Author/s:
Xiong, C;Zhu, Y-G;Wang, J-T;Singh, B;Han, L-L;Shen, J-P;Li, P-P;Wang, G-B;Wu, C-F;Ge, A-H;Zhang, L-M;He, J-Z

Title:
Host selection shapes crop microbiome assembly and network complexity

Date:
2021-01

Citation:
Xiong, C., Zhu, Y. -G., Wang, J. -T., Singh, B., Han, L. -L., Shen, J. -P., Li, P. -P., Wang, G. -B., Wu, C. -F., Ge, A. -H., Zhang, L. -M. & He, J. -Z. (2021). Host selection shapes crop microbiome assembly and network complexity. New Phytologist, 229 (2), pp.1091-1104. https://doi.org/10.1111/nph.16890.

PROF. BRAJESH K SINGH (Orcid ID : 0000-0003-4413-4185)

PROF. LI-MEI ZHANG (Orcid ID : 0000-0002-7383-8475)

Article type : MS - Regular Manuscript

Host selection shapes crop microbiome assembly and network complexity

Chao Xiong[1,2], Yong-Guan Zhu[1,3], Jun-Tao Wang[1], Brajesh Singh[4,5], Li-Li Han[1,2], Ju-Pei Shen[1,2], Pei-Pei Li[6], Gui-Bao Wang[7], Chuan-Fa Wu[1,6], An-Hui Ge[1,2], Li-Mei Zhang[1,2*], and Ji-Zheng He[1,2]

[1]State Key Laboratory of Urban and Regional Ecology, Research Center for Eco-Environmental Sciences, Chinese Academy of Sciences, Beijing 100085, China

[2]University of Chinese Academy of Sciences, Beijing 100049, China

[3]Key Laboratory of Urban Environment and Health, Institute of Urban Environment, Chinese Academy of Sciences, Xiamen 361021, China

[4]Global Centre for Land-Based Innovation, Western Sydney University, Penrith South DC, NSW 2751, Australia

[5]Hawkesbury Institute for the Environment, Western Sydney University, Penrith, New South Wales 2751, Australia

[6]College of Resource and Environmental Sciences, Henan Agricultural University, Zhengzhou 450002, China

[7]Soil and Fertilizer Station of Qilin District, Qujing, Yunnan Province, Qujing, 655000, China

*Author for correspondence: Li-Mei Zhang
Email: zhanglm@rcees.ac.cn, Tel: +86-10-6295 3251

Received: 28 March 2020
Accepted: 13 August 2020

ORCID:
Chao Xiong (0000-0002-3023-0494)
Yong-Guan Zhu (0000-0003-3861-8482)
Jun-Tao Wang (0000-0002-1822-2176)
Brajesh Singh (0000-0003-4413-4185)
Li-Li Han (0000-0002-8105-1672)
Li-Mei Zhang (0000-0002-7383-8475)
Ji-Zheng He (0000-0002-9169-8058)

Summary

- Plant microbiomes are essential to host health and productivity but the ecological processes that govern crop microbiome assembly are not fully known.
- Here we examined bacterial communities across 684 samples from soils (rhizosphere and bulk soil) and multiple compartment niches (rhizoplane, root endosphere, phylloplane, and leaf endosphere) in maize (*Zea mays*)-wheat (*Triticum aestivum*)/barley (*Hordeum vulgare*) rotation system under different fertilization practices at two contrasting sites.
- Our results demonstrate that microbiome assembly is shaped predominantly by compartment niche and host species rather than by site or fertilization practice. From soils to epiphytes to endophytes, host selection pressure sequentially

increased and bacterial diversity and network complexity consequently reduced, with the strongest host effect in leaf endosphere. Source tracking indicates that crop microbiome is mainly derived from soils and gradually enriched and filtered at different plant compartment niches. Moreover, crop microbiomes were dominated by a few dominant taxa (~0.5% of bacterial phylotypes), with Bacilli identified as the important biomarker taxa for wheat and barley and *Methylobacteriaceae* for maize.

- Our work provides comprehensive empirical evidence on host selection, potential sources and enrichment processes for crop microbiome assembly, and has important implications for future crop management and manipulation of crop microbiome for sustainable agriculture.

Keywords: Plant microbiomes, Community assembly, Plant-microbe interactions, Co-occurrence networks, Compartment niches, Soil-plant continuum, Fertilization practices

Introduction

Plants are heavily inhabited by diverse microbes in every compartment, including root, stem, leaf, flower, and fruit (Lindow & Brandl, 2003; Hacquard *et al.*, 2015). Plant and their inhabiting microbiome collectively constitute a "holobiont" in which they have been interacting and evolving with each other (Vandenkoornhuyse *et al.*, 2015; Hassani *et al.*, 2018). Similar to the human microbiome, the plant microbiome acts as a secondary genome and is linked with host fitness (Vandenkoornhuyse *et al.*, 2015; Muller *et al.*, 2016; Martin *et al.*, 2017). Harnessing the plant microbiome to maximize crop production is increasingly considered as a viable sustainable future approach (Geddes *et al.*, 2015; Qiu *et al.*, 2019). One of the pre-requisites for the

development of such effective technologies is to improve the fundamental understanding of ecological processes that govern microbiome assembly (Singh & Trivedi, 2017; Toju *et al.*, 2018).

Plant bacterial community is diverse and contributes to numerous aspects of the plant health (Bulgarelli *et al.*, 2013; Liu *et al.*, 2019). For example, previous studies on *Arabidopsis thaliana* (Durán *et al.*, 2018), grapevine (Rolli *et al.*, 2015), and citrus (Zhang *et al.*, 2017) have demonstrated that bacterial community plays an essential role for plant growth through a variety of mechanisms, including increasing nutrient acquisition, promoting plant hormone production, and protecting plant against pathogen attacks (Ritpitakphong *et al.*, 2016; Alvarez-Perez *et al.*, 2017; Hassani *et al.*, 2018). These studies suggest that plant and its microbiomes have co-evolved for millions of years and most of these interactions are mutually beneficial (Foster *et al.*, 2017; Cordovez *et al.*, 2019) but influenced by environmental conditions (Philippot *et al.*, 2013; Hassani *et al.*, 2018). The theoretical framework of co-evolution implies that plants attract and select beneficial microbiomes via releasing signal molecules to attract target microbiota and then apply selection pressure using immune systems and provision of specialized nutrient and habitat types (Foster *et al.*, 2017; Martin *et al.*, 2017; Cordovez *et al.*, 2019). Microbes able to recognize the signal molecules and colonize specific compartment niches are preferentially enriched while others are filtered out (Hacquard *et al.*, 2015; Muller *et al.*, 2016). However, how host and environment shape microbiome assemblies and co-occurrence patterns across rhizosphere, phyllosphere and endosphere remains largely unknown.

Host plants provide microbes with multiple microhabitats, including diverse host organs, plant external tissues (epiphytes), and internal tissues (endophytes) (Muller *et al.*, 2016; Vorholt *et al.*, 2017; Edwards *et al.*, 2018a). Soil habitats serve as an extremely rich microbial reservoir for host microbial selection (Bulgarelli *et al.*, 2013; Zarraonaindia *et al.*, 2015; Mangeot-Peter *et al.*, 2020). Previous studies have demonstrated that different plant compartments are colonized by distinct microbial

communities (Uroz *et al.*, 2010; Bulgarelli *et al.*, 2012; Hamonts *et al.*, 2017; Cregger *et al.*, 2018). The successful colonization of microbes in a specific niche requires microbes have the capability to overcome host immune system and abiotic stresses in different microhabitats (Vorholt, 2012; Bulgarelli *et al.*, 2013). Most recent studies focusing on root and rhizosphere have suggested that microbial communities are shaped by multiple factors, including plant genetics and age, soil types, and nutrients (Barnes *et al.*, 2016; Walters *et al.*, 2018; Yu *et al.*, 2018; Emmett *et al.*, 2020). Yet we still know little about the plant microbiome assemblies along the soil-plant continuum including epiphytes and endophytes, particularly for crops growing in fields in manipulative management setting.

In addition to exposure to biotic stress from host, microbial communities in agricultural ecosystems are usually exposed to various climate, edaphic conditions and agronomic management regimes, which have an inevitable influence on crop microbiomes. For instance, a couple of studies have suggested that different soil properties (e.g. soil pH and nutrients) and agricultural regimes (e.g. management type and tillage intensity) could significantly influence the community composition of root and soil microbiomes (Peiffer *et al.*, 2013; Hartman *et al.*, 2018; Toju *et al.*, 2018; Yu *et al.*, 2018). All these suggested that crop-associated microbiomes were shaped by multiple biotic and environmental factors, while most of the related studies mainly considered one or a few of aspects. A systematic understanding of how compartment niche, host species, edaphic factors and agricultural management perturbations interactively drive the assembly of crop microbiome is still scarce. Observations based on field practice suggested that straw retention (Sun *et al.*, 2015), application of nitrification inhibitor (Wu *et al.*, 2018), biochar amendment (Edwards *et al.*, 2018b), and transplantation of beneficial microbes (Van der Heijden *et al.*, 1998; Christian *et al.*, 2019) could effectively improve soil fertility and plant productivity. These practices are proposed as complimentary measures to reduce the use of inorganic N fertilizers for sustainable agriculture. However, little is known about the effects of these practices on

the assemblies of crop-associated microbiomes, particularly on the main cereal crops.

