

Leonardo Hiroyuki Santos Momo

CARACTERIZAÇÃO DA FAMÍLIA DE REGULADORES  
DE ABSORÇÃO DE METAIS E RESPOSTA A ESTRESSE  
EM *LEPTOSPIRA INTERROGANS* SOROVAR  
COPENHAGENI

Tese apresentada ao  
Programa de Pós-Graduação  
Interunidades em Biotecnologia  
USP/Insituto Butantan/IPT  
para obtenção do Título  
de Mestre em Biotecnologia.

São Paulo  
2015

Leonardo Hiroyuki Santos Momo

CARACTERIZAÇÃO DA FAMÍLIA DE  
REGULADORES DE ABSORÇÃO DE  
METAIS E RESPOSTA A ESTRESSE EM  
*LEPTOSPIRA INTERROGANS* SOROVAR  
COPENHAGENI

Tese apresentada ao  
Programa de Pós-Graduação  
Interunidades em Biotecnologia  
USP/Insituto Butantan/IPT  
para obtenção do Título  
de Mestre em Biotecnologia.

Área de concentração: Biotecnologia  
Orientador: Paulo Lee Ho  
Co-Orientadora: Renata Maria Augusto da  
Costa

Versão corrigida. A versão original eletrônica encontra-se disponível tanto na biblioteca do ICB quanto na Biblioteca Digital de Teses e Dissertações da USP (BDTD).

São Paulo  
2015

DADOS DE CATALOGAÇÃO NA PUBLICAÇÃO (CIP)  
Serviço de Biblioteca e Informação Biomédica do  
Instituto de Ciências Biomédicas da Universidade de São Paulo

© reprodução total

Momo, Leonardo Hiroyuri Santos.

Caracterização da família de reguladores de absorção de metais e resposta a estresse em *leptospira interrogans* sorovar copenhageni / Leonardo Hiroyuki Santos Momo. -- São Paulo, 2015.

Orientador: Prof. Dr. Paulo Lee Ho.

Dissertação (Mestrado) – Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação Interunidades em Biotecnologia USP/IPT/Instituto Butantan. Área de concentração: Biotecnologia. Linha de pesquisa: Mutagênese e reparo de DNA.

Versão do título para o inglês: Characterization of the metal absorption regulator family and stress response in *leptospira interrogans* serovar copenhageni.

1. Ferric uptake regulator 2. Sistema SOS 3. Reparo de DNA 4. Mutagênese 5. Expressão Gênica I. Lee Ho, Prof. Dr. Paulo II. Universidade de São Paulo. Instituto de Ciências Biomédicas. Programa de Pós-Graduação Interunidades em Biotecnologia USP/IPT/Instituto Butantan III. Título.

ICB/SBIB017/2015

UNIVERSIDADE DE SÃO PAULO  
**Programa de Pós-Graduação Interunidades em Biotecnologia**  
Universidade de São Paulo, Instituto Butantan, Instituto de Pesquisas Tecnológicas

---

Candidato(a): Leonardo Hiroyuki Santos Momo.

Título da Dissertação: Caracterização da família de reguladores de absorção de metais e resposta a estresse em *leptospira interrogans* sorovar copenhageni.

Orientador(a): Prof. Dr. Paulo Lee Ho.

A Comissão Julgadora dos trabalhos de Defesa da **Dissertação de Mestrado**, em sessão pública realizada a ...../...../....., considerou

**Aprovado(a)**

**Reprovado(a)**

Examinador(a): Assinatura: .....  
Nome: .....  
Instituição: .....

Examinador(a): Assinatura: .....  
Nome: .....  
Instituição: .....

Presidente: Assinatura: .....  
Nome: .....  
Instituição: .....



Universidade Federal do ABC

Comissão de Ética em Uso de Animais - CEUA  
Avenida dos Estados, 5001 - 1º andar - Torre 1 do Bloco A  
CEP: 09210-170 - Santo André, SP  
Te: (11) 4996-7938  
E-mail: ceua@ufabc.edu.br

## CERTIFICADO

Certificamos que o protocolo registrado sob nº **006/2012** nas fls. **11** do livro **01** para uso de animais em experimentação, sob a responsabilidade da Profa. Dra. Renata Maria Augusto da Costa, Coordenadora da linha de pesquisa **Caracterização da família de reguladores de absorção de ferro em leptospira interrogans Sorovar Copenhageni** do qual participam os pesquisadores: **Josefa Bezerra da Silva; Paulo Lee Ho e Elizabeth A. E. Martins**, está de acordo com os Princípios Éticos de Experimentação Animal adotado pela Sociedade Brasileira de Ciência de animais de Laboratório (SBCAL) e foi aprovada a solicitação de isenção pela COMISSÃO DE ÉTICA EM USO DE ANIMAIS (CEUA) no dia **02.05.2012**.

Com validade a partir de **Abril/2012** à **Abril/ 2015**.

Santo André-SP, 03 de maio de 2012.

Prof. Dra. Renata Simões  
Coordenadora  
CEUA - UFABC

## **AGRADECIMENTOS**

Agradeço aos meus orientadores Paulo Lee Ho e Renata Maria Augusto da Costa por seu empenho e dedicação. Sem as respectivas orientações, este trabalho não teria sido realizado.

Agradeço ao CNPq pela bolsa concedida.

Agradeço aos membros dos laboratórios de biotecnologia I e de organismos patogênicos, do Instituto Butantã e da Universidade Federal do ABC, respectivamente.

Agradeço a Deus pela oportunidade apresentada, e por todas as bênçãos recebidas.

Agradeço aos meus pais e amigos pelo apoio dedicado ao longo destes anos.

E, por fim, agradeço aos Instituto Butantã e à Universidade Federal do ABC, locais onde este trabalho foi realizado.

## RESUMO

MOMO, L. H. S. **Caracterização da família de reguladores de absorção de metais e resposta a estresse em *Leptospira Interrogans* sorovar Copenhageni**. 2015. 101 f. Dissertação (Mestrado em Biotecnologia) – Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2015.

A leptospirose é uma zoonose de importância mundial, e é causada por bactérias patogênicas do gênero *Leptospira*, pertencente à ordem Spirochaetales. Os seres humanos são hospedeiros acidentais e os surtos de leptospirose ocorrem em grandes centros urbanos após enchentes contaminadas por urina de ratos. Existem poucas informações a respeito de como *Leptospira spp* lida com situações de estresse induzidas pelo hospedeiro e pelo ambiente. O ferro é um íon essencial para a maioria dos seres vivos. A regulação de genes envolvidos por seu aporte e estoque na célula bacteriana é mediada por proteínas da família de reguladores transcricionais, Fur (*ferric uptake regulator*). *L. interrogans* sorovar Copenhageni possui quatro ortólogos para Fur, que foram alvo de estudo deste trabalho. A caracterização destes genes foi realizada através de estudos evolutivos, determinação do seu padrão de expressão em modelo animal e análise de modelagem estrutural. Durante o andamento do mestrado, ensaios paralelos revelaram resultados promissores na análise de expressão de genes relacionados ao sistema SOS, um mecanismo de resposta bacteriano a danos no material genético. Assim sendo, o estudo de caracterização de expressão em modelo animal suscetível e resistente à doença foi ampliado. Ensaios de qRT-PCR de cDNAs provenientes de pulmão, rim e fígado permitiram a identificação de dois genes que foram expressos quase que constitutivamente ao longo de toda a infecção em todos os órgãos e organismos estudados: *fur979* e *recA*. Os demais foram requeridos em dias específicos da infecção. Quanto aos componentes do sistema SOS, observamos padrão de expressão específico para o rim, no quinto dia após a infecção. Para os estudos evolutivos de Fur foi gerada uma árvore filogenética que revelou o agrupamento de duas sequências da família Fur de *Leptospira interrogans* sorovar Copenhageni em ramos fechados com sequências muito similares a proteína Fur e Zur de *Escherichia coli*. Os outros dois ortólogos agruparam com as proteínas correspondentes nas demais espécies de *Leptospira*. Uma destas sequências apresentou padrão evolutivo específico dentre as espécies patogênicas. A modelagem da estrutura terciária, confirmou o padrão evolutivo obtido em nossa inferência filogenética.

**Palavras-chave:** Ferric Uptake Regulators. Sistema SOS. Reparo de DNA. Mutagênese. Expressão gênica.

## ABSTRACT

MOMO, L. H. S. **Characterization of the metal absorption regulator family and stress response in *Leptospira interrogans* serovar Copenhageni**. 2015. 101 p. Master thesis (Biotechnology) - Instituto de Ciências Biomédicas, Universidade de São Paulo, São Paulo, 2015.

Leptospirosis is a worldwide zoonosis caused by pathogenic bacteria from the genus *Leptospira*, order Spirochetales. Human beings are accidental hosts, and leptospirosis outbreaks occur in large urban centers after contact with contaminated water by rodent urine. There are few informations concerning the mechanisms employed by *Leptospira spp* to deal with the stress induced by the host and the environment. Iron is an essential ion to most of living beings. The regulation of genes involved in its uptake and maintenance in the bacterial cell is mediated by the transcriptional regulator family proteins, Fur (*ferric uptake regulators*). *L. interrogans* serovar Copenhageni possesses four Fur orthologues, which were the focus of this work. The characterization of *Leptospira* Fur genes was done through evolutive studies, determination of their expression pattern on animal model and structural modeling analysis. In parallel, some experiments presented promising results for the expression analysis of genes related to the SOS system, a bacterial response mechanism to DNA damage. Therefore, the gene expression characterization on susceptible and resistant animal model was amplified. qRT-PCR experiments of cDNA from lung, kidney and liver allowed the identification of two genes expressed almost constitutively during the infection in all organs and organisms: *fur979* and *recA*. The others were required in specific days of the infection. Curiously, the SOS system components showed specific expression pattern in the fifth day after inoculation, in kidney. For the Fur evolutive studies, a phylogenetic tree was inferred, revealing the clustering of two Fur family sequences from *Leptospira interrogans* serovar Copenhageni in closed branches with very similar sequences to Fur and Zur proteins from *Escherichia coli*. The other two orthologues clustered with corresponding proteins in the other *Leptospira* species. One of these sequences presented a specific evolutive pattern among pathogenic species. The tertiary structure modeling confirmed the evolutive pattern obtained in our phylogenetic inference.

**Keywords:** Oxidative stress. Genetic Expression. Mutagenesis. DNA Repair. SOS System. Ferric Uptake Regulators.



## LISTA DE ABREVIATURAS E SIGLAS

µg – Micrograma

ATP – Adenosina trifosfato

cDNA – DNA complementar

ChIP-seq – método de análise de interação de proteínas a DNA

Ct – Cycle threshold (limiar de fluorescência que poder ser detectado)

DNA – Ácido desoxirribonucleico

DPI – Pontos por polegada

FliY – *flagellar motor switch protein*

Fur – *Ferric uptake regulators*

HHpred – Site de buscas baseado na cadeia oculta de Markov

LPS – Lipopolissacarídeo

MAT – Teste de aglutinação microscópica

MIT – Instituto de Tecnologia de Massachusetts

MUSCLE – metodologia de alinhamento de sequências

NCBI – Centro nacional de informações biotecnológicas

Ng – Nanograma

OMS – Organização Mundial da Saúde

PB – Pares de base (Mpb 1000 pares de bases)

PCR – Reação em cadeia pela polimerase

PDB – Banco de dados protéicos

PerR – Fur relacionada a peróxido de hidrogênio

pH – potencial hidrogeniônico

qPCR – PCR quantitativo

rDNA – DNA ribossomal

RNA – Ácido ribonucleico

ROS – Espécies reativas de oxigênio

Toll – “toll like”

Zur – “Zinc uptake regulator”

## LISTA DE TABELAS

TABELA 1: CASOS DE LEPTOSPIROSE NO BRASIL ENTRE OS ANOS 2004 E 2008. ....	15
TABELA 2: SEQUÊNCIAS DE INICIADORES UTILIZADAS NESTE TRABALHO .....	34
TABELA 3: PADRÃO DE EXPRESSÃO DOS QUATRO GENES QUE CODIFICAM PARA FUR NO PULMÃO.....	56
TABELA 4: PADRÃO DE EXPRESSÃO DE QUATRO GENES QUE PARTICIPAM DO SISTEMA SOS NO PULMÃO.....	57
TABELA 5: PADRÃO DE EXPRESSÃO DOS QUATRO GENES QUE CODIFICAM PARA FUR NO RIM .....	57
TABELA 6: PADRÃO DE EXPRESSÃO DE QUATRO GENES QUE PARTICIPAM DO SISTEMA SOS NO RIM. ....	57
TABELA 7: PADRÃO DE EXPRESSÃO DOS QUATRO GENES QUE CODIFICAM PARA FUR NO FÍGADO .....	58
TABELA 8: PADRÃO DE EXPRESSÃO DE QUATRO GENES QUE PARTICIPAM DO SISTEMA SOS NO FÍGADO. ....	58

## LISTA DE FIGURAS

FIGURA 1: MECANISMO DE AÇÃO DE FUR.....	21
FIGURA 2: DISTRIBUIÇÃO DAS PROTEÍNAS FUR EM DIVERSOS ORGANISMOS. ....	22
FIGURA 3: MECANISMO DE AÇÃO DO SISTEMA SOS EM <i>E. COLI</i> . ....	26
FIGURA 4: ÁRVORE FILOGENÉTICA "FURTREE1" .....	39
FIGURA 5: ÁRVORE FILOGENÉTICA "FURSUBTREE1" .....	41
FIGURA 6: ALINHAMENTO ENTRE FUR DE <i>ESCHERICHIA COLI</i> COM AS 4 FURS.....	46
FIGURA 7: ALINHAMENTO ENTRE ZUR DE <i>ESCHERICHIA COLI</i> E AS 4 FURS .....	47
FIGURA 8: RESÍDUOS CONSERVADOS EM FUR3538 .....	48
FIGURA 9: ALINHAMENTO ENTRE 2XIG E AS 4 FURS.....	50
FIGURA 10: PREDIÇÕES DE <i>FUR1128</i> , <i>FUR1972</i> E <i>FUR3538</i> .....	51
FIGURA 11: ALINHAMENTO DAS PREDIÇÕES DAS QUATRO FURS .....	52

## SUMÁRIO

<b>1 INTRODUÇÃO</b> .....	12
1.1 Leptospira.....	12
1.2 Leptospirose .....	13
1.3 Estudos moleculares.....	18
1.4 Processos fisiológicos envolvidos na resposta ao estresse.....	19
1.4 Sistema de resposta a danos ao DNA .....	23
1.5 Modelo animal em leptospirose.....	26
<b>2 JUSTIFICATIVA</b> .....	29
<b>3 OBJETIVOS</b> .....	30
<b>4 MATERIAIS E MÉTODOS</b> .....	31
4.1 Análise Filogenética.....	31
4.2 Análise de Expressão Gênica .....	32
4.2.1 Reações de PCR, desenho e análise de oligonucleotídeos.....	32
4.2.2 Amplificação por PCR em tempo real .....	35
4.2.3 Predição da estrutura terciária dos ortólogos de Fur.....	36
<b>5 RESULTADOS E DISCUSSÃO</b> .....	38
5.1 Inferência filogenética .....	38
5.2 Modelagem da estrutura terciária dos ortólogos de Fur.....	44
5.3 Caracterização do padrão de expressão dos genes da família Fur e de genes componentes do sistema SOS....	53
<b>6 CONCLUSÃO</b> .....	59
<b>REFERÊNCIAS*</b> .....	60
<b>ANEXOS</b> .....	66
Anexo 1: Tabela adicional contendo um total de 52 sequencias coletadas manualmente da árvore FurTree1.....	66
Anexo 2: Alinhamento entre as sequências LIC11158 e LIC12034 .....	67
Anexo 3: Alinhamento entre as sequências LA_1857 e LIC12034 .....	67
Anexo 4: Correlação entre os números de acesso de genes e proteínas utilizadas. ....	68
Anexo 5: Alinhamento entre p71086 e as 4 Furs.....	69
Anexo 6: Genomic survey and expression analysis of DNA repair genes in the genus <i>Leptospira</i> .....	71

# 1 INTRODUÇÃO

## 1.1 Leptospira

A leptospirose é uma zoonose de importância mundial, e é causada por bactérias patogênicas do gênero *Leptospira*, pertencente à ordem *Spirochaetales*. O gênero *Leptospira* agrupa espécies de vida livre e patogênicas, e são bactérias aeróbias, gram-negativas, com alta motilidade, devido à presença de dois endoflagelos localizados em suas extremidades (EVANGELISTA; COBURN, 2010). Suas dimensões celulares são de 6-18 µm de comprimento e 0,1 µm de diâmetro. Foi documentado que em condições de estresse nutricional, suas células podem falhar em se separar durante a divisão celular, resultando em correntes de leptospiras (LODISH, A et al., 1999). Ácidos graxos presentes no ambiente são sua maior fonte de energia, enquanto açúcares não são comumente observados como fonte de carbono (JOHNSON; GARY, 1963).

O gênero *Leptospira* foi inicialmente dividido em duas espécies sorológicas, as patogênicas, agrupadas dentro do complexo “interrogans” (*Leptospira interrogans*) e as demais, agrupadas no complexo “biflexa” (*Leptospira biflexa*) (BHARTI; NALLY et al., 2003). Atualmente, as leptospiras são classificadas de acordo com a heterogeneidade da estrutura na composição dos carboidratos dos lipopolissacarídeos (LPS) da membrana da bactéria (WORLD HEALTH ORGANIZATION, 2009). Existem cerca de 250 sorovares de leptospiras patogênicas e saprófitas descritos em 24 sorogrupos (FAINE; ADLER et al, 1999). Esta classificação fenotípica é amplamente utilizada, devido à sua importância epidemiológica e clínica, embora, não esteja relacionada com a classificação genética atual (ADLER; DE LA PENÃ MOCTEZUMA, 2010; LEVETT, 2003).

Análises filogenéticas baseadas em sequências de rDNA do gene 16S, utilizando *Leptonema* como grupo externo, mostram que as leptospiras são divididas em dois grupos monofiléticos, um deles formado por 8 espécies patogênicas: *L. interrogans*, *L. alexanderi*, *L.*

*borgpetersenii*, *L. kirschneri*, *L. noguchii*, *L. santarosai*, *L. weilii* e *L. alstoni* e o outro grupo é formado por espécies saprófitas; *L. biflexa* e a *L. wolbachii*. Existe ainda um grupo não resolvido filogeneticamente com cinco espécies, dentre as quais, *L. inadai* e *L. fainei*, cujas patogenicidades não foram confirmadas (ADLER et al., 1999; BRENNER; KAUFMANN et al., 1999; KO; GALVÃO REIS et al., 1999).

## 1.2 Leptospirose

A infecção por bactérias patogênicas do gênero *Leptospira* ocorre através de contato direto com a urina do animal, ou indiretamente através de exposição à água contaminada. A leptospirose é mantida através da colonização contínua de túbulos renais proximais em animais hospedeiros. Em geral, cada sorovar de *Leptospira* está adaptado a uma espécie particular de hospedeiro, conhecido também como hospedeiro de manutenção, ou reservatório. Vários mamíferos são conhecidos como importantes reservatórios de *Leptospira spp*, sendo os ratos os principais vetores da leptospirose humana. As chances de infecção do animal aumentam com a idade, porém é possível que ocorra precocemente em seu ciclo de vida. A persistência de focos de leptospirose se deve a estes animais infectados, convalescentes e assintomáticos, os quais se comportam como fonte contínua de contaminação ambiental para as outras espécies animais durante um longo tempo ou por toda a vida. Estes hospedeiros acidentais, tais como os seres humanos, desenvolvem a forma clínica da doença, (FAINE; ADLER et al., 1999; GALVAO R et al., 1999; LEONARD; QUINN et al, 1993).

Os animais domésticos e de criação mais comumente acometidos são cães, bovinos, suínos e eqüinos. As manifestações da leptospirose canina são septicemia e comprometimento hepático e renal. *L. interrogans* sorovar Hardjo (FAINE; ADLER et al., 1999) afeta principalmente bovinos e suínos, sendo que as perdas econômicas estão principalmente relacionadas a aumento nas taxas de aborto, seguida de casos de nascimentos prematuros, natimortos e queda na produção de leite em

bovinos (DA SILVA; CARVALHO et al., 2012). A uveíte é a ocorrência mais comum em eqüinos (FAINE; ADLER et al., 1999).

Os seres humanos são hospedeiros acidentais e os surtos de leptospirose ocorrem em grandes centros urbanos após enchentes contaminadas por urina de ratos. Os principais hospedeiros de manutenção são os roedores sinantrópicos das espécies *Rattus norvegicus* (rato de esgoto) e *Rattus rattus* (rato de telhado), contaminados por *L. interrogans*, a espécie que mais causa leptospirose em seres humanos em todo o mundo (LEVETT, 2003).

A incidência de leptospirose é significativamente mais alta em países tropicais do que em países de clima temperado, pois o clima quente e úmido favorece a sobrevivência das leptospiras. Em países desenvolvidos, a infecção em humanos está principalmente relacionada a exposições ocupacionais, recreativas e esportivas, já que profissionais como veterinários, agricultores e trabalhadores de serviços de água e esgoto, por exemplo, ou praticantes de esportes aquáticos, podem ter contato direta ou indiretamente com a urina de mamíferos hospedeiros (LEVETT, 2003). Em países subdesenvolvidos, a doença está diretamente relacionada às condições inadequadas de moradia, ausência de saneamento básico e à alta infestação por roedores infectados.

No Brasil, a leptospirose é considerada um problema de saúde pública, sendo de notificação compulsória. É bastante disseminada principalmente em épocas de chuvas, pois muitas áreas estão sujeitas a alagamentos e possuem saneamento básico deficiente. São relatados mais de 3.000 casos de leptospirose confirmadas todo ano (veja tabela 1) (REIS; RIBEIRO et al., 2008). Entretanto, estima-se que estas taxas sejam subestimadas. Isto porque seus sintomas são muito parecidos com outras doenças comuns, e a falta de um método de diagnóstico rápido e eficiente dificulta sua identificação. A identificação de leptospiras por testes imunoquímicos depende de exames laboratoriais de qualidade, e do número de organismos presentes no tecido. Em geral, testes imunoquímicos não identificam o sorovar infectante, e os resultados devem ser

interpretados em conjunto com resultados sorológicos (BHARTY; NALLY et al., 2003; KO; GALVAO REIS et al., 1999; LOMAR; DIAMENT et al., 2000).

**Tabela 1: Casos de leptospirose no Brasil entre os anos 2004 e 2008.**

*Casos confirmados de Leptospirose no Brasil de 1997 a 2012.*

Fonte: [http://portal.saude.gov.br/portal/saude/profissional/area.cfm?id\\_area=1562](http://portal.saude.gov.br/portal/saude/profissional/area.cfm?id_area=1562)

**Casos confirmados de Leptospirose. Brasil, Grandes Regiões e Unidades Federadas. 1997 a 2012\***

Região e UF	1997	1998	1999	2000	2001	2002	2003	2004	2005	2006	2007	2008	2009	2010	2011	2012*
<b>Região Norte</b>	<b>484</b>	<b>584</b>	<b>340</b>	<b>391</b>	<b>142</b>	<b>227</b>	<b>248</b>	<b>224</b>	<b>272</b>	<b>752</b>	<b>245</b>	<b>338</b>	<b>359</b>	<b>263</b>	<b>484</b>	<b>184</b>
Rondônia	0	4	0	1	2	4	1	2	4	10	3	17	28	15	56	2
Acre	111	19	2	30	8	18	15	4	18	467	24	39	69	44	122	108
Amazonas	46	14	22	30	28	25	27	34	44	60	47	50	59	41	76	32
Roraima	5	0	0	0	0	0	0	0	0	2	2	4	1	2	0	1
Pará	240	440	227	255	102	167	110	158	164	132	111	138	106	96	131	17
Amapá	80	107	89	73	0	13	91	26	37	80	58	89	95	63	95	24
Tocantins	2	0	0	2	2	0	4	0	5	1	0	1	1	2	4	0
<b>Região Nordeste</b>	<b>847</b>	<b>514</b>	<b>194</b>	<b>1006</b>	<b>651</b>	<b>638</b>	<b>514</b>	<b>807</b>	<b>746</b>	<b>679</b>	<b>569</b>	<b>643</b>	<b>925</b>	<b>719</b>	<b>890</b>	<b>90</b>
Maranhão	20	39	26	9	33	27	20	25	14	52	17	60	62	38	46	4
Piauí	0	0	0	1	0	0	0	0	1	2	0	1	14	0	1	0
Ceará	97	44	33	43	52	52	83	101	61	103	70	92	303	37	104	9
Rio Grande do Norte	15	8	13	18	8	10	10	12	5	9	3	18	40	21	35	5
Paraíba	35	15	4	32	2	18	21	45	17	16	15	14	13	8	22	2
Pernambuco	265	123	32	589	324	307	201	371	335	224	205	193	205	272	383	25
Alagoas	96	31	36	186	70	81	47	98	70	78	49	79	77	72	85	11
Sergipe	61	20	35	54	48	23	12	25	31	41	85	72	56	69	49	10
Bahia	258	234	15	74	114	120	120	130	212	154	125	114	155	202	165	24
<b>Região Sudeste</b>	<b>944</b>	<b>1242</b>	<b>1102</b>	<b>948</b>	<b>1222</b>	<b>957</b>	<b>999</b>	<b>1319</b>	<b>1363</b>	<b>1693</b>	<b>1223</b>	<b>1085</b>	<b>1523</b>	<b>1529</b>	<b>1762</b>	<b>536</b>
Minas Gerais	52	41	43	67	43	44	168	87	94	70	79	80	110	94	108	58
Espírito Santo	21	25	6	20	104	40	27	219	179	298	147	139	237	276	278	89
Rio de Janeiro	470	272	209	206	266	209	245	293	310	265	247	273	317	279	422	92
São Paulo	401	904	844	655	809	664	559	720	780	1060	750	593	859	880	954	297
<b>Região Sul</b>	<b>863</b>	<b>1084</b>	<b>782</b>	<b>1094</b>	<b>1649</b>	<b>907</b>	<b>1192</b>	<b>673</b>	<b>1088</b>	<b>1175</b>	<b>1260</b>	<b>1561</b>	<b>1096</b>	<b>1230</b>	<b>1673</b>	<b>367</b>
Paraná	353	195	247	118	185	251	316	204	334	282	372	197	204	325	448	79
Santa Catarina	180	257	274	168	330	199	306	304	407	346	370	953	427	440	695	167
Rio Grande do Sul	330	632	261	808	1134	457	570	165	347	547	518	411	465	465	530	121
<b>Região Centro-Oeste</b>	<b>160</b>	<b>25</b>	<b>15</b>	<b>48</b>	<b>44</b>	<b>40</b>	<b>52</b>	<b>74</b>	<b>65</b>	<b>70</b>	<b>34</b>	<b>52</b>	<b>45</b>	<b>49</b>	<b>23</b>	<b>17</b>
Mato Grosso do Sul	4	4	3	5	8	5	9	4	15	10	2	7	4	2	0	1
Mato Grosso	2	0	3	4	1	11	3	14	11	12	3	13	1	4	8	1
Goiás	2	3	3	10	7	5	7	16	11	15	7	10	12	14	4	8
Distrito Federal	152	18	6	29	28	18	33	40	28	33	22	22	28	29	11	7
<b>Brasil</b>	<b>3298</b>	<b>3449</b>	<b>2433</b>	<b>3487</b>	<b>3708</b>	<b>2769</b>	<b>3005</b>	<b>3097</b>	<b>3534</b>	<b>4369</b>	<b>3331</b>	<b>3679</b>	<b>3948</b>	<b>3790</b>	<b>4832</b>	<b>1.194</b>

Fonte: Sinan/SVS/MS

\*Dados obtidos em 25 de maio de 2012 sujeitos a alteração

A infecção por *Leptospira* ocorre devido à sua capacidade de rápida invasão dos tecidos do hospedeiro através de abrasões na pele ou por membranas mucosas, tais como mucosas da boca, narina e olhos. Devido à sua grande capacidade de locomoção e invasão, espalham-se rapidamente no organismo pela corrente sanguínea, atingindo vários órgãos, sendo os rins, o pulmão e o fígado os órgãos mais frequentemente envolvidos que, em casos graves, produzem hemorragia e necrose tecidual, resultando na disfunção desses órgãos (FAINE; ADLER et al., 1999). A severidade da doença varia com a espécie de *Leptospira*, com a saúde e a condição imunológica do paciente (CINCO, 2010).

As manifestações clínicas da leptospirose humana ocorrem em duas fases distintas. Depois de um tempo médio de incubação de 7 a 14 dias, a fase aguda, na qual ocorre bacteremia, aparece com sintomas de febre, dores de cabeça intensas, dor muscular e náusea, que persistem por



aproximadamente sete dias. A maioria dos casos de leptospirose apresenta-se sob esta forma leve ou moderada. A seguir, pode ocorrer um período com aparente melhora. Posteriormente, alguns pacientes desenvolvem a fase imune, caracterizada pelo retorno do estado febril, produção de anticorpos aglutinantes e a excreção de leptospiras pela urina. A leptospirose grave raramente apresenta-se bifásica. Inicia-se com sintomas bastante intensos, como dor nas panturrilhas e febre e evolui para doença icterícia com insuficiência renal e hemorragia pulmonar, representando de 5 a 10% dos casos, podendo atingir a taxa de 40% de letalidade (LEVETT, 2003). O contraste entre os sintomas clínicos causados pela doença e a escassez de lesões histológicas sugere que a maioria do dano deve ocorrer em nível celular. A característica mais notável da leptospirose severa é o prejuízo renal e hepático, sendo a falha renal a causa mais comum de morte em indivíduos infectados. A resposta imunológica do hospedeiro à *Leptospira* é caracterizada como a responsável pelas lesões na fase mais tardia desta doença, o que ajuda a explicar a ineficiência do emprego de antibióticos, quando os sintomas da doença estão presentes por mais de quatro dias.

O tratamento do paciente com leptospirose é feito com antibióticos, em especial dexociclina e penicilinas, além de amoxicilina e cefalosporinas que reduzem as chances de evolução para a forma grave, e só são efetivos quando aplicados no início da doença. Os pacientes com leptospirose sem icterícia podem ser tratados em casa, mas aqueles que desenvolvem a forma grave da doença necessitam de internação com tratamento intensivo (BHARTI; NALLY et al., 2003; LEVETT, 2003).

