Characterization of the leaf microbiome from whole-genome sequencing data of the 3000 rice genomes project.

Veronica Roman-Reyna  
The Ohio State University  
https://orcid.org/0000-0003-0072-8096

Dale Pinili  
International Rice Research Institute

Frances Nikki Borja  
International Rice Research Institute

Ian Lorenzo Quibod  
International Rice Research Institute

Simon C. Groen  
New York University

Nikolai Alexandrov  
International Rice Research Institute

Ramil Mauleon  
Southern Cross University

Ricardo Oliva (r.oliva@irri.org)  
https://orcid.org/0000-0001-9517-5249

---

Short communication

**Keywords:** Oryza sativa, leaf microbiome, abundance network, GWAS, functional profile

**DOI:** https://doi.org/10.21203/rs.3.rs-35287/v3

**License:** This work is licensed under a Creative Commons Attribution 4.0 International License.  
Read Full License
Abstract

**Background:** The crop microbial communities are shaped by interactions between the host, microbes and the environment, however, their relative contribution is beginning to be understood. Here, we explore these interactions in the leaf bacterial community across 3,024 rice accessions.

**Findings:** By using unmapped DNA sequencing reads as microbial reads, we characterized the structure of the rice bacterial microbiome. We identified central bacteria taxa that emerge as microbial “hubs” and may have an influence on the network of host-microbe interactions. We found regions in the rice genome that might control the assembly of these microbial hubs. To our knowledge this is one of the first studies that uses raw data from plant genome sequencing projects to characterize the leaf bacterial communities.

**Conclusion:** We showed, that the structure of the rice leaf microbiome is modulated by multiple interactions among host, microbes, and environment. Our data provide insight into the factors influencing microbial assemblage in the rice leaf and also opens the door for future initiatives to modulate rice consortia for crop improvement efforts.

**Findings**

Plant colonization of terrestrial and aquatic habitats ignited the formation of biodiverse systems, termed phytobiomes. In phytobiomes, plants are in constant interaction with microbial communities that adapted to colonize plant tissues, termed microbiomes (Hassani et al. 2018). Microbial communities that live in association with plants carry a great diversity of metabolic capabilities and often influence broad aspects of plant biology. In agricultural environments, the composition of these communities affects overall crop performance by contributing to important plant functions such as nutrient uptake, environmental responses, and host development (Klein et al. 2012; Naylor et al. 2017; Edwards et al. 2018). For instance, rice seeds studies showed the microbiome as a potential source of plant beneficial bacteria and a source of microbes that could be vertically transmitted (Cottyn et al. 2009; Eyre et al. 2019). For roots and paddy soil microbiomes, several studies identified microbial clusters involved in methane metabolism and nitrogen fixation (Butler et al. 2003; Sessitsch et al. 2011; Bao et al. 2014). Reports on rice rhizosphere microbiome demonstrated associations with vegetative and reproductive host stages and as potential source of biocontrol agents (Spence et al. 2014; Edwards et al. 2015, 2018). Most rice microbiome studies use rRNA gene sequencing and a small sample size. This approach introduces a bias towards diversity and abundance of microorganisms. It also limited statistical power to identify the importance of factors that shape host-microbe interactions (Louca et al. 2018). An alternative to rRNA gene sequencing is shotgun sequencing with large-scale databases. This approach provides more information about the composition and structure of microbiome and also increases confidence about correlations among microbes, environment, and host.
As part of the 3,000 Rice Genomes Project (3K-RGP), 3,024 rice accessions were sequenced (The 3,000 rice genomes project 2014; Wang et al. 2018). The 3K-RGP panel has been successfully used to identify the rice genetic architecture underlying several complex morphological and phenological traits (Wang et al. 2018). These whole-genome-shotgun sequenced reads capture rice reads and also total DNA of resident microbial communities. The use of non-plant reads provides access to an impressive microbiome dataset in which one can systematically probe the role of environment vs. genotype in dictating microbial abundance. Based on that hypothesis, we extracted non-plant sequences from the 3K-RGP raw sequence data to characterize the rice leaf microbiome (Additional file 1: Figure S1. Additional file 2). We used bacterial and archaea reads since it constitutes 86% of the total reads.

