DNA Research, 2025, **32**, dsaf032
https://doi.org/10.1093/dnares/dsaf032
Advance access publication 6 November 2025
Research Article

OXFORD

Research Article

Strain-level dissection of complex rhizoplane and soil bacterial communities using single-cell genomics and metagenomics

Masako Kifushi[1], Yohei Nishikawa[2,3], Masahito Hosokawa[1,3], Toyoaki Anai[4], and Haruko Takeyama[1,3,*]

[1]Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan
[2]Biomanufacturing and Process Research Center (BPRC), National Institute of Advanced Industrial Science and Technology, Tsukuba 305-8566, Japan
[3]Research Organization for Nano and Life Innovation, Waseda University, 513 Wasedatsurumaki-cho, Shinjuku-ku, Tokyo 162-0041, Japan
[4]Faculty of Agriculture, Kyushu University, 744 Motooka, Nishi-ku, Fukuoka 819-0395, Japan

[*]Corresponding author. Graduate School of Advanced Science and Engineering, Waseda University, 2-2 Wakamatsu-cho, Shinjuku-ku, Tokyo 162-8480, Japan (Email: haruko-takeyama@waseda.jp).

Abstract

Root exudates shape root-associated microbial communities that differ from those in soil. Notably, specific microorganisms colonize the root surface (rhizoplane) and strongly associate with plants. Although retrieving microbial genomes from soil and root-associated environments remains challenging, single amplified genomes (SAGs) and metagenome-assembled genomes (MAGs) are essential for studying these microbiomes. This study compared SAGs and MAGs constructed from short-read metagenomes of the same soil samples to clarify their advantages and limitations in soil and root-associated microbiomes, and to deepen insights into microbial dynamics in rhizoplane. We demonstrated that SAGs are better suited than MAGs for expanding the microbial tree of life in soil and rhizoplane environments, due to their greater gene content, broader taxonomic coverage, and higher sequence resolution of quality genomes. Metagenomic analysis provided sufficient coverage in the rhizoplane but was limited in soil. Additionally, integrating SAGs with metagenomic reads enabled strain-level analysis of microbial dynamics in the rhizoplane. Furthermore, SAGs provided insights into plasmid-host associations and dynamics, which MAGs failed to capture. Our study highlights the effectiveness of single-cell genomics in expanding microbial genome catalogues in soil and rhizosphere environments. Integrating high-resolution SAGs with comprehensive rhizoplane metagenomes offers a robust approach to elucidating microbial dynamics around plant roots.

Keywords: metagenome-assembled genomes; rhizosphere microbiome; single-cell genomics; soil microbiome; soybean.

1. Introduction

Soil is one of the most diverse microbial environments, with 10^8 to 10^{10} microorganisms and an estimated diversity of 4,000 to 50,000 species per gram.[1,2] Especially, the rhizosphere is a unique environment where plants and microorganisms interact. Plants release up to 40% of their photosynthetic products into the rhizosphere,[3,4] promoting microbial activity and forming even denser microbial communities (10^{10} to 10^{12} microorganisms per gram[1]) than bulk soil (here: soil without direct root effects). This high microbial density increases the probability of physical cell-to-cell contact, facilitating horizontal gene transfer.[5,6] Additionally, root exudates contain plant species-specific metabolites, and their attractive and repellent effects reduce microbial diversity around the root compared to bulk soil.[4,7] This regulation occurs in a highly selective manner; for example, soybeans recruit their symbiotic *Bradyrhizobium* selectively at the species level.[8] Notably, some rhizosphere microorganisms colonize the root surface, forming the rhizoplane microbiome that interacts closely with the plants.[9,10] Thus, root-associated microbiomes exhibit distinct characteristics from those in bulk soil.

Due to the abundance of unculturable microorganisms, genome recovery is the key for understanding microbial dark matter in the rhizosphere.[11] The development of metagenome-assembled genomes (MAGs) and single amplified genomes (SAGs) has enabled the cultivation-independent recovery of microbial genomes, significantly expanding our understanding of environmental microbiomes.[12–14] However, obtaining high-qualified microbial genomes from soil and the rhizosphere remains challenging for both approaches.[14–17] MAGs are reconstructed by assembling metagenomic reads and binning contigs inferred to originate from the same species. Recently, a catalogue of over 40,000 soil microbial MAG was generated from public soil metagenomes.[15] Although

Received 5 August 2025; revised 27 October 2025; accepted 30 October 2025

metagenomics recovers genomes in proportion to their abundance in the environment, MAGs have some limitations, including the inability to bin highly conserved sequences, such as 16S rRNA genes, and plasmids, and the collapse of intraspecies diversity into population genomes.[18–22] Furthermore, soil-specific challenges include the need for extensive sequencing due to the high species diversity and the computational demands for data processing.[23,24] Additionally, extracellular DNA, comprising up to 40% of metagenomic reads, complicates analysis.[25,26] Although these challenges have limited the number of soil metagenomic studies,[11,27] recovering microbial genomes from the rhizosphere and soil is essential for understanding plant–microorganisms interactions. SAGs are constructed by physically isolating individual microbial cells from a suspension using FACS or microdroplets, followed by cell lysis, whole genome amplification, and independent assembly of sequenced reads. Theoretically, SAGs enable the reconstruction of individual genomes; however, they typically exhibit lower genome completeness than MAGs.[11,28,29] In soil, FACS-based single-cell genomics yielded <10% amplification efficiency.[28,30] Recent advancements utilizing gel beads have improved sorting efficiency,[12,31–33] facilitating applications in soil, and rhizosphere microbiomes.[12,34,35] Given that MAGs and SAGs undergo distinct experimental and computational workflows and encounter unique challenges in soil and rhizosphere environments, a comparison of their respective advantages and limitations is necessary.

In this study, we evaluate the advantages and limitations of MAGs and SAGs in reconstructing draft genomes of soil and root-associated microbiomes. Using samples collected from the same soybean field, we compared the performance of metagenomics and single-cell genomics in terms of genome recovery, microbiome coverage, and taxonomic collection. We also assessed the data processing challenges associated with each approach. Furthermore, by integrating SAGs and metagenomic reads, we resolved strain-level abundance dynamics in the rhizoplane. Additionally, as part of our analysis of horizontal gene transfer in the root-associated environment, we investigated the plasmid dynamics of *Bradyrhizobium*, a typical soybean symbiont.

2. Materials and methods

2.1. Soybean cultivation and sampling

Soybeans were grown at the experimental yard of Saga University, Japan (33°14′34″N, 130°17′24″E) in 2021. Eight plots (20 × 4 m) were established for the experiment. In 7 plots, winter cover crops, hairy vetch (Visia villosa Roth subsp. Villosa), onion (Allium cepa L.), barley (Hordeum vulgare), mustard (Brassica rapa), wheat (Triticum aestivum), oat (Avena sativa), and rye (Secale cereale), were sown in December 2020. The remaining plot (control plot) was maintained without applying fertilizer. Cover crops were allowed to grow until May 2021. In May 2021, cover crops were incorporated into the soil by rotary cultivation to a depth of ~30 cm. Control plots were also ploughed but without any additional plant materials. In the summer, soybean seeds (Glycine max (L.) Merr., cv. Fukuyutaka), were sown in July 2021. The soil fraction was collected between soybean plant rows at approximately 20 cm away from soybean plants and 20 cm depth. Three soybean plants, except the side row, were randomly selected and excavated with roots and the surrounding soil at each plot. Plants and soils were transported by refrigerated shipping to Waseda University on August 23rd (nodulation (V5) stage), September 6th (full bloom (R2) stage), and October 13th (full seed (R6) stage) in 2021.

2.2. Soil sample collection

Three types of soil samples were collected: the soil, rhizosphere, and rhizoplane. Soil fractionation was performed according to a previously described protocol.[34,36–38] A total of 216 samples were collected and split each for DNA extraction and single-cell genome analysis. Samples for DNA extraction were stored at −30 °C and subsequently processed using an Extrap Soil DNA Kit Plus ver.2 (BioDynamics Laboratory Inc., Tokyo, Japan). 16S rRNA gene sequencing was conducted on all 216 samples, and shotgun metagenomics on 52 samples (Table S1). Single-cell genome analysis was performed on 7 samples (Table S1). Three replicates per sample were combined into 1.5 mL tubes and stored at 4 °C.

2.3. 16S rRNA gene sequencing and bioinformatic analysis

The V3–V4 hypervariable regions of 16S rRNA genes were analysed according to the Illumina protocol for 16S Metagenomic Sequencing Library Preparation with 341F and 806R primers. Barcoded amplicons were sequenced using the Illumina MiSeq 2 × 300 bp platform with the MiSeq Reagent Kit v3 (Illumina Co.), according to the manufacturer's instructions. The construction of operational taxonomic units (OTUs) for 16S rRNA gene analysis followed a previously described protocol.[34] Downstream analyses were performed using phyloseq[39] and nyankomicro packages (https://github.com/xvtyzn/nyankomicro). Permutational multivariate analysis of variance (PERMANOVA) was performed using the vegan package. Differential OTU abundances were analysed using the R package ALDEX2,[40] which infers abundance from counts using a Dirichlet multinomial model.