Devido à variedade de sintomas e sua semelhança com doenças de sintomas febris, ela é muitas vezes confundida e diagnosticada como febre amarela, dengue, malária, hepatite e gripe, resultando em medicação errônea. O diagnóstico pode ser confirmado por meio de testes sorológicos como a aglutinação microscópica (MAT), técnica de referência indicada pela Organização Mundial da Saúde (OMS) para leptospirose humana ou animal que emprega suspensões de leptospiras vivas como antígeno (LEVETT, 2003). Normalmente, o MAT demora

até duas semanas para confirmar o diagnóstico clínico da doença. O diagnóstico errado pode levar à progressão da forma mais grave, inclusive com hemorragia pulmonar, que geralmente leva o paciente ao óbito. Pacientes com altas concentrações do patógeno no sangue ( $\sim 10^4$  patógenos  $\text{ml}^{-1}$ ) já são considerados casos graves, geralmente com resultados fatais (SERAIS; TRUCCOLO et al., 2001). Durante este período de produção de anticorpos, as *Leptospiras* são rapidamente eliminadas de muitos órgãos do corpo, porém permanecem latentes no cérebro, nos olhos e nos rins. As que residem no cérebro podem se multiplicar, bem lentamente, enquanto as que sobrevivem nos rins se multiplicam rapidamente e são em parte, eliminadas na urina, consistindo em um caso clínico de constante eliminação da bactéria através da urina, tornando o hospedeiro um possível transmissor da doença (BROWN, 2002). Pode-se chegar ao óbito, mesmo em casos de diminuição de patógenos na circulação sanguínea por ação de antibióticos, caso órgãos-chave sejam seriamente lesados (rins e fígado) (TRUCCOLO; SERAIS et al., 2001).

O controle da leptospirose envolve medidas difíceis de serem adotadas, como a eliminação de roedores e a melhoria das condições de moradia da população. Devido às dificuldades de implementação de tais medidas, houve nos últimos anos um aumento da incidência da doença em países tropicais, o que reforça a necessidade de desenvolvimento de uma vacina eficiente (FAINE; ADLER et al., 1999).

As vacinas comerciais disponíveis até o momento consistem em preparações de culturas de leptospiras de diferentes sorovares de *L. interrogans* inativadas pela ação do calor e/ou formol (bacterinas). São amplamente usadas na pecuária e, licenciadas para uso humano somente em Cuba, Rússia, China e Argentina. Todas apresentam baixa eficiência, pois promovem proteção, usualmente, contra os sorovares presentes na preparação e falham em induzir imunidade de longa duração, o que requer administração anual ou semestral (BHARTI; NALLY et al., 2003; FAINE; ADLER et al., 1999). Além disso, elas contêm uma série de contaminantes oriundos do processo

de obtenção, como componentes do meio e lipopolissacarídeos (LPS), que são associados aos efeitos adversos observados (ADLER; DE LA PEÑA MOCTEZUMA, 2010; LEVETT, 2003).

### 1.3 Estudos moleculares

Devido aos graves danos à saúde e aos custos econômicos que provoca, novos esforços foram desenvolvidos com o objetivo de conhecer melhor a bactéria que transmite a leptospirose. Os projetos genoma chinês e brasileiro foram desenvolvidos como parte de uma grande iniciativa nacional de cara respectivo país, para o desenvolvimento de vacinas eficientes que controlem esta doença. Recentemente, foram sequenciados seis genomas de *Leptospira spp*, sendo de dois sorovares de *L. interrogans* (Lai e Copenhageni) (NASCIMENTO; VERJOVSKI-ALMEIDA et al., 2004; REN; FU, et al., 2003), de duas estirpes de *L. borgpetersenii* sorovar Hardjo-bovis (L550 e JB197) (BULACH; ZUERNER et al., 2006) e de duas estirpes da espécie saprófita *L. biflexa* sorovar Patoc I (Paris e Ames) (PICARDEAU; BULACH et al., 2008). As informações geradas refletem progresso significativo na melhor compreensão da biologia da *Leptospira* e dos prováveis mecanismos empregados em sua patogenicidade.

O genoma do gênero *Leptospira* pode compreender de 3.9 à 4.6 Mpb, dependendo da espécie, e é composto por dois cromossomos circulares nas espécies patogênicas e três na espécie saprófita. O menor genoma corresponde à *L. borgpetersenii*, espécie de vida estritamente patogênica. Alguns genes necessários para responder ao ambiente, assim como transporte de metabólitos e sua utilização foram perdidos nesta espécie, o que justifica sua incapacidade de sobreviver fora do hospedeiro (BULACH; ZUERNER et al., 2006) . Por outro lado, *L. interrogans* não mostra este nível de redução genômica, e como consequência, a bactéria pode sobreviver na água ou lodo por semanas e ainda reter a habilidade em infectar o hospedeiro (NASCIMENTO; VERJOVSKI-ALMEIDA et al., 2004; REN; FU et al., 2003).

Não obstante, a manipulação genética de *Leptospira* ainda não está consolidada. A ausência de plasmídeos naturais nas espécies patogênicas e seu crescimento fastidioso em cultura

dificultam estudos funcionais. Vale ressaltar que o sorovar Copenhageni é o que impõe maiores desafios para sua manipulação (EVANGELISTA; COBURN 2010). Desta forma, os estudos convencionais que empregam mutantes são extremamente limitados nas espécies patogênicas.

Os genomas de *Leptospira interrogans* possuem mais lipoproteínas do que genomas de outras bactérias, com aproximadamente 145 genes que codificam supostas lipoproteínas. Estas lipoproteínas de membrana externa parecem exercer papel importante na virulência das leptospiros nos tecidos de seus hospedeiros. Ensaio de mutagênese realizados até o momento identificaram somente quatro fatores de virulência em *Leptospira*: as proteínas Loa22 (RISTOW; BOURHY et al., 2008), heme oxigenase Murray, Srikrum et al., 2009), *FliY* (*flagellar motor switch protein*) (LIAON; SUN et al., 2009) e LPS (LO; MURRAY et al., 2010). Entretanto, estas informações ainda não são suficientes para traduzir estes avanços em medidas de saúde pública.

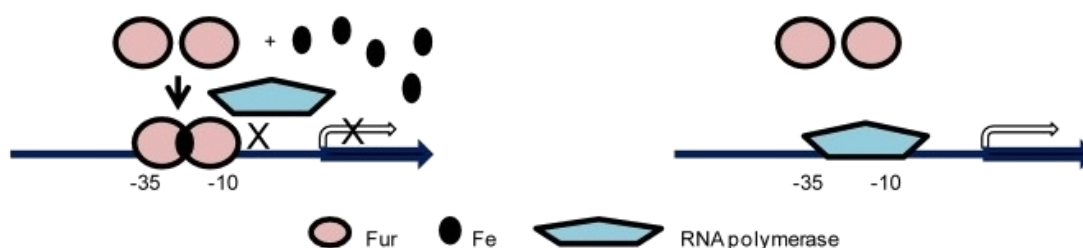
#### **1.4 Processos fisiológicos envolvidos na resposta ao estresse**

Sugere-se que *Leptospira interrogans* esteja extremamente adaptada a mudanças ambientais, visto que seu ciclo de vida apresenta períodos de transição entre o hospedeiro e o ambiente, de maneira sucessiva (BROWN, 2002), fato que pode ser acompanhado por diferenças nos padrões de expressão gênica. Um conhecimento mais aprofundado dos genes regulados pelos diversos parâmetros ambientais, e das respectivas sequências responsáveis por sua regulação, pode fornecer esclarecimentos fundamentais para a patogênese da leptospirose.

Íons metálicos participam em uma gama enorme de funções celulares, incluindo a respiração, atuam como cofatores de enzimas, catalisando funções básicas, são essenciais para a manutenção da pressão osmótica das células e para a composição de estruturas de proteínas. A homeostase intracelular de íons deve ser mantida, não somente por causa dos benefícios destes nutrientes, mas também para proteger as células dos efeitos tóxicos destes metais quando em excesso. O influxo e efluxo de íons metálicos são controlados por uma combinação de mecanismos dependentes de ATP e de prótons (LOUVEL; BOMMEZZADRI et al., 2006).

O ferro é um destes íons essenciais para a maioria dos seres vivos, salvo algumas exceções. Ele é um cofator de diversas enzimas envolvidas em processos biológicos importantes. Entretanto, os níveis de ferro livre no ambiente são baixos devido à sua baixa solubilidade sob condições aeróbicas em pH neutro ou devido ao sequestro de ferro por eucariontes como mecanismo de defesa contra patógenos microbianos (SCHAIBLE; KAUFFMANN, 2004). Além disso, quantidade excessiva de ferro livre intracelularmente é prejudicial às células devido à geração do radical hidroxila via reação de Fenton (IMLAY, 2008). É, portanto, vital para um dado organismo que os níveis intracelulares tanto das espécies reativas de oxigênio (ou *Reactive Oxygen species* - ROS) quanto de ferro (II) sejam estritamente reguladas através de expressão gênica. Desta forma, durante a evolução os organismos vivos desenvolveram uma ampla gama de estratégias para manter os níveis de ferro dentro de um patamar fisiológico.

A disponibilidade de ferro é um estímulo ambiental pelo qual os organismos respondem através da regulação da expressão de inúmeros genes, sendo seu conjunto denominado *regulon* do ferro (BAICHO; WANG et al., 2002). Algumas bactérias patogênicas aproveitam este sinal para coordenar a expressão de genes de virulência (ESCOLAR; PEREZ-MARTIN et al., 1999). Estes genes são regulados por diferentes reguladores transcricionais, dos quais a proteína Fur (do inglês *Ferric Uptake Regulator*) é a melhor caracterizada (HELMANN; LEE, 2007). Fur foi primeiramente descrita como repressora de sistemas de transporte em *Escherichia coli*, em modelo mutante, sensível ao ferro (para revisão, veja Helmann e Lee, 2007). O modelo resultante dos estudos realizados nesta bactéria propõe a repressão de genes alvos para Fur em condições abundantes de ferro no ambiente. Nestas condições, um átomo de ferro capacita a formação de dímeros da proteína Fur, que então, se liga a uma sequência específica de cerca de 19 pb no DNA, chamada *Fur box*, na região promotora dos genes alvo, prevenindo o acesso da RNA polimerase (figura 1). Estas sequências promotoras reconhecidas por Fur variam para cada espécie.



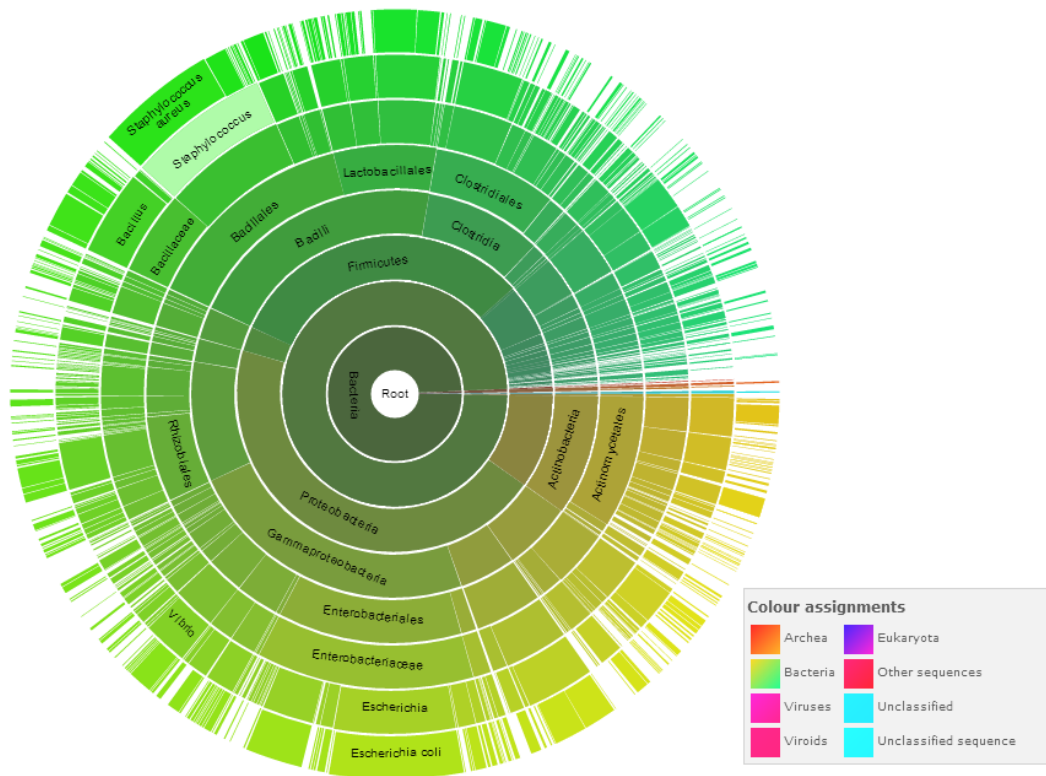
**Figura 1: Mecanismo de ação de Fur.**

*Fur* previne a ação da RNA polimerase ao se ligar à região operadora de seus genes alvo, denominada *Fur box*. Ferro inorgânico se liga preferencialmente a outras proteínas que necessitam de ferro, portanto quando a disponibilidade de ferro intracelular é baixa, os dímeros de *Fur* não são formados, e a sequência promotora não é reprimida, resultando na expressão dos genes regulados por *Fur*. Adaptado de (Liang, Hasturk et al. 2011).

Os genes sob o controle de *Fur* constituem o regulon *Fur*, cujo número varia de acordo com a espécie. Em *E. coli* foram identificados mais de 90 genes (MCHUGH; RODRIGUEZ-QUIMONES et al., 2003), em *Pseudomonas aeruginosa*, 87 (OCHSNER; WILDERMAN et al., 2002) e em *Bacillus subtilis*, 46 (BAICHO; WANG et al., 2002). Além de genes envolvidos com a aquisição e estoque de ferro, *Fur* pode regular genes envolvidos na produção de fatores de patogenicidade, resposta ao estresse ácido e ao estresse oxidativo (TOUATI, 2000).

Na verdade, *Fur* representa uma família de proteínas com estrutura similar, responsável não somente pelo controle da aquisição de ferro, mas em diversas linhagens de bactérias relata-se o controle da aquisição de zinco, níquel, manganês ou cobalto. Alguns membros da família *Fur* de proteínas adquiriram outra função durante a evolução: o sítio de ligação da proteína ao metal é sensível a peróxidos. O primeiro membro identificado foi a proteína *Peroxide Regulon Repressor* (*PerR*) de *Bacillus subtilis* (CHEN; JAMES et al., 1993; BSAT; HELMANN, 1999). Ele é o principal regulador da resposta ao estresse induzido por peróxidos em *B. subtilis* e ortólogos são encontrados tanto em bactérias Gram positivas quanto Gram negativas. Além de *PerR*, *B. subtilis* apresenta dois outros genes que codificam para *Fur*, os quais respondem a diferentes estímulos ambientais e regulam grupos de genes diferentes. *Fur* reprime um regulon de 40 genes em resposta à suficiência ao ferro (BAICHO; WANG et al., 2002), enquanto que *Zur* reprime dois operons que codificam transportadores de zinco (HELMANN; GABALLA, 1998).

As proteínas desta família são pequenas, de massa molecular de cerca de 17 kDa, e análises de bioinformática revelam que nos bancos de dados até a atualidade constam 4483 seqüências com similaridade ao padrão Fur, encontradas, em sua grande maioria em bactérias, sendo pouco encontradas em eucariotos (Figura 2).



**Figura 2: Distribuição das proteínas Fur em diversos organismos.**

Pode-se observar a prevalência das proteínas Fur em bactérias (Verde-amarelado) e suas raras ocorrências em outros grupos (vermelho). (Fonte: <http://pfam.sanger.ac.uk/family/fur#tabview=tab7>, acessado em 2013).

O ferro é um nutriente essencial para o crescimento de *Leptospira*. Os projetos genoma de *L. interrogans* revelaram a presença de quatro ortólogos para Fur (NASCIMENTO; VERJOVSKI-ALMEIDA et al., 2004; REN; FU et al., 2003), entretanto, poucos estudos foram conduzidos até o momento com o objetivo de compreender melhor a atuação destes quatro ortólogos em *Leptospira*. Um único mutante foi obtido através de mutagênese aleatória, com o emprego do transposon *HimarI* (pSC189ColE1) em *Leptospira interrogans* sorovar Lai, cedido pelo Instituto Nacional de Controle e e Prevenção de Doenças (NICDCP), Beijing, China (MOREL; MURRAY et al., 2009). Este mutante é oito vezes mais resistente ao estresse oxidativo, o que foi interpretado

pelos autores como indicação de sua participação no controle da resposta de *Leptospira* a esta condição (LO; MURRAY et al., 2010). Neste mesmo estudo foram conduzidos ensaios de microarranjo para identificar genes responsivos à baixa disponibilidade de ferro e estudar os efeitos desta mutação na expressão gênica. Os genes identificados como responsivos ao ferro na linhagem selvagem não tiveram sua expressão alterada na linhagem mutante, indicando que este ortólogo não participa no controle da homeostase de ferro em *Leptospira*.

#### 1.4 Sistema de resposta a danos ao DNA

Durante seu ciclo de vida, a leptospira está exposta a diferentes condições ambientais que podem danificar seu material genético. Leptospiras, entre outros organismos, possuem um mecanismo de resposta a danos no material genético denominado sistema SOS. Este sistema foi inicialmente identificado em *E. coli* e depois verificou-se estar presente em quase todos os grupos bacterianos. Condições ambientais altamente estressantes, tais como radiação ultravioleta, fatores produzidos pelos hospedeiros, como é o caso das concentrações de oxigênio e ferro e alguns antibióticos podem induzir lesões cromossômicas e ativar o sistema SOS (FRIEDBERG; AGUILERA et al., 2006). A presença de danos no DNA resulta no bloqueio da forquilha de duplicação devido à inabilidade da DNA polimerase em lidar com estas lesões (FRIEDBERG; AGUILERA et al., 2006). Sendo assim, o objetivo primário da resposta SOS é recomeçar a replicação antes que a célula morra.

O sistema SOS é bem documentado em *E. coli* e envolve a participação de mais de 40 genes (COURCELLE; KHODURSKY et al., 2001), cujas expressões são reguladas pelo repressor transcricional LexA e pela recombinase RecA .

O repressor LexA se liga nas adjacências do promotor dos genes controlados (em uma sequência chamada *SOS box*), impedindo a sua transcrição. Em *E. coli*, os genes regulados por LexA estão envolvidos em diferentes processos celulares, tais como a inibição da divisão celular, síntese translesão e reparo por excisão. Quando a célula é submetida a qualquer condição que

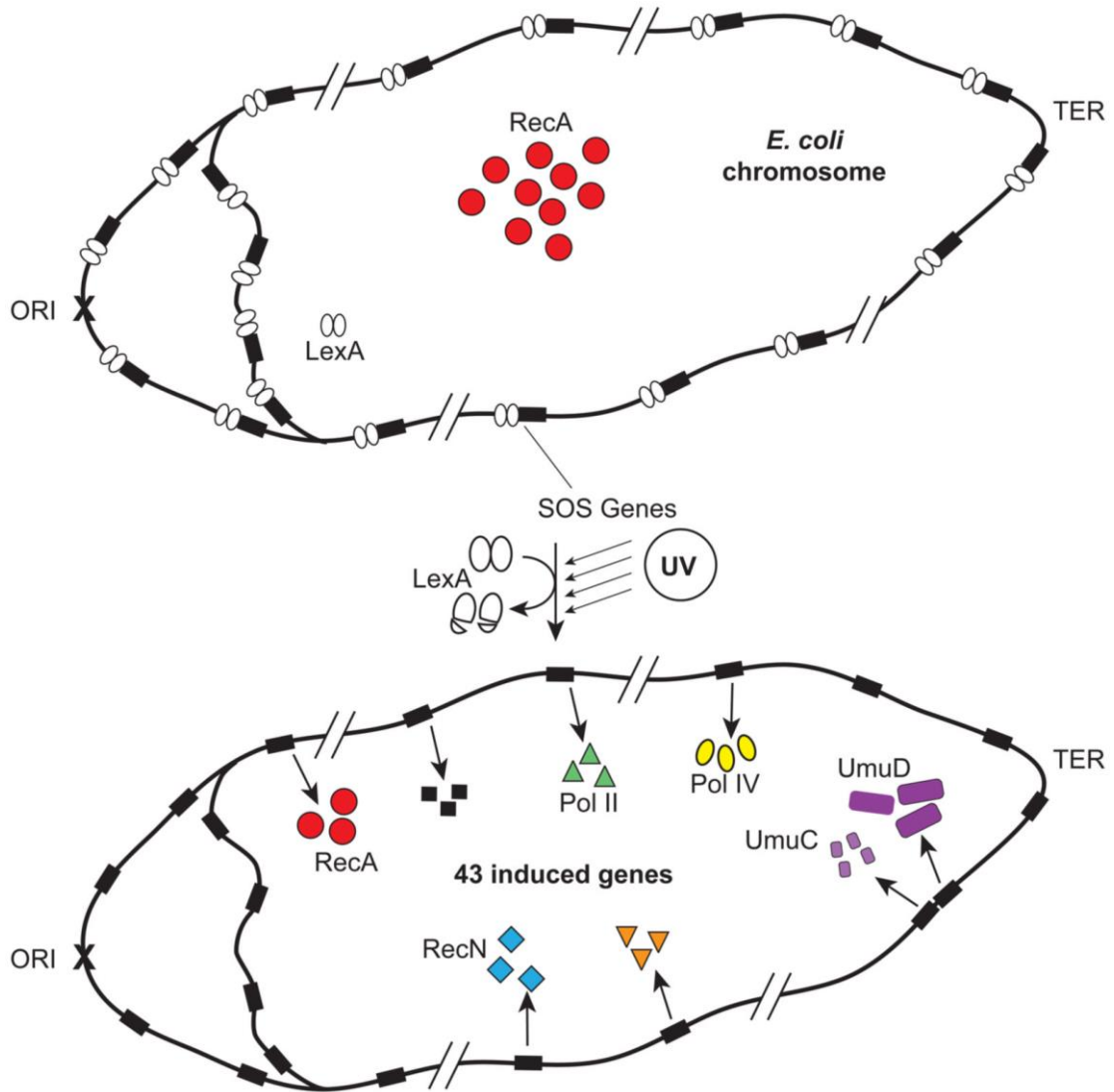


resulte em indução de danos no DNA, a proteína RecA, que atua como sensora do sistema SOS, se liga aos fragmentos de DNA simples fita, gerados por bloqueios da forquilha de duplicação, e adquire seu estado ativo RecA\* (ERILL; CAMPOY et al., 2007). Esta conformação de RecA é responsável por sua associação com LexA, promovendo a clivagem autocatalítica do repressor, que perde a capacidade de ligação ao SOS box (ERILL; CAMPOY et al., 2007). Desta forma, o acesso da RNA polimerase à sequência de DNA é liberado e os genes em questão são transcritos (veja figura 3). Quando os danos ao DNA são reparados, o nível de RecA\* cai, e o sistema volta ao estado não-induzido. As consequências diretas da indução da resposta SOS são a pausa no ciclo celular, a síntese de DNA translesão, o reparo de lesões no DNA e a mobilização de elementos genéticos móveis (AERTSEN; MICHIELS, 2006). As consequências indiretas são o aumento no nível de mutações, rearranjos gênicos e ativação de genes de virulência, de extrema relevância para organismos patogênicos. Além de contribuir para a sobrevivência da célula após situações de estresse, o sistema SOS também desempenha função extremamente importante na evolução procariótica, pois é responsável pela introdução de mutações em ambientes desfavoráveis ao crescimento (CIRZ; O'NEIL et al., 2006).

O gene *lexA* é encontrado em quase todas as bactérias, com raras exceções, entretanto o grupo de genes diretamente reprimido por *lexA* varia consideravelmente. O sistema SOS de *Leptospira interrogans* sorovar Copenhageni é objeto de estudo de nosso grupo. Nossos estudos identificaram a presença de um segundo gene ortólogo para LexA somente em *Leptospira interrogans* sorovar Copenhageni (FONSECA; DA SILVA et al., 2013). Esta segunda cópia foi denominada de *lexA2*, para diferenciar do ortólogo *lexA1*, correspondente ao das demais *Leptospiras*. Estudos realizados por nosso grupo mostraram a ativação do sistema SOS após exposição de leptospira à luz ultravioleta (FONSECA; DA SILVA et al., 2013).. A ativação do sistema SOS foi confirmada após verificação da depleção de LexA1, provavelmente devido à sua atividade auto-proteolítica, ativada nesta condição. Curiosamente, LexA2 apresentou o mesmo

comportamento. Além disso, observamos o aumento da expressão de *lexA1* e *lexA2*, assim como de mais sete genes, *recA*, *recN*, *dinP* e quatro genes que codificam para proteínas hipotéticas. Este grupo de genes é potencialmente regulado por LexA1, pois a proteína mostrou capacidade de ligação nas respectivas regiões promotoras. Por outro lado, LexA2 foi capaz de se ligar somente na região promotora de seu próprio gene.

A caracterização do regulon SOS de *L. interrogans* sorovar Copenhageni se encontra em fase final. Para tal, realizamos ensaios de imunoprecipitação de cromatina seguida por sequenciamento em larga escala (ChIP-seq), a fim de mapear as interações DNA/proteína, em colaboração com o Dr. Alan Grossman do MIT. Os resultados obtidos confirmaram a presença dos nove genes citados acima como membros do regulon SOS de *L. interrogans* e apresentaram mais quatorze genes alvo (FONSECA; DA SILVA et al., 2013).



**Figura 3: Mecanismo de ação do sistema SOS em *E. coli*.**

Sob condições normais de crescimento, os genes SOS são reprimidos pelo repressor transcricional LexA. Quando o DNA é lesado e a forquilha de duplicação é bloqueada RecA assume a conformação ativada, promovendo a autoclivagem catalítica de LexA. Desta forma, o sistema deixa de se apresentar reprimido, e os genes componentes do sistema SOS são transcritos. Acima estão representadas apenas algumas das proteínas transcritas quando o sistema SOS deixa de ser reprimido. (fonte:(Patel, Jiang et al. 2010))

### 1.5 Modelo animal em leptospirose

A ocorrência de uma determinada doença infecciosa envolve uma complexa relação patógeno/hospedeiro. Nesta relação, a resistência do hospedeiro envolve o desencadeamento de uma complexa resposta imunológica, enquanto que o patógeno emprega diferentes fatores de virulência na tentativa de anular o desafio imposto pelo hospedeiro, assim como de promover a colonização dos seus tecidos.

A interação patógeno/hospedeiro em *Leptospira* é pouco compreendida. Até hoje, não existe uma explicação plausível para a grande susceptibilidade de alguns pacientes enquanto que outros são assintomáticos. Com o objetivo de compreender melhor os mecanismos moleculares que *Leptospira* emprega no hospedeiro, iniciamos uma colaboração com o grupo dos professores Elizabeth Martins, Paulo Lee Ho e Josefa Bezerra da Silva, do Instituto Butantã. Este grupo estabeleceu um modelo animal extremamente interessante para estudar estas relações em *Leptospira* (DA SILVA; RAMOS et al., 2009; DA SILVA; CARVALHO et al., 2012). Eles dispõem de linhagens de camundongos com diferentes graus de sensibilidade à bactéria estudada. Camundongos são resistentes à infecção por *Leptospira*, no entanto, a linhagem C3H/HeJ, é mutante para o receptor *Toll4*, o que resulta em sensibilidade, culminando em morte nos animais, em processo infeccioso semelhante ao humano. A linhagem progenitora de C3H/HeJ, denominada C3H/HePas, é relativamente sensível à bactéria, uma vez que adoece, mas não sucumbe à doença.

Dois grupos de doze animais para cada uma das linhagens descritas acima foram inoculados intraperitonealmente com  $10^7$  de *L. interrogans* sorovar Copenhageni. Os camundongos foram acompanhados ao longo de sete dias e a taxa de sobrevivência deles foi calculada (DA SILVA; CARVALHO et al., 2012).

Observou-se que na linhagem C3H/Hej, foram encontrados baixos níveis de quimiocinas MIP-1 $\alpha$ , MCP-1, IL-8, TNF- $\alpha$  especialmente nos primeiros dias de infecção, o que possivelmente sugere a relação destas quimiocinas na susceptibilidade destes animais. Este fato pode ser observado devido ao aumento ou diminuição que foi observado em cada uma das linhagens de camundongos. Algumas quimiocinas sofreram um aumento em determinados órgãos na linhagem C3H/Hej, porém estes eventos não foram suficientes para controlar a doença.

A inoculação dos camundongos por *Leptospira interrogans* sorovar Copenhageni foi realizada no Centro de Biotecnologia-USP, pela Dra. Josefa Bezerra da Silva (CARVALHO; DA

SILVA et al., 2012). Foram utilizados camundongos das linhagens C3H/HeJ, C3H/HePas e BALB/c, todos com 3-4 semanas de idade. As linhagens BALB/c não foram objeto de estudo deste trabalho. As linhagens C3H/Hej, C3H/HePas foram procedentes do departamento de patologia da Faculdade de medicina Veterinária e Zootecnia, e do Instituto de Ciências Biomédicas da Universidade de São Paulo. A linhagem BALB/c foi proveniente do Centro de Criação de Animais do Instituto Butantan. Os procedimentos de manuseio, manutenção e experimentação com animais foram aprovados pelo Comitê de ética em Pesquisa Animal do Instituto Butantan. (Protocolo n °:560/08). Foram realizadas duas inoculações, independentemente, em grupos contendo três animais de cada linhagem. Os órgãos rim, fígado e pulmão foram extraídos nos dias 1, 3, 5 e 7 após a inoculação. O RNA dos órgãos foi extraído e este serviu de molde para reação de síntese de DNA complementar (CARVALHO; DA SILVA et al., 2012). Estas amostras de cDNA, quantificadas e estocadas, foram gentilmente cedidas para nossos estudos.

## 2 JUSTIFICATIVA

Ainda restam muitas informações a respeito de como *Leptospira spp* lida com situações de estresse induzidas pelo hospedeiro e pelo ambiente. Nosso grupo acredita que este conhecimento é de extrema importância, pois os resultados gerados podem contribuir por novas alternativas de tratamento e prevenção da doença. Desta forma, este projeto teve o objetivo de caracterizar a família de proteínas Fur de *Leptospira interrogans*, sorovar Copenhageni. Esta caracterização visa compreender a sua evolução, padrão de expressão em modelo animal e propor mecanismos de atuação na bactéria.

Durante o andamento do mestrado, ensaios paralelos de *Real Time* PCR nas amostras utilizadas de camundongos com os oligonucleotídeos para os genes *lexA* e *recA* do sistema SOS revelaram resultados preliminares promissores. Desta forma, ampliamos as análises de padrão de expressão de alguns componentes do sistema SOS no modelo animal em estudo.

### 3 OBJETIVOS

1. Inferir relações evolutivas sobre os quatro ortólogos de Fur de *Leptospira interrogans* sorovar Copenhageni através de análise filogenética.
2. Analisar o padrão de expressão dos ortólogos de Fur em modelo animal com diferentes níveis de sensibilidade à infecção por *L. interrogans* sorovar Copenhageni.
3. Analisar o padrão de expressão de alguns genes componentes do sistema SOS em modelo animal com diferentes níveis de sensibilidade à infecção por *L. interrogans* sorovar Copenhageni.
4. Caracterizar a estrutura dos ortólogos de Fur de *Leptospira interrogans* sorovar Copenhageni por modelagem da estrutura terciária.

