Integration of omics data to unravel root microbiome recruitment

Anouk Zancarini, Johan Westerhuis, Age Smilde, Harro Bouwmeester

- **Title**

Integration of omics data to unravel root microbiome recruitment

- **Author names and affiliations**

Anouk Zancarini [a,b] (ORCID 0000-0002-8209-5255)
[a] Plant Hormone Biology group, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands
[b] Biosystems Data Analysis group, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands
a.a.c.zancarini@uva.nl

Johan A. Westerhuis [b] (ORCID 0000-0002-6747-9779)
[b] Biosystems Data Analysis group, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands
j.a.westerhuis@uva.nl

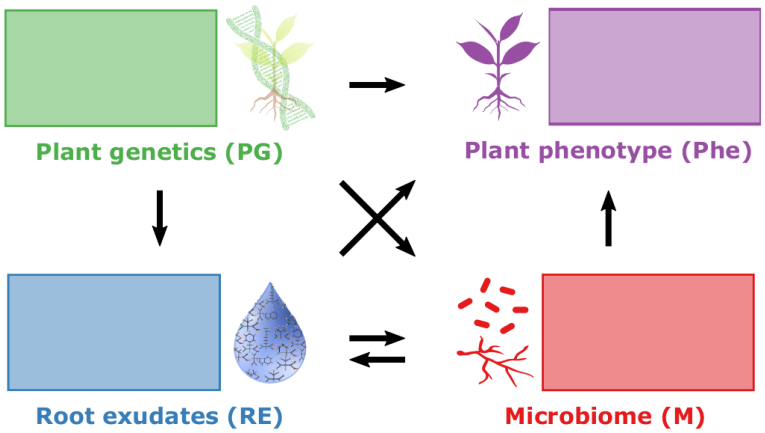
Age K. Smilde [b] (ORCID 0000-0002-3052-4644)
[b] Biosystems Data Analysis group, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands
a.k.smilde@uva.nl

Harro J. Bouwmeester [a] (ORCID 0000-0003-0907-2732)
[a] Plant Hormone Biology group, Swammerdam Institute for Life Sciences, University of Amsterdam, Science Park 904, 1098 XH, Amsterdam, The Netherlands
h.j.bouwmeester@uva.nl

- **Corresponding author**

Anouk Zancarini
Postal address: Plant Hormone Biology group, Swammerdam Institute for Life Sciences, University of Amsterdam, Postbus 1210, 1000 BE Amsterdam, The Netherlands
Email: a.a.c.zancarini@uva.nl
Telephone: (+31) 205 255 963

Graphical abstract



Direct and indirect relationships among plant genetics, root exudates, root microbiome and plant phenotype. Multi-omics data integration can help to unravel the mechanisms that drive microbiome recruitment.

Introduction

The current challenge in agriculture is to be able to increase crop yield under sustainable conditions to feed the growing world population without harming the environment. The plant microbiome could play an essential role in achieving this challenge, as it is becoming increasingly clear that it plays an essential role in supporting plant growth and health. Advances in data analysis - such as multivariate analyses, differential abundance testing methods and machine learning methods - now enable us to link candidate microbes to a phenotype of interest (*e.g.* plant growth, yield, nutrient uptake efficiency, tolerance to disease) [1–3]. However, to be able to select plants that recruit these beneficial microbes, it is essential that the molecular mechanisms underlying microbiome recruitment are unravelled. With the advent of the omics technologies, we can characterise plants in great detail using (epi-)genomics, transcriptomics, proteomics and metabolomics. Through the development of new data analysis paradigms, in principle, these omics data could be related to the associated microbiome (Figure 1). In this review we will discuss the different advanced statistical approaches that have been developed to analyse and integrate plant -omics with microbiome data to propose new mechanistic hypotheses for root microbiome recruitment and its effect on plant phenotype.