2.4. Shotgun metagenome sequencing

Three replicates from each of the 52 samples were pooled in equal DNA amounts. Metagenomic sequencing libraries were constructed with 10 µL (1/5 volume) reactions using the QIAseq FX DNA Library Kit (QIAGEN), and sequenced using the DNBSEQ-G400 2 × 150 bp configuration (MGITech Co., Ltd., Beijing, China) using the MGIEasy Universal Library Conversion Kit. Adapter sequences and low-quality reads were eliminated from raw sequence reads using BBDuk v38.90[41] (options: qtrim = r trimq = 10 minlength = 40 maxns = 1 minavgquality = 15 ktrim = r ref = adapters k = 23 mink = 11 hdist = 1 tpe tbo). The quality-controlled reads were analysed using Nonpareil v3.401[42] with default parameters to estimate average community coverage and the diversity of metagenomes. Metagenome reads were individually assembled using SPAdes with the following options (–meta).

MAGs were constructed using metaWRAP v1.3[43] with 3 binning tools, including CONCOCT,[44] MaxBin2,[45] and MetaBAT2,[46] with default options, and the metaWRAP-Bin_refinement module was used to refine the binning results. For the construction of MAGs, all metagenomic reads from different plots were mapped to the contigs to acquire multi-coverage information for binning. This approach was

employed because the use of multicoverage information from multiple samples is highly recommended for generating high-quality MAGs.[44,47,48] Furthermore, to obtain MAGs with higher genome completeness, we selected the MAG with the highest genome completeness as the representative genome for all highly similar genomes derived from the same plot.

2.5. Single-cell genome sequencing with SAG-gel

SAGs were obtained by the SAG-gel method[47] using the bit-MAP platform (bitBiome, Inc., Japan). Frozen soil was suspended in Dulbecco's phosphate-buffered saline (DPBS; Thermo Fisher Scientific, MA, United States), and microbial fractions were obtained by density gradient centrifugation using 80% (w/v) Nycodenz (Serumwerk Bernburg AG, Bernburg, Germany) at 13,000 × *g* for 40 min at 4 °C. The microbial fraction was collected into a 1.5 mL tube and centrifuged at 13,000 × *g* for 15 min at 4 °C. This was followed by 3 wash steps, each involving centrifugation at 8,000 × *g* for 3 min. The concentration of live cells was determined using LIVE/DEAD BacLight bacterial viability assay (Thermo Fisher Scientific), and the cells were then suspended in DPBS with 1.5% low-gelling-temperature agarose (Sigma-Aldrich, MO, United States) at 0.3 to 1.5 particles/droplet. DNA-amplified gel beads were isolated using a FACS melody (BD Biosciences, Franklin) on two 384-well plates for soil samples and one 384-well plate for rhizoplane samples. SAG libraries were prepared using the QIAseq FX DNA Library Kit (QIAGEN). The ligation adaptors were modified using TruSeq Compatible Full-length Adapters UDI (Integrated DNA Technologies, Inc., IA, United States). Each SAG library was sequenced with the DNBSEQ-G400 2 × 150 bp configuration (MGITech Co., Ltd., Beijing, China) using the MGIEasy Universal Library Conversion Kit. The quality control of reads was performed using the same procedure as in metagenome analysis. Quality-controlled reads were assembled de novo into contigs using SPAdes v3.15.2[49] (options for SAG: -sc -careful -disable-rr -disable-gzip-output -t 4 -m 128). Contigs shorter than 1,000 bp were excluded from the SAG assemblies.

2.6. Sequence analysis of MAG and SAG

2.6.1. Quality evaluation. The following procedures were applied to both MAGs and SAGs. Assembled contigs were qualified using QUAST 4.5[50] (options: –no-plots -m 200). The completeness and contamination were evaluated using lineage_wf in CheckM 1.1.3,[51] and genome quality was classified following the Genomic Standards Consortium's criteria,[52] CDS, rRNAs, and tRNAs were extracted using Prokka 1.14.5[53] (options: –rawproduct –mincontiglen 1,000 –cpus 4 –force –compliant –addgenes –rnammer). Taxonomy identification was performed using classify_wf in GTDB-Tk 2.2.3[54] with the default option, using the Release207 database. The unrooted tree generated by GTDB-Tk infer was rooted by midpoint rooting using Gotree v0.4.4[55] and visualized using iTOL 6.5.8.[56] Euler diagrams were drawn using VennDiagram package.[57] Average nucleotide identity (ANI) was calculated using fastANI v1.32.[58] For the comparison of the strain number, all MAGs and SAGs were dereplicated at 99% ANI value using dRep[59] with the options '-comp 50-con 10' at each sample. All the visualization was conducted using the ggplot2 package[60] and ggsci package[61] in R Studio.[62]

2.6.2. Comparison of microbiome coverage. The overlap between metagenomic reads, MAGs and SAGs was represented by the proportion of reads mapped to above Medium-quality MAGs and SAGs using MINIMAP2 v2.17.[63] The overlap between MAGs and SAGs was calculated by aligning SAG contigs to MAG contigs using MINIMAP2 v2.17.

The proportion of MAGs is calculated as the mapping rate of metagenomic reads to MAGs.

The overlap between the metagenome and SAGs is defined as the mapping rate of metagenomic reads to the SAGs.

The proportion of SAGs is calculated by dividing the mapping rate of metagenomic reads to the SAGs by the fraction of the total SAG sequence length that is covered by metagenomic reads; that is,

$$\text{Proportion of SAGs} = (\text{Mapping rate to SAGs}) / (\text{Mapped SAG length} / \text{Total SAG length}).$$

The overlap between MAGs and SAGs is calculated by multiplying the fraction of the MAG sequence length covered by SAG contigs by the mapping rate of metagenomic reads to the MAGs; that is,

$$\text{Overlap} = (\text{Mapped MAG length by SAG contigs} / \text{Total MAG length}) \times (\text{Mapping rate to MAGs})$$

For visualization, the average values of the rhizoplane and soil were used (Table S5 and Fig. S2).

2.6.3. Estimation of strain-level abundance. We removed strain-level redundancy for SAGs and MAGs in 3 rhizoplane samples. According to a previously described protocol,[34,64,65] SAGs above medium-quality were composited. MAGs were dereplicated at 99% ANI value using dRep[59] with the options '-comp 50-con 10'. To determine the relative abundances of MAGs and SAGs, the coverage and read per kilobase per million reads of each genome were calculated based on the mapping of short reads from all samples to MAGs and SAGs using CoverM v0.6.1[66] with parameters '–min-read-percent-identity 97 -p minimap2-sr –min-covered-fraction 2'.

2.6.4. Identification of plasmid. Metagenome assemblies, MAGs, and SAGs above medium-quality were used for mobilome analysis. Plasmids were predicted using Platon (version 1.6)[67] with default parameters. The list was filtered with '#Plasmid Hits' = 1 (True). The closely related plasmids were identified using blastn v2.9.0+[68] against a plasmid database downloaded from National Centre for Biotechnology Information (n = 59,206, downloaded on 2024-10-10). The sequence variants were visualized by IGV 2.8.13.[69]

3. Results and discussion

3.1. Microbiome analysis by 16S rRNA gene sequencing

Bacterial species diversity (OTU count) increased in the order of rhizoplane, rhizosphere, and soil. Within the rhizoplane, diversity was relatively higher in the early growth stages and declined as plant growth progressed (Table S5 and Fig. S1), consistent with previous studies.[34,37,70,71] In all rhizoplane samples, *Proteobacteria* accounted for over 50%, and together with *Bacteroidota*, comprised 70% to 80% of the

community. At the genus level, *Bradyrhizobium* exhibited a sharp increase during the flowering and ripening stages. In contrast, *Proteobacteria* and *Acidobacteriota* were dominant in the soil, each contributing ~50%, but no single genus showed marked dominance (Fig. S2a,b). Additionally, species-level identification remained challenging with 16S rRNA gene sequencing (Fig. S2c).

The rhizoplane-enriched bacterial OTUs were detected at each growth stage by differential abundance analysis. A total of 154 OTUs were enriched in the rhizoplane compared to the rhizosphere (90 OTUs at the nodulation stage, 97 OTUs at the flowering stage, and 62 OTUs at the ripening stage). The trend that the number of rhizoplane-enriched bacteria is the highest at the flowering stage and the lowest at the ripening stage is consistent with our previous study. These bacteria most frequently belonged to *Burkholderiales*, which was also consistent with the previous study.[34]

PERMANOVA demonstrated that bacterial community structures differed most significantly among the soil fractions ($R2 = 0.73$, $P < 0.0001$). The effects of soybean growth stage and cropping history on the bacterial community varied among soil fractions. In soil fractions close to the root, such as the rhizoplane and rhizosphere, the soybean growth stage had a greater impact. Conversely, in the bulk soil, cropping history had a greater effect (Table S2). From these results, it was shown that in the rhizosphere and rhizoplane, which are the focus of this study, the bacterial community is minimally influenced by differences in cropping history.