**Diversity Of The Rice Leaf Microbiome**

We were able to capture and classify Bacterial and Archaea reads from the 3K-RGP raw database (Additional file 3: Table S1). On average rice leaves of each accession harbor 212 ± 111 genera (Figure 1A). The accumulation curve indicated that unmapped reads captured most of the expected taxa. The accumulation curve reached a plateau around 520 bacteria after the first 100 rice accessions (Figure 1B). The average Shannon index was 3.64 ± 1.2 similar to seed and root rice microbiome (Edwards et al. 2015; Eyre et al. 2019) (Figure 1C).

We assessed the taxonomic composition at phylum and Genus levels. We found nine Phyla and 23 genera with an abundance higher than 1% (Figure 1D). The phylum Proteobacteria, common in other leaf microbiomes, was the most abundant group (Rastogi et al. 2012; Bodenhausen et al. 2013; Wallace et al. 2018). Interestingly, the phylum Euryarchaeota, which includes methanogenic Archaea, was marginally present in the aerobic phyllosphere (Knief et al. 2012; Edwards et al. 2015) (Figure 1D). The aerobic conditions in the phyllosphere could also explain why we did not detect members of taxa commonly found in the soil or rice rhizosphere (Bao et al. 2014; Edwards et al. 2015). The genera *Pseudomonas*, *Xanthomonas*, *Mycoplasma*, and *Burkholderia* were the most abundant bacteria in the leaf microbiome (Figure 1D). *Pseudomonas*, *Xanthomonas*, and *Burkholderia* have rice pathogenic species. This could indicate that the rice microbiome has a mixture of commensals, beneficial and pathogenic species (Klein et al. 2012). The presence of *Mycoplasma* and other four genera associate with human or animal hosts might be explained by animal and human interaction with the rice cultivation (Additional file 3: Table S2). These interactions should be consider as a way of microbial horizontal acquisition (Sasaki et al. 1999; Cottyn et al. 2009; Campisano et al. 2014).

The 3K-RGP contained accessions grown in the Philippines (agPh) and in China (agCh). To further dissect the microbial community composition, we assessed taxa richness in each location. We found that the Shannon index for agCh was 4.08 while for agPh the index was 3.39 (Additional file 1: Figure S2). For genus richness, we found that that agCh contains on average 186 genera while agPh contained 152 genera. Then we compared the relative abundance of Bacteria and Archaea at different taxonomical levels. The differences between locations were mainly explained by genera. Forty-eight genera contributed to 70% of the differences between locations; 25 of them had a relative abundance higher than
1% (Additional file 3: Table S3). Other studies have shown that environmental variation appears to be the major driver of microbiome diversity (Peiffer et al. 2013; Okubo et al. 2014; Copeland et al. 2015; Edwards et al. 2015, 2018; Wagner et al. 2016; Moronta-Barrios et al. 2018). Thus, it is likely that the differences between agCh and agPh might be associated with the exposed to an distinct array of microbes.

We were aware that the manipulation of samples during gDNA extraction or sequencing might include bacteria not commonly present in rice. To rule out that abundant genera were not artificially introduced during sample collection, we used qPCR to detect 11 highly abundant genera in 17 rice accessions from our 3K-RGP panel (Additional file 2; Additional file 3: Table S4). We used new plants with the idea that the genera we found in 3K panel are common members of the rice leaf microbiome. We were able to quantify the presence of all taxa and observed a similar distribution across accessions (Additional file 3: Table S4). Similar to our previous findings, the genera *Pseudomonas*, *Xanthomonas*, and *Mycoplasma* were the most abundant genera, ruling out that highly abundant genera were introduced artificially. Overall, our data showed that environmental variation play a key role in determining variation in microbial community composition (Hartmann et al. 2015).

