

Exploring sagebrush leaf microbial metagenomes from deep, host-derived sequencing

Adedotun Adedayo Arogundade,¹ Carlos Dave C. Dumaguít,^{1,2} Anthony Melton,^{1,3} Sven Buerki,¹ Leonora S. Bittleston¹

AUTHOR AFFILIATIONS See affiliation list on p. 16.

ABSTRACT Advanced sequencing technologies and improvements in bioinformatics have provided a new way to study plant-associated microbial communities, including the use of host genomic sequencing. Our study focuses on the leaf microbiome of basin big sagebrush (*Artemisia tridentata* subsp. *tridentata*), a foundational shrub of western North America. We analyzed Illumina shotgun sequences from sagebrush leaves to investigate the metagenomes of leaf-associated microbes that were sequenced alongside their plant hosts. We aimed to profile the leaf microbiome across different sample sources (magenta box, greenhouse, and field/wild), reconstruct metagenome-assembled genomes (MAGs) where possible, and investigate functional gene annotations of the resulting MAGs, specifically with regard to the potential metabolism of sagebrush chemicals. To achieve this, Illumina shotgun sequence reads (containing both host and associated microbial reads) were mapped to the reference genomes of *Artemisia tridentata*, *Artemisia annua*, and the human reference genome to remove plant host and human-associated sequences. Host-cleaned reads were then analyzed using microbial metagenomics techniques. Taxonomic profiling revealed that *Phyllobacterium* and *Sphingomonas* were the most abundant microbial genera in greenhouse-grown plants, with very little variation among the samples. Wild, field-collected samples were much more variable and were dominated by *Klebsiella* and *Aureobasidium* species. From the co-assembly of greenhouse samples, we reconstructed two high-quality MAGs (a *Phyllobacterium* species and a *Sphingomonas* species) with >98% completion and <1% contamination. Functional annotation of these MAGs uncovered genes associated with the degradation and metabolism of camphor and other essential oils such as pinene, geraniol, and limonene, which are part of sagebrush leaf chemistry.

IMPORTANCE Big sagebrush (*Artemisia tridentata*), the foundation species of the sagebrush steppe, has broad ecological importance because its evergreen leaves offer nutrients and shade that facilitate the establishment of diverse understory plants in arid environments. Sagebrush leaves contain various secondary metabolites, including terpenoids, flavonoids, and phenolic compounds. These chemicals contribute to the plant's defense mechanisms against herbivores and pathogens. Despite this, sagebrush hosts diverse bacterial and fungal communities. We found that the microbial metagenome-assembled genomes (MAGs) we recovered contained genes that have the potential to degrade some of the chemical compounds in sagebrush leaves that could inhibit the growth of other microbes. This is the first study to mine plant genome data using host-derived sequences to generate microbial MAGs. Our results showed that MAGs can be recovered from plant host-derived sequence data, providing a new way to explore the identity and functional capabilities of difficult-to-culture microbes.

KEYWORDS microbial metagenomics, metagenome-assembled genomes (MAGs), sagebrush, phyllosphere microbiome, microbial bioinformatics, leaf microbiome,

Editor Dirk Tischler, Ruhr-Universität Bochum, Bochum, Germany

Address correspondence to Adedotun Adedayo Arogundade, adedotunarogunda@u.boisestate.edu.

The authors declare no conflict of interest.

See the funding table on p. 17.

Received 13 August 2025

Accepted 2 March 2026

Published 31 March 2026

Copyright © 2026 Arogundade et al. This is an open-access article distributed under the terms of the [Creative Commons Attribution 4.0 International license](https://creativecommons.org/licenses/by/4.0/).

plant-host derived microbial profiling, plant-associated microbial communities, functional annotation, Illumina reads, microbial ecology

The traditional method for characterizing microbes and sequencing their genomes begins with the isolation of individual strains from laboratory cultures (1, 2). However, many microbes are difficult to culture because their natural habitat is too complex to be reproduced in the laboratory or because they rely on other species to successfully grow. This has limited the scope of microbial studies and has left many microbes uncharacterized (3, 4). However, metagenomic sequencing provides a culture-independent approach to studying complex communities of microbes (5, 6). This culture-independent approach supports taxonomic and functional profiling of microbes, as well as reconstruction of metagenome-assembled genomes (MAGs). A MAG refers to a collection of similar scaffolds that are grouped together from a metagenome assembly based on tetranucleotide frequencies, abundances, complementary marker genes (7), taxonomic alignments (8), and codon usage (9) to represent a microbial genome.

The plant microbiome, defined as the community of microorganisms living in association with a plant, is increasingly recognized as essential to plant health, ecosystem functioning, and sustainable agriculture (10). Microbial assembly in this habitat is governed by complex interactions among microbes, their plant hosts, and the environment (11, 12). Numerous recent studies of the plant microbiome have explored microbial functional potential and composition using metagenomics techniques. For example, analyzing microbial communities across different plant species and conditions has helped to uncover how microbial assemblage and function vary based on factors like geography, climate, and plant genotype (13–15). Most plant metagenomics studies wash the microbes from the leaves and concentrate them prior to DNA extraction (16). A much rarer approach is to start with whole-leaf DNA sequencing and remove the host plant DNA to focus on the microbial component (17). This approach allows us to use existing whole-genome sequencing for plant genomes to uncover additional information about their microbial associations and thus their extended phenotype (17).

Our study focuses on the leaf microbiome of basin big sagebrush (*Artemisia tridentata* subsp. *tridentata*; Asteraceae), a foundational shrub species that dominates much of western North America (18). A reference-quality genome assembly was recently completed for this species, and deep sequencing is available across multiple individuals (19, 20). The ecological attributes of sagebrush affect community composition, ecosystem processes, and wildlife habitat over vast portions of western North America (18). Its evergreen leaves offer nutrients and shade that facilitate the establishment of diverse understory plants even in arid environments (21). Sagebrush foliage is relatively unpalatable to livestock and other large herbivores due to high levels of terpenes and tannins, thereby buffering more palatable understory plants from intensive grazing and allowing sensitive species like perennial bunchgrasses to persist (18). Greater sage-grouse, pygmy rabbits, pronghorn, and mule deer rely on sagebrush for shelter and their primary winter food source (21). Unfortunately, sagebrush ecosystems are critically threatened by wildfires, invasive species, climate change, land conversion, and urban development (22), which has resulted in the loss of more than 50% of the historic sagebrush range (23). The major ecological role of sagebrush underscores the need to protect the remaining intact sagebrush communities through ecological restoration and sustainable management in order to conserve biodiversity and ecosystem services across the western USA (18).

Although there is minimal research into the sagebrush phyllosphere microbiome (in the above-ground parts of plants), recent sequencing and culture-based studies have shown that, despite the high levels of terpenoids, phenolics, and other antibiotic metabolites in the foliage, sagebrush plants host diverse bacterial and fungal communities in and on their leaves (24, 25). Therefore, it is likely that sagebrush microbes are able to degrade many of the chemical compounds in the leaves, while other species are likely inhibited by the leaf-associated chemicals. Sagebrush-associated microbes

could potentially confer stress tolerance, nutrient acquisition, and pathogen protection benefits to their host plant, as has been found for other plant species (26, 27). Furthermore, establishing linkages between sagebrush chemistry and microbiome composition might offer new insights into plant-microbe coevolution. Reports have shown that terpene profiles vary across sagebrush subspecies and populations, which may drive corresponding shifts in associated microbiomes (28). Thus, comparative analyses of microbiomes among sagebrush chemotypes could uncover specialist microbial taxa adapted to particular compounds.

Culture-based experiments and amplicon sequencing (e.g., 16S for bacteria and ITS for fungi) have provided insight into the diversity and profile of microbes associated with the sagebrush phyllosphere (24, 25). However, these analyses are limited compared to shotgun-based metagenomics because they do not give insight into functional capabilities (29). In addition to taxonomic profiling, shotgun metagenomics provides a platform for genome-focused functional analysis if MAGs can be recovered (30). Characterizing genomes of prevalent or specialized sagebrush-associated microbes may reveal metabolic pathways involved in the detoxification or utilization of terpenoids, polyphenols, or other secondary metabolites that are produced by sagebrush. Furthermore, advances in sequencing technologies and bioinformatics analyses have led to the development of methods that allow more complete analyses of microbes that are not yet cultivated (31).

The objectives of this study were to profile the microbial taxa associated with basin big sagebrush leaves across different sample sources, reconstruct MAGs where possible, and explore the potential functions of the resulting MAGs. We asked the following questions: (i) Which microbial taxa can we profile from deeply sequenced sagebrush shotgun reads? (ii) How does host genotype or phenotype affect microbial composition and diversity in greenhouse and wild settings? (iii) Are there differences in the leaf-associated microbial community composition between greenhouse-grown vs wild plants? (iv) Can we reconstruct draft microbial genomes (MAGs) directly from sagebrush sequencing data? And if yes, (v) do these MAGs possess genes capable of metabolizing sagebrush chemicals?

MATERIALS AND METHODS

Genomic samples

We used existing sequence data from previous studies of sagebrush by researchers at Boise State University who developed *in vitro* propagation protocols to assemble the genome (32) and propagated seedlings from different populations (33). These data are divided into three categories based on the environment in which the plants were grown prior to sequencing. Sagebrush has a genome size of ~4.2 Gb (20).

Sterile magenta box

We analyzed Illumina HiSeq paired-end sequencing data from two plants grown in axenic or sterile conditions (32) that were sequenced at ~160× read depth (20). We expect that these plants should have little to no plant-associated microbes.

Greenhouse

Twelve plants were grown in a controlled, greenhouse environment. DNA was extracted from plant leaf tissues and was sent for Illumina HiSeq paired-end sequencing at ~20× read depth. This includes three groups of plants with origin locations in Idaho, Utah, and Nevada (34). Each region has four representative plants. Previous studies have shown that these plants have varying genotypes, and they are the focus of our research question to investigate the relationship between microbial composition and host genotypes.

Field

Illumina HiSeq data at ~10–12× read depth from 14 wild plants were also analyzed. All plants are likely to be the same general genotype, as they came from a single population but with different phenotypes, specifically, leaf-dropping phenotypes. Samples 14, 15, 26, 29, 34, 35, and 36 drop their leaves early, while samples 2, 21, 24, 3, 32, 33, and 8 drop their leaves late. These data enabled us to compare the diversity of microbes between wild plants with different phenotypes and also between wild plants and those grown in the greenhouse.

DNA extraction and sequencing

As reported by Melton et al. (20), DNA extraction for all the samples was done using the QIAGEN DNeasy Plant Mini Kit (Hilden, Germany; catalog # 69204) using approximately 20 mg of plant leaf tissue per sample. The samples were then processed for library preparation and Illumina HiSeq (San Diego, CA, USA) whole-genome paired-end sequencing (2 × 150 bp) by Azenta Life Sciences (New Jersey, NJ, USA), formerly Genewiz Inc.

