



COMPARATIVE GENOMIC PROFILING OF PLANT-ASSOCIATED BACTERIA WITH AGRICULTURAL IMPORTANCE

Neelakanth S. Wali¹, Vishwajit B Darekar^{2*}

¹Department of Zoology, M.S.I. Degree College, Kalaburagi-585102, Karnataka, India

^{2*} Department of Post Graduate Studies and Research in Zoology, Sharnbasva University, Kalaburgi-585103, Karnataka, Email ID: vishwajeetdarekar@gmail.com , ORCID: 0009-0005-5190-758X

Corresponding Author: Vishwajit B Darekar, vishwajeetdarekar@gmail.com

ABSTRACT

Plant-associated bacteria have significant contributions to plant-growth promotion, nutrient cycling, stress tolerance, rhizosphere adaptation and biological control thus are a valuable part of sustainable agriculture. This study was performed to comparatively estimate the genomic characteristics of the plant-associated bacterial genera, which were based on the genome size, GC content, coding sequence abundance, scaffold number, and multivariate genomic patterns. A total of 944 bacterial genome records of 11 genera were studied. The results of descriptive analysis revealed significant differences in the size of the genomes between the genera, with Paraburkholderia, Streptomyces and Bradyrhizobium having relatively large genomes and Streptomyces having the highest mean GC content. The genus-level differences in genome size, GC content, coding sequences, and scaffold number were significant ($p < 0.001$). We found that abundance of coding sequences (abundance of proteins) was positively correlated with genome size ($\rho = 0.964$; $p < 0.001$), suggesting that larger genomes have more protein-coding potential. PCA revealed that the 1st and 2nd principal components accounted for 90.71% of the total genomic variation, indicating a clear distinction among the different groups of bacteria. In general, the results show high genomic diversity within the prokaryotic populations associated with plants that are germplasm resources for agricultural applications, including the identification of taxa that may be useful in biofertilization, biostimulation, nutrient acquisition and biocontrol.

Keywords: plant-associated bacteria; comparative genomics; PGPR; genome size; GC content; sustainable agriculture

1.Introduction

Plant-associated bacteria are an important part of the phytomicrobiome that are vital for plant health, soil fertility, nutrient availability and productivity in agriculture. These microorganisms occupy a variety of plant-associated microhabitats such as the rhizosphere, the phyllosphere and the internal tissues of the plants, in which they have complex interactions that can have direct or indirect effects on the performance of the crop. The beneficial bacteria can enhance nutrient uptake, root development, reduce phytopathogens, regulate the production of phytohormones and boost plant stress tolerance. Therefore, the use of plant-associated bacteria has gained considerable importance in decreasing the use of chemical fertilizers and pesticides and contributing to environmentally sustainable systems for crop production (Gupta et al., 2021; Chouhan et al., 2021).

Plant growth-promoting rhizobacteria (PGPR) have been the subject of significant research due to their ability to affect several physiological and biochemical processes in plants. The biological nitrogen fixation, phosphate solubilisation, production of siderophores, synthesis of phytohormones and modulation of plant stress responses are the mechanisms of PGPR improve crop growth. Their properties render them good candidates for the development of microbial inoculants and biofertilizers to be used as supplements or partial replacements of conventional agricultural inputs (Aloo et al., 2022). The rhizosphere is a dynamic root-soil ecosystem with a high density of microorganisms that is influenced by root exudates and environmental factors and can affect nutrient cycling and plant establishment (Bhadrecha et al., 2023). Soil-plant-microorganism interactions, hence, can be termed as a basic biological mechanism of enhancing the efficiency of use of nutrients and sustainable crop productivity (Das et al., 2022).

Crop-associated bacteria also make significant contributions to crop resilience in response to changing environmental conditions. Abiotic stress factors, such as salinity, drought, heavy metals, and other stressors, can impact crop yields by disrupting nutrient uptake, water relations, and cell metabolism. These effects can be mitigated by beneficial bacteria, which change hormone levels, increase concentrations of osmoprotective metabolites, increase nutrient availability, and activate antioxidant defenses. The potential application of PGPR in improving plants' tolerance to saline environments has been gaining traction in the field of climate-resilient agriculture (Mishra et al., 2021; Kumawat et al., 2022). In contaminated soils, PGPR can also boost the phytoremediation process by raising plant biomass, metal mobilization/immobilization, and stress tolerance, and thus, combining microbial biotechnology with

environmental restoration and agricultural land management (Bhanse et al., 2022).

Although these bacteria are important for providing services, agriculturally beneficial taxa are genetically and biologically diverse. Genome size, nucleotide composition, coding capacity, genome structure and accessory genome elements can impact adaptation to plant-associated environments and the function of bacterial strains in agriculture-relevant applications. Knowledge of such genomic variation is thus crucial in understanding the separation of the bacterial groups with different ecological strategies and potential applications. However, the study of plant microbiota has been undergoing a paradigm shift, from the characterization of known cultivated species to the genomic and systems biology approaches that allows a more comprehensive characterization of the microbial composition, function, and adaptation (Chialva et al., 2022). The integration of genomic data with data on plant expression (transcriptomics), protein expression (proteomics), and metabolites (metabolomics) has also enhanced the capacity to study plant–microbe interactions (Diwan et al., 2022).

