data

MDPI

Data Descriptor

Draft Genome Sequence Data of *Lysinibacillus sphaericus* Strain 1795 with Insecticidal Properties

Maria N. Romanenko [1,2], Maksim A. Nesterenko [1], Anton E. Shikov [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; m.romanenko@arriam.ru (M.N.R.); a.shikov@arriam.ru (A.E.S.); a.nizhnikov@arriam.ru (A.A.N.)
[2] Faculty of Biology, St. Petersburg State University, 199034 St. Petersburg, Russia
* Correspondence: k.antonets@arriam.ru

Abstract: *Lysinibacillus sphaericus* holds a significant agricultural importance by being able to produce insecticidal toxins and chemical moieties of varying antibacterial and fungicidal activities. In this study, the genome of the *L. sphaericus* strain 1795 is presented. Illumina short reads sequenced on the HiSeq X platform were used to obtain the genome's assembly by applying the SPAdes v3.15.4 software. The genome size based on a cumulative length of 23 contigs reached 4.74 Mb, with a respective N50 of 1.34 Mb. The assembled genome carried 4672 genes, including 4643 protein-encoding ones, 5 of which represented loci coding for insecticidal toxins active against the orders Diptera, Lepidoptera, and Blattodea. We also revealed biosynthetic gene clusters responsible for the synthesis of secondary metabolites with predicted antibacterial, fungicidal, and growth-promoting properties. The genomic data provided will be helpful for deepening our understanding of genetic markers determining the efficient application of the *L. sphaericus* strain 1795 primarily for biocontrol purposes in veterinary and medical applications against several groups of blood-sucking insects.

Dataset: The raw genome sequencing data of Illumina HiSeq X were submitted to the NCBI SRA database in a FASTQ format with BioSample SAMN37209907, under BioProject PRJNA1011199. The assembled genome is available in the NCBI GeneBank under ASM3119793v1.

Keywords: *Lysinibacillus sphaericus*; draft genome; illumina sequencing; economically valuable species

Citation: 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. https://doi.org/10.3390/data8110167

Academic Editor: Juraj Gregáň

Received: 8 September 2023
Revised: 1 November 2023
Accepted: 1 November 2023
Published: 3 November 2023

1. Summary

Lysinibacillus sphaericus (formerly called *Bacillus sphaericus*) is a spore-forming bacterium first described as an insect pathogen nearly six decades ago by Kellen et al. [1,2]. Despite being initially perceived as a highly effective mosquito control agent [3–5], this species was later shown to exhibit a wide range of other characteristics, including insecticidal activity against species other than Diptera, and bactericidal, fungicidal, plant growth-promoting, and bioremediation activities, among others, thus being potentially useful in agriculture [1,6,7]. The majority of *L. sphaericus* strains produce spore-associated larvicidal binary toxins comprised of two subunits called Tpp1 and Tpp2. These subunits were formerly known as BinA and BinB, respectively [1,8]. Some strains are also capable of producing the 3-domain cry-toxin Cry48, requiring binary Tpp49 protein to activate toxicity [9]. An extensive usage of the spore–crystal complex does not fully exploit the insecticidal potential of strains secreting other toxins during the vegetative stage, leading to the emergence of resistant insects [1,7,10]. Given the aforementioned information, there is a high demand for isolating, characterizing, and testing novel strains, especially those synthesizing previously unreported proteins and compounds with agriculturally valuable

activities [11]. In this context, the ongoing accumulation of genomic data provides insights into possible mechanisms delineating the potential usefulness of the isolates and could ease the selection of promising strains.

2. Data Description

2.1. Isolation of Lysinibacillus sphaericus Strain 1795 and Characterization of Its Morphology and Insecticidal Activity

The *Lysinibacillus sphaericus* strain 1795 was isolated from a freshwater pond inhabited by *Aedes* sp. larvae located in Babolovsky park, Pushkin, St. Petersburg, Russia. When cultivated on a Lysogeny broth agar nutrient medium [12], the strain forms yellow-white, smooth, flat, shiny, and circular colonies (Figure 1a). The vegetative cells are rod-shaped, 0.6–1.0 × 1.5–5.0 µm in size, and capable of forming subterminal spores (Figure 1b).

