Stochastic processes shape the biogeographic variations in core bacterial communities between aerial and belowground compartments of common bean

Running title: Stochastic processes shape core bacteria

Yang Liu, Da Li, Jiejun Qi, Ziheng Peng, Weimin Chen, Gehong Wei*, Shuo Jiao*

State Key Laboratory of Crop Stress Biology for Arid Areas, Shaanxi Key Laboratory of Agricultural and Environmental Microbiology, College of Life Science, Northwest A&F University, Yangling, Shaanxi 712100, People's Republic of China

* Corresponding author. State Key Laboratory of Crop Stress Biology for Arid Areas, Shaanxi Key Laboratory of Agricultural and Environmental Microbiology, College of Life Science, Northwest A&F University, Yangling, Shaanxi 712100, People's Republic of China.

Tel.: +86 29 87091175; Fax: +86 29 87091175; *E-mail addresses:* weigehong@nwsuaf.edu.cn (G. Wei) and shuojiao@nwsuaf.edu.cn (S. Jiao)

This article has been accepted for publication and undergone full peer review but has not been through the copyediting, typesetting, pagination and proofreading process which may lead to differences between this version and the Version of Record. Please cite this article as doi: 10.1111/1462-2920.15227

Originality-Significance Statement

Plant tissues provide unique ecological habitats for microorganisms as similar as rhizosphere soil. However, little is known about the biogeographic distributions and assembly processes of bacterial communities in plant endosphere. Hence, comprehensive sampling types (belowground: bulk and rhizosphere soil, root endosphere; aerial: stem and leaf endosphere) were included in the present study at a regional scale. Core bacterial communities were selected by taking the intersections of the whole communities among all soil-plant associated compartments, which might have plant filtering effects from each other. The stochastic processes increased to higher degree on the stem and leaf endosphere, where there were distinct biogeographic distributions with lower alpha diversities and less compositional variations across the distinct sampling sites compared with belowground compartments. Additionally, core communities had various distance-decay relationships and were affected by different environmental factors among the five compartments. Thus, their coexistences also varied due to niche habitability. Importantly, this study provides comprehensive understanding of different biogeographic and coexistence patterns of core bacterial communities among five soil-plant associated compartments and the predominance of stochastic processes in aerial compartment. It will be easier to predictably manipulate the core bacterial communities using synthetic microflora techniques availably to produce better soil sustainability outcomes and crop yields.

Summary

Although studies of biogeography in soil bacterial communities have attracted considerable attention, the generality of these patterns along with assembly processes and underlying drivers are poorly understood in the inner tissues of plants. Plant tissues provide unique ecological habitats for microorganisms, which play an essential role in plant performance. Here, we compared core bacterial communities among five soil-plant associated compartments of common bean across five sampling sites in China. Neutral and null modeling consistently suggested that stochastic processes dominated the core community assembly processes and escalated from the belowground compartments to the inner tissues of aerial plant parts. The multiple distance-decay relationships also varied and had flattened patterns in the stem endosphere, which were shaped by distinct environmental factors in each compartment. Coexistence patterns also varied in topological features, in addition with the sparsest networks in the stem endosphere resulted from the interaction with the stochastic processes. This study considerably expanded our understanding of various biogeographic patterns, assembly processes, and the underlying mechanisms of core bacterial communities between aerial and belowground compartments of common bean. That will provide a scientific basis for the reasonable regulation of core bacterial consortia to get better plant performance.

Keywords: Core bacteria, Stem and leaf endosphere, Biogeography, Environmental factor, Community assembly, Species coexistence

Introduction

Biogeography is the study of the distribution of biodiversity over space and time and is designed to reveal the habitats and abundances of organisms. Gaining knowledge about the biogeographic patterns of microorganisms provides important key insights into the mechanisms that generate and maintain microbial diversity (Martiny et al., 2006; Tedersoo L et al., 2014). The random and non-random distributions of microbial communities are not only affected by environmental factors (i.e., soil pH or organic carbon) (Rousk et al., 2010; Lin et al., 2015), but also can be significantly correlated to geographic distance (Xiong et al., 2012). Therefore, the distance-decay relationships (DDRs) are widely investigated as frequently used methods (Liu et al., 2015; Shuo et al., 2016; Wang et al., 2017; Jiao et al., 2020) and include multiple types of distance, such as geographical and environmental distances (Chen et al., 2017). It is also important to predict biogeographic variations of microbial community by unraveling the drivers of microbial dynamics in response to different environmental conditions. Currently, microbial biogeography studies are turning their attention toward the microbial community assembly processes (Nemergut et al., 2013), which have been investigated in some agricultural and natural habitats (Shi et al., 2018; Wang et al., 2019).

The formation of microbial biogeography is simultaneously determined by two general factors that, to varying degrees, include local environmental factors (i.e., contemporary factor/deterministic processes) and regional process of geographic distance (i.e., historical factor/stochastic processes) (Fierer and Jackson, 2006; Nemergut et al., 2011; Liu et al., 2014). Similarly, the effects of neutral and niche-based theory in microbial community assemblages and their interaction on shaping biogeographic patterns have been widely debated (Hanson et al., 2012). The deterministic processes include selection imposed by biotic factors that trigger interactions among species and abiotic factors via environmental filtering (Lozupone and Knight, 2007), and the stochastic processes include probabilistic dispersal, ecological drift, and random birth-death events (Zhou and Ning, 2017). Recently, a new method has been developed to separate the relative effects of deterministic and stochastic processes according to the framework of Stegen et al. (Stegen et al., 2015). The biogeography of agricultural bulk and rhizosphere soil bacterial communities has been increasingly investigated (Fan et al., 2017; Fan et al., 2018; Shen et al., 2019), while the biogeographic variations and DDRs of plant-tissue associated endophytic bacterial communities are rarely studied.

Plant tissues provide unique ecological habitats for the diverse microorganisms that live in them, and the plant-associated microbiome and complex plant-microbe interactions play an essential role in plant performance and resistance to abiotic stress (Ikeda et al., 2010; Bulgarelli et al., 2013). Some studies have investigated the biogeographic patterns and their assembly processes of soil-root associated microbial communities (Zhang et al., 2018a; Zhang et al., 2018b). There are

also studies that have briefly examined the biogeographic distribution of the plant microbiome in *Agave* species (Coleman-Derr et al., 2016) and in *Cycas panzhihuaensis* (Zheng and Gong, 2019), but the assembly processes of the plant shoot-associated endosphere (stem and leaf) remain unknown. Such that bacterial biogeographic patterns and community assembly processes have not been compared among the bulk soil, rhizosphere soil, and plant-associated endosphere. Considering the critical roles of these communities in agricultural fields, understanding their community assembly processes could provide explanations for biogeographic patterns, facilitate their management for enhanced agricultural production, and also enable predictions of ecosystem sustainability in response to environmental changes (Chen et al., 2017). Therefore, more sampling types associated with distinct soil-plant associated compartments are needed to compare with each other. The coexistence of species provides insights into the interactions of complex microbial communities (Barberan et al., 2012; Shuo et al., 2016). The interactions of species as biotic factors and the community assembly processes interact with each other (Stegen et al., 2012; Jiao et al., 2020). However, the variation in coexistence patterns and how community assembly processes are related to the ecological strategies of species interactions have not been comprehensively characterized among different soil-plant associated compartments.

Applying a core bacteria approach, in which the focus is only on the persistent members of the microbial community that appear in almost all assemblages associated with a particular host, is becoming increasingly popular (Kumar et al., 2017; Lemanceau et al., 2017; Hamonts et al., 2018; Shuo et al., 2019). The plant core microbiome is critical for plant performance based on

evolutionary processes and strong plant filtering effects that result in selection and enrichment of the microbiota among different plant compartments (Lemanceau et al., 2017; Hamonts et al., 2018). Therefore, identification of the core bacteria communities may be an important step in identifying key members of the bacterial community that sustain plant heath and development. Defining the biogeographic variations and assembly processes of the core bacterial community in plants could thus help expand our knowledge of bacterial communities in general.

Common bean (*Phaseolus vulgaris* L.) is a legume species that represents an important source of protein in the human diet (Yêyinou et al., 2018), and it is grown as a major crop in the worldwide region (Rendon-Anaya et al. 2017), particularly in China because of dense population. This enables sampling across a broad range (~1,000 km) of geographical locations and environmental gradients in an agroecosystem at a regional scale. In this study, five soil-plant associated compartments (bulk soil, rhizosphere soil, root, stem and leaf endosphere) across five sampling sites were subjected to high quality, long-read single-end Ion S5^{TM}XL DNA sequencing (Mehrotra et al., 2017) of the bacterial nuclear 16S rRNA gene with the goals of 1) assessing the geographical variations in alpha and beta diversity; 2) characterizing the distance-decay relationships and dynamics regulated by environmental factors; and 3) elucidating distinction of community assembly processes and coexistences of the core bacterial communities among the five compartments. Accordingly, we hypothesized that 1) biogeographic patterns vary markedly among the five compartments, and 2) the effect of stochastic processes on the aerial compartments are stronger due to specific niche adaptation to plant habitats.

