Whole-genome sequencing of *Bacillus thuringiensis* strain SY49.1 reveals the detection of novel candidate pesticidal and bioactive compounds isolated from Turkey

Semih Yılmaz^{1,2*}, Abeer Babiker Idris³, Abdurrahman Ayvaz⁴, Rıdvan Temizgül⁴,
Mohammed A. Hassan^{2,5,6,7}.

***Corresponding author:** ylmazsemh@yahoo.com

¹ Department of Agricultural Biotechnology, Faculty of Agriculture, Erciyes University, Kayseri, Turkey.

² Erciyes Teknopark, Promoseed biotechnology A.Ş., Kayseri, Turkey.

³ Department of Agricultural Science and Technology, Institute of Natural and Applied Sciences, Erciyes University, Kayseri, Turkey.

⁴ Department of Biology, Faculty of science, Erciyes University, Kayseri, Turkey.

⁵ Department of Bioinformatics, Africa City of Technology, Khartoum, Sudan

⁶ Department of Bioinformatics, DETAGEN Genetic Diagnostics Center, Kayseri, Turkey

⁷ Department of Translation Bioinformatics, Detavax Biotech, Kayseri, Turkey

Authors' E-mails: SY: ylmazsemh@yahoo.com; ABI: abeer.babiker89@gmail.com; AA: ayvaza@erciyes.edu.tr; MAH: altwoh20002002@yahoo.com.

Abstract

Background: *Bacillus thuringiensis* SY49.1 (*Bt* SY49.1) strain is a new strain isolated from a soil sample in Adana city which is nestled in the most fertile agricultural area in Turkey. This strain has insecticidal activity against insect pests from different orders. Also, it is characterized by its significant activity against plant fungal infections and as plant growth promotion.

Aims: To report the whole genome sequencing (WGS) and annotation of the *Bt* SY49.1 strain.

Materials and Methods: The *Bt* SY49.1 strain was isolated from the soil sample in Adana city by using a sodium acetate enriched medium. Bacterial DNA was extracted then sequenced using Illumina HiSeq technology. After data filtering and quality control, de novo assembly and genomic annotation were applied. Evolutionary and comparative genomic analysis and variant calling were performed using various in silico tools.

Results: The WGS of strain *Bt* SY49.1 is 6.32 Mbp long with a GC content of 34.68%. It reveals large numbers of single nucleotide polymorphisms (SNPs) and InDels. The assembled genome contains 6,562 protein-encoding genes. In addition, it encodes various gene inventories for the biosynthesis of bioactive compounds such as insecticidal delta endotoxins, growth stimulatory deaminase and phosphatases, fungicidal thuricin, bacillibactin, petrobactin, fengycin/plipastin, and zwittermycin A.

Conclusion: The strain *Bt* SY49.1 could have several potential utilities as a source of antibiotics compounds, plant growth promoting metabolites, and biocontrol agents for fungal phytopathogens, and insects. We expect that the draft genome of the *Bt* SY49.1 strain may provide a model for proper understanding and studying of antimicrobial compound mining, genetic diversity among the *B. cereus* group, and pathogenicity against insect pests and plant diseases, and growth-promoting activity.

Keywords: *Bacillus thuringiensis*, *Bt* SY49.1, whole-genome sequencing, annotation, pesticide, fungicide, PGPR

1. Introduction

Bacillus thuringiensis (*Bt*) is a ubiquitous rod-shaped, Gram-positive, spore-forming and aerobic or facultative anaerobic bacterium found usually in a variety of ecological niches including soil, grain dust, dead insect, and aquatic environments, among many others (1, 2). It is characterized by producing parasporal crystals, during its sporulation or stationary growth phase (3). These crystals contain protein toxins (e.g. δ -endotoxins or Cry proteins) which act as powerful intestinal toxins against various insect hosts (4, 5). The toxins are active against larvae of lepidopteran, dipteran, and/or coleopteran species and have been successfully introduced to genetically modified crops. Their lack of toxicity against mammals, high specificity, and environmental safety led to advancing in search for new strains with different spectra of pesticide activities (4, 6). In addition, *Bt* strains have been found to exhibit antibacterial, antibiofilm, antifungal, and emulsifying activities (7-9). They are known to be significant sources for antimicrobial compounds, such as bacteriocins and lipopeptides, and antifungal compounds, such as zwittermycin, lipopeptides, and chitinase (8, 10, 11).