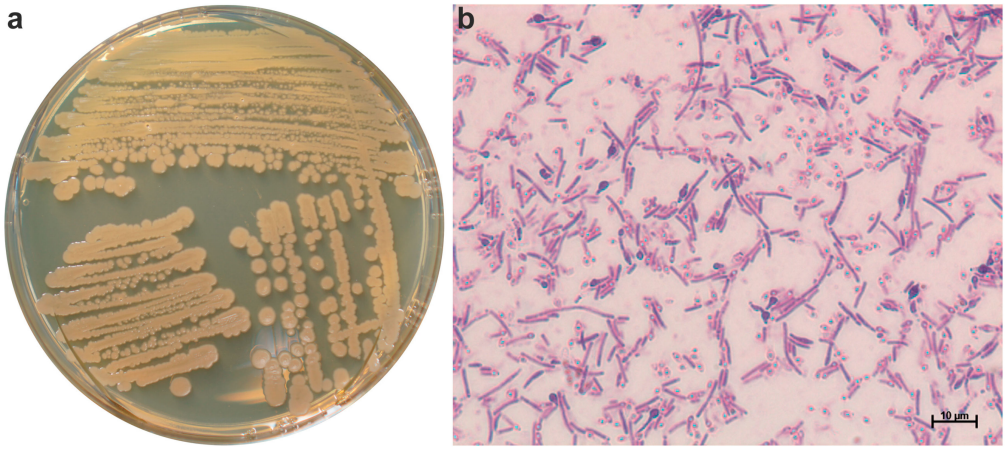

Figure 1. The morphology of the *L. sphaericus* strain 1795's colonies (**a**) and sporulating culture stained with Coomassie brilliant blue (100× objective) (**b**) after four days of cultivation on an LB nutrient medium.

The strain was deposited in the joint Russian Collection of Agricultural Microorganisms (RCAM) at the All-Russia Research Institute for Agricultural Microbiology, Saint-Petersburg (http://62.152.67.70/cryobank/login.jsp, accessed on 30 August 2023), in 2014, under the registration number RCAM02787 (Supplementary Data S1).

According to the information given in the certificate of deposition, the strain is highly toxic to the second instar larvae of a set of harmful mosquito species: *Aedes caspis*, *Aedes communis*, *Aedes dorsalis*, *Aedes dorsalis*, *Aedes flavescens*, *Aedes leucomelas*, *and Culex pipiens molestus* (Supplementary Data S1).

It does not affect hydrobionts accompanying mosquito larvae, including flatworms (*Planaria torva*), gastropods (*Limnaea palustris*, *Limnaea stagnalis*), crustaceans (*Asellus aquaticus*, *Cyclops strenuus*, *Cypridopsis* sp., *Cypris* sp., *Daphnia magna*, *Daphnia pulex*), springtails (*Podura aquatica*), and insects (*Cloeon dipterum*, *Cloeon inscriptum*, *Lestes nympha*, *Lestes sponsa*, *Lestes virens*, *Nemoura cinerea*, *Sympetrum danae*, *Sympetrum flaveolum*) (Supplementary Data S1).

These data suggest the suitability of the *Lysinibacillus sphaericus* strain 1795 for the development of biological preparations against blood-sucking insects and determine the importance of studying its genome to decipher the molecular determinants of insecticidal properties.

2.2. Genome Assembly and Annotation

The whole genome of the 1795 strain was sequenced using the Illumina HiSeq X platform with 150 bp paired-end reads. According to the quality control reported using the FastQC v0.12.1 [13] and fastP v0.23.2 [14] programs, the sequencing data of the short-read DNA libraries, both raw and devoid of adapters, were of high quality, i.e., presented uniform distributions of quality scores, GC content, etc. The draft, de novo genome

assembly generated using the SPAdes v3.15.4 software [15] consisted of 23 contigs with a total size and N50 of 4.74 Mb and 1.34 Mb, respectively. The genome's completeness was equal to 99%, while contamination constituted only 1%, as revealed with CheckM v1.2.2 [16] (Supplementary Data S2). Other properties of the assembly are presented in Table 1.

