

GENOMA RASCUNHO DE ISOLADO *LEPTOSPIRA INTERROGANS* CEPA PISCINA

CHRISTIAN DOMINGUES SANCHEZ¹; FREDERICO SCHMITT KREMER²;
SÉRGIO JORGE³; VINICIUS FARIAS CAMPOS⁴; ODIR ANTÔNIO
DELLAGOSTIN⁵; LUCIANO DA SILVA PINTO⁶

¹ Biotecnologia - Universidade Federal de Pelotas - cdsanchez@inf.ufpel.edu.br

² Biotecnologia - Universidade Federal de Pelotas – fred.s.kremer@gmail.com

³ Biotecnologia - Universidade Federal de Pelotas – sergiojorgevet@hotmail.com

⁴ Biotecnologia - Universidade Federal de Pelotas - fariascampos@gmail.com

⁵ Biotecnologia - Universidade Federal de Pelotas - odirad@gmail.com

⁶ Biotecnologia - Universidade Federal de Pelotas – ls_pinto@hotmail.com

1. INTRODUÇÃO

Com distribuição em escala global, a leptospirose é uma zoonose causada por bactérias espiroquetas do gênero *Leptospira*, grande parte de seus casos ocorre em regiões tropicais e subtropicais aonde estima-se que acometa de 10-100 de cada 10.000 pessoas (GUERRA, 2013)

A análise comparativa de genomas de cepas patogênicas e não patogênicas (saprofitas) de *Leptospira* spp. demonstrou que estes organismos apresentam genoma núcleo mandatório para a sobrevivência e um conjunto de aproximadamente 900 genes presentes apenas nas bactérias patogênicas (ADLER et al., 2011). A análise genômica de patogenicidade pode ajudar no desenvolvimento de novas abordagens profiláticas, como desenvolvimento de vacinas de subunidade recombinante (DELLAGOSTIN et al., 2011)

Atualmente, 8 genomas finalizados de isolados de *Leptospira interrogans*, montados e anotados, estão disponíveis no GenBank, e aproximadamente 284 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. Desta forma, novos isolados devem ser sequenciados para a análise de suas divergências e regiões conservadas.

No artigo que descreve o isolado “piscina”, na parte de virulência, foi identificado que o LD⁵⁰ (dose necessária para matar 50% dos indivíduos em uma população) é de 2 leptospiros, isso significa que é uma bactéria bastante letal, o que justifica o sequenciamento do seu genoma, pois a partir disso pode-se realizar novas análises para entender melhor como ela desencadeia a doença e abre espaço para o desenvolvimento de novas medidas de controle (FORSTER et al., 2013)

O presente trabalho teve por objetivo o sequenciamento, montagem e anotação do genoma de um novo isolado de *Leptospira interrogans*.

2. METODOLOGIA

O isolado denominado piscina (LPool), foi obtido a partir da amostra de água coletada de uma piscina abandonada, na cidade de Pelotas, RS, sul do Brasil (FORSTER et al., 2013). O isolado foi cultivado em 1 ml Ellinghausen-McCullough-Johnson-Harris (EMJH) meio líquido suplementado com *Leptospira* Enrichment EMJH (Difco, BD Diagnostics, Sparks, MD, EUA) e as culturas foram incubadas a 30 °C. Após sete dias, a cultura foi centrifugada a 5000 x g durante 5 min, o sobrenadante foi passado através de um filtro de 0,22 µm (Millipore,

Billerica, MA, EUA) e o filtrado foi utilizado para inocular um outro tubo de meio líquido EMJH. A preparação do DNA foi realizada com o kit *illustra* bacteria genomicPrep Mini Spin (GE Healthcare).

O sequenciamento foi realizado com a plataforma IonTorrent PGM, usando biblioteca "*fragment*" (*single-end*) de 400 bp. Foram utilizadas as ferramentas "BamToFastq" do pacote Bedtools para converter os arquivos BAU do IonTorrent Server em formato FASTQ, e as leituras foram filtradas usando o pacote Fastx-Toolkit, mantendo as com Phred score ≥ 20 (Q20) para pelo menos 80% das bases.

A montagem *de novo* foi realizada com a ferramenta Newbler (www.rocke.com/), mira (<http://www.chevreux.org/>) e SPAdes (BANKEVICH et al., 2012), e depois os resultados destes foram integrados com o programa CISA (LIN; LIAO, 2013). A anotação foi realizada com a *pipeline* de anotação Genix (labbioinfo.ufpel.edu.br/cgi-bin/genix_index.py), que integra os programas Prodigal, BLAST, RNAmmer, tRNAscan, Aragorn e INFERNAL.

