Title: Diversity and biogeography of plant leaf bacteria along a latitudinal gradient

Authors: Zihui Wang[1,2], Yuan Jiang[4], Minhua Zhang[2], Chengjin Chu[4], Yongfa Chen[4], Shuai Fang[10], Guangze Jin[5], Mingxi Jiang[6], Juyu Lian[7], Yanpeng Li[11], Yu Liu[2], Keping Ma[8], Xiangcheng Mi[8], Xiujuan Qiao[6], Xihua Wang[9], Xugao Wang[10], Han Xu[11], Wanhui Ye[7], Li Zhu[8], Yan Zhu[8], Fangliang He[2, 3*] and Steven W. Kembel[1,2*]

Affiliations:

[1.] Département des sciences biologiques, Université du Québec à Montréal, Montréal, Québec, H2X 1Y4, Canada

[2.] ECNU-Alberta Joint Lab for Biodiversity Study, Tiantong Forest Ecosystem National Observation and Research Station, School of Ecology and Environmental Sciences, East China Normal University, Shanghai, 200241, China

[3.] Department of Renewable Resources, University of Alberta, Edmonton, Alberta, T6G 2H1, Canada

[4.] State Key Laboratory of Biocontrol, School of Life Sciences / School of Ecology, Sun Yat-sen University, Guangzhou, 510275, China

[5.] Center for Ecological Research, Northeast Forestry University, Harbin, 150040, China

[6.] Key Laboratory of Aquatic Botany and Watershed Ecology, Wuhan Botanical Garden, Chinese Academy of Sciences, Wuhan, 430074, China

[7.] Key Laboratory of Vegetation Restoration and Management of Degraded Ecosystems, South China Botanical Garden, Chinese Academy of Sciences, Guangzhou, 510650, China

[8.] State Key Laboratory of Vegetation and Environmental Change, Institute of Botany, Chinese Academy of Sciences, Beijing, 100093, China

[9.] Zhejiang Tiantong Forest Ecosystem National Observation and Research Station, School of Ecology and Environmental Sciences, East China Normal University, Shanghai, 200241, China

[10.] CAS Key Laboratory of Forest Ecology and Management, Institute of Applied Ecology, Chinese Academy of Sciences, Shenyang, 110016, China

[11.] Research Institute of Tropical Forestry, Chinese Academy of Forestry, Guangzhou, 510520, China

*Corresponding author. Email: fhe@ualberta.ca (FH) & kembel.steven_w@uqam.ca (SWK)

Abstract:

Plant-associated microbes influence plant ecology, evolution, and ecosystem function, while explaining the diversity and abundance of these microbes at broad geographic scales remains problematic. Here we quantified the contribution of biotic and abiotic factors to the diversity and biogeography of leaf-associated bacteria on woody plants across a latitudinal gradient. We show that host plant attributes, environments, space and neighborhood effects jointly explained the majority of variation in leaf bacterial diversity. With increasing latitude, the importance of host plant attributes for bacterial community assembly increased while the host specificity of bacteria decreased. Broad-scale plant-bacteria association networks were

interconnected by bacterial hub taxa whose abundance was controlled by climate, host, and spatial factors. These findings imply a latitude-dependent response of the coupled host-microbial system to global change.

One-Sentence Summary:

Latitude-dependent mechanisms structure the diversity and biogeography of plant leaf-associated bacteria.

Main Text:

Plant microbiomes are essential for plant growth, health and stress resilience, and play key roles in mediating biogeographical patterns and biogeochemical processes (*1*). The ecology and evolution of plant microbiomes and their roles in ecosystem function are being impacted by global change with cascading effects on plant communities (*2*). Although global patterns of microbial diversity and biogeography have been quantified in systems such as soils, oceans and animal guts (*3–5*), we lack broad-scale data on the biogeography of plant-microbe associations, which limits our understanding of the drivers of plant-microbe interactions and ability to predict microbial responses to global change (*6*).

The aboveground surfaces of plants, termed the phyllosphere, is a key habitat for microorganisms (*7*), and bacteria in the phyllosphere are a major component of the plant microbiome that influence plant fitness and ecosystem functioning (*8*, *9*). The ecological and evolutionary processes that drive community assembly of phyllosphere bacteria include host plant attributes such as phylogeny and functional traits, environmental factors such as temperature, and the composition of neighboring plant communities (*10–12*). However, current knowledge of the drivers of plant-microbe associations is almost exclusively based on

laboratory studies or small-scale studies at one or a few sampling locations (*13*, *14*), which consequently prevents predicting large-scale patterns and understanding the varying strength of constraints across spatial gradients. Furthermore, drivers of phyllosphere diversity have been generally examined separately in previous studies, which limited the ability to understand the relative importance of different drivers and to disentangle the independent and confounding effects of factors.

Plant-associated bacteria exhibit differing degrees of host specialization in natural forests, with contrasting impacts on the dynamics of both plant and bacteria communities (*12*, *15*). Host-specialized bacteria including plant pathogens can be a major force maintaining plant diversity and structuring plant biogeography (*16–18*), while bacteria with broad host and geographic ranges connect the dynamics of local communities through dispersal and gene flow, acting as metacommunity keystone taxa that affect ecological and evolutionary dynamics at large scales (*19–21*). Identifying the drivers of host specialization and mapping the biogeography of microbial keystone taxa in the phyllosphere are essential for predicting the bacterial response to global change and the subsequential effects on plant communities, but our knowledge in this area is limited by a lack of large-scale standardized datasets.

Here, we addressed these knowledge gaps in a comprehensive study of the drivers of phyllosphere bacterial diversity and biogeography in a network of stem-mapped forest sites across China. We quantified bacterial communities using high-throughput sequencing of the 16S rRNA gene for a total of 1453 leaf samples from 329 plant species from 10 forest sites across major biomes in China. Specifically, we aimed to 1) disentangle the effects of host, environment, plant neighborhoods and spatial factors on the community assembly of phyllosphere bacteria; 2) determine the ecological and evolutionary drivers of host

specificity; and 3) identify metacommunity keystone taxa and the factors that control their biogeography.

Results

We identified 49254 16S rRNA gene amplicon sequence variants (ASV) from 1453 leaf samples, belonging to 237 bacterial orders and 386 families. *Proteobacteria* were the dominant phylum, accounting for 70% of the total abundance of sequences across all samples. *Beijerinckiaceae* (*Rhizobiales*) were most abundant in tropical and subtropical regions while *Sphingomonadaceae* (*Sphingomonadales*) were most abundant in temperate forests (Fig S1).

We quantified the variation in bacteria ASV richness and composition explained by multiple variables that have been proposed as drivers of host-associated microbial diversity using generalized linear models and redundancy analysis respectively (see Methods). The variables included host-related factors (phylogenetic eigenvectors + plant traits), abiotic environment (soil condition + topography + climate), diversity and abundance of neighborhood plants, and spatial eigenvectors. At the local scale of individual sites, these models explained 26%-76% of the variation in ASV richness and 19%-36% of the variation in community composition of phyllosphere communities. At regional levels, these models explained 27%-68% of the variation in species richness, and 33%-42% of the variation in community composition respectively. At a ‘global’ scale (defined as the scale including all study sites), the model explained 43% of the total variation in species richness and 47% of the variation in community composition across all sites (Fig. 1 & S2).

We used variation partitioning to disentangle the effects of these variables and their interactions on the bacterial diversities. Generally, the independent effect of host was the

most important in explaining both phyllosphere bacterial ASV richness and community composition, accounting for 15-89% of the explained variance, followed by abiotic environments and spatial eigenvectors (contributions of up to 29% and 25% respectively), while the attributes of plant neighborhoods were less important (only contributing 9% of explained variance). There was a strong interactive effect of all four groups of variables, which accounted for 8-72% of the explained variation in bacterial composition (Fig. 1). The relative importance of these factors varies across geographic scales and was highly latitude dependent; for example, the independent effect of host was positively correlated with latitude and showed less importance at large scale while the shared effect among these four variables decreased with latitude and was more pronounced at large scale, and the overall contribution of host-related factors increased while that of other factors decreased with latitude (Fig. 1 & S3).