## 4 MATERIAIS E MÉTODOS

### 4.1 Análise Filogenética

Todas as sequências de aminoácidos foram obtidas a partir do banco de dados *Protein Databank* (PDB) (FINN; MISTRY et al., 2010). Através da plataforma pFAM (FINN; MISTRY et al., 2010) foi obtido e organizado um banco de dados local (acessadas em 04/05/2012) em formato Stockholm com um total de 4483 sequências que possuem similaridade superior a 95% com a família Fur, utilizando-se datamining a partir das ferramentas locais disponibilizadas pelo pFAM. Estas sequências foram separadas em grupos por similaridade (*clusters*) utilizando uma taxa mínima de 70% de similaridade através do software CD-HIT. Todas as sequências com similaridade superior a 70% foram condensadas em arquivos diferentes, sem qualquer perda de dados relativos ao procedimento. O resultado parcial foi um arquivo com 1411 sequências com representantes de cada Fur depositada, entretanto que apresentem uma diferença de similaridade de ao menos 30%. Este arquivo foi nomeado como multifasta\_c1.FASTA. Neste, tomou-se o cuidado de procurar entre os mesmos, se havia alguma das sequências-alvo de *L. interrogans* agrupada a outra pelo programa de separação em grupos, que dificultasse o processo posterior de análise. Desta forma, garantiu-se também a presença dos quatro ortólogos dentro do banco de dados.

Para o alinhamento foram utilizados os métodos *MUSCLE* (EDGAR, 2004) e *T-coffee* (NOTREDAME; HIGGINS et al., 2000). Após o agrupamento e alinhamento, as sequências foram reformatadas para um formato MSF (*Multiple Sequences Format*), que permite compatibilidade com o programa *GeneDoc* (JR; NICHOLAS et al., 1997). Paralelamente, estes quatro ortólogos de *L. interrogans* foram alinhados com sequências da família Fur de *E. coli* e *B. subtilis*, a fim de verificar quais regiões são melhor conservadas. As sequências foram editadas manualmente com o emprego do *GeneDoc*, permitindo a retirada de regiões não informativas, com o cuidado de



manter as regiões conservadas dos quatro ortólogos e de outras Fur de interesse baseadas no alinhamento anterior evitando-se assim desemparelhamento do processamento computacional.

O resultado final foi um arquivo chamado multifasta\_f1.FASTA, editado e com 1411 sequências, em formato multifasta. A inferência filogenética foi realizada por máxima verossimilhança através do programa *PhyML* (ANISIMOVA; GASCUEL, 2006; GUINDON; GASCUEL 2003), com uma matriz WAG, suporte de ramos aLRT (ANISIMOVA; GASCUEL, 2006), sem sítios invariáveis, com o melhor de NNI e SPR e árvores iniciais otimizadas como parâmetros. A melhor árvore foi nomeada “*FurTree1*” (veja anexo 5).

Posteriormente, foram geradas subárvores coletando-se manualmente 11 a 12 sequências próximas a cada um dos ortólogos para a inferência de outra árvore filogenética. Além destas, foram adicionadas também as sequências já caracterizadas de *E. coli* e *B. subtilis*, formando-se assim um banco de dados adicional para a segunda análise filogenética. Estas sequências com suas respectivas entradas nos bancos de dados são apresentadas no anexo 2. A inferência filogenética foi feita através de uma inferência pelo método de Máxima verossimilhança (GUINDON; GASCUEL 2003; GOLDMAN; WHELAN, 2001). A melhor árvore foi nomeada *FurSubTree1*, e a tabela com as informações sobre as sequências utilizadas nesta árvore encontra-se nos anexos (anexo 1).

## 4.2 Análise de Expressão Gênica

### 4.2.1 Reações de PCR, desenho e análise de oligonucleotídeos

Os oligonucleotídeos foram desenhados no programa Primer Blast (<http://www.ncbi.nlm.nih.gov/tools/primer-blast/>). Para garantir a integridade dos iniciadores às variações de temperatura obrigatórias das reações de PCR realizadas, foi feita uma análise de formação de estruturas secundárias, através da plataforma *mfold* versão 2.3 (ZUKER, 2003), do RNA Institute (<http://mfold.rna.albany.edu>). A estrutura secundária foi prevista por bioinformática

como uma análise adicional, com o objetivo principal de se analisar complementaridade entre os iniciadores (dímeros ou primer-dimers) e formação de estruturas autoligantes nas extremidades (*hairpins*).

**Tabela 2: Sequências de iniciadores utilizadas neste trabalho**

Sequências de iniciadores utilizadas nos experimentos de PCR em tempo real. Na segunda coluna estão apresentados os números de acesso dos genes referentes a cada iniciador.

Iniciador	Locus Tag	Orientação (5'→3')	Sequência	Fragmento (n° pb)
16s	LIC11010	→	CATGCCGCGGTGAATACGTT	137
		←	CGACTTCACCCCCTTCACGA	
<i>fur1128</i>	LIC11158	→	GTATGTACTCTGTGAAGCGGATC	135
		←	TCCCTCAATAATCCAGCTGAAAC	
<i>fur1972</i>	LIC12034	→	ACCGTACAAAGATTACAAATGGC	106
		←	GAGAAGCGTTTGGCATATGTTC	
<i>fur979</i>	LIC11006	→	GCCACAAACATCACGATCATATC	127
		←	CTATGACCTGTGATTTTGAAACCG	
<i>fur3538</i>	LIC20147	→	TCCGATTCTCTGGACAAAAG	77
		←	CGGAACAAGTCCCATAGAGAATG	
<i>LexA1</i>	LIC12305	→	GCGGACCCGGTAGAACGAAT	129
		←	TGCAATCACCAAGTCCCCATCT	
<i>LexA2</i>	LIC12654	→	TTGTGGCCTGAACGGGAAGT	131
		←	GGCAAAGGGTGCCTACGATC	
<i>recA</i>	LIC11745	→	TTGATTCTGTGGCGGCTTTGG	101
		←	TCTAAGCGCCTGGGACATGAG	
<i>dinP</i>	LIC13052	→	GAAACGCGCAAATCATTTCATGT	133
		←	CAGCCGAAACAACCGATCTACTATT	

A Tabela 2 apresenta as sequências selecionadas e utilizadas neste estudo. A determinação das condições de amplificação de cada par de oligonucleotídeos foi realizada em reação de PCR de gradiente (T100<sup>TM</sup>-Thermal Cycler, Bio Rad) utilizando como molde o DNA genômico de *L. interrogans* serovar Copenhageni. Cada reação de PCR utilizou 100ng de DNA genômico de *Leptospira interrogans* serovar Copenhageni (gentilmente cedidos pela Dra. Josefa B. da Silva).

O volume dos reagentes utilizados seguiu as especificações do fabricante (GoTaq®Hot Start Polymerase, Promega), O ciclo utilizado foi; 94 °C/3 min, seguido de 40 ciclos de 94 °C/30 s, 55-59 °C/30s e 72 °C/ 45s. Uma etapa de extensão de 72 °C/5 min foi realizada ao fim do ciclo. Os produtos de PCR foram visualizados em gel de eletroforese em concentração 2% de agarose.

#### 4.2.2 Amplificação por PCR em tempo real

A análise da expressão gênica foi realizada por PCR em tempo real (qPCR ou PCR quantitativo) usando o equipamento Step One Plus, Applied Biosystems. Em cada placa de tempo real (MicroAmp® Fast 96-Well Reaction Plate, 0.1 mL, Invitrogen) foram utilizados 90 dos 96 poços, com 9 pares de oligonucleotídeos diferentes e 3 amostras em cada placa. São estes nomeados *16s*, *Fur-1128*, *Fur-1972*, *Fur-979*, *Fur-3538*, *LexA1*, *LexA2*, *recA* e *dinP*. Cada reação consistiu de 12,5 µl do reagente goTaq qPCR master Mix (Promega) mais 0,5 µl do corante de referência carboxil-X-rodamina (CXR), 1 µl de água DNase free, 0,5 µl de cada iniciador à concentração de 10 pmol/ µl e 10 µl de cDNA (40 ng) para um volume final de 25 µl por reação. O ciclo utilizado foi: 95 °C/2 min. seguido de 40 ciclos a 95 °C/3 min. e 60 °C seguido da extensão de 95 °C/1 min e o ciclo selecionado foi o *Fast*. A expressão relativa de cada gene selecionado foi analisada usando o método  $2^{-\Delta\Delta Ct}$  descrito por Livak, (LIVAK; SCHMITTGEN, 2001). \*\*\*

Calculou-se inicialmente o  $\Delta Ct$  de cada amostra, utilizando-se de validação por triplicatas, e subsequente normalização, subtraindo-se os valores de Ct do gene controle (*16s*) de Ct do gene alvo. Após determinação do  $\Delta Ct$  normalizado de cada amostra, escolheu-se utilizar uma análise através de cinética dos dias 3, 5 e 7 em relação ao primeiro dia, e o negativo desta diferença foi elevada ao quadrado, para determinação do *fold change*. Os dados para determinação do Ct (*cycle Threshold*) foram coletados na fase exponencial de fluorescência da reação. Para um gene ser considerado como transcrito ele deveria apresentar Ct inferior a 37. Os valores do Ct específico para cada gene foram normalizados com os valores do Ct do gene *16s* analisado com a mesma amostra de cDNA e na mesma corrida. Todas as reações foram analisadas em triplicatas e os resultados representam a média de dois experimentos consecutivos. A eficiência de reação (E) foi calculada para cada par de iniciadores utilizando o programa LingRegPCR (RAMAKERS; RUIJTER et al., 2003), considerando um intervalo ótimo de eficiência entre 1,8 a 2,2, de acordo com o manual do usuário.

#### 4.2.3 Predição da estrutura terciária dos ortólogos de Fur

Foram obtidas as sequências primárias de cada uma das quatro cópias de genes que codificam proteínas da família Fur existentes em *Leptospira interrogans* sorovar Copenhageni em formato FASTA, e estes foram analisados através da ferramenta HHpred (SODING; BIEGERT et al., 2005).

Os códigos de acesso do banco de dados internacional NCBI são: LIC11006, LIC12034, LIC11158 e LIC20147. Todas as sequências analisadas apresentaram a mesma sugestão de modelos. Cada uma das entradas foi analisada individualmente, e a tabela dos resultados obtidos encontra-se no anexo 3. A partir dos modelos selecionados 2xig\_A e 2fe3\_A foi criado, através do programa *Modeller*, um modelo para cada uma das sequências Fur de *Leptospira interrogans* sorovar Copenhageni. Todas as figuras foram exportadas com resolução de 300 pontos por polegada (DPI) (SCHRÖDINGER, 2006).

Para validação dos modelos, foi utilizada a ferramenta de análise por diagrama de Ramachandran através da ferramenta Molprobity (CHEN; ARENDALL et al., 2010) com as configurações os diagramas de Ramchandran, avaliação dos rotâmeros e avaliação da geometria da estrutura.

Os resíduos importantes de membros da família Fur em bactérias foram averiguados.. Três grupos de resíduos foram encontrados para a proteína 2XIG\_A que está relacionada a tomada de zinco e resposta a ácidos (GABALLA; HELMANN, 1998), três grupos de resíduos foram encontrados para P71086, identificada como Fur presente em *Bacillus subtilis* (GABALLA; HELMANN, 1998). As sequências de aminoácidos para 2XIG\_A e P71086 foram alinhadas com as 4 Furs de *L. interrogans* e a partir deste alinhamento, os aminoácidos citados pelos respectivos autores como importantes resíduos foram identificados no alinhamento gerado, seguido de posterior identificação de resíduos correspondentes encontrados em nossas Furs. Somente foram

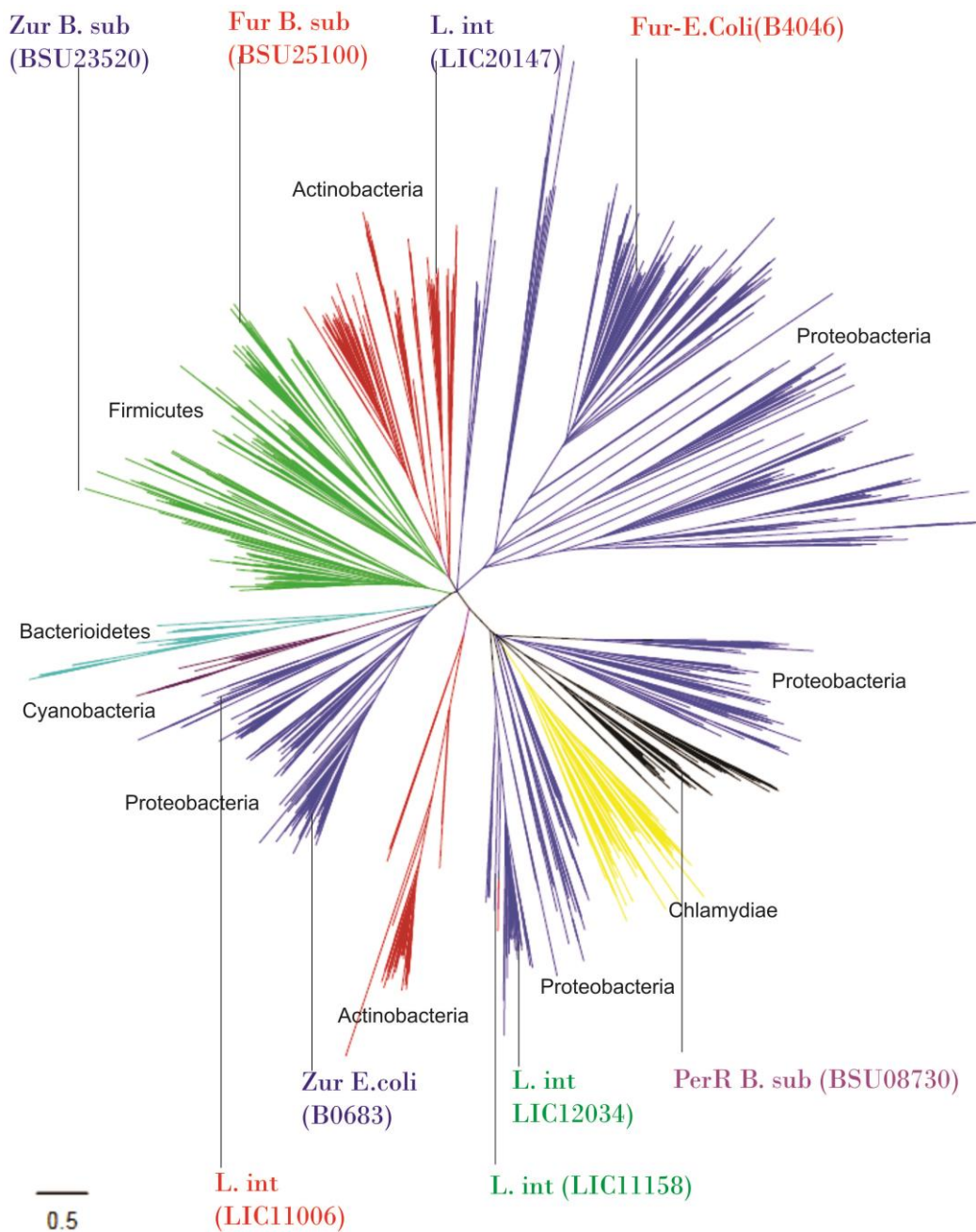
considerados como sítios possíveis de possuir a mesma função aqueles que mantiveram todos os resíduos correspondentes devido a alta importância de cada aminoácido na constituição de cada sítio.

## 5 RESULTADOS E DISCUSSÃO

### 5.1 Inferência filogenética

A árvore filogenética com 1411 sequências obtida por máxima verossimilhança foi nomeada “FurTree1” e é apresentada na Figura 4. Os ramos maiores foram colapsados em relação a porcentagem de organismos de cada filo e coloridos em relação ao filo com a maior porcentagem de representantes.

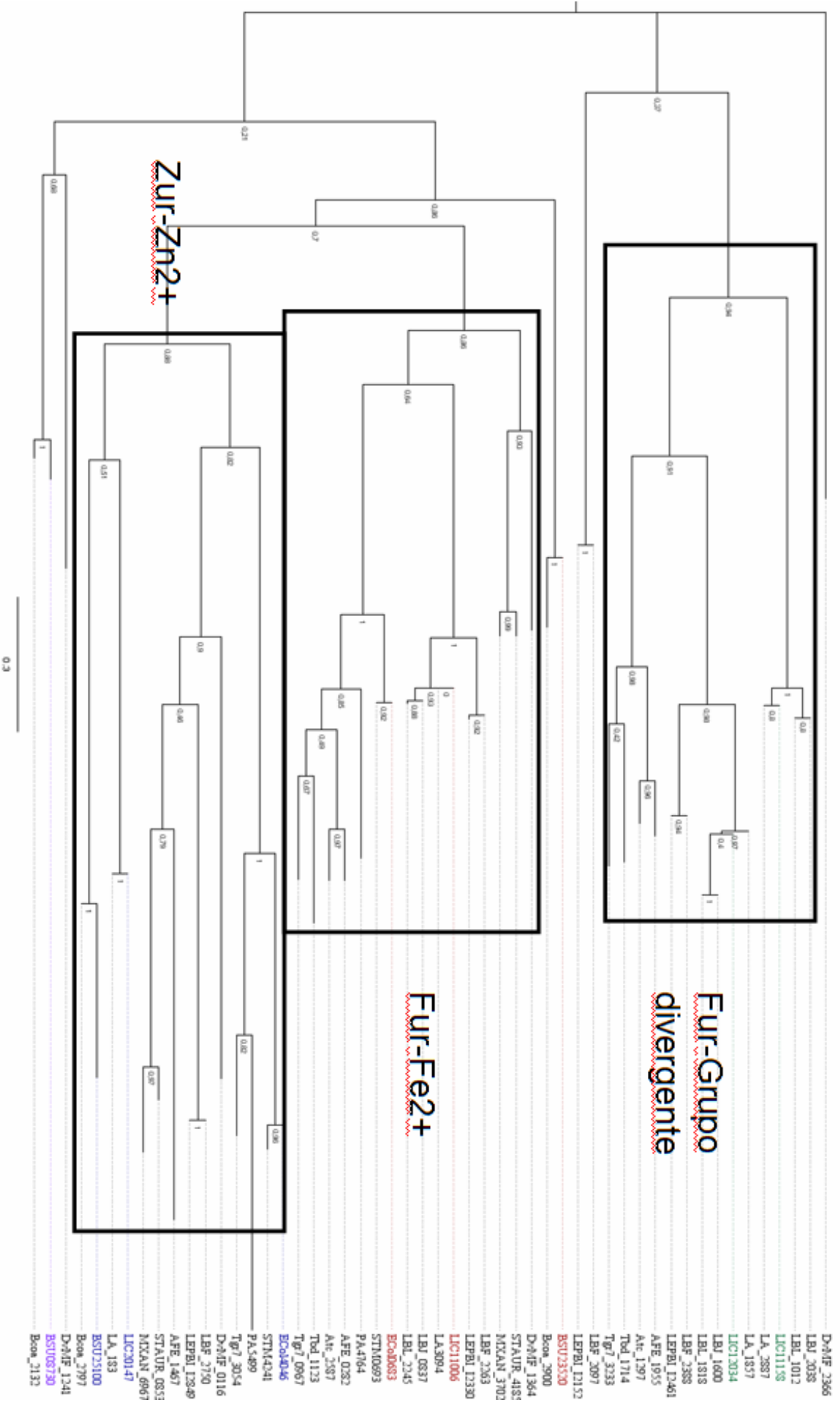
Nesta árvore foi possível observar que todas as sequências de interesse se mostraram distantes entre si, e agrupadas em ramos separados. Pode-se observar que LIC11006 agrupa com b0683, e que LIC12034 estão agrupadas com LIC11158, entretanto o grande número de sequências obtidas dificultou a visualização da relação entre as sequências de interesse, o que nos levou ao próximo passo, a inferência de uma subárvore, utilizando-se de ramos retirados da árvore FurTree1 para compor a árvore “FurSubTree1”, apresentada na Figura 5. Ambas as árvores possuem uma sequência Fur de Archea como raiz.



**Figura 4: Árvore Filogenética “FurTree1”**

Árvore filogenética FurTree 1. A árvore foi inferida a partir da matriz WAG, com as seguintes especificações; suporte de ramos aLRT, sem sítios invariáveis, com o melhor de NNI e SPR e árvores iniciais otimizadas como parâmetros. Os ramos principais foram colapsados e coloridos em relação a maioria de representantes de cada filo. O filo Bacteroidetes se encontra na cor ciano, Cyanobacteria se encontra na cor roxa, o filo Proteobacteria se encontra na cor azul marinho, o filo Actinobacteria se encontra na cor vermelha, o filo Firmicutes se encontra na cor verde e representantes do filo Chlamydiae se encontram em amarelo. Esta divisão por filios bacterianos teve por limiar 80% de representantes em grupos muito heterogêneos e os filios representados em preto apresentaram uma proporção de representantes conflitantes. Em destaque se apresentam as sequências de interesse provenientes de *B. subtilis*, *E. coli* e *L. interrogans* sorovar Copenhageni.





**Figura 5: Árvore filogenética "FurSubTree1".**

Subárvore inferida a partir da matriz WAG, com as seguintes especificações; suporte de ramos aLRT, sem sítios invariáveis, com o melhor de NNI e SPR e árvores iniciais otimizadas como parâmetros. Em azul, encontram-se as sequências provenientes de *Leptospira interrogans* sorovar *Copenhageni*, em vermelho as sequências relacionadas a absorção de ferro, em verde as sequências relacionadas a tomada de zinco e em roxo as sequências relacionadas a peróxido de hidrogênio. Estão retratados através dos retângulos os possíveis agrupamentos encontrados através deste trabalho, com 3 divergências, em um estão agrupadas as sequências relativas a absorção de ferro, em outra, sequências relativas a absorção de zinco e por último sequências sem agrupamento com outras sequências de função conhecida.

A árvore gerada (Figura 5) mostra que duas dentre as quatro sequências da família Fur de *Leptospira interrogans* sorovar *Copenhageni* (LIC11006 e LIC20147) estão agrupadas em ramos fechados com sequências muito similares a proteínas cristalizadas e extensivamente estudadas de *Escherichia coli* e *Bacillus subtilis* (BSAT; HELMANN, 1999; GABRIEL; HELMANN 2009).

Fur foi inicialmente descrita como um sistema de repressão de transporte de ferro, responsivo a ferro em *Escherichia Coli* (HANTKE, 1981; JABOUR; HAMED, 2009). Estes estudos levaram a criação de um modelo conciso para a Fur sob condições de depleção de ferro. Vários estudos validam este modelo, entretanto resultados recentes indicam que a regulação das FURs pode ser muito mais complexa, entre um dos fatores é que Fur consegue utilizar metais diferentes de ferro como co-fatores, e portanto o número de genes regulados que se conhece é maior do que se pensava inicialmente (DELANY; SPOHN et al., 2001; HOLMES; MULHOLLAND et al., 2005). A Fur LIC11006, assim como suas correspondentes nas outras espécies de *Leptospira*, encontra-se próxima filogeneticamente da proteína Fur de *E. coli* (ECol0683), que se liga ao ferro e responde ao seu excesso ou limitação no ambiente (MCHUGH; RODRIGUEZ-QUINONES et al., 2003). O suporte de ramo que agrupa LIC11006 e ECol0683 é de 0,86, valor que garante que estas duas sequências possuem uma grande proximidade filogenética. Examinando-se os resultados experimentais deste trabalho, nota-se que LIC11006 se mostra o gene mais expresso comparativamente, tanto em relação a tecido quanto a animal, com uma expressão mais elevada em rim (veja tabela 6). Neste órgão o estresse imposto sobre a

bactéria se mostra mais intenso de forma a requisitar uma maior expressão desta Fur, e sugere-se, relacionando-se os dados experimentais com os dados filogenéticos, que esta Fur esteja mais expressa devido à necessidade da bactéria de obter ferro do meio externo, em que hospedeiros proporcionam uma menor disponibilidade de ferro em comparação ao ambiente de vida livre.

Outro caso de proximidade filogenética se apresenta com a outra proteína Fur, LIC20147, que agrupou em um ramo com valor igualmente forte para suporte de ramos (0,88). Neste agrupamento estão incluídas as proteínas Zur de *E. coli* e *B. subtilis* (ECol14046 e BSU25100, respectivamente). Como anteriormente, aponta-se uma possível função de LIC20147, em relação a tomada de zinco, da mesma forma que ocorre com LIC11006. Vale ressaltar que este ortólogo está ausente nas duas cepas de *L. borgpetersenii*, provavelmente como resultado do processo de redução genômica desta espécie.

Os outros dois ortólogos para Fur, LIC12034 e LIC11158, agruparam com as proteínas correspondentes nas demais espécies de *Leptospira*. A distribuição das distâncias evolutivas e a topologia da árvore sugerem que um evento de duplicação ocorreu antes da divergência do gênero. Como os dois parálogos são bastante divergentes, pois apresentam somente 32% de identidade, é razoável propormos que este evento de duplicação seja muito antigo, e ao longo do tempo as duas sequências acumularam muitas diferenças. Vale comentar que LIC11158 parece ter sofrido pressão de seleção relacionada ao estilo de vida patogênico, pois o correspondente em *L. biflexa* agrupou em ramo divergente. Como LIC11158 agrupou em um ramo que contém apenas organismos patogênicos, ela é uma boa candidata a estudos futuros sobre FURs relacionadas a patogênese.

Este é um alto indício que esta Fur possua uma pressão evolutiva para se manter em organismos patogênicos, pois a leptospira saprófita (*Leptospira biflexa*), que possui quatro cópias de Fur não possui um ortólogo neste ramo. Entretanto, no ramo abaixo, onde se encontra a LIC12034 encontram-se agrupadas outras três sequências de leptospiros patogênicos, LA1857,

LBJ1600 e LBL1818 além de outras duas sequências de leptospiros saprófitas, LBF2388 e LEPBII2461. Uma hipótese sugerida é o surgimento de um evento de duplicação entre estas duas Fur entre estes dois ramos em um ancestral muito distante, datando da separação entre os dois organismos. Contudo, estas sequências são altamente divergentes (somente 32% de identidade veja anexo 2) ambas possuem o domínio Fur. Estes eventos de duplicação foram provavelmente seguidos de perdas específicas e mutações que modularam a capacidade de resposta de cada organismo para o grupo de pressões evolutivas que caracterizam seus estilos de vida. Sugere-se que a função de LIC12034, bem como das outras sequências deste ramo esteja relacionada a regulação de genes independentes da presença ou ausência de um hospedeiro. Outro ponto importante é que LIC12034 possui uma chance maior de estar relacionada a fatores de patogenicidade devido à presença dos genes homólogos LBJ1600 e LBL\_1818, pertencentes a *leptospiras Hardjo-Bovis* de sorovares diferentes, que infectam principalmente bovinos e não possuem fase de vida livre.

Quanto ao grupo de LIC12034, é provável que estes parálogos exerçam função importante no controle da resposta ao estresse oxidativo. Isto porque o mutante para o gene correspondente em *L. interrogans* sorovar Lai não apresentou deficiência na expressão de genes responsivos ao ferro, além de ser mais resistente ao estresse induzido por peróxido de hidrogênio (LO et al., 2010). Além disso, o mutante em questão diferia da cepa selvagem em relação à maior expressão do gene que codifica para a catalase e de outros envolvidos na via biosintética de heme. Como as duas sequências, LIC12034 e LA\_1857, provém de leptospiros que só são separadas por sorovares diferentes mas possuem alta similaridade entre os genomas (por volta de 99% para rrs de 16S) (NASCIMENTO; VERJOVSKI-ALMEIDA et al., 2004), e o suporte de ramo confirma que neste ramo estas sequências são homólogas (ramo de suporte ALRT 0.98) portanto é muito provável que LIC12034 possua características muito semelhantes a LA1857. Nenhuma das outras sequências incluídas neste ramo possui estudos mais aprofundados em relação à sua função,

entretanto, BSU08730 é uma sequência que codifica PerR de *Bacillus subtilis* muito estudada (JACQUAMET; TRAORE et al., 2009; LEE; HELMANN, 2006) que se encontra em outro ramo, separado de LIC11158. Este distanciamento entre BSU08730 e LIC12034 pode ser devido a um distanciamento filogenético, por diferença de função ou devido à maior proximidade de outras sequências (DvMF\_1241 e Bcoa\_2132) que pode alterar a estrutura da árvore gerada.

Como os ortólogos muito provavelmente retêm suas funções entre as espécies, é razoável argumentarmos que estas análises filogenéticas não só contribuem para melhor compreensão da aquisição dos ortólogos de Fur em leptospira, como permitem uma primeira avaliação da sua provável função exercida. No caso de LIC11006 e LIC20147, devido ao seu agrupamento com ortólogos bem caracterizados, é possível propor que estas proteínas estejam envolvidas na resposta ao ferro e zinco, respectivamente. Este cenário é uma proposta teórica baseada na relação entre as filogenias e a comparação das estruturas preditas com as proteínas com atividade e estrutura já determinadas experimentalmente. Ele carece de uma confirmação experimental, entretanto já aponta diversas sugestões que encaminham hipóteses e experimentos.

## 5.2 Modelagem da estrutura terciária dos ortólogos de Fur

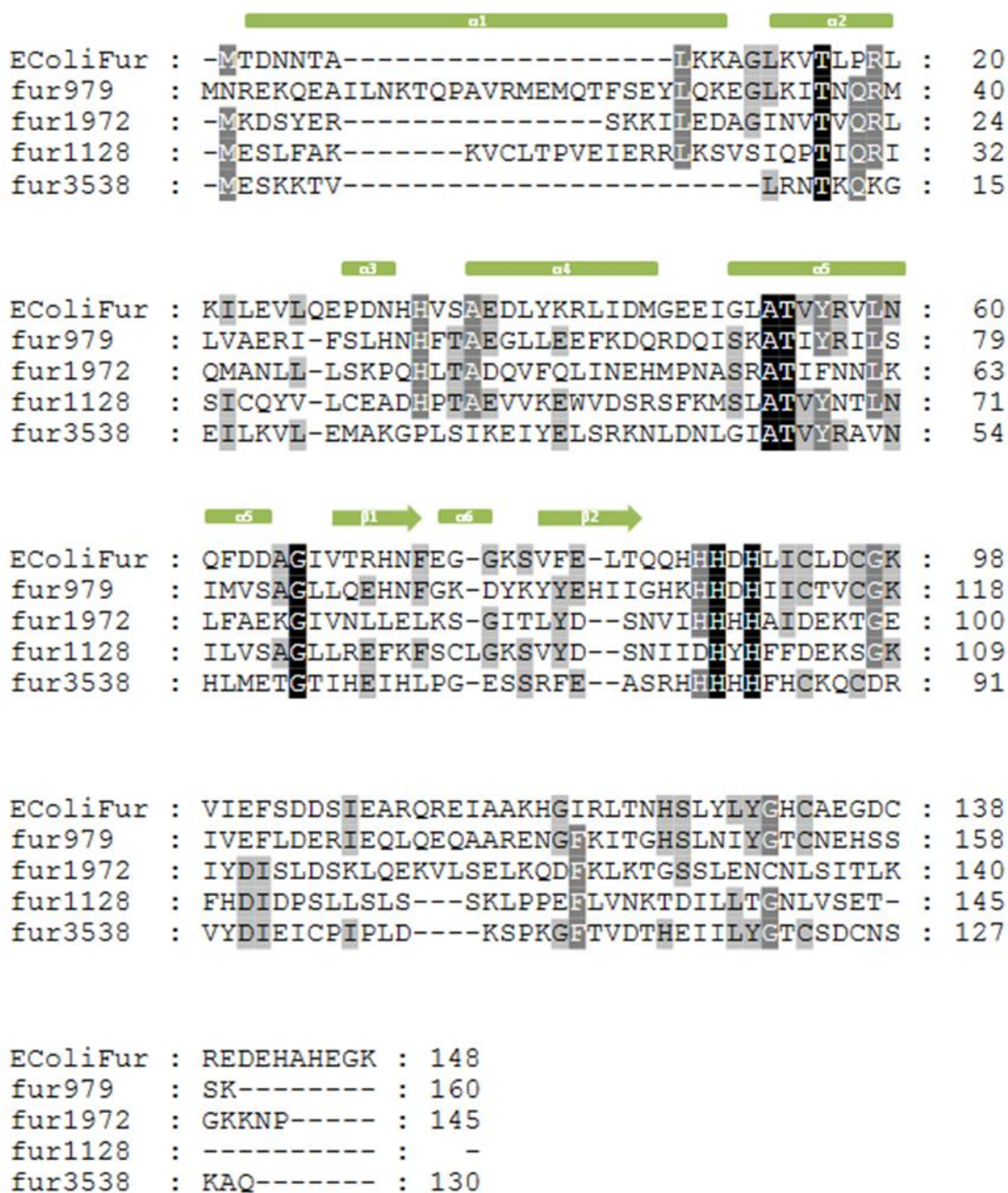
Inicialmente, este trabalho tinha o objetivo de caracterizar a função de um ortólogo para Fur através de sua clonagem e posterior análise de função nas sequências promotoras de prováveis genes alvo. Os quatro ortólogos foram clonados em vetor de expressão pAE e as respectivas expressões foram testadas em diversas linhagens de *E. coli* BL21. Infelizmente, não obtivemos sucesso, pois nenhuma das sequências apresentou expressão em *E. coli*. Desta forma, optamos por uma análise teórica, através de modelagem da estrutura terciária, a fim de verificar se, juntamente com as análises filogenéticas, é possível propor alguma função para os ortólogos. O modelo gerado a partir do gene LIC11006 recebeu a nomenclatura *fur979*, o modelo gerado a partir do gene LIC11158 recebeu a nomenclatura *fur1128*, o modelo gerado a partir do gene LIC12034

recebeu a nomenclatura *fur1972* enquanto o modelo gerado a partir do gene LIC20147 recebeu a nomenclatura *fur3538*.