Comparative genomic profiling offers a useful tool to study the conserved and variable genomic features of agriculturally important bacteria. Genome size can be used to infer the overall genetic capacity and ecological flexibility, and the GC content reflects fundamental variations in the composition of the four nucleotides and evolutionary history. Likewise, the abundance of coding sequences can yield insights into the protein-coding potential, while the characteristics of scaffolds and replicons can offer clues regarding the differences in genome organization and assembly structure. Comparative analysis of these parameters could then serve to discover genomic patterns that distinguish the major groups of plant-associated bacteria and serve as a basis for future functional and evolutionary studies.

The growing amount of genomic data has also propelled the development of the concept of shaping and creating useful microbial communities for farming. The engineering of the microbiome and synthetic biology has opened the door to possibilities for enhancing functions like biological nitrogen fixation, nutrient uptake, and rhizosphere competence (Han & Yoshikuni, 2022). In the future, microbial technologies for crop production could be further developed using engineered plant-associated microbial communities, which will render the technologies more predictable and efficient (Ke et al., 2021). To be successful, however, it is essential to have a clear idea of the genomic diversity found within the candidate bacterial taxa. Additionally, understanding the beneficial microorganisms in the microbiome of plants is crucial to characterize microorganisms with agricultural potential and solve challenges related to field performance and establishment of the microbial community (Ayilara et al., 2023).

Hence, the present study aims to comparatively depict some of the important characteristics of the genomes of the selected genera of bacteria associated with plants with recognized significance in agriculture. Using genome size, GC composition, abundance of coding sequences, scaffold structure and multivariate patterns of genomes, it aims to interpret the genomic diversity of bacterial groups related to plant growth promotion, nutrient cycling, rhizosphere adaptation, and related agricultural functions.

Objectives

- 1.To compare the genome size, GC content, abundance of coding sequences, and the characteristics of scaffolds among the agriculturally important plant-associated bacterial genera.
- 2.To assess the interdependency of the key genomic traits and identify important genomic differences between the genera.
- 3.To detect if there are multivariate patterns and/or clustering trends that can discriminate between the main groups of agriculturally relevant plant-associated bacteria.

2. Methodology (12pt, Bold)

2.1 Research Design and Dataset Source

The design of this study was quantitative and comparative genomic and was based on secondary genomic data that was publicly available in the “Genome Information for Sequenced Organisms” dataset (Nugent, 2018). The prokaryotic subset was chosen since the study was on agriculturally relevant plant-associated bacteria. There were 9,351 prokaryotic genome records included in the dataset, and each record had information describing the organism’s name, strain, BioSample, BioProject, assembly accession, assembly level, genome size, GC content, number of replicons, number of scaffolds, number of coding sequences, release date and GenBank/RefSeq links.

2.2 Selection of Agriculturally Important Bacteria

The data were first processed to find genera of bacteria commonly associated with plant growth promotion, nitrogen fixation, nutrient cycles, rhizosphere colonization, biocontrol and agricultural biotechnology. The genera selected were *Rhizobium*, *Bradyrhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Azospirillum*, *Azotobacter*, *Pseudomonas*, *Bacillus*, *Paenibacillus*, *Paraburkholderia* and *Streptomyces*. Where appropriate, records lacking sufficient genomic information were excluded, as were duplicate assembly records.

2.3 Data Extraction and Preparation

Relevant genomic variables were retrieved and set up for the comparison analysis. These consisted of the bacterial genus, species, strain, accession number of the assembly, assembly level, genome size, GC content, number of replicons, number of scaffolds, and number of coding sequences. Taxonomic labels were standardized and invalid numerical data were removed from analyses that relied on the missing numerical data.

2.4 Classification of Bacterial Groups

Selected bacteria were classified in the following groups based on their main agricultural functions. Nitrogen-fixing genera, including *Rhizobium*, *Bradyrhizobium*, *Sinorhizobium*, *Mesorhizobium*, *Azospirillum*, and *Azotobacter*, and other genera such as *Pseudomonas*, *Bacillus*, *Paenibacillus*, *Paraburkholderia*, and *Streptomyces* were classified as primarily plant-growth-promoting, rhizosphere-associated or biocontrol-related bacteria.

2.5 Comparative Genomic Profiling

Comparative genomic profiling was done based on the Genome size, GC content, abundance of coding sequences, number of scaffolds, replicon information, and assembly characteristics. Descriptive statistics such as mean, standard deviation, median, minimum and maximum values were computed. The genome architecture of different groups of bacteria within the agricultural sector and among genera was then explored.

2.6 Statistical and Multivariate Analysis

Genomic variation was assessed by making statistical comparisons among bacterial groups. When the parametric assumptions were met, one-way ANOVA was used; otherwise, appropriate non-parametric methods were considered. The relationships between the genome size, GC content, number of scaffolds and number of coding sequences were evaluated by correlation analysis. Multivariate genomic patterns were detected by means of principal component analysis and hierarchical clustering. A p-value of < 0.05 was deemed to be statistically significant.

2.7 Data Visualization

The abundance of coding sequences, GC content and genome size were compared among bacterial genera

using box plots. The relationships between major genomic variables were evaluated using scatter plots, and relationships among agriculturally important bacteria were visualized using heatmaps and principal component plots for genomic similarity, clustering and differentiation.

2.8 Genomic Accession and Data Traceability

Information from BioSample, BioProject, assembly accession, GenBank and RefSeq were also preserved during analysis to allow the traceability of individual bacterial genomes. These identifiers also enable access to genome assemblies, which will facilitate future functional annotation, phylogenomic analysis, pan-genome analysis, or identification of agriculturally relevant genes.

