

SEQUENCIAMENTO, MONTAGEM E ANOTAÇÃO DO GENOMA DE *Leptospira interrogans* CEPA TANDE

GBRIELLE DE OLIVEIRA SANCHES VALERIO NAVARRO¹; FREDERICO SCHIMMITT KREMER¹; SÉRGIO JORGE¹; VINICIUS FARIAS CAMPOS¹; ODIR ANTÔNIO DELLAGOSTIN¹; LUCIANO DA SILVA PINTO¹

¹*Biotecnologia – Universidade Federal de Pelotas*
gabi.oliveira19@hotmail.com

1. INTRODUÇÃO

Leptospira são bactérias espiroquetas classificadas em espécies não patogênicas, intermediárias, ou patogênicas que habitam solo e água doce contaminados, predominantemente nas regiões tropicais (ESHGHI et al., 2015). Classificadas em 21 espécies e cerca de 300 sorovares organizados em 29 sorogrupos. Entre essas 21 espécies estabelecidas, 9 são caracterizadas como patogênicas e são frequentemente isoladas em humanos e animais (MORENO et al., 2016), dentre essas se encontra a *Leptospira interrogans* que é uma bactéria espiroqueta, gram negativa, (GOMES, 2013) responsável pela maior causa de morbidade e mortalidade associada à leptospirose em humanos (LEHMANN et al., 2016).

A crescente disponibilidade de sequências genômicas de espécies pertencentes aos três grupos permitiu a identificação dos processos evolutivos do genoma envolvidos na transição de uma forma não patogênica e de vida livre a um estilo de vida patogênica e adaptada ao hospedeiro. Por exemplo, genômica comparativa revelaram que *L. interrogans* (grupo I - patogênicas) tem um genoma maior que *L. biflexa* (grupo III – não patogênicas), provavelmente refletindo características genéticas adicionais necessárias para a sobrevivência tanto em solo / água quanto a hospedeiros mamíferos. É importante ressaltar que o fato de que *L. interrogans* reteve a capacidade de sobreviver no ambiente como um organismo de vida livre impacta diretamente na ecologia e epidemiologia da leptospirose (IRAOLA et al., 2016).

Para melhor entender o mecanismo de ação da *Leptospira* o sequenciamento e anotação tem se demonstrado uma importante ferramenta a fim de entender a relação genoma-patogenicidade, entender o mecanismo patogênico em manifestações clínicas e identificar抗ígenos comuns (FOUTS et al., 2016).

O isolamento e caracterização da *L. interrogans* sorovar Canicola cepa Tande fora previamente descrito por Brod et al. (Brod et al, 2015), sendo obtido a partir de amostras de um cão infectado na cidade de Pelotas. O presente trabalho teve como objetivo o sequenciamento, montagem e caracterização por MLST da cepa.

2. METODOLOGIA

O sequenciamento do isolado de *Leptospira* foi realizado utilizando a plataforma *Ion Torrent*TM. O sequenciamento foi realizado utilizando uma biblioteca *single-end* de 400 pares de bases, onde a identificação das bases ocorre pela detecção da mudança depH por um sensor que realiza a montagem da sequência correta.

Após o sequenciamento, foi realizada a montagem desse genoma utilizando as ferramentas Newbler, SPAdes (BANKEVICH et al., 2012) e MIRA, e utilizamos o programa CISA (LIN; LIAO, 2013) para integrar as montagens anteriores e obter uma montagem consenso.

A anotação foi realizada utilizando a plataforma Genix (http://labbioinfo.ufpel.edu.br/cgi-bin/genix_index.py). Esta plataforma online integra varias ferramentas de anotação, como por exemplo, Prodigal (que identifica genes demicrorganismos)(HYATT et al., 2010), BLAST (que compara sequências de nucleotídeos de sequências de DNA)(ALTSCHUL et al., 1990), RNAmmer (que identifica RNAsribosomais)(LAGESEN et al., 2007), Aragorn (que identifica RNAs transportadores)(LASLETT, 2004), entre outras. A anotação é realizada em diferentes etapas – cada uma empregando uma ferramenta diferente – e ao final é gerada uma anotação consenso.

Por fim, para a caracterização precisa deste isolado foi realizada uma análise de *multi locus sequence typing* (MLST) *in silico* através de dados obtidos a partir do repositório PubMLST (<http://pubmlst.org/>), utilizando-se o esquema 1 de MLST de *Leptospira*(BOONSILP et al., 2013).O MLST permite discriminar os isolados baseando-se em fragmentos de 450-500 pares de bases de uma sequência de DNA de genes *housekeeping*, em que durante o processo se realiza uma tipagem contra os alelos (*housekeeping genes*). A comparação das sequências do genoma com o banco de dados foi realizada com a ferramenta BLAST. Esta mesma análise foi também realizada para outras cepas do sorovar Canicola disponíveis no GenBank, se forma a se avaliar a concordância entre perfis genotípicos e resposta humoral induzida.

3. RESULTADOS E DISCUSSÃO

Através do sequenciamento obtivemos 644,064 single-endreads, sendo ao todo 131.939,645 bases sequenciadas, onde 106.304.059 tiveram um score de Q20 caracterizando uma cobertura de até 30x cada base e uma precisão de 99%.