Here we established a field experiment with seven different fertilization regimes in two major agricultural areas, corresponding to two contrasting soil types in China, and examined bacterial community across 684 samples from soils (rhizosphere and bulk soil) and multiple compartment niches (rhizoplane, root endosphere, phylloplane, and leaf endosphere) of three crops (maize, *Zea mays*; wheat, *Triticum aestivum;* barley, *Hordeum vulgare*). We aim to (1) assess how host (compartment niche and host species) and environmental factors (site and fertilization practice) interactively shape crop microbiome assembly and co-occurrence patterns across rhizosphere, phyllosphere and endosphere; (2) to identify the potential sources and keystone taxa of crop microbiomes along the soil-plant continuum. We hypothesized that (1) the relative contribution of host effect and environmental effect on crop microbiome assembly will shift across soils to epiphytes to endophytes, and the microbial diversity and network complexity will decrease correspondingly due to increasing host effect, and (2) crop microbiome will be gradually enriched and filtered from bulk soil to endosphere and endosphere niches may select microbial taxa from nearby species pool.

Materials and methods

Field trial and treatments

Two field trials were established in Xuchang (XC, 34°08'20.4"N, 113°48'34.9"E, 79 meters above sea level, masl) in Henan province (northern China), and Qujing (QJ, 25°09'40.8"N, 104°01'51.5"E, 1925 masl) in Yunnan province (southwest China) in spring of 2016. The two study sites are about 1800 km far away, and soil at XC was classified as calcaric cambisols with pH at about 7.5 and QJ as chromic cambisols with pH at about 5.0, according to the FAO soil classification system (IUSS Working Group WRB, 2007). XC site has a temperate monsoonal climate with an annual mean precipitation (AMP) of 704 mm, and a month mean temperature (MMT) of 26.1 °C in summer and 3.6 °C in winter. QJ site has a subtropical monsoon climate with an AMP of 1,008 mm, a MMT of 19.8 °C in summer and 9.3 °C in winter.

The fertilization trials were established in spring of 2016, including seven treatments in both sites: (1) control without nitrogen fertilizer (Control); (2) fertilization with urea (N), superphosphate (P) and potassium sulfate (K) in light of local regime (N); (3) 20% N reduction on the basis of N treatment (80%N); (4) 20% N reduction plus straw addition at a rate of 3000 kg ha^{-1} (80%NS); (5) 20% N reduction plus nitrification inhibitor chlorinated pyridine (CP) application at a rate of 735 g ha^{-1} (80%NI); (6) 20% N reduction plus foliar spraying of an asymbiotic nitrogen-fixing bacteria *Klebsiella variicola* W12 at a rate of 500 L ha^{-1} (1×10^{12} CFU/ml) (80%NKle); (7) 80%NS treatment plus biochar addition at a rate of 30,000 kg ha^{-1} (80%NSB). For both sites, the doses of P and K fertilizer were kept identical for all treatments at a rate of 90 kg ha^{-1} P_2O_5 and 90 kg ha^{-1} K_2O for each crop season. N fertilizer was applied at a rate of 200 kg N ha^{-1} in N treatment and a rate of 160 kg N ha^{-1} in all 80%N treatments. Each treatment had three replicate plots, and each plot was about 30 m^2 and surrounded by a buffer strip of about 1.5 m. All plots were randomly arranged in field and managed according to local habits. Agronomic practices for fertilization were detailed in supplementary materials "Method S1".

Due to the contrasting climate and edaphic features in two sites, field trials were managed following the local regime, with maize-wheat rotation in XC site and maize-barley rotation in QJ site. The crop varieties planted were the same as those used by local farmers, with *Zea mays* cultivar Zhengdan 958 and *Triticum aestivum* cultivar Qiule 2122 sowed in XC, and *Zea mays* cultivar Shidan 8 and *Hordeum vulgare* cultivar V43 in QJ.

Sample collection

Maize sampling was performed in September 2016 (n = 216, including the first 6 treatments) and in August 2017 (n = 252, including all 7 treatments) for both study sites, respectively. In wheat/barley growing-season, sampling was performed in March 2017 (n = 216, including the first 6 treatments) (Fig. S1). For leaves and roots sampling, about 5 individual maize plants and 20-30 individual wheat/barley plants were

randomly selected from each plot, and 1-2 leaves at mid-upper position of the plant were clipped and immediately placed on ice bag (totally 50-100 g fresh weight for each plot). Rhizosphere soil (defined as those tightly attached to the roots) and roots (10-20 g fresh weight for each plot) of the same plant were collected afterwards. The topsoil (0-15 cm) ~20cm away from roots was collected as bulk soil, with five subsamples thoroughly mixed as a biological sample for each plot. A total of 684 samples were collected from the two sites (see detailed sample information in Fig. S1). All leaf, root, and soil samples were transported to the laboratory on dry ice, and stored at -80 °C before DNA extraction. Soil physicochemical characteristics including pH, NH_4^+-N, NO_3^--N, dissolved organic carbon (DOC), and dissolved organic nitrogen (DON) were measured according to standard protocols (Zhao *et al.*, 2019), and these parameters together with yield data are listed in Table S1.

DNA extraction and bacterial 16S rRNA gene amplification

For epiphytic DNA extraction, microbial cells were dislodged and collected from 10-15 g of leaves and 3-5 g of roots according to previous method (Bodenhausen *et al.*, 2013; Ruiz-Perez *et al.*, 2016) with some modifications, and subjected to DNA extraction using the PowerSoil DNA Isolation Kit (MO BIO Laboratories, Carlsbad, CA, USA). Endophytic DNA was extracted from the same leaves and roots used for epiphytic DNA after further surface sterilization following the procedure reported by previous works (Ruiz-Perez *et al.*, 2016; Samad *et al.*, 2017). The rhizosphere and bulk soil DNA were extracted from 0.4 g soil using the PowerSoil DNA Isolation Kit. Chloroplast excluding primers targeting the V5-V6 region of the bacterial 16S rRNA gene（799F/1115R) (Kembel *et al.*, 2014) were used for the bacterial 16S rRNA gene amplification. PCR products were purified using the QIAquick gel extraction kit (Qiagen, USA). All samples were pooled in equimolar concentrations and then were sequenced on the Illumina MiSeq platform with a Paired-End protocol. More detailed descriptions on DNA extraction and bacterial 16S rRNA gene amplification are available in Supplementary Information (Method S1).

Bioinformatic analysis

Primer sequences and low-quality reads ends with a quality score (Q) below 30 were trimmed. Paired-end sequences were merged to a single sequence of length of approximately 300 bp, and the resultant sequences were quality-filtered (maximum expected error 0.5) and singletons were removed in USEARCH v10 software (Edgar, 2010). All correct biological reads (i.e. zero-radius operational taxonomic units, or ZOTUs) were picked using UNOISE3 (Edgar, 2016) with default parameters in USEARCH (Edgar, 2010). Representative sequences were classified using BLAST algorithm with the SILVA reference database (V12.8) in QIIME 1.91 (Caporaso *et al.*, 2010). Non-bacterial ZOTUs (chloroplast, mitochondrial and viridiplantae) and rare bacteria (< 20 reads) were then removed (Laforest-Lapointe *et al.*, 2016). Totally, paired-end sequencing resulted in 18,884,458 high-quality reads, and these reads could be assembled into 20179 ZOTUs (i.e. phylotypes; analogous to amplicon sequence variants) at 100% identity. The ZOTU table was rarefied to 570 reads (lowest number in leaf endosphere) for bacterial alpha diversity estimates among different niches. We also rarefied the ZOTU table based on the lowest reads in each niche but the results were almost same. MetagenomeSeq's cumulative sum scaling (CSS) was used as a normalization method for beta-diversity analyses. Both alpha-diversity and beta-diversity were calculated in QIIME. The raw sequencing data have been submitted to the Sequence Read Archive (SRA) under the accession number PRJNA559707.

Statistical analysis

Bacterial beta-diversity was assessed by computing weighted UniFrac distances matrices and then ordinated using non-metric multi-dimensional (NMDS) ordinations. The significance of different factors on community dissimilarity was tested with PERMANOVA or nested PERMANOVA using the *adonis* function of the *vegan* R package (Oksanen *et al.*, 2007) based on weighted UniFrac distances. We defined "host-environment effects index" to evaluate the effects degree of crop species and site within each niche, and the higher HEEI value represents greater host species effect. The

network analysis was performed using the routine CoNet (Faust *et al.*, 2012) in Cytoscape v3.5 (Shannon *et al.*, 2003) and visualized in Gephi (Bastian *et al.*, 2009). Spearman correlation scores were calculated and only robust (Spearman's $r > 0.6$ or $r < -0.6$) and statistically significant ($p < 0.01$) correlations were kept. The network complexity was defined according to previous studies (Pimm, 1984; Wagg *et al.*, 2019), and nodes with high degree and closeness centrality values were identified as hub nodes in co-occurrence networks (Agler *et al.*, 2016; van der Heijden & Hartmann, 2016). The linear mixed model was used to identify the major drivers of bacterial alpha-diversity (Nakagawa *et al.*, 2013). We used SourceTracker (v1.0) (Knights *et al.*, 2011) based on Bayesian approach to estimate the sources of the crop bacterial communities in each host niche. Differential abundance analysis was performed using EdgeR's generalized linear model (GLM) approach (Robinson *et al.*, 2010), and "Depleted index" (DI) and "Dissimilarity index" (DSI) were defined to estimate bacterial selection processes from bulk soil to other niches. The linear discriminant analysis effect size (LEfSe) (Segata *et al.*, 2011) was applied (Wilcoxon *p*-value: 0.05, LDA > 2) to identify the biomarker of different host niches and crops. Phylogenetic tree was annotated and visualized in iTOL software (Letunic & Bork, 2019).

