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

Semih Yılmaz[1,2], Abeer Babiker Idris[3]*, Abdurrahman Ayvaz[4], Aysun Çetin[5], Funda Ülgen[6], Mustafa Çetin[2], Berkay Saraymen[7], Mohamed A. Hassan[2,8,9,10]

* Corresponding author: abeer.babiker89@gmail.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 Medical Biochemistry, Faculty of Medicine, Erciyes University, Kayseri, Turkey
[6] Department of Biology, Institute of Natural and Applied Sciences, Erciyes University, Kayseri, Turkey
[7] ERNAM Nanotechnology Research and Application Center, Erciyes University, Kayseri, Turkey
[8] Department of Bioinformatics, Africa City of Technology, Khartoum, Sudan
[9] Department of Bioinformatics, DETAGEN Genetic Diagnostics Center, Kayseri, Turkey
[10] Department of Translation Bioinformatics, Detavax Biotech, Kayseri, Turkey

SY: ylmazsemh@yahoo.com; ABI: abeer.babiker89@gmail.com; AA: ayvaza@erciyes.edu.tr; AÇ: aysuncetin@yahoo.com; FÜ: fundaulgen@gmail.com; MÇ: mcetin2000@yahoo.com; BS: berkaysaraymen@gmail.com; MAH: altwoh20002002@yahoo.com

Abstract

Aims

This study aimed to analyze the evolutionary relationship between *Bacillus* species isolated from agricultural soil using in-silico tools.

Methods and Results

Across-sectional study was conducted in Adana province, in Turkey. A total of 120 *Bacillus* species were isolated from 80 soil samples. However, the phylogenetic tree diverged into two lineages; one belongs to *B. subtilis* group while the other belongs to *B. cereus* group. Interestingly, three native strains (*SY27.1A*, *SY35.3A*, and *SY58.5A*), which produce Cry proteins, shared high similarity with *B. subtilis* group (over 99%) and less than 95% similarity with known *B. thuringiensis* and other species of *B. cereus* group. Furthermore, 11 canonical SNPs (canSNPs) were identified in strains that belong to *B. pumilus* group when compared with *B. subtilis* reference sequences.

Conclusions

Phylogenetic analysis of *16S rRNA* sequences was found valuable for differentiation between *Bacillus* species isolated from soil samples. In addition, SNPs analysis provided more intra-specific information in the cases of *B. subtilis* group.

Significance and Impact of Study

A detailed analysis was provided for the SNPs present in a conserved region of *16S rRNA* gene of *Bacillus* species. Also, we proposed three novel *Bacillus* strains that produce Cry proteins and belong to *B. subtilis* group.

Keywords: *Bacillus* species, *16S rRNA* gene, evolutionary tree, canonical SNPs, Turkey.

1. Introduction

Bacillus is agriculturally important insecticidal bacterial genus that naturally inhabit the phyllosphere and rhizosphere (1). It consists of a heterogeneous group of Gram-positive, endospore-forming, aerobic or facultative anaerobic organisms (2). Most members of the genus *Bacillus* have the ability to produce antibiotics, enzymes, vitamins, proteins, or secondary metabolites that are capable to induce defense mechanisms and promote growth in animals and plants (3, 4). Benefiting from their metabolic diversity and spore dispersal, *Bacillus* is ubiquitous in various natural environments especially terrestrial environments (5). At the time of writing, the genus *Bacillus* consisted of more than 408 species with validly published names (LPSN, http://www.bacterio.net), only 54 species of them were reported before 2000 (6).

Analysis of *16S rRNA* gene, the "ultimate molecular chronometer", has been extensively applied for bacterial phylogeny and taxonomy which resulted eventually in the establishment of large public-domain databases (7-11). The *16S rRNA* gene characterizes by several properties which include being present in all bacteria, thus it is a universal target for bacterial identification and characterization (12, 13). In addition, the function of *16S rRNA* has not changed over a long period, i.e. random sequence changes are more likely to reflect the microbial evolutionary change (phylogeny) (11), and any introduction of selected

changes in one domain does not greatly affect sequences in other domains (13, 14). Based on phylogenetic analysis of the *16S rRNA* gene, the species and strains in *Bacillus* are divided into five groups: *B. cereus*, *B. megaterium*, *B. subtilis*, *B. circulans* and *B. brevis* groups (15).

The *B. subtilis* group is a tight assemblage of closely related species which includes *B. subtilis*, *B. amyloliquefaciens*, *B. atrophaeus*, *B. axarquiensis*, *B. malacitensis*, *B. mojavensis*, *B. sonorensis*, *B. tequilensis*, *B. vallismortis* and *B. velezensis* (16). These species share high genetic homogeneity (over 99.5%) and cannot differentiate on the basis of phenotypic or biochemical characteristics (15, 16). In addition, *B. pumilus* and their relatives belong to the *subtilis* group (17). The *B. pumilus* group, which is a large group of *Bacillus*, composed of *B. pumilus*, *B. altitudinis*, *B. safensis*, *B. zhangzhouensis*, *B. xiamenensis*, and *B. australimaris* (5).

The bacteria of *B. cereus* group share high genetic homogeneity despite their phenotypic diversity, with over 97% *16S rRNA* sequence similarity among *B. cereus*, *B. anthracis*, *B. thuringiensis*, *B. weihenstephanensis*, *B. mycoides*, *B. pseudomycoides*, *B. cytotoxicus*, *B. gaemokensis* and *B. manliponensis* (18). Moreover, this group is of interest to researchers, especially *B. thuringiensis*, because of their significance in agriculture, industry and medicine (19). *B. thuringiensis* acts as a biological control agent against different phytopathogenic organisms due to their ability to produce insecticidal proteins (Cry, Vip, Sip, Bin, etc), fungicidal metabolites (iturin, fengycin, surfactin, zwittermycin, etc) (20, 21). Also, *B.*

thuringiensis can promote plant growth by producing ACC deaminase, phosphatases, siderophore, etc (22, 23). However, the isolation and characterization of native *Bacillus* species or strains, especially from agricultural soil, should receive a good attention because of their wide potential biological products with immense applications. In addition, there are several studies on the isolation and characterization of native *Bacillus* strains of soil in order to identify novel toxins with high level of toxicity and effective against agricultural pests (24-26). Finding of novel toxins produce from *Bacillus* species, especially *B. thuringiensis,* will delay the resistance within the pests due to the use of existing *Bacillus* toxins. Therefore, in this study, we aimed to characterize and establish a phylogenetic relationship between *Bacillus* species isolated from agricultural soil in Adana, one of the most fertile agricultural area in Turkey, by reconstructing *16S rRNA* phylogenetic trees using in silico tools. Also, a detailed analysis was provided, for the SNPs present in a conserved region of *16S rRNA* gene of *B. cereus* group, *B. Subtilis* group and *B. pumilus* group. The results of canonical SNPs (canSNPs) are of great significance for the design of primers or probes specific to a strain, species, or group of species.

2. Materials and Methods

2.1 Study Settings and Sample collection

A cross-sectional study was conducted in Adana province, which is located in the southern region of Turkey. Adana province is divided into 13 districts with different texture. For the isolation of bacterial strains, 80 different soil samples were collected throughout Adana province from different altitudes ranging from 0 to 1582 meters. Soil samples have been taken in a depth of 2-10 centimeters and stored in sterile tubes at 4°C for the studies.

