Vavilovskii Zhurnal Genetiki i Selektsii
Vavilov Journal of Genetics and Breeding. 2025;29(8):1304-1317

doi 10.18699/vjgb-25-137

Study of insecticidal and fungicidal potential of endophytic bacteria of wheat, soybean and rapeseed by bioinformatic analysis methods

T.N. Lakhova [1, 2], A.I. Klimenko [1], G.V. Vasiliev [1], E.Yu. Gyrnets [3], A.M. Asaturova [3], S.A. Lashin [1, 2]

[1] Institute of Cytology and Genetics of the Siberian Branch of the Russian Academy of Sciences, Novosibirsk, Russia
[2] Novosibirsk State University, Novosibirsk, Russia
[3] Federal State Budgetary Scientific Institution "Federal Research Center of Biological Plant Protection", Krasnodar, Russia
tlakhova@bionet.nsc.ru

Abstract. Endophytic bacteria play a key role in agricultural ecosystems, as they can affect the availability of various compounds, crop yield and growth, and provide resistance to diseases and pests. Therefore, the study of endophytes of agriculturally important crop plants is a promising task in the field of biological plant protection. Understanding the mechanisms of interaction between endophytic bacteria and plants will allow the use of these microorganisms as bioagents in the future and thus reduce dependence on chemical pesticides. In this paper, samples obtained from the leaves and/or roots of wheat, rapeseed and soybean are considered. Whole-genome sequencing of the isolates was performed. Using an analytical pipeline, the genomes of 15 strains of endophyte bacteria of cultivated plants were assembled and characterized. Their insecticidal and fungicidal potential was analyzed. Gene repertoire analysis performed with GenAPI showed a high degree of correspondence between the gene repertoires of strain BZR 585 against *Alcaligenes phenolicus*, BZR 762 and BZR 278 against *Alcaligenes* sp., BZR 588 and BZR 201P against *Paenochrobactrum pullorum*. All strains, with the exception of BZR 162, BZR 588 and BZR 201P, were found to contain genes encoding proteins with fungicidal activity, such as iturins, fengycins and surfactins. All strains also contained genes encoding proteins with insecticidal activity, namely GroEL, Spp1Aa1, Spp1Aa2, Vpb1Ab1, Vpb4Ca1, HldE, mycosubtilin, fengycin and bacillomycin. The obtained genomic data are confirmed by the results of previous experimental studies: high insecticidal activity of a number of strains (BZR 1159, BZR 936, BZR 920, etc.) against *Galleria mellonella*, *Tenebrio molitor* and *Cydia pomonella*, as well as fungicidal properties against *Fusarium*, *Alternaria*, *Trichothecium*, was demonstrated. This shows the practical significance of the identified genetic determinants for the creation of new biocontrol agents.
Key words: bioinformatics; comparative genomics; endophytes; biocontrol

For citation: Lakhova T.N., Klimenko A.I., Vasiliev G.V., Gyrnets E.Yu., Asaturova A.M., Lashin S.A. Study of insecticidal and fungicidal potential of endophytic bacteria of wheat, soybean and rapeseed by bioinformatic analysis methods. *Vavilovskii Zhurnal Genetiki i Selektsii = Vavilov J Genet Breed*. 2025;29(8):1304-1317. doi 10.18699/vjgb-25-137

Funding. This research was funded by Russian Science Foundation grant No. 23-16-00260, https://rscf.ru/en/project/23-16-00260/. Genome sequencing was performed at the Center for Genomic Research, ICG SB RAS. High-performance computing was performed using resources of the "Bioinformatics" Joint Computational Center ICG SB RAS supported by the budget project FWNR-2022-0020.

Исследование инсектицидного и фунгицидного потенциала бактерий эндофитов пшеницы, сои и рапса методами биоинформатического анализа

Т.Н. Лахова [1, 2], А.И. Клименко [1], Г.В. Васильев [1], Е.Ю. Гырнец [3], А.М. Асатурова [3], С.А. Лашин [1, 2]

[1] Федеральный исследовательский центр Институт цитологии и генетики Сибирского отделения Российской академии наук, Новосибирск, Россия
[2] Новосибирский национальный исследовательский государственный университет, Новосибирск, Россия
[3] Федеральное государственное бюджетное научное учреждение «Федеральный научный центр биологической защиты растений», Краснодар, Россия
tlakhova@bionet.nsc.ru

Аннотация. Эндофитные бактерии способны влиять на доступность различных соединений, урожайность и рост сельскохозяйственных растений, а также обеспечивать устойчивость к болезням и вредителям. Поэтому исследование эндофитов сельскохозяйственно значимых культурных растений является перспективной задачей в области биологической защиты растений. В данной работе рассмотрены изоляты штаммов бактерий, полученные из листьев и/или корней пшеницы, рапса и сои. Было произведено полногеномное секвенирование

изолятов. С помощью аналитического конвейера собраны и охарактеризованы геномы 15 штаммов бактерий-эндофитов культурных растений, проанализирован их инсектицидный и фунгицидный потенциал. Анализ генного репертуара с помощью программы GenAPI показал высокую степень соответствия между генным репертуаром штамма BZR 585 относительно *Alcaligenes phenolicus*, BZR 762 и BZR 278 относительно *Alcaligenes* sp., BZR 588 и BZR 201P относительно *Paenochrobactrum pullorum*. Во всех штаммах, за исключением BZR 162, BZR 588 и BZR 201P, найдены гены, кодирующие белки, обладающие фунгицидной активностью, такие как итурины, фенгицины и сурфактины. Также во всех штаммах найдены гены, кодирующие белки с инсектицидной активностью, а именно: GroEL, Spp1Aa1, Spp1Aa2, Vpb1Ab1, Vpb4Ca1, HldE, фенгицин, микосубтилин и бацилломицин. Полученные геномные данные подтверждены экспериментальными испытаниями: ранее показана высокая инсектицидная активность штаммов BZR 1159, BZR 936, BZR 920 и др. против *Galleria mellonella*, *Tenebrio molitor* и *Cydia pomonella*, а также фунгицидные свойства против *Fusarium*, *Alternaria*, *Trichothecium*. Это демонстрирует практическую значимость выявленных генетических детерминант для создания новых агентов биоконтроля.
Ключевые слова: биоинформатика; сравнительная геномика; эндофиты; биоконтроль

Introduction

Endophytic bacteria are components of the plant microbiome that can colonize roots, stems, or leaves, where they obtain stable sources of nutrients and, in turn, increase plant resistance to both biotic and abiotic stresses. These symbionts have growth-stimulating activity due to nitrogen fixation, phosphorus mobilization, and phytohormone synthesis, and also produce a wide range of metabolites, hydrolytic enzymes, and volatile compounds involved in the biological control of pests and diseases of agricultural crops. In addition, endophytes compete for ecological niches and induce systemic resistance in plants, creating a multi-level defense that can be comparable in effectiveness to chemical pesticides, but is safe for humans and the environment (Hamane et al., 2023; Ali et al., 2024). The study of endophytes in agriculturally important crops opens up prospects for the search for biocontrol agents for various pests and pathogens (insects, fungi, and others).

Endophytic bacteria are actively being studied using bioinformatics methods and comparative genomics approaches (Wang Z. et al., 2023; Zhang M. et al., 2023; Pellegrinetti et al., 2024; Yang K. et al., 2024; Dai et al., 2025). Sequencing their complete genomes provides a large array of information for identifying genes and their products that have insecticidal and fungicidal potential in terms of solving biocontrol problems. In addition, functional annotation of genes of interest can be performed based on well-characterized related strains of microorganisms, such as *Bacillus velezensis*, *B. thuringiensis*, and others.

For example, in the rhizosphere of plants, *B. velezensis* creates a favorable nutritional and physicochemical environment for root microbiota by forming a biofilm, which promotes plant growth and protection against phytopathogens, both through the secretion of antimicrobial compounds and through the formation of phytoimmune potential in plants. The study (Rabbee et al., 2019) summarized information on strain-specific gene clusters of *B. velezensis* associated with the biosynthesis of secondary metabolites, which play an important role in both suppressing pathogens and stimulating plant growth. For *B. velezensis* BRI3 strains and related lines, comparative analysis confirmed the preservation of the "core" gene clusters of lipopeptide synthetases (iturin, fengycin, and surfactin), with the simultaneous appearance of unique type III polyketide synthases, which allows these strains to exhibit broad fungicidal potential *in vitro* (Liu et al., 2024).

Pangenomic studies of *Burkholderia* bacteria, including a comparison of 18 endophytic and pathogenic strains, revealed the loss of classical virulence determinants and enrichment of antimicrobial compound synthesis gene clusters in endophytic symbionts. These changes are a sign of adaptation to an intracellular lifestyle and serve as an indirect marker of biological control potential (Liu et al., 2024).

For the representative of Pseudomonadota, *Ochrobactrum quorumnocens* A44 is capable of disrupting quorum sensing (QS) in Gram-negative bacteria by inactivating N-acyl-homoserine lactones (AHLs) and protecting plant tissues from soft rot pathogens, the virulence of which is regulated by QS. For this strain, isolated from the potato rhizosphere, and six related type strains of the genus *Ochrobactrum*, comparative genomics showed that the core genome contains 50–66 % of genes, and the variable part for each genome accounts for 8 to 15 % (Krzyżanowska et al., 2019).

The entomopathogenic strain *B. thuringiensis* ser. *israelensis*, a well-known source of δ-endotoxins (*cry*, *cyt*), plays a key role (occupies a special place) in research. The complete sequence of the 127-kb pBtoxis plasmid showed that the cluster of genes encoding toxins remains stable, while regulatory and mobile elements are actively reorganized, ensuring horizontal transfer of insecticidal protein genes and expanding the adaptive potential of the strain (Berry et al., 2002; Bolotin et al., 2017).