Table 1. The main characteristics of the draft genome assembly of the *L. sphaericus* strain 1795 obtained using the QUAST v5.2.0 [17] and CheckM v1.2.2 [16] utilities.

Total amount of contigs	23
Largest contig (number of nucleotides)	1,459,849
Total length (number of nucleotides)	4,737,839
GC-content (%)	37
N50 value	1,336,327
N90 value	144,621
L50 value	2
L90 value	9
Number of properly paired reads (%)	99.09
Average depth of coverage	226
Assembly completeness (%)	99.34
Suspected contamination (%)	0.99

When utilizing the BUSCO v5.4.2 program [18], we found that the number of fully assembled single-copy orthologues was at least 99.8% percent, when compared with both the Bacillales_odb10 and Bacilli_odb10 databases (Table 2). Therefore, the results indicate the high quality and completeness of the genome assembly.

Table 2. Estimation of the presence of BUSCO v5.4.2 markers [18] in the protein-coding genes presented in the assembly. Presented are the number of orthologs found in the assembly, coupled with their respective percentages.

Database	Bacillales_Odb10	Bacilli_Odb10
Single-copy orthologues assembled completely	449 (99.8%)	302 (100.0%)
Orthologues present in one copy	445 (98.9%)	301 (99.7%)
Multi-copies orthologues	4 (0.9%)	1 (0.3%)
Fragmented sequences	0 (0.0%)	0 (0.0%)
Orthologues missing from the assembly	1 (0.2%)	0 (0.0%)
Total number of single-copy orthologues in the database	450	302

We then picked 10 phylogenetically closest genome assemblies of bacteria belonging to the *Bacillaceae* family deposited in the NCI Refseq database [19], based on the ANI (average nucleotide identity) values calculated using the fastANI v1.33 tool [20]. The genomes with the highest overall similarity of nucleotide sequences almost exclusively belonged to the *L. sphaericus* strains, thus corroborating the valid taxonomic attribution of the strain 1795 (Table 3).

The annotation with the Prokka v1.14.6 tool [21] showed that the genome of the studied strain contained 4672 genes, 4643 of which were coding sequences, with 1128 of them marked as hypothetical proteins and lacking annotations (Supplementary Data S3). The BtToxin_Digger v1.0.10 tool indicated that the genome analyzed housed loci coding for insecticidal toxins, namely, Mtx1Aa1, Mpp3Aa1, Tpp1Aa2, Tpp2Aa2, and Spp1Aa1. The respective toxins were shown to exert an effect on a wide range of insects belonging to the orders Diptera, Lepidoptera, and Blattodea. According to the inferences obtained with the BtToxin_Digger v1.0.10 [22] and CryProcessor v1.0 [23] utility, the genome did not contain *cry* genes (Table 4).

Table 3. The list of the phylogenetically closest assemblies relative to the genome of the studied strain according to the ANI value calculated with the fastANI v1.33 software [20].

NCBI RefSeq Assembly	Taxon	ANI
GCF_001598075.1	*L. sphaericus*	99.9977
GCF_001581875.1	*L. sphaericus*	99.997
GCF_015335425.1	*L. sphaericus*	99.9962
GCF_000568835.1	*L. sphaericus*	99.9962
GCF_015845635.1	*L. sphaericus*	99.9956
GCF_015845625.1	*L. fusiformis*	99.995
GCF_024753545.1	*L. sphaericus*	99.9947
GCF_015845595.1	*L. sphaericus*	99.9942
GCF_001629735.1	*L. sphaericus*	99.9705
GCF_001623495.1	*L. sphaericus*	99.9702

Table 4. The repertoire of insecticidal toxins identified in the genome of the analyzed strain using the BtToxin_Digger v1.0.10 program [22]. The target species describe experimentally derived data deposited in the BPPRC [24] (Bacterial Pesticidal Protein Resource Center) specificity database.