Several strains of *B. thuringiensis* have raised a global interest for various bio-pest applications because of their specific pesticidal activities (12-16). However, the development of insect resistance to *B. thuringiensis* toxin proteins has been widely reported (17, 18). This leads to urgent searching for novel *B. thuringiensis* with novel insecticide toxins as an effective management strategy (19). In this study, we report the complete genome sequence of a novel *B. thuringiensis* SY49.1 strain which was isolated from a soil sample collected from Adana city in Turkey. This strain has insecticidal activity against insect pests from different orders; and was found to carry various types of insecticidal cry genes which include *cry1Aa/Ad*, *cry1B*, *cry1C*, *cry5*, *cry9A*, and *cry9C* (20-23). Also, it is characterized by its significant activity against plant fungal infections (24).

2. Materials and Methods

2.1 Bacterial isolation and identification

Bt SY49.1 strain was isolated from a soil sample collected from Adana city which is nestled in the most fertile agricultural area in the south of Turkey. The isolation of the *Bt* SY49.1 strain was done by using a sodium acetate enriched medium, as previously described (25). In brief, one gram of soil sample was inoculated in LB broth (pH 6.8±2) which includes 0.25M sodium acetate, then incubated in a shaking incubator at 200 rpm at 30°C for 4 hours. After that, 1.5 ml of the inoculated LB broth has been transferred to a sterile Eppendorf tube and exposed to high temperature (80°C) for 10 minutes to kill the vegetative bacterial form. Then, 20-50 ul from the Eppendorf tube were spread on LB agar plates and incubated overnight at 30°C. Finally, colonies with a morphological appearance that resembles *Bacillus* have been spread on agar plates to obtain pure colonies which were re-incubated in 5 ml LB broth (pH 6.8±2) in 50 ml tubes at 200 rpm at 30°C overnight. Further bacterial identifications were performed by colonial morphology on Potato Dextrose Agar (PDA), gram stain, and malachite green stain as previously described in (26), (27), and (28), respectively.

For para-sporal inclusion bodies characterization, the bacteria were inoculated in 150 ml of 3T medium and incubated at 30 °C for 7 days to induce sporulation (20). After that, the mixtures were centrifuged at 15.000 rpm and 4 °C for 10 min and suspended in dH₂O on microscope slides and fixed. Then, the slides were sputter-coated with 10 150 nm Au/Pd using an SC7620 Mini-sputter coater. Finally, they were viewed using a LEO440 scanning electron microscope at 20kV beam current (29).

2.2 Bacterial DNA extraction and sequencing

The bacterial DNA was extracted according to a previously described method of Jensen *et al.* and Porcar *et al.* with some modifications (30, 31). In brief, a loopful of cells from a

single colony, that has been cultured in LB medium overnight, was placed into 400 μ l sterile dH₂O. Then, to lyse the cells, the mixture was boiled for 10 min. After that, the resulting cell lysate was centrifuged for 10 sec at 10,000 rpm. Finally, the supernatant was used as DNA templates for PCR reactions.

The whole-genome sequencing (WGS) was commercially conducted using Illumina's HiSeq technology at BGI-HongKong Co., Ltd. company.

2.3 Genome assembly and genome annotation

A 760.2 Mbp paired-end reads with a 500-bp insert size library were obtained from the DNA sample. However, to ensure accuracy of the follow-up analysis 10-20% of the data was eliminated after filtering the raw data and removing reads with low-quality, reads with a certain proportion of Ns' bases, adapter contamination, and duplication reads. For K-mer analysis (32), regions in the reads with high-quality scores were intercepted and, then, cut into 15-bp fragments (15-mer). The depth of 15-mer was 16x and the genome coverage depth was 18.96x. In addition, by using SOAPaligner (version 2.21) software (33), all read sequences obtained by sequencing were initially aligned with the reference sequence (NC_017208) to calculate the average depth and coverage ratio in order to determine the differences between sequencing species and the reference sequence. The Coverage depth and coverage ratio of *Bt* SY49.1 strain to reference sequence (*Bt* chinensis-CT-43 serovar) were 48.79x and 80.29%, respectively.

De novo assembly to the short filtered reads was performed using SOAPdenovo (version 1.05) (34), which is an in-house assembler. De Bruijn graph was constructed, firstly, and erroneous connections were removed to build the contig and then scaffold. For filling gaps, reads are mapped to scaffold and a paired-end reads relationship was used. In addition, the assembly error was corrected by aligning the reads to the assembly result using SOAPaligner (version 2.21) tool and counting the mapping information of reads. Then the single base error of the assembly result was corrected based on mapping information. Also, GC content and depth correlative analysis, and K-

mer analysis were performed to evaluate whether heterogeneous sequences exist in the sample or not.

Genome annotation was done using the NCBI Prokaryotic Genome Annotation Pipeline (PGAP) and Rapid Prokaryotic Genome Annotation (Prokka), and gene functions were predicted using Rapid Annotation using Subsystems Technology (RAST) and the Pathosystems Resource Integration Center (PATRIC) (35-38). A circular map was generated using PATRIC sever (38).