Bacterial ASV richness was lowest at mid-latitudes and increased towards tropical and temperate regions (Fig. 2A). There was strong evidence that ASV richness was positively correlated with the diversity of neighborhood plants, negatively with slope and linked with elevation and annul temperature in quadratic relationships (Fig. S4). We use global non-metric multidimensional scaling ordinations to visualize the variation in bacterial community composition and identify clusters of samples related to sites and host taxonomy (Fig. 2B & S5). We then tested the drivers of this variation using permutational multivariate analysis of variance and showed strong evidence that climate, soil conditions, topography, neighborhood plant diversity and host traits all impact bacterial community assembly (Table S1). We detected a signal of spatial autocorrelation and host phylogenetic structure in the community composition of phyllosphere bacteria at both local and large scales (Fig S6).

We measured the host specialization of bacteria using the scaled distance-based specialization index (DSI*) for each ASV, where larger DSI* values indicate specialists and lower values indicate generalists (see Methods). Most phyllosphere bacterial ASVs were specialists, associated with more closely phylogenetically related host plants than expected by chance (Fig S7). The degree of host specialization decreased with increasing latitude (Fig. 2C), and was influenced by the annual mean temperature, soil conditions, host plant traits and density after accounting for effects of site and host taxa (Table S2). There was strong evidence for phylogenetic signal in DSI* (Pagels Lambda= 0.922, p= 0.001), indicating that closely related bacteria are similar in terms of their degree of specialization vs. generalism. We collected several bacterial functional traits from published data (see Methods) and showed strong evidence for traits as determinants of host specialization; for example, host-generalist bacterial taxa were linked to smaller genome and cell size and motility ability (phylogenetic generalized least squares; Fig. 3).

We showed very strong evidence for a negative relationship between mean DSI* and the local abundance of host plant species (Mixed-effect model, estimates: -0.005, $p<0.001$). The relationship was supported in most sites, indicating that locally dominant plant species had more host-generalized phyllosphere bacteria (Table S3). We also asked if this relationship held for community uniqueness, i.e. whether dominant plants hosted a more generalist bacterial community composition than rare plant species. By calculating the local contribution to beta diversity index and the mean Sørensen dissimilarity index, we quantified the ecological uniqueness of each sample. We found very strong evidence that the community uniqueness indexes were negatively correlated with host plant relative abundance, consistent with the hypothesis that less abundant plant species are associated with more unique bacterial communities (Table S3). After partitioning the Sørensen dissimilarity

index into turnover- and nestedness-induced components, we showed that the observed patterns were mainly driven by species turnover among samples but not nestedness, suggesting that locally rare plant species have a unique composition of phyllosphere bacteria due to the replacement of bacterial ASVs across samples rather than loss of bacteria species.

In order to identify keystone ASVs whose population dynamics can affect large-scale patterns of plant microbiomes, we constructed metacommunity-level networks of plant-bacterial associations at regional and global scales and quantified the topological positions of ASVs within each co-occurrence network by calculating the betweenness centrality of ASVs (Fig 4). We identified the hub ASVs within each network as keystone taxa (see Methods), which differed across regions; in subtropical and tropical regions, keystone taxa were primarily ASVs belonging to the *Rhizobiales* genera *1174-901-12*, *Methylocella* and *Methylobacterium*, while in temperate networks the keystone ASVs included *Sphingomonas* (*Sphingomonadales*) and *Hymenobacter* (*Cytophagales*) (Fig. S8).

The distribution of keystone bacterial ASVs across sites showed clear latitudinal patterns, with clusters of keystone ASVs associated with one or more regions (Fig 2D). In order to determine the relative importance of different factors structuring the biogeography of these keystone ASVs, we applied a joint species distribution modeling approach where nine habitat variables were taken as fixed effects and two distance matrices representing the spatial and host phylogenetic structure as random effects in a model to predict keystone ASV abundance (see Methods). The model explained up to 90% and 73% of the variation in bacterial ASV presence-absence and abundance, respectively (Fig 5&S9). Overall, climate (including temperature, precipitation and isothermality) was the most important factor explaining the distribution of phyllosphere bacterial keystone ASVs. There was also a major contribution of the random effects of spatial factors and host phylogeny on keystone ASV

abundance, highlighting the importance of taking spatial and host phylogenetic structure into account for unbiased inference and to improve the prediction of the distribution of host microbiomes. The abundance of keystone ASVs in the tropics was more influenced by the random effect of spatial factors but less affected by the fixed effect of host plant traits than that of temperate keystone ASVs (Fig 5). Additionally, we found a strong phylogenetic signal in the responses of ASVs to habitat covariance, indicating that phylogenetically close ASVs have similar habitat preferences (Fig S10).

Discussion

Here we quantified the community composition of phyllosphere bacteria on woody plants across multiple biomes along a latitudinal gradient and identified the principal drivers of the diversity and biogeography of phyllosphere bacteria. We have shown that plant host attributes are the primary driver of phyllosphere bacterial diversity and revealed a latitudinal pattern of community assembly processes in the plant microbiome. Most phyllosphere bacteria were host-specialized and host specialists were more prevalent at lower latitudes and on locally rare plant species. Metacommunity-level keystone bacteria play a pivotal role in structuring large-scale plant-bacterial association networks, and their abundance was controlled by climate, host attributes and space with varying importance of these factors across a broad latitudinal gradient. Taken together, our findings provide the most comprehensive understanding to date of the patterns and mechanisms of plant-associated bacterial diversity, and suggest avenues for predicting and managing plant microbiomes in natural ecosystems under global change (*2*).

Although host attributes, environment, climate, plant neighborhoods and spatial factors have been shown to influence the diversity of phyllosphere bacteria in separate

studies (*10–12*), the relative importance of these factors was previously unknown, which made it difficult to infer the mechanisms that explain biogeographic patterns of host-microbe associations (*6*). By simultaneously evaluating a broad range of factors across spatial scales, we have shown that host-related factors were consistently the most important drivers of phyllosphere bacterial diversity at different geographic scales, and that the effects of multiple community assembly drivers were highly confounded in structuring the bacteria community. These results are consistent with previous studies on rhizosphere microbes (*3*, *22*, *23*), confirming that the ecological filtering process exerted by plant hosts dominates during community assembly, and that different processes including environmental selection and dispersal interact in the assembly of the plant microbiome.

We observed a latitudinal gradient in the assembly processes structuring bacterial phyllosphere communities, with a progressive shift from strong host filtering and a lesser role for the other drivers of community assembly in temperate regions, to less host filtering and a stronger role for multiple interacting drivers of community assembly in the tropics. This result is consistent with the prediction of the niche-neutral continuum hypothesis that the importance of non-neutral processes such as host filtering should increase with latitude due to low diversity and harsh environmental conditions such as cold temperatures in temperate regions (*24–26*). The highly confounded effects of host, environment and spatial factors emphasized the importance of considering multiple factors potentially influencing community assembly in order to understand the actual underlying mechanisms of plant-microbial interactions, especially in diverse tropical regions and at large geographic scales.

The host specificity of phyllosphere bacteria varied along latitude with a higher prevalence of host specialization at lower latitudes. The degree of host specificity was related to the morphological and physiological traits of bacteria such as genome size and motility,

and was phylogenetically conserved, with some bacterial clades containing primarily either specialists or generalists. Although the mechanisms behind these relationships are unclear, they match the hypothesized ecological importance of these traits; for example, bacteria with smaller genomes are less likely to be able to afford the metabolic cost of being generalists (*27*), while motility could enable bacteria to select their preferred host microhabitats and exhibit higher host specificity (*27*, *28*). The phylogenetic conservatism and trait-dependence of bacterial host specificity provides a framework for predicting plant-microbe associations, for example by allowing identification of potential hosts and the potential for host shifts for ecologically or economically important bacteria such as pathogens or plant growth promoting bacteria (*29*, *30*).

Host-specialized plant microbiomes have been recognized to play key roles in structuring plant community composition and maintaining plant diversity (*1*, *17*, *31*, *32*). We found that the degree of host specialization of phyllosphere bacteria varies across plant species; less abundant plant species were associated with more host-specialized phyllosphere bacteria and had more compositionally unique bacterial communities. This finding suggests microbe-mediated mechanisms to explain how so many plant species can coexist in diverse ecosystems such as tropical forests. On the one hand, considering beneficial bacteria, plants with unique bacterial communities could be more likely to lack mutualist bacteria, leading to reduced host fitness and resulting in species rarity, so that the high prevalence of unique phyllosphere bacterial communities in the tropics may contribute to the numerous rare plant species compared to that in temperate regions (*1*, *32*). On the other hand, with harmful bacteria, associating with more host-specialized pathogens could enable frequency-dependent enemy pressures, giving rare plant species an advantage at low density to avoid stochastic extinction. Due to limited knowledge on the ecological function of most bacterial taxa, we

cannot explicitly determine the impact of each bacterial ASV on plant fitness and community assembly; however, the strong correlation between plant-associated bacterial communities and local plant abundance provided a solid basis for understanding the role of microbes in regulating plant diversity via density-dependent effects. Future experimental evidence is needed to confirm and understand the degree to which these associations impact plant abundance and community dynamics.