Bioinformatic workflow

Preprocessing (quality assessment and removal of host reads)

The overview of the bioinformatic workflow used in this study is shown in Fig. 1. Preprocessing is the first step in any metagenomic analysis: removing adapter sequences, filtering low-quality nucleotides/reads, and generating quality assessment metrics. This was done using FastQC (35), and the Illumina sequences with Phred scores <20 were filtered with Trimmomatic (36). Our sequences were host-derived as they were extracted from plant leaf tissues and contained sequences from the host and the collective DNA of resident microbial communities (17), and with our focus on the associated microbial reads, we consider host genomes as contaminants that could impact our analyses. Using Bowtie2 (37), we mapped the trimmed reads to the reference genome of *Artemisia tridentata* subsp. *tridentata* and used Samtools to filter the alignment file to retain the unmapped reads, which are considered “non-host” reads as described by Roman-Reyna et al. (17). However, downstream analysis showed that there were still sagebrush reads in our “non-host” reads; therefore, we used BMTagger (38), a tool specifically designed to remove the human genome from metagenomic data sets but that can be adapted to remove other host genomes. In order to improve the clean-up process, we indexed reference genomes from *Artemisia tridentata* subsp. *tridentata*, *Artemisia annua*, and the human reference genome GRCh38/hg38 (NCBI Accession # [GCA_000001405.15](#)) and mapped them with our data sets. Table 1 shows sample information and the amount of “host” or “human” reads that were removed from each of our sequencing data sets. See the Supplementary Information for additional bioinformatic tools that were used but not included in the final workflow.

Taxonomic profiling of “host”-cleaned short Illumina reads

The Department of Energy Systems Biology Knowledgebase (KBase) System was used for most of our downstream metagenomic analyses (39) with default parameters for all software. We primarily used Kaiju v.1.9.0 (the taxonomic classifier on KBase) to classify host-cleaned short reads. Kaiju uses a protein-level comparison approach, which is more sensitive than nucleotide-based methods by comparing protein sequences from the reads against a reference database (40). However, we also used other tools not found in KBase in order to enable us to compare the results of these programs.

The other taxonomic tools we used were Kraken2 and DIAMOND/MEGAN (41, 42). Kaiju and DIAMOND programs used the NCBI non-redundant protein database, and Kraken2 used the “Standard Kraken2 plus RefSeq protozoa & fungi” database. Kaiju was run using protein-level sequence classification with stringent filtering thresholds.

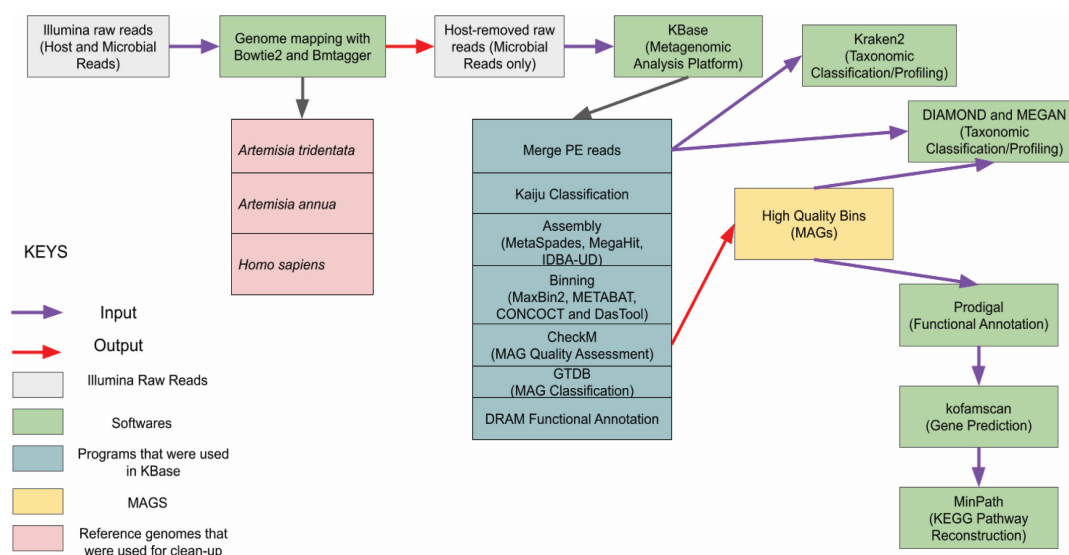


FIG 1 Overview of the bioinformatic workflow for the recovery of metagenome-assembled genomes from host-derived sequences of basin big sagebrush *Artemisia tridentata* subsp. *tridentata*

In greedy matching mode, up to three amino acid substitutions were allowed during match extension (-e 3), and only matches with a minimum alignment score of 65 were considered for taxonomic assignment (-s 65), ensuring sufficient match length and quality. An E-value cutoff of 0.01 (-E 0.01) was specified for maximum exact match searches, corresponding to a $\leq 1\%$ probability of random matches (40). DIAMOND alignments were filtered using an E-value threshold of $1e-3$, and only hits with a minimum bit score of 50 were retained (42). MEGAN taxonomic assignment was performed using the LCA algorithm with a 10% top-score cutoff to minimize spurious classifications. Kraken2 classification uses exact k-mer matching, and taxonomic assignments were filtered using a confidence threshold of 0.1, requiring at least 10% of k-mers within a read to support the assigned taxon, thereby reducing spurious classifications (41).

Co-assembly and reconstruction of MAGs

We built co-assemblies of sequence data from plants in the same category; that is, microbial sequences from all greenhouse-grown plants were co-assembled, and those from all the field-grown plants were co-assembled. KBase provides three assembly tools: metaSPAdes v.3.15.3 (43), MEGAHIT v.1.2.9 (44), and IDBA-UD v.1.1.3 (45), and we used all three of these programs to assemble the metagenomic reads into contiguous sequences. Assembling with multiple different tools is a common way to find which one produces the best results. Using Compare Assembled Contig Distributions v.1.1.2 software on KBase, we saw that metaSPAdes has the best N50 and N75, and it produced the longest contigs. Just like with the assembly process, KBase also provides three binning programs: MaxBin2 v.2.2.4 (46), MetaBAT2 v.1.7 (47), and CONCOCT v.1.1 (48), which were all used for MAG reconstruction. The individual outputs (bins) from these programs were passed into DASTools v.1.1.2 (49) to produce high-quality, non-redundant MAGs.

Classification of MAGs

According to literature, the differences between read-based taxonomic profiles and MAG taxonomy arise from a combination of biological complexity and methodological bias (50). Read-based approaches capture the total pool of sequenced DNA and therefore reflect community composition across both abundant and low-abundance taxa, whereas MAG recovery is constrained by assembly and binning efficiency, which preferentially

TABLE 1 Sample information including amount of host or human reads that were removed from the sequencing data

Sample ID	Category	Initial gzipped file size	Approximate size of host/human data removed from cleaned reads			Final gzipped file size	Number of cleaned PE raw reads (basepairs)
			<i>A. tridentata</i>	<i>A. annua</i>	<i>Homo sapien</i>		
IDT3	Magenta	38.0 Gb	24.0 Gb	4.4 Gb	970 Mb	8.4 Gb	25,738,784
UTT2	Magenta	58.0 Gb	34.0 Gb	4.6 Gb	1.31 Gb	18.2 Gb	55,350,386
29-ID	Greenhouse	4.4 Gb	2.2 Gb	658 Mb	0 Mb	1.48 Gb	4,207,818
143-ID	Greenhouse	3.6 Gb	1.83 Gb	550 Mb	0 Mb	1.29 Gb	3,683,012
163-ID	Greenhouse	3.4 Gb	1.84 Gb	540 Mb	0 Mb	1.11 Gb	3,158,346
192-ID	Greenhouse	3.4 Gb	1.66 Gb	456 Mb	0 Mb	1.31 Gb	3,717,942
246-NV	Greenhouse	4.2 Gb	2.20 Gb	602 Mb	0 Mb	1.58 Gb	4,490,656
326-NV	Greenhouse	3.6 Gb	1.98 Gb	576 Mb	0 Mb	1.15 Gb	3,267,482
379-NV	Greenhouse	3.8 Gb	2.02 Gb	560 Mb	0 Mb	1.29 Gb	3,666,774
399-NV	Greenhouse	4.0 Gb	2.40 Gb	580 Mb	0 Mb	1.21 Gb	3,440,062
422-UT	Greenhouse	5.8 Gb	3.20 Gb	780 Mb	0 Mb	2.02 Gb	5,794,018
443-UT	Greenhouse	4.0 Gb	2.02 Gb	562 Mb	0 Mb	1.47 Gb	4,186,566
541-UT	Greenhouse	4.0 Gb	2.20 Gb	620 Mb	0 Mb	1.25 Gb	3,558,232
579-UT	Greenhouse	3.8 Gb	1.98 Gb	564 Mb	0 Mb	1.30 Gb	3,696,426
2-Wild	Field/Wild	3.2 Gb	1.8 Gb	464 Mb	4.4 Mb	911 Mb	2,601,426
3-Wild	Field/Wild	3.4 Gb	2.0 Gb	474 Mb	3.6 Mb	907 Mb	2,590,152
8-Wild	Field/Wild	4.0 Gb	2.0 Gb	524 Mb	2.6 Mb	2.3 Gb	6,519,242
14-Wild	Field/Wild	3.0 Gb	2.0 Gb	502 Mb	1.0 Mb	853 Mb	2,436,802
15-Wild	Field/Wild	4.0 Gb	2.2 Gb	588 Mb	48 Mb	1.2 Gb	3,604,450
21-Wild	Field/Wild	4.0 Gb	2.2 Gb	540 Mb	12.8 Mb	1.3 Gb	3,559,682
24-Wild	Field/Wild	3.0 Gb	1.7 Gb	470 Mb	1.9 Mb	800 Mb	2,283,504
26-Wild	Field/Wild	3.6 Gb	2.2 Gb	548 Mb	13.4 Mb	1.0 Gb	2,960,114
29-Wild	Field/Wild	3.6 Gb	2.2 Gb	448 Mb	2.2 Mb	853 Mb	2,433,302
32-Wild	Field/Wild	3.4 Gb	2.2 Gb	510 Mb	1.5 Mb	911 Mb	2,590,152
33-Wild	Field/Wild	3.8 Gb	2.2 Gb	512 Mb	3.6 Mb	1.2 Gb	3,528,676
34-Wild	Field/Wild	3.4 Gb	1.9 Gb	488 Mb	8.4 Mb	1.1 Gb	3,125,510
35-Wild	Field/Wild	4.0 Gb	1.8 Gb	234 Mb	556 Mb	1.3 Gb	3,839,712
36-Wild	Field/Wild	3.6 Gb	2.0 Gb	518 Mb	3.0 Mb	965 Mb	2,754,048

reconstructs genomes from abundant, low-diversity populations (50). Prior to classification, CheckM v.1.0.18 (51) was used to assess the quality of the MAGs, and then the Genome Taxonomy Database (GTDB) toolkit, GTDB-Tk v.1.7.0 (52), was used to classify them. We chose GTDB because it is the most common classification tool for MAGs that has been reported (52). Although traditional methods of 16S rRNA classification have helped to understand the diversity of microbes in metagenomic data sets, they are often limited in characterizing metagenome-assembled genomes because small subunits such as 16S rRNAs are poorly represented in MAGs (53, 54). To overcome this issue, GTDB-Tk uses HMMER (55) to identify a set of 120 bacterial marker genes (<https://data.ace.uq.edu.au/public/gtdb/data/releases/latest/>) in the reference genomes on the GTDB (56) to create multiple sequence alignments, which are then used to create an inference-based phylogenomic tree (57). The classification of the queried MAG is determined by its position on the tree. In addition to GTDB classification, we exported the fasta files of these bins from KBase and classified them using Kraken2 and DIAMOND. We found that the Kraken2 and DIAMOND results were consistent with the GTDB-generated taxonomy.