3.2. Comparison of MAGs and SAGs obtained from soil microbiomes

In this study, we followed previous research and obtained the amount of sequencing reads recommended for soil metagenomic analysis (see Supplementary Materials).[15] A total of 52 soil, rhizosphere, and rhizoplane samples yielded 1,016 Gbp of quality-controlled metagenomic reads. The number of reads ranged from 69 to 135 million for rhizoplane samples, 121 to 135 million for rhizosphere samples, and up to 200 million for bulk soil samples. Metagenome reads were individually assembled and binned using coverage across all metagenomes. To ensure a fair comparison of SAG and MAG, we used metagenomes from all plots with different cropping histories to improve MAG quality.[44,47,48] Although metagenomes were collected from 8 plots, we focused on the 7 samples from 3 plots shared with the single-cell genome analysis (Table S1, 169 Gbp of quality-controlled reads), 3 high-quality (rhizoplane: 2 MAGs, soil: 1 MAGs), 257 medium-quality (rhizoplane: 99 MAGs, soil: 158 MAGs), and 1,093 low-quality genomes were obtained (Fig. 1a; Table S2). The average completeness of medium- to high-quality MAGs was 75% for soil and 77% for rhizoplane samples, with contamination levels of 2.9% and 2.5%, respectively (Fig. 1b)—comparable to previously reported soil MAGs with an average completeness of 76.1% and contamination of 3.6%.[30,70]

SAGs were obtained from 7 rhizoplane and soil samples using the SAG-gel method, generating 4,224 amplified DNA samples and 606 Gbp of quality-controlled reads. Of the 4,006 SAGs with contigs ≥1,000 bp, 3,335 passed quality filtering based on CheckM. We recovered 38 high-quality (rhizoplane: 9 SAGs, soil: 29 SAGs), 993 medium-quality (rhizoplane: 259 SAGs, soil: 734 SAGs), and 2,304 low-quality genomes (Fig. 1a; Table S2). The average completeness was 67% for soil SAGs and 65% for rhizoplane SAGs, while the average contamination was 2.0% and 2.1%, respectively. Compared to MAGs, SAGs exhibited slightly lower completeness but also lower contamination (Fig. 1b). In this study, we obtained SAGs from rhizoplane samples, which were limited in quantity (~1 g). The proportion of medium- to high-quality genomes in the analysed sequence data ranged 16%–33%, indicating efficiency comparable to or higher than previous soil SAG-gel studies.[30,72] Notably, FACS-based soil single-cell genomics has reported recovery rates of just 0.41%–0.67% for medium- to high-quality SAGs,[71] while the SAG-gel approach used here achieved higher recovery efficiency regardless of soil type.

The variation in recovered SAG numbers across samples likely reflects differences in genome completeness and contamination (Fig. 1a). Fluorescence-positive rates of gel beads following multiple displacement amplification (MDA) were lower than theoretical expectations (Table S3), suggesting incomplete cell lysis or amplification. Additional biases introduced by MDA and DNA fragmentation may have further reduced genome completeness.[73] On average, 63% ± 17% of SAGs were classified as low-quality. Incomplete lysis or genome amplification during MDA and cell encapsulation exceeding 1.5 cells/droplet were associated with increased contamination rates, although these were consistent with theoretical expectations. Despite potential challenges introduced by biofilm formation on the rhizoplane, appropriate encapsulation control enabled recovery of a high proportion of medium- to high-quality SAGs from both soil and rhizoplane samples. This underscores the importance of controlling cell encapsulation rates to improve SAG recovery efficiency.

From all 52 metagenomes, 5,862 MAGs (24 high-quality, 1,496 medium-quality, and 4,342 low-quality genomes) were constructed using 1.6 times the sequencing data compared to single-cell genome analysis. However, the 7 SAG samples yielded a higher number of high-quality genomes, highlighting the greater efficiency of the SAG-gel method for recovering high-quality microbial genomes from both soil and rhizoplane environments.

We compared the genomic statistics of medium- to high-quality SAGs and MAGs (Fig. 1b; Table S4). MAGs had more contigs and shorter N50, reflecting fragmented genomes. The average GC content was 52.1% for soil SAGs and 55.3% for rhizoplane SAGs, compared to 59.5% for soil MAGs and 62.5% for rhizoplane MAGs. Considering that the average GC content of soil bacteria is 61%[74] and that SAGs and MAGs were constructed from the same samples, the GC content of SAGs is likely biased towards lower values. The average GC content of quality-controlled reads was 62.5% for metagenome reads and 52.0% for single-cell genome reads. This is likely due to the reduced amplification efficiency of high-GC regions by the phi29 polymerase employed in MDA.[73] The number of protein-coding sequences (CDSs) and types of tRNA genes were higher in SAGs. A significant difference between SAGs and MAGs was observed in the recovery of rRNA genes. The recovery rate of full-set rRNA genes was 69% for soil SAGs and 78% for rhizoplane SAGs, whereas only 1.9% of soil MAGs and 2.0% of rhizoplane MAGs contained rRNA genes. This difference is likely due to the high conservation of rRNA sequences, which complicates the binning process.[75]

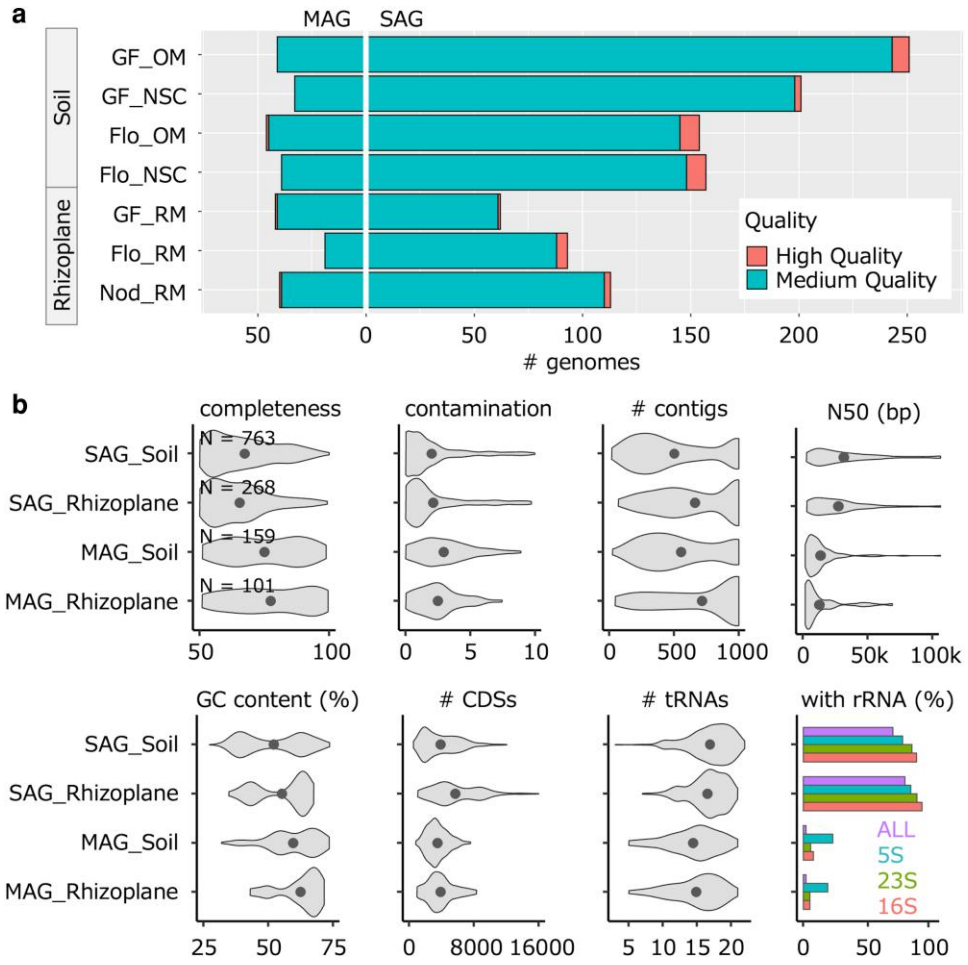

Fig. 1. Acquisition of soil microbial SAGs and MAGs. a) The number of medium- to high-quality genomes obtained through metagenome and single-cell genome analyses. The number of medium- to high-quality genomes is displayed for each sample and coloured by quality. Left: MAG; Right: SAG. Nod: nodulation stage; Flo: flowering stage; GF: ripening stage; OM: field with a winter cover crop barley; NSC: field with no winter cover crop; RM: field with a winter cover crop rye. b) Assembly Statistics for Both SAGs and MAGs. Genome assembly statistics for medium- to high-quality genomes (763 soil SAGs, 268 rhizoplane SAGs, 159 soil MAGs, and 101 rhizoplane MAGs). Grey dots indicate the average values. Metrics for medium- to high-quality genomes include completeness, contamination, contig count, N50, GC content, CDS count, tRNA repertoire, and rRNAs.