Results

Biogeographic distribution of bacterial communities

After filtering low quality reads, 7,773,758 clean sequencing reads remained with an average of 77,737 ± 7,527 reads/sample from 100 samples (20 samples per soil-plant associated compartment). In total, 5,748,450 reads could be classified as bacteria, and most (98.2%) could be classified at the phylum level. A total of 5,954 OTUs were clustered with an average of 962 ± 609 OTUs. After homogenization, subsequent analyses were performed based on a minimum of 38,957 reads per sample. The different soil-plant associated compartments lead to the adaptation of niche-specialized bacterial inhabitants. To identify the generalist (ubiquitous) taxa in each compartment, core microbial taxa were primarily selected from the OTUs appearing (100% of prevalence) in all of the 100 samples. In total, 611 core OTUs were shared among all plant compartments (Fig. 1A), and they accounted for 10.3% of all observed taxa. These core taxa accounted for 57.8% of all bacterial sequences on average, and they were also classified to different phylum, including abundant taxa of *Proteobacteria* (46.2%), *Actinobacteria* (29.9%), and *Bacteroidetes* (8.5%), with *Firmicutes* (4.9%) at lower relative abundance. Additionally, their relative abundance also varied across different compartments and sampling sites at various levels (Suppl. Figs. S2A and S3).

The core communities were not only strongly related to the other communities (exclusion of core communities) (r =0.84, P =0.001) (Suppl. Fig. S2B), but had similar alpha diversity and

composition with the overall communities (Fig. 1B). Specifically, the Shannon index was ordered BS>RS>R>S>L and differed significantly (P <0.05) among the bulk soil, root-associated (RS and R), and shoot-associated (S and L) compartments across the overall and core communities. The ACE index had similar varied patterns compared with the Shannon diversity (except that L>S) (Suppl. Fig. S2C). In addition, these two alpha diversity indices of the core communities were lower than for the overall communities and also varied from distinct sites in each compartment (Suppl. Fig. S4). Particularly, the factors of compartments and sites always had significantly varied effects on these two indices (Table S1). The composition of the overall and core communities were clearly clustered in accordance with the greatest influence of soil-plant associated compartments [core (overall): Adonis: R^2 = 0.466 (0.443), Anosim: R =0.752 (0.753), MRPP: δ = 0.595 (0.618), P <0.001] (Fig. 1C and Suppl. Fig. S2D and Tables S2 and S3). The whole community was then split into five compartments that were significantly separated into distinct sites to varying degrees in each compartment (Fig. 1D and Suppl. Fig. S2E) and were confirmed by three test methods (i.e., S: R^2 = 0.522 (0.521), R = 0.495 (0.505), δ = 0.436 (0.443), P <0.001). In general, the diversity and composition of the core and overall communities varied in a manner similar to the different biogeographic patterns in each compartment. Therefore, we focused on these core communities and further explored them.

Distance-decay relationships and dynamics of the core bacterial communities

The distance-decay relationships among the core community similarity and the geographic and environmental (including edaphic and climatic) distances were elucidated across all sites. We

found that community similarity in the different compartments had significantly distinct decay patterns with the increase in these four kinds of distances, except that the stem endosphere did not vary significantly across edaphic distance (P >0.05) (Fig. 2). The geographic decay trend gradually declined from the bulk soil into the leaf endosphere based on the flattened slopes of the model (BS: slope = -0.020, RS: slope = -0.009, R: slope = -0.006, S: slope = -0.005, L: slope = -0.004). The slopes of the environmental decay patterns were ranked in the order BS (-0.058) <RS (-0.051) <R (-0.048) <L (-0.042) <S (-0.017), which was similar to the edaphic and climatic decay patterns. Moreover, the degree of the geographic decay patterns was always lower than the environmental patterns in each compartment resulting from the higher slope values. The degree of the edaphic decay patterns was higher in the rhizosphere, root, and leaf endosphere; however, the bulk soil and stem endosphere had steeper decay patterns across climatic distance (Table 2).

Additionally, the distance-decay patterns were confirmed by the results of (partial) Mantel tests. In general, the environmental distance had greater effects on all of the communities (partial Mantel test: BS: r=0.420, RS: r=0.509, R: r=0.675, L: r=0.432, P=0.001) except for stem endosphere, which was more affected by geographic distance (Suppl. Table S4). Furthermore, the edaphic distance had stronger correlations with community dissimilarity in rhizosphere soil, root and leaf endosphere, and the climatic distance was closely related to bulk soil and stem endosphere (Table 3).

Because of the stronger multicollinearity among environmental variables, the multiple stepwise and random forest regression methods synthetically elucidated the best predictors driving

the dynamics of core bacterial communities. The environmental variables that showed significant correlations with the dissimilarity of bacterial community in each compartment (Suppl. Table S5) were selected for further regression analyses. We found that minimal temperature of coldest month (MIT) and mean annual temperature (MAT) were major factors in shaping the core community composition of the bulk soil and stem endosphere, respectively. Additionally, total carbon (TC) and available nitrogen (AN) were major factors in shaping the core community composition of the rhizosphere, leaf and root endosphere, separately (Table 4). TC and AN were not the most important factors in the rhizosphere and root endosphere from the results of random forest, but they always had significant effects (P <0.05) (Suppl. Table S6). In particular, the regulatory dynamics of the core bacterial communities were corresponded with the results of the (partial) Mantel tests. Thus, we concluded that there were stark differences in the dynamics among five soil-plant associated compartments, suggesting that distinct specific drivers have essential roles for community assembly in each compartment (Table 1).

Assembly processes and coexistence patterns of the core bacterial communities

The neutral and null models were combined to investigate the community assembly in each soil-plant associated compartment across all sites. Firstly, part of the relationship was estimated by the Sloan neutral model between the occurrence frequency of the OTUs and their relative variation in abundance, with 73% (BS), 56% (RS), 69% (R), 71% (S), and 63% (L) of the explained core community variance. The metacommunity size times immigration between communities (Nm-value) was higher for the leaf (Nm = 36206) and stem (Nm = 36477) endosphere than for the

bulk soil (Nm = 21762). The migration rates (m values) were enhanced from 0.09 (L) to 0.24 (R), which showed different variation tendencies with the fit of the neutral model (R^2) and manifested leaf and stem endosphere had more migrators between communities despite the low migration rate (Fig. 3A). Despite the fact that the model explained most of the microbial community variation, other community assembly mechanisms still existed at the same time, which caused the un-neutral distribution. For example, there were more root endosphere OTUs below the expectation (under-represented), and there were more rhizosphere OTUs above the expectation (over-represented) (Fig. 3B).

Secondly, we deeply elucidated the assembly processes based on the null model. The β-nearest taxon index (βNTI) scores of the core communities in the range of -2 to +2 accounted for 50.5% in bulk soil, 62.1% in rhizosphere soil, 77.9% in the root endosphere, 99.5% in the stem endosphere, and 100% in the leaf endosphere, and these scores are indicative of neutral processes that are greater in the stem and leaf endosphere (Fig. 4A). These results were further identified by combining values of βNTI with the Bray-Curtis-based Raup-Crick (RC_{Bray}). We found that the aerial plant tissues (stem and leaf) were more affected by undominated processes. Homogenous dispersal dominated in leaf and stem endosphere in the absence of undominated processes. Variation selection dominated in bulk soil, and dispersal limitation had greater effects on the root endosphere, rhizosphere soil, bulk soil, and stem endosphere communities to declining degrees (Fig. 4B). Consequently, stochastic processes dominated the assembly of the core community across different compartments to varying degrees (Suppl. Table S7). In addition, significantly

higher mean community-level B-values (*B*com) were observed in bulk and rhizosphere soil, while the root, stem and leaf endosphere had lower values (Fig. 4C). This indicated that wider microbial habitat breadth existed in bulk and rhizosphere soil, compared with plant inner tissues (root, stem and leaf endosphere).

The community assembly was also affected by biotic factors represented by species interactions. The networks topological features were similar based on the three methods, which were also identified by each other (Table 5 and Suppl. Tables S8 and S9). And we finally chose the Spearman correlated networks to conduct further investigation because of appropriate number of nodes and edges. There were significantly distinct interactions between neutral and un-neutral species based on the coexistence patterns of core communities among the five compartments (Fig. 5). The content of neutral species in the corresponding networks was higher in the stem (72.2%) and leaf (67.7%) endosphere, which indicated the contribution of species interactions to the stochastic processes. The soil of bulk and rhizosphere had more complex interactions that resulted from the edges and other network topographical features, such as average degree (Table 5). More competition relationships between species were manifested in the bulk soil according to the most negative edges (557). The ratio of positive links increased from bulk soil (77.9%) to stem endosphere (100%), which manifested the higher cooperation in the stem endosphere corresponding with the highest modularity values (0.748). It was also the largest and sparsest network, according to some specific network topological features in the stem, such as average degree (4.256), diameter (19), density (0.016) and average path distance (6.785). Moreover,

random networks were generated with the same nodes and edges in each compartment to verify that our empirical networks were non-random. Node-level topological properties also varied with the response to network-level topological features among the five compartments (Suppl. Fig. S5).

Discussion

Niche habitability leads to distinct environments (Coleman-Derr et al., 2016; Zheng and Gong, 2019) in the different soil-plant associated compartments. Hence, there were obvious separations of the bacterial community compositions. In the present study, core bacterial communities were particularly selected to elucidate their biogeographic patterns, assembly processes, and coexistences. In support of our hypotheses, stochastic processes dominated the assembly of the core bacterial communities among the different compartments, with a larger influence on the stem and leaf endosphere, which had lower alpha diversities and less compositional variations across the distinct sampling sites. The stem and leaf endosphere also had flattened environmental and climatic DDRs with respect to homeostasis and were less affected by soil conditions. Hence, they faced less dispersal limitation. There were distinct best predictors shaping the bacterial community composition because of niche adaptation to the different compartments. In general, coexistence patterns varied and had the sparsest structure in the stem endosphere, which were inferred to be interacted with stochastic processes each other.