At the same time, microorganisms that establish symbiotic relationships with their hosts were found to contain GroEL proteins (an ATP-dependent molecular chaperone that is present in all forms of life and is one of the most conservative proteins in living organisms), which acted as toxins (Horwich et al., 2007; Shi et al., 2012; Kupper et al., 2014). Its homologue, XnGroEL, has been described in the *Xenorhabdus nematophila*, which retains its folding function but acquires the ability to bind to the insect's chitin cuticle and suppress the host's immune responses (Horwich et al., 2007).

Thus, the creation of databases on proteins associated with the insecticidal and fungicidal properties of endophytic bacteria opens up broad opportunities for bioinformatic analysis and *in silico* screening of bacteria with protective properties that are significant for agriculture. For example, the BPPRC database (Panneerselvam et al., 2022) summarizes information on some proteins and peptides with insecticidal activity.

In this paper, we present assemblies and annotations of the genomes of 15 strains of endophytic bacteria isolated from

various organs of wheat, soybean, and rapeseed plants. Annotation was performed relative to close bacterial reference genomes. We also demonstrated the presence of proteins with fungicidal and insecticidal properties in the studied strains.

Materials and methods

Strains of endophytic bacteria. In this research, we used the scientific equipment "Technological line for obtaining microbiological plant protection products of a new generation" (http://ckp-rf.ru/usu/671367). The objects of the study were bacterial strains from the bioresource collection of the Federal State Budgetary Scientific Institution Federal Research Center for Biological Resources "State Collection of Entomophagous Acarids and Microorganisms" (https://fncbzr.ru/brk-i-unu/unique-installation-1/). The strains under study were isolated from the roots and leaves of wheat, soybeans, and rapeseed. Samples were collected in four districts of the Krasnodar Territory (Krylovsky, Vyselkovsky, Pavlovsky, and Krasnodar). The general characteristics of the microbial strains under study are presented in Supplementary Material 1[1].

Sequencing. Bacterial DNA was isolated from individual colonies grown on agarized medium in Petri dishes using the D-Cells kit (Biolabmix, Russia) according to the method for Gram-negative bacteria. The colony was transferred to a 1.5 ml tube and resuspended in 150 µl of PBS buffer. After adding 20 µl of proteinase K and 150 µl of lysis buffer, the cells were incubated for 10 min at 56 °C. After adding 500 µl of LB buffer, the lysate was applied to the column and centrifuged for 30 seconds at 12,000g. The column was sequentially washed with 500 µl of WB1 and WB2 buffers, followed by centrifugation. DNA was eluted from the column using 60 µl of EB buffer. Ultrasonic DNA fragmentation was performed on a Covaris M220 sonicator (Shelton, USA) in a volume of 50 µl using a protocol optimized for average fragment lengths of 300 bp. DNA concentration was measured using a Qubit4 fluorometer with the DNA HS kit (Thermo Fisher Scientific, USA). Genomic libraries were prepared using the KAPA Hyper Prep kit (Roche, Switzerland) with KAPA double indices according to the manufacturer's instructions. Fifty nanograms of fragmented DNA were used for the experiment, with nine cycles of final PCR. The quality and molarity of the obtained libraries were calculated after analysis on a BA2100 bioanalyzer (Agilent, USA) and measurement of concentrations on a Qubit4 fluorometer. After normalization, the resulting libraries were pooled to a concentration of 4 nM/µl. Sequencing was performed on a GenolabM device (GeneMind, China) using 2×150 bp paired-end reads with the GenolabM V2.0 kit (FCM 300) according to the manufacturer's protocol.

Bioinformatic analysis. The genome sequencing results were quality checked using FastQC (https://www.bioinformatics.babraham.ac.uk/projects/fastqc/), trimming was performed using fastp (Chen et al., 2018; Chen, 2023). Filtering for possible contamination by human reads (Hg38) was performed using BWA MEM (Li, 2013) and synthetic sequences (Univec (https://ftp.ncbi.nlm.nih.gov/pub/UniVec/, accessed April 2025) and blastn).

The taxonomic composition of the sequenced material was analyzed using the MetaPhlan4 tool (Blanco-Míguez et al., 2023). The initial taxonomic identification of the draft primary genome assemblies was performed by tetracorrelation search using the JSpeciesWS web service with the GenomesDB database (Richter et al., 2016). The draft primary assemblies were obtained in the same way as described below, except that Ragout scaffolding (Kolmogorov et al., 2014) and Pilon correction (Walker et al., 2014), were not used for the draft primary assemblies, and the results of the tetra-correlation search during primary taxonomic identification were used as reference genomes. The GFinisher assembler accepts multiple assemblies as input data (Guizelini et al., 2016). The authors of GFinisher note that the first assembly is the main one, and the rest are additional. In our case, depending on the quality control of all assemblies, the main assembly in GFinisher was one of the following: bins, the MinYS assembly (Guyomar et al., 2020), or the Spades assembly (Bankevich et al., 2012; Prjibelski et al., 2020) after mapping to the reference.

Next, taxonomic identity was refined in order to select the closest reference genomes for corrected bacterial genome assembly using a hybrid (*de novo* + reference-guided) cascade approach based on a set of closely related references (Fig. 1). To search for refined reference genomes, average nucleotide identity (ANI) analysis was performed using fastani (Jain et al., 2018) within the genus established during the initial taxonomic screening. Draft primary assemblies were compared by ANI with all publicly available genome assemblies within the same genus deposited in NCBI Genbank, which were downloaded in batch processing using ncbi-genome-download (https://zenodo.org/records/8192486, accessed April 2025).

Three draft assemblies were obtained. The first option was a *de novo* assembly of Spades. The assembly was binned using MaxBin2 (Wu et al., 2016). The resulting bins were filtered by completeness. If this parameter was <87 %, such bins were not considered further. The second assembly variant was a reference-guided assembly using MinYS. A closely related reference genome and filtered reads were fed into the input. Based on the results of the MinYS assembly, the main reference was selected from the closest reference genomes and used in the third assembly option and in the final assembly with correction. The third assembly option was performed by the Spades assembler based on mapped reads to the selected reference genome.

All draft assembly variants were scaffolded using Ragout, which can utilize information about a number of the closest reference genomes. Configuration files for each Ragout assembly variant were generated using a custom Python 3 script. The resulting draft assemblies at the contig and scaffold level were merged into a single assembly using the GFinisher program. Depending on the intermediate quality control results, the main assembly in GFinisher was either the Ragout assemblies or the bins obtained in the first draft assembly variant. GFinisher can output two assemblies, which were submitted for correction in Pilon. At each stage, the quality of the assemblies was controlled using QUAST (Gurevich et al., 2013). Based on the results of quality control, which was carried out after correction, the final genome assemblies were selected.

Genome annotation was performed using the prokka software pipeline (Seemann, 2014), additionally configured to

[1] Supplementary Materials 1–6 are available at:
https://vavilovj-icg.ru/download/pict-2025-29/appx52.zip

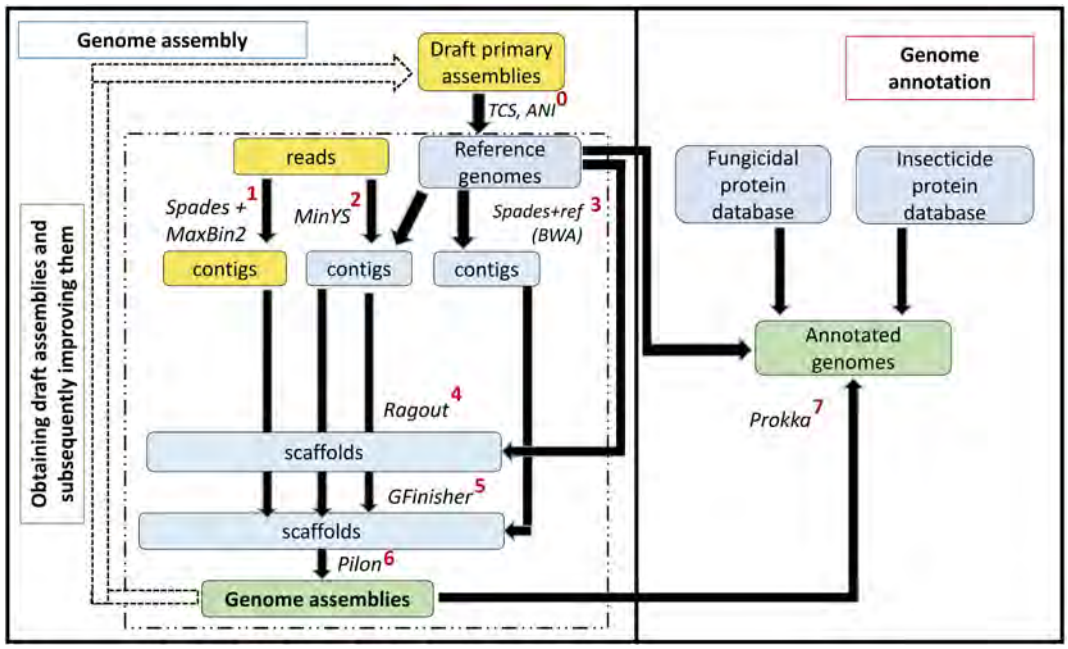

Fig. 1. Analytical pipeline for assembling bacterial genomes and annotating them.

