data

MDPI

Data Descriptor

Draft Genome Sequencing of the *Bacillus thuringiensis* var. Thuringiensis Highly Insecticidal Strain 800/15

Anton E. Shikov [1,2,†], Iuliia A. Savina [1,†], Maria N. Romanenko [1,2], Anton A. Nizhnikov [1,2] and Kirill S. Antonets [1,2,*]

[1] All-Russia Research Institute for Agricultural Microbiology, 196608 St. Petersburg, Russia
[2] Faculty of Biology, St. Petersburg State University, 199034 St. Petersburg, Russia
* Correspondence: k.antonets@arriam.ru
† These authors contributed equally to this work.

Abstract: The *Bacillus thuringiensis* serovar thuringiensis strain 800/15 has been actively used as an agent in biopreparations with high insecticidal activity against the larvae of the Colorado potato beetle *Leptinotarsa decemlineata* and gypsy moth *Lymantria dispar*. In the current study, we present the first draft genome of the 800/15 strain coupled with a comparative genomic analysis of its closest reference strains. The raw sequence data were obtained by Illumina technology on the HiSeq X platform and de novo assembled with the SPAdes v3.15.4 software. The genome reached 6,524,663 bp. in size and carried 6771 coding sequences, 3 of which represented loci encoding insecticidal toxins, namely, Spp1Aa1, Cry1Ab9, and Cry1Ba8 active against the orders Lepidoptera, Blattodea, Hemiptera, Diptera, and Coleoptera. We also revealed the biosynthetic gene clusters responsible for the synthesis of secondary metabolites, including fengycin, bacillibactin, and petrobactin with predicted antibacterial, fungicidal, and growth-promoting properties. Further comparative genomics suggested the strain is not enriched with genes linked with biological activities implying that agriculturally important properties rely more on the composition of loci rather than their abundance. The obtained genomic sequence of the strain with the experimental metadata could facilitate the computational prediction of bacterial isolates' potency from genomic data.

Dataset: The raw genome sequencing data were submitted to the NCBI with BioSample number SAMN38204081, under BioProject PRJNA1039177. The assembled genome is available in the NCBI Assembly database under ASM3376411v1.

Keywords: *Bacillus thuringiensis*; biovar Thuringiensis; *Bt*; biopreparations; draft genome; Illumina; insecticidal activity

Citation: Shikov, A.E.; Savina, I.A.; Romanenko, M.N.; Nizhnikov, A.A.; Antonets, K.S. Draft Genome Sequencing of the *Bacillus thuringiensis* var. Thuringiensis Highly Insecticidal Strain 800/15. *Data* **2024**, *9*, 34. https://doi.org/10.3390/data9020034

Academic Editor: Pufeng Du

Received: 21 November 2023
Revised: 19 January 2024
Accepted: 1 February 2024
Published: 10 February 2024

1. Summary

Bacillus thuringiensis (*Bt*) is a gram-positive spore-forming bacterium that is ubiquitously distributed in a variety of ecosystems, predominantly in soil [1,2]. *Bt* is commonly known for producing a range of pesticidal proteins, including Cry, Cyt, Vip, and Sip toxins with considerable host specificity [3]. Distinct *Bt* isolates are also capable of synthesizing secondary metabolites with bactericidal, fungicidal, nematocidal, and plant growth-promoting activities [3]. Since *Bt* strains exhibit multifunctional activities and are considered environmentally friendly [4], they have found a worldwide application as pest control agents making up the vast majority of biological preparations [5]. However, the limited number of *Bt* strains used in agriculture may spark the emergence of resistant populations of insects and other plant pathogens as well. The *Bacillus thuringiensis* serovar thuringiensis strain 800/15 has been shown to be effective against various insects and has been patented as a

biocontrol agent [6]; however, its genome has not previously been sequenced and analyzed. Moreover, experimentally verified metadata for publicly available genomes remain limited. Either sequences in databases lack concomitant descriptions of their biological properties or they are extensively studied and/or patented but their genome sequences are absent. In this regard, genomic research aimed at searching for genetic determinants that delineate agriculturally important properties can assist in the accurate prediction of strains' activities directly from genomic data easing the selection of new strains based on examining their genome sequences.

2. Data Description

2.1. Isolation and Morphology Characterization of Bt Strain 800/15

The 800/15 strain was isolated from the dead larvae of the Colorado potato beetle (*Leptinotarsa decemlineata*) collected in the Leningrad region (Russia) and deposited in the joint Russian Collection of Agricultural Microorganisms (RCAM) at the All-Russia Research Institute for Agricultural Microbiology in Saint Petersburg (http://62.152.67.70/cryobank/login.jsp, accessed on 10 November 2023) on 20 September 2006 under the registration number 611 [6].

According to the classification of H. De Barjac and A. Bonnefoi, the strain was identified as *B. thuringiensis* serovar thuringiensis serotype H1 [6]. The Serological attribution was confirmed by molecular systematics methods based on 16SrRNA and *gyrB* sequences [7]. After 24 h of cultivation at +28 °C on the LB (Luria–Bertani) Broth, Miller (VWR International Ltd., Poole, UK), the strain forms flat, matte, grayish-white, rounded colonies with undulate margins (Figure 1a). On the second day of cultivation at +28 °C on the CCY nutrient medium [8], rod-shaped vegetative cells coupled with oval spores and bipyramidal crystals are observed indicating the onset of the transition from the vegetative stage to the sporulating one (Figure 1b). According to the patent data, the strain can also be cultivated on solid meat peptone and fish agar media as well as on liquid yeast polysaccharide media, including a yeast extract tryptone medium [7].

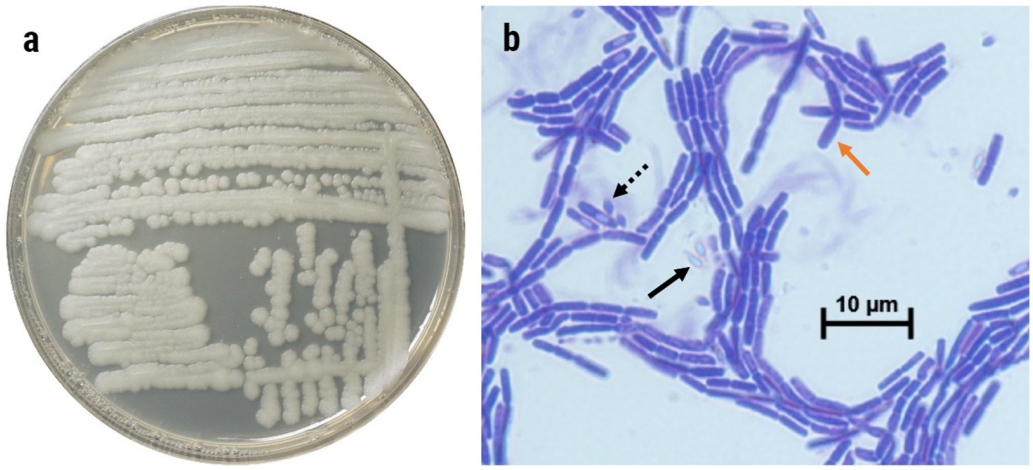

Figure 1. The morphology of the 800/15 strain's colonies after one day of cultivation on LB nutrient medium (**a**) and transition from vegetative to sporulating culture after two days of cultivation on CCY [8] nutrient medium stained with Coomassie brilliant blue (100× objective) (**b**). The black solid arrow shows the spore, the black dotted arrow—bipyramidal crystals of toxins, and the orange solid arrow—the vegetative cell.

