Bioinformatics, 38(1), 2022, 250–251
doi: 10.1093/bioinformatics/btab506
Advance Access Publication Date: 9 July 2021
Applications Note

Genome analysis

BtToxin_Digger: a comprehensive and high-throughput pipeline for mining toxin protein genes from *Bacillus thuringiensis*

Hualin Liu [1], Jinshui Zheng [1,2,*], Dexin Bo[1,2], Yun Yu[1], Weixing Ye[1], Donghai Peng[1] and Ming Sun [1,*]

[1]State Key Laboratory of Agricultural Microbiology, Huazhong Agricultural University, Wuhan 430070, China and [2]Hubei Key Laboratory of Agricultural Bioinformatics, Huazhong Agricultural University, Wuhan 430070, China

*To whom correspondence should be addressed.

Associate Editor: Inanc Birol

Received on January 3, 2021; revised on June 18, 2021; editorial decision on July 2, 2021; accepted on July 7, 2021

Abstract

Summary: *Bacillus thuringiensis* (Bt) has been used as the most successful microbial pesticide for decades. Its toxin genes are used for the development of genetically modified crops against pests. We previously developed a web-based insecticidal gene mining tool BtToxin_scanner. It has been frequently used by many researchers worldwide. However, it can only handle the genome one by one online. To facilitate efficiently mining toxin genes from large-scale sequence data, we re-designed this tool with a new workflow and the novel bacterial pesticidal protein database. Here, we present BtToxin_Digger, a comprehensive and high-throughput Bt toxin mining tool. It can be used to predict Bt toxin genes from thousands of raw genome and metagenome data, and provides accurate results for downstream analysis and experiment testing. Moreover, it can also be used to mine other targeting genes from large-scale genome and metagenome data with the replacement of the database.

Availability and implementation: The BtToxin_Digger codes and web services are freely available at https://github.com/BMBGenomics/BtToxin_Digger and https://bcam.hzau.edu.cn/BtToxin_Digger, respectively.

Contact: jszheng@mail.hzau.edu.cn or m98sun@mail.hzau.edu.cn

Supplementary information: Supplementary data are available at *Bioinformatics* online.

1 Introduction

The toxins produced by *Bacillus thuringiensis* (Bt) have insecticidal activity against many agricultural and forestry pests. Bt can produce several kinds of insect-targeting toxins, such as insecticidal crystal protein (Cry), vegetative insecticidal protein (Vip), cytotoxic protein (Cyt), etc. The reported target insects of these toxins include those from Lepidoptera, Diptera, Coleoptera, etc. The *cry* and *vip* genes are among the most important ones used for the development of genetically modified (GM) crops targeting insect pests. From 1996 to 2016, the planting of Bt maize and cotton had delivered $50.6 billion and $54 billion of extra farm income, respectively (Brookes and Barfoot, 2018). To fight against the Bt toxin resistant insects and the new emerging pests, the discovery of new Bt strains and novel toxin genes is one of the most important strategies (Sanahuja *et al.*, 2011). Previously, we developed an on-line tool BtToxin_scanner (Ye *et al.*, 2012) to predict *cry* genes from Bt genome sequences, and it was frequently used by researchers including those who are interested in plant protection, GM crops development or sustainable agriculture (Adang *et al.*, 2014; Carroll *et al.*, 2020; Prado *et al.*, 2014; Reyaz *et al.*, 2019, 2021). It can handle one assembled genome each time and provides comparative results between the predicted toxin and the reported ones. Here, we re-designed the previous tool to provide a novel and high-throughput software BtToxin_Digger which can be used to handle large-scale genomic and metagenomic data to predict all kinds of putative toxin genes that match the recently updated toxin database (Crickmore *et al.*, 2020), as well as other virulence factors which contribute to the pathogenicity but not lethality of Bt against its target insects, such as Sip (Donovan *et al.*, 2006), Chitinase (Zhang *et al.*, 2014), InhA (Dalhammar and Steiner, 1984), Bmp1 (Luo *et al.*, 2013), Enhancin (Fang *et al.*, 2009) and ZwA (He *et al.*, 1994). It also generates comprehensive and readable results to facilitate the downstream sequence analysis or experiment design.