Plant-phyllosphere bacteria association networks at local sites were interconnected by keystone bacterial taxa to form regional-scale networks, and these regional networks were further linked together through keystone bacteria at global scales (Fig 4). These metacommunity-level keystone taxa are hypothesized to mediate the ecological and evolutionary processes that structure plant-bacteria associations at large scales (*19–21*). We identified clear clusters of bacterial keystone ASVs within each geographic scale, and the abundance of these ASVs was mainly controlled by climate, host attributes and spatial factors. Given the importance of keystone taxa in affecting large-scale patterns of community assembly, this suggests that global change could influence the biogeography of phyllosphere bacteria through its impacts on these keystone taxa that play central roles in network assembly of microbial co-occurrence (*20*, *21*).

Here we have shown the potential for feedbacks between leaf bacterial communities and plant communities, as evidenced by the links between phyllosphere bacterial communities and the local abundance and rarity of host plant species. These relationships imply that processes that that affect one partner in the plant-bacterial association can cause cascading effect on the other at the level of entire communities of coexisting plants and bacteria. Our results further highlight a potential latitude-dependent response of plant microbiomes to global change; plant-bacteria associations in temperate forests (in terms of

diversity, community composition and the abundance of keystone bacteria) should be more affected by changes in host plant identity and phylogenetic diversity caused by plant species range shifts or human activities, while associations in tropical forests will likely be sensitive to a variety of factors including climate warming and atmospheric nutrient deposition, and the higher prevalence of host-specialized bacteria in tropical forests could further increase the sensitivity of plant-microbe associations to global change.

Materials and Methods

Study sites

We conducted the study at 10 forest sites across China with latitude ranging from 18.7°N to 51.8°N (Table S4). The sites are part of a global network of stem-mapped forest plots of the Forest Global Earth Observatory (ForestGEO http://forestgeo.si.edu/). For each site, we selected approximately 1/3 of the tree species and at least 3 individuals per species. The choice of these species considered both shade tolerant and intolerant, early- and late-successional, locally common and rare species, in order to have a gradient in host attributes that may explain the variation in bacterial community assemblages. In total, 1453 trees from 329 plant species of 148 genera and 59 families were sampled.

Sampling and processing

For each tree individual, 50-100g leaves from the sub-canopy (2-10 m above ground) were collected. Leaves were placed in a sterile plastic bag and washed using 100 mL of a 1:50 dilution of Redford buffer (1M Tris, 500 mM EDTA, 1.2% CTAB) within 24 hours after sampling (*10*). The suspension was collected and centrifuged at 3300 g for 25 minutes and the pellet was added into an extraction tube to store in -20 °C until DNA extraction.

Microbial DNA was extracted following the manufacturer's protocol (QIAGEN Powersoil DNA extraction kit) and the V5-V6 region of bacteria 16S rRNA genes was amplified using primers 799F and 1115R to determine the composition of the phyllosphere bacterial community. PCR products were normalized using a SequalPrep Normalization kit (Thermo Fisher Scientific), pooled and then purified using AMPure to remove contaminants. The DNA library was prepared by mixing equimolar concentrations of DNA from each sample, and then sequenced using Illumina MiSeq Reagent Kit v3 for 300 bp paired-end sequencing.

Bioinformatics

We used DADA2 to identify the amplicon sequence variants (ASV) observed in each sample(*33*). The base quality scores (Q) were calculated, and the nucleotides at 19-240 bp and 16-240 bp positions were kept for forward and reverse DNA sequences respectively. Then we inferred error rates and inferred ASVs using DADA2 default parameters (*33*). Paired read ends were merged with a minimum overlap of 20. Non-target sequences and chimeras were removed. Taxonomy was assigned to each ASV by comparison with the SILVA SSU r138 database.

We revealed around 116 million high-quality 16S rRNA gene sequences and identified 55693 amplicon sequence variants (ASVs) from 1453 samples. Sequence abundance per sample ranged from 15485 to 522574, which could bias estimates of ASV diversity (Fig S11a). Thus, we randomly rarefied the ASV table to 15000 sequences per sample using the R package 'vegan', and 49254 ASVs were retained after rarefaction. This eliminated only a small proportion of extremely rare ASVs and had little impact on beta diversity (e.g. community dissimilarity; Fig S11b). Most ASVs (82%) were identified to the family level, 58% were classified to genus and only 4% were identified to the species level.

Metadata

We collected metadata describing host plant attributes, abiotic environments, neighborhood plants and spatial location. Host plant identity and functional traits of leaf samples were collected from the base data of sites. Plant phylogenetic tree was derived from a phylogeny database for vascular plants according to the species identity using the *V.phyloMaker* function in R (*34*). Metadata for environmental conditions included topographic and edaphic variables, which were derived from the base data of each site. Climate variables including annual mean temperature, annual precipitation, and precipitation seasonality were collected from WorldClim database based on the geographical coordinates of sites. We calculated the richness, abundance, mean phylogenetic distance and mean nearest neighbor distance of neighborhood plants present within 10 meters around the sample, and included the density and local relative abundance of host plant species in the metadata (Table S5).

We collect two sets of metadata on bacterial attributes, their phylogeny and functional traits, to investigate the potential determinants of bacterial host specialization. The phylogenetic tree was built based on the 16S rRNA gene information of ASVs, which were aligned using the *AlignSeqs* function in R package DECIPHER and then input to FastTree 2.0 to infer an approximately-maximum-likelihood phylogenetic tree using the GTRCAT model of evolution. Functional traits of bacteria were collected from a synthesis database, which compiled data from well-established data repositories (e.g. GenBank, Bergey's Manual of Systematics of Archaea and Bacteria and published datasets) and condensed to species level (*35*). Although functional traits of bacteria are generally thought to vary within species or even within strains, certain traits show strong phylogenetic conservatism and can be predicted by the taxonomy identity (i.e. genetic traits, gram positive/negative). We

selected 9 traits that can be reliably predicted by taxonomy, i.e. the mean trait values of family can predict at least 85% of the traits of species within this family (Table S6). We then assigned trait values to each ASV based on their taxonomy identity: ASVs that were assigned to species were assigned a species-level trait value, while ASVs that were not assigned to species but to genus were assigned the mean trait value of the genus, and ASVs with no genus assignment were assigned a family-level trait value. ASVs that were not assigned to family were excluded from trait assignment and subsequent analysis. In total 10453 out of 14774 ASVs were assigned at least one bacterial functional trait.

Statistical analysis

We used R (version 3.6) for all statistical analyses. The diversity of phyllosphere bacteria was modelled as a function of variables of host, environment, plant neighborhood and spatial factors. To prepare these variables, Moran's Eigenvector Maps (MEM) was used to decompose geographic distance matrices into a series of orthogonal eigenvectors representing the spatial relationships among samples, and principal coordinates of neighbor matrices (PCNM) was applied on the phylogenetic distance matrix among host plants to generate a series of variables that representing the position of host plant in a multi-dimensional phylogenetical space. Both phylogenetic and spatial eigenvectors can be used in regression to take phylogenetic/spatial autocorrelation into account (*36*). Environment, host traits and diversity of plant neighborhoods variables were rescaled using the formula $(x - x_{min})/(x_{max} - x_{min})$ to standardize the effects of the variables. Pairwise Pearson's correlation coefficients were calculated to test the correlation among all variables, and for predictors with correlation coefficient > 0.9, one predictor was randomly removed to reduce multicollinearity.

Species richness was calculated as the number of ASVs with abundance > 1, which follows a negative binomial distribution (NBD). To model the effects of variables on richness, we used generalized linear model with NBD as the error structure, implemented with function *glm.nb* in the R package MASS. We first built a candidate set of generalized linear models to represent the individual and joint linear contributions for types of mechanisms (host, environment, neighborhoods and spatial proximity) using the predictors for each mechanism, resulting in a total of 15 models. For each model, a stepwise selection procedure was used for model selection based on Akaike's information criterion corrected for small sample size (AICc). Non-significant variables ($P < 0.05$) were further excluded using the *update* function. Variance inflation factor (VIF) was calculated and variables with VIF > 5 were also removed to avoid multicollinearity. To quantify the individual effects of four types of variables and their interactions on bacterial richness, we used variation partitioning where the independent effect of each factor was evaluated as the difference in R^2 between models with and without this factor, and the interactive effect was calculated by subtracting the independent effect from overall effects. We used a subset of samples to evaluate the effect of host plant traits as the trait data are available for most but not all samples (Table S5). The effect of plant traits was highly corrected with that of plant phylogeny (Fig S12a). Thus, we only included plant phylogenetic eigenvectors as host variables in the variation partitioning analysis.