Foi gerado um gráfico pelo programa QUAST (quast.bioinf.spbau.ru) para demonstrar a similaridade dos programas que se encontra na figura 1.

3. RESULTADOS E DISCUSSÃO

O sequenciamento com a plataforma IonTorrent resultou em 143,636,823 bases sequenciadas, sendo 117,775,157 pares mantidos após a filtragem com *Phred* 20, também foram obtidos 689,451 *reads* (fragmentos curtos de sequência). O resultado do processo de montagem pelas ferramentas MIRA, SPAdes e Newbler e da unificação das *contigs* pelo CISA está apresentado na tabela 1.

Tabela 1. Comparação dos resultados obtidos pelos montadores SPAdes, MIRA e Newbler e pelo integrador de montagens CISA.

| Software | Tamanho total (pb) | Número de Contigs | Conteúdo CG (%) | N50 |
|----------|--------------------|-------------------|-----------------|-------|
| CISA | 4.546.645 | 84 | 34,99 | 79705 |
| SPAdes | 4.693.155 | 749 | 35,21 | 10267 |
| Mira | 4.382.917 | 277 | 34,99 | 46313 |
| Newbler | 4.555.710 | 176 | 34,98 | 44889 |

Como demonstrado na tabela 1, a comparação dos resultados obtidos pelos montadores com os obtidos pelo integrador de montagens CISA demonstra uma diminuição expressiva em relação ao número de *contigs*. O montador SPAdes, por exemplo, gerou 749 *contigs*, enquanto que no processo final com o CISA foi possível chegar ao número de 84 *contigs*, reforçando assim a grande importância de utilizar abordagens de integração de montagem. A montagem gerada por cada ferramenta de programa tem sua importância e mesmo o SPAdes tendo gerado um número de *contigs* muito elevado, não significa que os dados gerados por ele são de baixa qualidade. Na Figura 1 está apresentada a relação entre número de *contigs* e o tamanho da montagem, onde o tamanho total da montagem é mantido e a diminuição do número total de *contigs* no processo final com o CISA é bastante expressiva, o que é de grande relevância, pois o genoma está menos fragmentado e com um tamanho muito próximo dos outros montadores.

