ uptake system and possesses an ATP/GTP binding motif at its N-terminal hydrophilic domain, therefore being probably energized by ATP or GTP hydrolysis (23).

Oxygen defense

The *Leptospira* genome contains several genes encoding enzymes with peroxidase activities, such as catalase, glutathione peroxidase and thiol peroxidase. However, superoxide dismutase orthologues and two important regulons, SoxRS and OxyR, normally responsible for the defense against oxidative stress, are absent (5). It is possible that metalloporphyrins (24) could provide defense against oxidative damage, since *L. interrogans* is competent for porphyrin biosynthesis and has metal uptake transport proteins. Genes coding for co-migratory bacterioferritin (LIC20093 and LIC10732), thiol peroxidase (LIC12765) and peroxiredoxin (LIC11219) with alkyl hydroperoxide reductase (AhpC) and thiol-specific antioxidant activities, respectively, were also identified. Two predicted coding sequences for bacterioferritin (LIC11310 and LIC13209),

which may have functions analogous to animal ferritin, are also present and may provide both iron detoxification and storage, that would prevent free iron in *Leptospira* from driving oxidative reactions.

Regulatory systems and signal transduction

There are many genes encoding signal transduction proteins in *Leptospira*, indicating that this organism has developed a vast array of regulatory systems that enable it to respond to environmental signals. This variety of regulatory domains is found in non-obligate parasitic bacteria, indicating a much greater need to interpret the signals from the environment in order to respond properly, while obligate parasites have a much lower number and variety of domains in their signal transduction proteins (25). There are 80 genes encoding components of the phosphorylation-mediated signal transduction pathway: 29 histidine kinases (HK), 30 response regulators (RR) and 18 hybrid kinase/regulators (HK/RR). These members of the two-component systems present several domains organized into different arrangements (Figure 5). Nineteen of the histidine kinases are located in the inner membrane, nine are cytoplasmic and one is probably found in the periplasm, as predicted by the PSORT program [<http://psort.nibb.ac.jp/>].

On the other hand, only a third of the hybrid HK/RR are membrane-bound, suggesting that these hybrid proteins are most likely to be involved in phosphorylation cascades, although some of them have the sensor PAS domain. The PAS domain has been reported to sense different environmental stimuli, like oxygen, redox state, nitrogen availability or light, and it may or may not be associated with co-factors such as heme or FAD (26). A two-component protein containing a PAS sensor domain was found to be required for virulence of *M. tuberculosis* in mice, probably because it senses radical oxygen intermediates generated by macro-

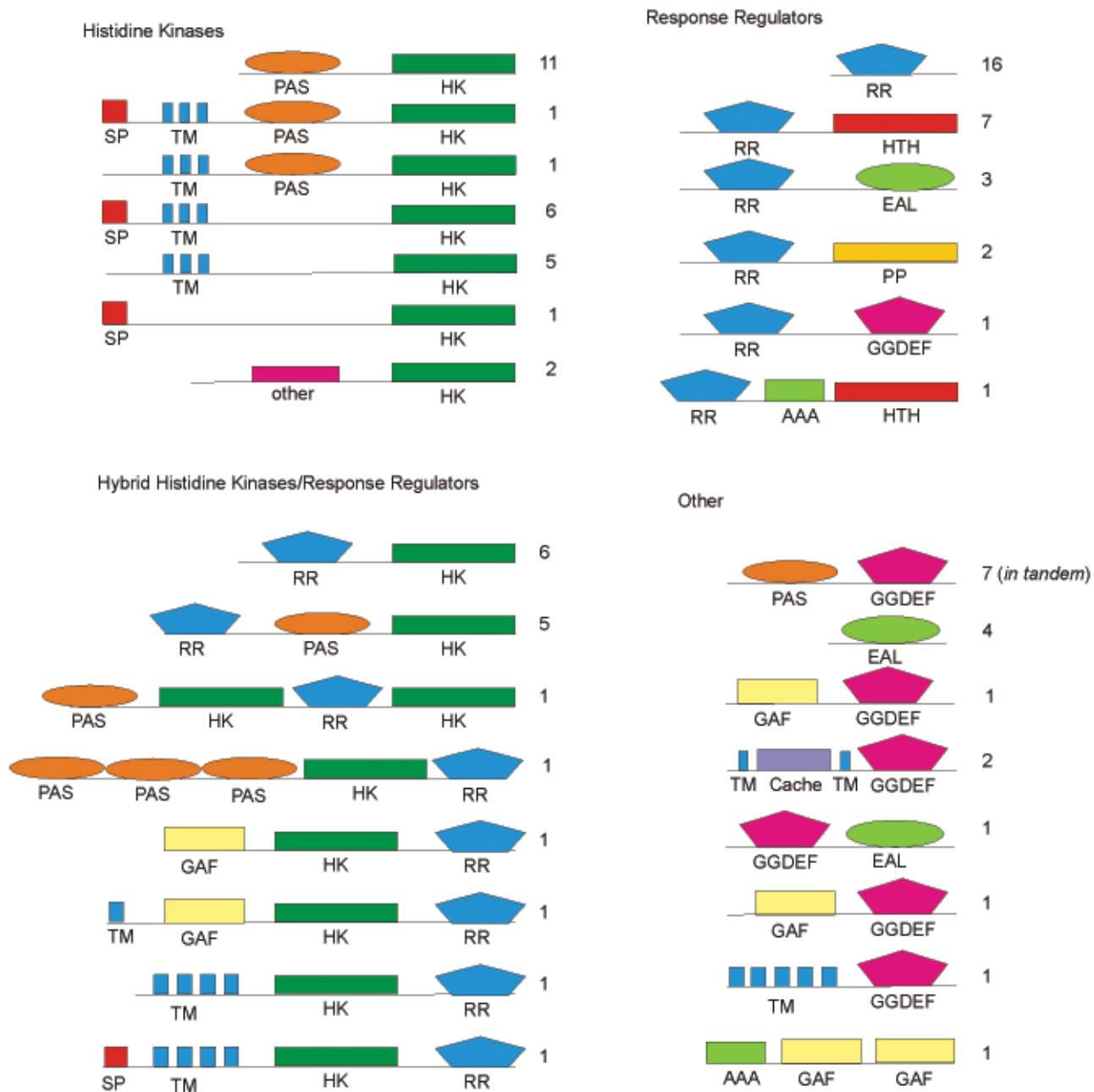


Figure 5. Domain architecture of putative signal transduction proteins from *Leptospira interrogans*. The domain organization of histidine kinases, response regulators, hybrid histidine kinase/response regulators and other peptides probably associated with signal transduction is shown schematically. The number of peptides containing each domain organization is shown on the right, and the genes are found distributed over the large or small replicons, except for the seven peptides containing the PAS/GGDEF arrangement, which are encoded by genes located in tandem in the large replicon. AAA = ATPase; Cache = small ligand binding domain; EAL = phosphodiesterase; GAF = cGMP-binding domain; GGDEF = diguanylate cyclase; HK = histidine kinase domain; HTH = helix-turn-helix; PAS = sensor domain (may include a PAS/PAC arrangement); PP = phosphatase; RR = response regulator domain; SP = signal peptide; TM = transmembrane region.

phage phagocytosis (27), and probably this is also the case for *Leptospira*.