Existe um consenso, até a presente data, para a existência de três subclasses de funções para as proteínas Fur relatadas na literatura (HALLER; POHL et al., 2003). A subclasse mais estudada está baseada em dados experimentais abundantes de diversos organismos e está relacionada a manutenção de ferro, porém pode também se encontrar envolvido com tolerância a ácidos e proteção a estresse oxidativo (HALL; FOSTER, 1996). Uma segunda classe controla a resposta de bactérias a estresse oxidativo, entretanto sem se envolver na resposta celular a ferro (BSAT; HERBIG et al., 1998; XIONG; SINGH, et al., 2000; ZAHRT; SONG et al., 2001). Uma terceira classe de homólogos de Fur recebe o nome de Zur, e controla o efluxo de zinco em *E.coli* (PATZER; HANTKE, 1998) e *B. subtilis* (GABALLA; HELMANN, 1998).

Para analisar as propriedades de cada uma das sequências, alinhamentos de estrutura primária foram feitos com diversas Furs com trabalhos relatando definição de função através de sítios com resíduos conservados.

A Figura 6 apresenta o alinhamento entre a Fur relacionada a tomada de ferro de *Escherichia Coli* K-12, de número de acesso NP\_415209.1, e as quatro Furs de *Leptospira interrogans* sorovar Copenhageni, alvo de estudo deste trabalho. São mostradas também a estrutura secundária de NP 415209.1, entretanto apenas a região de ligação a DNA foi submetida até o presente momento, portanto apenas ela é mostrada na Figura 6. Pode-se observar que a posição das estruturas terciárias de *E. coli* está de acordo com as previsões geradas neste trabalho, com a exceção de  $\alpha_6$ , que está presente em *E. coli* porém não está presente nas previsões das outras Furs.



**Figura 6: Alinhamento entre Fur de *Escherichia Coli* com as 4 Furs**

Alinhamento feito através do método MUSCLE com a sequência de aminoácidos de *Escherichia coli* (NCBI accession: NP\_415209.1) com as quatro Furs de *Leptospira interrogans* sorovar *Copenhageni*. As regiões em tons de cinza apresentam os aminoácidos conservados entre as sequências. Os tons mais escuros mostram as regiões consenso entre as sequências. Na parte superior, são apresentadas as estruturas secundárias de *EColiFur*, sendo as elipses representações de alfa-hélice e as setas de beta-folhas.

A Figura 7 apresenta o alinhamento entre a Fur relacionada a ligação a zinco de *Escherichia coli* K-12, de número de acesso NP\_415209.1, e as quatro Furs de *Leptospira interrogans* sorovar Copenhageni.

```

EColiZur : MEKTTTQEL-LA-----QAEKICAQRNVRLTPQR : 28
fur979 : MNREKQEAI-LNKTPAVRMEMQTFSEYLOKEGLKITNQR : 39
fur1972 : MKDSYERS-----KKILEDAGINVTVQR : 23
fur1128 : MESLFAKKVCLTP-----VEIERRLKSVSIQPTIQR : 31
fur3538 : MESKK-----TVLRNTKQK : 14

EColiZur : LEVLRIMSLQDGAISAYDILLDLLREAEPOAKPPTVYRALD : 68
fur979 : MLVAERIFSLHNHFTAEGLLLEFKDQRDQISKATIYRILS : 79
fur1972 : LQMANLLLSKPQHLTADQVFOLINEHMPNASRATIENNLIK : 63
fur1128 : ISICQYVLC EADHPTAEVVKKEWVDSRSFKMSLATVYNTLN : 71
fur3538 : GEILKVLEMAKGPLSIKEIYELSRKNLDNLGIATVYRAVN : 54

EColiZur : FLLEQGFVHKVESTN-SYVLCHLEFDQPT--HTSAMFICDR : 105
fur979 : IMVSAGLLQEHNF GK-DYK---YYEHIIGHKHHDHI ICTV : 115
fur1972 : LFAEKGI VNLLELKS-GIT---LYDSNV--IH HHHHAIDEK : 97
fur1128 : ILVSAGLLREFKFSCLGKS---VYDSNI--IDHYHFFDEK : 106
fur3538 : HLMETGTIHEIHLPG-ESS---RFEASR--HHHHHFHCKQ : 88

EColiZur : CCAVKEECAEGVEDIMHTL--AAKMGFALRH-NVIEAHGL : 142
fur979 : CCKIVEFLDERIEQLQEQA--ARENGFKITG-HSLNIYG- : 151
fur1972 : TGEIYDISLD--SKLQEKVLSLSELKQDFKIKTGSSLE---- : 131
fur1128 : SCKFHDIDPS-LLSISLSS----KLPPEFLV NK-TDILLTG- : 139
fur3538 : CDRVYDIEIC---PIPL-D--KSPKGETVDT-HEIILYGT : 121

EColiZur : CAACVEVEACRHPEQCQHDHSVQVKKKPR : 171
fur979 : -----TCN-----EHSSK : 160
fur1972 : -----NCNLSITL-KGKKNP : 145
fur1128 : -----N-----LVSET : 145
fur3538 : CSDC-----NSKAQ : 130

```

**Figura 7: Alinhamento entre Zur de *Escherichia coli* e as 4 Furs**

Alinhamento feito através do método MUSCLE com a sequência de aminoácidos de *Escherichia coli* (NCBI accession: NP\_418470.4) com as quatro Furs de *Leptospira interrogans* sorovar Copenhageni. As regiões em tons de cinza apresentam os aminoácidos conservados entre as sequências. Os tons mais escuros mostram as regiões consenso entre as sequências.

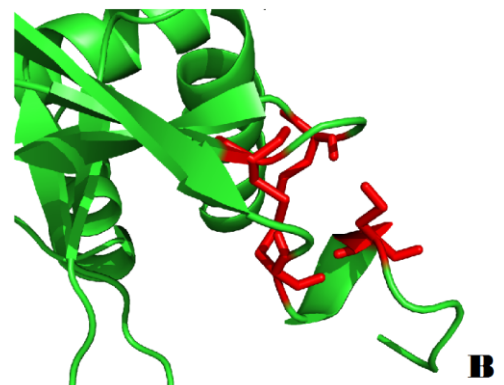


A Figura 8 apresenta o alinhamento por meio do método MUSCLE entre a proteína P71086, em que Hellmann et al (DIAN; VITALE et al., 2011) apresentam diversos grupos de resíduos com funções estudadas em *Bacillus subtilis*. O primeiro grupo de resíduos não possui uma função definida por Hellmann *et al.* porém é sugerido como um sítio de reconhecimento de íons, apresentado na Figura 8A com os aminoácidos destacados em verde em P71086 e seus respectivos aminoácidos também destacados nas quatro Furs de *Leptospira interrogans* sorovar Copenhageni, em que pode se observar que as quatro Furs de *L. interrogans* não compartilham todos os resíduos (veja anexo 6 para a figura completa). O segundo grupo de resíduos não foi completamente estabelecido por Hellmann et al, não se encontra presente nas Furs de *L. interrogans*, e é apresentado na Figura 8A em vermelho. O terceiro grupo de resíduos é sugerido por Hellmann et al como um sítio de ligação a zinco, requerido para a estabilidade da proteína, e encontra-se presente somente em fur3538, e é apresentado na Figura 8A com os aminoácidos destacados em amarelo em P71086 e seus respectivos aminoácidos também destacados nas Furs de *Leptospira interrogans* sorovar Copenhageni. Fur979 também apresenta 3 dos 4 aminoácidos do tipo cisteína, porém devido a importância de cada cisteína em formar a estrutura de ligação a metais, somente fur3538 considera-se passível de reter a função de ligação a zinco.

P71086	F	EV	TS	--D	H	A	C	E	N	C	G	K	I	V	F	H	P	G	L	D	E	V	Q	L	A	A	H	V	T	G	F	K	V	S	--H	R	L								
fur979	Y	E	H	I	G	H	K	H	D	H	I	C	T	V	C	G	K	I	V	E	F	L	D	E	R	I	E	Q	L	Q	E	A	A	R	E	N	G	F	K	I	T	G	--H	S	L
fur1972	Y	S	N	V	I	--H	H	H	A	I	D	E	K	T	G	E	I	Y	I	S	L	D	S	K	L	Q	E	K	V	L	S	E	L	K	Q	D	F	K	L	T	G	S	S	L	
fur1128	Y	S	N	I	I	--D	H	H	F	F	D	E	R	S	G	K	F	H	I	D	P	S	L	L	S	L	S	--S	K	L	P	E	F	L	V	N	K	--T	D	I					
fur3538	F	E	A	S	R	--H	H	H	F	C	K	Q	C	D	R	V	Y	I	E	I	C	P	I	P	L	--D	K	S	P	G	F	T	V	D	T	--H	E	I							
										*	*																																		

P71086	E	I	Y	G	V	C	Q	E	C	S	K	K	E	N	H
fur979	N	I	Y	G	T	C	N	E	--	H	S	S	S	K	
fur1972	E	N	C	N	L	S	I	T	L	K	G	K	K	N	P
fur1128	L	L	T	G	N	--	--	--	--	L	V	S	E	T	
fur3538	I	L	Y	G	T	C	S	D	C	--	N	S	K	A	Q
						*	*								



**Figura 8: Resíduos conservados em fur3538**

(A) Alinhamento entre P71086 de *Bacillus subtilis* e as 4 furs. Esta figura mostra apenas o alinhamento da região 150-165 entre os alinhamentos. Os aminoácidos em destaque são os aminoácidos destacados como essenciais por Hellmann *et al.* Os aminoácidos em verde representam o grupo relacionado a um possível sensor de íons. Os aminoácidos em vermelho são sugeridos como de ligação a ferro em *Pseudomonas aeruginosa*. Os aminoácidos em amarelo estão relacionados a ligação a zinco. Os aminoácidos com um asterisco estão presentes em P71086 e fur3538.

(B) Visão ampliada da hipotética região de ligação a zinco na predição realizada de fur3538. Em vermelho

*apresenta-se a representação em linhas da ligação a metal e em verde a representação caricaturada da predição de fur3538.*

Os resultados observados para fur3538 sugerem que a mesma esteja inserida na classe três citada anteriormente, ou seja, que possua uma provável ligação preferencial a zinco como cofator. Primeiramente, os resultados obtidos pela análise filogenética mostram a proximidade filogenética entre fur3538 com Ecol4046 e BSU25100, representantes de Fur de terceira classe (Figura 5). Segundo, fur3538 é a única Fur de *L. interrogans* que possui os resíduos correspondentes citados por Helmann et al. que estão relacionados a ligação a zinco e estabilidade da proteína (Figura 8A). Ao analisar-se a estrutura dimensional dos ligantes relacionados a estas quatro cisteínas, é possível observar que a estrutura de ligação a metais encontra-se correspondida (Figura 8B). Terceiro, fur3538 está validada como uma representante da família Fur, pois possui os cinco resíduos citados por Schrempf et al., e portanto possui os requisitos para ligação a DNA e apresenta pontes dissulfeto. Ademais, os resíduos C96, H92, H93 estão presentes, que são os resíduos citados por Schrempf et al. e por Normand et al. relacionados a coordenação de zinco. Não se pode afirmar sobre a função de uma proteína apenas através de um resultado obtido por bioinformática, entretanto, neste caso, temos vários resultados convergindo a mesma conclusão. Soma-se a isto os resultados experimentais (Tabela 5 e Tabela 7), em que nota-se que a expressão de fur3538 não possui um padrão claro, porém apenas picos de expressão isolados.

A Figura 9 apresenta o alinhamento por meio do método MUSCLE entre a proteína 2XIG, em que Terradot *et al* (DIAN; VITALE et al., 2011) apresentam diversos grupos de resíduos com funções estudadas em *Helicobacter pylori*. Terradot *et al.* sugerem que o primeiro grupo, destacado na figura em amarelo, possui função regulatória e está presente em fur979 e fur3538. O segundo grupo é sugerido que esteja envolvido em uma função regulatória, e se encontra presente somente em todas as Furs, por exceção de fur3538. Este grupo está apresentado na Figura 9 com aminoácidos destacados em vermelho. O terceiro grupo citado por Terradot et al. é um grupo de

quatro cisteínas, relacionado a regulação em caso de excesso de metais, e está presente somente em fur3538.

```

2XIG MKRLETLESI LER----- ----LRMSIK KNGLKNSKQR EEVSVLYRS
fur979 MNRE-KQEAI LNKTQPAVRM EMQTFSEYLQ KEGLKITNQR MLVAERIFSL
fur1972 MKDSYER--- ----- ----SKKILE DAGINVTVQR LQMANLLLSK
fur1128 MESLFAKVC LTPV----- ---EIERRLK SVSIQPTIQR ISICQYVLCE
fur3538 MES----- ----- ----K KTVLRNTRKQK GEILKVLEMA

2XIG GTHLSPEEIT HSIRQDKNT SISSVYRILN FLEKENFISV LETSKSG-RR
fur979 HNHFTAEGLL EEFKDQRDQI SKATIYRILS IMVSAGLLQE HNFQKDY-KY
fur1972 PQHLTADQVF QLINEHMPNA SRATIFNNLK LFAEKGIVNL LELKSGI-TL
fur1128 ADHPTAEVVK EWVDSRSFKM SLATVYNTLN ILVSAGLLRE FKFSCLGKSV
fur3538 KGPLSIKEIY ELSRKNLDNL GIATVYRAVN HLMETGTIHE IHLPGES-SR

2XIG YEIAAK--EH HDHII CLHCG KII EFAD-PE IENR-QNEVV KKYQAKLI-S
fur979 YE-HIIGHKH HDHIICTVCG KIVEFLD-ER IEQL-QEQAA RENGFKIT-G
fur1972 YD-SNV--IH HHHAI DEKGTG EIYDISLDSK LQEKVLSELK Q--DFKLKTG
fur1128 YD-SNI--ID HYHFFDEKSG KPHDIDP-SL -LSL-SSKLP P--EFLVN-K
fur3538 FE-ASR--HH HHHFHCKQCD RVDIEI-CP -IPL--DKSP K--GFTVD-T

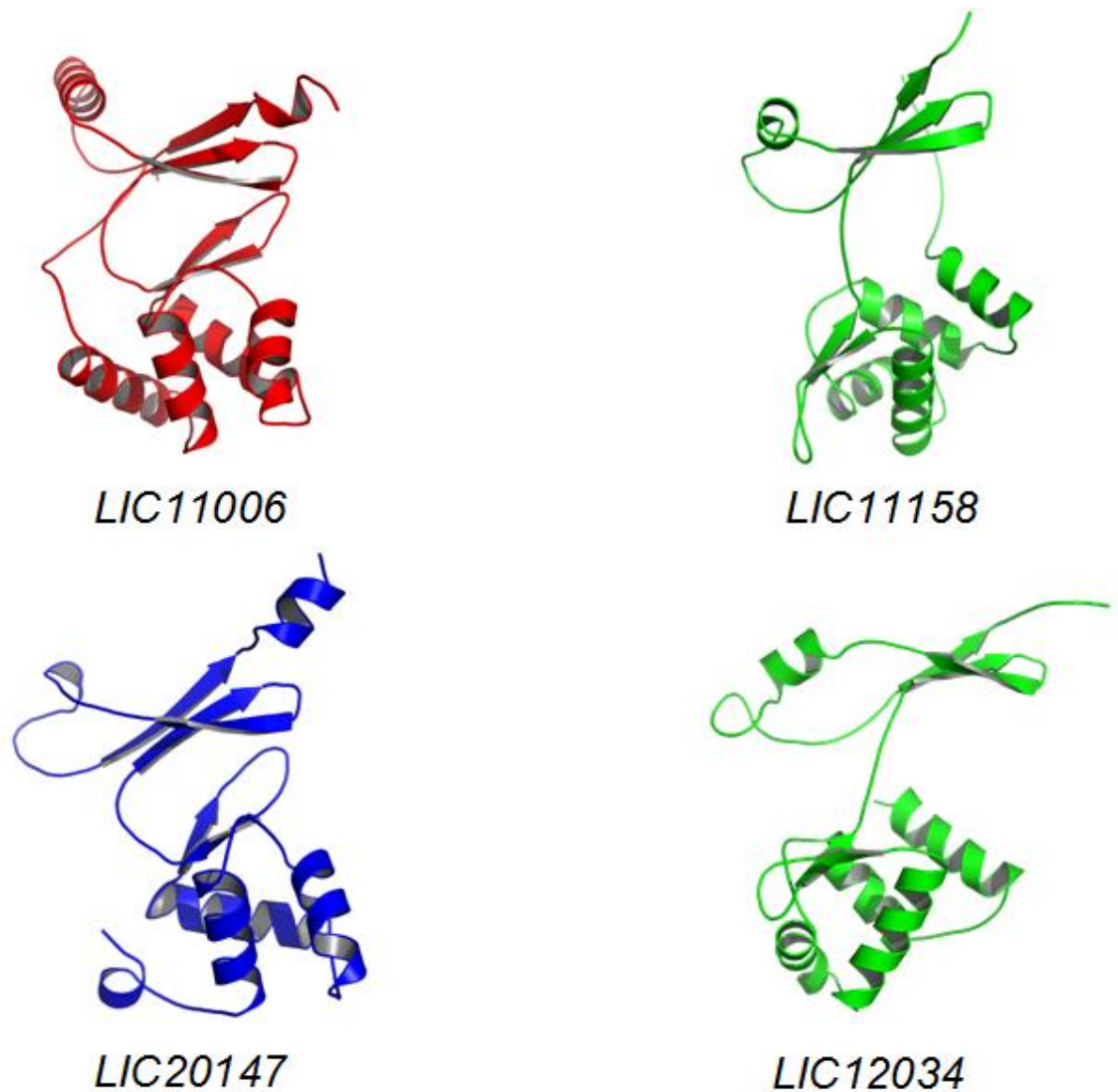
2XIG HDM-KMFVWC KECQESSES
fur979 HSL-NIYGTG NE-HSSSK
fur1972 SSLENCNLSI TLKGKKNP
fur1128 TDI-LLTGN- ---LVSET
fur3538 HEI-ILYGTG SDCNSKAQ

```

### Figura 9: Alinhamento entre 2XIG e as 4 Furs

Alinhamento entre 2XIG de *Helicobacter pylori* e as 4 furs. Os aminoácidos em destaque são os aminoácidos destacados como essenciais por Terradot *et al.* Os aminoácidos em verde representam o grupo relacionado a uma função regulatória. Os aminoácidos em vermelho são sugeridos como de função regulatória. Os aminoácidos em amarelo estão relacionados a excesso de metais.

A Figura 10 apresenta as predições para os quatro ortólogos de Fur de *L. interrogans*. Todas as predições apresentam cinco alfa-hélices e uma região bem definida com duas folhas-beta, como estabelecido como essencial por Normand *et al.*, para a composição de uma estrutura Fur (SANTOS; VIEIRA *et al.*, 2008).

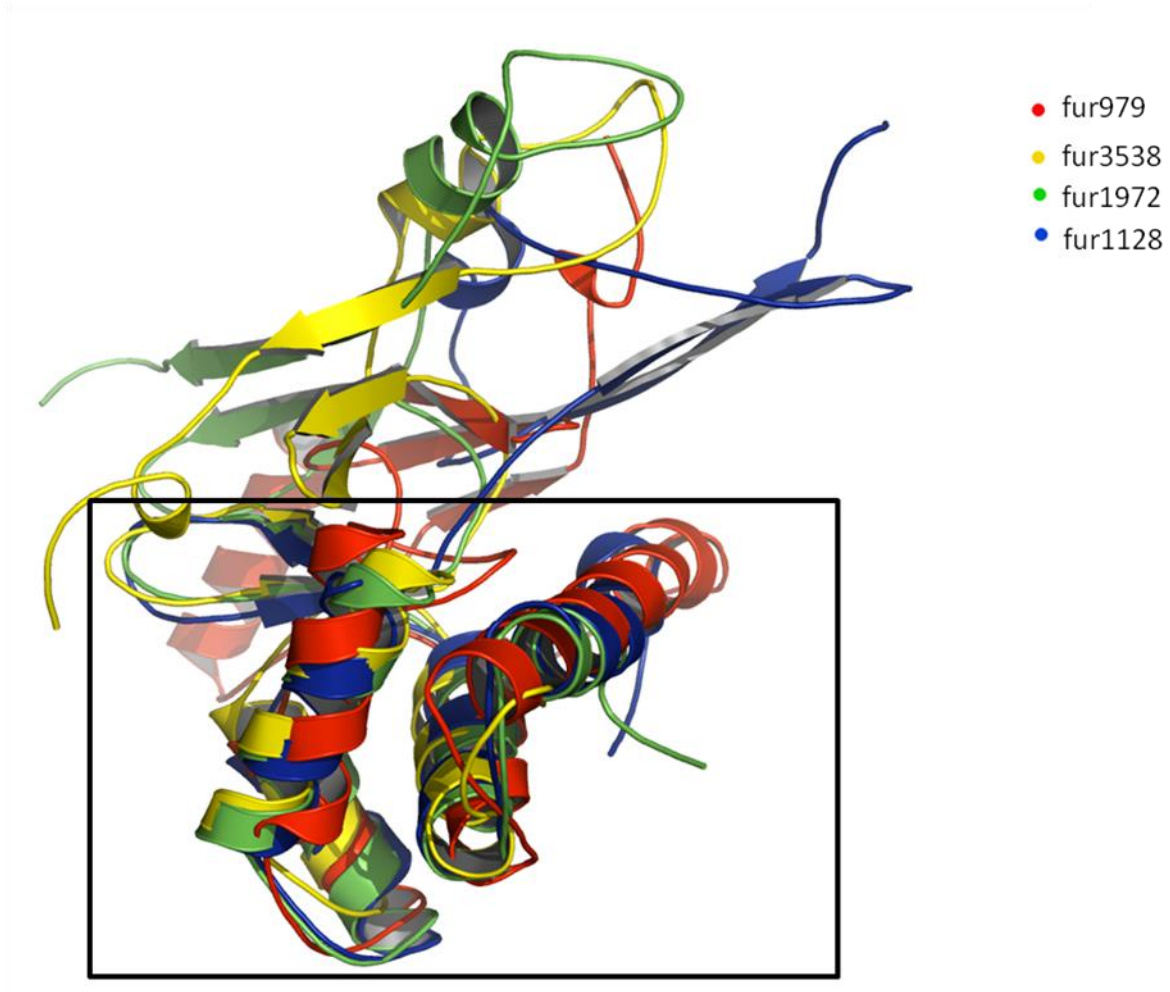


**Figura 10: Predições de fur979, fur1128, fur1972 e fur3538.**

*Predições de estrutura de quatro Furs de Leptospira interrogans sorovar Copenhageni. Todas as estruturas apresentam a região N-terminal na região superior e a C-terminal na região inferior. Todas as predições foram validadas por meio do diagrama de Ramachandran.*

Vale ressaltar a similaridade de estrutura de fur1128 e fur1972. Ambas possuem duas folhas-beta e uma alfa-hélice menor em seu domínio superior, porém os respectivos ângulos se alteram nesta região e criam a diferença de estrutura que é observada. Revela-se com isso uma semelhança geral entre seus enovelamentos, característica de processos de modelagem comparativa.

A Figura 11 apresenta o alinhamento realizado entre as quatro Furs e revela o centro comum a todas e o segundo domínio observado, que difere entre as Furs modeladas. Nesta figura é adicionada *fur979*, que apesar de possuir validação pobre para seu domínio, possui seu centro bem modelado.



**Figura 11: Alinhamento das predições das quatro Furs**

*Alinhamento através do programa pymol das 4 furs preditas neste trabalho. Em destaque apresenta-se a região do centro comum entre as Furs, que se apresenta bem modelada e que encaixa todas as predições. A região superior apresenta o segundo domínio, com conformações semelhantes mas em posições diferentes graças às suas diferenças de ângulos entre o centro e o segundo domínio. Em vermelho, apresenta-se *fur979*, em amarelo, *fur3538*, em verde *fur1972* e em azul, *fur1128*.*

As diferenças observadas no alinhamento entre a Fur presente em *Streptomyces reticuli* e as quatro Furs de *L. interrogans* sorovar *Copenhageni* revelam que dos cinco resíduos relatados por Schrempf et al., apenas três estão presentes em *fur1972* e *fur1128*. É possível que estas duas

Furs não estejam classificadas na definição feita anteriormente de três subclasses de Furs conhecidas, portanto possuam função não-definida já que não apresentam as cisteínas C93 e C94, responsáveis pela formação de pontes dissulfeto. Este resultado exclui estas duas Furs de modelos conhecidos de Furs comprovadamente funcionais *in vivo*. Este resultado também corrobora com nossa análise filogenética, em que *fur1128* e *fur1972* se encontram distantes filogeneticamente das outras duas Furs de *L.interrogans*, e distantes de Furs com funções definidas de subclasses de Fur. *Fur1128* e *fur1972* apresentam-se, portanto, sem função definida e são possivelmente sequências oriundas de duplicações, seguidas de mutações não-sinônimas em resíduos vitais para funções conhecidas, visto que não possuem os resíduos essenciais para funções bem estabelecidas por outras Furs de outros organismos (ver Figura 6, Figura 7 e Figura 9). É possível que com o estudo de outras Furs, ou com uma análise experimental destas duas Furs, seja possível estabelecer uma função para cada uma delas, diferente das funções conhecidas atualmente.

### **5.3 Caracterização do padrão de expressão dos genes da família Fur e de genes componentes do sistema SOS**

A verificação do padrão de expressão dos ortólogos para *fur*, assim como para alguns genes componentes do sistema SOS (*lexA1*, *lexA2*, *recA* e *dinP*) foi realizada através de ensaios de qRT-PCR. Como molde foram empregados os cDNAs provenientes de três órgãos, pulmão, rim e fígado, em modelo animal, extraídos ao longo da infecção (dias 1, 3, 5 e 7). Os ensaios foram realizados em réplica biológica, isto é, duas inoculações independentes (DA SILVA et al., 2009). Não são apresentados os resultados da primeira inoculação para fígado, pois o cDNA produzido apresentou problemas de amplificação. Para cada réplica foram utilizados doze animais para cada linhagem, divididos em grupos de três animais para o mesmo grupo de estudo, por exemplo, pulmão C3H/Hej-dia1. Camundongos deficientes para o receptor *Toll4* (C3H/Hej) são

extremamente sensíveis à doença, e sucumbem à infecção. Por sua vez, a sua linhagem progenitora, C3H/HePas, apesar de adoecer, é capaz de se recuperar.

Os resultados obtidos mostraram que o nível de expressão entre os três animais do mesmo grupo de estudo, é muito variável, dificultando sua análise quantitativa. Portanto, optamos pela realização de análise qualitativa, de forma que consideramos um gene expresso quando seu Ct apresentava valor inferior à 37. A variação da sua expressão ao longo da infecção, isto é, aumento ou diminuição, foi inferida pelo valor de *fold change*, quando presente nos três animais. Valores de *fold change* acima de 1,5 foram considerados significativos para aumento de expressão, enquanto que valores inferiores a 0,5 foram considerados significativos para diminuição da expressão. A análise qualitativa é apresentada nas tabelas 3 à 8. Uma das primeiras observações que destacamos está relacionada à diferenças no padrão de expressão dos genes em suas réplicas biológicas. Isto é, verificamos grande variação no padrão de expressão dos genes entre as duas inoculações. No geral, a inoculação 1 resultou em menor expressão dos genes estudados. Entretanto, algumas conclusões são comuns às duas infecções. Dois genes foram expressos quase que constitutivamente ao longo de toda a infecção em todos os órgãos e organismos estudados: *LIC11006* e *recA*.

*LIC11006* apresentou padrão de expressão constitutiva ao longo de toda a infecção. Entretanto, este gene parece ser recrutado em grandes concentrações mais tardiamente no pulmão de C3H/Hej (dia 3, comparado com dia 3 de Pas) (Tabela 3). Se compararmos os resultados do experimento 2, verificaremos que este gene é muito mais expresso no rim de Pas do que de Hej (Tabela 5). Entretanto, isto não se repete no experimento 1.

Os outros ortólogos de *fur* não apresentaram este padrão de expressão contínuo e houve picos de expressão específicos ao longo da infecção. Observamos que *LIC20147* (Tabela 3) não é expresso em pulmão, pois nenhum dos modelos animais apresentou presença de seu mRNA ao longo da infecção.

Uma diferença importante que observamos comparando os padrões de expressão para Fur nos modelos animais estudados foi a maior expressão dos ortólogos de Fur no sétimo dia após a inoculação nos animais sensíveis, corroborando os resultados clínicos. Camundongos C3H/HePas eliminam as bactérias e C3H/Hej ainda são portadores da bactéria por volta deste período (DA SILVA et al., 2009).