2.2. Entomocidal, Fungicidal, and Plant Growth-Promoting Properties of Bt Strain 800/15

The 800/15 strain was patented as the major component for preparing the entomocidal biopreparation (Russian Federation patent No. 2514211 dated 04.27.2014). The strain has high toxicity against the second instar larvae of the Colorado potato beetle (*Leptinotarsa decemlineata*, Coleoptera) and gypsy moth (*Lymantria dispar*, Lepidoptera) [6]. In field

conditions, when treating crops in various geographical areas of Russia, the biopreparation exerted considerable toxic effects on the red spider mite (*Tetranichus urtica*, Trombidiformes), glasshouse whitefly (*Trialeurodes vaporariorum*, Hemiptera), diamondback moth (*Plutella xylostella*, Lepidoptera), large white (*Pieris brassicae*, Lepidoptera), small white (*P. rapae*, Lepidoptera), cabbage moth (*Barathra brassicae*, Lepidoptera), gooseberry sawfly (*Pteronidea ribesii*, Hymenoptera), potato ladybird (*Henosepilachna vigintioctomaculata*, Coleoptera), and house fly (*Musca domestica*, Diptera) [7].

The strain inhibits the growth of fungi, including *Botrytis cinerea*, *Bipolaris sorokiniana*, *Verticillium dahlia*, and *Pythium* spp., by 18–54% and does not affect *Fusarium avenacrum* and *Fusarium solani* [7].

The 800/15 strain-containing preparation demonstrated plant growth-promoting properties, including an increase in seed germination, seedling height, and root length of the tomato (*Solanum lycopersicum*), cucumber (*Cucumis sativum*), zucchini (*Cucurbita pepo* serovar *giromontina*), beets (*Beta vulgaris*), pumpkins (*Cucurbita pepo*), and cabbage (*Brassica oleraceae*) [7].

On that account, studying the genome of the 800/15 strain is useful for detecting molecular factors contributing to its polyfunctional properties.

2.3. Genome Assembly and Annotation

According to the quality control procedure, the sequencing data used for de novo genome assembly were of high quality. The sequencing coverage averaged over contigs reached 773. The draft genome obtained with the SPAdes v3.15.4 tool [9] consisted of 423 contigs with a total size of 6,524,663 b.p. and GC-content of 34.84%. As revealed by CheckM v1.2.2 [10], the contamination level was less than 1% indicating the high quality of the resulting assembly. Other main characteristics of the draft genome are presented in Table 1.

Table 1. The main properties of the draft genome assembly of the *Bt* strain 800/15 obtained using QUAST v5.2.0 [11] and CheckM v.1.2.2 [10].

Genome Characteristics	Value
Genome coverage	773
Total amount of contigs	423
Total length (b.p.)	6,524,663
Largest contig (b.p.)	776,444
GC-content (%)	34.84
N50 value	159,667
L50 value	11
Assembly completeness (%)	99.34
Suspected contamination (%)	0.15

Using the BUSCO v5.4.2 program [12] with both the Bacillales_odb10 and Bacilli_odb10 databases, we determined that the genome assembly contained at least 99.5% of fully assembled single-copy orthologues (Table 2). Therefore, the results further prove the high quality and completeness of the genome assembly.

Table 2. The percentage of the BUSCO v5.4.2 [12] markers present in the protein-coding genes found in the assembly.

Database	Bacillales_odb10	Bacilli_odb10
Single-copy orthologues assembled completely	448 (99.5%)	301 (99.7%)
Orthologues present in one copy	443 (98.4%)	299 (99.0%)
Multi-copies orthologues	5 (1.1%)	2 (0.7%)
Fragmented sequences	0 (0.0%)	0 (0.0%)
Orthologues missing from the assembly	2 (0.5%)	1 (0.3%)
Total number of single-copy orthologues in the database	450	302

Next, we applied the fastANI v1.33 tool [13] to obtain average nucleotide identity (ANI) values to identify the 10 most similar genome assemblies from the Bacillaceae family in the NCBI RefSeq database [14]. The Btyper v3.4.0 utility [15] taxonomically assigned the *Bt* strain 800/15 and selected reference genomes to the group IV of *B. cereus sensu lato* (Table 3). According to the recently proposed classification, this group, now called *B. cereus sensu stricto*, encompasses the vast majority of biovar Thuringiensis isolates, including the type serovar Berliner strain ATCC 10792 [16].

Table 3. The list of the phylogenetically closest assemblies relative to the genome of the strain 800/15 according to the ANI (average nucleotide identity) values calculated with the fastANI v1.33 software [13]. Taxonomic assignment was attributed using the Btyper v3.4.0 utility [15].

NCBI RefSeq Assembly	Strain	ANI	Btyper3 Taxonomy
GCF_033071735.1	S908	99.8215	*B. cereus s.s.*biovar Thuringiensis
GCF_031337245.1	S1307	99.8465	*B. cereus s.s.*biovar Thuringiensis
GCF_017165575.1	S601	99.8275	*B. cereus s.s.*biovar Thuringiensis
GCF_002912115.1	Bt407	99.9022	*B. cereus s.s.*
GCF_002574115.1	AFS057829	99.8302	*B. cereus s.s.*biovar Thuringiensis
GCF_000341665.1	IS5056	99.8639	*B. cereus s.s.*biovar Thuringiensis
GCF_000306745.1	407	99.9176	*B. cereus s.s.*
GCF_000193355.1	CT-43	99.8311	*B. cereus s.s.*biovar Thuringiensis
GCF_000161495.1	Bt407	99.8506	*B. cereus s.s.*
GCF_000571955.1	NA205-3	99.8317	*B. cereus s.s.*biovar Thuringiensis

Using the Prokka v1.14.6 tool [17], we found a total of 6827 genes in the genome of the studied strain (Supplementary Data S1). Of these, 6771 are coding sequences with 3723 coding for hypothetical proteins. Furthermore, BtToxin_Digger v1.0.10 [18] identified loci in the genome that encode insecticidal toxins, specifically Spp1Aa1, Cry1Ab9, and Cry1Ba8, which have been found to exert an effect on a wide range of insects from the orders Lepidoptera, Blattodea, Hemiptera, Diptera, and Coleoptera (Table 4).

Table 4. The insecticidal toxins identified in the analyzed genome using the BtToxin_Digger v1.0.10 program [18]. The target species describes experimentally derived data deposited in the BPPRC (Bacterial Pesticidal Protein Resource Center) specificity database [19].

Toxin	Percent of Identity	Target Order	Target Species
Spp1Aa1	80.3	Lepidoptera	*Spodoptera litura*
		Blattodea	*Blattella germanica*
Cry1Ab9	100	Lepidoptera	*Anticarsia gemmatalis, Chilo partellus, Choristoneura fumiferana, Chrysodeixis includens, Conogethes punctiferalis, Ephestia kuehniella, Epinotia aporema, Helicoverpa armigera, Lymantria dispar, Mythimna separata, Ostrinia nubilalis, Pectinophora gossypiella, Plodia interpunctella, Plutella xylostella, Sesamia inferens, Spodoptera exigua, Spodoptera frugiperda*
		Hemiptera	*Acyrthosiphon pisum, Diaphorina citri, Nilaparvata lugens*
Cry1Ba8	100	Lepidoptera	*Chilo suppressalis, Spodoptera frugiperda, Epinotia*
		Hemiptera	*Diaphornia citri*
		Diptera	*Musca domestica*
		Coleoptera	*Xylotrechus arvicola, Chrysomela scripta, Acanthoscelides obtectus*

The *Bt* strain 800/15′s genome contained 48 biosynthetic gene clusters (BGCs) revealed by the DeepBGC v0.1.30 tool [20]. The antiSMASH v6.1.1 tool [21], in turn, identified 12 specific BGCs listed in Table 5. It is noteworthy that within these BGCs, petrobactin exhibited a 100% similarity to the known entity, while bacillibactin and fengycin showed a 46% and 40% similarity, respectively.

Table 5. Selected biosynthetic gene clusters revealed in the analyzed genome. The score reflects the accuracy of the prediction reported by DeepBGC v0.1.30 [20], while the similarity to the known clusters is calculated with the antiSMASH v6.1.1 [21] program. The "-" symbol in the last three columns indicates that the biosynthetic cluster was found by only one program or that the properties of the cluster were absent. The total length of the regions is given in brackets after the genomic coordinates.