Gene prediction/functional profiling of MAGs

Gene prediction helps to identify the genetic potential of the recovered MAGs. This step enables the identification of functional genes and pathways relevant to plant-microbe interactions and ecosystem functioning. Distilled and Refined Annotation of Metabolism

(DRAM) profiles microbial (meta)genomes for metabolisms known to impact ecosystem function (58). DRAM annotation on KBase assessed the completion of a few metabolic modules and electron transport chain complexes and also predicted the presence of carbohydrate-active enzymes in the metagenomes. We also annotated the MAGs using Prodigal (59), and the annotated genomes were then passed into KofamScan (https://github.com/takaram/kofam_scan), a command-line interface of the web-based KofamKOALA (60), a gene function annotation tool based on Kyoto Encyclopedia of Genes and Genomes (KEGG) Orthology and hidden Markov models. MinPath (61) was used to reconstruct a parsimonious KEGG (62) biological pathway using protein family predictions. This provides comprehensive reports on the metabolic modules and pathways that are found in MAGs. Sagebrush contains many secondary metabolites, including terpenes, flavonoids, phenolic compounds, and essential oils, which play essential roles in the plant's defense against herbivores, pathogens, and environmental stressors. Several studies have reported that the major chemical compounds in sagebrush are artemiseole, p-cymene, camphor, camphene, eucalyptol, limonene, geraniol, beta-pinene, and α -pinene (63, 64). They are also essential components of the plant's volatile organic compounds that influence the plant's interactions with its environment, including other organisms and ecosystems. Therefore, as part of our functional annotation, we screened the MAGs for genes that code for enzymes capable of metabolizing these compounds.

Diversity metrics

Microbial diversity metrics were generated and analyzed using R version 4.2.1 (65). Community matrix data that contain differentially assigned taxa and read counts for each category (greenhouse, wild, and magenta) were generated using the "dcast" function of the reshape2 package (66) and were rarefied using the "rrarefy" function in the vegan package (67).

We measured alpha diversity using the effective number of species (calculated as the exponential of the Shannon diversity) for the taxa assigned during metagenomic profiling. After checking the assumptions for parametric tests on the data, we used a simple analysis of variance (ANOVA) to test the differences in the effective number of species between the three host genotypes within the greenhouse-grown samples (Idaho, Nevada, and Utah) and between host phenotype among wild samples. We used the Kruskal-Wallis test to check for the difference between categories (greenhouse and wild samples) because the data failed the assumptions for a parametric test.

We calculated Bray-Curtis microbial community dissimilarities between samples using the "vegdist" function in the vegan package. Differences in community structure among host genotypes (Idaho, Nevada, and Utah) of the greenhouse-grown plants and phenotype-driven differences among wild plants were assessed using a permutational multivariate analysis of variance (PERMANOVA) on Bray-Curtis distances using the "adonis" function (also in the vegan package). Likewise, differences in microbial community composition between greenhouse-grown and field-grown plant categories were assessed using PERMANOVA on Bray-Curtis distances. For all PERMANOVA comparisons, we also tested for homogeneity of variance using the "betadisper" function.

Graphs

Figure 1 was created using Microsoft PowerPoint. Figures 2 and 3 were plotted using R version 4.2.1 (65). Alpha and beta diversity figures were plotted using the "ggplot" function in the ggplot2 package (68). Taxonomy bar plots were plotted using the "barplot" function in base R. Figures 4 and 5 were generated using SpeciesTree (v.2.2.0) and the Annotate and Distill Assemblies with DRAM (v.0.1.2) pipeline on KBase (39), respectively.

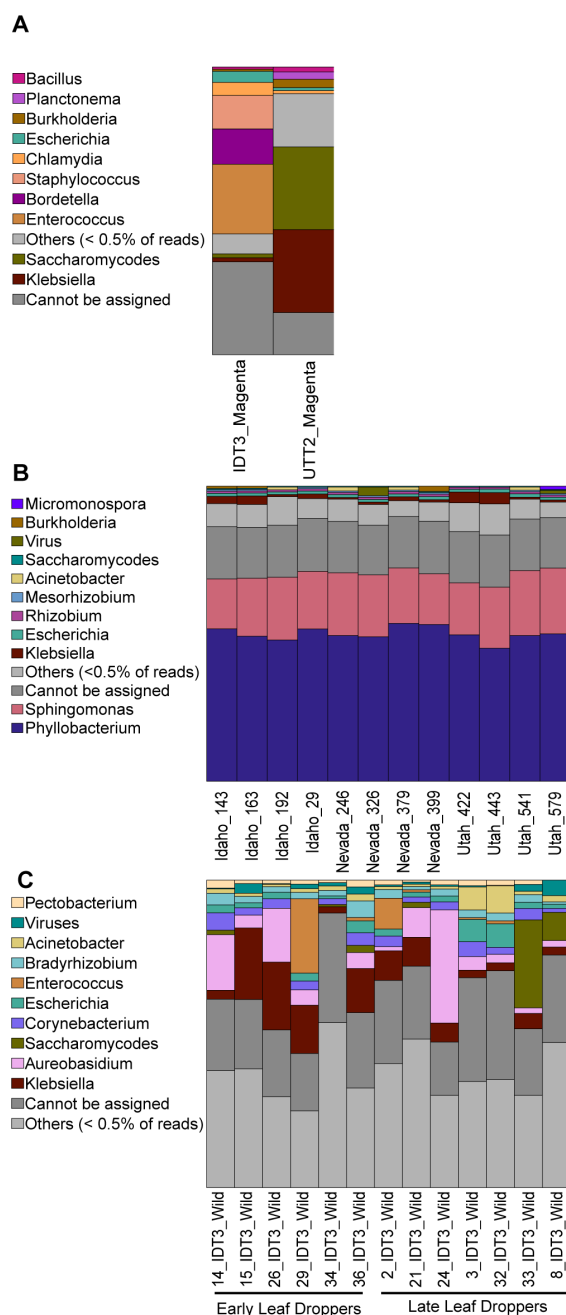


FIG 2 Taxonomic barplot showing the relative abundance of the top taxa at the genus level across (A) magenta box samples, (B) greenhouse samples, and (C) wild/field samples.

RESULTS

Preprocessing (quality assessment and removal of host reads)

As expected, a large proportion of the initial reads were removed after mapping sequences to the *Artemisia tridentata* genome (Table 1), and we saw further clean-up across all samples after mapping against the closely related *A. annua* genome. We also removed human reads from magenta and wild samples, but no human reads were detected in the greenhouse samples (Table 1). The large differences in the sizes of “host”-cleaned reads, final read sizes, and the number of cleaned paired-end reads across the samples (especially with the two magenta samples relative to the wild and greenhouse samples) were due to differences in sequencing depth for these samples.

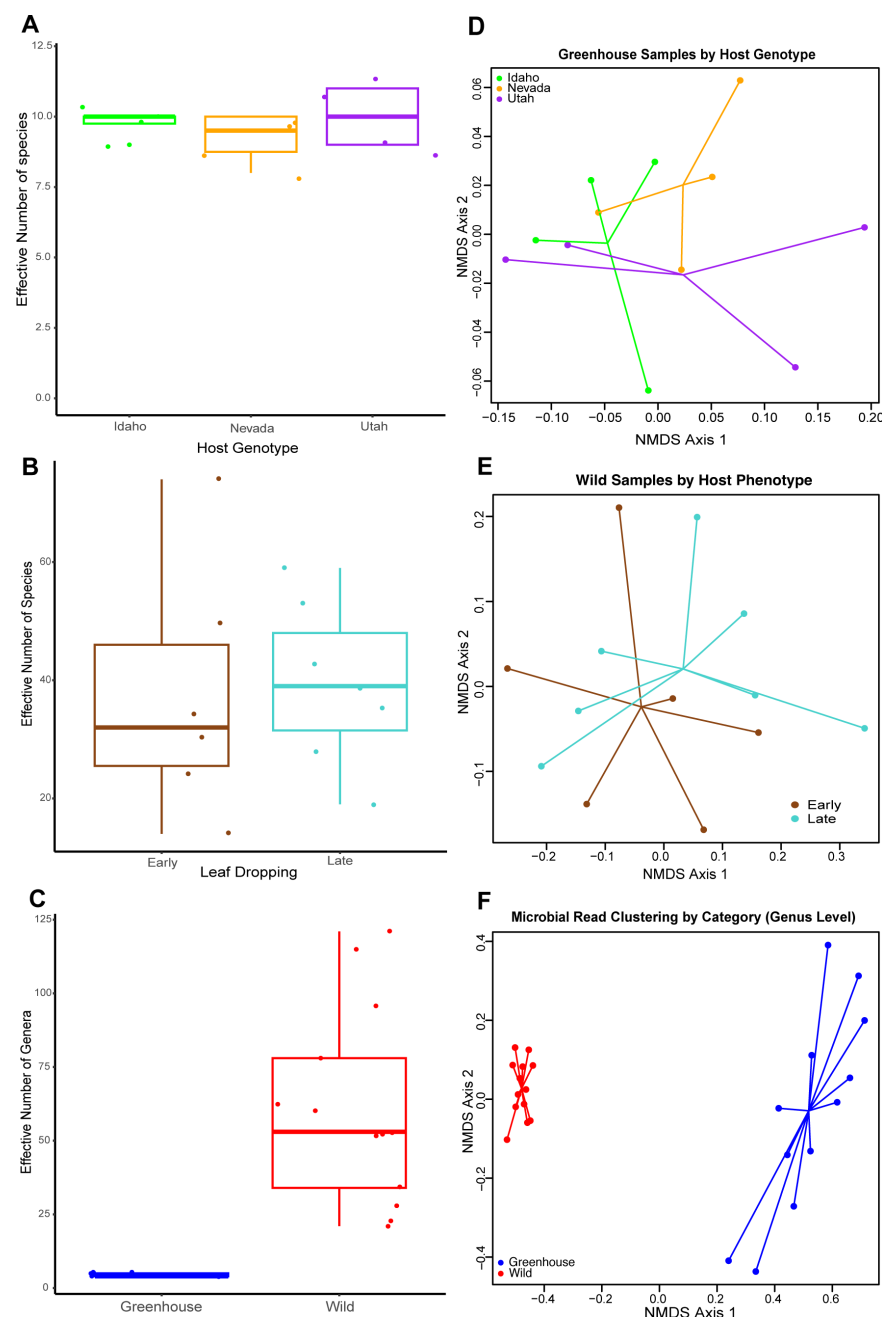


FIG 3 Wild samples had higher richness and different community composition than greenhouse samples. (A) Alpha diversity boxplot showing the effective number of species in greenhouse samples across genotypes from distinct origin populations. (B) Alpha diversity boxplot showing the effective number of species among wild samples with different leaf-dropping phenotypes. (C) Alpha diversity boxplot showing higher microbial diversity at the genus level in wild samples relative to greenhouse samples. (D) Non-metric multi-dimensional scaling (NMDS) plot showing no relationship between host genotype and microbial composition among the three populations of plants grown together in a greenhouse. (E) NMDS plot showing no relationship between host phenotype and microbial composition among wild samples. (F) NMDS plot showing strong clustering between microbial communities at the genus level in the categories of wild vs greenhouse plants.

