
Functional and Comparative Phylogenomics of Pathogen-Inhibiting Native Mexican *Bacillus velezensis* Across Regional and Global Scales Reveal Open Pan-Genome Architecture and Adaptive Divergence

[Hilda Mabel Sosa-Esquivel](#) , Atzin Fraire-Mayorga , Yadira Yumiko De la Cruz-Rodríguez ,
Angélica Judith Granados-López , Gloria Viviana Cerrillo-Rojas , [Jesús Adrián López](#) , [Saúl Fraire-Velázquez](#)
*

Posted Date: 6 May 2026

doi: 10.20944/preprints202605.0317.v1

Keywords: rhizobacteria; comparative genomics; PGPR traits; antifungal activity



Preprints.org is a free multidisciplinary platform providing preprint service that is dedicated to making early versions of research outputs permanently available and citable. Preprints posted at Preprints.org appear in Web of Science, Crossref, Google Scholar, Scilit, Europe PMC, OpenAlex.

Copyright: This open access article is published under a [Creative Commons CC BY 4.0 license](#), which permit the free download, distribution, and reuse, provided that the author and preprint are cited in any reuse.

Disclaimer/Publisher's Note: The statements, opinions, and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions, or products referred to in the content.

Article

Functional and Comparative Phylogenomics of Pathogen-Inhibiting Native Mexican *Bacillus velezensis* Across Regional and Global Scales Reveal Open Pan-Genome Architecture and Adaptive Divergence

Hilda-Mabel Sosa Esquivel ¹, Atzin Fraire Mayorga ¹, Yadira-Yumiko De la Cruz Rodríguez ¹, Angelica-Judith Granados López ¹, Gloria Viviana Cerrillo-Rojas ², Jesús Adrián López ¹ and Saúl Fraire Velázquez ^{1,*}

¹ Academic Unit of Biological Sciences, Autonomous University of Zacatecas, Zacatecas, PC.- 98068, Mexico

² Academic Unit of Chemical Sciences, Autonomous University of Zacatecas, Zacatecas, Zac. PC.- 98160, Mexico

* Correspondence: sfraire@uaz.edu.mx

Abstract

Bacillus velezensis is a rhizosphere-associated bacterium widely recognized for its roles in biological control and plant growth promotion; however, its functional diversity and evolutionary structure across scales remain incompletely understood. This study evaluated strains 2A-2B, 3A-6A, 2A-10A and 3A-25B from the Center-North of Mexico through integrated phenotypic assays and comparative genomics. Antagonistic activity was assessed via dual confrontation assays against major chili pepper phytopathogens, and plant–bacteria compatibility was examined in vitro. Genome-based analyses included pan-genome reconstruction, phylogenetic inference, and functional annotation, incorporating the screening of plant-associated genetic traits using the PLaBAs platform. All strains consistently inhibited phytopathogens (40–80%), with no significant differences among them, and displayed non-pathogenic interactions with the host plant. Genomic analyses revealed highly conserved core features alongside variation in accessory and strain-specific genes, with strain 3A-25B showing the highest divergence. Pan-genome analyses at regional and global scales indicated an open structure shaped by geography. Phylogenetically, three strains clustered together, whereas strain 3A-25B grouped with distant lineages. All genomes encoded extensive plant growth–promoting traits, while a substantial fraction of genes remained unannotated. These findings highlight functional consistency despite genomic divergence and support the ecological versatility and biotechnological potential of native *B. velezensis* strains.

Keywords: rhizobacteria; comparative genomics; PGPR traits; antifungal activity

1. Introduction

Bacillus velezensis is one of the most prominent plant growth–promoting rhizobacteria (PGPR) recognized by its biocontrol activity against a broad spectrum of phytopathogens, including fungi, bacteria, and nematodes [1,2]. As a Gram-positive species it is characterized by its remarkable ability to form endospores and by harboring an extensive repertoire of biosynthetic gene clusters responsible for the production of antimicrobial secondary metabolites, which enable the inhibition of competing microorganisms [2–4]. In addition, *B. velezensis* produces molecules with phytohormonal activity, such as indole-3-acetic acid (IAA) [5,6], and cytokinins [7,8], among others, that facilitate close interaction with the host plant and exert significant effects on key physiological processes including cell division regulation, thereby promoting plant growth [7]. Furthermore, this species plays a

relevant role in modulating plant immunity, notably through the activation of induced systemic resistance (ISR) which strengthens the plant's defense responses against phytopathogens [9,10].

Beyond its biocontrol capabilities, *B. velezensis* also functions as a bio-organic fertilizer due to its ability to solubilize phosphorus [11,12] and potassium [13,14], enhance sulfur assimilation [15,16] and mediate iron chelation and uptake [17,18]. Numerous studies highlight *Bacillus* spp., and particularly *B. velezensis*, as effective rhizobacteria against fungal phytopathogens affecting plant roots, as well as bacterial pathogens impacting both root systems and foliar tissues [19,20]. Some strains have already been developed into commercial agrobiotechnological formulations, including RhizoPlus®, RhizoVital® (Andermatt Group AG, Switzerland), Taegro® (Novozymes, México, SA. de CV) Botrybel® (Probelte, España), Fungifree AB® (Agrobiotics, México), VELEZIGEN® (Cergen SRL Bio Tecnología, Argentina), among others.

Whole sequencing of microorganisms, coupled with the rapidly expanding number of genome records available in public databases, has opened extensive opportunities for comprehensive and in-depth studies. These approaches enable the characterization of the full gene repertoire, or pangenome, that defines a given phylogenetic clade, as well as the identification of the core genome shared across all members of a microbial population. This core set comprises essential genes required for fundamental metabolic functions that underpin key phenotypic traits and are critical for microbial survival within their predominant ecological niche [21,22].

Isolates of *B. velezensis* have been reported across all continents, with the highest number of identified strains documented in Asia, particularly in China, where the largest number of genomes has been sequenced. Species within the genus *Bacillus* typically possess genomes ranging from approximately 3.73 to 5.79 Mb, with *B. velezensis* exhibiting an average genome size of around 3.99 Mb [23,24].

In north-central Mexico, particularly in the state of Zacatecas, rhizobacteria have been isolated for the inhibition and biocontrol of root pathogenic fungi in chili pepper plants, as well as root and foliar pathogens in common bean. Among these isolates, strains exhibiting the highest inhibitory potential against phytopathogens and plant growth-promoting traits have, for reasons yet unclear, consistently been identified as belonging to the species *B. velezensis* [25,26]. Four genomes from these strains have been sequenced, corresponding to strains 2A-2B, 3A-6A, 2A-10A and 3A-25B, with NCBI accession numbers NZ_MLCV00000000.1, MLCX00000000.1, MKWT00000000.1 and MLCW00000000.1 respectively.

The aim of the present study was to analyze the phenotypic variation among these four *B. velezensis* strains in terms of their ability to inhibit phytopathogens and their interaction with chili pepper plants following root inoculation, in relation to their genomic and genetic features directly or indirectly associated with plant interaction and plant growth promotion. Additionally, the study evaluated the flexibility of the pan-genome, core genome, and singleton gene content at regional, continental, and global scales.

2. Materials and Methods

2.1. Dual Culture Assay Between Bacterial Strains and Fungal or Oomycete Phytopathogens

Virulent strains of the pathogens *Phytophthora capsici*, *Rhizoctonia solani*, *Fusarium solani*, and *Fusarium oxysporum* on chili pepper plants of the cultivar Mirasol were cultured on PDA or KB agar plates; fungi were grown for 4 days and the oomycete for 8 days. *Bacillus velezensis* strains 2A-2B, 3A-6A, 2A-10A and 3A-25B were grown in liquid KB medium for 12 to 18 h and subsequently subcultured for an additional 40 to 8 h until reaching a OD_{600} of 2.0, equivalent to 2×10^7 UFC/mL. After confirming that both that fungi and the oomycete grew adequately on KB medium, dual confrontation assays between bacterial strains and pathogens were performed.

For the dual confrontation assay, 8 mL of bacterial suspension (2×10^7 CFU/mL) obtained from 4-6 h subcultures, were inoculated at four equidistant points on KB agar plates, positioned 5 mm from the edge of the Petri dish. Subsequently, each plate was centrally inoculated with the corresponding

fungal or oomycete pathogen by placing a 3.5 mm diameter mycelial plug, obtained with a cork borer from actively growing cultures previously established on PDA medium. As a control treatment, each pathogen was grown individually from the center of KB plates without bacterial confrontation. Plates were incubated at 28°C in darkness for 5-12 days. Four biological replicates were included per treatment.

Pathogen growth and inhibition were measured every 3 days up to 12 days. The percentage of inhibition was calculated using the following formula:

$$\text{Pathogen inhibition} = \frac{rc-rt}{rc} \times 100$$

where rc is the radial growth in the control treatment and rt is the radial growth in the respective treatment. Differences in inhibition percentages for each pathogen in confrontation with each bacterial strain were analyzed. Statistical analyses were performed using PSPP software version 2.0.1. [27].

For these bacterial strains, three genomes had been previously sequenced, whereas the fourth was sequenced more recently using illumina (MiSeq) technology. Comparative genomics and subsequent bioinformatic analyses indicated that all four strains are consistently assigned to the species *B. velezensis* [25,26]. Three of the genomes had been previously deposited in the NCBI database, while the fourth genome was submitted more recently.

2.2. In Vitro Confrontation of *B. velezensis* Bacteria with Chili Pepper Plants cv. Mirasol

Fifteen-day-old chili pepper plants (cv. Mirasol) grown in vitro on MS medium in Petri dishes were root-inoculated with each *B. velezensis* strain independently, using a liquid suspension at 2×10^7 CFU/mL. Inoculation was performed by immersing the roots in the bacterial suspension for 10 min with agitation every 3 min. After inoculation, the plants were transferred to fresh MS medium and maintained at 37°C under a 12-h light/12-h dark photoperiod for 8 days of observation.

The experiment consisted of six treatments with six replicates each. T1 = Control (plants without bacterial inoculation), T2 = Inoculation with strain S50-2A-1 (positive controls, hostile strain), T3 - T6 = Inoculation with strains 2A-2B, 2A-10A, 3A-25B and 3A-6A, respectively. Plant damage was assessed using a five-level scale, based on Sunwoo et al. [28] with minor modifications: 0 = no visible disease symptom (healthy plant); 1 = initial brownish lesions on root and stems, with slight leaf wilting, 2 = necrosis affecting the main root system and stem base, with 30-50% of the plant affected, 3 = necrosis affecting the main root system and stem base, with 50-70% of the plant affected, and 4 = necrosis affecting the main root system and stem base, with $\geq 70\%$ of the plant affected. Bootstrap analysis was performed using the boot package in R, with a 95% confidence interval.

2.3. Review and Analysis of the Genomic Features of Native *B. velezensis* Strains 2A-2B, 3A-6A, 2A-10A and 3A-25B.

Genome assembly quality was assessed using QUAST algorithm, while genes and coding sequences were annotated with the Prokka algorithm on the Galaxy server. This analysis allowed us to confirm, quantify, and summarize the basic genomic features, including genome size, GC content, number of coding sequences (CDs), RNA genes, and putative pseudogenes. The genomes of the four strains (2A-2B, 3A-6A, 2A-10A and 3A-25B) under study have been registered in the NCBI database and assigned accession numbers NZ_MLCV00000000.1, MLCX00000000.1, MKWT00000000.1 and MLCW00000000.1 respectively. (The accession numbers for the genomes of strains 3A-6A, 2A-10A in the NCBI database will be released next week).

2.4. Comparative Genomic and Functional Analyses of Four Native *B. velezensis* Strains from North-Central Mexico, Including Reference Genome

For the analyses of the pan-genome, core genome, dispensable genome and singleton gene of the four native *B. velezensis* strains (2A-10A, 3A-6A, 2A-2B, and 3A-25B) the genome of strain JS25R (accession NZ_CP009679) was included as a reference. All the genomic datasets were analyzed using

EDGAR platform [29]. Gene prediction and coding sequence (CDS) annotation were standardized within the platform to ensure comparability. The pan-genome, core genome, and singleton functions were computed using the corresponding EDGAR modules, and their proportions were determined based on the total number of gene families identified across the five genomes. Pan-genome development curves and core genome reduction trends were generated through iterative genome addition analysis. Gene cardinality and the distribution of shared and unique CDS among strains were assessed using the Geneset tools, including Venn diagram analyses.

