



OPEN ACCESS

EDITED BY

Long Jin,
Nanjing Forestry University, China

REVIEWED BY

Lauren M. Lui,
Berkeley Lab (DOE), United States
Chang Soo Lee,
Nakdonggang National Institute of Biological
Resources, Republic of Korea

*CORRESPONDENCE

Bahareh Sorouri
✉ bsorouri@uci.edu

SPECIALTY SECTION

This article was submitted to
Frontiers in Microbiology Evolutionary and
Genomic Microbiology,
a section of the journal
Frontiers in Microbiology

RECEIVED 17 January 2023

ACCEPTED 29 March 2023

PUBLISHED 17 April 2023

CITATION

Sorouri B, Rodriguez CI, Gaut BS and
Allison SD (2023) Variation in *Sphingomonas*
traits across habitats and phylogenetic clades.
Front. Microbiol. 14:1146165.
doi: 10.3389/fmicb.2023.1146165

COPYRIGHT

© 2023 Sorouri, Rodriguez, Gaut and Allison.
This is an open-access article distributed under
the terms of the [Creative Commons Attribution
License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use, distribution or
reproduction in other forums is permitted,
provided the original author(s) and the
copyright owner(s) are credited and that the
original publication in this journal is cited, in
accordance with accepted academic practice.
No use, distribution or reproduction is
permitted which does not comply with
these terms.

Variation in *Sphingomonas* traits across habitats and phylogenetic clades

Bahareh Sorouri^{1*}, Cynthia I. Rodriguez¹, Brandon S. Gaut¹ and Steven D. Allison^{1,2}

¹Department of Ecology and Evolutionary Biology, University of California Irvine, Irvine, CA, United States, ²Department of Earth System Science, University of California Irvine, Irvine, CA, United States

Whether microbes show habitat preferences is a fundamental question in microbial ecology. If different microbial lineages have distinct traits, those lineages may occur more frequently in habitats where their traits are advantageous. *Sphingomonas* is an ideal bacterial clade in which to investigate how habitat preference relates to traits because these bacteria inhabit diverse environments and hosts. Here we downloaded 440 publicly available *Sphingomonas* genomes, assigned them to habitats based on isolation source, and examined their phylogenetic relationships. We sought to address whether: (1) there is a relationship between *Sphingomonas* habitat and phylogeny, and (2) whether there is a phylogenetic correlation between key, genome-based traits and habitat preference. We hypothesized that *Sphingomonas* strains from similar habitats would cluster together in phylogenetic clades, and key traits that improve fitness in specific environments should correlate with habitat. Genome-based traits were categorized into the Y-A-S trait-based framework for high growth yield, resource acquisition, and stress tolerance. We selected 252 high quality genomes and constructed a phylogenetic tree with 12 well-defined clades based on an alignment of 404 core genes. *Sphingomonas* strains from the same habitat clustered together within the same clades, and strains within clades shared similar clusters of accessory genes. Additionally, key genome-based trait frequencies varied across habitats. We conclude that *Sphingomonas* gene content reflects habitat preference. This knowledge of how environment and host relate to phylogeny may also help with future functional predictions about *Sphingomonas* and facilitate applications in bioremediation.

KEYWORDS

Sphingomonas, pangenome, habitats, traits, phylogenetics

1. Introduction

Bacteria occur in a wide diversity of habitats, but the factors that control habitat preference are unclear (Fierer and Jackson, 2006; Martiny et al., 2006; Merino et al., 2019). Given that habitats vary in their abiotic and biotic conditions, different habitats may select for different organismal traits (Noble and Slatyer, 1977). These traits can be phylogenetically conserved (Martiny et al., 2013; Dolan et al., 2017; Isobe et al., 2019, 2020), horizontally transferred (Ochman et al., 2000), and reflect trade-offs underlying life-history strategies. For environmental microbes, one way to organize these trade-offs is the Y-A-S framework,

which posits that bacterial life-history strategies are driven by tradeoffs in resource allocation to growth Yield, resource Acquisition, and Stress tolerance responses (Malik et al., 2020). Investigating functional traits related to the Y-A-S strategies has the potential to yield insights into factors that affect the distributions of microbial taxa.

Sphingomonas is an excellent bacterial genus to investigate the distribution of habitat preference traits because it is found in a wide range of habitats. Within the Proteobacteria phylum, the *Sphingomonas* genus contains gram-negative, strictly aerobic, chemoheterotrophic, yellow-pigmented bacteria that possess glycosphingolipids in their cell envelope (Yabuuchi et al., 1990; Balkwill et al., 2006). *Sphingomonas* species have been isolated from soils, plant roots, water distribution systems, human samples, and hospital machines (White et al., 1996; Leys et al., 2004). Some species cause animal disease, while others are antagonistic toward phytopathogenic fungi that infect commercially important plants (White et al., 1996). Additionally, *Sphingomonas* species have also been used on the International Space Station to aid the extraction of rare earth elements (Cockell et al., 2020). On planet Earth, *Sphingomonas* serves as biocatalyst for bioremediation and can be found in soils that are contaminated with pollutants (Leys et al., 2004). Understanding the distribution of *Sphingomonas* is especially important because with appropriate management strategies, this lineage can be a tool to clean up polluted environments (Onder Erguven and Demirci, 2019). Furthermore, *Sphingomonas* is able to degrade cellulose and hemicellulose and is therefore involved in organic carbon decomposition (Koskinen et al., 2000). Hence, the distribution and functional abilities of *Sphingomonas* make it an ideal genus for investigating phylogenetic histories of habitat preference traits.

Despite the potential importance and widespread distribution of *Sphingomonas* species, there has not yet been a comprehensive, in-depth study of the comparative genomics and phylogenetics of the genus from a trait-based perspective. Most studies thus far look at the distribution and phylogeny of select genomes from 16S rRNA perspective, and often do not consider genome-based traits (Leung et al., 1999; Leys et al., 2004; Asaf et al., 2020). Moreover, the *Sphingomonas* genus classification is still evolving; *Sphingomonas* has five sub-genus classifications, and although additional strains continue to be identified, it is difficult to place them into specific clades (Takeuchi et al., 2001; Jogler et al., 2013; Asaf et al., 2020). Additionally, some *Sphingomonas* species have been shown to improve plant growth during stressful drought and salinity conditions (Halo et al., 2015; Asaf et al., 2017). Currently, there are knowledge gaps in the literature with respect to *Sphingomonas* phylogenetics, taxonomy, and genome mapping in the context of stress tolerance and bioremediation (Asaf et al., 2020). Therefore, it is useful to explore the phylogenomics of *Sphingomonas* from a whole-genome and trait-based perspective. Since *Sphingomonas* has important bioremediation qualities, understanding the genetics and distributions of these traits can provide preliminary knowledge toward harnessing *Sphingomonas* to rehabilitate natural habitats (Schmidt et al., 1992).

The knowledge of how environment and host correspond to traits may also help with future functional predictions. In this study, we downloaded over 400 available *Sphingomonas* sequences from public databases, assigned them to a habitat based on where they were isolated, and assessed their phylogenetic relationships.

With this information, we sought to address two questions. First, are there significant relationships between habitat and phylogeny? Second, do key, genome-based traits demonstrate phylogenetic clustering and correlate with habitat preference?