Magenta samples were sequenced at ~160× sequencing depth, while greenhouse and wild samples were sequenced at ~20× and ~10–12×, respectively.

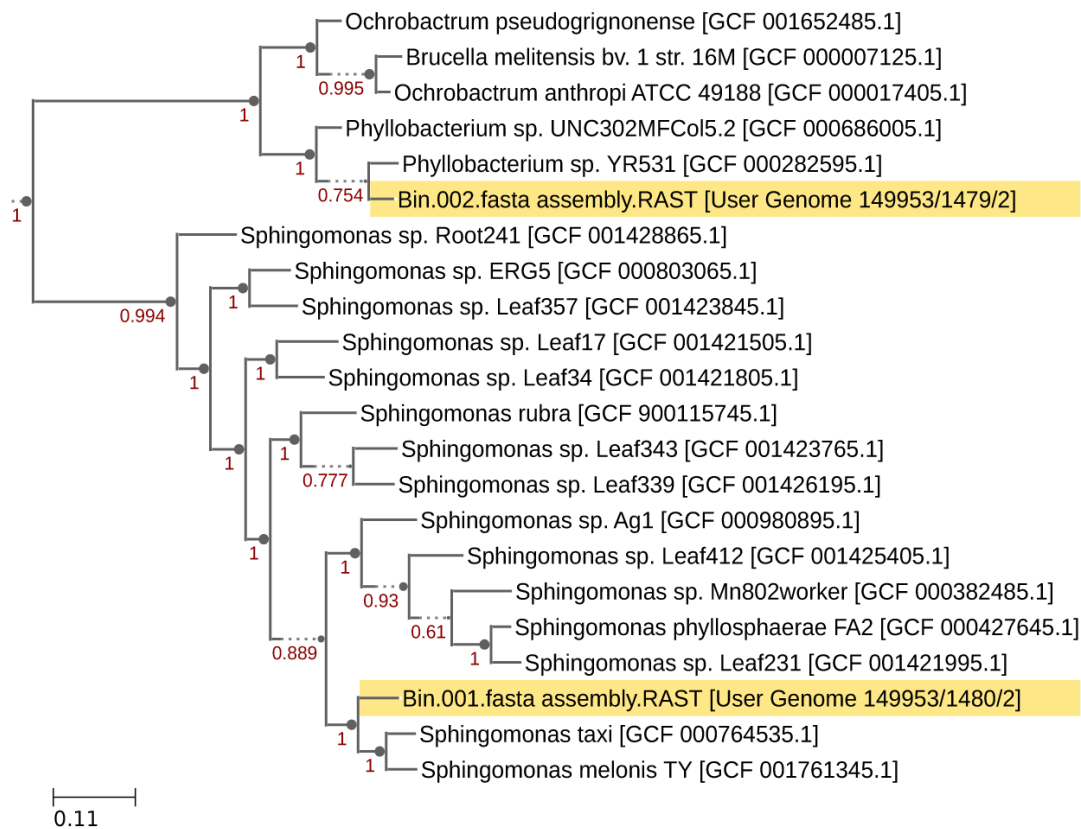


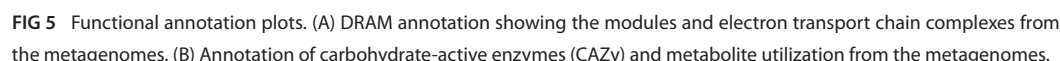
FIG 4 The placement of two high-quality MAGs in a phylogenomic tree supports their identities as a *Phyllobacterium* species and a *Sphingomonas* species.

Taxonomic assignment of short metagenome reads

Taxonomic profiling with different software packages (Kaiju, Kraken2, and DIAMOND) revealed that less than 20% of the reads across each sample could be classified, that is, assigned taxa, with the rest of the data not having a match across the different databases that were used. This might indicate that these unclassified reads were novel sequences; however, due to the complex nature of the sagebrush genome (with 78% being repetitive elements [20]), these are most likely to be repetitive host-related sequences that were not properly removed when mapping to the host genome and are not relevant for microbial analyses. In order to verify this, we re-mapped the “cleaned” reads from one sample back to the plant host reference genome again, and the result showed that many of the reads matched to regions of the plant genome at higher than 80% identity. This issue has also been seen in other studies (69–71).

The magenta box samples were dominated by human-associated microbes such as *Staphylococcus* sp., *Escherichia* sp., and *Enterococcus* sp., among others (Fig. 2A). These microbes were likely introduced while caring for the plants during cultivation or during handling prior to DNA extractions. Therefore, we excluded these two samples from further downstream analyses. We also excluded “Sample 35” of the wild samples from downstream analysis because it had a significant number of human-contaminated reads (Table 1) and was an outlier in terms of community composition.

The most abundant microbial genera across the greenhouse samples were *Phyllobacterium* and *Sphingomonas* (Fig. 2B). Species from these genera are known plant-growth promoters and are likely to be truly plant-associated. We also note that the greenhouse samples did not contain detectable human reads and so were likely to be less contaminated with human-associated microbes (Table 1). Other taxa in the greenhouse samples in low relative abundances included *Rhizobium* sp., *Mesorhizobium* sp., and *Acinetobacter* sp. (Fig. 2B).

10.1128/spectrum.02198-25 **11**

Alpha diversity

For greenhouse samples, which are composed of three host genotypes, the effective number of species per sample ranged from 8 to 11, with a median of 10 and a mean of 9.67. These samples were very similar and even, with no significant differences in the effective number of species among the host genotypes (ANOVA, P -value = 0.522) (Fig. 3A). Tukey's HSD *post hoc* tests showed no significant pairwise differences between any of the genotypes (all adjusted P -values >0.50). Confidence intervals for mean differences all included zero. For wild samples, the effective number of species per sample ranged from 16 to 84, with a median of 42.5 and a mean of 45.43. There was no significant difference in the effective number of species among wild samples with different phenotypes (ANOVA, P -value = 0.861110) (Fig. 3B). Tukey's HSD test also showed that the mean difference between the two phenotypes (9.86) was not significant, with a 95% confidence interval of (lwr = -19.69, upr = 39.40). However, there was higher microbial diversity (Kruskal-Wallis test, P -value = 1.544×10^{-5}) in the wild plants relative to the greenhouse plants (Fig. 3C). Because only two groups (greenhouse and wild plants) were compared, this result is equivalent to a Wilcoxon rank-sum test, which confirmed that greenhouse plants exhibited significantly lower values than wild/field plants ($W = 0$, $P = 1.75 \times 10^{-5}$). The Hodges-Lehmann estimate indicated a median difference of -49 (95% CI: -74 to -30), which shows a significant separation between the groups.

Beta diversity

NMDS using Bray-Curtis dissimilarity and an associated PERMANOVA showed no relationship between microbial composition and plant host genotypes among greenhouse samples ($R^2 = 0.16036$, P -value = 0.555, stress = 0.0382), also a non-significant (P -value = 0.5123) test for homogeneity of multivariate dispersions (betadisper) indicated that within-group variability was comparable across host genotypes. This is not surprising because of the close proximity and low microbial diversity of these plants while growing in the greenhouse (Fig. 3D).

Likewise, tests for homogeneity of multivariate dispersion were not significant (betadisper, P -value = 0.9887) for wild/field samples, indicating similar dispersion of microbial communities across host phenotypes. Consistent with this, PERMANOVA revealed no significant relationship between microbial community composition and host phenotype among wild plants ($R^2 = 0.08535$, P -value = 0.437, stress = 0.0614; Fig. 3E).

In contrast, when comparing greenhouse and wild samples, tests for homogeneity of dispersion were significant (betadisper, P -value = 1.889×10^{-12}), indicating differences in multivariate dispersion between environments. Although this suggests caution must be taken in interpreting PERMANOVA results, we noted clear separation of clusters (Fig. 3F), and a PERMANOVA conducted using Bray-Curtis distances nevertheless revealed a strong and significant difference in genus-level community composition between greenhouse and wild samples ($R^2 = 0.72119$, P -value = 0.001, stress = 0.0652; Fig. 3F), suggesting substantial environmental structuring of the microbial communities between these two categories.

Read assembly and reconstruction of MAGs

According to the Compare Assembled Contig Distributions and its metrics, metaSPAdes performed best because it produced the highest number of larger contigs that were between 25 and 75 Kb (Fig. S1). Therefore, we chose to move forward with it for our metagenomic binning. Although no bins were recovered from the wild sample co-assembly, metaSPAdes also performed better in that case.

We recovered three MAGs with >70% completion from the co-assembled contigs from the greenhouse samples. These MAGs were filtered using CheckM with set parameters of >50% completion and <10% contamination; this effectively removed one of the bins and retained two high-quality MAGs with >98% completion and <1% contamination/redundancy (Table 2). We recovered no bins from the co-assembly of the

TABLE 2 CheckM quality assessment of MAGs that were recovered from greenhouse plants^a

Bin name	Marker lineage	Genomes size (bp)	# Markers	# Marker sets	Completeness (%)	Contamination (%)
Bin.001	<i>o__Sphingomonadales</i>	3,972,965	569	293	98.62	0.9
Bin.002	<i>c__Alphaproteobacteria</i>	5,311,195	349	230	100	0.87
<i>Bin.003</i>	<i>k__Archaea</i>	<i>21,065,390</i>	<i>149</i>	<i>107</i>	<i>70.02</i>	<i>23.3</i>

^aThe italicized text shows the one low-quality bin that was filtered out with the set parameter of >50% completion and <10% contamination.

wild, field-collected samples. To check if this was due to the large variability in microbial composition and high diversity across these samples, we also tried assembling each wild sample individually. However, we got the same result when we assembled and binned the samples individually. Therefore, it is likely that the wild samples did not assemble well into potential MAGs due to the lower sequencing depth together with the higher diversity of these communities.

