

GENOME DRAFT DE UM NOVO ISOLADO *LEPTOSPIRA KIRSCHNERI*

FREDERICO SCHMITT KREMER¹; MARCUS REDÜ ESLABÃO²; CARLOS EDUARDO POUHEY DA CUNHA³; LUISA ZANOLLI MORENO⁴; ODIR ANTÔNIO DELLAGOSTIN⁵; LUCIANO DA SILVA PINTO⁶

¹ Universidade Federal de Pelotas - fred.s.kremer@gmail.com

² Universidade Federal de Pelotas - marcus.eslabao@yahoo.com.br

³ Universidade Federal de Pelotas - cpouey@gmail.com

⁴ Universidade de São Paulo - luzanolli@gmail.com

⁵ Universidade Federal de Pelotas - odirad@gmail.com

⁶ Universidade Federal de Pelotas – ls_pinto@hotmail.com

1. INTRODUÇÃO

A leptospirose é uma doença zoonótica emergente causada por bactérias espiroquetas do gênero *Leptospira*. Considerada a zoonose com maior distribuição em escala global, a grande maioria dos seus casos se concentra em regiões tropicais e subtropicais, aonde estima-se que acometa de 10-100 por 10.000 pessoas (GUERRA, 2013). Sua manifestação em humanos, apesar de muitas vezes ser assintomática, pode em alguns casos comprometer o funcionamento de diferentes órgãos, como pulmão, apresentando taxa de fatalidade de aproximadamente 40% (VIJAYACHARI *et al*, 2008). Além disso, na pecuária possui um grande impacto, sendo responsável por consideráveis perdas econômicas devido a complicações acarretadas na prenhez de diversos animais (PULIYATH *et al*, 2012).

A análise do genoma destas bactérias possibilita um maior entendimento dos fatores associados ao processo de patogênese e sua biologia (ADLER, 2014). A análise comparativa de genomas de variantes patogênicas e não patogênicas demonstrou que estes organismos apresentam um núcleo de genes mandatório para a sobrevivência e um conjunto de aproximadamente 900 genes presentes apenas nas bactérias patogênicas (ADLER *et al*, 2011). A análise de genes de patogenicidade pode servir de base para o desenvolvimento de novas abordagens profiláticas, como vacinas de subunidade recombinante (DELLAGOSTIN *et al*, 2011)

Atualmente, 7 genomas de isolados de *Leptospira* spp., montados e anotados, estão disponíveis no Genbank, e aproximadamente 250 estão disponibilizados na forma de *contigs* ou *scaffolds*, ainda não finalizados (<http://www.ncbi.nlm.nih.gov/genbank/>). Entretanto, este número, apesar de alto, ainda não representa toda a variedade genética encontrada neste gênero, que com base em ensaios sorológicos pode ser classificado em 300 sorovares (LEVETT, 2001). Desta forma, novos isolados devem ser sequenciados para a análise de suas divergências e regiões conservadas.

O presente trabalho teve por objetivo o sequenciamento, montagem e anotação do genoma de um novo isolado de *Leptospira* spp, assim como sua classificação taxonômica através de tipagem molecular *in silico*.

2. METODOLOGIA

O isolado analisado, denominado 61H, foi obtido a partir de amostras de sangue de um paciente residente na cidade de Pelotas com quadro clínico compatível com Leptospirose aguda, sendo a coleta do material feita com base no Termo de Consentimento Livre e Esclarecido (TCLE). As leptospirosas foram cultivadas em meio semi-sólido suplementado com 10% *Leptospira Enrichment EMJH (Difco)* e 200 µg/mL de 5-fluoruracil e 5% de soro fetal bovino. Os cultivos foram mantidos em estufa B.O.D a 30 °C sem agitação. A preparação do DNA foi realizada com o kit *illustra bacteria genomicPrep Mini Spin (GE Healthcare)* e este foi encaminhado para o sequenciamento por plataforma *Illumina Solexa* com tecnologia *paired-end* pela empresa *Macrogen* (<http://www.macrogen.com/>).

A montagem do genoma foi realizada através das ferramentas *Velvet* (ZERBINO *et al*, 2008), *Ray* (BOISVERT *et al*, 2010) e *CLC Genomics Workbench* (<http://www.clcbio.com>) a partir dos pares de sequências que apresentavam qualidade média mínima de Phred 20 para ambas as leituras. *Velvet* e *Ray* foram executados com tamanho de k-mer igual a 41 e tamanho mínimo de *contigs* de 800 e 500, respectivamente. O *CLC* foi executado com tamanho mínimo de *contig* igual a 500. Após a montagem, as *contigs* geradas pelos diferentes montadores foram integradas pelo software *CISA* (LIN *et al*, 2013) e ordenadas com base nos cromossomos I e II do genoma de *Leptospira interrogans* sorovar L1-130 (Genbank: AE016823 e AE016824) com a ferramenta *Mauve* (RISSMAN *et al*, 2009). O arquivo de alinhamento gerado foi pré-processado por scripts escritos em linguagem *Python* para geração de uma *scaffold* em formato FASTA aonde as *contigs* foram delimitadas por regiões com tamanho arbitrário de 1000 'N's. A ferramenta *IMAGE* (TSAI *et al*, 2010) foi utilizada para a extensão das *contigs* sendo rodada com 20 passos de iteração para cada um dos cromossomos. Para cada cromossomo as *contigs* foram concatenadas em dois arquivos diferentes: um contendo a concatenação intercala por regiões de 1000 Ns e um contendo a concatenação sem Ns. A anotação do genoma foi realizada com a ferramenta *Square* (<http://sourceforge.net/projects/sqgenome/>), para identificação dos genes codificantes de proteínas e tRNAs, e *RNAmmer* (LAGESEN *et al*, 2007) para busca por sequências de RNAs ribossômicos. A representação gráfica do genoma foi gerada pela ferramenta *DNAPlotter* (CARVER *et al*, 2009).