The response regulators are the cytoplasmic effectors of the message, which become functional after being phosphorylated at an aspartate residue by the cognate histidine kinase. The RRs may possess a second effector domain (Figure 5), which will perform its ultimate function, such as the DNA-binding helix-turn-helix domain that allows the RR to regulate transcription. Other noticeable domains found in *L. interrogans* RRs are the GGDEF and EAL motifs, which correspond to putative diguanylate cyclase and phosphodiesterase domains, respectively, and a phosphatase domain similar to the mammalian phosphatase 2C that may be involved in the phosphorelay.

Cyclic nucleotides are likely to have a major regulatory role in *Leptospira*. There are 19 homologues of adenylate/guanylate cyclases, two specific phosphodiesterases and 7 cyclic nucleotide-binding proteins that probably have a regulatory function. There are 12 genes encoding proteins containing the GGDEF motif, seven of which are organized in tandem in chromosome I (LIC11131 to LIC11125), and they also have a PAS/PAC sensor domain in the amino-terminal region, suggesting that they are the product of gene duplication. The diguanylate cyclase activity of the GGDEF domain is required for a novel regulatory mechanism involving bis-(2',5')-cyclic diguanylic acid (c-di-GMP) as an allosteric activator (28). Two of the GGDEF-containing proteins also have the Cache domain, which is involved in chemotaxis signal transduction (29), an important feature for *Leptospira*. There are eight putative diguanylate phosphodiesterases containing an EAL domain, with three of them being associated with an RR domain (Figure 5). *L. interrogans* also presents in its genome three serine/threonine kinases and two hybrid HK/RR with a GAF domain, which is a binding domain for cGMP (30).

Other interesting findings in the genome

include three related genes encoding putative DNA-binding proteins (LIC20104, LIC20105 and LIC20178), which contain 6 transmembrane domains in the amino-terminus and one helix-turn-helix domain of the AraC type at the carboxyl-terminus. These genes are present in chromosome II, and two of them are clustered. Orthologues with a C-terminal DNA-binding domain and a hydrophobic N-terminal region were described in other bacteria, including *Borrelia* and *Treponema* (31), but their function is unknown.

Motility and chemotaxis

The *L. interrogans* genome comprises a relatively large number of motility and chemotaxis genes. Enteric bacteria usually have about 50 genes coding for structural and functional proteins involved in motility (32). A similar number of genes have been identified for the spirochetes *T. pallidum* and *B. burgdorferi* (33,34). Apparently, the motility and chemotaxis apparatus of *L. interrogans* is more complex since its genome contains at least 79 putative motility-associated genes. All genes are well conserved among *L. interrogans*, *T. pallidum* and *B. burgdorferi* and 42 genes were found to be common to all three genera. However, the leptospiral genome contains multiple copies of a number of motility-associated genes, accounting partly for the higher number. For example, the genome of serovar Copenhageni contains 5 *flaB* genes, 4 *motB* genes and 2 *motA* genes compared to 3 *flaB* genes and one copy of both *motA* and *motB* in the *T. pallidum* genome and one copy each of *flaB*, *motA*, and *motB* in the *B. burgdorferi* genome. In addition, the *L. interrogans* genome contains 11 putative genes encoding methyl-accepting chemotaxis proteins, which is roughly twice as many as in *T. pallidum* and *B. burgdorferi*. Forty-eight of the 79 motility-associated genes are positioned into 14 gene clusters varying in size from 2 to 8 genes. Thus, like in *T. pallidum* and *B. burg-*

dorferi, the majority of the structural and functional motility genes are positioned in potential operons. However, the operons probably underwent extensive rearrangements as they are generally smaller, often corresponding to only parts of the major *Treponema* and *Borrelia* operons. For example, the *flgB* operon in *B. burgdorferi* consisting of 26 genes (35) has been disrupted into 6 fragments dispersed at 5 positions in the leptospiral genome. Differences in number of genes and operon organization might be associated with the high flexibility of pathogenic leptospires enabling them to survive and adapt to a variety of environments and hosts.

Hemolysins

The primary lesion caused by *Leptospira* is the damage to the endothelium of small blood vessels, leading to hemorrhage and localized ischemia in multiple organs. As a consequence, renal tubular necrosis, hepatocellular damage, meningitis, and myositis may occur in the infected host (1). Hemolysins may play a fundamental role in this toxic process. Several genes coding for predicted hemolysins were identified in the *L. interrogans* genome. Some of them are structurally related to sphingomyelinases C (LIC10657, LIC12631, LIC12632, LIC11040, and LIC13198). Although generically called sphingomyelinases, it is possible that these genes code for hydrolases that act not only on sphingomyelins but also on other sphingolipids. Erythrocytes may represent a target for these enzymes since they are rich in glycosphingolipids like the antigenic determinants of the ABO system. Interestingly, LIC12631 and LIC12632 are organized as an operon, which may suggest a concerted expression and action.

The other class of genes coding for hemolysins, *tlyABC*, was identified. Although they were assigned to the same TlyABC class, they do not present structural similar-

ity. These putative hemolysins were first identified in the spirochete *Brachyspira hyodysenteriae*. Hemolytic and cytotoxic activities of the recombinant TlyA, TlyB and TlyC proteins, expressed in *E. coli*, were detected (36). TlyB belongs to the family of Clp ATP-dependent proteases (37) and there are 3 genes coding for structurally related proteins (LIC10339, LIC11814 and LIC12017). Thus, 5 genes of the *tlyABC* class (LIC10284, LIC10339, LIC11814, LIC12017, and LIC13143) were identified in the *L. interrogans* genome.