### The Functional Profile Of The Leaf Microbiome

Networks of interactions among microbes further shape the establishment and maintenance of the microbial community (Horton et al. 2014; Layeghifard et al. 2017; Hassani et al. 2018). In those networks, highly connected genera or “hubs” play an important ecological role in the establishment of the community and the regulation of microbial assembly (Agler et al. 2016; Layeghifard et al. 2017; Hassani et al. 2018). To determine the structure of the rice leaf microbiome, we inferred microbial co-occurrence networks and identified critical hubs (Additional file 2). The criteria to select the hubs was based on a combination of the network properties such as weighted degree, betweenness centrality, modularity class, clustering, and eigen-centrality (Additional file 3: Table S5). We found that the rice microbe network can be defined by 12 hubs: *Clostridium*, *Mycoplasma*, *Bacillus*, *Buchnera*, *Prochlorococcus*, *Helicobacter*, *Methylobacterium*, *Chamaesiphon*, *Azotobacter*, *Kineococcus*, *Acidovorax*, and *Pseudomonas* (Figure 2A, Additional file 3: Table S5). Interestingly, connectivity of a genus within the network did not correlate with their abundance. For example, the highly abundant genera *Xanthomonas* and *Burkholderia* were not identified as hubs. The genera *Kineococcus* or *Helicobacter*, with less than 1% abundance, played a role in shaping the network of interactions (Agler et al. 2016).

The hubs *Methylobacterium*, *Pseudomonas*, *Bacillus*, *Kineococcus*, *Azotobacter*, *Acidovorax*, and *Clostridium* have also been reported as commensal, beneficials, or pathogens in other plants (Cottyn et al. 2009; Bian et al. 2012; Horino et al. 2015; Chen et al. 2018; Lai and Huang 2018). The hubs *Chamaesiphon* and *Prochlorococcus* could be part of the rice microbiome, as the paddy field conditions like shallow water and sunlight could create an ideal environment for algae growth. The hub *Buchnera*, could be part of aphids and rice interaction, as *Buchnera* is an aphid endosymbiont. The hubs *Mycoplasma* and *Helicobacter*, associated with human diseases, have not been associated with plant microbiomes. We hypothesized that these bacteria, generally present in animals and humans, are hubs in
the rice microbiome due to a strong influence of agricultural practices, like irrigation water or crop-human interaction.

The networks can be further sub-structured into modules, which represent a group of organisms that has more interactions within the group than with other members of the network (Layeghifard et al. 2017). We found seven modules. Based on the bacterial composition of some modules, microbial genetic ancestry or ecological niche seems to shape them (Figure 2A, Additional file 3: Table S6). For example, one module was enriched with Cyanobacteria, another module had bacteria associated to animals and other modules mainly had Alphaproteobacteria or Betaproteobacteria associated with plants or soils (Additional file 3: Table S6). Overall, taxonomy status or ecological niche rather than the abundance of individual taxa appears to define the interactions within the microbial community. Moreover the number of modules in the network suggested a highly stable network since a microbial community appears to reach an equilibrium when its network of interactions had a small number of modules (Layeghifard et al. 2017).