We used the genome-based traits as proxies for the Y-A-S life history categories (Figure 1; Malik et al., 2020). For growth yield, we investigated the distribution of genes underlying amino acid related enzymes, lipid biosynthesis proteins, and lipopolysaccharide biosynthesis proteins. Genes for carbohydrate-active enzymes (CAZymes) reflected resource acquisition strategies. Finally, for stress tolerance we explored genes associated with chaperones, folding catalysts, the prokaryotic defense system, as well as peptidoglycan biosynthesis and degradation proteins. Collectively, these traits underlie habitat preference. We hypothesize that *Sphingomonas* strains from similar habitats will cluster together in phylogenetic clades. Furthermore, key traits that improve fitness in specific environments should correlate with the isolation habitat. For example, CAZyme genes should be most prevalent in genomes of *Sphingomonas* associated with plants, and prokaryotic defense system genes would be the highest in *Sphingomonas* genomes found at locations with a contaminant. Ultimately, our findings will improve the understanding of *Sphingomonas* distribution across habitats, as well as illuminate the link between habitat preference and life history strategies.

2. Materials and methods

2.1. Library collection and curation

We downloaded 440 publicly available *Sphingomonas* genomes from the PATRIC database on 31 July 2020 (Wattam et al., 2014) and used the metadata for each strain to identify the isolation source (Table 1). The sequences were categorized by their isolation source and assigned to one of eight groups based on the strain description: animal ($n = 10$), clinical ($n = 43$), contaminated site ($n = 13$), industrial ($n = 13$), environmental ($n = 54$), plant ($n = 68$), water ($n = 34$), and other ($n = 17$; Table 1). More specifically, strains in the animal category were isolated from living, non-human sources. Strains in the clinical category came from hospital settings and included bodily samples from human beings, like blood. Any strain with the word “contaminated” in the description was placed in the contaminated site category. The environment category consisted of strains from abiotic, outdoor sources that were not water-based, like soils. The industrial category included samples from bioreactors, mines, and wastewater facilities (which contained the key phrase “activated sludge” in the description). Strains isolated from hosts in the plant kingdom were placed in the plant category; these strains were isolated from different plant parts such as the seed, root, stem, and leaf. The water category consisted of strains isolated from a water source and sediments that did not include “contaminated” in the description. Finally, strains that could not be assigned to one of the previous 7 distinct groups were placed in the other category, such as samples from lichens and dust (Table 1). Genomes with unspecified isolation sources were removed from our analyses.

Next, we checked the completeness of the genomes against the *Sphingomonadales* order using the BUSCO (Benchmarking

Universal Single Copy Orthologs) v4.1.4 program (Seppey et al., 2019). Genomes with a BUSCO completeness score of less than 95% were filtered out. We used the online QUAST (Quality Assessment Tool for Genome Assemblies) server v5.0.2 to investigate the quality of the remaining genomes (Gurevich et al., 2013). We also ran CheckM v1.2.2 against the *Sphingomonadales* and *Alphaproteobacteria* lineages to confirm the completeness of the remaining genomes and check for contamination (Parks et al., 2015).

From the initial genome library, 254 high quality *Sphingomonas* genome sequences remained for further analysis. These genomes consisted of 23 complete genomes and 231 fragmented genomes. All genomes were annotated with Prokka v1.14.6 default parameters and the *Sphingomonas* genus tag (Seemann, 2014). Core and accessory genes were identified with Roary v3.13.0 using a 50% blastp sequence identity, the default core gene identity of 99%, and a maximum gene cluster of 25,000,000 (Page et al., 2015). For comparison to the larger subset that included fragmented genomes, we also used Prokka and Roary to quantify the pangenome for just the 23 complete genomes (Seemann, 2014; Page et al., 2015).

2.2. Outgroup optimization

Zymomonas, *Rhizobium*, and *Rhodospirillum* are three closely related genera to *Sphingomonas* (Leys et al., 2004). To select the best outgroup or combinations of outgroups, we used Roary core gene counts. Specifically, we compared the core

TABLE 1 Classification descriptions of the isolation sources for *Sphingomonas* samples.

Classification	Description
Animal	Isolation source is from a living, non-human, non-plant source
Clinical	Samples from a hospital that caters toward human beings, includes blood samples
Contaminated site	Any sample that contains “contaminated” in the description
Environment	Isolated from abiotic, outdoor sources, such as soil
Industrial	Sources from bioreactors, mines, and wastewater facilities (contains “activated sludge”)
Plant	Samples isolated from the plant kingdom, can come from seeds, leaves, and roots
Water	Any sample from a water source and sediment that does not contain “contaminated” in the description
Other	Samples that do not fit in to the other categories, can be lichens, dust

TABLE 2 Pangenome analysis for the 252 *Sphingomonas* genomes.

Gene	Description	Frequency
Core genes	99% ≤ strains ≤ 100%	444
Soft core genes	95% ≤ strains ≤ 99%	304
Shell genes	15% ≤ strains ≤ 95%	4,070
Cloud genes	0% ≤ strains ≤ 15%	108,998
Total	0% ≤ strains ≤ 100%	113,186

genes of the *Sphingomonas*-only ingroup to the core genes of the ingroup with various combinations of outgroups. We also included *Escherichia coli* as a distantly related outgroup for further confirmation (Zhao et al., 2017). We selected *Rhodospirillum centenum* SW (GenBank Accession: CP000613) as an outgroup because it yielded a core gene count that was closest to the *Sphingomonas*-only ingroup. Furthermore, previous phylogenetic analysis (Leys et al., 2004) confirmed that *Rhodospirillum* is not part of the ingroup.

2.3. Reference tree visualization

We made a phylogenetic tree with core genes present in *Sphingomonas* genomes and the *Rhodospirillum* outgroup using methods from Rodriguez and Martiny (2020). In short, we ran Roary again with the same previously mentioned parameters for the *Sphingomonas* ingroup and *Rhodospirillum* outgroup. We identified 401 core genes and generated a bootstrapped maximum likelihood tree of the alignment with RAXML v8.2.12 with the PROTGAMMABLOSUM62 substitution model and 100 rapid bootstrap searches (Stamatakis, 2014). Two of the 254 *Sphingomonas* sequences were removed from the analyses since RAXML deemed them identical. Therefore, to minimize bias, we removed the duplicate sequences and re-ran Roary with the 252 *Sphingomonas* genomes to generate an alignment of 404 core genes (Supplementary material). We used the core gene alignment to construct a phylogenetic tree with RAXML and subsequently visualized the tree with the iTOL v6.5 interactive tool (Figure 3; Letunic and Bork, 2019).

2.4. Clade designation

We manually designated phylogenetic clades based on their divergence from the common ancestor. We marked the first clade by starting from the most distant, large monophyletic group. Subsequently, we moved along the tree until we came across another large, monophyletic group that was interpreted as another clade. Clades were defined in this manner until we identified a total of 12. Two strains that resembled an outgroup within two separate monophyletic clades were not included as part of the clade. We confirmed the clades and genome clusters by identifying pairwise average amino acid and nucleotide identities with the Enveomics tool (Rodriguez-R and Konstantinidis, 2016). Additionally, clades possessed a bootstrap identity of at least 86.