Contig	Tool	Type/Activity	Location (Relative Coordinate, b.p.)	Most Similar Known Cluster	Similarity	Score
4	antiSMASH	Siderophore	55,618-69,325 (13,708)	Petrobactin	100%	-
8	antiSMASH	NRPS-like	95,304-138,885 (43,582)	-	-	-
11	antiSMASH	NRPS	49,700-96,857 (47,158)	Bacillibactin	46%	-
23	antiSMASH	Terpene	19,851-41,704 (21,854)	Molybdenum Cofactor	17%	-
26	antiSMASH	RiPP-like	5610-15,876 (10,267)	-	-	-
35	antiSMASH	RRE-containing	18,991-40,094 (21,104)	-	-	-
42	antiSMASH	Betalactone	1-16,063 (16,063)	Fengycin	40%	-
6	antiSMASH/DeepBGC	LAP, RiPP-like	90,297-113,803 (23,507)	-	-	0.80372
17	antiSMASH/DeepBGC	NRPS/antibacterial	57,052-104,062 (47,011)	-	-	0.77436
19	antiSMASH/DeepBGC	NRPS/antibacterial	21,439-87,344 (65,906)	-	-	0.93123
32	antiSMASH/DeepBGC	RiPP-like/antibacterial	6249-16,578 (10,330)	-	-	0.88864
49	antiSMASH/DeepBGC	NRPS/antibacterial	1-19,638 (19,638)	-	-	0.71362
1	DeepBGC	antibacterial	547,383-563,304 (15921)	-	-	0.88938
2	DeepBGC	Saccharide/antibacterial	181,696-199,535 (17839)	-	-	0.81677
5	DeepBGC	Saccharide/antibacterial	16,626-34,094 (17468)	-	-	0.90965
8	DeepBGC	Saccharide/antibacterial	181,609-198,845 (17236)	-	-	0.91004
26	DeepBGC	antibacterial	64,334-77,440 (13106)	-	-	0.89632

Based on the genomic information obtained, it can be inferred that the strain has insecticidal properties along with potential bactericidal, fungicidal, and plant growth-promoting ones.

2.4. Comparative Genomic Analysis

We then compared our strain with similar assemblies. Notably, three of these genomes were representatives of the acrystalliferous Bt407 (also known as 407), the widely used strain derived from a wild-type isolate subjected to a high temperature [22]. While our strain was isolated from the dead larva, it groups with soil isolates (Table S1), as demonstrated by the ANI-base whole-genome clusterization (Figure 2a). The metadata from the BioSample database [23] coupled with the existing research articles (Table S2) indicated that most of the strains were toxic to the Lepidoptera order, whereas one isolate (S1307) belonging to one clade with 800/15 was active against Diptera and Coleoptera as well [24]. To reveal whether and how the assembly obtained differs from the closest genomes, we searched for insecticidal toxins and virulence factors from the VFDB (Virulence Factors Database) [25] and predicted pesticidal activity for these strains accordingly (Table S3). When considering the fourth rank of the existing nomenclature [19], the strain 800/15 possessed Cry1Ab1 (Figure 2b). Reduction to the third rank (Figure 2c) coupled with the summary of predicted insecticidal activities (Figure 2d) and homologs of known virulence factors (Figure 2e) indicated that strain 800/15 shared certain properties with at least one reference genome. We then compared the sets of identified homologs of insecticidal toxins and known virulence determinants between the strains applying the Jaccard coefficient and found that the strain Bt407 was the most similar to 800/15 in terms of the composition of these factors (Table S4).

Having shown the overall shared set of genomic determinants, we further inspected the total amount and distribution of these factors in a phylogeny-wise way. The sequenced strain was not enriched in insecticidal loci (Figure 3a) and did not show a higher number of putatively affected species (Figure 3b). At the same time, the composition of pesticidal toxins in our strain was unique (Figure 3d). It is noteworthy that all but two strains (S1307 and S601) possessed Spp1Aa (Table S3). All the genomes were substantially similar in terms of the abundance of known virulence determinants spanning from 106 to 116

(Figure 3c, Table S5). The tree-wise distribution patterns were generally consistent with the formed clades (Figure 3d–g). Nevertheless, there were noticeable inconsistencies even between the closest strains, such as S1307 and IS5056 or S908 and S601 (Figure 3d,e). This observation corroborates the fact that the distribution of determinants in *Bt* does not reflect taxonomic units and serological attributions [26]. Of all the homologs from VFDB, 98 were found in all the strains, and their abundance did not correlate with the number of loci coding for insecticidal moieties. Therefore, they can be considered core genes responsible for general virulence mechanisms but not the specificity which was shown for *Serratia marcescens* exhibiting insecticidal activity as well [27]. Despite the strain 800/15 being theoretically expected to infect a lower range of species (Figure 3f,g), it was efficient against various insects from multiple orders as revealed by experimental tests. We propose that such an inference might be explained by the quality but not the quantity of the insecticidal toxins. For instance, it was shown that the Cry1Ba and Cry1Ab that our strain produces work synergistically increasing the toxicity against *Chilo suppressalis* (Lepidoptera) 11 times [28]. That being said, one might expect that higher amounts of insecticidal genes could sometimes decrease toxicity through antagonistic effects. Therefore, an optimal combination of pesticidal determinants delineates the potency of our strain.

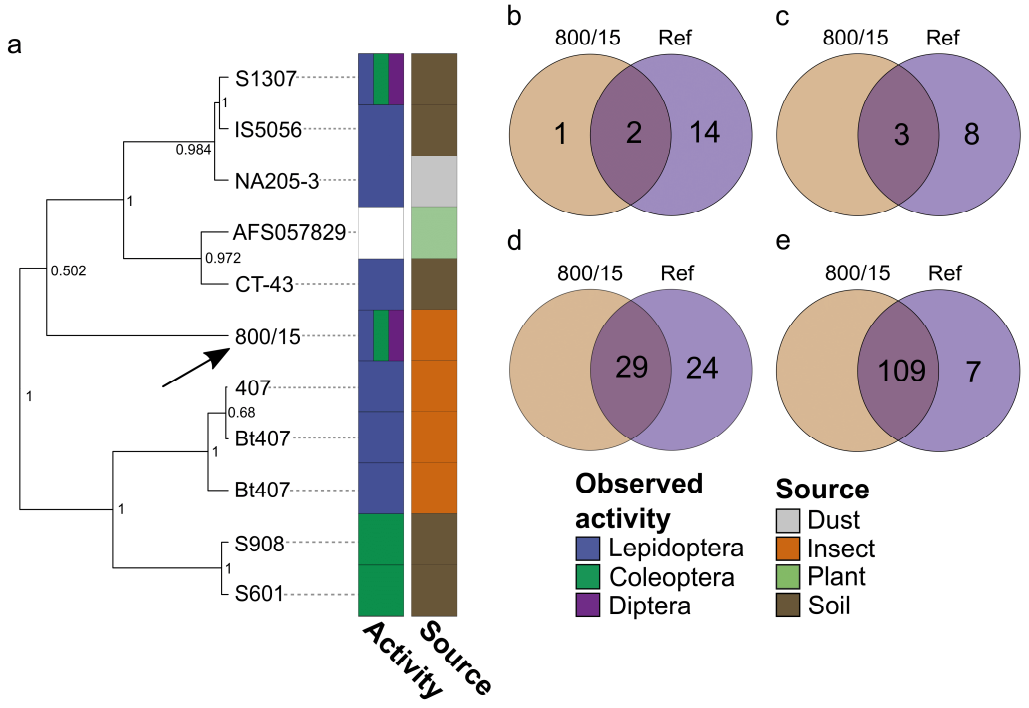

Figure 2. Comparisons between the studied strains in terms of genomic similarity and composition of insecticidal toxins and virulence determinants. (**a**) The bootstrapped hierarchal clustering tree of the genomes is based on the pair-wise ANI (average nucleotide identity) estimates. The numbers near branches correspond to support values, and the black arrow points to the strain 800/15. Adjacent strips are colorized according to experimentally verified insecticidal activities and isolation sources, respectively. (**b**) Shared and non-common insecticidal toxins within the fourth and third ranks (**c**) of the current structure-based nomenclature, predicted host species (**d**), and known virulence determinants from VFDB (Virulence Factors Database) [25] (**e**) within the strain 800/15 and closest reference genomes. Non-shared values imply that the respective homologs or host species are absent in the genomes of reference strains or strain 800/15.

