Trends in
Microbiology

CellPress

Review

Engineering agricultural soil microbiomes and predicting plant phenotypes

Chiara A. Berruto[1] and Gozde S. Demirer[2,*]

Plant growth-promoting rhizobacteria (PGPR) can improve crop yields, nutrient use efficiency, plant tolerance to stressors, and confer benefits to future generations of crops grown in the same soil. Unlocking the potential of microbial communities in the rhizosphere and endosphere is therefore of great interest for sustainable agriculture advancements. Before plant microbiomes can be engineered to confer desirable phenotypic effects on their plant hosts, a deeper understanding of the interacting factors influencing rhizosphere community structure and function is needed. Dealing with this complexity is becoming more feasible using computational approaches. In this review, we discuss recent advances at the intersection of experimental and computational strategies for the investigation of plant–microbiome interactions and the engineering of desirable soil microbiomes.

Brief history and a look towards the future

The addition of **PGPR** (see Glossary) to farmlands has been in practice since the 1895 patent for the use of nitrogen-fixing bacteria in agriculture by Nobbe and Hiltner [1]. These microbial additives, known as **biofertilizers**, are an attractive strategy because they harness the power of PGPR to boost crop yields while reducing environmental damage associated with the overapplication of certain agrochemicals, such as synthetic fertilizers and pesticides [1,2]. However, the addition of biofertilizers to soil does not guarantee successful plant colonization, persistence of inoculums in the soil ecosystem, or consistency of benefits conferred to the plant host [1]. This variability in the efficacy of biofertilizers poses a substantial challenge to their widespread use, which stands to be improved through a deeper understanding of the complex factors that impact soil microbial communities.

Both stochastic and deterministic factors influence the ultimate structure and function of microbiomes. This includes environmental conditions, host features, interspecies microbial competition, and many other factors [1]. Understanding these interactions and pinpointing the factors with outsized influence is being increasingly achieved through a combination of experimental and computational techniques. The use of these interdisciplinary approaches has accelerated as the need to better understand interactions between hosts and their microbiomes has become increasingly important for both medical and agricultural research [3,4]. **Holo-omics** has emerged as a framework for thinking about this as it involves the incorporation of multi-omics data from both the microbiome and the host [3,4]. Integration of these data types provides more comprehensive analysis into host–microbe interactions that influence host physiology [3,4]. Additionally, advances in the studies of plant–microbiome interactions and the engineering of desirable soil communities have recently been accelerated by the application of **machine learning** (**ML**) and **deep learning** (**DL**) strategies.

This review covers recent literature identifying factors with predictable influence over soil microbial communities as well as strategies to design plant microbiomes that will confer benefits to the host. We discuss different experimental and computational strategies to identify plant–microbe

Highlights

Engineering plant microbiomes presents a promising strategy for advancing sustainable agriculture practices.

Genome-wide association studies can identify plant host genomic regions that influence microbiome structure and function, thereby informing genetic engineering and targeted breeding strategies.

Co-inoculation experiments and metabolic modeling can identify critical microbial interspecies interactions.

The incorporation of machine learning algorithms into microbiome research has expanded the predictive capabilities.

Integration of experimental and computational tools has been successful in designing microbiomes with precise functions.

[1]Biology and Biological Engineering, California Institute of Technology, Pasadena, CA, USA
[2]Chemistry and Chemical Engineering, California Institute of Technology, Pasadena, CA, USA

*Correspondence:
gdemirer@caltech.edu (G.S. Demirer).

https://doi.org/10.1016/j.tim.2024.02.003

© 2024 Elsevier Ltd. All rights reserved.

and microbe–microbe interactions with influence over the structure and function of the soil microbiome. We also provide insights on how ML is being used to advance this field. Finally, we discuss recent examples of various plant microbiome engineering approaches and the challenges and current progress of translating them to the field. We end by highlighting areas where additional research is needed.

Identifying host factors dictating plant microbiome structure and function

While plant microbiome structure and function are greatly influenced by soil features such as moisture content, biomass, and soil organic matter, evidence suggests that the host's genotype significantly influences both microbial community assembly and its ultimate functional role [5,6]. Although the exact contribution of these different factors is debated and likely highly variable, it is commonly accepted that the plant host exerts an active influence on the soil microbial community through various mechanisms, such as the secretion of metabolites and the regulation of immune responses [7]. Identifying key host features with predictable influence over the soil microbiome is being increasingly accomplished at the intersection of experimental and computational methods, which integrate omics datasets from the plant host with the soil microbial community (Figure 1). Elucidating host genomic determinants of microbiome structure will accelerate targeted crop breeding efforts, supporting more desirable microbiome communities.

Host genotype influences microbiome structure and function

Genome-wide association studies (**GWAS**) are conducted to identify genomic loci responsible for a trait of interest. This can be used in the study of plant–microbiome interactions to identify host genomic loci responsible for functional responses to PGPR or those associated with microbiome features of interest. GWAS can be used to decipher the genetic elements behind desirable responses to bacterial colonization, which is critical for the successful application of PGPR, such as in the case of biofertilizers. This is done by surveying the natural variation of a plant species' response to inoculation by a given bacterial strain, then associating the various responses with the underlying genetic variation. Studies investigated the response of different *Arabidopsis* accessions to *Azoarcus olearius* DQS-4T [8] and *Pseudomonas simiae* WCS417r [9] colonization, and identified loci implicated in host responsiveness to PGPR. Similar studies were conducted in rice [10],

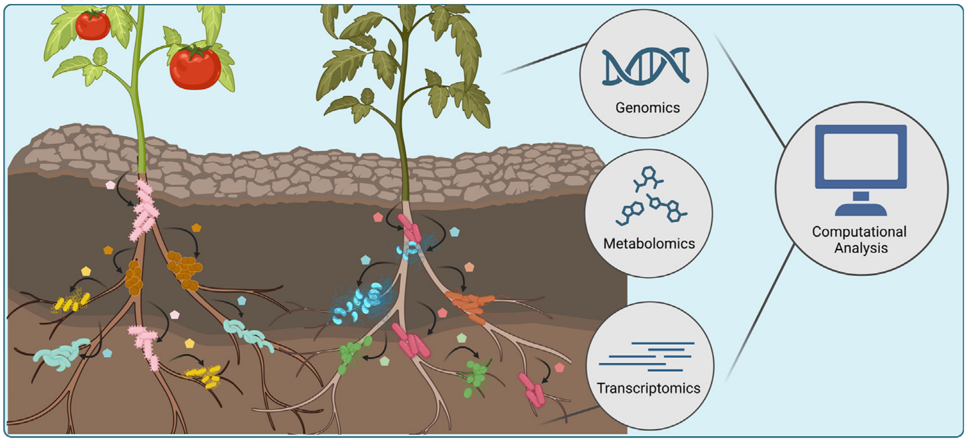


Figure 1. Host genotype and the chemical composition of the root exudates influence microbial community assembly. The integration of genomics, **metabolomics**, and transcriptomics help to disentangle the precise host features responsible for the colonization of a specific microbe and the functional role of the community. This involves genome-wide association studies (genomics and metagenomics), exudate composition analysis (metabolomics), and expression data (transcriptomics).

Glossary

Biofertilizers: products containing beneficial microorganisms that can be applied to soil to promote crop growth.
Chemotaxis: movement of motile cells in response to a chemical gradient.
Deep learning (DL): a type of machine learning that utilizes artificial neural networks to imitate human neural processes; DL requires at least three or more layers of neural connections in the network.
Genome-wide association studies (GWAS): a research approach to find genomic loci implicated with a given phenotype of interest. To do this, genomic variants are statistically associated with variability in a trait.
Heritable taxa: microorganisms whose abundances are significantly dictated by plant genotype.
Holo-omics: incorporation of omics data from a host and its microorganisms.
Machine learning (ML): the use of computer algorithms that learn to adapt without specific instructions based on patterns in the input data.
Metabolomics: the study of specific metabolites present in a given sample.
Metagenome-wide association studies (MWAS): a genome-wide association study where the trait of interest is a microbiome-related feature.
Metagenomics: the study of DNA isolated from all organisms in a given sample.
Plant growth-promoting rhizobacteria (PGPR): rhizosphere bacteria that can promote the growth and stress tolerance of their host plants.
Rhizosphere: the region of soil in the immediate vicinity of a plant root which is directly influenced by root secretions.
Transcriptome-wide association studies (TWAS): a research approach similar to GWAS; however, transcripts are probed rather than the genome in order to associate expression of genes to traits of interest.
Transcriptomics: the study of the total set of RNA molecules in a sample to investigate gene expression.

