FIOCRUZ

FUNDAÇÃO OSWALDO CRUZ
CENTRO DE PESQUISAS GONÇALO MONIZ

Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa

TESE DE DOUTORADO

LEPTOSPIRA INTERROGANS SOROVAR COPENHAGENI E ICTEROHAEMORRHAGIAE: RELAÇÃO EVOLUTIVA, DIFERENÇAS GENÉTICAS E ASSOCIAÇÃO COM DESFECHO CLÍNICO

LUCIANE AMORIM SANTOS

Salvador - Bahia
2015

FUNDAÇÃO OSWALDO CRUZ
CENTRO DE PESQUISAS GONÇALO MONIZ

Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa

LEPTOSPIRA INTERROGANS SOROVAR COPENHAGENI E ICTEROHAEMORRHAGIAE: RELAÇÃO EVOLUTIVA, DIFERENÇAS GENÉTICAS E ASSOCIAÇÃO COM DESFECHO CLÍNICO

LUCIANE AMORIM SANTOS

Orientador: Prof. Dr. Luiz Carlos Júnior Alcântara
Co-orientador: Prof. Albert Ko

Tese apresentada ao Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa para a obtenção do grau de Doutor.

Salvador - Bahia
2015

"LEPTOSPIRA INTERROGANS SOROVAR COPENHAGENI E ICTEROHAEMORRHAGIAE: DIFERENÇAS GENÉTICAS E ASSOCIAÇÃO COM O DESFECHO CLÍNICO."

LUCIANE AMORIM SANTOS

FOLHA DE APROVAÇÃO

COMISSÃO EXAMINADORA

Dr. Nuno Rodrigues Faria
Professor
OXFORD

Dr. Márcio Roberto Teixeira Nunes
Pesquisador Associado
IEC

Dra. Paula Carvalhal Lage Von Buettner Ristow
Professora Adjunta
UFBA

“E ele muda os tempos e as
estações; ele remove os reis e
estabelece os reis; ele dá
sabedoria aos sábios e
conhecimento aos entendidos.”

Daniel 2:21

AGRADECIMENTOS

A Deus, “ao único Deus, nosso Salvador, mediante Jesus Cristo, Senhor nosso, glória, majestade, império e soberania, antes de todas as eras, e agora, e por todos os séculos. Amém!” (Jd 1:25).

Ao meu orientador Albert Ko que me deu a oportunidade de fazer parte deste trabalho me ensinando a ter uma visão de cientista e a fazer as perguntas importantes a serem respondidas pela ciência e a buscar a melhor abordagem para responde-las. Por me receber em seu laboratório e sua casa, sempre preocupado com o meu bem estar a adaptação a nova cidade e ao laboratório.

Ao meu orientador Dr Luiz Carlos Junior Alcântara pela grande contribuição na minha formação científica, pela paciência e incentivo, dando-me a oportunidade de desenvolver este trabalho, e me ensinando em todo o tempo.

Ao amigo, colega de trabalho e “littel boss” Elsio Wunder. A sua amizade, companhia e conselhos científicos tornaram os dias longe de casa e da família muito mais fáceis e divertidos.

Ao meu “papai godinho”, Roberto Santos, pelo amor, apoio e incentivo, sendo um exemplo de vida pra mim.

Á minha mãe e melhor amiga, Ana Lúcia, pelo amor, amizade e pelas palavras de sabedoria que sempre vem na hora certa e por nunca deixar de orar por mim.

Aos meus irmãos, Juliana e Gustavo, e cunhados, Anderson e Ana Carolina, pelos momentos de “lezeiras” e por sempre acreditarem em mim.

Á minha avó, Celina, pelo seu grande amor por mim, pelos agrados, paparicações e lanchinhos das madrugadas longas de trabalho.

Aos amigos Fernanda, Lua, Iukary, Jess, Bina, Gisa, Carol e Lipinho pela grande ajuda, paciência e amizade, sempre acreditando que sou capaz de muito mais. É muito bom conviver todos os dias com vocês

Aos amigos do Dr. Ko’s Lab e da Yale School of Public Heath pela amizade, apoio, suporte e importantes contribuições no meu aprendizado e formação.

Aos meus amigos baianos, cariocas, mineiros e americanos, pelas orações, incentivo, e grande amizade em todo o tempo. Vocês fazem a minha vida muito mais divertida.

A todos os professores do Curso de Pós-Graduação em Biotecnologia em Saúde e Medicina Investigativa com os quais tive a oportunidade e o prazer de aprender.

A todos os co-autores dos trabalhos pela confiança e colaboração.

A equipe da biblioteca do CPqGM pelo suporte no desenvolvimento e na formataçãoo desta tese.

Ao CNPq, CAPES, Fundação Lemann e NIH pelo apoio financeiro.

SANTOS, Luciane Amorim. *Leptospira interrogans* sorovar Copenhageni e Icterohaemorrhagiae: relação evolutiva, diferenças genéticas e associação com desfecho clínico. 95 f. il. Tese (Doutorado) – Fundação Oswaldo Cruz, Centro de Pesquisa Gonçalo Moniz, Salvador, 2015.

RESUMO

A leptospirose é a zoonose mais disseminada mundialmente por infectar diversas espécies diferentes de animais mamíferos. Apresenta 22 espécies identificadas, sendo dez patogênicas, cinco intermediarias e sete saprofiticas, além de apresentar mais de 250 sorovares diferentes. Em Salvador, *Leptospira interrogans* sorovar Copenhageni é a causadora da epidemia urbana na cidade e apresenta ratos como seu hospedeiro reservatório. As formas clínicas da leptospirose podem variar de assintomática a formas graves. As manifestações clínicas mais graves envolve o desenvolvimento da síndrome Hemorrágica pulmonar severa, e óbito do paciente. Estudos para entender as diferenças genéticas entre as diferentes espécies e sorovares é de extrema importância para identificar fatores de virulência da bactéria, genes que possam está associado aos diferentes formas clinicas, e sua capacidade de se adaptar aos diferentes ambientes. Neste trabalho foi estudado o genoma de dois importantes serovares de *L. interrogans*, o sorovar Copenhageni e o serovar Icterohaemorrhagiae, e suas diferenças genéticas e associação com dados clínicos e epidemiológicos. Um total de 141 isolados tiveram seus genomas sequenciados. Foi construindo e validado um *pipeline* para a o mapeamento e construção dos genomas e a identificação de SNPs e Indels. Os resultados encontrados demostraram um alta similaridade entre os isolados dos dois serovares, de diferentes regiões geográficas e isolados em anos diferentes. As sequências deste estudo se mostram conservadas ao longo do tempo sem apresentar nenhuma mutação associada as diferentes forma clínicas da doença, indicando que outros fatores, tais como os do hospedeiro, podem estar envolvidos na diversidade de sintomatologia. Na comparação do genoma dos isolados de *L. interrogans*, sorovar Copenhageni e sorovar Icterohaemorrhagiae foi identificado apenas uma mutação que as difere geneticamente. Essa mutação está presente no gene LIC12008 que produz uma proteína hipotética, e que a sua avaliação *in silico* demostrou estar envolvida na síntese de LPS, justificando assim as diferenças encontradas no teste serológico. Além disto, também foram avaliadas as diferenças entre 20 das 22 espécies de *Leptospira*, para identificar possíveis fatores de virulência e genes que possam estar envolvidos na patogênese e adaptação da bactéria ao ambiente. Estudos de fatores genéticos da *Leptospira* pode auxiliar ao manejo da doença, com uma melhor assistência e terapia para os pacientes, desenvolvimento de vacinas e diagnostico desta doença negligenciada.

Palavras-chave: *Leptospira spp.*, Genoma, Leptospirose, patogênese.

SANTOS, Luciane Amorim. *Leptospira interrogans* serovar copenhageni and icterohaemorrhagiae: evolutionary relationship, genetic diferences and association with clinical outcomes. 95 f. il. Tese (Doutorado) – Fundação Oswaldo Cruz, Centro de Pesquisa Gonçalo Moniz, Salvador, 2015.

ABSTRACT

Leptospirosis is a zoonosis disseminated worldwide, infecting a wide range of mammals species. There are 22 different species of *Leptospira spp.* in which 10 are pathogenic, 5 intermediate and 7 saprophytic species. In Salvador the *Leptospira interrogans* sorovar Copenhageni is the main serovar detected, responsible for the urban epidemics, and has rats as their main host. The clinical manifestations of leptospirosis can vary from asymptomatic form to severe disease like pulmonary hemorrhagic syndrome, and death. Studies to understand de genetic differences among the species and serovars are of great importance to identify virulence factors, genes that could be related to the different clinical manifestations and its capacity to adapt in different environments. Here, the genome of two epidemiologically important serovar of the *L. interrogans*, the serovar Copenhageni and serovar Icterohaemorrhagiae, and their genetic differences and the association of these differences with epidemiological and clinical data were studied. A total of 141 strains were genome sequenced. A pipeline for the genome mapping and variant call were constructed and validated. The results showed a high similarity among the strains from both serovars from different geographic locations and year of isolation. The sequences from this study showed to be very conserved, not presenting any mutation associated with the different clinical outcome, indicating that other factors, like host factors, could be related to the diversity of clinical outcome. Only one genetic mutation was detected in the genome comparison of the strains belonging to the *L. interrogans* sorovar Copenhageni and sorovar Icterohaemorrhagiae. This mutation was found in the gene LIC12008 that produce a hypothetical protein, in which its *in silico* analysis reviled that this protein could be related to the LPS synthesis, justifying the serological test differences between the two serovar. Besides that, the differences between 20 of the 22 species of *Leptospira* identified were evaluated to detect possibly virulence factors and genes that could be involved in the pathogenesis and adaptation. Studies of the *Leptospira* virulence factors can give support to the disease management, giving a better assistance and treatment to the patients and developing vaccines and better diagnostic for the neglected disease

Key words: *Leptospira spp.*, Genome, Leptospirose, pathogenesis.

LISTA DE FIGURAS

LISTA DE TABELAS

LISTA DE ABREVIATURAS E SIGLAS

BLAST	*Basic Local Alignment Search Tool*
BEAST	*Bayesian Evolutionary Analysis Sampling Trees*
Big	*bacterial immunoglobulin-like*
CAAT	teste de soro aglutinação cruzada (*cross-agglutinin absorption test*)
DNA	Ácido desoxirribonucléico (*Desoxyribonucleic Acid*)
GC	Guanina e Citosina
Indels	Inserção e Deleção
kb	Kilobase
Lig	*leptospiral immunoglobulin-like*
LPS	Lipopolissacarídeo
Mb	Megabase
ML	Máxima Verossimilhança (*Maximum Likelihood*)
MLST	Tipagem multilocus de sequência (*multilocus sequence typing*)
NIH	*National Institute of Health*
NJ	Agrupamento de vizinhos (*Neighbor-Joining*)
SHPS	Síndrome Hemorrágica Pulmonar Severa
SNP	Polimorfismos de nucleotídeo único (*Single Nucleotide Polimorphisms*)
VNTR	Número variável de repetições em tandem (*multilocus variable-number tandem-repeat*)

SUMÁRIO

1 INTRODUÇÃO

1.1 LEPTOSPIRA SPP.

A leptospirose é uma zoonose de importância global causada por espiroquetas do gênero *Leptospira*, que apresentam uma grande diversidade e são representadas por 22 espécies genômicas, sendo dez patogênicas, cinco intermediárias e sete saprófitas, além de mais de 250 sorovares diferentes (HAAPALA 1969; YASUDA et al, 1987; BRENNER et al, 1999; FAINE et al, 1999; LEVETT, 2001; SLACK et al, 2008; SAITO et al, 2013; BOURHY et al, 2014).

Em fevereiro de 1915, foi relatado o primeiro isolamento de *Leptospira* a partir de humanos por dois japoneses, Inada e Ido, sendo a bacteria nomeado de *Spirochaeta icterohaemorrhagiae*. Um ano depois, 1916, na Alemanha foi também relatado o isolamento de *Leptospira* a partir de amostras humanas e sendo então considerado o agente etiológico da síndrome de Weil's (WEIL, 1886). Atualmente, estes isolados tem o nome de Ictero 1 e RGA, respectivamente e são pertencentes a *L. interrogas* serovar Icterohaemorrhagiae (KMETY e DIKKEN, 1993). Ao longo dos anos, novas espécies e serovares patogênicas, intermediárias e saprofíticas foram relatadas em diferentes partes do mundo (FAINE et al, 1999; LEVETT, 2001).

As espécies patogênicas podem infectar humanos e animais e possuem uma afinidade específica pelos diferentes mamíferos onde são encontradas. Podem ser encontradas colonizando rim de roedores, porém sem causar doença. Ratos são um importante reservatório e transmissores da bactéria. Em animais como cachorros e animais de produção como bovinos, porcos e equinos, *Leptospira* pode causar danos hepáticos e renais, e no caso de mães infectadas, pode levar à morte do feto. Em humanos, a infecção pelas espécie patogênicas de *Leptospira*, apresenta uma ampla variedade de formas clínicas podendo variar de apresentação assintomática a doença grave culminando em morte (LEVETT, 2001; BHARTI et al, 2003). Estima-se que por ano, aproximadamente, 500.000 pessoas desenvolvem leptospirose grave no mundo, tornando a leptospirose um importante problema de saúde pública [WHO, 1999].

1.2 EPIDEMIOLOGIA DA LEPTOSPIROSE

A leptospirose apresenta distribuição mundial, porém possui uma maior ocorrência em climas tropicais e em países em desenvolvimento. Os casos de leptospirose tem um característica ocupacional quando ocorrem na zona rural. Um exemplo é na Ásia onde muitos casos estão associados a plantações de arroz devido ao contato com água contaminada nas plantações. Nos últimos anos, devido à rápida urbanização, o número de bairros sem saneamento basico tem crescido nos países em desenvolvimento. Nestes locais, o saneamento básico não existe ou é muito precário, com esgoto a céu aberto passando próximo às casas. Este ambiente é propício para a presença de roedores, tornando a transmissão da bactéria mais frequente. Na temporada de chuvas, a água contaminada pela bactéria entra nas casas aumentando o número de casos de leptospirose e criando uma relação com as estações chuvosas do ano. Devido a essas condições houve um grande aumentodo número de casos de leptospirose urbana, tornando-se um problema de saúde pública comum em países em desenvolvimento (KO et al, 1999).

Os maiores número de casos de leptospirose se concentram no Caribe, América Centrais e do Sul, além de casos no sudoeste da Ásia e Oceania (Figura 1). A República das Seicheles e Trinidade e Tobago ocupam a primeira e secunda posições dos países com um maior índice de incidência da infecção reportado no mundo. O Brasil ocupa a decima sétima colocação com um índice de 12.8 casos por um milhão de habitantes (PAPPAS et al, 2008). Com a rápida urbanização o número de casos vem crescendo no Brasil. A região sul e sudestes do país apresenta um maior número de casos compondo cerca de 69,1% dos casos notificados no Brasil (BVS, 2011). O nordeste representa 20,4% dos casos, onde Salvador, Bahia apesenta um soroprevalência de 12,4% (BVS, 2011; DIAS, 2007).

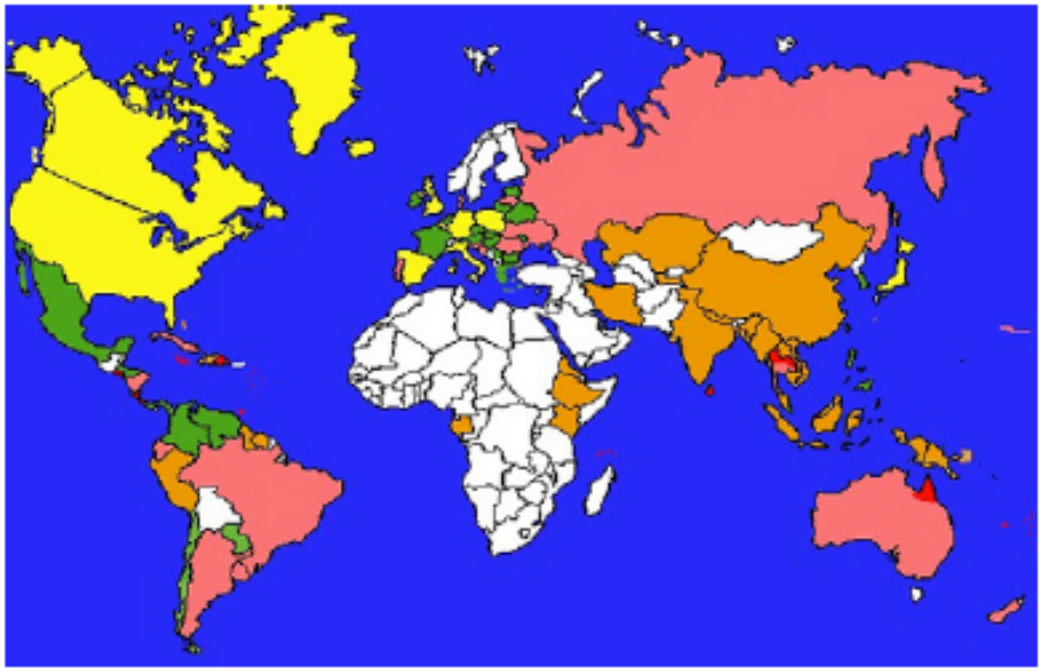


Figura 1: Incidência global de casos humanos de leptospirose. As cores refletem a incidência, em ordem decrescente: vermelho, rosa, verde, amarelo. Dourado representa áreas prováveis, porém não estimado, de apresentarem uma alta incidência. Branco reflete locais onde não se tem dados. Adaptado de PAPPAS et al, 2008.

Leptospira tem a capacidade de colonizar os rins do seu hospedeiro reservatório. Neste hospedeiro, como por exemplo os ratos, a *Leptospira* não causam doença e permanecem por um longo período nos rins destes animais sendo eliminada para o meio ambiente através da sua urina. A transmissão da bactéria ocorre pelo contato direto da pele e mucosas não intactas com a urina de animais infectados ou pelo contato com água e solo contaminados pela urina. Desta forma, a bactéria é transmitida por meio do ambiente contaminado ou de forma direta pelos hospedeiros reservatórios para os hospedeiros acidentais. Este tipo de ciclo de transmissão de *Leptospira* requer que a bactéria tenha capacidade de sobreviver por longos períodos de tempo no ambiente e de se adaptar às diferentes condições ambientais e do hospedeiro (KO et al, 1999; LEVETT, 2001) (Figura 1).

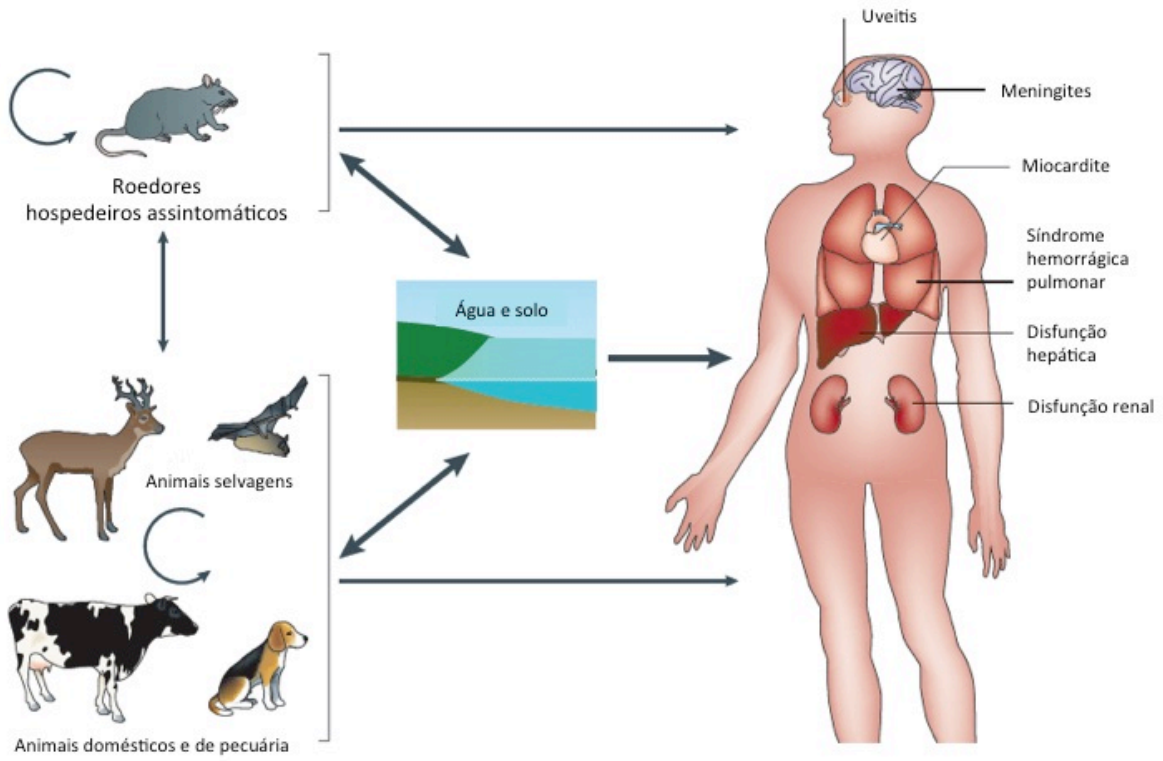


Figura 2: Ciclo de transmissão da *Leptospira spp*. e principais sintomas. Adaptado de KO et al, 1999.