2.2 Bacterial Isolation and Identification

Isolation processes has been performed according to the method of Travers *et al.* (27). One gram of soil sample was inoculated in LB medium (pH 6.8±2) including 0.25M sodium acetate in a shaking incubator at 200 rpm at 30°C for 4 hours. After the incubation step, 1.5 ml of liquid samples have been transferred to a sterile Eppendorf tube and exposed to 80°C for 10 minutes to kill the vegetative bacterial forms. A 20-50 ul of samples were spread on LB agar plates and incubated overnight at 30°C. The colonies with morphological differences have been spread on agar plates and pure colonies were obtained. Pure colonies were incubated in 5ml LB broth (pH 6.8±2) in 50 ml tubes at 200 rpm and 30°C overnight.

Then the colonies were homogenized in 400 µl of sterile dH_2O in microfuge tubes and 10 ul were added onto sterilized Watman no:1 paper disc with 0.4 mm diameter. The discs were then placed into Potato Dextrose agar (PDA) plates and incubated overnight at 30°C. Morphologically pale-yellow, grayish white, pale-pink, ciliated, or wrinkled ends, round-shaped outlines were selected for Gram stain (28). Gram- positive colonies were further investigated for spore production using Malachite green (5g /100ml) staining as previously described (29).

2.3 Characterization of para-sporal inclusions

To characterize para-sporal inclusions, the bacteria were incubated in 150 ml of 3T medium (2 g triptose, 3 g triptone, 1.5 g yeast extract, 6 g NaH_2PO_4, 0.005 g $MnCl_2$ and 7.1 g Na_2HPO_4) at 200 rpm and 30 °C for 7 days to induce sporulation (27). Then, to harvest spore-crystal mixtures, the suspensions were centrifuged at 15000 ×g and 4 °C for 10 min. After that, the mixtures were suspended in dH_2O on microscope slides and fixed. Finally, the slides were sputter coated with 10 nm Au/Pd using a SC7620 Mini-sputter coater and viewed using a LEO440 scanning electron microscope (SEM) at 20kV beam current (30, 31) in this study reference standard *B. thuringiensis* strains such as *Bt. kurstaki HD1*, *Bt. kurstaki HD73*, *Bt. aizawai* (Universidad Nacional Autonoma de Mexico Biyotechnology Institute), *Bt. morrisoni*, *Bt. israelensis* (Pasteur Institute, Paris, France) and, *Bt. tenebrionis* (Plant Genetic Systems, J. Plateaustroat 22, 900 Gent, Belgium) were used for comparison with the native local isolates.

2.4 Genetics analysis

2.4.1 Determination of the insecticidal *crystal* genes (*cry*) carrying isolates

Extraction of DNA was performed according to a previously described method (32, 33). Briefly, the bacteria were grown in LB medium for overnight, and then a loopful of cells was placed into 400 µl sterile dH_2O. Then the mixture was boiled for 10 min to lyse the cells. The resulting cell lysate was centrifuged for 10 sec at 10.000 rpm and the supernatant was used as DNA templates for PCR reactions. The extracted DNA from the isolates was used to determine the *cry* genes carrying strains. In our previous work, we have characterized the isolates using *cry* genes *cry1Ab/ Ac, cry1Aa/Ad, cry2, cry5, and cry9C, cry1C, cry1Ad, cry1Ac, cry1D, cry1B, cry3-7-8, cry4A, cry9A,* and *cry11A/B* (30, 31).

2.4. 2 Amplification and sequencing of *16S rRNA* gene

The extracted DNA was used to amplify the *16S rRNA* gene of Bacillus species using universal primers with the following sequences: F: 5'-AAA CTY AAA KGA ATT GAC GG-3' and R: 5'-ACG GGC GGT GTG TRC-3'. The thermal procedures were performed with ABI veriti Thermocycler and the PCR mixtures contained 2.3 mM MgCl2, 1x Taq buffer, 0.2 mM dNTP mix, 0.3 pmol for each primer, 0.5 U Taq DNA polymerase, and 30–100 ng template DNA. The PCR conditions were an initial denaturation 94°C 5 min, then 40 cycles of denaturation at 94°C for 1min, primer annealing at 48°C for 1 min, extension at 72°C for 2min, and then

177 additional extension step 72°C for 10min. The size of expected PCR products was
178 850 bp.

179 For sequencing, the DNA fragments were extracted from the gel using a
180 EasyPure® Quick Gel Extraction Kit (EG101-01) according to the manufacturer's
181 instruction. Then the PCR products of 21 samples, which have the clearest bands,
182 were sent for commercial DNA purification and Sanger dideoxy sequencing by
183 DETAGEN Genetic Diagnostics Center Inc., Turkey.

184 2.**5 Bioinformatics analysis**

185 2.5.1 Sequence and SNP analysis

186 The two chromatograms for each strain (forward and reverse) were visualized,
187 checked the quality, and analyzed using the Finch TV program version 1.4.0 (34).
188 The bacterial strains were identified by searching for their homology among
189 published reference sequences using the nucleotide Basic Local Alignment
190 Search Tool (BLASTn; https://blast.ncbi.nlm.nih.gov/) (35). To determine the
191 SNPs, multiple sequences alignment (MSA) was accomplished with reference
192 sequences of *Bacillus* species using BioEdit software (36) and MEGA version 7.0
193 software (37). This MSA facilitated the use of polymorphisms to detect potential
194 relationships between the *Bacillus* strains and species. In addition, the detected
195 SNPs were carefully reviewed by eye using the Finch TV software; and
196 polymorphisms present in both the forward and reverse strands were
197 considered.

198 2.5. 2 Molecular phylogenetic analysis

199 For building the phylogenetic tree, the studied sequences and their highly similar
200 references sequences that retrieved from the NCBI GenBank were subjected to
201 Gblocks software to eliminate poorly aligned positions and divergent regions of
202 aligned sequences, so the alignment becomes more suitable for phylogenetic
203 analysis (38). The molecular evolutionary analyses were conducted with MEGA
204 7.0 software (37) using the maximum likelihood (ML) method and neighbour-
205 joining (NJ) method (39, 40). The Kimura 2-parameter (K2+G) model from the
206 substitution (ML) model was used with 1000 bootstrap replicates to construct
207 distance-based trees (41, 42).

208

209

210

211 **3. Results**

212 **3.1 Bacterial isolation and characterization**

213 A total of 120 *Bacillus* species were isolated from 80 soil samples. Eighty-eight of *Bacillus*
214 isolates harbored *cry* genes. The *cry* genes were determined and characterized using
215 conventional PCR. Also, spore–crystals of some of the samples were examined under the
216 SEM. The results in details were presented in our previous works (30, 31). In the current
217 study, *16S rRNA* of 21 *Bacillus* species were amplified and sequenced to construct a

phylogenetic tree. Among them, six strains were characterized by producing Cry proteins. The morphology of spore–crystals and *cry* genes profiles of the native *Bacillus* isolates are illustrated in Table 1. The nucleotide sequences of the *16S rRNA* were deposited in the GenBank database under the following accession numbers: from OK428682 to OK428687 and from OK384678 to OK384692.

Table 1. The morphology of spore–crystals and *cry* gene profiles of the native *Bacillus* isolates

Isolates	**Cry forms**	***cry* genes**
SY49.1	bp	*cry1Aa/Ad, cry1B, cry1C, cry5, cry9A, cry9C*
SY27.1A	bp, cs, ss, pe	*cry1Ad, cry1Ac, cry1Ab/Ac, cry1B, cry2, cry9C*
SY35.3A	ss	*cry3-7-8*
SY58.5A	bp, cs	*cry1Ac, cry1Ab/Ac, cry2*
SY10.1A	bp	*cry1Ac*
SY25. 1A	bp	*cry1C, cry1Ab/Ac, cry1Aa/Ad*

bp: bipyramidal; ss: smooth spherical; pe: spherical crystal with pointy edges; cs, cuboidal

3.2 Sequencing analysis of *16S rRNA* gene

Twenty-one isolates of *Bacillus* species were subjected to PCR amplification and nucleotide sequencing using universal *16S rRNA* primers but specific for a conserved region of the gene which was located between 629 bp and 1552 bp. The sequences of the studied strains were aligned with reference sequences of *Bacillus* species retrieved from NCBI databases. The information about the retrieved strains is given in supplementary Table S1.

