

GENOMA RASCUNHO DE UM NOVO ISOLADO DE *L. INTERROGANS* ISOLADO DE AMOSTRAS DE CANINOS EM SÃO PAULO (SP)

RAFAEL CAGLIARI¹; FREDERICO SCHIMITT KREMER²; CHRISTIAN
DOMINGUES SANCHEZ³; SÉRGIO JORGE⁴;
LUCIANO DA SILVA PINTO⁵

¹ BioPro-Lab - Biotecnologia - Universidade Federal de Pelotas – rafael.cagliari22@gmail.com

² BioPro-Lab - Biotecnologia - Universidade Federal de Pelotas

³ BioPro-Lab - Biotecnologia - Universidade Federal de Pelotas

⁴ Laboratório de Vacinologia - Biotecnologia – Universidade Federal de Pelotas

⁵ BioPro-Lab - Biotecnologia - Universidade Federal de Pelotas

1. INTRODUÇÃO

A leptospirose é uma zoonose negligenciada de maior ocorrência em zonas de clima tropical, afetando especialmente países em desenvolvimento, por conta de fatores que auxiliam sua proliferação, como enchentes e condições precárias de saneamento. A leptospirose é causada por espiroquetas do gênero *Leptospira*, capazes de sobreviver tanto no ambiente quanto em mamíferos, nos túbulos proximais renais. O rato é seu principal hospedeiro e reservatório, podendo carregar altas concentrações da bactéria em seu organismo sem apresentar sintomas.

Humanos são considerados hospedeiros incidentais por desenvolverem sintomas nocivos e fatais (em até 10% dos casos) mediante infecção, como: febre, náusea, enxaqueca e até mesmo hemorragia pulmonar e disfunção renal em estágios mais severos da doença. Esta bactéria efetua sua entrada no organismo humano através de lesões (geralmente micro) na pele ou através das mucosas do olho, nariz e garganta. Devido aos seus sintomas primários relativamente comuns, o diagnóstico inicial da doença é falho, sendo o desenvolvimento de métodos eficientes de diagnóstico rápido e vacinas de extrema importância (EVANGELISTA et al., 2010).

Filogeneticamente, as leptospiras são divididas em 19 espécies, das quais 13 são patogênicas, expressando cerca de 900 genes relacionados à sua virulência. Atualmente, são conhecidos mais de 250 sorovares e 24 sorogrupo distintos dentre estas espécies, classificando-as sorologicamente em função dos lipopolissacarídeos (LPS) expostos em sua superfície. O sequenciamento do genoma de novas cepas de *Leptospira* é fundamental para o melhor entendimento da doença e seu agente etiológico, revelando mecanismos moleculares de infecção, defesa e de maneira geral novas estratégias e abordagens para o desenvolvimento de vacinas, testes diagnósticos e tratamentos para a leptospirose (PALANAPPIAN et al., 2007).

Para este trabalho realizou-se o sequenciamento, montagem de novo e anotação genoma de uma nova cepa da espécie patogênica *Leptospira interrogans* denominada DU114, previamente caracterizada como pertencente ao sorogrupo canicola e isolada a partir da urina de amostras caninas na cidade de São Paulo (SP). A caracterização genética e da distribuição sorológica de diferentes espécies da bactéria por região geográfica e vetor são essenciais para

o melhor entendimento da doença, além de auxiliar a traçar estratégias visando combate-la.

2. METODOLOGIA

A cepa DU114 teve seu genoma sequenciado utilizando a plataforma de sequenciamento *Ion Torrent personal genome machine* (PGM), chip 316 e bibliotecas do tipo single-end, gerando leituras de tamanho 400 pb em formato .BAM, convertidas para FASTQ através do utilitário bamTofastq do pacote bedtools. A qualidade das leituras foi então averiguada através do FASTX-Toolkit, filtrando sequências com Phred score => 20 para ao menos 80% das bases.