Os sintomas de leptospirose se iniciam com dor de cabeça, febre, mal estar e dores musculares, caracterizando um quadro clínico inespecífico que pode evoluir com dor abdominal e torácica e meningite asséptica. Nesta fase é muito importante o diagnóstico diferencial da leptospirose com doenças como dengue, gripe e meningite viral, possibilitando assim a implementação de terapia específica. A leptospirose pode evoluir para a síndrome de Weil, forma grave da doença caracterizada por insuficiência renal e hepática, miocardite e hemorragias, podendo levar a óbito 5 a 15% dos pacientes. Em alguns casos há evolução para a síndrome hemorrágica pulmonar severa (SHPS) associada à leptospirose, a qual é fatal em até 74% dos casos (FAINE et al, 1999; GOUVEIA et al, 2008).

Casos de SHPS associada a leptospirose foram relatados em diversas regiões geográficas diferentes (PARK et al, 1989; GONÇALVES et al, 1992; SEHGAL, 1995; ZAKI et al, 1995; VIEIRA e BRAUNER 2002; YERSIN et al, 2002; SEGURA et al, 2005), como no surto na

Nicarágua em 1995 (ZAKI et al, 1995), além de outras regiões do Brasil como Rio de Janeiro, São Paulo, Porto Alegre e Salvador (GONÇALVES et al, 1992; VIEIRA e BRAUNER 2002).

1.3 EPIDEMIOLOGIA DE LEPTOSPIROSE EM SALVADOR

A cidade de Salvador, Bahia, apresenta epidemias urbanas anuais concentradas em sua maioria nas comunidades pobres, onde durante a estação chuvosa, a ocorrência de alagamentos constantes, juntamente com as condições precárias de saneamento, favorecem a transmissão no ambiente domiciliar e peri-domiciliar, sendo *Leptospira interrogans* sorovar Copenhageni o agente da leptospirose mais importante em Salvador, que pode ser encontrado em diferentes partes do mundo e já foi isolado de diferentes animais (KO et al, 1999; MCBIRDE et al, 2005).

Em 2003, começaram a ser diagnosticados casos de leptospirose associada à SHPS na cidade de Salvador (GOUVEIA et al, 2008). Diante da gravidade da SHPS, tem se buscado entender fatores que contribuam para o desenvolvimento desta forma da doença. Até hoje não é conhecido se existe associação de SHPS com fatores climáticos ou comportamento de risco, ou se o grande número de casos de SHPS em ambientes urbanos seja devido à introdução de uma cepa mais virulenta no ambiente (Figura 2).

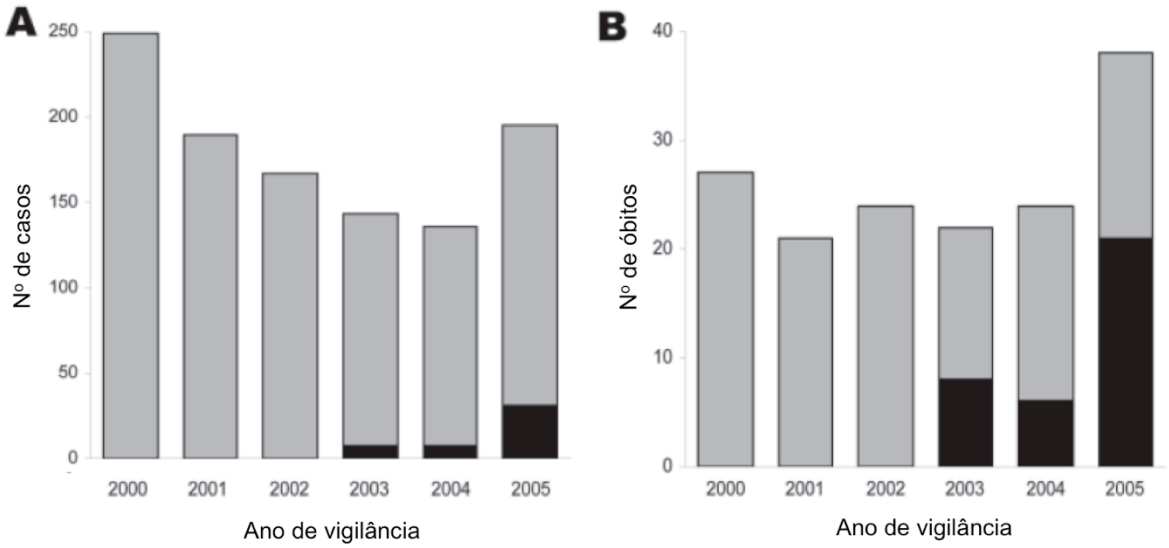


Figura 3: Vigilância hospitalar ativa de leptospirose em Salvador entre o ano de 2000 a 2005. A: número de casos de leptospirose por ano. B: número de óbitos por leptospirose. Os caso de leptospirose sem síndrome hemorrágica pulmonar severa (SHPS) então em cinza e com SHPS em preto. Adaptado de Gouveia et al, 2008.

1.4 TAXONOMIA

A taxonomia da *Leptospira* inicialmente era dividida em dois grupos de acordo com suas características fenotípicas: o patogênico chamado de *L. interrogans* s*ensu lato* e o saprofítico chamado de *L. biflexa sensu lato*. Além disso, cada uma das espécies apresentava seus diferentes sorogrupos e sorovares, determinados pela reação no teste de soro aglutinação cruzada (CAAT). Este teste é baseado nas diferenças no lipopolissacarídeo (LPS) presente nas membranas das células (DIKKEN e KMETY, 1978; KMETY e DIKKEN, 1993). Com a introdução de técnicas de identificação genéticas, como hibridização de DNA, nos anos 90, foi possível então identificar as diferentes espécies de *Leptospira* e perceber que a relação entre a informação genética e a sorológica era muito pequena. Com o avanço das técnicas moleculares já foram identificadas 22 espécies diferentes de *Leptospira*, sendo dez patogênicas (*L. alexanderi*, *L. alstoni*, *L. interrogans*, *L. borgpertersenii*, *L. kirschneri*, *L. kmetyi*, *L. noguchii*, *L. santarosai* e *L. weilli, L. mayottensis* sp), cinco intermediárias (*L. broomii*, *L. fainei*, *L. inadai*, *L. liscerasiae*, *L. wolffii*) e sete saprófitas (*L. biflexa*, *L. meyeri*, *L. terpstrae*, *L. vanthielii*, *L. wolbachii*, *L. yanagawae, L. idonii*) (HAAPALA et al, 1969; YASUDA et al, 1987; BRENNER et al, 1999; FAINE et al, 1999; LEVETT, 2001; SLACK et al, 2008; SAITO et al, 2013; BOURHY et al, 2014).

A classificação sorológica que identifica os mais de 250 sorovares e seus sorogrupos não é considerada como taxonomia oficial. Porém, devido ao uso desta técnica por um longo tempo, a classificação sorológica é amplamente usada por existir uma associação dos sorovares com dados clínicos e epidemiológicos. Diante desta dicotomia das classificações, hoje são usadas as classificações genéticas e sorologica, identificando a espécie e o sorovar da bactéria.

Técnicas de tipagem molecular como número variável de repetições em tandem (*multilocusvariable-number tandem-repeat* - VNTR) e tipagem multilocus de sequência (*multilocussequencetyping* - MLST) tem sido usadas para diferenciar as espécies e sorovares de *Leptospira* fornecendo informações epidemiológicas e auxiliando nas investigações de surtos. Porém, estas técnicas não são capazes de diferenciar todos os sorovares (SALAÜN et al, 2006; THAIPADUNGPANIT et al, 2007; BOURHY et al, 2010).

1.5 PROTEÍNAS ALVO DE VACINA E DIAGNÓSTICO

A alta diversidade de espécies e sorovares de *Leptospira* torna difícil o diagnóstico capaz de identificar as diferentes espécies patogênicas e seus sorovares, com um alta eficiência e com resultados rápidos. Além disto, torna o desenvolvimento de uma vacina que não seja espécie e sorovar especifica um desafio. Proteínas da membrana da *Leptospira* são foco de estudos por serem importantes alvos de vacina e diagnóstico. As proteínas do tipo Lig (*leptospiral immunoglobulin-like*) foram identificadas em espécies patogênicas de *Leptospira*. Estas proteínas apresentam repetições de domínios Big (*bacterial immunoglobulin-like*) que foram previamente caracterizadas como fatores de virulência de diferentes bactérias (HAMBURGER et al, 1999; LUO, 2000). Em *Leptospira* existem três proteínas do tipo Lig identificadas. A LigA e LigB são genes/proteínas identificados em diversas espécies patogênicas e a LigC foi caracterizado como um pseudogene (MATSUNAGA et al, 2003; MCBRIDE et al, 2009; CERQUEIRA et al, 2009). Outras proteínas de membrana foram identificadas apresentando uma alta reatividade a soro de paciente infectados com *Leptospira interrogans* sorovar Copenhageni (LESSA-AQUINO et al, 2013). Estas proteínas, juntamente com as Ligs são importantes alvos para desenvolvimento de vacina e diagnóstico.

1.6 ESTRUTURA GENÔMICA BACTERIANA

Até 2011, as únicas sequências do genoma completo publicadas foram aquelas pertencentes a três espécies de *Leptospira*: uma saprofítica *Leptospira biflexa* e duas patogênicas, *Leptospira interrogans* (sorovares Lai e Copenhageni L1 130) e *Leptospira borgpetersenii* (dois sorovares Hardjo) (REN et al, 2003; NASCIMENTO et al, 2004; BULACH et al, 2006; PICARDEAU et al, 2008). Em geral, o genoma de *Leptospira spp.* é composto de dois cromossomos circulares, cromossomo I com aproximadamente 4 Mb e o cromossomo II com 300 kb, e apresenta um conteúdo de GC de 35% a 41%. As espécies patogênicas, *L. interrogans e L. borgpetersenii*, apresentam aproximadamente 3400 e 2800 regiões codificantes, respectivamente, em seus genomas, onde 656 genes são específicos de espécies patogênicas e não são encontrados na espécie sprofitica *L. biflexa*. Além disto, a função de aproximadamente 59% dos genes é desconhecida, sugerindo mecanismos patogênicos específicos do gênero da *Leptospira* (REN et al, 2003; NASCIMENTO et al, 2004; BULACH et al, 2006; PICARDEAU et al, 2008; KO, Goarant e Picardeau, 2009) (Figura 3).

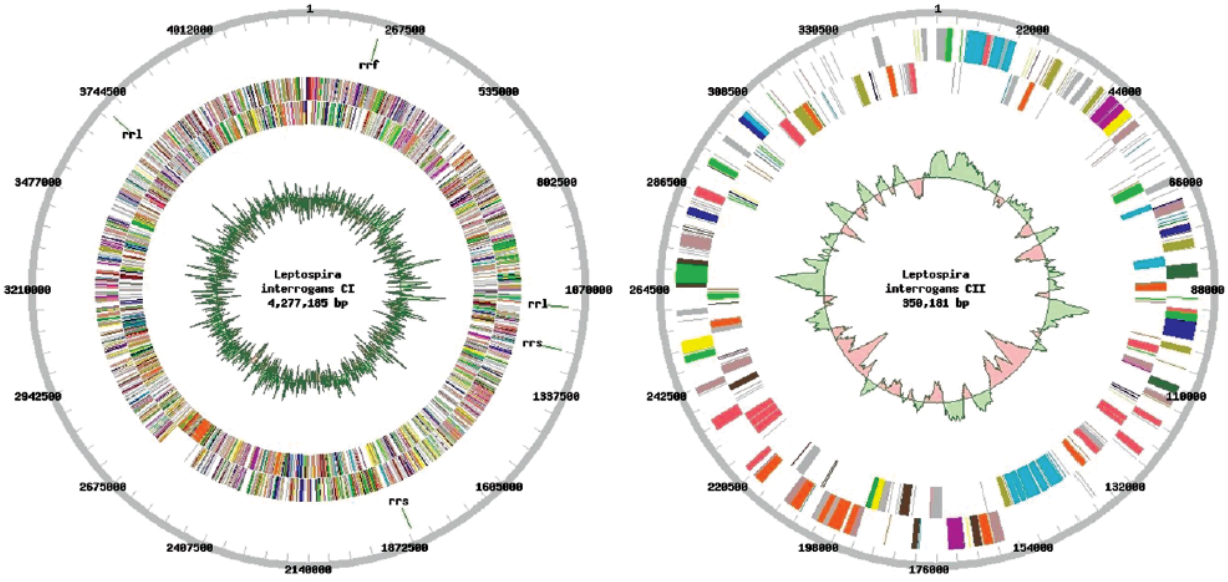


Figura 4: Representação circular dos cromossomos I e II de *Leptospira interrogans* sorovar Copenhageni. Os círculos 1 e 2 (de fora para dentro) representam todas as regiões codificantes de proteínas preditas (margem *forward* e *reverse*, respectivamente) com coloração por categoria; circulo 3: conteúdo C+G. Os números no circulo externo são os pares de base. Adaptado de REN et al, 2003.

Nos últimos anos, tem ocorrido grandes avanços nos métodos de sequenciamento. Novas plataformas de sequenciamento tem surgido e tornado o sequenciamento de genomas mais eficiente e com um custo mais baixo. Além disto, as sequências obtidas através destas plataformas proporcionam uma alta cobertura do genoma com uma qualidade melhor das sequências obtidas (MARDIS, 2008). Esta tecnologia tornou possível o estudo de diversos genomas e das mutações que diferenciam esses genomas (BROWN, FISHWICK e CHOKSHI, 2011; HARRIS et al, 2010; HOLT et al, 2008).

Existem diversos tipos de mutações que podem ocorrer no genoma. As mutações pontuais alteram uma base por outra em um determinado sítio do DNA e são chamadas polimorfismos de nucleotídeo único (*single nucleotide polimorphisms* – SNP) quando esta mutação está presente em mais de 1% da população, e as inserções e deleções de um ou mais nucleotídeos são chamados Indels. A maioria dessas mutações ocorre em regiões não gênicas e não são influenciadas pela pressão seletiva (BARREIRO et al, 2008), porém, as mutações em regiões gênicas podem levar a alteração ou não da proteína sintetizada e são classificadas de acordo com essa alteração. As mutações onde a mudança de nucleotídeo não altera o polipeptídio são chamadas de mutações sinônimas ou mutações silenciosas. Quando a mutação leva a alteração do polipeptídio ela é chamada de mutação não-sinônima e pode ser de dois tipos, as mutações com sentido trocado ou missense onde ocorre a alteração do aminoácido e as sem sentido ou nonsense onde a mutação resulta em um códon de parada (STENSON et al, 2008). Além disto, a ocorrência de um Indel múltiplo de três resultará na inserção ou deleção do aminoácido, porém se este Indel não for múltiplo de três, o mesmo leva a mudança do quadro de leitura (*frameshift*), resultando na alteração de toda a proteína.

Muitos SNPs e Indels podem estar associados a doenças e a ocorrência de diferentes quadros clínicos. Essas mutações podem levar a mudanças na proteína sintetizada ou até mesmo a não produção de uma determinada proteína. Como demonstrado em outros estudos, cepas que apresentam essas alterações em proteínas essenciais ou ligadas a patogênese e virulência do microrganismo podem estar associadas ao quadro clínico (DENBAKKER et al, 2011; PHAN et al, 2009). Além disto, mutações podem estar associada a genótipos diferentes de um microrganismo e a correlações epidemiológicas e de tempo e espaço (DENBAKKER et al, 2011, HARRIS et al, 2010).

Em bactérias, a taxa de substituições nucleotidicas pode ser diferente de uma espécie para outra variando também a diversidade dentro de cada espécie. A *Escherichia. Coli* apresenta uma diversidade alta com uma taxa de $5X10^{-5}$ mutações/geração, quando comparada com outras bactérias como *Salmonella enterica* sorovar Typhimurium com $1x10^{-6}$ mutações/geração (DENAMUR e MATIC , 2006). A taxa de mutação da *Leptospira* não foi calculada para as suas diferentes espécies e sorovares.

O estudo molecular de sequências do genoma completo da *L. interrogans* sorovar Copenhageni e Icterohaemorrhagiae é de grande importância, pois pode permitir identificar mutações associadas aos diferentes perfis clínicos e epidemiológicos, além de identificar diferenças entre os sorovares. Isto, juntamente com a identificação das relações evolutivas entre as cepas de diferentes regiões geográficas e isoladas em tempos diferentes, pode contribuir para o melhor entendimento da dinâmica da bactéria e de uma melhor assistência e prevenção a epidemias de leptospirose. Estudos entre as sequências das diferentes espécies de *Leptospira spp* pode contribuir para o melhor conhecimento desta bactéria e de seus fatores de virulência.

2 OBJETIVOS

2.1 OBJETIVO GERAL

Estudar a epdemioligia molecular de isolados de *Leptospira interrogans* serovares Copenhageni e Icterohaemorrhagiae e avaliar os diferentes fatores clínicos e associação com características genéticas.

2.2 OBJETIVOS ESPECÍFICOS

1. Sequenciar o genoma de isolados de *L. interrogans* serovar Copenhageni provenientes de Salvador, Bahia, Brasil, e estudar a epidemiologia molecular;
2. Avaliar se as mutações no genoma completo da *L. interrogans* sorovar Copenhageni estão associadas aos diferentes desfechos clínicos da leptospirose;
3. Identificar diferenças entre os genomas de *L. interrogans* serovares Copenhageni e Icterohaemorrhagiae;
4. Caracterizar as relações filogenéticas e a história evolutiva das cepas de *L. interrogans* serovar Copenhageni e Icterohaemorrhagiae;
5. Identificar e caracterizar as diferenças genéticas entre os genes das proteínas imunorreativas nas 20 espécies de *Leptospira*;

3 RESULTADOS

Os resultados estão descritos em três artigos, evidenciando os objetivos para cada estudo realizado.

3.1 EPIDEMIOLOGIA MOLECULAR DOS ISOLADOS DE *LEPTOSPIRA INTERROGANS* SOROVARS COPENHAGENI DE SALVADOR

Neste artigo intitulado "**Molecular epidemiology of *L. interrogans* serovar Copenhageni in Salvador, Bahia, Brazil",** isolados *Leptospira interrogans* sorovars Copenhageni de pacientes bem caracterizados clinicamente e de ratos foram sequenciados para avaliar a diversidade do sorovar nos isolados circulantes em Salvador, Bahia, Brasil, e a relação dessa diversidade molecular com os dados clínicos e epidemiológicos. Este manuscrito encontra-se em preparação.

Molecular epidemiology of *L. interrogans*serovarCopenhageni in Salvador, Bahia, Brazil

Luciane Amorim Santos, Xiting Yan, Elsio Augusto WunderJr, HarithaAdikarla, Jeff Townsend, Hongyu Zhao, Luiz Carlos Junior Alcantara, and Albert I. Ko.

Molecular epidemiology of *L. interrogans* serovar Copenhageni in Salvador, Bahia, Brazil

Luciane Amorim Santos[1,2], Xiting Yan[1], Elsio Augusto Wunder Jr[1]., Haritha Adikarla[1], Jeff Townsend[1] Hongyu Zhao[1], Luiz Carlos Junior Alcantara[2], and Albert I. Ko[1,2]

[1] Yale School of Public Health and Medicine, New Haven, USA

[2] Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, Brazilian Ministry of Health, Salvador, Brazil

* Corresponding author. Mailing address: Yale School of Public Health, Epidemiology of Microbial Disease Division, 60 College Street, LEPH Room 319B, P.O. Box 208034, New Haven, CT, 06520-8034 USA. Phone: +1 203 785 6292. Fax: +1 203 785 6193. E-mail: albert.ko@yale.edu

ABSTRACT

Leptospirois is a worldwide-distributed zoonosis cause by the Leptospira spp. In Salvador, Bahia Brazil, *L. interrogans* serovar Copenhageni is responsible for the majority of cases of the disease. The epidemics in Salvador are associated with rainy seasons and with a higher number of cases in patients that leaves in the slum areas. The absence of infrastructure is one of the main reasons for this high number of cases. The clinical manifestations can vary from asymptomatic to severe disease and death. In 2003 the first case of Leptospiral Pulmonary Hemorrhagic Syndrome (LPHS) was reported. To understand if there are any mutations in the strains circulating in Salvador that could explain the variety of the clinical outcomes, 96 clinical isolates of *L. interrogans* serovar Copenhageni were genome sequenced. Ten isolates from *Rattus norvergicus* were also included in to the analyses to study the diversity and evolutionary dynamics of *L. interrogans* serovar Copenhageni in Salvador. A total of 439 SNPs and 177 Indels were detected among the sequences. The mutations detected did not show any association with the different clinical outcomes, year of isolation or source of isolation, with statistical support. The SNPs detected showed a dN/dS of 2:1, indicating a high selective pressure. Phylogenetic reconstruction was performed using ML and Bayesian methods and no temporal structure was observed. The phylogeny and PCA analyses did not detected any cluster related to the epidemiological and clinical data. The sequence showed to be very close related. These findings indicate that the *L. interrogans* serovar Copenhageni in Salvador are very conserved and the detected mutations have no association with the analyzed clinical or epidemiological data.