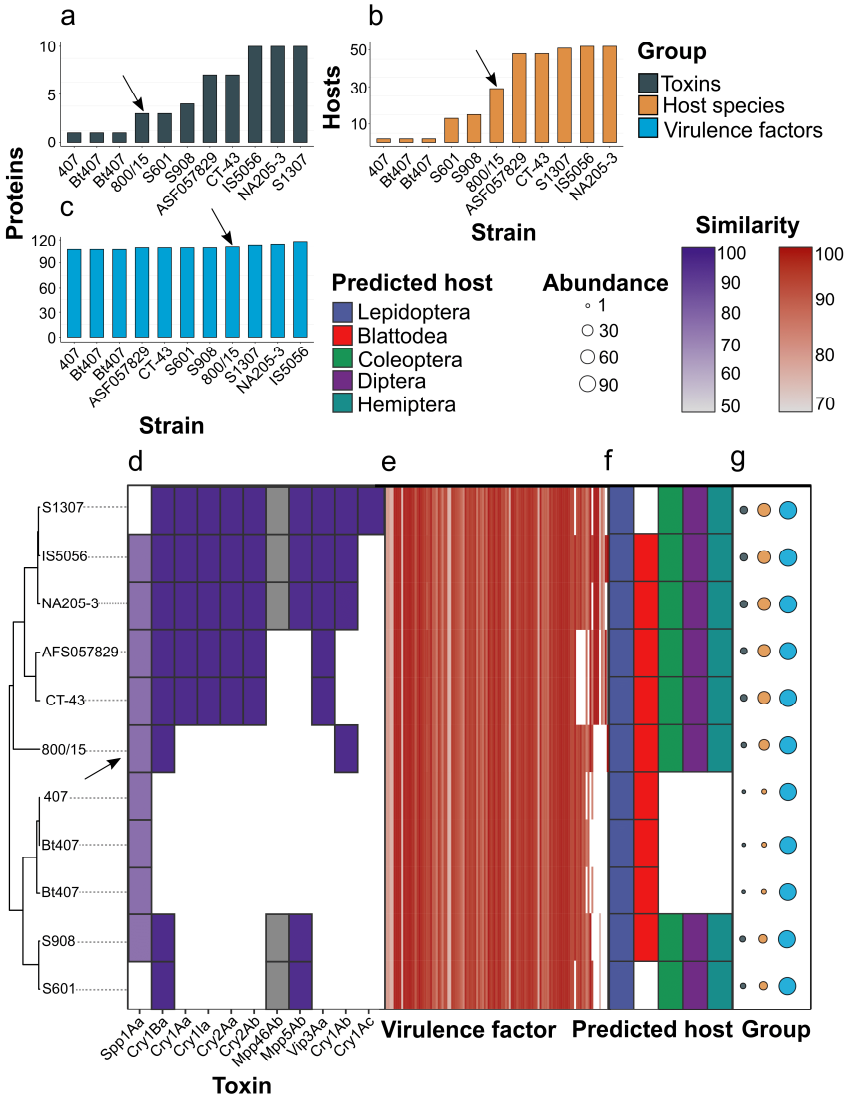

Figure 3. The overall number and phylogeny-wise distribution of genomic determinants and theoretically predicted pesticidal activities of the analyzed strains. (**a**) The total amount of insecticidal toxins, theoretically predicted sets of affected species (**b**), and homologs of well-studied virulence determinants (**c**) from VFDB (Virulence Factors Database) [25]. The arrow points to strain 800/15. (**d**) The heatmap displaying the distribution of insecticidal toxins and virulence factors (**e**), homologs, as well as the number of putative host species (**f**). For the distribution of homologs, the intensity of the color corresponds to the identity with the closest hit from the respective database. (**g**) The sum of insecticidal toxins, virulence determinants, and predicted host species found in the studied strains.

Since insecticidal loci are often located in plasmids, we classified contigs as fragments of plasmidic/chromosomal DNA. The predictions showed that all of the genes, except for *spp1Aa* in all of the isolates, were found on plasmids (Table S6). This observation explains a considerable diversity in terms of gene content indicating possible plasmid loss and acquisition. We then screened the studied genomes for the presence of diverse genomic loci, including mobile genomic elements (Table S7). First, we found that the strain 800/15 contains three CRISPR (clustered regularly interspaced short palindromic repeats) sequences which is close to the average among all the assemblies (Figure 4a). Notably, it harbored the lowest number of genetic islands (5) in contrast to other genomes (Figure 4b). Nevertheless, the total abundance of insertion sequences (51) was similar to the average (Figure 4c), and the strain was enriched with prophages (6) as well (Figure 4d).

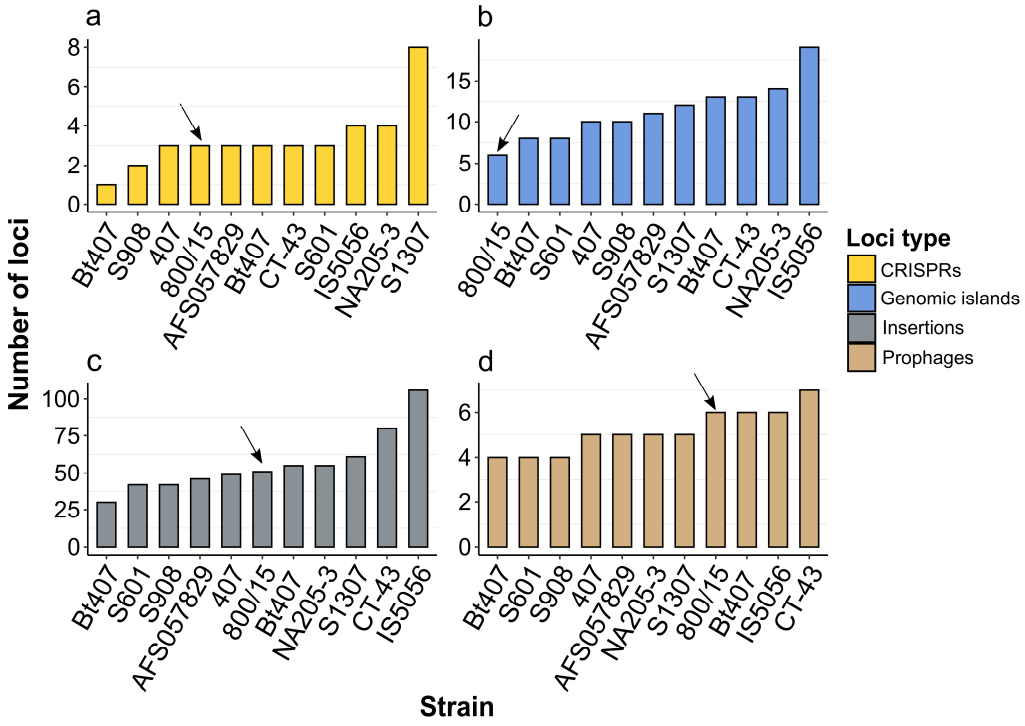

Figure 4. The total number of predicted CRISPR (clustered regularly interspaced short palindromic repeats) sequences (**a**), and mobile genetic elements, namely, genomic islands (**b**), insertion sequences (**c**), and prophages (**d**) presented in the *B. thuringiensis* strain 800/15 genome and the closest reference genome assemblies. Black arrows point to the bars related to the studied strain.