LIC12134 codes for an HlpA-related protein which was characterized as a hemolysin in *Prevotella intermedia*, a common oral bacterium associated with periodontitis (38). Another identified gene (LIC10325) is related to the *hlyX* gene which codes for a predicted hemolysin previously identified in *L. borgpetersenii* serovar Hardjo (Accession number AAF09252, unpublished results). LIC11352 should also be mentioned as the gene which codes for LipL32 or HAP-1, a ubiquitous lipoprotein of pathogenic *Leptospira* (see Lipoproteins section below) with hemolytic activity (39).

Surface-exposed proteins

Pathogenic leptospires require several types of surface-exposed proteins for the purpose of colonization and survival in the mammalian host. Surface-exposed proteins may be categorized as nonspecific porins, specific channels for nutrient acquisition, efflux channels, lipoproteins, adhesins, S-layer glycoproteins, peripheral membrane proteins, or surface-maintenance proteins. Aside from S-layer proteins and peripheral membrane proteins, these surface-exposed proteins would be expected to be integrated into the outer membrane via transmembrane regions or lipid modification. The leptospiral genome contains homologues of SecY and other secretory proteins involved in exporting proteins with signal peptides across

the cytoplasmic membrane. The leptospiral genome also contains both the standard signal peptidase and the lipoprotein signal peptidase. The standard signal peptidase is involved in the hydrolysis of signal peptides of transmembrane outer membrane proteins and periplasmic proteins, releasing them from the cytoplasmic membrane. Lipoprotein signal peptidase hydrolyzes the signal peptides of lipoproteins prior to lipidation of the N-terminal cysteine.

Transmembrane outer membrane proteins

Analysis of the *L. interrogans* genome reveals 83 beta-sheet transmembrane outer membrane proteins (OMPs) and all except one (OmpL1, LIC10973) (40) were previously unknown. An example of a newly identified leptospiral protein with a transmembrane beta-sheet structure is LIC10642 which is predicted to be a member of the OMP superfamily.

Among these newly identified OMPs is a family of predicted coding sequences (LIC20202, LIC10995, LIC11465, and LIC11103) belonging to the superfamily of alpha/beta hydrolases, which includes bacterial lipases. Another transmembrane OMP is LIC11623, a member of a family of highly conserved proteins of Gram-negative bacteria, including *N. meningitidis* Omp85, which is thought to be a chaperonin involved in the assembly of OMPs in the outer membrane (41). OmpL1 belongs to the class of nonspecific porins allowing passage of small molecules (<1000 Daltons) across the leptospiral outer membrane (40). Nonspecific porins allow both influx of nutrients and efflux of products of metabolism. Another example is LIC11458, which is predicted to be a member of the porin superfamily.

Specific channels

Survival in the mammalian host environment by bacterial pathogens requires the

acquisition of certain trace nutrients. For example, iron is essential for leptospiral growth and is bound by a number of high-affinity binding proteins in mammals, which restrict its availability. Bacteria scavenge trace nutrients, including iron, heme, and vitamin B12, from their environment utilizing the cytoplasmic membrane protein TonB and a series of "TonB-dependent" OMPs (Figure 6). After binding, transport of the nutrient across the outer membrane into the periplasm by this type of channel is an energy-dependent step requiring interaction of the TonB-dependent OMP with TonB. The *L. interrogans* genome contains a TonB orthologue (LIC10889) and a large number of TonB-dependent OMPs: LIC20151, LIC20214, LIC10998, LIC10964, LIC10714, LIC12374, LIC12898, LIC12998, LIC11694, LIC11345, LIC10882, LIC10881, and LIC10896.

The leptospiral genome also contains OMPs involved in efflux pathways (Figure 6). The type I efflux system is represented by an orthologue of TolC (LIC13135), the outer membrane exporter of hemolysin and drugs, along with an orthologue of CusC (LIC11941), an outer membrane exporter of copper ion. A second type of efflux pathway found in the leptospiral genome belongs to the resistance-nodulation-cell division superfamily, a three-member complex that catalyzes substrate efflux via an H⁺ antiport mechanism. The three-member complex consists of an resistance-nodulation-cell division transporter, a membrane fusion protein, and an outer membrane factor involved in the export of proteins, carbohydrates, drugs or toxic heavy metals (Figure 6). The leptospiral orthologues are most closely related to those of the *Alcaligenes eutrophus* cobalt/cadmium/zinc export system consisting of the resistance-nodulation-cell division transporter CzcA (LIC12224, LIC15510 and LIC11938), CzcB (LIC12306 and LIC11940), and CzcC (LIC12307 and LIC11941). In addition, *L. interrogans* has

two orthologues of the cation efflux system protein CzcD (LIC11714 and LIC13205), which are members of the cation diffusion facilitator family. CzcD of *Bacillus subtilis* has been shown to catalyze divalent cation (Zn^{2+} or Cd^{2+}) efflux in exchange for the uptake of two monovalent cations (K^+ and H^+) in an electroneutral process energized by the transmembrane pH gradient (42).

Lipoproteins

Experimental evidence for fatty acid modification of leptospiral lipoproteins has been described for the outer membrane lipoproteins LipL32 (LIC11352) (43), LipL36 (LIC13060) (44), and LipL41 (LIC12966) (45). The cytoplasmic membrane also contains lipoproteins, as demonstrated for

LipL31 (LIC11456) and LipL71 (LIC11003) (46). A total of 184 predicted coding sequences in the *L. interrogans* genome were found to have a lipoprotein signal peptidase cleavage site (5). All proposed lipoprotein-coding sequences conform to the rule of having an L, I, V, or F in the -3 and/or -4 position relative to cysteine and most of them have A, G, S, or N in the -1 position relative to cysteine (47).