As presented in Figure 1, 11 strains revealed high similarity with the *B. cereus* group (over 99%), see Figure 1. Among them, seven strains, which were found to be homogenous, revealed two nucleotide variations (A1015C and A1146T). Numbers are given in all sequences in accordance with numbering in the *B. thuringiensis* genome (NZ_CM000747). Six of these strains (*SY49.1, SY10.1A*, and *SY25.1A*) were characterized by the production of Cry proteins. While six and four of the studied strains exhibited high similarities to *B. subtilis* group and *B. pumilus* group, respectively (Figure 3). Interestingly, three of native strains (*SY27.1A, SY35.3A, SY58.5A*), which produce Cry proteins, shared high similarity with *B. subtilis* group (over 99%) and less than 95% similarity with known *B. thuringiensis* and other species of *B. cereus* group, see Figure 2.

3.3 Molecular phylogenetic analysis

The evolutionary analysis of 21 native Turkish *Bacillus* strains, based on *16S rRNA* gene, was conducted with reference sequences of *Bacillus* species retrieved from NCBI GenBank databases. Neighbor-joining (NJ), maximum-likelihood (ML) analyses were performed with MEGA 7 (37). The topology of the ML and NJ trees was similar, and the bootstrap supports of the NJ tree were approximately higher than those of ML. The phylogenetic tree diverged into two lineages; one belongs to *B. subtilis* group while the other belongs to *B. cereus* group (Figure 4).

The lineage of *B. subtilis* group branched into two major clades (I and II). In clade I, six strains (SY27.1A, SY35.3A, SY58.5A, SY80, SY81 and SY85) were clustered with *B. subtilis*, *B. amyloliquefaciens*, *B. velezensis* and *B. mojavensis*. However, SY27.1A, SY35.3A and SY85 strains were closely related to *B. velezensis* strain CBMB205 and they shared nucleotide variations at TA1461-1462AT. While in clade II, four strains (*SY61.6, SY82*,

SY83 and *SY84*) were grouped with *B. pumilus* group which comprises *B. pumilus*, *B. safensis*, and *B. altitudinis*. In addition, all strains in clade II shared 11 nucleotide variations with *B. pumilus* group (T1017C, A1030G, G1032T, C1034T, G1045A, A1265T, A1270G, G1276A, A1282T, T1432C and A1485G), see Figure 3a. Four mutations (C971G, C1316T, C1330T and A1543C) in strain A made it a separate minor clade. Interestingly, three native strains (*SY27.1A*, *SY35.3A* and *SY58.5A*) that characterized by producing Cry proteins, which is often considered as a feature of *B. thuringiensis*, were clustered with *B. subtilis* group (Figure 2). The lineage of *B. cereus* group involved 11 strains. Three of them (*SY49.1*, *SY10.1A*, and *SY25.1A*) were characterized by producing Cry proteins. Moreover, seven strains shared a common ancestor and were characterized by two nucleotide variations (A1015C and A1146T). However, strain *SY25.1A* and *B. nitratireducens* strain *MCCC 1A00732* were sisters with a bootstrap value of 63%, see Figure 4 for more illustration.

4. Discussion

In this study, we observed three novel Gram- positive bacilli (*SY27.1A*, *SY35.3A*, and *SY58.5A*) which produce Cry proteins but, based on the analysis of *16S rRNA* gene, they are unlikely to belong to the known *B. thuringiensis* or other species of *Bacillus cereus* group. Although the defining feature of the *B. thuringiensis* species is the ability to express Cry proteins (20), the analysis of *16S rRNA* gene sequences of these strains showed that they were sharing over 97% similarity with *B. subtilis* group and less than

95% similarity with the known *B. thuringiensis* and other species of *B. cereus* group. However, in prokaryotes taxonomy, *16S rRNA* gene sequence identity of 97 % is generally used as a threshold value for species definition, therefore, strains with less than 97.5% identity are unlikely to be related at the species level (43, 44). This finding is partially in agreement with a previous study that systematically searched for Cry proteins expressed by *Bacillus* species, other than *B. thuringiensis*, genomes using conserved sequences from the C-terminal half of reported Cry proteins in hidden Markov Model (HMM) profiles (45). Interestingly, there were 174 Cry protein sequences were observed, as expected most of them were in *B. thuringiensis* genomes, but 42 were found in other species. In addition, several studies reported the presence of Cry proteins in other Bacilli, such as *P. popilliae*, *C. bifermentans*, *L. sphaericus* and *P. lentimorbus* (46-49). Nevertheless, the great diversity of Cry proteins may indicate that this family of proteins may not be restricted only to the *B. thuringiensis* species (20), and their dispersion and role in nature are could be much wider. Therefore, further studies, either based on an in-silico procedure or the use of a large data collection of different species of bacilli, are recommended to search for new Cry proteins with higher toxicity or different mode of action, which may render alternatives in case of resistance development. However, the development of resistance to insecticidal *B. thuringiensis* proteins has been documented which raise concerns about the adequacy of current resistance management strategies (54). Hence, continuous searching for novel *Bacillus* species with novel insecticidal genes to delay the development of insect resistance is of the utmost importance. In this connection, to isolated novel *Bacillus* strains, we collected soil samples throughout Adana province. This region is rich in biodiversity due to its unique climate and geographical location that is situated on the fertile and watery delta of Seyhan and

Ceyhan rivers, furthermore in this study, the genetic identity based on *16S rRNA* sequences indicated that three native strains of *Bacillus* (SY27.1A, SY35.3A, and SY58.5A), which produce Cry proteins were close relatives of the *B. subtilis* group and appeared to be discrete from the *B. cereus* group. However, according to the low discrimination of the *16S rRNA* gene between *Bacillus* species, it cannot be assigned accurately as a certain species (55), therefore, complete genome sequencing of these bacterial strains is recommended. While, the other ten *Bacillus* strains, which clustered with *B. cereus* group, showed diversity into two minor clades. Seven of them shared a common ancestor and were characterized by two nucleotide variations (A1015C and A1146T). This finding is in agreement with other studies conducted in different countries which showed diversity in *B. cereus* / *B. thuringiesis* strains isolated from soil and other natural sources (24, 26, 56-58).

Furthermore, in the phylogenetic tree of the native strains with global reference sequences of *Bacillus* species, the lineage of *B. subtilis* group were branched into two major clades. Clade I contained *B. amyloliquefaciens, B. velezensis, B. mojavensis* along with *B. subtilis* which is not entirely unexpected since these species share a remarkably high level of *16S rRNA* gene sequence similarity to *B. subtilis* (often 99 % or greater) (16). While in clade II, the members of *B. pumilus* group (*B. pumilus, B. altitudinis,* and *B. safensis*) were clustered together along with four strains (*SY61.6A, SY82, SY83,* and *SY84*). They shared 11 nucleotide variations when their *16S rRNA* sequences compared with *B. subtilis* reference sequences, see Figure 3a. In 1973, Gordon *et al.* speculated that *B. pumilus* might one day be considered a variety of *B. subtilis* rather than a separate species once more data were collected (59). Intriguingly, our findings are in agreement

with Rooney *et al.* results which clearly indicated that the *B. pumilus* forms a clade distinct from *B. subtilis* (16).