Next, we attempted to examine which functional features distinguished strain 800/15 from the reference isolates. In this regard, we carried out the over-representation tests using the topGO v.3.15 [29] package with all sets of protein sequences defined as the background and strain-specific proteins perceived as the target set (Table S8). The enrichments in biological processes (Figure 5a) highlight DNA-related processes, possibly associated with the dispersal of mobile genetic elements (MGEs). Another noticeable group of enrichments belonged to the cell wall and the development of bacterial cells. The two remaining categories were compliant with reported biological processes. For instance, cellular components in which the proteins of interest reside are related to cellular surfaces, including endospores and flagellum (Figure 5b). Other terms were the synthesis of amino acids, translation, and transcription regulation. Over representations within the molecular function annotation system further showed the role of transcription modulation, and the cell-to-cell contact, namely, mannose and fibronectin binging (Figure 5c). The results are consistent with the current views on the genetic control of pathogenesis and host specificity. Forasmuch as insecticidal loci in *Bt* genomes are frequently embedded into MGEs, their dissemination orchestrates the evolution of *Bt* strains when adapting to certain hosts [30]. The prominent role of transcriptional regulation was demonstrated for a plethora of bacterial pathogens, including *Bt* [31,32]. Given that transcription factors can attune the pathogenesis process, the possession of distinct virulence determinants does not delineate the strains' efficacy as the regulation of expression patterns is essential as well [31]. Finally, extracellular structures serve as key players in virulence and specificity through initiating and maintaining contact with host cells. This role was reported for cell wall structural units [33], flagellum [34], lectin-binding [35], and fibronectin-binding proteins [36].

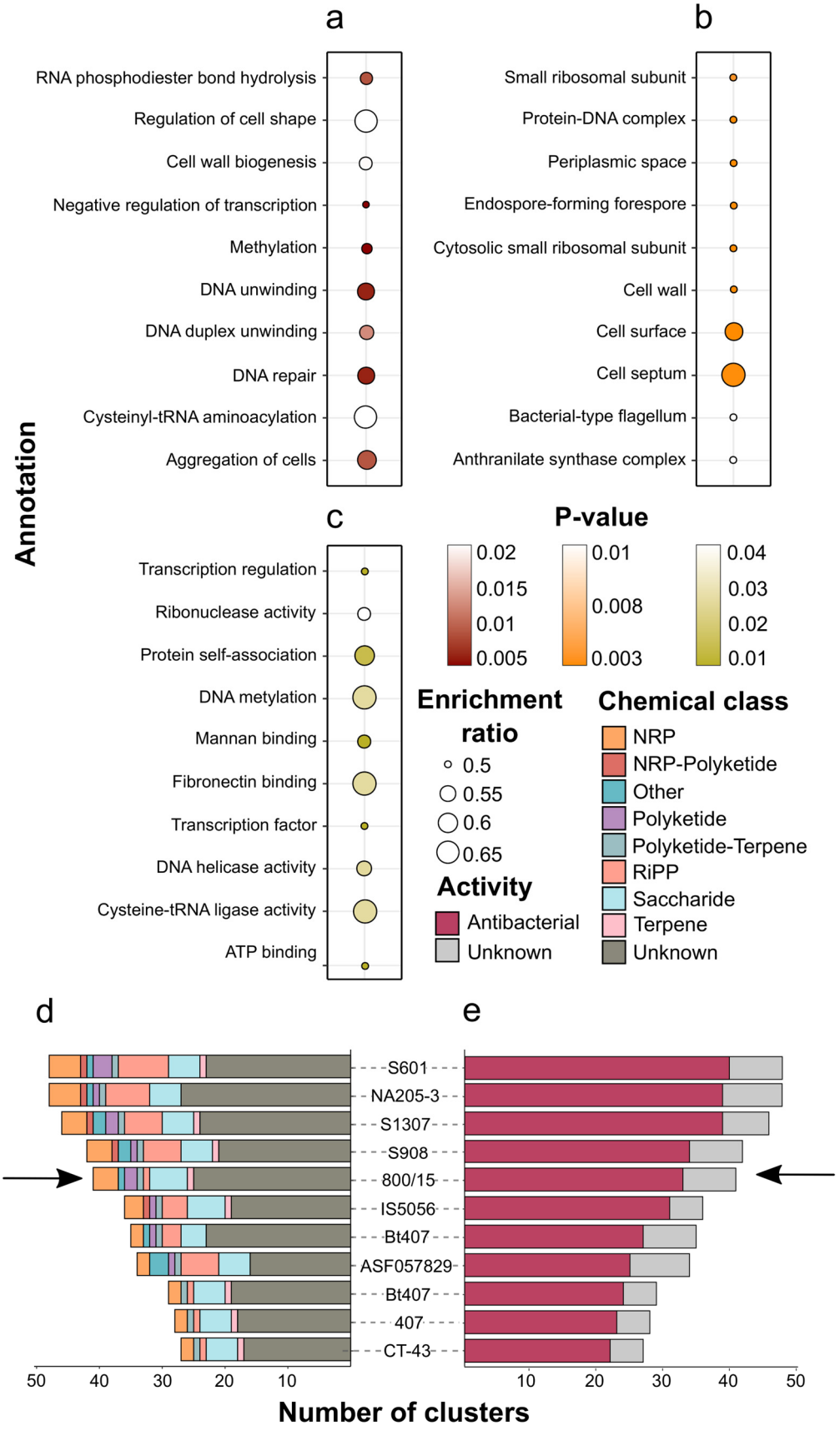

Figure 5. Functional annotation of the strain 800/15 and the composition of biosynthetic gene clusters in the studied genomic dataset. (**a**) Over-representation tests of functional terms in the analyzed strain

within the Gene Ontology (GO) system in Biological Process (BP), (**b**) Cellular Component (CC), and Molecular Function (MF) (**c**) categories. The size of the circles denotes the enrichment ratio, i.e., the ratio of terms belonging to the strain 800/15 to the total number of the respective terms in the universe (the sum of the terms for all proteins in the dataset). The color indicates a *p*-value obtained from Fisher's exact text corrected with the "weight0" algorithm. (**d**) The total number of biosynthetic gene clusters (BGCs) identified using DeepBGC v0.1.30 [20] software concerning the chemical class of potentially synthesized metabolites. (**e**) The distribution of BGCs displaying the percentage of BGCs with predicted antibiotic activity.

We continued to explore the diversity of BGCs in the genomic dataset. Given that the machine learning approach allows DeepBGC v0.1.30 [20] to predict the biological activities of the BGCs, we launched it on the set of genomes (Table S9). According to the distribution of chemical classes, our strain was enriched in saccharides (Figure 5d) with 6 clusters vs. 4.5 on average. However, for all of the isolates, most of the BGCs were of the unknown type. The higher fraction of saccharides, putatively embedded into the cell wall, is consistent with the abovementioned functional annotation results. We next summarized the activity spectrum. There were no clusters marked as antifungal, whereas antibacterial action prevailed in all the strains (Figure 5e). The percentage of BGCs with antimicrobial activity in strain 800/15 accounted for up to 80.5% which is close to the average of 81%. The absence of antifungal clusters might be explained by the low predictive power due to the scarcity of underlying data. Nonetheless, as revealed by antiSMASH v6.1.1 [21], the strain can produce fengycin—an antifungal lipopeptide drastically inhibiting the growth of both fungi and bacteria [37]. The tool also identified bacillibactin, a strong siderophore-based antibiotic that can combat even multi-drug resistant pathogens [38]. As a siderophore, petrobactin acts as a peptide that binds to metals, reducing the availability of iron to pathogens, and thereby contributing to the reduction in pathogenic microorganisms within the soil [37,39,40]. Siderophores could also exert a potential plant growth-promoting effect, providing essential iron [41]. A large collection of unexplored secondary metabolites whose production is implied by DeepBGC reports suggests *Bt* strain 800/15 contains a source of potential fungicidal and bactericidal moieties.

In summary, the draft genome sequence of the highly insecticidal *Bt* strain 800/15 was obtained. It was shown that the 800/15 strain possesses a wide range of agriculturally important activities, ranging from pesticidal to antifungal. The comparative genomic analysis revealed that quantitative genomic features, i.e., the number of virulence factors, do not delineate the activity spectrum of the strain *per se*. It is more likely that the content of the respective loci is more crucial. Therefore, summarizing the co-occurrence of genomic determinants is a better alternative for the accurate prediction of the isolates' properties directly from the sequencing data. On that account, the provided genome assembly holds significance for the field being a useful source of genetic markers contributing to the efficacy of *Bt* strains in agriculture.