The functional and nutritional capacities of the microbes partly define the networks of interactions among microbes (Agler et al. 2016; Layeghifard et al. 2017; Hassani et al. 2018). To understand if biological functions in the bacterial communities associate to the ecological network, we predicted functional categories for the microbial taxa (Additional file 2). We found 85 predicted KEGG functional categories, where 28 had more than 1% abundance (Figure 2B, Additional file 3: Table S7). Transcription, translation, primary metabolism, flagellar assembly, environmental adaptation (ECM receptor interaction) and secondary metabolism (terpenoids, antibiotics, and xenobiotics) were the most abundant pathways. The enrichment of these KEGG pathways linked with the abundant genera as well as the bacterial hubs. For example, the enrichment of pathways related to human diseases reflected the presence of animal pathogens in the rice microbiome. Moreover, *Pseudomonas*, facultative anaerobes, methanogenic bacteria, and some gram-positive bacteria, like *Bacillus* and *Streptomyces*, have the capacity to metabolize xenobiotic, terpenoids and antibiotics compounds (Additional file 3: Table S2). The xenobiotics metabolism KEGG category suggested an adaptation of the bacterial to an environment where chemicals, like pesticides, might be applied to crop fields (Tipayno et al. 2017). The presence of functional categories common to other leaf microbiomes, restated the idea that functions and communities are not random (Delmotte et al. 2009; Xiao et al. 2017). Moreover, our results aligned with the idea that metabolic functions and interaction within the microbiome regulate the establishment of the community (Agler et al. 2016; Louca et al. 2017).

We then determined the microbial co-occurrence networks and functional categories for each agCh and agPh dataset. Interestingly, agCh and agPh assembled networks with similar average degree and modularity (Additional file 1: Figure S3A). We identified seven highly connected hubs in each dataset, five of which were common to agCh and agPh networks (Additional file 1: Figure S3A, Figure 2). Furthermore, we found no significant differences between agCh and AgPh functional profiles, where both datasets shared 22 of 24 KEGG level pathways (Additional file 1: Figure S3B). The data suggest that rice leaf microbiome assemble communities with similar structure, independently from the available microbial diversity. Overall, the similarity of the network structures and the functional redundancy of the leaf
microbiome dataset supported the idea that key microbial groups might regulate the establishment by providing essential functions in the community (Hassani et al. 2018). This result also aligned with other studies where core biological functions of the microbiome are associated to different plant tissues or plant substrates (Vorholt 2012; Agler et al. 2016; Louca et al. 2017; Xiao et al. 2017).

**Rice Genetic Associations With The Leaf Microbiome Composition**

The interactions of environment, microbes, and hosts modulate the assembly of the microbial community (Layeghifard et al. 2017; Hassani et al. 2018; Wallace et al. 2018). The host might play a key role, as genetic control on the microbiome assembly has been reported in Arabidopsis and *Nicotiana benthamiana* (Long et al. 2010; Lebeis et al. 2015). To identify rice genetic factors that control the recruitment and establishment of specific microbial players, we conducted a genome-wide association study (GWAS) on 3,024 rice accessions, using the genomic information from 6.5 million SNPs and the relative abundance of the 12 hubs (Additional file 2). To avoid association bias due to lack of information for most of the Chinese accessions, we kept agCh and agPh together for the analysis.

Overall, we found 22 significant SNPs shared among the 12 hubs abundance (P-value < 1E-16, ci = 0.95), distributed across six chromosomes (Figure 3A). Twenty of the 22 SNPs located within 11 annotated rice genes (Additional file 3: Table S7). Seven genes had at least one SNPs with a missense effect. The seven genes were catalase isozyme A, malate synthase, inorganic H+ pyrophosphatase, endo-1,4-beta-xylanase, similar to ClpC, similar to ATP-dependent Clp protease, and 5-methyltetrahydropteroyltriglutamate–homocysteine methyltransferase. Most of the seven genes had annotations for stress responses, carbon metabolism and regulation of gene expression (Additional file 3: Table S8). We then determined the relationships between significant SNPs and hubs abundance (Additional file 2). The average of the 12 hub abundances (centered log-ratio normalized) showed significant differences in four haplotypes, Chr4-32927447, Chr5-14856070, Chr5-14856078, and Chr5-25806324 (Additional file 1: Figure S4A). The SNPs related to chromosome five were the genes endo-1,4-beta-xylanase and a heat shock protein (Additional file 3: Table S8). We also looked at the relationship between significant SNPs and the hubs with higher betweenness centrality, *Mycoplasma, Clostridium* and *Bacillus*. The three bacteria had significant differences for the SNP Chr4-32927447. *Mycoplasma* had significant differences with 18 more haplotypes. *Clostridium* with one more and *Bacillus* two more. Based on the 22 significant SNPs and the differences found with hub abundances, it seems the regulation of the microbial community could be associated with less toxic or carbon-enriched host environments.