A caracterização molecular do isolado foi feita com base na análise no *Multi Locus Sequence Typing* (<http://leptospira.mlst.net/>) a partir de busca de cada locus a partir da ferramenta *BLASTn* do pacote *NCBI-BLAST+* (CAMACHO *et al*, 2009) e bancos de dados gerados a partir de sequências obtidas do repositório *MLST* de *Leptospira spp*, sendo o perfil encontrado comparado com os já previamente descritos.

3. RESULTADOS E DISCUSSÃO

O sequenciamento com a plataforma *Illumina* resultou em 960.626 pares de leituras de 249 nucleotídeos divididas em 2 arquivos FastQ e codificação Q33, sendo 848.321 pares mantidos após a filtragem com por Phred 20. O resultado do processo de montagem pelos montadores *CLC*, *Velvet* e *Ray* e da unificação das *contigs* pelo *CISA* está apresentado na tabela 1. O ordenamento das *contigs* usando o *Mauve* resultou em uma montagem de 3.738.643 pares de base para o cromossomo I e 335.634 para o cromossomo II, contendo

155 e 20 *gaps*, respectivamente. O número de *gaps* foi reduzido para 113 no cromossomo 1 e 18 no cromossomo 2.

Tabela 1. Comparação dos resultados obtidos pelos montadores CLC, Velvet e Ray e pelo integrador e montagens CISA.

Software	Nº de Contigs	Menor Contig	Maior Contig	N50	Tamanho
CLC	422	500	100646	69	4344851
Velvet	1139	800	3668	429	1369013
Ray	898	810	29647	191	389346
CISA	406	114	100999	73	4132512

A anotação realizada pelo *Square* resultou em 3398 genes codificantes para proteínas identificados, sendo 3104 presentes no cromossomo 1 e 294 no cromossomos 2, e 39 sequências de RNAs transportadores. A busca por RNAs ribossômicos resultou em na identificação de duas regiões, sendo uma codificantes para a subunidade 16s e outra para a subunidade 5s. A representação dos genomas foi gerada pelo software *DNAPlotter* e está apresentada na figura 2.

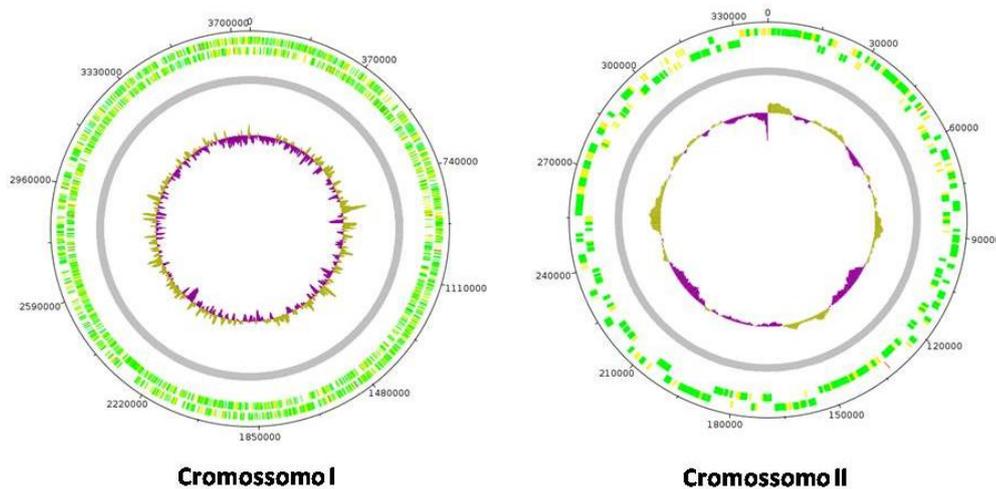


Figura 1. Representação gráfica dos cromossomos 1 e 2 gerado pelo *DNAPlotter* a partir dos arquivos de anotação gerados pelo *Square*.