Keywords

Leptospira, genome, SNPs, LPHS, Clinical manifestations

INTRODUCTION

Leptospirosis is a widespread zoonosis caused by *Leptospira* spp., a bacteria of the Spirochaetales order. There are 22 different species in which ten are pathogenic, five intermediate and seven non-pathogenic [FAINE, 2009; LEVETT, 2001]. The pathogenic species that causes the highest number of cases of diseases in humans worldwide is *L. interrogans* [EVANGELISTA, 2010; ADLER, 2010]. Different species of rats (*Rattus norvergicus* and *Rattus rattus*) are reservoir of the bacteria that colonizes the kidneys of the animals, spreading bacteria in the environment by urine. The transmission to humans occurs through the direct contact of the host skin or mucosa membrane with infected urine or tissues, or by contact with contaminated water or soil in the environment [KO, 2009; LEVETT, 2001].

The number of cases worldwide has increase in the last 20 years due to the changes in the epidemiology of the disease transmission, with more than 500.000 severe cases of leptospirosis worldwide every year [WHO, 2009]. Before, the transmission of *Leptospira* occurs more often in rural settings. With the rapid urbanization growth and the development of urban slum areas in developing countries like Brazil, the number of urban cases of leptospirosis increased dramatically. This increasing is associated with the lack of sewer systems and infrastructure in those areas, which in the raining seasons the contaminating water flood the houses, increasing the chance of transmission. Studies in slum areas have shown that the number of cases of the disease increases with the increasing of rainfalls. The epidemics in Salvador, Bahia, Brazil are caused mainly by one serovar, *L. interrogans* serovar Copenhageni, and are associated with the increase of rain and leaving close to the open sewers in the slum communities [KO, 1999].

Leptospirosis infected with the *L. interrongas* serovar Copenhageni can vary from asymptomatic to more severe cases. The symptoms can start with fever, headache, nausea, muscle pain, which are not specific symptoms, which can be misdiagnosed with dengue, yellow fever, flue or other viral infection. Leptospirosis can became more severe developing the Weli syndrome, which is characterized by renal and hepatic failure, myocarditis and hemorrhage, and has a 5% to 15% chance of death. In xx% of the cases the patient develop a Leptospiral Pulmonary Hemorrhagic Syndrome (LPHS) that has a fatality rate of 75% [MCBRIDE, 2005; YERSIN, 2000; GOUVEIA, 2008].

The first case of LPHS was in South Korean in 1987 [PARK, 1989]. In Salvador, the first case of LPHS was reported in 2003. Before this date, many cases of *Leptospira* infection were reported, but none were associated to hemorrhagic syndrome [GOUVEIA, 2008].

With the increasing number cases of leptospirosis every year and of cases of hemorrhagic syndrome leading to death, and considering that the only serovar isolated in Salvador is the *L. interrogns* serovar Copenhageni, it raised a question: “what makes some patients have mild symptoms and other develop severe forms leading to death?”. With that question in mind we hypothesized that mutations in the genomes of different *Leptospira interrongans* serovar Copenhagen strains are associated with the clinical outcome and the development of LPHS. To test this hypothesis, we sequenced the whole genome of 97 well characterized strains of *L. interrogas* serovar Copenhageni, isolated from humans in different time points and with different clinical outcomes, to detect the mutations that differ one strain from the other, and test if there was any association of the genotypes and the clinical history of the disease in Salvador, Bahia, Brazil. This is the first study to investigate the genetic diversity of the *L. interrogas* serovar

Copenhageni strains that circulate in Salvador, understanding the dynamic and epidemiological history of the epidemic.

MATERIALS AND METHODS

Leptospira isolates

A total of 96 strains of *L. interrogans* serovar Copenhageni clinical isolates obtained from Salvador, Bahia, Brazil, were included in this study. The strains were isolated from well-characterized clinical patients from the reference hospital of infectious disease from Salvador, Hospital Couto Maia. The patients signed an informed consent, clinical evaluation and serological diagnostic was performed. All patients answered a questionnaire to collect epidemiological data. These strains were isolated from epidemics of different years, form 1996 to 2012. Also ten isolates from rats (*Rattus norvergicus*) obtained in 1998 were included in the study. The clinical and epidemiological information from each isolate is listed in Table 1.

Bacterial Culture, Genomic DNA extraction and sequencing

Leptospira strains were cultured in liquid Ellinghausen-McCullough-Johnson-Harris (EMJH) media incubated at 29^{o} C with moderate aeration (shaking at 100 rpm). DNA was then extracted from late-log cultures using the Maxwell 16 cell DNA purification kit along with the Maxwell DNA extraction system (Promega). The quality and concentration of DNA was measured by spectrophotometry using the NanoDrop system (Thermo Scientific, DE, USA) and by fluorometic assay using the Quanti-iT PicoGreen dsDNA assay kit (Invitrogen).

The genomes of the isolates were sequenced at the J. Craig Venter Institute (JCVI) using Illumina/Solexa Genome Analyzer II technology and at the Yale Center for Genome Analysis (YCGA) using the Illumina HiSeq 2000 sequencing system (pair end of 100bp fragment). Whole

genome sequences reads for each isolate are available for download from the NCBI The Sequence Read Archive (SRA) database. Accession numbers can be found in table 1.

Sequence analysis pipeline

For the SNP detection, the reads from each strain were mapped to the *L. interrogans* serovar Copenhageni strain L1-130 [NASCIMENTO, 2004] reference sequence using Stampy [LUNTER, 2011]. For a better mapping quality and variant call, reads duplicates were removed and local re-alignment was performed using Samtools [LI, 2009]. Samtools was also used for identification of the SNPs. The complex SNPs (SNPs in heterozygosis) and SNPs with quality score lower than 30 where excluded for further analyses (Figure 1).

For the Indel detection the reads were mapped to the reference and Indels call using CLC genomic workbench v.4. The Indels with coverage lower than 5x in the Indel site were excluded for further analyses.

Phylogenetic analyses

The 106 isolates, along with the reference strains L1-130 were included in the phylogenetic analysis. *L. interrogans* serovar Lai was used as outgroup. Only the SNPs sites for each genome were used to construct the phylogenetic relationship. The length of the sequence alignment consisted of 1731 variable sites. Maximum Likelihood (ML) phylogeny was inferred using PAUP* [SWOFFORD, 2002] applying the GTR with gamma model of nucleotide substitution. Bootstrap analysis (1000 replicates) was used to calculate the statistical support of the tree branches. Bayesian trees were also inferred including the years of isolation in the tree construction parameters using BEAST software [DRUMMOND, 2007]. The strict molecular

clock with constant population size prior and the relaxed molecular clock with the constant population size and exponential growth priors were tested. Using TreeAnnutator v1.4.8 program, included in the BEAST package, the maximum clade credibility tree were selected from the posterior tree distribution after a 50% burn-in, for each dataset and all trees were visualized using FigTree v1.2.2 graphic viewer.

Statistical analyses

To detect if there were any SNPs or Indels associated with the clinical outcomes, fisher's exact test was performed using four clinical outcomes, Acute Respiratory Distress (ARD), Oligo-anuric Renal Failure (ORF), Massive Pulmonary Hemorrhage (MPH) and Death. The Likelihood ratio test was used to detect if any SNP or Indel were associated to the host of isolation, humans or rats.

In order to detect the presence of any cluster in the data a Principal Component Analysis (PCA) was performed using the SNPs data. All of the statistical analyses were performed using R.

RESULTS

In this study 106 strains isolated from the city of Salvador, Bahia, Brazil, were genome sequenced. Of those 10 isolates were from rats and 96 from well-characterized clinical patients. A total number of SNPs detected by the pipeline in this group of sequences were 439 SNPs, which 153 were in non-coding region and 286 in coding regions (Table 2). Of the mutations found in the genes, 89 were synonymous mutation (do not change in the amino acid) and 197 were non-synonymous mutation (change the amino acid), presenting a 2:1 dN/dS ratio. These SNPs are distributed in 239 different genes. Of those 30 genes had two or more SNPs in the same gene. The genes with the highest number of SNPs were LIC11095 (adenylate/guanylate cyclase) with nine and LIC11218 (hypothetical protein) with five SNPs.

The Indels detected were 177, which 105 (54 deletion and 51 insertions) were found in gene region and 72 (27 deletion and 45 insertions) in non-gene region. The 105 Indels are found in 63 different genes with 16 genes presenting two or more Indels. The genes with the highest number of Indels are LIC12627 (histidine kinase response regulator hybrid protein) with seven Indels and LIC10672 (hypothetical protein), LIC10900 (adenylate/guanylate cyclase), LIC12097 (histidine kinase sensor protein) and LIC13379 (CAAX protease) with five Indels each.

The L1-130 strain was used as a reference for the SNPs and Indels calling. This strain was previously sequenced using the shotgun full genome sequencing and could have some errors. To detect some of these errors we re-sequenced the L1-130 strain using Illumina sequencing method and compared to the reference L1-130 strain for the SNP and Indel calling. 66 SNPs and 62 Indels were detected between L1-130 reference strain and the L1-130 Illumina strain. Out of

those 46 SNPs and 46 Inldes were found in more than 97% of the strains. These high frequency mutations are an indication of possible errors in the L1-130 reference strain.

Phylogenetic analyses were performed in order to identify if there is any cluster related to temporal, clinical or spatial characteristics of the strains. The low diversity among the strains and the non-informative characteristic of the SNPs is shown on the phylogeny were the strains is very conserved and no spatial, temporal or clinical structure was detected (Figure 2). A Bayesian phylogenetic analyses was also performed incorporating the year of isolation as a parameter in the analyses. No temporal structure was observed in this analysis. The tree topology was very similar to the ML tree. PCA analysis was also performed to detect clusters, but no outlier strains or cluster were detected.

Likelihood ratio test to detect the association of any SNPs or Indel o the host of isolation, human or rats, was performed. No mutation was found to be associated to the isolation source of the strains.

Four clinical outcomes, Acute Respiratory Distress (ARD), Oligo-anuric Renal Failure (ORF), Massive Pulmonary Hemorrhage (MPH) and Death, were analyzed to search for association to any SNPs or Indels. No mutations were fount to be associated with the clinical outcomes with statistical significance.

DISCUSSION

The genetic diversity of the different strains of *Leptospira* is unknown. The development of new genome sequence technology with lower cost, generating good quality sequence in a sort time, made the study of its genome diversity possible. In this study, 96 clinical and 10 rat isolates of *L. interrogans* serova Copenhageni were genome sequenced. The strains were isolated from well-characterized clinical patients with a broad spectrum of outcomes. Also the isolates were from different years of epidemics in Salvador, from 1996 to 2012. With this range of epidemiological information, it was expected to detect a higher diversity among the sequences with informative mutations associated with different clinical outcomes, host of isolation and time. Instead, the genomes showed to be very conserved with a relative low number of mutations that were non-informative or associated with the epidemiological data (MORELLI, 2010, JOSHI, 2012).

The genome comparison analyses results showed that the *L. interrogans* serovar Copenhageni strains from Salvador are very conserved with low diversity. The phylogenetic tree and PCA analyses, together with the low number of mutations related to the genome size, gives support to how conserved the sequences are. The isolate sequences did not form any cluster related to host, year or clinical data, and showed a very close relation. It is possibly to identify isolates from 1996 clustering, with bootstrap support, with isolates from 2010, as well as form different clinical outcomes. The dN/dS (non-synonymous/ synonymous) ratio of 2:1 is indicative that the organism is under positive selective pressure, which was not expected for the whole genome. This ratio could also be an indication that the mutation occurred randomly and that are not under selective pressure. This could also indicate a recent population expansion, that there were a rapid population size expansion with not enough time to permit selection of the strains. Studies in other

bacteria like *Mycobacterium tuberculosis, Mycobacterium bovis* and *Staphylococcus aureus* has also identified this ratio (JOSHI, 2012; GUTACKER, 2002 and 2006; HARRISON, 2013). Another hypothesis is that the serovar Copenhageni presents redundancy in the genome. This means that when non-synonymous mutations occur in a specific gene, there are other genes in which the proteins products presents similar functions, not changing the bacteria metabolism and adaptation. The inclusion of more recent isolates (2012) could be a limitation for the study and the inclusion of older strains would make possible to detect a high diversity and a higher temporal structure of the SNPs. Theses results are also found in studies using *L. interrogans* serovar Copenhagen from different geographic locations and with sequences from strains isolated from 1915 to 2012 indicating that this serovar has a slow evolution and that it is well adapted to the different environments (SANTOS, data not published).

Using genotype data to identify the relation with disease severity, spatial and temporal information, showing the evolutionary history of the different strains has been used in different bacteria [BAKER, 2010; FIERER, 2001]. The genome variants among the strains of other bacteria like *Yersinia pestis*, *Staphylococcus aureus*, *Samonela typhi*, among others, has been shown to be informative [BAKER, 2010, BAKER, 2008, MORELLI, 2010, BOS, 2011]. The relations of these variants to the different epidemiological information can help understanding the dynamic of the pathogen in the different location and through time, contributing to the appropriate intervention and better assistance for treatment and control of the spread of the disease. Differing from other bacteria, the variants detected among the *L. interrogans* serovar Copenhageni strains have shown to be non-informative, with no association with the clinical, temporal or host associated data. One of the reasons of this could also be the presence of redundant genes, making it possible to adept to different host (humans and rats) and

environmental conditions, and not been under selective pressure, making the mutations random and non-informative.

The factors involved in the wide range of symptoms that can vary from asymptomatic to severe disease and death is unknown [MCBRIDE 2005]. The hypothesis is that there are: different dose of infection; differences in the *Leptospira*; and host factors. Here in this study, we tested the second hypothesis and identified that there is no mutation in the genome of the bacteria that are associated to the different clinical outcomes. There could be a difference in the expression of some genes that could be associated with the development of disease severity, but no study has been done evaluating the transcriptome of the bacteria in different clinical manifestations. Also, the increase in the sample size and the inclusion of isolates of asymptomatic patients would make possible to detect mutation associated with clinical outcomes with a statistic support. The problem with this approach is that there is no *Leptospira* isolate from asymptomatic patients. Host factors are a strong factor that could lead to different outcomes since the immune system can react to the infection differently from one individual to the other. Proteomic study in guinea pigs that developed LPHS has suggests that the change in the host protein expression could be involved in adhesion and cellular architecture, leading to increased of alveolar wall leakage, seen in LPHS [SCHULLE, 2015].

The understanding of the *Leptospira* evolutionary dynamics and the detection of mutations that could be associated with disease severity would help to better assist patient treatment and disease control. Based on the results of this study the *L. interrogans* serovar Copenhageni strains sequence from Salvador are very conserved with low diversity. Also the development of different clinical outcomes, like LPHS and death, are not associated with any mutation on the genome.

Other studies with different approaches and testing different hypothesis for the differences in the outcome need to be performed for a better understanding of the *Leptospira* pathogenesis.

REFERENCES

Adler B, de la Pena MA. (2010) Leptospira and leptospirosis. Vet Microbiol. 140(3–4):287–96.

Baker S, Hanage WP and Holt KE. (2010) Navigating the future of bacterial molecular epidemiology. Current Opinion in Microbiology. 13:640–645.

Baker S, Holt K, van de Vosse E, Roumagnac P et al. (2008) High-Throughput Genotyping of Salmonella enterica Serovar Typhi Allowing Geographical Assignment of Haplotypes and Pathotypes within an Urban District of Jakarta, Indonesia. Journal Of Clinical Microbiology. p. 1741–1746.

Bos KI, Schuenemann VJ, Golding GB, et al. (2011) A draft genome of Yersinia pestis from victims of the Black Death. Nature 478(7370):506-10.

Drummond, A. J.; Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol, 7:214.

Evangelista KV, Coburn J. (2010) Leptospira as an emerging pathogen: a review of its biology, pathogenesis and host immune responses. Future Microbiol. 5(9):1413–25.

Faine SB, Adler B, Bolin C and Perolat P. *Leptospira* and leptospirosis. (Melbourne A, ed. MediSci), 1999

Fierer J and Guiney DG. (2001) Diverse virulence traits underlying different clinical outcomes of Salmonella infection. 107(7):775-780.

Gouveia EL, Metcalfe J, Carvalho ALF, et al. (2008) Leptospirosis-associated Severe Pulmonary Hemorrhagic Syndrome, Salvador, Brazil. Emerging Infectious Diseases. 14:505-508.

Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, (2006) Single-Nucleotide Polymorphism–Based Population Genetic Analysis of Mycobacterium tuberculosis Strains from 4 Geographic Sites. JID. 193:121-128.

Gutacker MM, Smoot JC, Migliaccio CA, et al. (2002) Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics. 162:1533–43.

Harrison ME, Paterson GK, Holden MTG, Larsen J, Stegger M. (2013) Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel mecA homologue mecC. EMBO Mol Med. 5:509–515.

Joshi D, Harris NB, Waters R, Thacker T, Mathema B. (2012) Single Nucleotide Polymorphisms in the Mycobacterium bovis Genome Resolve Phylogenetic Relationships. Journal of Clinical Microbiology. 50(12):3853-3861.

Ko AI, Galvão Reis M, Ribeiro Dourado CM, Johnson Jr WD and Riley LW. (1999) Urban epidemic of severe leptospirosis in Brazil. Lancet. 354:820-825.

Levett PN. (2001) Leptospirosis. Clin Microbiol Rev, 14:296-326.

Li H, Handsaker B, Wysoker A, et. al. (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9.

Lunter and Goodson. (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21:936-939.

Maciel EA, de Carvalho ALF, Nascimento SF, de Matos RB, Gouveia EL, et al. (2008) Household transmission of Leptospira infection in urban slum communities. PLoS Negl Trop Dis 2: e154.

McBride AJ, Athanazio DA, Reis MG, Ko AI. (2005) Leptospirosis. Curr. Opin. Infect. Dis. 18, 376–386.

Morelli G**,** Song Y, Mazzoni CJ, Eppinger M, Roumagnac P et al. (2010) Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nature Genetics. 42(12):1140-45.

Morelli G, Song Y, Mazzoni CJ, Eppinger M, Roumagnac P, et al. (2010) Yersinia pestis genome sequencing identifies patterns of global phylogenetic diversity. Nature Genetics. 42(12):1140-1145.

Nascimento, Alto et al. (2004) Genome features of Leptospira interrogans serovar Copenhageni. Brazilian Journal of Medical and Biological Research. 37:459-478.

Oliviera DS, Guimaraes MJ, Portugal JL, Medeiros Z. (2009) The socio-demographic, environmental and reservoir factors associated with leptospirosis in an urban area of north-eastern Brazil. Ann Trop Med Parasitol 103: 149–157.

Park SK, Lee SH, Rhee YK, Kang SK, Kim KJ, Kim MC, et al. (1989) Leptospirosis in Chonbuk Province of Korea in 1987: a study of 93 patients. Am J Trop Med Hyg. 41:345–51.

Reis RB, Ribeiro GS, Felzemburgh RD, Santana FS, Mohr S, et al. (2008) Impact of environment and social gradient on Leptospira infection in urban slums. PLoS Negl Trop Dis 2: e228.

Sarkar U, Nascimento SF, Barbosa R, Martins R, Nuevo H, et al. (2002) Population-based case-control investigation of risk factors for leptospirosis during an urban epidemic. Am J Trop Med Hyg 66: 605–610.

Schuller S, Sergeant K, Renaut J et al, (2015) Comparative proteomic analysis of lung tissue from guinea pigs with leptospiral pulmonary haemorrhage syndrome (LPHS) reveals a decrease in abundance of host proteins involved in cytoskeletal and cellular organization. J Prot. 122:55-72.

Swofford DL. (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4.0. Sinauer Associates, Sunderland, MA

World Health Organization. Leptospirosis worldwide, 1999. **Weekly Epidemiol. Rec**. 74, 237-242.

Yersin C, Bovet P, Merien F, Clément J, Laille M, Van Ranst M, et al. (2000) Pulmonary haemorrhage as a predominant cause of death in leptospirosis in Seychelles. Trans R Soc Trop Med Hyg. 94:71–6.