Para os componentes do sistema SOS, encontramos resultados intrigantes. O gene *recA* apresentou um padrão de expressão praticamente constante ao longo da infecção, com exceção de pulmão em C3H/HePas, cujo aumento de expressão foi significativo no último dia de infecção do experimento 1 e dos dias 3-7, do experimento 2 (Tabela 4). Em fígado também houve aumento significativo de expressão para *recA*, entretanto este aumento foi pontual: dia 5 para C3H/Hej (tabela 8). *RecA* é componente do sistema SOS de *L. interrogans*, como demonstrado em estudos realizados por nosso grupo (FONSECA; DA SILVA et al., 2013). Além de controlar a ativação da resposta SOS, RecA desempenha função central na recombinação homóloga (WANG et al., 2008). Desta forma, é possível que sua expressão contínua seja importante para que a bactéria possa completar a duplicação de seu DNA corretamente. Este resultado é consistente com o modelo de infecção causado por UPEC (uropathogenic *Escherichia coli*) (LI et al., 2010). UPEC mutante para *recA* foi incapaz de manter os níveis de colonização equivalentes aos do mutante para *lexA<sub>AT355G</sub>*, indicando que as funções de RecA independentes de LexA são recrutadas durante a infecção. Outra conclusão que podemos associar aos nossos resultados é referente ao controle da expressão de *recA* que não deve estar restrito à LexA,

Este padrão contínuo de expressão não foi observado para os demais componentes do sistema SOS. Pelo contrário, observamos seu recrutamento em períodos específicos, tais como o quinto dia após a inoculação no rim (Tabela 6). Todos os componentes estudados estavam expressos no quinto dia após a infecção, com exceção de C3H/HePas, experimento 1. O que pode



ser explicado pelo fato de termos utilizado um fator bastante restritivo para o Ct. Os resultados brutos obtidos para esta amostra sugerem que estes genes possam estar expressos nesta amostra.

A importância do sistema SOS durante infecção do trato urinário foi demonstrada em modelo animal para UPEC (JUSTICE et al., 2006; LI et al., 2010). Mutantes para *recA*, *lexA* e *sulA* (responsável pela inibição da divisão celular) apresentaram virulência atenuada em camundongos imunocompetentes. Por outro lado, para os camundongos imuno-comprometidos *recA*, *lexA* e *sulA* foram dispensáveis. De acordo com os autores, os radicais oxidativos produzidos pelo sistema imunológico do hospedeiro relacionados à via de sinalização TLR4 ativa a resposta SOS, que é essencial para a viabilidade da bactéria. Curiosamente, em nosso modelo, o sistema SOS também é ativado no camundongo imuno-comprometido. Portanto, em infecção por leptospira outros fatores devem agir na sua indução.

Curiosamente, encontramos indução de genes relacionados ao SOS no pulmão de camundongos sensíveis à doença, sugerindo sua ativação nestes animais no início da infecção (tabela 7). Em fígado, a ativação do SOS não foi verificada (Tabela 8).

Por último, gostaríamos de ressaltar a diferença no padrão de transcrição entre os dois ortólogos para *lexA*. *LexA2* é exclusivo à *L. interrogans* sorovar Copenhageni, e provavelmente foi adquirido por evento de transferência horizontal (FONSECA et al., 2013). Curiosamente, o nível de expressão deste ortólogo foi superior ao de *lexA1* na maioria das amostras estudadas. Não podemos especular a respeito do significado biológico deste achado, mas é curioso observar que um gene presente em região de prófago apresente este padrão de expressão.

**Tabela 3: Padrão de expressão dos quatro genes que codificam para Fur no pulmão.**

*HEJ*: animais sensíveis; *Pas*: resistentes. Os valores são comparativos ao primeiro dia de infecção. *NE*: não expresso; *+*: expresso; *++*: significativo aumento de expressão. *Exp1*: experimento 1; *Exp2*: experimento 2.

PHeJ-EXP1	dia 1	dia 3	dia 5	dia 7	PPas-EXP1	dia 1	dia 3	dia 5	dia 7
LIC11158	NE	NE	+	NE	LIC11158	NE	NE	+	NE
LIC12034	NE	+	+	NE	LIC12034	NE	NE	NE	NE
LIC11006	+	+	+	+	LIC11006	+	++	+	NE
LIC20147	NE	NE	NE	NE	LIC20147	NE	NE	NE	NE
PHeJ-EXP2	dia 1	dia 3	dia 5	dia 7	PPas-EXP2	dia 1	dia 3	dia 5	dia 7
LIC11158	NE	NE	NE	NE	LIC11158	+	NE	NE	NE
LIC12034	+	NE	+	+	LIC12034	+	NE	NE	NE
LIC11006	+	+	+	+	LIC11006	+	+	+	+
LIC20147	NE	NE	NE	NE	LIC20147	NE	NE	NE	NE

**Tabela 4: Padrão de expressão de quatro genes que participam do sistema SOS no pulmão.**

HEJ: animais sensíveis; Pas: resistentes. Os valores são comparativos ao primeiro dia de infecção. NE: não expresso; +: expresso; ++: significativo aumento de expressão. Exp1: experimento 1; Exp2: experimento 2.

PHeJ-EXP1	dia 1	dia 3	dia 5	dia 7	PPas-EXP1	dia 1	dia 3	dia 5	dia 7
LexA1	NE	NE	NE	NE	LexA1	NE	NE	NE	NE
LexA2	NE	NE	NE	NE	LexA2	NE	NE	+	NE
recA	+	+	+	+	recA	+	+	+	++
dinP	NE	NE	+	NE	dinP	NE	NE	NE	NE
PHeJ-EXP2	dia 1	dia 3	dia 5	dia 7	PPas-EXP2	dia 1	dia 3	dia 5	dia 7
LexA1	+	++	NE	NE	LexA1	NE	NE	NE	NE
LexA2	+	+	++	+	LexA2	NE	NE	NE	NE
recA	+	+	+	+	recA	+	++	+	+
dinP	NE	NE	NE	NE	dinP	NE	NE	NE	NE

**Tabela 5: Padrão de expressão dos quatro genes que codificam para Fur no rim**

HEJ: animais sensíveis; Pas: resistentes. Os valores são comparativos ao primeiro dia de infecção. NE: não expresso; +: expresso; ++: significativo aumento de expressão. Exp1: experimento 1; Exp2: experimento 2.

RHeJ-EXP1	dia 1	dia 3	dia 5	dia 7	RPas-EXP1	dia 1	dia 3	dia 5	dia 7
LIC11158	NE	NE	+	NE	LIC11158	NE	NE	NE	NE
LIC12034	NE	NE	+	+	LIC12034	NE	NE	+	NE
LIC11006	NE	+	+	+	LIC11006	NE	+	+	NE
LIC20147	NE	+	+	NE	LIC20147	NE	NE	NE	NE
RHeJ-EXP2	dia 1	dia 3	dia 5	dia 7	RPas-EXP2	dia 1	dia 3	dia 5	dia 7
LIC11158	NE	NE	NE	NE	LIC11158	NE	NE	NE	NE
LIC12034	+	NE	++	NE	LIC12034	+	NE	++	NE
LIC11006	+	++	+	++	LIC11006	+	++	+	++
LIC20147	NE	+	+	NE	LIC20147	NE	+	+	NE

**Tabela 6: Padrão de expressão de quatro genes que participam do sistema SOS no rim.**

HEJ: animais sensíveis; Pas: resistentes. Os valores são comparativos ao primeiro dia de infecção. NE: não expresso; +: expresso; ++: significativo aumento de expressão. Exp1: experimento 1; Exp2: experimento 2.

RHeJ-EXP1	dia 1	dia 3	dia 5	dia 7		RPas-EXP1	dia 1	dia 3	dia 5	dia 7
<i>LexA1</i>	NE	NE	+	NE		<i>LexA1</i>	NE	NE	NE	NE
<i>LexA2</i>	+	+	+	+		<i>LexA2</i>	NE	NE	NE	+
<i>recA</i>	+	+	+	+		<i>recA</i>	+	+	+	++
<i>dinP</i>	NE	NE	NE	NE		<i>dinP</i>	NE	NE	NE	NE
RHeJ-EXP2	dia 1	dia 3	dia 5	dia 7		RPas-EXP2	dia 1	dia 3	dia 5	dia 7
<i>LexA1</i>	+	++	+	NE		<i>LexA1</i>	+	+	+	NE
<i>LexA2</i>	NE	NE	+	NE		<i>LexA2</i>	NE	NE	+	NE
<i>recA</i>	NE	NE	+	NE		<i>recA</i>	NE	NE	+	NE
<i>dinP</i>	NE	NE	NE	NE		<i>dinP</i>	NE	NE	NE	NE

**Tabela 7: Padrão de expressão dos quatro genes que codificam para Fur no fígado**

*HEJ*: animais sensíveis; *Pas*: resistentes. Os valores são comparativos ao primeiro dia de infecção. *NE*: não expresso; +: expresso; ++: significativo aumento de expressão. *Exp1*: experimento 1; *Exp2*: experimento 2.

FHeJ-EXP2	dia 1	dia 3	dia 5	dia 7		FPas-EXP2	dia 1	dia 3	dia 5	dia 7
<i>LIC11158</i>	+	+	+	+		<i>LIC11158</i>	NE	NE	+	NE
<i>LIC12034</i>	+	+	+	NE		<i>LIC12034</i>	+	+	+	NE
<i>LIC11006</i>	+	+	+	+		<i>LIC11006</i>	+	+	+	NE
<i>LIC20147</i>	NE	+	+	NE		<i>LIC20147</i>	NE	NE	NE	NE

**Tabela 8: Padrão de expressão de quatro genes que participam do sistema SOS no fígado.**

*HEJ*: animais sensíveis; *Pas*: resistentes. Os valores são comparativos ao primeiro dia de infecção. *NE*: não expresso; +: expresso; ++: significativo aumento de expressão. *Exp1*: experimento 1; *Exp2*: experimento 2.

FHeJ-EXP2	dia 1	dia 3	dia 5	dia 7		FPas-EXP2	dia 1	dia 3	dia 5	dia 7
<i>LexA1</i>	NE	NE	NE	NE		<i>LexA1</i>	NE	NE	NE	NE
<i>LexA2</i>	NE	NE	NE	NE		<i>LexA2</i>	NE	NE	NE	NE
<i>recA</i>	NE	NE	+	+		<i>recA</i>	NE	NE	NE	NE
<i>dinP</i>	+	+	NE	NE		<i>dinP</i>	+	NE	NE	NE

## 6 CONCLUSÃO

Através da análise filogenética, é possível identificar uma possível função para LIC11006 relacionada ao metal ferro, com o suporte dos resultados experimentais, que mostram que LIC11006, ou *fur979* é a Fur recrutada em *L. interrogans* nas condições experimentais utilizadas.

Através da análise filogenética também é possível identificar uma possível função para LIC20147, relacionada a zinco, que é reafirmado através dos resultados obtidos por alinhamento de modelos e posterior identificação de grupos de resíduos. Portanto, através de análises de bioinformática, é possível se obter uma hipótese forte sobre a possível função de LIC20147.

LIC11158 e LIC12034 se mostraram sequencias altamente diferentes das outras Furs estudadas, e LIC11158 se mostrou um potencial alvo de futuros estudos em relação a Furs relacionadas a patogênese.

Os modelos *fur1128* e *fur1972* se mostraram bastante divergentes de LIC20147, o que concorda com as análises filogenéticas realizadas.

È possível observar em que órgãos dos animais utilizados notou-se a ativação do sistema SOS, e que *recA* mostra-se expresso de maneira constante, enquanto *lexA1* não. Notou-se quais mudanças podem ser realizadas no desenho de futuros experimentos para otimizar estudos moleculares com as amostras de tecidos de modelos animais utilizados neste experimento. Observou-se que existe também um padrão de expressão tecido-específico.

## REFERÊNCIAS\*

- ADLER, B. A.; DE LA PEÑA MOCTEZUMA Leptospira and leptospirosis, **Vet. Microbiol.**, 2010 doi: 10.1016/j.vetmic.2009.03.012. 2010.
- AERTSEN, A.; MICHIELS, C. W. "Upstream of the SOS response: figure out the trigger." **Trends Microbiol.**, v.14, n.10, p.421-423, 2006.
- ANISIMOVA, M.; GASCUEL, O. "Approximate likelihood-ratio test for branches: A fast, accurate, and powerful alternative." **Syst. Biol.**, v.55, n.4, p.539-552, 2006.
- BAICHO, N.; WANG, T. et al., "Global analysis of the Bacillus subtilis Fur regulon and the iron starvation stimulon." **Mol. Microbiol.**, v.45 n. 6, p.1613-1629, 2002.
- BHARTI, A. R.; NALLY, J. E. et al., "Leptospirosis: a zoonotic disease of global importance." **Lancet Infect. Dis.**, vol.3 n.12, p.757-771, 2003.
- BRENNER, D. J.; KAUFMANN, A. F. et al., Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for Leptospira alexanderi sp. nov. and four new Leptospira genomospecies. **Int. J. Syst. Bacteriol.**, vol.49, Pt 2, p.839-858, 1999.
- BROWN, T. A. **Genomes**. New York: Science, 2002, 736 p.
- BSAT, N.; HELMANN J. D. "Interaction of Bacillus subtilis Fur (ferric uptake repressor) with the dhb operator in vitro and in vivo." **J. Bacteriol.**, vol.181, n. 14, p.4299-4307, 1999.
- BSAT, N.; HERBIG, A. et al., "Bacillus subtilis contains multiple Fur homologues: identification of the iron uptake (Fur) and peroxide regulon (PerR) repressors." **Mol. Microbiol.**, vol. 29 n.1, p.189-198, 1998.
- BULACH, D. M.; ZUERNER, R. L. et al., "Genome reduction in Leptospira borgpetersenii reflects limited transmission potential." **Proc. Natl. Acad. Sci. USA**, vol. 103, n.39, p. 14560-14565, 2006.
- CHEN, L.; JAMES, L. P. et al., "Metalloregulation in Bacillus subtilis: isolation and characterization of two genes differentially repressed by metal ions." **J. Bacteriol.**, vol. 175 n.17, p. 5428-5437, 1993.
- CHEN, V. B.; ARENDALL, W. B. et al., "MolProbity: all-atom structure validation for macromolecular crystallography." **Acta Crystallogr. D. Biol. Crystallogr.**, vol. 66, Pt 1, p.12-21, 2010.

\*De acordo com: ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. **NBR 6023**: informação e documentação: referências: elaboração. Rio de Janeiro, 2002.

CINCO, M. "New insights into the pathogenicity of leptospire: evasion of host defences." **New Microbiol.**, vol. 33, n.4, p. 283-292, 2010

CIRZ, R. T.; O'NEILL, B. M. et al., "Defining the *Pseudomonas aeruginosa* SOS response and its role in the global response to the antibiotic ciprofloxacin." **J. Bacteriol.**, vol. 188, n.20, p.7101-7110, 2006.

COURCELLE, J.; KHODURSKY, A. et al., "Comparative gene expression profiles following UV exposure in wild-type and SOS-deficient *Escherichia coli*." **Genetics.** vol. 158, n.1, p.41-64, 2001.

DA SILVA, J. B.; CARVALHO, E. et al., "Induction of TNF- $\alpha$  and CXCL-2 mRNAs in different organs of mice infected with pathogenic *Leptospira*." **Microb. Pathog.**, vol. 52 n.4, p. 206-216, 2012.

DA SILVA, J. B.; RAMOS, T. M. et al., "Chemokines expression during *Leptospira interrogans* serovar Copenhageni infection in resistant BALB/c and susceptible C3H/HeJ mice." **Microb. Pathog.**, vol. 47, n. 2, p. 87-93, 2009.

DELANY, I.; SPOHN, G. et al., "The Fur repressor controls transcription of iron-activated and -repressed genes in *Helicobacter pylori*." **Mol. Microbiol.**, vol. 42, n. 5, p.1297-1309, 2001.

DIAN, C.; VITALE, S. et al., "The structure of the *Helicobacter pylori* ferric uptake regulator Fur reveals three functional metal binding sites." **Mol. Microbiol.**, vol. 79, n. 5, p. 1260-1275, 2011.

EDGAR, R. C. "MUSCLE: multiple sequence alignment with high accuracy and high throughput." **Nucleic Acids Res.**, vol. 32, n. 5, p.1792-1797, 2004.

ERILL, I.; CAMPOY, S. et al., "Aeons of distress: an evolutionary perspective on the bacterial SOS response." **FEMS Microbiol. Rev.**, vol. 31, n. 6, p. 637-656, 2007.

ESCOLAR, L.; PEREZ-MARTIN, J. et al., "Opening the iron box: transcriptional metalloregulation by the Fur protein." **J. Bacteriol.**, vol. 181, n. 20, p. 6223-6229, 1999.

EVANGELISTA, K. V.; COBURN J. "Leptospira as an emerging pathogen: a review of its biology, pathogenesis and host immune responses." **Future Microbiol.**, vol. 5, n. 9, p.1413-1425, 2010.

FAINE, S.; ADLER, B. et al., **Leptospira and Leptospirosis**. Melbourne, Australia: MedSci, 1999.

FINN, R. D.; MISTRY, J. et al., "The pFAM protein families database." **Nucleic Acids. Res.**, vol. 38, p. 211-222, 2010.

FONSECA, L. S.; DA SILVA, J. B. et al., "Leptospira interrogans serovar Copenhageni Harbors Two *lexA* Genes Involved in SOS Response." **PLoS One**, vol. 8, n. 10, p. 76419, 2013.

FRIEDBERG, E. C.; AGUILERA, A. et al., "DNA repair: from molecular mechanism to human disease." **DNA Repair**, v. 5, n. 8, p. 986-996, 2006.

- GABALLA, A.; HELMANN J. D. "Identification of a zinc-specific metalloregulatory protein, Zur, controlling zinc transport operons in *Bacillus subtilis*." **J. Bacteriol.**, v. 180, n. 22, p. 5815-5821, 1998.
- GABRIEL, S. E.; HELMANN J. D. "Contributions of Zur-controlled ribosomal proteins to growth under zinc starvation conditions." **J. Bacteriol.**, v. 191, n. 19, p. 6116-6122, 2009.
- GUINDON, S.; GASCUEL O. "A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood." **Syst. Biol.**, 52, n. 5, p.696-704, 2003.
- HALL, H. K.; FOSTER J. W. "The role of fur in the acid tolerance response of *Salmonella typhimurium* is physiologically and genetically separable from its role in iron acquisition." **J. Bacteriol.**, v. 178, n. 19, 5683-5691, 1996.
- HANTKE, K. "Regulation of ferric iron transport in *Escherichia coli* K12: isolation of a constitutive mutant." **Mol. Gen. Genet.**, v. 182, n. 2, p. 288-292, 1981.
- HOLMES, K.; MULHOLLAND, F. et al., "Campylobacter jejuni gene expression in response to iron limitation and the role of Fur." **Microbiology**, v.151, Pt 1, p.243-257, 2005.
- IMLAY, J. A. "Cellular defenses against superoxide and hydrogen peroxide." **Annu. Rev. Biochem.**, v. 77, p. 755-776, 2008.
- JABOUR, S.; HAMED M. Y. "Binding of the Zn<sup>2+</sup> ion to ferric uptake regulation protein from *E. coli* and the competition with Fe<sup>2+</sup> binding: a molecular modeling study of the effect on DNA binding and conformational changes of Fur." **J. Comput. Aided Mol. Des.**, v. 23, n. 4, p.199-208, 2009.
- JACQUAMET, L.; TRAORE, D. A. et al., "Structural characterization of the active form of PerR: insights into the metal-induced activation of PerR and Fur proteins for DNA binding." **Mol. Microbiol.**, v. 73, n. 1, p.20-31, 2009.
- JOHNSON, R. C.; GARY N. D. "Nutrition of *Leptospira Pomona*. Ii. Fatty Acid Requirements." **J. Bacteriol.**, v. 85, p. 976-982, 1963.
- KO, A. I.; GALVAO REIS, M. et al., "Urban epidemic of severe leptospirosis in Brazil. Salvador Leptospirosis Study Group." **Lancet**, v. 354, n. 9181, p.820-825, 1999.
- LEE, J. W.. HELMANN J. D. "Biochemical characterization of the structural Zn<sup>2+</sup> site in the *Bacillus subtilis* peroxide sensor PerR." **J. Biol. Chem.**, v. 281, n. 33, p. 23567-23578, 2006.
- LEE, J. W.; HELMANN J. D. "Functional specialization within the Fur family of metalloregulators." **Biometals**, v. 20, n. 3-4, p. 485-499, 2007.
- LEONARD, F. C.; QUINN, P. J. et al., "Association between cessation of leptospiruria in cattle and urinary antibody levels." **Res. Vet. Sci.**, v. 55, n. 2, p.195-202, 1993.
- LEVETT, P. N. "Usefulness of serologic analysis as a predictor of the infecting serovar in patients with severe leptospirosis." **Clin. Infect. Dis.**, v. 36, n. 4, p.447-452, 2003.

- LIANG, Y.; HASTURK, H. et al., "Toll-like receptor 2 induces mucosal homing receptor expression and IgA production by human B cells." **Clin. Immunol.** v. 138, n. 1, p. 33-40, 2011.
- LIAO, S.; SUN, A. et al., "Inactivation of the *fliY* gene encoding a flagellar motor switch protein attenuates mobility and virulence of *Leptospira interrogans* strain Lai." **BMC Microbiol.**, v. 9, p. 253, 2009.
- LIVAK, K. J.; SCHMITTGEN T. D. "Analysis of relative gene expression data using real-time quantitative PCR and the  $2^{-(\Delta\Delta C(T))}$  Method." **Methods**, v. 25, n. 4, p. 402-408, 2001.
- LO, M.; MURRAY, G. L. et al., "Transcriptional response of *Leptospira interrogans* to iron limitation and characterization of a PerR homolog." **Infect. Immun.**, v. 78, n. 11, p.4850-4859, 2010.
- LODISH, H. et al., **Molecular Cell Biology**. New York: W. H. F. Co, 1999. 1464 p.
- LOMAR, A. V.; DIAMENT, D. et al., "Leptospirosis in Latin America." **Infect. Dis. Clin. North Am.**, v.14, n. 1, p. 23-39, 2000.
- LOUVEL, H.; BOMMEZZADRI, S. et al., "Comparative and functional genomic analyses of iron transport and regulation in *Leptospira* spp." **J. Bacteriol.**, v.188, n. 22, p.7893-7904, 2006.
- MCHUGH, J. P.; RODRIGUEZ-QUINONES, F. et al., "Global iron-dependent gene regulation in *Escherichia coli*. A new mechanism for iron homeostasis." **J. Biol. Chem.**, v. 278, n. 32, p. 29478-29486, 2003.
- MURRAY, G. L.; MOREL, V. et al., "Genome-wide transposon mutagenesis in pathogenic *Leptospira* species." **Infect. Immun.**, v. 77, n. 2, p.810-816, 2009.
- MURRAY, G. L.; SRIKRAM, A. et al., "*Leptospira interrogans* requires heme oxygenase for disease pathogenesis." **Microbes Infect.**, v. 11, n. 2, p.311-314, 2009.
- NASCIMENTO, A. L.; VERJOVSKI-ALMEIDA, S. et al., "Genome features of *Leptospira interrogans* serovar Copenhageni." **Braz. J. Med. Biol. Res.**, v. 37, n. 4, p. 459-477, 2004.
- NICHOLAS, K. B.; JR., N. H. B. et al., "GeneDoc: Analysis and Visualization of Genetic Variation." **Embnew. News.**, v.4, n. 14, 1997.
- NOTREDAME, C.; HIGGINS, D. G. et al., "T-Coffee: A novel method for fast and accurate multiple sequence alignment." **J. Mol. Biol.**, v. 302, n. 1, p. 205-217, 2000.
- OCHSNER, U. A.; WILDERMAN, P. J. et al., "GeneChip expression analysis of the iron starvation response in *Pseudomonas aeruginosa*: identification of novel pyoverdine biosynthesis genes." **Mol. Microbiol.**, v. 45, n. 5, p. 1277-1287, 2002.
- PATEL, M.; JIANG, Q. et al., "A new model for SOS-induced mutagenesis: how RecA protein activates DNA polymerase V." **Crit. Rev. Biochem. Mol. Biol.**, v. 45, n. 3, p.171-184, 2010.
- PATZER, S. I.; HANTKE K. "The ZnuABC high-affinity zinc uptake system and its regulator Zur in *Escherichia coli*." **Mol. Microbiol.**, v. 28, n. 6, p.1199-1210, 1998.



PICARDEAU, M.; BULACH, D. M. et al., "Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis." **PLoS One**, v. 3, n. 2, p. 1607, 2008.

POHL, E.; HALLER, J. C. et al., "Architecture of a protein central to iron homeostasis: crystal structure and spectroscopic analysis of the ferric uptake regulator." **Mol. Microbiol.**, v. 47, n. 4, p. 903-915, 2003.

RAMAKERS, C.; RUIJTER, J. M. et al., "Assumption-free analysis of quantitative real-time polymerase chain reaction (PCR) data." **Neurosci. Lett.**, v. 339, n. 1, p. 62-66, 2003.

REIS, R. B.; RIBEIRO, G. S. et al., "Impact of environment and social gradient on *Leptospira* infection in urban slums." **PLoS Negl. Trop. Dis.**, v. 2, n. 4, p. 228, 2008.

REN, S. X.; FU, G. et al., "Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing." **Nature**, v. 422, n. 6934, p. 888-893, 2003.

RISTOW, P.; BOURHY, P. et al., "Biofilm formation by saprophytic and pathogenic leptospires." **Microbiology**, v. 154, Pt 5, p.1309-1317, 2008.

SANTOS, C. L.; VIEIRA, J. et al., "On the nature of fur evolution: a phylogenetic approach in Actinobacteria." **BMC Evol. Biol.**, v. 8, p. 185, 2008.

SCHAIBLE, U. E.; KAUFMANN S. H. "Iron and microbial infection." **Nat. Rev. Microbiol.**, v. 2, n. 12, p.946-953, 2004.

SCHRÖDINGER, L. "The PyMOL Molecular Graphics System, Version 1.5.0.4 Schrödinger, LLC."

SODING, J.; BIEGERT, A. et al., "The HHpred interactive server for protein homology detection and structure prediction." **Nucleic Acids. Res.**, v. 33, p. 244-248, 2005.

TOUATI, D. "Iron and oxidative stress in bacteria." **Arch. Biochem. Biophys.**, v.373, n. 1, p. 1-6, 2000.

TRUCCOLO, J.; SERAIS, O. et al., "Following the course of human leptospirosis: evidence of a critical threshold for the vital prognosis using a quantitative PCR assay." **FEMS Microbiol. Lett.**, v. 204, n. 2, p. 317-321, 2001.

WHELAN, S.; GOLDMAN N. "A General Empirical Model of Protein Evolution Derived from Multiple Protein Families Using a Maximum-Likelihood Approach." **Mol. Biol. Evol.**, v. 18, n. 5, p.691-9, 2001.

WORLD HEALTH ORGANIZATION. **Website da Organização Mundial de Saúde.** Disponível em <http://www.who.int/en/> (página em inglês) Acesso em 23 de Junho de 2009.

XIONG, A.; SINGH, V. K. et al., "Molecular characterization of the ferric-uptake regulator, fur, from *Staphylococcus aureus*." **Microbiology**, v. 146, Pt 3, p. 659-668, 2000.

ZAHRT, T. C.; SONG, J. et al., "Mycobacterial FurA is a negative regulator of catalase-peroxidase gene katG." **Mol. Microbiol.**, v.39, n. 5, p.1174-1185, 2001.

ZUKER, M. "Mfold web server for nucleic acid folding and hybridization prediction." **Nucleic Acids. Res.**, v. 31, n. 13, p. 3406-3415, 2003.

## ANEXOS

### Anexo 1: Tabela adicional contendo um total de 52 sequencias coletadas manualmente da árvore FurTree1.

*O primeiro campo (Organismos) da tabela mostra os organismos inclusos, e o último campo (Código) mostra a nomenclatura utilizada no arquivo multifasta. Cada campo de formato fasta teve de ser renomeado com um código, devido a um requerimento do programa PhyML.*

Organismos	Strain	Número de cópias	Locus tag	YP	Código
Leptospira interrogans serovar Copenhageni	Fiocruz L1-130	4	LIC11006	YP_000979.1	LIC11006
			LIC12034	YP_001972.1	LIC12034
			LIC11158	YP_001128.1	LIC11158
			LIC20147	YP_003538.1	LIC20147
Leptospira interrogans serovar Lai	56601	4	LA_3094	NP_713274.1	LA_3094
			LA_1857	NP_712038.1	LA_1857
			LA_2887	NP_713068.1	LA_2887
			LB_183	NP_714727.1	LA_183
Leptospira biflexa serovar Patoc	Patoc 1 (Ames)'	4	LBF_2388	YP_001963456.1	LBF_2388
			LBF_2750	YP_001963808.1	LBF_2750
			LBF_2263	YP_001963334.1	LBF_2263
			LBF_2097	YP_001963170.1	LBF_2097
Leptospira biflexa serovar Patoc	Patoc 1 (Paris)	4	LEPBI_12461	YP_001839826.1	LEPBI_12461
			LEPBI_12152	YP_001839530.1	LEPBI_12152
			LEPBI_12330	YP_001839699.1	LEPBI_12330
			LEPBI_12849	YP_001840195.1	LEPBI_12849
Leptospira borgpetersenii serovar Hardjo-bovis	JB197	3	LBJ_1600	YP_800923.1	LBJ_1600
			LBJ_0837	YP_800251.1	LBJ_0837
			LBJ_2038	YP_801300.1	LBJ_2038
Leptospira borgpetersenii serovar Hardjo-bovis	L550	3	LBL_1818	YP_798193.1	LBL_1818
			LBL_2245	YP_798581.1	LBL_2245
			LBL_1012	YP_797475.1	LBL_1012
Escherichia coli	K-12 substr. MG1655	2	b0683	NP_415209.1	ECol0683
Bacillus subtilis	168	3	b4046	NP_418470.4	ECol4046
			BSU23520	NP_390233.2	BSU23520
			BSU25100	NP_390389.2	BSU25100
Acidithiobacillus ferrooxidans	ATCC 23270	3	BSU08730	NP_388753.1	BSU08730
			AFE_0282	YP_002424788.1	AFE_0282
			AFE_1955	YP_002426362.1	AFE_1955
Pseudomonas aeruginosa		1 ou 2	AFE_1467	YP_002425895.1	AFE_1467
			PA4764	NP_253452.1	PA4764
			PA5499	NP_254186.1	PA5499
Bacillus coagulans	36D1	3	Bcoa_2797	YP_004860753.1	Bcoa_2797
			Bcoa_2132	YP_004860096.1	Bcoa_2132
			Bcoa_2900	YP_004860856.1	Bcoa_2900
Salmonella enterica subsp. enterica serovar Typhimurium	LT2	2	STM0693	NP_459678.1	STM0693
			STM4241	NP_463106.1	STM4241
Stigmatella aurantiaca	DW4/3-1	2	STAU_4185	YP_003953794.1	STAU_4185
			STAU_0853	YP_003950484.1	STAU_0853
Desulfovibrio vulgaris	Miyazaki F	4	DvMF_1241	YP_002435659.1	DvMF_1241
			DvMF_2366	YP_002436775.1	DvMF_2366
			DvMF_1364	YP_002435781.1	DvMF_1364
			DvMF_0116	YP_002434545.1	DvMF_0116
Myxococcus xanthus	DK 1622	2	MXAN_6967	YP_635082.1	MXAN_6967
			MXAN_3702	YP_631888.1	MXAN_3702
Acidithiobacillus caldus	SM-1	2	Atc_2587	YP_004749936.1	Atc_2587
			Atc_1297	YP_004748646.1	Atc_1297
Thiobacillus denitrificans	ATCC 25259	2	Tbd_1123	YP_314881.1	Tbd_1123
			Tbd_1714	YP_315472.1	Tbd_1714
			Tgr7_0967	YP_002513042.1	Tgr7_0967
Thioalkalivibrio sulfidophilus	HL-EbGr7	3	Tgr7_3054	YP_002515110.1	Tgr7_3054
			Tgr7_3233	YP_002515288.1	Tgr7_3233

Range 1: 20 to 113 [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
64.3 bits(155)	2e-18	Compositional matrix adjust.	30/94(32%)	53/94(56%)	1/94(1%)
Query 12	LEDAGINVTVQRLQMANLLLSKPOHLTADQVFQLINEHMPNASRATIFNNLKLFAEKGIV				71
	L+ I T+QR+ + +L+ H TA+ V + ++ S AT++N L + G++				
Sbjct 20	LKSVSIQPTIQRISICQYVLCEADHPTAEVVKWVDSRSFKMSLATVYNTLNILVSAGLL				79
Query 72	NLLELKS-GITLYDSNVIIHHHHAIDEKTGEIYDI				104
	+ G ++YDSN+I H+H DEK+G+ +DI				
Sbjct 80	REFKFSCLGKSVYDSNIIIDHYHFFDEKSGKFHDI				113

## Anexo 2: Alinhamento entre as sequências LIC11158 e LIC12034

Alinhamento utilizando-se BLAST align (NCBI/bl2seq) utilizando-se LIC11158 como entrada e LIC12034 como comparativo.