Yellow blocks indicate data obtained at the preliminary stage of primary genome assembly, blue blocks indicate data obtained during genome assembly refinement, and green blocks indicate target results. Arrows indicate data interdependencies, and labels next to the arrows indicate the tools and methods used for data analysis or conversion. 0 – tetracorrelation analysis (JSpeciesWS and GenomesDB) and average nucleotide identity analysis (fastani) to establish the closest (and best annotated) reference genomes from GenBank; 1 – *de novo* assembly using Spades followed by binning in MaxBin2 (bin selection filter – completeness > 87 %); 2 – reference-guided assembly in MinYS; 3 – assembly using Spades of reads pre-mapped to the reference using BWA; 4 – scaffolding of previously obtained draft genomes (contigs) in Ragout, using information about the closest reference genome; 5 – genome polishing in GFinisher, combining the results of previously obtained draft assemblies into one common assembly; 6 – final correction in Pilon; 7 – annotation of final assemblies in Prokka, taking into account the most complete known genome annotations of the corresponding species/genus, as well as specially prepared databases of amino acid sequences of fungicidal and insecticidal proteins, the latter including BPPRC.

take into account the databases of protein sequences with insecticidal and fungicidal properties prepared within the framework of this study, as well as the annotation of refined reference genomes.

For each genome, a separate annotation database was created, consisting of the following parts: annotated protein sequences from the closest reference genome, protein sequences with insecticidal and fungicidal properties collected in this study (described below), and the BPPRC database (https://www.bpprc-db.org/, accessed June 2025).

To search for insecticidal proteins in the genomes under study, a database of protein sequences was compiled manually. It included the amino acid sequences of Cry and Vip proteins. The sequences were selected from the UniprotDB (Bateman et al., 2021) and NCBI Protein (https://www.ncbi.nlm.nih.gov/protein/, accessed October 2024) databases by genus name and protein name/function. In the complete genomic sequences of bacteria from the GenBank card, the protein product was found and the corresponding protein sequence was downloaded. Then, a search was performed using blastp on the formed database. The results were filtered by a threshold for two parameters: identity > 50 % and e-value < 0.001, where identity corresponds to the percentage of matching amino acids, and e-value corresponds to the statistical significance of the results. The BPPRC database, containing protein sequences of bacterial pesticidal proteins downloaded from the Bacterial Pesticidal Protein Database project, was also added to this variant.

Similarly, a database of protein sequences was compiled manually for certain proteins with fungicidal activity: iturins, fengycins, and surfactins. Using blastp, the results of alignment of proteins encoded in the genomes of the analyzed strains to the sequences from the compiled database were obtained. Initial filtering of the results was performed according to two parameters: identity > 50 % and e-value < 0.001.

Figure 1 shows a graphical diagram of the corresponding analytical pipeline. Venn diagrams for comparing the gene repertoire of the analyzed strains with that of the reference genomes were constructed using the Draw Venn Diagram web service (http://bioinformatics.psb.ugent.be/webtools/Venn/, accessed May 2025).

Subsequent analysis of the alignment results of the collected genomes based on insecticidal and fungicidal proteins was performed by visualizing information about the proteins found in the studied strains in the form of a heat map based on the iScore value (see Equation 1), which represents the percentage of identity weighted by the proportion of the aligned region to the total length of the reference protein. This representation allows us to take into account not only the percentage of identity of the local alignment found, but also the extent to which this alignment covers the length of the original protein.

The formula for calculating iScore is as follows:

$$\text{iScore} = \text{identity} \cdot \frac{\text{la}}{\text{ls}}, \quad (1)$$

where identity is the percentage of matching amino acids, la is the alignment length, and ls is the initial length of the amino acid sequence of the protein from the collected database.

Gene repertoire analysis was performed using GenAPI (Gabrielaite, Marvig, 2020).

Multiple alignment of concatenated amino acid sequences was performed using the GTDB-Tk tool (Chaumeil et al., 2022) based on a search for 120 bacterial marker genes. Multiple alignment was used to construct a phylogenetic tree in PhyML (Guindon et al., 2010). In PhyML, the default support level for internal branches is estimated using a Bayesian test. The iTOL web service (Letunic, Bork, 2024) (https://itol.embl.de/, accessed May 2025) was used to visualize the tree.

Program versions and launch parameters are listed in Supplementary Material 2.

Results

Taxonomic identification

A preliminary assessment of the taxonomic composition of sequenced genetic material using Metaphlan revealed heterogeneity in most of the analyzed samples to varying degrees (Table 1). Since not all reads for all samples belonged to a single organism, it was decided to perform binning of the initial draft genome assemblies to isolate the genomic fragments most fully represented in the microorganism sample. Next, taxonomic identification was performed for the bins using tetranucleotide frequency analysis and correlation coefficients (tetra-correlation search, TCS).

The taxonomic identity of the assembled genomes (see subsection Assembly and annotation) was refined during the analysis of average nucleotide identity by comparison with publicly available reference genomes of the same genus in NCBI Genbank (see Supplementary Material 3). The results of the identification are summarized in Figure 2 and Table 2.

Table 1. Results of the initial analysis of the taxonomic composition of the samples studied and subsequent binning

Sample	Species by Metaphlan, share of reads %	Bin number, bin completeness %	Species according to TCS
BZR 1159	*Brevundimonas naejangsanensis*, 100	1) 001 – 38.3 2) 002 – 57.9	1) *Brevundimonas naejangsanensis* DSM 23858 2) *Brevundimonas naejangsanensis* DSM 23858
BZR 162	*Ochrobactrum quorumnocens*, 78.9 Unclassified, 21.1	2) 002 – 87.9	2) *Ochrobactrum quorumnocens* A44
BZR 201	*Leucobacter aridicollis*, 56.1 Unclassified, 43.5 *Paenochrobactrum gallinarii*, 0.4	1) 001 – 95.3 2) 002 – 99.1	1) *Leucobacter aridicollis* L-9 2) *Paenochrobactrum glaciei* JCM 15115
BZR 206	*Leucobacter aridicollis*, 90.5 Unclassified, 9.5	1) 001 – 95.3 2) 002 – 75.7	1) *Leucobacter aridicollis* DSM 17380 2) *Leucobacter chinensis* NC76-1
BZR 278	*Alcaligenes faecalis*, 92.7 *Brevundimonas naejangsanensis* 0.4 Unclassified, 7.3	1) 001 – 100	1) *Alcaligenes nematophilus* A-TC2
BZR 466	*Leucobacter aridicollis*, 69.8 *Brevundimonas naejangsanensis*, 19.3 Unclassified, 11.5	1) 001 – 95.3 2) 002 – 95.3	1) *Leucobacter aridicollis* DSM 17380 2) *Brevundimonas naejangsanensis* DSM 23858
BZR 585	*Alcaligenes faecalis*, 91.1 Unclassified, 8.9	1) 001 – 100	1) *Alcaligenes nematophilus* A-TC2
BZR 588	Unclassified, 95.8 *Paenochrobactrum gallinarii* SGB85919, 4.2	1) 001 – 99.1 2) 002 – 94.4 3) 003 – 84.1	1) *Paenochrobactrum glaciei* JCM 15115 2) *Leucobacter chinensis* NC76-1 3) *Leucobacter komagatae* DSM 8803
BZR 635	Unclassified, 99.9 *Mesorhizobium hungaricum* SGB11031, 0.1	1) 001 - 38.3 2) 002 – 57.9	1) *Leucobacter* sp. G161 2) *Leucobacter* sp. G161 (*Leucobacter komagatae*)
BZR 736	*Bacillus cereus*, 87 Unclassified, 13	1) 002 – 99.1	1) *Bacillus cereus* BAG1O-3
BZR 762	*Alcaligenes faecalis*, 92.1 Unclassified, 7.9	1) 001 – 64.5 2) 002 – 35.5	1) *Alcaligenes nematophilus* A-TC2 2) *Alcaligenes nematophilus* A-TC2
BZR 920	*Bacillus velezensis*, 100	1) 001 – 51.4 2) 002 – 46.7	1) *Bacillus amyloliquefaciens* Bs006 2) *Bacillus velezensis* OB9
BZR 926	*Achromobacter* sp. MYb9, 47.9 Unclassified, 52.1	1) 001 – 74.8 2) 002 – 26.2	1) *Achromobacter marplatensis* B2 2) *Achromobacter marplatensis* HLE
BZR 936	*Bacillus velezensis*, 100	1) 001 – 45.8 2) 002 – 54.2	1) *Bacillus amyloliquefaciens* EGD-AQ14 2) *Bacillus velezensis* OB9

Note. The correspondence of the bin numbers in the third and fourth columns has been preserved.

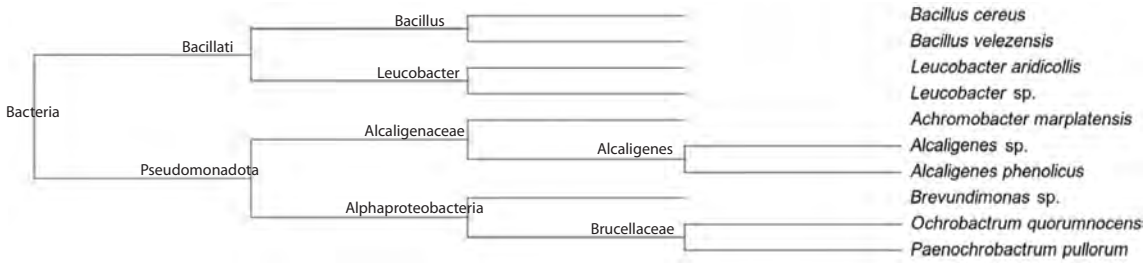

Fig. 2. General taxonomic tree of species represented in samples and selected for further analysis.