Variations in alpha diversity and composition of the core bacterial communities

The core taxa were selected among five soil-plant associated compartments on account of

niche differentiation. They are more representative and have rarely been investigated by other studies that usually pay more attention to the overall bacterial communities (Zhang et al., 2018a; Zheng and Gong, 2019) or to the abundant and rare microbial communities in the distinct ecosystems (Chen et al., 2019; Jiao and Lu, 2020). The variations in relative abundance of *Proteobacteria* and *Actinobacteria* were similar with the results of previous studies (Edwards et al., 2015; Fan et al., 2017). In particular, the *Cyanobacteria* phylum and the *Heliobacteriaceae* family of *Firmicutes* could help plants fix CO_2 due to their ability to photosynthesize (Meeks and Elhai, 2002; Tanai and Cardona, 2015). Thus, they were more abundant in the leaf and stem endosphere. The representativeness of the core bacterial communities was further confirmed by similar alpha and beta variation tendencies with the overall communities. Generally, the alpha diversity decreased with the root proximity, which is consistent with previous studies (Fan et al., 2017; Zhang et al., 2018a). However, the stem and leaf endosphere manifested lower diversity, possibly because aerial plant tissues may have a more homogeneous homeostasis without soil disturbances or have plant filtering effects such as the ‘root enrichment process’ (Edwards et al., 2015), which needs to be further studied.

The core bacterial communities were mostly clustered by compartment, which is similar with a study showing that niche differentiation shapes the diversity and composition of the *Cycas panzhihuaensis* microbiome (Zheng and Gong, 2019). In addition, the differences in certain environmental factors (e.g., TC, TN, MAT, and MIT) among the sampling sites were significant in our study. The composition was significantly separated by sampling sites to varying degrees in

each compartment but with less variation in the stem. This is identified by a previous study that reported how biogeography affects the microbiome composition in three species of *Agave* from distinct sites (Coleman-Derr et al., 2016). Therefore, the interactions among different geographical and environmental conditions can shape the distinct microbial community composition (Lopez et al., 2020).

Effects of geographical and environmental factors on the core bacterial communities

Four types of distance-decay relationships (DDRs) were compared. This enriched the general distance-decay models that were reported by other studies (Chen et al., 2017; Chen et al., 2019; Feng et al., 2019), because environmental factors also varied with geographical distance. The flattened geographic DDRs were mostly exhibited in the stem and leaf endosphere. This was not surprising because the stem and leaf faced less dispersal limitation in our study according to Hubbell's neutral theory (Hubbell, 2001). However, decreasing geographic DDRs were found from the bulk soil to the rhizosphere, this was contrary to other studies (Zhang et al., 2017; Zhang et al., 2018b). These differences could possibly result from distinct microbial taxa or sampling scales. The most flattened environmental and climatic DDRs occurred in the stem endosphere where community compositions are homogenized to single homeostasis. Moreover, the stem endosphere was more affected by geographical distance. In general, these results confirmed our initial prediction that biogeographic patterns remarkably varied among the five compartments.

There were distinct best predictors shaping the core bacterial community compositions among the five compartments, and the results were consistent with (partial) Mantel correlations.

Despite of significantly different soil pH among our sampling sites, the climatic factors (e.g. MIT) were important in the bulk soil in our study, which is different from many studies where soil pH was a dominant factor (Fan et al., 2018; Feng et al., 2019; Shen et al., 2019). That may be due to the inhomogeneous biogeographic distribution of sampling sites in our datasets. And there is another published study showing that climatic factors have significant effects on bulk soil bacterial communities (Zhang et al., 2018a). Soil TC played a major role in the leaf endosphere and rhizosphere soil, and this may be due to photosynthesis in the leaf and more root secretions of carbon sources into the rhizosphere. Soil AN dominated in the root endosphere, which could correlate with the role of root nodules in common bean (Ikeda et al., 2010).

Effects of stochastic processes and species interactions on the core community assembly

The influence of stochastic processes on community assembly should not be inferred only from the neutral model (Hanson et al. 2012); thus, the combined results of two popular models were considered in the current study. The results synthetically elucidated that stochastic processes dominated the assembly of core communities among the different compartments. This supports our second hypothesis and is related to more homogeneous homeostasis in the leaf and stem endosphere. Not only were the stem and leaf less affected by the soil, but the stem endosphere was more stable compared with the leaf based on the fact that leaves have larger areas of contact with the outside and are sensitive to changes in the environment. It was inferred that microbes in the air can invade aerial compartments (i.e., especially in the leaf) at random, but that soil properties and root exudates exert stronger selection forces in the belowground compartments (Mendes et al.,

2014). Additionally, decreased niche differentiation among species in a more homogeneous environment occurs in plant inner tissues and could lead to increased neutrality in community assembly (Bar-Massada et al., 2014). This is similar to the phenomenon that stochastic processes in planktonic microeukaryotes are higher in the wet seasons where well-stocked rivers make the environment more uniform (Chen et al., 2019).

Correspondingly, the dominance of environmental filtering and soil heterogeneity selection was less in the stem and leaf endosphere resulting from DDRs and Mantel tests in our study. Hence, homogenous dispersal dominated in the leaf and stem endosphere. Some studies have demonstrated that deterministic processes dominate the assemblage of microbial communities in the soil of bulk and rhizosphere (Fan et al., 2017; Zhang et al., 2018b). While, dissimilar results emerged in our study showing that variation in selection occurred in the bulk soil, but the stochastic processes still dominated in the compartments. This might require further research and discussion based on a wider range of sampling sites in future. The degree of influence of stochastic processes was also confirmed by the Nm and migration rates (m) values. The core OTUs were regarded as generalist in each compartment, and the wider habitat breadth (*B*com) means higher metabolic flexibility (Wu et al. 2018), suggesting that generalist may be more distributed in the soil of bulk and rhizosphere where there were rich nutrients (Pandit et al. 2009). We inferred that lower habitat breadth may limit the migrated spaces of some core species in the stem and leaf endosphere, and thus dispersal limitation may contribute to the assembly of these core bacterial communities. Despite the clear division of community assembly processes, there is

still debate about their mechanisms (Hanson et al., 2012; Zhou and Ning, 2017). The niche differentiation was significant, which was different from other studies including the consistency of sample types (Fierer and Jackson, 2006; Chen et al., 2017). Therefore, more in-depth studies and clear evidence are needed in future.

The neutrally distributed taxa were mostly derived from stochastic processes (Chen et al., 2019), and had different coexistence patterns and occurrence rates in our networks. The coexistences among species should also be responsible for the community assembly (Hanson et al., 2012; Hu et al., 2020). Thus, coexistence patterns were inferred to have a certain effect on the stochastic processes as biotic factors in our study, but the specific quantification of the effect is still challenging (Feng et al., 2019). Niche habitability leads to distinct environmental (Coleman-Derr et al., 2016) and coexistence patterns among the different compartments. The soil of bulk and rhizosphere had more complicated networks with respect to heterogeneous bulk soil properties and the rapid need for high speed element cycling and signal transmission in the rhizosphere soil (Hinsinger, 2001). The competition was robust in bulk soil, which is consistent with the results of a study finding bulk soil has complicated environments (Fan et al., 2017). Greater resource availability is thought to reduce competition in microbial communities (Hubbell, 2005). Hence the stem had the largest and sparsest network with higher cooperated links according to their stably enriched nutrients. And then, the endophytic bacteria may be transmitted from other compartments. Moreover, this could result from stronger stochastic processes in the stem endosphere according to the mass effect (Lindström ES, 2012), and that the sparse network is

likely to be associated with stochastic processes (Cornwell et al., 2006). Overall, we inferred that species coexistences may interact with stochastic processes resulting from the sparser network structure and higher occurrence rates of neutrally distributed taxa appeared in the stem and leaf endosphere.

Conclusions

The biogeographic patterns of core bacterial communities in aerial plant compartments (stem and leaf endosphere) were significantly different than the belowground compartments (root endosphere, soil of bulk and rhizosphere) based on the results of varied alpha diversity and community composition under regional sampling scales. The different distance-decay relationships were manifested from the gradient of spatial, edaphic, and climatic factors among the five compartments. Differential niche adaptations were found for bacterial communities in each compartment where there were distinct environmental drivers specifically dominated. We have considerably advanced this research by showing that stochastic processes dominated the community assembly to varying degrees among the five compartments based on the combined results from neutral and null modes. In addition, coexistence patterns were distinct from the five compartments and partially interacted with stochastic processes. This study provides an integrated understanding of biogeographic patterns, assembly processes, coexistence patterns, and the underlying mechanisms of core bacterial communities among different compartments. Further studies are needed with more extensive sampling sites to dissect the specific functional roles of

core species, which could make it easier to predictably manipulate the core bacterial communities to produce better soil sustainability outcomes and crop yields.