Random forest analyses using the "randomForest" R package (Liaw & Wiener, 2002) were conducted to identify the most important dominant taxa in predicting the crop yield. The significance of the model was assessed with 5,000 permutations of the response variable using the "A3" R package, and the significance of the importance of each predictor on crop yield was assessed by using the "rfPermute" R package (Fortmannroe, 2015; Delgado-Baquerizo *et al.*, 2016). Bacterial functional profiles were predicted using functional annotation of prokaryotic taxa (FAPROTAX) (Louca et al., 2016). All statistical analyses were carried out in R (http://www.r-project.org). Nonparametric statistical test (Kruskal-Wallis test) was used to evaluate the alpha-diversity difference and the taxonomical difference observed among different niches.

More information on the definition of HEEI, DSI, and DI, and the methods of network analysis and source tracking analysis are detailed in the Supplementary Material (Method S1). Scripts employed in the computational analyses and plotting figures are available at https://github.com/ChaoXiong2020/Xiong2020_NewPhytologist.git.

Results

Crop microbiome assembly was shaped more strongly by host than by environmental factors

Our results based on NMDS ordinations and nested PERMANOVA analysis of the complete dataset suggested that the variation in bacterial community was mainly explained by compartment niche (R^2 = 39.8%, p < 0.001) and host species (R^2 = 7.8%, p < 0.001), then by site (R^2 = 2.7%, p < 0.001) and fertilization practice (R^2 = 0.4%, p = 0.07) (Fig. 1a, Table S2). Bacterial communities within each crop season showed similar patterns, with the variations in microbial communities mainly explained by compartment niche (Fig. 1a, Table S3). Moreover, hierarchical clustering analysis revealed a clear and separate clustering among leaf, root, and soil samples (Fig. S2a). The sampling year had significant influence on maize bacterial community but only accounted for slight variation (3.3%) (Fig. 1a, Table S2). Fertilization practice did not show significant influence on bacterial community structure for the first season (maize in 2016) (R^2 = 0.7%, p = 0.83), but did for the second season (wheat/barley in 2017) (R^2 = 0.6%, p = 0.03) and the third season (maize in 2017) (R^2 = 1.2%, p < 0. 001) (Table S3). Within each compartment niche, bacterial community was distinctly separated according to crop species and sites, with varied intensity of host effects and environment effects (Fig. 1b, Table 1).

We defined "Host-environment effects index" (HEEI) to evaluate the relative contribution of crop species and site on bacterial community in different compartments. The HEEI values gradually increased from soils (bulk soil: 0.13, rhizosphere: 0.34) to epiphytes (rhizoplane: 3.36, phylloplane: 0.83) and then to endophytes (root endosphere:

4.93, leaf endosphere: 10.79), demonstrating an increasing host effect from soils to epiphytes to endophytes, with the strongest host effect in leaf endosphere (host species explained 60.4% of variation) (Fig. 1b, Table 1). Although the two study sites are thousands of miles far away, the endophytic samples of maize from two sites (2 cultivars) formed a close cluster but distinctly separated from wheat and barley samples (Fig. 1b). In contrast, environment effect sequentially decreased from soils to epiphytes to endophytes, with site explained 48.3% and 56.5% of variation in rhizosphere and bulk soil, respectively (Fig. 1b, Table 1). Additionally, fertilization practice also showed significant influence on bacterial communities in rhizosphere soil (R^2 = 6.7%, p < 0. 001), bulk soil (R^2 = 4.6%, p = 0.041), leaf endosphere (R^2 = 4.4%, p = 0.011), and rhizoplane (R^2 = 2.7%, p = 0.049), while the influence was marginal in comparison to crop species and site (Table 1).

Host selection effect reduced bacterial diversity and network complexity

To further characterize the host selection effect on crop microbiomes, particularly for endophytes, we assessed the alpha-diversity and co-occurrence patterns of bacterial communities along the soil-plant continuum. Our results suggested that crop host had a strong effect on bacterial diversity (Shannon diversity and Chao1 richness) and network complexity (a higher average degree representing a greater network complexity) (Fig. 2a-c). Bacterial richness and network complexity gradually decreased from soils (with an average degree of 32.46 in bulk soil and 26.55 in rhizosphere) to epiphytes (19.02 in rhizoplane and 16.40 in phylloplane) and then to endophytes (8.15 in root endosphere and 5.33 in leaf endosphere), with the lowest richness and network complexity in leaf endosphere (Fig. 2a-c). Moreover, the lowest network complexity was found in leaf endosphere in functional network (Fig. S3a). The number of "Hub nodes" (nodes with high values of degree (> 60) and closeness centrality (> 0.3) in the network) gradually decreased from soils to epiphytes to endophytes (Fig. 2d, Table 2). The taxonomic composition of the networks differed between soil and plant compartments, with more nodes belonged to Acidobacteria in soil niches and Firmicutes in plant niches (Fig. 2b).

335 Remarkably, the genus *Carnobacterium* within Firmicutes was identified as an
336 important network hub in rhizoplane (Fig. 2b, d). In addition, higher modularity and
337 average path distance were found in phylloplane and rhizoplane (Fig. 2a, Table 2).
338 The linear mixed model analysis based on all samples showed that bacterial alpha
339 diversity was mainly influenced by compartment niche (Shannon diversity, $F5, 581 = 360.1$, $p < 0.0001$) and crop species ($F2, 589 = 14.6$, $p < 0.0001$) (Table S4), further
341 supporting the above observation. Moreover, the fertilization practice had significant
342 influence on bacterial Shannon diversity in bulk soil ($F6, 29 = 5.8$, $p = 0.0005$),
343 rhizosphere soil ($F6, 29 = 3.4$, $p = 0.01$), and phylloplane ($F6, 29 = 4.0$, $p = 0.005$)
344 (Table S5). Additionally, host showed strong selection effect on the distribution
345 patterns of the core taxa (ZOTUs present in at least 80% of samples in each niche) and
346 the dominant phyla/class (relative abundance ≥ 1%) among different compartment
347 niches (Fig. S2b). The numbers of core taxa gradually decreased from soils (bulk soil:
348 571, rhizosphere: 359) to epiphytes (root: 54, leaf: 67) to endophytes (root: 23, leaf: 2)
349 (Table S6), with the lowest values in leaf endophytes.
350 **The potential sources and selection processes of crop microbiome**
351 Source-tracking analysis was conducted to identify the potential sources of observed
352 bacterial communities in each host niche. The Source Model of Plant Microbiome
353 (SMPM) suggested that crop-associated bacterial communities were mainly derived
354 from bulk soils and gradually filtered at different plant compartment niches (Fig. 3a).
355 Specifically, rhizoplane, root endosphere, and leaf endosphere potentially selected
356 majority of taxa from nearby species pool, with the known source values more than
357 78%, and root endophytes were the main potential sources of leaf endophytes (Fig. 3a).
358 Similar source patterns were also observed in each crop species (Fig. S3b). In contrast,
359 the neighbor compartments accounted for less proportion of the derivation in
360 phylloplane microbiome in comparison to rhizoplane and endosphere niches, with 51-
361 66% of known source from rhizosphere and rhizoplane for wheat and barley, but only
362 ~35% for maize (Fig. 3a, Fig. S3b), indicating other environmental sources might

363 contribute to phylloplane microbiome.

364 The differential abundance analysis indicated that 4.6% of ZOTUs (310 out of 6744
365 ZOTUs), mainly from the families *Enterobacteriaceae* (40 ZOTUs),
366 *Pseudomonadaceae* (36 ZOTUs) and *Methylobacteriaceae* (30 ZOTUs), were
367 significantly enriched and overlapped in 4 plant niches (epiphyte and endophyte of leaf
368 or root) (Fig. 3b, c). In contrast, 21.2% of ZOTUs (1427 out of 6744 ZOTUs) were
369 significantly depleted in 4 plant niches, and these ZOTUs mainly belonged to
370 *Chitinophagaceae* (101 ZOTUs) (Fig. 3b, c). Additionally, rhizoplane and phylloplane
371 possessed the greatest numbers (249， 470) of specific enriched ZOTUs, while leaf
372 endosphere possessed the greatest numbers (352) of depleted ZOTUs (Fig. 3b, c). We
373 defined "Depleted index (DI)" and "Dissimilarity index (DSI)" to further evaluate the
374 species filtration and selection processes from bulk soil to other niches. DI value was
375 found to be gradually increased from rhizosphere (0.08) to epiphytes (0.90-1.32) and to
376 endophytes (1.60-2.12), with the highest DI and DSI values in leaf endosphere (Fig.
377 3b).