3. Methods

3.1. DNA Extraction and Whole Genome Sequencing

The total DNA was extracted according to the protocol described by Romanenko et al. (2023) [42]. The 800/15 strain was cultured at 28 °C overnight on liquid Spizizen nutrient medium, and the precipitate was collected via centrifugation and washed three times with an EDTA/NaCl buffer. The cells were resuspended in the above buffer with serial incubations in the presence of Ribonuclease A, lysozyme, and mutanolysin solutions; proteinase K; and 10% SDS with a CTAB/NaCl solution, respectively. Then the DNA was purified with phenol/chloroform, precipitated with isopropanol, and washed three times with 70% ethanol. The pellet was dissolved in a Tris–EDTA buffer. Final genomic DNA concentration and quality control were performed with a Qubit 3.0 fluorimeter, a ClarioSTAR Plus multi-mode microplate reader, and under 1% agarose gel. The whole genome sequencing was carried out on the Illumina HiSeq X platform in the paired-end mode by Macrogen (Seoul, Republic of Korea).

3.2. De Novo Genome Assembly and Annotation

The genome assembly and annotation were carried out according to the aforementioned protocol with slight modifications. The quality of the raw data was analyzed with FastQC v0.12.1 [43] with further read filtering using fastp v0.23.2 [44]. The genome assembly was performed with SPAdes v3.15.4 [9], followed by quality control with QUAST v5.2.0 [11]. Next, we evaluated the taxonomy completeness of the assembly with BUSCO v5.4.2 [12] by calculating the percentage of one-copy orthologs derived from the "Bacillales_odb10" and "Bacilli_odb10" databases and checked the taxonomical attribution with CheckM v1.2.2 [10]. The fastANI v1.33 [13] tool was utilized to identify the closest genomes of the *Bacillus* spp. genomes downloaded from the NCBI RefSeq database [14]. Following that, 10 genomes with the highest ANI values were used for training a model for Prodigal v2.6.3 [45] which was further used for gene prediction using Prokka v1.14.6 [17]. To attribute genomes to taxonomic units within genomospecies from the *B. cereus sensu lato* group, we applied Btyper v3.4.0 [15] with default settings. The genes encoding insecticidal toxins were mined with BtToxin_Digger v1.0.10 [18] and CryProcessor v1.0 [46] with the target insect species derived from the BPPRC specificity database (https://www.bpprc-db.org/, accessed on 10 November 2023) [19]. The biosynthetic gene clusters and the spectrum of their activities were predicted using the DeepBGC v0.1.30 [20] and antiSMASH v6.1.1 [21] tools.

3.3. Comparative Genomic Analysis of Closest Reference Strains

The metadata for the ten closest strains determined in accordance with the highest ANI estimate, namely, hosts and isolation source, were retrieved from the BioSample database [23]. In addition, an experimentally verified toxicity spectrum was obtained from literature sources. The phylogenetic relationships between the genomes were based on ANI values calculated via the Mash v2.3 utility [47] with a k-mer size of 21 and a sketch size of 100,000 was specified. Then, we built the distance matrix (Table S10) with a custom Python 3.7 script and clustered genomes using the "Bclust" function from the shipunov v.1.17.1 package [48]. We then followed the aforementioned methodology to examine genomes for the presence of insecticidal toxins and subsequently conducted a theoretical insecticidal activity spectrum analysis with the prediction of BGCs. Furthermore, we mined for well-known virulence factors deposited in the VFDB (http://www.mgc.ac.cn/VFs/, accessed on 15 November 2023) [25] utilizing the MMseqs2 v.14.7 software [49] on the full dataset of protein sequences. Homologs were then selected if jointly meeting the following criteria: not less than 70% of inter-sequence similarity and mutual coverage. The genomic regions harboring insecticidal loci were classified as plasmidic or chromosomal with the PlasmidHunter v1.0 [50] tool. For finding CRISPR (clustered regularly interspaced short palindromic repeats) loci, we used MinCED v0.4.2 [51]. Genomic islands, insertion sequences, and prophages were identified with IslandPath-DIMOB v1.0.6 [52], ISEScan v1.7.2.3 [53], and Phigaro v2.3.0 [54], respectively. Finally, we functionally characterized the 800/15 strain genome using functional annotations made by the eggNOG standalone tool v2.0.1b-2-g816e190 [55] in the "mmseqs2" search mode. Over-representation functional terms were evaluated in the topGO v.3.15 package [29] within all three domains of the GO system (Gene Ontology) [56], e.g., Cellular Component (CC), Molecular Function (MF), and Biological Process (BP).

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/data9020034/s1, Supplementary Data S1: The GBK-formatted file with annotation results with Prokka v1.14.6, using the Prodigal v2.6.3 model trained on the 10 closest assemblies; Supplementary Table S1: The metadata from the Biosample database for the ten closest genome assemblies according to ANI values; Supplementary Table S2: Experimentally assessed toxicity spectrum for the aforementioned strains obtained from literature sources; Supplementary Table S3: Identified homologs of insecticidal toxins and known virulence determinants from VFDB coupled with predicted toxic activities inferred from the BPPRC specificity database; Supplementary Table S4: Comparisons between the 800/15 strain with reference genomes in terms of the content of insecticidal toxins and known virulence determinants; Supplementary Table S5: The total amount

of pesticidal toxins, virulence determinants, and affected host species; Supplementary Table S6: The classification of contigs harboring insecticidal loci; Supplementary Table S7: The total number of predicted CRISPR (clustered regularly interspaced short palindromic repeats) sequences, and mobile genetic elements; Supplementary Table S8: Significant over-represented functional terms within the GO annotation system determined using the topGO package; Supplementary Table S9: The number of biosynthetic gene clusters found by the DeepBGC software regarding the chemical class of the synthesized products and their activity; Supplementary Table S10: The similarity matrix incorporating pair-wise ANI comparisons between genomes calculated with the mash utility.

Author Contributions: Conceptualization, A.E.S. and K.S.A.; methodology, A.E.S. and M.N.R.; software, A.E.S.; validation, I.A.S.; formal analysis, I.A.S.; investigation, A.E.S. and M.N.R.; resources, A.A.N.; data curation, A.E.S.; writing—original draft preparation, I.A.S. and A.E.S.; writing—review and editing, A.E.S., I.A.S., M.N.R., A.A.N. and K.S.A.; visualization, A.E.S. and M.N.R.; supervision, K.S.A.; project administration, K.S.A.; funding acquisition, A.A.N. All authors have read and agreed to the published version of the manuscript.