Genomic similarity among strains was evaluated through Average Nucleotide Identity (ANI) analysis using FastANI algorithm implemented in EDGAR environment. Pairwise ANI values were calculated for all genome combinations to determine intra- and interspecific relationships. Additionally, the Percentage of Conserved Proteins (POCP) was computed to assess evolutionary relatedness and functional conservation at the proteome level.

Functional annotation and classification of genes were performed using both KEGG and COG frameworks. KEGG-based analyses categorized genes into metabolic pathways, genetic information processing, environmental information processing, and cellular processes, using KEGG Sunburst tool [30]. In parallel, Clusters of Orthologous Groups (COG) classification was applied to assign genes into functional categories such as motility, secondary metabolite biosynthesis, and defense mechanisms [31]. The distribution and abundance of genes within each functional category were quantified for each genome and comparatively analyzed across strains.

All analyses, including pan-genome modeling, ANI, POCP, functional categorization, and gene distribution assessments, were performed under default settings unless otherwise specified.

2.5. Multi-Scale Comparative Pan-Genome and Phylogenomics Analyses of *B. velezensis* Across Regional, Continental, and Global Populations

To evaluate pan-genome structure, genomic diversity, and phylogenetic relationships of *B. velezensis* across increasing geographic scales, a hierarchical comparative genomics approach was implemented using EDGAR platform. Genome datasets were structured into four levels: (i) regional (north-central Mexico, including four native strains plus one reference genome), (ii) North America (19 genomes from México, the United States, and Canada), (iii) intercontinental comparison (North America vs. South America; 9 genomes each), and (iv) global scale (30 genomes from Americas, Europe, Asia and Africa). Genome selection was based on availability of complete genomes in the NCBI database and geographic representativeness.

All genomes were processed within EDGAR under standardized annotation pipelines to ensure consistency in coding sequences (CDS) prediction and ortholog identification. Pan-genome, core genome, dispensable genome, and singleton fractions were computed using EDGAR modules. Pan-genome development and core genome decay were assessed through sequential genome addition analyses, allowing evaluation of genome openness on Heaps' law [32]. Gene distribution patterns and cardinality were analyzed to determine shared and strain-specific genomic content across datasets.

Comparative analyses between geographic regions were conducted by normalizing genome samples sizes where necessary (e. g., equal number of genomes for North and South America) to reduce sampling bias. Pan-genome structure and dynamics were evaluated independently for each region and comparatively interpreted across scales.

Phylogenetic relationships were inferred using EDGAR phylogenomics tools based on conserved gene sets. Phylogenetic trees were constructed to assess clustering patterns, lineage divergence using ANI and POCP, as implemented in EDGAR. These metrics provided quantitative support for intra-species relationships and protein-level conservation across strains.

Finally, comparative analyses of pan-genome structure across scales (regional, continental, and global) were performed by integrating absolute gene counts and relative proportions of core, dispensable, and singleton fractions. Observed patterns were interpreted in the context of genome plasticity, geographic structure, and evolutionary processes such as gain and horizontal gene transfer. Default parameters were used in all EDGAR analyses unless otherwise specified.

2.6. Analysis of Bacterial Gene Groups Associated with Plant-Microbe Interaction and Plant Growth Promotion

To analyse genes involved in plant-bacteria interactions and plant growth promotion, the PLaBAs platform was used. Within PLaBAs the PGPT-Pred tool was employed to predict plant growth-promotion traits, enabling the identification of gene groups organized into eight functional categories: colonizing plant system, competitive exclusion, stress control/biocontrol, biofertilization, phytohormone and plant signal production, bioremediation, plant immune response stimulation, and putative functions [33,34].

For the analysis of genes associated with plant-bacteria interactions, the PIFAR-Pred tool was utilized. This tool classifies genes into 13 functional categories, including exopolysaccharide synthesis, hormone biosynthesis, antibiotic biosynthesis, detoxification, motility, volatile compound production, adhesion, plant cell wall-degrading enzymes (PWDE), proteases, among others [34].

Both PGPT_Pred and PIFAR-Pred operate through the identification of marker genes and are supported by a database comprising approximately 6,900 plant growth-promoting traits (PGPTs) associated with 6,965,955 protein sequences [33,34].

3. Results

3.1. Phenotypic Variation Among Bacterial Strains in Phytopathogen Inhibition and Root Interaction in Chili pepper cv. Mirasol

In a screening of 150 rhizobacterial isolates obtained from wild plants in forest soils and cultivated plants in the state of Zacatecas, Mexico, 21 strains exhibited $\geq 35\%$ inhibition of each of the pathogens *Rhizoctonia solani*, *Phytophthora capsici*, *Fusarium solani*, and *Fusarium oxysporum* in *in vitro* dual confrontation assays. Among the most effective rhizobacteria, isolates 2A-2B, 3A-6A, 2A-10A and 3A-25B consistently stood out, showing inhibition levels ranging from 38% to 83% against all four phytopathogens. (Figures 1 and 2). In the comparative phytopathogen inhibition analysis, strain 3A-25B inhibited *R. solani* in a 37.8%, while a 40.8%, 44.2% and 45.5% inhibition was achieved with 3A-6A, 2A-10A and 2A-2B strains respectively, however, it exhibited a major inhibiting effect over *P. capsici* than 3A-6A strain, with an 81.2 and 69.7% respectively, and of similar manner over *F. solani*, with a 67.9 and 61.7% respectively. On the other hand, strain 2A-10A briefly surpasses the other 3 strains in pathogens *P. capsici* and *F. solani* inhibition. No statistically significant differences were observed in any case (Figures 1 and 2).

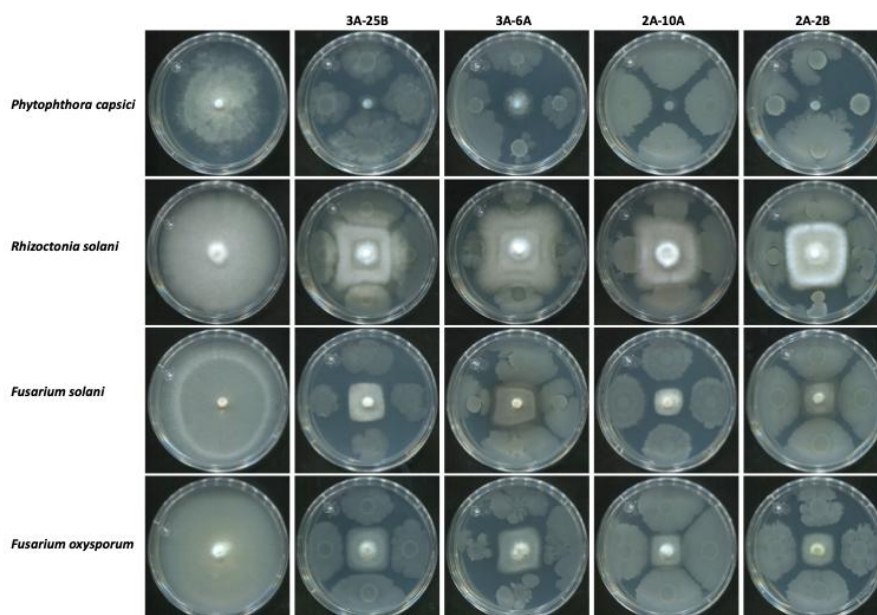


Figure 1. Pathogen inhibition performance of bacterial strains 3A-25B, 3A-6A, 2A-10A and 2A-2B against four pathogens causing root rot in chili pepper. The pathogens *P. capsici*, *Rhizoctonia solani*, *Fusarium solani* and *Fusarium oxysporum*. Four replicates each bacteria-pathogen dual confrontation.

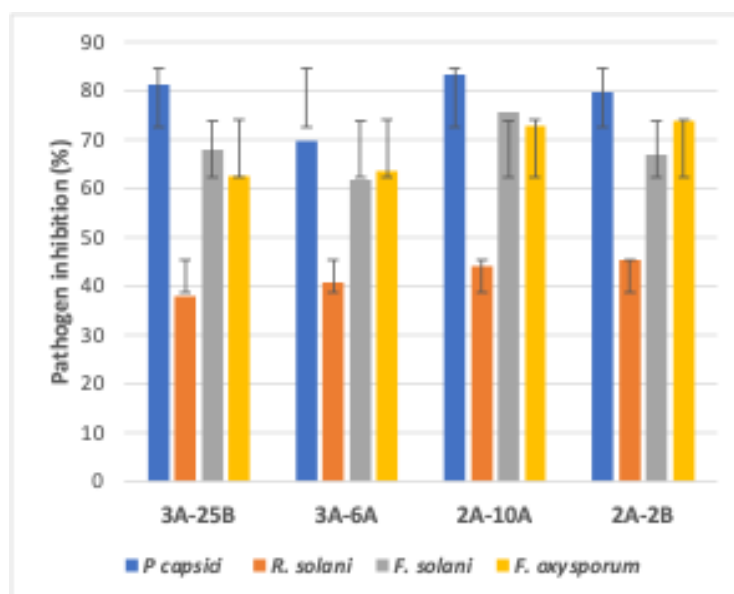


Figure 2. Percentage inhibition of each of the four pathogens by each of the four strains of *B. velezensis*. Four replicates each bacteria-pathogen dual confrontation.

In root inoculation assays evaluating bacterial pathogenicity in chili pepper plants under in vitro conditions, observations at 8 days post-inoculation showed that bootstrap analysis (10,000 iterations) identified the treatment inoculated with strain S50-2A-1 (T2), corresponding to the positive control (pathogenic bacterium causing necrosis in the primary and lateral roots), as having a disease index of 87.5%. This value was significantly higher than those of the remaining treatments, which ranged from 6.25% to 18.75%. Confidence intervals showed substantial overlap among these latter treatments, indicating no significant differences between them. Bootstrap analysis was performed using the boot package in R, with 95% confidence intervals (Figure 3).

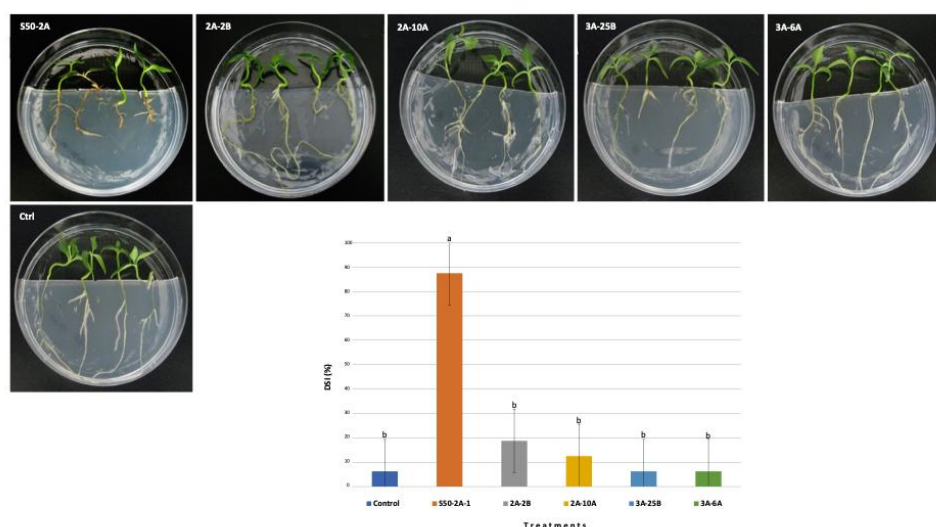


Figure 3. Behavior of native bacterial strains inoculated into the roots of 15-days-old chili plants, 8 days post-inoculation. The strain S50-2A-1 as positive control treatment with high root necrosis damage. DSI (%) with IC 95% (Bootstrap).