In addition, CRISPR repeats were predicted by using the CRISPRfinder web server (39). AntiSMASH version 4.0 server was used to predict, annotate, and analyze secondary metabolite biosynthesis gene clusters in the bacterial genome (40). PHASTER web server is used for the identification and annotation of prophage sequences within *Bt* SY49.1 genome and plasmids (41). Also, IslandViewer 4 webserver was applied to predict and interactively visualize genomic islands (GI) in *Bt* SY49.1 genome (42).

2.4 Evolutionary and comparative genomic analyses of *Bt* SY49.1 strain

The species was established using *16S rRNA* and phylogenic approach. Sequences of *16S rRNA* genes of thirteen *Bacillus* strains and SY49.1 were selected for phylogenetic analysis. The selected species represent the members of the *B. cereus sensu lato* group. The sequence of the strain *B. subtilis* IAM 12118 was selected as an outgroup. The Kimura 2-parameter (K2+G) model from the substitution (ML) model was chosen with 1000 bootstrap replicates to construct a maximum likelihood (ML) phylogenetic tree using MEGA7 software (43). Also, phylogenetic analysis of the strain SY49.1 with the closest reference and representative genomes was conducted using PATRIC sever (38). The closest reference and representative genomes to SY49.1 were determined by Mash/MinHash (44). PATRIC global protein families (PGFams) were selected from these genomes to identify the phylogenetic placement of *Bt* SY49.1 genome (45). Moreover, the strain was established using the Genome-to-Genome Distance Calculator (GGDC) webserver (46).

Functional comparison (sequence-based) of *Bt* SY49.1 genome was performed with closely related *Bacillus* species, *B. thuringiensis* str. 97-27 (NC_005957), *B. thuringiensis* str. Al Hakam (NC_008600), *B. cereus* ATCC 10987 (NC_003909), *B. cereus* ATCC 14579 (NC_004722), *B. cereus* E33L (NC_006274), and *B. anthracis* str. Ames Ancestor (NC_007530). While *B. subtilis* subsp. *subtilis* str.186 (NC_000964) was used as an outgroup in the map. The comparison was done by using bidirectional and unidirectional best hit (protein Blast) implemented in the RAST server (35). Furthermore, variation analyses for SNP and InDel calling were conducted using MUMmer (version 3.22) package and LASTZ (version 1.01.50) (47, 48).

3. Results

3.1 Genome properties of strain *Bt* SY49.1

The complete genome sequence of the *Bt* SY49.1 strain is 6,322,230 bp long with a GC content of 34.68%, containing 270 scaffolds with N₅₀ of 79,855 bp. The *Bt* SY49.1 genome contains a chromosome and seven circular plasmids bigger than 15kb. The GC contents of the seven plasmids ranged from 31.6% to 36.9%. In total, 6,562 protein-encoding genes (PEGs), 3 *rRNAs*, 10 *tRNAs*, 5 *ncRNAs*, and 429 pseudogenes were annotated by NCBI-PGAP. Two sequences were predicted with CRISPR repeats and one sequence with Cas cluster. The *Bt* SY49.1 whole-genome sequencing project has been deposited at NCBI GenBank under the accession number NZ_JAHKEZ000000000, assembly number GCF_018791885.1 BioProject accession number PRJNA734785, and BioSample accession number SAMN19533886. The 6.32 Mbp draft genome map of *Bt* SY49.1 is presented in Figure 1.

3.2 Insights from the genome sequence

Bt SY49.1 was found to be flagellated, sporulating with a subterminal endospore and producing the insecticidal parasporal inclusions, see Figure 2. These phenotypes are supported by gene inventories observed in the genome of *Bt* SY49.1. The RAST annotation has sorted the 6,562 PEGs into 342 functional subsystem pathways using SEED subsystems. The most abundant of them are genes that are associated with amino acids and derivatives metabolism 392 (18.8%), followed by carbohydrates metabolisms 264 (12.67%), cofactors, pigment, and prosthetic groups 158 (7.59%), protein metabolism 156 (7.5%), and dormancy and sporulation 109 (9.23%), for more illustration, see Figure 3. In addition, the antiSMASH 4.0 server predicted that the *Bt* SY49.1 genome encodes gene clusters that are responsible for the biosynthesis of secondary metabolites which include lantipeptide, antimicrobial peptides, and siderophores. The *Bt* SY49.1 genome was found to carry gene clusters with high homology to the biosynthetic gene clusters of the thuricin, iron-siderophore (bacillibactin and petrobactin), fengycin/plipastin, and antifungal compound (zwitermycin A) (Figure 4). Moreover, the PHASTER server identified ten prophage regions, of which one region is intact (score > 90), seven regions are incomplete (score < 70), and two regions are questionable (score 70-90), see Figure 5.