Medicago truncatula [11], and foxtail [12], verifying the importance of host genotype in determining the functional role of plant microbiomes across various plant species. This highlights an important consideration for the successful application of biofertilizers: ensuring crop genotype is capable of realizing the benefits of the applied PGPR. This is, however, complicated by the wide natural diversity even within the members of the same plant genotype or accession, making the practical prediction of plant responses to PGPR colonization challenging.

Building on the GWAS approach, and keeping with the idea of holo-omics data integration, **metagenome-wide association studies** (**MWAS**) has emerged as a strategy to integrate host genomic and microbial **metagenomic** data (16S or shotgun sequences) to identify host genomic loci that influence microbiome assembly. This involves treating microbiome features, such as the abundance of a given species or community diversity, as distinct traits and using statistical methods to identify correlated regions of the host genome [13]. These approaches can help to identify targets for the breeding of plants with improved microbial associations.

Recently, MWAS have identified host loci correlated with differential rhizobacteria abundances. One of the first studies compared 200 genotypes of field-grown sorghum and found different loci associated with distinct groups of bacteria [13]. Interestingly, they found that highly **heritable taxa** were enriched in the microbiome, including Verrucomicrobiales, Flavobacteriales, and Planctomycetales [13]. MWAS in maize under nitrogen deprivation identified many distinct microbiome-associated plant loci (MAPL) responsible for community assembly. It also showed that microbiome structure is influenced by host genotype more directly under nutrient deprivation compared with optimal growth, suggesting active microbiome selection by the plant host [14]. MWAS have also been conducted in grapevine [15], barley [16], switchgrass [17,18], foxtail [12], poplar trees [19], and tea [20]. With current advances in sequencing and the increased attention on plant microbiomes, we expect the number of crop MWAS to continue to grow, which may reveal conserved MAPL and provide a greater understanding of how plants regulate their microbiomes.

Traditionally, analysis of GWAS has involved univariate statistical testing of individual SNPs to identify those significantly correlated with the trait of interest. The utility of GWAS is greatly improved by the incorporation of ML, which can better manage the high-dimensional data allowing for more robust analysis. Rather than investigating each SNP individually, ML models can identify complex interactions between multiple loci, thereby identifying multi-gene interactions associated with the desired trait or phenotype [21]. The use of ML for GWAS analysis is not yet widespread and has largely been used to identify disease-associated loci in humans. The field of plant biology and microbiology will benefit from the adoption of these tools. To carry out this analysis, various ML algorithms have been utilized, all showing improvements compared with the traditional approaches [21]. A noteworthy example is ensemble learning algorithms, which utilize a series of decision trees and have been shown to be a robust analytical tool for identifying gene interactions in GWAS [21]. For more information on statistical models and specific ML algorithms, we refer the reader to Enoma *et al*. [21] and Sun *et al*. [22].

In sum, through the use of MWAS, microbiome omics data can be combined with plant omics data to identify potential host genomic loci influencing the **rhizosphere** community structure and function. Coupling GWAS with **transcriptome-wide association studies** (**TWAS**) may improve efforts to identify gene targets by also considering gene expression data. One challenge to the use of MWAS is that the data generated is correlative, and inferences made regarding genetic variations that may implicated in a given trait can only be confirmed through the generation of specific plant mutant lines, which is often time-intensive and costly.

Host exudate composition influences microbiome structure

Estimates suggest that plants secrete on average 17% of the carbon they fix into the soil in the form of root exudates [23]. This exuded carbon is of great interest for sustainability research, such as in efforts to increase carbon capture. These rhizodeposits also form the nutritional basis by which the rhizosphere microbiome is built and is one of the major ways in which plants modulate their soil microbial community [7]. Microbial substrate utilization of exuded material determines niche segregation and competitive exclusion, ultimately influencing soil microbiome species composition and abundance in the rhizosphere [24]. Plant root exudate composition is highly variable and dynamically influenced by plant genotype and developmental stage [24], as well as environmental conditions [25,26]. Thus, complete characterization of exudate composition and substrate utilization by different microorganisms will help in the elucidation of the precise mechanisms by which the plant host influences the rhizosphere microbiome. By comparing differences in exudate components and microbiome structure between different native or mutant plants, developmental stages, and stress conditions, researchers can uncover the relationship among exuded chemicals and the microbes that utilize them. Recent studies exploring how different exudate components shape the rhizosphere are summarized in Table 1.

Researchers can also identify key exudate components mediating bacterial colonization through targeted bacterial **chemotaxis** and growth assays. This allows for one-to-one testing of a single strain against a single exudate compound and can be an important way of characterizing plant–microbe communication. This type of experimental work has been done for many key PGPR, such as *Azorhizobium caulinodans* ORS571 [45], *Azospirillum brasilense* [46], *Bacillus amyloliquefaciens* [47–49], *Bacillus flexus* KLBMP 4941 [50], *Bacillus subtilis* [51,52], *Bacillus velezensis* B26 [53], *Paenibacillus polymyxa* SQR-21 [54], and *Pseudomonas putida* [55,56]. Understanding the colonization cues for desirable species can inform microbiome engineering strategies that focus on increasing the colonization of PGPR.

Information from MWAS and exudate composition studies can then inform plant genetic engineering approaches focused on modifying the rhizosphere microbiome. One viable strategy for microbiome engineering is through the modification of plant root exudates for the specific recruitment of desirable PGPR. For example, overexpression of *Arabidopsis thaliana* ALMT1 transporter resulted in increased exudation of malate and increased recruitment of *B. subtilis* FB17 [57]. Similarly, an inward amino acid transporter knock-out in arabidopsis resulted in increased amino acid concentration in root exudates and increased colonization of *P. simiae* WCS417r [58]. However, this did not lead to growth benefits typically conferred by *P. simiae* colonization, suggesting that increased colonization is not always sufficient for beneficial plant phenotypes and that more attention should be given to stimulating precise metabolic functions in PGPR [58]. In rice, upregulation of apigenin secretion caused increased colonization and nitrogen fixation by *Gluconacetobacter diazotrophicus* [59], which increased rice yields and highlighted the feasibility of exudate engineering as a strategy for microbiome engineering.

While these studies corroborate the feasibility of identifying key exudate components as targets for plant genetic engineering, they neglect the microbe–microbe interactions that may impact these results. Complex cross-feeding and competitive interactions play key roles in structuring multispecies microbial communities. Therefore, responses observed in exudate composition versus microbiome studies may be dependent on the bacteria present in the starting soil. Studying the specificity of colonization when conducted in the presence of diverse native soil communities is required for more conclusive identification of plant metabolites mediating species-specific colonization of soil microbes. The adoption of computational tools may help to mitigate some of these

Table 1. Recent studies on exudate composition versus microbiome

Plant species	Plant variables	Exudate component studied [a]	Responsive microbes discussed	Refs
Wheat *Triticum*	Plant development and nitrogen fertilizers	Acetic acid, oxalic acid, succinic acid, tartaric acid	*Arthrobacter*, *Bacillus*, and *Devosia*	[26]
	Transporter mutants and root type	Citrate, malate, GABA	Complete microbiome characterization and spatial analysis – no key responsive microbes were discussed in detail	[27]
Corn *Zea mays*	Mutant lines and transcriptomic gradients along root	Flavones	*Oxalobacteraceae*	[28]
	Natural variation and developmental stage	GABA, DIMBOA	*Xanthomonadaceae*, *Ktedonobacteraceae*, *Oxalobacteraceae*	[29]
	Natural variation and developmental stage	Sugars Jasmonic acid	*Massilia, Bacillus, Burkholderia, Sphingomonas, Streptomyces*	[30]
	Natural variation and developmental stage	IAA, ABA	*Burkholderia, Phytofirmans, Stenotrophomonas, Variovorax, Sphingobacterium, Streptomyces, Pseudoxanthomonas, Microvirga*, and *Sphingobium*	[31]
Sorghum *Sorghum bicolor*	Mutant lines	Sorgoleone	*Nocardia* and *Methylobacillus*	[32]
	Genotype accessions	Organic acids Sugars	Complete microbiome characterization – no key responsive microbes were discussed in detail	[33]
	Mutant lines	Orobanchol	*Chthoniobacter*, *Pirellula*, *Pedosphaeraceae*, *Haliangium*	[34]
	Natural variation and nitrogen application	Trehalose Shikimic acid Phytol	*Pseudomonas*, *Streptomyces*, *Rhizobiaceae*, *Oxalobacteraceae*, *Comamonadaceae*, and *Burkholderiaceae*	[35]
Thale cress *Arabidopsis thaliana*	Pathogen exposure	Amino acids Long-chain organic acids	*Fictibacillus* and *Sphingomonas*	[36]
	Mutant lines	Salicylate	*Streptomyces*	[37]
	Mutant lines	Myo-inositol	*Pantoea* sp.	[19]
Tomato *Solanum lycopersicum*	Intercropping	Taxifolin	*Verticillium dahliae*, *Bacillus* sp.	[38]
	Diseased versus healthy	Metabolite mixture	*Ralstonia solanacearum, Bacillus velezensis*	[39]
Switchgrass *Panicum virgatum*	Intercropping	Malic acid	*Sphingomicrobium, Zymomonas, Methylotenera, Caulobacter, Methylophilus, Flavobacterium*	[40]
Tea plant *Camellia sinensis*	Diseased versus healthy	Phenolic acids Flavonoids	*P. camelliaesinensis*, *Bacillus*, *Pseudomonas*, *Streptomyces*, *Burkholderia*	[41]
Black cottonwood *Populus trichocarpa*	Genotype accessions	Salicylic acid	Betaproteobacteria, Firmicutes, Acidobacteria, Verrucomicrobia, Chloroflexi, Gemmatimonadetes	[42]
Field mustard *Brassica rapa*	Recombinant inbred lines	Glucosinolates	*Chryseobacterium, Pedobacter, Stenotrophomonas, Pseudomonas, Phyllobacterium, Paenarthrobacter*	[43]
Melon *Cucumis melo* Watermelon *Citrullus lanatus*	Diseased versus healthy	Cucurbitacins	*Enterobacter* and *Bacillus*	[44]