Proteolytic functions can be assigned to some of the newly identified lipoproteins: four lipoproteins are thermolysin homologues (LIC10715, LIC13320, LIC13321, and LIC13322), and one is a predicted metalloprotease (LIC11834). Several lipoproteins may be involved in hemolysis: two lipoproteins are sphingomyelinase homologues (LIC10657 and LIC12632), and one is a

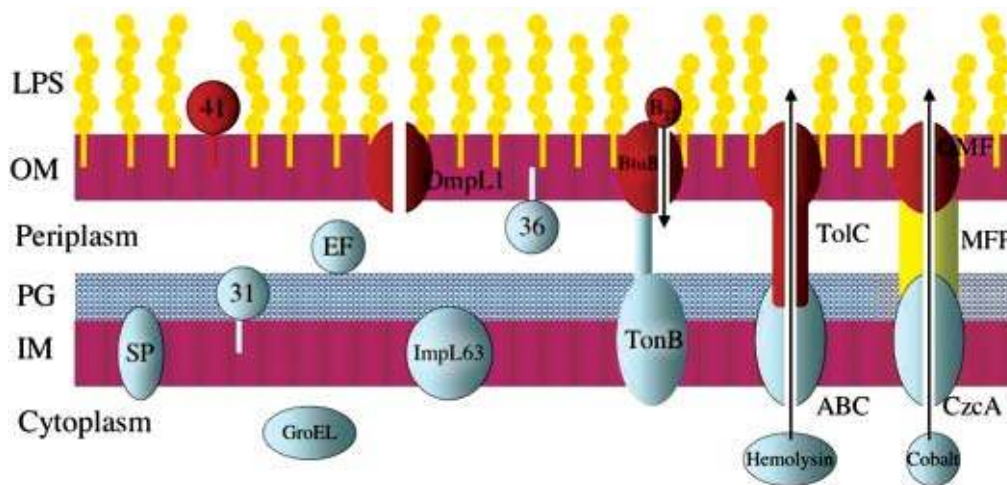


Figure 6. Model of leptospiral membrane architecture. Leptospire has two membranes, an outer membrane (OM) and a cytoplasmic or inner membrane (IM). As in Gram-positive bacteria, the peptidoglycan (PG) cell wall is closely associated with the IM. The leptospiral surface is dominated by lipopolysaccharide (LPS) carbohydrate side chains. Subsurface proteins include the cytoplasmic protein, GroEL, and the periplasmic endoflagella (EF). The IM contains lipoproteins such as LipL31 and transmembrane proteins such as signal peptidase (SP) and ImpL63. The OM contains lipoproteins including LipL41 and LipL36, and transmembrane proteins including the porin, OmpL1. Genomic sequence analysis reveals several novel types of outer membrane proteins (OMPs), including TonB-dependent OMPs involved in nutrient acquisition. BtuB is an example of a TonB-dependent OMP involved in uptake of vitamin B₁₂. The type I efflux system is represented by TolC, which forms a complex with ATP-binding cassette (ABC) transporters to export hemolysins and other cytoplasmic components. The leptospiral genome also contains genes involved in a three-component efflux system consisting of an outer membrane factor (OMF), membrane fusion protein (MFP), and an inner membrane transporter, in this case CzcA, which is involved in heavy metal detoxification.

phospholipase D homologue (LIC11754). Lipoprotein LIC10972 is predicted to be located in the outer membrane and is a MauG homologue belonging to a family of cytochrome c peroxidases that may be involved in defense against hydrogen peroxide. Other lipoproteins with putative enzymatic functions are homologues of glucose dehydrogenase (LIC12294) and cholesterol oxidase (LIC13202).

S-layer and peripheral membrane proteins

S-layers are thought to provide structural integrity to the bacterial surface (48). Although an S-layer has not been observed in pathogenic leptospires, the genome contains at least two proteins with S-layer homology (LIC10868 and LIC12952). Experimental evidence is needed to determine whether these proteins are actually S-layer components.

Like S-layer proteins, peripheral membrane proteins are not integral membrane proteins. LipL45 is processed to a peripheral membrane associated with the outer membrane, P31_{LipL45} (49). P31_{LipL45} expression is dramatically up-regulated in stationary phase cultures, and for this reason may have a membrane-stabilizing function. At this time it is unclear whether P31_{LipL45} is surface-exposed. Interestingly, the leptospiral genome contains a number of predicted coding sequences with homology to LipL45 (LIC20102, LIC20114, LIC13414, and LIC10123).

Surface-maintenance proteins

Bacteria are likely to have a variety of proteins involved in maintaining the surface structural integrity. One such protein, glycerophosphoryl diester phosphodiesterase, is a protein belonging to this category which has been found in all spirochetal genomes studied to date. *E. coli* has two forms of this enzyme, a cytosolic form, *ugpQ*, and a peri-

plasmic form, *glpQ*. The leptospiral genome contains two homologues of this enzyme, LIC13182 and LIC10293. The former should be the cytosolic form because it lacks a signal peptide while the latter should be the exported form because it has an N-terminal signal peptide. In other spirochetes, *GlpQ* is associated with the outer membrane (50).

Cytoplasmic membrane proteins

The leptospiral genome contains a number of proteins that belong to a large family of prokaryotic proteins with homology to the peptidoglycan-associated portion of *E. coli* OmpA (51). These proteins are predicted to be either cytoplasmic membrane (LIC20250, LIC10592, LIC13479, and LIC10050) or periplasmic proteins (LIC10537 and LIC10191), rather than OMPs, which is consistent with experimental evidence that the spirochetal cell wall is more closely associated with the cytoplasmic membrane than the outer membrane. An interesting protein family is the mechanosensitive ion channel (LIC20069 and LIC12671). Two members of this protein family of *M. jannaschii* have been functionally characterized and both form mechanosensitive ion channels (52). Therefore, this family is likely to also encode mechanosensitive channel proteins in *L. interrogans*, playing a physiological role in bacterial osmoregulation.

Lipopolysaccharides

Lipopolysaccharides distinguish the leptospiral surface from those of the other invasive spirochetes. Changes in genes involved in the LPS biosynthesis apparatus are thought to account for serovar diversity among leptospires (53). It has been shown that the leptospiral LPS resembles a typical LPS from Gram-negative bacteria, containing pentoses, hexoses and heptoses. Although the predominant sugar is rhamnose, a large variety of other sugar residues are found (1). Anti-

bodies raised against LPS from different *Leptospira* strains during infections are related to polysaccharide structure in terms of its sugar composition, number, repetitiveness, and ramification (1). In *Leptospira*, as in many other bacteria, at least part of the genes coding for enzymes of the polysaccharide biosynthesis pathway are found clustered in a region of chromosomes named O-antigen gene cluster (*rfb* locus) (53).