3.3 Result of evolutionary and comparative genomic analyses

The phylogenetic tree of complete genome sequence of SY49.1 using PGFams and maximum likelihood tree of thirteen strains of *Bacillus* supports the placement of strain *Bt* SY49.1 within the *B. thuringiensis* group, see Figure 6. In addition, the genome of *Bt* SY49.1 is highly similar to *B. thuringiensis* serovar kurstaki strain HD 1 (CP010005) based on digital DNA: DNA Hybridization (DDH) estimate (GLM-based): 99.60%.

The Comparison of *Bt* SY49.1 genomes with the six closely related strains of *B. thuringiensis*, *B. cereus*, and *B. anthracis*, revealed strain-specific genes which encode hypothetical proteins, phage-like proteins, prophages, mobile genetic elements, and transposases. These findings were cross-validated and visualized in IslandViewer 4

server which showed these genes grouped into genomic islands, see Figure 7 and supplementary file 1: Table S1.

Regarding variants calling, alignment of the *Bt* SY49.1 scaffolds to the reference sequence (NC_017208) revealed 122,508 SNPs and 160 InDels (6,206 bp), which accounts for 1.94% and 2.58%, respectively. Mutation types caused by InDels are presented in Table 1, and InDel statistics of their different size are shown in Figure 8. Among the 122,508 SNPs, 104,06 (84.94%) were in CDS region, while 18, 45 (15.06%) were identified in intergenic region. The distribution of SNPs in the CDS region is illustrated in Table 1.

Table 1. Mutation types cause by InDels

Type	Frame shifted	Damaged start codon	Damaged stop codon	Premature stop	Other changes	CDS with InDel	Total CDS
No. (#)	48	5	26	37	44	160	6,206
Rate (%)	0.773	0.081	0.419	0.596	0.709	2.578	100

Frame shifted, InDels induce the open reading frameshift mutation; Damaged Start Codon, the start codon is damaged by InDels; Damaged Stop Codon, the stop codon is damaged by InDels; Premature Stop, frameshifted cause a premature stop codon to occur, shortening the actual length by InDels; Other Change, there is no significant influence for the open reading frame with InDels.

Table 2. Distribution of SNPs in CDS region

Type	Start_syn	Stop_syn	Start_nonsyn	Stop_nonsyn	Premature_stop	Synonymous	Nonsynonymous
No. (#)	5	115	63	36	124	78,207	25,557
Rate (%)	0.004	0.094	0.051	0.029	0.101	63.838	20.861

Start_Syn, start codon synonymous mutation; Stop_Syn, stop codon synonymous mutation; Start_nonsyn, start codon non-synonymous mutation; Stop_nonsyn, stop codon non-synonymous mutation; Premature_stop, a triplet codon mutate into a stop codon; Synonymous, synonymous in the coding region; Nonsynonymous, non-synonymous in the coding region.

4. Discussion

In this study, we provided a standard genome draft of strain *Bt* SY49.1 which may contribute to the evolution and comparative genomics studies among the *B. cereus sensu lato* group. The draft genome of the *Bt* SY49.1 strain is 6.32 Mbp with 34.68% GC content, similar to the genomes of other *B. thuringiensis* strains (15)(16). In addition, the comparative analysis of strain *Bt* SY49.1 with closely related *Bacillus* genomes, the phylogenetic tree of the 16S rRNA sequences, and detection of gene inventories for the insecticidal Cry proteins biosynthesis placed *Bt* SY49.1 under *B. thuringiensis*. However, the *Bt* SY49.1 strain was selected for complete genome sequencing, annotation, and analysis as it was determined to be one of the most promising strains for the discovery of compounds with strong antibacterial, antifungal, and iron acquisition abilities (Figure 4).

Natural compounds produced by microorganisms form the basis of many drugs such as antibiotics, antivirals, and others (49). However, the rapid emergence of drug-resistant to recent antimicrobials and quick evolution through genetic alteration are great threats to control microbial infection (50). Therefore, research strategies like applying naturally occurring antimicrobial peptides (AMPs), and structural modification of antimicrobials drugs are extensively studied (51, 52). In this study, the *Bt* SY49.1 genome was found to carry gene clusters including lantibiotics, antimicrobial peptides, and siderophores. The antifungal compounds (fengycin and zwittermycin A) were found in the *Bt* SY49.1 genome. Fengycin, like most natural AMPs, acts by increasing the plasma membrane permeability of the target cell and exhibits strong fungi toxic activity (53, 54). Also, the aminopolyol antifungal compound (zwittermycin A) was previously shown to suppress oomycete diseases in plants (55, 56). These findings indicate that the *Bt* SY49.1 has antifungal activity besides chitinase, as indicated previously (24). Therefore, further experimental studies are also required to investigate the effectiveness of these compounds (fengycin and zwittermycin A) in different fungal diseases of plants. In addition, the *Bt* SY49.1 genome carries gene clusters of siderophores, like petrobactin and bacillibactin, which suggests its iron acquisition abilities. However, these gene clusters are not exclusive to *B. thuringiensis* genomes but are also detected in the

genomes of other members of the *Bacillus cereus* group and other *Bacillus* species (57, 58).