[a] Abbreviations: ABA, abscisic acid; DIMBOA, 2,4-dihydroxy-7-methoxy-1,4-benzoxazin-3-one; GABA, γ-aminobutyric acid; IAA, indole-3-acetic acid.

challenges through the use of community metabolic modeling to identify interspecies metabolic interactions with substantial impact on community structure. This is discussed in greater detail later in 'Microbe–microbe interactions influence microbiome structure'.

Identifying microbial factors dictating plant microbiome structure and function

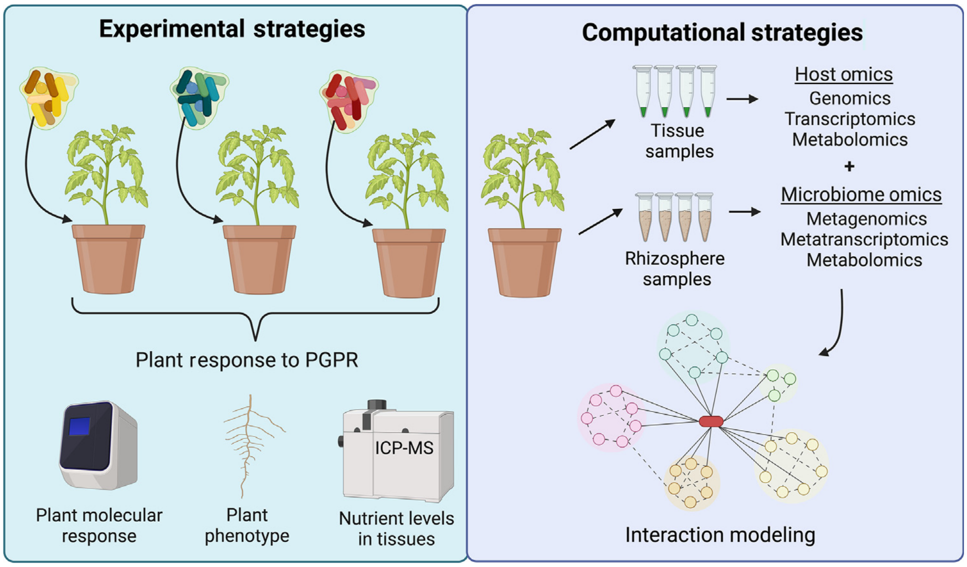
Studying interspecies interactions in the rhizosphere is important for understanding community dynamics, which can have direct impacts on plant phenotypes. Probing the influence of microbe–microbe interactions on the host phenotype can be done by conducting inoculation experiments.

Computational tools can also be used to understand interspecies relationships and predict their impact on community assembly (Figure 2).

Microbe–microbe interactions influence microbiome function

In vitro bacterial competition studies, those conducted in a laboratory setting, provide insight into microbe–microbe interactions that may influence community structure. To do this, two or more strains are grown together in coculture and the survival of each strain is assessed at different timepoints. This provides pair-wise information about competitive or synergistic microbe–microbe interactions but is far removed from native soil contexts. Additionally, coculture competition assays neglect the plant host. Collecting information regarding interspecies interactions in the rhizosphere therefore requires inoculation experiments in which small groups of bacteria are introduced to a sterile plant and various features of both the microbiome and host physiology are probed over time to determine the influence of the different interactions on the plant host.

Co-inoculation of species with complimentary traits can amplify desirable microbiome functions. For instance, enhancing nutrient use efficiency of crops can be achieved by applying bacteria capable of phosphate solubilization and nitrogen fixation. The joint inoculation of multiple phosphate-solubilizing bacteria outperformed inoculation with a single organism to enhance soil nutrients and promote growth of wheat [60] and maize [61]. Similar results were shown for dual inoculation of phosphate-solubilizing fungi with bacteria [62]. The positive impact of inoculation by nitrogen-fixing strains can also be boosted by the addition of other known PGPR and fungi, as shown in common bean [63], barley [64], soybean [65–67], and cowpea [68]. Additionally, consortia of bacteria have been shown to improve host plant growth phenotypes and resilience to a variety of different environmental stressors compared with inoculation with a single strain [69–72]. These pronounced effects on host phenotype have also been recorded for inoculations of PGPR in combination with fungi [73–77]. These studies underscore the importance of considering synergistic



Trends in Microbiology

Figure 2. Experimental and computational strategies for identifying key microbe–microbe interactions. Experimental strategies involve inoculating sterile plants with subsets of bacteria and studying the impact of colonization on the plant host. Computational approaches such as metabolic modeling can be used to integrate multi-omics data to create community models through which key interactions between species can be elucidated. Abbreviations: ICP-MS, inductively coupled plasma mass spectrometry; PGPR, plant growth-promoting rhizobacteria.

microbial interactions when designing microbiomes with a desired function and represents a critical consideration in the design of more effective, more stable biofertilizers [78].

Testing the impact of different combinations of bacterial species, such as in co-inoculation experiments, can be labor-intensive, but the data generated are very valuable. These data can be fed into ML algorithms to predict the function of untested combinations or more diverse synthetic communities. This will greatly accelerate the design of optimal microbiome compositions and synthetic communities. This was shown in the recent integration of ML analysis into co-inoculation experiments which facilitated the construction of microbiomes that confer pathogen protection [79]. The study screened 136 randomly assembled five-member synthetic communities for their ability to confer protection to the plant host in the presence of *Pseudomonas syringae* DC3000 [79]. Using ML for data analysis, the researchers were able to predict the resulting host phenotype of untested combinations and identify the key strains necessary for the protective response [79], this is one example of the power of incorporating ML into data analysis of co-inoculation experiments.

These studies provide much-needed insight into the impact of interspecies interactions on microbiome function but are severely limited by the number of testable strains and associations. Increasing the number of introduced bacterial strains, such as in the case of synthetic communities, increases the complexity of the system, thereby limiting the conclusions that can be made regarding the role of specific interspecies interactions. While these pairwise tests are critical, findings must be further validated in the presence of diverse soil communities to assess if the relationships observed hold true in more soil realistic environments. Computational modeling can be used to expand the power of these studies allowing for the precise investigation of metabolic interactions between many species, which is discussed in the following section.

Microbe–microbe interactions influence microbiome structure

Co-occurrence models (Box 1) have been used to elucidated specific plant–microbe associations, such as the differences in microbial associations between domestic and wild tomato varieties [80], interactions between rhizobacteria and wheat [81], host range and impact of colonization of *Caulobacter* [82], identification of core microbial communities [83,84], and keystone species [85]. However, drawing accurate conclusions on species interactions from co-occurrence data has challenges [86–88]. These challenges can arise from inadequacies in sample collection, where the number of samples required for appropriate interpretation is in many cases not feasible [86,87]. An additional challenge is the difficulty of discriminating interspecies interactions from the influence of abiotic factors. Last, resolving interactions between two species when many other direct and indirect interactions are present is difficult [87]. Therefore, computational strategies that address these limitations are needed before conclusions can be made solely from co-occurrence patterns.

Box 1. Co-occurrence modeling to investigate microbial interactions

In macroecology, co-occurrence and co-abundance patterns have been investigated to identify interspecies interactions that influence community structure [89]. This allows for the identification of keystone species and the inference of positive and negative interactions between organisms. These tools can also be applied to microbial communities to investigate the pairwise correlations between species abundances across samples.

Using co-occurrence modeling of bacterial communities challenged by various stressors – such as soil toxicity [90], microgravity [91], domestication [92,93], and climate change [94] – researchers have shown increased cooperative interactions in communities challenged by stress. This supports the stress gradient hypothesis in which cooperative interactions dominate stressful environments and competitive interactions are more common in benign environments. Co-occurrence networks and the interactions that create community stability under stress conditions can then inform synthetic community design for the construction of more stable consortia or biofertilizers better able to persist when challenged by soil abiotic conditions.