2.5. Genome-based traits

We quantified the abundances of genome-based traits involved in high growth yield, resource acquisition, and stress tolerance strategies. To identify the traits, we used the CAZy (Cantarel et al., 2009) and KEGG databases (Kanehisa and Goto, 2000). For CAZymes we determined glycoside hydrolase and carbohydrate binding module abundances. Specifically we identified cellulase and glycoside hydrolase genes from Prodigal protein annotations using dbCAN2, a metaserver based on the CAZy database (Hyatt et al.,

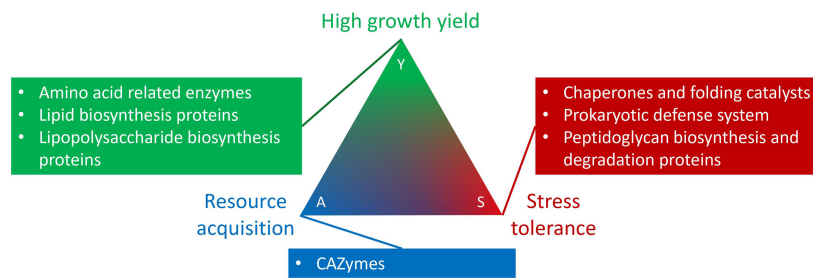


FIGURE 1

Genome-based trait groupings into the Y-A-S life history strategy framework developed by Malik et al. (2020).

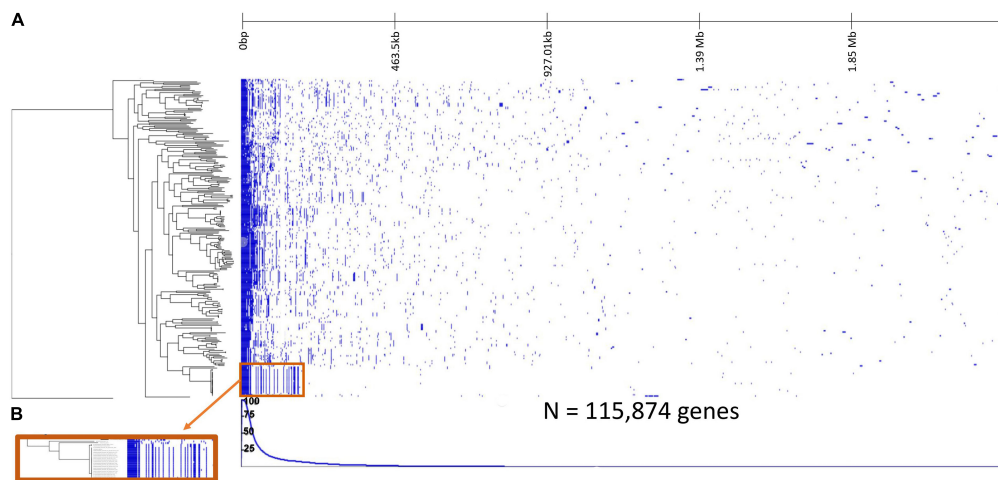


FIGURE 2

Pangenome analysis of 252 *Spingomonas* genomes and the *Rhodospirillum centenum* SW outgroup. (A) Gene presence-absence heatmap where vertical blue lines represent presence of a gene within rows corresponding to the *Spingomonas* genome, and white reflects gene absence. The line graph underneath indicates the percentage of strains possessing the corresponding gene. (B) Close-up of the gene patterns within a clade shows how clades contain similar gene clusters.

2010; Zhang et al., 2018). In our analysis, we only selected genes that were found with all three tools available on dbCAN2: HMMER, DIAMOND, and Hotpep. Additionally, we used the GhostKOALA v2.2 automatic annotation server to annotate the remaining genes based on KEGG Orthology (Kanehisa et al., 2016). We selected these genome-based traits for further analyses: lipopolysaccharide biosynthesis proteins ($n = 66$), lipid biosynthesis proteins ($n = 29$), amino acid related enzymes ($n = 52$), prokaryotic defense system ($n = 77$), peptidoglycan biosynthesis and degradation proteins ($n = 34$), and finally chaperones and folding catalysts ($n = 42$). These genes were grouped into the Y-A-S microbial life history trait-based framework developed by Malik et al. (2020) based on their role in growth yield, resource acquisition, and stress tolerance strategies (Supplementary material). We calculated the average relative abundance of each trait for each clade and visualized them with the ggpubr v0.4.0 R package (Kassambara, 2020).

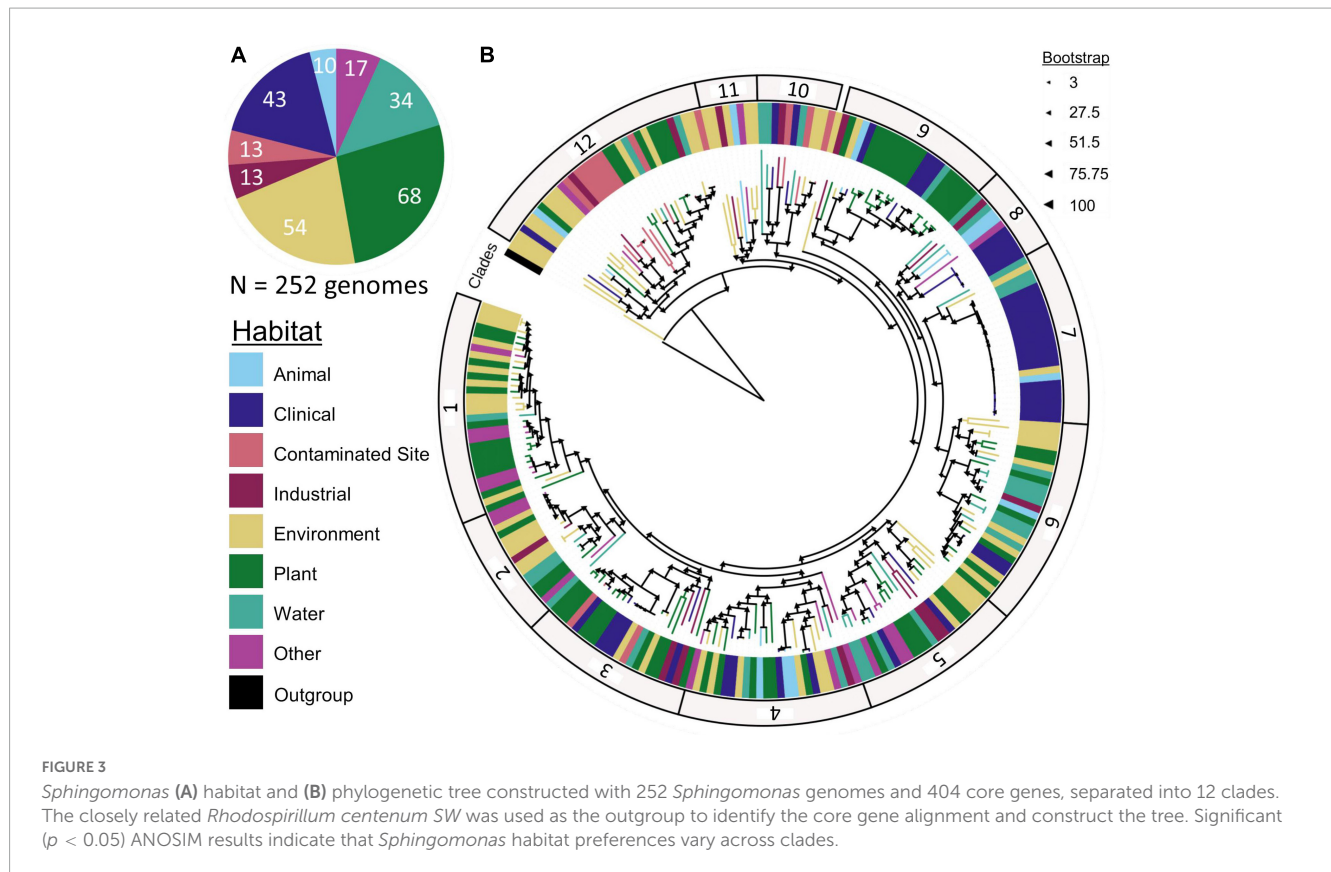
2.6. Statistical analyses

After quantifying gene abundances, we natural log transformed the gene counts of the genome-based traits. Subsequently, we

confirmed the normality of residuals using histograms and the Shapiro–Wilk tests, then ran Kruskal–Wallis rank sum tests to identify differences across habitats. We performed Kruskal–Wallis tests since not all the functional gene data were normally distributed. Additionally, we conducted phylogenetic generalized least squares (PGLS) statistical analyses to test whether there was an association between the habitat and the genome-based traits, independent of phylogenetic history (Mundry, 2014). We also used PGLS statistics to limit statistical bias by confirming if significant Kruskal–Wallis results were influenced by phylogenetic relatedness.