3.3. Limitations in soil microbial SAGs and MAGs

We assessed the comprehensiveness of each microbiome by evaluating metagenomic read coverage. Nonpareil curves estimated the average metagenome coverage of 74% ± 11% for the rhizoplane, 49% ± 5.5% for the rhizosphere, and 41% ± 3.4% for bulk soil (Fig. S4 and Table S5). Considering metagenome coverage above 60% is recommended for reliable analysis,[75] rhizosphere and soil coverage in this study was insufficient despite obtaining a sequencing depth comparable to that of previous studies.[74] In contrast, reduced species diversity in the rhizoplane (see Supplementary materials), allowed the metagenome to achieve an average coverage exceeding 60%. These findings highlight the need to consider that the coverage of soil metagenomes varies with sample complexity and microbial diversity.[76]

In general, reconstructing microbial genomes from soil is considered computationally challenging. Therefore, we evaluated information loss during the metagenomic workflow (Fig. S4b). Read mapping rate, which reflects the proportion of reads utilized for contig assembly, serves as an indicator of assembly difficulty; lower rates suggest greater challenges. The mapping rates were 56% ± 17% for the rhizoplane, 20% ± 7.5% for the rhizosphere, and 29% ± 1.5% for the bulk soil. In previous studies, mapping rates ranged from 0.17% to 27%,[16,75] and even with a terabase-scale metagenome, the mapping rate reached only 51.2%.[16,77] High microbial diversity is recognized as a major factor complicating soil metagenome analysis,[78] further supported by the increased computational difficulty in our study (Table S5 and Fig. S1). In contrast, single-cell genome analysis exhibited stable mapping rates to SAGs (47% ± 3.3%) across sample types, unaffected by diversity (Table S2). Next, the proportion of contig length binned into MAGs was 54% ± 4.9% in the rhizoplane, 39% ± 5.4% in the rhizosphere, and 16% ± 2.7% in the bulk soil, respectively. This demonstrates that metagenome coverage in soil is often insufficient, resulting in fragmented genome sequences.

Information content comparisons (Fig. S5) revealed that, with metagenomic reads set at 100, SAGs accounted for 20 in the rhizoplane and 11 in the soil; MAGs accounted for 21 and 4, respectively. Thus, conversion from metagenomic read to MAGs resulted in a 5-fold reduction in information content in the rhizoplane and a 25-fold reduction in the soil. In the rhizoplane, 75% of SAG sequences overlapped with metagenomic reads, whereas in the soil, <20% overlapped, suggesting that metagenome coverage was insufficient to

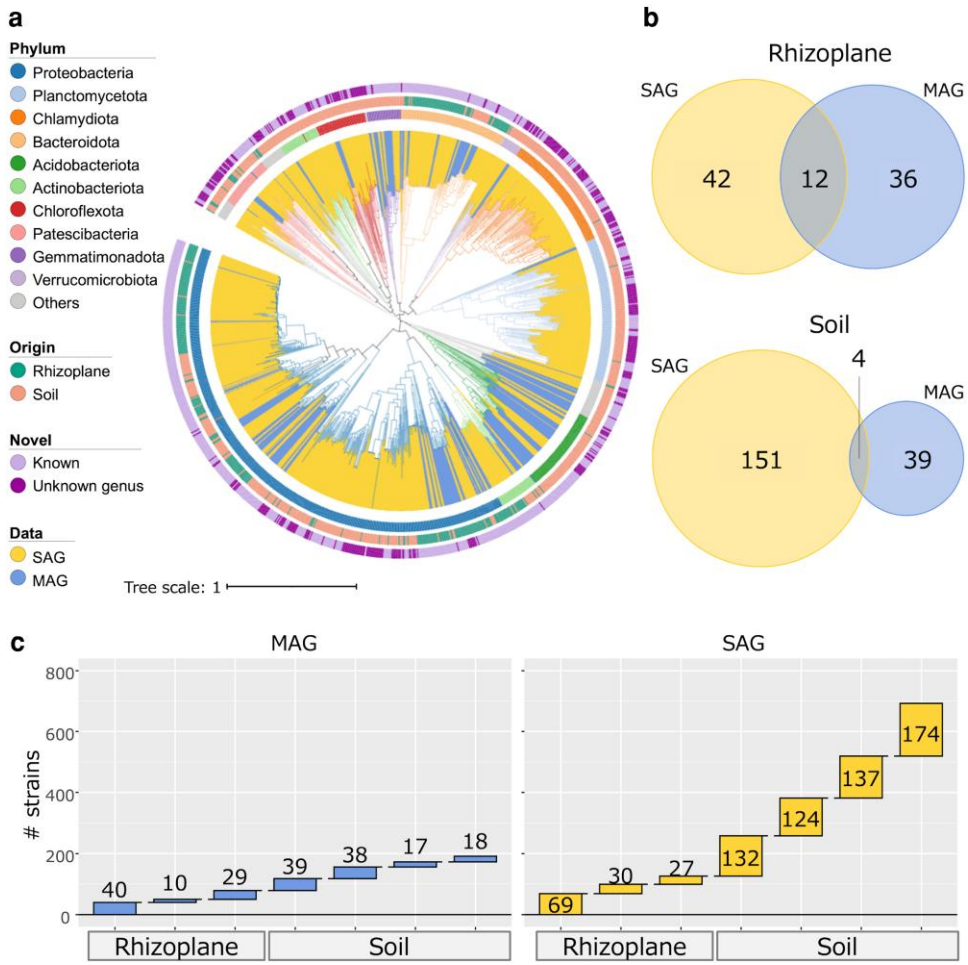

Fig. 2. Phylogenic collection of MAGs and SAGs collected from the soil and the rhizoplane. a) The GTDB-Tk phylogenetic tree of medium- to high-quality bacterial genomes of 973 SAGs and 254 MAGs. Phylogenic tree edges are coloured by taxonomy at the phylum level. The background colours of clades represent SAGs (yellow) or MAGs (blue). Inner circle, phylum level taxonomy; middle circle, genome origin part of plants; outer circle, a taxonomic novelty at the genus level. b) Venn diagrams visualizing the number of genera collected by SAGs (yellow) and MAGs (blue). c) The increase in the cumulative number of strains collected as additional samples were incorporated. Rhizoplane samples were arranged in the order of 3 growth stages. Soil samples were arranged as follows: Flo from NSC, Flo from OM, GF from NSC, and GF from OM.

detect certain microorganisms. The information content of SAGs was 1.3 times greater than that of MAGs in the rhizoplane and 2.8 times greater in the soil. Furthermore, the overlap between SAGs and MAGs in the soil was <1%. Our method of single-cell isolation by density gradient centrifugation may have excluded bacteria bound to soil particles, while MAGs can include extracellular DNA fragments. We hypothesize that these methodological differences caused the limited overlap between SAGs and MAGs. Collectively, in this study, the information richness followed the hierarchy: metagenomic reads > SAGs > MAGs. We also discussed these results in the context of other comparative studies between MAG and SAG, extending beyond the soil environment (see Supplementary materials).

3.4. Comparing taxonomic variety captured by MAGs and SAGs

Taxonomic annotation based on GTDB resulted in 973 bacterial SAGs, 58 archaeal SAGs, 254 bacterial MAGs, and 6 archaeal MAGs. In both the rhizoplane and bulk soil, major taxa detected by 16S rRNA gene sequencing were also recovered in SAGs and MAGs (see Supplementary materials). However, the phylogenetic tree revealed that SAGs covered a broader range of bacterial lineages compared to MAGs, which were more phylogenetically biased (Fig. 2a). In rhizoplane samples, SAGs encompassed 11 phyla, while MAGs represented 5 phyla. In both approaches, *Proteobacteria* was predominant, followed by *Bacteroidota*, consistent with the taxonomic composition by 16S rRNA gene sequencing (Fig. S2). In the bulk soil, SAGs captured 22 phyla, while MAGs covered 17 phyla. Single-cell genome analysis retrieved over 100 SAGs from *Proteobacteria*, *Planctomycetota*, and *Chlamydiota*, whereas MAGs were primarily reconstructed from *Acidobacteriota*. *Chlamydiota*, one of the 9 phyla uniquely captured by SAGs has a small genome (~1.5 Mbp) and ~1% relative abundance in soil microbiomes ('Others' in Fig. S2).[25,26,79] Its low representation in metagenomic reads likely made MAG construction challenging.