3.2. Genomic Features of Four *Bacillus velezensis* Strains Native to North-Central Mexico

Genome sequencing achieved approximately 99% coverage of the total genome length. Genome sizes ranged from 3.93 to 4.01 Mb, values very close to the reported average for this species (~3.9 Mb). GC content, gene number, and coding sequences were highly similar across the four strains. However, variation was observed in the number of rRNAs, tRNAs, miscellaneous RNAs (miscRNAs), and tmRNAs. Strain 3A-25B exhibited the greater divergence, particularly in miscRNAs and tmRNAs, where reduced counts were detected, whereas the other three strains showed high similarity in these features (Table 1).

Table 1. Statistic data of genomes of the four *B. velezensis* native strains.

strain	genome size	GC content (%)	genes	CDS	rRNAs	tRNAs	misc RNAs	tmRNAs
3A-25B	4.01	46.34	3,948	3,786	10	68	5	0
3A-6A	3.93	46.36	3,896	3,778	10	70	82	1
2A-10A	3.94	46.37	3,921	3,804	7	28	81	1
2A-2B	3.95	46.36	3,987	3,822	9	72	83	1

3.3. Pan-Genome of Four *B. velezensis* Strains from North-Central Mexico with a Reference Genome

The pan-genome of the four *B. velezensis* strains native to north-central México, including strain JS25R (accession NZ_CP009679) as the reference genome, comprises 4,258 genes. Of these, 77.9% correspond to the core genome (3,345 genes), 14.4% to the dispensable genome (582 genes), and 7.7% to singleton genes (331 genes) (Figure 4-A). In the pan-genome development, the sequential addition of genomes showed that the inclusion of strain 3A-25B resulted in a marked increase in pan-genome size, while the core genome exhibited a gradual decline (Figure 3-B). The core genome encompasses highly conserved genes present in most of the analysed genomes, presumably associated with essential cellular functions, including central metabolic processes, maintenance and processing of genetic information, and other components of the fundamental bacterial cellular machinery.

In terms of gene cardinality, the five genomes displayed a relatively balanced number of coding sequences (CDS), forming a conserved core genome composed of 3,337 CDS shared among all strains. Each genome contributed a variable number of unique genes, with strain 3A-25B being the most prominent, harboring 149 singleton genes, exceeding the combined total of the three strains (138 singleton genes) (Figure 4-C).

Genomic similarity among these bacterial strains was assessed using Average Nucleotide Identity (ANI) calculated with the FastANI algorithm, which is considered accurate for both complete and draft genomes. Consistent with established thresholds, ANI values >95% indicate intraspecific relationships, whereas values <83% suggest interspecific divergence [35]. The analysis revealed that strain 3A-25B exhibits the highest similarity to strain JS25R, with an ANI of 99.15%, while ANI values between 3A-25B and the other three native strains ranged from 97.8% to 97.9%. In contrast, ANI values among strains 2A-2B, 2A-10A and 3A-6A ranged between 99.6% and 99.8% (Figure 5-B).

To further assess evolutionary relatedness and potential phenotypic differences in terms of protein conservation, ultimately responsible for cellular functional processes, the Percentage of Conserved Proteins (POCP) was calculated [36]. POCP analysis performed on EDGAR server showed that strain 3A-25B is evolutionarily closer to both the reference strain JS25R and native strain 2A-2B, with values of 95.6% and 95.5%, respectively, whereas POCP values among the other three native strains ranged from 96.9% to 98.8% (Figure 5-C).

Functional characterization based on metabolic pathways, signaling, and other biological processes using KEGG, together with lineage and protein function-based analysis using COG, revealed that the genomes of the four *B. velezensis* strains, in comparison with the reference genome, display highly similar functional profiles. According to KEGG classification, the most abundant gene

categories correspond to genetic information processing (534–551 genes) and signaling and cellular processes (559–574 genes), whereas the least represented categories include terpenoid and polyketide metabolism, as well as organismal systems. Notably, a substantial number of genes remain unannotated, ranging from 1,353 to 1,472, predominantly in strain 2A-2B and, to a lesser extent, in strain 3A-6A (Figure 6). Comparative analysis of gene counts across functional categories among the four native strains, as well as among all five strains including the reference, indicates an overall high degree of similarity (Figure 6).

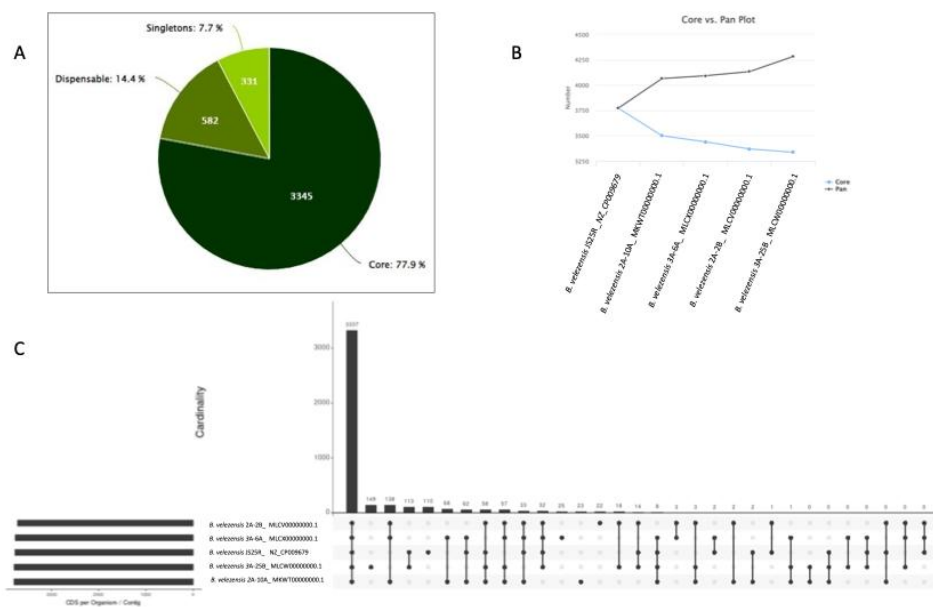


Figure 4. Pan-genome, genome core and cardinality by number of CDS and singletons of the four genomes of *B. velezensis* strains, plus JS25R as reference genome. In A, percentage of core genome, dispensable genome and singletons. In B, gradual opening of the core genome and pangenome with the addition of each genome, with 3A-25B at the end. In C, cardinality comparing the size of the genomes expressed in CDS number, the nucleus genes (3,337) and contribution of each genome in the number of singletons.

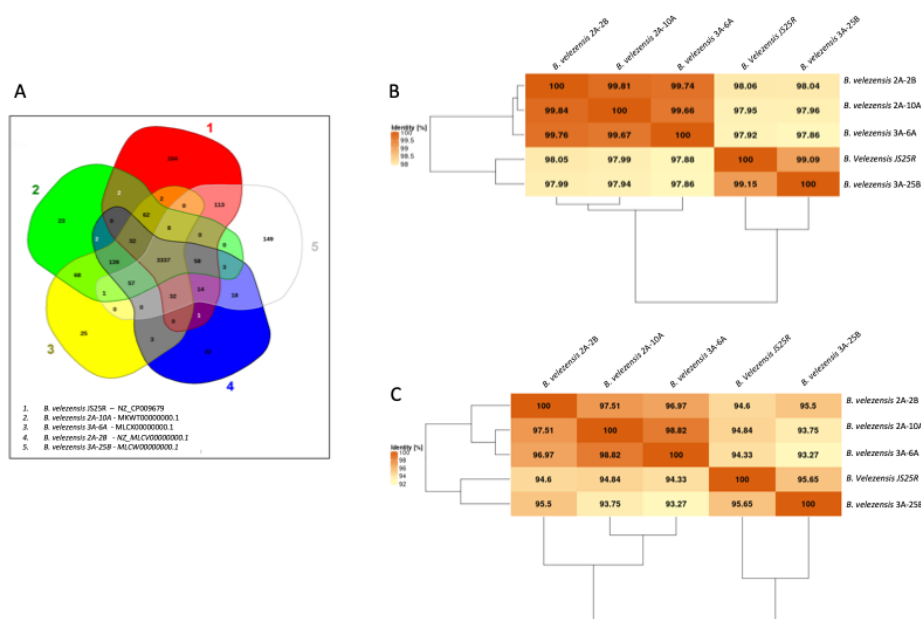


Figure 5. Venn diagram of the conserved genes and the whole genome similarity metrics between the *B. velezensis* genomes under study, plus a reference genome. In A, Venn diagram of the four native *B. velezensis* genomes plus

the *B. velezensis* JS25R strain as reference. In B, average nucleotide identity (FastANI) between these same five genomes. In C, the percentage of conserved proteins (POCP), with the same five genomes. All these calculus performed with the respective bioinformatic packages in the EDGAR server.

Based on Clusters of Orthologous Groups (COG) classification, these bacteria possess 46-48 genes associated with motility, an important trait for chemotaxis during rhizosphere colonization. They also harbor 87-91 genes involved in secondary metabolite biosynthesis and 105-113 genes related to defense mechanisms, both relevant for antibiosis and inhibition of competing microorganisms. Additionally, the native *B. velezensis* strains from Mexico contain between 28 and 29 transposase genes, in contrast to the reference strain JS25R, which contains only 15. As observed with KEGG, COG analysis also revealed a high number of unannotated genes, ranging from 609 to 684, again predominantly in strain 2A-2B and secondarily in strain 3A-6A (Figure 6, Table 2).

Table 2. Singleton genes identified in four genomes of bacterial strains native to north-central Mexico compared with genomes of *B. velezensis* at North American, South American and global scales. Values indicate the number of singleton genes annotated as hypothetical proteins, except in the columns corresponding to strains TrigoCor1448, JS25R, and 3A-25B, where the numbers in parentheses represent, respectively, the number of hypothetical proteins (first value) and genes with known function (second value).

	FZB42	TrigoCor1448	JS25R	2A-2B	2A-10A	3A-6A	3A-25B	Total, strains CNM
North America	3			12	8	10	40 (30/10)	70
South America		13 (5/8)		10	7	10	72 (51/21)	99
Global			13 (4/9)	5	7	10	29 (24/5)	51

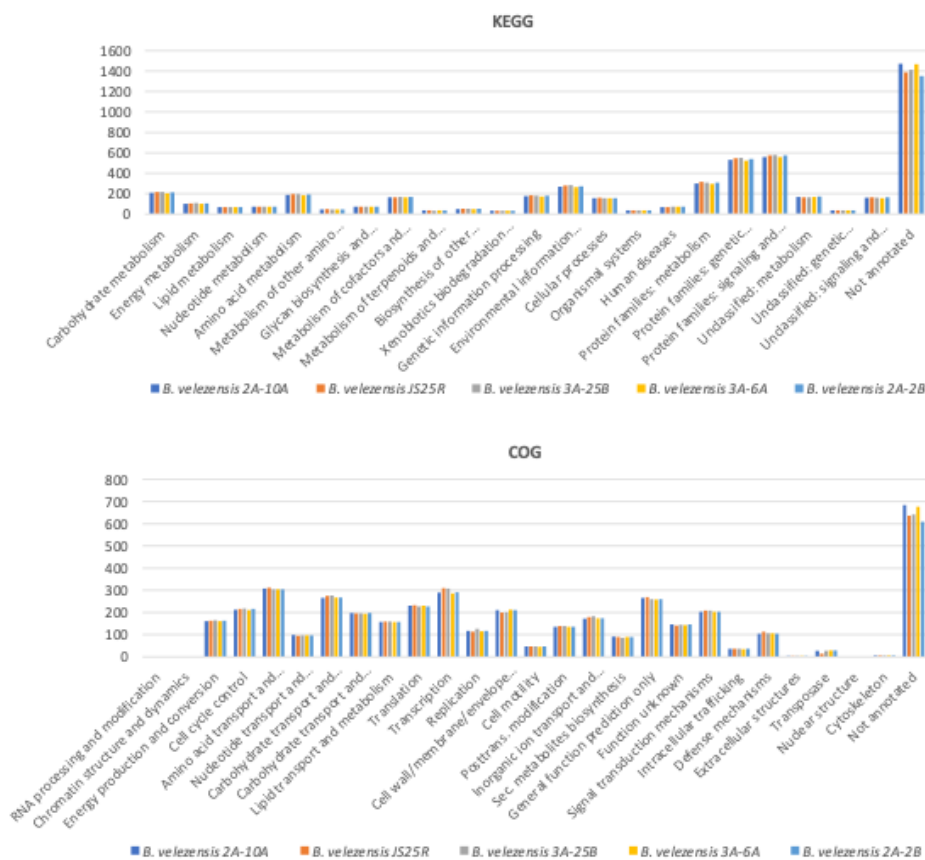


Figure 6. Number of gene families or functional groups based on KEGG and COG criteria in the genomes of the four native *B. velezensis* strains (2A-10A, 3A-6A, 2A-2B and 3A-25B) and the reference strain JS25R. The upper

panel corresponds to KEGG classification, and the lower panel to COG classification. Analysis was performed on the EDGAR server using the Funccats Strains tool.