Table 2. Results of final taxonomic identification with indication of the NCBI identifier of the closest reference genome

No.	Strain	Taxonomic identity	Nearest related strain	Nearest reference genome
1	BZR 588	*Paenochrobactrum pullorum*	LMG 28095	GCF_041929845.1
2	BZR 1159	*Brevundimonas* sp.	LPMIX5	GCF_003576505.1
3	BZR 920	*Bacillus velezensis*	X-Bio_1	GCF_014230015.1
4	BZR 936	*Bacillus velezensis*	X-Bio_1	GCF_014230015.1
5	BZR 206	*Leucobacter aridicollis*	JUb111	GCF_024806925.1
6	BZR 926	*Achromobacter marplatensis*	LMG 26219	GCF_902859635.1
7	BZR 466	*Leucobacter aridicollis*	JUb111	GCF_024806925.1
8	BZR 736	*Bacillus cereus*	CMCC P0011	GCF_001635955.1
9	BZR 635	*Leucobacter* sp.	G161	GCF_001482305.1
10	BZR 762	*Alcaligenes* sp.	YSL9	GCF_030515105.1
11	BZR 278	*Alcaligenes* sp.	YSL9	GCF_030515105.1
12	BZR 201	*Leucobacter aridicollis*	JUb111	GCF_024806925.1
13	BZR 201P*	*Paenochrobactrum pullorum*	LMG 28095	GCF_041929845.1
14	BZR 162	*Ochrobactrum quorumnocens*	A44	GCF_002278035.1
15	BZR 585	*Alcaligenes phenolicus*	JC896	GCF_040430305.1

Note. Since two genomes were isolated from sample BZR 201, the second strain belonging to *Paenochrobactrum pullorum* from sample BZR 201 is designated BZR 201P (* in the second column).

As a result, genomes were assembled and 15 bacterial strains were taxonomically identified. Some of the samples had readings that could be attributed to different bacteria, which may indicate contamination of the samples prior to sequencing or that the culture was mixed.

For example, since *Paenochrobactrum pullorum* is found in samples taken from wheat roots (BZR 588) and winter rapeseed roots (BZR 201P), we conclude that this is not accidental contamination, but a normal representative of the rhizosphere of flowering plants. Thus, in the study (Hussain et al., 2025), this species was used as a part of a community of bacteria that affect phosphorus availability, yield, and wheat growth.

Assembly and annotation

As a result of the hybrid (*de novo* + reference-guided) approach, assemblies of 15 bacterial genomes were obtained. The main characteristics of the final assemblies obtained after QUAST quality control are presented in Table 3. Additionally, for comparison, the table shows such reference characteristics as assembly length (reference genome sequence length), CG composition of the reference genome, and reference-dependent indicators: the proportion of the reference represented in the studied genome (Genome fraction) and the duplication ratio (Duplication ratio).

However, the approach presented was not suitable for samples BZR 635 and BZR 926. Although closely related references were found for these samples, the Genome fraction indicator was low during assembly (4.274 and 1.511 %, respectively), which made it impossible to perform successful scaffolding in Ragout. Therefore, for BZR 635 and BZR 926, the assembly scheme consisted of the following steps. The primary assembly was a *de novo* Spades assembly without binning (the bins did not pass the set threshold). Rough assemblies were performed using MinYS and Spades with mapping to the reference. The steps with Ragout and Pilon were

Table 3. Characteristics of the final genome assembly and reference genome metrics

No.	Strain	Reference assembly length	CG content of reference, %	Genome fraction, %	Duplication ratio	Number of contigs	The largest contig, bp	Total length of assembly, bp	N50	L50	CG content, %
1	BZR 1159	3144717	66.41	87.420	1.000	239	82719	2772387	19497	39	67.17
2	BZR 162	4593578	53.41	88.176	1.002	60	487021	4079562	225616	6	53.23
3	BZR 201	3459535	67.32	93.609	1.000	10	783893	3424442	486624	3	67.45
4	BZR 201P	3664841	48.45	89.886	1.001	53	510935	3452129	153809	7	48.47
5	BZR 206	3459535	67.32	93.600	1.000	13	1015594	3387765	424194	3	67.51
6	BZR 278	4098772	56.49	95.093	1.000	12	1448095	4159417	406907	3	56.45
7	BZR 466	3459535	67.32	93.743	1.005	15	562204	3459663	424178	4	67.37
8	BZR 585	4121371	56.49	94.610	1.000	7	1900810	4183792	871829	2	56.42
9	BZR 588	3664841	48.45	85.527	1.029	60	412949	3238581	144220	8	48.54
10	BZR 635	3554188	65.32	4.274	1.001	69	181203	3024800	66004	16	66.43
11	BZR 736	6097988	34.94	85.014	1.010	16	1408346	5366477	760581	3	35.03
12	BZR 762	4098772	56.49	95.136	1.000	63	486891	3921546	118193	10	56.60
13	BZR 920	3861135	46.54	95.678	0.999	70	267487	3714855	123313	10	46.84
14	BZR 926	6885835	65.11	1.511	1.001	43	737359	6214265	318995	7	64.92
15	BZR 936	3861135	46.54	95.657	1.000	74	283481	3714091	123326	10	46.85

skipped, since they require the locations of the regions in the assembly relative to the reference. The resulting assemblies were submitted to Gfinisher, the result of which was selected as the final assembly.

Genome annotation was performed using the Prokka tool with a user database of annotated amino acid sequences taken from the reference genome, as well as databases of fungicidal and insecticidal proteins. A total of 10 custom databases were used, as the reference genomes coincided for a number of strains. Thus, for strains BZR 201, BZR 206, and BZR 466, the reference is GCF_024806925.1; for BZR 201P and BZR 588 – GCF_041929845.1; for BZR 278 and BZR 762 – GCF_030515105.1; and for BZR 920 and BZR 936 – GCF_014230015.1. Individual user databases were used for the remaining samples.

Table 4 shows info on the annotation of bacterial genomes compared to the corresponding reference genome. The "Predicted proteins" column shows the percentage of predicted proteins relative to all genes predicted in the genome, with the rest annotated as "hypothetical protein". Although the Genome fraction was low for the BZR 635 and BZR 926 genomes, we used the reference data as annotation.

Analysis of the gene repertoire using GenAPI revealed differences between the analyzed strains and the reference genomes used (Fig. 3).

The genomes of the bacteria were annotated, and the gene repertoire of each genome was analyzed relative to the corresponding reference genomes. Analysis of the intersections of gene sets for different samples showed varying results. For example, strains BZR 920 and BZR 936 showed high similarity to *B. velezensis*, and strains BZR 206, BZR 466, and BZR 201 showed high similarity to *L. aridicollis*. However, the fact that 160 genes were found only in the reference genome of *B. velezensis* but not in the studied strains of this species may be due to both the fragmentation of the BZR 920 and BZR 936 genome assemblies and the accumulated changes in the sequences, based on the homology of which the absence and presence of genes is assessed. It is also possible that these strains do not actually have a number of genes relative to the selected reference genome. These considerations apply to all results presented in Figure 3.

BZR 585 showed a high degree of similarity between the gene repertoire of the studied strains and the corresponding reference genomes in relation to *A. phenolicus*; BZR 762 and BZR 278 in relation to *Alcaligenes* sp.; and BZR 588 and BZR 201P in relation to *P. pullorum*. In addition, BZR 736 showed a good degree of correspondence with *B. cereus*, BZR 1159 with *Brevundimonas* sp., and BZR 162 with *O. quorumnocens*. At the same time, BZR 926 showed only an average level of correspondence with the reference genome of *A. marplatensis*, and BZR 635, when compared with the closest reference genome, *Leucobacter* sp., showed only a small number of intersections, which may be due both to different taxonomic identity and to the fragmentation of the genome assembly of this strain.

Comparison of strains, analysis of their insecticidal and fungicidal potential

Below are the results of the analysis of the insecticidal (Fig. 4) and fungicidal (see figure in Supplementary Material 5) potential of the studied strains based on a comparison of their protein repertoires with the corresponding functional activity.

Table 4. Characteristics of bacterial genome annotations from the analyzed samples compared with the corresponding reference genomes

Strain/reference	Identifier	Total length of assembly, bp	Genes	CDS	Predicted proteins, %
Strain	BZR 1159	2772387	2692	2655	79.16
Reference	GCF_003576505.1	3144717	3026	2953	83.58
Strain	BZR 162	4079562	3831	3786	78.86
Reference	GCF_002278035.1	4593578	5344	5118	95.92
Strain	BZR 201	3424442	3088	3034	80.47
	BZR 206	3387765	3052	2997	81.26
	BZR 466	3459663	3147	3092	79.41
Reference	GCF_024806925.1	3459535	3152	3079	81.63
Strain	BZR 201P	3452129	3277	3227	84.50
	BZR 588	3238581	3031	2991	86.26
Reference	GCF_041929845.1	3664841	3523	3355	89.64
Strain	BZR 278	4159417	3862	3807	85.81
	BZR 762	3922022	3606	3551	89.57
Reference	GCF_030515105.1	4098772	3753	3677	93.15
Strain	BZR 585	4184220	3863	3807	85.63
Reference	GCF_040430305.1	4121371	3752	3679	93.58
Strain	BZR 635	3024800	2707	2652	77.39
Reference	GCF_001482305.1	3554188	3404	3285	85.40
Strain	BZR 736	5366477	5456	5407	79.73
Reference	GCF_001635955.1	6097988	6148	5781	87.64
Strain	BZR 920	3715092	3643	3579	88.28
	BZR 936	3714897	3640	3570	88.41
Reference	GCF_014230015.1	3861135	3828	3691	92.74
Strain	BZR 926	6214265	5705	5638	85.17
Reference	GCF_902859635.1	6885835	6393	6280	90.94

Heat maps reflect the results of the search for fungicidal and insecticidal proteins among our strains in the form of iScore, which is the percentage of identity weighted by the proportion of the aligned region of the entire length of the reference protein (see Materials and methods).