Regarding novel taxa, SAGs were particularly effective in recovering genomes from unannotated genera (Fig. 2a). In bulk soil, 43% of SAGs (331/705) were unannotated at the genus level, compared to only 18% (27/153) of MAGs. In the rhizoplane, the proportions were 8% (21/268) for SAGs and 14% (14/101) for MAGs. These results indicate that SAGs—especially from bulk soil—are more efficient at capturing

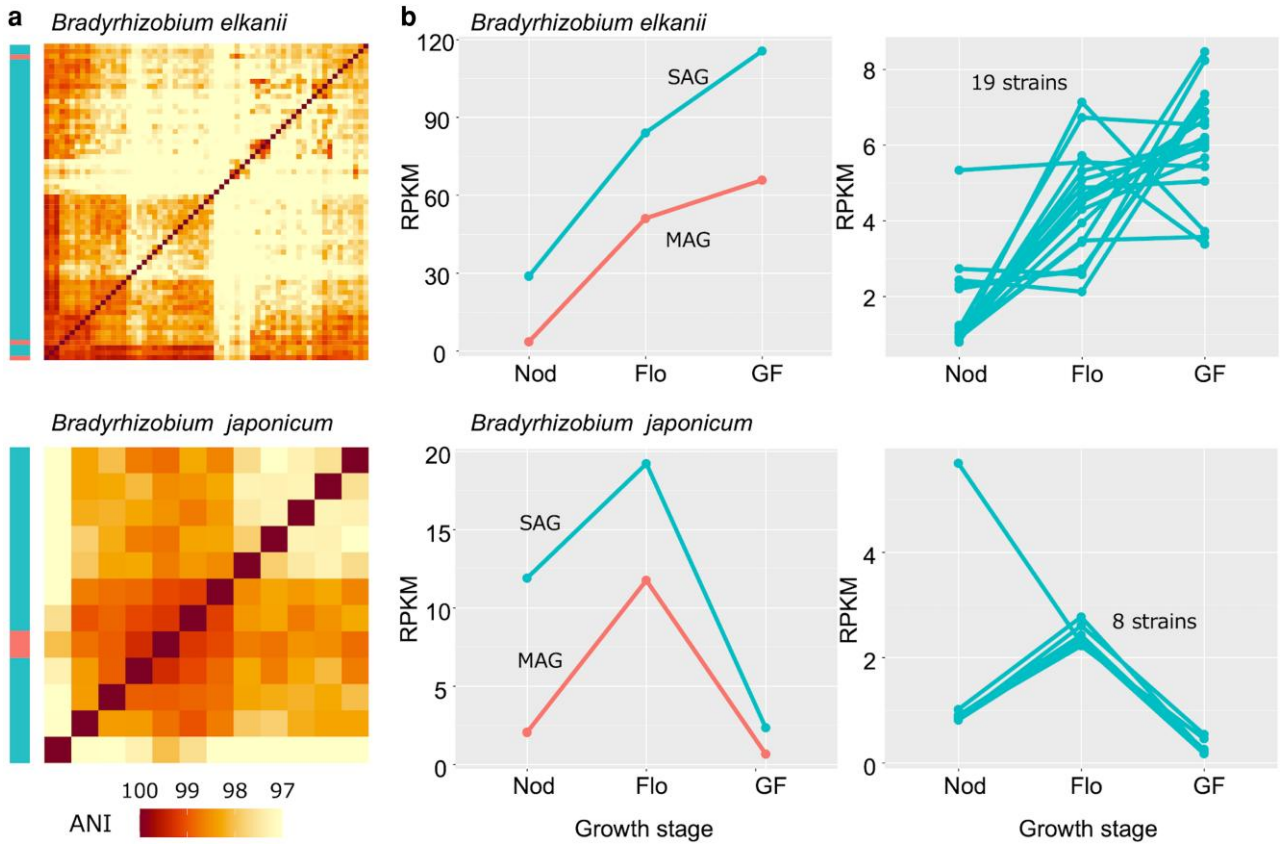

Fig. 3. Comparison of strain-resolved characteristics between MAGs and SAGs. a) ANI heatmap across SAGs (blue) and MAGs (red). b) Abundance transition along with soybean growth stages of total abundance for SAGs (blue) and MAGs (red) (left) and strain-level abundance of SAGs (right). Upper panel: *Bradyrhizobium elkanii*; lower panel: *Bradyrhizobium japonicum*.

novel taxonomic diversity. The high species diversity in soil likely explains this observation, and efficient recovery of novel taxa from soil highlights SAGs' superiority. At the genus level, both approaches yielded comparable results in the rhizoplane (Fig. 2b, SAGs: 54 genera; MAGs: 48 genera), with 12 shared genera. SAGs predominantly recovered *Bradyrhizobium*, while MAGs were reconstructed *Novosphingobium* (Others in Fig. S2). In soil, SAGs recovered 155 genera compared to only 43 by MAGs, with limited overlap (Fig. 2b). MAGs were heavily influenced by the abundant genera such as *PSRF01* (*RB41* in Fig. S2) and *Sphingomicrobium* (*Sphingomonas* in Fig. S2), likely due to differences in cell selection and DNA extraction methods between approaches. While MAGs can capture extracellular DNA, SAGs isolate live cells, reducing environmental DNA interference but potentially introducing biases in cell recovery.[35,80]

At the strain level, SAGs excelled, recovering 1.6 times more strains than MAGs in rhizoplane samples (Fig. 2c, SAGs: 126 strains from 54 genera; MAGs: 79 strains from 48 genera). Across all 7 samples, SAGs recovered 693 strains from 195 genera, while MAGs recovered 191 strains from 93 genera. These results emphasize the superior strain-level resolution of SAGs.

3.5. Strain-resolved genome analysis to reveal abundance dynamics

After removing strain-level redundancy in 3 rhizoplane samples (see Materials and Methods), SAGs recovered multiple strains from 11 species, while MAGs did so for only 4 species (Table S6). In 5 species represented by multiple strains, 4 showed higher strain counts in SAGs, showing their superior resolution for capturing intraspecies diversity[81] (Fig. 3a; Fig. S6a).

Integrating SAGs with metagenome reads enabled the estimation of abundance dynamics for 5 species in the rhizoplane (Table S6; Fig. 3b). In 3 of them, SAGs and MAGs exhibited similar total abundance trends across soybean growth, indicating that both methods captured species-level dynamics (Fig. 3b, left; Fig. S6b, left). On the other hand, composite SAGs tended to show higher total abundance, indicating their ability to capture strains or genomic regions missed in the MAG construction process. SAGs revealed diverse strain-level abundance patterns. For instance, *Bradyrhizobium elkanii* and *B. japonicum*, key nitrogen-fixing rhizobia for soybean, showed distinct species-level abundance patterns: the former increased towards ripening, while the latter peaked at flowering and declined thereafter, as indicated by MAGs and the total abundance of SAGs (Fig. 3b, left). SAG-based strain-level analysis revealed intraspecies variation, with some *B. elkanii* strains decreasing at ripening and *B. japonicum* strains gradually declining throughout development (Fig. 3b, right). These results show that SAGs resolve strain-level dynamics beyond the genus-level of 16S rRNA gene sequencing (Fig. S2) and the species-level of most MAGs (Fig. 3; Fig. S6).

Furthermore, we focused on *B. japonicum* and compared the gene content between a highly abundant strain in the nodulation stage (S_FY_GF_ES_RM-00020) and 7 others.

Mapping 8 SAG reads to the closely related *B. japonicum* USDA 6 genome revealed that S_FY_GF_ES_RM-00020 retained part of the symbiotic island missing in the others (Fig. S7). This region is rich in insertion sequences (IS) and prone to gene transfer.[82] A transposase was found at the 3′ end of the deleted region. The deleted region encoded Gamma-aminobutyric acid (GABA) transporters, suggesting higher GABA uptake in S_FY_GF_ES_RM-00020 than in other strains. Soybeans produce GABA immediately after germination.[82] Therefore, this strain likely imported GABA efficiently, converted it to succinate, and enhanced energy production via the citric acid cycle, supporting its proliferation during the nodulation stage.

Further, we also considered the spatial dynamics of *Bradyrhizobium* by mapping metagenomes from the rhizosphere and bulk soil to rhizoplane-derived *Bradyrhizobium* SAGs. As a result, the majority of *Bradyrhizobium* increased their abundance in the rhizosphere and rhizoplane compared to the bulk soil. We also revealed that the dynamics of individual strains observed on the rhizoplane tended to show a similar pattern in the rhizosphere (Table S8).

Collectively, these findings demonstrate that SAGs not only provide high-resolution genomes from root-associated environments but also enable tracking of strain-level abundance dynamics and functional interpretation of genomic differences, revealing strain-specific responses to plant-driven environmental changes.