We repeated the same variance partition analysis on bacterial composition. To do that, the ASV table was Hellinger transformed to represent the community composition of bacteria. We used redundancy analysis (RDA) to model the effect of those variables on community composition. Variable selection was done for each model using forward selection implemented with the function *forward.sel* in R package packfor following the recommended

stopping rules, and variation partitioning was implemented by *varpart* function in the 'vegan' package (*37*). As for the analysis of ASV richness, we only used plant phylogenetic eigenvectors as host variables in the variation partitioning because most of the plant trait effects overlapped with the plant phylogeny (Fig S12b).

The analyses of richness and community composition were conducted at local scales for the 10 sites, at regional scales for tropical, subtropical and temperate forests, and at the global scale by combining all the sites. Specifically, Heishiding (HSD) and Dinghushan (DHS) sites (located at 23.27°N, 23.17°N respectively) were on the northern boundary of the tropics. Although classified as subtropical sites, the two forests are more similar in vegetation composition and climate conditions with the tropical forest Jianfengling (JFL, 18.73°N) than those typical subtropical forests (Table S6). We thus included HSD and DHS as 'tropical' sites, resulting in 3 tropical, 4 subtropical and 3 temperate sites. The relative importance of each factor and their interactive effects were evaluated as the proportion of explained variance, which was further related to latitude using a linear regression.

To estimate the host specificity in phyllosphere bacteria, we calculated the distance-based specialization index (DSI) for each bacterial ASV (*38*). To do that, we first calculated the abundance-weighted mean phylogenetic distance (MPD) between all host species on which a bacterial ASV occurred, and then compared the observed MPD to a null distribution obtained by randomly sampling hosts from the null model (n=999). The index was then scaled by the maximum or minimum value obtained for that bacterial ASV (*39*). The rescaled index, DSI*, ranges between -1 and 1, with -1 representing a full generalist and 1 a full specialist. This estimate considers plant phylogenetic relatedness and accounts for differences in abundance and sampling effort of bacteria so that it enables robust comparison

of host specificity among different taxa and sites/regions. To estimate community-level host specificity, we calculated the abundance averaged mean value of DSI* for each sample. We also calculated the local contribution to beta diversity (LCBD) to indicate the uniqueness of the community composition(*40*). Both mean DSI* and LCBD indicate community-level specialization but from different perspectives. To further illustrate the mechanisms of community uniqueness, we calculated total community dissimilarity Sørensen index (Sor) using presence-absence information of ASVs and partitioned the Sor to turnover- (Sim) and nestedness-induced community dissimilarity (Sne) using *beta.pair* function in betapart package. The averaged values of dissimilarity matrix were calculated for each sample to represent the ecological uniqueness of samples.

To investigate the evolutionary factors of host specificity. We modeled the DSI* as a function of bacteria traits using phylogenetic generalized least squares (PGLS), where the phylogenetic tree of bacteria was treated as a random term to control for the effect of phylogenetic trait conservatism. All variables were tested for normality and Box-Cox transformation was used for non-normal variables. The model was fitted using the generalized least squares (*gls*) function in 'nlme' package. Variable selection was conducted based on AIC in stepwise algorithm using *stepAIC* function in MASS package. To exam the ecological drivers of host specificity, we modeled the mean DSI* as a function of the ecological variables including annual mean temperature, soil conditions and host plant attributes using a mixed-effect model with site and host species as random terms. We further tested for relationships between host specialization and local plant abundance in each site by linking the host specialization indexes (mean DSI*, LCBD, Sor, Sim and Sne) to log-transformed relative abundance of host species in a linear regression.

To construct networks that represent plant-bacteria associations at different geographical scales, we first transformed ASV abundances to presence-absence with a threshold of 15 reads per sample (0.1% of the total read count of each sample) to be considered present, in order to minimize the effects of extreme rare species and sequencing errors on network construction. The occurrence of ASV in samples was then converted to occurrences on plant species according to the plant host identity of each sample, where a cell entry represented the number of leaf samples that were observed to have the ASV occurring on a plant species. In this step, the same plant species from different sites were represented in different rows as we were interested in how metacommunity bacteria interconnect local plant-bacteria networks across sites. The species-ASV table comprises 490 rows and 12065 columns, and the binary (presence/absence) data of this table were used to construct the network. We constructed 10 local networks representing plant-bacteria associations at site level, 3 regional networks for temperate, subtropical and tropical regions, and a global network using 'igraph' package in R. The regional and global networks were visualized using program GePhi with the 'ForceA-tlas2' layout algorithm, where plant species from different sites were highlighted using warm colors. For these networks, we identified hub ASVs in core positions of the networks by calculating the betweenness centrality score of each node/ASV. Betweenness centrality was standardized using the formula $(x - x_{min})/(x_{max} - x_{min})$ to make it range from 0 to 1, and ASVs with scaled betweenness above 0.5 were recognized as hubs(*19*). A total of 115 ASVs were identified as metacommunity-level hubs at regional and/or global scales.

The biogeography of the hub ASVs was first explored using a heatmap that mapped the frequency of ASVs (the proportion of samples in which the ASVs occurred) for each site. We then applied a joint species distribution modeling approach, called Hierarchical

Modelling of Species Composition (HMSC)(*41*), to predict the abundance of the 115 hub ASVs at the global scale. We selected a set of habitat variables that indicate important mechanisms potentially structuring bacterial community assembly, including 3 soil variables (total nitrogen, total phosphorus and water content), 3 climate variables (temperature, precipitation and isothermality), and 3 plant leaf traits (nitrogen, phosphorus and specific leaf area). As plant trait data were not available for all samples, 1242 out of 1453 samples with available plant trait data were used in this analysis. The 9 variables were treated as fixed terms to evaluate the effect of habitat conditions on ASVs, and two distance matrices representing the spatial and host phylogenetic relationships between samples were used as random terms to control for the effects of spatially and phylogenetically clustered data.

We applied a hurdle model, i.e., one model for presence-absence and another one for abundance conditional on presence, in order to quantify the drivers of both presence-absence and abundance. We applied probit regression in the presence-absence model, and linear regression for transformed count data in the abundance model. The count data were transformed by declaring zeros as missing data, log-transforming, and then scaling the data to zero mean and unit variance within each species(*41*). We fitted the HMSC model with the R-package Hmsc assuming the default prior distributions. We sampled the posterior distribution with five Markov Chain Monte Carlo (MCMC) chains, each of which was run for 37,500 iterations, of which the first 12,500 were removed as burn-in. We examined MCMC convergence by examining the potential scale reduction factors of the model parameters. The MCMC convergence of our model was satisfactory: the potential scale reduction factors for the β-parameters (that measures the responses of the species to environmental covariates) were on average 1.013 and 1.04 for abundance and presence-absence model respectively. To compare the species distribution patterns of ASV among regions, we calculated the relative

importance of each mechanism: the fixed effects of host traits, soil condition and climate, and the random effects of space and host phylogeny, which were compared among temperate, subtropical and tropical hubs using ANOVA and paired samples Wilcoxon tests.

References and Notes

1. C. S. Delavaux, P. Weigelt, W. Dawson, J. Duchicela, F. Essl, M. van Kleunen, C. König, J. Pergl, P. Pyšek, A. Stein, M. Winter, P. Schultz, H. Kreft, J. D. Bever, Mycorrhizal fungi influence global plant biogeography. *Nat. Ecol. Evol*. **3**, 424–429 (2019).

2. R. Cavicchioli, W. J. Ripple, K. N. Timmis, F. Azam, L. R. Bakken, M. Baylis, M. J. Behrenfeld, A. Boetius, P. W. Boyd, A. T. Classen, T. W. Crowther, R. Danovaro, C. M. Foreman, J. Huisman, D. A. Hutchins, J. K. Jansson, D. M. Karl, B. Koskella, D. B. M. Welch, J. B. H. Martiny, M. A. Moran, V. J. Orphan, D. S. Reay, J. V. Remais, V. I. Rich, B. K. Singh, L. Y. Stein, F. J. Stewart, M. B. Sullivan, M. J. H. van Oppen, S. C. Weaver, E. A. Webb, N. S. Webster, Scientists' warning to humanity: microorganisms and climate change. *Nat. Rev. Microbiol*. **17**, 569–586 (2019).