2.9 Ethical Considerations

This study used only publicly available secondary genomic data and did not involve human participants, animals, or direct biological sampling. Therefore, ethical approval and informed consent were not required. All genomic records were used strictly for academic purposes, with appropriate acknowledgement of the original dataset and public repositories. Data integrity was maintained by preserving original accession information and avoiding fabrication or inappropriate modification of genomic values.

3. Results

3.1 Distribution and General Genomic Characteristics

The Genome Information for Sequenced Organisms dataset was screened, and 944 genome records of 11 genera of bacteria associated with agricultural plants were retained. The genus *Bacillus* was the most abundant ($n = 392$), and *Pseudomonas* ($n = 273$) and *Streptomyces* ($n = 90$) were the next most abundant. The genera with the lowest numbers were *Azospirillum* ($n = 6$) and *Azotobacter* ($n = 5$). Table 1 shows that there was considerable variation in the genome size, GC content, coding sequences and scaffold numbers among the selected genera of bacteria.

Table 1. Genomic characteristics of selected agriculturally important bacterial genera

Genus	n	Mean Genome Size (Mb)	Mean GC (%)	Mean CDS	Mean Scaffolds
<i>Bacillus</i>	392	4.72	40.79	4,626.46	2.36
<i>Pseudomonas</i>	273	6.28	63.33	5,643.27	1.24
<i>Streptomyces</i>	90	8.62	72.01	7,185.61	1.61
<i>Rhizobium</i>	54	6.77	60.88	6,248.96	5.28
<i>Paenibacillus</i>	53	6.29	48.29	5,327.62	1.45
<i>Sinorhizobium</i>	30	6.89	62.11	6,349.57	3.83
<i>Bradyrhizobium</i>	16	8.60	64.08	7,698.88	1.38
<i>Paraburkholderia</i>	15	8.71	62.37	7,505.47	3.67
<i>Mesorhizobium</i>	10	6.85	62.88	6,396.50	2.20
<i>Azospirillum</i>	6	6.54	67.97	5,628.33	5.83
<i>Azotobacter</i>	5	5.25	65.76	4,658.20	2.80

All the selected records had a mean genome size of 5.98 ± 1.57 Mb, with the range extending from 3.07 to 12.01 Mb. Mean GC content was $53.80 \pm 12.41\%$, while the mean CDS count was $5,473.98 \pm 1,237.57$. The median value for scaffold number was 1, and the highest value reached was 15, as shown in Table 2, which demonstrates that there was a moderate level of dispersion for this data.

Table 2. Overall descriptive statistics of major genomic variables

Genomic Variable	Valid n	Mean	SD	Median	Minimum	Maximum
Genome size (Mb)	944	5.98	1.57	6.07	3.07	12.01
GC content (%)	943	53.80	12.41	59.11	34.63	74.80

Scaffolds	944	2.16	1.97	1.00	1.00	15.00
Coding sequences	935	5,473.98	1,237.57	5,643.00	2,251.00	10,009.00

The histogram of genome size distribution of the bacterial genera is provided in Figure 1. The average genome size of *Bacillus* was the smallest of the studied genera, while *Paraburkholderia*, *Streptomyces* and *Bradyrhizobium* had the largest genome sizes.