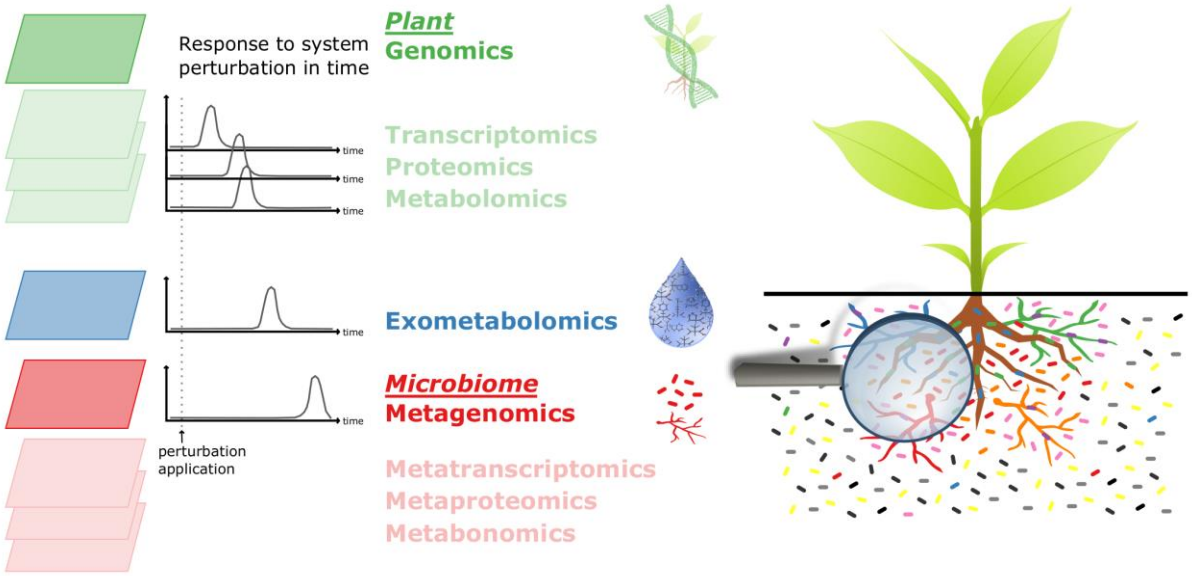


Figure 1: Use of omics data to unravel plant microbiome recruitment
Integration of multiple omics data sets, such as genomics, epigenetics, transcriptomics, proteomics, metabolomics (defined here as the metabolite profile in any given organism), exometabolomics (defined here as the metabolite profile secreted by an organism), metagenomics, metatranscriptomics, metaproteomics and metabonomics (defined here as the metabolite profile from complex systems, such as microbial communities) to unravel the plant-microbiome interaction.

Plant genetics underlying plant microbiome recruitment

Traditional Genome Wide Association Studies (GWAS)

Development of the next-generation sequencing technologies and their decreasing costs have allowed high-throughput plant genotyping using large numbers of single nucleotide polymorphisms (SNPs). This has enabled the use of mapping approaches to identify genes underlying plant traits of interest, through QTL mapping and GWAS [4–6]. Since a number of years, also the plant microbiome is being used as a quantitative plant trait in GWAS to find plant genes underlying microbiome recruitment using mixed linear models [1,7–10]. This confirmed the notion that the plant genotype

drives its associated microbial communities, and linked plant genes involved in stress response, kinase activity, cell wall integrity, root development and carbohydrate metabolism to the occurrence of specific taxa [7–10]. Interestingly, in maize, the predicted bacterial metabolic functions displayed a higher and more significant heritability than the diversity and relative abundance of individual taxa [1]. In future studies, the use of shotgun metagenomics data will further improve the mapping of microbial functions, as was recently demonstrated for the rice phyllosphere microbiome [10].

Nevertheless, identifying the underlying plant loci involved in the microbiome recruitment remains challenging. First, only a small percentage of the variation in the microbiome is generally explained by the plant genotype and just few microbiome traits are usually heritable. Moreover, microbiome recruitment seems to mostly be a polygenic trait. So, the current GWAS models, even with enough power, often fail to detect the microbiome recruitment loci, as discussed elsewhere in this issue [11]. If candidate genes are identified, reproducibility and validation of these candidates using plant mutants and synthetic communities are challenging. In human-microbiome GWAS, results are often difficult to compare between studies [12,13]. For plants, Beilsmith and collaborators proposed a workflow, including thorough quantification and standardized protocols [14]. Also, as environmental conditions are a major component of the variability, GWAS will need to be done across different environmental conditions to test the effect of the environment on candidate genes. Recently, Brachi and collaborators were able to identify heritable microbial hubs that are affected by plant genomics traits across different environmental conditions [15].

Perspectives

There are several recent methodological advances in association studies. First, the use of *k*-mers instead of the commonly used SNPs confirmed associations previously found, but also pinpointed new associations with gene variants missing from reference genomes [16]. Second, to increase the mapping power, Beilsmith and collaborators proposed using multi-traits GWAS modelling SNP associations with many traits rather than with each trait individually, although these models are computationally challenging [14]. Third, to overcome the difficulties of experimental validation, causal inference methods [17], such as genetic structural equation model (GSEM), were proposed, which have been applied as covariance models in multi-traits GWAS to improve power [18].