3.4. Pan-Genome and Phylogenetic Placement of Four *B. velezensis* Strains Native to North-Central Mexico at the North American Scale

Pan-genome analysis of the native strains from north-central Mexico, expanded to include 19 *B. velezensis* strains from North America, comprising 5 strains from Canada, 6 from United States, and 9 from Mexico (the four under study plus 5 additional strains) with strain FZB24 as the reference genome, revealed a total pan-genome of 5,759 genes. Of these, 2,584 correspond to the core genome (44.9%), 2,183 to the dispensable genome (37.9%), and 992 to singleton genes (17.2%) (Figure 7-A). The high proportion of accessory and singleton genes (55.1%) indicates substantial genomic plasticity in this *B. velezensis* species, at least within the northern region of the American continent.

Pan-genome development analysis showed that the inclusion of the ninth genome, originating from Montreal, Canada, resulted in a marked reduction of the core genome from 3,197 to 2,699 genes. Thereafter, the core genome remained relatively stable with the addition of the remaining genomes, ultimately reaching 2,583 genes. This suggests that approximately 2,580-2,700 core genes are sufficient for these bacteria to maintain standard growth under natural conditions. In contrast, the pan-genome exhibited a gradual upward trend with the sequential addition of genomes. According to Heaps' law, the continuous increase in pan-genome size indicates an exponent $\alpha < 1$, characteristic of an open pan-genome [37] (Figure 6-B).

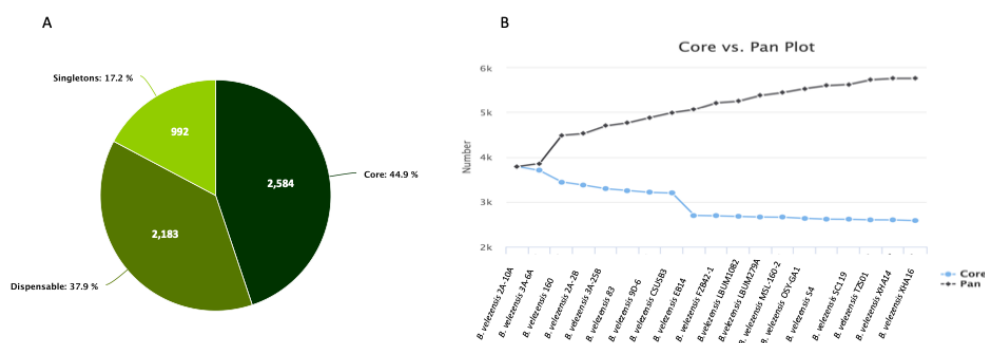


Figure 7. Pan-genome of 19 *B. velezensis* strains representative of North America, including the 4 strains from north-central Mexico. (A) Proportion of core, dispensable, and singleton genes. (B) Core and pan-genome accumulation curves, showing continuous pan-genome growth consistent with an open pan-genome.

Phylogenetic analysis based on the genomes of the same 19 bacterial strains revealed that the native Mexican strains do not form a single clade. While 2A-2B, 3A-6A, and 2A-10A cluster closely together, indicating high genomic similarity, strain 3A-25B is positioned on a separate phylogenetic branch (Figure 8).

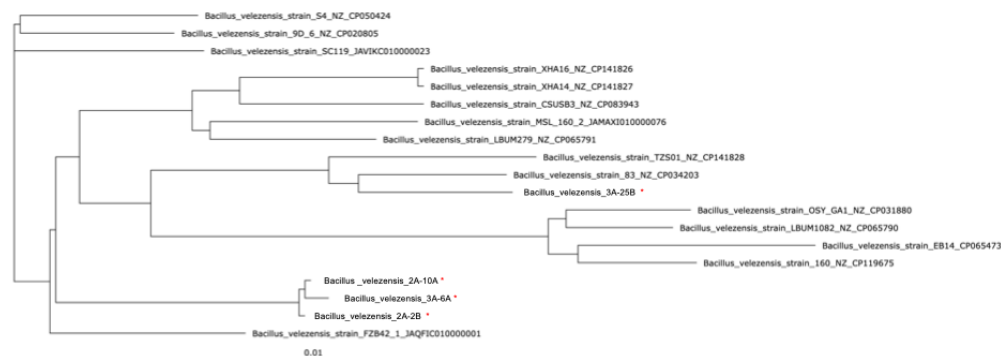


Figure 8. Phylogenomic tree of *Bacillus velezensis* strains from North America. Strains marked with a red asterisk represent the Center-North Mexican isolates analyzed in this study, with FZB42 included as the reference genome. Strains 2A-2B, 3A-6A, and 2A-10A cluster closely together, whereas strain 3A-25B is positioned on a separate phylogenetic branch, indicating intraspecific diversity. The scale bar represents 0.01 substitutions per site. The tree was constructed using the default pipeline implemented in the EDGAR server, applying the Neighbor-Joining algorithm followed by FastTree.

3.5. Comparative Distribution, Pan-Genome Dynamics, and Phylogeny of *B. velezensis* Strains Across North and South America

In this comparative analysis, it should be noted that the number of complete genomes available in the NCBI database for *B. velezensis* from South America is limited. Therefore, nine accessions, mostly from Brazil, were included, and, for balance, nine genomes from North America were included, comprising three from Mexico (including strain 2A-2B), three from the United States and three from Canada.

The North American group exhibits a predominant core genome representing 58.6% of the pan-genome (2,762 CDS), along with smaller fractions of dispensable genes (27.2%) and singletons (14.2%). The pan-genome accumulation curve shows a continuous increase without reaching a plateau, while the core genome gradually declines and stabilizes at approximately 2,700 CDS (Figure 9-A). This pattern indicates an open pan-genome alongside a high degree of genomic conservation, consistent with a relatively homogeneous population.

In contrast, South American strains display a higher proportion of variable (dispensable) genes. The core genome accounts for only 32.3% of the pan-genome (1,615 CDS), whereas dispensable genes represent 50.3% and singletons 17.5%, with the dispensable fraction being the largest component. As additional genomes are incorporated, the pan-genome shows sustained growth, accompanied by a pronounced reduction in the core genome, which stabilizes at approximately 1,600 CDS (Figure 9-B), also consistent with an open pan-genome.

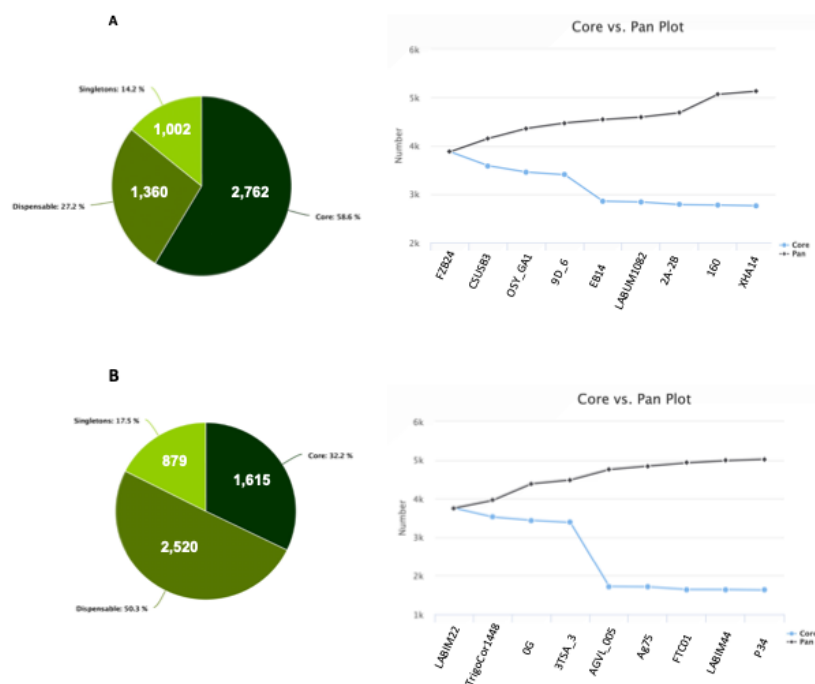


Figure 9. Pan-genome of *B. velezensis* by geographic region. (A) Strains from North America. (B) Strains from South America. The composition of the pan-genome and the accumulation curves of the core and pan-genome are shown, evidencing regional differences in genomic diversity and an open pan-genome in both sets.

In the phylogenetic analysis of strains from the South American region, including the four native strains from north-central Mexico, strains 2A-10A, 3A-6A and 2A-2B from this region clustered together, whereas strain 3A-25B is positioned in a separate subclade alongside Brazilian strains LABIM22, FTCO1 and 3TSA-3 (Figure 10).



Figure 10. Phylogenetic tree of South American strains including the four native strains from north-central Mexico (2A-2B, 3A-6A, 2A-10A and 3A-25B). Strains 2A-2B, 3A-6A and 2A-10A remain clustered together, whereas strain 3A-25B, as observed in North America, is phylogenetically separated and, in this case, clusters closely with strain LABIM22 which is native to Paraná, Brazil.

3.6. Comparative Analysis of *B. velezensis* Pan-Genome Structure Across Increasing Levels of Geographic Sampling and Population Size

Considering 30 genomes of *B. velezensis* strains at a global scale, originating from the American and Afro-Eurasian continents, including the four strains from north-central Mexico, the pan-genome comprises 6,535 genes. Of these, the core genome consists of 2,462 genes (37.7%), the dispensable genome includes 2,614 genes (40.0%), and the singleton fraction accounts for 1,459 genes (22.3%). These results indicate marked intraspecific genetic diversity (Figure 11-A).

In pan-genome development, the initial gene set comprises 3,187 genes. The core genome shows a sharp decline upon incorporation of the third genome, followed by a more gradual decrease from the fourth genome onward, as each of the remaining 27 genomes is sequentially added, ultimately stabilizing at 2,462 genes. In contrast, the pan-genome trajectory exhibits several modest yet noticeable increments, reaching a final total of 6,535 genes. Overall, the progressive addition of genomes results in a continuous increase in the total number of genes without approaching saturation, while the core genome gradually declines and stabilizes at approximately 2,500–2,700 genes (Figure 11-B). This pattern is indicative of an open pan-genome, in which each newly incorporated genome contributes additional genes.

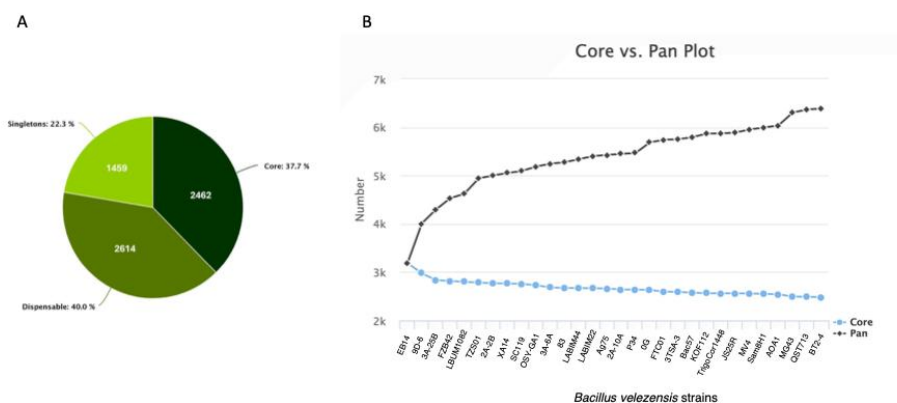


Figure 11. Pan-genome and pan-genome development of *B. velezensis* strains native to north-central Mexico within a representative set of 30 strains at a global scale, encompassing the American and Afro-Eurasian continents. In A, accumulated pan-genome showing core genome and singleton fractions. In B, Pan-genome development with the successive addition of each of the 30 genomes.