The complete genome analysis reveals a large portion of about 119 kb (genome position 2.538.470-2.554.255), in which all the 108 predicted coding sequences are transcribed in the same orientation (Figure 7). Interestingly, this region is dense in genes related to LPS biosynthesis (Table 1) and includes the O-antigen gene cluster previously described in *L. interrogans* serovar Copenhageni (53,54). In the first 14 kb of this region the predicted genes seem not to be related to LPS biosynthesis, but to DNA replication, and some genes code for ribosomal proteins. In the other 105 kb of this region there are 94 predicted genes 56 of which are clearly related to LPS biosynthesis. These predicted genes encode enzymes for nucleotide sugar biosynthesis, such as the enzymes for dTDP-rhamnose biosynthesis (54) for CMP-N-acetylneuraminic acid and for perosamine synthase. In addition, many genes coding for sugar transferases, sugar modifications and for the O-antigen processing, Wzx-flippase and Wzy-polymerase, involved in oligosaccharide exporting and assembly of the LPS, were identified. Some genes encoding proteins of the lipid A biosynthesis (*lpxD*) and transport (*msbA*) are also found in this region.

The comparison with the corresponding region in the genome of *L. interrogans* serovar Lai (4) revealed only three distinct predicted coding sequences: two extra copies of genes for aminotransferase (LIC12197 and LIC12198) and the absence of a gene encoding galactoside O-acetyltransferase (LA1622).

Seventy-seven other genes, probably related to LPS biosynthesis, were identified along the genome. All the genes related to lipid A biosynthesis described in *E. coli* were identified in the *Leptospira* genome (*lpxA*, *lpxC*, *lpxD*, *lpxB*, *lpxK*). However, the predominant fatty acids in lipid A of *L. interrogans* are dodecanoic and hexadecanoic acid instead of hydroxymyristoyl (14 carbons), which is the signature of Gram-negative bacteria (1). Comparative analysis between LpxA of *E. coli* and *P. aeruginosa* (55) showed that few structural differences in this enzyme determine changes in the fatty acid chain size incorporated during lipid A biosynthesis. The lower endotoxicity of leptospiral LPS as compared to LPS from Gram-negative bacteria has been reported (56). It was also reported that leptospiral LPS induces a TLR2-dependent cell activation, instead of LTR4, the receptor frequently involved in the LPS immune response (57). The lower endotoxicity and the difference in the mechanism of cell immune response activation are supposed to be related to differences in the lipid A structure.

Genes encoding enzymes of Kdo biosynthesis (*kdsA*, *kdsB*) such as Kdo-transferase (*waaA* or *kdtA*), which catalyzes the binding of Kdo to lipid IVA, were identified. Although typical Kdo was found in *Leptospira* LPS, it was reported to be substituted at different carbon positions (1).

Four paralogues encoding MsbA, the protein that transfers the flippase of lipid A in

Table 1. Number of genes at the lipopolysaccharide 119-kb locus.

| Predicted protein function | |
|--------------------------------------|-------------------|
| Nucleotide sugar biosynthesis | 22 |
| Sugar transferase | 17 |
| Sugar epimerase | 10 |
| Lipid A and polysaccharide transport | 4 |
| Hypothetical | 26 (17 conserved) |
| Aminoacyl tRNA synthase | 3 |
| DNA metabolism | 4 |
| Ribosomal protein | 3 |
| Others | 19 |