We used R v4.1.0 to run all the statistical analyses, and specially incorporated the “nlme,” “geiger,” “phytools,” and “ape” packages (Revell, 2012; Pennell et al., 2014; Paradis and Schliep, 2019; Pinheiro et al., 2021). We also used the *cor.test* function of the base R “stats” package to calculate the Pearson’s product-moment correlation to determine whether genome size and gene counts were correlated (R Core Team, 2021).

Additionally, we ran ANOSIM tests to determine whether phylogeny was related to habitat preference. Using the “ape” package in R, we called the tree in R and subsequently used the “cophenetic” function in the “stats” package to calculate a distance matrix (Paradis and Schliep, 2019; R Core Team, 2021). Then,



we used the “anosim” function in the R package “vegan” to run ANOSIM tests (Oksanen et al., 2020).

3. Results

3.1. Pangenome

We downloaded 440 publicly available *Sphingomonas* genomes, selected 252 high-quality genomes, and carefully curated them into 8 habitat categories based on the isolation source. The minimum N50 value was 17,936 bases and the maximum was 6,205,897 bases. The minimum GC percent content was 61.98% and the maximum was 70.01%. We selected genomes with a BUSCO completeness score of at least 95%. Most of the sequences had a CheckM completeness score of at least 99% ($n = 220$) and only 1 genome had a completeness score under 96.5% with the lowest score of 94.2%. Additionally, CheckM contamination scores revealed a mode of 0.0, an average score of 1.35, and a maximum score of 13.27 (Supplementary material).

Roary and Prokka pangenome analysis for the 252 *Sphingomonas* genomes revealed a total of 113,816 genes. Specifically, there were 444 core genes found in at least 99% of the genomes, 304 soft core genes found in 95 to 99% of genomes, 4,070 shell genes found in 15 to 95% of genomes, and 108,998 cloud genes present in less than 15% of genomes (Table 2).

When the *Rhodospirillum centenum* SW outgroup was included in the pangenome analysis, there was a total of 115,874 genes with 404 core genes, 321 soft core genes, 4,091 shell genes, and 111,058

cloud genes (Supplementary Table 1; Figure 2). Some of the core gene functions include but are not limited to those associated with ribosomes, transcription factors, translation factors, and ATP synthases.

The pangenome analysis of just the 23 complete *Sphingomonas* genomes revealed a total of 33,131 genes comprised of 758 core genes, 184 soft core genes, 4,452 shell genes, and 27,737 cloud genes (Supplementary Table 2).

3.2. Phylogenetic tree

Phylogenetic analysis of 252 *Sphingomonas* sequences with a *R. centenum* SW outgroup yielded a phylogenetic tree assembled from an alignment of 404 core genes (Figure 3). The tree leaves clustered into 12 clades with a minimum bootstrap value of 86. After running Enveomics, pairwise comparisons within clades revealed a minimum average amino acid identity of 33.24% and a minimum average nucleotide identity of 76.37%.

Significant ANOSIM tests ($p < 0.05$) showed that *Sphingomonas* strains from the same habitat clustered together based on phylogeny, meaning that taxa within a clade had similar habitat preferences. For example, clade 7 was mostly composed of clinical samples that were highly similar to each other, although it also contained representatives isolated from other habitats, such as water and the environment. Clade 12 was dominated by strains from contaminated regions (Figure 3; Supplementary Figure 1). Some known lineages clustered in specific clades. Clade 2 contained *Sphingomonas melonis*, which is

a pathogen of yellow Spanish melon fruits and causes brown spots (Buonaurio et al., 2002). Clade 3 included *Sphingomonas sanguinis*, which causes dry rot of mango (Liu et al., 2018). *Sphingomonas naasensis* was found in clade 6 and was first isolated from forest soil in South Korea (Kim et al., 2014). Clade 7 contained *Sphingomonas koreensis*, which was first isolated from natural mineral water and can be a human pathogen in patients with meningitis (Lee et al., 2001; Marbjerg et al., 2015). Clade 8 included strains of *Sphingomonas japonica* (Supplementary Figure 2) that were isolated from the red king crab from the Sea of Japan (Romanenko et al., 2009). Moreover, *Sphingomonas* strains within the same clade shared similar clusters of accessory genes (Figure 2B).

3.3. Functional genes

We identified 3,615 unique genes from the KEGG database and 269 CAZymes. A subset of KEGG Orthology genes were chosen to investigate genome-based habitat preference traits based on their classification within the Y-A-S framework, specifically for growth ($n = 147$) and stress tolerance ($n = 153$). Genes within the growth category consisted of many tRNA synthetases, and several genes in the stress tolerance category were involved with antitoxins, CRISPR, and binding to antibiotics such as penicillin (Supplementary material). Kruskal–Wallis rank sum tests on habitat and genome-based trait counts yielded significant ($p < 0.05$) differences for all traits except peptidoglycan biosynthesis and degradation proteins (Figure 4). Similarly, analysis of variance tests (ANOVA) on the phylogenetic generalized least squares (PGLS) models indicated that trait frequencies differed significantly ($p < 0.05$) by habitat for all traits except the prokaryotic defense system.

As expected, there was a significant correlation ($p < 0.05$) between genome size and gene counts for most of the traits we analyzed (Supplementary Figure 3). The largest genome with 6,899,075 bases belonged to a strain isolated from a contaminated site, and the shortest genome came from the animal classification with 2,861,323 bases. On average, genomes from contaminated sites were the largest and those from animals were the smallest (Supplementary Table 3). Therefore, genomes from contaminated sites typically had a higher enrichment of genome-based traits, whereas strains from animals often had the lowest gene enrichment when compared to the other habitats (Figure 4). The prokaryotic defense system gene group was highest within contaminated habitats. Additionally, as we anticipated, CAZyme gene frequencies were highest in strains from plants. There was also a high enrichment of chaperones and folding catalysts within genomes isolated from the clinical habitat; on average, the genome size of clinical strains was the second largest (Figure 4; Supplementary Table 3).

We also calculated the relative abundances of the habitat preference traits for each clade (Supplementary Figure 4). For the genome-based traits associated with high growth yield, the amino acid related enzymes and lipopolysaccharide biosynthesis proteins were the most abundant in clade 11, whereas lipid biosynthesis proteins were most abundant in clade 12. CAZymes linked to the resource acquisition strategy were abundant overall, with clades 1 and 3 having the highest relative abundances compared to other

clades. With respect to the genome-based stress tolerance traits, chaperones and folding catalysts were most abundant in clade 7 and clade 11 had the highest abundance of genes for the prokaryotic defense system, as well as peptidoglycan and biosynthesis proteins (Supplementary Figure 4).