A comparative assessment of *B. velezensis* pan-genome structure was conducted across hierarchical levels of geographic sampling and population size, ranging from lower to higher complexity. The lowest level (4CNM+1ref) represents north-central Mexico (5 genomes), intermediate levels correspond to North America (14 genomes) and South America (11 genomes) analyzed separately, and the highest level represents the global dataset (30 genomes). Both absolute gene counts and relative proportions of genomic fractions were evaluated (Figure 12).

In terms of absolute gene numbers (Figure 12, left panel), a progressive increase is observed as both the number and geographic diversity of genomes expand. The 4CNM+1ref dataset exhibits the smallest pan-genome, whereas North and South America show intermediate expansion, and the global dataset displays the largest pan-genome. This pattern is characteristic of an open pan-genome, in which the incorporation of additional genomes continuously contributes new genes. The core genome correspondingly decreases as sampling expands, a predictable mathematical effect whereby increasing genome numbers reduce the conserved fraction. In contrast, the dispensable genome and singleton fractions increase with broader geographic scale, indicating greater genetic diversity and potential regional specialization.

In relative terms (Figure 12, right panel), the pan-genome is normalized to 100% as a reference. The core genome proportion decreases from regional to continental and global scales, whereas the dispensable and singleton fractions increase, particularly at the global level and in the ProPan database (prokaryotic pan-genome database) for *B. velezensis* species. This trend reflects a lower proportion of conserved genes and a higher contribution of accessory and unique genes. According to ProPan post-processed datasets [34], the reported distribution for *B. velezensis* consists of 8.13% core genome, 55.36% dispensable genome, and 36.51% singletons. The results obtained in the present study deviate somewhat from these values; however, the South American dataset shows the closest correspondence to the ProPan distribution (Figure 12).

Overall, the regional, continental, and global pan-genomes of *B. velezensis* analyzed in this study consistently exhibit an open structure, with genetic diversity strongly influenced by geographic structuring, high evolutionary plasticity, and expansion of accessory genome associated with ecological niche adaptation.

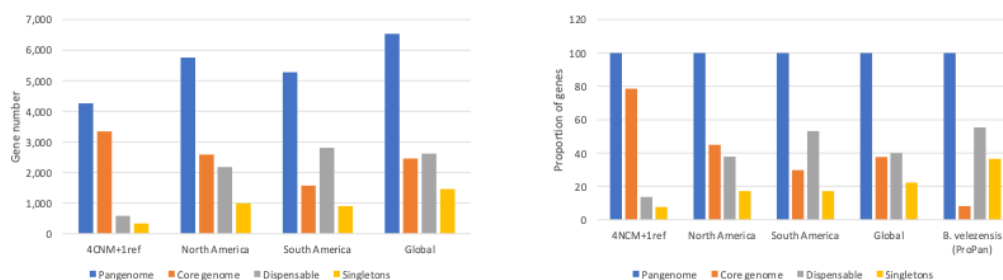


Figure 12. Comparative pan-genome structure of *Bacillus velezensis* at regional and global scales. Left: absolute number of genes in the pan-genome, core, dispensable, and singleton fractions for 4CNM+1ref (four Mexican strains plus one reference), North America (15 genomes), South America (9 genomes), Global (30 genomes), and ProPan (worldwide dataset). Right: proportional distribution of genomic fractions. The expansion of the pan-genome and reduction of the core fraction with broader geographic sampling indicate an open pan-genome structure.

With respect to the global evolutionary history of the four *B. velezensis* strains native to north-central Mexico, phylogenetic relationships inferred from a dataset of 30 representative strains across continents, including 13 from North America (among them the four Mexican strains), 8 from South America, 4 from Asia, 3 from Europe, and 2 from Africa, indicate that strains 2A-10A, 3A-6A and 2A-2B consistently cluster together, closely related to a strain from Brazil. In contrast, strain 3A-25B is

more distantly positioned and shows closer evolutionary affinity to strains from Europe and Asia, specifically from Italy and China (Figure 13).

This pattern may be associated with genomic differences, particularly the presence of a transposase gene unique to strain 3A-25B. A comparative search for transposase genes revealed that all four strains share up to six transposase genes, including types such as ISAa1, ISSu9, ISCac2, and ISBs1, among other variants. However, strain 3A-25B harbors a transposase gene associated with the Tn916 transposon, which is absent in the other three strains. Tn916 belongs to a group of mobile genetic elements capable of integrating into the genome, mediating conjugative transfer to recipient bacterial cells, and undergoing intracellular transposition to different genomic loci. Additionally, these elements can readily acquire accessory genes and promote genome rearrangements [34]. This distinctive feature of strain 3A-25B may underlie genome restructuring over time, contributing to its phylogenetic divergence from the other native *B. velezensis* strains.

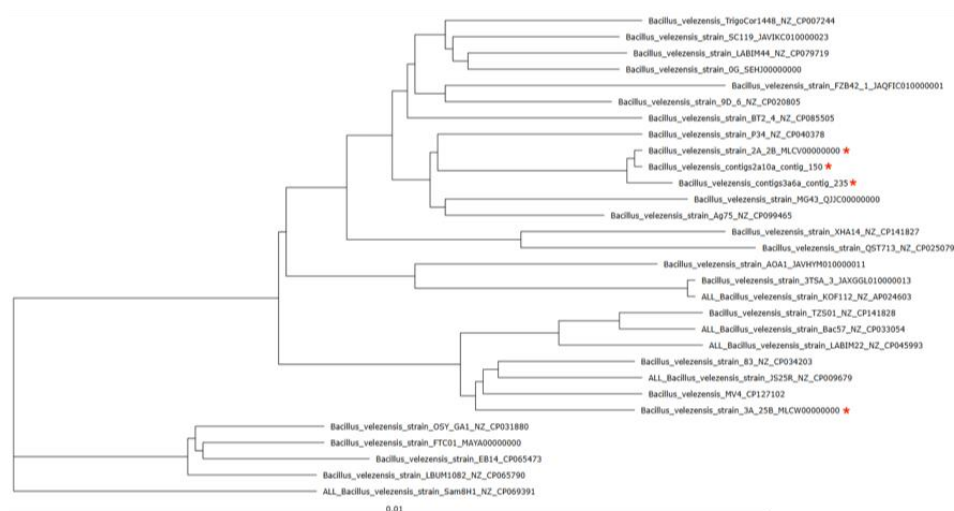


Figure 13. Phylogenetic tree on a global scale with evolutionary positioning of the four native strains of north-central Mexico. Red asterisk highlights the four native strains under study.

3.7. Bacterial Gene Groups Associated with Microbe-Plant Interaction and Plant Growth Promotion

For bacteria intended for biocontrol of phytopathogens, it is essential to assess their genomic composition in terms of gene involved in plant interaction and plant growth promotion. A useful resource for this purpose is the PLaBase platform, which provides tools such as PGPT-Pred for the prediction of plant growth-promoting traits (PGPTs) and PIFAR-Pred for the identification of genes involved in plant-bacteria associations. Both tools rely on marker gene identification and are supported by a database comprising approximately 6,900 PGPTs associated with 6,965,955 protein sequences [33,34].

Key biological functions involved in interaction with plant tissues in the rhizosphere include chemotaxis, adhesion, synthesis of exopolysaccharides and lipopolysaccharides, plant cell wall-degrading enzymes, hormone biosynthesis, and volatile compound production, among others. Analysis using PIFAR-Pred tool across the four genomes under study revealed a high degree of similarity among these *B. velezensis* strains. All strains possess gene repertoires associated with exopolysaccharide synthesis, hormone biosynthesis, and antibiotic production, including compounds such as syringomycin, syringopeptin, toxoflavin, and tagetitoxin, as well as genes related to motility, adhesion, and detoxification.

The main differences among the four strains were observed in toxin-related categories. Notably, strain 2A-10A exhibited a slight increase (~1%) in genes associated with biosynthesis of three antibiotic types. Strain 3A-25B also showed moderate enrichment in genes related to metabolism, pigment production, motility and antibiotic biosynthesis (Figure 14).

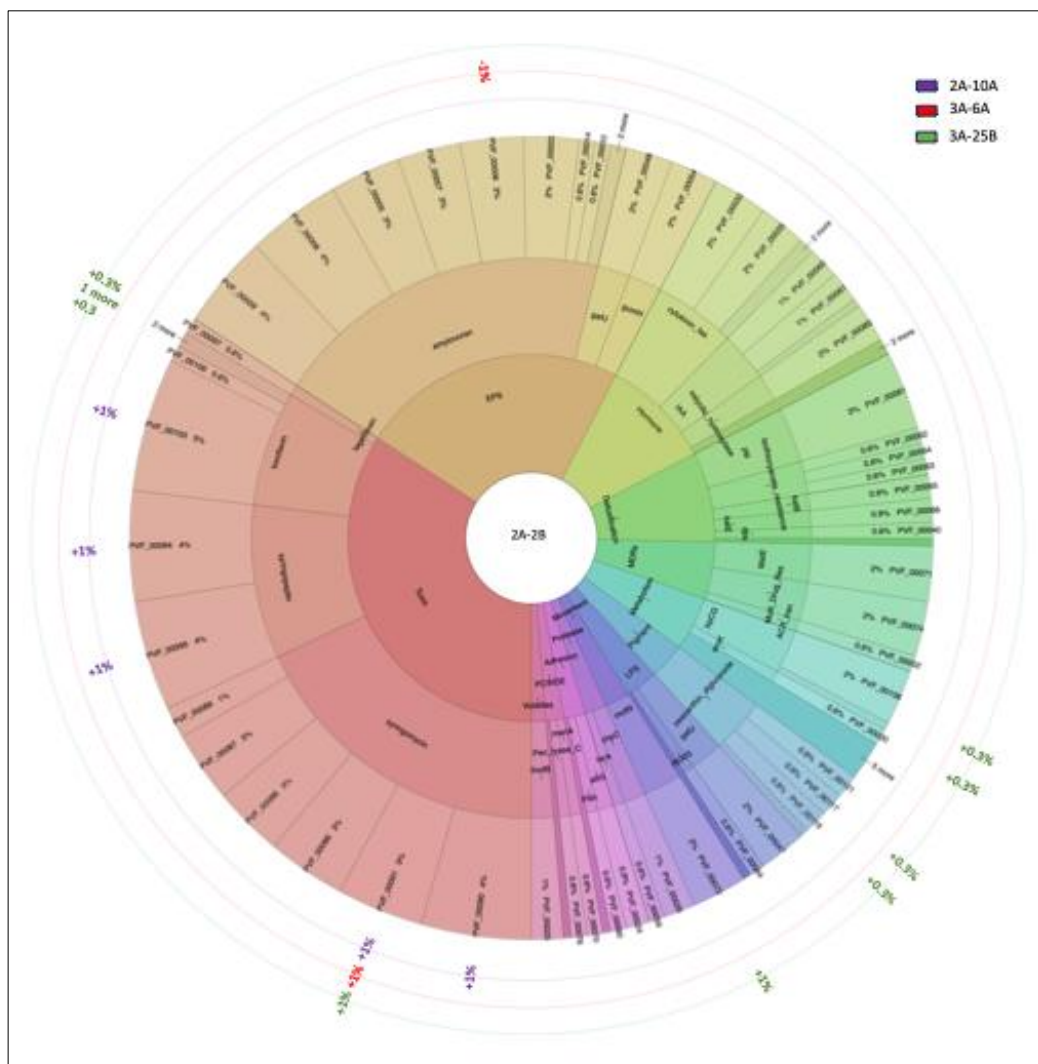


Figure 14. Prediction of genes that bacteria use to interact with the plants. The 2A-2B genome is shown in the center, and the variations in more or less number of genes in the genomes of strains 2A-10A, 3A-6A and 3A-25B respectively are shown in the outside, in purple, red and green circles and numbers. The prediction was performed with the PIFAR-Pred tool using the option blastp-hmmer.