Finally, a recent promising method development in plant-microbe interaction association studies is the addition of the plant phenotype. Since the microbiome can be considered as a host phenotype but also contribute to the host phenotype, Oyserman and collaborators recently proposed an extended model that includes the microbiome into the traditional GE model, *i.e.* GEM [19]. While the traditional GE model considers the effect of the genotype (G), the environment (E) and their interaction (G:E) on the phenotype (Y or here M for the microbiome), *i.e.* M=G+E+G:E+e, the new GEM model considers the effect of the genotype, the microbiome, the environment and their interaction to determine the plant phenotype (Y), *i.e.* Y=G+E+M+G:E+G:M+E:M+G:E:M+e. A future challenge will be to apply this model to complex natural communities, consisting of hundreds to thousands of species, and taking into consideration also covariance between host genotype, environment and microbiome. Finally, a method called SICOMORE (Selection of Interaction effects in COmpressed Multiple Omics REpresentations) was recently developed and applied in a plant-microbiome study. The authors detected interactions

between plant genomic markers (SNPs) in *Medicago truncatula* and rhizosphere bacterial genera that are linked to a plant phenotype (*e.g.* specific nitrogen uptake) [20].

Root exudates shape the root and rhizosphere microbiome

In a number of studies, it was shown that metabolites in the root exudate play a role in shaping the composition of the root and rhizosphere microbiome [21–28]. We postulate that the discovered relationships between root exudates and the microbiome represent just the tip of the iceberg and propose that data integration methods can be used to unravel new signalling relationships. Pang *et al.* reviewed the integration of plant specialized metabolites and microbiome data [29]. Many methods have been suggested, but most do not take into account the zero-inflated count distribution nor the compositional aspect of these microbiome data. Solutions for these problems include using transformation, imputation and normalization of the data, or using distance-based models. M2IA (automated microbiome and metabolome integrative analysis pipeline), a web-based application combines such pre-processing with standard data integration methods [30].

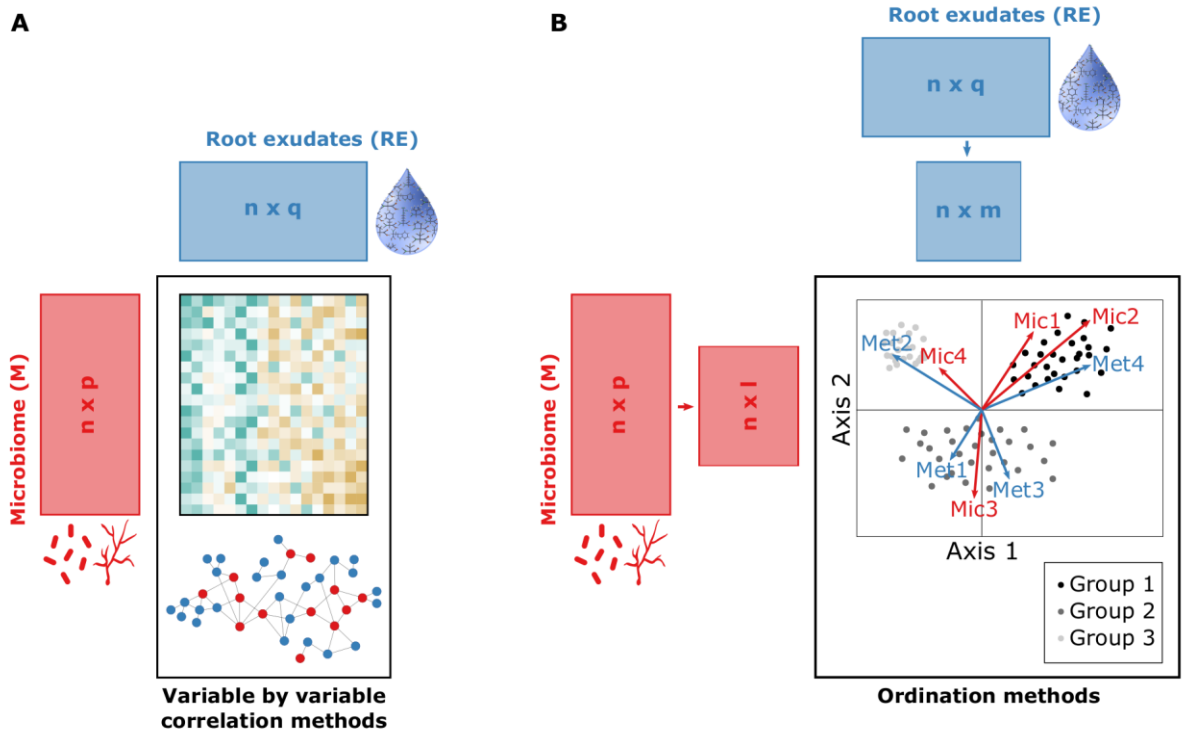


Figure 2: Metabolomics and metagenomics data integration approaches

A graphical summary of the main approaches for metabolomics and metagenomics data integration, including: A) variable-by-variable analyses, such as Pearson, Spearman correlations, sparCC, or neural network approaches, where the outputs can be represented as heatmap and/or networks; B) supervised and unsupervised ordination methods, for which an ordination plot can be rendered and/or features explaining variance extracted.

Abbreviations: n, number of samples; p, number of microbiome variables; q, number of root exudate variables; l, number of latent variables for the microbiome data; m, number of latent variables for the root exudates data; Mic: microbe; Met: root exudate metabolite.