Taxonomic classification of recovered MAGs

The two high-quality MAGs were classified using GTDB-Tk v.1.7.0 of the GTDB as *Sphingomonas* sp. and *Phyllobacterium* sp. Corroborating this taxonomic assignment, phylogenomic placement using SpeciesTree v.2.2.0 placed these bins with related species as shown in Fig. 4.

Functional annotation of genes

Our analyses using DRAM and Carbohydrate-Active Enzymes Database (CAZy) showed that our two recovered MAGs contain basic functions that are necessary for survival (Fig. 5). The MAGs have close to 100% completion for life-essential modules/pathways such as Embden-Meyerhof, Entner-Doudoroff, Pentose Phosphate, TCA cycle. They are also very similar in their annotations for the ETC (Electron Transport Chain) complexes, with the exception of Cytochrome bd ubiquinol oxidase, which is only found in Bin.002 (*Phyllobacterium*). Cytochrome bd is a ubiquinol: oxygen oxidoreductase of many prokaryotic respiratory chains, and it catalyzes the reduction of oxygen (O₂) to water (H₂O) by using quinols as a reducing substrate (72). It enables bacterial O₂ consumption to promote specific metabolic pathways and/or to prevent O₂ toxicity and other stress conditions, including high temperature and exposure to toxic compounds (73–75). KEGG annotation recovered multiple genes in our two MAGs associated with degradation pathways of chemicals found in sagebrush leaves (Table 3). We note that per-gene coverage for these degradation pathways was not calculated, but would increase confidence that these pathways were well-represented and functional in the genomes.

TABLE 3 KEGG annotation table showing the number of genes in categories related to the degradation of chemicals found in sagebrush leaves

KEGG pathway	Number of pathway genes found in		Subcategory related to degradation of sagebrush chemicals
	MAGs		
	Bin.001 <i>Sphingomonas</i>	Bin.002 <i>Phyllobacterium</i>	
Geraniol degradation	16	16	K19653 geoA; geraniol dehydrogenase (NAD+) (EC:1.1.1.347)
Limonene and pinene degradation	12	12	K21569 camC, CYP101A1; camphor 5-monooxygenase (EC:1.14.15.1)
Nitrotoluene degradation	18	20	K21683 dntB; 4-methyl-5-nitrocatechol 5-monooxygenase (EC:1.14.13.210)

DISCUSSION

Our culture-independent approach, using sagebrush high-throughput whole-genome sequencing to explore the leaf microbiome, was able to characterize associated taxa. Previous microbial metagenomics studies have highlighted how low coverage of microbial genomes and high microbial diversity can hinder the recovery of MAGs (76, 77). For example, one study recovered 20 MAGs from soil with 4.3 Gb of sequencing data (and soil is known to be one of the environments with the highest bacterial richness), while another identified 12 MAGs from olive metagenomes from 3.47 Gb of sequencing data, with the majority obtained from samples exhibiting higher average coverage and lower community complexity (78, 79). The sequencing depths were unequal across our sample types because the data used in this study originated from previous sagebrush studies conducted by researchers at Boise State University, which involved different initial sequencing efforts. According to Orellana et al. (80), differences in sequencing depth directly impact the analysis of microbial diversity and the successful recovery of MAGs because lower sequencing depth, especially when coupled with high microbial diversity, can impede the reconstruction of high-quality MAGs. In line with this, our results were best from the greenhouse-grown plant samples, which had deep enough host genome sequencing ($\sim 20\times$) and simple enough richness (~ 10 effective number of species per sample) to generate high-quality MAGs of the two most abundant bacterial taxa. The wild-collected samples, with ~ 10 – $12\times$ host genome sequencing and much higher species richness, were too complex to build high-quality MAGs. This coverage and diversity information will likely be useful for other studies aiming to use plant host genome sequencing to better characterize the associated microbiome.

To the best of our knowledge, this is the first study to use this approach of mining plant genomes to generate microbial MAGs. Roman-Reyna et al. (17) used whole-genome-shotgun sequences from rice that encompassed both rice and the collective DNA of resident microbial communities to leverage the “unmapped” non-plant reads as microbiome data set. However, they did not generate MAGs. Numerous previous studies have used computational filtering of human genome-mapped reads, as it is a common preprocessing step in human microbiome research (81) and other studies on plant microbiomes have also cleaned metagenomes with the plant host genome. For example, Bulgarelli et al. (82) mapped metagenome samples to barley annotated genomic sequences using bwa-mem. In a more recent study, Regalado et al. (15) removed *Arabidopsis thaliana* reads from their metagenomic data by aligning against the *A. thaliana* reference genome and used only read pairs that were not mapped with the plant reference genome for classification.

One issue we faced is that only a small percentage of our post-host-removal reads were able to be classified. All of the classification programs we used showed that less than 20% of the reads across each sample could be classified. Further analysis suggested that many of these unclassified reads were host-associated repetitive sequences that were not properly removed when mapping against the host genome. However, having high numbers of unclassified reads is common in microbiome research, where multiple studies using shotgun data have found that more than 50% of reads cannot be annotated against known databases (15, 50, 77, 83). Regalado et al. (15) checked the efficiency of host-read removal by mapping the unclassified reads with *A. thaliana* reference genome, but less than 1% mapped with the reference genome, and they reported that only a small percentage of unclassified reads was likely to come from the plant. Thus, their unmapped reads were mostly from portions of microbial genomes not present in the NCBI nr database used with DIAMOND. This is likely partially the case for our study; however, the big sagebrush genome is one of the most highly repetitive plant genomes to be sequenced and assembled, with repeat elements comprising 77.99% of the genome (20) and it is very large when compared with *Arabidopsis thaliana* (~ 0.13 Gb for *A. thaliana* relative to ~ 4.2 Gb for *A. tridentata*). Therefore, in our study, many of these unmapped reads are likely to be host-derived and just not yet properly assembled into the highly complex host genome.

The most abundant taxa in all our greenhouse samples were *Phyllobacterium* and *Sphingomonas* species (Fig. 2B). These two taxa showed up reliably, and with very little variation across individual plants. Both genera contain known plant-growth-promoting species. *Phyllobacterium* is known plant-growth-promoting bacteria on several plants like *Picea* spp., *Brassica napus*, and *Arabidopsis thaliana* (84, 85). Some *Phyllobacterium* species have been observed to enhance plant tolerance to abiotic stresses such as drought. Bresson et al. (86) reported that inoculation of *Phyllobacterium brassicacearum* strain STM196 improved the drought tolerance in *Arabidopsis thaliana* under a long-term water deficit. *Sphingomonas* are also known to produce indole acetic acid, gibberellins, and essential amino acids (87, 88). *Sphingomonas* species are known for their ability to degrade complex compounds such as aromatic hydrocarbons and other xenobiotics (89). Although their ability to degrade cellulose and hemicellulose could be beneficial for carbon decomposition and recycling, they can become opportunistic pathogens in certain conditions (90). To understand more about the capabilities of the strains we recovered, we annotated the functional genes of our MAGs. Analysis of carbohydrate-active enzymes highlighted differences in hydrolytic capabilities of the two bacteria, with the *Sphingomonas* species having the genes to degrade xyloglucan, mixed-linkage glucans, and arabinan while the *Phyllobacterium* species has genes for degrading amorphous cellulose (Fig. 5B). Interestingly, from our KEGG annotation results, we found that specific genes associated with the degradation of chemicals found in sagebrush leaves were abundant in our two MAGs (Table 3). Geraniol, limonene, and pinene are naturally occurring terpenes in sagebrush (63, 64). They contribute to the characteristic scent of the plant and potentially serve as protective agents against herbivores and pathogens. Therefore, it makes sense that these MAGs might contain genes capable of metabolizing these toxic compounds. Nitrotoluene has not been reported to be present in sagebrush; however, detoxification enzymes are often capable of metabolizing other substrates with similar chemical structures; therefore, it is possible that the products of these genes could degrade similar compounds that are found in sagebrush. Thus, a previous study by Kohl et al. (91) that investigated the genomic potential of microbes in the gut of sage-grouse fed with 100% sagebrush also looked into the presence of nitrotoluene-degrading genes. These chemicals are key parts of the complex sagebrush leaf chemistry and suggest that the *Phyllobacterium* and *Sphingomonas* species we recovered have the potential to metabolize secondary metabolites which are generally considered antimicrobial.

Relative to the greenhouse samples, the wild samples were more variable with high alpha diversity compared to the greenhouse plants (Fig. 3C). The reduced diversity of the greenhouse samples was likely because the plants were grown in the same controlled environment and were exposed to fewer microbial species and less environmental variability than they would be in their natural environment. Greenhouse-grown plants often have reduced leaf microbiomes (14), so it makes sense that our wild samples were more diverse. Our beta-diversity analyses also showed significant differences in community structure among greenhouse and wild samples (Fig. 3F). Several factors contribute to microbial community assembly in plants, including host genotype, where plant hosts exert some control over the abundance and composition of their microbiome as a result of the differing chemical and physical features of resources provided on their phyllosphere (92), a phenomenon known as genotype effects (12). The greenhouse-grown plants we used had different genotypes (34), but our results showed no significant relationship between microbial composition and plant host genotypes (Fig. 3D). This could be due to the reduced microbial diversity in the greenhouse, particularly for plants that had been grown in a sterile environment as seedlings before entering the greenhouse.

Our field-collected samples were likely all of similar genotype, but with different leaf-dropping phenotypes. While previous studies have shown that there is a significant relationship between microbial communities and plant phenotypes (93, 94), our test for

phenotype-driven differences in the microbiome of these more natural samples showed no significant relationship (Fig. 3E).

Although taxonomic composition varied strongly across wild samples, the most abundant genera were *Klebsiella* and *Aureobasidium*. Recent research from the Bittleston lab group using both sequencing and culture-based methods to characterize fungal associates has found *Aureobasidium pullulans* to be the most abundant fungal species on sagebrush leaves (25). This cosmopolitan species often associates with plants (95–97) and can also be found in water or in harsh environments. It is a black yeast, and the high melanin content may provide some protection against UV light damage. *A. pullulans* produces siderophores that provide iron to plants and the antifungal pullulan that can inhibit pathogens (98, 99). Exopolysaccharides from *A. pullulans* can also increase soil water retention capacity and cation exchange capacity, which promotes nutrient availability and increases plant yield (100). The most abundant bacterium identified in our wild samples was *Klebsiella pneumoniae*, a common plant-associated bacterium capable of a broad range of functions such as nitrogen fixation, production of phytohormones like indole acetic acid (IAA), solubilization of inorganic phosphates, and compound degradation (101). Kumar et al. (102) reported that *K. pneumoniae* strain M6, which was isolated from mango rhizosphere, is a potent plant growth promoter, biocontrol, and paclobutrazol (a chemical that inhibits gibberellin biosynthesis in plants) degrading agent. Thus, both the *Klebsiella* and *Aureobasidium* taxa we identified here have the potential to benefit the sagebrush host plant.