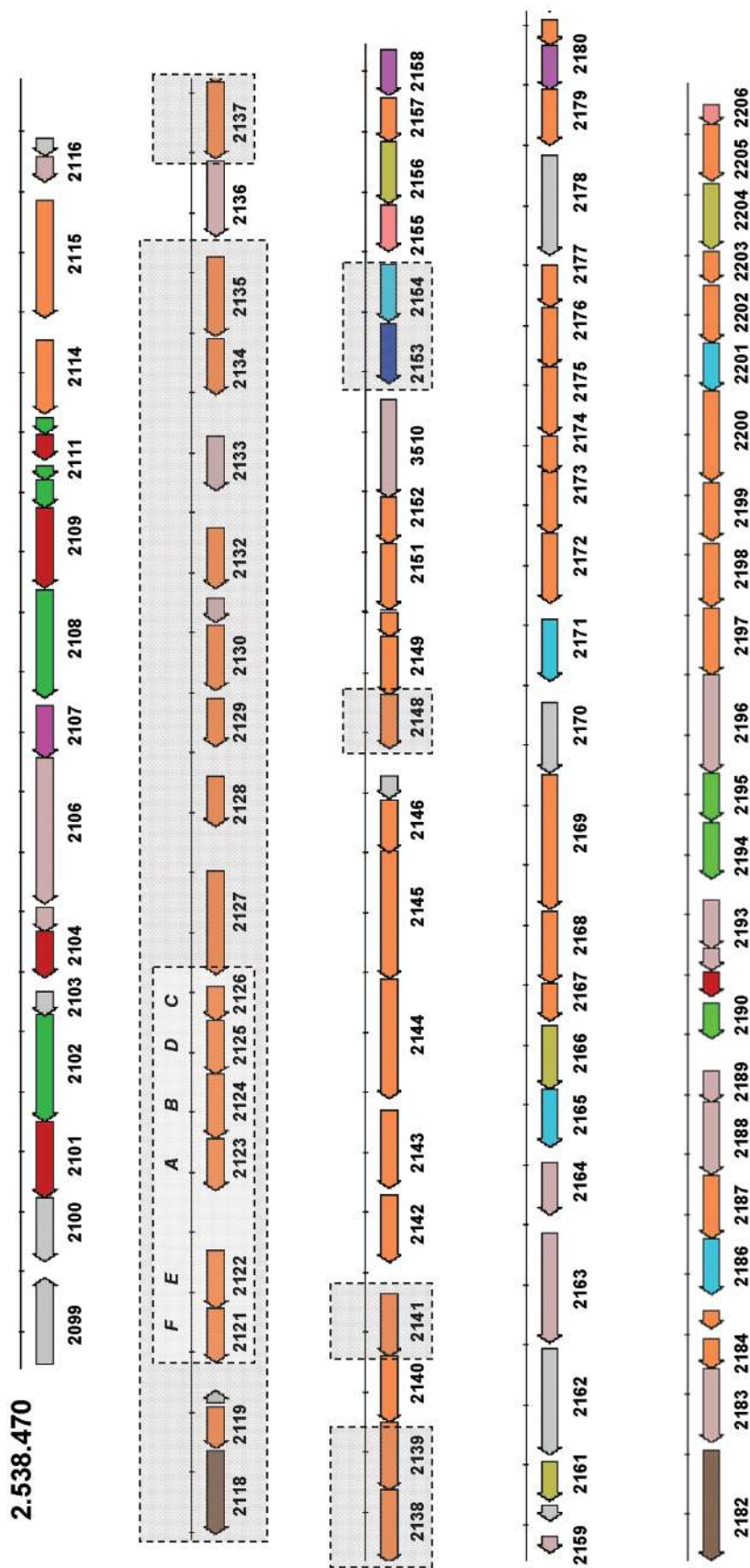


Figure 7. Lipopolysaccharide region of *Leptospira interrogans*. The 119-kb region of *L. interrogans* serovar Copenhageni genome with 108 predicted coding sequences transcribed in the same orientation. Orange arrows indicate predicted genes encoding polysaccharide biosynthesis-related proteins. Shaded regions indicate the homologous genes from *L. borgpetersenii* serovar Hardjovis (53). The region containing the predicted coding sequences, indicated by the letters F, E, A, B, D, C, was first indicated as the *rfb* genes related to rhamnose biosynthesis in *L. interrogans* Copenhageni (54).

Table 2. Insertion sequence (IS) elements present in *Leptospira interrogans* serovar Copenhageni.

| Element | Family | Copies | Length (bp) | Transposase | Terminal inverted repeats |
|---------------|--------|--------|-------------|------------------|--|
| <i>ISlin1</i> | IS110 | 11 | 1,423 | 2 (126aa + 195) | 5'end AAACTCAACAtctCGCTCTTTAcGAATCGC 3'end GCGATTCaTAAAGAGCGttgTGTTGAGTTT |
| <i>IS1500</i> | IS3 | 8 | 1,234 | 2 (282aa + 99aa) | 5'end TAACCTaGtgACGaAttAttGGACACaTTTT 3'end AAAAgGTGTCCgtTttTaCGTagCcAGGTTA |
| <i>IS1501</i> | IS3 | 5 | 1,230 | 2 (132aa + 85aa) | 5'end TGAagTAGTcACATAAAAGTgGACAgCcaTTT 3'end AAAaaGgTGTCtACTTTTATGTtACTAggTCA |

the inner membrane from the cytoplasmic side to the periplasmic side, were identified. One of the *msbA* genes is located in the 119-kb region. Another *msbA* gene is located upstream of *lpxK*, as reported in many other genomes (55). It will be interesting to compare the set of genes coding for enzymes related to LPS biosynthesis identified in the *Leptospira* genome to the orthologues in other microorganisms, in order to correlate LPS structural differences and endotoxic activity.

Comparative genomics and insertion sequences

A three-way comparison between the *L. interrogans* genome and the genomes of *B. burgdorferi* and *T. pallidum* yields the following results: 1167 (31%) of the genes in *L. interrogans* Copenhageni are found in *B. burgdorferi* and/or in *T. pallidum*, 666 (41%) of the genes in *B. burgdorferi* are found in the Copenhageni genome, and 589 (57%) of the genes in *T. pallidum* are found in the *L. interrogans* genome. A total of 362 predicted genes were found to be shared by all three spirochetes, 45 of which are hypothetical (detailed list in our project website: <http://aeg.lbi.ic.unicamp.br/world/lic/>). These results show that a thorough analysis of the genome can significantly contribute to the understanding of spirochete biology.

Our analysis of the *L. interrogans* genome revealed the presence of the previ-

ously described insertion sequence (IS) elements *IS1500*, *IS1501* (58), reminiscent insertions of *IS1533* (59) and *IS1502* (60) and an IS recently identified that we designated *ISlin1* (5) (Table 2). *IS1500*, *IS1501* and *ISlin1* account for 24 insertions distributed throughout chromosome I. Together, IS elements and transposases comprise 2% of the genome. These elements are related to major IS families such as IS110 and IS3 that are defined by their conserved amino acid motif (DDE) in the transposase. So far all the insertions were found in intergenic regions, having no mutagenic effects on *L. interrogans* genes.

The *L. interrogans* genome provides new insights into biosynthesis pathway, transport families, environmental response, and pathogenesis. A broad array of regulatory system proteins enable this organism to respond to environmental signals. A new group of genes involved in LPS biosynthesis may contribute to the elucidation of serovar diversity among leptospires. Several categories of surface-exposed proteins required for colonization and survival in the mammalian host were identified. These proteins may have a role in mechanisms of leptospiral pathogenesis and protective immunity. Available vaccines are serovar-specific and have low efficacy (1). Surface-exposed proteins that are conserved among pathogenic serovars may be used for vaccine development for the prevention of leptospirosis.

Acknowledgments

We are deeply indebted to Drs. I. Raw (Fundação Butantan, São Paulo, SP, Brazil) for use of laboratory facilities and valuable support. We thank Dr. Albert I. Ko (Fiocruz, Salvador, BA, Brazil) for the strain Fiocruz L1-130, the Agronomical and Environmental Genomes (AEG)

Consortium of the network Organization for Nucleotide Sequencing and Analysis (ONSA) for the genome sequencing data, and Dr. L.C.C. Leite (Instituto Butantan) for a critical reading of this manuscript.

Linear representation of the *Leptospira interrogans* serovar Copenhageni chromosomes (CI and CII). Genes are colored according to their biological role. Arrows indicate the direction of transcription. The pie representation indicates the distribution of the number of genes according to their biological role. The numbers below protein-producing genes correspond to gene identification numbers (IDs).