3. S. van der Linde, L. M. Suz, C. D. L. Orme, F. Cox, H. Andreae, E. Asi, B. Atkinson, S. Benham, C. Carroll, N. Cools, B. D. Vos, H.-P. Dietrich, J. Eichhorn, J. Gehrmann, T. Grebenc, H. S. Gweon, K. Hansen, F. Jacob, F. Kristöfel, P. Lech, M. Manninger, J. Martin, H. Meesenburg, P. Merilä, M. Nicolas, P. Pavlenda, P. Rautio, M. Schaub, H.-W. Schröck, W. Seidling, V. Šrámek, A. Thimonier, I. M. Thomsen, H. Titeux, E. Vanguelova, A. Verstraeten, L. Vesterdal, P. Waldner, S. Wijk, Y. Zhang, D. Žlindra, M.

I. Bidartondo, Environment and host as large-scale controls of ectomycorrhizal fungi. *Nature* **558**, 243–248 (2018).

4. T. Větrovský, P. Kohout, M. Kopecký, A. Machac, M. Man, B. D. Bahnmann, V. Brabcová, J. Choi, L. Meszárošová, Z. R. Human, C. Lepinay, S. Lladó, R. López-Mondéjar, T. Martinović, T. Mašínová, D. Morais, D. Navrátilová, I. Odriozola, M. Štursová, K. Švec, V. Tláskal, M. Urbanová, J. Wan, L. Žifčáková, A. Howe, J. Ladau, K. G. Peay, D. Storch, J. Wild, P. Baldrian, A meta-analysis of global fungal distribution reveals climate-driven patterns. *Nat. Commun*. **10**, 5142 (2019).

5. M. Bahram, F. Hildebrand, S. K. Forslund, J. L. Anderson, N. A. Soudzilovskaia, P. M. Bodegom, J. Bengtsson-Palme, S. Anslan, L. P. Coelho, H. Harend, J. Huerta-Cepas, M. H. Medema, M. R. Maltz, S. Mundra, P. A. Olsson, M. Pent, S. Põlme, S. Sunagawa, M. Ryberg, L. Tedersoo, P. Bork, Structure and function of the global topsoil microbiome. *Nature* **560**, 233–237 (2018).

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

7. B. Koskella, The phyllosphere. *Curr. Biol*. **30**, R1143–R1146 (2020).

8. I. Laforest-Lapointe, A. Paquette, C. Messier, S. W. Kembel, Leaf bacterial diversity mediates plant diversity and ecosystem function relationships. *Nature* **546**, 145–147 (2017).

9. T. Chen, K. Nomura, X. Wang, R. Sohrabi, J. Xu, L. Yao, B. C. Paasch, L. Ma, J. Kremer, Y. Cheng, L. Zhang, N. Wang, E. Wang, X.-F. Xin, S. Y. He, A plant genetic network for preventing dysbiosis in the phyllosphere. *Nature* **580**, 653–657 (2020).

10. S. W. Kembel, T. K. O'Connor, H. K. Arnold, S. P. Hubbell, S. J. Wright, J. L. Green, Relationships between phyllosphere bacterial communities and plant functional traits in a neotropical forest. *Proc. National. Acad. Sci.* **111**, 13715–13720 (2014).

11. M. W. Horton, N. Bodenhausen, K. Beilsmith, D. Meng, B. D. Muegge, S. Subramanian, M. M. Vetter, B. J. Vilhjálmsson, M. Nordborg, J. I. Gordon, J. Bergelson, Genome-wide association study of Arabidopsis thaliana leaf microbial community. *Nat. Commun*. **5**, 5320 (2014).

12. G. Lajoie, S. W. Kembel, Host neighborhood shapes bacterial community assembly and specialization on tree species across a latitudinal gradient. *Ecol. Monogr*. **91**, e01443 (2021).

13. J. M. Whipps, P. Hand, D. Pink, G. D. Bending, Phyllosphere microbiology with special reference to diversity and plant genotype. *J. Appl. Microbiol*. **105**, 1744–1755 (2008).

14. C. Vacher, A. Hampe, A. J. Porté, U. Sauer, S. Compant, C. E. Morris, The Phyllosphere: Microbial Jungle at the Plant–Climate Interface. *Annu. Rev. Ecol. Evol. Syst*. **47**, 1–24 (2016).

15. B. E. Sedio, A. M. Ostling, How specialised must natural enemies be to facilitate coexistence among plants. *Ecol. Lett*. **16**, 995–1003 (2013).

16. D. H. Janzen, Herbivores and the number of tree species in tropical forests. *Am. Nat*. **104**, 501–528 (1970).

17. M.-S. Benítez, M. H. Hersh, R. Vilgalys, J. S. Clark, Pathogen regulation of plant diversity via effective specialization. *Trends Ecol. Evol*. **28**, 705–711 (2013).

18. C. E. Mitchell, A. G. Power, Release of invasive plants from fungal and viral pathogens. *Nature* **421**, 625–627 (2003).

19. H. Toju, M. Yamamichi, P. R. Guimarães, J. M. Olesen, A. Mougi, T. Yoshida, J. N. Thompson, Species-rich networks and eco-evolutionary synthesis at the metacommunity level. *Nat. Ecol. Evol*. **1**, 0024 (2017).

20. M. A. Leibold, M. Holyoak, N. Mouquet, P. Amarasekare, J. M. Chase, M. F. Hoopes, R. D. Holt, J. B. Shurin, R. Law, D. Tilman, M. Loreau, A. Gonzalez, The metacommunity concept: a framework for multi-scale community ecology. *Ecol. Lett*. **7**, 601–613 (2004).

21. J. N. Thompson, *The Geographic Mosaic of Coevolution* (University of Chicago Press, Chicago, IL, 2005)

22. Z. Wang, Y. Jiang, D. C. Deane, F. He, W. Shu, Y. Liu, Effects of host phylogeny, habitat and spatial proximity on host specificity and diversity of pathogenic and mycorrhizal fungi in a subtropical forest. *New Phytol*. **223**, 462–474 (2019).

23. T. Yang, L. Tedersoo, P. S. Soltis, D. E. Soltis, J. A. Gilbert, M. Sun, Y. Shi, H. Wang, Y. Li, J. Zhang, Z. Chen, H. Lin, Y. Zhao, C. Fu, H. Chu, Phylogenetic imprint of woody plants on the soil mycobiome in natural mountain forests of eastern China. *Isme J*. **13**, 686–697 (2019).

24. H. Stephen, *The Unified Neutral Theory of Biodiversity and Biogeography* (Princeton University Press, Princeton, NJ, 2001).

25. X. Qiao, F. Jabot, Z. Tang, M. Jiang, J. Fang, A latitudinal gradient in tree community assembly processes evidenced in Chinese forests. *Global Ecol. Biogeogr*. **24**, 314–323 (2015).

26. X. Zhang, S. Liu, J. Wang, Y. Huang, Z. Freedman, S. Fu, K. Liu, H. Wang, X. Li, M. Yao, X. Liu, J. Schuler, Local community assembly mechanisms shape soil bacterial β diversity patterns along a latitudinal gradient. *Nat. Commun*. **11**, 5428 (2020).

27. L. G. Barrett, M. Heil, Unifying concepts and mechanisms in the specificity of plant-enemy interactions. *Trends Plant Sci*. **17**, 282–292 (2012).

28. P. R. Ehrlich, P. H. Raven, Butterflies and plants: a case study in coevolution. *Evolution* **18**, 586–608 (1964).

29. K. Wells, N. J. Clark, Host specificity in variable environments. *Trends Parasitol*. **35**, 452–465 (2019).

30. A. Corrales, S. A. Mangan, B. L. Turner, J. W. Dalling, An ectomycorrhizal nitrogen economy facilitates monodominance in a neotropical forest. *Ecol. Lett*. **19**, 383–392 (2016).

31. S. A. Mangan, S. A. Schnitzer, E. A. Herre, K. M. L. Mack, M. C. Valencia, E. I. Sanchez, J. D. Bever, Negative plant-soil feedback predicts tree-species relative abundance in a tropical forest. *Nature* **466**, 752–755 (2010).