A gene encoding the chaperonin GroEl was found in all analyzed strains (Fig. 4). At the same time, the samples belonging to the genus *Bacillus* (BZR 736, BZR 920, BZR 936) had the highest number of genes encoding insecticidal proteins. While strains BZR 920 and BZR 936, which are representatives of *B. velezensis*, demonstrate the presence of fengycin, mycosubtilin, and bacillomycin synthetases, BZR 736, belonging to *B. cereus*, possesses the genes Spp1Aa1 and Spp1Aa2, as well as Vpb1Ab1 and Vpb4Ca1. At the same time, all analyzed strains of the genus *Bacillus* also possess a gene encoding the enzyme N-acetylmuramic acid 6-phosphate esterase (EC 4.2.1.126). In addition, strains BZR 278, BZR 585, and BZR 762, belonging to the genus *Alcaligenes*, have fragments homologous to the gene for the bifunctional protein HldE, while strains BZR 201, BZR 206, and BZR 466, representing *Leucobacter aridicollis*, contain a number of fragments homologous to bacillomycin synthetase genes.

Proteins with fungicidal activity were identified (see Supplementary Material 5; the heat map was constructed based on the values specified in Supplementary Material 6) in samples belonging to the genus *Bacillus* (BZR 736, BZR 920, BZR 936). It can be seen that the most complete proteins matching the sequences from the database are present in two samples: BZR 920 and BZR 936, belonging to *B. velezensis*. These samples also lead in terms of the number of proteins

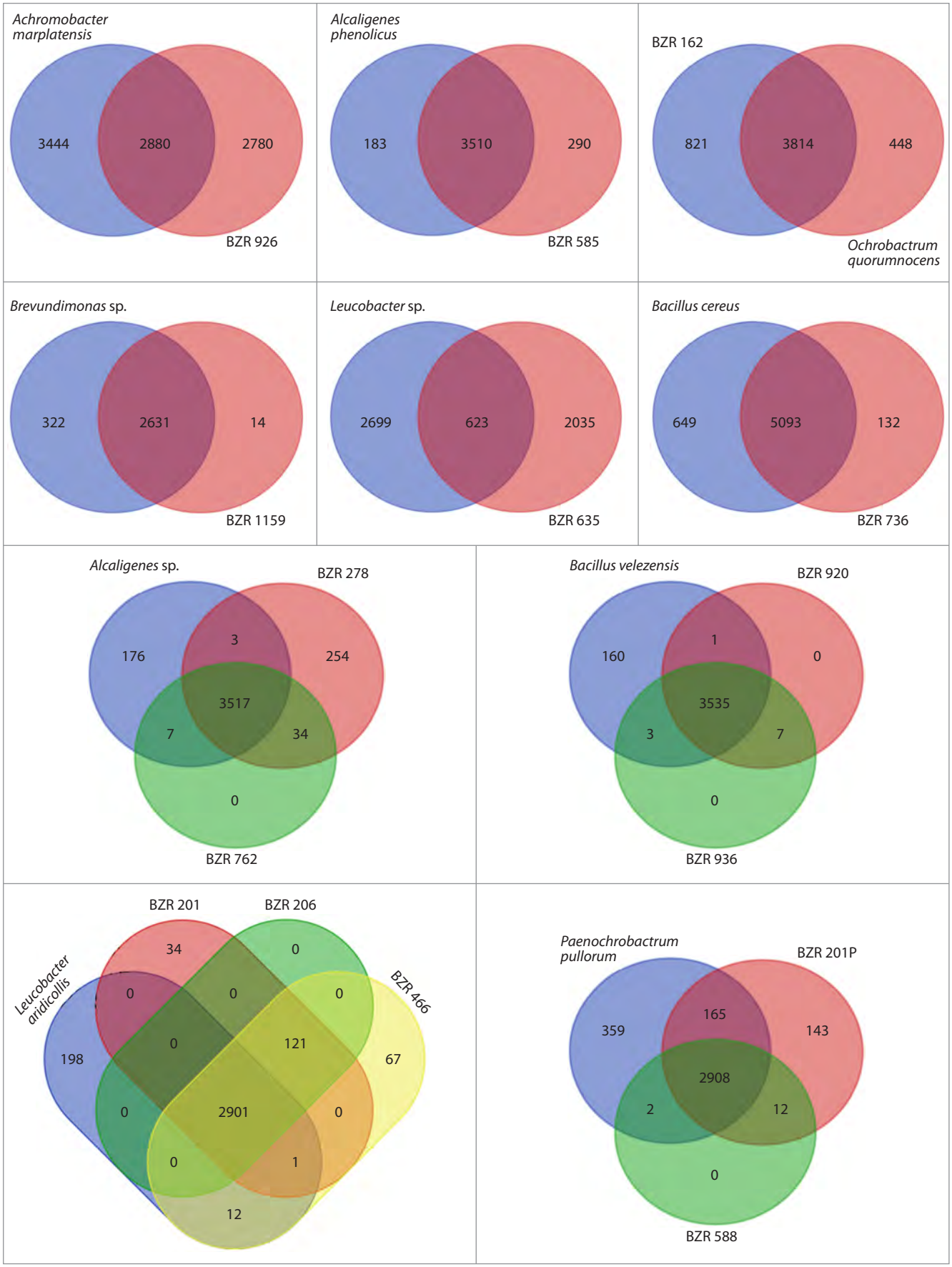

Fig. 3. Venn diagrams for gene sets of the studied strains, grouped by species, analyzed together with gene sets from the corresponding reference genomes.

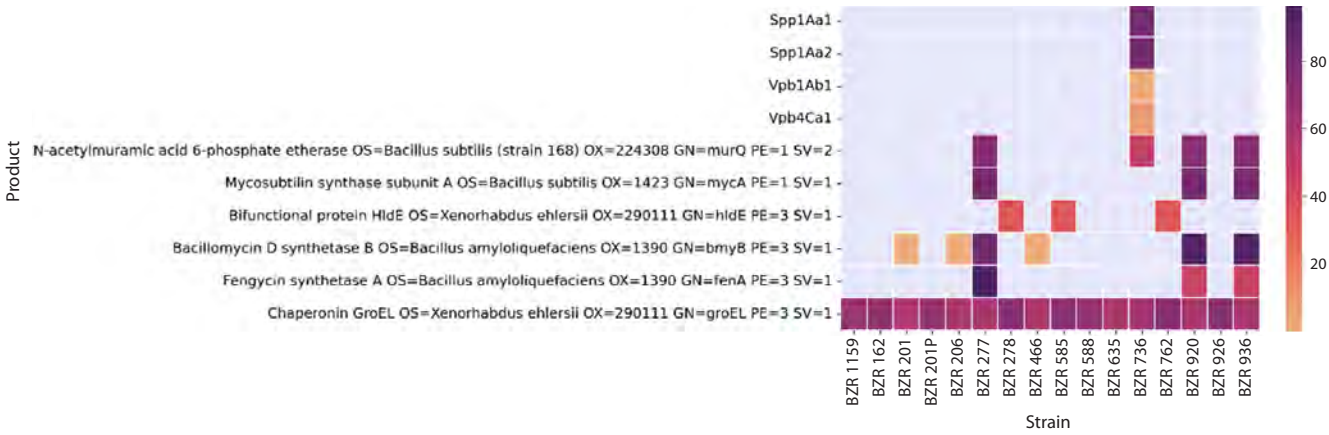

Fig. 4. Repertoire of proteins selected during pairwise alignment of translated genomes of the studied strains to the insecticidal protein database, where the color scale reflects the iScore parameter (see Equation 1). That is, the more intense the color, the more completely the sequence from the studied genome was aligned with the amino acid sequence of the protein from the database. Missing proteins are marked in grey. The heat map is based on the values (see Supplementary Material 4).

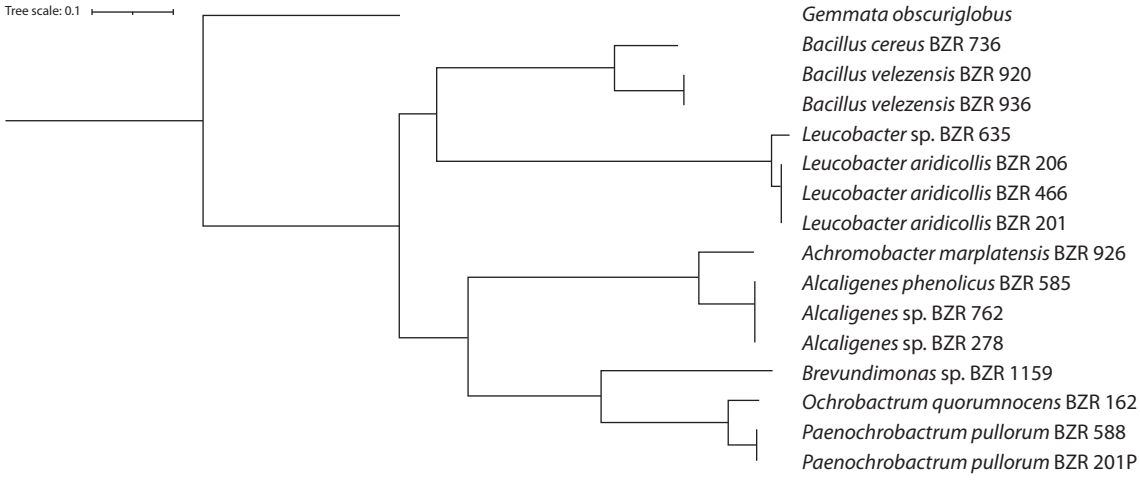

Fig. 5. Phylogenetic tree of the studied strains, constructed based on 120 marker genes found in the bacteria. *Gemmata obscuriglobus* serves as the outgroup.

found. Since the vast majority of sequences in the database of amino acid sequences of fungicidal proteins belonged to the species *B. velezensis* and *B. amyloliquefaciens*, the discovery of proteins in these two samples is a logical consequence. The predominance of data in the collected database of sequences related to *B. velezensis* and *B. amyloliquefaciens* was due to the predominance of data on iturins, fengycins, and surfactins of *B. velezensis* and *B. amyloliquefaciens* relative to other species of the genus *Bacillus* in the UniprotDB and NCBI Protein databases. However, genes encoding synthetases, as well as *YxjF* and *YxjC*, were also found in a strain belonging to *B. cereus* (BZR 736), although unlike the other strains analyzed, BZR 736 does not have the *ScoA* gene, which, along with *ScoB*, is present in all the strains analyzed. In addition, strains BZR 201, BZR 206, and BZR 466, representing *L. aridicollis*, contain a number of fragments homologous to the genes of iturin and fengycin synthetases. It is worth mentioning that the genomes of the entire clade of representatives of the Brucellaceae family (strains BZR 162, BZR 588, and BZR 201P) do not contain genes encoding fungicidal proteins represented in our database.