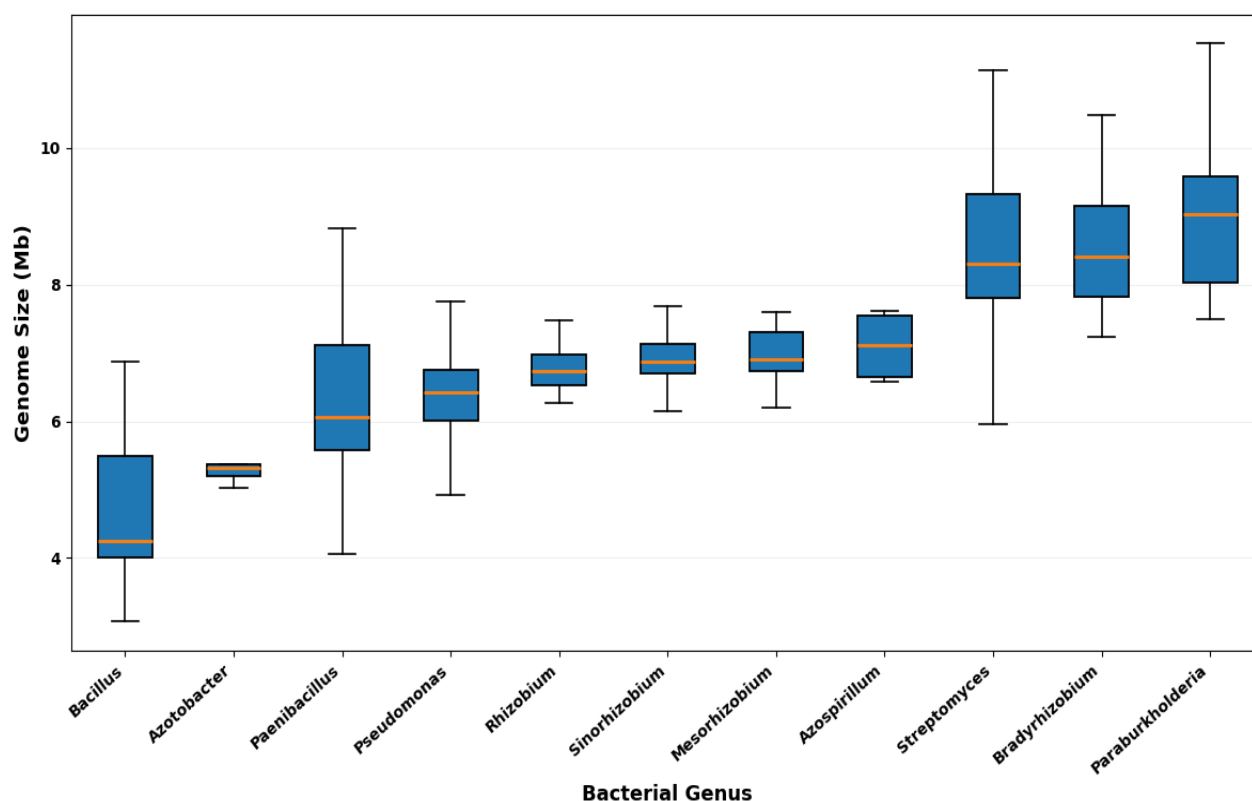


Figure 1. Distribution of genome size among selected agriculturally important plant-associated bacterial genera

As illustrated in Figure 2, substantial variation in mean GC content was observed among the selected bacterial genera. *Streptomyces* exhibited the highest mean GC content (72.01%), followed by *Azospirillum* (67.97%) and *Azotobacter* (65.76%), whereas *Bacillus* showed the lowest mean value (40.79%). This pattern demonstrates pronounced differences in genomic nucleotide composition among the agriculturally important bacterial taxa.

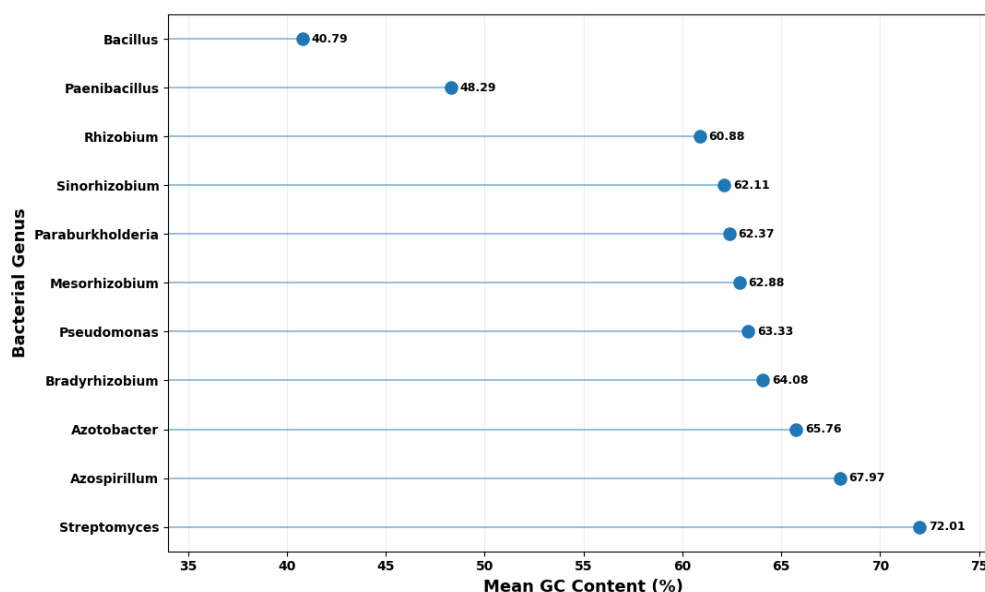


Figure 2. Mean GC content across selected agriculturally important plant-associated bacterial genera

3.2 Genus-Level Differences in Genomic Characteristics

For all major genomic variables, significant differences were found between the genera of bacteria. As presented in Table 3, there was a significant variation in genome size among the genera ($H = 630.558$, $p < 0.001$). GC content also differed markedly ($H = 795.477$, $p < 0.001$), while significant variation was observed for CDS abundance ($H = 486.400$, $p < 0.001$) and scaffold number ($H = 319.746$, $p < 0.001$).

Table 3. Kruskal–Wallis comparison of genomic characteristics among bacterial genera

Genomic Variable	Kruskal–Wallis H	p-value	Interpretation
Genome size (Mb)	630.558	<0.001	Significant
GC content (%)	795.477	<0.001	Significant
Coding sequences	486.400	<0.001	Significant
Scaffold number	319.746	<0.001	Significant

The findings indicate that the genomic architecture of the different bacterial genera is significantly divergent and thereby important for agriculture. The great differences in genome size and GC content indicate the strong taxonomic division of genome composition and coding capacity.

3.3 Relationships Among Genomic Features

As listed in Table 4, there were several statistically significant correlations between the genomic variables. Several statistically significant correlations were found between the genomic variables with Spearman correlation analysis as shown in Table 4. There was a very strong positive correlation between genome size and CDS abundance ($\rho = 0.964$, $p < 0.001$), which suggested that the greater the genome size, the higher the number of protein-coding sequences in the genome. The GC content was also positively correlated with genome size ($\rho = 0.629$, $p < 0.001$). There was a moderate positive correlation between GC content and CDS abundance ($\rho = 0.514$, $p < 0.001$). Scaffold number was weakly positively associated with genome size ($\rho = 0.209$, $p < 0.001$) and CDS abundance ($\rho = 0.305$, $p < 0.001$), whereas it was negatively correlated with GC content ($\rho = -0.255$, $p < 0.001$).

Table 4. Spearman correlations among major genomic characteristics

Variables	Genome Size	GC Content	Scaffolds	CDS
Genome Size	1.000	0.629***	0.209***	0.964***
GC Content	0.629***	1.000	-0.255***	0.514***
Scaffolds	0.209***	-0.255***	1.000	0.305***
CDS	0.964***	0.514***	0.305***	1.000

*Note: ** $p < 0.001$.