Experimental Procedures

Soil sampling and characteristics collection

Five sampling sites across from the northeast to the northwest of China, extending between 38.81–49.76 °N and 111.65–126.85 °E (Suppl. Fig. S1), were selected based on their larger scale distribution and long-term planting of cultivated common bean. One large field (1-1.5 ha) was selected in each site, and four plots (15 * 15 m^2 per plot) in each field were sampled based on a Z-shaped pattern as biological replicates in July 2017 during the flowering period of *P. vulgaris*. And the entire fields in our study were subjected to similar fertilization and management practices. The samples consisted of five soil cores (0-20 cm) and the corresponding six intact and growing well plants that were collected randomly and pooled as one biological replicate of bulk soil (BS) and plant samples in each plot, respectively.

The soil and plant samples were transported to the laboratory on dry ice as quickly as possible. The bulk soil was sieved through a 2 mm mesh to remove visible roots, residues, and stones, and the plants were carefully divided into the roots, stems, and leaves to further obtain rhizosphere soil (RS), root (R), stem (S), and leaf (L) endosphere samples based on a routine sampling method, the details of which are given in Zheng and Xiao et al. (Xiao et al., 2017; Zheng and Gong, 2019). Consequently, a total of 100 samples (five sites × five compartments × four

plots) were obtained and stored at -80ºC prior to further processing.

A subset of bulk soil samples was air-dried and analyzed for edaphic properties using standard test methods (Bao 2005). The tests conducted were soil pH, soil water content (SWC), soil organic matter (SOM), total nitrogen (TN), total carbon (TC), available nitrogen (AN), total phosphorus (TP), available phosphorus (AP), and available potassium (AK) as our previously described (Liu et al., 2020). The climatic data, namely mean annual temperature (MAT), temperature seasonality (TS), max temperature of warmest month (MT), min temperature of coldest month (MIT), annual temperature range (TR), mean annual precipitation (MAP), and precipitation seasonality (PS) for all sampling sites were taken from the WorldClim database (http://www.worldclim.org) (Hijmans et al., 2005). The information of all characteristics across all five sampling sites in China was given in Table 1.

DNA extraction, sequencing and bioinformatic analyses

The total DNA was extracted from bulk and rhizosphere soil samples (0.5 g each) using the Fast DNA® SPIN Kit (MP Biochemicals, Solon, USA) and from the root, stem, and leaf endosphere samples (0.5 g each) using a DNA Secure Plant Kit (Tiangen Biotech, Beijing, China) following the manufacturers' procedures. The DNA concentration and purity were estimated using a Nanodrop 1000 spectrophotometer (Thermo Fisher Scientific, Waltham, MA, USA) and electrophoresis in 1% (w/v) agarose gels (Xiao et al., 2017).

Because of the sample heterogeneity, we selected the hypervariable V5-V7 region of the 16S rRNA gene for amplification using the primers pair 799F2 (5′- GAT GGC CAT TAC GGC C-3′)

and 1193R (5′- ACG CAT CCC CAC CTT CCT C-3′) (Bai et al., 2015). The polymerase chain reaction (PCR) amplifications were performed in triplicate for each sample, and the PCR products were mixed in equidensity ratios. The mixed PCR products were then purified using a GeneJET™ Gel Extraction Kit (Thermo Scientific). DNA sequencing libraries were generated using the Ion Plus Fragment Library Kit (48 rxns; Thermo Fisher Scientific). The library quality was assessed using the Qubit@ 2.0 Fluorometer (Thermo Fisher Scientific). Finally, the library was sequenced on an Ion S5™XL platform (Thermo Fisher Inc., Waltham, MA, USA) and 400 bp single-end reads were generated by Novogene (Beijing, China). Sequencing datasets are available on the NCBI website under the accession numbers SRR12276901 through SRR12276999. The processed data and scripts are available at https://github.com/YangLiu0910/Stochastic-processes-shape-the-biogeographic-variations-in-core-bacterial-communities.

Cleaned sequence reads were obtained from the process of shearing of low-quality sequences using Cutadapt (Martin, 2011), followed by quality-filtering using the QIIME pipeline (v1.7.0) (Caporaso et al., 2010) and removal of chimeric sequences using the USEARCH tool in the UCHIME algorithm (Edgar et al., 2011). The cleaned sequence reads were then assigned to operational taxonomic units (OTUs) at similarities of 97% using the UPARSE pipeline (Edgar et al., 2011). OTUs lacking more than two sequences were removed. Taxonomic information was annotated for a representative sequence of each OTU using the RDP classifier at a confidence level of 80% (Wang et al., 2007) using the SILVA database release 132.

Statistical analyses

All statistical analyses were conducted in the R software environment (v3.6.1; http://www.r-project.org/). Most of the results were visualized using the 'ggplot2' package (Ginestet, 2011), unless otherwise indicated. And all of the *P* values were adjusted using the false discovery rate method (Benjamini and Hochberg, 1995). The corresponding variables were log or log (x + 1) transformed across distinct sampling sites and soil-plant associated compartments, respectively. The ANOVA and multiple comparisons (Tukey's HSD test) were then performed using the 'multcomp' package (Torsten et al., 2008) after the normality of residues and homogeneity of variance were checked using the Shapiro–Wilk and Bartlett tests, respectively. The core communities were selected by identifying the commonly shared OTUs among five compartments in 20 samples (5 sites × 4 plots per each compartment) and visualized using the 'VennDiagram' package (Chen and Boutros, 2011). The overall and core bacterial community compositions were illustrated by a principle coordinate analysis (PCoA) based on Bray-Curtis distances using the 'ape' package (Paradis et al., 2004) and tested by three different but complementary non-parametric multivariate statistical analysis methods; permutational multivariate analysis of variance (PERMANOVA), analysis of similarities (Anosim), and the multiple response permutation procedure (MRPP) using the 'vegan' package (Dixon, 2003). Heatmaps were illustrated based on Z-score-normalized relative abundance of taxa using the 'pheatmap' package (Kolde, 2015).

The distance-decay relationships (DDRs) were investigated between the core community

similarity and geographic, environmental, edaphic, and climatic distances using liner regression analysis. The slopes of the DDRs can vary based on habitat, which reflects different rates of species turnover. The effects of geographic, environmental, edaphic, and climatic factors on core community dissimilarity were conducted with the Spearman Mantel and partial Mantel tests using the 'vegan' package in each compartment. The significantly correlated environmental factors were selected by the Spearman Mantel test. These environmental factors were then scaled and further selected as best predictors for the corresponding core communities by combining two model analyses of multiple stepwise regression and random forest using the 'relaimpo' and 'rfPermute' packages (Shuo et al., 2019), respectively. The significance of the random forest model was tested using the 'randomForest' and 'A3' packages (Svetnik V et al. 2003).

The assembly processes of core communities were investigated by both neutral and null models among the five soil-plant associated compartments across all sites. Firstly, to determine the potential importance of stochastic processes on community assembly, we used the Sloan neutral community model to predict the relationship between OTU detection frequency and their relative abundance (Sloan et al. 2006). In this model, Nm and migration rates (m) were defined as the metacommunity size times immigration and estimates of dispersal rate between communities. And metacommunity size (N) is the total number of reads of core OTUs in each compartment. The parameter R^2 represents the overall fit to the neutral model. Calculation of 95% confidence intervals around all fitting statistics was done by bootstrapping with 1,000 bootstrap replicates. All of them were fitted using a custom program reported previously by Burns et al. (Burns et al.,

2016). In addition, the core species from each compartment were subsequently separated into three partitions depending on whether they occurred more frequently than (over-represented OTUs), less frequently than (under-represented OTUs), or within (neutrally distributed OTUs) the 95% confidence interval of the neutral model predictions. Secondly, a null-modeling-based statistical framework (Stegen et al., 2015) was used to quantify the contributions of various ecological processes to bacterial community structure and biogeography. The model expectation was generated using 999 randomizations. In this framework, the variation in both phylogenetic diversity and taxonomic diversity was measured using null-model-based phylogenetic and taxonomic β-diversity metrics, namely β-nearest taxon index (βNTI) and the Bray-Curtis-based Raup-Crick (RC_{Bray}) using the 'picante' (Kembel et al., 2010) and 'ecodist' packages (Goslee and Urban, 2007), respectively. Four assembly processes were then quantified as variation (heterogeneous) selection, homogenous selection, dispersal limitation, homogenous dispersal, and undominated based on the threshold of the absolute values of βNTI (2) and RC_{Bray} (0.95) (Zhou and Ning, 2017). The 'niche breadth' approach (Levins, 1970) was used to quantify habitat specialization, and the community-level B-value (*Bcom*) was calculated as the average of B-values from all species occurring in one community to reveal the contributions of species selection and dispersal limitation to microbial community assembly using the 'spaa' package (Zhang et al., 2016).

Coexistence patterns of core bacterial communities (homogenized absolute sequences or relative abundance matrixes) in each soil-plant associated compartment were constructed based on

robust Spearman correlation ($|r| > 0.8$, $P < 0.01$) ['igraph' package (Csardi and Nepusz, 2006)], SparCC [SparCC module ($|r| > 0.8$, $P < 0.01$) in python (Friedman and Alm, 2012)], and SPIEC-EASI [SpiecEasi' package (method = 'glasso', λ.min.ratio=0.01 and nλ=30) (Kurtz et al., 2015)] methods, separately. Topological features of nodes (i.e., degree, betweenness, and closeness centrality) and networks (i.e., diameter, density, average degree, average path distance, average clustering coefficient, and modularity) were also calculated. And only Spearman correlated networks were visualized with the interactive platform Gephi (Heymann, 2014). Meanwhile, 1000 Erdös-Réyni random networks of equal size were constructed as real networks for each compartment based on Spearman correlated networks (Rényi, 2012).