Phylogenetic analysis

To construct a phylogenetic tree (Fig. 5), a search for a set of 120 bacterial marker genes was performed using GTDB-Tk in all studied samples, and multiple alignment of protein sequences was performed, which was fed into the tree construction. Proteins from *Gemmata obscuriglobus* (GCF_003149495.1) were used as an outgroup.

It should be noted that the constructed phylogenetic tree (see Fig. 5) fully reflects the topology of the previously presented taxonomic tree (see Fig. 2), which relatively confirms the correctness of the established taxonomic identity. At the same time, the gene sequences used to construct the tree are identical in strains BZR 762 and BZR 278, as well as in the trio of strains BZR 466, BZR 206, BZR 201, indicating that the former belong to one species of the genus *Alcaligenes*, and the latter, to *L. aridicolis*.

Discussion

The study of endophytic bacteria in agriculturally important crops is a pressing issue, the resolution of which will enable the regulation of pest populations and effective control of plant diseases arising from interaction with pathogens. Therefore, it is particularly important to identify strains that possess fungicidal and insecticidal potential.

To date, the following antagonistic properties of endophytic bacteria are known for the studied genera. Here, we list the genera of bacteria for which taxonomic identification of genomes was performed in this study.

***Bacillus* spp.** The genomes of bacteria of the genus *B. velezensis* contain nine key NRPS/PKS clusters, synthesized lipopeptides (iturin, fengycin, surfactin) and polyketides (deficidin, macrolactin) capable of suppressing phytopathogenic fungi of the genera *Fusarium*, *Bipolaris*, *Exserohilum*, and others, both *in vitro* and *in vivo* (Wang S. et al., 2024; Yeo et al., 2024). Surfactin, iturin, and fengycin can act as entomicides and nematicides, as evidenced by the high mortality (up to 100 %) of *Aedes aegypti* mosquito larvae and *Agriotes lineatus* click beetles, while pure surfactin causes systemic metabolic disorders in the caterpillars of the Asian cotton bollworm *Spodoptera litura* (Falqueto et al., 2021; Zhang F. et al., 2024; Knežević et al., 2025).

***Alcaligenes* spp.** A binary protein, AfIP-1A/1B, has been discovered in *A. faecalis* bacteria that is capable of forming pores in the intestines of western corn rootworm larvae *Diabrotica virgifera*, including in insect populations resistant to *B. thuringiensis*. Some strains produce an exoprotease that kills the root-knot nematode *Meloidogyne incognita* and the soil nematode *Caenorhabditis elegans*. The *A. faecalis* N1-4 strain is capable of producing dimethyl disulfide and methyl isovalerate, which inhibit the growth of the fungus *Aspergillus flavus* and reduce the amount of mycotoxins in grain during storage (Ju et al., 2016; Gong et al., 2019; Pérez Ortega et al., 2021).

***Achromobacter* spp.** The *A. xylosoxidans* soil strain causes 95 % mortality of larvae and 100 % mortality of adults of the housefly *Musca domestica*. Volatile esters (S-methylthiobutyrate, acetates) reduce the population of the gall nematode *M. javanica* and suppresse gall formation on tomatoes by 60 %. The endophytic strain CTA8689 reduces melon wilt caused by *Fusarium oxysporum*/*F. solani* by 60 % in a greenhouse thanks to siderophores and esterases (Yamaç et al., 2010; Dhaouadi et al., 2019; Deng et al., 2022; Mohamadpoor et al., 2022).

***Brevundimonas* spp.** *B. diminuta* YYC02 produced 42 volatile compounds, of which butyl-2-methylbutanoate and isoamyl butyrate caused 90–100 % mortality of the root-knot nematode *Meloidogyne javanica* within 48 hours, while soil treatment reduced the number of galls by 37 % and increased the mass of cucumber shoots (Sun et al., 2023). *Brevundimonas* is part of the core microbiome of entomopathogenic nematodes (Steinernema, Meloidogyne) and actively adheres to the cuticle of J2 larvae, increasing their mortality and reducing egg hatching, which indirectly confirms the anti-PPN activity of the genus (Topalović et al., 2019). Although no classical Cry/Cyt toxins have been found in *Brevundimona*, a membrane organophosphate hydrolase has been studied in detail in *B. diminuta*, allowing the strain to use organophosphate insecticides as a source of phosphorus; the enzyme is localized in the periplasm by means of a Tat signal (Parthasarathy et al., 2016).

***Leucobacter* spp.** Two strains, Verde1/Verde2, cause nematode death through a rare mechanism in which a sticky exopolymer sticks the nematode tails together into “stars”, leading to the death of the colony. Nematode resistance to these strains is controlled by individual genes, which emphasizes the specialized virulence of the genus. The *L. aridicollis* SASBG215 strain inhibits cucumber anthracnose *Colletotrichum orbiculare* and causes lysis of hyphae, presumably with the help of unknown polyketides (Hodgkin et al., 2013; Abdul Salam et al., 2022).

***Ochrobactrum* spp.** Strain BS-206 synthesizes the glycolipid ochrozin, which kills 90–100 % of storage pests (*Tribolium*, *Sitophilus*, *Callosobruchus*) and has an insecticidal effect against the corn borer *Spodoptera*. The *O. pseudogrignonense* NC1 strain produces dimethyl disulfide and benzaldehyde, which cause up to 100 % mortality of young *M. incognita* individuals and reduce tomato gall formation by more than 60 % (Kumar et al., 2014; Yang T. et al., 2022).

As can be seen from the brief summary above, there is a variety of active substances for different pathogens, which can complicate mass analysis, but offer prospects for future research on little-studied endophytes that may have similar properties.

In this article, we presented 15 genomes of endophytic bacteria isolated from various parts of wheat, soybeans, and rapeseed. We have demonstrated the presence of genes responsible for insecticidal and fungicidal activity in the studied strains, with the largest number of genes encoding insecticidal and fungicidal proteins found in strains BZR 736, BZR 920, and BZR 936 of the genus *Bacillus*. However, for strains BZR 162, BZR 588, and BZR 201P of the Brucellaceae family, no genes encoding fungicidal proteins present in our database were identified.

This may indicate several factors: firstly, bacteria of the genus *Bacillus* are more widely studied and published, and secondly, the database contains a bias in data on insecticidal and fungicidal proteins for this genus. Therefore, when interpreting the results obtained for less represented genera, we make the reservation that the presence of such proteins is possible in these genomes, but their sequences differ significantly from the publicly available sequences that we were able to aggregate into the database. Nevertheless, the bias in the data for the genus *Bacillus* does not affect the search for proteins with insecticidal and fungicidal properties. There is also little aggregated information in the literature on insecticidal and fungicidal proteins, which confirms the need to create databases of proteins associated with insecticidal and fungicidal properties.

Nevertheless, the genomic data obtained in this study are consistent with the results of bioassays previously conducted for some of the strains studied. According to the initial screening of the bioresource collection, strains with pronounced entomopathogenic activity against the wax moth *Galleria mellonella* were identified. The strains BZR 1159, BZR 588, BZR 936, BZR 206, BZR 920, BZR 926, and BZR 277 (65–95 % mortality on the third day and 83–95 % on the fifth day) (Gyrnets (Bondarchuk), Asaturova, 2022).

With regard to the large mealworm beetle *Tenebrio molitor*, strains BZR 201, BZR 278, BZR 1159, BZR 635, BZR 762, BZR 736 showed themselves to be effective (72–98 % on the third day and 81–99 % on the fifth day). In addition, the potential multifunctionality of a number of strains was assessed: BZR 1159, BZR 936, and BZR G3, which showed insecticidal activity against the natural population of the codling moth *Cydia pomonella*, reaching 95.5 %, as well as fungicidal activity against apple disease pathogens of the genera *Fusarium*, *Alternaria*, *Trichothecium*, with mycelium inhibition of up to 84.8 % (Gyrnets (Bondarchuk), Asaturova, 2022; Gyrnets, Asaturova, 2023). The BZR 936 strain is worth noting: it has both insecticidal and fungicidal properties, which directly correlates with its identified lipopeptide synthase genes (iturin, fengycin, surfactin) and other biocontrol markers. Thus, comparison of the genomic composition with experimental biotests confirms that the presence of specific insecticidal protein genes and lipopeptide synthetase clusters is a reliable indicator of the biocontrol potential of strains. This opens up prospects for their further inclusion in programs for the development of biological products for the protection of agricultural crops.

Conclusion

In this article, we present the taxonomic identification, assemblies, and annotations of 15 endophytic bacterial genomes, the samples of which were obtained from the roots and/or leaves of wheat, soybean, and rapeseed.