Funding: The article was made with the support of the Ministry of Science and Higher Education of the Russian Federation in accordance with agreement № 075-15-2022-320, date 20 April 2022 on providing a grant in the form of subsidies from the Federal budget of Russian Federation. The grant was provided for state support for the creation and development of a World-class Scientific Center "Agrotechnologies for the Future".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw genome sequencing data were submitted to the NCBI with BioSample number SAMN38204081, under BioProject PRJNA1039177. The assembled genome is available in the NCBI Assembly database under ASM3376411v1.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Martin, P.A.W.; Travers, R.S. Worldwide Abundance and Distribution of *Bacillus thuringiensis* Isolates. *Appl. Environ. Microbiol.* **1989**, *55*, 2437–2442. [CrossRef]
2. Raymond, B.; Johnston, P.R.; Nielsen-LeRoux, C.; Lereclus, D.; Crickmore, N. *Bacillus thuringiensis*: An Impotent Pathogen? *Trends Microbiol.* **2010**, *18*, 189–194. [CrossRef] [PubMed]
3. Palma, L.; Muñoz, D.; Berry, C.; Murillo, J.; Caballero, P. *Bacillus thuringiensis* Toxins: An Overview of Their Biocidal Activity. *Toxins* **2014**, *6*, 3296–3325. [CrossRef] [PubMed]
4. Belousova, M.E.; Malovichko, Y.V.; Shikov, A.E.; Nizhnikov, A.A.; Antonets, K.S. Dissecting the Environmental Consequences of *Bacillus thuringiensis* Application for Natural Ecosystems. *Toxins* **2021**, *13*, 355. [CrossRef] [PubMed]
5. Lacey, L.A.; Grzywacz, D.; Shapiro-Ilan, D.I.; Frutos, R.; Brownbridge, M.; Goettel, M.S. Insect Pathogens as Biological Control Agents: Back to the Future. *J. Invertebr. Pathol.* **2015**, *132*, 1–41. [CrossRef] [PubMed]
6. Tikhonovich, I.; Romanova, T.; Ermolova, V.; Grishechkina, S. Bacterial Strain *Bacillus thuringiensis* Var. Thuringiensis N800/15 as Agent for Preparation Entomocidal Biopreparation. RU Patent 2514211 C1, 27 April 2014.
7. Grishechkina, S.D.; Ermolova, V.P.; Kovalenko, T.K.; Antonets, K.S.; Belousova, M.E.; Yakhno, V.V.; Nizhnikov, A.A. Polyfunctional Properties of the *Bacillus thuringiensis* Var. thuringiensis Industrial Strain 800/15. *Agric. Biol.* **2019**, *54*, 494–504. [CrossRef]
8. Stewart, G.S.; Johnstone, K.; Hagelberg, E.; Ellar, D.J. Commitment of Bacterial Spores to Germinate a Measure of the Trigger Reaction. *Biochem. J.* **1981**, *198*, 101–106. [CrossRef]
9. Bankevich, A.; Nurk, S.; Antipov, D.; Gurevich, A.A.; Dvorkin, M.; Kulikov, A.S.; Lesin, V.M.; Nikolenko, S.I.; Pham, S.; Prjibelski, A.D.; et al. SPAdes: A New Genome Assembly Algorithm and Its Applications to Single-Cell Sequencing. *J. Comput. Biol.* **2012**, *19*, 455–477. [CrossRef]
10. Parks, D.H.; Imelfort, M.; Skennerton, C.T.; Hugenholtz, P.; Tyson, G.W. CheckM: Assessing the Quality of Microbial Genomes Recovered from Isolates, Single Cells, and Metagenomes. *Genome Res.* **2015**, *25*, 1043–1055. [CrossRef]
11. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]
12. Manni, M.; Berkeley, M.R.; Seppey, M.; Simão, F.A.; Zdobnov, E.M. BUSCO Update: Novel and Streamlined Workflows along with Broader and Deeper Phylogenetic Coverage for Scoring of Eukaryotic, Prokaryotic, and Viral Genomes. *Mol. Biol. Evol.* **2021**, *38*, 4647–4654. [CrossRef]
13. Jain, C.; Rodriguez-R, L.M.; Phillippy, A.M.; Konstantinidis, K.T.; Aluru, S. High Throughput ANI Analysis of 90K Prokaryotic Genomes Reveals Clear Species Boundaries. *Nat. Commun.* **2018**, *9*, 5114. [CrossRef] [PubMed]

14. O'Leary, N.A.; Wright, M.W.; Brister, J.R.; Ciufo, S.; Haddad, D.; McVeigh, R.; Rajput, B.; Robbertse, B.; Smith-White, B.; Ako-Adjei, D.; et al. Reference Sequence (RefSeq) Database at NCBI: Current Status, Taxonomic Expansion, and Functional Annotation. *Nucleic Acids Res.* **2016**, *44*, D733–D745. [CrossRef]
15. Carroll, L.M.; Cheng, R.A.; Kovac, J. No Assembly Required: Using BTyper3 to Assess the Congruency of a Proposed Taxonomic Framework for the *Bacillus cereus* Group with Historical Typing Methods. *Front. Microbiol.* **2020**, *11*, 580691. [CrossRef] [PubMed]
16. Carroll, L.M.; Wiedmann, M.; Kovac, J. Proposal of a Taxonomic Nomenclature for the *Bacillus cereus* Group Which Reconciles Genomic Definitions of Bacterial Species with Clinical and Industrial Phenotypes. *mBio* **2020**, *11*, 10–128. [CrossRef] [PubMed]
17. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef] [PubMed]
18. Liu, H.; Zheng, J.; Bo, D.; Yu, Y.; Ye, W.; Peng, D.; Sun, M. BtToxin_Digger: A Comprehensive and High-Throughput Pipeline for Mining Toxin Protein Genes from *Bacillus thuringiensis*. *Bioinformatics* **2021**, *38*, 250–251. [CrossRef] [PubMed]
19. Panneerselvam, S.; Mishra, R.; Berry, C.; Crickmore, N.; Bonning, B.C. BPPRC Database: A Web-Based Tool to Access and Analyse Bacterial Pesticidal Proteins. *Database* **2022**, *2022*, baac022. [CrossRef]
20. Hannigan, G.D.; Prihoda, D.; Palicka, A.; Soukup, J.; Klempir, O.; Rampula, L.; Durcak, J.; Wurst, M.; Kotowski, J.; Chang, D.; et al. A Deep Learning Genome-Mining Strategy for Biosynthetic Gene Cluster Prediction. *Nucleic Acids Res.* **2019**, *47*, e110. [CrossRef]
21. Blin, K.; Shaw, S.; Kloosterman, A.M.; Charlop-Powers, Z.; van Wezel, G.P.; Medema, M.H.; Weber, T. AntiSMASH 6.0: Improving Cluster Detection and Comparison Capabilities. *Nucleic Acids Res.* **2021**, *49*, W29–W35. [CrossRef]
22. Lereclus, D.; Arantès, O.; Chaufaux, J.; Lecadet, M.-M. Transformation and Expression of a Cloned δ-Endotoxin Gene in *Bacillus thuringiensis*. *FEMS Microbiol. Lett.* **1989**, *60*, 211–217. [CrossRef] [PubMed]
23. Barrett, T.; Clark, K.; Gevorgyan, R.; Gorelenkov, V.; Gribov, E.; Karsch-Mizrachi, I.; Kimelman, M.; Pruitt, K.D.; Resenchuk, S.; Tatusova, T.; et al. BioProject and BioSample Databases at NCBI: Facilitating Capture and Organization of Metadata. *Nucleic Acids Res.* **2012**, *40*, D57–D63. [CrossRef]
24. Togawa, R.; Martins, É.; Queiroz, P.; Grynberg, P.; Monnerat, R. Draft Genome Sequence of *Bacillus thuringiensis* Strain S1307, an Isolate Toxic for Lepidoptera. *Braz. Appl. Sci. Rev.* **2022**, *6*, 942–946. [CrossRef]
25. Chen, L.; Yang, J.; Yu, J.; Yao, Z.; Sun, L.; Shen, Y.; Jin, Q. VFDB: A Reference Database for Bacterial Virulence Factors. *Nucleic Acids Res.* **2005**, *33*, D325–D328. [CrossRef]
26. Shikov, A.E.; Malovichko, Y.V.; Lobov, A.A.; Belousova, M.E.; Nizhnikov, A.A.; Antonets, K.S. The Distribution of Several Genomic Virulence Determinants Does Not Corroborate the Established Serotyping Classification of *Bacillus thuringiensis*. *Int. J. Mol. Sci.* **2021**, *22*, 2244. [CrossRef]
27. Shikov, A.E.; Merkushova, A.V.; Savina, I.A.; Nizhnikov, A.A.; Antonets, K.S. The Man, the Plant, and the Insect: Shooting Host Specificity Determinants in *Serratia marcescens* Pangenome. *Front. Microbiol.* **2023**, *14*, 1211999. [CrossRef]
28. Gao, Y.; Hu, Y.; Fu, Q.; Zhang, J.; Oppert, B.; Lai, F.; Peng, Y.; Zhang, Z. Screen of *Bacillus thuringiensis* Toxins for Transgenic Rice to Control *Sesamia inferens* and *Chilo suppressalis*. *J. Invertebr. Pathol.* **2010**, *105*, 11–15. [CrossRef]
29. Alexa, A.; Rahnenfuhrer, J. *TopGO: Enrichment Analysis for Gene Ontology. R Package Version 2.48.0 2022*; Bioconductor: Boston, MA, USA, 2022.
30. Méric, G.; Mageiros, L.; Pascoe, B.; Woodcock, D.J.; Mourkas, E.; Lamble, S.; Bowden, R.; Jolley, K.A.; Raymond, B.; Sheppard, S.K. Lineage-Specific Plasmid Acquisition and the Evolution of Specialized Pathogens in *Bacillus thuringiensis* and the *Bacillus cereus* Group. *Mol. Ecol.* **2018**, *27*, 1524–1540. [CrossRef] [PubMed]
31. Lereclus, D.; Agaisse, H.; Grandvalet, C.; Salamitou, S.; Gominet, M. Regulation of Toxin and Virulence Gene Transcription in *Bacillus thuringiensis*. *Int. J. Med. Microbiol.* **2000**, *290*, 295–299. [CrossRef]
32. Deng, C.; Peng, Q.; Song, F.; Lereclus, D. Regulation of *Cry* Gene Expression in *Bacillus thuringiensis*. *Toxins* **2014**, *6*, 2194–2209. [CrossRef]
33. Peña, G.; Miranda-Rios, J.; de la Riva, G.; Pardo-López, L.; Soberón, M.; Bravo, A. A *Bacillus thuringiensis* S-Layer Protein Involved in Toxicity against *Epilachna Varivestis* (Coleoptera: Coccinellidae). *Appl. Environ. Microbiol.* **2006**, *72*, 353–360. [CrossRef]
34. Ghelardi, E.; Celandroni, F.; Salvetti, S.; Beecher, D.J.; Gominet, M.; Lereclus, D.; Wong, A.C.L.; Senesi, S. Requirement of FlhA for Swarming Differentiation, Flagellin Export, and Secretion of Virulence-Associated Proteins in *Bacillus thuringiensis*. *J. Bacteriol.* **2002**, *184*, 6424–6433. [CrossRef]
35. Akao, T.; Mizuki, E.; Yamashita, S.; Saitoh, H.; Ohba, M. Lectin Activity of *Bacillus thuringiensis* Parasporal Inclusion Proteins. *FEMS Microbiol. Lett.* **1999**, *179*, 415–421. [CrossRef]
36. Martínez-Zavala, S.A.; Barboza-Pérez, U.E.; Hernández-Guzmán, G.; Bideshi, D.K.; Barboza-Corona, J.E. Chitinases of *Bacillus thuringiensis*: Phylogeny, Modular Structure, and Applied Potentials. *Front. Microbiol.* **2020**, *10*, 3032. [CrossRef] [PubMed]
37. Vanittanakom, N.; Loeffler, W.; Koch, U.; Jung, G. Fengycin—A Novel Antifungal Lipopeptide Antibiotic Produced by *Bacillus subtilis* F-29-3. *J. Antibiot.* **1986**, *39*, 888–901. [CrossRef] [PubMed]
38. Dimopoulou, A.; Theologidis, I.; Benaki, D.; Koukounia, M.; Zervakou, A.; Tzima, A.; Diallinas, G.; Hatzinikolaou, D.G.; Skandalis, N. Direct Antibiotic Activity of Bacillibactin Broadens the Biocontrol Range of *Bacillus amyloliquefaciens* MBI600. *mSphere* **2021**, *6*, e0037621. [CrossRef] [PubMed]
39. Beneduzi, A.; Ambrosini, A.; Passaglia, L.M.P. Plant Growth-Promoting Rhizobacteria (PGPR): Their Potential as Antagonists and Biocontrol Agents. *Genet. Mol. Biol.* **2012**, *35*, 1044–1051. [CrossRef] [PubMed]
40. Ongena, M.; Jacques, P. *Bacillus lipopeptides*: Versatile Weapons for Plant Disease Biocontrol. *Trends Microbiol.* **2008**, *16*, 115–125. [CrossRef] [PubMed]