Conclusion

We used a culture-independent method to study the plant microbiome directly from host-targeted Illumina shotgun sequence data. As might be expected, wild samples were much more diverse than samples grown in a greenhouse and/or sterile magenta boxes. The top taxa from our greenhouse samples were *Phyllobacterium* sp. and *Sphingomonas* sp., while *Klebsiella* sp. and *Aureobasidium* sp. were the top taxa in wild samples. All these genera contain species known to support plant health. We recovered two high-quality MAGs from the co-assembly of the greenhouse samples, and functional annotation showed that these MAGs contain genes with the potential to degrade complex secondary compounds found in sagebrush leaves, further supporting the likelihood that these are long-term associations of this foundational plant species.

ACKNOWLEDGMENTS

This research is partially supported by the NSF Idaho EPSCoR Program and by the National Science Foundation under award number OIA-1757324. The authors acknowledge support from Dr. Kevin Feris, Peggy Martinez, the Buerki lab, and the Bittleston lab group.

The authors thank Boise State University for providing the resources needed, and the Department of Biological Science for providing the best atmosphere for learning and research.

AUTHOR AFFILIATIONS

¹Department of Biological Sciences, Boise State University, Boise, Idaho, USA

²Department of Nutrition and Integrative Physiology, University of Utah, Salt Lake City, Utah, USA

³Department of Biology, Chemistry, Mathematics and Computer Sciences, University of Montevallo, Montevallo, Alabama, USA

AUTHOR ORCIDs

Adedotun Adedayo Arogundade  <http://orcid.org/0009-0006-3893-4716>

Leonora S. Bittleston  <http://orcid.org/0000-0003-4007-5405>

FUNDING

Funder	Grant(s)	Author(s)
National Science Foundation	OIA-1757324	Leonora S. Bittleston Sven Buerki

AUTHOR CONTRIBUTIONS

Adedotun Adedayo Arogundade, Formal analysis, Investigation, Methodology, Visualization, Writing – original draft, Writing – review and editing | Carlos Dave C. Dumaquit, Data curation, Resources, Writing – review and editing | Anthony Melton, Data curation, Resources, Writing – review and editing | Sven Buerki, Data curation, Funding acquisition, Methodology, Resources, Supervision, Writing – review and editing | Leonora S. Bittleston, Conceptualization, Funding acquisition, Investigation, Methodology, Resources, Supervision, Validation, Writing – review and editing

DATA AVAILABILITY

All host-cleaned microbial sequences that were pulled out of the Illumina sequences have been deposited in the NCBI BioProject database under accession number [PRJNA1256099](https://doi.org/10.1093/bioinformatics/bty519). The complete analysis pipeline, data processing workflows, and reproducibility documentation, isare publicly available on Zenodo at: (<https://doi.org/10.5281/zenodo.16601103>).

ADDITIONAL FILES

The following material is available [online](#).

Supplemental Material

Supplemental material (Spectrum02198-25-s0001.docx). Details of other computational tools that we also used with our data; Fig. S1.

REFERENCES

- Andersen SB, Schluter J. 2021. A metagenomics approach to investigate microbiome sociobiology. *Proc Natl Acad Sci USA* 118:e2100934118. <https://doi.org/10.1073/pnas.2100934118>
- Gulati M, Plosky B. 2020. As the microbiome moves on toward mechanism. *Mol Cell* 78:567. <https://doi.org/10.1016/j.molcel.2020.05.006>
- Bharti R, Grimm DG. 2021. Current challenges and best-practice protocols for microbiome analysis. *Brief Bioinform* 22:178–193. <https://doi.org/10.1093/bib/bbz215>
- Berg G, Rybakova D, Fischer D, Cernava T, Vergès M-CC, Charles T, Chen X, Cocolin L, Eversole K, Corral GH, et al. 2020. Microbiome definition revisited: old concepts and new challenges. *Microbiome* 8:103. <https://doi.org/10.1186/s40168-020-00875-0>
- Nayfach S, Shi ZJ, Seshadri R, Pollard KS, Kyrpides NC. 2019. New insights from uncultivated genomes of the global human gut microbiome. *Nature* 568:505–510. <https://doi.org/10.1038/s41586-019-1058-x>
- Browne HP, Forster SC, Anonye BO, Kumar N, Neville BA, Stares MD, Goulding D, Lawley TD. 2016. Culturing of “unculturable” human microbiota reveals novel taxa and extensive sporulation. *Nature* 533:543–546. <https://doi.org/10.1038/nature17645>
- Lin H-H, Liao Y-C. 2016. Accurate binning of metagenomic contigs via automated clustering sequences using information of genomic signatures and marker genes. *Sci Rep* 6:24175. <https://doi.org/10.1038/srep24175>
- Wang Z, Wang Z, Lu YY, Sun F, Zhu S. 2019. SolidBin: improving metagenome binning with semi-supervised normalized cut. *Bioinformatics* 35:4229–4238. <https://doi.org/10.1093/bioinformatics/btz253>
- Yu G, Jiang Y, Wang J, Zhang H, Luo H. 2018. BMC3C: binning metagenomic contigs using codon usage, sequence composition and read coverage. *Bioinformatics* 34:4172–4179. <https://doi.org/10.1093/bioinformatics/bty519>
- Trivedi P, Leach JE, Tringe SG, Sa T, Singh BK. 2020. Plant-microbiome interactions: from community assembly to plant health. *Nat Rev Microbiol* 18:607–621. <https://doi.org/10.1038/s41579-020-0412-1>
- Carlström CI, Field CM, Bortfeld-Miller M, Müller B, Sunagawa S, Vorholt JA. 2019. Synthetic microbiota reveal priority effects and keystone strains in the *Arabidopsis* phyllosphere. *Nat Ecol Evol* 3:1445–1454. <https://doi.org/10.1038/s41559-019-0994-z>
- Morella NM, Weng F-H, Joubert PM, Metcalf CJE, Lindow S, Koskella B. 2020. Successive passaging of a plant-associated microbiome reveals robust habitat and host genotype-dependent selection. *Proc Natl Acad Sci USA* 117:1148–1159. <https://doi.org/10.1073/pnas.1908600116>
- Vannier N, Mesny F, Getzke F, Chesneau G, Dethier L, Ordon J, Thiergart T, Hacquard S. 2023. Genome-resolved metatranscriptomics reveals conserved root colonization determinants in a synthetic microbiota. *Nat Commun* 14:8274. <https://doi.org/10.1038/s41467-023-43688-z>
- Meyer KM, Porch R, Muscettola IE, Vasconcelos ALS, Sherman JK, Metcalf CJE, Lindow SE, Koskella B. 2022. Plant neighborhood shapes diversity and reduces interspecific variation of the phyllosphere microbiome. *ISME J* 16:1376–1387. <https://doi.org/10.1038/s41396-021-01184-6>
- Regalado J, Lundberg DS, Deusch O, Kersten S, Karasov T, Poersch K, Shirsekar G, Weigel D. 2020. Combining whole-genome shotgun sequencing and rRNA gene amplicon analyses to improve detection of microbe-microbe interaction networks in plant leaves. *ISME J* 14:2116–2130. <https://doi.org/10.1038/s41396-020-0665-8>
- Bandla A, Pavagadhi S, Sridhar Sudarshan A, Poh MCH, Swarup S. 2020. 910 metagenome-assembled genomes from the phytobiomes of three urban-farmed leafy Asian greens. *Sci Data* 7:278. <https://doi.org/10.1038/s41597-020-00617-9>

