

ANÁLISE GENÔMICA E FUNCIONAL DE UM NOVO ISOLADO DE *LYSINIBACILLUS SPHAERICUS* OBTIDO NO ESTADO DE PERNAMBUCO, BRASIL

LUDMILA DUDA VICENTE FERREIRA; TÚLIO DE LIMA CAMPOS; JOÃO PITTA;
CLÁUDIA MARIA FONTES DE OLIVEIRA; TATIANY PATRÍCIA ROMÃO

RESUMO

Avanços na tecnologia de sequenciamento de DNA e a ascensão da bioinformática permitiram com que genomas de diferentes espécies pudessem ser estudados e catalogados, com o auxílio de bancos de dados biológicos. Apesar dos avanços, ainda existem desafios em relação ao estudo de novos genomas e dos já existentes no que se refere a sua caracterização funcional e biotecnológica, sobretudo para organismos procariotos. *Lysinibacillus sphaericus* é uma espécie de bactéria gram-positiva e de importância para o controle biológico de insetos, visto que ela produz toxinas de atividade larvívora, além de possuir outras aplicações biotecnológicas. Diante da importância de uma caracterização genômica mais apropriada de novos isolados, mas também de cepas de referência, o objetivo deste trabalho é caracterizar geneticamente e funcionalmente um novo isolado de Ls obtido no estado de Pernambuco, enfatizando-se sua função entomopatogênica e comparando-a com a cepa de referência 2362. Amostras do isolado nomeado I20A39 tiveram seu DNA extraído e sequenciado, permitindo a montagem e a anotação dos seus genomas por bioinformática. A caracterização genética se baseou na construção de gráficos de grupos de genes ortólogos e mineração de genes de importância biotecnológica, além da predição estrutural através de *docking* molecular de proteínas com potencial atividade tóxica. As análises genômicas revelaram que o isolado I20A39 possui as toxinas clássicas de ação entomopatogênica típicas de Ls. Além disso, o isolado destacou-se por apresentar cópias de genes relacionados a proteínas processadoras de metais pesados, podendo ser utilizadas para ações de biorremediação de ambientes poluídos por metais. Os resultados mostram o potencial desta cepa em comparação à referência para aplicações biotecnológicas diversas, esperamos que estes possam ser utilizados nas estratégias de controle populacional de insetos e para ações de biorremediação.

Palavras-chave: *Bacillus*; *Docking* molecular; Genomas; Predição estrutural; Sequenciamento;

1 INTRODUÇÃO

Lysinibacillus sphaericus (Ls) é uma espécie de bactéria gram-positiva, cosmopolita, aeróbia e esporulante (LACEY, 2007). Anteriormente conhecida como *Bacillus sphaericus*, essa bactéria possui destaque como agente larvívora devido a ação de toxinas produzidas durante a fase de esporulação da bactéria contra mosquitos dos gêneros *Culex* e *Anopheles* (AHMED *et al.*, 2007, PEÑA-MONTENEGRO; DUSSÁN, 2013). O principal fator larvívora das cepas comerciais é a toxina binária (bin), composta por duas subunidades que agem de forma sinérgica em células do epitélio intestinal dos mosquitos, provocando efeitos citopáticos tais como rupturas de membranas citoplasmática, vacuolização intensa e alterações mitocondriais (SILVA-FILHA *et al.*, 2014).

A toxina binária é um heterodímero de natureza proteica composto pelas subunidades BinA (42 kDa) e BinB (51 kDa) e transcrita por um óperon altamente conservado em diferentes cepas (BAUMANN; BAUMANN, 1989, PARK *et al.*, 2010). A subunidade BinB, de 448 aminoácidos, é responsável pela interação da toxina com receptores presentes no epitélio intestinal de *Culex* por meio da sua região N-terminal (ELANGO VAN *et al.*, 2000, ROMÃO *et al.*, 2011). Por outro lado, a subunidade BinA, de 370 aminoácidos, possui a função de provocar os efeitos citopáticos nas células intestinais, interagindo com a subunidade BinB pela sua região C-terminal (LIMPANAWAT *et al.*, 2009).