32. E. Laliberté, H. Lambers, T. I. Burgess, S. J. Wright, Phosphorus limitation, soil-borne pathogens and the coexistence of plant species in hyperdiverse forests and shrublands. *New Phytol*. **206**, 507–521 (2015).

33. B. J. Callahan, P. J. McMurdie, M. J. Rosen, A. W. Han, A. J. A. Johnson, S. P. Holmes, DADA2: High-resolution sample inference from Illumina amplicon data. Nat. Methods **13**, 581–583 (2016).

34. Y. Jin, H. Qian, V.PhyloMaker: an R package that can generate very large phylogenies for vascular plants. Ecography **42**, 1353–1359 (2019).

35. J. S. Madin, D. A. Nielsen, M. Brbic, R. Corkrey, D. Danko, K. Edwards, M. K. M. Engqvist, N. Fierer, J. L. Geoghegan, M. Gillings, N. C. Kyrpides, E. Litchman, C. E. Mason, L. Moore, S. L. Nielsen, I. T. Paulsen, N. D. Price, T. B. K. Reddy, M. A. Richards, E. P. C. Rocha, T. M. Schmidt, H. Shaaban, M. Shukla, F. Supek, S. G. Tetu, S. Vieira-Silva, A. R. Wattam, D. A. Westfall, M. Westoby, A synthesis of bacterial and archaeal phenotypic trait data. Sci. Data **7**, 170 (2020).

36. S. Dray, P. Legendre, P. R. Peres-Neto, Spatial modelling: a comprehensive framework for principal coordinate analysis of neighbour matrices (PCNM). Ecol. Model **196**, 483–493 (2006).

37. P. R. Peres-Neto, P. Legendre, S. Dray, D. Borcard, Variation partitioning of species data matrices: estimation and comparison of fractions. Ecology **87**, 2614–2625 (2006).

38. L. R. Jorge, P. I. Prado, M. Almeida‐Neto, T. M. Lewinsohn, An integrated framework to improve the concept of resource specialisation. Ecol. Lett. **17**, 1341–1350 (2014).

39. L. R. Jorge, V. Novotny, S. T. Segar, G. D. Weiblen, S. E. Miller, Y. Basset, T. M. Lewinsohn, Phylogenetic trophic specialization: a robust comparison of herbivorous guilds. Oecologia **185**, 551–559 (2017).

40. P. Legendre, M. Cáceres, Beta diversity as the variance of community data: dissimilarity coefficients and partitioning. Ecol. Lett. **16**, 951–963 (2013).

41. O. Ovaskainen, G. Tikhonov, A. Norberg, F. G. Blanchet, L. Duan, D. Dunson, T. Roslin, N. Abrego, How to make more out of community data? A conceptual framework and its implementation as models and software. Ecol. Lett. **20**, 561–576 (2017).

Acknowledgments:

We thank the research crew from the forest stations of Jianfengling, Heishiding, Dinghushan, Badagongshan, Tiantongshan, Gutianshan, Baishanzu, Donglingshan, Lingshui and Changbaishan for their help during the field work and data collection. We also thank Dr. Otso Ovaskainen from the University of Helsinki and the NSERC CREATE Computational Biodiversity Science and Services (BIOS²) training program for their respective training on the use of some analysis tools.

Funding:

Natural Science and Engineering Research Council of Canada (FH, SWK)

Canada Research Chairs Program (SWK)

Author contributions:

Conceptualization: ZW, FH, SWK

Data collection: ZW, YJ, MZ, CC, YC, SF, GJ, MJ, JL, YL, YL, KM, XM, XQ, XW, XW, HX, WY, LZ, YZ

Analysis: ZW, YJ, SWK

Funding acquisition: FH, SWK

Writing – original draft: ZW, FH, SWK

Writing – review & editing: ZW, FH, SWK, YJ, MZ, CC, YC, SF, GJ, MJ, JL, YL, YL, KM, XM, XQ, XW, XW, HX, WY, LZ, YZ

Competing interests: Authors declare that they have no competing interests.

Data and materials availability: All data, code, and materials used in the analysis will be available once the manuscript being reviewed or accepted.

Supplementary Materials

Materials and Methods

Figs. S1 to S12

Tables S1 to S6

Fig. 1. Map of study sites indicating the percentage of variation in community composition explained by driving factors. The factors include host phylogeny, abiotic environment (soil + topography), spatial variables and the structure of plant neighborhoods at (**A**) local and (**B**) regional scales, with (**C**) demonstrating the latitude patterns of their relative contributions. The Venn diagrams show the absolute variance explained by each factor and the histograms on the right of Venn diagrams show the relative contribution of the pure and confounding effects of four variables to the explained variance. Explained variation less than 1% was not shown in the

Veen diagram. Abbreviation; H: host, E: abiotic environments; S: spatial variables; N: neighborhood plant communities.

Fig. 2. The biogeographic patterns of phyllosphere bacteria, in regarding to (**A**) alpha diversity, (**B**) community composition (**C**) host specialization and (**D**) occurrence of metacommunity-level hub ASVs versus latitude. First- and second order polynomial fits are shown in black (dashed) and red (solid) respectively in (**A**), Bacterial community composition was visualized using nonmetric multidimensional scaling ordination with each ellipse representing site in (**B**), where arrows inside indicate significant variables correlated with ordination scores and the explained power (R^2) of the correlation. Abundance-weighted average DSI* was correlated with latitude in (**C**). Heatmap in (**D**) show the relative frequency of hub ASVs clustering regarding to (1) temperate (2) subtropical (3) tropical hubs, (4) hubs that connect temperate and subtropical, (5) subtropical and tropical sites, and (6) global hubs that connect all sites.

Fig. 3. Effect size and 95% confidence interval of bacterial traits that determine their host specificity. Phylogenetic generalized least squares was applied to predict the DSI* with the bacteria traits as fixed terms and phylogenetic tree as random ones. Sample size = 6567.

Fig. 4. Global and regional networks representing the associations between plants and phyllosphere bacteria. Circles represent either bacteria ASVs (dark blue) or plant species (warm colors). A link was established if the plant species was associated with more than 15 reads of an ASV (see Methods). Size of the circles indicates the betweenness centrality of nodes. Plant

species were colored according to latitude of site, including tropical: JFL, DHS, HSD; subtropical: BSZ, GTS, BDGS, TT; and temperate sites: DLS, CBS, LS.

Fig. 5. The results of joint species distribution modeling. Showing (**A**) the absolute variation in bacterial abundance explained by each factor, the relative contribution of (**B**) host plant traits and (**C**) spatial factors on determining the abundance of ASVs in different regions. 115 ASVs were included in the model, and 9 habitat covariates and two distance matrices representing spatial and host phylogenetic relationship among samples were used as fixed and random term respectively. Each bar in (**A**) represents a bacterial ASV sorted according to the total variance explained, and the caption above the barplot shows the averaged variation explained by each factor. In in (**B**) and (**C**), Anova tests show significant difference in the relative importance of factors among regions (host factors: $F = 10.62$, $p < 0.001$; spatial factor: $F = 5.4$, $p = 0.005$). Numbers above the paired boxs symbolize the p values of pairwise Wilcox test.