Where co-occurrence models rely on presence–absence data, genome-scale metabolic models (GSMMs, Box 2) provide a more complete investigation into the metabolic mechanisms that mediate various interspecies interactions. GSMMs are a widely used strategy to model microbial communities and are of interest for researchers attempting to integrate holo-omics data into a comprehensive model. While the holo-omics approach is still in its early days, scientists stand to learn a lot from these integrated models that provide more representative and context-specific information. Through simulating the metabolic interplay between microorganisms and the host, holobiont GSMMs can vastly accelerate the field of microbiome engineering and the design of prebiotic molecules to stimulate desired microbiome functions; this is discussed later in 'Plant microbiome engineering'.

Additionally, **transcriptomics** and proteomics data can be integrated into GSMMs to provide constraints related to the activation of specific genetic circuits, further refining the model's validity [4,95]. The power of integrating transcriptomics data into GSMMs was highlighted in the investigation into *Phytophthora infestans*, a pathogen responsible for late blight disease [96]. Integrating transcriptomics data into the model allowed for investigation into the metabolic changes of the pathogen at various stages of infection, providing insights on host–pathogen interactions [96]. By combining the multi-omics data, one can study the effects of environmental perturbations on microbiome composition and function, generating insights into microbial dynamics in complex environments. This was highlighted in the study of the collective response, or the metaphenome, of prairie microbiomes to moisture perturbations finding a more pronounced effect of drying on the community dynamics compared with wetting [97].

To date, there are only a few examples of host–microbiome integrated metabolic modeling in the context of plant–microbe interactions, but we anticipate this to be a growing area of interest in the field. To our knowledge, the first group to do this was in 2020 where the symbiotic interaction between *M. truncatula* and the nodule-forming bacterium *Sinorhizobium meliloti* was modeled, resulting in the construction ViNE: 'Virtual Nodule Environment' [98]. ViNE can simulate nutrient exchange between species, investigate spatial metabolic differences, emergent properties of the interaction, and the effect of various perturbations on the system [98]. The functional capabilities of this model highlight the power of creating holobiont metabolic models, which, once expanded to include more species, will drastically improve our understanding of the key processes mediating microbiome assembly and the impacts of various perturbations on the system. Recently, an additional genome-scale model of the fungal–bacterial–viral holobiont (iHol) has been developed, representing the first GSMM of a fungal holobiont, and highlighting the expanding power of modeling approaches in the context of host–microbe interactions [99].

Box 2. Genome-scale metabolic models (GSMMs) for investigating microbial interactions

GSMMs involve reconstruction of an organism or a community's metabolic capabilities based on genome sequence, allowing for the investigation of microbial communities at a network level, which can be useful in identifying emergent community properties and behaviors. Community-level GSMMs, made by combining individuals GSMMs from different microbial species, are used to investigate host–microbe [98,99], microbe–microbe [100], and microbe–environment interactions [101]. Simulations made using GSMMs provide the starting point to perform flux balance analysis (FBA), which involves incorporation of constraint-based mathematical representations of possible metabolic reactions, elucidating the flow of metabolites through a community [102]. These models can be used to simulate the metabolic interplay between strains and the influence of varying substrate availability, which is important for picking apart the influence of various plant root exudate components on microbiome dynamics. This allows for the identification of functional redundancies and metabolites mediating interspecies interactions. In this way, GSMMs can be used to identify 'keystone metabolites' or the minimal strains required to perform a desired community function, which can inform synthetic community design and microbiome engineering. Publicly available tools have emerged to automate this process, such as MetaGEM [103] to predict the impact of perturbations on a microbial community directly from metagenomic data; MDPbiome [104] to inform microbiome engineering through prescriptive perturbations; and FLYCOP [105] to specifically inform the design of microbial consortia for a given function.

The integration of ML is also expanding the scope and utility of GSMMs by improving predictive performance, automated assembly, and data coverage [106]. ML and DL are powerful tools to aid in the reconstruction of metabolic models resulting in more rapid and accurate model assembly. For example, DL has been used to rapidly predict enzyme functions from protein sequences, thereby improving model reconstruction [106]. ML algorithms can be trained on flux data in combination with GSMMs to predict specific community level responses to environmental perturbations [107]. This has been successful for the prediction of unobserved interactions in soil microbial communities [108], yields of desirable products in the context of microbial factories [109,110], and novel context-dependent microbial interactions from spatially distributed data [111]. Integrating ML algorithms with multi-omics approaches to predict metabolic functions of whole microbial communities will accelerate our ability to engineer complex microbiomes or design consortia with precise outputs, enabling the identification of key factors influencing emergent functional roles of multi-species communities. While the integration of ML into GSMMs is not yet widely adopted, these early efforts are highly promising.

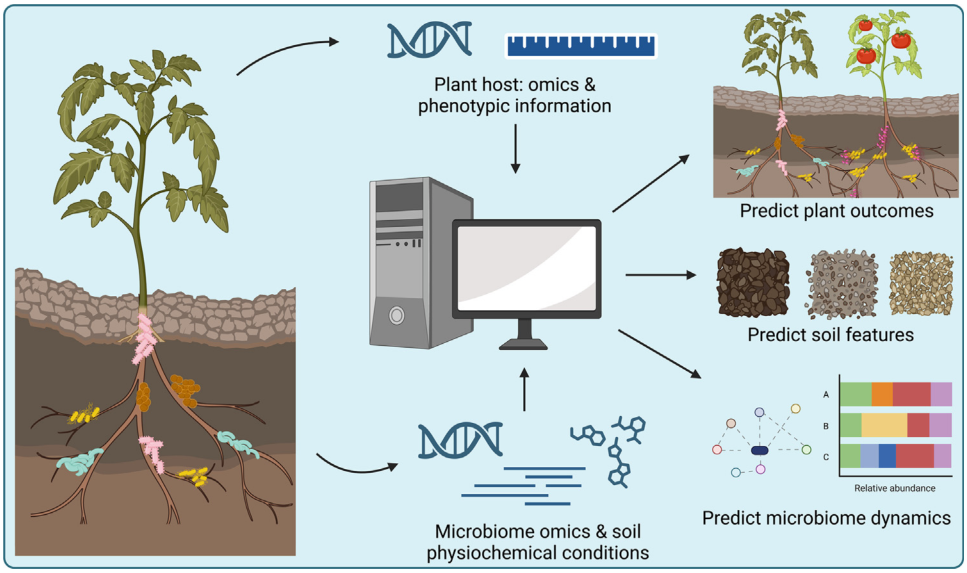
In summary, GSMMs provide a platform to investigate microbial communities and increase our understanding of microbial community assembly and the factors that mediate desirable emergent properties. When combined with additional omics data and ML algorithms, these models are greatly improved and can be very powerful. While this type of modeling can be useful, it is not without limitations. A major challenge is the dependence on genomic sequences and annotations, which for many species remain incomplete. GSMM is also limited in its ability to simulate negative interactions other than resource competition, such as the production of toxins or antimicrobials that may limit the growth of some species [112]. Last, conclusions from modeling data may be improved through the integration of ecological theory that can help contextualize predictions generated by computational design [95,102].

Designing communities with predictable microbiome-associated phenotypes

Using combinations of previously described methods, scientists are making progress towards gaining a predictive understanding of microbial communities. The application of this lies in the ability to design optimized communities for desirable microbiome-associated phenotypes (MAPs). Microbiome engineering is still in its infancy, but the tools described earlier have begun to facilitate the design of synthetic communities with predictable host-associated phenotypes and the modification of plant microbiome structure for desirable MAPs. Future progress in this area may be achieved through the integration of computational and experimental tools (Figure 3).

Predicting plant outcomes based on soil microbiome

Microbes play a significant role in geochemical cycling and plant stress defenses, and therefore, are intimately tied to the soil health and the outcome of plants grown in it. Different microbial species respond to different ranges of environmental conditions. Therefore, probing the microbial species present in a given soil sample can provide insights into the soil's physicochemical properties as well as its future conditions [113]. The power of this has expanded greatly through the availability of large datasets and the incorporation of ML algorithms. These algorithms are trained on metagenomic datasets containing information about soil physicochemical features and other data of interest, such plant productivity. This has proven successful in the prediction of soil quality and plant outcomes of samples based on the structure of the soil microbial community [113]. Similar approaches have predicted land use, as well as certain soil physiochemical features with varying range of accuracy [114,115]. With better model training and optimization, this could represent a novel approach to monitoring soil health.