Author Contributions

Y.L. conducted the experiments, analyzed the data, and wrote the manuscript. G.W. and S.J. conceived and designed the experiments and revised the manuscript. Both authors read and approved the final manuscript.

Acknowledgments

This work was funded by the National Natural Science Foundation of China (No. 41830755 and 41807030). We thank Prof. Baili Feng, Dr. Xianchen Li and Xianxing Meng for assistance in soil sampling. We also thank the anonymous reviewers for comments on the manuscript.

Conflict of interest

The authors have no conflicts of interest to declare.

References

Bai, Y., Muller, D.B., Srinivas, G., Garrido-Oter, R., Potthoff, E., Rott, M. et al. (2015) Functional overlap of the *Arabidopsis* leaf and root microbiota. *Nature* **528**: 364-369.

Bar-Massada, A., Kent, R., and Carmel, Y. (2014) Environmental heterogeneity affects the location of modelled communities along the niche-neutrality continuum. *Proceedings Biological Sciences* **281**: 20133249.

Barberan, A., Bates, S.T., Casamayor, E.O., and Fierer, N. (2012) Using network analysis to explore co-occurrence patterns in soil microbial communities. *ISME J* **6**: 343-351.

Benjamini, Y., and Hochberg, Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**: 289-300.

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

Burns, A.R., Zac Stephens, W., Stagaman, K., Wong, S., Rawls, J.F., Guillemin, K., and Bohannan, B.J. (2016) Contribution of neutral processes to the assembly of gut microbial communities in the zebrafish over host development. *ISME J* **10**: 655-664.

Caporaso, J.G., Kuczynski, J., Stombaugh, J., Bittinger, K., Bushman, F.D., Costello, E.K. et al. (2010) QIIME allows analysis of high-throughput community sequencing data. *Nat Methods* **7**: 335-336.

Chen, H., and Boutros, P.C. (2011) VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R. *BMC Bioinformatics* **12**: 35-30.

Chen, R., Zhong, L., Jing, Z., Guo, Z., Li, Z., Lin, X., and Feng, Y. (2017) Fertilization decreases compositional variation of paddy bacterial community across geographical gradient. *Soil Biology & Biochemistry* **114**: 181-188.

Chen, W., Ren, K., Isabwe, A., Chen, H., Liu, M., and Yang, J. (2019) Stochastic processes shape microeukaryotic community assembly in a subtropical river across wet and dry seasons. *Microbiome* **7**: 138.

Coleman-Derr, D., Desgarennes, D., Fonseca-Garcia, C., Gross, S., Clingenpeel, S., Woyke, T. et al. (2016) Plant compartment and biogeography affect microbiome composition in cultivated and native *Agave* species. *New Phytol* **209**: 798-811.

Cornwell, W.K., Schwilk, D.W., and Ackerly, D.D. (2006) A trait-based test for habitat filtering: convex hull volume. *Ecology* **87**: 1465-1471.

Csardi, G., and Nepusz, T. (2006) The igraph software package for complex network research. *InterJournal, complex systems* **1695**: 1-9.

Dixon, P. (2003) Vegan, a package of R functions for community ecology. *Journal of Vegetation Science* **14**: 927-930.

Edgar, R.C., Haas, B.J., Clemente, J.C., Quince, C., and Knight, R. (2011) UCHIME improves sensitivity and speed of chimera detection. *Bioinformatics* **27**: 2194.

Edwards, J., Johnson, C., Santos-Medellín, C., Lurie, E., and Sundaresan, V. (2015) Structure, variation, and assembly of the root-associated microbiomes of rice. *Proc Natl Acad Sci U S A* **112**: E911-920.

Fan, K., Weisenhorn, P., Gilbert, J.A., Shi, Y., Bai, Y., and Chu, H. (2018) Soil pH correlates with the co-occurrence and assemblage process of diazotrophic communities in rhizosphere and bulk soils of wheat fields. *Soil Biology & Biochemistry* **121**: 185-192.

Fan, K., Cardona, C., Li, Y., Shi, Y., Xiang, X., Shen, C. et al. (2017) Rhizosphere-associated bacterial network structure and spatial distribution differ significantly from bulk soil in wheat crop fields. *Soil Biology & Biochemistry* **113**: 275-284.

Feng, M., Tripathi, B.M., Shi, Y., Adams, J.M., Zhu, Y.G., and Chu, H. (2019) Interpreting distance-decay pattern of soil bacteria via quantifying the assembly processes at multiple spatial scales. *Microbiologyopen* **8**: e00851.

Fierer, N., and Jackson, R.B. (2006) The diversity and biogeography of soil bacterial communities. *Proc Natl Acad Sci U S A* **103**: 626-631.

Ginestet, C. (2011) Ggplot2: Elegant graphics for data analysis. *Journal of the Royal Statistical Society* **174**: 245-246.

Goslee, S., and Urban, D. (2007) The ecodist package for dissimilarity-based analysis of ecological data. *Journal of Statistical Software* **22**: 1-19.

Hamonts, K., Trivedi, P., Garg, A., Janitz, C., Grinyer, J., Holford, P. et al. (2018) Field study reveals core plant microbiota and relative importance of their drivers. *Environ Microbiol* **20**: 124-140.

Hanson, C.A., Fuhrman, J.A., Horner-Devine, M.C., and Martiny, J.B. (2012) Beyond biogeographic patterns: Processes shaping the microbial landscape. *Nat Rev Microbiol* **10**: 497-506.

Heymann, S. (2014) Gephi encyclopedia of social network analysis and mining. *Springer New York*.

Hijmans, R.J., E., C.S., and ., P.J.L. (2005) Very high resolution interpolated climate surfaces for global land areas. *International Journal of Climatology* **25**: 1965-1978.

Hinsinger, P. (2001) Bioavailability of soil inorganic P in the rhizosphere as affected by root-induced chemical changes: a review. *Plant & Soil* **237**: 173-195.

Hu, J., Wei, Z., Kowalchuk, G.A., Xu, Y., Shen, Q., and Jousset, A. (2020) Rhizosphere microbiome functional diversity and pathogen invasion resistance build up during plant development. *Environ Microbiol* **22**: 1011-1024.

Hubbell, S.P. (2001) Unified neutral theory of biodiversity and biogeography. *Scholarpedia*

Hubbell, S.P. (2005) Neutral theory in community ecology and the hypothesis of functional equivalence. *Functional Ecology* **19**: 166-172.

Ikeda, S., Okubo, T., Kaneko, T., Inaba, S., Maekawa, T., Eda, S. et al. (2010) Community shifts of soybean stem-associated bacteria responding to different nodulation phenotypes and N levels. *ISME J* **4**: 315-326.

Jiao, S., and Lu, Y. (2020) Soil pH and temperature regulate assembly processes of abundant and rare bacterial communities in agricultural ecosystems. *Environ Microbiol* **22**: 1052-1065.

Jiao, S., Yang, Y., Xu, Y., Zhang, J., and Lu, Y. (2020) Balance between community assembly processes mediates species coexistence in agricultural soil microbiomes across eastern China. *ISME J* **14**: 202-216.

Kembel, S.W., Cowan, P.D., Helmus, M.R., Cornwell, W.K., Helene, M., Ackerly, D.D. et al. (2010) Picante: R tools for integrating phylogenies and ecology. *Bioinformatics* **26**: 1463-1464.

Kolde, R. (2015) Pheatmap: pretty heatmaps. *R package version* 061.

Kumar, M., Brader, G., Sessitsch, A., Maki, A., van Elsas, J.D., and Nissinen, R. (2017) Plants assemble species specific bacterial communities from common core taxa in three arcto-alpine climate zones. *Front Microbiol* **8**: 12.

Tedersoo L, Bahram M, Põlme S, et al. (2014) Global diversity and geography of soil fungi. *Science* **346**.

Lemanceau, P., Blouin, M., Muller, D., et al. (2017) Let the core microbiota be functional. *Trends in Plant Science* **22**: 583-595

Levins, R. (1970) Ordering the phenomena of ecology. *Science* **167**: 1478-1480.

Lin, Y.B., Guo, Y.Q., Di, X.H., Fan, M.C., Dong, D.H., and Wei, G.H. (2015) *Saccharothrix stipae* sp. nov., an actinomycete isolated from the rhizosphere of *Stipa grandis*. *Int J Syst Evol Microbiol* **66**: 1017-1021.

Lindström ES, L.S. (2012) Local and regional factors influencing bacterial community assembly. *Environ Microbiol Reports* **4**: 1-9.

Liu, J., Sui, Y., Yu, Z., Shi, Y., Chu, H., Jin, J. et al. (2014) High throughput sequencing analysis of biogeographical distribution of bacterial communities in the black soils of northeast China. *Soil Biology & Biochemistry* **70**: 113-122.

Liu, L., Yang, J., Yu, Z., and Wilkinson, D.M. (2015) The biogeography of abundant and rare bacterioplankton in the lakes and reservoirs of China. *ISME J* **9**: 2068-2077.

Liu, Y., Zhang, L., Lu, J., Chen, W., Wei, G., and Lin, Y. (2020) Topography affects the soil conditions and bacterial communities along a restoration gradient on Loess-Plateau. *Applied Soil Ecology* **150**: 103471.

Lopez, S., van der Ent, A., Sumail, S., Sugau, J.B., Buang, M.M., Amin, Z. et al. (2020) Bacterial community diversity in the rhizosphere of nickel hyperaccumulator plant species from Borneo Island (Malaysia). *Environ Microbiol* **22**: 1649-1665.