Fig. 1

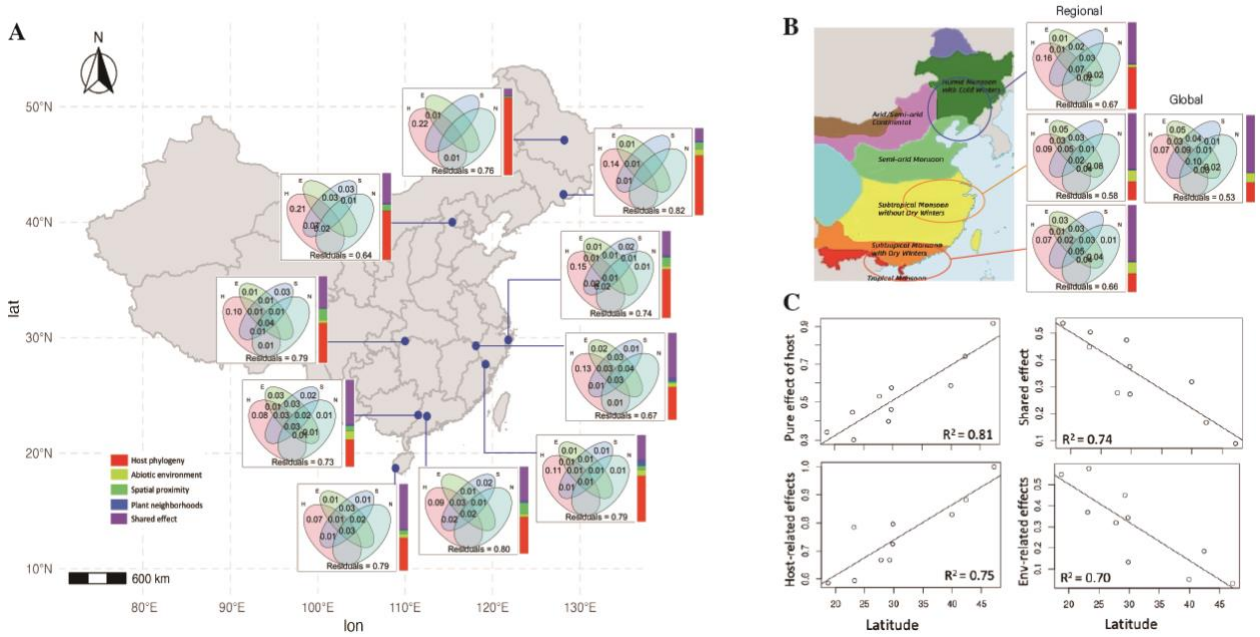

A
N
50°N
40°N
30°N
20°N
10°N
lat
80°E
90°E
100°E
110°E
120°E
130°E
lon
Host phylogeny
Abiotic environment
Spatial proximity
Plant neighborhoods
Shared effect
600 km
B
Regional
Global
C
Pure effect of host
Shared effect
Host-related effects
Env-related effects
R² = 0.81
R² = 0.74
R² = 0.75
R² = 0.70
Latitude
Latitude

Fig. 2

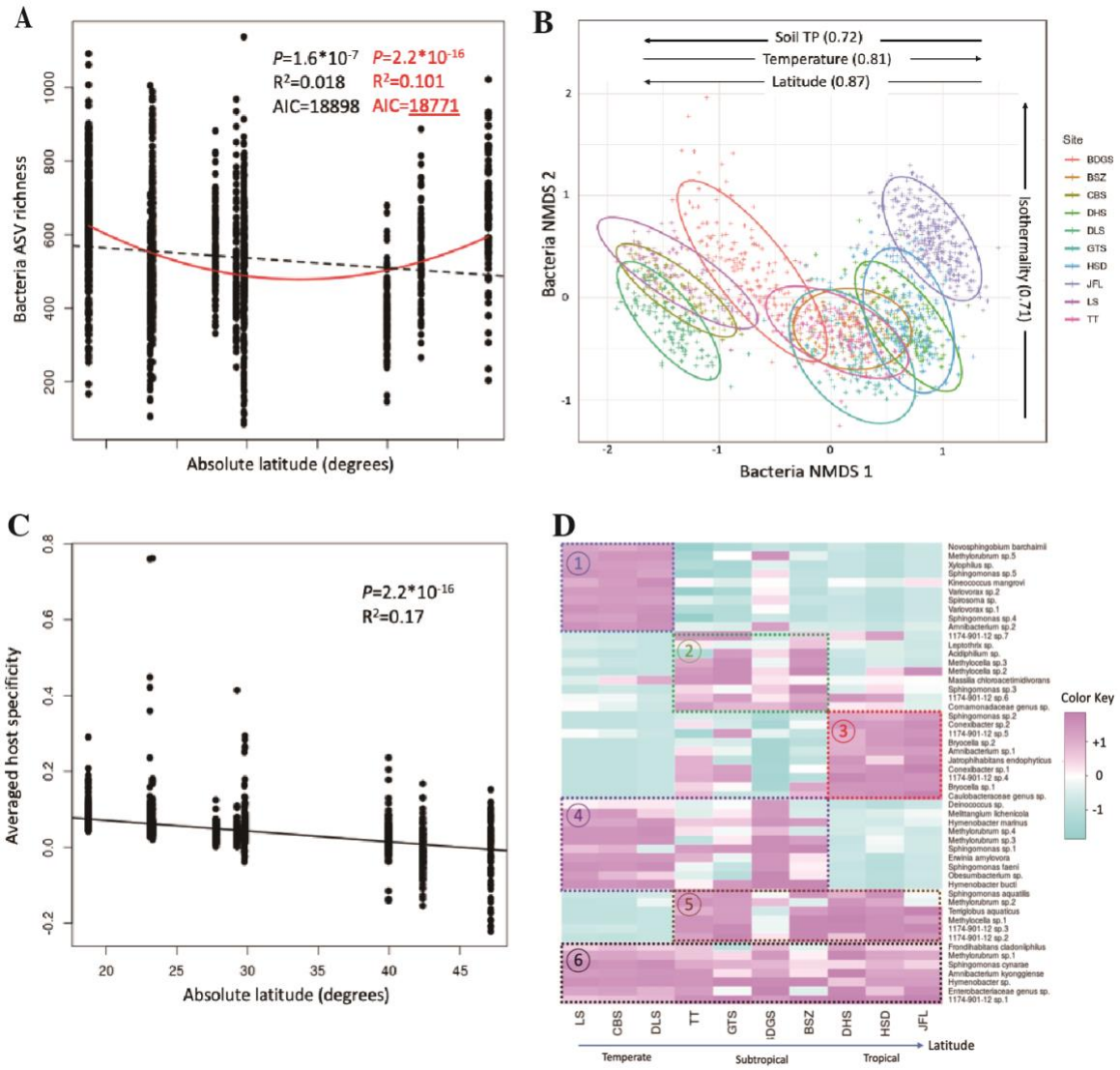

A
P=1.6*10-7
R2=0.018
AIC=18898
P=2.2*10-16
R2=0.101
AIC=18771
Bacteria ASV richness
Absolute latitude (degrees)
B
Soil TP (0.72)
Temperature (0.81)
Latitude (0.87)
Isothermality (0.71)
Bacteria NMDS 2
Bacteria NMDS 1
Site
BDGS
BSZ
CBS
DHS
DLS
GTS
HSD
JFL
LS
TT
C
P=2.2*10-16
R2=0.17
Averaged host specificity
Absolute latitude (degrees)
D
Color Key
+1
0
-1
LS
CBS
DLS
TT
GTS
IDGS
BSZ
DHS
HSD
JFL
Latitude
Temperate
Subtropical
Tropical

Fig. 3

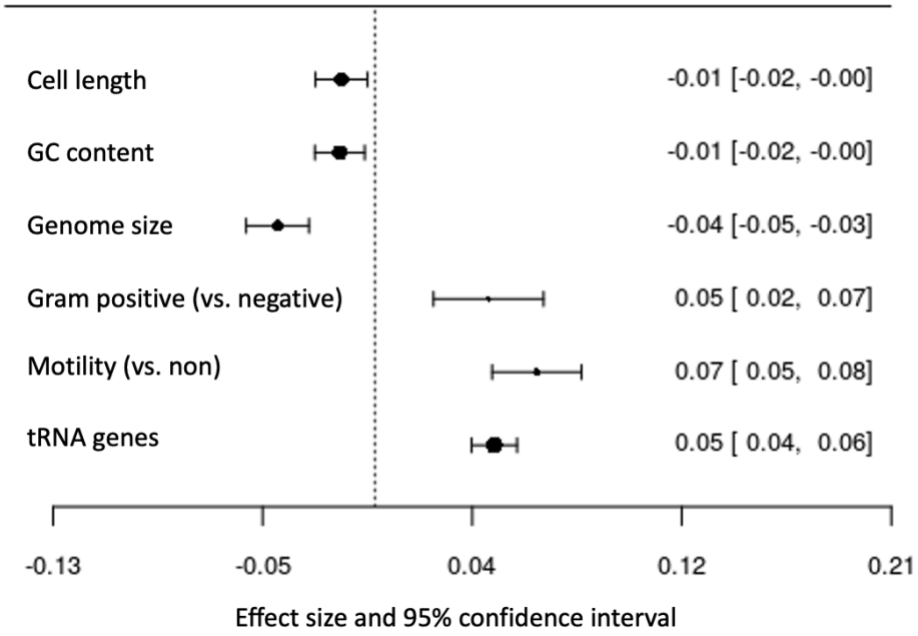

Cell length
GC content
Genome size
Gram positive (vs. negative)
Motility (vs. non)
tRNA genes
-0.01 [-0.02, -0.00]
-0.01 [-0.02, -0.00]
-0.04 [-0.05, -0.03]
0.05 [0.02, 0.07]
0.07 [0.05, 0.08]
0.05 [0.04, 0.06]
-0.13
-0.05
0.04
0.12
0.21
Effect size and 95% confidence interval