3.6. Plasmid detection from SAGs and MAGs

MAGs are constructed using nucleotide composition and abundance patterns; however, the sequence composition of plasmids often differs from that of host genomes, making them difficult to bin correctly.[81,82] In contrast, SAGs recover all DNA from individual cells, including plasmids, enabling more comprehensive genome reconstruction.[83,84] In this study, plasmid contigs were identified in 110 SAGs (11%) and 21 MAGs (8.1%) (Fig. 4). The efficiency of plasmid recovery in MAGs varied widely, ranging from 0% to 37% depending on the sample (Table S7 and Supplementary materials). SAGs recovered plasmids from 5 phyla (*Proteobacteria*, *Actinobacteriota*, *Armatimonadota*, *Patescibacteria*, and *Planctomycetota*) and 19 genera, whereas MAGs recovered plasmids from only 2 phyla and 11 genera. Notably, while MAGs included representatives of *Patescibacteria* and *Planctomycetota*, no plasmids were detected from these lineages. *Armatimonadota* plasmids were exclusively detected in soil-derived SAGs. Most plasmid-containing genomes belonged to *Proteobacteria* (98 SAGs [89%], 20 MAGs [95%]), a phylum well known for plasmid diversity.[67] Detection via Platon,[85] which relies on known plasmid protein databases, may be biased towards well-characterized taxa. Plasmids were more frequently detected in rhizoplane samples than in bulk soil by both approaches, suggesting more active horizontal plasmid transfer near plant roots (see Supplementary Materials).

We further focused on *Bradyrhizobium*, a key soybean symbiont. High-quality SAGs and MAGs were obtained for *B. diazoefficiens*, *B. japonicum*, and *B. elkanii*, allowing comparative plasmid analysis. Plasmids were detected in *B. japonicum* only by SAGs. In *B. diazoefficiens* and *B. elkanii*, MAG-derived plasmid contigs harboured mobile genetic elements (MGEs) such as insertion sequences, transposases, and integrases (Table S8). As MGEs can facilitate gene exchange between plasmids and chromosomes,[86] their sequence features may have enabled their binning despite compositional differences. SAG-derived plasmids commonly encoded genes for conjugative transfer, MGEs, plasmid stability, and replication machinery across all 3 *Bradyrhizobium* species (Table S9). Acyl-homoserine lactone synthases (AHLs), which mediate quorum sensing, were frequently detected on plasmids in SAGs of all species, while MAGs detected them only in *B. diazoefficiens*. Chromosomal AHL synthases were detected by both methods. Since AHL synthases encoded on plasmids often respond to environmental cues such as biofilm formation and nutrient availability,[68] their detection supports the hypothesis that *Bradyrhizobium* species use plasmid-mediated signalling for early stage adaptation in the rhizoplane.

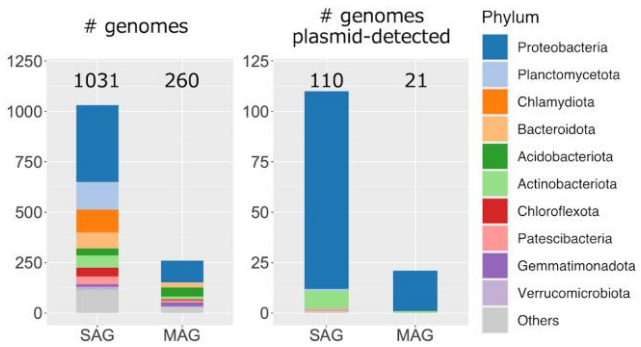

Fig. 4. Plasmid detection profiles from SAGs and MAGs. Stacked bar charts showing the total number of SAGs and MAGs (left) and those with plasmids (right). All charts are coloured by taxonomy at the phylum level.

BLASTn[87] analysis of SAG-derived plasmids revealed a shared top hit (plasmid CP126011.1 from *B. japonicum* N03G-Bj) across all 3 species. Mapping reads revealed conserved regions across species (Fig. S8, blue box) and sequence variation within *B. japonicum* strains (orange box), suggesting active interspecies plasmid transfer.

These results indicate that SAGs effectively capture plasmid-containing genomes from diverse taxa, while MAGs often fail to do so. In *Bradyrhizobium*, plasmids appear to mediate adaptation to the rhizoplane environment and are likely involved in interspecies gene transfer.

4. Conclusion

This study demonstrates the effectiveness of single-cell genomics in expanding the catalogue of soil and root-associated microbial genomes. By comparing SAGs and MAGs derived from the same soybean field samples, we showed that SAGs outperform MAGs in recovering high-quality genomes with greater gene content, broader taxonomic representation—including previously uncharacterized lineages—and more comprehensive plasmid detection. Although metagenomic sequencing achieved higher overall coverage, the conversion efficiency from reads to MAGs was markedly low, especially in the complex bulk soil environment, leading to considerable information loss. In contrast, SAGs retained more genomic information, resolved greater strain-level diversity, and offered superior insights into plasmid-host associations.

Integrating high-resolution SAGs with metagenomic data enabled strain-resolved analyses of microbial dynamics in the rhizoplane and uncovered strain-specific responses to

environmental changes, as exemplified by *Bradyrhizobium* species. The detection of plasmid-encoded genes, including those involved in conjugative transfer and quorum sensing, highlighted the ecological significance of plasmids in shaping root-associated microbiomes.

However, our study also reveals several limitations. Single-cell genomics, while powerful, requires further optimization in terms of encapsulation efficiency, amplification bias, and recovery of biofilm-associated microbes, particularly from the rhizoplane. Metagenomics, although more accessible and increasingly enhanced by long-read sequencing[30,77] and Hi-C technologies,[88] faces challenges related to high sequencing depth requirements, computational complexity,[88] and interference from extracellular DNA, especially in highly diverse bulk soil environments. Furthermore, our analyses were conducted on samples from a single geographic location and crop type, limiting the generalizability of our findings.

Future research should focus on expanding sampling to diverse soils, crops, and environmental conditions to assess the broader applicability of these approaches. Incorporating complementary technologies such as long-read sequencing, Hi-C, and functional assays (eg, plasmid transfer experiments, plant colonization studies) will be essential to further elucidate microbial functions and interactions in complex environments.

In conclusion, this study establishes that single-cell genomics, especially when integrated with metagenomics, is a robust approach for dissecting complex soil and rhizosphere microbial communities. This combined strategy holds promise for uncovering hidden microbial diversity, resolving strain-level dynamics, and advancing our understanding of microbial contributions to plant–microbe interactions.

Acknowledgments

The Human Genome Center at University of Tokyo provided supercomputing resources. This work was supported by Cabinet Office, Government of Japan, Moonshot R&D Program for Agriculture, Forestry and Fisheries (JPJ009237, funding agency: Bio-oriented Technology Research Advancement Institution).

Supplementary material

Supplementary data are available at *DNARES* online.

Funding

This work was supported by Cabinet Office, Government of Japan, Moonshot R&D Program for Agriculture, Forestry and Fisheries (JPJ009237, funding agency: Bio-oriented Technology Research Advancement Institution).

Conflicts of interest

M.H. and H.T. are shareholders in bitBiome, to which the patents pertaining to the SAG-gel workflow were transferred.

Data availability

The raw data produced in this study were deposited at NCBI under BioProject ID PRJNA 1231735.