Insert 

Additional information is contained in the project website
<http://aeg.lbi.ic.unicamp.br/world/lic/>

References

- Faine S, Adler B, Bolin C & Perolat P (1999). *Leptospira and Leptospirosis*. MediSci, Melbourne, Australia.
- Bharti AR, Nally JE, Ricaldi JN et al. (2003). Leptospirosis: a zoonotic disease of global importance. *Lancet Infectious Diseases*, 3: 757-771.
- Romero EC, Bernardo CC & Yasuda PH (2003). Human leptospirosis: a twenty-nine-year serological study in São Paulo, Brazil. *Revista do Instituto de Medicina Tropical de São Paulo*, 45: 245-248.
- Ren SX, Fu G, Jiang XG et al. (2003). Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature*, 422: 888-893.
- Nascimento ALTO, Ko AI, Martins EAL et al. (2004). Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *Journal of Bacteriology* (in press).
- Simpson AJ, Reinach FC, Arruda P et al. (2000). The genome sequence of the plant pathogen *Xylella fastidiosa*. *Nature*, 406: 151-157.
- Van Sluys MA, de Oliveira MC, Monteiro-Vitorello CB et al. (2003). Comparative analyses of the complete genome sequences of Pierce's disease and citrus variegated chlorosis strains of *Xylella fastidiosa*. *Journal of Bacteriology*, 185: 1018-1026.
- Swofford DL (2000). *PAUP*-Phylogenetic Analysis Using Parsimony (*and other methods)*. Version 4. Sinauer Associates, Sunderland, MA, USA.
- Ochman H, Elwyn S & Moran NA (1999). Calibrating bacterial evolution. *Proceedings of the National Academy of Sciences, USA*, 96: 12638-12643.
- Messer W (2002). The bacterial replication initiator DnaA. DnaA and oriC, the bacterial mode to initiate DNA replication. *FEMS Microbiology Reviews*, 26: 355-374.
- Lobry JR (1996). Asymmetric substitution patterns in the two DNA strands of bacteria. *Molecular Biology and Evolution*, 13: 660-665.
- Fukunaga M & Mifuchi I (1989). Unique organization of *Leptospira interrogans* rRNA genes. *Journal of Bacteriology*, 171: 5763-5767.
- Paster BJ & Dewhirst FE (2000). Phylogenetic foundation of spirochetes. *Journal of Molecular Microbiology and Biotechnology*, 2: 341-344.
- Postic D, Riquelme-Sertour N, Merien F, Perolat P & Baranton G (2000). Interest of partial 16S rDNA gene sequences to resolve heterogeneities between *Leptospira* collections: application to *L. meyeri*. *Research in Microbiology*, 151: 333-341.
- Heidelberg JF, Eisen JA, Nelson WC et al. (2000). DNA sequence of both chromosomes of the cholera pathogen *Vibrio cholerae*. *Nature*, 406: 477-483.
- Salanoubat M, Genin S, Artiguenave F et al. (2002). Genome sequence of the plant pathogen *Ralstonia solanacearum*. *Nature*, 415: 497-502.
- Bourhy P & Saint Girons I (2000). Localization of the *Leptospira interrogans* metF gene on the CII secondary chromosome. *FEMS Microbiology Letters*, 191: 259-263.
- Guegan R, Camadro JM, Saint Girons I & Picardeau M (2003). *Leptospira* spp. possess a complete haem biosynthetic pathway and are able to use exogenous haem sources. *Molecular Microbiology*, 49: 745-754.
- Rodionov DA, Vitreschak AG, Mironov AA & Gelfand MS (2003). Comparative genomics of the vitamin B12 metabolism and regulation in prokaryotes. *Journal of Biological Chemistry*, 278: 41148-41159.
- Spencer JB, Stolowich NJ, Roessner CA & Scott AI (1993). The *Escherichia coli* cysG gene encodes the multifunctional protein, siroheme synthase. *FEBS Letters*, 335: 57-60.
- Paulsen IT, Nguyen L, Sliwinski MK, Rabus R & Saier Jr MH (2000). Microbial genome analyses: comparative transport capabilities in eighteen prokaryotes. *Journal of Molecular Biology*, 301: 75-100.
- Meidanis J, Braga MD & Verjovski-Almeida S (2002). Whole-genome analysis of transporters in the plant pathogen *Xylella fastidiosa*. *Microbiology and Molecular Biology Reviews*, 66: 272-299.
- Kammler M, Schon C & Hantke K (1993). Characterization of the ferrous iron uptake system of *Escherichia coli*. *Journal of Bacteriology*, 175: 6212-6219.
- Vujaskovic Z, Batinic-Haberle I, Rabbani ZN, Feng QF, Kang SK, Spasojevic I, Samulski TV, Fridovich I, Dewhirst MW & Anscher MS (2002). A small molecular weight catalytic metalloporphyrin antioxidant with superoxide dismutase (SOD) mimetic properties protects lungs from radiation-induced injury. *Free Radical Biology and Medicine*, 33: 857-863.
- Galperin MY, Nikolskaya AN & Koonin EV (2001). Novel domains of the prokaryotic two-component signal transduction systems. *FEMS Microbiology Letters*, 203: 11-21.
- Taylor BL & Zhulin IB (1999). PAS domains: internal sensors of oxygen, redox potential, and light. *Microbiology and Molecular Biology Reviews*, 63: 479-506.
- Rickman L, Saldanha JW, Hunt DM, Hoar DN, Colston MJ, Millar JB & Buxton RS (2004). A two-component signal transduction system with a PAS domain-containing sensor is required for virulence of *Mycobacterium tuberculosis* in mice. *Biochemical and Biophysical Research Communications*, 314: 259-267.
- Ausmees N, Mayer R, Weinhouse H, Volman G, Amikam D, Benziman M & Lindberg M (2001). Genetic data indicate that proteins containing the GGDEF domain possess diguanylate cyclase activity. *FEMS Microbiology Letters*, 204: 163-167.
- Anantharaman V & Aravind L (2000). Cache - a signaling domain common to animal Ca(2+)-channel subunits and a class of prokaryotic chemotaxis receptors. *Trends in Biochemical Sciences*, 25: 535-537.
- Ho YS, Burden LM & Hurley JH (2000). Structure of the GAF domain, a ubiquitous signaling motif and a new class of cyclic GMP receptor. *EMBO Journal*, 19: 5288-5299.
- Subramanian G, Koonin EV & Aravind L (2000). Comparative genome analysis of the pathogenic spirochetes *Borrelia burgdorferi* and *Treponema pallidum*. *Infection and Immunity*, 68: 1633-1648.
- Bourret RB, Charon NW, Stock AM & West AH (2002). Bright lights, abundant operons-fluorescence and genomic technologies advance studies of bacterial locomotion and signal transduction: review of the BLAST meeting, Cuernavaca, Mexico, 14 to 19 January 2001. *Journal of Bacteriology*, 184: 1-17.
- Fraser CM, Casjens S, Huang WM et al. (1997). Genomic sequence of a Lyme disease spirochaete, *Borrelia burgdorferi*. *Nature*, 390: 580-586.
- Fraser CM, Norris SJ, Weinstock GM et al. (1998). Complete genome sequence of *Treponema pallidum*, the syphilis spirochete. *Science*, 281: 375-388.
- Ge Y, Old IG, Saint Girons I & Charon NW (1997). Molecular characterization of a large *Borrelia burgdorferi* motility operon which is initiated by a consensus sigma70 promoter. *Journal of Bacteriol-*