From the 22 SNPs, we identified 12 haplotype blocks spanning 120 candidate genes using linkage disequilibrium analysis (Additional file 2, Additional file 3: Table S8). Gene ontology (GO) enrichment analysis, with the 120 genes, indicated physiological process, intracellular, membrane-bound organelle and catalytic activity as the most abundant (>50%) GO terms (Additional file 2, Additional file 1: Figure S5). To explore if the 120 gene were enriched and previously described to be associated with rice agronomic traits, we used the rice quantitative trait locus (QTL) database, Q-TARO. Overall, the 120 candidate genes mapped to 42 QTLs distributed in different categories: resistance or tolerance (15 QTLs),...
morphological traits (16 QTLs), and physiological traits (11 QTLs) (Figure 3B, Additional file 3: Table S9). Drought tolerance was the most abundant trait in resistance or tolerance QTL. The presence of drought QTL followed by other traits related to abiotic stimuli validated the idea that host responses to the environment can be associated with the microbiome composition (Long et al. 2010; Lebeis et al. 2015; Naylor et al. 2017). The morphological traits Culm/leaf, Panicle/flower, and seed traits were equally abundant. The traits associated with different plant tissues validate the idea of a dynamic microbiome that shifts in different host tissue or developmental conditions, like vegetative or reproductive (Edwards et al. 2018). The seeds trait, the eating quality as physiological trait, and the QTL associated with grain chalkiness, suggested that seeds quality are a key factor in the vertical transmission and shaping of the microbiome (Cottyn et al. 2009; Eyre et al. 2019). Similar to other studies, we found that host responses to stress, primary pathways, and plant tissue morphology, are perhaps, common host genetic factors related to leaf microbiome assembly (Horton et al. 2014; Wallace et al. 2018). In that scenario, is likely that allelic variation of certain rice genes influences the composition of plant microbiome or of particular groups.

Conclusions

Our study explored the idea that information about the composition of leaf microbial communities of rice plants can be extracted from the raw host genome sequence data. Using the 3K-RGP dataset, we were able to describe the composition and structure of the rice leaf microbiome. We validated the idea that microbiomes do not assemble randomly but that their formation is governed by complex interactions among microbes, host, and environment. Given the scale of the 3K-RGP dataset, we took the first steps in unearthing the factors behind rice leaf microbiome assembly by using GWAS and microbiome abundance as a trait. The next steps will be to understand how the microbiome from roots, soil, seeds, and leaves transmit among tissues and the interaction with fungal and virus microbiome. This study leaves open questions on the benefit of these hubs but also on the host mechanisms that can be used to modulate the community for crop improvement purposes.

Abbreviations

GWAS: Genome-wide association studies.

rRNA: Ribosomal ribonucleic acid.

3K-RGP: 3000 Rice Genomes Project.

qPCR: real-time polymerase chain reaction.

SNPs: single-nucleotide polymorphisms.

QTL: quantitative trait locus.
KEGG: Kyoto Encyclopedia of Genes and Genomes.

FDR: False discovery rate.

GO: gene ontology

**Declarations**

**Ethical Approval and Consent to participate**

Not applicable.

**Consent for publication**

Not applicable.

**Availability of supporting data**

All data generated or analyzed during this study are included within figures, tables and the supplemental material. The 3K SNPs were downloaded from [https://snp-seek.irri.org/](https://snp-seek.irri.org/)

**Competing interests**

The authors declare no competing interests.

**Funding**

DP was funded by the DOST-ASTHRDP program. VRR was funded by the Newton Fund. S.C.G. was supported by a grant from the Gordon and Betty Moore Foundation (Life Sciences Research Foundation Postdoctoral Fellowship Grant GBMF2550.06 to S.C.G.)