Figure legends

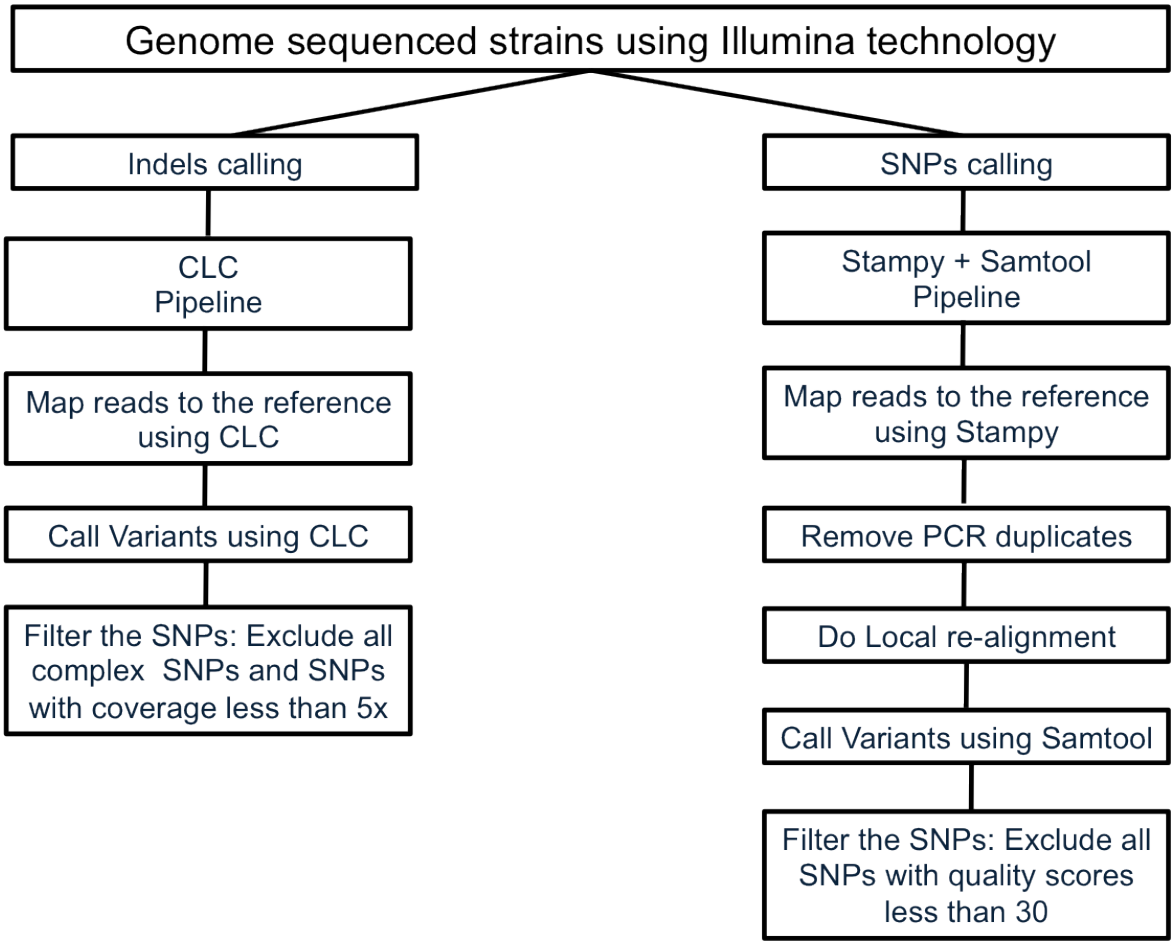
Figure 1: SNPs and Indels calling pipeline

Figure 2: Maximum likelihood tree representing phylogenetic relationship among the *Leptospira* strains serovar Copenhageni from Salvador, Bahia, Brazil. Asterisk (*) represents Bootstrap support higher than 70%.

Tables

Table 1: Epidemiological and clinical information of the strains in the study

Table 2: SNPs and Indels characteristics


Genome sequenced strains using Illumina technology
Indels calling
CLC Pipeline
Map reads to the reference using CLC
Call Variants using CLC
Filter the SNPs: Exclude all complex SNPs and SNPs with coverage less than 5x
SNPs calling
Stampy + Samtool Pipeline
Map reads to the reference using Stampy
Remove PCR duplicates
Do Local re-alignment
Call Variants using Samtool
Filter the SNPs: Exclude all SNPs with quality scores less than 30

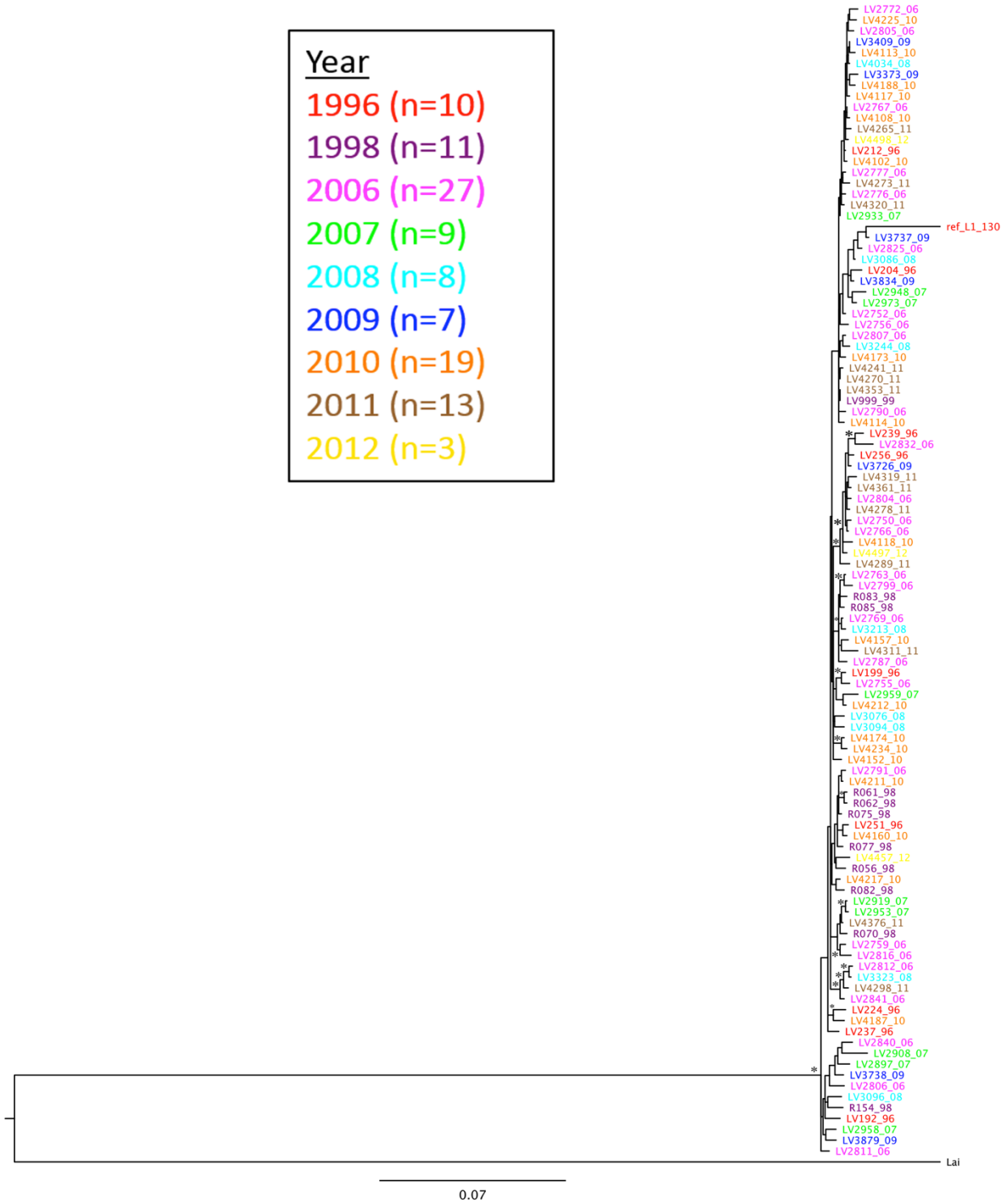

Year
1996 (n=10)
1998 (n=11)
2006 (n=27)
2007 (n=9)
2008 (n=8)
2009 (n=7)
2010 (n=19)
2011 (n=13)
2012 (n=3)
LV2772_06
LV4225_10
LV2805_06
LV3409_09
LV4113_10
LV4034_08
LV3373_09
LV4188_10
LV4117_10
LV2767_06
LV4108_10
LV4265_11
LV4498_12
LV212_96
LV4102_10
LV2777_06
LV4273_11
LV2776_06
LV4320_11
LV2933_07
ref_L1_130
LV3737_09
LV2825_06
LV3086_08
LV204_96
LV3834_09
LV2948_07
LV2973_07
LV2752_06
LV2756_06
LV2807_06
LV3244_08
LV4173_10
LV4241_11
LV4270_11
LV4353_11
LV999_99
LV2790_06
LV4114_10
LV239_96
LV2832_06
LV256_96
LV3726_09
LV4319_11
LV4361_11
LV2804_06
LV4278_11
LV2750_06
LV2766_06
LV4118_10
LV4497_12
LV4289_11
LV2763_06
LV2799_06
R083_98
R085_98
LV2769_06
LV3213_08
LV4157_10
LV4311_11
LV2787_06
LV199_96
LV2755_06
LV2959_07
LV4212_10
LV3076_08
LV3094_08
LV4174_10
LV4234_10
LV4152_10
LV2791_06
LV4211_10
R061_98
R062_98
R075_98
LV251_96
LV4160_10
R077_98
LV4457_12
R056_98
LV4217_10
R082_98
LV2919_07
LV2953_07
LV4376_11
R070_98
LV2759_06
LV2816_06
LV2812_06
LV3323_08
LV4298_11
LV2841_06
LV224_96
LV4187_10
LV237_96
LV2840_06
LV2908_07
LV2897_07
LV3738_09
LV2806_06
LV3096_08
R154_98
LV192_96
LV2958_07
LV3879_09
LV2811_06
Lai
0.07

Table 1: Epidemiological and clinical information of the strains in the study

Species	Serovar	Strain	Host	Year	ARD*	ORF+	MPH#	Death
L. interrogans	Copenhageni	Fiocruz LV192	Human	1996	No	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz LV199	Human	1996	No	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz LV204	Human	1996	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV212	Human	1996	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV224	Human	1996	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV237	Human	1996	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV239	Human	1996	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV251	Human	1996	Yes	No	No	No
L. interrogans	Copenhageni	Fiocruz LV256	Human	1996	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV2750	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2752	Human	2006	No	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz LV2755	Human	2006	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV2756	Human	2006	Yes	Yes	Yes	Yes
L. interrogans	Copenhageni	Fiocruz LV2759	Human	2006	No	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz LV2763	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2766	Human	2006	No	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz LV2767	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2769	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2772	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2776	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2777	Human	2006	Yes	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz LV2787	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2790	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2791	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2799	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2804	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2805	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2806	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2807	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2811	Human	2006	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV2812	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2816	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2825	Human	2006	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV2832	Human	2006	No	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz LV2840	Human	2006	Yes	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2841	Human	2006	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2897	Human	2007	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV2908	Human	2007	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV2919	Human	2007	Yes	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV2933	Human	2007	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV2948	Human	2007	Yes	Yes	Yes	No
L. interrogans	Copenhageni	Fiocruz LV2953	Human	2007	Yes	No	Yes	Yes

L. interrogans	Copenhageni	Fiocruz	LV2958	Human	2007	No	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz	LV2959	Human	2007	Yes	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz	LV2973	Human	2007	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV3076	Human	2008	Yes	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV3086	Human	2008	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV3094	Human	2008	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV3096	Human	2008	Yes	Yes	Yes	Yes
L. interrogans	Copenhageni	Fiocruz	LV3213	Human	2008	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV3244	Human	2008	Yes	Yes	Yes	No
L. interrogans	Copenhageni	Fiocruz	LV3323	Human	2008	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV3373	Human	2009	Yes	Yes	Yes	Yes
L. interrogans	Copenhageni	Fiocruz	LV3409	Human	2009	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV3726	Human	2009	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV3737	Human	2009	Yes	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV3738	Human	2009	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV3834	Human	2009	Yes	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV3879	Human	2009	Yes	Yes	Yes	No
L. interrogans	Copenhageni	Fiocruz	LV4034	Human	2008	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4102	Human	2010	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4108	Human	2010	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4113	Human	2010	Yes	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4114	Human	2010	Yes	Yes	No	Yes
L. interrogans	Copenhageni	Fiocruz	LV4117	Human	2010	Yes	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4118	Human	2010	Yes	Yes	Yes	No
L. interrogans	Copenhageni	Fiocruz	LV4152	Human	2010	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4157	Human	2010	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4160	Human	2010	Yes	Yes	Yes	Yes
L. interrogans	Copenhageni	Fiocruz	LV4173	Human	2010	Yes	Yes	Yes	No
L. interrogans	Copenhageni	Fiocruz	LV4174	Human	2010	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4187	Human	2010	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4188	Human	2010	Yes	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4211	Human	2010	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4212	Human	2010	Yes	Yes	Yes	No
L. interrogans	Copenhageni	Fiocruz	LV4217	Human	2010	Yes	Yes	Yes	Yes
L. interrogans	Copenhageni	Fiocruz	LV4225	Human	2010	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4234	Human	2010	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4241	Human	2011	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4265	Human	2011	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4270	Human	2011	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4273	Human	2011	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4278	Human	2011	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4289	Human	2011	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4298	Human	2011	Yes	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4311	Human	2011	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz	LV4319	Human	2011	No	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4320	Human	2011	Yes	No	No	No
L. interrogans	Copenhageni	Fiocruz	LV4353	Human	2011	No	Yes	No	No

L. interrogans	Copenhageni	Fiocruz LV4361	Human	2011	Yes	Yes	Yes	No
L. interrogans	Copenhageni	Fiocruz LV4376	Human	2011	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV4457	Human	2012	No	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV4497	Human	2012	No	No	No	No
L. interrogans	Copenhageni	Fiocruz LV4498	Human	2012	Yes	Yes	No	No
L. interrogans	Copenhageni	Fiocruz LV999	Human	1998	No	No	No	No
L. interrogans	Copenhageni	Fiocruz R056	Rat	1998	NA	NA	NA	NA
L. interrogans	Copenhageni	Fiocruz R061	Rat	1998	NA	NA	NA	NA
L. interrogans	Copenhageni	Fiocruz R062	Rat	1998	NA	NA	NA	NA
L. interrogans	Copenhageni	Fiocruz R070	Rat	1998	NA	NA	NA	NA
L. interrogans	Copenhageni	Fiocruz R075	Rat	1998	NA	NA	NA	NA
L. interrogans	Copenhageni	Fiocruz R077	Rat	1998	NA	NA	NA	NA
L. interrogans	Copenhageni	Fiocruz R082	Rat	1998	NA	NA	NA	NA
L. interrogans	Copenhageni	Fiocruz R083	Rat	1998	NA	NA	NA	NA
L. interrogans	Copenhageni	Fiocruz R085	Rat	1998	NA	NA	NA	NA
L. interrogans	Copenhageni	Fiocruz R154	Rat	1998	NA	NA	NA	NA

*Acute respiratory distress; +Oligo renal Failure; #Massive pulmonary hemorrhage

Table 2: SNPs and Indels characteristics

	Intragenic region	**Intergenic region**	**Total**
SNPs			
Synonymous	89	NA	NA
Non-synonymous	197	NA	NA
Total	286	153	439
Indels			
Insertion	51	45	96
Deletion	54	27	81
Total	105	72	177

NA = Not Applicable

3.2 IDENTIFICAÇÃO DE DIFERENÇAS ENTRE OS GENOMAS DOS ISOLADOS DE *L. INTERROGANS* SEROVAR ICTEROHAEMORRHAGIAE E COPENHAGENI

O artigo intitulado "**Analysis of genome-wide variations among global *L. interrogans* Icterohaemorrhagiae isolates**" realizou análise de sequências do genoma completo de cepas de *Leptospira interrogans* serovars Icterohaemorrhagiae e Copenhageni isoladas de diferentes partes do mundo em diferentes anos, para avaliar a diversidade entre os dois sorovares. Neste estudo foi identificada uma grande similaridade genética entre os dois sorovares, apresentando apenas uma única mutação que as diferenciam. Foi realizado também análise da relação evolutiva das diferentes cepas e sua ligação com tempo e espaço. Este manuscrito está pronto para submissão.

Analysis of genome-wide variations among global *L. interrogans* Icterohaemorrhagiae isolates

Luciane Amorim Santos, Haritha Adikarla, Xiting Yan, Elsio Augusto Wunder Jr., Derrick E Fouts, Joseph M Vinetz, Luiz Carlos Junior Alcantara, Mitermayer Galvão Reis, Jeff Townsend, Hongyu Zhao and Albert I. Ko.

Analysis of genome-wide variations among global *L. interrogans* Icterohaemorrhagiae isolates

Luciane Amorim Santos[1,2§], Haritha Adhikarla[1§], Xiting Yan[1], Elsio A. Wunder Jr[1]., Derrick E Fouts[3], Joseph M. Vinetz[4], Luiz Carlos Junior Alcantara[2], Mitermayer G. Reis[2], Jeff Townsend[1], Hongyu Zhao[1] and Albert I. Ko[1,2*]

[1]Department of Epidemiology of Microbial Diseases, Yale School of Public Health, New Haven, USA

[2]Gonçalo Moniz Research Center, Oswaldo Cruz Foundation, Salvador, Brazil.

[3]J. Craig Venter Institute, Rockville, Maryland, United States of America

[4]Division of Infectious Diseases, Department of Medicine, University of California San Diego School of Medicine, La Jolla, California, United States of America.

[§]These authors contributed equally to this work.

* Corresponding author.

*Correspondence: 60 College Street, LEPH Room 319B, P.O. Box 208034, New Haven, CT, 06520-8034 USA. albert.ko@yale.edu

Background

Leptospirosis is the most wide spread zoonosis in the world with the pathogenic species of the genus *Leptospira spp* being the etiologic agent of the disease. *Leptospira* can be classified by serological and genotypic approaches. *L. interrogans* serovar Copenhageni and serovar Icterohaemorrhagiae are the most widespread and genetically indistinguishable pathogenic strains capable of causing severe leptospirosis.

Results

In this study we investigated the influence of spatial and temporal variations on sequence diversity of *L. interrogans* Copenhageni and Icterohaemorrhagiae serovars by sequencing the corresponding strains, and to identify the genomic differences that could possibly explain their antigenic variations. The total number of SNPs and Indels detected were low when compared to the size of the genome. The phylogenetic analyses showed that both serovars are closely related and did not cluster separately. Comparative genomic analyses identified only one insertion in a hypothetical protein (LIC12008) that differentiated *L. interrogans* serovar Icterohaemorrhagiae strains from serovar Copenhageni. *In silico* analysis indicated the possible role of LIC12008 in LPS synthesis, explaining the serological differences between genetically similar serovars.

Conclusions

This study showed that the *L. interrogans* serovar Copenhageni and Icterohaemorrhagiae are genetically similar with only one point mutation in LIC12008 that can differentiate both serovars. This mutation is presumably related to LPS synthesis and might explain the serological differences between *L. interrogans* serovar Copenhageni and Icterohaemorrhagiae. There is no difference in the host adaptation and virulence among these strains. The phylogeny and dN/dS

ratio support the conclusion that the *L. interrogans* serovar Copenhageni and Icterohaemorrhagiae strains are highly conserved among time and with a little spatial structure.

Keywords

Leptospira, virulence, whole-genome sequencing, Single nucleotide polymorphisms (SNPs), Insertions and deletions (Indels), phylogeny.

Background

Leptospirosis is a zoonosis with worldwide distribution caused by a spirochete from the genus *Leptospira* and is endemic in developing countries and tropical regions [1, 2]. Transmission of leptospirosis requires continuous enzootic circulation of the pathogen among animal reservoirs. This zoonosis is maintained in nature through chronic renal infection of carrier animals, with rodents and other small mammals being the most important reservoirs. Leptospirosis occurs by direct contact with infected animals or contact with contaminated water or soil [1,3]. More than 800,000 severe cases of leptospirosis are reported every year [4, 5]. Weil's disease is the classic presentation of severe leptospirosis which is characterized by jaundice and acute renal failure with a 10% fatality rate. Additionally, Leptospirosis-associated pulmonary haemorrhage syndrome (LPHS) is the severe disease form for which the case fatality rate is more than 50% [6, 7].

With the increased use of genomic information for the classification of *Leptospira* the genus has been reorganized. Currently this genus is classified into 22 genomospecies, (ten pathogenic, five intermediate and seven non-pathogenic) comprising more than 250 serovars, which indicates a high diversity of the genus [10 - 14]. The structural differences in the carbohydrate moiety of surface - exposed LPS determine antigenic diversity among the numerous serovar groups [8, 9]. Several genotyping methods such as Pulsed-field gel electrophoresis (PFGE), Multilocus variable-number tandemrepeat (VNTR) multispacer sequence typing (MST) analysis have been developed to tentatively identify the isolates to the serovar level. However, the differentiation of certain serovars such as *L. interrogans* serovars Icterohaemorrhagiae and Copenhageni of the most prevalent Icterohaemorrhagiae serogroup, remains difficult [15 – 17].

Genomic differences contributing to the diverse distribution of Leptospira serovars and the underlying genetic variations remain poorly understood. Employing DNA polymorphisms such as single nucleotide polymorphisms (SNPs), insertions and deletions (indels), and other larger rearrangements were successfully employed to study sequence diversity among closely related but distinct populations [22, 23]. The use of short-read next generation sequencing (NGS) data to detect DNA polymorphisms in the context of whole-genome analysis have been previously reported in pathogenic bacteria like *Salmonella typhi*. Unlike the taxonomically informative or canonical SNP-based approaches, whole-genome sequencing served as a robust and unbiased method to resolve intraspecies relationships in closely related species such as Brucella spp. [12,13] and Bacillus anthracis [14,15].

Genome-wide identification of SNP's and Indels in *L. interrogans* serovar Copenhageni and Icterohaemorrhagiae serovars will enable us to identify the genetic relatedness of these strains isolated from various geographic locations. Studying these variations at genomic level will have important implications for development of new molecular markers to differentiate pathogenic serovars from epidemiological settings and to understand their evolutionary relationships [24]. In this study we performed whole-genome sequencing of 67 different strains of *L. interrogans* serovar Copenhageni and serovar Icterohaemorrhagiae and conducted sequence analyses to identify genome-wide DNA-based variation/s presumably critical for strain divergence and pathogenicity.