Trends in Microbiology

Figure 3. Predictive power of computer algorithms for microbiome-associated phenotypes. Using machine-learning (ML) algorithms trained on multi-omics data from plants and microbial communities, one can make predictions about plant and microbiome features in novel settings. This can be extended to incorporate data from the environment where the samples are collected.

Information regarding crop productivity in the presence of different microbial communities can be used to identify key species implicated in desirable growth phenotypes. For example, bacterial species associated with increased potato yield were identified, supporting the feasibility of this approach [116]. The incorporation of ML has since expanded the power of these approaches allowing for predictability of crop productivity based on 16S data [117,118] as well as the disease-suppressive ability of soil based on the microbial community structure [119,120]. Additionally, the impact of pesticides on soil microbial communities has recently been investigated using ML for the design of less harmful pesticides [121]. These applications of ML represent a powerful way of identifying microbial species involved in desirable plant phenotypes.

Plant microbiome engineering

One approach to designing microbial communities with a specific function is through host-centric directed evolution known as host-mediated microbiome engineering. This involves challenging plants with a particular stressor, sampling the microbiome from the plant with the best phenotypic response, then using the microbial sample as the starting inoculum for the next generation, then repeating the process. This was a successful strategy used to select microbiomes that are able to confer benefits to wheat challenged by drought stress [122], maize grown in insoluble phosphorus [123], and *Brachypodium distachyon* challenged by salinity stress [124]. This strategy allows for the creation of optimal microbiome communities and provides useful information regarding beneficial communities structures. However, this is limited in application as many of the native species in the resulting community may be unculturable and therefore would not make translatable biofertilizers.

Identifying the key species responsible for desirable host phenotypes requires elucidation of the core bacteria within the starting community. This was done for the identification of a minimal bacterial consortia involved in inhibition of seed-borne *Fusarium* infection in maize [125]. The process used involved a progressive dilution of the starting community to remove low abundance taxa, followed by chemotaxis screenings to narrow down microbial targets, which were then tested in small synthetic communities for their ability to protect against plant infection [125]. Through this approach, authors were able to identify two key species – *B. amyloliquefaciens* and *Burkholderia cenocepacia* – which could individually protect against *Fusarium* infection. However, this study also found that the protective effects of these species were most pronounced when included in an optimal eight-member synthetic community [125]. In this way, ideal consortia of bacteria can be designed for a desired MAP.

There is additional power in the ability to rationally construct communities from the bottom up, which allows for better characterization of the community and the identification of minimal consortia capable of conferring benefits. For instance, synthetic communities that improve plant phosphate nutrient acquisition were built and tested using small 'building blocks' of species and testing their individual and combined impact on the host's physiology [126]. The impact of these communities on host phenotype was used to train ML algorithms that could then predict novel combinations of microbes resulting in enhanced host effects [126]. This research verifies the feasibility of rationally designing plant microbiomes for desired host phenotypes.

The next efforts lie in engineering *in situ* microbial communities and translating synthetic community-informed discoveries to agricultural field settings. This requires predictable perturbation of native microbial soil communities, which presents many additional challenges compared with controlled greenhouse experiments. Understanding how plant microbiomes develop and stabilize in the presence of diverse communities of native bacteria and changing environmental conditions is essential for successful plant microbiome engineering in the field. The strategies discussed in this review are accelerating the ability of scientists to engineer plant microbiomes both in a laboratory setting and in the field. This includes microbiome-focused selective breeding efforts, community modeling to simulate perturbation strategies, and identifying the best approach to alter the microbial community for improved plant outcomes. Additionally, tools developed for human microbiome engineering are being adapted to soil communities, such as INTEGRATE, a microbiome engineering tool that uses CRISPR RNA-guided integrases for targeted genetic manipulation of bacteria in complex communities [127]. Crosstalk like this between microbiome engineering in health and agriculture sciences will accelerate advancements in both fields. New technologies to introduce or remove species from native soil communities are discussed below and are continuing to emerge, highlighting the rapid advancements being made and hinting at the exciting future of this field. Successful agricultural application of these microbiome engineering strategies will ultimately depend on scalability, cost, consistency, and regulatory policies.

The introduction of beneficial microbes, through the use of biofertilizers, to agricultural soils has been long sought after. Improved understanding of microbial community dynamics and the factors mediating community assembly have informed novel strategies to increase the abundance of plant growth-promoting microbes in the soil. Priority effects influence colonization efforts [7] and, therefore, seed treatments represent an efficient new delivery strategy for bioinoculants. This involves coating seeds in beneficial microbes prior to germination and has already demonstrated success in field studies [128]. Additionally, the application of exogenous carbon sources can stimulate the growth of plant beneficial bacteria in the rhizosphere acting as a soil prebiotic [39,129,130]. This rationale can also be extended to inform the creation of engineered plants with altered exudate composition to support improved microbial communities, which is becoming increasingly possible with advances in plant synthetic biology and bioengineering.

The aforementioned strategies focus on introducing or stimulating the growth of desired strains, but in some cases the removal of unwanted species, such as pathogens, may be desired. This can be achieved using alternative approaches, including phage treatments and nanoparticle-mediated microbiome engineering. Phage treatments involve the application of bacteriophage that target the unwanted species. The phage can be naturally existing or engineered and have shown promising results as a biocontrol agent in many different contexts [131]. Additionally, nanomaterials are being explored as a tool for microbiome engineering, wherein application of specific nanomaterials has been shown to alter the soil microbial community in predictable ways [132], posing an alternative route to suppress plant pathogens or alter *in situ* community composition.

Concluding remarks

The development of computational tools capable of integrating host and microbiome omics data, such as MWAS and GSMM, have expanded the potential for successful microbiome engineering for predictable plant host phenotypes. We anticipate that well-established computational frameworks for the integration of holo-omics data to create comprehensive host–microbiome models will be established soon. While this field is still in its infancy, tools to do this have begun to emerge, such as TransNet, a data-driven analysis protocol to integrate multi-omics data [133]. Currently, researchers are largely reliant on inferences made from individual analyses of different holo-omics datasets from which correlative conclusions regarding host–microbiome relationships are made. The past 20 years have marked a major paradigm shift, where researchers can no longer deny the active role of microorganisms on host physiology, forcing a trend towards models that integrate the holo-biont. The approaches discussed in this review shed light on some of the early strategies and current progress in this area, but improvements are still needed to push the field forward and additional research will be needed to address outstanding questions, some of which are highlighted below (see Outstanding questions).

A big hurdle in the application of computationally driven strategies to identify key genetic loci in plant hosts or to model metabolic interactions relies on sequencing data. Poor annotation coverage of many plant and bacterial genomes creates challenges in drawing conclusions from MWAS and GSMM. The polygenic nature of many MAPL creates additional difficulty in identifying targets for genetic manipulation, which is further challenged by the functional redundancy observed in many plant species. Improvements in our ability to utilize these tools are therefore dependent on progress in fundamental plant genomics research and improvement of computational tools for genome annotation. Data-collection strategies also put limitations on the ability to create computational models of plant microbiomes, which often require absolute abundance rather than relative abundance data generated by 16S sequencing. Collection of host omics data in combination with microbiome omics data will improve our understanding of plant–microbiome interactions, but the development of analytical frameworks for the integration of holo-omics data is still in early stages. Further, the improvement of data collection standards and analytical tools will help to accelerate the creation of more robust computational models of host–microbe relationships, informing strategies to manipulate and select for desirable plant host–microbiome interactions.

Rapid advancement in the use of ML in biological science has been key in gaining predictive understandings of microbial communities. This may be further improved by the incorporation of DL algorithms that have advantages when compared with traditional ML for microbiome studies [134]. Additionally, the use of ML for plant microbiome studies is currently challenged by the noise, compositionality, and sparsity of current datasets; for more information on these challenges and how to address them, we refer the reader to Busato *et al*. [135]. Improving the predictive capabilities of microbiome models and our fundamental understanding of community dynamics will expedite rational microbiome engineering. These tools will inform the construction of synthetic communities, perturbation of naturally occurring communities, and breeding and/or genetic engineering of plants best suited to reap the benefits of their symbionts. Advancements addressed in the section ‘Plant microbiome engineering’ highlight the progress made thus far in both building

Outstanding questions

Can we use ML to determine the main factors influencing the dynamics of a given microbial community ?

How conserved are MAPL?

Once microbial communities are engineered to have a desirable structure or function, how do they change over time?

synthetic communities and manipulating *in situ* ones to promote desirable microbiome-associated plant phenotypes. We anticipate that the continued advancement of these technologies will lead to more successful and widespread field applications, and positively impact the environment by decreasing use of environmentally harmful agrochemicals, such as synthetic fertilizers and pesticides.

Acknowledgments

The authors acknowledge support from Caltech, the Resnick Sustainability Institute (RSI), and the Center for Environmental Microbial Interactions (CEMI). We also thank Catherine Griffin for critically reading this manuscript and providing feedback. All figures were prepared via BioRender.