Lozupone, C.A., and Knight, R. (2007) Global patterns in bacterial diversity. *Proc Natl Acad Sci U S A* **104**: 11436-11440.

Martin, M. (2011) Cutadapt removes adapter sequences from high-throughput sequencing reads. *Embnet Journal* **17**: 10-12.

Martiny, J.B.H., Bohannan, B.J.M., Brown, J.H., Colwell, R.K., Fuhrman, J.A., Green, J.L. et al. (2006) Microbial biogeography: putting microorganisms on the map. *Nature Reviews Microbiology* **4**: 102-112.

Meeks, J.C., and Elhai, J. (2002) Regulation of cellular differentiation in filamentous *Cyanobacteria* in free-living and plant-associated symbiotic growth states. *Microbiology & Molecular Biology Reviews* **66**: 94-121.

Mehrotra, M., Duose, D.Y., Singh, R.R., Barkoh, B.A., and Luthra, R. (2017) Versatile ion S5XL sequencer for targeted next generation sequencing of solid tumors in a clinical laboratory. *Plos One* **12**: e0181968.

Mendes, L., W., Kuramae, E., E. et al. (2014) Taxonomical and functional microbial community selection in soybean rhizosphere. *ISME J* **8**: 1577-1587.

Nemergut, D.R., Costello, E.K., Hamady, M., Lozupone, C., Jiang, L., Schmidt, S.K. et al. (2011) Global patterns in the biogeography of bacterial taxa. *Environ Microbiol* **13**: 135-144.

Nemergut, D.R., Schmidt, S.K., Fukami, T., O'Neill, S.P., Bilinski, T.M., Stanish, L.F. et al. (2013) Patterns and processes of microbial community assembly. *Microbiol Mol Biol Rev* **77**: 342-356.

Paradis, E., Claude, J., and Strimmer, K. (2004) Ape: analyses of phylogenetics and evolution in R language. *Bioinformatics* **20**: 289-290.

Rényi, P.E.a.A. (2012) On the evolution of random graphs. *Publ Res I Math Sci* **38**: 17-61.

Rousk, J., Baath, E., Brookes, P.C., Lauber, C.L., Lozupone, C., Caporaso, J.G. et al. (2010) Soil bacterial and fungal communities across a pH gradient in an arable soil. *ISME J* **4**: 1340-1351.

Shen, C., Shi, Y., Fan, K., He, J.S., Adams, J.M., Ge, Y., and Chu, H. (2019) Soil pH dominates elevational diversity pattern for bacteria in high elevation alkaline soils on the Tibetan Plateau. *FEMS Microbiol Ecol* **95:** fiz003.

Shi, Y., Li, Y., Xiang, X., Sun, R., Yang, T., He, D. 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.

Shuo, Jiao, Zhenshan, Liu, Yanbing, Lin et al. (2016) Bacterial communities in oil contaminated soils: Biogeography and co-occurrence patterns. *Soil Biology & Biochemistry* **98**: 64-73.

Shuo, Jiao, Yiqin, Xu, Jie, Zhang et al. (2019) Core microbiota in agricultural soils and their potential associations with nutrient cycling. *mSystems* **4**:e00313–18.

Stegen, J.C., Xueju, L., and Konopka A E , e.a. (2012) Stochastic and deterministic assembly processes in subsurface microbial communities. *ISME J* **6**: 1653-1664.

Stegen, J.C., Xueju, L., K., F.J., and E., K.A. (2015) Estimating and mapping ecological processes influencing microbial community assembly. *Frontiers in Microbiology* **6**: 370.

Tanai, and Cardona (2015) A fresh look at the evolution and diversification of photochemical reaction centers. *Photosynthesis Research* **126**: 111-134.

Torsten, Hothorn, Frank, Bretz, Peter, and Westfall (2008) Simultaneous inference in general parametric models. *Journal of Mathematical Methods in Biosciences* **50**: 346-363.

Svetnik V, Liaw A, Tong C, et al. (2003) Random forest : A classification and regression tool for compound classification and QSAR modeling. *J Chem Inf Comput Sci* **43**: 1947-1958.

Wang, K., Hu, H., Yan, H., Hou, D., Wang, Y., Dong, P., and Zhang, D. (2019) Archaeal biogeography and interactions with microbial community across complex subtropical coastal waters. *Mol Ecol* **28**:

3101-3118.

Wang, Q., Garrity, G.M., Tiedje, J.M., and Cole, J.R. (2007) Naive bayesian classifier for rapid assignment of rRNA sequences into the new bacterial taxonomy. *Applied & Environmental Microbiology* **73**: 5261-5267.

Wang, X.B., Lu, X.T., Yao, J., Wang, Z.W., Deng, Y., Cheng, W.X. et al. (2017) Habitat-specific patterns and drivers of bacterial beta-diversity in China's drylands. *ISME J* **11**: 1345-1358.

Xiao, X., Chen, W., Zong, L., Yang, J., Jiao, S., Lin, Y. et al. (2017) Two cultivated legume plants reveal the enrichment process of the microbiome in the rhizocompartments. *Mol Ecol* **26**: 1641-1651.

Xiong, J., Liu, Y., Lin, X., Zhang, H., Zeng, J., Hou, J. et al. (2012) Geographic distance and pH drive bacterial distribution in alkaline lake sediments across Tibetan Plateau. *Environ Microbiol* **14**: 2457-2466.

Yêyinou, L.L.E., Joelle, T., Arlette, A., Joel, A.A., Azize, O., and Alexandre, D. (2018) Folk taxonomy and traditional uses of common bean (*Phaseolus vulgaris* L.) landraces by the sociolinguistic groups in the central region of the Republic of Benin. *Journal of Ethnobiology & Ethnomedicine* **14**: 52.

Zhang, B., Zhang, J., Liu, Y., Guo, Y., Shi, P., and Wei, G. (2018a) Biogeography and ecological processes affecting root-associated bacterial communities in soybean fields across China. *Sci Total Environ* **627**: 20-27.

Zhang, J., Ding, Q., and Huang, J. (2016) Spaa: Species association analysis. *R package version 0.2* **1**.

Zhang, J., Zhang, B., Liu, Y., Guo, Y., Shi, P., and Wei, G. (2018b) Distinct large-scale biogeographic patterns of fungal communities in bulk soil and soybean rhizosphere in China. *Sci Total Environ* **644**: 791-800.

Zhang, K., Adams, J.M., Shi, Y., Yang, T., Sun, R., He, D. et al. (2017) Environment and geographic distance differ in relative importance for determining fungal community of rhizosphere and bulk soil. *Environ Microbiol* **19**: 3649-3659.

Zheng, Y., and Gong, X. (2019) Niche differentiation rather than biogeography shapes the diversity and composition of microbiome of *Cycas panzhihuaensis*. *Microbiome* **7**: 152.

Zhou, J., and Ning, D. (2017) Stochastic community assembly: Does it matter in microbial ecology? *Microbiology & Molecular Biology Reviews* **81**: e00002-00017.

Table 1 Summary of the locations, edaphic, and climatic characteristics across the five sampling sites in China.

Site	D	N	K	H	X
GPS coordinates					
Latitude (° N)	49.76	49.23	48.01	45.83	38.81
Longitude (° E)	124.62	125.34	125.83	126.85	111.65
Edaphic characteristics					
pH	5.11±0.12d	6.29±0.07c	7.16±0.11b	7.19±0.16b	8.37±0.34a
SWC（%）	0.51±0.03a	0.3±0.01c	0.37±0.01b	0.23±0.01d	0.12±0.02e
SOM（g/kg）	117.23±3.94a	53.81±0.77b	41.68±0.27c	29.63±1.92d	11.99±2.6e
TN (g/kg)	6.01±0.2a	2.56±0.21b	2.16±0.07c	1.38±0.04d	0.84±0.23e
TC (g/kg)	67.21±2.89a	29.73±1b	22.93±0.58c	13.35±0.43d	13.76±1.17d
AN（mg/kg）	418.99±7.8a	171.01±1.67b	149.49±11.46c	99.61±10.42d	57.96±16.63e
TP (g/kg)	1.41±0.18a	0.89±0.04b	0.63±0.02c	0.48±0.03d	0.62±0.01c
AP（mg/kg）	18.85±4.85ab	29.11±7.77a	32.03±13.54a	13.33±10.07b	7.91±1.88b
AK（mg/kg）	220.42±35.08c	333.98±23.61b	396.99±33.81a	194.63±3.8d	131.39±11.62e
Climatic characteristics					
MAT(°C)	-0.8	0	1.3	3.3	5.2
TS (°C)	1569.3	1569.6	1528.7	1468.1	1086.5
MT(°C)	25.8	26.1	26.4	27.7	25.8
MIT (°C)	-32.3	-30.6	-28.2	-26	-18.4
TR (°C)	58.1	56.7	54.6	53.7	44.2
MAP (mm)	493	495	493	568	486

PS (mm)	108	107	111	103	103

Notes: D: Dayangshu, Neimeng; N: Nenjiang, Heilongjiang; K: Keshan, Heilongjiang; H: Harbin, Heilongjiang; X: Xinzhou, Shanxi. The coordinates of each site were recorded with a Global Positioning System unit. pH, soil pH; SWC, soil water content; SOM, soil organic matter; TN, total nitrogen; TC, total carbon; AN, available nitrogen; TP, total phosphorus; AP, available phosphorus; AK, available potassium; MAT, Mean Annual Temperature; TS, Temperature Seasonality; MT, Max Temperature of Warmest Month; MIT, Min Temperature of Coldest Month; TR, Annual Temperature Range; MAP, Mean Annual Precipitation; PS, Precipitation Seasonality. Values for soil chemical properties within the same row followed by different letters indicate significant differences (P <0.05, ANOVA, Tukey's HSD test).