A montagem *de novo* foi realizada por meio de três montadores: SPAdes (BANKEVICH et al., 2012), MIRA e Newbler (KISAND et al., 2013), sendo as contigs geradas por este último utilizadas para o resto do processo devido à sua maior qualidade. As contigs resultantes foram alinhadas contra os cromossomos de uma cepa de *L. Interrogans* referência, a 56609 (genbank: CP006723.1, CP006724.1). Por fim alguns *gaps* remanescentes foram fechados por meio das ferramentas FGap e GMCLOSER. O genoma rascunho foi submetido à plataforma de anotação do NCBI e então à *pipeline* de anotação Genix (KREMER et al., 2016), que reúne as ferramentas Prodigal, RNAmmer, tRNAscan, Aragorn e INFERNAL.

3. RESULTADOS E DISCUSSÃO

As contigs geradas pelas plataformas MIRA e Newbler apresentaram resultados de montagem superiores aos do assembler SPAdes, MIRA gerou um contigs de maior N50, enquanto o Newbler montou um número menor de contigs, indicando menor incidência de *gaps*. O tamanho total do genoma gerado pela plataforma Newbler, com aproximadamente 4,64 mega pares de base foi o mais próximo da sequência da cepa de *Leptospira interrogans* utilizada posteriormente como referência, com 4,3 megabases. Além disto, MIRA foi o único assembler a incorporar nucleotídeos desconhecidos ("N") em sua montagem. O conteúdo CG, entretanto, foi similar para os três softwares, conforme consta na tabela 1.

Tabela 1. Comparação entre os dados obtidos para a montagem *de novo* do isolado DU114 através de três distintas plataformas de assembly.

Software	Tamanho total (pb)	Número de contigs	Conteúdo CG (%)	N50	Nº de "N" por 100 kpb
MIRA	5376950	721	34.91	46894	4.14
Newbler	4657258	244	34.93	37970	0.00
SPAdes	4790497	544	34.95	15919	0.00

Na figura 1 está apresentada a distribuição do tamanho das contigs com o número total de contigs geradas. Apesar de apresentar tamanho ligeiramente maior, a montagem gerada pelo MIRA se estende por quase o triplo de contigs em relação ao estabelecido pelo Newbler, indicando alta fragmentação deste.

Dados obtidos a partir da *pipeline* de anotação de genomas bacterianos Genix permitiram a identificação de regiões funcionais no genoma da *L.*

interrogans cepa DU114, incluindo 3792 sequencias codificantes (CDS), 37 sítios de origem para RNA's transportadores (tRNA), 2 sequências para RNA ribossomal (rRNA), 1 sítio para RNA transfer-messenger (tmRNA) e 27 para sequências de RNA não-codificante diversas, responsáveis em sua maioria pela regulação dos níveis de expressão proteica. Dados ilustrados na tabela 2.