Fig. 4

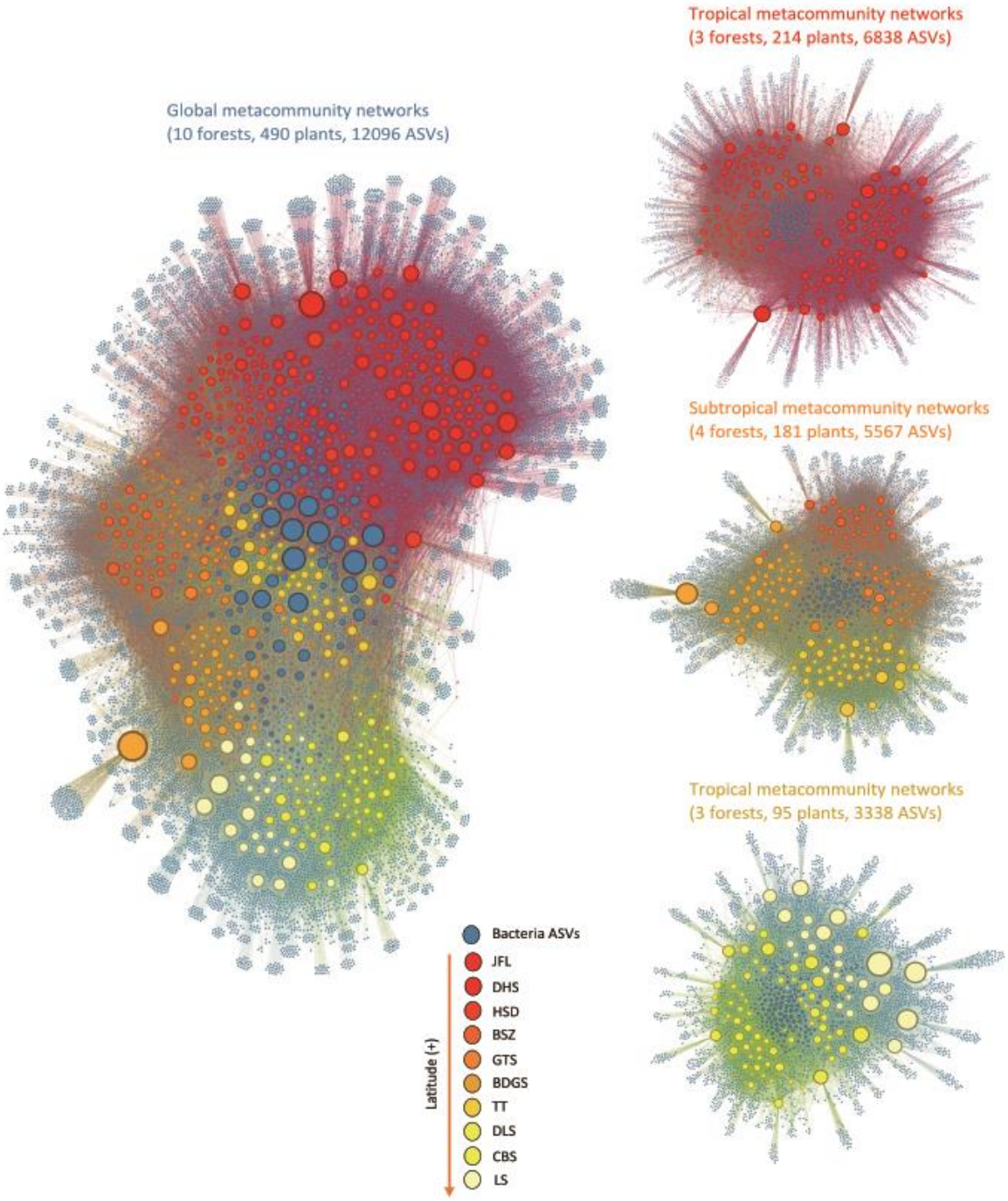
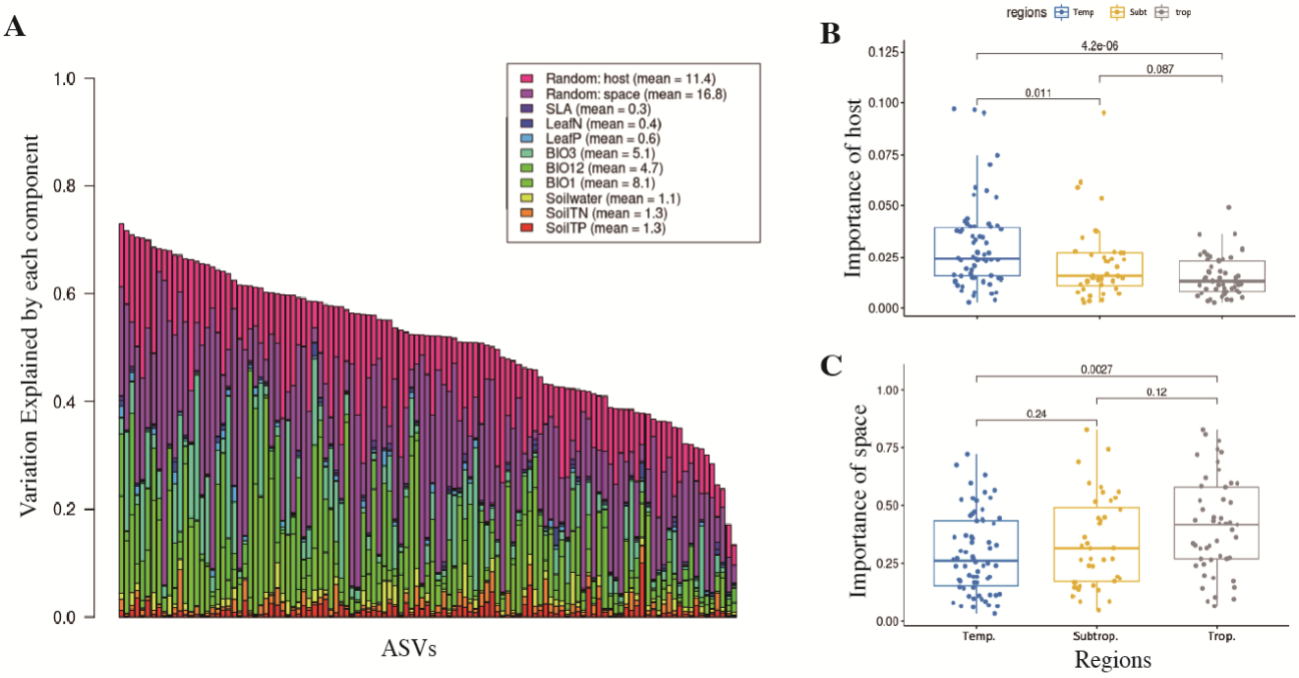

Global metacommunity networks
(10 forests, 490 plants, 12096 ASVs)
Tropical metacommunity networks
(3 forests, 214 plants, 6838 ASVs)
Subtropical metacommunity networks
(4 forests, 181 plants, 5567 ASVs)
Tropical metacommunity networks
(3 forests, 95 plants, 3338 ASVs)
Bacteria ASVs
JFL
DHS
HSD
BSZ
GTS
BDGS
TT
DLS
CBS
LS
Latitude (+)

Fig. 5


A
Variation Explained by each component
1.0
0.8
0.6
0.4
0.2
0.0
ASVs
Random: host (mean = 11.4)
Random: space (mean = 16.8)
SLA (mean = 0.3)
LeafN (mean = 0.4)
LeafP (mean = 0.6)
BIO3 (mean = 5.1)
BIO12 (mean = 4.7)
BIO1 (mean = 8.1)
Soilwater (mean = 1.1)
SoilTN (mean = 1.3)
SoilTP (mean = 1.3)
B
regions
Temp
Subt
trop
Importance of host
4.2e-06
0.087
0.011
0.125
0.100
0.075
0.050
0.025
0.000
C
Importance of space
0.0027
0.12
0.24
1.00
0.75
0.50
0.25
0.00
Temp.
Subtrop.
Trop.
Regions