Os genomas das cepas de Ls com ação entomopatogena também possuem outras características comuns como por exemplo a presença de genes codificantes para toxinas de fase vegetativa, como as toxinas mosquitocidas (Mtx), sphaericolisinas e proteínas S-Layer (LACEY, 2007, BERRY, 2012, ALLIEVI *et al.*, 2014). De modo geral, os genomas de Ls disponíveis nos bancos de dados possuem aproximadamente 4.67 Mb, conteúdo GC de 37,3% e um pouco mais que 4.500 genes (HERNÁNDEZ-SANTANA *et al.*, 2016). Em relação ao número de genomas conhecidos, a plataforma do *National Center for Biotechnology Information* (NCBI), por exemplo, possui ao menos 50 genomas de diferentes isolados de Ls catalogados, no entanto, a grande maioria não foi avaliada em relação ao seu potencial biotecnológico (LEE *et al.*, 2017, GENG *et al.*, 2021).

Ls é uma espécie interessante no que se refere ao seu potencial biotecnológico, visto que seus diferentes isolados apresentam grande heterogeneidade funcional (LOZANO; DUSSÁN, 2013). Dentre as suas diferentes aplicações destaca-se a sua atividade biomineralizadora para formação de concreto, biofertilizante em solos degradados e biorremediadora em ambientes poluídos por metais pesados (CRICHTON, 2012, MARTÍNEZ; DUSSÁN, 2018). Na era do sequenciamento genômico e da ascensão da bioinformática tornou-se mais comum a quantidade de trabalhos que buscam analisar as características gerais de novas cepas, de tal forma a entender a evolução destes grupos e potenciais aplicações não antes vistas (REY *et al.*, 2016, GÓMEZ-GARZÓN *et al.*, 2017).

A análise genômica do novo isolado I20A39 se faz necessária para a exploração de novas aplicações que podem ser exclusivas deste novo genoma. Neste contexto o objetivo deste projeto é caracterizar e analisar o genoma de um novo isolados de Ls obtido no estado de Pernambuco, Brasil. Além da análise genômica também objetivou-se analisar proteínas de potencial biotecnológico codificadas neste genoma como uma forma de prever a sua funcionalidade *in vivo*, sobretudo para o controle biológico de mosquitos.

2 MATERIAIS E MÉTODOS

2.1 REATIVAÇÃO DOS CULTIVOS

O isolado I20A39 foi obtido em amostras de valetas na Região Metropolitana do Recife, armazenado a -80 °C e reativado em meio Luria-Bertani líquido. A amostra foi acondicionada em agitador orbital nas condições de 200 rpm, temperatura de 30 °C e durante 48 h. Após o referido período o isolado foi levado a centrífuga a fim de se obter os sedimentos esporulados que foram utilizados para a extração de DNA.

2.2 EXTRAÇÃO DE DNA E SEQUENCIAMENTO

O material genético do isolado foi extraído a partir de uma alíquota de 60 µl com células esporuladas pelo kit Dneasy Blood and Tissue Kit (Quiagen®), seguindo recomendações do fabricante. A amostra foi quantificada em Qubit dsDNA BR Assay Kit (ThermoFisher®) e suas bibliotecas genômicas foram preparadas pelo Nextera XT DNA Library Prep Kit (Illumina). O

sequenciamento genômico dos isolados foi realizado pelo sequenciador MiSeq Sequencer (Illumina). O genoma sequenciado foi filtrado pelo Trimmomatic 0.38 e montado pelo PathoSystems Resource Integration Center (PATRIC) pela ferramenta Unicycler. A anotação dos genomas foi realizada pelo software Prokka.

2.3 CARACTERIZAÇÃO GENÉTICA DOS ISOLADOS

A partir dos dados de sequenciamento e anotação genômica foi realizada a prospecção de proteínas de importância biotecnológica, fatores de entomopatogenicidade clássicos e fatores adicionais contidos nos genomas por meio do Blast Local Alignment Search Tool (BLAST) pelo algoritmo Needleman-Wunsch. Os resultados foram descritos de forma comparativa com o genoma de referência Ls 2362, disponível DDBJ/EMBL/GenBank sob número de acesso CP015224.