Declaration of interests

No interests are declared.

References

1. Mitter, E.K. *et al.* (2021) Rethinking crop nutrition in times of modern microbiology: innovative biofertilizer technologies. *Front. Sustain. Food Syst.* 5, 606815
2. Nosheen, S. *et al.* (2021) Microbes as biofertilizers, a potential approach for sustainable crop production. *Sustainability* 13, 1868
3. Nyholm, L. *et al.* (2020) Holo-omics: integrated host–microbiota multi-omics for basic and applied biological research. *iScience* 23, 101414
4. Xu, L. *et al.* (2021) Holo-omics for deciphering plant–microbiome interactions. *Microbiome* 9, 69
5. Chai, Y.N. *et al.* (2021) High-resolution phenotyping of sorghum genotypic and phenotypic responses to low nitrogen and synthetic microbial communities. *Plant Cell Environ.* 44, 1611–1626
6. Yang, C. *et al.* (2022) Influence of plant genotype and soil on the cotton rhizosphere microbiome. *Front. Microbiol.* 13, 1021064
7. Trivedi, P. *et al.* (2020) Plant–microbiome interactions: from community assembly to plant health. *Nat. Rev. Microbiol.* 18, 607–621
8. Plucani Do Amaral, F. *et al.* (2023) Mapping genetic variation in *Arabidopsis* in response to plant growth-promoting bacterium *Azoarcus olearius* DQS-4T. *Microorganisms* 11, 331
9. Wintermans, P.C.A. *et al.* (2016) Natural genetic variation in *Arabidopsis* for responsiveness to plant growth-promoting rhizobacteria. *Plant Mol. Biol.* 90, 623–634
10. Singh, A. *et al.* (2022) Influence of host genotype in establishing root associated microbiome of indica rice cultivars for plant growth promotion. *Front. Microbiol.* 13, 1033158
11. Batstone, R.T. *et al.* (2022) Genome-wide association studies across environmental and genetic contexts reveal complex genetic architecture of symbiotic extended phenotypes. *mBio* 13, e01823-22
12. Wang, Y. *et al.* (2022) GWAS, MWAS and mGWAS provide insights into precision agriculture based on genotype-dependent microbial effects in foxtail millet. *Nat. Commun.* 13, 5913
13. Deng, S. *et al.* (2021) Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome. *ISME J.* 15, 3181–3194
14. Meier, M.A. *et al.* (2022) Association analyses of host genetics, root-colonizing microbes, and plant phenotypes under different nitrogen conditions in maize. *eLife* 11, e75790
15. Berlanas, C. *et al.* (2019) The fungal and bacterial rhizosphere microbiome associated with grapevine rootstock genotypes in mature and young vineyards. *Front. Microbiol.* 10, 1142
16. Escudero-Martinez, C. *et al.* (2022) Identifying plant genes shaping microbiota composition in the barley rhizosphere. *Nat. Commun.* 13, 3443
17. Sutherland, J. *et al.* (2022) Host genomic influence on bacterial composition in the switchgrass rhizosphere. *Mol. Ecol.* 31, 3934–3950
18. Edwards, J.A. *et al.* (2023) Genetic determinants of switchgrass-root-associated microbiota in field sites spanning its natural range. *Curr. Biol.* 33, 1926–1938.e6
19. O'Banion, B.S. *et al.* (2023) Plant myo-inositol transport influences bacterial colonization phenotypes. *Curr. Biol.* 33, 3111–3124.e5
20. Tan, X. *et al.* (2022) Host genetic determinants drive compartment-specific assembly of tea plant microbiomes. *Plant Biotechnol. J.* 20, 2174–2186
21. Enoma, D.O. *et al.* (2022) Machine learning approaches to genome-wide association studies. *J. King Saud Univ. Sci.* 34, 101847
22. Sun, S. *et al.* (2021) Revisiting genome-wide association studies from statistical modelling to machine learning. *Brief. Bioinform.* 22, bbaa263
23. Nguyen, C. (2003) Rhizodeposition of organic C by plants: mechanisms and controls. *Agronomie* 23, 375–396
24. Zhalnina, K. *et al.* (2018) Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly. *Nat. Microbiol.* 3, 470–480
25. Luo, Z. *et al.* (2023) Divergent effects of fertilizer regimes on taxonomic and functional compositions of rhizosphere bacteria and fungi in *Phoebe bournei* young plantations are associated with root exudates. *Forests* 14, 126
26. Chen, S. *et al.* (2019) Root-associated microbiomes of wheat under the combined effect of plant development and nitrogen fertilization. *Microbiome* 7, 136
27. Kawasaki, A. *et al.* (2021) Manipulating exudate composition from root apices shapes the microbiome throughout the root system. *Plant Physiol.* 187, 2279–2295
28. Yu, P. *et al.* (2021) Plant flavones enrich rhizosphere Oxalobacteraceae to improve maize performance under nitrogen deprivation. *Nat. Plants* 7, 481–499
29. Wang, P. *et al.* (2022) Natural variation in root exudation of GABA and DIMBOA impacts the maize root endosphere and rhizosphere microbiomes. *J. Exp. Bot.* 73, 5052–5066
30. Lopes, L.D. *et al.* (2022) Sugars and jasmonic acid concentration in root exudates affect maize rhizosphere bacterial communities. *Appl. Environ. Microbiol.* 88, e00971-22
31. Lopes, L.D. *et al.* (2023) Root exudate concentrations of indole-3-acetic acid (IAA) and abscisic acid (ABA) affect maize rhizobacterial communities at specific developmental stages. *FEMS Microbiol. Ecol.* 99, fiad019
32. Wang, P. *et al.* (2021) The *Sorghum bicolor* root exudate sorgoleone shapes bacterial communities and delays network formation. *mSystems* 6, e00749-20
33. Seitz, V.A. *et al.* (2022) Variation in root exudate composition influences soil microbiome membership and function. *Appl. Environ. Microbiol.* 88, e00226-22
34. Hao, J. *et al.* (2023) CRISPR/Cas9-mediated mutagenesis of *Carotenoid Cleavage Dioxygenase* (*CCD*) genes in sorghum alters strigolactone biosynthesis and plant biotic interactions. *Phytobiomes J.* 7, 297–411
35. Chai, Y.N. *et al.* (2024) Root-associated bacterial communities and root metabolite composition are linked to nitrogen use efficiency in sorghum. *mSystems* 9, e01190-23
36. Yuan, J. *et al.* (2018) Root exudates drive the soil-borne legacy of aboveground pathogen infection. *Microbiome* 6, 156
37. Worsley, S.F. *et al.* (2021) Investigating the role of root exudates in recruiting *Streptomyces* bacteria to the *Arabidopsis thaliana* microbiome. *Front. Mol. Biosci.* 8, 686110