17. Roman-Reyna V, Pinili D, Borja FN, Quibod IL, Groen SC, Alexandrov N, Mauleon R, Oliva R. 2020. Characterization of the leaf microbiome from whole-genome sequencing data of the 3,000 rice genomes project. *Rice (N Y)* 13:72. <https://doi.org/10.1186/s12284-020-00432-1>
18. Remington TE, Deibert PA, Hanser SE, Davis DM, Robb LA, Welty JL. 2021. Sagebrush conservation strategy—Challenges to sagebrush conservation, 2020–1125. Open-file report. U.S. Geological Survey
19. Melton AE, Novak SJ, Buerki S. 2024. Utilizing a comparative approach to assess genome evolution during diploidization in *Artemisia tridentata*, a keystone species of western North America. *Am J Bot* 111:e16353. <https://doi.org/10.1002/ajb2.16353>
20. Melton AE, Child AW, Beard RS, Dumaguit CDC, Forbey JS, Germino M, de Graaff M-A, Kliskey A, Leitch IJ, Martinez P, Novak SJ, Pellicer J, Richardson BA, Self D, Serpe M, Buerki S. 2022. A haploid pseudochromosome genome assembly for a keystone sagebrush species of western North American rangelands. *G3 (Bethesda)* 12:jkac122. <https://doi.org/10.1093/g3journal/jkac122>
21. Shipley LA, Davila TB, Thines NJ, Elias BA. 2006. Nutritional requirements and diet choices of the pygmy rabbit (*Brachylagus idahoensis*): a sagebrush specialist. *J Chem Ecol* 32:2455–2474. <https://doi.org/10.1007/s10886-006-9156-2>
22. Miller RF, Knick ST, Pyke DA, Hanser SE, Wisdom MJ, Hild AL. 2011. Characteristics of sagebrush habitats and limitations to long-term conservation, p 145–184. In Knick ST, Connelly JW (ed), *Greater Sage-Grouse: ecology and conservation of a landscape species and its habitats*. Studies in Avian Biology (vol. 38), University of California Press, Berkeley, CA.
23. Rigge M, Homer C, Cleaves L, Meyer DK, Bunde B, Shi H, Xian G, Schell S, Bobo M. 2020. Quantifying Western U.S. rangelands as fractional components with multi-resolution remote sensing and *in situ* data. *Remote Sens* 12:412. <https://doi.org/10.3390/rs12030412>
24. Heil JA, Geisler M, Skinner M, DeFehr C, Arogundade AA, Finney B, Bittleston LS. 2025. Context-dependent effects of microbial inoculation on sagebrush seedlings during drought stress. *Phytobiomes Journal* 9:574–585. <https://doi.org/10.1094/PBIOMES-04-25-0029-R>
25. Heil JA, Simler - Williamson A, Striluk ML, Trawick D, Capezza R, DeFehr C, Osorio A, Finney B, Turner KG, Bittleston LS. 2025. Weather and leaf age separately contribute to temporal shifts in phyllosphere fungal community structure in sagebrush. *Ecosphere* 16:e70295. <https://doi.org/10.1002/ecs2.70295>
26. Bodenhausen N, Somerville V, Desirò A, Walser J-C, Borghi L, van der Heijden MGA, Schlaeppi K. 2019. *Petunia*- and *Arabidopsis*-specific root microbiota responses to phosphate supplementation. *Phytobiomes Journal* 3:112–124. <https://doi.org/10.1094/PBIOMES-12-18-0057-R>
27. Xu J, Zhang Y, Zhang P, Trivedi P, Riera N, Wang Y, Liu X, Fan G, Tang J, Coletta-Filho HD, et al. 2018. The structure and function of the global citrus rhizosphere microbiome. *Nat Commun* 9:4894. <https://doi.org/10.1038/s41467-018-07343-2>
28. Thacker ET, Gardner RD, Messmer TA, Guttery MR, Dahlgren DK. 2012. Using gas chromatography to determine winter diets of greater sage - grouse in Utah. *J Wildl Manag* 76:588–592. <https://doi.org/10.1002/jwm.g.273>
29. Brooks JP, Edwards DJ, Harwich MD Jr, Rivera MC, Fettweis JM, Serrano MG, Reris RA, Sheth NU, Huang B, Girerd P, Vaginal Microbiome Consortium, Strauss JF 3rd, Jefferson KK, Buck GA. 2015. The truth about metagenomics: quantifying and counteracting bias in 16S rRNA studies. *BMC Microbiol* 15:66. <https://doi.org/10.1186/s12866-015-0351-6>
30. Luo C, Tsementzi D, Kyrpidis NC, Konstantinidis KT. 2012. Individual genome assembly from complex community short-read metagenomic datasets. *ISME J* 6:898–901. <https://doi.org/10.1038/ismej.2011.147>
31. Su P, Wicaksono WA, Li C, Michl K, Berg G, Wang D, Xiao Y, Huang R, Kang H, Zhang D, Cernava T, Liu Y. 2022. Recovery of metagenome-assembled genomes from the phyllosphere of 110 rice genotypes. *Sci Data* 9:254. <https://doi.org/10.1038/s41597-022-01320-7>
32. Barron R, Martinez P, Serpe M, Buerki S. 2020. Development of an *in vitro* method of propagation for *Artemisia tridentata* subsp. *tridentata* to support genome sequencing and genotype-by-environment research. *Plants (Basel)* 9:1717. <https://doi.org/10.3390/plants9121717>
33. Melton AE, Moran K, Martinez P, Ellestad P, Milliken E, Morales W, Child AW, Richardson BA, Serpe M, Novak SJ, Buerki S. 2023. A genotype × environment experiment reveals contrasting response strategies to drought between populations of a keystone species (*Artemisia tridentata*; Asteraceae). *Plant Environ Interact* 4:201–214. <https://doi.org/10.1002/pei3.10119>
34. Dumaguit CD. 2023. Genomic underpinning of drought-associated phenotypes in a keystone shrub species of western North America, Boise State University Theses and Dissertations. <https://doi.org/10.18122/td.2101.boisestate>
35. Andrews S. 2010. FastQC a quality control tool for high throughput sequence data. Scientific Research Publishing. Available from: <https://www.scirp.org/reference/referencespapers?referenceid=4024153>. Retrieved 16 Feb 2026.
36. Bolger AM, Lohse M, Usadel B. 2014. Trimmomatic: a flexible trimmer for Illumina sequence data. *Bioinformatics* 30:2114–2120. <https://doi.org/10.1093/bioinformatics/btu170>
37. Langmead B, Salzberg SL. 2012. Fast gapped-read alignment with Bowtie 2. *Nat Methods* 9:357–359. <https://doi.org/10.1038/nmeth.1923>
38. BMTagger: best match tagger for removing human reads from metagenomics datasets. Policy Commons. Available from: <https://policycommons.net/artifacts/30861577/bmtagger/31761403/view>. Retrieved 17 Feb 2026.
39. Arkin AP, Cottingham RW, Henry CS, Harris NL, Stevens RL, Maslov S, Dehal P, Ware D, Perez F, Canon S, et al. 2018. KBase: The United States department of energy systems biology knowledgebase. *Nat Biotechnol* 36:566–569. <https://doi.org/10.1038/nbt.4163>
40. Menzel P, Ng KL, Krogh A. 2016. Fast and sensitive taxonomic classification for metagenomics with Kaiju. *Nat Commun* 7:11257. <https://doi.org/10.1038/ncomms11257>
41. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. *Genome Biol* 20:257. <https://doi.org/10.1186/s13059-019-1891-0>
42. Buchfink B, Xie C, Huson DH. 2015. Fast and sensitive protein alignment using DIAMOND. *Nat Methods* 12:59–60. <https://doi.org/10.1038/nmeth.3176>
43. Nurk S, Meleshko D, Korobeynikov A, Pevzner PA. 2017. metaSPAdes: a new versatile metagenomic assembler. *Genome Res* 27:824–834. <https://doi.org/10.1101/gr.213959.116>
44. Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. 2015. MEGAHIT: an ultra-fast single-node solution for large and complex metagenomics assembly via succinct de Bruijn graph. *Bioinformatics* 31:1674–1676. <https://doi.org/10.1093/bioinformatics/btv033>
45. Peng Y, Leung HCM, Yiu SM, Chin FYL. 2012. IDBA-UD: a *de novo* assembler for single-cell and metagenomic sequencing data with highly uneven depth. *Bioinformatics* 28:1420–1428. <https://doi.org/10.1093/bioinformatics/bts174>
46. Wu Y-W, Simmons BA, Singer SW. 2016. MaxBin 2.0: an automated binning algorithm to recover genomes from multiple metagenomic datasets. *Bioinformatics* 32:605–607. <https://doi.org/10.1093/bioinformatics/btv638>
47. Kang DD, Li F, Kirton E, Thomas A, Egan R, An H, Wang Z. 2019. MetaBAT 2: an adaptive binning algorithm for robust and efficient genome reconstruction from metagenome assemblies. *PeerJ* 7:e7359. <https://doi.org/10.7717/peerj.7359>
48. Alneberg J, Bjarnason BS, de Bruijn I, Schirmer M, Quick J, Ijaz UZ, Lahti L, Loman NJ, Andersson AF, Quince C. 2014. Binning metagenomic contigs by coverage and composition. *Nat Methods* 11:1144–1146. <https://doi.org/10.1038/nmeth.3103>
49. Sieber CMK, Probst AJ, Sharrar A, Thomas BC, Hess M, Tringe SG, Banfield JF. 2018. Recovery of genomes from metagenomes via a dereplication, aggregation and scoring strategy. *Nat Microbiol* 3:836–843. <https://doi.org/10.1038/s41564-018-0171-1>
50. Arumugam K, Bessabab I, Haryono MAS, Liu X, Zuniga-Montanez RE, Roy S, Qiu G, Drautz-Moses DI, Law YY, Wuertz S, Lauro FM, Huson DH, Williams RBH. 2021. Recovery of complete genomes and non-chromosomal replicons from activated sludge enrichment microbial communities with long read metagenome sequencing. *NPJ Biofilms Microbiomes* 7:23. <https://doi.org/10.1038/s41522-021-00196-6>
51. Parks DH, Imelfort M, Skennerton CT, Hugenholtz P, Tyson GW. 2015. CheckM: assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res* 25:1043–1055. <https://doi.org/10.1101/gr.186072.114>
52. Chaumeil P-A, Mussig AJ, Hugenholtz P, Parks DH. 2019. GTDB-Tk: a toolkit to classify genomes with the Genome Taxonomy Database. *Bioinformatics* 36:1925–1927. <https://doi.org/10.1093/bioinformatics/btz848>