To assess similarities and differences in plant interaction-related bacterial factors between these strains and those from other geographic regions, strain JS25R was included as reference genome; this strain originates from Yancheng province, Jiangsu, China. The results indicate that the native Mexican strains exhibit broadly similar profile of plant interaction factors compared to the reference strain. However, differences were observed in specific functional categories: the reference strain showed reduced representation of genes associated with metabolism and plant cell-wall degrading enzymes (PCWDE), whereas it displayed comparatively higher proportions of genes related to exopolysaccharide production and toxin synthesis.

Additionally, native strain 3A-25B showed greater similarity to the reference strain JS25R than to other three native strains, particularly in the proportion of genes associated with exopolysaccharide production, metabolism, and motility (Figure 15).

With respect to plant growth-promoting traits (PGPTs), analysis of the four *B. velezensis* strains from north-central Mexico revealed a high degree of similarity among them. An exception was strain 3A-25B, which showed a higher number of genes associated with bioremediation and biofertilization, but a lower representation of genes related to competitive exclusion (Figure 16). Notably, all four strains harbor the *rbcL* and *cbbR* genes, which are associated with the regulation of CO₂ fixation through alternative pathways or those related to RUBISCO.

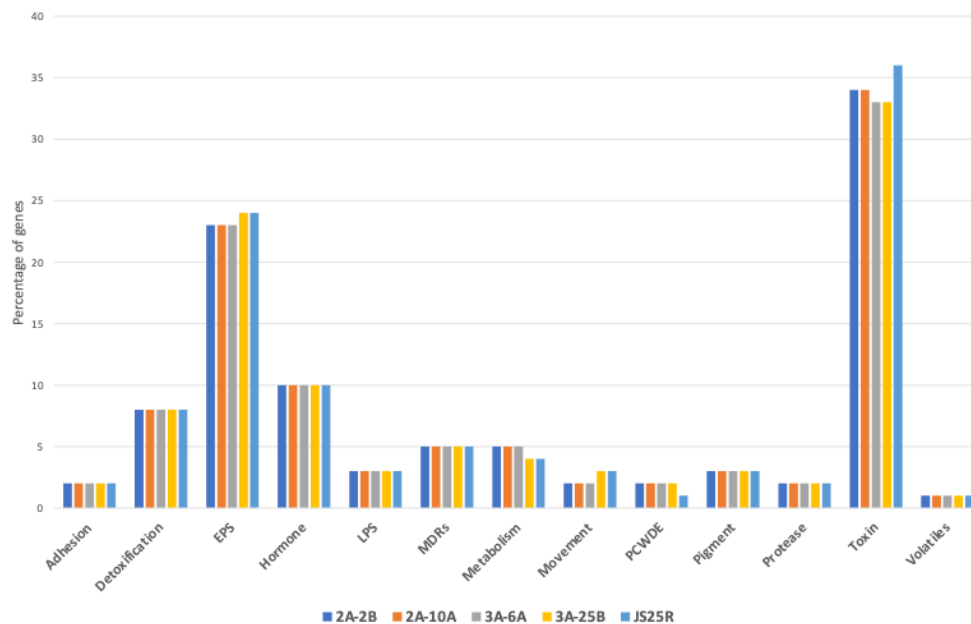


Figure 15. Prediction of bacterial factors involved in the interaction plant-bacteria. The four bacterial genomes under study are included, along with the JS25R strain as a reference genome. The work was carried out with the PIFAR-Pred tool using blastp+hmmer.

Within the biofertilization category, the strains possess approximately 40 genes involved in iron acquisition, 51 in nitrogen acquisition, more than 100 in phosphate solubilization, fewer than 100 in potassium solubilization, and around 14 in sulfur assimilation. Gene-level comparisons indicate the highest similarity between strains 2A-2B and 3A-6A. In the phytohormone signaling category, approximately 40 genes are associated with indole-3-acetic acid (IAA), 7 with cytokinins, and smaller numbers with abscisic acid (ABA) and gibberellins, along with 13 genes related to acetoin/butanediol production. Approximately 825 genes are involved in plant tissue colonization, with slightly higher representation in strains 2A-2B and 2A-10A, although gene-level similarity is again highest between strains 2A-2B and 3A-6A.

In competitive exclusion category, around 630 genes were identified, with a slightly lower number in strain 3A-25B; gene-level similarity was again highest between strains 2A-2B y 3A-6A. For plant immune response induction, 66-67 genes were identified, and for stress control/biocontrol, between 585 and 599 genes, with the highest similarity consistently observed between strains 2A-2B and 3A-6A, including three genes with insecticidal function.

Across the seven functional PGPT categories, the total number of genes identified was 4,095, 4,070, 4,043, and 4,111 for strains 2A-2B, 2A-10A, 3A-6A and 3A-25B, respectively. These values are slightly higher than the total gene counts obtained from genome annotation, likely because annotation relies on predictive approaches that do not always capture all functional elements and are primarily based on coding sequences (CDS).

To assess the degree of similarity or divergence among native strains from north-central Mexico in the proportion of plant growth-promotion traits, both among themselves and relative to strains from other geographic regions, strain JS25R from China was included as a reference. The proportional distribution of genes associated with plant growth-promoting functions is highly consistent among the native Mexican strains. Again, strain 3A-25B represents an exception, exhibiting a reduced proportion of genes associated with competitive exclusion (21% vs. 22%) and a slightly higher proportion associated with bioremediation (7% vs. 6%).

In comparison with the reference strain, the proportions of genes related to bioremediation, plant colonization, biocontrol, and stress response are identical, at least for strains 2A-2B, 2A-10A and 3A-6A. However, strain 3A-25B shows a 1% reduction in the competitive exclusion category,

while simultaneously matching the reference strain with a 1% higher proportion in the bioremediation category (Figure 17).

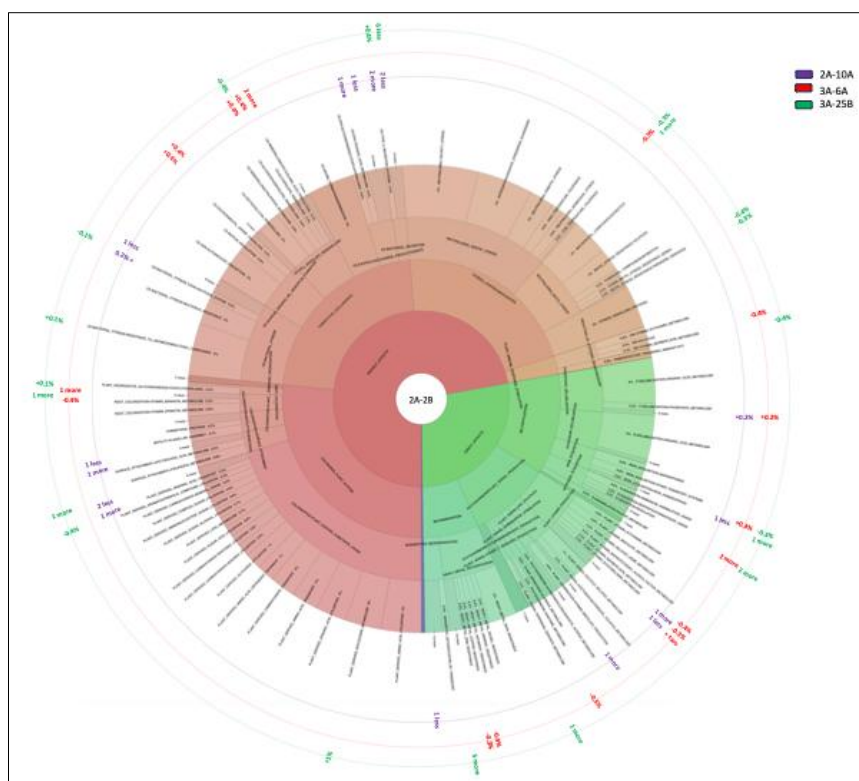


Figure 16. Prediction of genes in bacteria related to plant-growth promoting traits. The 2A-2B genome is shown in the center, and the variations in the genomes of strains 2A-10A, 3A-6A and 3A-25B respectively are shown in the outside, in purple, red and green numbers. The prediction was performed with the PGPT-Pred tool using the option blastp-hmmer.

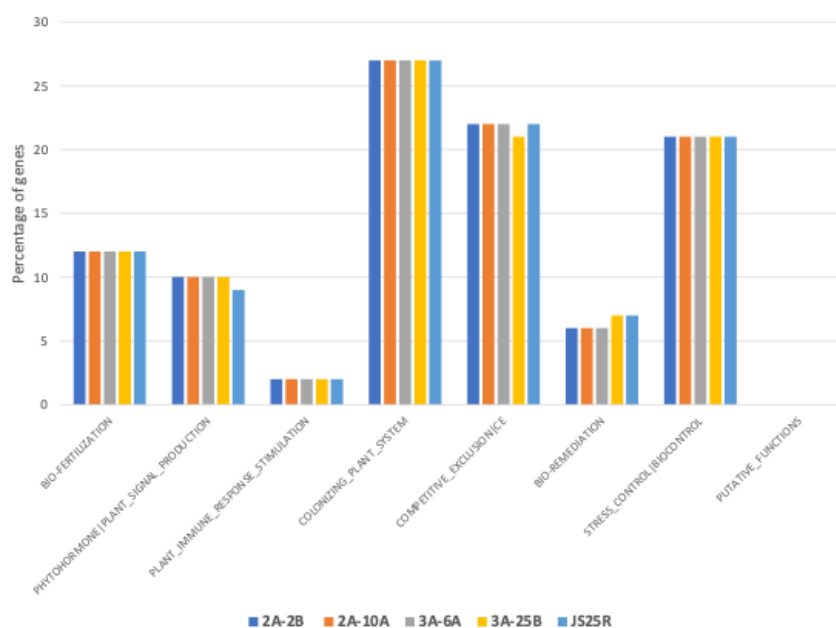


Figure 17. Prediction of bacterial traits involved in the promotion of plant growth. The four bacterial genomes under study are included, along with the JS25R strain as a reference genome. The work was carried out with the PGPT-Pred tool using blastp+hmmer.

3.8. Strain-Specific Genes (Singletons) in *B. velezensis* Across Regional and Global Scales

The presence of singleton genes was analyzed in bacterial genomes across regional (North America and South America) and global scales. In the North American comparison, which included 20 genomes from Mexico, the United States, and Canada, the four strains from north-central Mexico collectively harbored 70 singleton genes distributed as follows: 40 in strain 3A-25B, 12 in strain 2A-2B, 10 in strain 3A-6A, and 8 in strain 2A-10A. In strains 2A-2B, 3A-6A and 2A-10A, all singleton genes were annotated as hypothetical proteins. In contrast, among the 40 singleton genes in strain 3A-25B, 8 correspond to proteins with known functions and 32 to hypothetical proteins; among the characterized genes, three are associated with phage structure or recombination. As a reference genome, strain FZB42, native to Wisconsin, USA, was included; this strain contains only three singleton genes, all encoding hypothetical proteins (Table 2).

For South America, where genome availability in public databases is more limited, a total of 13 genomes were included. In this comparison, the four strains from north-central Mexico collectively accounted for 99 singleton genes, with the majority (72) found in strain 3A-25B, followed by 10 in strain 2A-2B, 10 in strain 3A-6A, and 7 in strain 2A-10A. In strains 2A-2B, 3A-6A and 2A-10A, all singleton genes were hypothetical proteins, whereas in strain 3A-25B, 50 were hypothetical and the remainder corresponded to proteins with known functions, including roles in phage-related processes, manganese transport, and DNA packaging. The reference genome for this dataset was strain TrigoCor1448, native to Brazil, which contains 13 singleton genes, of which 5 encode hypothetical proteins and 8 have known biological functions. Among these are members of the TetR/AcrR family, which act as transcriptional regulators involved in gene expression associated with antibiotic resistance and responses to environmental stresses such as osmotic and oxidative stress [38].