References

1. McNear DH Jr. The Rhizosphere—Roots, Soil and Everything in Between. Nature Education Knowledge. 2013:4. https://www.nature.com/scitable/knowledge/library/the-rhizosphere-roots-soil-and-67500617/
2. Raynaud X, Nunan N. Spatial ecology of bacteria at the microscale in soil. PLoS One. 2014:9:e87217. https://doi.org/10.1371/journal.pone.0087217
3. Bais HP, Weir TL, Perry LG, Gilroy S, Vivanco JM. The role of root exudatesin rhizosphere interactions with plants and other organisms. Annu Rev Plant Biol. 2006:57:233–266. https://doi.org/10.1146/annurev.arplant.57.032905.105159
4. Berendsen RL, Pieterse CMJ, Bakker PAHM. The rhizosphere microbiome and plant health. Trends Plant Sci. 2012:17:478–486. https://doi.org/10.1016/j.tplants.2012.04.001
5. Maheshwari M, Abulreesh HH, Khan MS, Ahmad I, Pichtel J. Horizontal gene transfer in soil and the rhizosphere: impact on ecological fitness of bacteria. In: Meena VS, Mishra PK, Bisht JK, Pattanayak A, editors. Agriculturally important microbes for sustainable agriculture. Springer Singapore; 2017. p. 111–130.
6. Sánchez-Salazar AM et al. An overview of plasmid transfer in the plant microbiome. Plasmid. 2023:127:102695. https://doi.org/10.1016/j.plasmid.2023.102695
7. Sugiyama A. The soybean rhizosphere: metabolites, microbes, and beyond—a review. J Adv Res. 2019:19:67–73. https://doi.org/10.1016/j.jare.2019.03.005
8. Sarao SK, Boothe V, Das BK, Gonzalez-Hernandez JL, Brözel VS. Bradyrhizobium and the soybean rhizosphere: species level bacterial population dynamics in established soybean fields, rhizosphere and nodules. Plant Soil. 2024:508:515–530. https://doi.org/10.1007/s11104-024-06814-4
9. Zhang Y et al. Huanglongbing impairs the rhizosphere-to-rhizoplane enrichment process of the citrus root-associated microbiome. Microbiome. 2017:5:97. https://doi.org/10.1186/s40168-017-0304-4
10. Edwards J et al. Structure, variation, and assembly of the root-associated microbiomes of rice. Proc Natl Acad Sci U S A. 2015:112:E911–E920. https://doi.org/10.1073/pnas.1414592112
11. Arikawa K, Hosokawa M. Uncultured prokaryotic genomes in the spotlight: an examination of publicly available data from metagenomics and single-cell genomics. Comput Struct Biotechnol J. 2023:21:4508–4518. https://doi.org/10.1016/j.csbj.2023.09.010
12. Nishikawa Y et al. Validation of the application of gel beads-based single-cell genome sequencing platform to soil and seawater. ISME Commun. 2022:2:1–11. https://doi.org/10.1038/s43705-022-00179-4
13. Hosokawa M, Nishikawa Y. Tools for microbial single-cell genomics for obtaining uncultured microbial genomes. Biophys Rev. 2024:16:69–77. https://doi.org/10.1007/s12551-023-01124-y
14. Nayfach S et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2020:39:499–509. https://doi.org/10.1038/s41587-020-0718-6
15. Ma B et al. A genomic catalogue of soil microbiomes boosts mining of biodiversity and genetic resources. Nat Commun. 2023:14:7318. https://doi.org/10.1038/s41467-023-43000-z
16. Lombard N, Prestat E, van Elsas JD, Simonet P. Soil-specific limitations for access and analysis of soil microbial communities by metagenomics: limitations in soil metagenomics. FEMS Microbiol Ecol. 2011:78:31–49. https://doi.org/10.1111/j.1574-6941.2011.01140.x
17. Choi J et al. Strategies to improve reference databases for soil microbiomes. ISME J. 2017:11:829–834. https://doi.org/10.1038/ismej.2016.168
18. Mise K, Iwasaki W. Unexpected absence of ribosomal protein genes from metagenome-assembled genomes. ISME Commun. 2022:2:118. https://doi.org/10.1038/s43705-022-00204-6
19. Maguire F et al. Metagenome-assembled genome binning methods with short reads disproportionately fail for plasmids and genomic

islands. Microb Genom. 2020:6:mgen000436. https://doi.org/10.1099/mgen.0.000436
20. Kawano-Sugaya T et al. A single amplified genome catalog reveals the dynamics of mobilome and resistome in the human microbiome. Microbiome. 2024:12:1–13. https://doi.org/10.1186/s40168-024-01903-z
21. Chen L-X, Anantharaman K, Shaiber A, Eren AM, Banfield JF. Accurate and complete genomes from metagenomes. Genome Res. 2020:30:315–333. https://doi.org/10.1101/gr.258640.119
22. Arikawa K et al. Recovery of strain-resolved genomes from human microbiome through an integration framework of single-cell genomics and metagenomics. Microbiome. 2021:9:202. https://doi.org/10.1186/s40168-021-01152-4
23. Jansson J. Towards "Tera-Terra": terabase sequencing of terrestrial metagenomes: microbial ecologists are taking a metagenomics approach to analyze complex and diverse soil microbial communities. Microbe (Washington, DC). 2011:6:309–315. https://doi.org/10.1128/microbe.6.309.1
24. Riley R et al. Terabase-scale coassembly of a tropical soil microbiome. Microbiol Spectr. 2023:11:e0020023. https://doi.org/10.1128/spectrum.00200-23
25. Carini P et al. Relic DNA is abundant in soil and obscures estimates of soil microbial diversity. Nat Microbiol. 2016:2:16242. https://doi.org/10.1038/nmicrobiol.2016.242
26. Bairoliya S, Koh Zhi Xiang J, and Cao B. Extracellular DNA in environmental samples: occurrence, extraction, quantification, and impact on microbial biodiversity assessment. Appl Environ Microbiol. 2022:88:e0184521. https://doi.org/10.1128/aem.01845-21
27. Richardson L et al. MGnify: the microbiome sequence data analysis resource in 2023. Nucleic Acids Res. 2023:51:D753–D759. https://doi.org/10.1093/nar/gkac1080
28. Rinke C et al. Obtaining genomes from uncultivated environmental microorganisms using FACS-based single-cell genomics. Nat Protoc. 2014:9:1038–1048. https://doi.org/10.1038/nprot.2014.067
29. Alneberg J et al. Genomes from uncultivated prokaryotes: a comparison of metagenome-assembled and single-amplified genomes. Microbiome. 2018:6:173. https://doi.org/10.1186/s40168-018-0550-0
30. Myeong NR et al. Genomic profiling of Antarctic geothermal microbiomes using long-read, Hi-C, and single-cell techniques. Sci Data. 2024:11:1023. https://doi.org/10.1038/s41597-024-03875-z
31. Hosokawa M et al. Droplet-based microfluidics for high-throughput screening of a metagenomic library for isolation of microbial enzymes. Biosens Bioelectron. 2015:67:379–385. https://doi.org/10.1016/j.bios.2014.08.059
32. Nishikawa Y et al. Large-scale single-virus genomics uncovers hidden diversity of river water viruses and diversified gene profiles. ISME J. 2024:18:wrae124. https://doi.org/10.1093/ismejo/wrae124
33. Chijiiwa R et al. Single-cell genomics of uncultured bacteria reveals dietary fiber responders in the mouse gut microbiota. Microbiome. 2020:8:5. https://doi.org/10.1186/s40168-019-0779-2
34. Kifushi M et al. Analysis of microbial dynamics in the soybean root-associated environments from community to single-cell levels. J Biosci Bioeng. 2024:137:429–436. https://doi.org/10.1016/j.jbiosc.2024.02.007
35. Aoki W, Kogawa M, Matsuda S, Matsubara K, Hirata S. Massively parallel single-cell genomics of microbiomes in rice paddies. Front Microbiol. 2022:13:1024640. https://doi.org/10.3389/fmicb.2022.1024640
36. Bulgarelli D et al. Revealing structure and assembly cues for Arabidopsis root-inhabiting bacterial microbiota. Nature. 2012:488:91–95. https://doi.org/10.1038/nature11336
37. Liu Y et al. Rhizobium symbiotic capacity shapes root-associated microbiomes in soybean. Front Microbiol. 2021:12:709012. https://doi.org/10.3389/fmicb.2021.709012
38. Simmons T, Caddell DF, Deng S, Coleman-Derr D. Exploring the root microbiome: extracting bacterial community data from the soil, rhizosphere, and root endosphere. J Vis Exp. 2018:135:e57561. https://doi.org/10.3791/57561
39. McMurdie PJ, Holmes S. Phyloseq: an R package for reproducible interactive analysis and graphics of microbiome census data. PLoS One. 2013:8:e61217. https://doi.org/10.1371/journal.pone.0061217
40. Fernandes AD, Macklaim JM, Linn TG, Reid G, Gloor GB. ANOVA-like differential expression (ALDEx) analysis for mixed population RNA-Seq. PLoS One. 2013:8:e67019. https://doi.org/10.1371/journal.pone.0067019
41. Bushnell, B. BBMap: a fast, accurate, Splice-Aware aligner. 2014. https://escholarship.org.
42. Rodriguez-R LM, Gunturu S, Tiedje JM, Cole JR, Konstantinidis KT. Nonpareil 3: fast estimation of metagenomic coverage and sequence diversity. mSystems. 2018:3:9. https://doi.org/10.1128/mSystems.00039-18
43. Uritskiy GV, DiRuggiero J, Taylor J. MetaWRAP-a flexible pipeline for genome-resolved metagenomic data analysis. Microbiome. 2018:6:158. https://doi.org/10.1186/s40168-018-0541-1
44. Alneberg J et al. Binning metagenomic contigs by coverage and composition. Nat Methods. 2014:11:1144–1146. https://doi.org/10.1038/nmeth.3103
45. Wu Y-W, Simmons BA, Singer SW. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. Bioinformatics. 2016:32:605–607. https://doi.org/10.1093/bioinformatics/btv638
46. Kang DD, Froula J, Egan R, Wang Z. MetaBAT, an efficient tool for accurately reconstructing single genomes from complex microbial communities. PeerJ. 2015:3:e1165. https://doi.org/10.7717/peerj.1165
47. Mattock J, Watson M. A comparison of single-coverage and multi-coverage metagenomic binning reveals extensive hidden contamination. Nat Methods. 2023:20:1170–1173. https://doi.org/10.1038/s41592-023-01934-8
48. Han H, Wang Z, Zhu S. Benchmarking metagenomic binning tools on real datasets across sequencing platforms and binning modes. Nat Commun. 2025:16:2865. https://doi.org/10.1038/s41467-025-57957-6
49. Bankevich A et al. SPAdes: a new genome assembly algorithm and its applications to single-cell sequencing. J Comput Biol. 2012:19:455–477. https://doi.org/10.1089/cmb.2012.0021
50. Gurevich A, Saveliev V, Vyahhi N, Tesler G. QUAST: quality assessment tool for genome assemblies. Bioinformatics. 2013:29:1072–1075. https://doi.org/10.1093/bioinformatics/btt086
51. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. Checkm: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. Genome Res. 2015:25:1043–1055. https://doi.org/10.1101/gr.186072.114
52. Bowers RM et al. Minimum information about a single amplified genome (MISAG) and a metagenome-assembled genome (MIMAG) of bacteria and archaea. Nat Biotechnol. 2017:35:725–731. https://doi.org/10.1038/nbt.3893
53. Seemann T. Prokka: rapid prokaryotic genome annotation. Bioinformatics. 2014:30:2068–2069. https://doi.org/10.1093/bioinformatics/btu153
54. Parks DH et al. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. Nat Biotechnol. 2018:36:996–1004. https://doi.org/10.1038/nbt.4229
55. Lemoine F, Gascuel O. Gotree/Goalign: toolkit and Go API to facilitate the development of phylogenetic workflows. NAR Genom Bioinform. 2021:3:lqab075. https://doi.org/10.1093/nargab/lqab075
56. Letunic I, Bork P. Interactive Tree Of Life (iTOL) v5: an online tool for phylogenetic tree display and annotation. Nucleic Acids Res. 2021:49:W293–W296. https://doi.org/10.1093/nar/gkab301
57. Chen H, Boutros PC. VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. BMC Bioinformatics. 2011:12:35. https://doi.org/10.1186/1471-2105-12-35