53. Almeida A, Nayfach S, Boland M, Strozzi F, Beracochea M, Shi ZJ, Pollard KS, Sakharova E, Parks DH, Hugenholtz P, Segata N, Kyrpides NC, Finn RD. 2021. A unified catalog of 204,938 reference genomes from the human gut microbiome. *Nat Biotechnol* 39:105–114. <https://doi.org/10.1038/s41587-020-0603-3>
54. Yuan C, Lei J, Cole J, Sun Y. 2015. Reconstructing 16S rRNA genes in metagenomic data. *Bioinformatics* 31:i35–43. <https://doi.org/10.1093/bioinformatics/btv231>
55. Eddy SR. 2011. Accelerated profile HMM searches. *PLoS Comput Biol* 7:e1002195. <https://doi.org/10.1371/journal.pcbi.1002195>
56. Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil P-A, Hugenholtz P. 2018. A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of life. *Nat Biotechnol* 36:996–1004. <https://doi.org/10.1038/nbt.4229>
57. Matsen FA, Kodner RB, Armbrust EV. 2010. pplacer: linear time maximum-likelihood and Bayesian phylogenetic placement of sequences onto a fixed reference tree. *BMC Bioinformatics* 11:538. <https://doi.org/10.1186/1471-2105-11-538>
58. Shaffer M, Borton MA, McGivern BB, Zayed AA, La Rosa SL, Solden LM, Liu P, Narrow AB, Rodríguez-Ramos J, Bolduc B, Gazitúa MC, Daly RA, Smith GJ, Vik DR, Pope PB, Sullivan MB, Roux S, Wrighton KC. 2020. DRAM for distilling microbial metabolism to automate the curation of the microbiome function. *Nucleic Acids Res* 48:8883–8900. <https://doi.org/10.1093/nar/gkaa621>
59. Hyatt D, Chen G-L, Locascio PF, Land ML, Larimer FW, Hauser LJ. 2010. Prodigal: prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. <https://doi.org/10.1186/1471-2105-11-119>
60. Aramaki T, Blanc-Mathieu R, Endo H, Ohkubo K, Kanehisa M, Goto S, Ogata H. 2020. KofamKOALA: KEGG Ortholog assignment based on profile HMM and adaptive score threshold. *Bioinformatics* 36:2251–2252. <https://doi.org/10.1093/bioinformatics/btz859>
61. Ye Y, Doak TG. 2009. A parsimony approach to biological pathway reconstruction/inference for genomes and metagenomes. *PLoS Comput Biol* 5:e1000465. <https://doi.org/10.1371/journal.pcbi.1000465>
62. Kanehisa M, Goto S. 2000. KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* 28:27–30. <https://doi.org/10.1093/nar/28.1.27>
63. Swor K, Satyal P, Timsina S, Setzer WN. 2022. Chemical composition and terpenoid enantiomeric distribution of the essential oil of *Artemisia tridentata* subsp. *tridentata* from southwestern Idaho. *Nat Prod Commun* 17. <https://doi.org/10.1177/1934578X221117417>
64. Zheljazkov VD, Cantrell CL, Jeliakova EA, Astatkie T, Schlegel V. 2022. Essential oil yield, composition, and bioactivity of sagebrush species in the Bighorn Mountains. *Plants (Basel)* 11:1228. <https://doi.org/10.3390/plants11091228>
65. R Core Team. 2022. R a language and environment for statistical computing. Scientific Research Publishing. Available from: <https://www.scrip.org/reference/referencespapers?referenceid=3456808>. Retrieved 17 Feb 2026.
66. Wickham H. 2025. reshape2: flexibly reshape data: a reboot of the reshape package (1.4.5)
67. Oksanen J, Simpson GL, Blanchet FG, Kindt R, Legendre P, Minchin PR, O'Hara RB, Solymos P, Stevens MHH, Szoecs E, et al. 2001. vegan: community ecology package
68. Wickham H. 2009. ggplot2: elegant graphics for data analysis. Springer, New York, NY. <https://link.springer.com/10.1007/978-0-387-98141-3>.
69. Gao Y, Luo H, Lyu H, Yang SM, Ziusuf S, Huang S, Liu Y-X. 2025. Benchmarking short-read metagenomics tools for removing host contamination. *Gigascience* 14:giaf004. <https://doi.org/10.1093/gigascience/giaf004>
70. Xu R, Rajeev S, Salvador LCM. 2023. The selection of software and database for metagenomics sequence analysis impacts the outcome of microbial profiling and pathogen detection. *PLoS One* 18:e0284031. <https://doi.org/10.1371/journal.pone.0284031>
71. Lucaci R, Pelikan C, Gerner SM, Zioutis C, Köstlbacher S, Marx H, Herbold CW, Schmidt H, Rattei T. 2019. A bioinformatics guide to plant microbiome analysis. *Front Plant Sci* 10:1313. <https://doi.org/10.3389/fpls.2019.01313>
72. Borisov VB, Verkhovsky MI. 2015. Oxygen as acceptor. *EcoSal Plus* 6. <https://doi.org/10.1128/ecosalplus.ESP-0012-2015>
73. Borisov VB, Siletsky SA, Paiardini A, Hoogewijs D, Forte E, Giuffrè A, Poole RK. 2021. Bacterial oxidases of the cytochrome *bd* family: redox enzymes of unique structure, function, and utility as drug targets. *Antioxid Redox Signal* 34:1280–1318. <https://doi.org/10.1089/ars.2020.8039>
74. Borisov VB, Forte E, Siletsky SA, Arese M, Davletshin AI, Sarti P, Giuffrè A. 2015. Cytochrome *bd* protects bacteria against oxidative and nitrosative stress: a potential target for next-generation antimicrobial agents. *Biochim Biophys Acta* 185:565–575. <https://doi.org/10.1134/S0006297915050077>
75. Giuffrè A, Borisov VB, Arese M, Sarti P, Forte E. 2014. Cytochrome *bd* oxidase and bacterial tolerance to oxidative and nitrosative stress. *Biochim Biophys Acta* 1837:1178–1187. <https://doi.org/10.1016/j.bbabi.2014.01.016>
76. Parks DH, Rinke C, Chuvochina M, Chaumeil P-A, Woodcroft BJ, Evans PN, Hugenholtz P, Tyson GW. 2017. Recovery of nearly 8,000 metagenome-assembled genomes substantially expands the tree of life. *Nat Microbiol* 2:1533–1542. <https://doi.org/10.1038/s41564-017-0012-7>
77. Howe AC, Jansson JK, Malfatti SA, Tringe SG, Tiedje JM, Brown CT. 2014. Tackling soil diversity with the assembly of large, complex metagenomes. *Proc Natl Acad Sci USA* 111:4904–4909. <https://doi.org/10.1073/pnas.1402564111>
78. Soto-Giron MJ, Kim J-N, Schott E, Tahmin C, Ishoe T, Mincer TJ, DeWalt J, Toledo G. 2021. The Edible Plant Microbiome represents a diverse genetic reservoir with functional potential in the human host. *Sci Rep* 11:24017. <https://doi.org/10.1038/s41598-021-03334-4>
79. Carrión O, Gibson L, Elias DMO, McNamara NP, van Alen TA, Op den Camp HJM, Supramaniam CV, McGenity TJ, Murrell JC. 2020. Diversity of isoprene-degrading bacteria in phyllosphere and soil communities from a high isoprene-emitting environment: a Malaysian oil palm plantation. *Microbiome* 8:81. <https://doi.org/10.1186/s40168-020-00860-7>
80. Orellana LH, Krüger K, Sidhu C, Amann R. 2023. Comparing genomes recovered from time-series metagenomes using long- and short-read sequencing technologies. *Microbiome* 11:105. <https://doi.org/10.1186/s40168-023-01557-3>
81. Shi Y, Wang G, Lau H-H, Yu J. 2022. Metagenomic sequencing for microbial DNA in human samples: emerging technological advances. *Int J Mol Sci* 23:2181. <https://doi.org/10.3390/ijms23042181>
82. Bulgarelli D, Garrido-Oter R, Münch PC, Weiman A, Dröge J, Pan Y, McHardy AC, Schulze-Lefert P. 2015. Structure and function of the bacterial root microbiota in wild and domesticated barley. *Cell Host Microbe* 17:392–403. <https://doi.org/10.1016/j.chom.2015.01.011>
83. Qin J, Li R, Raes J, Arumugam M, Burgdorf KS, Manichanh C, Nielsen T, Pons N, Levenez F, Yamada T, et al. 2010. A human gut microbial gene catalogue established by metagenomic sequencing. *Nature* 464:59–65. <https://doi.org/10.1038/nature08821>
84. Mantelin S, Desbrosses G, Larcher M, Tranbarger TJ, Cleyet-Marel J-C, Touraine B. 2006. Nitrate-dependent control of root architecture and N nutrition are altered by a plant growth-promoting *Phyllobacterium* sp. *Planta* 223:591–603. <https://doi.org/10.1007/s00425-005-0106-y>
85. Larcher M, Muller B, Mantelin S, Rapior S, Cleyet-Marel J-C. 2003. Early modifications of *Brassica napus* root system architecture induced by a plant growth-promoting *Phyllobacterium* strain. *New Phytol* 160:119–125. <https://doi.org/10.1046/j.1469-8137.2003.00862.x>
86. Bresson J, Varoquaux F, Bontpart T, Touraine B, Vile D. 2013. The PGPR strain *Phyllobacterium brassicacearum* STM196 induces a reproductive delay and physiological changes that result in improved drought tolerance in *Arabidopsis*. *New Phytol* 200:558–569. <https://doi.org/10.1111/nph.12383>
87. Khan AL, Al-Harrasi A, Al-Rawahi A, Al-Farsi Z, Al-Mamari A, Waqas M, Asaf S, Elyassi A, Mabood F, Shin J-H, Lee I-J. 2016. Endophytic fungi from Frankincense tree improves host growth and produces extracellular enzymes and indole acetic acid. *PLoS One* 11:e0158207. <https://doi.org/10.1371/journal.pone.0158207>
88. Khan AL, Waqas M, Kang S-M, Al-Harrasi A, Hussain J, Al-Rawahi A, Al-Khiziri S, Ullah I, Ali L, Jung H-Y, Lee I-J. 2014. Bacterial endophyte *Sphingomonas* sp. LK11 produces gibberellins and IAA and promotes tomato plant growth. *J Microbiol* 52:689–695. <https://doi.org/10.1007/s12275-014-4002-7>
89. Leys N, Ryngaert A, Bastiaens L, Verstraete W, Top EM, Springael D. 2004. Occurrence and phylogenetic diversity of *Sphingomonas* strains in soils contaminated with polycyclic aromatic hydrocarbons. *Appl Environ Microbiol* 70:1944–1955. <https://doi.org/10.1128/AEM.70.4.1944-1955.2004>
90. Lladó S, López-Mondéjar R, Baldrian P. 2017. Forest soil bacteria: diversity, involvement in ecosystem processes, and response to global

- change. *Microbiol Mol Biol Rev* 81:e00063-16. <https://doi.org/10.1128/MMBR.00063-16>
91. Kohl KD, Connelly JW, Dearing MD, Forbey JS. 2016. Microbial detoxification in the gut of a specialist avian herbivore, the Greater Sage-Grouse. *FEMS Microbiol Lett* 363:fnw144. <https://doi.org/10.1093/femsle/fnw144>
 92. Ottesen AR, Gorham S, Reed E, Newell MJ, Ramachandran P, Canida T, Allard M, Evans P, Brown E, White JR. 2016. Using a control to better understand phyllosphere microbiota. *PLoS One* 11:e0163482. <https://doi.org/10.1371/journal.pone.0163482>
 93. Agler MT, Ruhe J, Kroll S, Morhenn C, Kim S-T, Weigel D, Kemen EM. 2016. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biol* 14:e1002352. <https://doi.org/10.1371/journal.pbio.1002352>
 94. Whitham TG, Young WP, Martinsen GD, Gehring CA, Schweitzer JA, Shuster SM, Wimp GM, Fischer DG, Bailey JK, Lindroth RL, Woolbright S, Kuske CR. 2003. Community and ecosystem genetics: a consequence of the extended phenotype. *Ecology* 84:559–573. [https://doi.org/10.1890/0012-9658\(2003\)084\[0559:CAEGAC\]2.0.CO;2](https://doi.org/10.1890/0012-9658(2003)084[0559:CAEGAC]2.0.CO;2)
 95. Di Francesco A, Zajc J, Stenberg JA. 2023. *Aureobasidium* spp.: diversity, versatility, and agricultural utility. *Horticulturae* 9:59. <https://doi.org/10.3390/horticulturae9010059>
 96. Pinto C, Pinho D, Sousa S, Pinheiro M, Egas C, Gomes AC. 2014. Unravelling the diversity of grapevine microbiome. *PLoS One* 9:e85622. <https://doi.org/10.1371/journal.pone.0085622>
 97. Barata A, Malfeito-Ferreira M, Loureiro V. 2012. The microbial ecology of wine grape berries. *Int J Food Microbiol* 153:243–259. <https://doi.org/10.1016/j.ijfoodmicro.2011.11.025>
 98. Rai M, Wypij M, Ingle AP, Trzcińska-Wencel J, Golińska P. 2021. Emerging trends in Pullulan-based antimicrobial systems for various applications. *Int J Mol Sci* 22:13596. <https://doi.org/10.3390/ijms222413596>
 99. Di Francesco A, Baraldi E. 2021. How siderophore production can influence the biocontrol activity of *Aureobasidium pullulans* against *Monilinia laxa* on peaches. *Biol Control* 152:104456. <https://doi.org/10.1016/j.biocontrol.2020.104456>
 100. Rajeeva G, Ranjan S, Monika G, Manogya KG. 2010. *Aureobasidium pullulans*, an economically important polymorphic yeast with special reference to pullulan. *Afr J Biotechnol* 9:7989–7997. <https://doi.org/10.5897/AJB10.948>
 101. Bhardwaj G, Shah R, Joshi B, Patel P. 2017. *Klebsiella pneumoniae* VRE36 as a PGPR isolated from *Saccharum officinarum* cultivar Co99004. *J App Biol Biotech*:047–052. <https://doi.org/10.7324/JABB.2017.50108>
 102. Kumar G, Lal S, Maurya SK, Bhattacharjee AK, Chaudhary P, Gangola S, Rajan S. 2021. Exploration of *Klebsiella pneumoniae* M6 for paclitaxel degradation, plant growth attributes, and biocontrol action under subtropical ecosystem. *PLoS One* 16:e0261338. <https://doi.org/10.1371/journal.pone.0261338>