Results

Whole-genome mapping and variant call of *L. interrogans* serogroup Icterohaemorrhagiae serovar Copenhageni and serovar Icterohaemorrhagiae isolates

The pipeline selected for read mapping and SNP identification was Stampy and Samtool, respectively. CLC genome workbench was selected for both mapping and identification of Indels. The pipeline used for identification of both SNPs and Indels was validated by re-sequencing of the seven isolates and was based on its ability to identify the highest percentage of SNPs and/or Indels in both sequences of the same isolate (Table S2).

L. interrogans serovar Copenhageni strain Fiocruz L1-130, sequenced using shotgun technology [30], was used as the reference sequence. Comparison of resequenced *L. interrogans* serovar Copenhageni strain Fiocruz L1-130 with the previously published sequence resulted in identification of 66 SNPs and 62 Indels. Out of these, 45 SNPs and 46 Indels had a distribution frequency of 97% or higher in all the strains sequenced in this study. Based on this frequency we included these mutations as sequence errors and did not consider them in our analyses.

Characteristics of the mutations detected in *L. interrogans* serovar Copenhageni and serovar Icterohaemorrhagiae strains

In this study a total of 67 strains were completely genome sequenced, out of which 55 were *L. interrogans* serovar Copenhageni and 12 were *L. interrogans* serovar Icterohaemorrhagiae isolates. These strains were isolated from different geographic location, hosts and years, including the first *Leptospira* ever isolated, Ictero 1, isolated in 1915 in Japan (Supplementary Table 3). Serogrouping of all the isolates was tested initially and confirmed to

be Icterohaemorrhagiae. Subsequently these isolates were typed to serovar level by MAT with monoclonal antibodies (MAbs) against the serovars Icterohaemorrhagiae and Copenhageni:
Cumulatively, we identified 1072 SNPs in 67 isolates, of which 276 were in non-coding region and 796 in coding regions (Table 1 and Table S4). Of the 796 mutations found in genes, 258 were synonymous and 538 were non-synonymous, showing a 2:1 dN/dS (number of non-synonymous mutation / number of synonymous mutations) ratio. These SNPs were distributed in 594 different genes, and 115 of those had two or more SNPs in the same gene. Genes displaying highest number of SNPs were LIC12896 - hypothetical protein (17 SNPs) and LIC10502 - cytoplasmic membrane protein (eight SNPs).

We identified 235 Indels, out of which 178 (98 deletion and 80 insertions) were found in coding regions and 54 (24 deletion and 33 insertions) in non-coding regions. The coding-region Indels were distributed in 140 different genes, whereas 24 genes presented two or more Indels (Table 1). Genes displaying highest number of INDELs were found in LIC10900 - adenylate/guanylate cyclase (six), LIC10674 - hypothetical protein, LIC10902 - hypothetical protein and LIC13017 - acriflavine resistance (four Indels each).

Phylogenetic analysis

We performed phylogenetic analysis in order to identify if there were any cluster related to serovar, temporal or spatial characteristics of the strains. A maximum likelihood (ML) tree of the 67 strains, using *L. interrogans* serovar Lai as an out-group, presented a topology with no cluster related to the different serovars, showing both serovar Copenhageni and Icterohaemorrhagiae clustering together, which is an indication of their genetic relatedness. However, the phylogeny did showed four clusters, with bootstrap support higher than 70%. These cluster were associated

with geographic location of the isolates from Japan, Egypt, Hawaii (USA) and Colombia (Figure 2). Bayesian tree were also constructed using the year of isolation as a prior and showed to be very similar with the ML tree. PCA analysis was then performed to confirm the clusters found, but it only confirmed the two sequences from Egypt separated from the rest of the sequences, which due to the high similarity of the strains.

A novel InDel identified in serovar Icterohaemorrhagiae strains

In order to identify possible genetic differences between the two serovars a likelihood ratio test (LRT) was performed. The test showed that there were no SNPs that could differentiate both serovars. However, the test with the Indel data showed one nucleotide insertion in all Icterohaemorrhagiae strains that was absent in all the Copenhageni strains ($p=0.039$). The insertion of a thymine nucleotide on gene LIC12008, which encodes a Hypothetical Protein with 242 amino acids, caused a frame shift mutation at the amino acid 45 of the protein, generating a stop codon at the amino acid 48 (Figure 3 C). The insertion is in a homopolymeric track were in the Copenhageni strains there is eight thymine and nine in the Icterohaemorrhagiae strains. Sanger sequence of the specific mutation region confirmed the presence of the insertion at all the Icterohaemorrhagiae strains and the insertion was not detected in the Copenhageni strains.

Analysis of genomic region with predicted InDel

Nucleotide and protein BLAST search was employed to study the distribution of LIC12008 gene in other species and serovars. The analysis at nucleotide level demonstrated that this gene was absent in non-pathogenic and intermediate *Leptospira* species, but present only in 4 of the pathogenic species: *L. interrogans, L. kirschneri, L. santarosai* and *L. noguchii*. The BLAST search of the protein sequence showed identity with the UDP-3-O-(3-hydroxymyristoyl)

glucosamine N-acyltransferase and sugar O-acyltransferase, sialic acid O-acetyltransferase belonging to the NeuD family.

Domain analysis of hypothetical protein encoded by LIC12008 showed that the region spanning amino acids 36 to 233 (out of 242 aa) is comprised of the Putative Acyltransferase (AT), Left-handed parallel beta-Helix (LbH) domain (E-value: 6.33e-48 and domain accession no: cd03360) (Figure 3B). Members of LbetaH superfamily contain a N-terminal helical subdomain followed by an LbH domain. The alignment contains 6 turns, each containing three imperfect tandem repeats of a hexapeptide repeat motif (X-[STAV]-X-[LIV]-[GAED]-X). Due to the presence of hexapeptide repeat, LIC12008 could presumably have acyltransferase activity. The presence of hexapeptide repeat in LIC12008 was also confirmed by domain search in Pfam database where the best hit to this protein was bacterial transferase hexapeptide (Hexapep2 family and E-value:0.0004).

BLAST search of *Leptospira interrogans* Copenhageni Fiocruz L1130 genome employing LIC12008 as a query sequence identified a paralogous protein LIC12184, annotated as an acetyl transferase gene (with identity of 74.1% and similarity of 89.2%).

Functional analysis

Real time PCR was performed to detect differences in the gene expression, synthesis of mRNA. Two pair of primers was used. The first one was in the region of the mutation and the second one after the mutation region. In both fragments the LIC12008 gene was.

To investigate if the mutation identified in LIC12008 had an overall impact on the phenotype of these serovars, an *in vivo* experiment was performed employing the hamster model of infection.

Representative strains from *L. interrogans* serovar Icterohaemorrhagiae and serovar Copenhageni strains were used to infect groups of six hamsters at doses of 100 (IP infection) and 10^8 (conjunctival). Both serovars were virulent at the indicative dose of *Leptospira*, showing that there is no difference in virulence between the serovars.

Discussion

Whole genome sequencing enabled us to study to study the genome- wide variations and of *L. interrogans* Copenhageni and Icterohaemorrhagiae serovars. In this study, we analyzed strains isolated from different geographical locations and at different time points, to have a better understanding of spatial and temporal diversity of these serovars.

The genetic relatedness of the *L. interrogans* Copenhageni and Icterohaemorrhagiae serovars is reflected in the phylogenetic reconstruction where the isolates of both serovars clustered Furthermore, our phylogenetic analyses indicated that there is no temporal clustering in the phylogeny.

Given the size of the genome, the number of mutations detected in individual strains was lower when compared to other bacteria [23, 25], indicating a lower diversity and a high sequence similarity between both serovars. Besides that, we identified a 2:1 ratio of non-synonymous to synonymous mutations comparable to that obtained for other bacteria like *Mycobacterium tuberculosis, Mycobacterium bovis* and *Staphylococcus aureus* [25-28]. One plausible explanation for this finding could be a rapid expansion of the population size, with insufficient evolutionary time to permit selection, indicating a recent population expansion. The time period of the *L. interrogans* serovars Copenhageni and Icterohaemorrhagiae strains included in this study, from 1915 to 2012, was not enough to permit purifying selection of the mutations. This indicates that this population is not under high selective pressure [28]. A different explanation for this finding is that when a mutations occur in a specific gene, altering the protein, other genes in which the proteins products presents similar functions will be expressed, not changing the bacterial metabolism and adaptation.

Despite the high conservancy between Copenhageni and Icterohaemorrhagiae we identified phylogenetic clusters related to a few geographic locations, like Japan, Egypt, Hawaii (USA) and Colombia. The population of *L. interrogans* serovar Copenhageni from Salvador, Brazil, has a larger representation in the phylogeny (higher number of sequences) and it clustered with isolates from other countries. One possibility could be the introduction of various *Leptospira* strains in Salvador by the ships that arrived from Europe, Africa and other parts of the world related to the discovery of the new world. The inclusion of a higher number sequences from other geographic locations would probably give a better understanding of the relation of this strains and maybe giving an insight into the history of the strains.

The molecular tests have been unsuccessful in distinguishing the differences between the *L. interrogans* serovar Copenhageni and Icterrohaemorrhagiae. Reason for this difficulty in distinguishing between two strains could be attributed to lack of genetic variations. Here we showed why that was not possible. In this study we identified the single point mutation in hypothetical protein/LIC12008 gene. Domain analysis revealed the presence of LbH_AT domain in this protein, which can be correlated with a sialic acid (Sia) O-acetyltransferase (NeuD protein) that is required for Sialic acid synthesis. Modification of Sia residues by *O*-acetylation is common in several pathogenic bacteria [34 - 38]; this could also alter the immunogenicity of polysaccharide epitopes [39 – 41]. Therefore this protein could probably explain the observed differences between two *L. interrogans* serovars in CAAT test. This result revaled how one gene can have a great impact in the serological assay, indicating that the serological phenotype does not necessarily differentiate the severity of the disease caused by the strain and their epidemiological information.

The identification of the serovar by the serological classification is known to be important for providing the relationship between serovars and their clinical and epidemiological data. Although these two serovars were treated differently, there is only one genetic mutation that differentiates the two serovars. Recent studies had identified the distribution of both the serovars in *Rattus norvergicus* and *Rattus rattus* species showing they are not host specific [19, 31-33]. I*n vivo* test performed to study differences in their virulence pattern did not yield any different two serovars. These evidences show that the epidemiological studies of these two serovars should not be separate and that the serology has no value in differentiating the two serovars. These two serovars study here does not show difference in virulence, host or time of isolation. This indicates a possible functional redundancy for LIC12008 and implies the importance of this gene in *Leptospira* metabolism.

The genomic data and the phenotypic evidences indicate that both serovars are the same with only genetic and phenotype differences being the LPS, influencing in the CAAT. If there is only one genetic difference and no differences in the virulence, host specificity or place of isolation between the *L. interrogans* serovar Copenhageni and Icterohaemorrhagiae, should these two serovars be considered different in epidemiological studies?

CONCLUSIONS

This study showed that the *L. interrogans* serovar Copenhageni and Icterohaemorrhagiae are genetically similar with only one point mutation in LIC12008 that can differentiate both serovars. This mutation is presumably related to LPS synthesis and might explain the serological differences between *L. interrogans* serovar Copenhageni and Icterohaemorrhagiae. There is no difference in the host adaptation and virulence among these strains. The phylogeny and dN/dS ratio support the conclusion that the *L. interrogans* serovar Copenhageni and Icterohaemorrhagiae strains are highly conserved among time and with a little spatial structure.

MATERIALS AND METHODS

Leptospira isolates

A total of 67 strains of *L. interrogans* serogroup Icterohaemorrhagiae, including fifty-five serovar Copenhageni isolates and twelve serovar Icterohaemorrhagiae isolates, were included in this study. These strains were isolated from different geographic locations and hosts, and the years of isolation ranged from 1915 to 2012 (Table 1).

Culture, Genomic DNA extraction and sequencing

The *Leptospira* strains were cultured in liquid Ellinghausen-McCullough-Johnson-Harris (EMJH) media incubated at 29^{o} C with moderate shaking at 100 rpm. DNA was then extracted from late-log cultures using the Maxwell 16 cell DNA purification kit along with the Maxwell DNA extraction system (Promega). The quality and concentration of DNA was measured by spectrophotometry using the NanoDrop system (Thermo Scientific, DE, USA) and by fluorometic assay using the Quanti-iT PicoGreen dsDNA assay kit (Invitrogen).

The genomes of the isolates were sequenced at the J. Craig Venter Institute (JCVI) using Illumina/Solexa Genome Analyzer II technology and at the Yale Center for Genome Analysis (YCGA) using the Illumina HiSeq 2000 sequencing system. Whole genome sequences reads for each isolate were deposited at NCBI in the Sequence Read Archive (SRA) database and accession numbers were provided (Table 1).

Serological characterization of isolates

The microscopic agglutination test (MAT) was used for antigenic characterization of Leptospira isolates, with a standard battery of rabbit antisera against reference serovars representing the 12 serogroups as previously described [18].

Sequence analysis pipeline

The sequenced reads were mapped to the *L. interrogans* serogroup Icterohaemorrhagiae serovar Icterohaemorrhagiae L1130 strain genome using Stampy tool [44]. The replicated alignment removal and local realignment were done using Samtools [45]. The processed mapping results were further analyzed for SNP calling using samtools. The called SNPs were filtered to have a quality score cut-off >30. CLC Genomics workbench (v 6.0.4) was used to call Indels [46]; Indels with coverage lower than 5x were filtered. The Samtools pipeline exhibited better rates of consistency for SNPs calling while CLC Bio was consistent for Indel calling (Figure 1, Table S1 and Table S2). All complex mutations (heterozygosis) were excluded in this pipeline.

Phylogenetic analyses

The whole-genome-based SNP phylogenetic analysis was used to infer the relationships between 67 isolates of *L. interrogans* serogroup Icterohaemorrhagiae serovars Icterohaemorrhagiae and Copenhageni collected from diverse geographical locations.The reference strain *L. interrogans* serogroup Icterohaemorrhagiae serovar Icterohaemorrhagiae L1-130 was included in the phylogenetic analysis. *L. interrogans* serovar Lai was used as out-group. Only the SNPs for each genome were used to construct the phylogenetic tree. The length of the sequence alignment consisted of 1731 variable sites. Maximum Likelihood (ML) phylogeny was inferred using the online tool PAUP* [24] applying the GTR model of nucleotide substitution and gamma shape

parameter. Bootstrap analysis (1000 replicates) was used to calculate the statistical support of the tree branches. Bayesian trees were also inferred including the years of isolation in the tree construction parameters using BEAST [47] software. Tree visualization and editing was done using FigTree v.1.2.2.

Statistical analyses

The genotypes of strains from Icterohaemorrhagiae and Copenhageni were compared based on the log likelihood ratio test. Basically for any given SNP or INDEL, two binomial distributions were fitted for the number of alternative alleles observed in strains from Icterohaemorrhagiae and Copenhageni separately by maximizing the likelihood. Suppose the maximized likelihood of Icterohaemorrhagiae and Copenhageni are denoted by L_I and L_C. Then another binomial distribution was fitted for the number of alternative allele by treating all strains from Icterohaemorrhagiae and Copenhageni as one group. If the maximized likelihood of this binomial distribution is $L_{I\&C}$, then the log likelihood ratio is calculated as $-2\log\left(L_{I\&C}/(L_I \cdot L_C)\right)$. The p value was then calculated by comparing this observed log likelihood ratio to the Chi-squared distribution with 1 degree of freedom. In order to detect the presence of any cluster in the data a Principal Component Analysis (PCA) was performed using the SNPs data. All of the statistical analyses were performed using R.

Identification and characterization of serovar Icterohaemorrhagiae specific Indel (in LIC12008)

To confirm the mutation found in the LIC12008 gene, PCR followed by Sanger sequencing of the region was performed using specific primers (forward 5'TAGGTTGGCACGAAGGTTCT3' and reverse 5'TTTTTCCGGGAACTCCAAC3'). Sequencher 5.2 and was employed to conduct the

sequence analysis and subsequent alignment with the reference strain to identify the presence of the mutation. A total of 16 Icterohaemorrhagiae and 16 Copenhageni strains were Sanger sequenced. BLAST was employed to identify homologous sequences of LIC12008 at nucleotide and protein level. The domain analyses of LIC12008 protein was performed employing NCBI CD-search and Pfam 27.0 sequence search tools.

Isolation of RNA and Quantitative reverse transcription PCR (RT-qPCR)

L. interrogans serogroup Icterohaemorrhagiae serovars Icterohaemorrhagiae and Copenhageni () were cultured to a density of 10^8 bacteria per ml at 30°C with shaking. Cultures were harvested via centrifugation at 3,200x*g* and RNA was extracted for two biological replicates using the TRIzol (Invitrogen) method, as previously described. Ambion® TURBO DNA-free™ DNase Treatment kit was employed to remove contaminating DNA from RNA preparations. The concentration of RNA was determined using a Spectrophotometer (NanoDrop). The High capacity cDNA reverse transcription kit (Life Technologies) is employed for conversion of total RNA to single-stranded cDNA. Two sets of primers were designed to study the impact of mutation in representative *L. interrogans* serovars Icterohaemorrhagiae and Copenhageni isolates. First primer set amplified a fragment of 126bp in region encompassing nucleotides from 29 to 155 before the mutation. A second primer set detected a fragment of 133 bp after the mutation of a size of 133bp (from nucleotide 335 to 468).

The qPCR was carried out on 7500 fast real-time PCR (ABI, USA) using iQ^{TM} SYBR^{R} Green supermix (Biorad) according to manufacturer's instructions. The thermal cycling conditions used in the qPCR were 95°C for 3 min, followed by 40 cycling of 95°C for 5 s and 60°C for 1 min. The specificity of the SYBR green PCR signal was confirmed by meltcurve analysis. In RT-

qPCR experiments, *rplH* gene encoding ribosomal subunit protein was used as an endogenous control and reference strain employed was *L. interrogans* serovar Manilae strain L495. A relative quantification analysis was performed using the comparative CT method, and the relative gene expression was calculated by using the $2^{-\Delta\Delta Ct}$ method [46].

In vivo characterization

To test if the mutation in LIC12008 gene impacts virulence phenotype, in vivo experiments were performed using the hamster model of infection. Two *L. interrogans* serovar Icterrohaemorrhagiae and two *L. interrogans* serovar Copenhageni strain at doses of 100 (IP infection) and 10^8 (conjunctival) were used to infect groups of six hamsters.

REFERENCES

1. Levett PN (2001) Leptospirosis. Clin. Microbiol. Rev. 14, 296–326.
2. Faine S. (1993) Leptospira and leptospirosis. Baton Raton: CRC Press.
3. Faine SB, Adler B, Bolin C and Perolat P. (1999) *Leptospira* and leptospirosis. (Melbourne A, ed. MediSci).
4. Ko AI, Galvão Reis M, Ribeiro Dourado CM, Johnson Jr WD and Riley LW. (1999) Urban epidemic of severe leptospirosis in Brazil. Lancet, 354:820-825.
5. Picardeau M, Bertherat E, Jancloes M, Skouloudis AN, Durski K, Hartskeerl RA. Rapid tests for diagnosis of leptospirosis: current tools and emerging technologies. Diagn Microbiol Infect Dis. 2014; 78(1). 1-8. doi: 10.1016/j.diagmicrobio.2013.09.012 PMID: 24207075
6. McBride AJ, Athanazio DA, Reis MG, Ko AI (2005) Leptospirosis. Curr. Opin.Infect. Dis. 18, 376–386.
7. Gouveia EL, Metcalfe J, Carvalho ALF, et al. (2008) Leptospirosis-associated Severe Pulmonary Hemorrhagic Syndrome, Salvador, Brazil. Emerging Infectious Diseases. 14:505-508.
8. Thaipadungpanit J, Wuthiekanun V, Chierakul W, Smythe LD, Petkanchanapong W, et al. (2007). A dominant clone of *Leptospira interrogans* associated with an outbreak of human leptospirosis in Thailand. PLoS Negl Trop Dis. 31(1), e56.

9. Dikken H and Kmety E. (1978) Serological typing methods of leptospires, p. 259-307. In T. Bergan and J. R. Norris (ed.), Methods in Microbiology, vol. 11. Academic Press, London, UK.
10. Kmety E and Dikken H. (1993) Classification of the species *Leptospira interrogans* and history of its serovars. University Press Groningen, Groningen, The Netherlands.
11. Haapala DK, Rogul M, Evans LB and Alexander AD. (1969) Deoxyribonucleic acid base composition and homology studies of Leptospira. J. Bacteriol. 98(2):421.
12. Yasuda PH, Steigerwalt AG, Sulzer KR, Kaufmann AF, Rogers FC, Brenner DJ (1987) Deoxyribonucleic acid relatedness between serogroups and serovars in the family Leptospiraceae with proposals for seven new Leptospira species. Int. J.Syst. Bacteriol. 37, 407-415.
13. Brenner DJ, Kaufmann AF, Sulzer KR, Steigerwalt AG, Rogers FC, Weyant RS. (1999) Further determination of DNA relatedness between serogroups and serovars in the family Leptospiraceae with a proposal for Leptospira alexanderi sp. nov. and four new Leptospira genomospecies. Int. J. Syst. Bacteriol. 49, 839–858.
14. Slack AT, Kalambaheti T, Symonds ML, Dohnt MF, Galloway RL, et al. (2008) Leptospira wolffii sp. nov., isolated from a human with suspected leptospirosis in Thailand. Int. J. Syst. Evol. Microbiol. 58, 2305–2308.
15. Slack AT, Khairani-Bejo S, Symonds ML, Dohnt MF, Galloway RL, et al. (2009) Leptospira kmetyi sp. nov., isolated from an environmental source in Malaysia. Int. J. Syst. Evol. Microbiol. 59, 705–708.
16. Salaün L, Mérien F, Gurianova S, Baranton G, Picardeau M. (2006). Application of multilocus variable-number tandem-repeat analysis for molecular typing of the agent of

leptospirosis. J. Clin. Microbiol. 44 (11), 3954- 3962.