Range 1: 20 to 113 [Graphics](#) ▼ Next Match ▲ Previous Match

Score	Expect	Method	Identities	Positives	Gaps
63.5 bits(153)	3e-18	Compositional matrix adjust.	29/94(31%)	53/94(56%)	1/94(1%)
Query 12	LEDAGINVTVQRLQMANLLLSKPOHLTADQVFQLINEHMPNASRATIFNNLKLFAEKGIV				71
	L+ I T+QR+ + +L+ H TA+ V + ++ S AT++N L + G++				
Sbjct 20	LKSVSIQPTIQRISICQYVLCEADHPTAEVVKWVDSRSFKMSLATVYNTLNILVSAGLL				79
Query 72	NLLELKS-GITLYDSNVVHHHHAIDEKTGEIYDI				104
	+ G ++YDSN++ H+H DEK+G+ +DI				
Sbjct 80	REFKFSCLGKSVYDSNIIIDHYHFFDEKSGKFHDI				113

## Anexo 3: Alinhamento entre as sequências LA\_1857 e LIC12034

Alinhamento utilizando-se BLAST align (NCBI/bl2seq) utilizando-se LA\_1857 como entrada e LIC12034 como comparativo.

```

P71086  MAAHELK--- ----- ---EALETLK ETGVRI TPQR  AILEYLVNS
fur979  MNREKQE-AI LNKTQPAVRM EMQTFSEYLQ KEGLKITNQR MLVAERIFSL
fur1972 MKDSYER--- ----- ---SKKILE DAGINVTVQR LQMANLLLSK
fur1128 MESLFAKVC LT-----PV ---EIERRLK SVSIQPTIQR ISICQYVLCE
fur3538 MES----- ----- -----K KTVLRNTRKQK GEILKVLEMA

```

```

P71086  MAHPTADDIY KALEGKFPNM SVATVYNNLR VFRSGLVKE LTYGD-ASSR
fur979  HNHFTAEGLL EEFKDQRDQI SKATIYRILS IMVSAGLLQE HNFQK-YKY
fur1972 PQHLTADQVF QLINEHMPNA SRATIFNNLK LFAEKGIVNL LELKS-GITL
fur1128 ADHPTAEVVK EWVDSRSFKM SLATVYNTLN ILVSAGLLRE FKFSCLGKSV
fur3538 KGPLSIKEIY ELSRKNLDNL GIATVYRAVN HLMETGTIHE IHLPG-ESSR

```

```

P71086  F FVTS--DE YHAICENCGK IV FHYPLD EVEQLAAHVT GFKVSH-HRL
fur979  YEHIIGHKHE DHI ICTVCGK IVEFLDERIE QLQEQAAAREN GFKITG-HSL
fur1972 Y SNVI--HE HHAIDEKTGE IY I SLDSKL QEKVLSSELKQ DFKLKTGSSL
fur1128 Y SNII--DE YHFFDEKSGK FH IDPSLL- SLS--SKLPP EFLVNK-TDI
fur3538 FEASRH--HE HFFHC KQCDR VY IEICPI- PL---DKSPK GFTVDT-HEI

```

```

P71086  EIYGV CQEC S KKENH
fur979  NIYGT CNE-- HSSSK
fur1972 ENC NLSITLK GKKNP
fur1128 LLTGN----- LVSET
fur3538 ILYGT CSDC- NSKAQ

```

#### Anexo 4: Correlação entre os números de acesso de genes e proteínas utilizadas.

As entradas da primeira coluna mostram os nomes utilizados para cada primer.

Nome	Locus Tag (gene)	Número de acesso (proteína)
16s	LIC11010	
1128	LIC11158	YP_001128.1
1972	LIC12034	YP_001972.1
979	LIC11006	YP_000979.1
3538	LIC20147	YP_003538.1
LexA1	LIC12305	YP_002239.1
LexA2	LIC12654	YP_002577.1
recA	LIC11745	YP_001697.1
dinP	LIC13052	YP_002964.1

## Anexo 5: Alinhamento entre p71086 e as 4 Furs

*Alinhamento entre P71086 de Bacillus subtilis e as 4 furs. Os aminoácidos em destaque são os aminoácidos destacados como essenciais por Hellmann et al. Os aminoácidos em verde representam o grupo relacionado a um possível sensor de íons. Os aminoácidos em vermelho são sugeridos como de ligação a ferro em Pseudomonas aeruginosa. Os aminoácidos em amarelo estão relacionados a ligação a zinco. Os aminoácidos com um asterisco estão presentes em P71086 e fur3538.*

Nome	Locus Tag (gene)	Accession number (proteína)
16s	LIC11010	
1128	LIC11158	YP_001128.1
1972	LIC12034	YP_001972.1
979	LIC11006	YP_000979.1
3538	LIC20147	YP_003538.1
LexA1	LIC12305	YP_002239.1
LexA2	LIC12654	YP_002577.1
recA	LIC11745	YP_001697.1
dinP	LIC13052	YP_002964.1

**Anexo 6: Genomic survey and expression analysis of DNA repair genes in the genus *Leptospira***

## Genomic survey and expression analysis of DNA repair genes in the genus *Leptospira*

Marinalva Martins-Pinheiro<sup>1,2, #</sup>, Luciane Schons Fonseca<sup>3,4, #</sup>, Josefa B. da Silva<sup>3</sup>, Leonardo Hiroyuki Santos Momo<sup>5</sup>, Ana Carolina Quirino Simões<sup>6</sup>, Paulo Lee Ho<sup>3,4</sup> and Renata M. A. da Costa<sup>\*,5</sup>

<sup>1</sup> Departamento de Microbiologia, Instituto de Ciências Biomédicas, Universidade de São Paulo, 05508-900, Brazil

<sup>2</sup> Departamento de Engenharia Química, Escola Politécnica, Universidade de São Paulo, São Paulo, São Paulo, Brazil

<sup>3</sup> Centro de Biotecnologia, Instituto Butantan, São Paulo, 05503-900, Brazil

<sup>4</sup> Departamento de Bioquímica, Instituto de Química, Universidade de São Paulo, São Paulo, 05508-000, Brazil

<sup>5</sup> Centro de Ciências Naturais e Humanas, Universidade Federal do ABC, Santo André, 09210-170, Brazil

<sup>6</sup> Centro de Engenharia, Modelagem e Ciências Sociais Aplicadas, Universidade Federal do ABC, Santo André, 09210-170, Brazil

# These authors contributed equally to this work.

\* To whom correspondence should be addressed: Renata M.A. da Costa, Centro de Ciências Naturais e Humanas, Universidade Federal do ABC. Tel: 55-11-4996 8369; Fax: 55-11-4996 8369; E-mail: renata.costa@ufabc.edu.br

Email addresses:

MMP: marimart@usp.br; LSF: luschons@gmail.com; JBS: josefa.silva@butantan.gov.br; LHSM: leonardo.momo@gmail.com; ACQS: ana.simoes@ufabc.edu.br; PLH: paulo.ho@butantan.gov.br; RMAC: renata.costa@ufabc.edu.br



## ABSTRACT

Leptospirosis is an emerging zoonosis with important economic and public health consequences and is caused by pathogenic leptospires. The genus *Leptospira* belongs to the order Spirochaetales and comprises saprophytic (*L. biflexa*), pathogenic (*L. interrogans*) and host-dependent (*L. borgpetersenii*) members. Here we present an *in silico* search for DNA repair pathways in *Leptospira* spp. The relevance of such DNA repair pathways was assessed through the identification of mRNA levels of some genes during infection in animal model. The search was performed by comparison of available *Leptospira* spp genomes in public databases with known DNA repair-related genes. Leptospires exhibit some distinct and unexpected characteristics, for instance the existence of a redundant mechanism for repair a chemically diverse spectrum of alkylated nucleobases, a new *MutS*-like gene and a new shorter version of UvrD. Although gram-negative, *Leptospira* spp shares some characteristics from gram-positive, as the presence of PcrA, two RecQ paralogs and two SSB proteins; the latter is considered a feature shared by naturally competent bacteria. We did not find a significant reduction in the number of DNA repair-related genes in both pathogenic and host-dependent species. Pathogenic leptospires were enriched for genes dedicated to base excision repair and to non-homologous end joining. Their evolutionary history reveals a remarkable importance of lateral gene transfer events for the evolution of the genus. Up-regulation of specific DNA repair genes, including components of SOS regulon, during infection in animal model validates the critical role of DNA-repair mechanisms for the complex interplay between host/pathogen.

Keywords: *Leptospira*, DNA repair, bacterial genomics, mutagenesis, SOS response.

## BACKGROUND

The genus *Leptospira* belongs to the order Spirochaetales, and the genetically based classification indicates the existence of at least 19 species (13 pathogenic and six saprophytic). *Leptospira* species are categorized into 24 serogroups and 250 serovars, based on the expression of surface-exposed lipopolysaccharide (LPS) (Evangelista and Coburn 2010). Pathogenic leptospires are the etiologic agents of leptospirosis, an emerging and worldwide zoonosis with important economic and public health consequences, in particular to developing tropical countries (Ko et al. 2009). Leptospires can infect a great range of animal species that compose their natural reservoir, colonizing kidneys and being shed in the urine during the whole life of these animals (Levett 2001). Humans are considered incidental hosts, and infection occurs through direct contact with infected animal urine or indirectly through contaminated water.

Functional studies are hampered by the lack of consolidated techniques for genetic manipulation of pathogenic leptospires. As consequence, the molecular mechanisms of *Leptospira* pathogenesis are still poorly understood. However, this scenario is changing due to the availability of genomic information. Whole-genome sequences of the pathogenic *L. interrogans* serovars Lai and Copenhageni, and *L. borgpetersenii* serovar Hardjo, as well as two strains of the free-living *L. biflexa* are driving the knowledge of their host-pathogen relation. For example, it became clear that the restricted pathogenic life-style from *L. borgpetersenii* results from the detriment of some genes necessary for environment sensing as well as metabolite transport and utilization (Bulach et al. 2006). On the other hand, *L. interrogans* does not show this level of genome reduction, and, as consequence, the bacteria can survive in water or mud for weeks, after what they are still able to infect the host (Ren et al. 2003; Nascimento et al. 2004). *L. biflexa* shares 60% of gene content with the pathogenic species, therefore it is considered an excellent model for the study of *Leptospira* evolution (Picardeau et al. 2008). Lateral gene transfer (LGT), considered a minor contributor to the overall genetic composition of the genus (Picardeau et al. 2008), might have an important role in this process.

As any living organism, leptospires have to deal with endogenous and exogenous agents that threaten its genetic material, resulting in DNA alterations, or lesions, that may disturb several metabolic processes (Friedberg et al., 2006) DNA repair pathways are the main mechanisms for DNA safeguarding, and are able to remove virtually all known kinds of lesions. The remaining lesions that were not removed

and block the progress of replication forks during DNA replication are tolerated by cells through DNA damage tolerance pathways. These latter efficient mechanisms differ from DNA repair because rather than restoring DNA to its proper sequence and structure, the lesion is still present in the DNA.

DNA repair capacity of *Leptospira* spp was barely addressed until now. The sensitivity to UV-C irradiation and to the DNA cross-linking agent mitomycin C was compared among the saprophytic and pathogenic species. Although *L. interrogans* and *L. biflexa* have a photoreactivation DNA repair mechanism, the pathogenic *Leptospira* was the most sensitive to both agents (Stamm and Charon 1988). Recently, our group showed the activation of SOS response following UV-C-induced DNA damage in *L. interrogans* serovar Copenhageni (Fonseca et al. 2013). Surprisingly, the typical DNA repair genes were not regulated by LexA, supporting the notion that the gene content from different species might differ in their expression regulation. In addition, this serovar has a second LexA ortholog (LexA2), which most probably was acquired by lateral gene transfer (LGT) (Fonseca et al. 2013). Moreover, the importance of genomic stability in maintaining the virulent status of *L. interrogans* serovar Lai was suggested by genomic and proteogenomic analyses (Zhong et al. 2011; Lehmann et al. 2013). Low expression of DNA repair genes in the virulence-attenuated IPAV strain may be responsible for insertions, deletions and single-nucleotide variations found in 101 genes in comparison to its highly virulent ancestral strain, 56601.

This work presents a thorough bioinformatics search for DNA repair pathways in the genus *Leptospira* based on genomic information. ORFs from *Leptospira* spp genome with significant similarity to known DNA repair-related genes were searched and analyzed. The genes involved in DNA repair are, in general, part of the cell core metabolism and they maintain a strong similarity in different bacterial genomes. However, this search revealed intriguing differences in *Leptospira*, suggesting a distinct biological diversity in response to DNA damage. *Leptospira* spp is provided with an unusual number of genes involved in reversion and removal of alkylated lesions and for members of the superfamily 1A helicase. So far, these activities appear to be the most extensively verified among other bacteria. In addition, the DNA repair core genes from saprophytic and pathogenic bacteria differed in some extension. Pathogenic leptospire were enriched for genes dedicated to base excision repair, responsible for the removal of small, non-helix-distorting base lesions from the genome. Surprisingly, lateral gene transfer events appear to be responsible for most of these acquisitions and provide evidence for the genome

dynamic of the genus. The relevant activation of some DNA repair pathways and the SOS response was assessed through *in vivo* expression analysis during leptospirosis in animal model. DNA repair genes showed up-regulation in kidney from susceptible mice during the acute phase of the infection. On the other hand, components from SOS regulon were up-regulated in both intermediate and susceptible mice.

## METHODS

- **Identification of DNA repair genes**

BLAST similarity searches using the DNA repair genes were first performed with *E. coli* as seed. Whenever an important gene was absent in *E. coli* other organisms were used: *Bacillus subtilis*, *Bacillus cereus*, *Mycobacterium tuberculosis* or *Caulobacter crescentus*. The searches were conducted against the non-redundant (nr) protein sequence database (BlastP) (Altschul et al. 1990) restricted individually for each complete leptospira genomes (*Leptospira biflexa* serovar Patoc str. 'Patoc1 (Paris)', RefSeq NC\_010602.1, NC\_010843.1 and NC\_010844.1; *Leptospira biflexa* serovar Patoc str. 'Patoc 1 (Ames)', RefSeq NC\_010842.1, NC\_010845.1 and NC\_010846.1; *Leptospira interrogans* serovar Copenhageni strain Fiocruz L1-130, RefSeq NC\_005823.1 and NC\_005824.1; *Leptospira interrogans* serovar Lai str. 56601, RefSeq NC\_004342.2 and NC\_004343.2; *Leptospira borgpetersenii* serovar Hardjo-bovis strain JB197, RefSeq NC\_008510.1 and NC\_008511.1; and *Leptospira borgpetersenii* serovar Hardjo-bovis str. L550, RefSeq NC\_008508.1 and NC\_008509.1). Significant results were considered only if E-values were lower than  $10^{-4}$  (as of July 2013). Candidate genes were thus confirmed both by sequence similarity searches (BlastP with default parameters, against nr database) and domain analysis (Pfam (Punta et al. 2012) and HHpred (Söding et al. 2005)). The list of organisms used in this study, as well as the seed sequences used to perform the similarity searches when absent in the phylogenetic trees, are shown in Table S1.

- **Phylogenetic analyses**

For the phylogenetic analyses, protein sequences of genes were aligned using ClustalX 2.0 algorithm (Thompson et al. 1997) and all gaps were eliminated from the dataset. Phylogenetic reconstructions were conducted in Mega5 (Tamura et al. 2007) and performed with Maximum Likelihood (ML) and Distance Based Neighbor-Joining (NJ). Similar topologies were found for both algorithms

employed, but only one is displayed. Bootstrap assessment of tree topology (1000 replicates) was generated and only bootstrap values greater than 50% are shown. The organisms and the accession codes of genes investigated in the phylogenetic analysis are shown in Table S2.

- **Animals and *Leptospira* strain.**

The virulence of *L. interrogans* serovar Copenhageni (ATCC BAA-1198) was maintained by passages in golden hamster (*Mesocricetus auratus*) as described (da Silva et al. 2012). Leptospire isolates from animals were cultured in liquid Ellinghausen McCullough Johnson and Harris (EMJH) medium at 30°C under aerobic conditions. After 5-6 days of culture, bacteria were counted in Petroff-Hausser counting chamber and suspensions were used for infection of mice. The hamsters and the strains C3H/HeJ and C3H/HePas of mice were provided by Instituto Butantan and Instituto de Ciências Biomédicas of Universidade de São Paulo (SP Brazil). Samples used in this study were collected previously (da Silva et al. 2012).

- **Experimental infection of animals**

We used material from kidneys of infected mice, during the acute phase of the disease, i.e. five days after mice infection. The experimental design was described previously (da Silva et al. 2012). The strains of C3H/HeJ and C3H/HePas mice were infected intraperitoneally with  $1 \times 10^7$  leptospire (*L. interrogans* serovar Copenhageni) and the remaining five animals of each strain were kept uninfected as control. In parallel, four 3-4 weeks-old hamsters were infected with the same dose to confirm the virulence of the bacteria. Samples of six mice were collected at days 0 (un-infected animals) and 5 after inoculation. Samples were kept at -80°C until analysis. The protocol used in this study was approved by the Ethical Committee for Animal Research of Butantan Institute (number: 560/08).

- **RNA manipulation and quantitative PCR**

Organs from infected mice were collected in liquid nitrogen and kept to -80°C, as previously described (da Silva et al. 2012). Total RNA was isolated using Trizol reagent (Invitrogen). An aliquot of 2 g RNA of each sample was transcribed to cDNA using RevertAid™ H Minus Reverse Transcriptase (Fermentas). The oligonucleotide sequences for *lexA1*, *recA*, *recN*, LIC\_12362 (*alkA2*), LIC11717 (*uvrA*) and LIC13064 (*tag*) genes were previously described (Fonseca et al., 2013). Additional oligonucleotide sequences were: LIC12677 (*DNA lig*) forward (AGCTCTCTTCAAACCGGAAC) and reverse (CTGTGTATCCTTTCGCCATCT), LIC\_10238 (*alkA1*)

forward (CCTTCGCCCCGAACAGATTT) and reverse (GCCTCAGCGATACGTTTGAT), LIC\_13402 (*ada*) forward (TTCTCTCTAAACCTCAAACCACTT) and reverse (CGGAGTCATCCCTTCAATCTTTA), and *16S* forward (TTCAGTTGGGCACTCGTAAG) and reverse (CGTGTGTTGCCCTAGACATAA).

Quantitative PCR reaction (qPCR) was carried out with the Syber Green Master Mix (Applied Biosystems, USA) using 100 ng of cDNA for samples of kidney and 10 ng for samples of bacteria from EMJH cultured 5 days at 30°C, 5 moles of each forward and reverse primer in 12 L final reaction. The PCR thermal cycle conditions were described previously (da Silva et al. 2012). qPCR analysis was performed using the Applied Biosystems 7300 Real-Time PCR System. Cycle threshold (Ct) values for specific genes were normalized to the Ct values of the leptospiral 16S mRNA gene and expression fold change compared to gene expression levels in vitro EMJH cultured *Leptospira*. The relative levels of mRNA from each selected gene were analyzed using  $2^{-CT}$  method described by (Livak et al. 2013). All real-time PCR reactions were performed in triplicate and the results represent the data of two individual experiments. Any values outside one absolute deviation around the median (MAD-median method) were considered outliers (Leys et al. 2013) and were not considered for analysis. P values were calculated through two-tailed Student's t-test. Statistics and plotting of data were done in Prism software (GraphPad).

## RESULTS

Genomic information from saprophytic and pathogenic leptospires was screened for open reading frames with significant similarity to genes known to play roles in DNA repair and/or tolerance. This search was performed only in genomes from leptospires whose sequence is available at the RefSeq collection. Five main categories of DNA repair were investigated: 1. direct repair; 2. excision repair; 3. recombinational repair; 4. DNA damage tolerance and 5. other repair related proteins. Fig. 1 summarizes our findings. Activation of some DNA repair pathways was tested in experimental intermediate and susceptible mice models through the detection of the corresponding transcripts by qRT-PCR.

### 1. Direct repair

Direct repair reverts DNA damage to its original structure, in an error-free single-step reaction. It can be achieved by photoreactivation, alkyltransfer and oxidative demethylation.

#### 1.1. Photoreactivation

Photoreactivation is performed by photolyase enzymes that revert pyrimidine dimers induced by UV irradiation through light absorption. The photolyase family shares a flavin adenine dinucleotide (FAD) as catalytic cofactor and includes CPD and 6-4 photolyases, that repair CPD (cyclobutane pyrimidine dimer) and 6-4 lesions, respectively, and the cryptochromes, that function as photoreceptors (Menck 2002). The presence of photolyase-encoding genes in leptospira genomes has a direct correlation with their respective life-style (Fig. 1 and 2a). *L. borgpetersenii*, the obligate pathogen, has no single ortholog (Table S3), while *L. interrogans* and *L. biflexa* encode a protein with similarity to photolyase family II (called *photolyase b*), which is closer phylogenetically to eukarya photolyases and cryptochromes (Table S3 and Fig. 2a, shaded). In addition to *photolyase b*, *L. biflexa* possesses another copy, similar to *phra-like* gene from *Rhodobacter sphaeroides* (Fig. 2a, asterisks). Leptospires also present another putative photolyase, encoded by the *splB* gene (Table S3). The homologous protein from *B. subtilis* repairs the exclusive UV-induced lesion 5-thymine-5,6-dihydrothymine from spores, when DNA is highly compacted. The role of this putative protein in *Leptospira* spp remains to be determined.

#### 1.2. Alkyltransferases and DNA glycosylases for removal of N-alkylated bases

Alkylated damages occur as result of interaction of environmental or endogenous compounds with cellular macromolecules, such as lipid peroxidation, subsequently generating a variety of adducts with different potential of genotoxicity. The most specific and genetically regulated system for repairing alkylated damage is the direct repair, through the alkyltransferases coded by the *ada* and *ogt* genes in *E. coli*. The products of these genes, *O*<sup>6</sup>-alkylguanine-DNA alkyltransferase I and II (*O*<sup>6</sup> AGT I and *O*<sup>6</sup> AGT II), respectively, act as a single agent, removing the alkyl groups in a non-enzymatic suicide mechanism (Pegg and Byers 1992). While *ogt* is absent in *Leptospira* spp., all species have *ada* and an alkyltransferase-like (*atl*) gene, whose identities were determined by phylogenetic analysis (Fig. 1 and 2b).

### 1.3. Oxidative demethylation

The unique gene involved in oxidative demethylation pathway in *E. coli* is *alkB*, and only *L. biflexa* shares an orthologous (Table S3). AlkB protects cells against the cytotoxic effects of S<sub>N</sub>2 methylating agents, through an unprecedented oxidative dealkylation mechanism that requires alpha-cetoglutarate, oxygen and iron (II) (Falnes and Rognes 2003).

## 2. Excision repair

Although efficient, direct reversion of lesions is limited due to the high enzyme/substrate specificity. Therefore, organisms have broader mechanisms that remove damaged bases and allow the polymerization of the correct nucleotide sequence in an error free manner.

### 2.1. Base excision repair

**Base excision repair (BER)** protects the genome from small modifications into nitrogenated bases from DNA. Such lesions are generated by spontaneous chemical modifications in DNA or by chemical assaults, as those induced by macrophages, resulting in oxidation or alkylation of the base. Excision of the damaged base requires the sequential activity from DNA glycosylases, apurinic/apyrimidinic (AP) endonucleases and an exonuclease. The repair is finalized by a repair polymerase, such as Pol I, and ligated, restoring the nucleotide to its undamaged form (Dalhus et al. 2009). The main genes involved in BER are present in *Leptospira* spp (Fig. 1). The monofunctional DNA-glycosylases diverged considerably among pathogenic and saprophytic species (Table S4). The only exception is the adenine DNA glycosylase (MutY), responsible for removing the adenine mispaired with 8-oxoguanine (8-oxo-G), common to all



leptospire species. On the other hand, all leptospire species contain only one bifunctional DNA glycosylase (endonuclease III), which also contain a lyase activity that cleaves the backbone immediately 3' to the AP site, and a single AP endonuclease, the exonuclease III (Table S4).

The glycosylases Tag and AlkA also remove alkylated damages. Orthologues for *alkA* are absent in *L. biflexa*, while there is only one in *L. borgpetersenii* and two in *L. interrogans*. Both LIC10238 (termed here AlkA1) and LIC12362 (AlkA2) from *L. interrogans* show a great degree of similarity in amino acids sequence and inferred secondary structure with the corresponding proteins from *E. coli* and *B. subtilis* (Labahn et al. 1996; Yamagata et al. 1996) (Fig. 3a). The product from LIC10238/LA0281 gene (marked as AlkA1) clusters with *L. borgpetersenii* AlkA and other pathogenic leptospire species, while LIC12362/LA1370 (AlkA2, shaded in Fig. 3b) clusters with a very distant group, suggesting its acquisition by a lateral gene transfer event (LGT). LGT might also explain the presence of *tag* orthologs only in *L. interrogans* (Table S4). This hypothesis is reinforced by its genomic neighborhood, with adjacent transposase genes (Fig. S1).

Curiously, orthologues for the new structural DNA glycosylases AlkC and D, with narrow phyletic distribution, are found in the genus. The reduced-genome *L. borgpetersenii* possesses one ortholog for *alkD*, while *L. biflexa* has *alkC* and *D* orthologues. However *L. interrogans* presents more diversity, with two *alkD* and one *alkC* genes. (Table S4).

Although uracil-DNA glycosylases (UDG) are present in all leptospire species, the occurrence of two UDGs is exclusive to pathogenic leptospire species. UDGs are universal DNA repair enzymes that catalyze the removal of mismatched uracil from DNA, originated from deamination of cytosine or by misincorporation of dUTP (Friedberg et al. 2006). UDGs from pathogenic and saprophytic leptospire species belong to different groups, as phylogenetic inference reveals (Table S4 and Fig. 4). UDG from saprophytic bacteria belongs to family 1, which is the most efficient Uracil-DNA glycosylase. On the other hand, UDG from pathogenic *Leptospira* represented by LIC\_10548 belongs to family 4, found in thermophilic organisms (Fig. 4). LIC\_10548 clusters with UDGs from other spirochaetes, suggesting this is the original ortholog for the genus, while LIC\_13102 and LEPBI\_I0821 were acquired through a LGT event. We could not assign a family for LIC\_13102 neither comparing its domains nor by phylogenetic analysis (Fig. 4). Nevertheless, this protein has the characteristics of and clusters with thermostable proteins from families 4 and 5.

The genus *Leptospira* showed remarkable differences among nudix-hydrolase superfamily members, the MutT-like proteins (Table S4). We found four, six and seven proteins from this family in *L. biflexa*, *L. borgpetersenii* and *L. interrogans*, respectively. They perform nucleotide pool sanitization, which is essential to maintain genomic stability. For example, MutT is able to remove 8-oxo-(d)GTP from the nucleotide pool, an oxidated form of (d)GTP which is highly mutagenic if incorporated in DNA (Cox and Yanofsky 1969; Foti et al. 2012, Bhatnagar and Bessman 1988).

## 2.2. Nucleotide excision repair

Nucleotide excision repair (NER) is considered as the most flexible and versatile DNA repair pathway, since it is responsible for removing any lesion causing distortion in the double helix (Truglio et al. 2006). This function is performed by UvrA, UvrB, endonuclease UvrC and helicase UvrD proteins. Polymerase I and ligase proteins complete the repair reaction (Truglio et al. 2006). Transcription-coupled nucleotide excision repair (TC-NER), that requires the TRCF protein (transcription repair coupling factor, encoded by *mfd* gene), permits faster repair of the transcribed strand in active genes (Ganesan et al. 2012). This pathway is also complete in *Leptospira* spp (Fig. 1 and Table S4).

Leptospire present four different UvrD-like proteins (Table S4). UvrD protein is a member of the superfamily 1A helicase, which includes the very similar PcrA and Rep helicases (Petit and Ehrlich 2002). Their domains are shown in Fig. 5, with the phylogenetic analysis that revealed the identity of all three helicases. The N-terminal region of UvrD consists of the ATPase module of the protein (P-loop NTPase domain) and its C-terminus is the domain of UvrD\_C\_2. PcrA is encoded by Gram-positive bacteria while Rep and UvrD are exclusive to Gram-negative bacteria. Phylogenetic analysis of UvrD-like proteins revealed the presence of all three helicases (Fig. 5). In addition, there is a shorter version of UvrD, which seems to be specific to Leptospiraceae, with homologues only in *L. biflexa*, *L. interrogans*, *Leptonema illini* and *Turneriella parva* (see Fig. 5 and Table S4). Phylogenetic tree generated for this gene family shows the proximity of the short UvrD with Archea UvrD, a member from the new nuclease superfamily characterized by the presence of the PDDEXK domain (Laganeckas et al. 2011).

## 2.3. Mismatch repair

Mismatch repair (MMR) is a post replicative repair that handles with some types of DNA damage that leads to incorrect base pairing. As such, it acts by preventing mutations that arise from errors of replication or

recombination, ensuring the fidelity of DNA replication and repair. Recognition of a mismatch is done by the highly conserved MutS and MutL proteins, which directs the endonuclease activity of MutH in *E. coli*. *MutH* shows limited distribution among microorganisms and, as expected, is not found in leptospires (Fig. 1). At least in *B. subtilis*, MutL could take the role of strand discrimination and lesion removal (Guarné 2012). The repair is completed by UvrD, DNA polymerase III and DNA ligase (reviewed in Kunkel and Erie 2005). *MutS* and *mutL* genes are present in all three species of leptospira, indicating that mismatch repair must be functional (Fig 1). Unexpectedly, we found a new member from MutS subfamily (*mutS3*) in all leptospires (Fig. 1 and Table S4) (Lin, et al. 2007). Functional activity of this paralogue is not documented and its role in DNA metabolism remains to be elucidated.