Figure 2 illustrates two different integration approaches of which one uses a variable-by-variable analysis, in which correlations between variables of both data sets are compared for their linear (Pearson), rank (Spearman), or other types of correlations or

co-occurrence. As an example, calculating Pearson correlations, Huang and collaborators [31] identified and linked rhizosphere bacterial OTUs and flavonoids that could explain bitterness in sugarcane. Using Pearson correlation on log transformed data sets, Chaparro *et al.* found that root exudates phenolics and amino acids correlated to bacterial communities composition and transcriptional changes in *Arabidopsis thaliana* [32]. Korenblum *et al.* [33] used self-organizing maps to cluster metabolites and OTUs that highly correlate in 16 clusters and revealed that abundance of specific taxa are related to systemic root metabolome and root exudate changes. However, Morton *et al.* [34] showed that these standard correlation approaches provide a huge number of false high correlations. An alternative approach is to consider co-occurrence probabilities instead of correlations. A new neural network approach method was recently developed, *mmvec* (microbe–metabolite vectors), which is able to identify microbe-metabolome pairs based on co-occurrence while considering compositionality of the data. While *mmvec* was shown to be superior to other correlation approaches, the statistical significance of the interactions remains unclear. For datasets with compositional restrictions, Fang *et al.* [35] introduced CCLasso (Correlation inference for Compositional data through Lasso) that uses the concept of sparsity to find relevant interactions between variables.

A second approach uses restricted ordination methods (Figure 2B) in which only the variation in the microbiome data is explored that is due to variations in the metabolite levels. Examples are Redundancy Analysis (RDA), Canonical Correspondence Analysis (CCA), and, especially for count data, Constrained Analysis of Principal coordinates analysis (CAP) [36]. The ordination is visualized in a biplot or triplot, where the samples (as scores), and the response variables of both datasets as loadings have their respective position on the ordination axes. Potential relationships between metabolites in the rhizosphere and the associated microbial community were thus highlighted using CCA in lettuce under different fertilization regimes, using log10 transformed relative abundance of bacterial/archaeal and fungal communities [37]. Likewise, in *Phragmites australis* relationships between rhizosphere metabolites and associated fungal communities in polluted soils were determined using CCA [38]. Moreover, CAP was applied to centered log-ratio transformed OTU counts with an Euclidean distance measure using plant specialized metabolites as constraining variables and showed that the microbial community was influenced by salicylic acid or its derivatives [39]. For these ordination methods, model significance is commonly tested using permutation of the metabolite's levels over the different samples to break the sample to sample relationship between the microbiome and the metabolome.

Furthermore, a more advanced set of data fusion methods uses canonical variables of both data sets that optimally correlate (Canonical Correlation Analysis CCorA) or have maximum covariance (Diablo). New methods, such as O2PLS (two-way orthogonal partial least squares), JIVE (joint and individual variation explained), DISCO (distinct and common simultaneous component analysis) not only focus on what is in common between the datasets, but also what is systematic within each set. Such methods are often used in medical metabolome-microbiome integration studies [40], but not yet in plants. Most of these methods can handle additional phenotypic data, as nicely discussed by Chu *et al.* [41], and so root exudates, microbiome and plant phenotype could be linked.

Finally, generalized linear models, which can model the data taking into account their specific error distribution, have been used for data fusion using generalized simultaneous component analysis methods. These methods are available for many distribution functions such as Poisson and (zero-inflated) negative binomial. R

packages make use of such models, such as edgeR, Deseq2 and pscl [42–44]. Recently, Song [45,46] introduced generalized simultaneous component analysis to fuse binary copy number aberration data with normally distributed gene expression data to look for their common variation. A similar generalization to include (zero-inflated) negative binomial models in data fusion would be very useful for the integration of metabolomics and microbiome data sets.

Plant genetics, root exudates and microbiome data integration to predict plant phenotype

To model the relationship between multiple actors - such as Plant Genetics (PG), Root Exudates (RE), Microbiome (M) and Phenotype (Phe) - Structural Equation Models (SEMs), introduced in the 1930s by Sewell Wright [47], can be used.

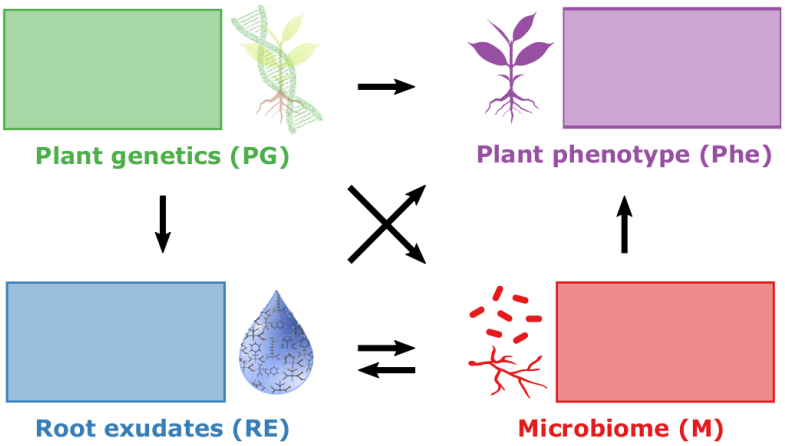


Figure 3: Multi- omics data integration to unravel plant microbiome recruitment
A graphical summary showing potential direct and indirect relationships among plant genetics, root exudates, microbiome and plant phenotype.