4. Discussion

Using comparative genomics, we investigated the association between *Sphingomonas* habitat and phylogeny. Our hypothesis that *Sphingomonas* strains from similar habitats would cluster together in phylogenetic clades was supported as depicted in the phylogenetic tree with a significant association between habitat and phylogeny (Figure 3, ANOSIM test $p < 0.05$). Furthermore, within clades, strains shared similar accessory genes (Figure 2). Moreover, we found partial support for the hypothesis that key, genome-based traits related to fitness in specific environments would correlate with the isolation habitat (Figure 3). A closer investigation of functional genes associated with life history strategies revealed significant differences in gene counts across habitats (Figure 3). Some of the patterns reflected what we anticipated, while others did not. Ultimately, these findings bring us one step closer toward understanding the relationship between habitat preference and phylogeny.

The phylogenetic tree indicates that there is an association between *Sphingomonas* habitat and phylogeny, supporting our hypothesis that strains from similar habitats are more closely related (Figure 3). These findings are also supported in other bacterial systems such as *Bifidobacteria*, *Curtobacterium*, and *Xylella fastidiosa* (Chase et al., 2018; Rodriguez and Martiny, 2020; Batarseh et al., 2022). It appears that abiotic factors as well as biological conditions, such as hosts, contribute to the environmental filtering and evolution of *Sphingomonas* within each habitat (Martiny et al., 2006; Kraft et al., 2015).

Although there was a significant association, the match between habitat and phylogeny was not a perfect. It is possible that our phylogeny could be improved by incorporating accessory genes. However, previous research suggests that the phylogeny produced from an alignment of accessory genes did not substantially differ from a phylogeny constructed with core genes (Batarseh et al., 2022; Scales et al., 2022). It is also possible that our 8 habitat categories (Table 1) may be too broad or too narrow, or perhaps dispersal between sources influences the evolutionary history (Finlay, 2002). Most of the environmental samples consisted of soils, while the plant samples could be separated into root, stem, and leaf subcategories (Supplementary Figure 5). The rhizosphere consists of soils in the vicinity of plant roots, and could include lineages that are selected by both soil and plant properties (Berendsen et al., 2012). Additionally, dispersal between habitats could bring together *Sphingomonas* strains from different sources in the same location (Finlay, 2002; Albright et al., 2019; Walters et al., 2022). Dispersal is particularly likely across environment, plant, water, and contaminated site habitats. Finally, we note that we eliminated some viable genomes from our analysis because we could not determine their isolation source. Future studies like ours would benefit from a standardized approach to metadata reporting about isolation methods in microbial genomics databases.

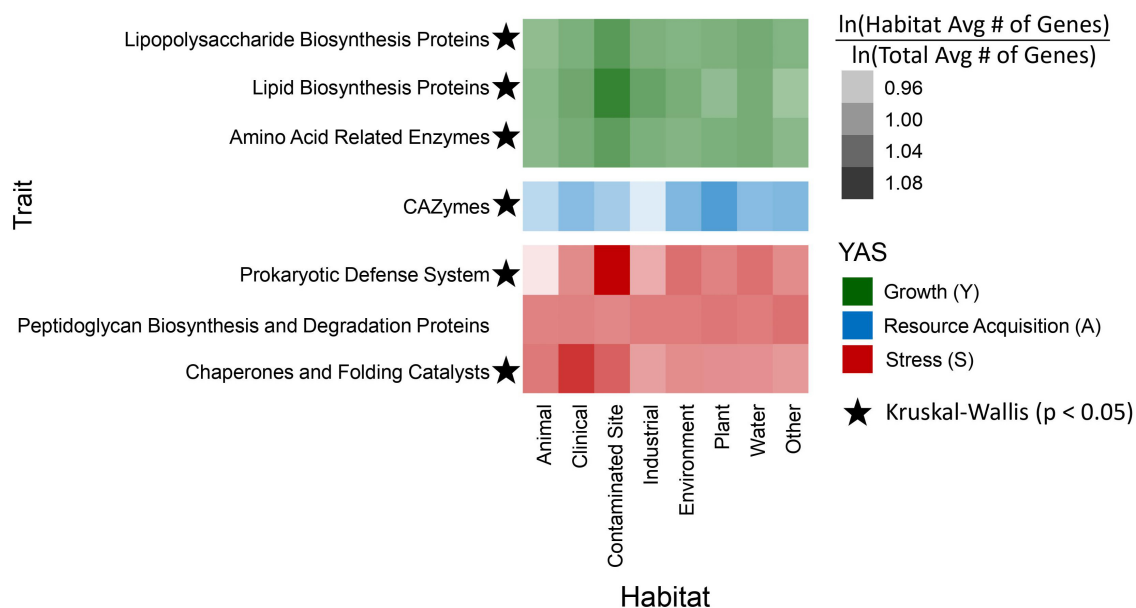


FIGURE 4

Heatmap depicting the enrichment of genome-based traits by habitat. Each heatmap box was calculated by taking the natural log of the average number of genes within a habitat for a specific trait and dividing it by the natural log of the total gene average in all habitats for the same trait. Traits are grouped together based on their Y-A-S classification: top green rows are growth traits, the middle blue CAZymes row is a resource acquisition trait, and the bottom, red rows are stress tolerance traits. Traits with stars indicate significant (Kruskal–Wallis, $p < 0.05$) differences of natural log transformed gene counts between habitats.

Since we found that habitat preference is phylogenetically conserved, we sought to disentangle potential genome-based traits that underlie habitat preference. *Sphingomonas* clades share similar accessory genes (Figure 2), and genome-based trait counts varied by habitat, together suggesting that adaptation to the local environment has shaped habitat preference (Figure 4). Strains from contaminated sites had more genes associated with the prokaryotic defense system, while clinical strains had higher averages for chaperones and folding catalysts (Figure 4). It is possible that in the Y-A-S life history framework, strains from both of these habitats may depend on stress tolerance strategies for survival (Figure 1; Malik et al., 2020). Chaperones and folding catalysts serve as signaling molecules to blood cells to promote immunity and inflammation (Henderson and Pockley, 2010), two common processes in clinical settings. Immune responses are stressful to bacterial infectious agents, and bacterial stress proteins such as chaperones may even trigger the immune response of hosts (Henderson et al., 2006). Moreover, compared to the other habitats, contaminated sites also had more genome-based traits associated with high growth yield (Figure 4). Since *Sphingomonas* can break down pollutants (Schmidt et al., 1992), it is possible that strains in contaminated habitats invested in resource use efficiency rather than stress tolerance. Additionally, we found that *Sphingomonas* strains isolated from plant habitats had more CAZymes (Figure 4), which suggests that they use the resource acquisition strategy to breakdown complex carbohydrates found in plant material (Hervé et al., 2010).

For traits that did not differ significantly across habitats, such as peptidoglycan biosynthesis and degradation proteins, there are two potential possibilities (Figure 4). These traits may be part of the

core genome and are required by all strains for basic functioning. Alternatively, there may be finer-scale differences in specific genes that are not detected because our traits are defined as broad sums of multiple genes. Moreover, proteins may have overlapping functions in metabolic pathways, making it difficult to assign them to a single life history strategy.