**Authors' contributions**

NA, RO designed research; VR-R, DL, FNB, ILQ, RM performed research; VR-R, DL, RM, RO analyzed data; VR-R, SCG, RO wrote the paper.

**Acknowledgements**

Authors would like to thank the Rice CRP and Department of Science and Technology - Advanced Science and Technology Institute (DOST-ASTI) for the free access of high-performance computing services.

**References**

Agler MT, Ruhe J, Kroll S, et al (2016) Microbial hub taxa link host and abiotic factors to plant microbiome variation. PLOS Biol 14:e1002352. https://doi.org/10.1371/journal.pbio.1002352
Bao Z, Okubo T, Kubota K, et al (2014) Metaproteomic Identification of Diazotrophic Methanotrophs and Their Localization in Root Tissues of Field-Grown Rice Plants. Appl Environ Microbiol 80:5043–5052. https://doi.org/10.1128/AEM.00969-14

Bian G-K, Feng Z-Z, Qin S, et al (2012) Kineococcus endophytica sp. nov., a novel endophytic actinomycete isolated from a coastal halophyte in Jiangsu, China. Antonie Van Leeuwenhoek 102:621–628. https://doi.org/10.1007/s10482-012-9757-4

Bodenhausen N, Horton MW, Bergelson J (2013) Bacterial communities associated with the leaves and the roots of Arabidopsis thaliana. PLoS ONE 8:. https://doi.org/10.1371/journal.pone.0056329

Butler JL, Williams MA, Bottomley PJ, Myrold DD (2003) Microbial Community Dynamics Associated with Rhizosphere Carbon Flow. Appl Environ Microbiol 69:6793–6800. https://doi.org/10.1128/AEM.69.11.6793-6800.2003

Campisano A, Antonielli L, Pancher M, et al (2014) Bacterial endophytic communities in the grapevine depend on pest management. PLOS ONE 9:e112763. https://doi.org/10.1371/journal.pone.0112763

Chen S-L, Tsai M-K, Huang Y-M, Huang C-H (2018) Diversity and characterization of Azotobacter isolates obtained from rice rhizosphere soils in Taiwan. Ann Microbiol 68:17–26. https://doi.org/10.1007/s13213-017-1312-0

Copeland JK, Yuan L, Layeghifard M, et al (2015) Seasonal community succession of the phyllosphere microbiome. Mol Plant Microbe Interact 28:274–285. https://doi.org/10.1094/MPMI-10-14-0331-FI

Cottyn B, Debode J, Regalado E, et al (2009) Phenotypic and genetic diversity of rice seed-associated bacteria and their role in pathogenicity and biological control. J Appl Microbiol 107:885–897. https://doi.org/10.1111/j.1365-2672.2009.04268.x

Delmotte N, Knief C, Chaffron S, et al (2009) Community proteogenomics reveals insights into the physiology of phyllosphere bacteria. Proc Natl Acad Sci U S A 106:16428–16433. https://doi.org/10.1073/pnas.0905240106

Edwards J, Johnson C, Santos-Medellín C, et al (2015) Structure, variation, and assembly of the root-associated microbiomes of rice. Proc Natl Acad Sci 112:E911–E920. https://doi.org/10.1073/pnas.1414592112

Edwards JA, Santos-Medellín CM, Liechty ZS, et al (2018) Compositional shifts in root-associated bacterial and archaeal microbiota track the plant life cycle in field-grown rice. PLOS Biol 16:e2003862. https://doi.org/10.1371/journal.pbio.2003862

Eyre AW, Wang M, Oh Y, Dean RA (2019) Identification and Characterization of the Core Rice Seed Microbiome. Phytobiomes J 3:148–157. https://doi.org/10.1094/PBIOMES-01-19-0009-R
Hartmann M, Frey B, Mayer J, et al (2015) Distinct soil microbial diversity under long-term organic and conventional farming. ISME J 9:1177–1194. https://doi.org/10.1038/ismej.2014.210