Toxin	Percent of Identity	Target Order	Target Species
Mtx1Aa1	98.5	Diptera	*Aedes aegypti*, *Chironomus riparius*, *Culex quinquefasciatus*, *Toxorhynchites splendens*
Mpp3Aa1	100	Diptera	*Aedes aegypti*, *Culex quinquefasciatus*
Tpp1Aa2	100	Diptera	*Aedes aegypti*, *Aedes atropalpus*, *Anopheles albimanus*, *Anopheles gambiae*, *Anopheles quadrimaculatus*, *Anopheles stephensi*, *Culex pipiens*, *Culex quinquefasciatus*
Tpp2Aa2	100	Diptera	*Aedes aegypti*, *Aedes atropalpus*, *Anopheles albimanus*, *Anopheles gambiae*, *Anopheles quadrimaculatus*, *Anopheles stephensi*, *Culex pipiens*, *Culex quinquefasciatus*
Spp1Aa1	100	Lepidoptera Blattodea	*Spodoptera litura* *Blattella germanica*

The biosynthetic gene clusters in the *L. sphaericus* strain 1795's genome revealed with the DeepBGC v0.1.30 tool [25] belonged to seven gene clusters responsible for the synthesis of secondary metabolites with predicted bactericidal properties. The usage of the antiSMASH v6.1.1 tool [26], in turn, resulted in the eight biosynthetic gene clusters listed in Table 5. The clusters with the highest similarity to the known entities were fencing and petrobactin. The former is known for its strong fungicidal activity, whereas the latter, being a siderophore, serves as a metal-chelating peptide that diminishes iron accessibility to pathogens, thereby contributing to the reduction in pathogenic microorganisms within the soil [27–29]. Siderophores could also exert a potential growth-promoting effect on plants, providing them with essential iron [30].

Table 5. Biosynthetic gene clusters harbored in the genomic assembly predicted with the antiSMASH v6.1.1 [26] and DeepBGC v0.1.30 [25] programs. The score reflects the accuracy of cluster prediction obtained with the DeepBGC v0.1.30 program, while the similarity to the known clusters is calculated with the antiSMASH v6.1.1 program. The "–" symbol indicates that the biosynthetic cluster was found by only one program out of the two used.

Contig	Tool	Type/Activity	Location (Relative Coordinate, b.p.)	Most Similar Known Cluster	Similarity	Score
1	antiSMASH	Terpene	329,114–349,935 (total: 20,822)	–	–	–
	DeepBGC	Antibacterial	561,520–562,288 (total: 768)	–	–	0.87
	DeepBGC	Antibacterial	1,323,667–1,325,923 (total: 2256)	–	–	0.60
2	antiSMASH/DeepBGC	Siderophore/ Antibacterial	345,655–359,195 (total: 13,541)	Petrobactin	33%	0.56
	antiSMASH	Non-ribosomal peptide synthetase	532,592–578,510 (total: 4519)	–	–	–
	antiSMASH	Non-ribosomal peptide synthetase	628,420–690,347 (total: 61,928)	Molybdenum cofactor	23%	–
	antiSMASH/DeepBGC	agrD-like cyclic lactone autoinducer peptides, linear azol(in)e-containing peptides, thiopeptide/ Antibacterial	1,302,324–1,334,433 (total: 32,109)	–	–	0.76
	DeepBGC	Antibacterial	792,182–805,500 (total: 13,318)	–	–	0.64
	DeepBGC	Antibacterial	807,568–810,289 (total: 2721)	–	–	0.56
3	antiSMASH	Type III PKS	114,878–155,960 (total: 41,083)	–	–	–
4	antiSMASH	Beta-lactone containing protease inhibitor	1–17,474 (total: 17,474)	Fengycin	46%	–
5	antiSMASH	NRPS-like fragment	15,196–58,363 (total: 43,168)	Kijanimicin	4%	–
6	DeepBGC	Antibacterial	192–1770 (total: 1578)	–	–	0.92

Having analyzed the gathered genomic data, we might conclude that the strain possesses insecticidal efficacy, along with potential bactericidal and fungicidal properties. It follows, therefore, that it may find its further application primarily, but not limited to, as a biological control agent against blood-sucking insects.