As shown in Figure 3, all the selected bacterial genomes showed a strong positive correlation between genome size and CDS abundance, with the increase in number being evident.

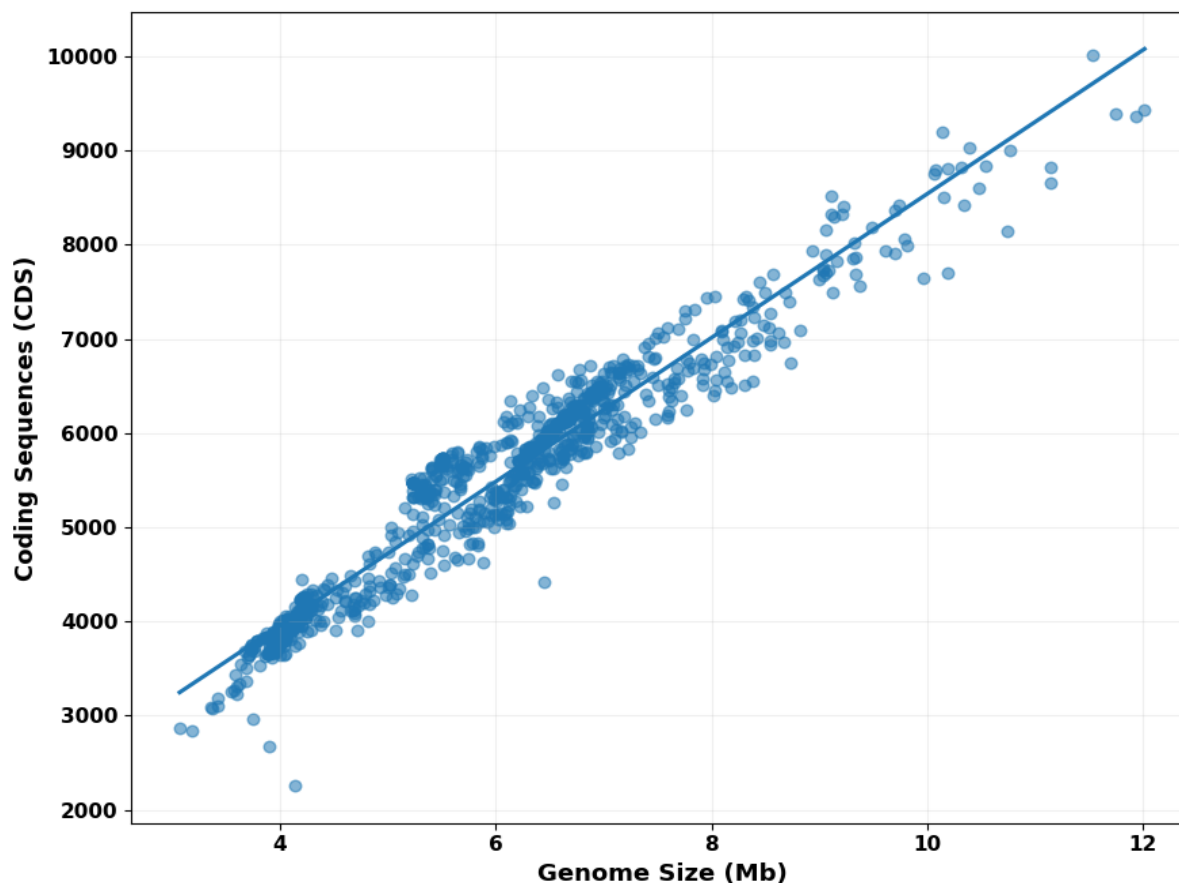


Figure 3. Relationship between genome size and coding sequence abundance among selected plant-associated bacterial genomes

3.4 Multivariate Genomic Profiling

Genome size, GC content, scaffold number and CDS abundance were analyzed by principal component analysis. There were 934 complete observations. Table 5 indicates that the first two PCs accounted for 90.71% of the total genomic variance, 61.51% by PC1 and 29.20% by PC2.

Table 5. Principal component loadings of genomic characteristics

Variable	PC1 Loading	PC2 Loading
Genome size	0.625	0.031
GC content	0.471	-0.478
Scaffolds	0.134	0.865

CDS	0.608	0.147
Variance explained (%)	61.51	29.20
Cumulative variance (%)	61.51	90.71

Overall genomic capacity was best indicated by genome size and CDS abundance, contributing most to PC1. The most prominent variable affecting PC2 was scaffold number, and the moderate negative loading was for GC content. The PCA ordination in Figure 4 also highlighted the distinct genomic positioning between the selected genera, indicative of genome size, coding capacity, nucleotide composition and assembly characteristics.