Hassani MA, Durán P, Hacquard S (2018) Microbial interactions within the plant holobiont. Microbiome 6:58. https://doi.org/10.1186/s40168-018-0445-0

Horino H, Ito M, Tonouchi A (2015) Clostridium oryzae sp. nov., from soil of a Japanese rice field. Int J Syst Evol Microbiol 65:943–951. https://doi.org/10.1099/ijs.0.000042

Horton MW, Bodenhausen N, Beilsmith K, et al (2014) Genome-wide association study of Arabidopsis thaliana's leaf microbial community. Nat Commun 5:5320. https://doi.org/10.1038/ncomms6320

Klein E, Ofek M, Katan J, et al (2012) Soil suppressiveness to Fusarium disease: shifts in root microbiome associated with reduction of pathogen root colonization. Phytopathology 103:23–33. https://doi.org/10.1094/PHYTO-12-11-0349

Knief C, Delmotte N, Chaffron S, et al (2012) Metaproteogenomic analysis of microbial communities in the phyllosphere and rhizosphere of rice. ISME J 6:1378–1390. https://doi.org/10.1038/ismej.2011.192

Lai Y-R, Huang C-J (2018) First report of Acidovorax avenae subsp. avenae causing bacterial brown stripe disease of rice in Taiwan. J Plant Pathol 100:595–595. https://doi.org/10.1007/s42161-018-0100-4

Layeghifard M, Hwang DM, Guttman DS (2017) Disentangling interactions in the microbiome: A network perspective. Trends Microbiol 25:217–228. https://doi.org/10.1016/j.tim.2016.11.008

Lebeis SL, Paredes SH, Lundberg DS, et al (2015) Salicylic acid modulates colonization of the root microbiome by specific bacterial taxa. Science 349:860–864. https://doi.org/10.1126/science.aaa8764

Liu L, Mei Q, Yu Z, et al (2013) An Integrative Bioinformatics Framework for Genome-scale Multiple Level Network Reconstruction of Rice. J Integr Bioinforma 10:94–102. https://doi.org/10.1515/jib-2013-223

Long HH, Sonntag DG, Schmidt DD, Baldwin IT (2010) The structure of the culturable root bacterial endophyte community of Nicotiana attenuata is organized by soil composition and host plant ethylene production and perception. New Phytol 185:554–567. https://doi.org/10.1111/j.1469-8137.2009.03079.x

Louca S, Doebeli M, Parfrey LW (2018) Correcting for 16S rRNA gene copy numbers in microbiome surveys remains an unsolved problem. Microbiome 6:41. https://doi.org/10.1186/s40168-018-0420-9

Louca S, Jacques SMS, Pires APF, et al (2017) High taxonomic variability despite stable functional structure across microbial communities. Nat Ecol Evol 1:0015. https://doi.org/10.1038/s41559-016-0015

Moronta-Barrios F, Gionechetti F, Pallavicini A, et al (2018) Bacterial microbiota of rice roots: 16S-based taxonomic profiling of endophytic and rhizospheric diversity, endophytes isolation and simplified endophytic community. Microorganisms 6:14. https://doi.org/10.3390/microorganisms6010014
Naylor D, DeGraaf S, Purdom E, Coleman-Derr D (2017) Drought and host selection influence bacterial community dynamics in the grass root microbiome. ISME J 11:2691–2704. https://doi.org/10.1038/ismej.2017.118

Okubo T, Ikeda S, Sasaki K, et al (2014) Phylogeny and functions of bacterial communities associated with field-grown rice shoots. Microbes Environ 29:329–332. https://doi.org/10.1264/jsme2.ME14077

Peiffer JA, Spor A, Koren O, et al (2013) Diversity and heritability of the maize rhizosphere microbiome under field conditions. Proc Natl Acad Sci 110:6548–6553. https://doi.org/10.1073/pnas.1302837110