These methods originated in the social sciences, but find increasingly use in the natural science as well [48]. The basic idea is to summarize blocks of manifest variables (*e.g.* one block of microbiome data and one block of metabolomics data) into latent variables. These latent variables are now connected through an assumed pathway defining their connectivity. This part of the model is called the inner model, and the part describing the manifest variables in terms of their latent variables is called the outer model. The elegance of SEMs is that they can distinguish direct from indirect effects. Figure 3 shows that there is a direct effect from PG to Phe, but also an indirect effect through the path PG, RE, M and Phe. SEMs are capable of disentangling these effects. Such SEMs can also be extended to deal with genetic effects [18]. Special versions of SEMs (called Structural Causal Models) are used in causal analysis [49]. SEMs are starting to be used in microbiome research, *e.g.*, in ecological applications [50–53]. In some cases, summaries of the microbiome (*e.g.* alpha-diversity measures) can be used as the outer model, *i.e.*, they are used as latent variables in the SEM model. These examples show that SEMs are indeed powerful models to study complex systems.

The real challenge of the use of SEMs in microbiome research is in keeping the notion of latent variables since that allows for modelling simultaneously multiple blocks of multivariate (manifest) variables. This may encompass, *e.g.*, many SNPs for the PG, many OTUs/ASVs (amplicon sequence variants) for the microbiome, and many

metabolites for the RE. There are (at least) three challenges to overcome. The first is to extend the traditional SEMs to handle more than one latent variable per block. This is not trivial, but some ideas on how to do this are available, *e.g*. using sequential and orthogonalized partial least square regression for path analysis (SO-PLS-Path) models [54,55]. Another challenge is to extend the SEMs to handle data of different measurement types. In the example above, SNPs and OTUs/ASVs consist of (limited) count data, while RE consists of quantitative data. One avenue to explore may be the use of nonlinear generalized structured component analysis, which can handle both quantitative and qualitative data [56] or extensions of generalized simultaneous component analysis [45]. Although both extensions can handle high-dimensional blocks in the SEM models, in each block there may still be variables/features that are not important but may obscure the relations. Hence, the final challenge is to select variables to overcome this problem. This may be done in each block before any SEM modelling using techniques from machine learning [57]. Alternatively, this can be done by carefully studying the outcome of a SEM model and interrogating the model for variable importance, *e.g.*, by studying the loadings of the variables in the outer relationships. If these challenges are tackled then the rewards are high: a full description of the system on the level of the measured variables relevant to the biological system.

Conclusions and outlook

By now there is substantial evidence that plant genetics affects the root microbiome although it often explains just a small part of the total variation. It is becoming clear that, to really expand our knowledge on the plant microbiome interaction, the microbiome should not only be considered as a phenotype but should also be part of the explanatory variables that predict the plant phenotype. Moreover, there are many indications that specific metabolites in the root exudate drive microbiome selection and/or assembly. Multi-omics data integration could help to identify the molecular mechanisms underlying microbiome recruitment also considering metabolite-metabolite, microbe-microbe, and metabolite-microbe interactions. Furthermore, modelling, using SEM, could help us to go beyond finding more associations and causation, integrating all the drivers, including plant genetics, root exudates and the microbiome to predict the plant phenotype, and identify direct and indirect effects among the drivers. This knowledge will allow us to shape the microbiome through breeding, possibly through changes in the root exudate, and optimise plant/crop growth under the desired conditions.

Acknowledgements

The authors acknowledge funding by the University of Amsterdam, especially through the Research Priority Area Systems Biology program, the Dutch Research Council (NWO/OCW) for the MiCRop Consortium programme, Harnessing the second genome of plants (to HB; Grant number 024.004.014) and funding by the European Research Council through ERC Advanced grant CHEMCOMRHIZO (to HB; 670211).

References

Paper of particular interest, published within the period of review (from 2019 to 2021), have been highlighted as:

• of special interest

•• of outstanding interest

1. Wallace JG, Kremling KA, Kovar LL, Buckler ES: **Quantitative Genetics of the Maize Leaf Microbiome**. *Phytobiomes J* 2018, **2**:208–224.