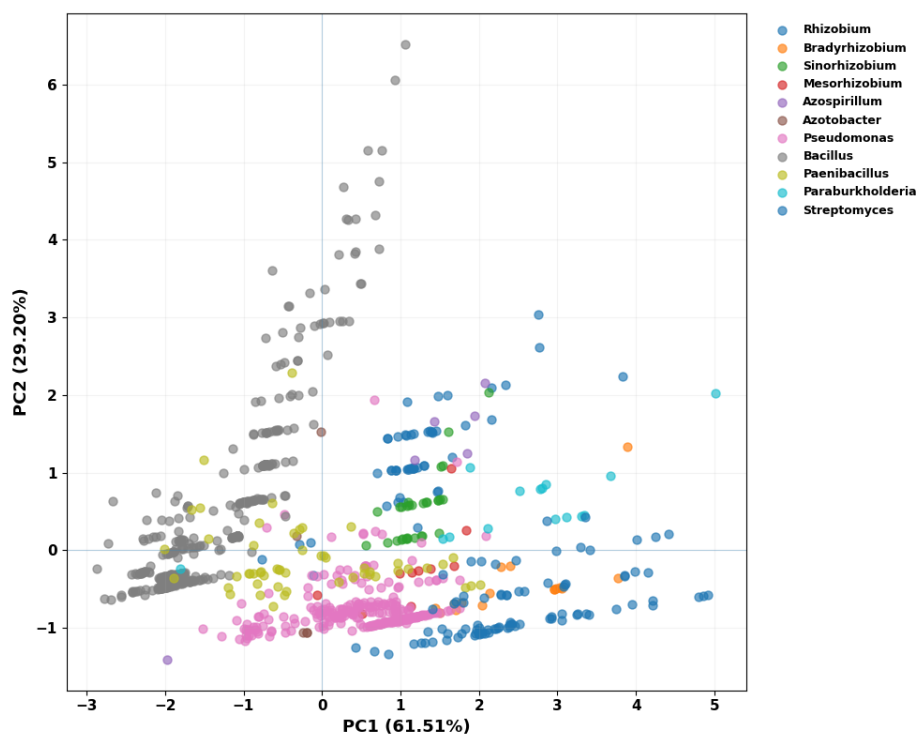


Figure 4. Principal component analysis showing multivariate genomic differentiation among agriculturally important plant-associated bacterial genera

3.5 Summary of Results

The comparative analysis showed that the selected agriculturally important plant-associated bacteria have significant genomic variations. The genera showed significant differences in genome size, GC content, coding sequence abundance and scaffold number (all $p < 0.001$). The genomes of Paraburkholderia and

Streptomyces, and Bradyrhizobium were relatively large, and Streptomyces had the highest GC content. Coding sequence abundance was strongly correlated with genome size ($p = 0.964$), suggesting genome expansion is closely linked to the amount of coding sequences. The multivariate analysis also revealed that the combination of genome size, the abundance of CDS, GC composition, and scaffold structure was able to distinguish bacterial genera, with a maximum of 90.71% of the total variation explained by the first two principal components. The results presented suggest different architectures of the genomes in bacterial groups that are important for plant growth promotion, nutrient cycling, rhizosphere adaptation, and use in agriculture.

5. Discussion

The present comparative genomic analysis showed that the genome sizes, GC content, abundance of coding sequences, number of scaffolds, and multivariate genome structure were all highly variable across the agriculturally relevant genera of plant-associated bacteria. These differences suggest that plant-associated beneficial bacteria have different genomic architectures, which could be due to differences in ecological adaptation, metabolic ability and/or plant-associated function. The substantial intergenus diversity also suggests that plant-associated bacteria are not necessarily genomically uniform, even if they share, in broad terms, similar types of functions in the plant, including nutrient uptake, plant growth promotion or rhizosphere colonization.

The selected genera differed significantly in genome size, with Paraburkholderia, Streptomyces and Bradyrhizobium having relatively large genomes, and Bacillus having smaller genomes. Larger genomes could mean more functional flexibility, as it enables the presence of more genes for environmental sensing, substrate utilization, stress response, and plant-associated interactions. Whole-genome studies of beneficial rhizobacteria also revealed that the genomes of species correlated with a variety of metabolic and ecological properties (Aloo et al., 2020). The variability in genome structure among closely related strains of plant growth-promoting bacteria associated with maize has also been shown through a genomic analysis of these microbes (Babalola et al., 2022). These observations are consistent with the present result that genomic complexity has a great range between different agriculturally relevant taxa.

One of the most significant observations of the study was the highly positive correlation between genome size and abundance of coding sequence. The correlation indicates that bigger bacteria genome sizes are linked to larger protein-coding potential. This is biologically feasible since larger genomes of bacteria tend to have more diverse sets of metabolic, regulatory, transport, and environmental-response genes. Similarly,

the genome of nitrogen-fixing and plant growth-promoting bacteria has been shown to have large coding potential associated with various plant-associated traits (Bandeppa et al., 2022). The coding sequence analysis for understanding plant-growth-promoting potential (PGP) was also useful for the whole-genome characterisation of *Caulobacter segnis* CBR1 (Berrios, 2021). Likewise, genomic evaluation of *Isophtericola* sp. Through complete genome characterization, AK164 revealed a wide range of genetic adaptations to benefit plant growth that occur in the rhizosphere (Alghamdi et al., 2023).

A significant difference was also observed between genera with *Streptomyces* having higher GC content and *Bacillus* comparatively low GC content. These differences are indicative of highly divergent taxonomical and evolutionary relationships and may also be related to genome stability and codon usage, as well as ecological specialization. Comparative genomic analysis of rhizosphere bacteria has revealed that even in bacteria with a common association with plants, there can be considerable differences in nucleotide composition (Leontidou et al., 2020). This positive correlation between coding sequence abundance and GC content in the current study may thus be an indirect consequence of the general genomic structure, or because the two factors are correlated in some regions of the genome.