2.4 PREDIÇÃO ESTRUTURAL E *DOCKING* MOLECULAR

As proteínas detectadas no genomas do novos isolado que apresentaram potencial atividade entomopatogênica foram levadas ao Alphafold 2 complementado pelo ColabFold 1.5.2 para a predição estrutural. O Alphafold é um sistema de inteligência artificial para a obtenção de estruturas tridimensionais de proteínas, podendo ser utilizado para predição funcional. As proteínas preditas estruturalmente também foram avaliadas em relação a sua capacidade de formar heterodímeros, por meio do servidor ClusPro 2.0. Os modelos gerados foram manipulados e visualizados no PyMol com o auxílio do PDBsum para a identificação da quantidade e tipos de interação intermolecular.

3 RESULTADOS E DISCUSSÃO

3.1 CARACTERÍSTICAS GERAIS DOS GENOMAS E GRUPOS DE GENES ORTÓLOGOS

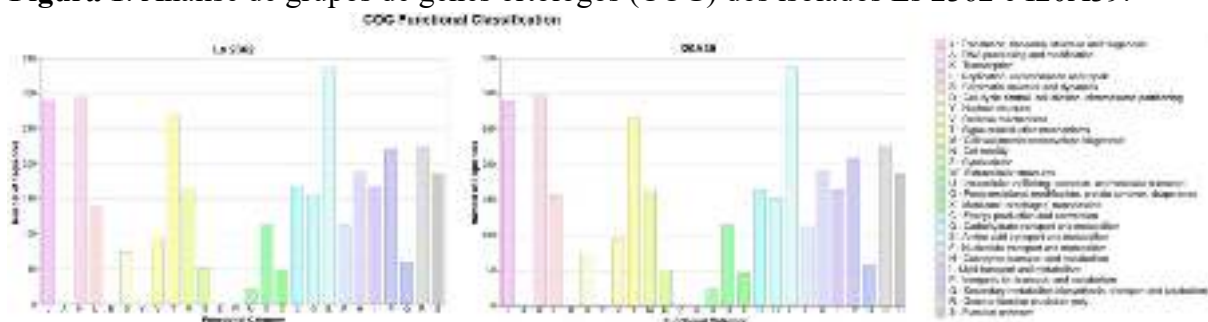
A análise do genoma total da cepa I20A39 mostra que o isolado apresenta aproximadamente 4,7 Mb que se distribuem em 70 sequências de RNA transportador, 1 sequências de RNA ribossomal e um pouco mais de 4.600 sequências codificantes (Tabela 1).

Tabela 1. Características gerais dos genomas dos novos isolados de *Lysinibacillus sphaericus*, obtidos em Pernambuco, Brasil.

<i>Lysinibacillus sphaericus</i>	2362	I20A39
Contigs	1	78
Bases	4692801	4728418
RNA _t	107	71
RNA _r	25	1
CDS	4504	4621

A análise de Grupos de Genes Ortólogos (*Cluster of Orthologous Genes* - COG) utilizada para determinar as categorias funcionais de genes destes genomas está representada na figura 1. Nela, fica evidente que as maiores famílias de genes estão distribuídas nas funções de metabolismo de aminoácidos e transcrição. Aproximadamente 5% dos genes destes genomas ainda não possuem função conhecida pela plataforma COG-NCBI enquanto 6,7% dos genes (225 sequências) participam de funções gerais com potencial biotecnológico.

Figura 1. Análise de grupos de genes ortólogos (COG) dos isolados Ls 2362 e I20A39.