3. Methods

3.1. DNA Extraction

For total DNA isolation, the bacterial culture was grown for 12 h on a liquid Spizizen nutrient medium [31,32] with aeration at +28 °C. It was then centrifuged and washed three times with the buffer (EDTA 0.01M, NaCl 0.15 M pH 8.0). Next, we added 500 µL of the above buffer and 15 µL of a 10 mg/mL Ribonuclease A solution (VWR International Ltd., Poole, UK) to the washed cells. To perform cell lysis, the samples were incubated for 60 min at +37 °C with 10 µL of lysozyme (PanReac AppliChem, Darmstadt, Germany) and 5 µL of mutanolysin (Sigma Chemical Co., St. Louis, MO, USA) added to the solution. The lysozyme had previously been diluted in a buffer (20 mM TrisCl pH 8.0, 2 mM EDTA, 1.2% Triton X-100) to a 10 mg/mL concentration. A mutanolysin working solution (1 mg/mL) was also prepared using the buffer with the following chemical composition: 0.05 M of TES, 1 mM of $MgCl_2$, a pH of 7.0. The purification of the sample from polysaccharides and proteins was carried out by adding 3 µL of proteinase K (600 U/mL; ThermoFisher Scientific, Bremen, Germany) to the cell lysate (3 µL, 30 min incubation at +37 °C). The

samples were then incubated with 10% of sodium dodecyl sulfate (50 µL, 10 min incubation at +65 °C) and 80 µL of cetyltrimethylammonium bromide (CTAB) and NaCl solution in a ratio of 1:10 to achieve the effective denaturation of the proteins. A further DNA purification was performed through phenol-chloroform extraction, without the addition of isoamyl alcohol. The DNA was precipitated by adding isopropanol to the samples, followed by washing three times with 70% of freshly prepared ethanol solution. At the last stage, 30 µL of Tris-EDTA (TE) buffer (pH 8.0) was added to dissolve the DNA, and the samples were left in a refrigerator at +4 °C for 18–24 h.

3.2. DNA Quality Control

The concentration of the isolated genomic DNA was assessed using a Qubit® 3.0 fluorimeter and a Qubit dsDNA BR Assay kit (Life Technologies, Eugene, ON, USA). The contamination with proteins, phenol, or other contaminating agents was evaluated using 260 nm/280 nm and 260 nm/230 nm absorbance ratios, with a value of ≥1.8 indicating the purity of the sample. The measurements were carried out on a CLARIOstar Plus multimodal reader (BMG labtech, Germany). Additional qualitative and quantitative analyses of the DNA samples were performed using electrophoresis in 1% of agarose gel stained with 0.002% of ethidium bromide via comparison with the λ DNA/HindIII marker (Thermo Fisher Scientific, Inc., Waltham, MA, USA).

3.3. Whole Genome Sequencing, De Novo Genome Assembly, and Annotation

The whole genome sequencing was conducted on the Illumina HiSeq X platform (Illumina) in the paired-end mode, with a read length of 2 × 150 bp, by Macrogen Inc. (Seoul, Republic of Korea). Quality control of the short nucleotide reads was executed with FastQC v0.12.1 [13]. The removal of the adapter sequences and the additional quality control of the reads was performed using fastp v0.23.2 [14]. The *de novo* assembly of the genome was made using SPAdes v3.15.4 [15] in the "--careful" mode. The obtained assembly was then quality-controlled with QUAST v5.2.0 [17]. The taxonomy-wise completeness of the assembly was evaluated by calculating the percentage of the one-copy orthologs from the "Bacillales_odb10" and "Bacilli_odb10" databases using the BUSCO v5.4.2 software [18]. The benchmarking datasets used in the analysis were based on the v10 release of the OrthoDB database [33]. We also verified the taxonomical attribution by assessing the completeness and contamination level utilizing CheckM v1.2.2 [16].

In the next stage, we used fastANI v1.33 [20] to reveal the phylogenetically closest genomes belonging to the *Lysinibacillus* spp. downloaded from the NCBI RefSeq database [19] by picking ten genomes with the highest average nucleotide identity (ANI) values when compared with our assembly. The selected dataset was then applied to train a model for Prodigal v2.6.3 [34], which was further used for the accurate prediction of coding sequences with the Prokka v1.14.6 [21]. To increase the number of meaningful annotations, we included more than 700,000 protein sequences of the *Bacillus* spp. from the Identical Protein Groups database [35] as the most-trusted proteins for Prokka-derived annotations.