41. Saha, M.; Sarkar, S.; Sarkar, B.; Sharma, B.K.; Bhattacharjee, S.; Tribedi, P. Microbial Siderophores and Their Potential Applications: A Review. *Environ. Sci. Pollut. Res.* **2016**, *23*, 3984–3999. [CrossRef] [PubMed]
42. Romanenko, M.N.; Nesterenko, M.A.; Shikov, A.E.; Nizhnikov, A.A.; Antonets, K.S. Draft Genome Sequence Data of *Lysinibacillus sphaericus* Strain 1795 with Insecticidal Properties. *Data* **2023**, *8*, 167. [CrossRef]
43. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 20 November 2023).
44. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef]
45. Hyatt, D.; Chen, G.-L.; LoCascio, P.F.; Land, M.L.; Larimer, F.W.; Hauser, L.J. Prodigal: Prokaryotic Gene Recognition and Translation Initiation Site Identification. *BMC Bioinform.* **2010**, *11*, 119. [CrossRef]
46. Shikov, A.E.; Malovichko, Y.V.; Skitchenko, R.K.; Nizhnikov, A.A.; Antonets, K.S. No More Tears: Mining Sequencing Data for Novel Bt Cry Toxins with CryProcessor. *Toxins* **2020**, *12*, 204. [CrossRef] [PubMed]
47. Ondov, B.D.; Treangen, T.J.; Melsted, P.; Mallonee, A.B.; Bergman, N.H.; Koren, S.; Phillippy, A.M. Mash: Fast Genome and Metagenome Distance Estimation Using MinHash. *Genome Biol.* **2016**, *17*, 132. [CrossRef] [PubMed]
48. Shipunov, A. Shipunov: Miscellaneous Functions from Alexey Shipunov. Available online: https://cran.r-project.org/web/packages/shipunov/index.html (accessed on 11 November 2023).
49. Steinegger, M.; Söding, J. MMseqs2 Enables Sensitive Protein Sequence Searching for the Analysis of Massive Data Sets. *Nat. Biotechnol.* **2017**, *35*, 1026–1028. [CrossRef] [PubMed]
50. Tian, R.; Imanian, B. PlasmidHunter: Accurate and Fast Prediction of Plasmid Sequences using Gene Content Profile and Machine Learning. *bioRxiv* **2023**, preprint. [CrossRef]
51. Bland, C.; Ramsey, T.L.; Sabree, F.; Lowe, M.; Brown, K.; Kyrpides, N.C.; Hugenholtz, P. CRISPR Recognition Tool (CRT): A Tool for Automatic Detection of Clustered Regularly Interspaced Palindromic Repeats. *BMC Bioinform.* **2007**, *8*, 209. [CrossRef]
52. Bertelli, C.; Brinkman, F.S.L. Improved Genomic Island Predictions with IslandPath-DIMOB. *Bioinformatics* **2018**, *34*, 2161–2167. [CrossRef]
53. Xie, Z.; Tang, H. ISEScan: Automated Identification of Insertion Sequence Elements in Prokaryotic Genomes. *Bioinformatics* **2017**, *33*, 3340–3347. [CrossRef]
54. Starikova, E.V.; Tikhonova, P.O.; Prianichnikov, N.A.; Rands, C.M.; Zdobnov, E.M.; Ilina, E.N.; Govorun, V.M. Phigaro: High-Throughput Prophage Sequence Annotation. *Bioinformatics* **2020**, *36*, 3882–3884. [CrossRef]
55. Huerta-Cepas, J.; Forslund, K.; Coelho, L.P.; Szklarczyk, D.; Jensen, L.J.; von Mering, C.; Bork, P. Fast Genome-Wide Functional Annotation through Orthology Assignment by EggNOG-Mapper. *Mol. Biol. Evol.* **2017**, *34*, 2115–2122. [CrossRef] [PubMed]
56. Ashburner, M.; Ball, C.A.; Blake, J.A.; Botstein, D.; Butler, H.; Cherry, J.M.; Davis, A.P.; Dolinski, K.; Dwight, S.S.; Eppig, J.T.; et al. Gene Ontology: Tool for the Unification of Biology. The Gene Ontology Consortium. *Nat. Genet.* **2000**, *25*, 25–29. [CrossRef]