Although the genomics field and sequencing technologies have made tremendous advancement (Heather and Chain, 2016), there are still challenges with assembling complete genomes. In publicly available data, there will be differences in the quality of the genomes since sampling and sequencing methods vary across studies. Therefore, to mitigate variability, we were very selective with the *Sphingomonas* genomes that we decided to investigate further. Even though we included fragmented genomes, all sequences had a minimum BUSCO score of 95% from the *Sphingomonadales* order (Seppey et al., 2019). Still, fragmented genomes may reduce the total core gene count in *Sphingomonas* pangenome analysis due to missing genes. Therefore, for comparison, we performed pangenome analysis on the 23 complete *Sphingomonas* genomes in our dataset (Supplementary Table 2), revealing 758 core genes. This analysis indicates that our core gene count of 404 for the genus is reasonable. As the diversity and frequency of genomes increases, the number of core genes should decrease.

We investigated the genomic variation and phylogeny of *Sphingomonas* across different habitats. Additionally, we used a trait-based framework to explore differences in genome-based traits and life history strategies. We found that strains from similar habitats group together in clades and share accessory genes. Although our results did not reveal distinct life history strategies for all habitats, genome-based trait counts varied by habitat. These findings indicate that *Sphingomonas* genome

content reflects habitat preference. Considering the relationships between habitat, genomics, and phylogeny may help us predict *Sphingomonas* habitat preference and better exploit its potential for bioremediation.

Data availability statement

The original contributions presented in this study are included in the article/[Supplementary material](#), further inquiries can be directed to the corresponding author.

Author contributions

BS performed data analysis and wrote the manuscript. BS, BG, and SA contributed to the data interpretation. All authors developed and designed the study, contributed to the article, and approved the submitted version.

Funding

This study was funded by the US Department of Energy Office of Science, Biological and Environmental Research, under award DE-SC0020382 to SA.

Acknowledgments

Thanks to the following people for discussion and feedback on the manuscript: Adam C. Martiny, Alex B. Chase, Brittni L. Bertolet, Edwin Solares, Elsa Abs, Galen T. Martin, Jennifer B. H. Martiny, José M. Murúa Royo, Lucas Ustick, Luciana Chavez Rodriguez, Nicole Hemming-Schroeder, Nicholas C. Scales, Renaud Berlemont, and Tiffany N. Batarseh. Thanks to Nadya Williams and the HPC Team for technical support and maintenance of the high-performance computing clusters.

References

- Albright, M. B. N., Chase, A. B., and Martiny, J. B. H. (2019). Experimental evidence that stochasticity contributes to bacterial composition and functioning in a decomposer community. *MBio* 10, e00568–19. doi: 10.1128/mBio.00568-19
- Asaf, S., Khan, A. L., Khan, M. A., Imran, Q. M., Yun, B. W., and Lee, I. J. (2017). Osmoprotective functions conferred to soybean plants via inoculation with *Sphingomonas* sp. LK11 and exogenous trehalose. *Microbiol. Res.* 205, 135–145. doi: 10.1016/j.micres.2017.08.009
- Asaf, S., Numan, M., Khan, A. L., and Al-Harrasi, A. (2020). *Sphingomonas*: From diversity and genomics to functional role in environmental remediation and plant growth. *Crit. Rev. Biotechnol.* 40, 138–152. doi: 10.1080/07388551.2019.1709793
- Balkwill, D. L., Fredrickson, J. K., and Romine, M. F. (2006). *Sphingomonas and related genera*. New York, NY: Springer New York.
- Batarseh, T. N., Morales-Cruz, A., Ingel, B., Roper, M. C., and Gaut, B. S. (2022). Using genomes and evolutionary analyses to screen for host-specificity and positive selection in the plant pathogen *Xylella fastidiosa*. *Appl. Environ. Microbiol.* 88:e0122022. doi: 10.1128/aem.01220-22
- Berendsen, R. L., Pieterse, C. M. J., and Bakker, P. A. H. M. (2012). The rhizosphere microbiome and plant health. *Trends Plant Sci.* 17, 478–486. doi: 10.1016/j.tplants.2012.04.001
- Buonauro, R., Stravato, V. M., Kosako, Y., Fujiwara, N., Naka, T., Kobayashi, K., et al. (2002). *Sphingomonas melonis* sp. nov., a novel pathogen that causes brown spots on yellow Spanish melon fruits. *Int. J. Syst. Evol. Microbiol.* 52, 2081–2087. doi: 10.1099/00207713-52-6-2081
- Cantarel, B. L., Coutinho, P. M., Rancurel, C., Bernard, T., Lombard, V., and Henrissat, B. (2009). The carbohydrate-active EnZymes database (CAZy): An expert resource for glycogenomics. *Nucleic Acids Res.* 37, D233–D238. doi: 10.1093/nar/gkn663
- Chase, A. B., Gomez-Lunar, Z., Lopez, A. E., Li, J., Allison, S. D., Martiny, A. C., et al. (2018). Emergence of soil bacterial ecotypes along a climate gradient. *Environ. Microbiol.* 20, 4112–4126. doi: 10.1111/1462-2920.14405
- Cockell, C. S., Santomartino, R., Finster, K., Waajen, A. C., Eades, L. J., Moeller, R., et al. (2020). Space station biomineral experiment demonstrates rare earth element extraction in microgravity and Mars gravity. *Nat. Commun.* 11:5523. doi: 10.1038/s41467-020-19276-w

We thank the two reviewers whose comments improved the manuscript clarity.

Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fmicb.2023.1146165/full#supplementary-material>

SUPPLEMENTARY TABLE 1

Sphingomonas genomes isolation source and habitat classification descriptions (see [Table 1.XLSX](#)).

SUPPLEMENTARY TABLE 2

CheckM *Sphingomonas* genome output table (see [Table 2.XLSX](#)).

SUPPLEMENTARY DATA SHEET 2

Roary gene presence and absence output file for *Sphingomonas* genomes and the *Rhodospirillum centenum* SW outgroup.

SUPPLEMENTARY DATA SHEET 3

Sphingomonas clades, habitat classification, genome length, Y-A-S trait frequencies, and natural log trait frequencies.

SUPPLEMENTARY DATA SHEET 4

GhostKoala gene descriptions and Y-A-S classifications.