378 Correspondingly, large variation in bacterial community composition was observed
379 between plants (leaves and roots) and soils, with Gammaproteobacteria (25.7%) and
380 Actinobacteria (23.9%) as the most abundant phylum in plants and soils, respectively
381 (Fig. S2b, Fig. S4). Specifically, leaf endosphere and root endosphere possessed more
382 abundant Alphaproteobacteria (~34.6%) and Gammaproteobacteria (~43.7%)
383 respectively, than any other niches (Fig. S4). The heat map of the dominant functional
384 groups (top 40) showed a strong clustering of functional groups dependent on three
385 compartment niches (Fig. S5a). The linear discriminant analysis effect size (LEfSe)
386 identified *Methylobacterium* in leaves, *Enterobacteriaceae* in roots, and
387 *Chitinophagaceae* in soils as the most significant biomarker taxa (Fig. S5b).

388 **Dominant taxa, biomarker taxa, and potential function of bacterial microbiome**
389 **among three crops**

390 To better characterize the host effect at crop species level, we identified the dominant

taxa (ZOTUs present in at least 60% of plant samples and with a relative abundance ≥ 0.1%) and biomarker taxa for each crop (leaves and roots). Though more than 10 thousand of ZOTUs were retrieved from each crop, only 43 (0.2%), 91 (0.7%), and 65 (0.6%) ZOTUs were identified as the dominant taxa for maize, wheat, and barley, respectively, and these ZOTUs accounted for ~50% (31.1, 62.5, and 49.1%) of the total sequences from each crop (Fig. 4a-c). For all three crops, these dominant ZOTUs were mainly affiliated within Gammaproteobacteria, with a relative abundance of 39%-50% within each crop (Fig. 4a-c). In total, there were 9 dominant taxa shared by three crops and 8 of them belonged to Gammaproteobacteria, with the top 4 abundant dominant taxa affiliated within the family *Enterobacteriaceae* (Fig. 4a-c, Table S7).

The class Bacilli accounted for a high proportion in the dominant taxa of wheat (19.0%) and barley (9.0%), but only accounted for 0.8% in maize (Fig. 4a-c). Remarkably, the LEfSe further identified *Bacillus* as the biomarker taxa for wheat, *Lactobacillaceae* (belonging to class Bacilli) for barley and *Methylobacteriaceae* for maize (Fig. S6a). Coincidently, the Random Forest Model identified Gammaproteobacteria and Bacilli as the main predictors of maize and wheat/barley yield, respectively (Fig. S6b).

Bacterial function prediction analyses for these dominant taxa showed that the top 5 abundant functional groups (accounting for more than 70% of total groups) included chemoheterotrophy, fermentation, ureolysis, symbionts, and nitrate reduction for maize, wheat and barley (Fig. S6c). Additionally, a greater relative abundance of functional group methylotrophy was observed in maize samples (3.8%) than in wheat (0.8%) and barley (0) samples (Fig. S6c), further supporting the LEfSe analysis result of *Methylobacteriaceae* as the biomarker taxa for maize (Fig. S6a).

Discussion

In this study, a scenario survey on bacterial communities along the soil-plant continuum in field-grown crops demonstrates that crop microbiomes are shaped largely by host selection at the plant level, with marginal influence from site-dependent environmental

factors and fertilization practice. Our results further showed that host had a strong effect on bacterial diversity and network complexity, with host selection sequentially increased from soils to epiphytes to endophytes. These findings provide comprehensive and empirical evidences for the theoretical host selection and niche occupation theory for crop microbiome assembly under different environmental conditions. Moreover, we identified the potential sources, selection processes, and keystone taxa of crop microbiome through multiple machine learning methods, which provided critical information for future crop microbiome manipulation.

The assemblies and maintenance of crop-associated bacterial microbiomes

Our results based on the complete dataset show that microbiome assembly at the plant level is primarily determined by compartment niche and crop species rather than by environmental factors. Although sampling year showed significant effect on maize bacterial community, the effect was slight in comparison to plant compartment. These results suggest that crop can select niche compliant microbes to occupy various compartments even when crops are planted in two contrasting soil types across long geographical distance and under different fertilization practices. Similarly, some previous studies on root microbiome of rice, microbiome in roots and aboveground compartments of poplar trees and sugarcane suggested that different root or leaf niches harbored distinct microbial communities (Edwards *et al.*, 2015; Beckers *et al.*, 2017; Hamonts *et al.*, 2017; Cregger *et al.*, 2018). We further found that within each plant compartment, the variation in bacterial community was explained more strongly by crop species than by site or fertilization practice. Consistently, some previous studies working on tree phyllosphere and maize rhizosphere separately showed that host genetics (e.g. species and subspecies level) played an important role in shaping bacterial microbiome in terms of a specific compartment (Laforest-Lapointe *et al.*, 2016; Walters *et al.*, 2018). All these suggest that host selection (i.e. compartment niche and species) has determining effect over environmental factors in shaping plant microbiome assembly. Together these results provide comprehensive empirical evidence for host

selection and theoretical framework of coevolution between the host and microbes, in which the host plant employs exudations to recruits, filters, and enriches microbial taxa with specific functions in different niches (Muller *et al.*, 2016; Martin *et al.*, 2017; Sasse *et al.*, 2018), and microbes competing for these resources drive their rapid evolutionary radiations thus diverge and adapt to different niches to reduce competition (Foster *et al.*, 2017).

We further found that host selection sequentially increased while bacterial richness and network complexity decreased from soils to epiphytes to endophytes, with highest HEEI, DI and DSI values and lowest bacterial diversity recorded in leaf endosphere. All these suggest that host species has the strongest selection effect on crop endophytes particular for leaf endophytes. It could be explained as there are intensive selection pressures caused by host immune system and plant exudates in endosphere and thus only relatively few bacteria are able to colonize therein (Guttman *et al.*, 2014; Kembel *et al.*, 2014; Hacquard *et al.*, 2015). By contrast, studies on *Arabidopsis thaliana* suggested that root endosphere bacterial communities were strongly influenced by soil type and soil properties, while host genotype had limited effect on the root microbiome (Bulgarelli *et al.*, 2012; Lundberg *et al.*, 2012; Thiergart *et al.*, 2020). This inconsistency could be due to the less difference in host effect at genotype level within *Arabidopsis thaliana* than that at species level in our study. Further, our data showed that endosphere samples from two maize cultivars in two sites formed a close cluster, but distinctly separated from wheat and barley samples. All these suggest that the relative contribution of host effect and environmental effect on plant endosphere microbiome could vary depending on host genetic difference, and host effect would be much stronger at species level than at genotype level.

Moreover, we showed that both host species and site had significant effects on bacterial community in rhizoplane and phylloplane (site explained 18-44% of variation), suggesting that the microbes inhabiting rhizoplane and phylloplane were shaped by both host and environment. It's reasonable as the surfaces of leaf or root were supposed

to be the interface between host and environment (Lindow & Brandl, 2003; Barnes *et al.*, 2016; Vacher *et al.*, 2016; Remus-Emsermann & Schlechter, 2018). Corresponding to the strong host-environment interaction effects in rhizoplane and phylloplane, microbial network modules were highly differentiated and much higher numbers of network node and specific enriched ZOTUs were observed in the two niches (Table 2 and Fig. 3c), further indicating that rhizoplane and phylloplane act as a hotspot of plant-microbes-environment interactions. Our work therefore further advanced current understanding of the crop microbiome assembly by providing systematic evaluation on the relative contribution of host species and multiple environmental factors on bacterial communities across different compartment niches.

We showed that bacterial microbiomes in soils were distinct between two sites (explained ~50% variation), which could be explained by large geographic distance and distinct environmental characteristics (Bahram *et al.*, 2018; Shi *et al.*, 2018). In our study, the site effect represented the effect of site-dependent environmental factors (including climate and edaphic factors) and the effect of cultivar used in each site, which may co-influence microbiome composition (Bulgarelli *et al.*, 2013; Walters *et al.*, 2018). Furthermore, fertilization practice had greater effects on bacterial Shannon diversity and community structure in soils than in other plant niches (Table 1 and Table S5). These results suggest that microbiomes in soils are more sensitive to different fertilization practices than those in plant niches in which microbiomes remained relatively stable potentially due to the strong selection pressure from the crop host. These findings provide critical message for developing effective strategy for future manipulation of the crop microbiomes.