- ogy, 179: 2289-2299.
36. ter Huurne AA, Muir S, van Houten M, van der Zeijst BA, Gaastra W & Kusters JG (1994). Characterization of three putative *Serpulina hydysenteriae* hemolysins. *Microbial Pathogenesis*, 16: 269-282.
 37. Squires C & Squires CL (1992). The Clp proteins: proteolysis regulators or molecular chaperones? *Journal of Bacteriology*, 174: 1081-1085.
 38. Beem JE, Nesbitt WE & Leung KP (1999). Cloning of *Prevotella intermedia* loci demonstrating multiple hemolytic domains. *Oral Microbiology and Immunology*, 14: 143-152.
 39. Lee SH, Kim KA, Park YG, Seong IW, Kim MJ & Lee YJ (2000). Identification and partial characterization of a novel hemolysin from *Leptospira interrogans* serovar lai. *Gene*, 254: 19-28.
 40. Shang ES, Exner MM, Summers TA, Martinich C, Champion CI, Hancock RE & Haake DA (1995). The rare outer membrane protein, OmpL1, of pathogenic *Leptospira* species is a heat-modifiable porin. *Infection and Immunity*, 63: 3174-3181.
 41. Voulhoux R, Bos MP, Geurtsen J, Mols M & Tommassen J (2003). Role of a highly conserved bacterial protein in outer membrane protein assembly. *Science*, 299: 262-265.
 42. Guffanti AA, Wei Y, Rood SV & Krulwich TA (2002). An antiport mechanism for a member of the cation diffusion facilitator family: divalent cations efflux in exchange for K⁺ and H⁺. *Molecular Microbiology*, 45: 145-153.
 43. Haake DA, Chao G, Zuerner RL, Barnett JK, Barnett D, Mazel M, Matsunaga J, Levett PN & Bolin CA (2000). The leptospiral major outer membrane protein LipL32 is a lipoprotein expressed during mammalian infection. *Infection and Immunity*, 68: 2276-2285.
 44. Haake DA, Martinich C, Summers TA, Shang ES, Pruetz JD, McCoy AM, Mazel MK & Bolin CA (1998). Characterization of leptospiral outer membrane lipoprotein LipL36: downregulation associated with late-log-phase growth and mammalian infection. *Infection and Immunity*, 66: 1579-1587.
 45. Shang ES, Summers TA & Haake DA (1996). Molecular cloning and sequence analysis of the gene encoding LipL41, a surface-exposed lipoprotein of pathogenic *Leptospira* species. *Infection and Immunity*, 64: 2322-2330.
 46. Haake DA & Matsunaga J (2002). Characterization of the leptospiral outer membrane and description of three novel leptospiral membrane proteins. *Infection and Immunity*, 70: 4936-4945.
 47. Haake DA (2000). Spirochaetal lipoproteins and pathogenesis. *Microbiology*, 146 (Part 7): 1491-1504.
 48. Schuster B & Sleytr UB (2000). S-layer-supported lipid membranes. *Journal of Biotechnology*, 74: 233-254.
 49. Matsunaga J, Young TA, Barnett JK, Barnett D, Bolin CA & Haake DA (2002). Novel 45-kilodalton leptospiral protein that is processed to a 31-kilodalton growth-phase-regulated peripheral membrane protein. *Infection and Immunity*, 70: 323-334.
 50. Cameron CE, Castro C, Lukehart SA & Van Voorhis WC (1998). Function and protective capacity of *Treponema pallidum* subsp. *pallidum* glycerophosphodiester phosphodiesterase. *Infection and Immunity*, 66: 5763-5770.
 51. De Mot R & Vanderleyden J (1994). The C-terminal sequence conservation between OmpA-related outer membrane proteins and MotB suggests a common function in both gram-positive and gram-negative bacteria, possibly in the interaction of these domains with peptidoglycan. *Molecular Microbiology*, 12: 333-334.
 52. Kloda A & Martinac B (2001). Structural and functional differences between two homologous mechanosensitive channels of *Methanococcus jannaschii*. *EMBO Journal*, 20: 1888-1896.
 53. de la Pena-Moctezuma A, Bulach DM, Kalambaheti T & Adler B (1999). Comparative analysis of the LPS biosynthetic loci of the genetic subtypes of serovar Hardjo: *Leptospira interrogans* subtype Hardjoprajitno and *Leptospira borgpetersenii* subtype Hardjobovis. *FEMS Microbiology Letters*, 177: 319-326.
 54. Mitchison M, Bulach DM, Vinh T, Rajakumar K, Faine S & Adler B (1997). Identification and characterization of the dTDP-rhamnose biosynthesis and transfer genes of the lipopolysaccharide-related rfb locus in *Leptospira interrogans* serovar Copenhageni. *Journal of Bacteriology*, 179: 1262-1267.
 55. Raetz CR & Whitfield C (2002). Lipopolysaccharide endotoxins. *Annual Review of Biochemistry*, 71: 635-700.
 56. de Souza L & Koury MC (1992). Isolation and biological activities of endotoxin from *Leptospira interrogans*. *Canadian Journal of Microbiology*, 38: 284-289.
 57. Werts C, Tapping RI, Mathison JC et al. (2001). Leptospiral lipopolysaccharide activates cells through a TLR2-dependent mechanism. *Nature Immunology*, 2: 346-352.
 58. Boursaux-Eude C, Saint Girons I & Zuerner R (1995). IS1500, an IS3-like element from *Leptospira interrogans*. *Microbiology*, 141 (Part 9): 2165-2173.
 59. Zuerner RL (1994). Nucleotide sequence analysis of IS1533 from *Leptospira borgpetersenii*: identification and expression of two IS-encoded proteins. *Plasmid*, 31: 1-11.
 60. Zuerner RL & Huang WM (2002). Analysis of a *Leptospira interrogans* locus containing DNA replication genes and a new IS, IS1502. *FEMS Microbiology Letters*, 215: 175-182.