However, single nucleotide polymorphism (SNP) analysis has emerged as one of the most useful molecular methods proposed for microbial characterization and improving discrimination among closely related species (60-62). Accordingly, we used universal primers of *16S rRNA* in order to produce a mixture of amplicons from all *rRNA* operons in the genome. Although there are small differences exhibit by the multiple *rRNA* gene copies in each genome, these differences do not invalidate bacterial identification and characterization based on *16S rRNA* sequences (63, 64). Moreover, these differences between *rRNA* operons appear in multiple peaks (two or more) at a single nucleotide position in the case of SNPs. But in the case of InDels variations, the sequence loses synchronicity and makes an abrupt change, from clean to dirty, following an InDel mutation (65), for more illustration see Figure3b. In the present study, we provided a detailed analysis of the SNPs present in the *16S rRNA* gene of *Bacillus* species isolated from soil samples. Of great significance, dual peaks (A and T) at position 1146, which were previously reported to be specific to *B. anthrac*is (66), were detected in three *B. thuringiensis / B. cereus* strains (*PSY6.1A, PSY6.2A* and *SY10.1A*). But in all other strains, a single peak (either A or T) was detected. This finding is in accordance with a study conducted by Hakovirta *et al.* which found five *B. cereus* strains and three *B. thuringiensis* strains also had A and T peaks (65). Hence, the SNP at nucleotide position 1146 is not necessarily reliable for identifying *B. anthracis*. While, in this study, all *B. thuringiensis* had only G at position 1139 which is another SNP proposed by Hakovirta *et al.* to be unique to *B. anthracis* (65), and it appears to be more reliable for identifying *B. anthracis*.

Regarding the *B. subtilis* group, we identified 11 canonical SNPs (canSNPs) in strains that belong to *B. pumilus* groups (clade II) when compared with *B. subtilis* reference sequences. Canonical SNPs (canSNPs) are useful and diagnostic SNPs that used for identifying long branches or key phylogenetic positions (67). In addition, G1268A and C1294T SNPs were specific to *B. safensis* and *B. altitude*, respectively. Also, strain *SY83* characterized by four nucleotide variations SNPs (C971G, C1316T, C1330T and A1543C). These findings are partially in agreement with a study conducted by Moorhead *et al.* in which strains of *Listeria monocytogenes* were partitioned into three previously described clonal lineages using a phylogenetic approach to detect a small number of SNPs in the *sigB* gene (68). However, a number of researchers have found that a small number of SNPs can be used to effectively identify genetic groups (61, 67, 69). Also, Keim *et. al.* proposed canSNPs to define the *B. anthracis* lineage that contains the Ames strain (67).The limitations of the present study include the relatively small sample size and the phylogenetic tree was built based on the *16S rRNA* gene only. Hence, further studies with large sample size and molecular techniques, such as multilocus sequence analysis (MLST) and whole genome sequencing (WGS), that are used to differentiate the closely related microbial species like *Bacillus* species are recommended. However, in this work, the SNPs analysis provided more intra-specific information than phylogenetic analysis in the cases of *B. subtilis* group. Eleven canSNPs were identified in strains that belong to *B. pumilus* groups when compared with *B. subtilis* reference sequences. In addition, these canSNPs in the conserved region of *16S rRNA* gene may provide important information for the design of primers and probes for Real-Time PCR, multiplex-PCR and microarray systems which is widely used for detection and typing purposes.

In conclusion, the phylogenetic tree diverged into two lineages; one belongs to *B. subtilis* group while the other belongs to *B. cereus* group. Interestingly, three of native strains (*SY27.1A, SY35.3A,* and *SY58.5A*), which produce Cry proteins, shared high similarity with *B. subtilis* group (over 99%). An 11 canSNPs were identified in strains that belong to *B. pumilus* groups when compared with *B. subtilis* reference sequences. These canSNPs in the conserved region of *16S rRNA* gene may provide important information for the design of primers and probes which is widely used for detection and typing purposes.

Supplementary file

Table S1. Information of the reference sequences of *Bacillus* species that were retrieved from NCBI databases.

Acknowledgment

We gratefully acknowledge the Genome and stem cell (GenKök) research center for their cooperation and supporting the study experiments.

Authors' contributions

Semih Yılmaz: Conceptualization, funding acquisition, methodology, writing - review & editing, and supervision; **Abeer Babiker Idris**: conceptualization, methodology, investigation, software, data curation and formal analysis, writing

- original draft; and writing - review & editing; **Abdurrahman Ayvaz:** funding acquisition, methodology and supervision; **Aysun Çetin:** methodology; **Funda Ülgen:** methodology and writing - original draft; **Mustafa Çetin:** funding acquisition, and supervision; **Berkay Saraymen:** methodology and data curation and formal analysis; **Mohamed A. Hassan:** conceptualization, methodology, writing - review & editing, and supervision.

Availability of data

All data generated or analyzed during this study are included in the manuscript.

Competing of interests

The authors have no competing of interests to declare.

Ethical Approval

This study was approved by Erciyes University, Faculty of Agriculture, Department of Agricultural Biotechnology.

Funding

This work was supported by Erciyes University Scientific Project Unit under the codes of FBD11-3634; and Ministry of Industrial Science and Technology, Turkey (TGSD-0802).

Consent of participants

Not applicable

409

410

411

412

413

414

415 **References**

416 1. Huang T-P, Tzeng DD-S, Wong ACL, Chen C-H, Lu K-M, Lee Y-H, et al.
417 DNA polymorphisms and biocontrol of Bacillus antagonistic to citrus bacterial
418 canker with indication of the interference of phyllosphere biofilms. PLoS One.
419 2012;7(7):e42124-e.

420 2. Liu Y, Lai Q, Dong C, Sun F, Wang L, Li G, et al. Phylogenetic diversity of
421 the Bacillus pumilus group and the marine ecotype revealed by multilocus
422 sequence analysis. PLoS One. 2013;8(11):e80097-e.

423 3. Radhakrishnan R, Hashem A, Abd Allah EF. Bacillus: A Biological Tool
424 for Crop Improvement through Bio-Molecular Changes in Adverse
425 Environments. Front Physiol. 2017;8:667-.

426 4. Garbeva P, van Veen JA, van Elsas JD. Predominant Bacillus spp. in
427 Agricultural Soil under Different Management Regimes Detected via PCR-
428 DGGE. Microbial Ecology. 2003;45(3):302-16.
429 5. Fu X, Gong L, Liu Y, Lai Q, Li G, Shao Z. Bacillus pumilus Group
430 Comparative Genomics: Toward Pangenome Features, Diversity, and Marine
431 Environmental Adaptation. Frontiers in microbiology. 2021;12(1084).
432 6. Parte AC, Sardà Carbasse J, Meier-Kolthoff JP, Reimer LC, Göker M. List
433 of Prokaryotic names with Standing in Nomenclature (LPSN) moves to the
434 DSMZ. International journal of systematic and evolutionary microbiology.
435 2020;70(11):5607-12.
436 7. Woese CR, Kandler O, Wheelis ML. Towards a natural system of
437 organisms: proposal for the domains Archaea, Bacteria, and Eucarya.
438 Proceedings of the National Academy of Sciences of the United States of
439 America. 1990;87(12):4576-9.
440 8. Van de Peer Y, Jansen J, De Rijk P, De Wachter R. Database on the
441 structure of small ribosomal subunit RNA. Nucleic Acids Research.
442 1997;25(1):111-6.
443 9. Maidak BL, Olsen GJ, Larsen N, Overbeek R, McCaughey MJ, Woese CR.
444 The Ribosomal Database Project (RDP). Nucleic Acids Research. 1996;24(1):82-5.
445 10. Drancourt M, Bollet C, Carlioz A, Martelin R, Gayral JP, Raoult D. 16S
446 ribosomal DNA sequence analysis of a large collection of environmental and