Table 2 Model parameters of distance-decay relationships of core community similarity (1-Bray-Curtis distance) among the five soil-plant associated compartments.

Compartment	BS			RS			R			S			L		
	Slope	R^2	*P*	Slope	R^2	*P*	Slope	R^2	*P*	Slope	R^2	*P*	Slope	R^2	*P*
Geographic	-0.020	0.573	**<0.001**	-0.009	0.075	**<0.001**	-0.006	0.040	**0.003**	-0.005	0.018	**0.036**	-0.004	0.016	**0.046**
Environmental	-0.058	0.656	**<0.001**	-0.051	0.389	**<0.001**	-0.048	0.395	**<0.001**	-0.017	0.036	**0.005**	-0.042	0.248	**<0.001**
Edaphic	-0.057	0.450	**<0.001**	-0.057	0.352	**<0.001**	-0.061	0.465	**<0.001**	-0.003	-0.005	0.724	-0.052	0.266	**<0.001**
Climatic	-0.061	0.497	**<0.001**	-0.039	0.158	**<0.001**	-0.029	0.096	**<0.001**	-0.037	0.130	**<0.001**	-0.036	0.121	**<0.001**

Notes: The parameters originated from the ordinary least square linear regression among the five different plant compartments. Bold *P*-values denotes that the slopes were significantly less than zero by permutation tests. BS, bulk soil; RS, rhizosphere soil; R, root endosphere; S, plant stem endosphere; L, plant leaf endosphere.

Table 3 The results of Spearman correlation tests between the core bacterial community dissimilarity (Bray–Curtis distances) and edaphic distance or climate distance for the different compartment samples using Mantel and partial Mantel tests.

Correlation between bacterial dissimilarity and:	BS		RS		R		S		L	
	r	*P*	r	*P*	r	*P*	r	*P*	r	*P*
Edaphic distance	0.664	**0.001**	0.591	**0.001**	0.711	**0.001**	-0.025	0.526	0.520	**0.001**
Climate distance	0.777	**0.001**	0.354	**0.003**	0.291	**0.01**	0.221	**0.038**	0.276	**0.009**
Controlling for:										
Climate distance	0.488	**0.002**	0.510	**0.002**	0.684	**0.001**	-0.166	0.9	0.459	**0.001**
Edaphic distance	0.678	**0.001**	0.071	0.213	-0.125	0.894	0.273	**0.031**	0.011	0.429

Notes: Edaphic and climate distances are the soil chemical properties and climate characteristics heterogeneity matrix calculated using Euclidean distance. Bold *P*-values indicate $P < 0.05$. *P*-values are one-tailed tests based on 999 permutations. For abbreviations see Table 2.

Table 4 Multiple stepwise regression analysis of the effect of environmental factors on bacterial communities (Beta-PCoA1) in each compartment across all sites.

BS (R^2=98.36 %)		RS (R^2=95.89 %)		R (R^2=91.62 %)		S (R^2=75.26 %)		L (R^2=84.66 %)	
Variable	Contribution (%)	Variable	Contribution (%)	Variable	Contribution (%)	Variable	Contribution (%)	Variable	Contribution (%)
MIT	19.31	TC	21.42	AN	40.17	MAT	38.85	TC	14.80
TR	18.77	SOM	18.98	MT	23.66	PS	17.06	TN	14.15
TS	18.04	MT	13.29	MIT	15.01	MAP	12.51	SOM	13.46
MAT	16.77	TS	10.07	MAT	14.54	MT	12.05	AN	13.40
pH	13.21	MIT	9.48	-	-	-	-	pH	11.43
SWC	12.78	pH	9.14	-	-	-	-	MAT	8.80
-	-	MAT	7.62	-	-	-	-	MIT	7.22
-	-	AK	7.61	-	-	-	-	MT	6.15
-	-	-	-	-	-	-	-	PS	2.51

Notes: Only significantly correlated variables were used in the analysis based on the Mantel test, and only significantly contributing variables are shown in the table

based on *P*-values ($P < 0.01$). pH, soil pH; SWC, soil water content; SOM, soil organic matter; TN, total nitrogen; TC, total carbon; AN, available nitrogen; TP, total phosphorus; AK, available potassium; MAT, Mean Annual Temperature; TS, Temperature Seasonality; MT, Max Temperature of Warmest Month; MIT, Min Temperature of Coldest Month; TR, Annual Temperature Range; MAP, Mean Annual Precipitation; PS, Precipitation Seasonality. For other abbreviations see Table 2.

Table 5 Topological features of networks in the different soil-plant associated core bacterial communities across all sites based on Spearman correlation method.

		BS	RS	R	S	L
Empirical networks	Nodes	380	464	379	273	297
	Edges	2518	2453	1264	581	892
	Edges (+)	1961	2245	1188	581	889
	Edges (-)	557	208	76	0	3
	Edges (+)/Total edges	77.88%	91.52%	93.99%	100%	99.66%
	Diameter	15	11	13	19	15
	Density	0.035	0.023	0.018	0.016	0.02
	Average degree (avgK)	13.253	10.573	6.67	4.256	6.007
	Average path distance (GD)	4.347	4.243	4.874	6.785	5.764
	Average clustering coefficient (avgCC)	0.501	0.431	0.436	0.347	0.442
	Modularity	0.430	0.590	0.622	0.748	0.679
Random networks (n=1,000)	Average path distance (GD)	2.583±0.002	2.848±0.003	3.335±0.009	3.989±0.033	3.366±0.014
	Average clustering coefficient (avgCC)	0.036±0.002	0.023±0.002	0.017±0.002	0.015±0.003	0.022±0.004

Notes: n=1,000, numbers of Erdös-Réyni random networks. For abbreviations see Table 2.

Figure legends

Fig. 1. Distribution patterns from aspects of alpha and beta diversity of core bacterial communities among the five soil-plant associated compartments across the five sampling sites. BS, bulk soil; RS, rhizosphere soil; R, root endosphere; S, plant stem endosphere; L, plant leaf endosphere. D: Dayangshu, Neimeng; N: Nenjiang, Heilongjiang; K: Keshan, Heilongjiang; H: Harbin, Heilongjiang; X: Xinzhou, Shanxi. **(a)** Venn diagrams of the core taxa shared among the five compartments across all sites and their relative abundance in the dominant phyla. **(b)** Boxplots of the alpha-diversity (Shannon and ACE index) of core bacterial communities among the five compartments. Different lowercase letters above indicate significant differences (P <0.05, ANOVA, Tukey's HSD test). **(c)** Principal coordinate analysis (PCoA) of the core bacterial community composition across all compartments and sites based on Bray-Curtis distances. **(d)** Principal coordinate analysis (PCoA) of core bacterial community composition across all sites in each compartment based on Bray-Curtis distances

Fig. 2. Distance-decay relationships between the core community similarity (1-Bray-Curtis distance) and the geographic, environmental, edaphic, and climatic distance were elucidated among the five soil-plant associated compartments across the five sampling sites (n=190, C_{20}^2). The environmental, edaphic, and climatic distances were calculated as Euclidean distances. The shaded region represents the 95% confidence limits of the regression estimates. BS, bulk soil; RS, rhizosphere soil; R, root endosphere; S, plant stem endosphere; L, plant leaf endosphere.

Fig. 3. Sloan neutral modeling of core bacterial communities among the five soil-plant associated compartments across five sampling sites. BS, bulk soil; RS, rhizosphere soil; R, root endosphere; S, plant stem endosphere; L, plant leaf endosphere. **(a)** Operational taxonomic units (OTUs) that occur more frequently than the value predicted by the model are shown in orange, while those that occur less frequently than predicted are shown in purple. Gray points represent the frequency of occurrence within the 95% confidence interval (blue dotted lines) ranging around the model prediction (red line). Values of m and Nm indicate the estimates of dispersal rate between communities and the metacommunity size times immigration, respectively; R^2 indicates the fit to this model**. (b)** Relative abundances of the core bacterial communities in neutrally distributed, over-represented, and under-represented OTUs among the five compartments.

Fig. 4. Null model of core bacterial communities among five soil-plant associated compartments across five sampling sites. BS, bulk soil; RS, rhizosphere soil; R, root endosphere; S, plant stem endosphere; L, plant leaf endosphere. **(a)** Distribution of the beta nearest-taxon index (βNTI) according to the spatial distance among the five compartments. Positive (negative) βNTI values indicate greater (less) than expected turnover in the phylogenetic composition. The horizontal dotted red line (above -2 or below +2 are statistically significant) shows the 95% confidence intervals around the expectation under neutral community assembly (above -2 or below +2 are

statistically significant). The explained percent of the stochastic processes is shown in red numbers. The gray line and shaded region represent the 95% confidence limits of the regression estimates. **(b)** The average relatively explained degree of core community assembly processes based on the values of βNTI and $RC_{Bray.}$ **(c)** Boxplots of mean habitat niche breadths (*B*com) among the five compartments. Different lowercase letters above the boxes indicate significant differences (P <0.05, ANOVA, Tukey's HSD test).