The potential sources of crop-associated bacterial communities

Understanding the potential sources and enrichment process of crop-associated microbiomes could provide critical information on the interactions among plants, soil, and microbes (Bulgarelli *et al.*, 2012; Edwards *et al.*, 2015; Zhang *et al.*, 2017). Although previous studies have reported that plant aboveground compartments shared

a large proportion of microbial taxa with belowground compartments (Bai *et al.*, 2015; Zarraonaindia *et al.*, 2015; Wagner *et al.*, 2016), we know little about the enrichment process of crop microbiome. Our source-tracking analysis reveals that crop-associated bacterial communities are derived primarily from bulk soils and are gradually filtered and enriched by different compartment niches. Further, we found that some members within *Enterobacteriaceae*, *Pseudomonadaceae*, and *Methylobacteriaceae* were significantly enriched in 4 plant niches, whereas some members of family *Chitinophagaceae* were significantly depleted in these niches. This finding is supported by previous works, which reported that these taxa were significant enriched in plant compartments (Hassani *et al.*, 2018; Cernava *et al.*, 2019). Furthermore, some members of *Enterobacteriaceae* are believed to be tolerant to a broad range of plant-specific metabolites and be able to adapt to plant-associated habitats (Solana *et al.*, 2014; Cernava *et al.*, 2019). These results suggest that soil habitats are the major sources for crop microbial selection, and certain taxa are gradually enriched while others are filtered out by different host niches. Soil microbes were supposed to colonize plant compartments through early inoculation, roots, and other external force like wind, rain splashes, and crawling insects (Vorholt *et al.*, 2017; Mangeot-Peter *et al.*, 2020; Tkacz *et al.*, 2020). Intriguingly, we found that the main source of leaf endophytes is root endophytes rather than phylloplane, and root endophytes are mainly derived from rhizoplane, indicating that the rhizosphere microbiomes can be transported to plant above-ground compartments via internal plant tissue transmission, among which rhizoplane act as an important transitional boundary (Hacquard *et al.*, 2015; Vandenkoornhuyse *et al.*, 2015). Moreover, lower unknown source was detected in endophytes than in epiphytes, further supported the strong host selection in endophytes. In contrast, higher unknown source in leaf epiphytes suggests that other environmental sources like air and rainwater potentially contribute to the formation of phylloplane microbiome (Remus-Emsermann & Schlechter, 2018). All these further supported that epiphytes are shaped by both host and environmental effects (Agler *et al.*, 2016).

Compared to wheat and barley, higher unknown source was observed in maize leaf epiphytes, which may be due to the difference in plant functional traits like leaf area and height, and season-dependent factors like precipitation between maize and wheat/barley (Vandenkoornhuyse *et al.*, 2015; Muller *et al.*, 2016; Vacher *et al.*, 2016). Overall, the relative contribution of other important sources (such as seed, air and rainwater) to crop microbiome formation remains unclear and requires further research (Shade *et al.*, 2017).

The keystone taxa for crop microbiome

The dominant taxa, biomarker taxa and network hubs were considered as potential keystone taxa, which have an important ecological role in microbiome assembly and ecosystem functions (Banerjee *et al.*, 2018; Delgado-Baquerizo *et al.*, 2018). We found that only ~0.5% of bacterial phylotypes consistently accounted for ~50% of bacterial community in each crop, suggesting that only few bacterial taxa dominate in plant compartment niches though there are highly diverse bacteria inhabiting. This is consistent with the finding that only 2% of bacterial phylotypes accounted for almost half of the soil bacterial communities at the global scale (Delgado-Baquerizo *et al.*, 2018). Our results demonstrated that the most dominant taxa belonged to Gammaproteobacteria, and that *Enterobacteriaceae* and *Pseudomonadaceae* within Gammaproteobacteria were significantly enriched in plant niches. The members of Gammaproteobacteria are capable of colonizing a wide range of niches like phyllosphere and rhizosphere, and play a key role in modulating host fitness, pathogen suppression and plant tolerance to stress (Mendes *et al.*, 2011; Alvarez-Perez *et al.*, 2017), which well explained their dominant distribution in various plant compartments.

A greater relative abundance of Bacilli was detected in wheat and barley samples than in maize samples in this study. Consistently, it was reported that more Bacilli-affiliated strains were isolated in wheat rhizoplane than in canola rhizoplane (Germida *et al.*, 1998). We further found that the relative abundance of Bacilli was a good predictor of wheat and barley yield, which is not surprising as members of Bacilli like

559 *Bacillus cereus* were frequently reported as the antagonist bacteria of soil borne
560 disease (Fira *et al.*, 2018). A recent study also reported that wheat may respond to the
561 elevated N input by recruiting *Bacillus* through secretion of organic acids (Chen *et al.*,
562 2019). Together these results suggest that wheat and barley may preferentially enrich
563 members of class Bacilli, which may be greatly beneficial to host performance.
564 Moreover, *Methylobacteriaceae* was identified as biomarker taxa for maize by LEfSe,
565 and function prediction showed a higher relative abundance of functional group
566 methylotrophy in maize than in wheat and barley. This might be attributed to the
567 difference in host selection pressure from host immune system and secondary
568 metabolites among different plant species (Guttman *et al.*, 2014; Laforest-Lapointe *et*
569 *al.*, 2016; Kudjordjie *et al.*, 2019), while the molecular mechanisms for these selection
570 effects remain to be explored.
571 Irrespective of crop species, *Methylobacterium* was identified as the most significant
572 biomarker taxa in leaves by LEfSe (Fig. S5b). *Methylobacterium* is facultative
573 methylotrophs capable of growing on one-carbon compounds like methanol (Green,
574 2006). Previous studies have shown its ubiquitous occurrence in the phyllosphere, and
575 some members of *Methylobacterium* were able to fix nitrogen in leaves of *Jatropha*
576 *curcas* and promote host biomass (Knief *et al.*, 2010; Madhaiyan *et al.*, 2015). For roots,
577 the genus *Morganella*, *Enterobacter*, and *Serratia* within family *Enterobacteriaceae*
578 (Gammaproteobacteria) were identified as the biomarker taxa (Fig. S5b), further
579 confirming that some members of family *Enterobacteriaceae* are able to utilize or
580 tolerate plant metabolites like root exudates (Solana *et al.*, 2014; Cernava *et al.*, 2019).
581 In addition, the distribution patterns of the "Hub nodes" varied among compartment
582 niches, and Firmicutes was identified as a key network hub in rhizoplane. We further
583 found that plant niches possessed more network nodes affiliating within Firmicutes in
584 comparison to soils, indicating that Firmicutes are able to adapt to plant niches as
585 well (Bulgarelli *et al.*, 2013; Bai *et al.*, 2015). Together these results demonstrated that
586 plant compartments likely create distinct niches for specific microbiota with capability

of recognizing the signal molecules and adapting to immune systems in each niche (Foster *et al.*, 2017; Cordovez *et al.*, 2019). The identification of these dominant taxa, biomarker taxa and network hubs could provide essential information for developing strategies to manipulate crop-associated microbiomes and facilitate the application of bioinoculants for crop health (Agler *et al.*, 2016; Toju *et al.*, 2018; Qiu *et al.*, 2019).

Conclusions

In this study, we provide comprehensive and empirical evidences on the relative contribution of host and environmental factors to microbiome assembly in maize, wheat and barley. Our results demonstrate that crop microbiome assembly is shaped predominantly by host niche and host species, rather than by site or fertilization practice. Further, we revealed that host selection sequentially increased and had a strong effect on reducing bacterial diversity and network complexity from soils to epiphytes to endophytes. These findings significantly advance our current understanding of the bacterial community assembly in maize-wheat/barley rotation system under different environment selection pressures, and highlight the importance of host selection effect.

Moreover, we provide empirical evidence of gradual filtration from bulk soils to various plant compartments and selective enrichment of specific microbial taxa like *Enterobacteriaceae*, *Pseudomonadaceae* and *Methylobacteriaceae* in plant. Crop endophytes could select a majority of taxa from nearby species pool, with root endophytes as the main potential sources of leaf endophytes. We further revealed that crop microbiomes were dominated by few dominant taxa (mainly belonging to Gammaproteobacteria) which accounted for only ~0.5% of bacterial phylotypes, and identified Bacilli as the important biomarker taxa for wheat and barley and *Methylobacteriaceae* for maize. Our work advances the discipline for eco-evolutionary theory of host microbiome assembly. Results from this study have implications for future crop management by providing baseline data to inform translational research to harness plant microbiome to sustainably increase agriculture productivity.

Acknowledgements

We would like to thank Prof. Yan-Lai Han, Mr. Qin-Bing Zhang and Hao-Tian Tong for their cooperation and assistance in the field work, and Lu-Qi Zhang, Ya-Qi Wang, Zi-Yang He, Zhi-Bo Zhao, and Chang-Chun Zhai for their generous helps in soil sampling and laboratory work. We also thank Dr. Hang-Wei Hu for his assistance in data analysis. This work was financially supported by the Strategic Priority Research Program (B) of Chinese Academy of Sciences (XDB15020200) and National Key R&D Program (2017YFD0200600). Research on plant and soil microbiomes in BKS lab is supported by Australian Research Council (DP170103628; DP190103714). Li-Mei Zhang was supported by the Youth Innovation Promotion Association, Chinese Academy of Sciences.

Author contributions

LMZ, JZH, YGZ, and CX designed the study. LMZ, LLH, JPS, GBW, and PPL managed the field trial stations. CX, CFW, and AHG collected samples. CX conducted the laboratory analyses. CX, JTW, BS and LMZ performed the data processes. CX, LMZ, BS, YGZ, JTW and JZH wrote the manuscript. All authors read and approved the final manuscript.