- Dolan, K. L., Peña, J., Allison, S. D., and Martiny, J. B. H. (2017). Phylogenetic conservation of substrate use specialization in leaf litter bacteria. *PLoS One* 12:e0174472. doi: 10.1371/journal.pone.0174472
- Fierer, N., and Jackson, R. B. (2006). The diversity and biogeography of soil bacterial communities. *Proc. Natl. Acad. Sci. U.S.A.* 103, 626–631. doi: 10.1073/pnas.0507535103
- Finlay, B. J. (2002). Global dispersal of free-living microbial eukaryote species. *Science* 296, 1061–1063. doi: 10.1126/science.1070710
- Gurevich, A., Saveliev, V., Vyahhi, N., and Tesler, G. (2013). QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* 29, 1072–1075. doi: 10.1093/bioinformatics/btt086
- Halo, B. A., Khan, A. L., Waqas, M., Al-Harrasi, A., Hussain, J., Ali, L., et al. (2015). Endophytic bacteria (*Sphingomonas* sp. LK11) and gibberellin can improve *Solanum lycopersicum* growth and oxidative stress under salinity. *J. Plant Interact.* 10, 117–125. doi: 10.1080/17429145.2015.1033659
- Heather, J. M., and Chain, B. (2016). The sequence of sequencers: The history of sequencing DNA. *Genomics* 107, 1–8. doi: 10.1016/j.ygeno.2015.11.003
- Henderson, B., Allan, E., and Coates, A. R. M. (2006). Stress wars: The direct role of host and bacterial molecular chaperones in bacterial infection. *Infect. Immun.* 74, 3693–3706. doi: 10.1128/IAI.01882-05
- Henderson, B., and Pockley, A. G. (2010). Molecular chaperones and protein-folding catalysts as intercellular signaling regulators in immunity and inflammation. *J. Leukoc. Biol.* 88, 445–462. doi: 10.1189/jlb.1209779
- Hervé, C., Rogowski, A., Blake, A. W., Marcus, S. E., Gilbert, H. J., and Knox, J. P. (2010). Carbohydrate-binding modules promote the enzymatic deconstruction of intact plant cell walls by targeting and proximity effects. *Proc. Natl. Acad. Sci. U.S.A.* 107, 15293–15298. doi: 10.1073/pnas.1005732107
- Hyatt, D., Chen, G. L., LoCascio, P. F., Land, M. L., Larimer, F. W., and Hauser, L. J. (2010). Prodigal: Prokaryotic gene recognition and translation initiation site identification. *BMC Bioinformatics* 11:119. doi: 10.1186/1471-2105-11-119
- Isobe, K., Allison, S. D., Khalili, B., Martiny, A. C., and Martiny, J. B. H. (2019). Phylogenetic conservation of bacterial responses to soil nitrogen addition across continents. *Nat. Commun.* 10:2499. doi:10.1038/s41467-019-10390-y
- Isobe, K., Bouskill, N. J., Brodie, E. L., Sudderth, E. A., and Martiny, J. B. H. (2020). Phylogenetic conservation of soil bacterial responses to simulated global changes. *Philos. Trans. R. Soc. B Biol. Sci.* 375:20190242. doi:10.1098/rstb.2019.0242
- Jogler, M., Chen, H., Simon, J., Rohde, M., Busse, H. J., Klenk, H. P., et al. (2013). Description of *Sphingorhabdus planktonica* gen. nov., sp. nov. and reclassification of three related members of the genus *Sphingopyxis* in the genus *Sphingorhabdus* gen. nov. *Int. J. Syst. Evol. Microbiol.* 63, 1342–1349. doi:10.1099/ijs.0.43133-0
- Kanehisa, M., and Goto, S. (2000). KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res.* 28, 27–30. doi: 10.1093/nar/28.1.27
- Kanehisa, M., Sato, Y., and Morishima, K. (2016). BlastKOALA and GhostKOALA: KEGG tools for functional characterization of genome and metagenome sequences. *J. Mol. Biol.* 428, 726–731. doi: 10.1016/j.jmb.2015.11.006
- Kassambara, A. (2020). *ggpubr: "ggplot2" based publication ready plots*. Available online at: <https://cran.r-project.org/package=ggpubr> (accessed February 11, 2023).
- Kim, S.-J., Moon, J.-Y., Lim, J.-M., Ahn, J.-H., Weon, H.-Y., Ahn, T.-Y., et al. (2014). *Sphingomonas aerophila* sp. nov. and *Sphingomonas naasensis* sp. nov., isolated from air and soil, respectively. *Int. J. Syst. Evol. Microbiol.* 64, 926–932. doi: 10.1099/ijs.0.055269-0
- Koskinen, R., Ali-Vehmas, T., Kämpfer, P., Laurikkala, M., Tsitko, I., Kostyal, E., et al. (2000). Characterization of *Sphingomonas* isolates from Finnish and Swedish drinking water distribution systems. *J. Appl. Microbiol.* 89, 687–696. doi: 10.1046/j.1365-2672.2000.01167.x
- Kraft, N. J. B., Adler, P. B., Godoy, O., James, E. C., Fuller, S., and Levine, J. M. (2015). Community assembly, coexistence and the environmental filtering metaphor. *Funct. Ecol.* 29, 592–599. doi: 10.1111/1365-2435.12345
- Lee, J. S., Shin, Y. K., Yoon, J. H., Takeuchi, M., Pyun, Y. R., and Park, Y. H. (2001). *Sphingomonas aquatilis* sp. nov., *Sphingomonas koreensis* sp. nov. and *Sphingomonas taëjonensis* sp. nov., yellow-pigmented bacteria isolated from natural mineral water. *Int. J. Syst. Evol. Microbiol.* 51, 1491–1498. doi:10.1099/00207713-51-4-1491
- Letunic, I., and Bork, P. (2019). Interactive tree of life (iTOL) v4: Recent updates and new developments. *Nucleic Acids Res.* 47, W256–W259. doi:10.1093/nar/gkz239
- Leung, K. T., Chang, Y. J., Gan, Y. D., Peacock, A., Macnaughton, S. J., Stephen, J. R., et al. (1999). Detection of *Sphingomonas* spp. in soil by PCR and sphingolipid biomarker analysis. *J. Ind. Microbiol. Biotechnol.* 23, 252–260. doi: 10.1038/sj.jim.2900677
- Leys, N. M. E. J., Rynjaert, A., Bastiaens, L., Verstraete, W., Top, E. M., and Springael, D. (2004). Occurrence and phylogenetic diversity of *Sphingomonas* strains in soils contaminated with polycyclic aromatic hydrocarbons. *Appl. Environ. Microbiol.* 70, 1944–1955. doi:10.1128/aem.70.4.1944-1955.2004
- Liu, F., Zhan, R. L., and He, Z. Q. (2018). First report of bacterial dry rot of mango caused by *Sphingomonas sanguinis* in China. *Plant Dis.* 102:2632. doi: 10.1094/PDIS-04-18-0589-PDN
- Malik, A. A., Martiny, J. B. H., Brodie, E. L., Martiny, A. C., Treseder, K. K., and Allison, S. D. (2020). Defining trait-based microbial strategies with consequences for soil carbon cycling under climate change. *ISME J.* 14, 1–9. doi: 10.1038/s41396-019-0510-0
- Marbjerg, L. H., Gaini, S., and Justesen, U. S. (2015). First report of *Sphingomonas koreensis* as a human pathogen in a patient with meningitis. *J. Clin. Microbiol.* 53, 1028–1030. doi: 10.1128/JCM.03069-14
- Martiny, A. C., Treseder, K., and Pusch, G. (2013). Phylogenetic conservatism of functional traits in microorganisms. *ISME J.* 7, 830–838. doi: 10.1038/ismej.2012.160
- Martiny, J. B. H., Bohannan, B. J. M., Brown, J. H., Colwell, R. K., Fuhrman, J. A., Green, J. L., et al. (2006). Microbial biogeography: Putting microorganisms on the map. *Nat. Rev. Microbiol.* 4, 102–112. doi: 10.1038/nrmicro1341
- Merino, N., Aronson, H. S., Bojanova, D. P., Feyhl-Buska, J., Wong, M. L., Zhang, S., et al. (2019). Living at the extremes: Extremophiles and the limits of life in a planetary context. *Front. Microbiol.* 10:780. doi: 10.3389/fmicb.2019.00780
- Mundry, R. (2014). “Statistical Issues and assumptions of phylogenetic generalized least squares,” in *Modern phylogenetic comparative methods and their application in evolutionary biology: Concepts and practice*, ed. L. Z. Garamszegi (Berlin: Springer Berlin Heidelberg), 131–153. doi: 10.1007/978-3-662-43550-2_6
- Noble, I. R., and Slatyer, R. O. (1977). Post-fire succession of plants in Mediterranean ecosystems [*Eucalyptus*]. *USDA For. Serv. Gen. Tech. Rep. WO* 3, 27–36.
- Ochman, H., Lawrence, J. G., and Grolsman, E. A. (2000). Lateral gene transfer and the nature of bacterial innovation. *Nature* 405, 299–304. doi:10.1038/35012500
- Oksanen, J., Blanchet, F. G., Friendly, M., Kindt, R., Legendre, P., McGlenn, D., et al. (2020). *vegan: Community ecology package*. Available online at: <https://cran.r-project.org/package=vegan> (accessed April 13, 2022).
- Önder Ergüven, G., and Demirci, U. (2019). Statistical evaluation of the bioremediation performance of *Ochrobactrum thiophenivorans* and *Sphingomonas melonis* bacteria on Imidacloprid insecticide in artificial agricultural field. *J. Environ. Heal. Sci. Eng.* 18, 395–402. doi:10.1007/s40201-019-00391-w
- Page, A. J., Cummins, C. A., Hunt, M., Wong, V. K., Reuter, S., Holden, M. T. G., et al. (2015). Roary: Rapid large-scale prokaryote pan genome analysis. *Bioinformatics* 31, 3691–3693. doi: 10.1093/bioinformatics/btv421
- Paradis, E., and Schliep, K. (2019). ape 5.0: An environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* 35, 526–528. doi: 10.1093/bioinformatics/bty633
- Parks, D. H., Imelfort, M., Skennerton, C. T., Hugenholtz, P., and Tyson, G. W. (2015). CheckM: Assessing the quality of microbial genomes recovered from isolates, single cells, and metagenomes. *Genome Res.* 25, 1043–1055. doi: 10.1101/gr.186072.114
- Pennell, M. W., Eastman, J. M., Slater, G. J., Brown, J. W., Uyeda, J. C., Fitzjohn, R. G., et al. (2014). geiger v2.0: An expanded suite of methods for fitting macroevolutionary models to phylogenetic trees. *Bioinformatics* 30, 2216–2218. doi: 10.1093/bioinformatics/btu181
- Pinheiro, J., Bates, D., DebRoy, S., Sarkar, D., and R Core Team (2021). *nlme: Linear and nonlinear mixed effects models*. Available online at: <https://cran.r-project.org/package=nlme> (accessed April 13, 2022).
- R Core Team (2021). *R: A language and environment for statistical computing*. Vienna: R Core Team.
- Revell, L. J. (2012). phytools: An R package for phylogenetic comparative biology (and other things). *Methods Ecol. Evol.* 3, 217–223.
- Rodriguez, C. I., and Martiny, J. B. H. (2020). Evolutionary relationships among bifidobacteria and their hosts and environments. *BMC Genomics* 21:26. doi: 10.1186/s12864-019-6435-1
- Rodriguez-R, L. M., and Konstantinidis, K. T. (2016). The envomics collection: A toolbox for specialized analyses of microbial genomes and metagenomes. *PeerJ* 4:e1900v1. doi: 10.7287/peerj.preprints.1900v1
- Romanenko, L. A., Tanaka, N., Frolova, G. M., Mikhailov, V. V., and Lyudmila Romanenko, C. A. (2009). *Sphingomonas japonica* sp. nov., isolated from the marine crustacean *Paralithodes camtschatica*. *Int. J. Syst. Evol. Microbiol.* 59(Pt 5), 1179–1182. doi: 10.1099/ijs.0.003285-0
- Scales, N. C., Chase, A. B., Finks, S. S., Malik, A. A., Weihe, C., Allison, S. D., et al. (2022). Differential response of bacterial microdiversity to simulated global change. *Appl. Environ. Microbiol.* 88:e0242921. doi:10.1128/aem.02429-21