38. Zhou, X. *et al.* (2023) Interspecific plant interaction via root exudates structures the disease suppressiveness of rhizosphere microbiomes. *Mol. Plant* 16, 849–864
39. Wen, T. *et al.* (2023) Tapping the rhizosphere metabolites for the prebiotic control of soil-borne bacterial wilt disease. *Nat. Commun.* 14, 4497
40. Ulbrich, T.C. *et al.* (2022) Plant root exudates and rhizosphere bacterial communities shift with neighbor context. *Soil Biol. Biochem.* 172, 108753
41. Wang, Q. *et al.* (2021) Tea plants with gray blight have altered root exudates that recruit a beneficial rhizosphere microbiome to prime immunity against aboveground pathogen infection. *Front. Microbiol.* 12, 774438
42. Veach, A.M. *et al.* (2019) Rhizosphere microbiomes diverge among *Populus trichocarpa* plant-host genotypes and chemotypes, but it depends on soil origin. *Microbiome* 7, 76
43. DeWolf, E. *et al.* (2023) The rhizosphere microbiome and host plant glucosinolates exhibit feedback cycles in *Brassica rapa*. *Mol. Ecol.* 32, 741–751
44. Zhong, Y. *et al.* (2022) Root-secreted bitter triterpene modulates the rhizosphere microbiota to improve plant fitness. *Nat. Plants* 8, 887–896
45. Liu, X. *et al.* (2019) Oxalic acid from *Sesbania rostrata* seed exudates mediates the chemotactic response of *Azorhizobium caulinodans* ORS571 using multiple strategies. *Front. Microbiol.* 10, 2727
46. O'Neal, L. *et al.* (2020) Specific root exudate compounds sensed by dedicated chemoreceptors shape *Azospirillum brasilense* chemotaxis in the rhizosphere. *Appl. Environ. Microbiol.* 86, e01026-20
47. Feng, H. *et al.* (2018) Identification of chemotaxis compounds in root exudates and their sensing chemoreceptors in plant-growth-promoting rhizobacteria *Bacillus amyloliquefaciens* SQR9. *Mol. Plant Microbe Interact.* 31, 995–1005
48. Gao, T. *et al.* (2023) Watermelon root exudates enhance root colonization of *Bacillus amyloliquefaciens* TR2. *Curr. Microbiol.* 80, 110
49. Feng, H. *et al.* (2019) Recognition of dominant attractants by key chemoreceptors mediates recruitment of plant growth-promoting rhizobacteria. *Environ. Microbiol.* 21, 402–415
50. Xiong, Y.-W. *et al.* (2020) Root exudates-driven rhizosphere recruitment of the plant growth-promoting rhizobacterium *Bacillus flexus* KLBMP 4941 and its growth-promoting effect on the coastal halophyte *Limonium sinense* under salt stress. *Ecotoxicol. Environ. Saf.* 194, 110374
51. Zhang, N. *et al.* (2014) Effects of different plant root exudates and their organic acid components on chemotaxis, biofilm formation and colonization by beneficial rhizosphere-associated bacterial strains. *Plant Soil* 374, 689–700
52. Rekha, K. *et al.* (2018) Plant-growth-promoting rhizobacteria *Bacillus subtilis* RR4 isolated from rice rhizosphere induces malic acid biosynthesis in rice roots. *Can. J. Microbiol.* 64, 20–27
53. Sharma, M. *et al.* (2020) A crosstalk between brachypodium root exudates, organic acids, and *Bacillus velezensis* B26, a growth promoting bacterium. *Front. Microbiol.* 11, 575578
54. Ling, N. *et al.* (2011) Identification and role of organic acids in watermelon root exudates for recruiting *Paenibacillus polymyxa* SQR-21 in the rhizosphere. *Eur. J. Soil Biol.* 47, 374–379
55. Neal, A.L. *et al.* (2012) Benzoxazinoids in root exudates of maize attract *Pseudomonas putida* to the rhizosphere. *PLoS ONE* 7, e35498
56. Li, T. *et al.* (2019) 1-Aminocyclopropane-1-carboxylate: a novel and strong chemoattractant for the plant beneficial rhizobacterium *Pseudomonas putida* UW4. *Mol. Plant-Microbe Interact.* 32, 750–759
57. Kobayashi, Y. *et al.* (2013) Overexpression of AtALMT1 in the *Arabidopsis thaliana* ecotype Columbia results in enhanced Al-activated malate excretion and beneficial bacterium recruitment. *Plant Signal. Behav.* 8, e25565
58. Agorsor, I.D.K. *et al.* (2023) The *Arabidopsis* LHT1 amino acid transporter contributes to *Pseudomonas simiae*-mediated plant growth promotion by modulating bacterial metabolism in the rhizosphere. *Plants* 12, 371
59. Yan, D. *et al.* (2022) Genetic modification of flavone biosynthesis in rice enhances biofilm formation of soil diazotrophic bacteria and biological nitrogen fixation. *Plant Biotechnol. J.* 20, 2135–2148
60. Emami, S. *et al.* (2020) Consortium of endophyte and rhizosphere phosphate solubilizing bacteria improves phosphorous use efficiency in wheat cultivars in phosphorus deficient soils. *Rhizosphere* 14, 100196
61. Ribeiro, V.P. *et al.* (2022) Co-inoculation with tropical strains of *Azospirillum* and *Bacillus* is more efficient than single inoculation for improving plant growth and nutrient uptake in maize. *Arch. Microbiol.* 204, 143
62. Xu, H. *et al.* (2023) Combined phosphate-solubilizing microorganisms jointly promote *Pinus massoniana* growth by modulating rhizosphere environment and key biological pathways in seedlings. *Ind. Crop. Prod.* 191, 116005
63. Pastor-Bueis, R. *et al.* (2021) Yield response of common bean to co-inoculation with *Rhizobium* and *Pseudomonas* endophytes and microscopic evidence of different colonised spaces inside the nodule. *Eur. J. Agron.* 122, 126187
64. Plaza, A. *et al.* (2022) The effect of the nitrogen-fixing bacteria and companion red clover on the total protein content and yield of the grain of spring barley grown in a system of organic agriculture. *Agronomy* 12, 1522
65. Kumawat, K.C. *et al.* (2019) Synergism of *Pseudomonas aeruginosa* (LSE-2) nodule endophyte with *Bradyrhizobium* sp. (LSBR-3) for improving plant growth, nutrient acquisition and soil health in soybean. *World J. Microbiol. Biotechnol.* 35, 47
66. Kumawat, K.C. *et al.* (2022) Co-inoculation of indigenous *Pseudomonas oryzihabitans* and *Bradyrhizobium* sp. modulates the growth, symbiotic efficacy, nutrient acquisition, and grain yield of soybean. *Pedosphere* 32, 438–451
67. Zveushe, O.K. *et al.* (2023) Effects of co-inoculating *Saccharomyces* spp. with *Bradyrhizobium japonicum* on atmospheric nitrogen fixation in soybeans (*Glycine max* (L.)). *Plants* 12, 681
68. Do Nascimento, T.R. *et al.* (2021) Co-inoculation of two symbiotically efficient *Bradyrhizobium* strains improves cowpea development better than a single bacterium application. *3 Biotech* 11, 4
69. Kumawat, K.C. *et al.* (2021) Dual microbial inoculation, a game changer? – bacterial biostimulants with multifunctional growth promoting traits to mitigate salinity stress in spring mungbean. *Front. Microbiol.* 11, 600576
70. Barua, N. *et al.* (2023) Screening the maize rhizobiome for consortia that improve *Azospirillum brasilense* root colonization and plant growth outcomes. *Front. Sustain. Food Syst.* 7, 1106528
71. Kumar, A. *et al.* (2021) Prospecting catabolic diversity of microbial strains for developing microbial consortia and their synergistic effect on Lentil (*Lens esculenta*) growth, yield and iron biofortification. *Arch. Microbiol.* 203, 4913–4928
72. Swiontek Brzezinska, M. *et al.* (2022) Consortium of plant growth-promoting rhizobacteria enhances oilseed rape (*Brassica napus* L.) growth under normal and saline conditions. *Arch. Microbiol.* 204, 393
73. Jabborova, D. *et al.* (2022) Dual inoculation of plant growth-promoting *Bacillus endophyticus* and *Funneliformis mosseae* improves plant growth and soil properties in ginger. *ACS Omega* 7, 34779–34788
74. Yu, L. *et al.* (2022) Cooperation between arbuscular mycorrhizal fungi and plant growth-promoting bacteria and their effects on plant growth and soil quality. *PeerJ* 10, e13080
75. Moreira, H. *et al.* (2020) Synergistic effects of arbuscular mycorrhizal fungi and plant growth-promoting bacteria benefit maize growth under increasing soil salinity. *J. Environ. Manag.* 257, 109982
76. Wu, T.-Y. *et al.* (2023) Dual inoculation with rhizosphere-promoting bacterium *Bacillus cereus* and beneficial fungus *Peniophora cinerea* improves salt stress tolerance and productivity in willow. *Microbiol. Res.* 268, 127280
77. Raklami, A. *et al.* (2019) Use of rhizobacteria and mycorrhizae consortium in the open field as a strategy for improving crop nutrition, productivity and soil fertility. *Front. Microbiol.* 10, 1106
78. Liu, X. *et al.* (2023) Inoculated microbial consortia perform better than single strains in living soil: a meta-analysis. *Appl. Soil Ecol.* 190, 105011