References

Agler MT, Ruhe J, Kroll S, Morhenn C, Kim ST, Weigel D, Kemen EM. **2016**. Microbial hub taxa link host and abiotic factors to plant microbiome variation. *PLoS Biology* **14**: e1002352.

Alvarez-Perez JM, Gonzalez-Garcia S, Cobos R, Olego MA, Ibanez A, Diez-Galan A, Garzon-Jimeno E, Coque JJR. **2017**. Use of endophytic and rhizosphere Actinobacteria from grapevine plants to reduce nursery fungal graft infections that lead to young grapevine decline. *Appl. Environ. Microbiol.* **83**: e01564-01517.

Bahram M, Hildebrand F, Forslund SK, Anderson JL, Soudzilovskaia NA, Bodegom PM, Bengtsson-Palme J, Anslan S, Coelho LP, Harend H *et al.* 2018. Structure and function of the global topsoil microbiome. *Nature* **560**: 233-237.

Bai Y, Mueller DB, Srinivas G, Garrido-Oter R, Potthoff E, Rott M, Dombrowski N, Muench PC, Spaepen S, Remus-Emsermann M *et al.* 2015. Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* **528**: 364-369.

Banerjee S, Schlaeppi K, van der Heijden MGA. **2018**. Keystone taxa as drivers of microbiome structure and functioning. *Nature Reviews Microbiology* **16**: 567-576.

Barnes CJ, van der Gast CJ, Burns CA, McNamara NP, Bending GD. **2016**. Temporally variable geographical distance effects contribute to the assembly of root-associated fungal communities. *Frontiers in Microbiology* **7**: 195-195.

Bastian M, Heymann S, Jacomy M. **2009**. Gephi: an open source software for exploring and manipulating networks. *Icwsm* **8**: 361-362.

Beckers B, Op De Beeck M, Weyens N, Boerjan W, Vangronsveld J. **2017**. Structural variability and niche differentiation in the rhizosphere and endosphere bacterial microbiome of field-grown poplar trees. *Microbiome* **5**: 25.

Bodenhausen N, Horton MW, Bergelson J. **2013**. Bacterial communities associated with the leaves and the roots of *Arabidopsis thaliana*. *PLoS One* **8**: e56329.

Bulgarelli D, Rott M, Schlaeppi K, Ver Loren van Themaat E, Ahmadinejad N, Assenza F, Rauf P, Huettel B, Reinhardt R, Schmelzer E *et al.* 2012. Revealing

structure and assembly cues for *Arabidopsis* root-inhabiting bacterial microbiota. *Nature* **488**: 91-95.

Bulgarelli D, Schlaeppi K, Spaepen S, Ver Loren van Themaat E, Schulze-Lefert P. **2013**. Structure and functions of the bacterial microbiota of plants. *Annual Review of Plant Biology* **64**: 807-838.

Caporaso JG, Kuczynski J, Stombaugh J, Bittinger K, Bushman FD, Costello EK, Fierer N, Peña AG, Goodrich JK, Gordon JI *et al.* 2010. QIIME allows analysis of high-throughput community sequencing data. *Nat. Methods* **7**: 335-336.

Cernava T, Erlacher A, Soh J, Sensen CW, Grube M, Berg G. **2019**. *Enterobacteriaceae* dominate the core microbiome and contribute to the resistome of arugula (*Eruca sativa* Mill.). *Microbiome* **7**: 13.

Chen S, Waghmode TR, Sun R, Kuramae EE, Hu C, Liu B. **2019**. Root-associated microbiomes of wheat under the combined effect of plant development and nitrogen fertilization. *Microbiome* **7**: 136.

Christian N, Herre EA, Clay K. **2019**. Foliar endophytic fungi alter patterns of nitrogen uptake and distribution in *Theobroma cacao*. *New Phytologist* **222**: 1573-1583.

Cordovez V, Dini-Andreote F, Carrión VJ, Raaijmakers JM. **2019**. Ecology and evolution of plant microbiomes. *Annu. Rev. Microbiol.* **73**: 69-88.

Cregger MA, Veach AM, Yang ZK, Crouch MJ, Vilgalys R, Tuskan GA, Schadt CW. **2018**. The *Populus* holobiont: dissecting the effects of plant niches and genotype on the microbiome. *Microbiome* **6**: 31.

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

Delgado-Baquerizo M, Oliverio AM, Brewer TE, Benavent-Gonzalez A, Eldridge DJ, Bardgett RD, Maestre FT, Singh BK, Fierer N. **2018**. A global atlas of the dominant bacteria found in soil. *Science* **359**: 320-325.

Durán P, Thiergart T, Garrido-Oter R, Agler M, Kemen E, Schulze-Lefert P, Hacquard S. **2018**. Microbial interkingdom interactions in roots promote *Arabidopsis* survival. *Cell* **175**: 973-983.e914.

Edgar RC. **2010**. Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.

Edgar RC. **2016**. UNOISE2: improved error-correction for Illumina 16S and ITS amplicon sequencing. *BioRxiv*: 081257. doi: 10.1101/081257

Edwards J, Johnson C, Santos-Medellín C, Lurie E, Podishetty NK, Bhatnagar S, Eisen JA, Sundaresan V. **2015**. Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc. Natl. Acad. Sci. U S A* **112**: E911-E920.

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

Edwards JD, Pittelkow CM, Kent AD, Yang WH. **2018b**. Dynamic biochar effects on soil nitrous oxide emissions and underlying microbial processes during the maize growing season. *Soil Biology and Biochemistry* **122**: 81-90.

Emmett BD, Buckley DH, Drinkwater LE. **2020**. Plant growth rate and nitrogen uptake shape rhizosphere bacterial community composition and activity in an agricultural field. *New Phytologist* **225**: 960-973.

Faust K, Sathirapongsasuti JF, Izard J, Segata N, Gevers D, Raes J, Huttenhower C. **2012**. Microbial co-occurrence relationships in the human microbiome. *PLoS Computational Biology* **8**: e1002606.

Fira D, Dimkic I, Beric T, Lozo J, Stankovic S. **2018**. Biological control of plant pathogens by *Bacillus* species. *Journal of Biotechnology* **285**: 44-55.

Fortmannroe S. **2015**. A3: accurate, adaptable, and accessible error metrics for predictive models. *Journal of Statistical Software* **66**: 1-23.

Foster KR, Schluter J, Coyte KZ, Rakoff-Nahoum S. **2017**. The evolution of the

717 host microbiome as an ecosystem on a leash. *Nature* **548**: 43-51.
718 **Geddes BA, Ryu MH, Mus F, Garcia Costas A, Peters JW, Voigt CA, Poole P**.
719 **2015**. Use of plant colonizing bacteria as chassis for transfer of N(2)-fixation to
720 cereals. *Curr. Opin. Biotechnol.* **32**: 216-222.
721 **Germida JJ, Siciliano SD, Renato de Freitas J, Seib AM**. **1998**. Diversity of root-
722 associated bacteria associated with field-grown canola (*Brassica napus* L.) and wheat
723 (*Triticum aestivum* L.). *FEMS Microbiol. Rev.* **26**: 43-50.
724 **Green PN**. **2006**. The Prokaryotes. In: Dworkin M, Falkow S, Rosenberg E, Schleifer
725 KH, Stackebrandt E, eds. *Methylobacterium*. New York, USA: Springer, 257-265.
726 **Guttman DS, McHardy AC, Schulze-Lefert P**. **2014**. Microbial genome-enabled
727 insights into plant-microorganism interactions. *Nature Reviews Genetics* **15**: 797-813.
728 **Hacquard S, Garrido-Oter R, Gonzalez A, Spaepen S, Ackermann G, Lebeis S,**
729 **McHardy AC, Dangl JL, Knight R, Ley R *et al.* 2015**. Microbiota and host
730 nutrition across plant and animal kingdoms. *Cell Host Microbe* **17**: 603-616.
731 **Hamonts K, Trivedi P, Garg A, Janitz C, Grinyer J, Holford P, Botha FC,**
732 **Anderson IC, Singh BK**. **2017**. Field study reveals core plant microbiota and relative
733 importance of their drivers. *Environ. Microbiol.* **20**: 124-140.
734 **Hartman K, van der Heijden MGA, Wittwer RA, Banerjee S, Walser JC,**
735 **Schlaeppi K**. **2018**. Cropping practices manipulate abundance patterns of root and
736 soil microbiome members paving the way to smart farming. *Microbiome* **6**: 14.
737 **Hassani MA, Duran P, Hacquard S**. **2018**. Microbial interactions within the plant
738 holobiont. *Microbiome* **6**: 58.
739 **Kembel SW, O'Connor TK, Arnold HK, Hubbell SP, Wright SJ, Green JL**.
740 **2014**. Relationships between phyllosphere bacterial communities and plant functional
741 traits in a neotropical forest. *Proc. Natl. Acad. Sci. U S A* **111**: 13715-13720.
742 **Knief C, Ramette A, Frances L, Alonso-Blanco C, Vorholt JA**. **2010**. Site and
743 plant species are important determinants of the *Methylobacterium* community
744 composition in the plant phyllosphere. *ISME J* **4**: 719-728.