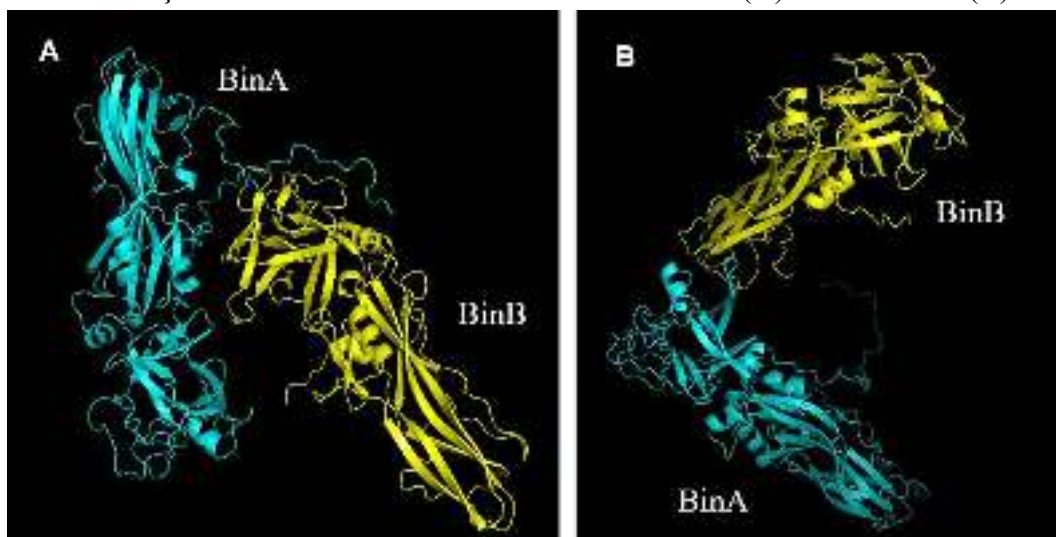
3.2 PROTEÍNAS DE POTENCIAL BIOTECNOLÓGICO

Das proteínas de potencial biotecnológico detectadas pela análise de COG destacamos inicialmente aquelas de função entomopatogênica que foram encontradas no genoma do isolado I20A39. O genoma do Ls I20A39 mostrou a presença de todos os fatores de entomopatogenicidade clássicos para a cepa de referência 2362, incluindo as subunidades BinA e BinB da toxina binária (BAUMANN; BAUMANN, 1989). Neste mesmo isolado também foram detectadas metaloproteínas processadoras de metais pesados como proteínas NikD, CbiO, MntC e TelA que metabolizam, respectivamente, níquel, cobalto, manganês e telúrio, (GÓMEZ-GARZÓN *et al.*, 2017). O isolado I20A39 apresentou ao menos 4 cópias de genes NikD, mostrando seu potencial biotecnológico para a biorremediação de ambientes poluídos por metais pesados.

3.1 PREDIÇÃO ESTRUTURAL DE TOXINAS E DOCKING MOLECULAR

Os modelos de interação molecular entre as subunidades da toxina binária detectadas no genoma do isolado I20A39 estão ilustradas na figura 2. No modelo hidrofóbico a energia livre de ativação para formação do heterodímero foi de $-936,6$ J enquanto no modelo balanceado a energia foi de $-876,3$ J. Visto que quanto menor a energia de ativação, maiores são as chances de interação intermolecular percebe-se que o melhor modelo foi o hidrofóbico (KESSEL; BENTALL, 2018).

Figura 2. Predição estrutural das toxinas BinA e BinB e *docking* molecular indicando os modelos de interação entre as toxinas em modelos hidrofóbico (A) e balanceado (B).



4 CONCLUSÃO

Lysinibacillus sphaericus, além de ser uma espécie importante para o controle biológico de culicídeos também apresenta uma plasticidade funcional relevante para a biotecnologia. A obtenção de novos isolados e a análise genômica destes foi necessária para o entendimento da sua evolução e aplicabilidade. O isolado I20A39 possui no seu genoma uma maior quantidade de sequências codificantes quando comparadas ao genoma de referência. O isolado I20A39, além das proteínas de ação entomopatogênica, também se destacou pela presença de genes relacionados a metaloproteínas, potencializando sua atuação em ações de biorremediação. Espera-se que os resultados possam ser utilizados para formulação de estratégias biotecnológicas destes isolados em diferentes contextos, a depender da necessidade de seu uso.