17. Anne-Laure Zilber, Mathieu Picardeau, Florence Ayral, Marc Artois, Pierre Demont, Angeli Kodjo, Zoheira Djelouadji (2014). High-Resolution Typing of Leptospira interrogans Strains by Multispacer Sequence Typing. J. Clin. Microbiol. 52(12), 564-571.
18. Thaipadungpanit J, Wuthiekanun V, Chierakul W, Smythe LD, Petkanchanapong W, et al. (2007). A dominant clone of *Leptospira interrogans* associated with an outbreak of human leptospirosis in Thailand. PLoS Negl Trop Dis. 31(1), e56.
19. Bourhy P, Collet L, Clément S, Huerre M, Ave P, et al. (2010). Isolation and characterization of new *Leptospira* genotypes from patients in Mayotte (Indian Ocean). PLoS Negl. Trop. Dis. 4 (6), e724.
20. Tamai T, Sada E, Kobayashi Y. (1988) Restriction endonuclease DNA analysis of Leptospira interrogans serovars Icterohaemorrhagiae and Copenhageni. Microbiol. Immunol. 32, 887–894.
21. Levett PN, Walton D, Waterman LD, Whittington CU, Mathison GE, Edwards CO. (1998). Surveillance of leptospiral carriage by feral rats in Barbados. West Indian Med. J. 47, 15–17.
22. Faria MT, Calderwood MS, Athanazio DA, McBride AJA, Hartskeerl RA, et al. (2008) Carriage of Leptospira interrogans among domestic rats from an urban setting highly endemic for leptospirosis in Brazil. Acta Tropica, 108:1-5.
23. Romero EC, Bernardo CCM, Yasuda PH. (2003) Human Leptospirosis: A twenty-nine-year serological study in São Paulo, Brazil. Rev. Inst. Med. Trop. 45(5):245-248.
24. Hayford AE, Mammel MK, Lacher DW, Brown EW. (2011) Single nucleotide polymorphism (SNP)-based differentiation of Shigella isolates by pyrosequencing.

Infection, Genetics and Evolution 11:1761–1768.

25. Harshavardhan Doddapaneni, Jiqiang Yao , Hong Lin, M Andrew Edwin L Walker, Civerolo. (2006) Analysis of the genome-wide variations among multiple strains of the plant pathogenic bacterium *Xylella fastidiosa*
26. Swofford DL. (2002) PAUP*: phylogenetic analysis using parsimony (* and other methods), version 4.0. Sinauer Associates, Sunderland, MA.
27. Joshi D, Harris NB, Waters R, Thacker T, Mathema B, et al. (2012) Single Nucleotide Polymorphisms in the Mycobacterium bovis Genome Resolve Phylogenetic Relationships. Journal of Clinical Microbiology 50(12):3853-3861.
28. Gutacker MM, Smoot JC, Migliaccio CA, et al. (2002) Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. Genetics 162:1533–43.
29. Gutacker MM, Mathema B, Soini H, Shashkina E, Kreiswirth BN, et al. (2006) Single-Nucleotide Polymorphism–Based Population Genetic Analysis of Mycobacterium tuberculosis Strains from 4 Geographic Sites. JID 193:121-128.
30. Harrison ME, Paterson GK, Holden MTG, Larsen J, Stegger M, et al. (2013) Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel mecA homologue mecC. EMBO Mol Med 5:509–515.
31. Kimura M. The neutral theory of molecular evolution. Cambridge: Cambridge University Press, 1983.
32. Nascimento ALTO, Verjovski-Almeida S, Van Sluys MA, et al. (2004) Genome features of *Leptospira interrogans* serovar Copenhageni. Brazilian Journal of Medical and

Biological Research. 37:459-478.

33. Suepaul SM, Carrington CV, Campbell M, Borde G and Adesiyun AA. (2014) Seroepidemiology of leptospirosis in dogs and rats in Trinidad. Tropical Biomedicine 31(4): 853-861.
34. Esfandiari B, Pourshafie MR, Gouya MM, Khaki P, Mostafavi E, et al. (2015) An epidemiological comparative study on diagnosis of rodent leptospirosis in Mazandaran Province, northern Iran. Epidemiol Health.
35. Loan HK, Cuong NV, Takhampunya R, Kiet BT, Campbell J, et al. (2015) How important are rats as vectors of leptospirosis in the mekong delta of Vietnam? Vector-Borne and Zoonotic Diseases 15(1): 56-64.
36. Orskov F. Orskov I, Sutton A, Schneerson R, Lin W, et al. (1979) Form variation in Escherichia coli K1: determined by O-acetylation of the capsular polysaccharide. JEM 149(3):669-685.
37. Bhattacharjee AK, Jennings HJ, Kenny CP, Martin A, Smith IC. (1976) Structural determination of the polysaccharide antigen of Neisseria meningidis serogroups Y, W-135, and BO1. Can. J. Biochem. 54, 1-8.
38. Knirel, Y. A., Vinogradov, E. V., Shashkov, A. S., Dmitriev, B. A., Kochetkov, N. K., Stanislavsky, E. S., and Mashilova, G. M. (1986) Somatic antigens of Pseudomonas aeruginosa. The structure of O-specific polysaccharide chains of P. aeruginosa O10 (Lányi) lipopolysaccharides. Eur. J. Biochem. 157,129-138.
39. Knirel, Y. A., Kocharova, N. A., Shashkov, A. S., Dmitriev, B. A., Kochetkov, N. K., Stanislavsky, E. S., and Mashilova, G. M. (1987) Somatic antigens of Pseudomonas aeruginosa. The structure of O-specific polysaccharide chains of the lipopolysaccharides

from P. aeruginosa O5 (Lányi) and immunotype 6 (Fisher)Eur. J. Biochem. 163, 639-652.

40. Thibault, P., Logan, S. M., Kelly, J. F., Brisson, J. R., Ewing, C. P., Trust, T. J., and Guerry, P. (2001) Identification of the carbohydrate moieties and glycosylation motifs in Campylobacter jejuni flagellin. J. Biol. Chem. 276, 34862-34870.
41. Berry DS. (2002) Effect of O Acetylation of Neisseria meningitidisSerogroup A Capsular Polysaccharide on Development of Functional Immune Responses. Infect. Immun 70:(7)3707-3713.
42. Fattom AI, Sarwar J, Basham L, Ennifar S, Naso R. (1998) Antigenic determinants of Staphylococcus aureus type 5 and type 8 capsular polysaccharide vaccines. Infect. Immun. 66, 4588-4592.
43. McNeely TB, Staub JM, Rusk CM, Blum MJ, Donnelly JJ. (2015) Antibody Responses to Capsular Polysaccharide Backbone and O-Acetate Side Groups of Streptococcus pneumoniae Type 9V in Humans and Rhesus Macaques. Infect. Immun 66:3705-3710.
44. Lunter and Goodson. (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res. 21:936-939.
45. Li H, Handsaker B, Wysoker A, et. al. (2009) The Sequence alignment/map (SAM) format and SAMtools. Bioinformatics, 25, 2078-9.
46. Drummond, A. J.; Rambaut, A. (2007) BEAST: Bayesian evolutionary analysis by sampling trees. BMC Evol Biol, 7:214.
47. Stampy: A statistical algorithm for sensitive and fast mapping of Illumina sequence reads**.** Genome Res. 2011 Jun; 21(6): 936–939.
48. The Sequence Alignment/Map format and SAMtools**.** Bioinformatics. 2009 Aug 15; 25(16): 2078–2079.

Figure legends

Figure 1: SNPs and Indels calling pipeline

Figure 2: Maximum likelihood tree representing phylogenetic relationship among the *Leptospira* strains serovar Copenhageni and Icterohaemorragiae from different geographical locations. Asterisk (*) represents Bootstrap support higher than 70%. The sequences from serovar Icterohaemorragiae are in a rectangle.

Figure 3: A) LIC12008 gene region organization. B) LIC12008 protein domain architecture analysis. C) Amino acid sequence fragment (1-55aa) showing the differences in the *L. interrogans* serovar Copenhageni (I) and *L. interrogans* serovar Icterohaemorrhagiae (II).

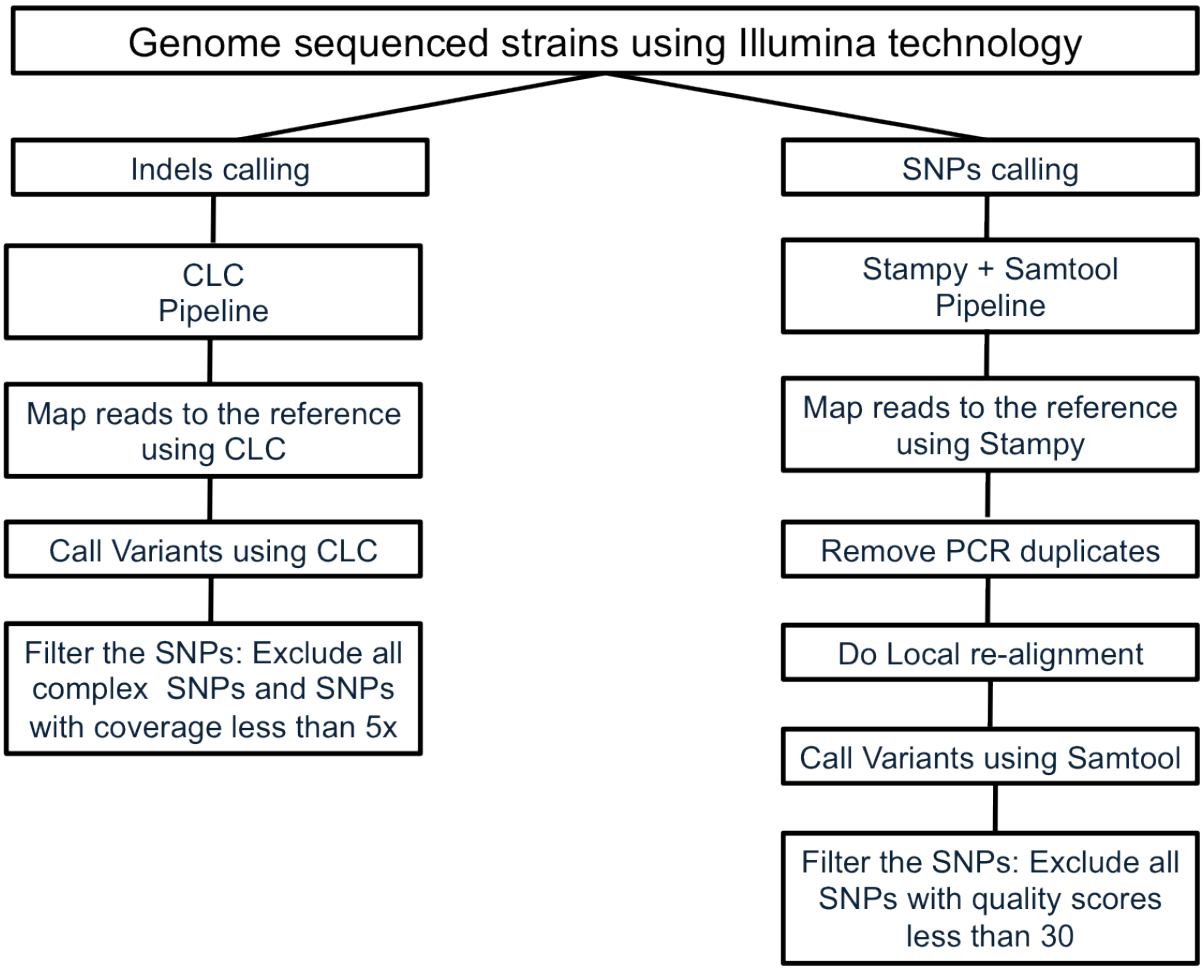
Tables

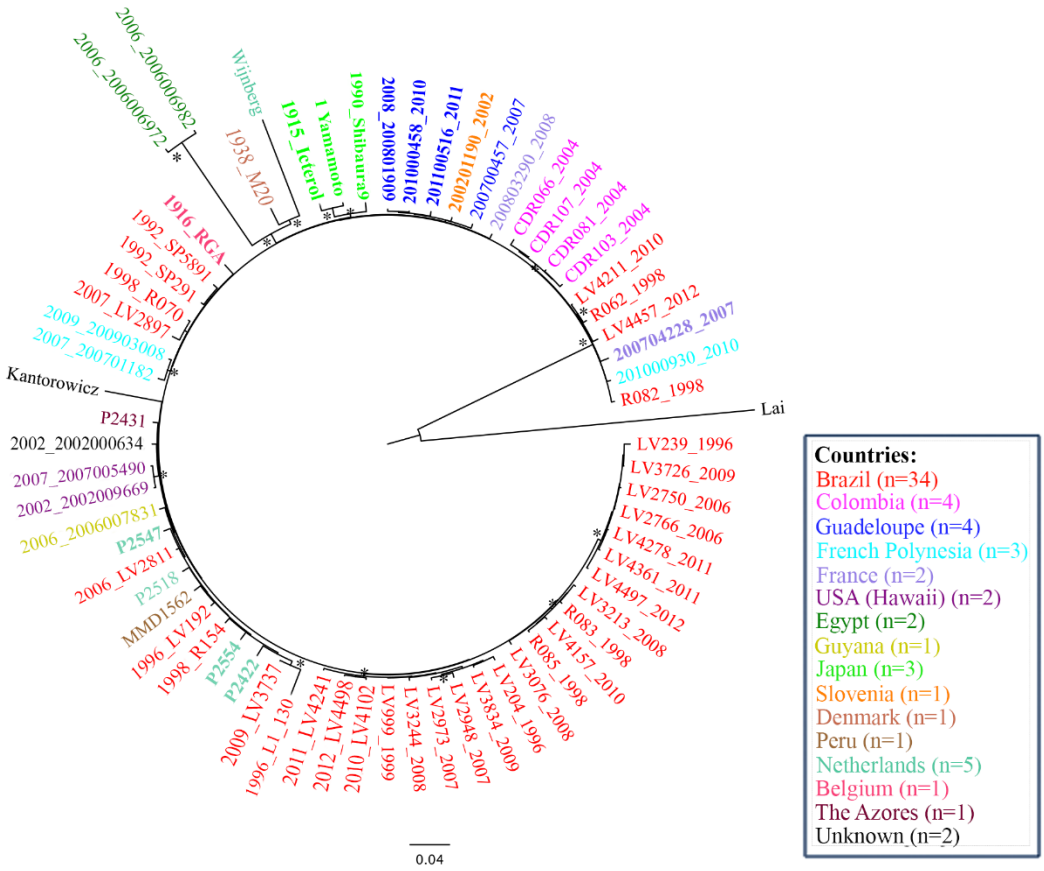
Table 1: Information of the strains in the study

Table 2: Number of SNPs and Indels detected among 55 *L. interrogans* serovar Copenhageni and 12 *L. interrogans* serovar Icterohaemorrhagiae

Table S1: Comparison of the overlapping rates for CLC and Samtools SNP calling pipeline

Table S2: Comparison of the overlapping rates for CLC and Samtools Indels calling pipelin


Genome sequenced strains using Illumina technology
Indels calling
CLC Pipeline
Map reads to the reference using CLC
Call Variants using CLC
Filter the SNPs: Exclude all complex SNPs and SNPs with coverage less than 5x
SNPs calling
Stampy + Samtool Pipeline
Map reads to the reference using Stampy
Remove PCR duplicates
Do Local re-alignment
Call Variants using Samtool
Filter the SNPs: Exclude all SNPs with quality scores less than 30


2006_2006006982
2006_2006006972
Wijnberg
1938_M20
1915_Icterol
1 Yamamoto
1990_Shibaura9
2008_2008010909
201000458_2010
201100516_2011
200201190_2002
200700457_2007
200803290_2008
CDR066_2004
CDR107_2004
CDR081_2004
CDR103_2004
LV4211_2010
R062_1998
LV4457_2012
200704228_2007
201000930_2010
R082_1998
Lai
LV239_1996
LV3726_2009
LV2750_2006
LV2766_2006
LV4278_2011
LV4361_2011
LV4497_2012
LV3213_2008
R083_1998
LV4157_2010
R085_1998
LV3076_2008
LV204_1996
LV3834_2009
LV2948_2007
LV2973_2007
LV3244_2008
LV999_1999
LV4102
2010_LV4498
2012_LV4241
2011_LV4241
1996_L1_130
2009_LV3737
P2422
P2554
1998_R154
1996_LV192
MMD1562
P2518
2006_LV2811
P2547
2006_2006007831
2002_2002009669
2007_2007005490
2002_2002000634
P2431
Kantorowicz
2007_200701182
2009_200903008
2007_LV2897
1998_R070
1992_SP291
1992_SP5891
1916_RGA
0.04
Countries:
Brazil (n=34)
Colombia (n=4)
Guadeloupe (n=4)
French Polynesia (n=3)
France (n=2)
USA (Hawaii) (n=2)
Egypt (n=2)
Guyana (n=1)
Japan (n=3)
Slovenia (n=1)
Denmark (n=1)
Peru (n=1)
Netherlands (n=5)
Belgium (n=1)
The Azores (n=1)
Unknown (n=2)

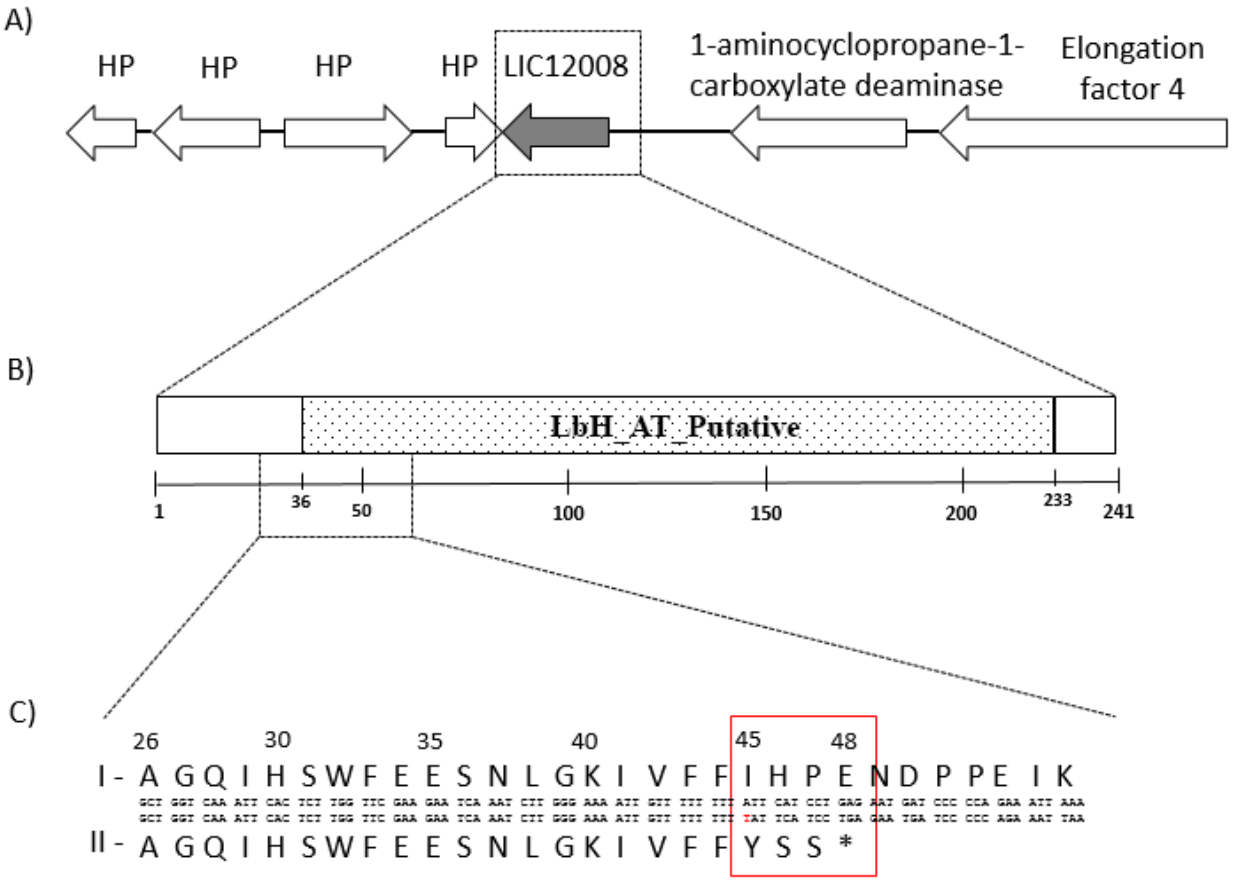

A)
HP
HP
HP
HP
LIC12008
1-aminocyclopropane-1-carboxylate deaminase
Elongation factor 4
B)
LbH_AT_Putative
1
36
50
100
150
200
233
241
C)
26
30
35
40
45
48
I- A G Q I H S W F E E S N L G K I V F F I H P E N D P P E I K
II- A G Q I H S W F E E S N L G K I V F F Y S S *