At the global scale, based on the analysis of 30 genomes distributed across North America, South America, Eurasia, and Africa, the four strains from north-central Mexico collectively harbor 51 singleton genes: 5 in strain 2A-2B, 7 in strain 2A-10A, 10 in strain 3A-6A, and 29 in strain 3A-25B. Most are annotated as hypothetical proteins, except for five genes with known functions identified in strain 3A-25B, including genes involved in phage structural assembly, viral genome packaging into capsids, autolysin activity, and DNA packaging (Table 2). The reference genome included at this scale was strain JS25R originating from Jiangsu Province, China, which contains 13 singleton genes, 4 encoding hypothetical proteins and 9 with known biological functions. Among these is a CbrC family protein, associated with bacterial tolerance to colicin E2, an antibiotic-like toxin [39], among other functions.

4. Discussion

The four *B. velezensis* strains (3A-25B, 3A-6A, 2A-10A and 2A-2B) native to north-central Mexico consistently inhibited the four major chili root pathogens, with inhibition percentages ranging from 38% to 80%. *Rhizoctonia solani* was the least affected pathogen, whereas *Phytophthora capsici* was the most susceptible, with no statistically significant differences among bacterial strains (Figures 1 and 2). These values are consistent with reports for *B. velezensis* strains from the Amazon region, where inhibition levels of 73–100% have been observed against phytopathogenic fungi such as *Botrytis cinerea*, *Colletotrichum acutatum*, and *Penicillium expansum* [40], as well as strains from Morocco showing inhibition of 67–78.4% against *Phytophthora infestans* [41]. This suggests that core antifungal mechanisms in *B. velezensis* may be conserved across geographically distant lineages.

In plant interaction assays conducted in vitro, strains 2A-2B, 2A-10A, 3A-6A, and 3A-25B exhibited low pathogenic impact, with the highest mean disease index reaching only 18.7%, in contrast to the pathogenic strain S50-2A-1, which reached 87% (Figure 3). These findings align with numerous studies demonstrating that *B. velezensis* strains are generally plant-friendly and capable of promoting plant growth [42,43].

At the genomic level, the four strains showed high similarity in genome size, GC content, gene number, and coding sequences (CDS), but differed in non-coding RNA elements, including rRNAs, tRNAs, miscRNAs, and tmRNAs. One strain exhibited nearly half the number of tRNAs compared to the others, while another contained only five miscRNAs and lacked tmRNAs entirely (Table 1). These differences in non-coding RNAs do not appear to directly influence antifungal activity, as all four strains remained statistically equivalent in this trait.

Pan-genome analysis of the four native strains plus reference strain JS25R revealed that strain 3A-25B harbors 149 singleton genes, exceeding both the 110 singletons in strain JS25R and the 22-25 singletons found in strains 3A-6A, 2A-10A and 2A-2B. Approximately 80% of shared genes constitute the core genome, with 14% classified as dispensable and 8% as singletons, indicating a moderately variable pan-genome with a large conserved functional backbone. Comparative genomic analyses based on average nucleotide identity (ANI) and percentage of conserved proteins (POCP) showed that strain 3A-25B is more divergent from the other native strains and clusters more closely with the reference strain JS25R from China. This pattern suggests non-uniform divergence among native strains, potentially reflecting distinct evolutionary trajectories, differential gene acquisition, or similar ecological selection pressures acting on strains 3A-25B and JS25R.

Functional annotation based on KEGG and COG categories revealed highly similar gene distributions among the four Mexican strains and the reference strain JS25R. The most represented categories corresponded to genetic information processing, signaling, and cellular processes, whereas terpene/polyketide metabolism and organismal systems were the least represented. Notably, approximately 17% of genes remained unannotated (~1,400 genes) (Figure 6), indicating a substantial proportion of “functional dark matter” with potential biotechnological relevance [44]. COG analysis further showed enrichment in amino acid transport, transcription, cell wall/membrane biogenesis, signal transduction, and general functional categories, alongside a similarly high proportion of uncharacterized genes (610–680) (Figure 6).

At the North American scale (19 genomes), the pan-genome exhibited continuous increase, while the core genome declined notably after incorporation of the ninth genome (from Montreal, Canada), consistent with an open pan-genome structure. Phylogenetic analysis revealed that strains 2A-10A, 3A-6A and 2A-2B cluster together, whereas strain 3A-25B occupies a more distant subclade alongside other Mexican strains (Figure 7). This configuration underscores genetic differentiation among bacteria sharing a common geographic origin and suggests partially independent evolutionary trajectories, potentially associated with distinct rhizosphere niches. Notably, strain 3A-25B harbors a transposase gene associated with the Tn916 transposon, absent in the other three strains, belonging to a group of mobile genetic elements capable of genome integration, conjugative transfer to recipient cells, and intracellular transposition across genomic loci, as well as acquisition of accessory genes and promotion of genome rearrangements [34]. This distinctive genomic feature may have facilitated structural genome remodeling over time, thereby contributing to the phylogenetic divergence of strain 3A-25B from the other tree *B. velezensis* strains.

In contrast, the South American pan-genome (13 genomes) displayed a markedly different structure, with a reduced core genome (29.9%) and an expanded dispensable genome (53%), while singleton proportions remained similar (~17%). This inverse relationship between core and accessory fractions suggests that geographic separation has contributed to increased evolutionary divergence and genomic plasticity. Phylogenetically, strain 3A-25B clustered closely with a Brazilian strain, whereas strains 2A-2B, 3A-6A, and 2A-10A remained grouped together (Figure 9). These results suggest greater intraspecific genomic diversity in South America, potentially associated with higher environmental heterogeneity.

At the global scale (30 genomes), the pan-genome structure resembled that observed for North America, with 38% core, 40% dispensable, and 22% singleton genes. The continuous expansion of the pan-genome and gradual decline of the core genome further support an open pan-genome model. Notably, South American strains exhibited a distinct genomic distribution, with a higher proportion of dispensable genes, aligning more closely with pan-genome structures reported in the ProPan

database (Figure 12). The relatively high proportions of dispensable genes and singletons suggest considerable genomic plasticity, likely linked to environmental interactions, adaptation to local ecological niches, microbial competition in the rhizosphere, and specific selective pressures.

Globally, phylogenetic relationships revealed that strains 3A-6A, 2A-10A y 2A-2B cluster with strains from diverse geographic origins, whereas strain 3A-25B form a distinct subclade with strains from Italy, China, and northern Mexico. This pattern may reflect the high dispersal capacity of *B. velezensis*, likely facilitated by endospore formation, enabling long-term survival under desiccation, radiation, and extreme temperatures [45], and long-distance transport via air currents [46], water systems [47], migratory organisms, or human-mediated activities [48]. Additionally, some *Bacillus* lineages exhibit relatively conserved pan-genomes with a stable core genome, supporting the occurrence of shared evolutionary origins among geographically distant strains [49]. In contrast, strain Sam8H1 forms a distinct phylogenetic branch; this strain originates from Botswana, Africa, where semi-arid to arid conditions (30–35 °C, high evaporation, nutrient-poor sandy soils) may have driven its marked evolutionary divergence and unique phylogenetic placement. Collectively, these findings suggest that the *B. velezensis* species possesses a conserved genetic backbone sufficient to maintain functional identity, complemented by a substantial proportion of variable genes that enhance its adaptive capacity and ecological success at a global scale. The significant contribution of accessory and unique (singleton) genes underscores the role of gene gain, including horizontal gene transfer, as a key evolutionary driver of diversity in this rhizobacterial species.

The analysis of plant growth-promoting traits (PGPTs) revealed a largely homogeneous functional profile among the four native strains, with genes associated with phytohormone biosynthesis (notably indole-3-acetic acid), motility, adhesion, detoxification, and plant cell wall degradation. Overall similarity with the Asian reference strain was high, although differences were observed in metabolic and plant cell wall-degrading enzyme categories. Notably, strain 3A-25B again showed greater similarity to the reference strain than to the other native strains.

Strain 3A-25B was distinguished by a higher abundance of genes related to biofertilization (iron acquisition, phosphate solubilization, sulfur assimilation) and bioremediation, suggesting enhanced ecological versatility. However, this was accompanied by a reduced number of genes associated with competitive exclusion, indicating a potential functional trade-off favoring host interaction over microbial competition.

A notable feature across all four strains was the presence of genes related to alternative CO₂ fixation pathways linked to Rubisco-like mechanisms, suggesting metabolic plasticity that may enhance survival under carbon-limited conditions [50]. Additionally, the high abundance of genes for nutrient acquisition (iron, nitrogen, phosphorus, potassium) and phytohormone signaling (IAA, cytokinins, ABA, gibberellins), along with volatile compound production (acetoin and butanediol), supports their classification as multifunctional plant growth-promoting rhizobacteria (PGPR).

Overall, the larger number of genes associated with plant colonization (~825) and microbial competitiveness (>600) indicates that plant-microbe interactions are governed by complex and redundant functional networks. Within this framework, strain 3A-25B stands out as functionally specialized, combining enhanced biofertilization and colonization capacity with reduced competitive exclusion, reinforcing its distinct adaptive trajectory among the four analyzed strains.

5. Conclusions

The four *Bacillus velezensis* native to north-central Mexico exhibit a functional and genomic profile that supports their strong potential as biocontrol agents and plant growth promoters, maintaining consistent antifungal activity against chili phytopathogens without causing significant adverse effects on the host plant. The absence of statistically significant differences in antagonistic activity, despite underlying genomic variation, suggest that key antifungal mechanisms are highly conserved within the species, even among lineages with distinct evolutionary histories and broad geographic distribution.

At the genomic level, these strains share a robust and stable core genome, while variation in dispensable and singleton genes, particularly pronounced in strain 3A-25B, reflects ongoing microevolutionary diversification and differential gene acquisition. Although these differences do not translate into major changes in antifungal capacity, they are associated with ecological specialization, especially in traits related to plant colonization, biofertilization, and metabolic adaptation to environmental conditions.

Pan-genome analyses at regional and global scales indicate that *B. velezensis* possesses an open pan-genome, consistent with its wide geographic distribution and adaptability to diverse environments. Differences observed between North and South America, together with phylogenetic proximity among geographically distant strains, support the notion of a highly dispersible species, likely facilitated by endospore formation and multiple natural and anthropogenic vectors. Within this framework, strain 3A-25B emerges as a more genetically divergent lineage within the Mexican group, showing evolutionary affinities with strains from other regions, suggesting shared selective pressures or convergent evolutionary trajectories.

From a functional perspective, the abundance of genes associated with biofertilization, nutrient acquisition, phytohormone signaling, and volatile metabolite production underscores the multifunctional nature of these strains as plant growth-promoting rhizobacteria (PGPR). The consistently high proportion of unannotated genes across all scales highlights the presence of substantial “functional dark matter” with unexplored biotechnological potential. Particularly noteworthy is the identification of genes linked to alternative Rubisco-like carbon fixation pathways, suggesting a previously underappreciated metabolic plasticity in *B. velezensis* that may confer adaptive advantages under carbon-limited conditions.

Overall, the integration of phenotypic, genomic, and phylogenetic evidence positions these strains as valuable biological resources for sustainable agricultural applications, while also emphasizing the evolutionary complexity of *B. velezensis* and need for further functional studies to elucidate the ecological and biotechnological roles of its variable and uncharacterized genes.

Author Contributions: Conceptualization, S.F.V. and H.M.S.E.; methodology, S.F.V., H.M.S.E. and Y.Y.D.R.; software, S.F.V., H.M.S.E. and A.F.M.; validation, J.A.L., V.C.R. and A.J.G.L.; formal analysis, S.F.V. and H.M.S.E.; investigation, H.M.S.E., S.F.V., A.F.M. and Y.Y.D.R.; data curation, S.F.V.; writing—original draft preparation, S.F.V.; writing—review and editing, A.J.G.L., and J.A.L.; visualization, V.C.R., J.A.L.; supervision, S.F.V.; project administration, S.F.V.; funding acquisition, S.F.V. All authors have read and agreed to the published version of the manuscript.”.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable.