447 clinical unidentifiable bacterial isolates. Journal of clinical microbiology.
448 2000;38(10):3623-30.
449 11. Woese CR. Bacterial evolution. Microbiol Rev. 1987;51(2):221-71.
450 12. Janda JM, Abbott SL. 16S rRNA gene sequencing for bacterial
451 identification in the diagnostic laboratory: pluses, perils, and pitfalls. Journal of
452 Clinical Microbiology. 2007;45(9):2761-4.
453 13. Patel JB. 16S rRNA gene sequencing for bacterial pathogen identification
454 in the clinical laboratory. Molecular diagnosis. 2001;6(4):313-21.
455 14. Idris AB, Hassan HG, Ali MAS, Eltaher SM, Idris LB, Altayb HN, et al.
456 Molecular Phylogenetic Analysis of 16S rRNA Sequences Identified Two
457 Lineages of Helicobacter pylori Strains Detected from Different Regions in Sudan
458 Suggestive of Differential Evolution. International Journal of Microbiology.
459 2020;2020:8825718.
460 15. Fan B, Blom J, Klenk H-P, Borriss R. Bacillus amyloliquefaciens, Bacillus
461 velezensis, and Bacillus siamensis Form an “Operational Group B.
462 amyloliquefaciens” within the B. subtilis Species Complex. Frontiers in
463 microbiology. 2017;8.
464 16. Rooney AP, Price NP, Ehrhardt C, Swezey JL, Bannan JD. Phylogeny and
465 molecular taxonomy of the Bacillus subtilis species complex and description of
466 Bacillus subtilis subsp. inaquosorum subsp. nov. International journal of
467 systematic and evolutionary microbiology. 2009;59(Pt 10):2429-36.

17. Logan NA. Modern Methods for Identification. Applications and Systematics of Bacillus and Relatives2002. p. 123-40.

18. Guinebretière MH, Auger S, Galleron N, Contzen M, De Sarrau B, De Buyser ML, et al. Bacillus cytotoxicus sp. nov. is a novel thermotolerant species of the Bacillus cereus Group occasionally associated with food poisoning. International journal of systematic and evolutionary microbiology. 2013;63(Pt 1):31-40.

19. Punina NV, Zotov VS, Parkhomenko AL, Parkhomenko TU, Topunov AF. Genetic Diversity of Bacillus thuringiensis from Different Geo-Ecological Regions of Ukraine by Analyzing the 16S rRNA and gyrB Genes and by AP-PCR and saAFLP. Acta naturae. 2013;5(1):90-100.

20. Tetreau G, Andreeva EA, Banneville AS, De Zitter E, Colletier JP. How Does Bacillus thuringiensis Crystallize Such a Large Diversity of Toxins? Toxins. 2021;13(7).

21. Garbeva P, van Veen JA, van Elsas JD. Predominant Bacillus spp. in agricultural soil under different management regimes detected via PCR-DGGE. Microbial ecology. 2003;45(3):302-16.

22. Wilson MK, Abergel RJ, Raymond KN, Arceneaux JE, Byers BR. Siderophores of Bacillus anthracis, Bacillus cereus, and Bacillus thuringiensis. Biochemical and biophysical research communications. 2006;348(1):320-5.

488 23. Raddadi N, Cherif A, Boudabous A, Daffonchio D. Screening of plant
489 growth promoting traits of Bacillus thuringiensis. Annals of Microbiology.
490 2008;58:47-52.
491 24. Rabha M, Sharma S, Acharjee S, Sarmah BK. Isolation and characterization
492 of Bacillus thuringiensis strains native to Assam soil of North East India. 3
493 Biotech. 2017;7(5):303.
494 25. Sauka DH, Benintende GB. Diversity and distribution of lepidopteran-
495 specific toxin genes in Bacillus thuringiensis strains from Argentina. Revista
496 Argentina de microbiologia. 2017;49(3):273-81.
497 26. Nair K, Al-Thani R, Al-Thani D, Al-Yafei F, Ahmed T, Jaoua S. Diversity
498 of Bacillus thuringiensis Strains From Qatar as Shown by Crystal Morphology,
499 delta-Endotoxins and Cry Gene Content. Frontiers in microbiology. 2018;9:708.
500 27. Travers RS, Martin PA, Reichelderfer CF. Selective Process for Efficient
501 Isolation of Soil Bacillus spp. Applied and environmental microbiology.
502 1987;53(6):1263-6.
503 28. Baig DN, Mehnaz S. Determination and distribution of cry-type genes in
504 halophilc Bacillus thuringiensis isolates of Arabian Sea sedimentary rocks.
505 Microbiological research. 2010;165(5):376-83.
506 29. Centralblatt fu□r Bakteriologie, Parasitenkunde und
507 Infektionskrankheiten. 1. Abt. Originale. Jena :: G. Fischer; 1902.

508 30. Yılmaz S, Ayvaz A, Azizoğlu U. Diversity and distribution of cry genes in
509 native Bacillus thuringiensis strains isolated from wild ecological areas of East-
510 Mediterranean region of Turkey. Tropical Ecology. 2017;58:605-10.
511 31. YILMAZ S. Molecular characterization of Bacillus thuringiensis strains
512 isolated from various habitats and their use against some harmful insects. Erciyes
513 University Research Information System: Erciyes University; 2010.
514 32. Cerón J, Ortíz A, Quintero R, Güereca L, Bravo A. Specific PCR primers
515 directed to identify cryI and cryIII genes within a Bacillus thuringiensis strain
516 collection. Applied and environmental microbiology. 1995;61(11):3826-31.
517 33. Bravo A, Sarabia S, Lopez L, Ontiveros H, Abarca C, Ortiz A, et al.
518 Characterization of cry genes in a Mexican Bacillus thuringiensis strain
519 collection. Applied and environmental microbiology. 1998;64(12):4965-72.
520 34. FinchTV. 1.4.0 ed. USA: Geospiza, Inc.; Seattle, WA,; 2012.
521 35. Altschul SF, Madden TL, Schäffer AA, J Zhang ZZ, Miller W, Lipman DJ.
522 Gapped BLAST and PSI-BLAST. A new generation of protein database search
523 programmes. Nucleic Acids Res. 1997;25(17):3389–402.
524 36. Hall TA. BioEdit: a user-friendly biologicalsequence alignment editor
525 and analysis programfor Windows 95/98/NT. Nucl Acids Symp Ser
526 1999;41:95-8.
527 37. Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics
528 Analysis Version 7.0 for Bigger Datasets. Molecular biology and evolution.
529 2016;33(7):1870-4.

530 38. Castresana J. Selection of Conserved Blocks from Multiple Alignments for
531 Their Use in Phylogenetic Analysis. Molecular Biology and Evolution.
532 2000;17(4):540-52.
533 39. Saitou N, Nei M. The neighbor-joining method: a new method for
534 reconstructing phylogenetic trees. Molecular biology and evolution.
535 1987;4(4):406-25.
536 40. Goldman N. Maximum Likelihood Inference of Phylogenetic Trees, with
537 Special Reference to a Poisson Process Model of DNA Substitution and to
538 Parsimony Analyses. Systematic Biology. 1990;39(4):345-61.
539 41. Nei M, Kumar S. Molecular Evolution and Phylogenetics. USA: Oxford
540 University Press; 2000.
541 42. Felsenstein J. CONFIDENCE LIMITS ON PHYLOGENIES: AN
542 APPROACH USING THE BOOTSTRAP. Evolution; international journal of
543 organic evolution. 1985;39(4):783-91.
544 43. Patel JB. 16S rRNA gene sequencing for bacterial pathogen identification
545 in the clinical laboratory. Molecular diagnosis : a journal devoted to the
546 understanding of human disease through the clinical application of molecular
547 biology. 2001;6(4):313-21.
548 44. Stackebrandt E, Goebel BM. Taxonomic Note: a Place for DNA-DNA
549 Reassociation and 16s rRNA Sequence Analysis in the Present Species Definition
550 in Bacteriology1994.