2 Materials and methods

The types of input data supported by BtToxin_Digger include raw Reads data (pair-end reads generated by different platforms of Illumina, long-reads from PacBio and ONT or hybrid-reads), genome or metagenome assemblies, coding sequences (CDSs) and

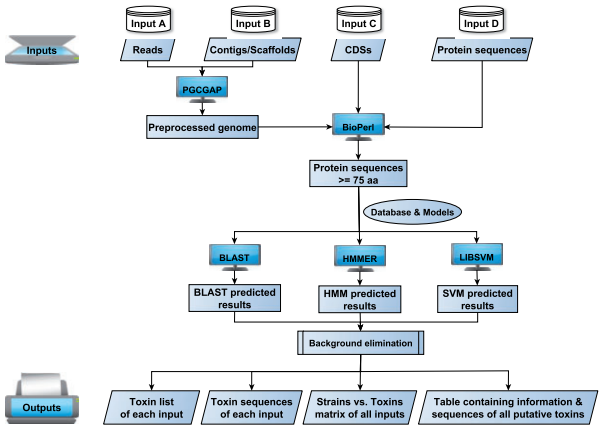

Fig. 1. A diagram of the BtToxin_Digger pipeline.

protein sequences. PGCGAP (Liu *et al.*, 2020) is used for genome assembly. ORFs finding and translation are performed by BioPerl (Stajich *et al.*, 2002). All protein sequences with a length above 115-aa are searched against the database and trained models by BLAST (Camacho *et al.*, 2009), HMMER (Eddy, 2011) and LIBSVM (Chang and Lin, 2011), respectively. The candidate proteins are blasted against a background database to filter out the false-positive records. Then several Perl scripts are used to parse the results to get the putative target protein genes (Fig. 1).

3 Results

BtToxin_Digger can be used online and easily installed on Linux, macOS and Windows Subsystem for Linux (WSL) platforms by the conda package manager (Grüning *et al.*, 2018) or docker container. We tested BtToxin_Digger on a laptop with an Intel CPU containing 8 threads of GHz-2.50 and 16 GB memory. It took about 14 min to process the 1.3-Gbp raw reads generated by Illumina Hiseq 2500 and less than 1 min for its assembled genome. In addition, BtToxin_Digger can be used to mine other interesting genes with the replacement of the toxin database by other target sequences.

Compared to the recent reported tool CryProcessor (Shikov *et al.*, 2020), BtToxin_Digger presents the following advantages, more flexible input file types, more comprehensive and accurate results, more readable outputs (Supplementary Table S1). We tested BtToxin_Digger and CryProcessor using the protein sequences of 601 *Bacillus thuringiensis* genomes retrieved from GenBank. Our tool identified 18 types of interesting genes, while CryProcessor just predicted one type (Supplementary Table S2). For Cry toxins, BtToxin_Digger output not only the 874 ones with 3-domain structure predicted by CryProcessor but also other 371 Crys with at least one domain.

Funding

This work was supported by the National Key R&D Program of China [2017YFD0201201] and National Natural Science Foundation of China [31670085, 31970003 and 31770003].

Conflict of Interest: none declared.

References

Adang,M.J. *et al.* (2014) Chapter two – diversity of *Bacillus thuringiensis* crystal toxins and mechanism of action. In: Dhadialla,T.S. and Gill,S.S. (eds.) *Advances in Insect Physiology, Insect Midgut and Insecticidal Proteins*, Vol. **47**. Academic Press, San Diego, Oxford, UK, pp. 39–87.

Camacho,C. *et al.* (2009) BLAST+: architecture and applications. *BMC Bioinformatics*, **10**, 421.

Carroll,L.M. *et al.* (2020) Proposal of a taxonomic nomenclature for the *Bacillus cereus* group which reconciles genomic definitions of bacterial species with clinical and industrial phenotypes. *mBio*, **11**, e00034–20.