Data Availability Statement: The original data presented in the study are openly available in the NCBI GeneBank database, which provides unrestricted public access to bacterial genome information, under the accession numbers NZ_MLCV00000000.1, MLCX00000000.1, MKWT00000000.1 and MLCW00000000.1 corresponding to strains 2A-2B, 3A-6A, 2A-10A and 3A-25B, respectively.

Acknowledgments: We acknowledge SECIHTI (Secretaría de Ciencia, Humanidades, Tecnología e Innovación) for providing graduate scholarship to HMSE. We also thank Dr. Raúl Rodríguez Guerra for kindly supplying the phytopathogenic fungi used in this study.

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

Abbreviations

The following abbreviations are used in this manuscript:

PGPTs	plant growth-promoting traits
PCWDE	plant cell-wall degrading enzymes
4CNM+1ref	four Mexican strains plus one reference

POCP Percentage of conserved protein

References

1. Figueiredo, J.E.F., et al., *Bacillus velezensis* CNPMS-22 as biocontrol agent of pathogenic fungi and plant growth promoter. *Front Microbiol*, 2025. **16**: p. 1522136.
2. Chen, M., et al., Lipopeptides from *Bacillus velezensis* induced apoptosis-like cell death in the pathogenic fungus *Fusarium concentricum*. *J Appl Microbiol*, 2024. **135**(3).
3. Xia, L., et al., Biosynthetic gene cluster profiling predicts the positive association between antagonism and phylogeny in *Bacillus*. *Nat Commun*, 2022. **13**(1): p. 1023.
4. De la Cruz-Rodriguez, Y., et al., Biosynthetic Gene Clusters in Sequenced Genomes of Four Contrasting Rhizobacteria in Phytopathogen Inhibition and Interaction with *Capsicum annuum* Roots. *Microbiol Spectr*, 2023. **11**(3): p. e0307222.
5. Jin, M., et al., Identification and Genomic Insights into the Biological Control and Growth-Promoting Mechanism of *Bacillus velezensis* L11-7, a Potential Biocontrol Agent of Passion Fruit Stem Basal Rot. *Microorganisms*, 2025. **13**(9).
6. Bayisa, R.A., Enhancing resistance of *Sesamum indicum* against *Alternaria sesami* through *Bacillus velezensis* AR1. *Pest Manag Sci*, 2020. **76**(11): p. 3577-3586.
7. Aunkam, P., et al., Mechanisms of Cannabis Growth Promotion by *Bacillus velezensis* S141. *Plants (Basel)*, 2024. **13**(21).
8. Ma, X., et al., *Bacillus velezensis* HR6-1 enhances salt tolerance in tomato by increasing endogenous cytokinin content and improving ROS scavenging. *Microbiol Res*, 2025. **296**: p. 128143.
9. Gao, S., et al., *Bacillus velezensis* QSE-21 cell-free supernatant primes resistance and outperforms live cells in controlling *Botrytis cinerea* on tomato. *Front Microbiol*, 2025. **16**: p. 1639396.
10. Chen, L., et al., *Bacillus velezensis* CLA178-Induced Systemic Resistance of *Rosa multiflora* Against Crown Gall Disease. *Front Microbiol*, 2020. **11**: p. 587667.
11. Yin, X., et al., Suppression of Grape White Rot Caused by *Coniella vitis* Using the Potential Biocontrol Agent *Bacillus velezensis* GSBZ09. *Pathogens*, 2022. **11**(2).
12. Chen, L., et al., Antimicrobial, plant growth-promoting and genomic properties of the peanut endophyte *Bacillus velezensis* LDO2. *Microbiol Res*, 2019. **218**: p. 41-48.
13. Wang, B., et al., Effects of two *Bacillus velezensis* strains isolated from different sources on the growth of *Capsicum annuum*. *Front Microbiol*, 2024. **15**: p. 1504660.
14. Wang, Z., et al., Solubilization of K and P nutrients from coal gangue by *Bacillus velezensis*. *Appl Environ Microbiol*, 2024. **90**(11): p. e0153824.
15. Lv, D., et al., Effect of *Bacillus velezensis* on the structure of the rhizosphere microbial community and yield of soybean. *BMC Plant Biol*, 2025. **25**(1): p. 1052.
16. Nandni, et al., Deciphering the Potential of Sulphur-Oxidizing Bacteria for Sulphate Production Correlating with pH Change. *Microb Ecol*, 2023. **86**(4): p. 2282-2292.
17. Wang, C., et al., Effects of *Bacillus velezensis* FKM10 for Promoting the Growth of *Malus hupehensis* Rehd. and Inhibiting *Fusarium verticillioides*. *Front Microbiol*, 2019. **10**: p. 2889.
18. Tan, T., et al., Siderophore-mediated iron enrichment in the biofilm matrix enhances plant iron nutrition. *Cell Rep*, 2025. **44**(11): p. 116481.
19. Rabbee, M.F., et al., *Bacillus velezensis*: A Valuable Member of Bioactive Molecules within Plant Microbiomes. *Molecules*, 2019. **24**(6).
20. Perez-Garcia, A., D. Romero, and A. de Vicente, Plant protection and growth stimulation by microorganisms: biotechnological applications of Bacilli in agriculture. *Curr Opin Biotechnol*, 2011. **22**(2): p. 187-93.
21. Vernikos, G., et al., *Ten years of pan-genome analyses*. *Curr Opin Microbiol*, 2015. **23**: p. 148-54.
22. Chun, B.H., et al., Genomic and metabolic features of the *Bacillus amyloliquefaciens* group- *B. amyloliquefaciens*, *B. velezensis*, and *B. siamensis*- revealed by pan-genome analysis. *Food Microbiol*, 2019. **77**: p. 146-157.

23. Ananev, A.A., et al., Whole Genome Sequencing of *Bacillus velezensis* AMR25, an Effective Antagonist Strain against Plant Pathogens. *Microorganisms*, 2024. **12**(8).
24. Ding, Y., S. Wei, and G. Zhang, Complete genome sequence analysis of a biosurfactant-producing bacterium *Bacillus velezensis* L2D39. *Mar Genomics*, 2024. **76**: p. 101113.
25. Martinez-Raudales, I., et al., Draft genome sequence of *Bacillus velezensis* 2A-2B strain: a rhizospheric inhabitant of *Sporobolus airoides* (Torr.) Torr., with antifungal activity against root rot causing phytopathogens. *Stand Genomic Sci*, 2017. **12**: p. 73.
26. Martinez-Raudales, I., et al., Draft Genome Sequence of *Bacillus velezensis* 3A-25B, a Strain with Biocontrol Activity against Fungal and Oomycete Root Plant Phytopathogens, Isolated from Grassland Soil. *Genome Announc*, 2017. **5**(39).
27. Foundation, F.S., GNU PSPP (Version 2.0.1) [Computer > Software]. 2015.
28. Sunwoo, J.Y., Lee, Y.K. & Hwang, B.K. , Induced resistance against *Phytophthora capsici* in pepper plants in response to DL- β -amino-n-butyric acid. *Eur J Plant Pathol*, 1996. **102**: p. 663-670.
29. Dieckmann, M.A., et al., EDGAR3.0: comparative genomics and phylogenomics on a scalable infrastructure. *Nucleic Acids Res*, 2021. **49**(W1): p. W185-W192.
30. Du, J., et al., KEGG-PATH: Kyoto encyclopedia of genes and genomes-based pathway analysis using a path analysis model. *Mol Biosyst*, 2014. **10**(9): p. 2441-7.
31. Galperin, M.Y., et al., COG database update: focus on microbial diversity, model organisms, and widespread pathogens. *Nucleic Acids Res*, 2021. **49**(D1): p. D274-D281.
32. Park, S.C., et al., Large-Scale Genomics Reveals the Genetic Characteristics of Seven Species and Importance of Phylogenetic Distance for Estimating Pan-Genome Size. *Front Microbiol*, 2019. **10**: p. 834.
33. Patz S., G.A., Ruppel S., Rodríguez-Palenzuela P., Huson DH. , PLaBAs: A comprehensive web resource for analyzing the plant growth-promoting potential of plant-associated bacteria. *bioRxiv*, 2021: p. 11-16.
34. Martinez-Garcia, P.M., et al., Prediction of bacterial associations with plants using a supervised machine-learning approach. *Environ Microbiol*, 2016. **18**(12): p. 4847-4861.
35. Jain, C., et al., High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. *Nat Commun*, 2018. **9**(1): p. 5114.
36. Qin, Q.L., et al., A proposed genus boundary for the prokaryotes based on genomic insights. *J Bacteriol*, 2014. **196**(12): p. 2210-5.
37. Tettelin, H., et al., Genome analysis of multiple pathogenic isolates of *Streptococcus agalactiae*: implications for the microbial "pan-genome". *Proc Natl Acad Sci U S A*, 2005. **102**(39): p. 13950-5.
38. Deng, W., C. Li, and J. Xie, The underlying mechanism of bacterial TetR/AcrR family transcriptional repressors. *Cell Signal*, 2013. **25**(7): p. 1608-13.
39. Cariss, S.J., et al., YieJ (CbrC) mediates CreBC-dependent colicin E2 tolerance in *Escherichia coli*. *J Bacteriol*, 2010. **192**(13): p. 3329-36.
40. Lanferdini, D.K., et al., Volatilome of Amazonian strains of *Bacillus velezensis* and its role on fungal biocontrol in grapes. *Food Res Int*, 2025. **220**: p. 117106.
41. Elhjouji, H., et al., Biocontrol Potential of *Bacillus velezensis* RS65 Against *Phytophthora infestans*: A Sustainable Strategy for Managing Tomato Late Blight. *Microorganisms*, 2025. **13**(3).
42. Li, H., et al., Inoculation of *Bacillus velezensis* Bv-116 and its bio-organic fertilizer serve as an environmental friendly biocontrol strategy against cucumber *Fusarium* wilt. *Front Plant Sci*, 2024. **15**: p. 1467265.
43. Shah, S., et al., Eco-friendly Biocontrol of *Ralstonia solanacearum* and Plant Growth Promotion in Tobacco Using Garbage Enzyme and *Bacillus velezensis* A1. *Mol Biotechnol*, 2025.
44. Escudeiro, P., C.S. Henry, and R.P.M. Dias, Functional characterization of prokaryotic dark matter: the road so far and what lies ahead. *Curr Res Microb Sci*, 2022. **3**: p. 100159.
45. Nicholson, W.L., et al., Resistance of *Bacillus* endospores to extreme terrestrial and extraterrestrial environments. *Microbiol Mol Biol Rev*, 2000. **64**(3): p. 548-72.
46. Smith, D.J., et al., Intercontinental dispersal of bacteria and archaea by transpacific winds. *Appl Environ Microbiol*, 2013. **79**(4): p. 1134-9.

47. Albert Leopold Müller, J.R.d.R., Casey R J Hubert, Kasper Urup Kjeldsen, Ilias Lagkouvardos, David Berry, Bo Barker Jørgensen, Alexander Loy,, *Endospores of thermophilic bacteria as tracers of microbial dispersal by ocean currents*. The ISME Journal, 214. 8(6): p. 1153-1165.
48. McDonald, B.A. and E.H. Stukenbrock, Rapid emergence of pathogens in agro-ecosystems: global threats to agricultural sustainability and food security. *Philos Trans R Soc Lond B Biol Sci*, 2016. 371(1709).
49. Kim, Y., et al., Pan-genome analysis of Bacillus for microbiome profiling. *Sci Rep*, 2017. 7(1): p. 10984.
50. Mingqi Li, et al., Molecular understanding of autotrophic CO₂-fixing bacterial communities in composting based on RuBisCO genes analysis. *Journal of Biotechnology*, 2020. 320: p. 8.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.