551 45. Castillo-Esparza JF, Hernández-González I, Ibarra JE. Search for Cry
552 proteins expressed by Bacillus spp. genomes, using hidden Markov model
553 profiles. 3 Biotech. 2019;9(1):13-.
554 46. Barloy F, Lecadet MM, Delécluse A. Cloning and sequencing of three new
555 putative toxin genes from Clostridium bifermentans CH18. Gene.
556 1998;211(2):293-9.
557 47. Zhang J, Hodgman TC, Krieger L, Schnetter W, Schairer HU. Cloning and
558 analysis of the first cry gene from Bacillus popilliae. Journal of bacteriology.
559 1997;179(13):4336-41.
560 48. Yokoyama T, Tanaka M, Hasegawa M. Novel cry gene from Paenibacillus
561 lentimorbus strain Semadara inhibits ingestion and promotes insecticidal activity
562 in Anomala cuprea larvae. Journal of invertebrate pathology. 2004;85(1):25-32.
563 49. Jones GW, Nielsen-Leroux C, Yang Y, Yuan Z, Dumas VF, Monnerat RG,
564 et al. A new Cry toxin with a unique two-component dependency from Bacillus
565 sphaericus. FASEB journal : official publication of the Federation of American
566 Societies for Experimental Biology. 2007;21(14):4112-20.
567 50. Ben-Dov E, Zaritsky A, Dahan E, Barak Z, Sinai R, Manasherob R, et al.
568 Extended screening by PCR for seven cry-group genes from field-collected
569 strains of Bacillus thuringiensis. Applied and environmental microbiology.
570 1997;63(12):4883-90.
571 51. Feitelson JS, Payne J, Kim L. Bacillus thuringiensis: Insects and Beyond.
572 Bio / Technology. 1992;10(3):271-5.

573 52. Shishir A, Roy A, Islam N, Rahman A, Khan SN, Hoq MM. Abundance
574 and diversity of Bacillus thuringiensis in Bangladesh and their cry genes profile.
575 Frontiers in Environmental Science. 2014;2(20).
576 53. Rabha M, Acharjee S, Sarmah BK. Multilocus sequence typing for
577 phylogenetic view and vip gene diversity of Bacillus thuringiensis strains of the
578 Assam soil of North East India. World journal of microbiology & biotechnology.
579 2018;34(7):103.
580 54. Gassmann AJ, Shrestha RB, Kropf AL, St Clair CR, Brenizer BD. Field-
581 evolved resistance by western corn rootworm to Cry34/35Ab1 and other Bacillus
582 thuringiensis traits in transgenic maize. Pest management science. 2020;76(1):268-
583 76.
584 55. Liu Y, Du J, Lai Q, Zeng R, Ye D, Xu J, et al. Proposal of nine novel species
585 of the Bacillus cereus group. International journal of systematic and evolutionary
586 microbiology. 2017;67(8):2499-508.
587 56. Zothansanga R, Senthilkumar N, Gurusubramanian G. Diversity and
588 Toxicity of Bacillus thuringiensis from Shifting Cultivation (Jhum) Habitat.
589 Biocontrol science. 2016;21(2):99-111.
590 57. Helgason E, Caugant DA, Lecadet MM, Chen Y, Mahillon J, Lovgren A, et
591 al. Genetic diversity of Bacillus cereus/B. thuringiensis isolates from natural
592 sources. Current microbiology. 1998;37(2):80-7.
593 58. Gdoura-Ben Amor M, Siala M, Zayani M, Grosset N, Smaoui S, Messadi-
594 Akrout F, et al. Isolation, Identification, Prevalence, and Genetic Diversity of

595 Bacillus cereus Group Bacteria From Different Foodstuffs in Tunisia. Frontiers in
596 microbiology. 2018;9(447).
597 59. Gordon REHWCPCH-NSNR. The genus Bacillus. Washington, D.C.:
598 Agricultural Research Service, U.S. Dept. of Agriculture : For sale by Supt. of
599 Docs., U.S. G.P.O.; 1973. Available from:
600 http://books.google.com/books?id=lXMpXssWFD0C.
601 60. Kwok P-Y. Single Nucleotide Polymorphisms (SNPs): Identification and
602 Scoring. eLS.
603 61. Dimauro C, Cellesi M, Steri R, Gaspa G, Sorbolini S, Stella A, et al. Use of
604 the canonical discriminant analysis to select SNP markers for bovine breed
605 assignment and traceability purposes. Animal Genetics. 2013;44(4):377-82.
606 62. Griffing SM, MacCannell DR, Schmidtke AJ, Freeman MM, Hyytiä-Trees
607 E, Gerner-Smidt P, et al. Canonical Single Nucleotide Polymorphisms (SNPs) for
608 High-Resolution Subtyping of Shiga-Toxin Producing Escherichia coli (STEC)
609 O157:H7. PLoS One. 2015;10(7):e0131967.
610 63. Fernandez-No IC, Bohme K, Caamano-Antelo S, Barros-Velazquez J, Calo-
611 Mata P. Identification of single nucleotide polymorphisms (SNPs) in the 16S
612 rRNA gene of foodborne Bacillus spp. Food microbiology. 2015;46:239-45.
613 64. Pillidge CJ, Sheehy LM, Shihata A, Pu Z-Y, Dobos M, Powell IB.
614 Intragenomic 16S rRNA gene heterogeneity in Lactococcus lactis subsp. cremoris.
615 International Dairy Journal. 2009;19(4):222-7.

616 65. Hakovirta JR, Prezioso S, Hodge D, Pillai SP, Weigel LM. Identification
617 and Analysis of Informative Single Nucleotide Polymorphisms in 16S rRNA
618 Gene Sequences of the Bacillus cereus Group. Journal of clinical microbiology.
619 2016;54(11):2749-56.
620 66. Sacchi CT, Whitney AM, Mayer LW, Morey R, Steigerwalt A, Boras A, et
621 al. Sequencing of 16S rRNA gene: a rapid tool for identification of Bacillus
622 anthracis. Emerging infectious diseases. 2002;8(10):1117-23.
623 67. Keim P, Van Ert MN, Pearson T, Vogler AJ, Huynh LY, Wagner DM.
624 Anthrax molecular epidemiology and forensics: using the appropriate marker for
625 different evolutionary scales. Infection, genetics and evolution : journal of
626 molecular epidemiology and evolutionary genetics in infectious diseases.
627 2004;4(3):205-13.
628 68. Moorhead SM, Dykes GA, Cursons RT. An SNP-based PCR assay to
629 differentiate between Listeria monocytogenes lineages derived from
630 phylogenetic analysis of the sigB gene. Journal of microbiological methods.
631 2003;55(2):425-32.
632 69. Van Ert MN, Easterday WR, Simonson TS, U'Ren JM, Pearson T, Kenefic
633 LJ, et al. Strain-specific single-nucleotide polymorphism assays for the Bacillus
634 anthracis Ames strain. Journal of clinical microbiology. 2007;45(1):47-53.