Lantibiotics are gene-encoded peptide antibiotics that imply great potential to be attractive options for medical, agricultural, and industrial applications. Because of their special mode of action and they don't raise substantial resistance. Thus, the discovery or detection of lantibiotics is urgent and appealing (59). However, the *Bt* SY49.1 genome was found to carry gene clusters with complete homology to the biosynthetic gene cluster of the thuricin (class II lantibiotic), see Figure 4E. The antimicrobial assay of thuricin exhibited its activity against *Micrococcus flavus*, and the conserved disulfide bridge was significant to the function of this bovicin HJ50-like lantibiotic (60). Interestingly, the antiSMASH 4.0 server predicted a number of novel secondary metabolites mainly belonging to the non-ribosomal peptide synthetase (NRPS) cluster and unspecified ribosomally synthesized and post-translationally modified peptide product (RiPP) cluster. Therefore, future investigations on *Bt* SY49.1 genome-encoded bioactive metabolites may be pursued.

Conclusion

In this study, we report a 6,32 Mbp draft genome with 34.68% GC content of strain *Bt* SY49.1 isolated from the soil sample in Adana, Turkey. The assembled genome contains 6,562 protein-encoding genes, of which the most abundant are genes that are associated with amino acids and derivatives metabolism 392 (18.8%), followed by carbohydrates metabolisms 264(12.67%), and cofactors, pigment, and prosthetic groups 158(7.59%). The antibacterial, antifungal, and insecticide properties of this strain could be inferred, in part, with several gene inventories encoded in the draft genome. This indicated that strain *Bt* SY49.1 could have several potential utilities as a source of antibiotics compounds, and a fungal phytopathogen and insect biocontrol agent. We expect that the draft genome of the *Bt* SY49.1 strain may provide a model for proper understanding

and studying of antimicrobial compound mining, genetic diversity among the *Bacillus cereus* group, and pathogenicity against insect pests and plant diseases.

Data availability

The *Bt* SY49.1 whole-genome sequencing project has been deposited at NCBI GenBank under the accession number NZ_JAHKEZ000000000, BioProject accession number PRJNA734785, and BioSample accession number SAMN19533886.

Acknowledgments

This work was supported by Erciyes University Scientific Project Unit under the codes of ÖNAP-3638 and FBD-08-540. We also thank to GenKök research center for their collaboration.

References

1. Ibrahim M, Griko N, Junker M, Bulla L. *Bacillus thuringiensis*. Bioengineered bugs. 2010;1:31-50.
2. Jouzani GS, Valijanian E, Sharafi R. *Bacillus thuringiensis*: a successful insecticide with new environmental features and tidings. Applied Microbiology and Biotechnology. 2017;101(7):2691-711.
3. Palma L, Muñoz D, Berry C, Murillo J, Caballero P. *Bacillus thuringiensis* Toxins: An Overview of Their Biocidal Activity. 2014;6(12):3296-325.
4. Federici B. *Bacillus thuringiensis* in biological control. Handbook of biological control: Elsevier; 1999. p. 575-93.
5. Schnepf E, Crickmore N, Rie JV, Lereclus D, Baum J, Feitelson J, et al. *Bacillus thuringiensis* and Its Pesticidal Crystal Proteins. 1998;62(3):775-806.
6. Höfte H, Whiteley HR. Insecticidal crystal proteins of *Bacillus thuringiensis*. Microbiological reviews. 1989;53(2):242-55.