Rastogi G, Sbodio A, Tech JJ, et al (2012) Leaf microbiota in an agroecosystem: spatiotemporal variation in bacterial community composition on field-grown lettuce. ISME J 6:1812–1822. https://doi.org/10.1038/ismej.2012.32

Sasaki K, Tajiri Y, Sata M, et al (1999) Helicobacter pylori in the natural environment. Scand J Infect Dis 31:275–279

Sessitsch A, Hardoim P, Döring J, et al (2011) Functional Characteristics of an Endophyte Community Colonizing Rice Roots as Revealed by Metagenomic Analysis. Mol Plant Microbe Interact 25:28–36. https://doi.org/10.1094/MPMI-08-11-0204

Spence C, Alff E, Johnson C, et al (2014) Natural rice rhizospheric microbes suppress rice blast infections. BMC Plant Biol 14:130. https://doi.org/10.1186/1471-2229-14-130

The 3,000 rice genomes project (2014) The 3,000 rice genomes project. GigaScience 3:7. https://doi.org/10.1186/2047-217X-3-7

Tipayno SC, Truu J, Samaddar S, et al (2017) The bacterial community structure and functional profile in the heavy metal contaminated paddy soils, surrounding a nonferrous smelter in South Korea. Ecol Evol 8:6157–6168. https://doi.org/10.1002/ece3.4170

Vorholt JA (2012) Microbial life in the phyllosphere. Nat Rev Microbiol Lond 10:828–40. http://dx.doi.org.virtual.anu.edu.au/10.1038/nrmicro2910

Wagner MR, Lundberg DS, del Rio TG, et al (2016) Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. Nat Commun 7:12151. https://doi.org/10.1038/ncomms12151

Wallace J, Kremling KA, Kovar LL, Buckler ES (2018) Quantitative genetics of the maize leaf microbiome. Phytobiomes. https://doi.org/10.1094/PBIOMES-02-18-0008-R

Wang W, Mauleon R, Hu Z, et al (2018) Genomic variation in 3,010 diverse accessions of Asian cultivated rice. Nature 557:43. https://doi.org/10.1038/s41586-018-0063-9
Figures

Figure 1

Composition of leaf bacterial community in the 3K rice genome project. A, B total leaf bacterial and archaea genera (A) and accumulation curves (B) in 3024 rice genome accessions. The shade on the curve represents the confidence interval of two in the curve points. C, Shannon index in all rice accessions. In the violin plots, the black box plot indicates the 75th and 35th percentile. The white dot
represents the media. D, Relative abundance at Phylum and genera level. The inner position of the sunburst chart represents Phylum and the outer position represents Genus. The chart shows average relative abundance higher than 1% across all samples.

Figure 2

Structure and functional profile of the rice leaf bacterial communities. A Microbial ecological network from the 3K microbiome with abundant genera present in at least 50% of all samples. The colors represent the seven modules in the network. Each node represents a genus and the circle size indicates betweenness centrality increment. For the network analysis the genus counts were center-log-transformed. B. KEGG pathways entries with more than 1% relative abundance in all rice accessions. The inner circle indicates KEGG class 1 pathways and the external circle indicates classes 2 or 3.
Rice metabolic pathways are associated with the microbiome structure. A Genome-wide association for the 12 microbial hubs simultaneously using a multivariate linear mixed model with GEMMA software. Left panel, Manhattan plot, using the rice dataset, indicated the major peaks (significant SNPs) distributed across six chromosomes, associated with microbial abundance of the 12 hubs. Right panel, Quartile–quartile plot for expected versus observed $-\log(P)$ values. P-values were adjusted with FDR and values lower than 1E-15 were considered significant (blue line). B, the significant SNPs found in this study co-localized with a number of agronomic QTLs retrieved from Q-TARO database, associated with resistance or tolerance; morphological trait; and physiological trait.

**Figure 3**

**Supplementary Files**

This is a list of supplementary files associated with this preprint. Click to download.

- Additionalfile1.pdf
- additionalfile2.pdf
- supplementaltablesV4.xls