Draft genome sequence of *Bacillus thuringiensis* strain 2193, an isolate toxic for lepidopteran

Projeto de sequência genómica de *Bacillus thuringiensis* strain 2193, um tóxico isolado para lepidopterano

DOI:10.34115/basrv6n3-013

Recebimento dos originais: 18/02/2022
Aceitação para publicação: 14/03/2022

Paulo Queiroz
Doutorado em Biologia Animal pela Universidade de Brasília (UNB)
Institution: Centro Universitário de Brasília (CEUB)
Address: 707/907, Campus Universitário, Asa Norte, Brasília, DF, CEP: 70790-075
E-mail: prqzqueiroz@gmail.com

Roberto Togawa
Doutorado pela University of Bedfordshire
Institution: Embrapa Recursos Genéticos e Biotecnologia
Address: Parque Estação Biológica, PqEB, Av. W5 Norte (final) Caixa Postal 02372, Brasília, DF, CEP: 70770-917
E-mail: roberto.togawa@embrapa.br

Priscila Grynberg
Doutorado em Bioinformática pela Universidade Federal de Minas Gerais (UFMG)
Institution: Embrapa Recursos Genéticos e Biotecnologia
Address: Parque Estação Biológica, PqEB, Av. W5 Norte (final) Caixa Postal 02372, Brasília, DF, CEP: 70770-917
E-mail: priscila.grynberg@embrapa.br

Érica Martins
Doutorado em Ciências Biológicas (Biologia Molecular) pela Universidade de Brasília
Institution: Universidade de Brasília
Address: SBN Quadra 02 Bloco C, Edifício Phenícia, CEP: 70040-020
E-mail: es.martins2@gmail.com

Rose Monnerat
Doutorado em Agronomia na Ecole Nacionale Agronomique de Montpellier
Institution: Embrapa Recursos Genéticos e Biotecnologia
Address: Parque Estação Biológica, PqEB, Av. W5 Norte (final) Caixa Postal 02372, Brasília, DF, CEP: 70770-917
E-mail: rose.monnerat@embrapa.br

ABSTRACT

Bacillus thuringiensis is a Gram-positive bacterium which produces insecticidal proteins effective to control several medical and agriculture insect pests. These proteins belong to the Cry, Cyt, Vip and Sip families. This work describes the draft genome sequence of B. thuringiensis S2193, which contains genes encoding the parasporal crystal cry1Ad, cry1Ca, cry1Da, cry1Fa, cry1Ia, cry2A, cry9Ea and vip3Ca. Genes coding for plant

growth promotion pathway were identified. And the operon which encodes the bacteriocin Thuricin 17 were annoted. We found the gene which encodes the toxin enhancin.

Keywords: bioinformatics, insecticidal toxins, enhancing.

RESUMO
Bacillus thuringiensis é uma bactéria Gram-positiva que produz proteínas insecticidas eficazes para controlar várias pragas de insectos médicos e agrícolas. Estas proteínas pertencem às famílias Cry, Cyt, Vip e Sip. Este trabalho descreve o projecto da sequência genómica de B. thuringiensis S2193, que contém genes que codificam o cristal parasporal cry1Ad, cry1Ca, cry1Da, cry1Fa, cry1Ia, cry2A, cry9Ea e vip3Ca. Foram identificados os genes que codificam o caminho de promoção do crescimento das plantas. E o ópero que codifica a bacteriocina Thuricin 17 foi anotado. Encontrámos o gene que codifica a toxina potenciadora.

Palavras-chave: bioinformática, toxinas inseticidas, melhoramento.

1 INTRODUCTION

Bacillus thuringiensis is a Gram-positive bacterium that occurs naturally in many environments including soil. The distribution of this bacterium is worlwide (1). *Bacillus thuringiensis* (Bt) produces insecticidal proteins classified into two families of membrane perforating toxins, crystalline (Cry) and cytolytic (Cyt) proteins. These Bt proteins were specific and environmentally sound insecticide, which were widely used to control several orders of insects in both agriculture and medicine (2-3).

Purified genomic and plasmid DNA`s from strain S2193 were extracted by Masterpure Gram-positive DNA purification kit (Epicentre) and QIAGEN Plasmid Maxi Kit (QiAGEN). Afterwards, these DNA`s were sequenced at Macrogen, Inc. (Seoul, Korea) using high-throughput HiSeq2000 and GS-FLX Plus platforms getting one lane of 100 bp and 1/8 region plate, respectively. The reads were assembled using SOAP de novo (version 1.05) and produced 447 contigs totaling 6,350,581 bp (Q20 = 99.33%), with a maximum scaffold size of 274,165 bp, an N_{50} length of 49,505 bp, and 34.70% G+C content and genome coverage depth was approximately 100x. De novo assembly was carried out using the Geneious version (8.0.4) (4). A total of 6,833 coding sequences (5) and 77 tRNA (6) and 8 rRNA operons (7), and 5 ncRNA were predicted. The chromosome was sized in approx. 5,356,319 bp and the plasmids sizes ranged from 585,993 bp to 2,061 bp. The functions of encoding genes were annotated by using the

NCBI nr, Swiss-Prot (8), Clusters of Orthologous Groups (COG) (9), KEGG (10), and InterProScan (11) databases.