7. Deepak R, Jayapradha R. Lipopeptide biosurfactant from *Bacillus thuringiensis* pak2310: A potential antagonist against *Fusarium oxysporum*. *Journal de Mycologie Médicale*. 2015;25(1):e15-e24.
8. Djenane Z, Nateche F, Amziane M, Gomis-Cebolla J, El-Aichar F, Khorf H, et al. Assessment of the Antimicrobial Activity and the Entomocidal Potential of *Bacillus thuringiensis* Isolates from Algeria. 2017;9(4):139.
9. Garner B, Brown E, Taplin M, Garcia A, Williams-Mapp B. Transferrin Impacts *Bacillus thuringiensis* Biofilm Levels. *BioMed Research International*. 2016;2016:3628268.
10. Cochrane SA, Vederas JC. Lipopeptides from *Bacillus* and *Paenibacillus* spp.: A Gold Mine of Antibiotic Candidates. 2016;36(1):4-31.
11. Datta SC, Wook YB, In-Cheol Y, Tae HY. Antimicrobial peptides of the genus *Bacillus*: a new era for antibiotics. 2015;61(2):93-103.
12. Tanada Y. *Bacillus thuringiensis*: Integrated Control — Past, Present, and Future. In: Cheng TC, editor. *Pathogens of Invertebrates: Application in Biological Control and Transmission Mechanisms*. Boston, MA: Springer US; 1984. p. 59-90.
13. Steinhaus EA. Possible use of *Bacillus thuringiensis* Berliner as an aid in the biological control of the alfalfa caterpillar. *Hilgardia*. 1951;20(18):359-81.
14. Nair K, Al-Thani R, Al-Thani D, Al-Yafei F, Ahmed T, Jaoua S. Diversity of *Bacillus thuringiensis* Strains From Qatar as Shown by Crystal Morphology, δ -Endotoxins and Cry Gene Content. 2018;9(708).
15. Daas MS, Rosana ARR, Acedo JZ, Douzane M, Nateche F, Kebbouche-Gana S, et al. Insights into the draft genome sequence of bioactives-producing *Bacillus thuringiensis* DNG9 isolated from Algerian soil-oil slough. *Standards in Genomic Sciences*. 2018;13(1):25.
16. Wang P, Zhang C, Guo M, Guo S, Zhu Y, Zheng J, et al. Complete genome sequence of *Bacillus thuringiensis* YBT-1518, a typical strain with high toxicity to nematodes. *Journal of Biotechnology*. 2014;171:1-2.
17. Kaze M, Brooks L, Sistro M. Antimicrobial resistance in *Bacillus*-based biopesticide products. *Microbiology*. 2021;167.

18. Gassmann AJ, Shrestha RB, Kropf AL, St Clair CR, Brenizer BD. Field-evolved resistance by western corn rootworm to Cry34/35Ab1 and other *Bacillus thuringiensis* traits in transgenic maize. *Pest management science*. 2020;76(1):268-76.
19. Rabha M, Sharma S, Acharjee S, Sarmah BK. Isolation and characterization of *Bacillus thuringiensis* strains native to Assam soil of North East India. *3 Biotech*. 2017;7(5):303.
20. Yilmaz S, Ayvaz A, Azizoglu U. Diversity and distribution of cry genes in native *Bacillus thuringiensis* strains isolated from wild ecological areas of East-Mediterranean region of Turkey. *Tropical Ecology*. 2017;58:605-10.
21. YILMAZ S. Molecular characterization of *Bacillus thuringiensis* strains isolated from various habitats and their use against some harmful insects. *Erciyes University Research Information System: Erciyes University*; 2010.
22. Yilmaz S, Ayvaz A, Akbulut M, Azizoglu U, Karabörklü S. A novel *Bacillus thuringiensis* strain and its pathogenicity against three important pest insects. *Journal of Stored Products Research*. 2012;51:33-40.
23. Azizoglu U, Ayvaz A, Yilmaz S, Karabörklü S, Temizgul R. Expression of cry1Ab gene from a novel *Bacillus thuringiensis* strain SY49-1 active on pest insects. *Braz J Microbiol*. 2016;47(3):597-602.
24. Azizoglu ZB, Yilmaz S, Azizoglu U, Karabörklü S, Temizgul R, Ayvaz A. Molecular characterization of the chitinase genes of native *Bacillus thuringiensis* isolates and their antagonistic activity against three important phytopathogenic fungi. *Biologia*. 2021;76(9):2745-55.
25. Travers RS, Martin PAW, Reichelderfer CF. Selective Process for Efficient Isolation of Soil Bacillus spp. *Applied and Environmental Microbiology*. 1987;53(6):1263.
26. Yilmaz S, Idris AB, Ayvaz A, Çetin A, Ülgen F, Çetin M, et al. New Insights into Molecular Basis Identification of Three Novel Strains of the *Bacillus Subtilis* Group Produce Cry Proteins Isolated from Soil Samples in Adana, Turkey. 2021:2021.12.24.474129.