Of particular interest are the genomic profiles of the taxa related to *Bacillus*; these include many plant growth-promoting and biocontrol strains that are widely studied. Carvalho et al. (2023) genome analyzed *Bacillus velezensis* BIB0110 that showed genomic properties related to plant growth promotion and antagonism potential. Similar evidence has been presented for the use of the strains of *Bacillus subtilis* BS87 and *Bacillus megaterium* BM89 for boosting crop productivity through whole-genome analysis (Chandra et al., 2021). Likewise, genomic analysis of *Bacillus altitudinis* LZP02 revealed the need for genome information to better understand the mechanism of rice-rhizosphere plant growth-promoting bacteria (Jiao et al., 2022). Another noteworthy feature of genome analysis of *B. velezensis* K1 was the significant potential for antagonistic and plant-beneficial activities encoded in its genome (Nanjani et al., 2022). The present observation that genome architecture and coding capacity can give important comparative indicators between agriculturally useful groups of bacteria are consistent with these findings.

The multivariate analysis also showed that these four parameters genome size, CDS abundance, GC content, and scaffold structure separate the bacterial genera. The first two principal components accounted for 90.71% of the total variation, showing that there was relatively little variation among the genomic variables. A comparable genome-based strategy has been applied in the characterization of plant-associated bacteria under stressful conditions. For instance, the genome of *Sphingomonas* sp. has been

analyzed. The drought-resistant features identified by Cra20 and the plant growth-promoting features identified by the psychrotrophic plant-beneficial bacteria from the Himalayan environment are identified. Drought resistance features were identified with Cra20 and features related to plant growth promotion were identified with the psychrotrophic plant-beneficial bacteria from the Himalayan environment. The whole-genome analysis of *Priestia filamentosa* AZC66 revealed the adaptive and agricultural potential of the rhizosphere-associated bacteria that can be explained by its genomic characteristics (Khalifa & Alsowayeh, 2023).

The overall results indicate that the genomes of agriculturally relevant plant-associated bacteria are very diverse. The variation of genome size, GC content, coding potential and multivariate structure is observed and is a useful genomic basis for delineating different groups of bacteria with varying ecological and agricultural potentials. Comparative genomic profiling is therefore an excellent initial step in the screening of bacterial taxa for promising future functional annotation, experimental validation, and use as biofertilizers, bio stimulants or biocontrols.

5. Conclusion

This study demonstrated that the genome size, GC content, coding sequence count, scaffold number, and multivariate genomic characteristics of the main genera of bacteria associated with plants are highly diverse in the agricultural context. There were significant intergeneric differences, suggesting that plant-associated bacteria have different genomic architectures, which may be related to their taxonomic origin, ecological niche and adaptive potential in the plant-associated environment. The relation between genome size and coding sequence abundance was particularly high and positive, indicating that larger bacterial genomes tend to have more coding sequences, and the differences in GC content further illustrated the compositional diversity of bacteria. The multivariate analysis confirmed these results, as it revealed that a combination of genome size, abundance of coding sequences, GC composition and scaffold structure explained most of the genomic variation. Comparative genomic profiling is a valuable tool to investigate agriculturally relevant bacteria, as the genera *Bacillus*, *Pseudomonas*, *Streptomyces*, *Rhizobium*, *Bradyrhizobium* and *Paraburkholderia* are clearly distinguishable. The general conclusion of the findings is that a genomic basis for the identification of bacterial taxa with potential applications in plant growth promotion, nutrient cycling, rhizosphere adaptation and biological control could be obtained. Future research efforts should aim to continue these comparisons using functional annotation, pan-genome analysis, phylogenomics and experimental validation of plant-beneficial genes. These integrated

approaches can help identify and optimize efficient and competitive microbial candidates for use in biofertilizers, biostimulants and sustainable management of crops.

REFERENCES

1. Alghamdi, A. K., Parween, S., Hirt, H., & Saad, M. M. (2023). Complete genome sequence analysis of plant growth-promoting bacterium, *Isoptericola* sp. AK164 isolated from the rhizosphere of *Avicennia marina* growing at the Red Sea coast. *Archives of Microbiology*, 205(9), 307.
2. Aloo, B. N., Mbega, E. R., Makumba, B. A., Friedrich, I., Hertel, R., & Daniel, R. (2020). Whole-genome sequences of three plant growth-promoting rhizobacteria isolated from *Solanum tuberosum* L. rhizosphere in Tanzania. *Microbiology Resource Announcements*, 9(20), 10-1128.
3. Aloo, B. N., Tripathi, V., Makumba, B. A., & Mbega, E. R. (2022). Plant growth-promoting rhizobacterial biofertilizers for crop production: The past, present, and future. *Frontiers in Plant Science*, 13, 1002448.
4. Ayilara, M. S., Adeleke, B. S., & Babalola, O. O. (2023). Bioprospecting and challenges of plant microbiome research for sustainable agriculture, a review on soybean endophytic bacteria. *Microbial ecology*, 85(3), 1113-1135.
5. Babalola, O. O., Agbodjato, N. A., Ayangbenro, A. S., Adjanohoun, A., & Baba-Moussa, L. (2022). Draft Genome Sequences of Four Strains of Plant Growth-Promoting Bacteria Associated with Maize Rhizosphere. *Microbiology Resource Announcements*, 11(11), e00730-22.
6. Bandeppa, S., Phule, A. S., Barbadikar, K. M., Govindannagari, R., B, P. B. M., Kavuru, V. P. B., ... & Chandran, L. P. (2022). Draft genome sequence of *Paenibacillus sonchi* Iirbnf1, a nitrogen-fixing and plant growth-promoting bacterium isolated from rice Rhizosphere. *Microbiology Resource Announcements*, 11(5), e00126-22.
7. Berrios, L. (2021). Complete genome sequence of the plant-growth-promoting bacterium *Caulobacter segnis* CBR1. *Current microbiology*, 78(8), 2935-2942.
8. Bhadrecha, P., Singh, S., & Dwibedi, V. (2023). 'A plant's major strength in rhizosphere': the plant growth promoting rhizobacteria. *Archives of Microbiology*, 205(5), 165.
9. Bhanse, P., Kumar, M., Singh, L., Awasthi, M. K., & Qureshi, A. (2022). Role of plant growth-promoting rhizobacteria in boosting the phytoremediation of stressed soils: Opportunities, challenges, and prospects. *Chemosphere*, 303, 134954.
10. Carvalho, I. D., Apaza-Castillo, G. A., Paz, I. C. P., Guimarães, A. M., Quecine, M. C., & Bonatelli, M. L.