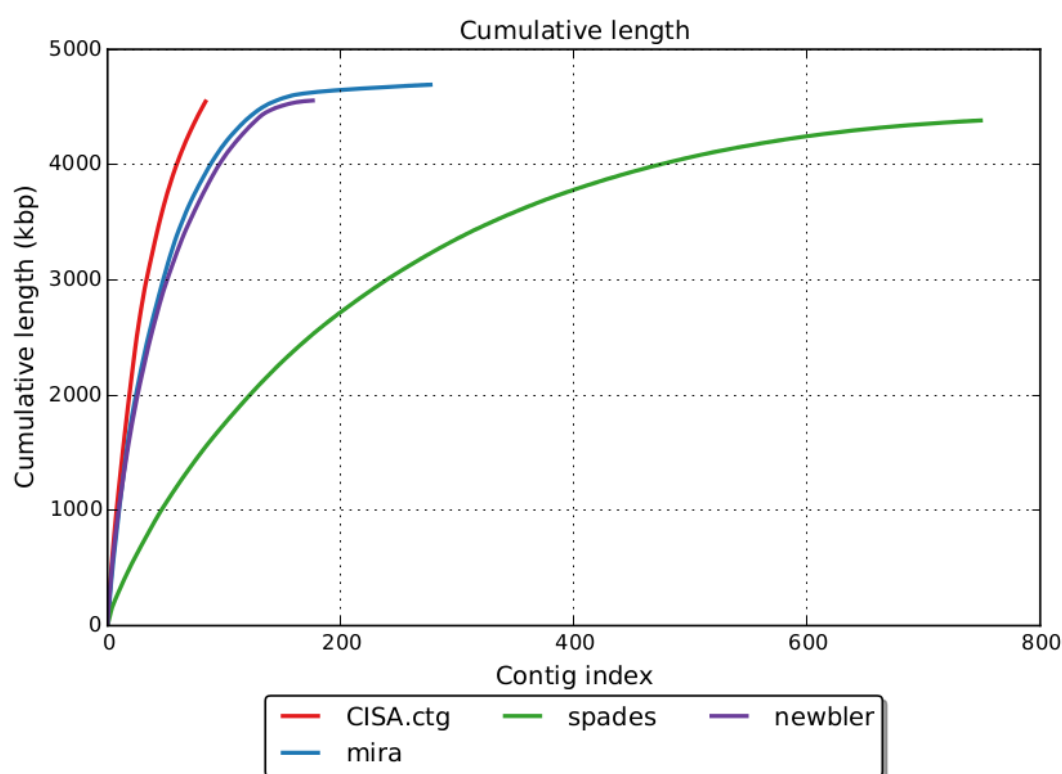


Figura 1. Representação da análise comparativa gerado pelo programa QUAST do *cumulative length* (tamanho cumulativo) dos diferentes montadores usados.

Tabela 2. Número de ocorrências das diferentes regiões funcionais encontradas no genoma de *Leptospira interrogans* cepa piscina,

| Nome | Número de ocorrências |
|-------------------------------------|-----------------------|
| Riboswitches | 5 |
| CRISPR | 7 |
| tRNA | 37 |
| tmRNA | 1 |
| rRNA | 2 |
| CDS | 4234 |
| Gene | 4275 |
| ncRNA (Rnase P classe A (Ribozima)) | 1 |

A partir da anotação, foi possível identificar várias regiões funcionais no genoma da cepa piscina. Foram identificados 4234 genes codificantes de proteínas (CDS), dos quais 5 são regulados por riboswitches, os Riboswitches são pequenas regiões localizadas normalmente na porção 5' dos mRNA e que servem como elementos regulatórios de expressão gênica. Podem ser ativados / inibidos por metabólitos (ex: cobalamina). Além das regiões codificantes, também foram identificadas features não codificantes, como tRNAs, rRNAs, tmRNAs, ncRNAs e CRISPRs. CRISPRs são regiões no genoma da bactéria que apresentam

sequências repetitivas de origem viral. Estas sequências são usadas como mecanismo de defesa contra fagos. As Ribozimas são enzimas não proteicas, e a RNA P é um pequeno RNA não codificante que auxilia na atividade de algumas RNases, atuando na regulação gênica também. Os tmRNAs são RNAs transportadores que também codificam para um pequeno peptídeo, que atua na regulação da síntese de algumas proteínas. O valor comum para uma leptospira em rRNAs é maior ou igual a 4 e encontramos apenas 2, o que nos diz que esse genoma está fragmentado e está faltando ao menos 2 rRNAs, também foi possível identificar um ncRNA que codifica para Rnase P classe A (Ribozima), como podemos observar na Tabela 2, essa foi uma anotação mais completa, o que auxilia a entender melhor diversos processos metabólicos e moleculares do organismo de interesse. Entretanto um menor número de sequências de RNAs ribossômicos do que o esperado torna necessário uma revisão do processo de montagem e possivelmente um re-sequenciamento.

4. CONCLUSÕES

Com a metodologia empregada, foi possível realizar a montagem e anotação do genoma do isolado piscina de *Leptospira interrogans*. 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. O genoma já foi enviado para o GenBank (<http://www.ncbi.nlm.nih.gov/nuccore/MACO000000000>) e está na versão 1.0, pode ser encontrado pelo *locus* MACO000000000 e é de livre utilização.

5. REFERÊNCIAS BIBLIOGRÁFICAS

- ADLER, B. et al. Pathogenesis of leptospirosis: The influence of genomics. **Veterinary Microbiology**, v. 153, n. 1-2, p. 73–81, 2011.
- 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.
- DELLAGOSTIN, O. A. et al. Recombinant vaccines against leptospirosis. **Human vaccines**, v. 7, n. 11, p. 1215–24, 2011.
- FORSTER, K. M. et al. Characterization of a virulent *Leptospira interrogans* strain isolate from an abandoned swimming pool. **Braz J Microbiol**, v. 44(1), p. 165–170, 2013.
- GUERRA, M. A. Leptospirosis: Public health perspectives. **Biologicals**, v. 41, n. 5, p. 295–297, 2013.
- LIN, S.-H.; LIAO, Y.-C. CISA: contig integrator for sequence assembly of bacterial genomes. **PloS one**, v. 8, n. 3, p. e60843, 2013.