2. Edwards JA, Santos-Medellín CM, Liechty ZS, Nguyen B, Lurie E, Eason S, Phillips G, Sundaresan V: **Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice**. *PLoS Biol* 2018, **16**:e2003862.

3. Moroenyane I, Tremblay J, Yergeau É: **Temporal and spatial interactions modulate the soybean microbiome**. *FEMS Microbiol Ecol* 2021, **97**.

4. Wang B, Lin Z, Li X, Zhao Y, Zhao B, Wu G, Ma X, Wang H, Xie Y, Li Q, et al.: **Genome-wide selection and genetic improvement during modern maize breeding**. *Nat Genet* 2020, **52**:565–571.

5. Ferrero-Serrano Á, Assmann SM: **Phenotypic and genome-wide association with the local environment of Arabidopsis**. *Nat Ecol Evol* 2019, **3**:274–285.

6. Tang W, Ye J, Yao X, Zhao P, Xuan W, Tian Y, Zhang Y, Xu S, An H, Chen G, et al.: **Genome-wide associated study identifies NAC42-activated nitrate transporter conferring high nitrogen use efficiency in rice**. *Nat Commun* 2019, **10**:5279.

7. Horton MW, Bodenhausen N, Beilsmith K, Meng D, Muegge BD, Subramanian S, Vetter MM, Vilhjálmsson BJ, Nordborg M, Gordon JI, et al.: **Genome-wide association study of Arabidopsis thaliana leaf microbial community**. *Nat Commun* 2014, **5**:5320.

8. Bergelson J, Mittelstrass J, Horton MW: **Characterizing both bacteria and fungi improves understanding of the Arabidopsis root microbiome**. *Sci Rep* 2019, **9**:24.
   • The authors showed that root microbiome is affected by plant candidate genes involved not only in plant immunity but also plant physiology and architecture, such as root hair development. This is the first GWAS study on root microbiome including both bacterial and fungal communities.

9. Deng S, Caddell DF, Xu G, Dahlen L, Washington L, Yang J, Coleman-Derr D: **Genome wide association study reveals plant loci controlling heritability of the rhizosphere microbiome**. *ISME J* 2021, doi:10.1038/s41396-021-00993-z.

10. Roman-Reyna V, Pinili D, Borja FN, Quibod IL, Groen SC, Mulyaningsih ES, Rachmat A, Slamet-Loedin IH, Alexandrov N, Mauleon R, et al.: **The rice leaf microbiome has a conserved community structure controlled by complex host-microbe interactions**. *bioRxiv* 2019, doi:10.1101/615278.

11. Bergelson J, Brachi B, Roux F, Vailleau F: **Assessing the potential to harness the microbiome through plant genetics.** *Curr Opin Biotechnol* 2021,

12. Gilbert JA, Quinn RA, Debelius J, Xu ZZ, Morton J, Garg N, Jansson JK, Dorrestein PC, Knight R: **Microbiome-wide association studies link dynamic microbial consortia to disease**. *Nature* 2016, **535**:94–103.

13. Awany D, Allali I, Dalvie S, Hemmings S, Mwaikono KS, Thomford NE, Gomez A, Mulder N, Chimusa ER: **Host and Microbiome Genome-Wide Association Studies: Current State and Challenges**. *Front Genet* 2018, **9**:637.

14. Beilsmith K, Thoen MPM, Brachi B, Gloss AD, Khan MH, Bergelson J: **Genome-wide association studies on the phyllosphere microbiome: Embracing complexity in host-microbe interactions**. *Plant J Cell Mol Biol* 2019, **97**:164–181.

15. Brachi B, Filiault D, Darme P, Mentec ML, Kerdaffrec E, Rabanal F, Anastasio A, Box M, Duncan S, Morton T, et al.: **Plant genes influence microbial hubs that shape beneficial leaf communities**. *bioRxiv* 2017, doi:10.1101/181198.

16. Voichek Y, Weigel D: **Identifying genetic variants underlying phenotypic variation in plants without complete genomes**. *Nat Genet* 2020, **52**:534–540.
••• The authors compared the use of SNP and k-mers in GWAS, and demonstrated that k-mer method allow to retrieve associations identified with SNP approach, but also allow finding new associations regardless of reference genome quality.

17. Hu P, Jiao R, Jin L, Xiong M: **Application of Causal Inference to Genomic Analysis: Advances in Methodology**. *Front Genet* 2018, **9**:238.

18. Kruijer W, Behrouzi P, Bustos-Korts D, Rodríguez-Álvarez MX, Mahmoudi SM, Yandell B, Wit E, van Eeuwijk FA: **Reconstruction of Networks with Direct and Indirect Genetic Effects**. *Genetics* 2020, **214**:781–807.
•• Although the focus was not on the microbiome, the authors showed a very nice extension of the standard SEM methodology. In this so-called GSEM models, genetic (G) effects can be modeled thereby also disentangling direct and indirect genetic effects. This method can also be very useful in applications connecting genetics to the microbiome.