A montagem gerou uma sequência genômica consenso final de 106 contigs que somam 4,522,715 nucleotídeos. Através da sequência final, a anotação foi realizada e identificamos: 4.330 CDS, 30 genes de tRNA, 3 genes de rRNA, 1 gene de tmRNA e 5 genes de RNA não codificante e 13 sítios de CRISPR. Os sítios CRISPR encontrados, já foram relatados em outros estudos e estão presentes somente em espécies patogênicas do gênero *Leptospira*, estando associados com a defesa (conduzindo a imunidade contra o invasor).

Por meio da caracterização realizada no MLST (Tabela 01) entre a cepa Tande e outras cepas presentes no sorovar Canicola, é possível observar que mesmo pertencendo à mesma sorovar as diferentes combinações alélicas dos *housekeeping genes* ocasionam diferentes cepas - definidas pelo perfil alélico representado pelas letras ST –.

Tabela 01 – Comparação de diferentes cepas presentes no sorovar Canicola mediante as análises realizadas no MLST.

Cepa	GenBank	Alelo							ST
		<i>tpiA</i>	<i>caIB</i>	<i>sucA</i>	<i>mreA</i>	<i>pntA</i>	<i>pfkB</i>	<i>glmU</i>	
LO-4	LIIYO1	3	5	3	5	3	4	3	37
Tande	MABU01	2	8	2	4	1	10	1	17
P2655	AOWJ01	3	5	3	5	3	4	3	37
LO-3	LIHE01	3	5	3	5	3	4	3	37
LT1962	AFMC02	1	9	62	2	1	6	1	11
Fiocruz LV133	AKWU02	3	5	3	5	3	4	3	37
HAI0024	AFLQ01	2	5	2	7	8	9	6	50

A comparação apresentada na Tabela 01 teve como objetivo confrontar diferentes cepas pertencentes ao sorovar Canicola. Como é possível observar a partir do resultado obtido no MLST, a cepa Tande, que pertence ao sorovar Canicola (pois induz uma resposta imune similar a este grupo), apresenta um perfil genético similar ao do sorovar *Icterohaemorrhagiae* devido ao padrão alélico (ST – 17), e difere das demais cepas mostradas na tabela.

Estes resultados demonstram que a caracterização precisa dos isolados é de suma importância, como demonstrado em outros estudos, uma vez que cada isolado através da combinação dos housekeeping (citados na Tabela 01) irá produzir funções essências para a sua sobrevivência, metabolismo, secreção, manutenção celular e estrutural e adaptação da bactéria ao hospedeiro.

4. CONCLUSÕES

Pode-se concluir com este trabalho que através do sequenciamento, montagem e anotação foi possível identificar a *Leptospira interrogans* sorovar Canicola cepa Tande e que através das análises realizadas pelo MLST nos permitiu uma caracterização precisa e consequentemente uma melhor compreensão evolutiva, filogeográfica e genética entre as diferentes cepas.

5. REFERÊNCIAS BIBLIOGRÁFICAS

ALTSCHUL, S. F. et al. Basic local alignment search tool. **Journal of molecular biology**, v. 215, n. 3, p. 403–10, 5 out. 1990.

BANKEVICH, A. et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **Journal of computational biology: a journal of computational molecular cell biology**, v. 19, n. 5, p. 455–77, maio 2012.

BOONSILP, S. et al. A single multilocus sequence typing (MLST) scheme for seven pathogenic *Leptospira* species. **PLoS neglected tropical diseases**, v. 7, n. 1, p. e1954, jan. 2013.

ESHGHI, A. et al. Pathogenic *Leptospira interrogans* Exoproteins Are Primarily Involved in Heterotrophic Processes. [s.d.].

FOUTS, D. E. et al. What Makes a Bacterial Species Pathogenic?:Comparative Genomic Analysis of the Genus *Leptospira*. **PLOS Neglected Tropical Diseases**, v. 10, n. 2, p. e0004403, 18 fev. 2016.

GOMES, M. J. . Gênero *Leptospira* spp. **FAVET - UFRGS**, p. 1 – 51, 2013.

HYATT, D. et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. **BMC bioinformatics**, v. 11, p. 119, jan. 2010.

IRAOLA, G. et al. Transcriptome Sequencing Reveals Wide Expression Reprogramming of Basal and Unknown Genes in *Leptospira biflexa* Biofilms. [s.d.].

LAGESEN, K. et al. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. **Nucleic acids research**, v. 35, n. 9, p. 3100–8, jan. 2007.

LASLETT, D. ARAGORN, a program to detect tRNA genes and tmRNA genes in nucleotide sequences. **Nucleic Acids Research**, v. 32, n. 1, p. 11–16, jan. 2004.

LEHMANN, J. S. et al. Whole Genome Shotgun Sequencing Shows Selection on *Leptospira* Regulatory Proteins during in vitro Culture Attenuation. **Am. J. Trop. Med. Hyg**, v. 94, n. 2, p. 302–313, 2016.

LIN, S.-H.; LIAO, Y.-C. CISA: contig integrator for sequence assembly of bacterial genomes. **PloS one**, v. 8, n. 3, p. e60843, jan. 2013.

MORENO, L. Z. et al. Profiling of *Leptospira interrogans*, *L. santarosai*, *L. meyeri* and *L. borgpetersenii* by SE-AFLP, PFGE and susceptibility testing-a continuous attempt at species and serovar differentiation. **Emerging microbes & infections**, v. 5, p. e17, 2016.