A tipagem molecular *in silico* com as sequências de *MLST* demonstrou que o isolado é pertencente a espécie *Leptospira kirschneri* sorovar *Mozdok*. A comparação dos dados do presente genoma e dos genomas já indexados no Genbank está apresentada na tabela 2. Uma pequena redução no tamanho do genoma foi observada, assim como uma alteração no conteúdo C+G em relação às demais cepas sequenciadas. Este fato, associada a um menor número de sequências de RNAs ribossômico torna necessário uma revisão do processo de montagem, apesar de um número próximo de proteínas preditas e sequências de tRNAs ser observado no isolado Tsaratsovo.

Tabela 2. Comparação dos dados do genoma do isolado 61H com o dos genomas de isolados de *Leptospira kirschneri* sorovar *Mozdok* atualmente indexados no Genbank.

Cepa	Tamanho (Mb)	C+G%	CDSs	tRNAs	rRNAs
61H	4.07	34.6%	3398	39	2
Vehlefans 2	4.41	35.9%	3576	37	6
Vehlefans 3	4.40	35.9%	3566	37	5
Brem 166	4.35	35.9%	3552	36	5
Tsaratsovo	4.25	35.9%	3394	37	3

4. CONCLUSÕES

A partir da metodologia empregada, foi possível realizar a montagem e anotação do genoma do isolado 61H de *Leptospira* spp, assim como a sua tipagem molecular e classificação com base no MLST como uma *L. kirschneri* sorovar *Mozdok*. Visando um refinamento do resultado, etapas adicionais, como sequenciamento direcionado e/ou novos passos de montagem e pós-montagem, poderiam ser empregadas para a resolução de regiões de *gaps* no genoma. Ao fim do processo, o presente genoma será submetido ao Genbank para sua livre utilização.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- ADLER, B. *Pathogenesis of leptospirosis: Cellular and molecular aspects*. **Veterinary Microbiology**, V. 172, p. 252-258, 2014.
- ADLER, B; LO, M; SEEMANN, T; MURRAY, G.L. *Pathogenesis of leptospirosis: the influence of Genomics*. **Veterinary Microbiology**, V. 153, p. 73-81, 2011.
- BOISVERT, S; LAVIOLETTE, F; CORBEIL, J. *Ray: simultaneous assembly of reads from a mix of high-throughput sequencing technologies*. **Journal of Computational Biology**, V. 17, N. 11, 2010.
- CAMACHO, C; COULOURIS, G; AVAGYAN, V; MA, N; PAPADOPOULOS, J; BEALER, K; MADDEN, T.L. *BLAST+: architecture and applications*. **BMC Bioinformatics**, V. 10, N. 421, 2009.
- CARVER, T; THOMPSON, N; BLEASBY, A; BERRIMAN, M; PARKHILL, J. *DNAPlotter: circular and linear interactive genome visualization*. **Bioinformatics**, V. 25, N. 1, p. 119-120, 2009.
- DELLAGOSTIN, O.A; GRASSMANN, A.A; HARTWIG, D.D; FÉLIX, S.R; SILVA, E.F MCBRIDE, A.J.A. *Recombinant vaccines against leptospirosis*. **Human Vaccines**, V. 7, N. 11, p. 1215-1224, 2011.
- GUERRA, M.A. *Leptospirosis: Public health perspectives*. **Biologicals**, V. 41, p. 295-297, 2013.
- LAGESEN, K; HALLIN, P; RODLAND, E. A; STAERFELDT, H; ROGNES, T; USSERY, D. *RNAmmer: consistent and rapid annotation of ribosomal RNA genes*. **Nucleic Acids Research**, V. 35, N. 9, p. 3100-3108, 2007.
- LEVETT, P.N. *Leptospirosis*. **Clinical Microbiology Reviews**, V. 14, N. 2, p. 296-326, 2001.
- LIN, S; LIAO, Y. *CISA: Contig Integrator for Sequence Assembly of Bacterial Genomes*. **PLOS ONE**, V. 8, N. 3, 2013.
- PULIYATH G; SINGH, S. *Leptospirosis in pregnancy*. **European Journal of Clinical Microbiology & Infectious Diseases**, V. 31, N. 10, p. 2491-2496, 2012.
- RISSMAN, A.I; MAU, B; BIEHL, B.S; DARLING, A.E; GLASNER, J.D; PERNA, N.T. *Reordering contigs of draft genomes using the Mauve Aligner*. **Bioinformatics**, V. 25, N. 16, p. 2071-2073, 2009.
- TSAI, I. J; OTTO, T.D; BERRIMAN, M. *Improving draft assemblies by iterative mapping and assembly of short reads to eliminate gaps*. **Genome Biology**, N. 11, 2010.
- VIJAVACHARI, P; SUGUNAN, A.P; SHRIRAM, A.N. *Leptospirosis: an emerging global public health problem*. **Journal of Biosciences**, V. 33, N. 4, p. 557-569, 2008.
- ZERBINO, D.R; BIRNEY, E. *Velvet: algorithms for de novo short read assembly using de Bruijn graphs*. **Genome Research**, V. 18, N. 5, p. 821-829, 2008.