Taxonomically, strains BZR 736, BZR 920, and BZR 936 belong to the genus *Bacillus*, strains BZR 635, BZR 466, BZR 206, and BZR 201 belong to the genus *Leucobacter*, and the remaining strains belong to the phylum Pseudomonadota. The phylogenetic tree constructed from a set of 120 bacterial marker genes fully reflects the topology of our taxonomic tree, which confirms the correctness of the established taxonomic identity at a relative level.

Genome assembly was performed in two stages: a preliminary stage of primary genome assembly and a hybrid (*de novo* + reference-guided) cascade approach based on a set of closely related references. Genome annotation was performed taking into account the databases of protein sequences with insecticidal properties prepared within the framework of this study, including the BPPRC database, and fungicidal properties, as well as the most complete known genome annotations of the corresponding species/genus.

Analysis of the gene repertoire revealed differences between the analyzed strains and the reference genomes used. A high degree of correspondence between the gene repertoire of the studied strains and the corresponding reference genomes was shown by BZR 585 in relation to *A. phenolicus*, by BZR 762 and BZR 278 in relation to *Alcaligenes* sp., by BZR 588 and BZR 201P with respect to *P. pullorum*, by BZR 920 and BZR 936 in relation to *B. velezensis*, and by BZR 206, BZR 466 and BZR 201 with respect to *L. aridicollis*. This indicator may also confirm the correctness of the established taxonomic identity. At the same time, when compared with the closest reference genome, *Leucobacter* sp., BZR 635 shows only a small number of intersections, which may be due to both its distinctive taxonomic identity and the fragmentation of the genome assembly of this strain.

We demonstrated the presence of genes encoding fungicidal and insecticidal proteins in all strains except BZR 162, BZR 588, and BZR 201P. No genes encoding fungicidal proteins present in our database were identified for these strains. However, the results obtained in this study indicate that the strains under study, which possess a complex of lipopeptide synthetase and insecticidal toxin genes, demonstrate an experimentally confirmed broad spectrum of biological activity against insects and phytopathogens.

Further targeted study of endophytic bacteria with fungicidal and insecticidal genes opens up prospects for identifying candidates for biocontrol agents of various pathogens and using bacteria to protect agricultural plants.

References

Abdul Salam S., Parthiban V.K., Paranidharan V., Johnson I., Karthikeyan M., Kavitha C. *Leucobacter aridicollis* strain SASBG215: a novel biocontrol agent against *Colletotrichum orbiculare*. *Biol Forum Int J.* 2022;14(2):905-911

Ali M.A., Ahmed T., Ibrahim E., Rizwan M., Chong K.P., Yong J.W.H. A review on mechanisms and prospects of endophytic bacteria in biocontrol of plant pathogenic fungi and their plant growth-promoting activities. *Heliyon*. 2024;10(11):e31573. doi 10.1016/j.heliyon.2024.e31573

Bankevich A., Nurk S., Antipov D., Gurevich A.A., Dvorkin M., Kulikov A.S., Lesin V.M., … Sirotkin A.V., Vyahhi N., Tesler G., Alekseyev M.A., Pevzner P.A. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. *J Comput Biol.* 2012;19(5):455-477. doi 10.1089/cmb.2012.0021

Bateman A., Martin M.J., Orchard S., Magrane M., Agivetova R., Ahmad S., Alpi E., … Vinayaka C.R., Wang Q., Wang Y., Yeh L.S., Zhang J. UniProt: the universal protein knowledgebase in 2021. *Nucleic Acids Res.* 2021;49(D1):D480-D489. doi 10.1093/nar/gkaa1100

Berry C., O'Neil S., Ben-Dov E., Jones A.F., Murphy L., Quail M.A., Holden M.T.G., Harris D., Zaritsky A., Parkhill J. Complete sequence and organization of pBtoxis, the toxin-coding plasmid of *Bacillus thuringiensis* subsp. *israelensis*. *Appl Environ Microbiol.* 2002;68(10):5082-5095. doi 10.1128/AEM.68.10.5082-5095.2002

Blanco-Míguez A., Beghini F., Cumbo F., McIver L.J., Thompson K.N., Zolfo M., Manghi P., … Franzosa E.A., Pasolli E., Asnicar F., Huttenhower C., Segata N. Extending and improving metagenomic taxonomic profiling with uncharacterized species using MetaPhlAn 4. *Nat Biotechnol.* 2023;41(11):1633-1644. doi 10.1038/s41587-023-01688-w

Bolotin A., Gillis A., Sanchis V., Nielsen-LeRoux C., Mahillon J., Lereclus D., Sorokin A. Comparative genomics of extrachromosomal elements in *Bacillus thuringiensis* subsp. *israelensis*. *Res Microbiol.* 2017;168(4):331-344. doi 10.1016/j.resmic.2016.10.008

Chaumeil P.-A., Mussig A.J., Hugenholtz P., Parks D.H. GTDB-Tk v2: memory friendly classification with the genome taxonomy database. *Bioinformatics*. 2022;38(23):5315-5316. doi 10.1093/bioinformatics/btac672

Chen S. Ultrafast one-pass FASTQ data preprocessing, quality control, and deduplication using fastp. *iMeta*. 2023;2(2):e107. doi 10.1002/imt2.107

Chen S., Zhou Y., Chen Y., Gu J. fastp: an ultra-fast all-in-one FASTQ preprocessor. *Bioinformatics*. 2018;34(17):i884-i890. doi 10.1093/bioinformatics/bty560

Dai R., Zhang J., Liu F., Xu H., Qian J.-M., Cheskis S., Liu W., … de Jonge R., Pieterse C.M.J., Levy A., Schlaeppi K., Bai Y. Crop root bacterial and viral genomes reveal unexplored species and microbiome patterns. *Cell*. 2025;188(9):2521-2539.e22. doi 10.1016/j.cell.2025.02.013

Deng X., Wang X., Li G. Nematicidal effects of volatile organic compounds from microorganisms and plants on plant-parasitic nematodes. *Microorganisms*. 2022;10(6):1201. doi 10.3390/microorganisms10061201

Dhaouadi S., Rouissi W., Mougou-Hamdane A., Nasraoui B. Evaluation of biocontrol potential of *Achromobacter xylosoxidans* against *Fusarium* wilt of melon. *Eur J Plant Pathol.* 2019;154(2):179-188. doi 10.1007/s10658-018-01646-2

Falqueto S.A., Pitaluga B.F., de Sousa J.R., Targanski S.K., Campos M.G., de Oliveira Mendes T.A., da Silva G.F., Silva D.H.S., Soares M.A. *Bacillus* spp. metabolites are effective in eradicating *Aedes aegypti* (Diptera: Culicidae) larvae with low toxicity to non-target species. *J Invertebr Pathol.* 2021;179:107525. doi 10.1016/j.jip.2020.107525

Gabrielaite M., Marvig R.L. GenAPI: a tool for gene absence-presence identification in fragmented bacterial genome sequences. *BMC Bioinformatics*. 2020;21(1):320. doi 10.1186/s12859-020-03657-5

Gong A.-D., Wu N.-N., Kong X.-W., Zhang Y.-M., Hu M.-J., Gong S.-J., Dong F.-Y., Wang J.-H., Zhao Z.-Y., Liao Y.-C. Inhibitory effect of volatiles emitted from *Alcaligenes faecalis* N1-4 on *Aspergillus flavus* and aflatoxins in storage. *Front Microbiol.* 2019;10:1419. doi 10.3389/fmicb.2019.01419

Guindon S., Dufayard J.-F., Lefort V., Anisimova M., Hordijk W., Gascuel O. New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol.* 2010;59(3):307-321. doi 10.1093/sysbio/syq010

Guizelini D., Raittz R.T., Cruz L.M., Souza E.M., Steffens M.B.R., Pedrosa F.O. GFinisher: a new strategy to refine and finish bacterial genome assemblies. *Sci Rep.* 2016;6(1):34963. doi 10.1038/srep34963

Gurevich A., Saveliev V., Vyahhi N., Tesler G. QUAST: quality assessment tool for genome assemblies. *Bioinformatics*. 2013;29(8):1072-1075. doi 10.1093/bioinformatics/btt086

Guyomar C., Delage W., Legeai F., Mougel C., Simon J.-C., Lemaitre C. MinYS: mine your symbiont by targeted genome assembly in symbiotic communities. *NAR Genom Bioinform.* 2020;2(3):lqaa047. doi 10.1093/nargab/lqaa047

Gyrnets (Bondarchuk) E.Yu., Asaturova A.M. Screening of bacterial strains on the criterion of entomopathogenic activity against *Galleria mellonella* L. and *Tenebrio molitor* L. *Dostizheniya Nauki i Tekhniki APK = Achievements of Science and Technology in Agro-Industrial Complex*. 2022;36(3):53-60. doi 10.53859/02352451_2022_36_3_53 (in Russian)

Gyrnets E.Yu., Asaturova A.M. Study of the polyfunctional properties of promising bacterial agents in relation to phytophages and pathogens of fruit cenosis. *Dostizheniya Nauki i Tekhniki APK = Achievements of Science and Technology in Agro-Industrial Complex*. 2023;37(5):39-45. doi 10.53859/02352451_2023_37_5_39 (in Russian)

Hamane S., El Yemlahi A., Hassani Zerrouk M., El Galiou O., Laglaoui A., Bakkali M., Arakrak A. Plant growth promotion and biocontrol potentiality of endophytes isolated from root nodules of *Sulla flexuosa* L. plants. *Int J Agron.* 2023;2023(1):1-9. doi 10.1155/2023/2451806

Hodgkin J., Félix M.-A., Clark L.C., Stroud D., Gravato-Nobre M.J. Two *Leucobacter* strains exert complementary virulence on *Caenorhabditis* including death by worm-star formation. *Curr Biol.* 2013;23(21):2157-2161. doi 10.1016/j.cub.2013.08.060