REFERÊNCIAS

AHMED *et al.* Proposal of *Lysinibacillus boronitolerans* gen. nov. sp. nov., and transfer of *Bacillus fusiformis* to *Lysinibacillus fusiformis* comb. nov. and *Bacillus sphaericus* to *Lysinibacillus sphaericus* comb. nov. **International Journal of Systematic and Evolutionary Microbiology**, v. 57, n. 5, p. 1117-1125. 2007.

ALLIEVI *et al.* Contribution of S-layer proteins to the mosquitocidal activity of *Lysinibacillus sphaericus*. **PLoS One**, v. 9, n. 10, p. e111114. 2014.

BAUMANN, L. I. N. D. A.; BAUMANN, P. A. U. L. Expression in *Bacillus subtilis* of the 51- and 42-kilodalton mosquitocidal toxin genes of *Bacillus sphaericus*. *Applied and environmental microbiology*, v. 55, n. 1, p. 252-253. 1989.

BERRY, C. The bacterium, *Lysinibacillus sphaericus*, as an insect pathogen. **Journal of invertebrate pathology**, v. 109, n. 1, p. 1-10. 2012.

CRICHTON, R. R. Biomineralisation. **Biological inorganic chemistry**, p. 359-378, 2012.

ELANGO VAN, G., *et al.* Identification of the functional site in the mosquito larvicidal binary toxin of *Bacillus sphaericus* 1593M by site-directed mutagenesis. **Biochemical and biophysical research communications**, v. 276, n. 3, p. 1048-1055. 2000.

GENG, P. *et al.* GenBank Reference Sequence: **CP071739.1**. 2021.

GOMEZ-GARZON, C. *et al.* A genome-scale metabolic reconstruction of *Lysinibacillus sphaericus* unveils unexploited biotechnological potentials. **PLoS One**, v. 12, n. 6, p. e0179666, 2017.

HERNÁNDEZ-SANTANA, A. *et al.* Complete genome sequence of *Lysinibacillus sphaericus* WHO reference strain 2362. **Genome announcements**, v. 4, n. 3, p. 10-1128. 2016.

KESSEL, A.; BEN-TAL, N. **Introduction to proteins: structure, function, and motion**. Crc Press, 2018.

LACEY, L. A. *Bacillus thuringiensis* serovariety *israelensis* and *Bacillus sphaericus* for

mosquito control. **Journal of the American Mosquito Control Association**, v. 23, p. 133-163. 2007.

LEE *et al.* GenBank Reference Sequence: **CP019980.1**. 2017.

LIMPANAWAT, S. *et al.* The C-Terminal Domain of BinA Is Responsible for *Bacillus sphaericus* Binary Toxin BinA–BinB Interaction. **Current microbiology**, v. 59, p. 509-513. 2009.

MARTÍNEZ, S. A.; DUSSÁN, J. *Lysinibacillus sphaericus* plant growth promoter bacteria and lead phytoremediation enhancer with *Canavalia ensiformis*. **Environmental Progress & Sustainable Energy**, v. 37, n. 1, p. 276-282, 2018.

PARK, H. W. *et al.* Properties and applied use of the mosquitocidal bacterium, *Bacillus sphaericus*. **Journal of Asia-Pacific Entomology**, v. 13, n. 3, p. 159-168. 2010.

PEÑA-MONTENEGRO, T. D.; DUSSÁN, J. Genome sequence and description of the heavy metal tolerant bacterium *Lysinibacillus sphaericus* strain OT4b. 31. **Standards in genomic sciences**, v. 9, n. 1, p. 42-56. 2013.

REY, A. *et al.* Complete genome sequencing and comparative genomic analysis of functionally diverse *Lysinibacillus sphaericus* III (3) 7. **Genomics data**, v. 9, p. 78-86, 2016.

ROMÃO, T. P. *et al.* The N-terminal third of the BinB subunit from the *Bacillus sphaericus* binary toxin is sufficient for its interaction with midgut receptors in *Culex quinquefasciatus*. **FEMS microbiology letters**, v. 321, n. 2, p. 167-174. 2011.

SILVA-FILHA, M. S. *et al.* Insects Midgut and Insecticidal Proteins. **Advances in Insect Physiology**. v. 1, p. 89-176. 2014.