We found one ortholog for Dam methylase only in the genome-reduced *L. borgpetersenii* (Table S4), suggesting its acquisition through LGT event.

### 3. Recombinational repair

Recombinational repair is essential to all living organisms, as it deals with double-strand break damages (DSB), single-strand gaps and it restores blocked or destabilized replication forks (Friedberg et al. 2006). In most bacteria these processes are performed by homologous recombination (HR) apparatus.

#### 3.1. Homologous recombination

Homologous recombination (HR) involves genetic exchanges between two DNA molecules that share an extended region of nearly identical sequence. There are two main pathways acting in HR: the RecBCD, which recognizes DSBs, while RecFOR, DNA nicks and gaps, and an alternative one, RecQJ (Wigley 2013). The three pathways are represented in all leptospires, (Fig 1 and Table S5).

RecA is the central protein in homologous recombination and it requires single-stranded DNA (SSD) to promote homology search and associated DNA strand exchange. SSDs are generated by the exonucleolytic activity of RecBCD (exonuclease V), resulting in 3' overhang remaining of the strand (Kowalczykowski SC, 2000). Single-stranded-DNA-binding protein (SSB) binds to this extremity and modulates the RecA-DNA filament formation. Leptospires possess two orthologs for *ssb* (they share 36% identity and 56% similarity among them). With few exceptions, naturally competent bacteria contain two SSB proteins while non-competent bacteria express only a single protein (Lindner, et al. 2004). Leptospires also show redundancy in the auxiliary DNA end-resection pathway, dependent on RecQ helicase and RecJ

exonuclease (Chu and Hickson 2009). However, as found in gram-positive, leptospire encode an additional shorter version of RecQ, named RecS (Lenhart et al., 2012). RuvABC complex and RecG protein, that promote branch migration (for review, see Friedberg et al., 2006), are fully represented in the genome of leptospire (Table S5).

SMC family members (structural maintenance of chromosomes) Mre11-Rad50 and RecN proteins are also found in leptospire (Table S5). Mre11-Rad50 proteins, known as SbcD and SbcC respectively, are highly conserved and they have the specific role to associate and stabilize DNA ends at a DSB (Ciccia and Elledge 2010).

### 3.2. Non homologous end joining

When a homologous chromosome is not available, DSBs can be repaired through a low-fidelity pathway, the non-homologous end joining (NHEJ). NHEJ catalyzes end joining of two broken DNA ends, using minimal or no sequence homology, and therefore it is highly mutagenic (Pitcher et al. 2007). Bacterial NHEJ is a two-component DNA repair pathway consisting of a Ku homolog and an ATP-dependent ligase (*ligD*). Orthologous for ATP-dependent ligase were found in *L. interrogans* (Table S5), suggesting its acquisition through LGT event. In addition, *L. interrogans* shares a Gam-like protein, as part of a prophage (Qin et al. 2008; Fonseca et al, not published data). Gam-like proteins are the orthologs of eukaryotic Ku, protect recombination intermediates from RecBCD and allow both phage rolling cycle replication (Hillyar, 2012) and bacterial non-homologous end joining to occur (d'Adda di Fagagna, Weller et al. 2003).

## 4. DNA damage tolerance

When the DNA replication machinery faces a remaining lesion that was not removed, replicative bypass over noncoding bases is recruited. Y-family translesion synthesis (TLS) polymerases have the special ability to bypass lesions that block progression of replicative DNA polymerase. We found a very limited distribution of Y-family TLS polymerases among leptospire (Table S6). *E. coli* PolV, the *umuDC* gene product, is absent in leptospire, as well as the error-prone processing cassette expressing a second copy of the catalytic subunit of Pol III, *dnaE2*, *imuA* and *imuB* (Galhardo et al. 2005; McHenry 2011). On the other hand, Pol IV, the *dinP* gene product, is present in *L. biflexa* and *L. interrogans* (Table S6). *L.*

*borgpetersenii* lost this ortholog, most probably due to genomic rearrangements, since a transposase insertion is found in its corresponding locus (Fig. S2).

*L. biflexa* encodes an ortholog for archaeal B family DNA polymerase (Table S6). This family was described as inactivated since it contains disrupted versions of the essential sequence motifs for the catalytic function (Rogozin et al. 2008). LGT event was associated with its presence in *L. biflexa* and two other bacteria. However, no clues to its functional activity were exposed until now.

## 5. Other repair related proteins

### 5.1. SOS system

Another important function of RecA protein is the regulation of SOS pathway. The SOS is an inducible and coordinated pathway comprised by a variable set of genes, and it is regulated by the repressor LexA and the sensor of ssDNA, RecA. (Friedberg et al. 2006). All *Leptospira* species have both *lexA* and *recA* genes, suggesting a functional SOS system. The serovar Copenhageni, exclusively, possesses a second ortholog for LexA (*lexA2*), acquired by LGT. It binds to its own promoter, while LexA1 controls at least nine genes (*lexA1*, *lexA2*, *recA*, *recN*, *dinP*, and four genes encoding hypothetical proteins) (Fonseca et al. 2013). Modulation of RecA activity is achieved by RecX in *E. coli* and *B. subtilis*. However, we did not find corresponding orthologs in *Leptospira* spp, suggesting a different mechanism for RecA modulation.

## 6. DNA repair induction during experimental leptospirosis

The relevance of DNA repair pathways in *Leptospira* biology was assessed through evaluation of *in vivo* gene expression in experimental intermediate and susceptible mice models. Previous studies from our group showed a different expression profile of some inflammatory mediators following inoculation with virulent *L. interrogans* serovar Copenhageni between C3H/HeJ mice and its parental lineage, C3H/HePas (da Silva et al. 2009; da Silva et al. 2012). The first is a TLR4 mutant, susceptible to leptospirosis, while the second shows an intermediate response. We used the same material from kidneys during the acute phase of infection (day 5), since this is the preferable organ for leptospira colonization, both lineages show similar levels of inflammatory chemokines and the amount of leptospires in this organ

is maximal (da Silva et al. 2009; da Silva et al. 2012). Transcripts from both *alkA1* (LIC\_10238) and *alkA2* (LIC\_12362), *ada* (LIC\_13402); *tag* (LIC\_13064); *uvrA* (LIC\_11717) and *ATP dependent DNA ligase* (LIC\_12677) were quantified by reverse transcription-quantitative PCR (RT-qPCR) and normalized to 16S rRNA control. Corresponding Ct values were compared to the *in vitro* cultured (EMJH) leptospires. The gene expression profile, expressed as  $\log_2$  of the fold change between the two conditions, is presented in Fig. 6. With the exception of *ada* and *alkA1*-LIC\_10238 (data not shown), which were barely expressed both in EMJH and *in vivo*, all other genes showed up-regulation in the infected organ from susceptible mice.

We also evaluated the recruitment of SOS response by measuring mRNA levels of the regulators *lexA1* (LIC\_12305) and *recA* (LIC\_11745), and the SMC family member, *recN* (LIC\_11620) (Fig.6). The LexA1 targets (*recA* and *recN*) were up-regulated in both susceptible and intermediate mice, while *lexA1* was not.

## DISCUSSION

This work assessed the genomic content for DNA repair-related genes in the genus *Leptospira* spp. Comparative analysis revealed similarities among the genus, most probably acquired by the common ancestral. The set of *Leptospira* DNA repair genes diverges considerably from model organisms, what renders it as an interesting organism for future studies. Some examples are the unusual number of genes related to reversion or removal of alkylation lesions and the presence of the uncommon *mutS3* and short *UvrD*. In addition, although gram-negative, leptospires share some orthologous from gram-positive bacteria involved in direct repair, NER and HR. Functional and evolutionary implications of these peculiarities should be further investigated.

We observed striking differences among the free-living *L. biflexa*, the facultative pathogen *L. interrogans* and the host-dependent *L. borgpetersenii*. Most of the differences were consequence of gene gain from pathogenic bacteria rather than gene loss. DNA repair genes are overrepresented in *L. borgpetersenii* in relation to its genomic content, as observed for specialized bacteria with tiny genomes (Merhej, Royer-Carenzi et al. 2009). In fact, representation of such genes was slightly higher in the pathogenic *L. interrogans*. Phylogenetic analysis for DNA repair genes showed a distinct evolutionary

history compatible to lateral gene transfer. As observed for others (Chou et al. 2012; Ricaldi et al. 2012; Fonseca et al. 2013; Lehmann et al. 2013) phylogenetic origin of laterally transferred genes suggests a special ability of leptospires to exchange genetic material with other bacteria. However, although its natural competence was not determined, the presence of two *ssb* genes indicates the occurrence of such phenomena. To date, the only system by which leptospires could acquire genetic information is the phage system in *L. biflexa* (Faine et al. 1999). Recently, Zeng et al., 2013 identified one of the SSB orthologs (LA\_1676) as component of the extracellular proteome. Extracellular proteins play important roles not only in competence, but also in pathogenicity, and this new finding opens the possibility for a new function for SSB protein in *Leptospira* physiology.

We observed some correlation among *Leptospira*'s life-style and the occurrence of some genes, for instance photolyase-encoding genes. While *L. borgpetersenii* does not exhibit any ortholog, *L. biflexa* is provided by a photolyase from family II and by a *phra*-like protein. Phylogenetic analysis clustered the *phra*-like to the same branch of *Rhodobacter sphaeroides* protein (Fig. 2a). This new photolyase is characterized by a [4Fe-4S] cluster within the catalytic domain and its expression is upregulated in response to singlet oxygen and hydrogen peroxide (Geisselbrecht et al. 2012). The existence of a catalase gene downstream of *L. biflexa phra*-like gene reinforces a similar mode of regulation.

*Leptospira* spp is provided with an unusual number of genes related to reversion or removal of alkylation lesions (Fig. 1). There are several natural sources of DNA-alkylating agents, from natural antibiotic to normal metabolic products, including intracellular S-adenosylmethionine (SAM) and alkylating stress generated by immune system (Friedberg et al. 2006). All species have *O*<sup>6</sup>-alkylguanine-DNA alkyltransferase I (*O*<sup>6</sup>AGT I) which in *E. coli* regulates four genes, *ada*, *aidB*, *alkA* and *alkB* (Friedberg et al. 2006). However, the existence of such mechanism in leptospires was not documented so far. The DNA binding motif is absent in *O*<sup>6</sup>AGT I from leptospires as already documented for other bacteria (Martins-Pinheiro et al. 2007). Our survey also revealed the existence of a novel class of alkyltransferases in leptospires, the alkyltransferase-like (ATLs) (see Fig.2b). They were described as homologues of alkyltransferases lacking the alkyl acceptor cysteine, which is replaced by other residues (Margison, Butt et al. 2007). Although these proteins seem to overlap substrate specificity with other alkyltransferases, the absence of intrinsic alkyltransferase, glycosylase, or endonuclease activities reinforces the idea of their interaction with other DNA repair proteins to remove the adducts (Pegg 2011; Schärer 2012). In fact, recent works suggest they bind alkylated DNA and shunt the damaged DNA into nucleotide excision repair pathway (Wilkinson et al. 2012).

In addition, leptospires possess several alkyl-DNA glycosylases, and *L. interrogans* harbors the higher number of genes with this function. Orthologues for AlkC and AlkD, with narrow phyletic distribution, were found in all three leptospires (Fig. 1). AlkC and D 3-methyladenine glycosylases are single domain proteins specific for removal of alkylated bases, with high specificity (Alseth et al. 2006). Beyond *alkC* and the two *alkD*, *L. interrogans* also bears *tag* and two *alkA*. Phylogenetic analysis suggested the acquisition of LIC12362/LA1370 (*AlkA2*, shaded in Fig. 3b) through a LGT event. Surprisingly, while *AlkA1* was barely expressed *in vitro* and *in vivo* conditions, *AlkA2* was up-regulated in kidneys from susceptible mice during acute phase of infection (Fig. 6).

Enrichment of orthologous for nucleotide pool sanitization genes in pathogenic species might be associated with oxidative burst promoted by host macrophages. According to this possibility, Luo and colleagues, 2011 identified a virulent-attenuated mutant for one nudix-hydrolase in Lai serovar (LA\_3977). However, they described the absence of the corresponding protein in saprophytic *L. biflexa*, which is not corroborated by our similarity searches. As shown in Table S4, LA\_3977 has orthologs in all three species.

Nucleotide excision repair is fully represented in leptospires, but the most unanticipated finding was the identification of four genes encoding UvrD-like proteins (Fig. 1 and 5): UvrD, a member of the superfamily 1A helicase, the similar gram-positive PcrA, the gram-negative Rep helicase and the exclusive short version of UvrD, specific to Leptospiraceae. Superfamily 1A helicases are highly processive 3'-5' ssDNA translocases and are involved in multiple cellular processes (Friedberg et al. 2006). To our knowledge *Leptospira* spp is the first organism that shows this apparent redundancy for such helicases and their functional interplay is subject for future studies. UvrD is required for releasing the excised oligonucleotide from the post-incision complex in both NER and MMR, and this action depends on its interaction with other proteins. Functional activity of the new short UvrD helicase is still an enigma and future experimental studies are required. In *E. coli*, the truncated form of UvrD obtained through deletion of 40 or 102 residues for the C-terminal region shows reduced affinity for single-stranded DNA, but retains helicase activity on a variety of substrates (Manelyte et al. 2009).

Post replicative mismatch repair has two special characteristics in leptospires: the occurrence of a second *MutS* gene and a *dam* methylase in *L. borgpetersenii*. *MutS3* belongs to a recently described family with restricted distribution (Lin, et al. 2007) and it is up-regulated after exposition of *L. interrogans* to UV-irradiation (Fonseca *et al.*, unpublished data). On the other hand, *dam* was identified only in host-restricted *L. borgpetersenii*. Functional activity of Dam methylase is well known in Gammaproteobacteria, as being



responsible for the epigenetic modification of DNA adenine through methylation. Dam plays special role in host-pathogen interactions through the regulation of virulence gene expression in *Escherichia coli* (Marinus and Casadesus, 2009). The impact of Dam activity in *L. borgpetersenii* was not investigated yet, but special attention should be directed to this aspect.

To repair DSB and destabilized replication forks, leptospire have the whole apparatus of homologous recombination (Fig. 1). In addition to RecQ, leptospire share RecS helicase with gram-positives. RecQ helicases are characterized by three conserved regions: 1. helicase domain, essential for ATP binding and hydrolysis; 2. RecQ carboxy-terminal (RQC) domain, restricted to RecQ family members and 3. Helicase and RNase D C-terminal (HRDC) domain, absent in RecS, might have an auxiliary role in nucleic acid binding (Chu and Hickson 2009).

During stationary phase, when a homologous chromosome is not available, bacteria rely on NHEJ. Only *L. interrogans* bears both components for NHEJ: a Gam-like protein and an ATP-dependent ligase. Gam-like is homologous to the eukaryotic Ku protein, relatively rare in bacteria. There is no obvious phylogenetic pattern between the bacteria that possess it. Therefore, NHEJ systems may most commonly be acquired by LGT events (Bowater and Doherty 2006). Transition from dividing cell to a more quiescent cellular state may be triggered by environmental changes, followed by cellular adaptation, such as sporulation or microfilm formation (Wilson et al 2003). Up-regulation of *ligD* in our experimental susceptible mice model suggests NHEJ requirement during infection. This could indicate a critical period for *Leptospira*'s adaptation during the transition phase from environment to host.

Despite the presence of such DNA repair mechanisms, it is likely that the replication machinery would still encounter misrepaired or even unrepaired lesions. In this scenario, low fidelity polymerases are recruited to promote translesion DNA synthesis (TLS). DNA pol IV was the sole TLS polymerase found in free-living *L. biflexa* and pathogenic *L. interrogans*. Absence of any TLS polymerases in *L. borgpetersenii* indicates its incapability to deal with DNA lesions blocking DNA replication. As consequence, reduced survival under such stressing situations is expected. The same proposition was done by Boscaro et al., 2013 for the genome-reduced symbiont *Polynucleobacter necessarius*.

To ensure genome stability, bacteria coordinate a complex network to ensure survival and adaptation to changing environments, the SOS response (Baharoglu and Mazel 2014). Recently, we identified the presence of a second ortholog for LexA only in the serovar Copenhageni (Fonseca et al. 2013). The so-called LexA2, probably acquired by LGT, is fully functional and able to control its own expression. Transcriptional analysis associated with electrophoretic mobility assays allowed us to expand our knowledge for SOS regulon composition in this serovar. LexA1 controls the expression of at least nine genes: *lexA1*, *lexA2*, *recA*, *recN*, *dinP*, and four genes encoding hypothetical proteins. Nevertheless, the components of SOS regulon might diverge among the genus, and this could explain the striking difference of sensitivity among *L. biflexa* and *L. interrogans* when exposed to UV irradiation. As discussed above, NER and HR, responsible for repair UV-induced DNA lesions, are complete in both species. However, we and others (Stamm and Charon 1988; Fonseca et al. 2013) showed strong sensitivity of *L. interrogans* to this genotoxic treatment. While *uvr* genes are not up-regulated following UV exposition in *L. interrogans* (Fonseca et al. 2013), the same is not true for *L. biflexa* (Bomediano et al., unpublished data). Therefore, it is reasonable to propose that this difference in gene expression regulation is responsible for the higher sensitivity of *L. interrogans* to this environmental mutagenic agent. This assumption is supported by our expression analysis, as SOS components and *uvrA* genes are up-regulated during the acute phase of infection in kidneys. Up-regulation of *uvrA* during infection implies a SOS-independent transcriptional control of NER components in *Leptospira interrogans* to deal with DNA insults introduced during infection. The existence of a second DNA damage response pathway was observed for *M. tuberculosis* (Rand et al. 2003) and this possibility is not excluded for leptospira.

Adaptation of leptospira to host requires DNA repair, as suggested by our expression analysis. During transition from a free-living to infective state, Caimano and colleagues, 2014 evaluated the transcriptional profile of leptospire to host adaptation. mRNA levels from SMC Mre11-Rad50 proteins were significantly up-regulated. In addition, LIC\_12653, regulated by both LexA repressors, was highly expressed within the peritoneal cavities of rats. Our investigation explored the recruitment of some DNA repair genes and SOS components during infection in animal model. Most of the tested genes showed up-regulation in the infected organ from susceptible mice. Therefore, as intermediate mice did not have the same expression pattern, DNA repair recruitment is not associated with immunological assaults that control

infection. Up-regulation of *tag* should be highlighted, as the induction of the *E. coli tag* gene was never reported. mRNA from genes acquired by LGT, as *Alk2*, *ligD* and *tag* were enriched, indicating their acquisition played special role during the evolution of the genus. On the other hand, SOS components were up-regulated in both, susceptible and intermediate animals. Thus, most probably, SOS activation is triggered by the environmental change (EMJH to kidney), which induces new challenges to the bacterium. SOS activation in the urinary tract was also observed for UPEC pathogenesis (Li et al. 2010). Requirement for RecA and LexA activity was essential for UPEC survival during infection, what is consistent with DNA damage induction by the host intracellular environment. Our results point out to similar destabilization of *Leptospira* genome during kidney infection.

#### **COMPETING INTERESTS**

The authors declare that they have no competing interests.

#### **AUTHORS' CONTRIBUTIONS**

Conceived and designed the experiments: MMP, LSF, JBS and RMAC. Performed the experiments: MMP, LSF, JBS, LHSM and RMAC. Analyzed the data: MMP, LSF, JBS, LHSM, ACQS, PLH and RMAC. Contributed reagents/materials/analysis tools: PLH and RMAC. Wrote the manuscript: MMP, LSF, JBS, ACQS, PLH and RMAC.

#### **ACKNOWLEDGMENTS**

This work was funded by FAPESP (projects 2010/51365-0; 2009/17123-2, [www.fapesp.br](http://www.fapesp.br)) and CNPq (project 482043/2009-1, [www.cnpq.br](http://www.cnpq.br)). We thank Dr. Carlos F. M. Menck for enlightening discussions.



## REFERENCES

- Alseth I, Rognes T, et al. (2006) A new protein superfamily includes two novel 3-methyladenine DNA glycosylases from *Bacillus cereus*, AlkC and AlkD. *Mol Microbiol* 59(5): 1602-1609. doi 10.1111/j.1365-2958.2006.05044.x
- Altschul SF, Gish W, et al. (1990) Basic local alignment search tool. *J Mol Biol* 215(3): 403-410. doi 10.1016/S0022-2836(05)80360-2
- Baharoglu Z and Mazel D (2014) SOS, the formidable strategy of bacteria against aggressions FEMS *Microbiol Rev* 38(6): 1126-1145. doi 10.1111/1574-6976.12077
- Bhatnagar SK and Bessman MJ (1988) Studies on the mutator gene, mutT of *Escherichia coli*. Molecular cloning of the gene, purification of the gene product, and identification of a novel nucleoside triphosphatase. *J Biol Chem* 263(18): 8953-8957
- Boscaro V, Felletti M, et al. (2013) *Polynucleobacter necessarius*, a model for genome reduction in both free-living and symbiotic bacteria. *Proc Natl Acad Sci U S A* 110(46): 18590-18595. doi 10.1073/pnas.1316687110
- Bowater R and Doherty AJ (2006) Making ends meet: repairing breaks in bacterial DNA by non-homologous end-joining *PLoS Genet* 2(2): e8. doi 10.1371/journal.pgen.0020008
- Bulach DM, Zuerner RL, et al. (2006) Genome reduction in *Leptospira borgpetersenii* reflects limited transmission potential. *Proc Natl Acad Sci U S A* 103(39): 14560-14565. doi 10.1073/pnas.0603979103
- Caimano MJ, Sivasankaran SK, et al. (2014) A Model System for Studying the Transcriptomic and Physiological Changes Associated with Mammalian Host-Adaptation by *Leptospira interrogans* Serovar Copenhageni. *PLoS Pathog* 10(3): e1004004. doi 10.1371/journal.ppat.1004004
- Chou LF, Chen YT, et al. (2012) Sequence of *Leptospira santarosai* serovar Shermani genome and prediction of virulence-associated genes. *Gene* 511(2): 364-370. doi 10.1016/j.gene.2012.09.074
- Chu WK and Hickson ID (2009) RecQ helicases: multifunctional genome caretakers. *Nat Rev Cancer* 9(9): 644-654. doi 10.1038/nrc2682
- Ciccia A and Elledge SJ (2010) The DNA damage response: making it safe to play with knives. *Mol Cell* 40(2): 179-204. doi 10.1016/j.molcel.2010.09.019
- Cox EC and Yanofsky C (1969) Mutator gene studies in *Escherichia coli*. *J Bacteriol* 100(1): 390-397
- Cuñé J, Cullen P, et al. (2005) The *Leptospira interrogans* *lexA* gene is not autoregulated. *J Bacteriol* 187(16): 5841-5845. doi 10.1128/JB.187.16.5841-5845.2005
- d'Adda di Fagagna F, Weller GR, et al. (2003) The Gam protein of bacteriophage Mu is an orthologue of eukaryotic Ku. *EMBO Rep* 4(1):47-52. DOI 10.1038/sj.embor.embor709
- da Silva JB, Carvalho E, et al. (2012) Induction of TNF- $\alpha$  and CXCL-2 mRNAs in different organs of mice infected with pathogenic *Leptospira*. *Microb Pathog* 52(4): 206-216. doi 10.1016/j.micpath.2012.01.002
- da Silva JB, Ramos TM, et al. (2009) Chemokines expression during *Leptospira interrogans* serovar Copenhageni infection in resistant BALB/c and susceptible C3H/HeJ mice. *Microb Pathog* 47(2): 87-93. doi 10.1016/j.micpath.2009.05.002
- Dalhus B, Laerdahl JK, et al. (2009) DNA base repair--recognition and initiation of catalysis. *FEMS Microbiol Rev* 33(6): 1044-1078. doi 10.1111/j.1574-6976.2009.00188.x
- Evangelista KV and Coburn J (2010) *Leptospira* as an emerging pathogen: a review of its biology, pathogenesis and host immune responses. *Future Microbiol* 5(9): 1413-1425. doi 10.2217/fmb.10.102
- Faine S, Adler B, et al. (1999) *Leptospira* and Leptospirosis. In. Melbourne, Australia: MediSci.
- Falnes P and Rognes T (2003) DNA repair by bacterial AlkB proteins. *Res Microbiol* 154(8): 531-538. doi 10.1016/S0923-2508(03)00150-5
- Fonseca LS, da Silva JB, et al. (2013) *Leptospira interrogans* serovar copenhageni harbors two *lexA* genes involved in SOS response. *PLoS One* 8(10): e76419. doi 10.1371/journal.pone.0076419
- Foti JJ, Devadoss B, et al. (2012) Oxidation of the guanine nucleotide pool underlies cell death by bactericidal antibiotics. *Science* 336(6079): 315-319. doi 10.1126/science.1219192

- Friedberg EC, Aguilera A, et al. (2006) DNA repair: from molecular mechanism to human disease. *DNA Repair (Amst)* 5(8): 986-996
- Galhardo RS, Rocha RP, et al. (2005) An SOS-regulated operon involved in damage-inducible mutagenesis in *Caulobacter crescentus*. *Nucleic Acids Res* 33(8): 2603-2614. doi 10.1093/nar/gki551
- Ganesan A, Spivak G, et al. (2012) Transcription-coupled DNA repair in prokaryotes. *Prog Mol Biol Transl Sci* 110: 25-40. doi 10.1016/B978-0-12-387665-2.00002-X
- Geisselbrecht Y, Frühwirth S, et al. (2012) CryB from *Rhodobacter sphaeroides*: a unique class of cryptochromes with new cofactors. *EMBO Rep* 13(3): 223-229. doi 10.1038/embor.2012.2
- Guarné A (2012) The functions of MutL in mismatch repair: the power of multitasking. *Prog Mol Biol Transl Sci* 110: 41-70. doi 10.1016/B978-0-12-387665-2.00003-1
- Ko AI, Goarant C, et al. (2009) *Leptospira*: the dawn of the molecular genetics era for an emerging zoonotic pathogen. *Nat Rev Microbiol* 7(10): 736-747. doi 10.1038/nrmicro2208
- Kunkel TA and Erie DA (2005) DNA mismatch repair. *Annu Rev Biochem* 74: 681-710. doi 10.1146/annurev.biochem.74.082803.133243
- Labahn J, Schärer OD, et al. (1996) Structural basis for the excision repair of alkylation-damaged DNA. *Cell* 86(2): 321-329
- Laganeckas M, Margelevicius M, et al. (2011) Identification of new homologs of PD-(D/E)XK nucleases by support vector machines trained on data derived from profile-profile alignments. *Nucleic Acids Res* 39(4): 1187-1196. doi 10.1093/nar/gkq958
- Landini P and Volkert MR (2000) Regulatory responses of the adaptive response to alkylation damage: a simple regulon with complex regulatory features. *J Bacteriol* 182(23): 6543-6549
- Lehmann JS, Fouts DE, et al. (2013) Pathogenomic inference of virulence-associated genes in *Leptospira interrogans*. *PLoS Negl Trop Dis* 7(10): e2468. doi 10.1371/journal.pntd.0002468
- Lenhart JS, Schroeder JW, et al. (2012) DNA repair and genome maintenance in *Bacillus subtilis*. *Microbiol Mol Biol Rev* 76(3): 530-564. doi 10.1128/MMBR.05020-11
- Levett PN (2001) Leptospirosis. *Clin Microbiol Rev* 14(2): 296-326. doi 10.1128/CMR.14.2.296-326.2001
- Ley C, Ley C, et al. (2013) Detecting outliers: Do not use standard deviation around the mean, use absolute deviation around the median. *Journal of Experimental Social Psychology* 49(4): 764-766. doi <http://dx.doi.org/10.1016/j.jesp.2013.03.013>
- Li B, Smith P, et al. (2010) SOS regulatory elements are essential for UPEC pathogenesis. *Microbes Infect* 12(8-9): 662-668. doi 10.1016/j.micinf.2010.04.009
- Lin Z, Nei M, et al. (2007) The origins and early evolution of DNA mismatch repair genes--multiple horizontal gene transfers and co-evolution. *Nucleic Acids Res* 35(22): 7591-7603. doi 10.1093/nar/gkm921
- Lindner C, Nijland R, et al. (2004) Differential expression of two paralogous genes of *Bacillus subtilis* encoding single-stranded DNA binding protein. *J Bacteriol* 186(4): 1097-1105
- Livak KJ, Wills QF, et al. (2013) Methods for qPCR gene expression profiling applied to 1440 lymphoblastoid single cells. *Methods* 59(1): 71-79. doi 10.1016/j.ymeth.2012.10.004
- Luo Y, Liu Y, et al. (2011) InvA protein is a Nudix hydrolase required for infection by pathogenic *Leptospira* in cell lines and animals. *J Biol Chem* 286(42): 36852-36863. doi 10.1074/jbc.M111.219931
- Manelyte L, Guy CP, et al. (2009) The unstructured C-terminal extension of UvrD interacts with UvrB, but is dispensable for nucleotide excision repair. *DNA Repair (Amst)* 8(11): 1300-1310. doi 10.1016/j.dnarep.2009.08.005
- Margison GP, Butt A, et al. (2007) Alkyltransferase-like proteins. *DNA Repair (Amst)* 6(8): 1222-1228. doi 10.1016/j.dnarep.2007.03.014
- Marinus MG and Casadesus J (2009) Roles of DNA adenine methylation in host-pathogen interactions: mismatch repair, transcriptional regulation, and more. *FEMS Microbiol Rev* 33(3): 488-503. doi 10.1111/j.1574-6976.2008.00159.x0
- Martins-Pinheiro M, Marques RC, et al. (2007) Genome analysis of DNA repair genes in the alpha proteobacterium *Caulobacter crescentus*. *BMC Microbiol* 7: 17. doi 10.1186/1471-2180-7-17
- McHenry CS (2011) Bacterial replicases and related polymerases. *Curr Opin Chem Biol* 15(5): 587-594. doi 10.1016/j.cbpa.2011.07.018
- Menck CF (2002) Shining a light on photolyases. *Nat Genet* 32(3): 338-339. doi 10.1038/ng1102-338