19. Oyserman BO, Cordovez V, Flores SS, Leite MFA, Nijveen H, Medema MH, Raaijmakers JM: **Extracting the GEMs: Genotype, Environment, and Microbiome Interactions Shaping Host Phenotypes**. *Front Microbiol* 2020, **11**:574053.
•• The authors proposed and tested with an in vitro experiment an extended GE model, i.e. GEM, that includes the contribution of the microbiome to the host phenotype, instead of considering the microbiome as the plant phenotype like in up-to-date plant-microbiome GWAS.

20. Guinot F, Szafranski M, Chiquet J, Zancarini A, Le Signor C, Mougel C, Ambroise C: **Fast computation of genome-metagenome interaction effects**. *Algorithms Mol Biol AMB* 2020, **15**:13.

21. Lebeis SL, Paredes SH, Lundberg DS, Breakfield N, Gehring J, McDonald M, Malfatti S, Glavina del Rio T, Jones CD, Tringe SG, et al.: **PLANT MICROBIOME. Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa**. *Science* 2015, **349**:860–864.

22. Stringlis IA, Yu K, Feussner K, de Jonge R, Van Bentum S, Van Verk MC, Berendsen RL, Bakker PAHM, Feussner I, Pieterse CMJ: **MYB72-dependent coumarin exudation shapes root microbiome assembly to promote plant health**. *Proc Natl Acad Sci U S A* 2018, **115**:E5213–E5222.

23. Hu L, Robert CAM, Cadot S, Zhang X, Ye M, Li B, Manzo D, Chervet N, Steinger T, van der Heijden MGA, et al.: **Root exudate metabolites drive plant-soil feedbacks on growth and defense by shaping the rhizosphere microbiota**. *Nat Commun* 2018, **9**:2738.

24. Huang AC, Jiang T, Liu Y-X, Bai Y-C, Reed J, Qu B, Goossens A, Nützmann H-W, Bai Y, Osbourn A: **A specialized metabolic network selectively modulates Arabidopsis root microbiota**. *Science* 2019, **364**.

25. Cotton TEA, Pétriacq P, Cameron DD, Meselmani MA, Schwarzenbacher R, Rolfe SA, Ton J: **Metabolic regulation of the maize rhizobiome by benzoxazinoids**. *ISME J* 2019, **13**:1647–1658.

26. Wang P, Chai YN, Roston R, Dayan FE, Schachtman DP: **The Sorghum bicolor Root Exudate Sorgoleone Shapes Bacterial Communities and Delays Network Formation**. *mSystems* 2021, **6**.

27. Zhalnina K, Louie KB, Hao Z, Mansoori N, da Rocha UN, Shi S, Cho H, Karaoz U, Loqué D, Bowen BP, et al.: **Dynamic root exudate chemistry and microbial substrate preferences drive patterns in rhizosphere microbial community assembly**. *Nat Microbiol* 2018, **3**:470–480.

28. Sasse J, Martinoia E, Northen T: **Feed Your Friends: Do Plant Exudates Shape the Root Microbiome?** *Trends Plant Sci* 2018, **23**:25–41.

29. Pang Z, Chen J, Wang T, Gao C, Li Z, Guo L, Xu J, Cheng Y: **Linking Plant Secondary Metabolites and Plant Microbiomes: A Review**. *Front Plant Sci* 2021, **12**:621276.
    •• Very thorough review that discusses plant secondary metabolites and their interaction with plant microbes.

30. Ni Y, Yu G, Chen H, Deng Y, Wells PM, Steves CJ, Ju F, Fu J: **M2IA: a web server for microbiome and metabolome integrative analysis**. *Bioinforma Oxf Engl* 2020, **36**:3493–3498.

31. Huang W, Sun D, Chen L, An Y: **Integrative analysis of the microbiome and metabolome in understanding the causes of sugarcane bitterness**. *Sci Rep* 2021, **11**:6024.

32. Chaparro JM, Badri DV, Vivanco JM: **Rhizosphere microbiome assemblage is affected by plant development**. *ISME J* 2014, **8**:790–803.

33. Korenblum E, Dong Y, Szymanski J, Panda S, Jozwiak A, Massalha H, Meir S, Rogachev I, Aharoni A: **Rhizosphere microbiome mediates systemic root metabolite exudation by root-to-root signaling**. *Proc Natl Acad Sci U S A* 2020, **117**:3874–3883.
• The authors demonstrated that the tomato rhizosphere microbiome affects the chemical composition of root exudation through a systemic root–root signaling mechanism, process, which they termed SIREM (systemically induced root exudation of metabolites).

34. Morton JT, Aksenov AA, Nothias LF, Foulds JR, Quinn RA, Badri MH, Swenson TL, Van Goethem MW, Northen TR, Vazquez-Baeza Y, et al.: **Learning representations of microbe-metabolite interactions**. *Nat Methods* 2019, **16**:1306–1314.