Table 1: Information of the strains in the study

Species	Serovar	Strain	Host	Country	Year	Accession number
L. interrogans	Copenhageni	Fiocruz LV130 3.7	Human	Brazil	1996	NA
L. interrogans	Copenhageni	Fiocruz LV192	Human	Brazil	1996	SRX055276
L. interrogans	Copenhageni	Fiocruz LV204	Human	Brazil	1996	SRX055274
L. interrogans	Copenhageni	Fiocruz LV239	Human	Brazil	1996	SRX055281
L. interrogans	Copenhageni	Fiocruz LV2750	Human	Brazil	2006	SRX237194
L. interrogans	Copenhageni	Fiocruz LV2766	Human	Brazil	2006	NA
L. interrogans	Copenhageni	Fiocruz LV2811	Human	Brazil	2006	SRX055230
L. interrogans	Copenhageni	Fiocruz LV2897	Human	Brazil	2007	SRX055235
L. interrogans	Copenhageni	Fiocruz LV2948	Human	Brazil	2007	SRX055240
L. interrogans	Copenhageni	Fiocruz LV2973	Human	Brazil	2007	SRX237225
L. interrogans	Copenhageni	Fiocruz LV3076	Human	Brazil	2008	NA
L. interrogans	Copenhageni	Fiocruz LV3213	Human	Brazil	2008	NA
L. interrogans	Copenhageni	Fiocruz LV3244	Human	Brazil	2008	SRX236775
L. interrogans	Copenhageni	Fiocruz LV3726	Human	Brazil	2009	SRX055253
L. interrogans	Copenhageni	Fiocruz LV3737	Human	Brazil	2009	SRX055255
L. interrogans	Copenhageni	Fiocruz LV3834	Human	Brazil	2009	SRX055256
L. interrogans	Copenhageni	Fiocruz LV4102	Human	Brazil	2010	NA
L. interrogans	Copenhageni	Fiocruz LV4157	Human	Brazil	2010	NA
L. interrogans	Copenhageni	Fiocruz LV4211	Human	Brazil	2010	SRX236694
L. interrogans	Copenhageni	Fiocruz LV4241	Human	Brazil	2011	NA
L. interrogans	Copenhageni	Fiocruz LV4278	Human	Brazil	2011	NA
L. interrogans	Copenhageni	Fiocruz LV4361	Human	Brazil	2011	NA
L. interrogans	Copenhageni	Fiocruz LV4457	Human	Brazil	2012	NA
L. interrogans	Copenhageni	Fiocruz LV4497	Human	Brazil	2012	NA
L. interrogans	Copenhageni	Fiocruz LV4498	Human	Brazil	2012	NA
L. interrogans	Copenhageni	Fiocruz LV999	Human	Brazil	1998	SRX237193
L. interrogans	Copenhageni	Fiocruz R062	Rat	Brazil	1998	NA
L. interrogans	Copenhageni	Fiocruz R070	Rat	Brazil	1998	NA
L. interrogans	Copenhageni	Fiocruz R082	Rat	Brazil	1998	NA
L. interrogans	Copenhageni	Fiocruz R083	Rat	Brazil	1998	NA
L. interrogans	Copenhageni	Fiocruz R085	Rat	Brazil	1998	NA
L. interrogans	Copenhageni	Fiocruz R154	Rat	Brazil	1998	SRX237044
L. interrogans	Copenhageni	SP 2/91	NA	Brazil	1992	NA
L. interrogans	Copenhageni	SP 58/91	NA	Brazil	1992	NA
L. interrogans	Copenhageni	CIDEIM R066	Rat	Colombia	2004	SRX055261
L. interrogans	Copenhageni	CIDEIM R081	Rat	Colombia	2004	NA
L. interrogans	Copenhageni	CIDEIM R103	Rat	Colombia	2004	SRX237045
L. interrogans	Copenhageni	CIDEIM R107	Rat	Colombia	2004	NA
L. interrogans	Copenhageni	M20	Human	Denmark	1938	NA

L. interrogans	Copenhageni	2006006972	Human	Egypt	2006	NA
L. interrogans	Copenhageni	2006006982	Human	Egypt	2006	NA
L. interrogans	Copenhageni	200803290	Human	France	2008	NA
L. interrogans	Copenhageni	200701182	Human	French Polynesia	2007	SRX236905
L. interrogans	Copenhageni	200903008	Human	French Polynesia	2009	NA
L. interrogans	Copenhageni	201000930	Human	French Polynesia	2010	NA
L. interrogans	Copenhageni	200700457	Human	Guadeloupe	2007	NA
L. interrogans	Copenhageni	2006007831	Human	Guyana	2006	NA
L. interrogans	Copenhageni	2002009669	Human	Hawaii	2002	NA
L. interrogans	Copenhageni	2007005490	Human	Hawaii	2007	NA
L. interrogans	Copenhageni	shibaura 9 (Yanagawa (1990))	Human	Japan	1990	NA
L. interrogans	Copenhageni	2002000634	Human	ND	2002	NA
L. interrogans	Copenhageni	P2518	Human	Netherlands	NA	SRX101365
L. interrogans	Copenhageni	Wijinberg	Human	Netherlands	NA	SRX236777
L. interrogans	Copenhageni	MMD1562	Bat	Peru	NA	SRX236211
L. interrogans	Copenhageni	P2431	Human	the Azores	NA	SRX236906
L. interrogans	Icterohaemorrhagiae	RGA	Human	Belgium	1916	NA
L. interrogans	Icterohaemorrhagiae	200704228	Human	France	2007	NA
L. interrogans	Icterohaemorrhagiae	200801909	Human	Guadeloupe	2008	NA
L. interrogans	Icterohaemorrhagiae	201000458	Human	Guadeloupe	2010	NA
L. interrogans	Icterohaemorrhagiae	201100516	Human	Guadeloupe	2011	NA
L. interrogans	Icterohaemorrhagiae	1 (yamamoto)	Human	Japan	NA	NA
L. interrogans	Icterohaemorrhagiae	Ictero 1	Guinea-Pig	Japan	1915	NA
L. interrogans	Icterohaemorrhagiae	KANTOROWICZ	ND	ND	ND	NA
L. interrogans	Icterohaemorrhagiae	P2422	Human	Netherlands	NA	SRX101371
L. interrogans	Icterohaemorrhagiae	P2547	Human	Netherlands	NA	SRX101367
L. interrogans	Icterohaemorrhagiae	P2554	Human	Netherlands	NA	SRX101366
L. interrogans	Icterohaemorrhagiae	200201190	Human	Slovenia (Eva)	2002	NA

Table 1: Number of SNPs and Indels detected among 55 *L. interrogans* serovar Copenhageni and 12 *L. interrogans* serovar Icterohaemorrhagiae.

	Coding region	**Non-coding region**	**Total**
SNPs			
Synonymous	258	NA	NA
Non-synonymous	538	NA	NA
Total	796	276	1072
Indels			
Insertion	80	33	113
Deletion	98	24	122
Total	178	57	235

NA = Not applied

Supplementary Table 1: Comparison of the overlapping rates for CLC and Samtools SNP calling pipeline

	CLC			Stampy+Samtools		
Strain	# of SNPs detected in the first sequence	# of SNPs detected in the second sequence	CLC validation rate	# of SNPs detected in the first sequence	# of SNPs detected in the second sequence	Samtools validation rate
CIDER103	102	81	0.77	91	95	1.00
LV2776	84	63	0.74	73	75	1.00
LV2791	93	73	0.77	77	83	0.99
LV2805	78	62	0.76	62	71	0.98
LV2953	88	77	0.78	76	87	0.97
LV3094	87	70	0.78	71	84	0.97
LV212	72	59	0.72	59	71	0.98

S**upplementary Table 2**: Comparison of the overlapping rates for CLC and Samtools Indel calling pipeline

Strain	CLC			Stampy+Samtools		
	# of Indels detected in the first sequence	# of Indels detected in the second sequence	CLC validation rate	# of Indels detected in the first sequence	# of Indels detected in the second sequence	Samtools validation rate
CIDER103	72	68	0.93	90	95	0.81
LV2776	68	61	0.90	70	75	0.80
LV2791	71	65	0.90	70	88	0.87
LV2805	68	64	0.94	46	75	1.00
LV2953	66	66	0.95	53	80	0.94
LV3094	67	68	0.96	53	87	0.96
LV212	61	63	0.98	42	73	1.00

3.3 ANÁLISE COMPARATIVA DO GENOMAS DAS DIFERENTES ESPÉCIES DO GÊNERO DA *LEPTOSPIRA*.

Estudos anteriores demostraram a presença de proteínas do tipo Lig (*Leptospiral Ig-like*) em espécie patogênicas de *Leptospira*. As proteínas do tipo Lig apresentam repetições em tamden de domínios de Big (*bacterial imunoglobuline-like*) que é um fator de virulência em outras bactérias (MATSUNAGA et al, 2003; MCBRIDE et al, 2009; CERQUEIRA et al). Além disto, Lessa-Aquino e colaboradores identificaram 27 proteínas imunodominantes, incluindo as Ligs, que apresentaram uma alta reatividade contra soro de pacientes infectados pela *L. interrogans* serovar Copenhageni (LESSA-AQUINO et al, 2013). Esses resultados sugerem que essas proteínas podem ser um importante alvo de vacinas e diagnóstico e para isso faz-se necessário a identificação e caracterização das diferenças genéticas entre os genes imunodominantes das 20 espécies de *Leptospira*.

Inicialmente foi realizada a comparação da identidade entre as sequências de aminoácido das proteínas imunodominantes com as proteínas pertencentes às demais espécies através da ferramenta *on line* BLAST. A LIC12180 (metiltransferase) só foi encontrada na espécie *L. interrogas* e LIC11222 (Proteína de repetição de Tetratricopeptideo) e LIC11352 (LipL32) não foram encontradas nas espécies não patogênicas. As proteínas LIC11335 (Chaperona GroEL), LIC10623 (Proteina flagelar MotB) e LIC10524 (chaperona molecular DnaK) foram encontradas em todas as espécies com identidade maior que 70% (Tabela 1). A LigB foi encontada em todas as espécies patogênicas enquanto que a LigA só foi encontrada nas espécies *L. alstoni* e *L. kirschneri,* além da *L. interrogan.* Já a LigC foi encontrada nas cinco espécies intermediárias, além de cinco das nove espécies patogênicas. Nenhuma das Ligs foram identificada nos genomas das espécies saprofíticas. A LigC foi previamente caracterizada como um pseudogene (MATSUNAGA et al, 2003; MCBRIDE et al, 2009; CERQUEIRA et al). Entretanto, a partir de análises mais completas foi encontrada uma cópia da LigC intacta, ou seja, com o genoma sem códons de parada. Vale ressaltar que nesse trabalho também foram identificados quatro genes adicionais que codificam domínios semelhantes aos encontrados nas Ligs em espécies não patogênicas (Tabela 2).

Os resultados acima descritos encontram-se citados no artigo intitulado "*What Makes a Bacterial Species Pathogenic?: Comparative Genomic Analysis of the Genus Leptospira Reveals Evolution from Saprophyte to Pathogen"* Este trabalho é fruto do projeto multicêntrico do *National Institute of Health* (NIH) que ocorreu com a colaboração de diversos grupos de pesquisa. Este manuscrito está em processo de submissão.

Tabela 1: Identidade das sequências da aminoácido das proteínas codificadas pelo gene dos genomas das 20 espécies que são genes ortologos a proteínas imunodominates de *L. interrogans* serovar Copenhageni.*

Nome da Proteína	Número de acesso GenBank	Função	AUC Agudo¶	AUC Convalescente§	*L. interrogans* sv. Copenhageni str. Fiocruz L1-130	*L. interrogans* sv. Copenhageni str. M20	*L. interrogans* sv. Lai str. 55601	*L. kirschneri* sv. Cynopteri str. 3522 CT	*L. noguchii* sv. Panama str. CZ 214T	*L. alstoni* sv. Pingchang str. 80-412	*L. weilii* sv. undetermined str. LNT 1234	*L. alexanderi* sv. Manhao str. 3 L 60T	*L. borgpetersenii* sv. Javanica str. UI 09931	*L. santarosai* sv. Shermani str. 1342KT	*L. kmetyi* sv. Malaysia str. Bejo-Iso9T	*L. fainei* sv. Hursibridge str. BUT 6T	*L. broomii* sv. Hurstbridge str. 5399T	*L. wolffii* sv. Khorat str. Khorat-H2T	*L. licerasiae* sv. Varillal str. VAR 010	*L. inadai* sv. Lyme str. 10T	*L. wolbachii* sv. Codice str. CDC	*L. yanagawae* sv. Saopaulo str. Sao Paulo	*L. bifletxa* sv. Patoc str. Patoc1	*L. vanthielii* sv. Holland str. Waz Holland	*L. terpstrae* sv. Hualin str. LT 11-33T	*L. meyeri* sv. Hardjo str. Went 5
					P	P	P	P	P	P	P	P	P	P	P	I	I	I	I	I	S	S	S	S	S	S
LigA7-13	NC_005823.1	LigA região não-idêntica	0.894	0.965	100	100	-	90	-	56	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LigB7-12	NC_005823.1	LigB região não-idêntica	0.857	0.968	100	100	99	93	80	48	62	64	61	62	63	-	-	-	-	-	-	-	-	-	-	-
LIC11352	NC_005823.1	Lipl32	0.841	0.986	100	100	100	100	100	98	94	94	99	99	94	71	71	69	67	71	-	-	-	-	-	-
LIC10215	NC_005823.1	TRAM domínio proteico	0.865	0.879	100	100	100	95	96	94	96	96	96	96	94	80	80	80	78	80	69	70	72	69	69	69
LIC11573	NC_005823.1	Sistema secreção putativo tipo II proteína G	0.775	0.926	100	100	100	99	98	92	92	93	94	92	94	64	64	56	56	64	52	52	51	-	52	50
LIC11456	NC_005823.1	Lipoproteína Putativa	0.763	0.851	100	100	100	98	97	93	90	91	92	93	89	58	54	48	48	54	51	49	51	51	51	53
LigA/B1-6	NC_005823.1	LigA and LigB região idêntica	0.785	0.956	100	99	94	92	81	68	63	64	63	65	62	-	-	-	-	-	-	-	-	-	-	-
LIC11335	NC_005823.1	Chaperona GroEL	0.724	0.749	100	100	99	99	99	97	96	96	93	97	98	92	92	91	91	92	86	86	86	86	87	86
LIC11222	NC_005823.1	Proteína de repetição de Tetratricopeptideo	0.711	0.568	100	100	99	88	87	61	62	61	59	-	59	36	37	35	29	38	-	-	-	-	-	-
LIC11389	NC_005823.1	Proteína flagelar FlbB	0.737	0.770	100	100	100	98	97	93	91	91	92	92	93	72	71	73	71	72	58	58	58	58	58	58
LIC11955	NC_005823.1	Proteína hipotética	0.687	0.477	100	100	100	98	97	97	96	96	84	96	81	79	79	77	77	75	48	48	48	48	48	49
LIC11271	NC_005823.1	Proteína hipotética	0.713	0.877	100	100	99	96	93	80	79	78	80	81	79	63	62	67	67	62	57	55	55	57	57	56
LIC10486	NC_005823.1	Proteína hipotética	0.677	0.831	100	100	99	98	97	94	90	93	90	90	89	57	57	61	58	57	37	36	38	36	36	37
LIC12180	NC_005823.1	Metiltransferase	0.663	0.534	100	100	100	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LIC10191	NC_005823.1	OmpA	0.678	0.655	100	100	99	98	98	96	95	94	93	94	95	71	69	71	71	68	54	56	56	53	54	54
LIC20042	NC_005824.1	BatC	0.672	0.525	100	100	99	91	88	70	66	66	70	68	68	47	47	51	48	45	32	30	33	32	33	35
LIC20087	NC_005824.1	Proteína de Membrana	0.699	0.948	100	100	99	96	95	87	85	86	85	85	83	37	38	36	42	39	29	28	27	28	28	28
LIC12544	NC_005823.1	Regulado transcricional	0.615	0.917	100	100	100	99	99	95	94	94	94	93	94	75	74	78	76	73	62	63	63	62	62	63
LIC11570	NC_005823.1	Sistema secreção putativo tipo II proteína D	0.750	0.892	100	100	100	98	99	93	92	91	92	93	90	78	77	77	77	76	62	63	63	62	62	63
LIC20301	NC_005824.1	Proteína hipotética	0.698	0.861	100	100	100	100	99	95	93	93	93	92	90	78	78	75	74	78	74	50	47	47	47	47
LIC10524	NC_005823.1	Chaperona molecular DnaK	0.781	0.846	100	100	99	98	98	95	96	96	96	96	95	87	87	86	87	88	80	81	81	80	81	80
LIC11437	NC_005823.1	Adenilato/guanilato ciclase	0.633	0.767	100	100	99	94	91	78	73	71	74	75	75	51	50	61	64	50	49	48	48	49	49	48
LIC10483	NC_005823.1	Família da Proteína de Membrana PF09851	0.575	0.715	100	100	100	99	98	96	97	95	96	95	95	77	81	78	78	76	57	58	55	57	57	56
LIC10623	NC_005823.1	Proteína de motor flagelar MotB	0.456	0.670	100	100	99	100	99	96	97	97	97	98	95	82	83	83	84	82	72	73	73	72	72	73

* As proteínas imunodominantes foram identificadas no estudo (Lessa-Aquino et al; PLoS Neglected Tropical Diseases; 2013; PMID: 24147173) que analisou a atividade sorológica de soro de pacientes com leptospirose e indivíduos controles saudáveis contra proteínas da L. interrogans sorovar Copenhageni cepa Fiocruz L1-130 *proteome array*. Marcado em marrom índia presença de ortologos com ORF completa e intacta e pertencentes ao mesmo cluster do Pantoc.

¶ Calculo da área sob uma curva (AUC) em análises de curva de operação do receptor da reatividade de soros da fase aguda contra antígenos de *L. interrogans* sorovar Copenhageni.

§ Calculo da área sob uma curva (AUC) em análises de curva de operação do receptor da reatividade de soros da fase convalescente contra antígenos de *L. interrogans* sorovar Copenhageni.

P = espécie patogênica; I = espécie intermediaria; S = espécie saprofítica

Tabela 2: Identidade de aminoácidos das proteínas Ligs e suas diferentes regiões as encontradas nos genomas das 20 espécies

Nome da Proteína	Região	No. AA	No. Domínios Big	*L. interrogans* sv. Copenhageni str. Fiocruz L1-130	*L. interrogans* sv. Copenhageni str. M20	*L. interrogans* sv. Lai str. 56601	*L. kirschneri* sv. Cynopteri str. 3522 CT	*L. noguchii* sv. Panama str. CZ 214T	*L. alstoni* sv. Pingchang str. 80-412	*L. weilii* sv. undetermined str. LNT 1234	*L. alexanderi* sv. Manhao str. 3 L 60T	*L. borgpetersenii* sv. Javanica str. UI 09931	*L. santarosai* sv. Shermani str. 1342KT	*L. kmetyi* sv. Malaysia str. Bejo-Iso9T	*L. fainei* sv. Hurstbridge str. BUT 6T	*L. broomii* sv. Hurstbridge str. 5399T	*L. wolffii* sv. Khorat str. Khorat-H2T	*L. licerasiae* sv. Varillal str. VAR 010	*L. inadai* sv. Lyme str. 10T	*L. wolbachii* sv. Codice str. CDC	*L. yanagawae* sv. Saopaulo str. Sao Paulo	*L. biflexa* sv. Patoc str. Patoc1	*L. vanthielii* sv. Holland str. Waz Holland	*L. terpstrae* sv. Hualin str. LT 11-33T	*L. meyeri* sv. Hardjo str. Went 5
				P	P	P	P	P	P	P	P	P	P	P	I	I	I	I	I	S	S	S	S	S	S
LigA	Proteína completa	1245	13	100	99	-	91	-	60	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Domínio Big 1-6	LigA e LigB região idêntica	519	6	100	99	-	92	-	68	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
Domínio Big 7-13	LigA região não-idêntica	600	7	100	100	-	90	-	56	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-
LigB	Proteína completa	1922	12	100	99	97	91	74	64	65	66	64	66	65	-	-	-	-	-	-	-	-	-	-	-
Domínio Big1-6	LigA and LigB região idêntica	519	6	100	99	94	92	81	68	66	66	66	66	62	-	-	-	-	-	-	-	-	-	-	-
Domínio Big 7-12	LigB região não-idêntica	633	6	100	100	99	93	80	48	66	66	66	65	63	-	-	-	-	-	-	-	-	-	-	-
C-terminal Não-Big	Região C-terminal	772	NA	100	100	96	91	74	67	72	74	66	71	70	-	-	-	-	-	-	-	-	-	-	-
LigC*	Proteína completa	1956	12	99¶	99	100§	95	95	-	-	-	88	-	-	63	63§	64	63	64$	-	-	-	-	-	-
Domínio Big 1-12	Região de repetição do domínio Big	1086	12	99¶	99	100	95	94	-	-	-	86	-	-	59	59	65	59	60	-	-	-	-	-	-
C-terminal Não-Big	Região C-terminal	790	NA	100¶	100	100	96	96	-	-	-	92	-	-	71	71	77	77	70	-	-	-	-	-	-
LIC13050	Proteína completa	366	2	100	100	100	95	95	87	88	88	84	88	89	46	46	54	49	46	48	50	50	47	51	49
LEP1GSC195_2902	Proteína completa	467	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100+	57	57	66	78	79
LEP1GSC047_3056	Proteína completa	324	1	-	-	-	-	-	-	-	-	-	-	-	-	100^	-	-	94	-	-	-	-	-	-
LEP1GSC195_2226	Proteína completa	450	1	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	100$	-	-	-	-	-

Big, *Bacterial immunoglobulin-like repeat*. Marcado em marrom índia presença de ortologos com ORF completa e intacta e pertencentes ao mesmo cluster do Pantoc, enquanto quemarcado em amarelo indica a presença de ortologo que são pseudogenes. P = espécie patogênica; I = espécie intermediaria; S = espécie saprofitica. AA = aminoácidos na ORF, NA = Não se aplica

* Sequencia proteica da LigC da *L. interrogans* sv. Pomona foi usada como referencia. Número de acesso do GenBank: AY327260.1

¶ LigC é um pseudogene in *L. interrogans* sv. Copenhagfeni cepa Fiocruz L1-130.