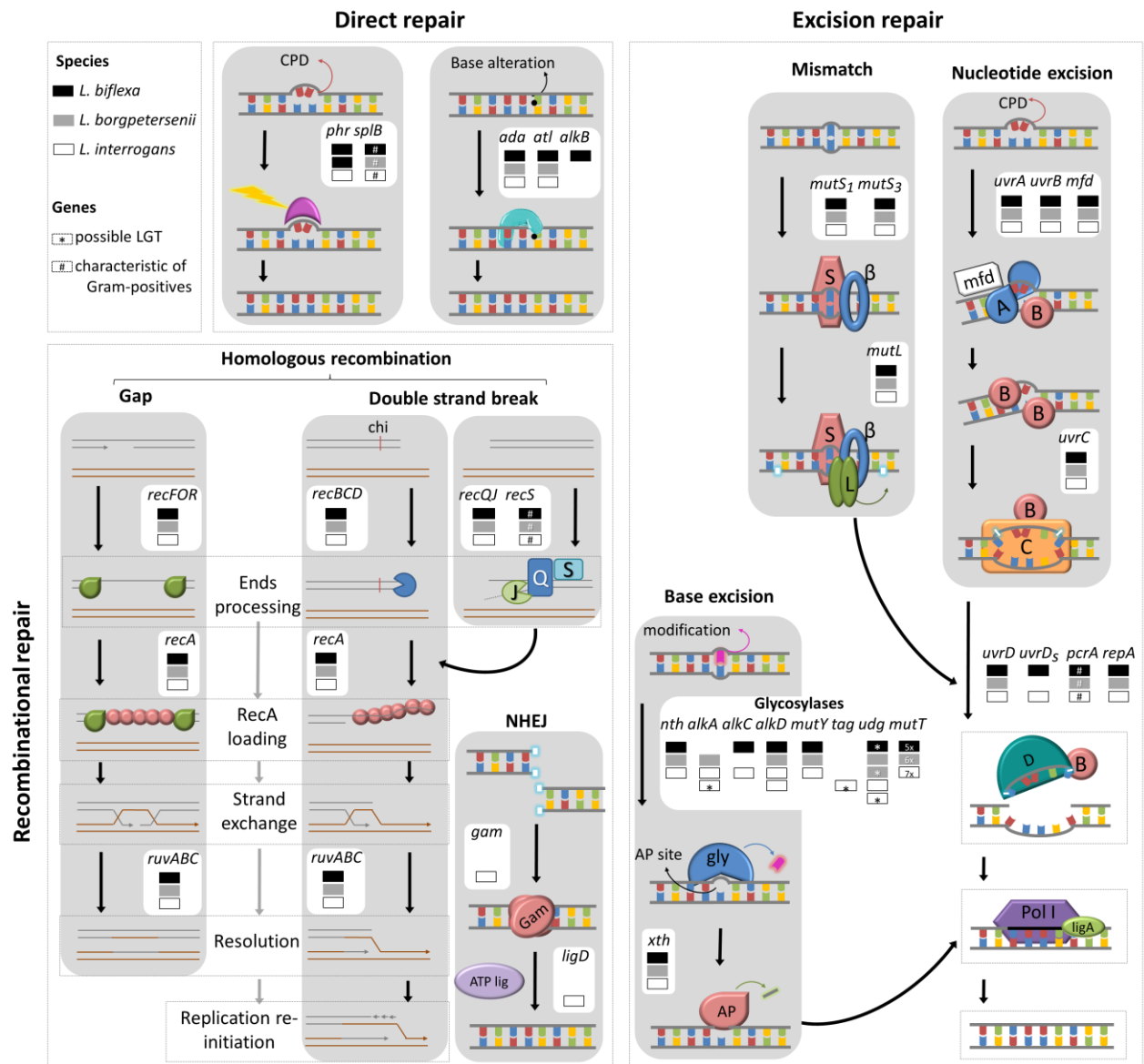
- Merhej V, Royer-Carenzi M, et al. (2009) Massive comparative genomic analysis reveals convergent evolution of specialized bacteria. *Biol Direct* 4: 13. doi 10.1186/1745-6150-4-13
- Nascimento AL, Ko AI, et al. (2004) Comparative genomics of two *Leptospira interrogans* serovars reveals novel insights into physiology and pathogenesis. *J Bacteriol* 186(7): 2164-2172
- Pegg AE (2011) Multifaceted roles of alkyltransferase and related proteins in DNA repair, DNA damage, resistance to chemotherapy, and research tools. *Chem Res Toxicol* 24(5): 618-639. doi 10.1021/tx200031q
- Pegg AE and Byers TL (1992) Repair of DNA containing O6-alkylguanine. *FASEB J* 6(6): 2302-2310
- Petit MA and Ehrlich D (2002) Essential bacterial helicases that counteract the toxicity of recombination proteins. *EMBO J* 21(12): 3137-3147. doi 10.1093/emboj/cdf317
- Picardeau M, Bulach DM, et al. (2008) Genome sequence of the saprophyte *Leptospira biflexa* provides insights into the evolution of *Leptospira* and the pathogenesis of leptospirosis. *PLoS One* 3(2): e1607. doi 10.1371/journal.pone.0001607
- Pitcher RS, Brissett NC, et al. (2007) Nonhomologous end-joining in bacteria: a microbial perspective. *Annu Rev Microbiol* 61: 259-282. doi 10.1146/annurev.micro.61.080706.093354
- Punta M, Coggill PC, et al. (2012) The Pfam protein families database. *Nucleic Acids Res* 40(Database issue): D290-301. doi 10.1093/nar/gkr1065
- Qin JH, Zhang Q, et al. (2008) Identification of a novel prophage-like gene cluster actively expressed in both virulent and avirulent strains of *Leptospira interrogans* serovar Lai. *Infect Immun* 76(6): 2411-2419. doi 10.1128/IAI.01730-07
- Rand L, Hind J, et al. (2003) The majority of inducible DNA repair genes in *Mycobacterium tuberculosis* are induced independently of RecA. *Mol Microbiol*. 50(3):1031-1042.
- Ren SX, Fu G, et al. (2003) Unique physiological and pathogenic features of *Leptospira interrogans* revealed by whole-genome sequencing. *Nature* 422(6934): 888-893. doi 10.1038/nature01597
- Ricaldi JN, Fouts DE, et al. (2012) Whole genome analysis of *Leptospira licerasiae* provides insight into leptospiral evolution and pathogenicity. *PLoS Negl Trop Dis* 6(10): e1853. doi 10.1371/journal.pntd.0001853
- Rogozin IB, Makarova KS, et al. (2008) A highly conserved family of inactivated archaeal B family DNA polymerases. *Biol Direct* 3: 32. doi 10.1186/1745-6150-3-32
- Schärer OD (2012) Alkyltransferase-like proteins: brokers dealing with alkylated DNA bases. *Mol Cell* 47(1): 3-4. doi 10.1016/j.molcel.2012.06.030
- Stamm LV and Charon NW (1988) Sensitivity of pathogenic and free-living *Leptospira* spp. to UV radiation and mitomycin C. *Appl Environ Microbiol* 54(3): 728-733
- Söding J, Biegert A, et al. (2005) The HHpred interactive server for protein homology detection and structure prediction. *Nucleic Acids Res* 33(Web Server issue): W244-248. doi 10.1093/nar/gki408
- Tamura K, Dudley J, et al. (2007) MEGA4: Molecular Evolutionary Genetics Analysis (MEGA) software version 4.0. *Mol Biol Evol* 24(8): 1596-1599. doi 10.1093/molbev/msm092
- Tchamedeu Kameni AP, Couture-Tosi E, et al. (2002) Inactivation of the spirochete recA gene results in a mutant with low viability and irregular nucleoid morphology. *Journal of bacteriology* 184(2): 452-458
- Thompson JD, Gibson TJ, et al. (1997) The CLUSTAL\_X windows interface: flexible strategies for multiple sequence alignment aided by quality analysis tools. *Nucleic Acids Res* 25(24): 4876-4882
- Truglio JJ, Croteau DL, et al. (2006) Prokaryotic nucleotide excision repair: the UvrABC system. *Chem Rev* 106(2): 233-252. doi 10.1021/cr040471u
- Wigley DB (2013) Bacterial DNA repair: recent insights into the mechanism of RecBCD, AddAB and AdnAB. *Nat Rev Microbiol* 11(1): 9-13. doi 10.1038/nrmicro2917
- Wilkinson OJ, Latypov V, et al. (2012) Alkyltransferase-like protein (At11) distinguishes alkylated guanines for DNA repair using cation- $\pi$  interactions. *Proc Natl Acad Sci U S A* 109(46): 18755-18760. doi 10.1073/pnas.1209451109
- Wilson TE, Topper LM et al. (2003) Non-homologous end-joining: bacteria join the chromosome breakdance. *Trends Biochem Sci* 28(2): 62-66 doi 10.1016/S0968-0004(03)00005-7
- Yamagata Y, Kato M, et al. (1996) Three-dimensional structure of a DNA repair enzyme, 3-methyladenine DNA glycosylase II, from *Escherichia coli*. *Cell* 86(2): 311-319
- Yang L and Li L (2013) The enzyme-mediated direct reversal of a dithymine photoproduct in germinating endospores. *Int J Mol Sci* 14(7): 13137-13153. doi 10.3390/ijms140713137

Zeng L, Zhang Y et al. (2013) Extracellular proteome analysis of *Leptospira interrogans* serovar Lai. *OMICS* 17(10): 527-535. doi 10.1089/omi.2013.0043

Zhong Y, Chang X, et al. (2011) Comparative proteogenomic analysis of the *Leptospira interrogans* virulence-attenuated strain IPAV against the pathogenic strain 56601. *Cell Res* 21(8): 1210-1229. doi 10.1038/cr.2011.46

## Figure Legends

**Fig.1 Main DNA repair pathways present in *Leptospira* spp.**

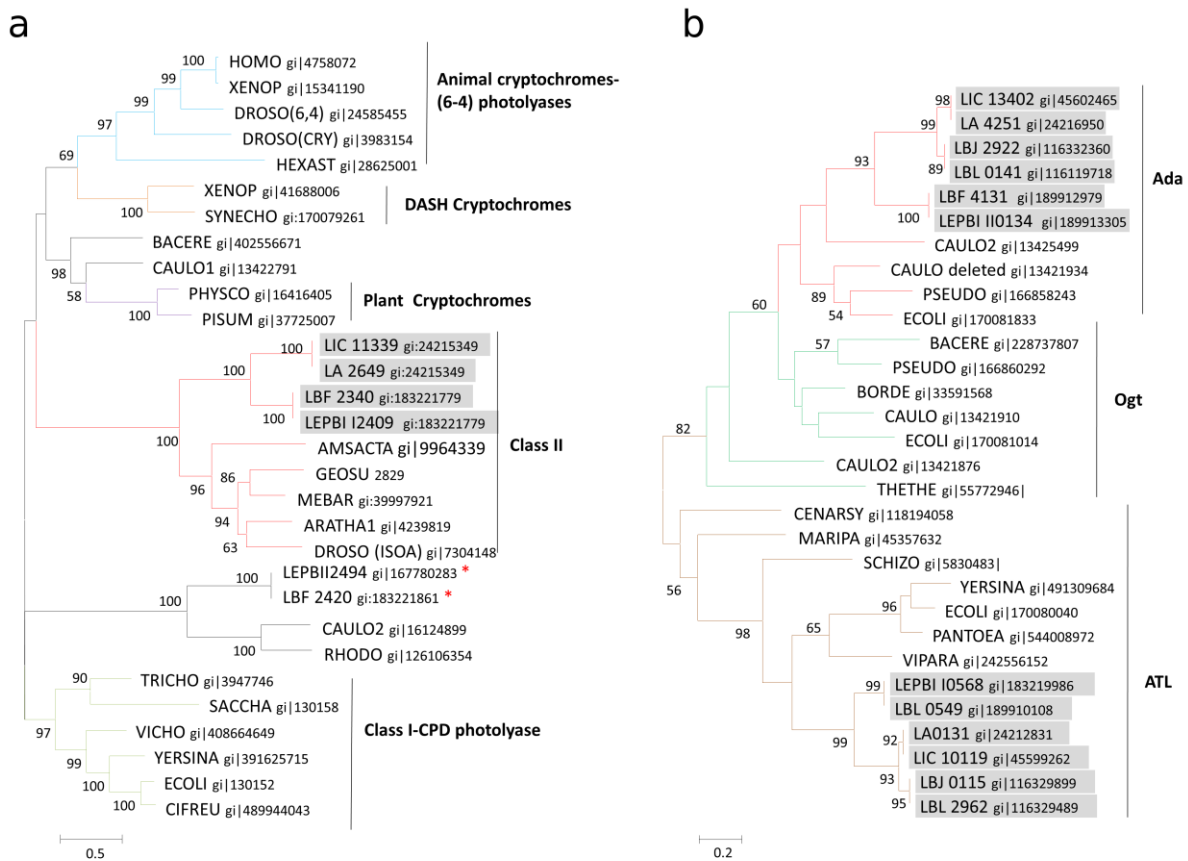


Schematic representation of DNA repair pathways found in *Leptospira* spp. In every pathway, the genes required for each step are shown, with black, gray or white boxes representing the genes identified in *L. biflexa*, *L. borgpetersenii* or *L. interrogans*, respectively. Asterisks indicate genes possibly acquired by LGT events and hashtags, genes found in Gram-positives. The direct repair is carried out by photolyases (*phra*, *phrb*, *splB*) or alkyltransferases (*ada*, *atl* and *alkB*). Excision repair requires *uvrABC* and in some



cases *mfd* for NER; mono (*alkA*, *alkC*, *alkD*, *mutY*, *tag*, *udg* or *mutY*) or bifunctional (*nth*) glycosylases and AP endonucleases (*xth*) for BER; and *mutSL*, interacting with the  $\beta$  clamp for MMR. The excision repair pathways are finished by one of the *uvrD*-like helicases *uvrD*, small *uvrD* (*uvrDs*), *pcrA* or *repA*, DNA polymerase I and DNA ligase (*ligA*). Homologous recombination requires RecFOR, RecBCD or RecJQ/S to process DNA end and load RecA. The products from *ruvABC* resolve the Holliday junctions. Finally, the two-component NHEJ pathway is found only in *L. interrogans* and consists of a Gam-like protein (*gam*) and an ATP-dependent ligase (*ligD*).

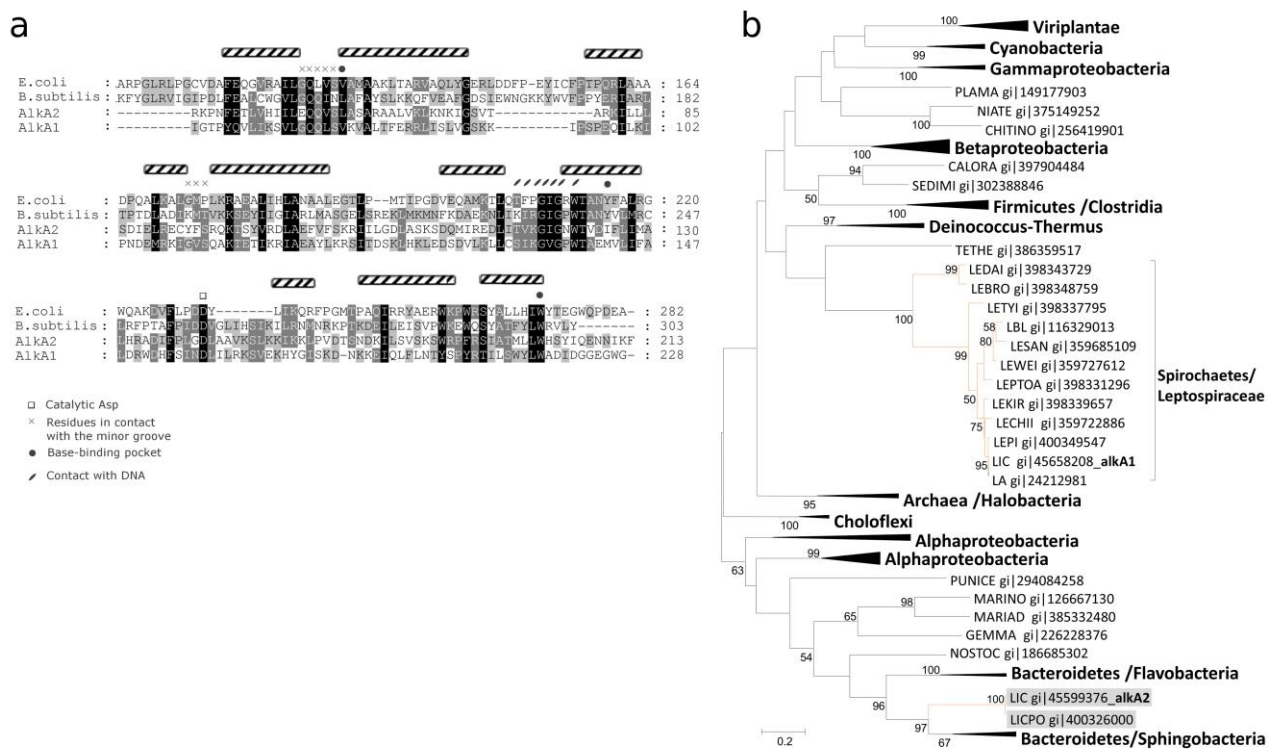
**Fig. 2 Phylogenetic relationships among photolyases (a) and putative alkyltransferase proteins (b)**



**(a)** Sequences were aligned by ClustalW program; gap-removed alignment was used for generating the phylogenetic tree using maximum likelihood inference. The common photolyase from *L. biflexa*, *L. interrogans* and close homologs are shaded. The *phra*-like from *L. biflexa* is highlighted by asterisks. Lines are colored to distinguish the phyletic pattern of the corresponding functional groups: animal cryptochromes and 6-4 photolyase (blue); DASH cryptochromes (orange); plant cryptochromes (violet); class II photolyase (red); Phra-like (black) and class I photolyase (green). **(b)** Relationships among alkyltransferase proteins were inferred using the maximum likelihood method. Branches are colored to

indicate the phyletic pattern of the corresponding functional groups: Ada (red); Ogt (green) and ATL (brown). The corresponding *Leptospira*'s orthologs are shaded. Both analyses were conducted in MEGA5. The bootstrap consensus trees were inferred from 1000 replicates and only bootstrap support levels  $\geq 50\%$  are shown. All position containing gaps and missing data were eliminated from the datasets. Accession codes of the proteins analyzed are presented after the abbreviated names of organisms. Organism names are shown in Table S2.

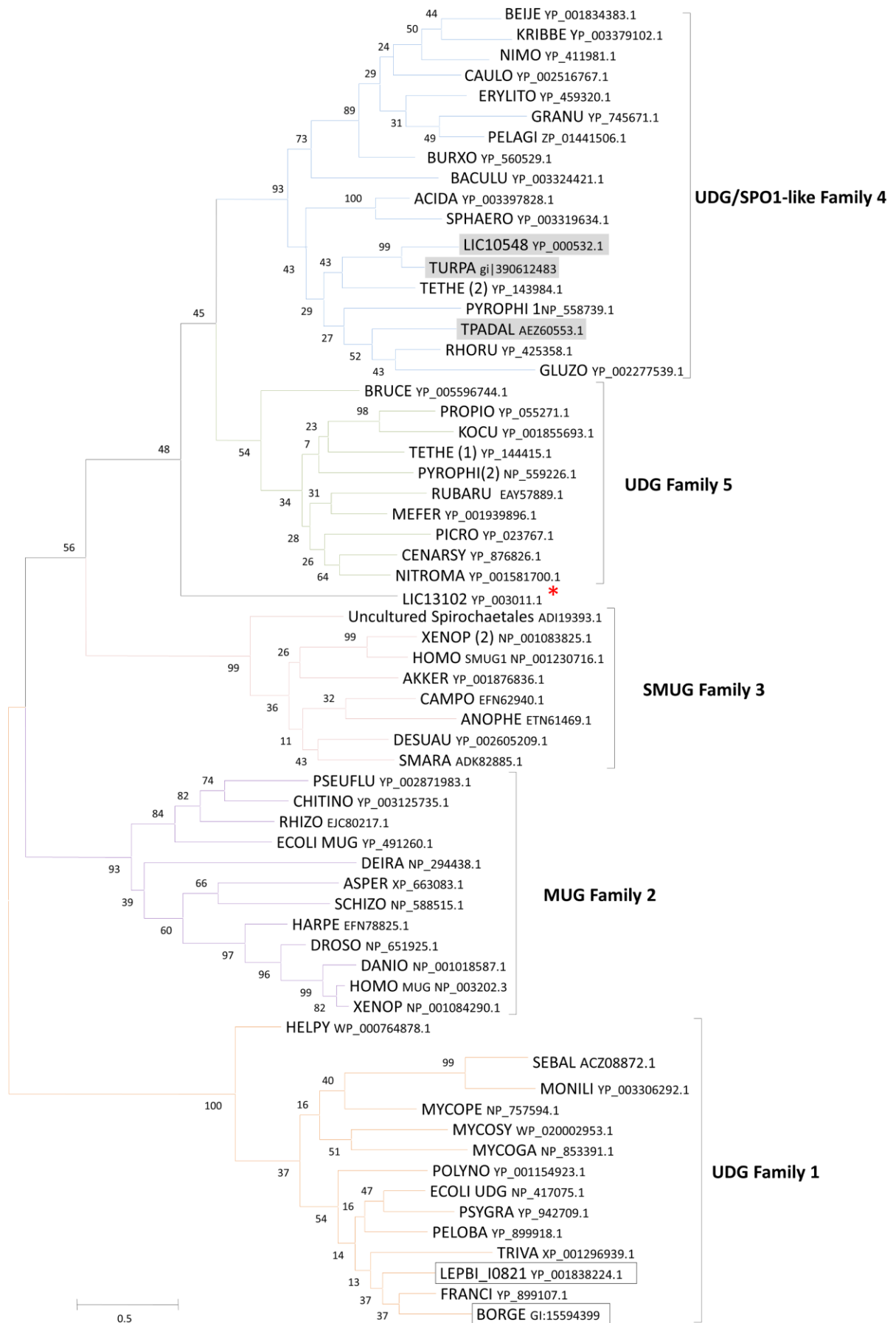
**Fig. 3 Comparison of AlkA amino acid sequences and phylogenetic analysis**



(a) Amino acid sequence alignment and secondary structure prediction of LIC10238/LA0281 (AlkA1) and LIC12362/LA1370 (AlkA2) were carried out using *E. coli* and *B. subtilis* AlkA as reference. The proteins are composed of 10 alpha helices (striped rectangles). The residues responsible for contact with DNA, forming the base-binding pocket and the catalytic Asp are indicated. (b) Phylogenetic analysis of AlkA amino acid sequences from several bacteria were done by maximum likelihood inference after alignment. The bootstrap consensus trees were inferred from 1000 replicates and only bootstrap support levels  $\geq 50\%$  are shown. *Leptospira* clade is colored in orange and indicated by the bracket. The two AlkA present in *L. interrogans* serovar Copenhageni are named alkA1 and alkA2 and the sequences exclusive to *L.*

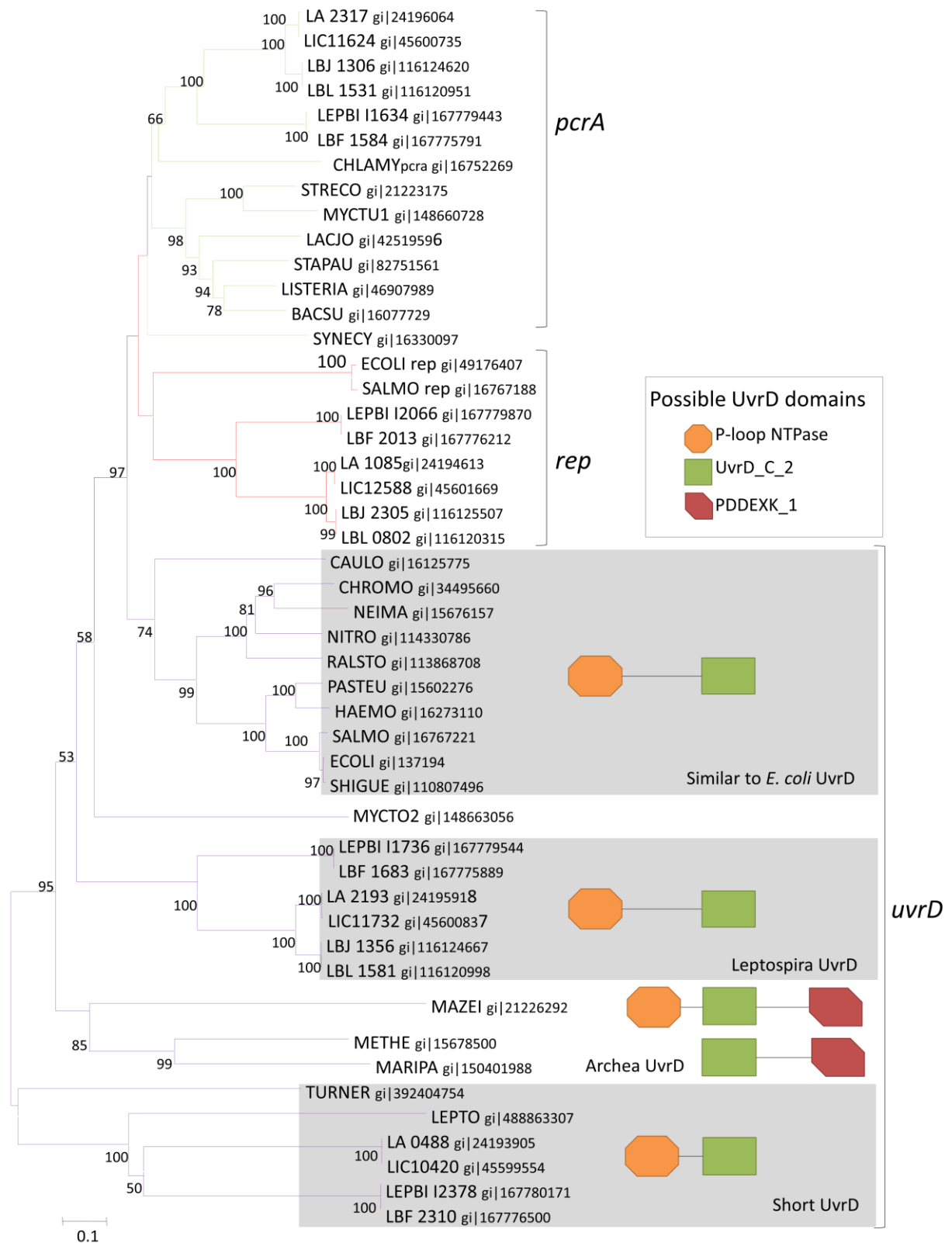
*interrogans* are shaded. Local bootstrap values are attached to the internal nodes and organism names are shown in Table S2.

**Fig. 4 Phylogenetic analysis of leptospiral UDGs**



Phylogenetic analysis was performed using amino acid sequences from several bacteria, including representatives from each of the five UDG families. Branches are colored to indicate the phyletic pattern of the corresponding functional groups: UDG/SPO1-like Family 4 (blue); UDG Family 5 (green); SMUG Family 3 (rose); MUG Family 2 (purple) and UDG Family 1 (orange). Proteins from pathogenic and saprophytic *Leptospira* spp clustered in different clades. Proteins from the Spirochaetes *T. pallidum*, *T. parva* and LIC\_10548, from *L. interrogans* serovar Copenhageni, were grouped in the same clade (UDG/SPO1-like Family 4) and are shaded. LIC\_13102 is indicated by asterisk. The proteins from *L. biflexa* and *B. burgdorferi* clustered with UDGs from Family 1 and are in rectangles. Local bootstrap values are attached to the internal nodes. Organism names are shown in Table S2

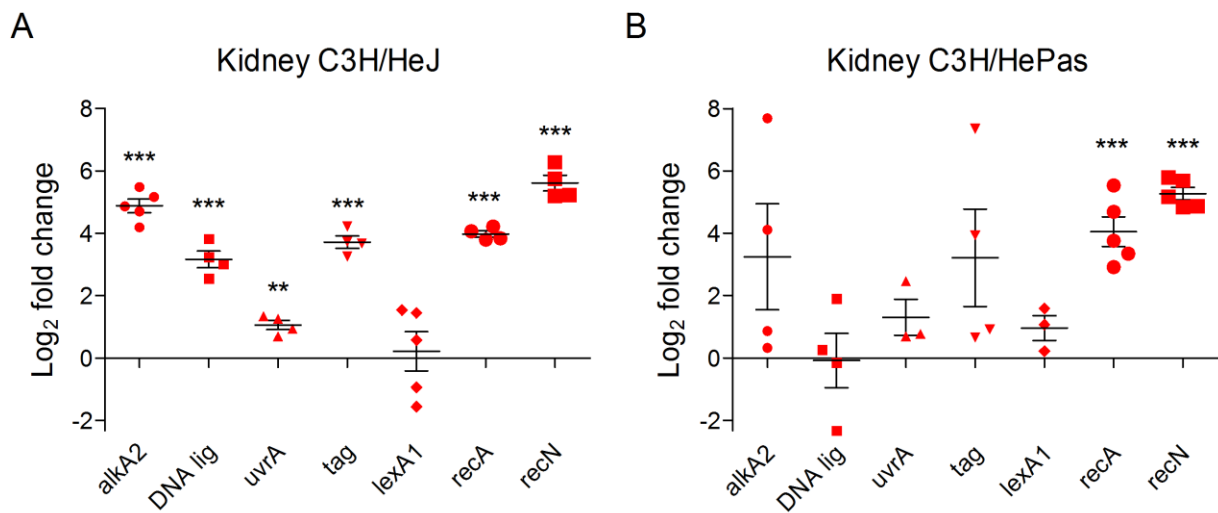
**Fig. 5 Phylogenetic tree for UvrD and related proteins**



The phylogenetic tree was constructed using Maximum likelihood implemented in MEGA5. Bootstrap values are reported in the figure only when higher than 50%. Schematic representations of domain structure present in UvrD and UvrD-like proteins are shown: the yellow boxes characterize P-loop NTPase domains;

the green box shows the C-terminal domain of UvrD\_C\_2; the PDDEXK domain (rose box) characterizes members from a new nuclease superfamily (Laganeckas, Margelevicius et al. 2011). *L. biflexa* and *interrogans*, together with the Spirochaetes *Leptonema illini* and *Turneriella parva* have a shorter UvrD helicase. The *Leptospira* members with this shorter UvrD have four genes coding for uvrD/rep helicase family. The accession codes of the proteins analyzed are presented after the abbreviated names of organisms. Organism names are shown in Table S2.

**Fig. 6** *In vivo* DNA repair-related genes expression



Intermediate (C3H/HePas) and susceptible (C3H/HeJ) mice were inoculated with virulent *L. interrogans* serovar Copenhageni. mRNAs from kidneys five days after mice infection were extracted, converted to cDNA and used as template for quantitative PCR analysis. mRNA transcript levels of *alkA2* (LIC\_12362), *ada* (LIC\_13402); *tag* (LIC\_13064); *uvrA* (LIC\_11717), *ATP dependent DNA ligase (DNA Lig; LIC\_12677)* and the SOS components *lexA1* (LIC\_12305), *recA* (LIC\_11745) and *recN* (LIC\_11620) were quantified. The corresponding Ct values were normalized to the 16S rRNA control and compared to the *in vitro* cultured (EMJH) leptospires. The gene expression profile is expressed logarithmically as the log<sub>2</sub> of the fold change between the two conditions. Experiments were performed in groups of three mice and were repeated on two separate occasions. The outlier values (calculated by MAD-median method) were excluded from the analyses. Two-tailed Student's t-test P values are shown as \*\*\* for P < 0.001, \*\* for P < 0.01 and \* for P < 0.05.