58. Jain C, Rodriguez-R LM, Phillippy AM, Konstantinidis KT, Aluru S. High throughput ANI analysis of 90K prokaryotic genomes reveals clear species boundaries. Nat Commun. 2018:9:1–8. https://doi.org/10.1038/s41467-018-07641-9
59. Olm MR, Brown CT, Brooks B, Banfield JF. Drep: a tool for fast and accurate genomic comparisons that enables improved genome recovery from metagenomes through de-replication. ISME J. 2017:11:2864–2868. https://doi.org/10.1038/ismej.2017.126
60. Wickham H. Ggplot2: elegant graphics for data analysis. Springer-Verlag; 2016.
61. Xiao N. Ggsci: scientific journal and sci-fi themed color palettes for "ggplot2" 2024. https://nanxstats.r-universe.dev/ggsci
62. RStudio Team. RStudio: integrated development environment for R. RStudio, PBC.; 2020.
63. Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018:34:3094–3100. https://doi.org/10.1093/bioinformatics/bty191
64. Ide K et al. Exploring strain diversity of dominant human skin bacterial species using single-cell genome sequencing. Front Microbiol. 2022:13:955404. https://doi.org/10.3389/fmicb.2022.955404
65. Kogawa M, Hosokawa M, Nishikawa Y, Mori K, Takeyama H. Obtaining high-quality draft genomes from uncultured microbes by cleaning and co-assembly of single-cell amplified genomes. Sci Rep. 2018:8:2059. https://doi.org/10.1038/s41598-018-20384-3
66. Aroney STN et al. CoverM: read coverage calculator for metagenomics. Zenodo. 2024.
67. Schwengers O et al. Platon: identification and characterization of bacterial plasmid contigs in short-read draft assemblies exploiting protein sequence-based replicon distribution scores. Microb Genom. 2020:6:mgen000398. https://doi.org/10.1099/mgen.0.000398
68. Camacho C et al. BLAST+: architecture and applications. BMC Bioinformatics. 2009:10:421. https://doi.org/10.1186/1471-2105-10-421
69. Robinson JT et al. Integrative genomics viewer. Nat Biotechnol. 2011:29:24–26. https://doi.org/10.1038/nbt.1754
70. Attia S, Russel J, Mortensen MS, Madsen JS, Sørensen SJ. Unexpected diversity among small-scale sample replicates of defined plant root compartments. ISME J. 2022:16:997–1003. https://doi.org/10.1038/s41396-021-01094-7
71. Yamazaki S et al. Field multi-omics analysis reveals a close association between bacterial communities and mineral properties in the soybean rhizosphere. Sci Rep. 2021:11:8878. https://doi.org/10.1038/s41598-021-87384-8
72. Rinke C et al. Insights into the phylogeny and coding potential of microbial dark matter. Nature. 2013:499:431–437. https://doi.org/10.1038/nature12352
73. Foerstner KU, von Mering C, Hooper SD, Bork P. Environments shape the nucleotide composition of genomes. EMBO Rep. 2005:6:1208–1213. https://doi.org/10.1038/sj.embor.7400538
74. Ahsanuddin S et al. Assessment of REPLI-g multiple displacement whole genome amplification (WGA) techniques for metagenomic applications. J Biomol Tech. 2017:28:46–55. https://doi.org/10.7171/jbt.17-2801-008
75. Rodriguez-R LM, Konstantinidis KT. Estimating coverage in metagenomic data sets and why it matters. ISME J. 2014:8:2349–2351. https://doi.org/10.1038/ismej.2014.76
76. Holland-Moritz H, Vanni C, Fernandez-Guerra A, Bissett A, Fierer N. An ecological perspective on microbial genes of unknown function in soil [preprint], bioRxiv. 2021. https://doi.org/10.1101/2021.12.02.470747
77. Anthony WE et al. From soil to sequence: filling the critical gap in genome-resolved metagenomics is essential to the future of soil microbial ecology. Environ Microbiome. 2024:19:56. https://doi.org/10.1186/s40793-024-00599-w
78. Collingro A, Köstlbacher S, Horn M. Chlamydiae in the environment. Trends Microbiol. 2020:28:877–888. https://doi.org/10.1016/j.tim.2020.05.020
79. Nagler M, Insam H, Pietramellara G, Ascher-Jenull J. Extracellular DNA in natural environments: features, relevance and applications. Appl Microbiol Biotechnol. 2018:102:6343–6356. https://doi.org/10.1007/s00253-018-9120-4
80. Bowers RM et al. Dissecting the dominant hot spring microbial populations based on community-wide sampling at single-cell genomic resolution. ISME J. 2022:16:1337–1347. https://doi.org/10.1038/s41396-021-01178-4
81. Kaneko T et al. Complete genome sequence of the soybean symbiont Bradyrhizobium japonicum strain USDA6T. Genes (Basel). 2011:2:763–787. https://doi.org/10.3390/genes2040763
82. Ryoichi M, Kiwamu S. Comparison of legumes GABA levels in different growth stages and maturity stages. Food Preserv. Sci. 2017:43:171–177. https://doi.org/10.5891/jafps.43.171
83. Douarre P-E, Mallet L, Radomski N, Felten A, Mistou M-Y. Analysis of COMPASS, a new comprehensive plasmid database revealed prevalence of multireplicon and extensive diversity of IncF plasmids. Front Microbiol. 2020:11:483. https://doi.org/10.3389/fmicb.2020.00483
84. Shintani M, Sanchez ZK, Kimbara K. Genomics of microbial plasmids: classification and identification based on replication and transfer systems and host taxonomy. Front Microbiol. 2015:6:242. https://doi.org/10.3389/fmicb.2015.00242
85. Kadibalban AS, Landan G, Dagan T. The extent and characteristics of DNA transfer between plasmids and chromosomes. Curr Biol. 2024:34:3189–3200.e5. https://doi.org/10.1016/j.cub.2024.06.030
86. Zeng Y, Wang Y, Yu Z, Huang Y. Hypersensitive response of plasmid-encoded AHL synthase gene to lifestyle and nutrient by Ensifer adhaerens X097. Front Microbiol. 2017:8:1160. https://doi.org/10.3389/fmicb.2017.01160
87. Valentin-Alvarado LE et al. Asgard archaea modulate potential methanogenesis substrates in wetland soil. Nat Commun. 2024:15:6384. https://doi.org/10.1038/s41467-024-49872-z
88. Chaudhari JK, Pant S, Jha R, Pathak RK, Singh DB. Biological big-data sources, problems of storage, computational issues, and applications: a comprehensive review. Knowl Inf Syst. 2024:66:3159–3209. https://doi.org/10.1007/s10115-023-02049-4