Horwich A.L., Fenton W.A., Chapman E., Farr G.W. Two families of chaperonin: physiology and mechanism. *Annu Rev Cell Dev Biol.* 2007;23(1):115-145. doi 10.1146/annurev.cellbio.23.090506.123555

Hussain H.S., Tabbasum S., Mahreen N., Yahya M., Ejaz K., Aslam Z., Imtiaz M., ul Islam E., Sajid Z.I., Yasmin S. Integrated application of drought-tolerant phosphate solubilizing bacteria and native P source improves P availability with sustainable wheat production. *J Soil Sci Plant Nutr.* 2025;25(2):4723-4746. doi 10.1007/s42729-025-02425-9

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(1):5114. doi 10.1038/s41467-018-07641-9

Ju S., Lin J., Zheng J., Wang S., Zhou H., Sun M. *Alcaligenes faecalis* ZD02, a novel nematicidal bacterium with an extracellular serine protease virulence factor. *Appl Environ Microbiol.* 2016;82(7):2112-2120. doi 10.1128/AEM.03444-15

Knežević M., Dervišević M., Jovković M., Jevđenović G., Maksimović J., Buntić A. Versatile role of *Bacillus velezensis*: biocontrol of *Fusarium poae* and wireworms and barley plant growth promotion. *Biol Control*. 2025;206:105789. doi 10.1016/j.biocontrol.2025.105789

Kolmogorov M., Raney B., Paten B., Pham S. Ragout – a reference-assisted assembly tool for bacterial genomes. *Bioinformatics*. 2014;30(12):i302-i309. doi 10.1093/bioinformatics/btu280

Krzyżanowska D.M., Maciąg T., Ossowicki A., Rajewska M., Kaczyński Z., Czerwicka M., Rąbalski Ł., Czaplewska P., Jafra S. *Ochrobactrum quorumnocens* sp. nov., a quorum quenching bacterium from the potato rhizosphere, and comparative genome analysis with related type strains. *PLoS One*. 2019;14(1):e0210874. doi 10.1371/journal.pone.0210874

Kumar C.G., Sujitha P., Mamidyala S.K., Usharani P., Das B., Reddy C.R. Ochrosin, a new biosurfactant produced by halophilic *Ochrobactrum* sp. strain BS-206 (MTCC 5720): purification, characterization and its biological evaluation. *Process Biochem.* 2014;49(10):1708-1717. doi 10.1016/j.procbio.2014.07.004

Kupper M., Gupta S.K., Feldhaar H., Gross R. Versatile roles of the chaperonin GroEL in microorganism-insect interactions. *FEMS Microbiol Lett.* 2014;353(1):1-10. doi 10.1111/1574-6968.12390

Letunic I., Bork P. Interactive Tree of Life (iTOL) v6: recent updates to the phylogenetic tree display and annotation tool. *Nucleic Acids Res.* 2024;52(W1):W78-W82. doi 10.1093/nar/gkae268

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv*. 2013. doi 10.48550/arXiv.1303.3997

Liu Y., Yin C., Zhu M., Zhan Y., Lin M., Yan Y. Comparative genomic analysis of *Bacillus velezensis* BRI3 reveals genes potentially associated with efficient antagonism of *Sclerotinia sclerotiorum* (Lib.) de Bary. *Genes* (*Basel*). 2024;15(12):1588. doi 10.3390/genes15121588

Mohamadpoor M., Amini J., Ashengroph M., Azizi A. Evaluation of biocontrol potential of *Achromobacter xylosoxidans* strain CTA8689 against common bean root rot. *Physiol Mol Plant Pathol.* 2022;117:101769. doi 10.1016/j.pmpp.2021.101769

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. doi 10.1093/database/baac022

Parthasarathy S., Parapatla H., Nandavaram A., Palmer T., Siddavattam D. Organophosphate hydrolase is a lipoprotein and interacts with P_i-specific transport system to facilitate growth of *Brevundimonas diminuta* using OP insecticide as source of phosphate. *J Biol Chem.* 2016;291(14):7774-7785. doi 10.1074/jbc.M116.715110

Pellegrinetti T.A., Monteiro G.G.T.N., Lemos L.N., dos Santos R.A.C., Barros A.G., Mendes L.W. *PGPg_finder*: a comprehensive and user-friendly pipeline for identifying plant growth-promoting genes in genomic and metagenomic data. *Rhizosphere*. 2024;30:100905. doi 10.1016/j.rhisph.2024.100905

Pérez Ortega C., Leininger C., Barry J., Poland B., Yalpani N., Altier D., Nelson M.E., Lu A.L. Coordinated binding of a two-component insecticidal protein from *Alcaligenes faecalis* to western corn rootworm midgut tissue. *J Invertebr Pathol.* 2021;183:107597. doi 10.1016/j.jip.2021.107597

Prjibelski A., Antipov D., Meleshko D., Lapidus A., Korobeynikov A. Using SPAdes de novo assembler. *Curr Protoc Bioinform.* 2020;70(1):e102. doi 10.1002/cpbi.102

Rabbee M.F., Ali M.S., Choi J., Hwang B.S., Jeong S.C., Baek K. *Bacillus velezensis*: a valuable member of bioactive molecules within plant microbiomes. *Molecules*. 2019;24(6):1046. doi 10.3390/molecules24061046

Richter M., Rosselló-Móra R., Oliver Glöckner F., Peplies J. JSpeciesWS: a web server for prokaryotic species circumscription based on pairwise genome comparison. *Bioinformatics*. 2016;32(6):929-931. doi 10.1093/bioinformatics/btv681

Seemann T. Prokka: rapid prokaryotic genome annotation. *Bioinformatics*. 2014;30(14):2068-2069. doi 10.1093/bioinformatics/btu153

Shi H., Zeng H., Yang X., Zhao J., Chen M., Qiu D. An insecticidal protein from *Xenorhabdus ehlersii* triggers prophenoloxidase activation and hemocyte decrease in *Galleria mellonella*. *Curr Microbiol*. 2012;64(6):604-610. doi 10.1007/s00284-012-0114-7

Sun Y., Ran Y., Yang H., Mo M., Li G. Volatile metabolites from *Brevundimonas diminuta* and nematicidal esters inhibit *Meloidogyne javanica*. *Microorganisms*. 2023;11(4):966. doi 10.3390/microorganisms11040966

Topalović O., Elhady A., Hallmann J., Richert-Pöggeler K.R., Heuer H. Bacteria isolated from the cuticle of plant-parasitic nematodes attached to and antagonized the root-knot nematode *Meloidogyne hapla*. *Sci Rep*. 2019;9(1):11477. doi 10.1038/s41598-019-47942-7

Walker B.J., Abeel T., Shea T., Priest M., Abouelliel A., Sakthikumar S., Cuomo C.A., Zeng Q., Wortman J., Young S.K., Earl A.M. Pilon: an integrated tool for comprehensive microbial variant detection and genome assembly improvement. *PLoS One*. 2014;9(11):e112963. doi 10.1371/journal.pone.0112963

Wang S., Jin P., Zheng Y., Kangkang W., Chen J., Liu J., Li Y. *Bacillus velezensis* B105-8, a potential and efficient biocontrol agent in control of maize stalk rot caused by *Fusarium graminearum*. *Front Microbiol*. 2024;15:1462992. doi 10.3389/fmicb.2024.1462992

Wang Z., Lu K., Liu X., Zhu Y., Liu C. Comparative functional genome analysis reveals the habitat adaptation and biocontrol characteristics of plant growth-promoting bacteria in NCBI databases. *Microbiol Spectr.* 2023;11(3):e0500722. doi 10.1128/spectrum.05007-22

Wu Y.-W., Simmons B.A., Singer S.W. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics*. 2016;32(4):605607. doi 10.1093/bioinformatics/btv638

Yamaç M., Şahin E., Ceyhan E., Amoroso M.J., Cuozzo S.A., Pilatin S. The screening of bacterial insecticides against *Musca domestica* L. (diptera: Muscidae). *Fresenius Environ Bull.* 2010;19(5):862-870

Yang K., Dai X., Maitikadir Z., Zhang H., Hao H., Yan C. Comparative genome analysis of endophytic *Bacillus amyloliquefaciens* MR4: a potential biocontrol agent isolated from wild medicinal plant root tissue. *J Appl Genet.* 2024;65(4):907-923. doi 10.1007/s13353-024-00905-9

Yang T., Xin Y., Liu T., Li Z., Liu X., Wu Y., Wang M., Xiang M. Bacterial volatile-mediated suppression of root-knot nematode (*Meloidogyne incognita*). *Plant Dis.* 2022;106(5):1358-1365. doi 10.1094/PDIS-06-21-1139-RE

Yeo Y.J., Park A.R., Vuong B.S., Kim J.-C. Biocontrol of *Fusarium* head blight in rice using *Bacillus velezensis* JCK-7158. *Front Microbiol.* 2024;15:1358689. doi 10.3389/fmicb.2024.1358689

Zhang F., Liu Q., Wang Y., Yin J., Meng X., Wang J., Zhao W., Liu H., Zhang L. Effects of surfactin stress on gene expression and pathological changes in *Spodoptera litura*. *Sci Rep.* 2024;14(1):30357. doi 10.1038/s41598-024-81946-2

Zhang M., Kong Z., Fu H., Shu X., Xue Q., Lai H., Guo Q. Rhizosphere microbial ecological characteristics of strawberry root rot. *Front Microbiol.* 2023;14:1286740. doi 10.3389/fmicb.2023.1286740

Conflict of interest. The authors declare no conflict of interest.
Received July 2, 2025. Revised October 17, 2025. Accepted October 18, 2025.