§ NCBI caracterizou estas proteínas como LigB-*like*, porém demonstrou alta identidade com a proteína LigC.

+ Sequencia proteica da *L. wolbachii* sv. Codice str. CDC foi usada como referencia. Número de acesso do GenBank: EOQ96200.1

^ Sequencia proteica da *L. broomii* sv. Hurstbridge str. 5399T foi usada como referencia. Número de acesso do GenBank: EQA46670.1

$ Sequencia proteica da *L. wolbachii* sv. Codice str. CDC foi usada como referencia. Número de acesso do GenBank: EOQ96544.1

Sequencia proteica da *L. kmetyi* sv. Malaysia str. Bejo-Iso9T foi usada como referencia. Número de acesso do GenBank: EQA53843.11

4 DISCUSSÃO

O gênero *Leptospira* apresenta uma grande diversidade com 22 espécies identificadas e mais 250 sorovares diferentes (HAAPALA 1969; YASUDA et al, 1987; BRENNER et al, 1999; FAINE et al, 1999; LEVETT, 2001; SLACK et al, 2008; SAITO et al, 2013; BOURHY et al, 2014). No inicio deste estudo só existiam três espécies com o seu genoma sequenciado (REN et al, 2003; NASCIMENTO et al, 2004; BULACH et al, 2006; PICARDEAU et al, 2008). Atualmente, com o desenvolvimento deste estudo 20 espécies de *Leptospira* tem o seu genoma sequenciado e seus principais sorovares, com mais de 400 sequências disponíveis. Com o surgimento de novas técnicas de sequenciamento com um custo menor e grande eficiência se tornou mais acessível a realização de estudo de diversidade genômica de bactéria. Neste trabalho, foi realizado o estudo da diversidade no genoma de diferentes isolados de *L. Interrogans* sorovares Copenhageni e Icterohaemorrhagiae e a correlação desses isolados com dados clínicos e epidemiológicos.

Os isolados de *L. interrogans* sorovares Copenhageni e Icterohaemorrhagiae se mostraram conservados neste estudo. O número de mutações encontradas entre esses isolados foi baixa em relação ao tamanho do genoma, mesmo com a utilização de isolados de localizações geográficas diferentes e com distância na data de isolamento de 100 anos entre a mais antiga e a mais recente. Outro resultado que dá suporte a este achado foi a reconstrução filogenética que demonstrou que não há um estrutura temporal nas árvores, porém foi possível observar alguns clados ligados ao local de isolamento.

Os SNPs detectados nestas sequências mostram não estar sofrendo pressão seletiva. A razão dN/dS encontrada foi de 2:1, o que é um indicativo de não estar sobre pressão seletiva. Outras espécies de bactérias também apresentaram esta razão e uma hipótese para este achado é que não houve tempo suficiente para sofrer seleção e/ou a população bacteriana sofreu uma expansão recente (GUTACKER et al, 2002 e 2006; JOSHI et al, 2012; HARRISON et al, 2013). Porém, neste estudo, foram incluídos isolados de 100 anos de história da infecção por *Leptospira* no mundo levantando o questionamento sobre quanto tempo seria necessário para detectar a seleção dos genes mutados. Outra hipótese é de que o genoma de *Leptospira* apresenta proteínas com funções redundantes, ou seja, na mutação de uma proteína que leve a perda da sua função,

outra proteína com função similar irá ser expressa não permitindo nenhum dano no metabolismo ou adaptação da bactéria aos diferentes ambientes. O grande número de sequências recentes de Salvador incluídas pode ser uma limitação deste estudo, tendo em vista que a inclusão de sequências de isolados mais antigos e de outras regiões geográficas poderia possibilitar a detecção de uma diversidade maior e de genes sobre pressão seletiva.

Neste estudo não foi encontrado nenhuma mutação associada a severidade da doença, desenvolvimento de SHPS ou morte, indicando que as diferenças genômicas do patógeno podem não ser a causa da diversidade clínica. Outro aspecto do patógeno que não foi estudado é a expressão gênica. A presença de um determinado gene não necessariamente indica a expressão deste gene, podendo existir uma relação entre o perfil de expressão gênica com a severidade da doença. Além disto, é difícil o isolamento de *Leptospira* de pacientes assintomáticos, sendo uma limitação de estudos que avaliam fatores do patógeno (FAINE et al, 1999; LEVETT, 2001). Outros possíveis fatores que podem ter relação com a sintomatologia é a carga bacteriana da infecção e fatores do hospedeiro (SCHULLE, 2015). Estudos avaliando estas duas hipóteses irão contribuir para o entendimento da patogênese da doença e assim promover uma melhor assistência ao paciente com leptospirose.

A sorologia sempre foi amplamente utilizada para identificar o sorovar do isolado por este apresentar uma associação com dados clínicos e epidemiológicos, com o hospedeiro e local de infecção (FAINE et al, 1999; LEVETT, 2001). Em algumas técnicas moleculares é possível diferenciar não só as diferentes espécies como também sorovares. Porém, nunca foi possível distinguir os sorovares Copenhageni e Icterohaemorrhagiae neste tipos de teste. No presente estudo foi realizada a comparação dos genomas dos dois sorovares mostrando uma grande similaridade genética, onde a única diferença é a inserção de um nucleotídeo no gene da proteína hipotética LIC12008. Análises *in silico* do perfil proteico demonstrou que este gene está ligado à síntese de LPS, explicando assim as diferenças na sorologia. Apesar do perfil de LPS diferente em espécies patogênicas e não patogênicas, a mutação na LIC12008 não demonstrou alterar a virulência do sorovar. A análise filogenética também demonstrou a similaridade entre os dois sorovares, não havendo formação de clado associado ao sorovar. Com este achado não há a necessidade de diferenciar os dois sorovares em estudos epidemiológicos, uma vez que não há diferença na resposta do hospedeiro a infecção e na virulência dos isolados.

Proteínas identificadas anteriormente como altamente reativas a soro de pacientes infectados pro *L. interrogans* serovar Copenhageni foram encontradas presentes em 20 espécies de *Leptospira* com similaridade maior do que 70% (LESSA-AQUINO, 2013). Essas proteínas podem ser usadas como foco para desenvolvimento de vacinas e diagnóstico que possam ser usadas para as diferentes espécies de *Leptospira*. Além disto a LigC que foi previamente descritas como pseudogene em cepas de *L. interrogas* serova Copenhageni L1-130, apresentou-se sem o códon de parada em outras cepas da mesma espécie (MATSUNAGA et al, 2003; MCBRIDE et al, 2009; CERQUEIRA et al). As funções e vias de atuação das proteínas imunorreativas devem ser estudadas pois podem ser possíveis fatores de virulência de *Leptospira*.

O número de sequências do genoma total das diferentes espécies e sorovares de *Leptospira* possibilita o melhor entendimento da relação entre essas cepas, identificação de fatores de virulência, entendendo melhor os diferentes mecanismos de patogênese e adaptação da bactéria aos seus diferentes hospedeiros e ambientes.

5 CONCLUSÃO

Os resultados encontrados nesse trabalho sugerem:

- Não existem mutações no genoma da *L. interrogans* serovar Copenhageni associadas aos diferentes desfechos clínicos da leptospirose;
- As cepas dos isolados de *L. interrogans* serovar Copenhageni e Icterohaemorrhagiae apresentam uma baixa diversidade sem apresentar associação de suas mutações a dados epidemiológicos;
- Os genomas dos isolados de *L. interrogans* serovar Copenhageni e Icterohaemorrhagiae são altamente conservados e similares com apenas uma mutação que as difere;
- O gene da proteína LigB é encontrado apenas em espécies patogênicas e o gene da proteína LigC é encontrada em espécies patogênicas e intermediárias, sem a presença de códons de parada;
- Sequências conservadas de proteínas imunogênicas foram identificadas nas vinte espécies de *Leptospira*, sugerindo potenciais alvos terapêuticos.

REFERÊNCIAS

BAKER, S. et al. High-Throughput Genotyping of Salmonella enterica Serovar Typhi Allowing Geographical Assignment of Haplotypes and Pathotypes within an Urban District of Jakarta, Indonesia. **J. Clin. Microbiol.,** p. 1741–1746, 2008.

BAKER, S.; HANAGE, WP; HOLT, KE. Navigating the future of bacterial molecular epidemiology. **Curr. Opin. Infec.t Dis.,** v. 13, p. 640–645, 2010.

BARREIRO, L.B. et al. Natural selection has driven population differentiation in modern humans. **Nat. Genet.,** v. 40, p. 340–345, 2008.

BHARTI, A.R. et al. Leptospirosis: A zoonotic disease of global importance. **Lancet Infect. Dis.** v. 3, p. 757-771, 2003.

BIBLIOTECA VIRTUAL EM SAÚDE (BVS). Incidência de leptospirose - D.1.17 – 2011. Disponível em: http://fichas.ripsa.org.br/2011/D-1-17/?l=pt_BR> Acesso em: 22 de setembro de 2015.

BOURHY, P. et al. Isolation and characterization of new Leptospira genotypes from patients in Mayotte (Indian Ocean). **PLoS Negl. Trop. Dis**., v. 4, n. 6, p. 724, 2010.

BOURHY, P. et al. Leptospira mayottensis sp. nov., a pathogenic species of the genus Leptospira isolated from humans. **Int. J. Syst. Evol. Microbiol.,** v. 64, p. 4061-4067, 2014.

BRENNER, D.J. 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.,** v. 49, p. 839–858, 1999.

BROWN, C.T.; FISHWICK, L.K.; CHOKSHI, B.M. Whole-genome sequencing and phenotypic analysis of Bacillus subtilis mutants following evolution under conditions of relaxed selection for sporulation. **Appl. Environ. Microbiol**., p. 6867–6877, 2011.

BULACH, D.M. et al. Genome reduction in Leptospira borgpetersenii reflects limited transmission potential. **Proc. Natl Acad. Sci. USA.,** v. 103, p. 14560–14565, 2006.

CERQUEIRA**,** G.M. et al. Distribution of the leptospiral immunoglobulin-like (*lig*) genes in pathogenic *Leptospira* species and application of *ligB* to typing leptospiral isolates. **J. Med. Microbiol.,** v. 58, p. 1173-1181, 2009.

DEN BAKKER, H.C. et al. A Whole-Genome Single Nucleotide Polymorphism-Based Approach To Trace and Identify Outbreaks Linked to a Common *Salmonella enterica* subsp. *enterica* Serovar Montevideo Pulsed-Field Gel Electrophoresis Type. **Appl. Environ. Microbiol.,** p. 8648–8655, 2011.

DENAMUR, E.; MATIC I. Evolution of mutation rates in bacteria. **Mol. Microbiol.,** v. 60, n. 4, p. 820-827, 2006.

DIAS, J.P.; TEIXEIRA, M.G.; COSTA, M.C.N. et al. Fatores associados à infecção por Leptospira sp em um grande centro urbano do Nordeste do Brasil. **Rev. Soc. Bras. Med.Trop**., v. 40, n. 5, p. 499-504, 2007.

DIKKEN, H.; KMETY, E. Serological typing methods of leptospires. In: BERGAN, T; NORRIS, JR. (Ed.). **Methods in Microbiology**. London: Academic Press, 1978. 259-307 p.

FAINE, S. B. et al. **Leptospira and leptospirosis**. Melbourne: MediSci, 1999.

GONÇALVES, A. J. et al. Hemoptysis and the adult respiratory distress syndrome as the causes of death in leptospirosis: changes in the clinical and anatomicopathological patterns. **Rev. Soc. Bras. Med. Trop.,** v. 25, p. 261–70, 1992.

GOUVEIA, E.L. et al. Leptospirosis-associated Severe Pulmonary Hemorrhagic Syndrome. Salvador, Brazil. **Emerg. Infect. Dis**., v. 14, p. 505-508, 2008.

GUTACKER, M.M. et al. Genome-wide analysis of synonymous single nucleotide polymorphisms in Mycobacterium tuberculosis complex organisms: resolution of genetic relationships among closely related microbial strains. **Genetics,** v. 162, p. 1533–43, 2002.

GUTACKER, M.M. et al. Single-Nucleotide Polymorphism–Based Population Genetic Analysis of Mycobacterium tuberculosis Strains from 4 Geographic Sites. **JID.** v. 193, p. 121-128, 2006.

HAAPALA, D.K. et al. **Bacteriol**. v. 98, n. 2, p. 421, 1969.

HAMBURGER, Z.A.; BROWN, S.M.; ISBERG, R.R; BJORKMAN, P.J. Crystal structure of invasin: a bacterial integrin-binding protein. **Science,** v.286, p.291–295, 1999.

HARRIS, S.R. et al. Evolution of MRSA During Hospital Transmission and Intercontinental Spread. **Science,** v. 327, n. 5964 p. 469-474*,* 2010.

HARRISON, M.E. et al. Whole genome sequencing identifies zoonotic transmission of MRSA isolates with the novel mecA homologue mecC. **EMBO Mol. Med.,** v. 5, p. 509–515, 2013.

HOLT, K.E. et al. High-throughput sequencing provides insights into genome variation and evolution in Salmonella Typhi. **Nat. Genet.,** v. 40, p. 987–993, 2008.

JOSHI, D. et al. Single Nucleotide Polymorphisms in the Mycobacterium bovis Genome Resolve Phylogenetic Relationships. **J. Clin. Microbiol.,** v. 50, n. 12, p. 3853-3861, 2012.

KMETY, E.; DIKKEN, H. **Classification of the species of Leptospira interrogans and history of its serovars.** Groningen: University Press Groningen, 1993.

KO, A.I. et al. Urban epidemic of severe leptospirosis in Brazil. **Lancet,** v. 354, p. 820-825, 1999.

KO, A.I.; GOARANT, C.; PICARDEAU, M. Leptospira: the dawn of the molecular genetics era for an emerging zoonotic pathogen. **Nat. Rev. Microbiol**., v. 7, p. 736-747, 2009.

LESSA-AQUINO, C. et al. Identification of seroreactive proteins of Leptospira interrogans serovar copenhageni using a high-density protein microarray approach. **PLoS Negl. Trop. Dis**, v. 7, n. 10, p. e2499, 2013.

LEVETT, P.N. Leptospirosis. **Clin. Microbiol. Rev**., v. 14, p. 296-326, 2001.

LUNTER; GOODSON. Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. **Genome Res**., v. 21, p. 936-939, 2011.

LUO, Y. et al. Crystal structure of enteropathogenic Escherichia coli intimin-receptor complex. **Nature,** v. 405, p.1073–1077, 2000.

MARDIS, E.R. Next-Generation DNA Sequencing Methods. **Annu. Rev. Genomics Hum. Genet**., v. 9, p. 387–402, 2008.

MATSUNAGA, J. et al. Pathogenic *Leptospira* species express surface-exposed proteins belonging to the bacterial immunoglobulin superfamily. **Mol. Microbiol.,** v. 49, p. 929-945, 2003.

MCBRIDE, A. J. A. Leptospirosis. **Curr. Opin. Infect. Dis.,** v.18 p. 376-386, 2005.

MCBRIDE, A. J. A. et al. Genetic diversity of the Leptospiral immunoglobulin-like (Lig) genes in pathogenic *Leptospira* spp. **Infect. Genet. Evol**., v. 9, p. 196-205, 2009.

NASCIMENTO, A.L.T.O. et al. Genome features of Leptospira interrogans serovar Copenhageni. **Braz. J. Med. Biol. Res**., v. 37, p. 459-478, 2004.

PAPPAS, G.; PAPADIMITRIOU, P.; SIOZOPOULOU, V.; CHRISTOU, L.; AKRITIDIS, N. The globalization of leptospirosis: worldwide incidence trends. **Int. J. Infect. Dis**., v. 12, p. 351-357, 2008.

PARK, S.K. et al. Leptospirosis in Chonbuk Province of Korea in 1987: a study of 93 patients. **Am. J. Trop. Med. Hyg.,** v. 41 p. 345–351, 1989.

PHAN, M.D. et al. Variation in Salmonella enterica serovar Typhi IncHI1 plasmids during the global spread ofresistant typhoid fever. **Anti Agent Chemother**., v. 53, n. 2, p. 716-727, 2009.

PICARDEAU, 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, e1607, 2008.

REN, S. et al. Unique and physiological pathogenic features of Leptospira interrogans revealed by whole genome sequencing. **Nature,** v. 422, p. 888–893, 2003.

SAITO, M. et al. Leptospira idonii sp. nov., isolated from environmental water. **Int. J. Syst. Evol. Microbiol.,** v. 63, p. 2457-2462, 2013.

SALAÜN, L. et al. Application of multilocus variable-number tandem-repeat analysis for molecular typing of the agent of leptospirosis. **J. Clin. Microbiol**. v. 44, n. 11, p. 3954- 3962, 2006.

SCHULLER, S. et al. Comparative proteomic analysis of lung tissue from guinea pigs with leptospiral pulmonary haemorrhage syndrome (LPHS) reveals a decrease in abundance of host proteins involved in cytoskeletal and cellular organization. **J. Prot.** v. 122, p. 55-72, 2015.

SEGURA, E.R. et al. Clinical spectrum of pulmonary involvement in leptospirosis in a region of endemicity, with quantification of leptospiral burden. **Clin. Infect. Dis.,** v. 40, p. 343-351, 2005.

SEHGAL, S.C. Outbreak of Leptospirosis with Pulmonary Involvement in North-Andaman. **Indian J. Med. Res.,** v. 102, p. 9-12, 1995.

SLACK, A.T. et al. Leptospira kmetyi sp. nov., isolated from an environmental source in Malaysia. **Int. J. Syst. Evol. Microbiol**., v. 59, p. 705–708, 2009.

SLACK, A.T. et al. Leptospira wolffii sp. nov., isolated from a human with suspected leptospirosis in Thailand. **Int. J. Syst. Evol. Microbiol**., v. 58, p. 2305–2308, 2008.

STENSON, P.D. et al. The Human Gene Mutation Database: 2008 update. **Genome Med.,** v. 1, n. 1, p. 13, 2009.

THAIPADUNGPANIT, J. et al. A dominant clone of *Leptospira interrogans* associated with an outbreak of human leptospirosis in Thailand. **PLoS Negl Trop Dis**., v. 31, n. 1, p. e56, 2007.

VIEIRA, S.R.R.; BRAUNER, J.S. Leptospirosis as a cause of acute respiratory failure: clinical features and outcome in 35 critical care patients. **Braz. J. Infect. Dis.,** v. 6, p. 135-139, 2002.

WEIL, A. Ueber eine eigentu¨mliche, mit Milztumor, Icterus und Nephritis einhergehende akute Infektionskrankheit. **Dtsche. Arch. Klin. Med.,** v. 39, p. 209–232, 1886.

WORLD HEALTH ORGANIZATION. Leptospirosis worldwide, 1999. **Weekly Epidemiol. Rec.**, v. 74, p. 237-242, 1999.

YASUDA, P.H. et al. Deoxyribonucleic acid relatedness between serogroups and serovars in the family Leptospiraceae with proposals for seven new Leptospira species. **Int. J. Syst. Bacteriol**., v. 37, p. 407-415, 1987.

YERSIN, C. et al. Pulmonary haemorrhage as a predominant cause of death in leptospirosis in Seychelles. **Trans. R. Soc. Trop. Med. Hyg.,** v. 94, p. 71-76, 2000.

ZAKI, S.R.; SHIEH, W.J. Leptospirosis associated with outbreak of acute febrile illness and pulmonary haemorrhage. Nicaragua, 1995. the Epidemic Working group. **Lancet,** v. 347, n. 9000, p. 535-36, 1996.