635

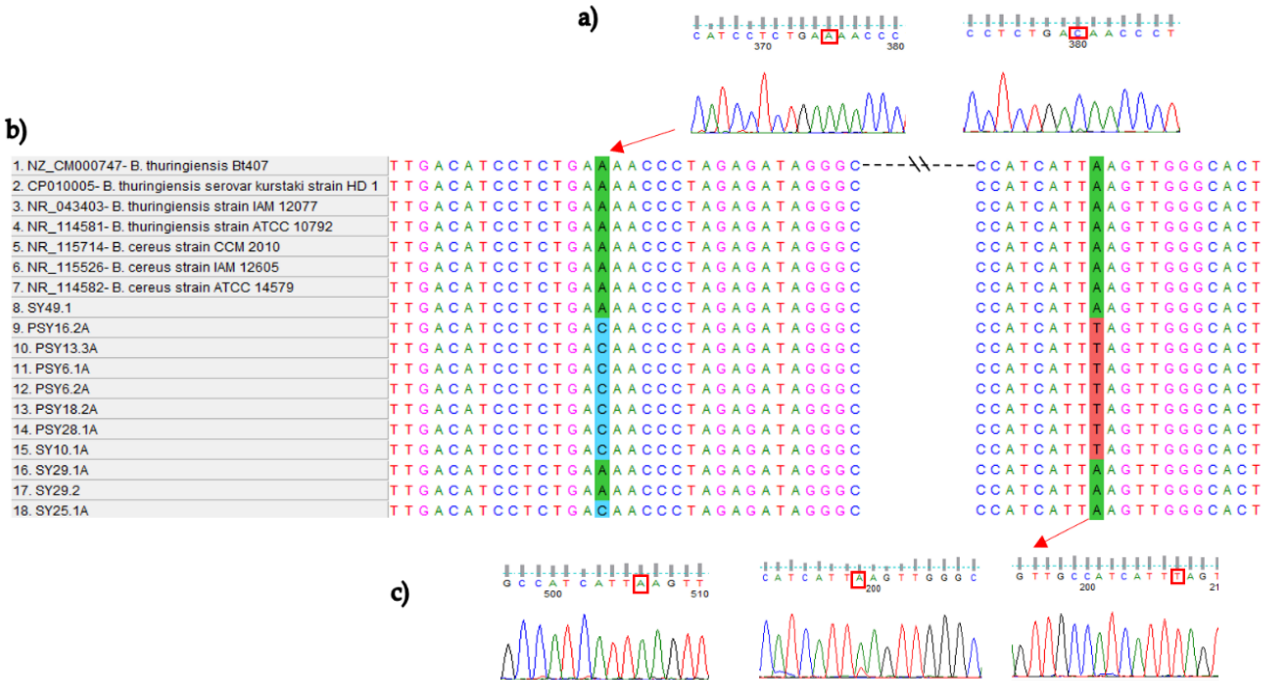

Figure 1. 1a) and 1b) Sequencing results of chromatograms using Finch TV software show nucleotide changes in the *16S rRNA* gene of *B. cereus* / *B. thuringiesis* strains illustrated by squares. 1c) Multiple sequences alignment of the native *Bacillus* strains that belong to *B. cereus* group using ClustalW.

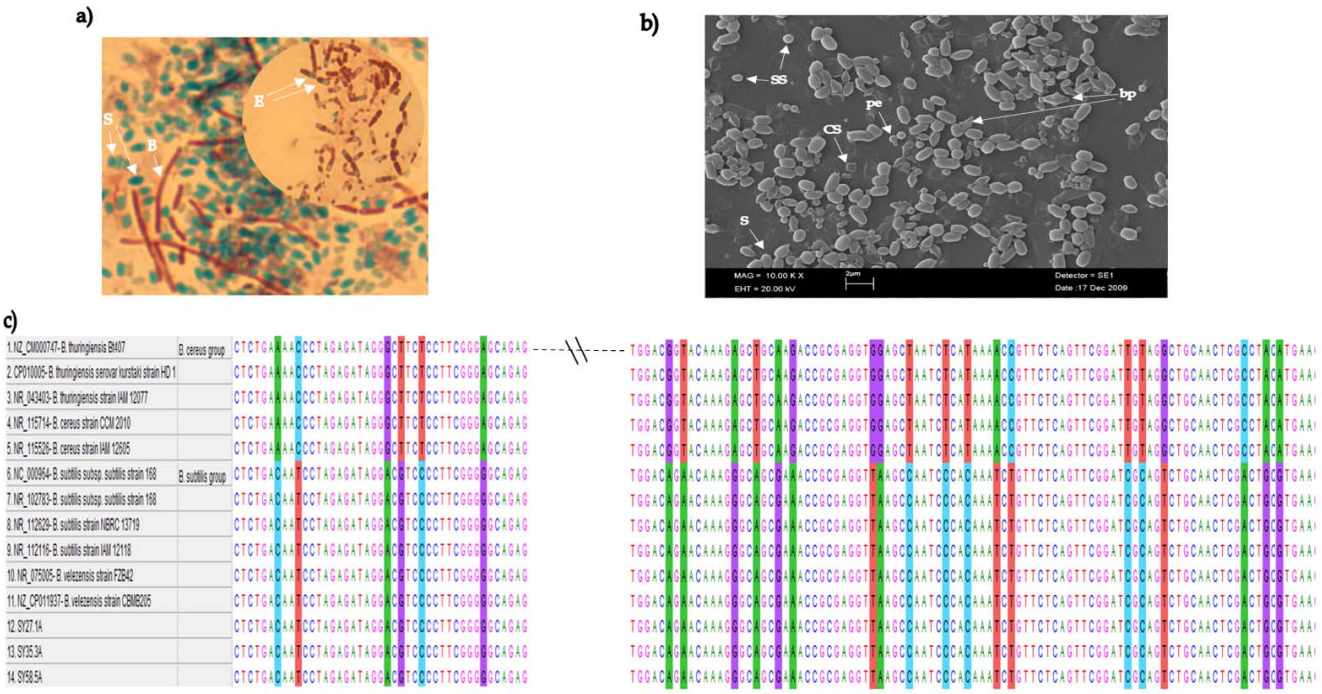

Figure 2. 2a) Microscopic view of native *Bacillus* strains after spore staining (1000x). B, *Bacillus* bacteria; S, spores; E, endospores. 2b) Scanning electron microscopy (SEM) image of the different types of crystal morphologies and the spores (S) produced by native *Bacillus* strain *SY27.1A* isolated from soil sample. SS, smooth spherical crystal; pe, spherical crystal with pointy edges; bp, bipyramidal crystal; cs, cuboidal crystal. 2c) Multiple Sequence Alignment (MSA) of *16S rRNA* sequences of the three native strains (*SY27.1A, SY35.3A*, *SY58.5A*) shows high similarity with *B. subtilis* group rather than species of *B. cereus* group.