79. Emmenegger, B. *et al.* (2023) Identifying microbiota community patterns important for plant protection using synthetic communities and machine learning. *Nat. Commun.* 14, 7983
80. Cordovez, V. *et al.* (2021) Successive plant growth amplifies genotype-specific assembly of the tomato rhizosphere microbiome. *Sci. Total Environ.* 772, 144825
81. Fan, K. *et al.* (2018) Wheat rhizosphere harbors a less complex and more stable microbial co-occurrence pattern than bulk soil. *Soil Biol. Biochem.* 125, 251–260
82. Berrios, L. (2022) The genus *Caulobacter* and its role in plant microbiomes. *World J. Microbiol. Biotechnol.* 38, 43
83. Machado, D. *et al.* (2021) Polarization of microbial communities between competitive and cooperative metabolism. *Nat. Ecol. Evol.* 5, 195–203
84. Toju, H. *et al.* (2020) Scoring species for synthetic community design: network analyses of functional core microbiomes. *Front. Microbiol.* 11, 1361
85. Kudjordjie, E.N. *et al.* (2019) Maize synthesized benzoxazinoids affect the host associated microbiome. *Microbiome* 7, 59
86. Hirano, H. and Takemoto, K. (2019) Difficulty in inferring microbial community structure based on co-occurrence network approaches. *BMC Bioinformatics* 20, 329
87. Blanchet, F.G. *et al.* (2020) Co-occurrence is not evidence of ecological interactions. *Ecol. Lett.* 23, 1050–1063
88. Freilich, M.A. *et al.* (2018) Species co-occurrence networks: can they reveal trophic and non-trophic interactions in ecological communities? *Ecology* 99, 690–699
89. Berry, D. and Widder, S. (2014) Deciphering microbial interactions and detecting keystone species with co-occurrence networks. *Front. Microbiol.* 5, 219
90. Luo, J. *et al.* (2021) Succession of the composition and co-occurrence networks of rhizosphere microbiota is linked to Cd/Zn hyperaccumulation. *Soil Biol. Biochem.* 153, 108120
91. Cui, J. *et al.* (2023) Microgravity stress alters bacterial community assembly and co-occurrence networks during wheat seed germination. *Sci. Total Environ.* 890, 164147
92. Abdullaeva, Y. *et al.* (2021) Domestication affects the composition, diversity, and co-occurrence of the cereal seed microbiota. *J. Adv. Res.* 31, 75–86
93. Yue, H. *et al.* (2023) Plant domestication shapes rhizosphere microbiome assembly and metabolic functions. *Microbiome* 11, 70
94. Hernandez, D.J. *et al.* (2021) Environmental stress destabilizes microbial networks. *ISME J.* 15, 1722–1734
95. Kessell, A.K. *et al.* (2020) Predictive interactome modeling for precision microbiome engineering. *Curr. Opin. Chem. Eng.* 30, 77–85
96. Botero, D. *et al.* (2018) A genome-scale metabolic reconstruction of *Phytophthora infestans* with the integration of transcriptional data reveals the key metabolic patterns involved in the interaction of its host. *Front. Genet.* 9, 244
97. Roy Chowdhury, T. *et al.* (2019) Metaphenomic responses of a native prairie soil microbiome to moisture perturbations. *mSystems* 4, e00061-19
98. diCenzo, G.C. *et al.* (2020) Genome-scale metabolic reconstruction of the symbiosis between a leguminous plant and a nitrogen-fixing bacterium. *Nat. Commun.* 11, 2574
99. Valadez-Cano, C. *et al.* (2023) Genome-scale model of *Rhizopus microsporus*: metabolic integration of a fungal holobiont with its bacterial and viral endosymbionts. *Environ. Microbiol.* 26, e16551
100. Mataigne, V. *et al.* (2022) Multi-genome metabolic modeling predicts functional inter-dependencies in the *Arabidopsis* root microbiome. *Microbiome* 10, 217
101. Saifuddin, M. *et al.* (2019) Microbial carbon use efficiency predicted from genome-scale metabolic models. *Nat. Commun.* 10, 3568
102. Beck, A.E. *et al.* (2022) Elucidating plant–microbe–environment interactions through omics-enabled metabolic modelling using synthetic communities. *Front. Plant Sci.* 13, 910377
103. Zorrilla, F. *et al.* (2021) metaGEM: reconstruction of genome scale metabolic models directly from metagenomes. *Nucleic Acids Res.* 49, e126
104. García-Jiménez, B. *et al.* (2018) MDPbiome: microbiome engineering through prescriptive perturbations. *Bioinformatics* 34, i838–i847
105. García-Jiménez, B. *et al.* (2018) FLYCOP: metabolic modeling-based analysis and engineering microbial communities. *Bioinformatics* 34, i954–i963
106. Kim, Y. *et al.* (2021) Machine learning applications in genome-scale metabolic modeling. *Curr. Opin. Syst. Biol.* 25, 42–49
107. Zampieri, G. *et al.* (2019) Machine and deep learning meet genome-scale metabolic modeling. *PLoS Comput. Biol.* 15, e1007084
108. DiMucci, D. *et al.* (2018) Machine learning reveals missing edges and putative interaction mechanisms in microbial ecosystem networks. *mSystems* 3, e00181-18
109. Oyetunde, T. *et al.* (2019) Machine learning framework for assessment of microbial factory performance. *PLoS ONE* 14, e0210558
110. Garcia, S. *et al.* (2020) Development of a genome-scale metabolic model of *Clostridium thermocellum* and its applications for integration of multi-omics datasets and computational strain design. *Front. Bioeng. Biotechnol.* 8, 772
111. Lee, J.-Y. *et al.* (2020) Deep learning predicts microbial interactions from self-organized spatiotemporal patterns. *Comput. Struct. Biotechnol. J.* 18, 1259–1269
112. van den Berg, N.I. *et al.* (2022) Ecological modelling approaches for predicting emergent properties in microbial communities. *Nat. Ecol. Evol.* 6, 855–865
113. Correa-Garcia, S. *et al.* (2023) The forecasting power of the microbiome. *Trends Microbiol.* 31, 444–452
114. Hermans, S.M. *et al.* (2020) Using soil bacterial communities to predict physico-chemical variables and soil quality. *Microbiome* 8, 79
115. Wilhelm, R.C. *et al.* (2022) Predicting measures of soil health using the microbiome and supervised machine learning. *Soil Biol. Biochem.* 164, 108472
116. Jeanne, T. *et al.* (2019) Using a soil bacterial species balance index to estimate potato crop productivity. *PLoS ONE* 14, e0214089
117. Chang, H.-X. *et al.* (2017) Metagenome-wide association study and machine learning prediction of bulk soil microbiome and crop productivity. *Front. Microbiol.* 8, 519
118. Jin, T. *et al.* (2017) Taxonomic structure and functional association of foxtail millet root microbiome. *Gigascience* 6, 1–12
119. Zhang, Z. *et al.* (2022) Composition identification and functional verification of bacterial community in disease-suppressive soils by machine learning. *Environ. Microbiol.* 24, 3405–3419
120. Yuan, J. *et al.* (2020) Predicting disease occurrence with high accuracy based on soil macroecological patterns of *Fusarium* wilt. *ISME J.* 14, 2936–2950
121. Ke, M. *et al.* (2022) Development of a machine-learning model to identify the impacts of pesticides characteristics on soil microbial communities from high-throughput sequencing data. *Environ. Microbiol.* 24, 5561–5573
122. Jochum, M.D. *et al.* (2019) Host-mediated microbiome engineering (HMME) of drought tolerance in the wheat rhizosphere. *PLoS ONE* 14, e0225933
123. De Zutter, N. *et al.* (2022) Innovative rhizosphere-based enrichment under P-limitation selects for bacterial isolates with high-performance P-solubilizing traits. *Microbiol. Spectr.* 10, e02052-22
124. Mueller, U.G. *et al.* (2021) Artificial selection on microbiomes to breed microbiomes that confer salt tolerance to plants. *mSystems* 6, e01125-21
125. Xun, W. *et al.* (2023) Sustained inhibition of maize seed-borne *Fusarium* using a *Bacillus* -dominated rhizospheric stable core microbiota with unique cooperative patterns. *Adv. Sci.* 10, 2205215
126. Herrera Paredes, S. *et al.* (2018) Design of synthetic bacterial communities for predictable plant phenotypes. *PLoS Biol.* 16, e2003962
127. Vo, P.L.H. *et al.* (2021) CRISPR RNA-guided integrases for high-efficiency, multiplexed bacterial genome engineering. *Nat. Biotechnol.* 39, 480–489
128. Rocha, I. *et al.* (2019) Seed coating: a tool for delivering beneficial microbes to agricultural crops. *Front. Plant Sci.* 10, 1357
129. Du, J. *et al.* (2021) Synergistically promoting plant health by harnessing synthetic microbial communities and prebiotics. *iScience* 24, 102918
130. Alahmad, A. *et al.* (2023) Prebiotics: a solution for improving plant growth, soil health, and carbon sequestration? *J. Soil Sci. Plant Nutr.* 23, 6647–6669

131. Holtappels, D. *et al.* (2021) The future of phage biocontrol in integrated plant protection for sustainable crop production. *Curr. Opin. Biotechnol.* 68, 60–71
132. Hussain, M. *et al.* (2023) Nano-enabled plant microbiome engineering for disease resistance. *Nano Today* 48, 101752
133. Rodrigues, R.R. *et al.* (2018) Transkingdom networks: a systems biology approach to identify causal members of host–microbiota interactions. In *Microbiome Analysis* (1849) (Beiko, R.G. *et al.*, eds), pp. 227–242, Springer
134. Deng, Z. *et al.* (2021) Application of deep learning in plant–microbiota association analysis. *Front. Genet.* 12, 697090
135. Busato, S. *et al.* (2023) Compositionality, sparsity, spurious heterogeneity, and other data-driven challenges for machine learning algorithms within plant microbiome studies. *Curr. Opin. Plant Biol.* 71, 102326