745 **Knights D, Kuczynski J, Charlson ES, Zaneveld J, Mozer MC, Collman RG,**
746 **Bushman FD, Knight R, Kelley ST**. **2011**. Bayesian community-wide culture-
747 independent microbial source tracking. *Nat. Methods* **8**: 761-763.
748 **Kudjordjie EN, Sapkota R, Steffensen SK, Fomsgaard IS, Nicolaisen M**. **2019**.
749 Maize synthesized benzoxazinoids affect the host associated microbiome. *Microbiome*
750 **7**: 59.
751 **Laforest-Lapointe I, Messier C, Kembel SW**. **2016**. Host species identity, site and
752 time drive temperate tree phyllosphere bacterial community structure. *Microbiome* **4**:
753 1-10.
754 **Letunic I, Bork P**. **2019**. Interactive Tree Of Life (iTOL) v4: recent updates and new
755 developments. *Nucleic Acids Research* **47**: W256-W259.
756 **Liaw A, Wiener M**. **2002**. Classification and regression by randomForest. *R news* **2**:
757 18-22.
758 **Lindow SE, Brandl MT**. **2003**. Microbiology of the phyllosphere. *Appl. Environ.*
759 *Microbiol.* **69**: 1875-1883.
760 **Liu H, Macdonald CA, Cook J, Anderson IC, Singh BK**. **2019**. An ecological
761 loop: host microbiomes across multitrophic interactions. *Trends in Ecology &*
762 *Evolution* **34**: 1118-1130.
763 **Louca S, Parfrey LW, Doebeli M**. **2016**. Decoupling function and taxonomy in the
764 global ocean microbiome. *Science*: 1272-1277.
765 **Lundberg DS, Lebeis SL, Paredes SH, Yourstone S, Gehring J, Malfatti S,**
766 **Tremblay J, Engelbrektson A, Kunin V, Del Rio TG *et al.* 2012**. Defining the core
767 *Arabidopsis thaliana* root microbiome. *Nature* **488**: 86-90.
768 **Madhaiyan M, Alex TH, Ngoh ST, Prithiviraj B, Ji L**. **2015**. Leaf-residing
769 *Methylobacterium* species fix nitrogen and promote biomass and seed production in
770 *Jatropha curcas*. *Biotechnol Biofuels* **8**: 222.
771 **Mangeot-Peter L, Tschaplinski TJ, Engle NL, Veneault-Fourrey C, Martin F,**
772 **Deveau A**. **2020**. Impacts of soil microbiome variations on root colonization by fungi

and bacteria and on the metabolome of *Populus tremula* × *alba*. *Phytobiomes Journal* **4**: 142-155.

Martin FM, Uroz S, Barker DG. **2017**. Ancestral alliances: plant mutualistic symbioses with fungi and bacteria. *Science* **356**: eaad4501.

Mendes R, Kruijt M, de Bruijn I, Dekkers E, van der Voort M, Schneider JH, Piceno YM, DeSantis TZ, Andersen GL, Bakker PA *et al.* 2011. Deciphering the rhizosphere microbiome for disease-suppressive bacteria. *Science* **332**: 1097-1100.

Muller DB, Vogel C, Bai Y, Vorholt JA. **2016**. The plant microbiota: systems-level insights and perspectives. *Annual Review of Genetics* **50**: 211-234.

Nakagawa S, Schielzeth H, O'Hara RB. **2013**. A general and simple method for obtaining r^2 from generalized linear mixed-effects models. *Methods in Ecology and Evolution* **4**: 133-142.

Oksanen J, Kindt R, Legendre P, O'Hara B, Stevens MHH, Oksanen MJ, Suggests M. **2007**. The vegan package. *Community Ecology Package* **10**: 631-637.

Peiffer JA, Spor A, Koren O, Jin Z, Tringe SG, Dangl JL, Buckler ES, Ley RE. **2013**. Diversity and heritability of the maize rhizosphere microbiome under field conditions. *Proc. Natl. Acad. Sci. U S A* **110**: 6548-6553.

Philippot L, Raaijmakers JM, Lemanceau P, van der Putten WH. **2013**. Going back to the roots: the microbial ecology of the rhizosphere. *Nat. Rev. Microbiol.* **11**: 789-799.

Pimm SL. **1984**. The complexity and stability of ecosystems. *Nature* **307**: 321-326.

Qiu Z, Egidi E, Liu H, Kaur S, Singh BK. **2019**. New frontiers in agriculture productivity: optimised microbial inoculants and in situ microbiome engineering. *Biotechnol Adv.* **37**: 107371.

Remus-Emsermann MNP, Schlechter RO. **2018**. Phyllosphere microbiology: at the interface between microbial individuals and the plant host. *New Phytologist* **218**: 1327-1333.

Ritpitakphong U, Falquet L, Vimoltust A, Berger A, Metraux JP, L'Haridon F.

2016. The microbiome of the leaf surface of *Arabidopsis* protects against a fungal pathogen. *New Phytologist* **210**: 1033-1043.

Robinson MD, McCarthy DJ, Smyth GK. **2010**. edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* **26**: 139-140.

Rolli E, Marasco R, Vigani G, Ettoumi B, Mapelli F, Deangelis ML, Gandolfi C, Casati E, Previtali F, Gerbino R *et al.* 2015. Improved plant resistance to drought is promoted by the root-associated microbiome as a water stress-dependent trait. *Environ. Microbiol.* **17**: 316-331.

Ruiz-Perez CA, Restrepo S, Zambrano MM. **2016**. Microbial and functional diversity within the phyllosphere of espeletia species in an andean high-mountain ecosystem. *Appl. Environ. Microbiol.* **82**: 1807-1817.

Samad A, Trognitz F, Compant S, Antonielli L, Sessitsch A. **2017**. Shared and host-specific microbiome diversity and functioning of grapevine and accompanying weed plants. *Environ. Microbiol.* **19**: 1407-1424.

Sasse J, Martinoia E, Northen T. **2018**. Feed your friends: do plant exudates shape the root microbiome? *Trends Plant Sci.* **23**: 25-41.

Segata N, Izard J, Waldron L, Gevers D, Miropolsky L, Garrett WS, Huttenhower C. **2011**. Metagenomic biomarker discovery and explanation. *Genome Biology* **12**: R60.

Shade A, Jacques MA, Barret M. **2017**. Ecological patterns of seed microbiome diversity, transmission, and assembly. *Curr. Opin. Microbiol.* **37**: 15-22.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. **2003**. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Research* **13**: 2498-2504.

Shi Y, Li Y, Xiang X, Sun R, Yang T, He D, Zhang K, Ni Y, Zhu YG, Adams JM *et al.* 2018. Spatial scale affects the relative role of stochasticity versus determinism in soil bacterial communities in wheat fields across the North China Plain. *Microbiome*

6: 27.

Singh BK, Trivedi P. **2017**. Microbiome and the future for food and nutrient security. *Microbial Biotechnology* **10**: 50-53.

Solana M, Boschiero I, Dall'Acqua S, Bertucco A. **2014**. Extraction of bioactive enriched fractions from *Eruca sativa* leaves by supercritical CO_2 technology using different co-solvents. *J Supercrit Fluid* **94**: 245-251.

Sun RB, Zhang XX, Guo XS, Wang DZ, Chu HY. **2015**. Bacterial diversity in soils subjected to long-term chemical fertilization can be more stably maintained with the addition of livestock manure than wheat straw. *Soil Biology and Biochemistry* **88**: 9-18.

Thiergart T, Duran P, Ellis T, Vannier N, Garrido-Oter R, Kemen E, Roux F, Alonso-Blanco C, Agren J, Schulze-Lefert P *et al.* 2020. Root microbiota assembly and adaptive differentiation among European *Arabidopsis* populations. *Nat. Ecol. Evol.* **4**: 122-131.

Tkacz A, Bestion E, Bo Z, Hortala M, Poole PS. **2020**. Influence of plant fraction, soil, and plant species on microbiota: a multikingdom comparison. *mBio* **11**: e02785-02719.

Toju H, Peay KG, Yamamichi M, Narisawa K, Hiruma K, Naito K, Fukuda S, Ushio M, Nakaoka S, Onoda Y *et al.* 2018. Core microbiomes for sustainable agroecosystems. *Nat. Plants* **4**: 247-257.

Uroz S, Buée M, Murat C, Frey‐Klett P, Martin F. **2010**. Pyrosequencing reveals a contrasted bacterial diversity between oak rhizosphere and surrounding soil. *Environ. Microbiol. Rep.* **2**: 281-288.

Vacher C, Hampe A, Porté AJ, Sauer U, Compant S, Morris CE. **2016**. The phyllosphere: microbial jungle at the plant–climate interface. *Annu. Rev. Ecol. Evol. S.* **47**: 1-24.

Van der Heijden MG, Klironomos JN, Ursic M, Moutoglis P, Streitwolf-Engel R, Boller T, Wiemken A, Sanders IR. **1998**. Mycorrhizal fungal diversity determines

plant biodiversity, ecosystem variability and productivity. *Nature* **396**: 69.

van der Heijden MGA, Hartmann M. **2016**. Networking in the plant microbiome. *PLoS Biology* **14**: e1002378.