- (2023). Draft genome sequence of the plant growth promoter and biocontrol agent *Bacillus velezensis* strain BIB0110. *Microbiology Resource Announcements*, 12(6), e00231-23.
11. Chandra, A., Chandra, P., & Tripathi, P. (2021). Whole genome sequence insight of two plant growth-promoting bacteria (*B. subtilis* BS87 and *B. megaterium* BM89) isolated and characterized from sugarcane rhizosphere depicting better crop yield potentiality. *Microbiological research*, 247, 126733.
 12. Chialva, M., Lanfranco, L., & Bonfante, P. (2022). The plant microbiota: composition, functions, and engineering. *Current Opinion in Biotechnology*, 73, 135-142.
 13. Chouhan, G. K., Verma, J. P., Jaiswal, D. K., Mukherjee, A., Singh, S., de Araujo Pereira, A. P., ... & Singh, B. K. (2021). Phytomicrobiome for promoting sustainable agriculture and food security: opportunities, challenges, and solutions. *Microbiological Research*, 248, 126763.
 14. Das, P. P., Singh, K. R., Nagpure, G., Mansoori, A., Singh, R. P., Ghazi, I. A., ... & Singh, J. (2022). Plant-soil-microbes: A tripartite interaction for nutrient acquisition and better plant growth for sustainable agricultural practices. *Environmental Research*, 214, 113821.
 15. Diwan, D., Rashid, M. M., & Vaishnav, A. (2022). Current understanding of plant-microbe interaction through the lenses of multi-omics approaches and their benefits in sustainable agriculture. *Microbiological research*, 265, 127180.
 16. Gupta, R., Anand, G., Gaur, R., & Yadav, D. (2021). Plant-microbiome interactions for sustainable agriculture: a review. *Physiology and Molecular Biology of Plants*, 27(1), 165-179.
 17. Han, S. W., & Yoshikuni, Y. (2022). Microbiome engineering for sustainable agriculture: using synthetic biology to enhance nitrogen metabolism in plant-associated microbes. *Current opinion in microbiology*, 68, 102172.
 18. Jiao, H., Xu, W., Chen, W., Hu, Y., Tian, R., & Wang, Z. (2022). Complete genome sequence data of *Bacillus altitudinis* LZP02, a bacterium from the rice rhizosphere, for studying the promotion of plant growth. *Molecular Plant-Microbe Interactions*, 35(5), 428-431.
 19. Ke, J., Wang, B., & Yoshikuni, Y. (2021). Microbiome engineering: synthetic biology of plant-associated microbiomes in sustainable agriculture. *Trends in Biotechnology*, 39(3), 244-261.
 20. Khalifa, A., & Alsowayeh, N. (2023). Whole-genome sequence insight into the plant-growth-promoting bacterium *Priestia filamentosa* strain AZC66 obtained from *Zygophyllum coccineum* rhizosphere. *Plants*, 12(10), 1944.

21. Kumawat, K. C., Nagpal, S., & Sharma, P. (2022). Potential of plant growth-promoting rhizobacteria-plant interactions in mitigating salt stress for sustainable agriculture: A review. *Pedosphere*, 32(2), 223-245.
22. Leontidou, K., Genitsaris, S., Papadopoulou, A., Kamou, N., Bosmali, I., Matsi, T., ... & Mellidou, I. (2020). Plant growth promoting rhizobacteria isolated from halophytes and drought-tolerant plants: genomic characterisation and exploration of phyto-beneficial traits. *Scientific reports*, 10(1), 14857.
23. Luo, Y., Zhou, M., Zhao, Q., Wang, F., Gao, J., Sheng, H., & An, L. (2020). Complete genome sequence of *Sphingomonas* sp. Cra20, a drought resistant and plant growth promoting rhizobacteria. *Genomics*, 112(5), 3648-3657.
24. Mishra, P., Mishra, J., & Arora, N. K. (2021). Plant growth promoting bacteria for combating salinity stress in plants—Recent developments and prospects: A review. *Microbiological Research*, 252, 126861.
25. Mukhia, S., Kumar, A., Kumari, P., & Kumar, R. (2022). Psychrotrophic plant beneficial bacteria from the glacial ecosystem of Sikkim Himalaya: genomic evidence for the cold adaptation and plant growth promotion. *Microbiological research*, 260, 127049.
26. Nanjani, S., Soni, R., Paul, D., & Keharia, H. (2022). Genome analysis uncovers the prolific antagonistic and plant growth-promoting potential of endophyte *Bacillus velezensis* K1. *Gene*, 836, 146671.
27. Nugent, C. (2018). Genome information for sequenced organisms [Data set]. Kaggle. *Genome Information for Sequenced Organisms – Kaggle*.