

Draft genome sequence of *Bacillus thuringiensis* strain s1785, an isolate showing insecticidal activity

Sequência do genoma de *Bacillus thuringiensis* strain s1785, um isolado que mostra actividade insecticida

DOI:10.34115/basrv6n3-020

Recebimento dos originais: 18/02/2022

Aceitação para publicação: 14/03/2022

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

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

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

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

Rose Monnerat

Doutorado em Agronomia na Ecole Nationale 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* S1785, which contains genes encoding the parasporal crystal cry, like Cry4Aa, Cry4Ba, Cry10Aa, Cry11Aa, Cry60Aa and Cry60Ba and Cyt genes, Mpp60Aa

and Mpp60Ba. The gene Spp1Aa was identified in the S1785 genome. Genes coding for plant growth promotion pathway were identified.

Keywords: ngs, plant promotion growth, cyt toxin, diptera.

RESUMO

Bacillus thuringiensis é uma bactéria Gram-positiva que produz proteínas inseticidas eficazes para controlar várias pragas de insetos médicos e agrícolas. Estas proteínas pertencem às famílias Cry, Cyt, Vip e Sip. Este trabalho descreve o projecto da sequência genómica de *B. thuringiensis* S1785, que contém genes que codificam o grilo de cristal parasporal, como Cry4Aa, Cry4Ba, Cry10Aa, Cry11Aa, Cry60Aa e Cry60Ba e genes Cyt, Mpp60Aa e Mpp60Ba. O gene Spp1Aa foi identificado no genoma S1785. Foram identificados os genes que codificam a via de promoção do crescimento das plantas.

Palavras-chave: ngs, crescimento de promoção de plantas, toxina cyt, díptera.

1 INTRODUCTION

Bacillus thuringiensis strains can produce many insecticidal proteins like Cry proteins. The Cry toxins have specific insecticidal activity against many kinds of insects i.e. Coleoptera, Diptera and Lepidoptera, for instance (1). The insecticidal activity of *B. thuringiensis* subsp. continues to grow through the isolation and identification of new strains with a wide variety of target pests (2).

Purified genomic and plasmid DNA's from strain S1785 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 218 contigs totaling 6,477,425 bp (Q20 = 99.31%), with a maximum scaffold size of 359,449 bp, an N₅₀ length of 115,122 bp, and 34.22% G+C content and genome coverage depth wasa approximately 100x. De novo assembly was carried out using the Geneious version (8.0.4) (3). A total of 6,833 coding sequences (4) and 77 tRNA (5) and 8 rRNA operons (6), 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 (7), Clusters of Orthologous Groups (COG) (8), KEGG (9), and InterProScan (10) 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.btnomenclature.info>). The strain S1785 draft genome sequence carries several insecticidal toxin genes showing identities to the *Cry4Aa*, *Cry4Ba*, *Cry10Aa*, *Cry11Aa*, *Cry60Aa* and *Cry60Ba* and *Cyt* genes, *Mpp60Aa* and *Mpp60Ba*. Additionally, were identified the *Cyt* genes, *Cyt1A* and *Cyt2B*. We found the *Spp1Aa* gene in the genome of the S1785 genome. 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 could not identify any operon related to the antimicrobial peptide bacteriocin (thuricin or nisin).

Nucleotide sequence accession numbers. The sequence of the *B. thuringiensis* strain S1785 has been deposited in GenBank with the accession number SAMN12211138.

Acknowledgment

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

REFERENCES

- 1 He J, Wang J, Yin W, Shao X, Zheng H, Li M, Zhao Y, Sun M, Wang S, Yu Z. 2011. Complete Genome Sequence of *Bacillus thuringiensis* subsp. *chinensis* Strain CT-43. *Journal of Bacteriology*. 193:3407–3408. doi: 10.1128/JB.05085-11.
- 2 Porcar M, Navarro L, Jimenez-Peydro R, 2006. Pathogenicity of intra-thoracically administered *Bacillus thuringiensis* spores in *Blatta orientalis*. *J Invertebr Pathol*. 93:63–66. doi: 10.1016/j.jip.2006.05.001.
- 3 - 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.
- 4 - 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.
- 5 – 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.
- 6 - 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](http://dx.doi.org/10.1093/nar/gkg095). [http:// dx.doi.org/10.1093/nar/gkg095](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](http://dx.doi.org/10.1093/nar/29.4.e22). [http:// dx.doi.org/10.1093/nar/29.4.e22](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>.