Next, we mined for the genes encoding various insecticidal toxins using the Bt-Toxin_Digger v1.0.10 [22] and CryProcessor v1.0 [23] tools. The target species of insects against which the revealed virulence factors are active were derived from the BPPRC (Bacterial Pesticidal Protein Resource Center) specificity database [24]. The biosynthetic gene clusters, as well as the spectrum of their activities, were predicted using the DeepBGC v0.1.30 [25] and antiSMASH v6.1.1 [26] tools.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/data8110167/s1, Supplementary Data S1: The registration certificate of the *L. sphaericus* strain 1795 in the RCAM collection (Russian Collection of Agricultural Microorganisms); Supplementary Data S2: The report from the CheckM v1.2.2 tool run on the annotation results; Supplementary Data S3: The annotation results in the GBK format made using Prokka v1.14.6, based on the Prodigal v2.6.3 model trained on the 10 closest assemblies.

Author Contributions: Conceptualization, K.S.A.; methodology, K.S.A. and M.A.N.; software, M.A.N.; validation, M.N.R. and A.E.S.; formal analysis, M.N.R. and A.E.S.; investigation, M.N.R.; resources, A.A.N.; data curation, M.A.N. and K.S.A.; writing—original draft preparation, M.N.R. and A.E.S.; writing—review and editing, M.N.R., A.E.S., K.S.A., M.A.N. and A.A.N.; visualization, 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 No. 075-15-2021-1055, date 28 September 2021, on providing a grant in the form of subsidies from the federal budget of the Russian Federation. The grant was provided for the implementation of the following project: "Mobilization of the genetic resources of microorganisms on the basis of the Russian Collection of Agricultural Microorganisms (RCAM) at the All-Russia Research Institute for Agricultural Microbiology (ARRIAM) according to the network principle of organization".

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The raw genome sequencing data of Illumina HiSeq X were submitted to the NCBI SRA database in a FASTQ format with BioSample SAMN37209907, under BioProject PRJNA1011199. The assembled genome is available in the NCBI GeneBank under ASM3119793v1.