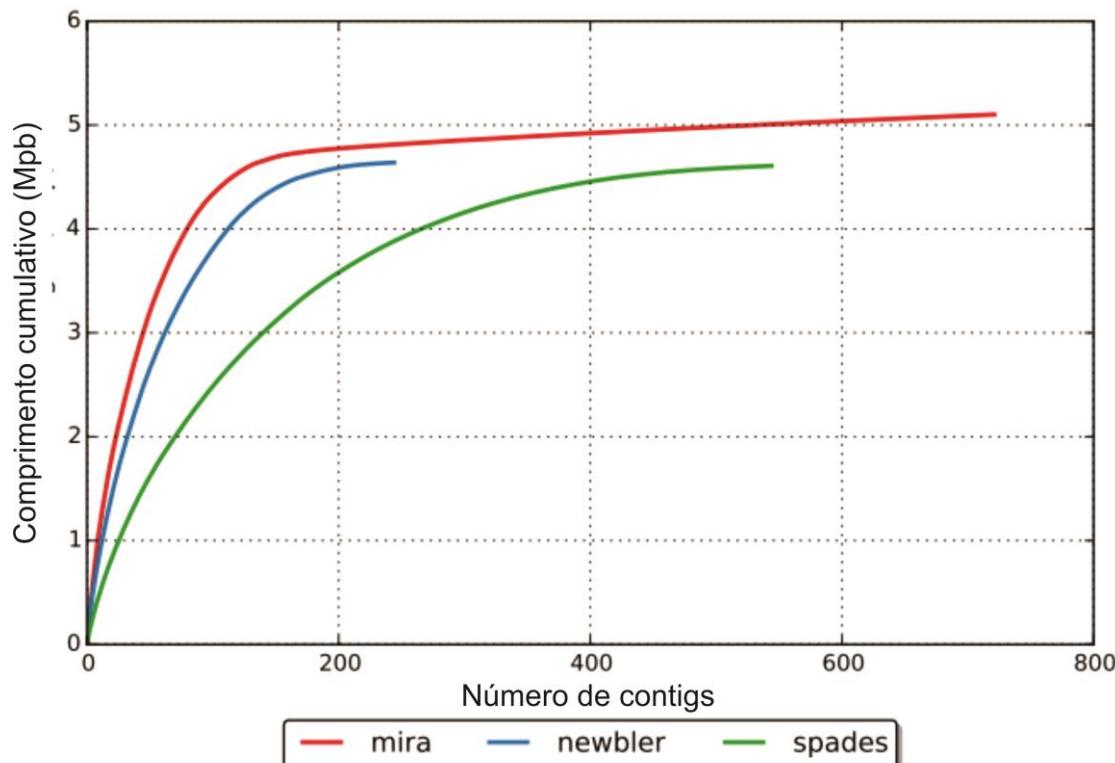


Figura 1. Análise comparativa entre a extensão cumulativa e número de *contigs* geradas durante as montagens. Foi utilizado o programa QUAST para a geração do gráfico.

Tabela 2. Regiões funcionais distintas e seu número de ocorrência de acordo com dados obtidos através de anotação do genoma da *Leptospira interrogans* cepa DU114.

Região Funcional	Número de Ocorrências
CDS	3972
Trna	37
ncRNA	27
rRNA	2
tmRNA	1

4. CONCLUSÕES

As leituras resultantes do sequenciamento da cepa DU114 de *Leptospira interrogans* foram ordenadas com sucesso em contigs pelo assembler Newbler, passando a seguir por anotação a fim de identificar possíveis regiões funcionais de interesse, como genes e RNA's diversos. Apesar de este ser um sólido genoma rascunho, a realização de mais passos é necessária para assegurar a qualidade das sequências. Perspectivas futuras incluem a correção manual de

genes quebrados em decorrência da inserção, deleção ou substituição de bases não correspondentes à sequência durante o processo de leitura. Uma versão prévia, sem correção, foi submetida ao GenBank para ser utilizada por outros pesquisadores da área (GenBank: CP022883, CP022884).

5. REFERÊNCIAS BIBLIOGRÁFICAS

- BANKEVICH, Anton et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. **Journal of computational biology**, v. 19, n. 5, p. 455-477, 2012.
- EVANGELISTA, Karen V.; COBURN, Jenifer. Leptospira as an emerging pathogen: a review of its biology, pathogenesis and host immune responses. **Future microbiology**, v. 5, n. 9, p. 1413-1425, 2010.
- KISAND, Veljo; LETTIERI, Teresa. Genome sequencing of bacteria: sequencing, de novo assembly and rapid analysis using open source tools. **BMC genomics**, v. 14, n. 1, p. 211, 2013.
- KREMER, Frederico Schmitt et al. Genix: a new online automated pipeline for bacterial genome annotation. **Microbiology Letters**, v. 363, n. 23, p. fnw263, 2016.
- PALANIAPPAN, Raghavan UM; RAMANUJAM, Subbupoongothai; CHANG, Yung-Fu. Leptospirosis: pathogenesis, immunity, and diagnosis. **Current opinion in infectious diseases**, v. 20, n. 3, p. 284-292, 2007.