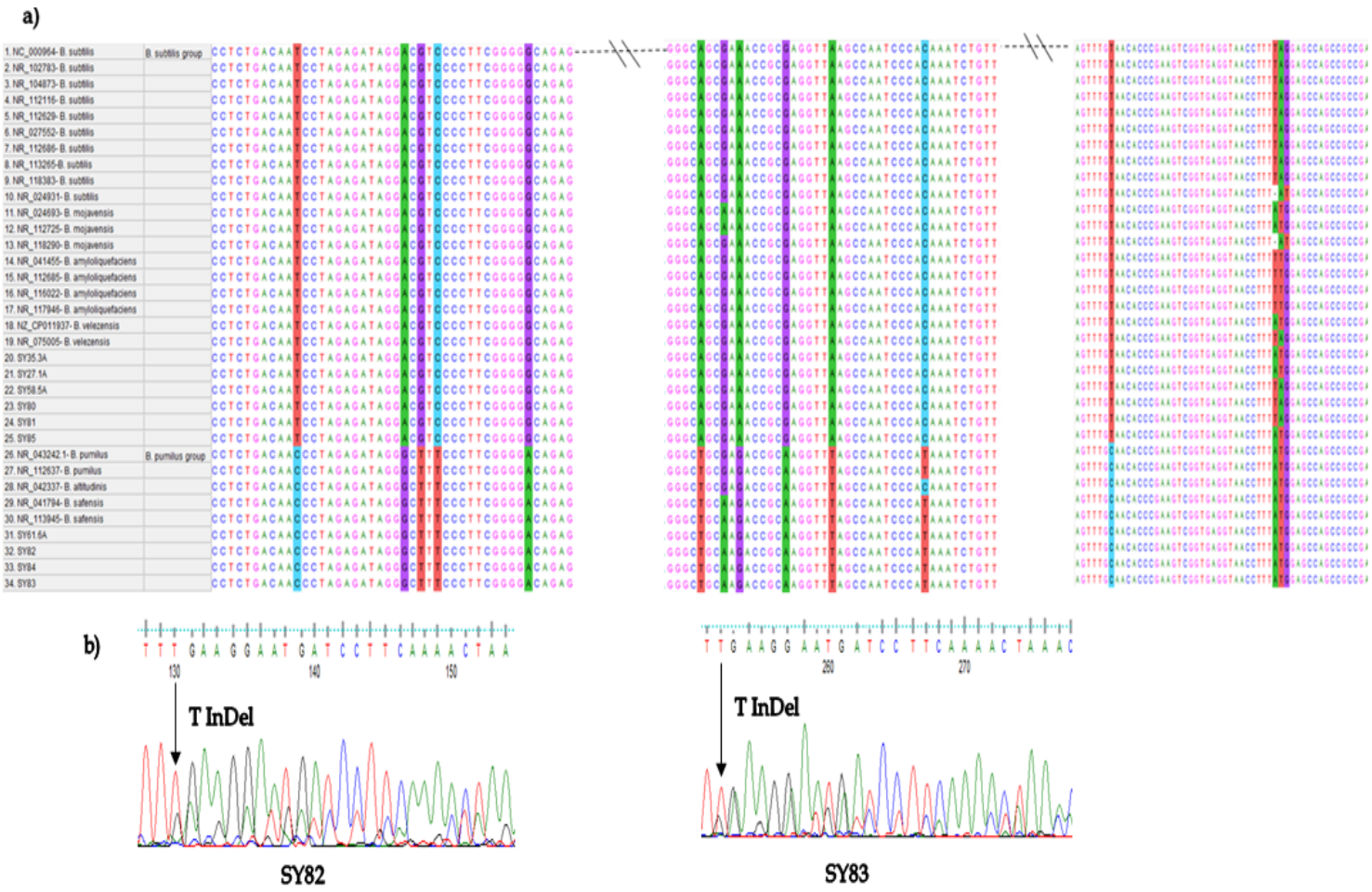

Figure 3. 3a) Multiple sequences alignment (MSA) of *B. subtilis* group shows 11 canonical SNPs (canSNPs) in strains that belong to *B. pumilus* groups when compared with *B. subtilis* reference sequences. 3b) Sequencing results of chromatograms illustrate the loss of synchronicity in *SY82* and *SY83* strains due to a nucleotide deletion or insertion within one or more of the *16S rRNA* genes among the multiple *rRNA* operons in the genome.

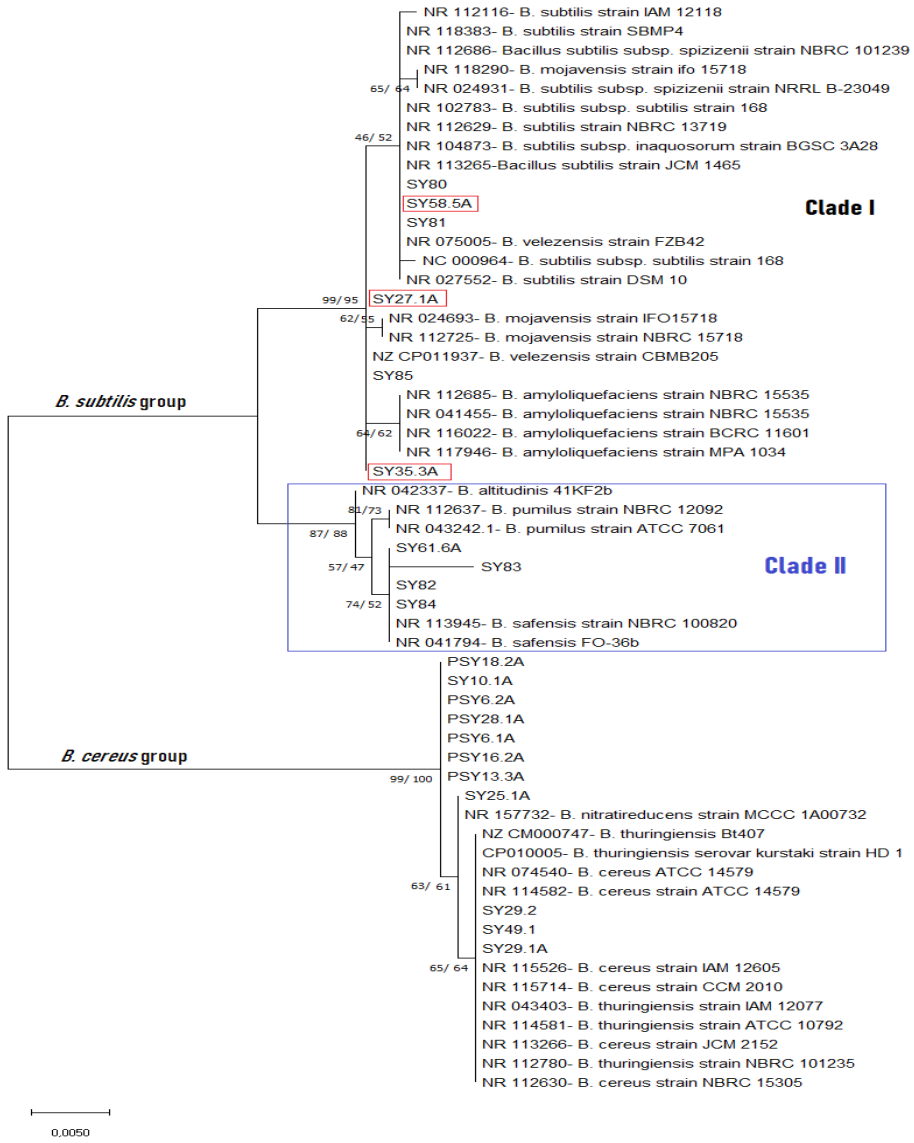

Figure 4. Maximum-Likelihood Phylogenetic tree of native *Bacillus* strains isolated from Turkey. The percentage of replicate trees (1000 replicates) are shown next to the branches from both NJ (before the slash '/') and ML (after the slash '/') analyses. The ML tree is only shown here, because the ML tree was very similar to the NJ tree. The evolutionary distance was computed using (K2+G) model and evolutionary analyses were conducted using MEGA7.

Highlights

- Isolation and characterization of native *Bacillus* strains from agricultural soil should receive a good attention because of their wide potential biological products with immense applications.
- Identification of three native strains (*SY27.1A*, *SY35.3A*, and *SY58.5A*), which produce Cry proteins, shared high similarity with *B. subtilis* group (over 99%) and less than 95% similarity with known *B. thuringiensis* and other species of *B. cereus* group.
- Eleven canonical SNPs (canSNPs) were detected in strains that belong to *B. pumilus* group when compared with *B. subtilis* reference sequences.
- The SNP at nucleotide position 1146 is not necessarily reliable for identifying *B. anthracis*, while G at position 1139, which is another SNP proposed to be unique to *B. anthracis*, appears to be more reliable for identifying *B. anthracis*.
- Phylogenetic analysis of *16S rRNA* sequences was found valuable for differentiation between *Bacillus* species isolated from soil samples.