35. Fang H, Huang C, Zhao H, Deng M: **CCLasso: correlation inference for compositional data through Lasso**. *Bioinforma Oxf Engl* 2015, **31**:3172–3180.

36. Legendre P, Legendre L: *Numerical Ecology*. Elsevier; 2012.

37. Windisch S, Sommermann L, Babin D, Chowdhury SP, Grosch R, Moradtalab N, Walker F, Höglinger B, El-Hasan A, Armbruster W, et al.: **Impact of Long-Term Organic and Mineral Fertilization on Rhizosphere Metabolites, Root-Microbial Interactions and Plant Health of Lettuce**. *Front Microbiol* 2020, **11**:597745.

38. Kalu CM, Oduor Ogola HJ, Selvarajan R, Tekere M, Ntushelo K: **Fungal and metabolome diversity of the rhizosphere and endosphere of Phragmites australis in an AMD-polluted environment**. *Heliyon* 2021, **7**:e06399.

39. Veach AM, Morris R, Yip DZ, Yang ZK, Engle NL, Cregger MA, Tschaplinski TJ, Schadt CW: **Rhizosphere microbiomes diverge among Populus trichocarpa plant-host genotypes and chemotypes, but it depends on soil origin**. *Microbiome* 2019, **7**:76.

40. Segal JP, Mullish BH, Quraishi MN, Acharjee A, Williams HRT, Iqbal T, Hart AL, Marchesi JR: **The application of omics techniques to understand the role of the gut microbiota in inflammatory bowel disease**. *Ther Adv Gastroenterol* 2019, **12**:1756284818822250.

41. Chu SH, Huang M, Kelly RS, Benedetti E, Siddiqui JK, Zeleznik OA, Pereira A, Herrington D, Wheelock CE, Krumsiek J, et al.: **Integration of Metabolomic and Other Omics Data in Population-Based Study Designs: An Epidemiological Perspective**. *Metabolites* 2019, **9**.

42. Robinson MD, McCarthy DJ, Smyth GK: **edgeR: a Bioconductor package for differential expression analysis of digital gene expression data**. *Bioinforma Oxf Engl* 2010, **26**:139–140.

43. Love MI, Huber W, Anders S: **Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2**. *Genome Biol* 2014, **15**:550.

44. Jackman S: *pscl: Classes and Methods for R Developed in the Political Science Computational Laboratory.* 2020.

45. Song Y, Westerhuis JA, Aben N, Wessels LFA, Groenen PJF, Smilde AK: **Generalized simultaneous component analysis of binary and quantitative data**. *J Chemom* 2021, **35**:e3312.

46. Song Y, Westerhuis JA, Smilde AK: **Separating common (global and local) and distinct variation in multiple mixed types data sets**. *J Chemom* 2020, **34**.

47. Wright S: **The Method of Path Coefficients**. *Ann Math Stat* 1934, **5**:161–215.

48. Kaplan D: *Structural equation modeling: Foundations and extensions.* Sage Publications; 2008.

49. Pearl J: **Causal inference in statistics: An overview**. *Stat Surv* 2009, **3**.

50. Delgado-Baquerizo M, Maestre FT, Reich PB, Jeffries TC, Gaitan JJ, Encinar D, Berdugo M, Campbell CD, Singh BK: **Microbial diversity drives multifunctionality in terrestrial ecosystems**. *Nat Commun* 2016, **7**:10541.

51. Mamet SD, Redlick E, Brabant M, Lamb EG, Helgason BL, Stanley K, Siciliano SD: **Structural equation modeling of a winnowed soil microbiome identifies how invasive plants re-structure microbial networks**. *ISME J* 2019, **13**:1988–1996.

52. Laforest-Lapointe I, Paquette A, Messier C, Kembel SW: **Leaf bacterial diversity mediates plant diversity and ecosystem function relationships**. *Nature* 2017, **546**:145–147.

53. Wang X, Wei Z, Yang K, Wang J, Jousset A, Xu Y, Shen Q, Friman V-P: **Phage combination therapies for bacterial wilt disease in tomato**. *Nat Biotechnol* 2019, **37**:1513–1520.

54. Romano R, Tomic O, Liland KH, Smilde A, Næs T: **A comparison of two PLS-based approaches to structural equation modeling**. *J Chemom* 2019, **33**:e3105.

55. Næs T, Romano R, Tomic O, Måge I, Smilde A, Liland KH: **Sequential and orthogonalized PLS (SO-PLS) regression for path analysis: Order of blocks and relations between effects**. *J Chemom* [date unknown], **n/a**:e3243.

56. Hwang H, Takane Y: **Nonlinear Generalized Structured Component Analysis**. *Behaviormetrika* 2010, **37**:1–14.

57. Murphy KP: *Machine learning: a probabilistic perspective.* MIT Press; 2012.