Vandenkoornhuyse P, Quaiser A, Duhamel M, Le Van A, Dufresne A. **2015**. The importance of the microbiome of the plant holobiont. *New Phytologist* **206**: 1196-1206.

Vorholt JA. **2012**. Microbial life in the phyllosphere. *Nat. Rev. Microbiol.* **10**: 828-840.

Vorholt JA, Vogel C, Carlstrom CI, Muller DB. **2017**. Establishing causality: opportunities of synthetic communities for plant microbiome research. *Cell Host Microbe* **22**: 142-155.

Wagg C, Schlaeppi K, Banerjee S, Kuramae EE, van der Heijden MGA. **2019**. Fungal-bacterial diversity and microbiome complexity predict ecosystem functioning. *Nat. Commun.* **10**: 4841.

Wagner MR, Lundberg DS, Del Rio TG, Tringe SG, Dangl JL, Mitchell-Olds T. **2016**. Host genotype and age shape the leaf and root microbiomes of a wild perennial plant. *Nat. Commun.* **7**: 12151.

Walters WA, Jin Z, Youngblut N, Wallace JG, Sutter J, Zhang W, Gonzalez-Pena A, Peiffer J, Koren O, Shi Q *et al.* 2018. Large-scale replicated field study of maize rhizosphere identifies heritable microbes. *Proc. Natl. Acad. Sci. U S A* **115**: 7368-7373.

Wu D, Zhao Z, Han X, Meng F, Wu W, Zhou M, Brüggemann N, Bol R. **2018**. Potential dual effect of nitrification inhibitor 3,4-dimethylpyrazole phosphate on nitrifier denitrification in the mitigation of peak N_2O emission events in North China Plain cropping systems. *Soil Biology and Biochemistry* **121**: 147-153.

Yu P, Wang C, Baldauf JA, Tai H, Gutjahr C, Hochholdinger F, Li C. **2018**. Root type and soil phosphate determine the taxonomic landscape of colonizing fungi and the transcriptome of field-grown maize roots. *New Phytologist* **217**: 1240-1253.

Zarraonaindia I, Owens SM, Weisenhorn P, West K, Hampton-Marcell J, Lax S, Bokulich NA, Mills DA, Martin G, Taghavi S *et al.* 2015. The soil microbiome influences grapevine-associated microbiota. *mBio* **6**: 1-10.

Zhang Y, Xu J, Riera N, Jin T, Li J, Wang N. **2017**. Huanglongbing impairs the rhizosphere-to-rhizoplane enrichment process of the citrus root-associated microbiome. *Microbiome* **5**: 97.

Zhao ZB, He JZ, Geisen S, Han LL, Wang JT, Shen JP, Wei WX, Fang YT, Li PP, Zhang LM. **2019**. Protist communities are more sensitive to nitrogen fertilization than other microorganisms in diverse agricultural soils. *Microbiome* **7**: 33.

Supporting Information

The following Supporting Information is available for this article:

Method S1 Detailed description of field trial, molecular methods and statistical analysis.

Fig. S1 Experimental design and the layout of samples in terms of treatments, sites, plant compartment niches, and crop seasons.

Fig. S2 Bacterial community structure and taxonomic composition varied among different niches.

Fig. S3 The functional networks and potential sources of crop-associated microbiomes.

Fig. S4 The relative abundance of dominant phyla/class varied among different niches.

Fig. S5 The potential functional groups and biomarker taxa of bacterial communities varied among different niches.

Fig. S6 The biomarker taxa and potential functions of bacterial communities in each crop.

Table S1 The soil properties and crop yields in two study sites

Table S2 Effects of host niche, crop species, site, and fertilization practice on the bacterial community based on PERMANOVA

Table S3 Effects of host niche, site, and fertilization practice on bacterial community in different seasons based on PERMANOVA

Table S4 Effects of host niche, crop species, site, fertilization practice on bacterial alpha diversity based on linear mixed model (LMM)

Table S5 Bacterial alpha diversity explained by different factors within each niche

Table S6 Core taxa of bacterial community in different niches

Table S7 Shared dominant taxa of three crops

Tables

Table 1 Effects of crop species, site, and fertilization practice on bacterial community structure in different compartment niches based on PERMANOVA

Niche	Crop species		Site		Fertilization practice		Explained variation (%)	HEEI [a]
	R^2 (%)	P (> F)	R^2 (%)	P (> F)	R^2 (%)	P (> F)		
Bulk soil	7.3	< 0.001	56.5	< 0.001	4.6	0.041	77.3	0.13
Rhizosphere soil	16.5	< 0.001	48.3	< 0.001	6.7	< 0.001	81.3	0.34
Rhizoplane	58.8	< 0.001	17.5	< 0.001	2.7	0.049	85.3	3.36
Phylloplane	36.5	< 0.001	44.2	< 0.001	1.9	0.193	87.3	0.83
Root endosphere	52.8	< 0.001	10.7	< 0.001	3.9	0.065	77.5	4.93
Leaf endosphere	60.4	< 0.001	5.6	< 0.001	4.4	0.011	80.3	10.79

[a] HEEI represents "host-environment effects index" (HEEI = relative contribution of crop species / relative contribution of site, based on nested PERMANOVA analysis). The higher HEEI value represents greater host species effect.

925 **Table 2** Bacterial co-occurrence network characteristics in each compartment niche

Niche	Node	Positive edge	Negative edge	Avg. degree	Modularity [a]	Avg. clustering coefficient [b]	Avg. path distance [c]	Hub node [d]
Bulk soil	243	1997	1947	32.46	0.229	0.612	2.118	42
Rhizosphere soil	296	2000	1929	26.55	0.297	0.586	2.545	39
Rhizoplane	419	1985	1999	19.02	0.542	0.533	3.333	18
Phylloplane	269	1989	217	16.40	0.325	0.565	3.337	9
Root endosphere	133	500	42	8.15	0.215	0.594	2.908	0
Leaf endosphere	18	45	3	5.33	0.124	0.761	1.856	0

926 In the network, modularity [a] means the degree of nodes tending to differentiate into different network modules while
927 clustering coefficient [b] represents the degree of nodes tending to cluster together. [c] Network path distance is the
928 length of the shortest path between two nodes within the network. [d] "Hub node" was defined as a node with high
929 values of degree (> 60) and closeness centrality (> 0.3) in the network.

930 **Figure legends**

931 **Fig. 1** Host dominates over environment in shaping crop microbiome. (a) Non-metric
932 multi-dimensional (NMDS) ordinations based on weighted UniFrac distances matrices
933 of bacterial communities for all samples (n = 684), maize samples in 2016 (n = 216)
934 and 2017 (n = 252), and wheat/barley samples in 2017 (n = 216). "XC" represents site
935 "Xuchang, "QJ" represents site "Qujing. (b) NMDS ordinations based on weighted
936 UniFrac distances of bacterial community in each compartment niche from samples in
937 2017 (n = 468). "HEEI" represents host-environment effects index, and the higher
938 HEEI value represents greater host species effect. "Env. Effect" represents
939 environmental effect.

940 **Fig. 2** Crop host has a strong effect on reducing bacterial diversity and network
941 complexity. (a, b) Bacterial co-occurrence networks along the soil-plant continuum
942 based on samples collected in 2017 (n = 468). (c) Bacterial alpha diversity in different
943 niches based on all samples (n = 684). "endo" represents endosphere. (d) The
944 distribution patterns of the "Hub nodes" of bacterial network in different niches.

945 **Fig. 3** Crop-associated bacterial communities are mainly derived from bulk soils and
946 gradually enriched and filtered by different compartment niches. (a) Source Model of
947 Plant Microbiome (SMPM) showing the potential sources of crop-associated bacterial
948 communities based on samples collected in 2017 (n = 468). "U" represents the unknown
949 source, and the thickness of lines equivalent to the source contribution. (b) MA plot
950 illustrating the enrichment and depletion patterns of the crop-associated bacterial
951 microbiomes in each compartment niche compared with bulk soil. Each point
952 represents a single ZOTU (RA > 0.1%, 6744 in total). Each red point represents an
953 individual enriched ZOTU, and green point represents an individual depleted ZOTU.
954 The position along the y-axis represents the abundance fold change compared with bulk
955 soil, and x-axis reports average ZOTU abundance (as count per million, CPM). "DI"
956 represents depleted index. "DSI" represents dissimilarity index. "All" represents the
957 numbers of the total ZOTUs (RA > 0.1%) in each niche. (c) Venn diagrams showing

the shared and specific bacterial ZOTUs in different niches within the significant enriched ZOTUs and depleted ZOTUs. For these shared differentially ZOTUs, only the top 3 taxonomies were shown.

Fig. 4 Phylogenetic tree, taxonomic composition, and distribution patterns of crop dominant taxa. (a) Identification of dominant taxa for maize (only considering leaf and root compartments, n = 312). (b) Identification of dominant taxa for wheat (n = 72). (c) Identification of dominant taxa for barley (n = 72). The dominant taxa were defined as ZOTUs with an average relative abundance ≥ 0.1% and present in more than 60 % of all samples. Low abundance classes with less than 2% of the total sequences of dominant taxa across three crops are grouped into "Other".