27. Baig DN, Mehnaz S. Determination and distribution of cry-type genes in halophilic *Bacillus thuringiensis* isolates of Arabian Sea sedimentary rocks. *Microbiological research*. 2010;165(5):376-83.
28. Centralblatt für Bakteriologie, Parasitenkunde und Infektionskrankheiten. 1. Abt. Originale. Jena :: G. Fischer; 1902.
29. Travers RS, Martin PA, Reichelderfer CF. Selective Process for Efficient Isolation of Soil *Bacillus* spp. *Applied and environmental microbiology*. 1987;53(6):1263-6.
30. Cerón J, Ortiz A, Quintero R, Güereca L, Bravo A. Specific PCR primers directed to identify cryI and cryIII genes within a *Bacillus thuringiensis* strain collection. *Applied and environmental microbiology*. 1995;61(11):3826-31.
31. Bravo A, Sarabia S, Lopez L, Ontiveros H, Abarca C, Ortiz A, et al. Characterization of cry genes in a Mexican *Bacillus thuringiensis* strain collection. *Applied and environmental microbiology*. 1998;64(12):4965-72.
32. Mapleson D, Garcia Accinelli G, Kettleborough G, Wright J, Clavijo BJ. KAT: a K-mer analysis toolkit to quality control NGS datasets and genome assemblies. *Bioinformatics*. 2017;33(4):574-6.
33. Gu S, Fang L, Xu X. Using SOAPaligner for Short Reads Alignment. *Current protocols in bioinformatics*. 2013;44:11.1-7.
34. Li R, Zhu H, Ruan J, Qian W, Fang X, Shi Z, et al. De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res*. 2010;20(2):265-72.
35. Aziz RK, Bartels D, Best AA, DeJongh M, Disz T, Edwards RA, et al. The RAST Server: Rapid Annotations using Subsystems Technology. *BMC Genomics*. 2008;9(1):75.
36. Li W, O'Neill KR, Haft DH, DiCuccio M, Chetvernin V, Badretdin A, et al. RefSeq: expanding the Prokaryotic Genome Annotation Pipeline reach with protein family model curation. *Nucleic acids research*. 2021;49(D1):D1020-d8.
37. Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics (Oxford, England)*. 2014;30(14):2068-9.

38. Wattam AR, Abraham D, Dalay O, Disz TL, Driscoll T, Gabbard JL, et al. PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Research*. 2013;42(D1):D581-D91.
39. Grissa I, Vergnaud G, Pourcel C. CRISPRFinder: a web tool to identify clustered regularly interspaced short palindromic repeats. *Nucleic acids research*. 2007;35(Web Server issue):W52-7.
40. Blin K, Shaw S, Kloosterman AM, Charlop-Powers Z, van Wezel GP, Medema Marnix H, et al. antiSMASH 6.0: improving cluster detection and comparison capabilities. *Nucleic Acids Research*. 2021;49(W1):W29-W35.
41. Arndt D, Grant JR, Marcu A, Sajed T, Pon A, Liang Y, et al. PHASTER: a better, faster version of the PHAST phage search tool. *Nucleic acids research*. 2016;44(W1):W16-W21.
42. Bertelli C, Laird MR, Williams KP, Lau BY, Hoad G, Winsor GL, et al. IslandViewer 4: expanded prediction of genomic islands for larger-scale datasets. *Nucleic Acids Res*. 2017;45(W1):W30-w5.
43. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. *Molecular biology and evolution*. 2016;33(7):1870-4.
44. Ondov BD, Treangen TJ, Melsted P, Mallonee AB, Bergman NH, Koren S, et al. Mash: fast genome and metagenome distance estimation using MinHash. *Genome Biology*. 2016;17(1):132.
45. Davis JJ, Gerdes S, Olsen GJ, Olson R, Pusch GD, Shukla M, et al. PATtyFams: Protein Families for the Microbial Genomes in the PATRIC Database. *Front Microbiol*. 2016;7:118.
46. Meier-Kolthoff JP, Auch AF, Klenk HP, Göker M. Genome sequence-based species delimitation with confidence intervals and improved distance functions. *BMC bioinformatics*. 2013;14:60.

47. Delcher AL, Salzberg SL, Phillippy AM. Using MUMmer to Identify Similar Regions in Large Sequence Sets. *Current protocols in bioinformatics*. 2003;00(1):10.3.1-3.8.
48. Delcher AL, Kasif S, Fleischmann RD, Peterson J, White O, Salzberg SL. Alignment of whole genomes. *Nucleic Acids Res*. 1999;27(11):2369-76.
49. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs over the Nearly Four Decades from 01/1981 to 09/2019. 2020;83(3):770-803.
50. Roy A, Mahata D, Paul D, Korpole S, Franco O, Mandal S. Purification, biochemical characterization and self-assembled structure of a fengycin-like antifungal peptide from *Bacillus thuringiensis* strain SM1. *Frontiers in Microbiology*. 2013;4(332).
51. Anantharaman A, Rizvi MS, Sahal D. Synergy with rifampin and kanamycin enhances potency, kill kinetics, and selectivity of de novo-designed antimicrobial peptides. *Antimicrobial agents and chemotherapy*. 2010;54(5):1693-9.
52. Roy A, Franco OL, Mandal SM. Biomedical exploitation of self assembled peptide based nanostructures. *Current protein & peptide science*. 2013;14(7):580-7.
53. Vanittanakom N, Loeffler W, Koch U, Jung G. Fengycin--a novel antifungal lipopeptide antibiotic produced by *Bacillus subtilis* F-29-3. *The Journal of antibiotics*. 1986;39(7):888-901.
54. Deleu M, Paquot M, Nylander T. Effect of Fengycin, a Lipopeptide Produced by *Bacillus subtilis*, on Model Biomembranes. *Biophysical Journal*. 2008;94(7):2667-79.
55. Silo-Suh LA, Stabb EV, Raffel SJ, Handelsman J. Target range of zwittermicin A, an aminopolyol antibiotic from *Bacillus cereus*. *Current microbiology*. 1998;37(1):6-11.
56. Emmert EA, Klimowicz AK, Thomas MG, Handelsman J. Genetics of zwittermicin a production by *Bacillus cereus*. *Applied and environmental microbiology*. 2004;70(1):104-13.
57. Wilson MK, Abergel RJ, Arceneaux JE, Raymond KN, Byers BR. Temporal production of the two *Bacillus anthracis* siderophores, petrobactin and bacillibactin.

Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine. 2010;23(1):129-34.

58. Koppisch AT, Dhungana S, Hill KK, Boukhalfa H, Heine HS, Colip LA, et al. Petrobactin is produced by both pathogenic and non-pathogenic isolates of the *Bacillus cereus* group of bacteria. *Biometals : an international journal on the role of metal ions in biology, biochemistry, and medicine*. 2008;21(5):581-9.

59. Willey JM, Donk WAvd. Lantibiotics: Peptides of Diverse Structure and Function. *Annual Review of Microbiology*. 2007;61(1):477-501.

60. Wang J, Ma H, Ge X, Zhang J, Teng K, Sun Z, et al. Bovicin HJ50-Like Lantibiotics, a Novel Subgroup of Lantibiotics Featured by an Indispensable Disulfide Bridge. *PLOS ONE*. 2014;9(5):e97121.

Figure 1. Circular visualization of the draft genome of *Bt SY49.1* represents relevant genome features. This includes, from outer to inner circles, the contigs, CDS on the forward strand, CDS on the reverse strand, RNA genes, CDS with homology to known antimicrobial resistance genes, CDS with homology to known virulence factors, GC content, and GC skew. The colors of the CDS on the forward and reverse strand indicate the subsystem that these genes belong to, as presented in B). The circular display has been limited to the 230 longest contigs of the 708 contigs in the genome.

Figure 2. Microscopical characteristic of *Bt SY49.1*. 2A. Microscopic view of *Bt SY49.1* after spore staining. Arrows indicate the endospores. 2B. Scanning Electron Microscopy (SEM) image of the different types of crystal proteins and spores (S) produced by *Bt SY49.1*. B: bipyramidal, C: cubic, Sp: spherical.

Figure 3. Functional annotation of the *Bt* SY49.1 strain genome sequence

Figure 4. Secondary metabolite biosynthetic gene cluster organization in *Bt* SY49.1 genomes as predicted by the antiSMASH 4.0 server.

Figure 5. Prophage regions predicted in *Bt* SY49.1 genomes. 4A. Shows intact prophage region (score >90). 4B and 4C. show questionable prophage regions (score 70-90). 4D. shows the color legend.

Figure 6. Molecular phylogenetic analysis of *Bt* SY49.1. The percentage of replicate trees (1000 replicates) is shown next to the branches. 5A. The maximum likelihood phylogenetic tree of *Bt* SY49.1 using *16S rRNA* sequences. The evolutionary distance was computed using the (K2+G) model. 5B. The phylogenetic tree of *Bt* SY49.1 complete genome using PATRIC global protein families (PGFams). The evolutionary analyses were conducted for *16S rRNA* and PGFams using MEGA7 and PATRIC software, respectively.

Figure 7. Genomic comparison of *Bt* SY49.1 to other *Bacillus* strains genomes. Each circle represents a pair-wise BLAST comparison between the open reading frames in query genome against those in *Bt* SY49.1 (Ref. = reference), with the percentage of similarity represented with different colors shown in the legend. Regions marked in the genomic map correspond to gene numbers presented in Additional File 1: Table S1, (A= 3993-4126) and (B= 4654-4792). Query genomes used in this analysis (outer ring to inner ring): *B. anthracis* str. Ames Ancestor (NC_007530), *B. cereus* ATCC 10987 (NC_003909), *B. cereus* ATCC 14579 (NC_004722), *B. cereus* E33L (NC_006274), *B. subtilis* subsp. *subtilis* str. 168 (NC_000964), *B. thuringiensis* str. 97-27 (NC_005957), *B. thuringiensis* str. Al-Hakam (NC_008600).

Figure 8. InDel statistics of different size