Fig. 5. Coexistence networks of the core bacterial communities based on correlation analysis among the five soil-plant associated compartments across the five sampling sites. Node colors represent the distributed types resulting from the Sloan neutral model; the size of each node is proportional to the degree. A connection stands for a robust (Spearman's r > |0.8|) and significant (P <0.01) correlation; the thickness of each connection between two nodes (edge) is proportional to the value of the Spearman's correlation coefficients. Positive and negative relationships are illustrated in pink and green, respectively. The occurrence rates of neutrally distributed nodes in the corresponding networks is illustrated by pink numbers in brackets. BS, bulk soil; RS, rhizosphere soil; R, root endosphere; S, plant stem endosphere; L, plant leaf endosphere.

Summary of Supporting information

Table S1 Two-way ANOVA for alpha diversity (Shannon and ACE indices) of the overall and core bacterial communities among the five compartments across all sampling sites.

Table S2 Three statistical analyses of the core bacterial community composition between compartment and site factors contributing to the whole and split variation based on Bray-Curtis distances.

Table S3 Three statistical analyses of the overall bacterial community composition between compartment and site factors contributing to the whole and split variation based on Bray-Curtis distances.

Table S4 Spearman correlations between the core bacterial community dissimilarity (Bray–Curtis distances) and geographic distance or environmental distances for the different compartment samples using Mantel and partial Mantel tests.

Table S5 Spearman Mantel correlations (r) and significance (P) between the core community and environmental factors across five different soil-plant associated compartments.

Table S6 Random forest mean predictor importance of the environmental factors for bacterial communities (Beta-PCoA1) in each compartment across all sites.

Table S7 Relatively explained degree of core community assembly processes based on the values of βNTI and $RC_{Bray.}$

Table S8 Topological features of networks in the different soil-plant associated core bacterial communities across all sites based on SPIEC-EASI method.

Table S9 Topological features of networks in the different soil-plant associated core bacterial communities across all sites based on SparCC method.

Fig. S1 Distribution of five sampling sites across three provinces in China. D: Dayangshu, Neimeng; N: Nenjiang, Heilongjiang; K: Keshan, Heilongjiang; H: Harbin, Heilongjiang; X: Xinzhou, Shanxi.

Fig. S2 (a) Heatmaps showing the relative abundance of the dominant phyla of the core bacterial communities among the five soil-plant associated compartments, and the five sampling sites in each compartment. **(b)** Correlations between core bacterial communities with the remaining communities across all sites. **(c)** Boxplots of the alpha-diversity (Shannon and ACE index) of the overall bacterial communities among the five compartments. Different lowercase letters above the boxes indicate significant differences (P <0.05, ANOVA, Tukey's HSD test). **(d)** Principal coordinate analysis (PCoA) of the overall bacterial community composition across all compartments and sites based on Bray-Curtis distances. **(e)** PCoA of the overall bacterial community composition across all sites in each compartment based on Bray-Curtis distances. BS, bulk soil; RS, rhizosphere soil; R, root endosphere; S, plant stem endosphere; L, plant leaf endosphere. D: Dayangshu, Neimeng; N: Nenjiang, Heilongjiang; K: Keshan, Heilongjiang; H: Harbin, Heilongjiang; X: Xinzhou, Shanxi.

Fig. S3 Heatmaps showing the relative abundance of the dominant (top ten) core bacterial communities among the five soil-plant associated compartments at class, order and family levels. For abbreviations see Fig. S2.

Fig. S4 Boxplots of the alpha-diversity (Shannon and ACE indices) of the overall and core bacterial communities among the five sampling sites in each soil-plant associated compartment. For abbreviations see Fig. S2.

Fig. S5 Topological features of nodes in the different soil-plant associated core bacterial communities across all sampling sites based on Spearman correlation method. For abbreviations see Fig. S2.

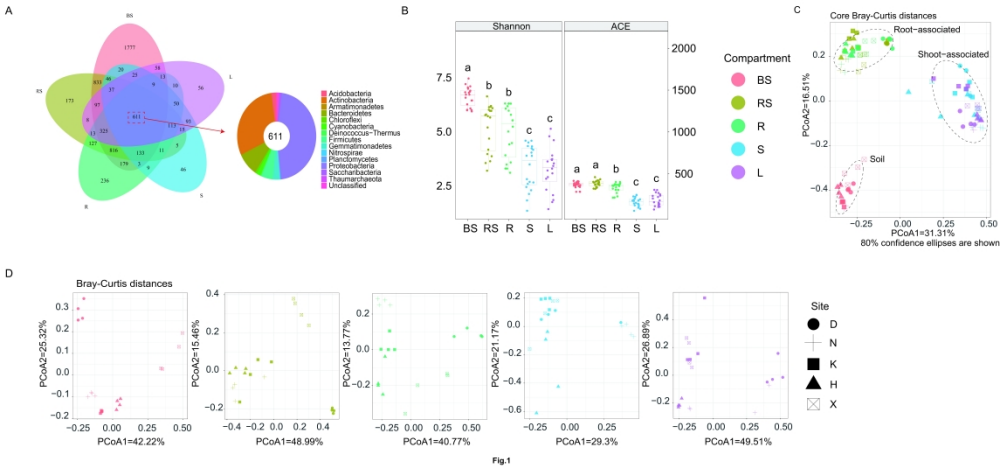

Fig.1

Figure 1-re-submitted version

1345x903mm (600 x 600 DPI)

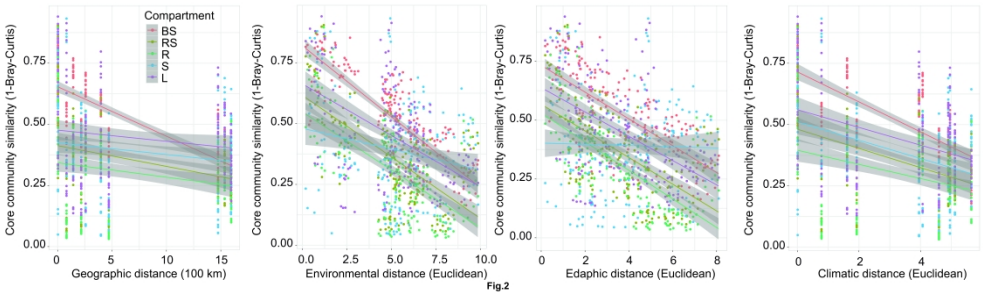

Figure 2-re-submitted version

1099x328mm (600 x 600 DPI)

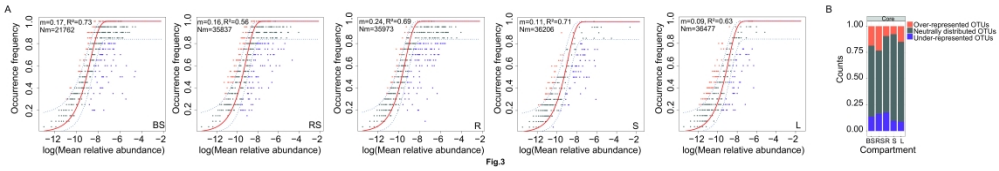

Figure 3-re-submitted version

1300x213mm (600 x 600 DPI)

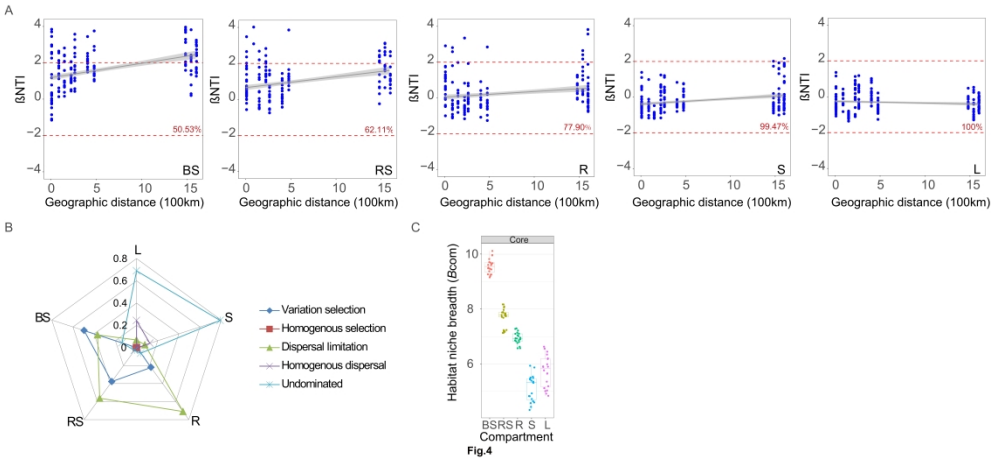

Figure 4-re-submitted version

988x451mm (600 x 600 DPI)

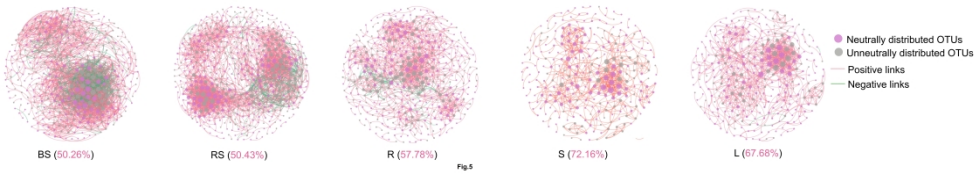

Figure 5-re-submitted version

1109x201mm (300 x 300 DPI)