Chang,C.-C. and Lin,C.-J. (2011) LIBSVM: a library for support vector machines. *ACM Trans. Intell. Syst. Technol.*, **2**, Article 27–27.

Crickmore,N. *et al.* (2020) A structure-based nomenclature for *Bacillus thuringiensis* and other bacteria-derived pesticidal proteins. *J. Invertebr. Pathol.*, 107438.

Dalhammar,G. and Steiner,H. (1984) Characterization of inhibitor A, a protease from *Bacillus thuringiensis* which degrades attacins and cecropins, two classes of antibacterial proteins in insects. *Eur. J. Biochem.*, **139**, 247–252.

Donovan,W.P. *et al.* (2006) Discovery and characterization of Sip1A: a novel secreted protein from *Bacillus thuringiensis* with activity against coleopteran larvae. *Appl. Microbiol. Biotechnol.*, **72**, 713–719.

Eddy,S.R. (2011) Accelerated Profile HMM Searches. *PLoS Comp. Biol.*, **7**, e1002195.

Fang,S. *et al.* (2009) *Bacillus thuringiensis* bel protein enhances the toxicity of Cry1Ac protein to *Helicoverpa armigera* larvae by degrading insect intestinal mucin. *Appl. Environ. Microbiol.*, **75**, 5237–5243.

Brookes,G. and Barfoot,P. (2018) *GM Crops: Global Socio-Economic and Environmental Impacts 1996–2016*. PG Economics Ltd., UK.

Grüning,B. *et al.*; Bioconda Team. (2018) Bioconda: sustainable and comprehensive software distribution for the life sciences. *Nat. Methods*, **15**, 475–476.

He,H. *et al.* (1994) Zwittermicin A, an antifungal and plant protection agent from *Bacillus cereus*. *Tetrahedron Lett.*, **35**, 2499–2502.

Liu,H. *et al.* (2020) Build a bioinformatics analysis platform and apply it to routine analysis of microbial genomics and comparative genomics. *Protoc. Exchange*, doi: 10.21203/rs.2.21224/v3.

Luo,X. *et al.* (2013) *Bacillus thuringiensis* metalloproteinase Bmp1 functions as a nematicidal virulence factor. *Appl. Environ. Microbiol.*, **79**, 460–468.

Prado,J.R. *et al.* (2014) Genetically engineered crops: from idea to product. *Annu. Rev. Plant Biol.*, **65**, 769–790.

Reyaz,A.L. *et al.* (2019) Genome sequencing of *Bacillus thuringiensis* isolate T414 toxic to pink bollworm (*Pectinophora gossypiella* Saunders) and its insecticidal genes. *Microb. Pathog.*, **134**, 103553.

Reyaz,A.L. *et al.* (2021) A novel *Bacillus thuringiensis* isolate toxic to cotton pink bollworm (*Pectinophora gossypiella* Saunders). *Microb. Pathog.*, **150**, 104671.

Sanahuja,G. *et al.* (2011) *Bacillus thuringiensis*: a century of research, development and commercial applications. *Plant Biotechnol. J.*, **9**, 283–300.

Shikov,A.E. *et al.* (2020) No more tears: mining sequencing data for novel *Bt* cry toxins with CryProcessor. *Toxins (Basel)*, **12**, 204.

Stajich,J.E. *et al.* (2002) The Bioperl toolkit: perl modules for the life sciences. *Genome Res.*, **12**, 1611–1618.

Ye,W. *et al.* (2012) Mining new crystal protein genes from *Bacillus thuringiensis* on the basis of mixed plasmid-enriched genome sequencing and a computational pipeline. *Appl. Environ. Microbiol.*, **78**, 4795–4801.

Zhang,L.L. *et al.* (2014) Biological activity of *Bacillus thuringiensis* (Bacillales: bacillaceae) Chitinase against *Caenorhabditis elegans* (Rhabditida: Rhabditidae). *J. Econ. Entomol.*, **107**, 551–558.