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

References

1. Berry, C. The Bacterium, *Lysinibacillus sphaericus*, as an Insect Pathogen. *J. Invertebr. Pathol.* **2012**, *109*, 1–10. [CrossRef] [PubMed]
2. Kellen, W.R.; Clark, T.B.; Lindegren, J.E.; Ho, B.C.; Rogoff, M.H.; Singer, S. *Bacillus sphaericus* Neide as a Pathogen of Mosquitoes. *J. Invertebr. Pathol.* **1965**, *7*, 442–448. [CrossRef] [PubMed]
3. Silva Filha, M.H.N.L.; Berry, C.; Regis, L. Lysinibacillus sphaericus: Toxins Mode of Action, Applications for Mosquito Control and Resistance Management. In *Advances in Insect Physiology*; Elsevier: Amsterdam, The Netherlands, 2014; pp. 89–176. ISBN 9780128001974.
4. Priest, F.G. Biological Control of Mosquitoes and Other Biting Flies by *Bacillus sphaericus* and *Bacillus thuringiensis*. *J. Appl. Bacteriol.* **1992**, *72*, 357–369. [CrossRef] [PubMed]
5. Charles, J.-F.; Nielson-LeRoux, C.; Delécluse, A. *Bacillus sphaericus* Toxins: Molecular Biology and Mode of Action. *Annu. Rev. Entomol.* **1996**, *41*, 451–472. [CrossRef] [PubMed]
6. Onajobi, I.B.; Idowu, E.O.; Adeyemi, J.O.; Samson, O.J.; Ogunyinka, P.I.; Fagade, O.E. In Vitro Antibacterial Activities and Molecular Characterization of Bacterial Species Isolated from Farmlands against Selected Pathogens. *Biotechnol. Rep.* **2020**, *27*, e00513. [CrossRef]
7. Nielsen-Leroux, C.; Charles, J.-F.; Thiery, I.; Georghiou, G.P. Resistance in a Laboratory Population of *Culex quinquefasciatus* (Diptera: Culicidae) to *Bacillus sphaericus* Binary Toxin Is Due to a Change in the Receptor on Midgut Brush-Border Membranes. *Eur. J. Biochem.* **1995**, *228*, 206–210. [CrossRef]
8. Hire, R.S.; Hadapad, A.B.; Dongre, T.K.; Kumar, V. Purification and Characterization of Mosquitocidal *Bacillus sphaericus* BinA Protein. *J. Invertebr. Pathol.* **2009**, *101*, 106–111. [CrossRef]
9. Glare, T.R.; Jurat-Fuentes, J.-L.; O'Callaghan, M. Basic and Applied Research: Entomopathogenic Bacteria. In *Microbial Control of Insect and Mite Pests*; Elsevier: Amsterdam, The Netherlands, 2017; pp. 47–67.
10. Wirth, M.C.; Georghiou, G.P.; Malik, J.I.; Abro, G.H. Laboratory Selection for Resistance to *Bacillus sphaericus* in *Culex quinquefasciatus* (Diptera: Culicidae) from California, USA. *J. Med. Entomol.* **2000**, *37*, 534–540. [CrossRef]
11. Lozano, L.C.; Dussán, J. Metal Tolerance and Larvicidal Activity of *Lysinibacillus sphaericus*. *World J. Microbiol. Biotechnol.* **2013**, *29*, 1383–1389. [CrossRef]
12. Bertani, G. Studies on lysogenesis I. *J. Bacteriol.* **1951**, *62*, 293–300. [CrossRef]
13. Andrews, S. FastQC: A Quality Control Tool for High Throughput Sequence Data. Available online: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 4 May 2023).
14. Chen, S.; Zhou, Y.; Chen, Y.; Gu, J. Fastp: An Ultra-Fast All-in-One FASTQ Preprocessor. *Bioinformatics* **2018**, *34*, i884–i890. [CrossRef] [PubMed]
15. 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] [PubMed]
16. 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] [PubMed]

17. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality Assessment Tool for Genome Assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef] [PubMed]
18. 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]
19. 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]
20. 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]
21. Seemann, T. Prokka: Rapid Prokaryotic Genome Annotation. *Bioinformatics* **2014**, *30*, 2068–2069. [CrossRef]
22. 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]
23. 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]
24. 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] [PubMed]
25. 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] [PubMed]
26. 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] [PubMed]
27. 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]
28. 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]
29. Ongena, M.; Jacques, P. *Bacillus* Lipopeptides: Versatile Weapons for Plant Disease Biocontrol. *Trends Microbiol.* **2008**, *16*, 115–125. [CrossRef]
30. 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]
31. Saleem, F.; Shakoori, A. The First Cry2Ac-Type Protein Toxic to *Helicoverpa armigera*: Cloning and Overexpression of *cry2ac7* Gene from SBS-BT1 Strain of *Bacillus thuringiensis*. *Toxins* **2017**, *9*, 358. [CrossRef]
32. Spizizen, J. Transformation of biochemically deficient strains of *Bacillus subtilis* by deoxyribonucleate. *Proc. Natl. Acad. Sci. USA* **1958**, *44*, 1072–1078. [CrossRef]
33. Kriventseva, E.V.; Kuznetsov, D.; Tegenfeldt, F.; Manni, M.; Dias, R.; Simão, F.A.; Zdobnov, E.M. OrthoDB V10: Sampling the Diversity of Animal, Plant, Fungal, Protist, Bacterial and Viral Genomes for Evolutionary and Functional Annotations of Orthologs. *Nucleic Acids Res.* **2019**, *47*, D807–D811. [CrossRef]
34. 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] [PubMed]
35. Agarwala, R.; Barrett, T.; Beck, J.; Benson, D.A.; Bollin, C.; Bolton, E.; Bourexis, D.; Brister, J.R.; Bryant, S.H.; Canese, K.; et al. Database Resources of the National Center for Biotechnology Information. *Nucleic Acids Res.* **2018**, *46*, D8–D13. [CrossRef]