The custom insecticidal toxin database was constructed with nucleotide and amino acid sequences of *cry*, *vip*, *cyt*, growth promotion, parasporin and bacteriocin genes using the complete sequences of other *B. thuringiensis* strains deposited in public databases (http://www.ncbi.nlm.nih.gov/genome/ or http://www. https://www.bpprc.org/). The strain S2193 draft genome sequence carries several insecticidal toxin genes showing identities to the *cry1Ad*, *cry1Ca*, *cry1Da*, *cry1Fa*, *cry1Ia*, *cry2A*, *cry9Ea* and *vip3Ca*. Using the new nomenclature of proteins toxic to insects proposed by Crickmore et al. (12) we found the *Spp1A*. The genes which encode the metabolic pathway of plant growth promotion 1-aminocyclopropane-1-carboxylate deaminase, indolepyruvate decarboxylase, putative acid phosphatase and siderophore biosynthesis protein were identified. The genomic analyses identified one operon which encodes antimicrobial peptide bacteriocin thuricin 17. As expected, the WGS procedure revealed more genes than reported previously by Macedo et al. (13) described this strain toxic against *Diatraea saccharalis* (Lepidoptera: Crambidae) and identified the genes *cry1Aa*, *cry1Ab*, *cry1E*, *cry1F*, *cry2Ab* by PCR method. Surprisingly, our NGS results don`t reveal the *cry1E* gene. This fact could be hypothesized by primers conditions. The bioinformatic procedure revealed one gene which codify the toxin enhancin.

Nucleotide sequence accession numbers. The sequence of the *B. thuringiensis* strain S2193 has been deposited in GenBank with the accession number SAMN12229590.

ACKNOWLEDGMENT

We would like to thank Dr. Lilian Praça for the technical assistance.

REFERENCES

1 Swamy HMM, Asokan R, Mahmood R. 2014. In silico Structural 3D Modelling of Novel Cry1Ib9 and Cry3A Toxins from Local Isolates of *Bacillus thuringiensis*. Indian J Microbiol. 54(1):94–103. doi: 10.1007/s12088-013-0364-5.

2 Pigott CR, Ellar DJ. 2007. Role of receptors in *Bacillus thuringiensis* crystal toxin activity. Microbiol. Mol. Biol. Rev. 71:255–281. doi: 10.1128/MMBR.00034-06.

3 Zhua, J, Zhanga, Q, Caoa, Y, Lia Q, Zhua Z, Wanga L, Lia P. 2016. The complete genome sequence of *Bacillus thuringiensis* serovar *Hailuosis* YWC2-8. Journal of Biotechnology. 219:38–39. doi: 10.1016/j.jbiotec.2015.12.016.

4 - Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, Buxton S, Cooper A, Markowitz S, Duran C, Thierer T, Ashton B, Meintjes P, Drummond A. (2012). Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 28:1647-1649. doi: 10.1093/bioinformatics/bts199.

5 - Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. 1999. Alignment of whole genomes. Nucleic Acids Res. 27:2369–2376. doi: 10.1093/nar/27.11.2369.

6 – Schattner P, Brooks AN, Lowe TM. 2005. The tRNAscan-SE, snoscan and snoGPS web servers for the detection of tRNAs and snoRNAs. Nucleic Acids Res. 33:W686-W689. doi: 10.1093/nar/gki366.

7 - Lagesen K, Hallin P, Rodland EA, Staerfeldt HH, Rognes T, Ussery DW. 2007. RNAmmer: consistent and rapid annotation of ribosomal RNA genes. Nucleic Acids Res. 35:3100–3108. doi: 10.1093/nar/gkm160.

8 - Boeckmann B, Bairoch A, Apweiler R, Blatter MC, Estreicher A, Gasteiger E, Martin MJ, Michoud K, O'Donovan C, Phan I, Pilbout S, Schneider M. 2003. The SWISS-PROT protein knowledgebase and its supplement TrEMBL in 2003. Nucleic Acids Res. 31:365–370. http:// dx.doi.org/10.1093/nar/gkg095.

9 - Tatusov RL, Natale DA, Garkavtsev IV, Tatusova TA, Shankavaram UT, Rao BS, Kiryutin B, Galperin MY, Fedorova ND, Koonin EV. 2001. The COG database: new developments in phylogenetic classification of proteins from complete genomes. Nucleic Acids Res. 29:22–28. http:// dx.doi.org/10.1093/nar/29.4.e22.

10 - Kanehisa M, Araki M, Goto S, Hattori M, Hirakawa M, Itoh M, Katayama T, Kawashima S, Okuda S, Tokimatsu T, Yamanishi Y. 2008. KEGG for linking genomes to life and the environment. Nucleic Acids Res. 36:D480–D484. http://dx.doi.org/10.1093/nar/gkm882.

11 - Quevillon E, Silventoinen V, Pillai S, Harte N, Mulder N, Apweiler R, Lopez R. 2005. InterProScan: protein domains identifier. Nucleic Acids Res. 33:W116 –W120. http://dx.doi.org/10.1093/nar/gni118.

12 – Crickmore N, Berry C, Panneerselvam S, Mishra R, Connor TR, Bonning BC. A structure-based nomenclature for *Bacillus thuringiensis* and other bacteria- derived pesticidal proteins. Journal of Invertebrate Pathology. https://doi.org/10.1016/j.jip.2020.107438.

13 - Macedo CL, Martins ES, Macedo LLP, Santos AC, Praça LB, de Góis LAB, Monnerat RG. 2012. Seleção e caracterização de estirpes de *Bacillus thuringiensis* eficientes contra a *Diatraea saccharalis* (Lepidoptera: Crambidae). Pesq. agropec. bras. 47(12):1759-1765. https://doi.org/10.1590/S0100-204X2012001200012.