- Schmidt, S., Wittich, R. M., Erdmann, D., Wilkes, H., Francke, W., and Fortnagel, P. (1992). Biodegradation of diphenyl ether and its monohalogenated derivatives by *Sphingomonas* sp. strain SS3. *Appl. Environ. Microbiol.* 58, 2744–2750. doi: 10.1128/aem.58.9.2744-2750.1992
- Seemann, T. (2014). Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30, 2068–2069. doi: 10.1093/bioinformatics/btu153
- Seppy, M., Manni, M., and Zdobnov, E. M. (2019). “BUSCO: Assessing genome assembly and annotation completeness BT-gene prediction: Methods and protocols,” in *Bioinformatics*, ed. M. Kollmar (New York, NY: Springer New York), 227–245. doi: 10.1007/978-1-4939-9173-0_14
- Stamatakis, A. (2014). RAxML version 8: A tool for phylogenetic analysis and post-analysis of large phylogenies. *Bioinformatics* 30, 1312–1313. doi: 10.1093/bioinformatics/btu033
- Takeuchi, M., Hamana, K., and Hiraishi, A. (2001). Proposal of the genus *Sphingomonas* sensu stricto and three new genera, *Sphingobium*, *Novosphingobium* and *Sphingopyxis*, on the basis of phylogenetic and chemotaxonomic analyses. *Int. J. Syst. Evol. Microbiol.* 51, 1405–1417. doi:10.1099/00207713-51-4-1405
- Walters, K. E., Capocchi, J. K., Albright, M. B. N., Hao, Z., Brodie, E. L., and Martiny, J. B. H. (2022). Routes and rates of bacterial dispersal impact surface soil microbiome composition and functioning. *ISME J.* 16, 2295–2304. doi: 10.1038/s41396-022-01269-w
- Wattam, A. R., Abraham, D., Dalay, O., Disz, T. L., Driscoll, T., Gabbard, J. L., et al. (2014). PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.* 42, D581–D591. doi: 10.1093/nar/gkt1099
- White, D. C., Sutton, S. D., and Ringelberg, D. B. (1996). The genus *Sphingomonas*: Physiology and ecology. *Curr. Opin. Biotechnol.* 7, 301–306. doi: 10.1016/S0958-1669(96)80034-6
- Yabuuchi, E., Yano, I., Oyaizu, H., Hashimoto, Y., Ezaki, T., and Yamamoto, H. (1990). Proposals of *Sphingomonas paucimobilis* gen. nov. and comb. nov., *Sphingomonas parapaucimobilis* sp. nov., *Sphingomonas yanoikuyae* sp. nov., *Sphingomonas adhaesiva* sp. nov., *Sphingomonas capsulata* comb. nov., and two genospecies of the genus *Sphingomonas*. *Microbiol. Immunol.* 34, 99–119. doi: 10.1111/j.1348-0421.1990.tb00996.x
- Zhang, H., Yohe, T., Huang, L., Entwistle, S., Wu, P., Yang, Z., et al. (2018). dbCAN2: A meta server for automated carbohydrate-active enzyme annotation. *Nucleic Acids Res.* 46:W95. doi: 10.1093/NAR/GKY418
- Zhao, Q., Yue, S., Bilal, M., Hu, H., Wang, W., and Zhang, X. (2017). Comparative genomic analysis of 26 *Sphingomonas* and *Sphingobium* strains: Dissemination of bioremediation capabilities, biodegradation potential and horizontal gene transfer. *Sci. Total Environ.* 609, 1238–1247. doi:10.1016/j.scitotenv.2017.07.249